INFORMATICA MEDICA SLOVENICA

Strokovni prispevki

Poročila

SDMI

Revija Slovenskega društva za medicinsko informatiko
Informatica Medica Slovenica
LETNIK 8, ŠTEVILKA 1
ISSN 1318-2129
ISSN 1318-2145 on line edition
http://lsd.uni-mb.si/ims

O REVIJI

Informatica Medica Slovenica je interdisciplinarna strokovna revija, ki objavlja prispevke s področja medicinske informatike, informatike v zdravstvu in zdravstveni negi, ter bioinformatike. Revija objavlja strokovne prispevke, znanstvene razprave, poročila o aplikacijah ter uvajanju informatike na področjih medicine in zdravstva, pregledne članke in poročila. Še posebej so dobrodošli prispevki, ki obravnavajo nove in aktualne teme iz naštetih področij.

Informatica Medica Slovenica je strokovna revija Slovenskega društva za medicinsko informatiko. Izhaja trikrat letno (en letnik, tri številke). Revija je dostopna internetu na naslovu http://lsd.uni-mb.si/ims. Avtorji člankov naj svoje prispevke v elektronski obliki pošiljajo odgovornemu uredniku po elektronski pošti na naslov kokol@uni-mb.si.

Revijo prejemajo vsi člani Slovenskega društva za medicinsko informatiko. Informacije o članstvu v društvu oziroma o naročanju na revijo so dostopne na tajništvu društva (doc. dr. Drago Rudel, drago.rudel@mf.uni-lj.si).

VSEBINA

Uvodnik ■

Dragi bralci,

Pred vami je nova številka Informatice Medice Slovenice. Prvi del (avtorji Povalej, Holobar, Mauratidis in Končar) predstavljajo najboljši študentski članki s konference Computer-Based Medical Systems, ki je bila ob sodelovanju Društva za medicinsko informatiko in pod pokroviteljstvom mednarodnega inženirskega združenja IEEE ter Univerze v Mariboru, Fakultete za elektrotehniko, računalništvo in informatiko izvedena v začetku junija 2002 v Mariboru. Drugi del vsebuje zanimive prispevke slovenskih in tujih avtorjev in poročila s konferenc.

prof. dr. Peter Kokol,
glavni urednik

■ **Infor Med Slov** 2003; 8(1): 1

Research Paper ■

Surface EMG Decomposition Using a Novel Approach for Blind Source Separation

Aleš Holobar, Damjan Zazula

Abstract. We introduce a new method to perform a blind deconvolution of the surface electromyogram (EMG) signals generated by isometric muscle contractions. The method extracts the information from the raw EMG signals detected only on the skin surface, enabling long-time noninvasive monitoring of the electromuscular properties. Its preliminary results show that surface EMG signals can be used to determine the number of active motor units, the motor unit firing rate and the shape of the average action potential in each motor unit.

■ **Infor Med Slov** 2003; 8(1): 2-14

Author's institution: Faculty of Electrical Engineering and Computer Science, University of Maribor.

Contact person: Aleš Holobar, Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova 17, 2000 Maribor. email: ales.holobar@uni-mb.si.

Introduction

The activity of muscles has been the subject of many studies of bioengineers, physiologists, neurophysiologists, and clinicians for more than 100 years. Many different methods of gathering and interpreting the physiological data and information have been developed. In the past few decades, the assessment of the electrical activity of muscles has proved to be very important. Using the computer based digital processing, many valuable knowledge has been extracted from the electromyographic signals enabling precise medical diagnosing and prevention of possible neurological and muscle disorders[1,2].

The muscle movement in all human beings is controlled by the central nervous system. It generates the electrical pulses that travel through the motor nerves to different muscles. The neuro-muscular junction is called innervation zone and is usually situated about the middle of the muscle body. Each muscle is composed of a large number of tiny muscle fibres, which are organized into so called motor units (MU). Each MU gathers all the fibres that are innervated by the same nerve, i.e. axon. When electrically excited, fibres produce a measurable electrical potential, called action potential (AP), which propagates along the fibres to both directions towards muscle tendons and causes the contraction of the fibres.

The electromyograms (EMGs) are taken with different kinds of electrodes whose uptake areas vary according to the electrode size. The majority of the EMG recordings is based on the invasive methods with the needle electrodes[3], whose invasive character prevents the long-time monitoring of the electromuscular parameters. The non-invasive method of the surface EMG (SEMG) has been developed recently and has numerous advantages. There is significantly less discomfort, no tissue damage and therefore no subsequent tissue scarring. This allows for unlimited repetition of tests in exactly the same place. Furthermore, recording of SEMG is inexpensive and gives global information about muscle activity[1]. Finally, linear surface electrode arrays can be used providing additional information about innervation zone location, fibre length and conduction velocity[4].

The main disadvantage of SEMG is poor morphological information about the MU action potentials (MUAP), caused by different filtering effects of several tissue layers (skin, fat, muscle, etc.)[5]. In the case of needle electrodes we can selectively observe the action potentials of only a few active motor units, or even of a single muscle fiber[3]. In the SEMG case, on the other hand, we deal with several tens of active motor units and the measurable contributions from other muscles not being under the clinical investigation, what is often referred to as muscle cross-talk. Many attempts to enhance the MUAP information and to suppress the cross-talk in SEMG were maid in the past as different surface recording technique, such as double differential, Laplacian, etc., were investigated[6]. Nevertheless, the manual decomposition of SEMG to separate MUAPs is, even with lower muscle contractions, virtually impossible and computer assisted decomposition is required.

The non-invasive assessment of muscle properties through the information extracted from the surface EMG signals has introduced new issues also in the field of the EMG signal processing. Despite the numerous efforts[7-11], no final technique for SEMG decomposition has been proposed yet. The main problem, as mentioned above, is a very high complexity of the SEMG signals. They are composed of a high number of individual, filtered MUAPs being superimposed into the surface signal. In addition, no a priori information on the number of active motor units and the nature of their mixture is available, either; hence the SEMG signals should be decomposed blindly.

The blind separation of the sources has been widely studied in the past and many solutions exist for both instantaneous and convolutive mixtures[12]. The problem of instantaneous mixtures is most often addressed by exploiting the Actually, this

possible non-Gaussianity of the sources. is the only possible route when the source signals are independently and identically distributed (i.i.d)[13,14]. When the first 'i' of 'i.i.d.' is not valid, i.e. when the source signals are correlated in time, another route is to exploit these correlations. Identifiability in such an approach is granted even when the signals are normally distributed, provided the source signals have different spectra[15,16]. The contributions from other authors[17-19] have explored the case where the second 'i' of 'i.i.d.' is failing, that is the non-stationary case. The latter can be successfully applied to the problem of muscular cross-talk[20].

The problem of convolutive mixtures is much more complex and although many different attempts (wavelets and neural networks classifications, time-frequency decomposition, etc.) [12] were made there is no general solutions for their complete deconvolution. However, the theoretical model of SEMG signals is, under the assumption of stable measurements and isometric muscle contractions, usually based on convolutive mixtures of the nerve train pulses and MUAPs detected on different electrodes. Considering their time-varying nature, the SEMG signals can also be modelled as non-stationary signals.

A. Belouchrani and M. Amin[18, 19] have addressed general convolutive mixtures of non-stationary signals and exploited the differences of energy locations of sources in time-frequency (t-f) domain. They have proposed to deal with the convolution in the form of mixing matrix by adding delayed repetitions of sources. New sources then form the block diagonal source *spatial t-f distribution (STFD) matrices*. Hence, with joint block diagonalization[24] of observation spatial time-frequency distribution matrices several versions of each source can be retrieved, but only up to a filtering effect[19].

In this paper, we present preliminary results of a new method for full deconvolution of surface EMG signals. Our approach is based on the work of A. Belouchrani and M. Amin, however, it additionally suggests the construction of diagonal source STFD matrices. These latter matrices are processed into a joint-diagonalization scheme (instead of joint block diagonalization), which provides an estimation of the transfer functions (MUAPs detected on different electrodes) and sources up to the scale factor and the phase shift.

Section 2 briefly reviews the algorithms and the problems of joint block diagonalization of spatial t-f distribution matrices for the instantaneous (convolutive) mixtures. In Section 3 we propose a new method for the construction of diagonal spatial t-f distribution matrices in the convolutive case. Section 4 reveals results of the proposed method on the synthetic SEMG signals. We end our paper with the conclusions and discussion in Section 5.

Separation of convolutive mixtures in t-f domain

Consider a general discrete convolutive multiple-input multiple-output (MIMO), linear, time invariant model given by

$$x_i = \sum_{j=1}^{N} \sum_{l=0}^{L} h_{ij}(l) s_j(t-l) + n_i(t). \qquad (1)$$

x_i $(i = 1,\ldots,M)$ is one of M observations and s_j $(j = 1,\ldots,N)$ is one of N sources ($M>N$) that are mutually independent for every time/lag and have different structures and localization properties in time-frequency domain. h_{ij} is the transfer function between the j-th source and the i-th sensor with the overall extent of $(L+1)$ taps. $n_i(t)$ $(i=1,\ldots,M)$ is additive i.i.d noise, independent from the sources defined by.

$$E[\mathbf{n}(t+\tau)\mathbf{n}(t)] = \delta(\tau)\sigma^2 \mathbf{I}_m, \qquad (2)$$

where E is mathematical expectation, $\mathbf{I}_M$ the $M \times M$ identity matrix, $\delta(\tau)$ the Dirac impulse and σ^2 the unknown variance of the noise.

Eq. (1) can be written as[19]:

$$\bar{\mathbf{x}}(t) = \mathbf{A}\bar{\mathbf{s}}(t), \tag{3}$$

where

$$\begin{aligned}&\bar{\mathbf{s}}(t) = [\bar{s}_1(t), \bar{s}_2(t), ..., \bar{s}_{L+K-1}(t), \bar{s}_{L+K}(t),\\ &\bar{s}_{L+K+1}(t), ..., \bar{s}_{N(L+K)}(t)]^T =\\ &= [s_1(t), s_1(t-1), ..., s_1(t-(L+K)+2,\\ &s_1(t-(L+K)+1, s_2(t), ...,\\ &s_N(t-(L+K)+1)]\end{aligned} \tag{4}$$

$$\begin{aligned}&\bar{\mathbf{x}}(t) = [\bar{x}_1(t), \bar{x}_2(t), ..., \bar{x}_K(t), \bar{x}_{K+1}(t),\\ &, ..., \bar{x}_{MK}(t)]^T =\\ &= [x_1(t), x_1(t-1), ..., x_1(t-K+1),\\ &x_2(t), ..., x_M(t-K+1)]\end{aligned} \tag{5}$$

are extended vectors of sources and observations, respectively, and

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{M1} & \cdots & \mathbf{A}_{MN} \end{bmatrix} \tag{6}$$

with

$$\mathbf{A}_{ij} = \begin{bmatrix} h_{ij}(0) & \cdots & h_{ij}(L) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & h_{ij}(0) & \cdots & h_{ij}(L) \end{bmatrix}. \tag{7}$$

$\boldsymbol{A}$ is $MK \times N(L+K)$ mixing matrix with full column rank, but is otherwise unknown. $\boldsymbol{A}_{ij}$ are $K \times (L+K)$ matrices where K is chosen such that $MK > N(L+K)$.

The covariance of vectors $\bar{\mathbf{s}}(t)$ and $\bar{\mathbf{x}}(t)$ is then[19]

$$\begin{aligned}&\mathbf{R}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}(t,\tau) = E[\bar{\mathbf{s}}(t+\tau)\bar{\mathbf{s}}(t)] =\\ &= diag[\mathbf{R}_{\tilde{\mathbf{s}}_1\tilde{\mathbf{s}}_1}(t,\tau), \ldots, \mathbf{R}_{\tilde{\mathbf{s}}_N\tilde{\mathbf{s}}_N}(t,\tau)]'\end{aligned} \tag{8}$$

$$\mathbf{R}_{\bar{\mathbf{x}}\bar{\mathbf{x}}}(t,\tau) = \mathbf{A}\mathbf{R}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}(t,\tau)\mathbf{A}^H + \delta(\tau)\sigma^2\mathbf{I}_{N(L+K)}, \tag{9}$$

where

$$\mathbf{R}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}(t,\tau) = \begin{bmatrix} \mathbf{R}_{\tilde{s}_1\tilde{s}_1}(t,\tau) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{\tilde{s}_N\tilde{s}_N}(t,\tau) \end{bmatrix} \tag{10}$$

is block diagonal, $\mathbf{0}$ is matrix with elements all equal to zero, and $\mathbf{R}_{\tilde{\mathbf{s}}_i\tilde{\mathbf{s}}_i}(t,\tau)$ denotes local $(L+K)\times(L+K)$ correlation matrix of vector $\tilde{\mathbf{s}}_i(t) = [s_i(t), ..., s_i(t-(L+K)+1]$

$$\mathbf{R}_{\tilde{\mathbf{s}}_i\tilde{\mathbf{s}}_i}(t,\tau) = E[\tilde{\mathbf{s}}_i(t+\tau)\tilde{\mathbf{s}}_i(t)]. \tag{11}$$

$\mathbf{R}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}(0,0)$ is generally block-diagonal since the correlations between $s_i(t+\tau_1)$ and $s_i(t+\tau_2)$ are not necessarily zero[19].

In the time-frequency plane, equation (9) becomes[19]:

$$\mathbf{D}^{\phi}_{\bar{\mathbf{x}}\bar{\mathbf{x}}}(t,f) = \mathbf{A}\mathbf{D}^{\phi}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}(t,f)\mathbf{A}^H + \delta(\tau)\sigma^2\mathbf{I}_{N(L+K)}, \tag{12}$$

where $\mathbf{D}^{\phi}_{\bar{\mathbf{x}}\bar{\mathbf{x}}}(t,f)$ is $MK \times N(L+K)$ STFD matrix of $\bar{\mathbf{x}}(t)$ whose *(k,l)* entry is the cross-(auto) t-f distribution of Cohen's class[22] for signals $\bar{x}_k(t)$ and $\bar{x}_l(t)$:

$$\begin{aligned}&D^{\phi}_{x_k x_l}(t,f) = \sum_{q=-\infty}^{\infty}\sum_{p=-\infty}^{\infty}\phi(p,k)x_k(t+p+q)\cdot\\ &x_l(t+p-q)e^{-j4\pi fq}.\end{aligned} \tag{13}$$

$\phi(p,q)$ denotes the kernel that characterizes the TF distribution.

In a low-noise environment the noise term in (12) can be neglected such that

$$\mathbf{D}^{\phi}_{\overline{\mathbf{xx}}}(t,f) \approx \mathbf{A}\mathbf{D}^{\phi}_{\overline{\mathbf{ss}}}(t,f)\mathbf{A}^{H}. \quad (14)$$

Let $\mathbf{W}$ denote a $N(L+K)\times MK$ whitening matrix, such that

$$\begin{aligned}&\mathbf{W}\mathbf{R}_{\overline{\mathbf{xx}}}(0,0)\mathbf{W}^{H} = \\ &= (\mathbf{W}\mathbf{A}\mathbf{R}^{\frac{1}{2}}_{\overline{\mathbf{ss}}}(0,0))(\mathbf{W}\mathbf{A}\mathbf{R}^{\frac{1}{2}}_{\overline{\mathbf{ss}}}(0,0))^{H} = \mathbf{I},\end{aligned} \quad (15)$$

where $\mathbf{R}^{\frac{1}{2}}_{\overline{\mathbf{ss}}}(t,\tau)$ denotes the block diagonal Hermitian square root matrix of source correlation matrix $\mathbf{R}_{\overline{\mathbf{ss}}}(t,\tau)$. Note that whitening matrix $\mathbf{W}$ can be obtained as an inverse square root of the observation autocorrelation matrix $\mathbf{R}_{\overline{\mathbf{xx}}}(0,0)$ [18]. More robust procedure for its calculation is described by Belouchrani et al.[23].

Pre- and post-multiplying the STDF matrices $\mathbf{D}^{\phi}_{\overline{\mathbf{xx}}}(t,f)$ by $\mathbf{W}$ leads to the *whitened STFD-matrices*, defined as:

$$\begin{aligned}&\mathbf{D}^{\phi}_{\mathbf{zz}}(t,f) = \mathbf{W}\mathbf{D}^{\phi}_{\overline{\mathbf{xx}}}(t,f)\mathbf{W}^{H} = \\ &= \mathbf{U}\mathbf{R}^{-\frac{1}{2}}_{\overline{\mathbf{ss}}}(0,0)\mathbf{D}^{\phi}_{\overline{\mathbf{ss}}}(t,f)\mathbf{R}^{-\frac{1}{2}}_{\overline{\mathbf{ss}}}(0,0)\mathbf{U}^{H},\end{aligned} \quad (16)$$

where $\mathbf{U} = \mathbf{W}\mathbf{A}\mathbf{R}^{\frac{1}{2}}_{\overline{\mathbf{ss}}}(0,0)$ is a $N(L+K)\times N(L+K)$ unitary matrix and $\mathbf{R}^{-\frac{1}{2}}_{\overline{\mathbf{ss}}}(0,0)\mathbf{D}^{\phi}_{\overline{\mathbf{ss}}}(t,f)\mathbf{R}^{-\frac{1}{2}}_{\overline{\mathbf{ss}}}(0,0)$ is block diagonal (each block is of size $(L+K)\times(L+K)$)[19].

According to (16), any whitened STFD matrix $\mathbf{D}^{\phi}_{\mathbf{zz}}(t,f)$ is block diagonal in the basis of the columns of the matrix $\mathbf{U}$ and the missing unitary matrix $\mathbf{U}$ can be retrieved by block diagonalization[24] of arbitrary $\mathbf{D}^{\phi}_{\mathbf{zz}}(t,f)$ matrix. To increase the robustness of determining of $\mathbf{U}$, we rather consider the joint block diagonalization of a combined set of several $\mathbf{D}^{\phi}_{\mathbf{zz}}(t,f)$ matrices[19].

Knowing the unitary mixing matrix $\mathbf{U}$, the sources can be retrieved by:

$$\hat{\mathbf{s}}(t) = \mathbf{U}^{H}\mathbf{W}\overline{\mathbf{x}}(t) \quad (17)$$

and the mixing matrix $\boldsymbol{A}$ as

$$A = \mathbf{W}^{\#}\mathbf{U}, \quad (18)$$

where superscript $^{\#}$ denotes the Moore-Penrose pseudoinverse.

According to (17) and (3), the recovered signals yield:

$$\hat{\mathbf{s}}(t) = \mathbf{D}\mathbf{R}^{\frac{1}{2}}_{\overline{\mathbf{ss}}}\overline{\mathbf{x}}(t), \quad (19)$$

where $\mathbf{D}$ is unknown, block diagonal unitary matrix coming from the inherent indeterminacy of the joint block diagonalization[24]. Hence, sources can be reconstructed only up to a filtering effect[19].

Assume now the matrices $\mathbf{R}_{\tilde{\mathbf{s}}_i\tilde{\mathbf{s}}_i}(0,0)$ and $\mathbf{D}^{\phi}_{\overline{\mathbf{ss}}}(t,f)$ diagonal in a particular *(t,f)* point. Then according to (16), the corresponding whitened STFD matrix $\mathbf{D}^{\phi}_{\mathbf{zz}}(t,f)$ is diagonal in the basis of the columns of the matrix $\boldsymbol{U}$ (the eigenvalues of $\mathbf{D}^{\phi}_{\mathbf{zz}}(t,f)$ being the diagonal entries of $\mathbf{D}^{\phi}_{\overline{\mathbf{ss}}}(t,f)$). If, for a *(t,f)* point in time-frequency plane, the diagonal elements of $\mathbf{D}^{\phi}_{\overline{\mathbf{ss}}}(t,f)$ are all distinct, the missing unitary matrix $\boldsymbol{U}$ can be uniquely retrieved (up to the order and the amplitude of the sources) by computing the eigendecomposition of $\mathbf{D}^{\phi}_{\mathbf{zz}}(t,f)$ [18]. In the case of degenerated eigenvalues, i.e., when the diagonal elements of $\mathbf{D}^{\phi}_{\mathbf{ss}}(t,f)$ are not all distinct, $\boldsymbol{U}$ can be retrieved by joint diagonalization[18, 19,21] of a combined set of $\mathbf{D}^{\phi}_{\mathbf{zz}}(t,f)$ matrices corresponding to *(t,f)* points for which $\mathbf{D}^{\phi}_{\mathbf{ss}}(t,f)$ is diagonal.

Joint diagonalization can recover the sources up to permutations, sign change and a constant factor (scale factor and phase shift in the complex case) [18,21,25-27], since these modifications can be balanced by the mixing matrix $\mathbf{A}$ to provide exactly the same observations. This property is often called

the *indeterminacy* of the blind source separation (BSS) approach[25-27].

Separation of surface EMG signals

Noninvasive nature of the SEMG recording enables several stable measurements of the motor-unit electrical activity during isometric muscle contraction, forming MIMO theoretical model of SEMG signals. In such a model the SEMG signals are most often considered as convolutive mixtures of pulse trains (triggering pulses from innervating motor neurons), with different MUAPs as system impulse responses.

Pulse trains in SEMG model can be interpreted as[30]:

$$s_i = \sum_{n=-\infty}^{\infty} \delta(t - nT_i + \Theta_{in}) \text{ for } i = 1,...,N, \quad (20)$$

where $\delta(\cdot)$ is the Dirac impulse, T_i are deterministic, and Θ_{in} random variables with Gaussian distribution. Again, the sources are assumed to have different localization properties in time-frequency domain, i.e. their pulses should not overlap in time. We will also assume that the average interpulse interval is longer then the length of the impulse response h_{ij}, i.e. L in (1).

Under these assumptions, the sources correlation matrix $\mathbf{R}_{\mathbf{ss}}(0,0)$ is diagonal. Moreover, the sources have well localized energy in time, which will become advantageous in the process of deconvolution, where we will try to make the STFD matrices $\mathbf{D}_{\mathbf{ss}}^{\phi}(t,f)$ diagonal, too.

Diagonal source STFD matrices

To preserve a high time resolution, the Wigner-Ville spectra defined by[22]

$$WV_{s_i s_j}(t,f) = \sum_{\tau=-\infty}^{\infty} s_i(t+\tfrac{1}{2}\tau) s_j(t-\tfrac{1}{2}\tau) e^{-j2\pi f\tau} \quad (21)$$

should be used as the time-frequency distribution in (13). Using any other distribution (kernel ϕ) would spread the energy in time and made the source STFD matrices $\mathbf{D}_{\mathbf{ss}}^{\phi}(t,f)$ block-diagonal.

The main disadvantage of using the Wigner-Ville spectra is its high sensitivity to the crossterms (non-zero cross Wigner-Ville spectra $WV_{s_i s_j}(t,f)$), which, again, make the source STFD matrices $\mathbf{D}_{\mathbf{ss}}^{\phi}(t,f)$ block-diagonal. The cross Wigner-Ville spectra in an optional (t_k, f_k) point is a summation of all the pulses from sources s_i and s_j that fulfil the following relation:

$$t_k = \frac{1}{2}(t_{in} + t_{jm}), \quad (22)$$

where t_{in} is the time of appearance of the n-th pulse in source s_i, t_{jm} is the time of appearance of the m-th pulse in source s_j, and $n,m = -\infty,...,\infty$. As a consequence, the source STFD matrix $\mathbf{D}_{\mathbf{ss}}^{\phi}(t_k, f_k)$ is not diagonal in such (t_k, f_k) point.

We can reduce the number of pulse contributions in $WV_{s_i s_j}(t,f)$ by calculating pseudo Wigner-Ville spectra[22], that is, by limiting τ in (21) to the finite interval $[-a, a]$:

$$PWV_{s_i s_j}(t,f) = \sum_{\tau=-a}^{a} s_i(t+\tfrac{1}{2}\tau) s_j(t-\tfrac{1}{2}\tau) e^{-j2\pi f\tau} \quad (23)$$

.

The number of pulses that contribute to crossterms in $PWV_{s_i s_j}(t,f)$ depends on the size of the interval $[-a, a]$ and is finite. Making the limit a small enough, all the crossterms in $PWV_{s_i s_j}(t,f)$ are left out, and the source STFD matrices $\mathbf{D}_{\mathbf{ss}}^{\phi}(t_k, f_k)$ begin to show their diagonal structure.

The time-frequency plane has now shrunk, because we decreased the resolution of the frequency content in the Wigner-Ville spectra. This made the blind source separation less noise resistant, because the energy of the noise is now less spread along the frequency axis (the time-frequency plane has the property of spreading the noise energy along the frequency axis while localizing the energy of the signal). As a consequence, we have to process longer signals to compensate the noise. However, the source STFD matrices $\mathbf{D}_{\mathbf{ss}}^{\phi}(t_k, f_k)$ are diagonal and we can now use the joint diagonalization [18,21,25-28] instead of joint block diagonalization[24] and, hence, retrieve the unfiltered version of the sources up to a scale factor.

Selection of diagonal STFD matrices in the case with overlapped pulses

Assume certain pulses of the sources overlap. Denote by $\{l_1,\ldots, l_p\}$ the positions of p sources in vector of sources $\bar{\mathbf{s}}(t) = [\bar{s}_1(t), \bar{s}_2(t), \ldots, \bar{s}_{N(L+K)}(t)]^T$ whose pulses overlap at certain time t_k. The source STFD matrix $\mathbf{D}_{\mathbf{ss}}^{\phi}(t_k, f_k)$ will then have p^2 non-zero elements at the positions (k_1,k_2) where $k_1, k_2 \in \{l_1,\ldots,l_p\}$ as follows:

$$\mathbf{D}_{\mathbf{ss}}^{\phi}(t_k, f_k) = M_d^p, \tag{24}$$

$$M_d^p(k_1, k_2) = \begin{Bmatrix} d & \text{if } k_1, k_2 \in \{l_1,\ldots,l_p\} \\ 0 & \text{otherwise} \end{Bmatrix}, \tag{25}$$

where d denotes the energy of one pulse, and the assumption that all pulses have the same energy, like in (20), has been considered. Only p of p^2 non-zero elements in (25) lie on the diagonal and M_d^p is far from being diagonal. Selecting the observation STFD matrices at time t_k strongly affects the performance of joint diagonalization[18,21,25-28].

Because $\mathbf{U}$ is a unitary matrix, the following relation is valid:

$$eig(\mathbf{D}_{\mathbf{zz}}^{\phi}(t,f)) = eig(\mathbf{U}\mathbf{D}_{\mathbf{ss}}^{\phi}(t,f)\mathbf{U}^H) = \\ = eig(\mathbf{D}_{\mathbf{ss}}^{\phi}(t,f)), \tag{26}$$

where $eig(\mathbf{M})$ denotes the eigenvalues of the matrix $\mathbf{M}$. Noting that the only non-zero eigenvalue of matrix M_d^p is $eig(M_d^p) = pd$, we can exclude the STFD source matrices that correspond to overlapped source pulses from the process of the joint diagonalization by simply comparing the maximum eigenvalues of the whitened observation STFD matrices $\mathbf{D}_{\mathbf{zz}}^{\phi}(t,f)$.

Criteria and algorithms for the selection of (t,f) points in which $\mathbf{D}_{\mathbf{ss}}^{\phi}(t,f)$ is diagonal are more in detail described by Holobar et al.[28] and by Nguyen et al.[29].

Results on synthetic SEMG signals

In this section, the preliminary results, as investigated via computer simulations, are reported. SEMG signals were generated by SEMG simulator[31], that allows to simulate the main features of the surface EMG signal, including the generation and extinction phenomena of the action potentials at the end-plate and tendon regions and the size and shape of the recording electrodes without approximation of the current density source. The simulator models the volume conductor as an anisotropic layered medium with muscle (anisotropic), fat (isotropic) and skin (isotropic) layers. The model allows simulation of multi-channel spatially filtered surface EMG signals and is based on efficient numerical algorithms, which implement the simulation of signals generated during voluntary contractions by the activity of a large number of MUs. The detection systems can be placed either along the fibres direction (usual linear electrode array configuration) or transversally with respect to the

muscle fibres. Motor units are placed randomly in the detection volume and are active at the selectable firing rates.

In our experiment the surface EMG signals from the biceps brachii muscle during isometric voluntary contraction at low force level were simulated. The main parameters applied in our simulation were the following:

1. MA(4,4) model was chosen, so 4 active MUs were assumed and 6 surface electrodes using the double differential recording technique were simulated what resulted in 4 measurements of SEMG.

2. The length of muscle fibres in all MUs was set of 70 mm. The first MU with 9 fibres was assumed 6 mm, the second with 12 fibres 4 mm, the third with 15 fibres 5 mm, and the fourth with 10 fibres 3 mm deep in the muscle layer.

3. Fibres of the MUs were inclined by 2, 8, 4 and 7 degrees and shifted 7, -5, -2 and 6 mm in the transversal (x in Fig. 1) direction from the centre of the electrode array, respectively.

4. The spread of the innervation zone was taken 6 mm for the first, 6 mm for the second, 5 mm for the third and 5 mm for the fourth MU.

5. The conduction velocity was assumed to be normally distributed with the mean of 4 m/s and standard deviation of 1 m/s (4.54 m/s for the first, 3.83 m/s for the second, 4.33 m/s for the third, and 3.56 m/s for the fourth MU.

6. The thickness of the skin layer was set to 2 mm and thickness of the fat layer to 7 mm.

7. Mean firing rate of the first, second and the third MU were set to 13 Hz, 14 Hz, 13Hz and 15Hz with the variance of 3 Hz, respectively.

8. Rectangular electrodes of 5 by 1 mm were simulated with the 5 mm interelectrode distance.

9. The electrode array was assumed placed between the innervation zone and the tendons of fibres.

10. Transversal orientation of detecting system with respect to the muscle fibres was simulated in order to emphasize the differences in contributions (impulse responses) of different motor units to observations, that is to say, to improve the rank of the mixing matrix **A**.

11. The sampling frequency of 1024 *Hz* was used for the generated EMG signals.

12. The length of synthetic surface EMG signals was set to 5 s (5120 samples).

13. Signal-to-noise ratio (SNR) was set to 15 dB.

The position and orientation of the detection system and the MUs is schematically depicted in Fig. 1, respectively. The generated SEMG signals are partially depicted in Fig. 2.

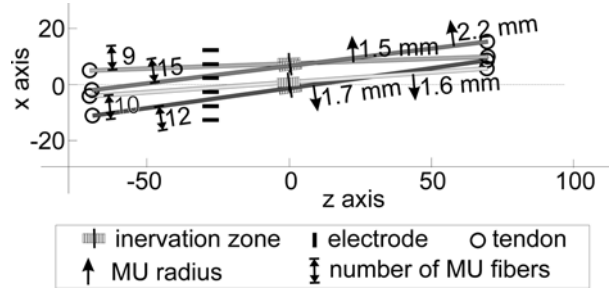


Figure 1 Position and orientation of the detecting system with respect to the simulated active motor units. The MUs are depicted by grey lines ending with circles (tendons), innervation zones by striped rectangular, electrodes by black rectangular. The depth, radius, inclination and the number of fibres in each MU is also depicted.

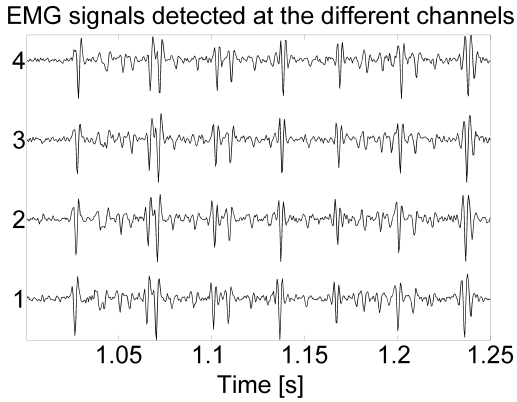
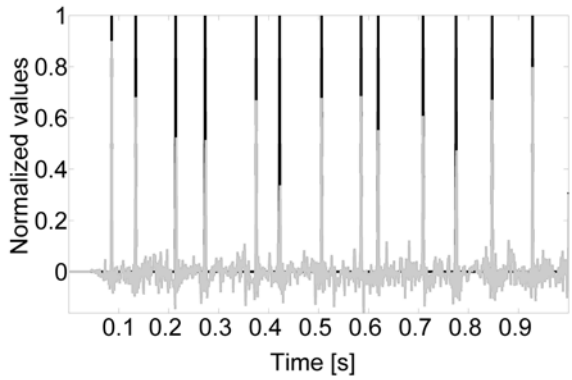
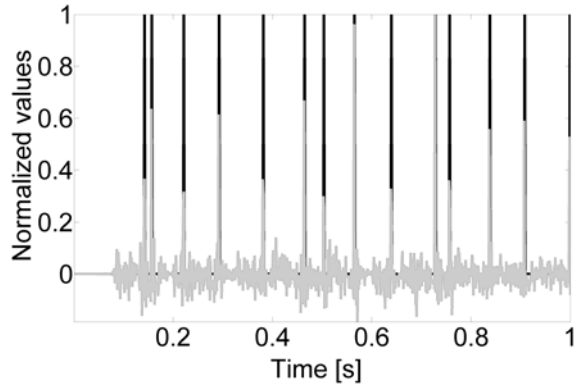


Figure 2 Parts of synthetic SEMG signals at four different channels. The detection system was placed transversally with respect to the muscle fibres. The interelectrode distance was set to 5 mm and double differential recording technique was selected.

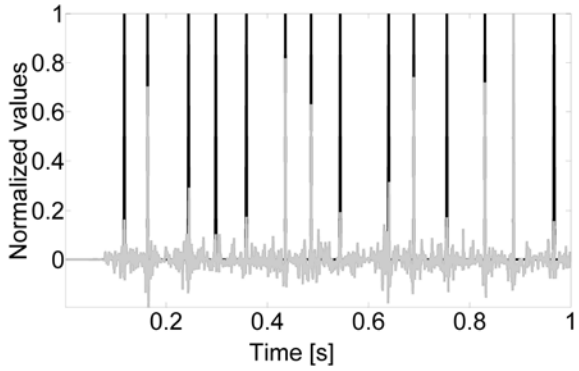
Setting the length of impulse response h_{ij}, to *26* samples, *26* estimations of each source were calculated. The estimations were then classified, aligned, normalized, and summed together. They showed almost perfect match with the original sources. Almost all the triggering pulses were successfully recovered. The mean normalized energy of recovered pulses was 0.57 with the variance of 0.21 and the minimum value of 0.14. The ground jitter stayed bellow 0.18, with the mean value of –0.03 and the variance of 0.11. Note the ground jitter is proportional to the nose and exceeds the recovered pulses at the SNR of 8 dB. The results for all 4 train pulses are partly depicted in Fig. 3.



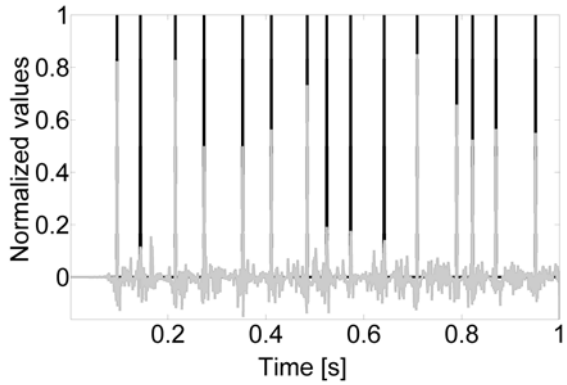
a)



b)



c)



d)

Figure 3 Comparison of the original innervation pulse trains (black) and corresponding retrieved sources (grey) over a 1s time interval for the first (a), second (b), third (c) and fourth (d) source, respectively. Notice the exact matching of the pulse trains.

The recovered MUAPs showed a good match with the original ones, too. The average absolute sample difference between the original MUAPs, detected by different electrodes, and decomposed MUAPs, expressed in percentage to the MAUP

amplitude span, was 7.3 %. The recovered MUAPs are partially depicted in Fig. 4.

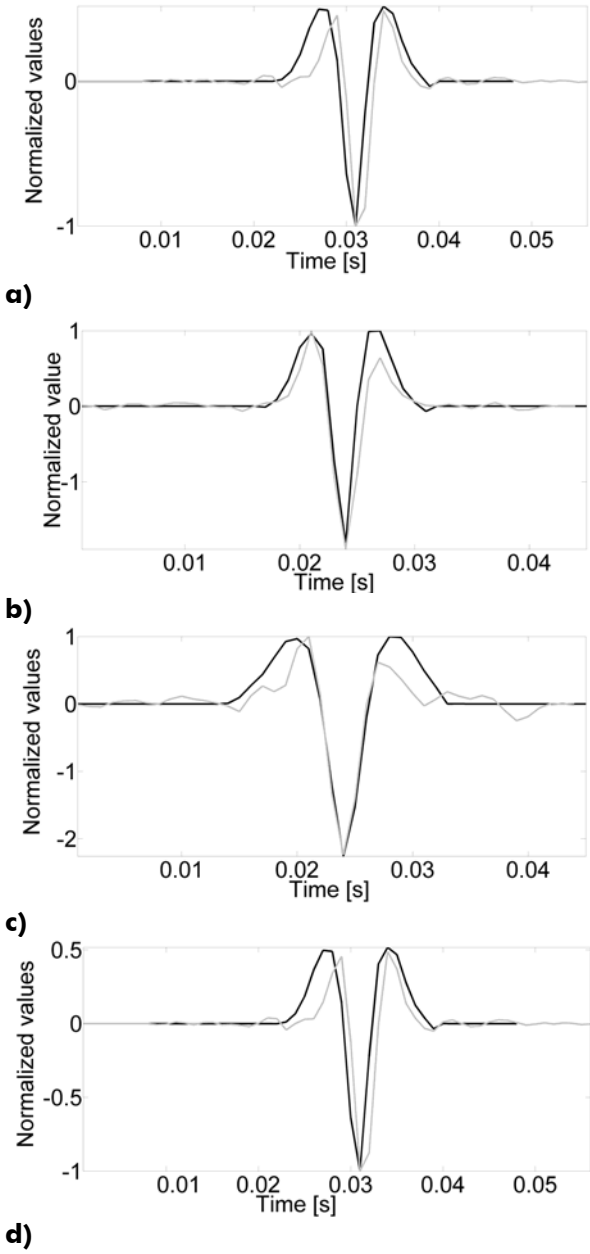


Figure 4 Comparison of the original MUAPs (black traces) and retrieved MUAPs (grey traces); the first MUAP in the third channel (a), the second MUAP in the fourth channel (b), the third MUAP in the second channel (c) and the fourth MUAP in the first channel (d). The impulse responses can be retrieved only up to a scaling factor; the amplitudes of the depicted impulse responses are normalized.

Conclusions

In this paper, a new approach for blind decovolution of surface EMG signals using pseudo Wigner-Ville time-frequency distributions is introduced. It takes the advantage of non-stationary characteristics of sources (the localization of energy in time) and is based on the joint diagonalization of a combined set of spatial time-frequency (STFD) matrices. A diagonal structure of the source STFD matrices is essential for the proposed approach and is enforced by incorporating only the (t,f) points corresponding to the autoterms (diagonal elements in STFD matrices) of one particular source. The off-diagonal elements in source STFD matrices are crossterms that become zero when calculating the Wigner-Ville spectra on the finite (short enough) interval.

The proposed method shows a number of attractive features. The expansion of convolutions to instantaneous mixture makes the STFD matrices very large. As a consequence, the separation can be time and space consuming. Our approach simplifies and fastens the calculation of auto (cross) time-frequency distributions. Moreover, since the time-frequency distributions are shrunk along the frequency axis, the approach is also space efficient. As a result, longer signals can be processed, which compensates the potential noise cancellation due to the effect of spreading the noise power while localizing the source energy in time-frequency domain as a whole. Finally, the K estimations of each source (train of pulses) and its transfer function (MUAPs detected by different electrodes) are retrieved up to a scaling factor by our approach. Hence, the calculated sources and impulse responses can further be improved by averaging the corresponding estimations, which makes the approach more noise resistant.

What are the limitations of our approach? Due to the uniqueness property of joint diagonalization[18, 21,25-28] and the structure of the searched source STFD matrices (only one non-zero diagonal

element), we have to find at least one source STFD matrix per source, i.e., each source (including the added delayed repetitions of sources) has to have at least one non-overlapped pulse. This situation is very probable when processing long enough signals and uncorrelated sources with low firing frequencies, i.e. surface EMG signals at low levels of muscle contraction. The performance drops at high contraction forces, due to the effects of motor unit synchronization[32]. A similar drop in performance is noticed at high firing rates (30Hz and more) when the interpulse intervals of sources become small in comparison to the length of impulse responses.

The impulse responses in our model are modelled with constant coefficients. As a consequence, the changes of the shapes of MUAPs in time, such as caused by fatigue, are not recognized by our method. Moreover, the number of active motor units is assumed to be constant. As a consequence, motor unit recruitment during the constant force and muscle contraction of long duration [33], and increasing force, respectively, is not recognized by our method, either. Longer SEMG signals must be broken into subsequent epochs and processed separately. The recommended length of each epoch depends on the muscle type and is generally inversely proportional to the level of muscle contraction. Processing the SEMG signals detected in biceps brachii at 30 % of its maximum voluntary contraction, for example, the recommended length of each epoch is approximately 10 s.

Due to possible permutations of source indices, caused by the indeterminacy of blind source separation, the reconstructed train pulses from each epoch may appear in the arbitrary order (two pulse trains, which are reconstructed from two different epochs and share the same index may correspond to different original pulse trains). In order to properly reconstruct the pulse sequences over all epochs, the subsequent epochs must share some common samples (we recommend each pair of subsequent epochs to share a half of common samples, i.e. 10 s long subsequent epochs should share 5 s of common SEMG signal). Aligning the common pulses in recovered innervation trains we easily identify the permutations of source indices and form the whole sequence of triggering pulses for each active MU. The MUAPs must retain constant only through the corresponding epoch, hence, the changes in the shape of MUAPs and the variation of the number of active MUs in time (subsequent epochs) can be monitored, respectively.

The analysis of individual motor units is quite important in clinical electromyography. The amplitude and duration of the motor unit action potential provide information on the type of muscle disorder incurred by the peripheral nervous system, the length of time since the disorder's onset, and the evidence for recovery. Furthermore, the reconstruction of the MUAP trains provides information on the firing rate of the individual MUs and on the change of this rate during an increase or decrease of the muscle contraction level. No SEMG decomposition technique has so far provided this information, which makes our approach unique.

Acknowledgement

This work was supported by the Slovenian Ministry of Education, Science, and Sport (Contract No. S2-796-010/21301/2000), and by the European funding in the 5th Framework project entitled NEW (Contract No. QLK6-2000-00139).

References

1. Merletti R: Surface electromyography: possibilities and limitations, J. of Rehab. Sci., Vol. 7, No. 3, 1994, pp. 25-34.
2. Knaflitz M, Balestra G: Computer analyisis of the myoelectric signal, IEEE Micro, No. 10, 1991, pp. 12-58.
3. Farina D, Colombo R, Merletti R, Olsen HB: Evaluation of intra-muscular EMG signal decomposition algorithms, Journal of Electromyography and Kinesiology, No. 11, 2001, pp. 175-187.

4. Farina D, Fortunato E, Merletti R: Noninvasive estimation of motor unit conduction velocity distribution using lnear electrode arrays, IEEE Trans. on biomed. eng., Vol 47, No. 3, 2000, pp. 380-388.
5. Farina D,Rainoldi A: Compensation of the effect of sub-cutaneous tissue layers on surface EMG: a simulation study, Med. eng. & Physics, No. 21, 1999, pp. 487-496.
6. Farina D, Merletti R: Effect of electrode shape on spectral features of surface detected motor unit action potentials, Acta Physiol. Pharmacol. Bulg., Vol. 26, pp. 63-66,2001.
7. Zazula D, Korže D, Šoštarič A, Korošec D: Study of Methods for Decomposition of Superimposed Signals with Application to Electromyograms, Pedotti et al. (Eds.), Neuroprosthetics, Berlin: Springer Verlag, 1996.
8. Gerber A, Studer RM, de Figueiredo RJP, Moschytz GS: A new framework and computer program for quantitative EMG signal analysis, IEEE Trans. on Biomedical Engineering, Vol. 31, No.12, Dec. 1984, pp.857-863.
9. Zazula D, Šoštarič A: Possible Approaches to Surface EMG Decomposition, SENIAM, Vol. 7, 1999, pp. 169-176.
10. Šoštarič A, Zazula D, Doncarli C: Time-Scale Decomposition of Compound (EMG) Signal, Electrotechnical Review, Vol. 67, 2000, pp. 69-75.
11. Zazula D, Plévin E: An Approach to Decomposition of Muscle and Nerve Signals, Int. Conf. on Signal, Speech, and Image Processing ICOSSIP '02, Skiathos, Greece, Sept. 2002, CD-ROM.
12. Mansour E, Barros AK, Ohnishi N: Blind separation of sources: methods, assumptions and applications, IEICE trans. Fundamentals, Vol. E83-A, No. 8, 2000, pp. 1498-1512.
13. Cardoso JF: Blind signal separation: statistical principles, Proc. of the IEEE. Special issue on blind identification and estimation, Vol. 9, No. 10, 1998,pp. 2009-2025.
14. Tong L, Liu R, Soon YHVC: Indeterminacy and identifiability of blind identification, IEEE Trans. Circuits Syst., Vol. 38, May 1991, pp. 499-509.
15. Tong L, Liu R: Blind estimation of correlated source signals, Proc. Asilomar Conf., 1990, pp. 161-164.
16. Belouchrani A, Meraim KA, Cardoso JF, Moulines E: A Blind Source Separation Technique Using Second-Order Statistics, IEEE trans. On Signal Porcessing, Vol. 45, No. 2, Feb. 1997, pp. 434-444.
17. Pham DT, Cardoso JF: Blind separation of instantaneous mixtures of non stationary sources, IEEE Transactions on Signal Processing, Vol. 49, No. 9, 2001, pp. 1837-1848.
18. Belouchrani A, Meraim KA: Blind Source Separation based on time-frequency signal representation, IEEE trans. On Signal Porcessing, Vol. 46, No. 11, Nov. 1998, pp. 2888-2898.
19. Boashash B, »Time-Frequency Signal Analysis and Processing«, Prentice Hall PTR, Englewood Cliffs, New Jersey, 2001.
20. Farina D, Merletti R, Indino B, Nazzaro M, Bottin A, Pozzo M, Caruso I: Surface EMG Cross-talk evaluated from experimental recordings and simulated signals, reflections on cross-talk interpretation, quantification and reduction, Proc. Of 4th International Workshop on Biosignal Interpretation BSI 2002, 2002, Como, Italy, pp. 163-166.
21. Cardoso F, Souloumiac A: Jacobi angles for simultaneous diagonalization, SIAM J. Mat. Anal. Appl., Vol. 17, No. 1, 1996, pp. 161-164.
22. Cohen L: Time-Frequency Analysis, Prentice Hall PTR, Englewood Cliffs, New Jersey, 1995
23. Belouchrani A, Cichocki A: A robust whitening procedure in blind source separation context, Electronics Letters, Vol. 36, 2000, pp. 2050-2051.
24. Belouchrani A, Meraim KA, Hua Y: Jacobi-like algorithms for joint block diagonalization: Application to Source Localization, Proc. ISPACS, 1998, Melbourne, Australia.
25. Cardoso JF: A least-squares approach to joint diagonalization, IEEE Signal Processing Lett., Vol. 4, Feb. 1997, pp. 52-53.
26. Cardoso JF, Souloumiac A: Jacobi angles for simultaneous diagonalization, SIAM J. Mat. Anal. Appl., Vol. 17, No. 1, 1996, pp. 161-164.
27. Cardoso JF: Perturbation of joint diagonalizers, technical report Ref\# 94D027, ftp://sig.enst.fr/pub/jfc/ Papers/joint_diag_pert_an.ps, 1994.
28. Holobar A, Fevotte C, Doncarli C, Zazula D: Single autoterm selection for blind source separation in time-frequency plane, Proc. of EUSIPCO, 2002, Toulouse, France, CD-ROM.
29. Nguyen LT, Belouchrani A, Meraim KA, Boashash B: Separating More Sources Than Sensors Using Time-Frequency Distributions, Proc. of 6th ISSPA, Vol. 2 , 2001, Kuala-Lumpur, Malaysia, pp. 583 -586.
30. Merletti R, Lo Conte L, Avignone E, Guglielminotti P: Modelling of surface myoelectric

signals; Part I: Model implementation, IEEE Trans. Biomed. Engng., Vol. 46, No. 7, pp. 810-820, 1999.
31. Farina D, Merletti R: A novel approach for precise simulation of the EMG signal detected by surface electrodes, IEEE Transactions on Biomedical Engineering, Vol. 48, 2001, pp. 637-646.
32. Weytjens JLF, van Steenberghe D: The effects of motor unit synchronization on the power spectrum of the electromyogram, *Biol. Cybern.*, No. 51, 1984 pp. 71-77.
33. Gazzoni M, Farina D, Merletti R: Motor unit recruitment during constant low force and long duration muscle contractions investigated with surface EMG, *IX International Symposium on Motor Control*, 2000, Varna, Bulgaria

Research Paper ■

Measurement and Analysis of Radial Artery Blood Velocity in Young Normotensive Subjects

Damjan Oseli, Iztok Lebar Bajec, Matjaž Klemenc, Nikolaj Zimic

Abstract. In our study we analyze the velocity contour of blood flow in the radial artery and compare it with the typical parameters of the diastolic function of the left ventricle in young normotensive subjects with and without familial predisposition to hypertension. The analysis of velocity contour shows significant shortening of the time delay for two typical points on the secondary velocity wave: the maximum and the end of the wave. This shift of the secondary wave towards the primary wave shows diminished compliance of small vessels.

According to the results our conclusion is that changes in small vessels begin earlier than impairment of the diastolic function.

■ **Infor Med Slov** 2003; 8(1): 15-19

Author's institution: Faculty of Computer and Information Science, University of Ljubljana (DO, ILB, NZ), Department of Cardiology, General Hospital of Gorica (MK).

Contact person: Damjan Oseli, Faculty of Computer and Information Science, University of Ljubljana, Tržaška 25, 1000 Ljubljana. E-mail: damjan.oseli@fri.uni-lj.si.

Introduction

Abnormalities in the peripheral circulation make a mayor contribution to the circulatory disturbances seen in essential hypertension.[1] Reduction of arterial distensibility leads to: an increase in arterial impedance and, thus, in cardiac afterload; a reduction in diastolic pressure and in perfusion of organs such as heart; and an acceleration of arteriosclerosis due to an increase in systolic and pulse pressure and to a greater traumatic effect on the vessel walls.[2-12] Beside the mentioned, changes in the diastolic function of the left ventricle are among the earliest signs of developing hypertensive heart disease.[13] In our study we compare typical parameters which are drawn from velocity wave analysis of blood flow in the radial artery with parameters that determine the proprieties of the left ventricle during diastole in young normotensive persons with and without familial predisposition to hypertension.

Methods

The study comprises 32 young normotensive subjects, 13 of them with positive familial predisposition to hypertension. Echocardiographic examination includes measurement of early (E wave) and atrial (A wave) flow velocity through the mitral valve, computation of the deceleration time of the E wave (dt) and isovolumetric relaxation time. Also measured are the flow velocity propagation (Vp), the time delay of the peak E wave velocity from mitral tips to apex and the flow in pulmonary veins: systolic, diastolic, reverse atrial wave and the duration of this wave. Ejection fraction and left ventricular mass index are calculated from long axis view of the left ventricle. Flow velocity is measured using 10 MHz Doppler effect ultrasound probe (see Figure 1). To efficiently establish individual velocity pulses, R peaks are extracted from the ECG signal. The ECG R waves provide precise time markers for the velocity signal. The measuring system consists of an Echocardiograph (ECG), a Doppler effect ultrasound velocity meter (Doppler) and a portable computer. The used ECG device has an analog output signal which is sampled using a parallel port attached A/D converter with 12 bit data conversion resolution. The Doppler has an integrated 12 bit A/D converter and serial data communication capabilities, thus constantly sending sampled data with a fixed frequency of 100 Hz. Both mentioned devices are connected to a portable computer. Whenever valid data is received from the Doppler, the ECG data is read. This way the data from Doppler and ECG are captured synchronously. Such events occur with the period of 10 milliseconds.

Figure 1 Custom designed arm attached Doppler velocity measurement probe holder

All analysis is carried out using the Matlab software package. The first step (pre-processing) of each measurement was to find and extract all R waves from the ECG signal. As presented in Figure 2, the R peaks are the splitting points of the synchronously captured velocity signal. Thus the result of every pre-processed measurement, is a large number of velocity wave sections (extracted between R peaks, see Figure 2). We call these sections 'velocity periods'. As the recording of each subject's data lasts five minutes, the number of extracted velocity periods varies from 280 to 400. The number of periods varies due to the heart rate frequency – with the R-R distance.

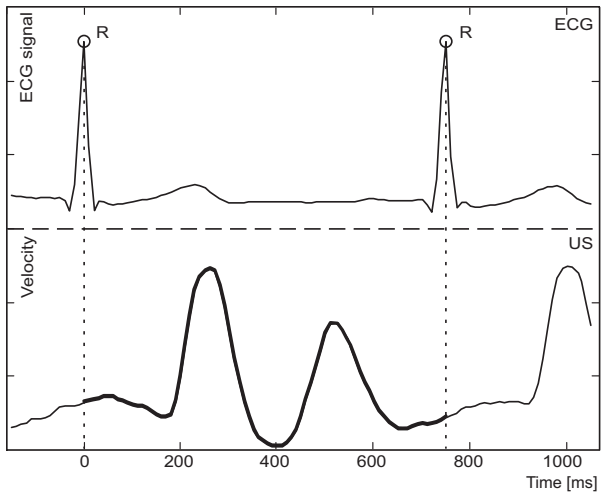


Figure 2 ECG signal and velocity measurement in radial artery

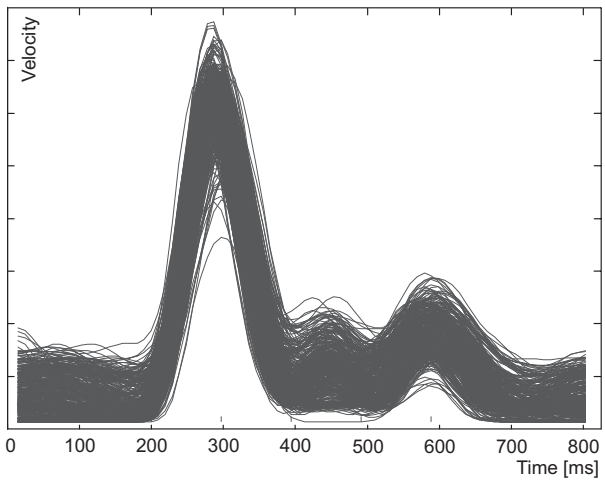


Figure 3 Extracted velocity periods aligned with starting points T0 (R peak); end points are shortened accordingly to the shortest R-R period

In Figure 3 all the extracted velocity periods from a single measurement are left aligned with their starting T_0 points (R peaks). As the R-R distance varies so does the length of individual velocity periods. To be able to display all the periods suitably they need to be right shortened, according to the shortest of the recorded periods. Because of the large dispersion of the velocity periods it is appropriate to calculate the mean period to read out the significant points easily (see Figure 4).

The mean velocity period is sufficiently described with six most significant points on the wave (T_0, …, T_5) and two inclinations of the primary wave (α_1, α_2) as presented in Figure 4.

The starting point on the mean velocity period is T_0 and matches exactly with R peaks. The point T_1 represents the start of the primary wave inclination. The maximum of the primary wave and also the global maximum of the mean velocity period is T_2. T_3 is the middle point between the primary and the secondary wave. The local maximum of the secondary wave is marked with T_4 and its ending point is marked with T_5.

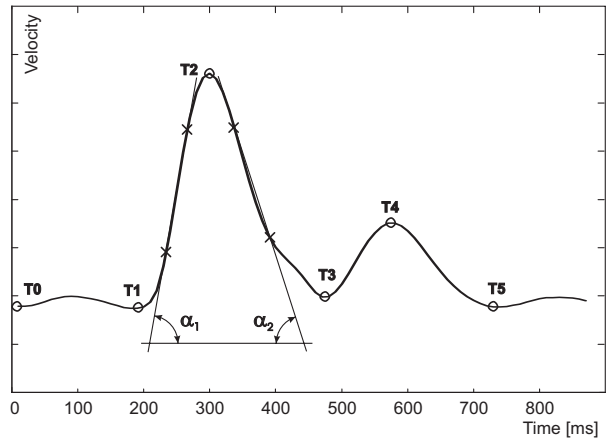


Figure 4 The characteristic points Ti and inclinations αj of the mean velocity period

The points T_1 and T_5, which mark the boundaries that separate the flat signal and the waves, are extracted by observing the slope change when the signal is close to its local minimum. Each characteristic point consists of two values, the value of the average velocity in that point and the time delay from the starting T_0 point. Further the inclinations of the rising and falling parts of the primary wave are calculated. The crosses in the Figure 4 represent the points that define the two inclinations. These points mark the 25% and 75% of the difference in amplitude between T_1 and T_2 for the rising and T_3 and T_2 for the falling edge of the primary wave. This way the inclinations of approximately linear segments of the primary wave are measured.

Results

The observed groups do not differ in parameters that describe diastolic function of the left ventricle i.e. E wave, A wave, dt of the E wave, Vp, IVRT and flow in pulmonary veins. However important differences in the time delay of points T_4 and T_5 regarding to starting point T_0 are found (see Table 1 and Figure 5).

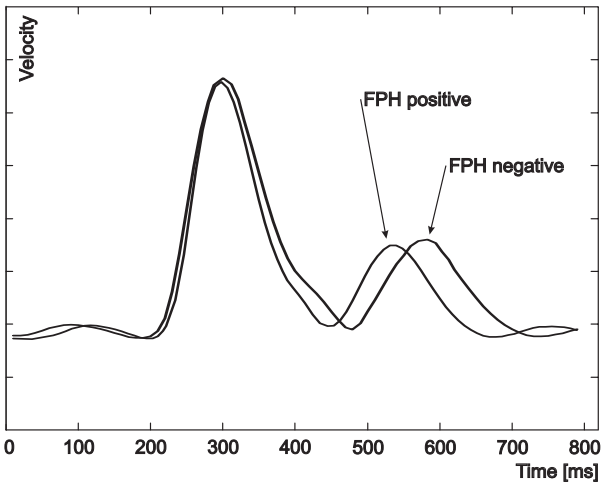


Figure 5 The shift of the secondary wave towards the primary in the FPH positive group

Table 1 Time delays for points T1, …, T5 according to T0 (ECG R peaks)

point	FPH negative	FPH positive	Sig.
T_1	180.8 ± 27.5	183.6 ± 15.7	NS
T_2	285.4 ± 29.3	280.0 ± 23.7	NS
T_3	457.7 ± 45.9	420.0 ± 46.5	0.059
T_4	573.1 ± 39.9	516.4 ± 68.2	0.019
T_5	744.6 ± 78.2	656.4 ± 105.2	0.028

FPH … familial predisposition to hypertension;
FPH column data … delay from the starting T_0 point, mean ± standard deviation [milliseconds];
Sig. … significance [t-test for equality of means].

Conclusions

Significantly shorter time delays for points T_4 and T_5 in the group with positive familial predisposition to hypertension are found. As the secondary velocity wave depends on the compliance of small vessels, the shift of the secondary wave towards the primary wave (i.e. smaller values T_4 and T_5) may be due to the structural or functional changes that are already present in small vessels of young normotensive subjects with familial predisposition to hypertension. The diastolic dysfunction of the left ventricle is composed of two main patophysiologic components: relaxation and compliance of the left ventricle.[14] Several authors found impaired relaxation and diminished compliance of the left ventricle in young hypertensive patients. Surprisingly enough our results do not show any difference in parameters that determine the diastolic function. We explain this by the fact that both groups are normotensive, so the left ventricle is not exposed to high pressure. On the other hand, the parameters, which are generally accepted and are used to assess the diastolic function, are not sensitive enough to distinguish both groups of normotensive subjects. On the base of our results we speculate that changes in small vessels occur before the alterations in relaxation and the compliance of the left ventricle in young normotensive subjects with familial predisposition to hypertension.

Medical aspect

Here we can stress again that there is a significant difference in the time delay of characteristic points on the secondary wave (T_4 and T_5) in the positive familial predisposition group. The shift of the secondary wave towards the primary wave is the consequence of diminished compliance of small arteries. There is no difference in parameters that determine the diastolic function. According to the presented results, we conclude that in young normotensive subjects with familial predisposition to hypertension changes of small vessels begin earlier than impairment of the diastolic function.

Technical aspect

During the research, we came across several ideas how to improve the system. The first step is to upgrade the system to be able to perform

significant point extraction and frequency analysis of velocity periods in real-time, during the data recording of each subject. In this way we intend to overcome the averaging of extracted velocity periods and perform the analysis on each extracted velocity period *on-the-fly*. Further we intend to add a new device that will enable us to perform blood pressure measurement synchronously with ECG and blood velocity measurement. This will provide our analysis with another data source that will improve the accuracy and broaden the field of our research.

References

1. John JN, Finkelstein SM: Abnormalities of vascular compliance in hypertension, ageing and heart failure. *Journal of Hypertension 1992*; 10(6): S61-S64.
2. Safar ME, London GM, Asmar R *et al.*: Recent advances on large arteries in hypertension. *Hypertension* 1998; 32: 156-161.
3. O'Rurke MF, Mancia G: Arterial stiffness. *J Hypertens* 1999; 17: 1-4.
4. Giannattasio C, Managoni AA, Failla M *et al.*: Impaired radial artery compliance in normotensive subjects with hyperholesterolemia. *Artherosclerosis* 1996; 124: 249-260.
5. Giannattasio C, Failla M, Grappiolo A *et al.*: Progression of large artery structural and functional alterations in type I diabetes. *Diabetologia* 2001; 44: 203-208.
6. Gatzka CD, Cameron JD, Kingwell BA *et al.*: Relation between coronary artery disease, aortic stiffness, and left ventricular structure in a population sample. *Hypertension* 1998; 32: 575-578.
7. Barenbrock M, Hausberg M, Kosch M *et al.*: Effect of hyperparathyroidism on arterial distensibility in renal transplant recipients. *Kidney Int* 1998; 54: 210-215.
8. Van der Heijden – Spek JJ, Steassen JA Fagard RH *et al.*: Effect of age on brachial artery wall properties differs from aorta and is gender dependent: a population study. *Hypertension* 2000; 35: 637-642.
9. Lehman ED, Gosling RG, Sonksen DH: Arterial wall compliance in diabetes. *Diabetic Med* 1992; 9: 114-119.
10. Makita S, Ohira A, Tachieda R *et al.*: Dilation and reduced distensibility of carotid artery in patients with abdominal aortic aneurysms. *Am Heart J.* 2000; 140: 297-302.
11. Guerin AP, Blacher J Pannier B *et al.*: Impact of aortic stiffness attenuation on survival on survival of patients in end-stage renal failure. *Circulation* 2001; 103: 987-992.
12. Van Dijk RA, Nijpels G, Twisk JW *et al.*: Change in common carotid artery diameter, distensibility and compliance in subjects with recent history of impaired glucose tolerance: a 3-year follow-up study. *J. Hypertension* 2000; 18: 293-300.
13. Klemenc M, Zimic N, Kranjec I, Jezeršek P: Diastolic function of the left ventricle and the activity of the autonomic nervous system in young hypertensive patients, *Med. Biol. Eng. Comp.* 1997; 34(1): 385-386.
14. Garcia MJ, Thomas JD, Klein AL: New Doppler echocardiographic applications for the study of diastolic function, *J. Am. Coll. Cardiol.* 1998; 32: 865-875.

Research Paper ■

PICAMS: Post Intensive CAre Monitoring System

Iztok Lebar Bajec, Damjan Oseli, Miha Mraz, Matjaž Klemenc, Nikolaj Zimic

Abstract. Medical staff on duty in hospitals provides high quality care for high-risk patients, especially those situated in intensive care units. Such care may present, in addition to great psychophysical pressure to the staff, considerable expenses for the hospital in question. In the paper we propose a Post Intensive CAre Monitoring System (PICAMS), based on the emerging technologies, which may eventually diminish the requirement for the doctor's non-interrupted presence, and allow for a faster transfer of the patient from the intensive care unit to the ordinary ward. What is more, it would also give patients the ability to move freely, which can be only positive for them in terms of better mood, quicker rehabilitation and indirectly lower costs for the hospital in question.

■ **Infor Med Slov** 2003; 8(1): 20-26

Faculty of Computer and Information Science, University of Ljubljana (ILB, DO, MM, NZ), Department of Cardiology, General Hospital of Gorica (MK).

Contact person: Iztok Lebar Bajec, Faculty of Computer and Information Science, University of Ljubljana, Tržaška 25, SI-1000 Ljubljana, Slovenia. tel. +386(0)14768785. email: iztok.bajec@fri.uni-lj.si.

Introduction

According to the American Heart Association[1] (AHA), the cardiovascular diseases are, with the ratio of one third, the leading cause of mortality in the developed world. Second most common cause are respiratory system diseases. More than 50% of hospitalized patients die of one of the two. What makes the costs higher are also the negative demographic trends and the constant changes of the nature of the diseases.

Patients with haemodynamically important cardiovascular and respiratory system diseases are usually hospitalized in Intensive Care Units (ICU). There they are connected to numerous devices (ECG, pulse oxymetry, non-invasive blood pressure and invasive monitors) that constantly measure the condition of their important physiological parameters. In this way constant supervision is provided, and in the case of emergency immediate help is available. Although the doctor on duty is in such a case usually alarmed by a pager, which allows them full mobility in the hospital, the mobility of the patient is neither desired nor possible, since they are connected to the monitoring devices. This does not at all help the patient's psychological condition. It makes them dependent upon the medical staff. And last but not least, providing constant 24h per day supervision increases the costs of treatment dramatically. This is why the health care policy is to find a way for faster transfers to ordinary wards of the patients hospitalized in the ICU without compromising the quality of the treatment. The moment of the patient's transfer from the ICU to the ordinary ward is of crucial importance for their health since it presents a consistent drop in the level of monitoring and cost of treatment (see line in Figure 1).

Short-Range Wireless (SRW) networks[2] (such as Bluetooth[3,4,5] or wireless LAN[6]), which are gradually becoming more and more widespread in modern information systems, enable us exactly that. Wearable computers, which used to be very rare in the past, are possible to get nowadays[7] and not far is the day when we will not even notice them since they will be so small and we will get accustomed to them[8]. These qualities exactly will enable us to set the patient free from the wires, thus allowing them full mobility. Later on, when the sensors are so small to be sewn in the clothes[9], it will be possible to measure their vital functions constantly. By adding some local processing and the use of a combination of SRW and LAN networks, we will enable the patient to move freely and yet be in constant high quality ICU type of supervision. Therefore the niche for our system is the phase when the patient is being transferred from the ICU to the ordinary ward. By using our system we can shorten the ICU hospitalization time and simultaneously lessen the drop in the level of monitoring (see curve in Figure 1).

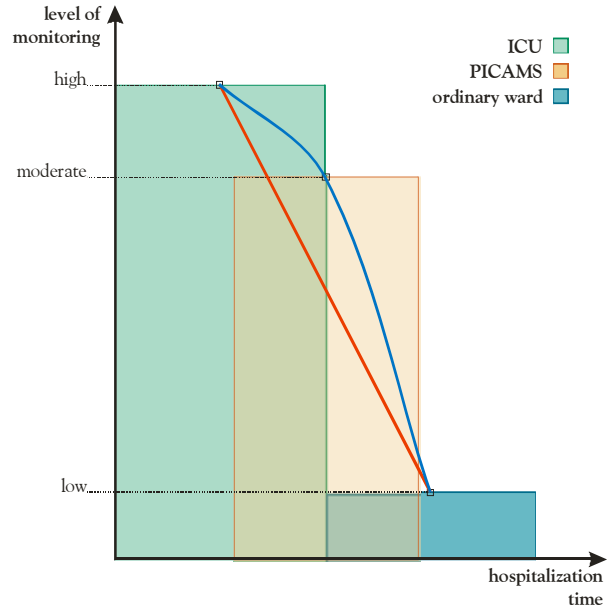


Figure 1 PICAMS niche

Not only that the PICAMS system will enable high quality ICU type of supervision of fully mobile patients; it will also allow virtual visits. For example, let us imagine a doctor on duty on one of their morning visits to the patients. They stop at a patient and check the display of their Personal Digital Assistant (PDA). There they can see the current values of the patient's vital functions (heart rate, blood pressure, blood oxygenation,

hemodinamically important arrhythmias, chest expansion, etc.), which are acquired by means of sensors sewn in the patient's clothes. Using the PDA the doctor can view the history of the patient's vital functions and a list of experienced critical situations. They can also view the progress of the medical treatment (list of prescribed medicines, digital medical imagery such as angiograms, test results, etc.). Such help enables the doctor to concentrate on the patients themselves, and not on the memorizing large quantities of data. All of the information shown via the PDA helps to illustrate the patient's current condition better.

To continue with, we try to explain the ideas behind the integration of the SRW technology into the information system of the hospitals in the form of a Post Intensive CAre Monitoring System (PICAMS), concentrating on a rough description of the concept of the system in its second part. In the third part we discuss the parameters, which are of the greatest challenge to the development of the system, and we finish with a short overview of the current status and future work on the system.

PICAMS

Our goal is to develop an intelligent monitoring system that would enable non-invasive monitoring of vital functions (heart rate, blood pressure, blood oxygenation, haemodinamically important arrhythmias and chest expansion) and at the same time allow the patient to move freely. The patient's ability to move freely would make them feel more secure and have a positive psychological effect. The latter, although welcome, represents a big challenge, particularly because the current state-of-the-art in the field of monitoring of vital functions is the following:

- *Clinical intensive care monitors*, the weight of which is more than 10kg and are thus not portable at all.
- *Holter monitors*, which are portable, but restricted only to ECG, and do not allow remote access of the acquired data.
- *Multi-channel patient recorders*, which weigh a little more than 1kg, and could thus be treated as portable, but have primarily a scientific function.

Besides the already mentioned patient's full mobility we want to give the doctor on duty a means of insight into the patient's current condition of vital functions regardless of the doctor's current location. Our aim is also to keep the ICU care standard and enable alarming of the medical staff in case of the patient's critical state. We have designed a modular system consisting of the following:

- PDD - Personal Diagnostic Device, which is dedicated to acquiring, local storage and basic analysis of the patients' vital functions.
- Combination of Bluetooth SRW and LAN networks used for data interchange.
- AP - Access Point server is intended for bridging between Bluetooth SRW and LAN networks.
- MD - Main Display is situated in the main monitoring room, enabling the medical staff on duty the insight into the current condition of the patient and representing a possible way of their alarming in case of emergency.
- IAS - Intelligent Analysing System is used for a thorough analysis of the received data of the patient's vital functions and setting off alarms in case of a detected critical state of one of the patients.
- PDA - Personal Digital Assistant enables the doctor on duty an insight into the patient's current condition and represents an efficient way of their alarming in case of a critical state of one of the patients.

- HIS - Hospital Information System is the existing information system of the hospital, used for storing the information about the hospitalized patients (treatment, list of prescribed medicines, digital medical imagery such as angiograms, test results, etc.).

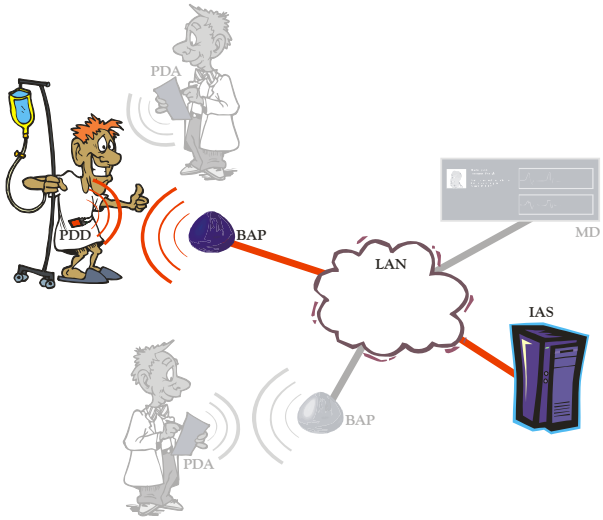


Figure 2 PDD detects unusual values of the patient's vital functions

A brief functional description

Patients with clearly expressed signs of cardiovascular or respiratory system diseases will be given a PDD and they will be hospitalized in ordinary wards. This will enable the patient the freedom of movement, which may have a positive influence on the patient's psychophysical condition and will accordingly shorten the hospitalization. A PDD will be constantly non-invasively acquiring the current values of the vital functions of the patient and storing them locally. Once stored the data will be analyzed by means of Soft Computing (SC) methodologies[10] based on Fuzzy Logic[11], Probabilistic Reasoning, Neural Networks[12,13] and Genetic Algorithms. Due to the limited processing power of the PDD this analysis will be above all dedicated to the detection of unusual values in the acquired data[14]. In case of such detection the PDD will, by means of AP, connect to the IAS and upload all of the stored data, followed by the continuous uploading of current values (see Figure 2).

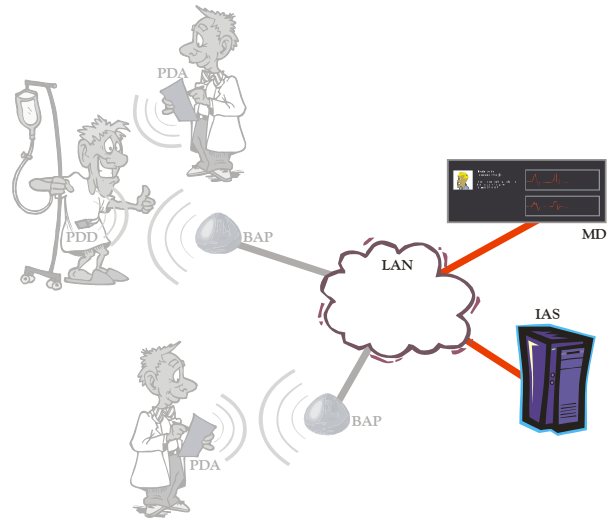


Figure 3 IAS alarms the medical staff on duty

IAS will store the uploaded information locally and with the help of SC methodologies analyze them thoroughly. Such analysis will be more accurate due to the processing power of IAS, and it will enable the detection and diagnoses of certain irregularities, e.g. arrhythmia, fibrillation and ischemic episodes, and detection of still unknown phenomena. In case of the detection of the aforementioned irregularities the IAS will, by means of AP, connect with the MD (see Figure 3). The MD will display the current condition of the patient, the history of irregularities and all additional information about the patient that is stored in the HIS (progress of medical treatment, list of prescribed medicines, digital medical imagery such as angiograms, test results, etc.).

At the same time the IAS will, by means of AP, establish a connection with the PDA of the doctor on duty. They will be able to see the alarm message with the patient's name and location (see Figure 4). They will then have the option to connect to the IAS and thus get the insight into the patients' current condition, history of irregularities and via HIS all additional information about the patient as well. So the doctor will, on the basis of the available

information, be able to make a decision about further actions concerning the patient.

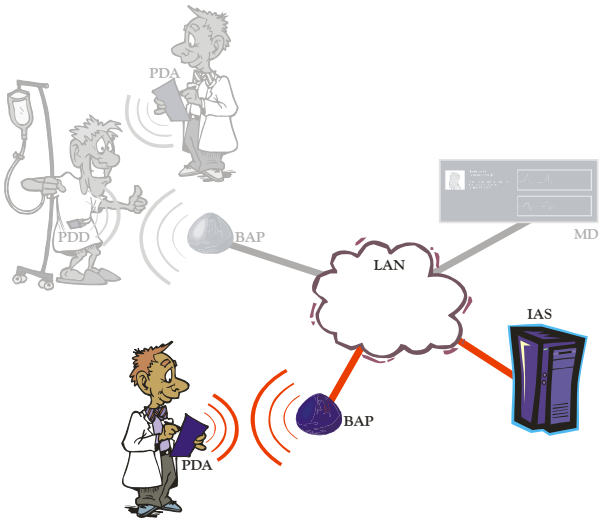


Figure 4 IAS alarms the doctor on duty

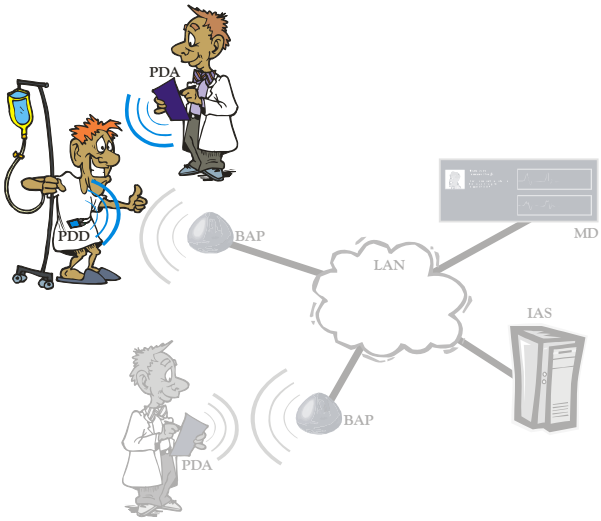


Figure 5 Doctor on duty performing a morning visit

On the other hand, on his morning visits the doctor on duty will have direct access to the PDD (see Figure 5) of the patient and see the current values of the patient's vital functions on their PDA. They will also be able to request the history of any irregularities as well as all the additional information concerning the patient. The latter will be received from IAS and HIS respectively.

Challenges

There will most definitely be challenging situations in the evolution of the PICAMS system. In the following paragraphs we enumerate the foreseen challenges with the possible solutions.

Integration of the doctor's knowledge into the PDD and IAS

Integrating the doctor's knowledge about the analysis of the fluctuations of the patient's vital functions presents a challenge by itself since it is usually based on "common-sense" reasoning. That is rather difficult to specify and thereafter to implement. It demands the use of SC methodologies[10], which consider inexactness and indefiniteness of the time dependent parameters and knowledge, and represent them in a humanly readable form.

Data safety

Safety in connection with remote monitoring may appear as a problem[16], but the Bluetooth SRW technology speaks in favour of this, because it per se enables authenticity and selection of the level of data encoding[4,5].

Reliability and range of the Bluetooth SRW network

At first we will reach the reliability by means of setting up periodical connections with the patient's PDD and checking for the proper functioning, the latter probably with the implementation of watch dog timers[17] (WDT) which are used to detect deadlocks. Later on we will concentrate on the range of the Bluetooth SRW network in the sense of Bluetooth aerials, spatial arrangement of AP servers and upgrading of the 7th level protocols of the TCP/IP protocol with special case handling algorithms (temporary loss of connection, implementation of Bluetooth roaming, etc.)

Reliability of the sensors

The largest obstacle here is how to ensure the patient's mobility but retain the reliability and accuracy of the measurements acquired. That was the reason we decided to try implanting (sewing in) all the aforementioned sensor systems into clothes the patients are wearing. In some cases (chest expansion measurements when the patient is talking) it can also happen that certain received values cannot be used, what expresses the demand for use of intelligent measurement acquiring algorithms.

False-positive and false-negative detection

The number of false-negative detections (the system does not trigger the alarm in case of a haemodynamically important arrhythmia) can be minimised to an acceptable level by means of parallel signal analysis (e.g. the analysis of multiple differentials of ECG signal) or multiple successive signal analysis.

The false-positive detection (the system determines a heart rhythm disorder when it did not occur) is not of vital importance for the patient, but rather a nuisance for the medical staff on duty. We can do away with it by using the classic medical monitoring systems with half the patients in the testing phase. By doing this we will attempt to achieve a shorter learning time of the system.

Testing

Testing of the system and the evaluation of its results is of great importance for the development of the project, due to it being of medical nature. The testing process will include the system tested in real life situations, for instance with patients with acute coronary syndromes, breathing irregularities with a considerate drop of oxygen level, etc. While testing, we will have to comply with all the required medical certificates and standards before actually deploying the system in hospitals.

Current status and future work

As mentioned before, the PICAMS system will enable non-interrupted, non-invasive monitoring of vital functions of the patient with expressed symptoms of haemodinamically important cardiovascular or respiratory diseases. It will enable the medical staff to transfer the patients from the ICU to ordinary wards faster without compromising the quality of treatment (see Figure 1), and thus reducing the cost of treatment while maintaining the ICU level of monitoring. The patients will therefore retain the best achievable level of treatment. What is more, a shorter reaction time of the medical staff may be achieved.

Currently we are working on signal analysis[14]. Following the Work-Centred Analysis (WCA) method[18] used in the design of information systems we are defining the system concept and the measurable characteristics and grading scales of the system that present the performance perspective of the WCA framework of the PICAMS system.

The PICAMS system is based on the current state-of-the-art technology and in the process of its evolution requires the development of yet unknown solutions. In spite of that we can already envisage the further expansion of the system. An optional GSM/GPRS mobile telephone with integrated Bluetooth support would enable the use of the PDD even outside the accordingly equipped hospital and thus enable the constant supervision and nearly ICU level of monitoring for the non-hospitalized patients.

References

1. American Heart Association, *2001 Heart and Stroke Statistical Update*, http://www.americanheart.org/statistics/pdf/hsstats 2001_1.0.pdf.
2. Leeper D.G., (2001), A Long-Term View of Short-Range Wireless, *IEEE Computer*, 34(6):39-44.
3. *Bluetooth Special Interest Group*, http://www.bluetooth.com.
4. Bluetooth Special Interest Group, *Specification of the Bluetooth system – CORE*,

http://www.bluetooth.com/developer/specification/Bluetooth_11_Specifications_Book.pdf.
5. Bluetooth Special Interest Group, *Specification of the Bluetooth system – PROFILES*, http://www.bluetooth.com/developer/specification/Bluetooth_11_Profiles_Book.pdf.
6. IEEE 802.11 WG, *IEEE 802.11 Wireless Local Area Networks*, http://www.ieee802.org/11/.
7. Ditela S., (2000), The PC goes ready-to-wear, *IEEE Spectrum*, 37(10):35-39.
8. Narayanaswami C., Kamijoh N., Roghunath M., et al., (2002), IBM's Linux Watch: The Challenge of Miniaturization, *IEEE Computer*, 35(1):33-41.
9. Schwiebert L., Sandeep G.K.S., Weinman J., (2001), Research Challenges in Wireless Networks of Biomedical Sensors, in *Proceedings of ACM SIGMOBILE*, Rome, Italy.
10. Yager R.R., Zadeh L.A., (1994), *Fuzzy Sets, Neural Networks and Soft Computing*, Van Nostrand Raiinhold, New York, USA.
11. Zimmerman H.J., (1992), *Fuzzy Set Theory & Its Applications*, Kluwer Academic Publishers, Dordrecht, Netherlands.
12. Miller A.S., Blott B.H., Hames T.K., (1992), Review of Neural Network Applications in Medical Imaging and Signal Processing, *Med. & Biol. Eng. & Comput.*, 30:449-464.
13. Micheli-Tzanakou E., (1995), *Neural Networks in Biomedical Signal Processing, The Biomedical Engineering Handbook*, (Bronzino J.D. ed.), CFC Press, IEEE Press, 917-932.
14. Klemenc M., Oseli D., Zimic N., (2002), Analysis of Velocity Contour and Diastolic Function of the Left Ventricle in Young Normatensive Subjects, To Appear in: *Proceedings of IFMBE 2002*, Reykjavik, Island.
15. Barro S., Marin R., Palacios F., Ruiz R., (2001), Fuzzy Logic in a Patient Supervision System, *Artificial Intelligence in Medicine*, 21:193-199.
16. Kara A., (2001), Protecting Privacy in Remote-Patient Monitoring, *IEEE Computer*, 34(5):24-27.
17. Mota A., Sampaio A., (1998), Model-checking CSP-Z, In: *Proceedings of FASE 1998*, 1382:205-220, Lisbon, Portugal.
18. Alter S., (1999), *Information Systems – A Management Perspective*, Addison-Wesley.

Research Paper ■

Concepts for Integrated Electronic Health Records Management System

Miroslav Končar, Sven Lončarić

Abstract. Due to a very sensitive nature of medical information, Electronic health records (EHCR) management systems are faced with a number of stringent requirements. More precisely, security problem that affects all the levels of communication architecture as well as all the medico/legal and ethical issues has been recognized as the primary step towards integrated health computing environment. This paper presents the solution for a functional EHCR management system that meets these strict requirements, but also follows the initiative taken by the Next Generation Network (NGN) approach, that addresses the problems of modularity and flexibility of medical information systems.

■ **Infor Med Slov** 2003; 8(1): 27-37

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia.

Contact Person: Miroslav Končar, Badalićeva 26, 10000 Zagreb, Croatia, email: miroslav_koncar@hotmail.com.

Introduction

Supported by the advances achieved in the computer science, communication technologies and especially by the growth of the Internet, medical information systems are becoming more and more present in physicians' practice. Telemedicine as the next step in medical care evolution has become possible in every aspect of its' essence. Distance in no longer a factor, and it is feasible to provide every person with high quality health care, independent of their current location.

However, when studying the requirements for medical information systems, the picture is somewhat different. Despite the fact that information systems used for different medical aspects raise diverse set of requirements and are evaluated by various performance factors, there are some basic issues that characterize practically every information and communication system used in medicine:

- Electronic health record (EHCR) management – every medical information system has to have either its' own implementation or a functional interface to the EHCR management.

- Security and data confidentiality – the system has to ensure that every piece of information is transferred, stored and retrieved in a secure way. This includes procedures like access control and the obligatory authorization at all levels in the health computing environment. Additionally, but not less important, the system has to respect patient's legal right to privacy, and ethical and legal policies required by the national regulations.

- Open system architecture – integration of different levels of medical care that would overcome today's boundaries is one common final goal set for medical information systems.

Security issues in medical information systems and EHCR management represent the starting point in the system design. Encryption of the communication channels, identification and authorization of users etc, solve just part of the problem. Confidentiality and privacy issues, legal, ethical and moral aspects of patients' personal and medical data as well as the integrity and professionalism of physicians' practice has to be preserved and supported by the medical information system management. EHCR management system is not just a computer science, and many other human disciplines take part in defining system requirements.

This paper describes the EHCR management system that successfully addresses all the issues mentioned above, and follows the ideas summarized in Next Generation Networks (NGN). Original software solution that supports the framework is also presented.

It is organized as follows: In Methods section basic set of requirements for the logical schema of the EHCR is presented; status and characteristics of information systems currently deployed for various medical purposes are discussed; communication architecture principles of the developed EHCR management system are introduced; the most important open R&D middleware issues are addressed; and details of the experimental laboratory implementation of EHCR management system are described. In Results section the performance results of developed encryption/decryption module are presented. Conclusion gives some final remarks as well as our plans and ideas for future development.

Methods

Open development issues for EHCR architecture and management systems

Health information systems today suffer from a number of significant problems.[1] Challenges that

need to be met by the systems of tomorrow include:

- support for a life-long health record
- interoperability among all the parties and systems used in patient care
- intelligent decision support
- domain size and rate of change
- systems obsolescence
- multi-contact healthcare system and mobile patients
- multiple medical cultures
- support for domain experts to have direct control over the information design and change management of their systems

Current work in health standards, notably by HL7, CEN, ISO and the OMG attempts to address some of these problems, as does implementation-based work including a number of policies-funded efforts around the globe. Very first efforts of EHCR architecture evolution were compromised by a number of factors that strongly influence the functionality and performance of the developed systems. Most of the organizations today agree on the basic concept, which is to separate context from the content even if the data is brought out of its original context,[2,3] in order to cope with a huge diversity between data recorded in the different departments of medical care. A clear context/content separation provides the users with medical data transfer without any loss of information by straightforward extract of the parts of patients' EHCR.

One of the cornerstones of the functional EHCR system is the security and confidentiality of patients' medical data. In the soul definition of EHCR it is stated, "the record is under control of the consumer and is stored and transmitted in the secure way",[4] which includes patients' ethical and legal rights to privacy and data confidentiality. Security includes obligatory authorization at all levels of the system, as well as the secure transfer of information between the end points of communication. Furthermore, the developers are advised to implement access-logging routines, which store all the transactions and data flows and are retrieved for auditing and legal purposes only.[5]

When considering the construction or review of good health information standards, insufficient attention was typically paid to the consequences for software construction and runtime systems. Many of the mayor problems of the past for information-intensive systems, including most EHCR and related systems, have to do with the inability to deal with change. This has led to an important turning point in the architecture design, which than significantly influences the development of the EHCR management systems. The EHCR specifications recommended by the standardization bodies are primarily focused on the logical health record architecture, i.e. the developers are provided with the formal model of the framework and generic features of the EHCR and there is no restriction regarding data formats in which the record are stored. More precisely, it is up to system architects decision to develop optimal EHCR management system that would suite their needs and requirements.

Today there are number of associations and standardization bodies that have organized task forces for developing EHCR architecture standards, some of which that were already mentioned above. In our case CEN standards and recommendations[6] have been the referral point when developing the EHCR management system. ENV 13606 recommendations with the title "Electronic Healthcare Record Communication" provide the principles, structures, terms, rules and formats for open and safe communication of EHCRs. Since management system framework and logical structure of health records are two separate problems, and although our focus was not on the CEN recommendations them self, we have respected the specifications and the developed

system offers a straightforward implementation of the ENV 13606 standard.

Communication platform for EHCR management system

The communication architectures and software solutions for medical information systems depend on number of parameters, such as the vendor of the hardware and software, requirements specification set for the particular example, the complexity of the services, type of processed information etc. Advanced, large-scale telemedicine applications are usually very performance sensitive, which could influence the developers decision for specific hardware solutions and programming techniques. As the result of these facts, at present most clinical software systems are "closed" with little or no interoperability between them.[7] Similar problem arise with the communication infrastructure solutions, which tend to be quite diverse. There are number of examples of teleconsultation or telesurgery systems that are based on communication protocols like ISDN or ATM.[8,9] Although these communication architectures fully satisfy their performance requirements, the services provided by the system are usually not transparent to the user, in sense that without specific equipment one is unable to use the application. In that case the application is not portable and cannot be used in other medical institutions without additional investments.

Integrated EHCR management systems should be able to manage patient information originating from various sources, and that is accessible independent of the users' current position and terminal. In that sense we have followed the guidelines of the communication networks convergence summarized in the NGN framework. The basic approach taken for NGN is one common network platform for transferring and serving different types of information, services and media. In this way it is intended to handle different media types and to use different services at the same time, with possible selection of well-defined Quality of Service (QoS) parameters. Today's concept of separate fixed and mobile network needs to be changed; it is assumed that a user is mobile within the system, and should be able to use all the provided services in a personalized and user-friendly way. The services developed for NGN tend to be inherently transparent, by which they assume IP based transfer protocol and are independent of a user's current position and terminal. As the result of this approach, the developed EHCR management system framework adopts multi-tier communication architecture with IP-based transfer and middleware layer that is able to satisfy the requirements of the NGN ideology, and the user does not require any special network equipment to access the medical data repository.

EHCR management system framework

Development of EHCR management system is based on distributed network architecture and CORBA[10] communication platform. The two dimensional view of the system architecture is shown in Figure 1. There are a number of advantages offered by the CORBA platform, which are of great importance to the application developers. In a distributed computing environment (DCE), where all system components are introduced as objects, CORBA provides a standard mechanism and tools for definition and implementation of the interfaces between objects. Communication between the components is accomplished using object references in such manner that the strict client/server distinction no longer exists. Also, by taking advantage of the common distributed object computing (DOC) communication platforms, we are provided with very important features such as complete platform and language independence, error handling, memory management etc.

Multi-tier communication architecture based on CORBA middleware platform is highly flexible and modular. Introduction of new features and addition of new object or modules usually does not require changes in the system in general, which is

very important in case of integration with other medical information systems. Since middleware communication layer contains most of the logic, potential upgrades of the system do not include changes and delivery of new client-side modules. The problems like diversity of data and media, localization based services and personalized delivery of information are addressed by classical middleware components in this communication model and comply with the concept of NGN architecture.

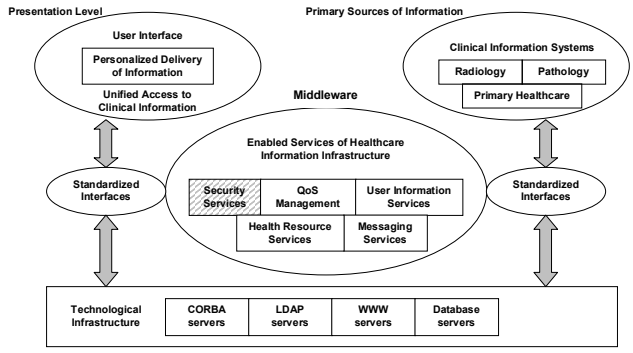


Figure 1 Multi-tier communication architectural framework for EHCR management system

In respect to sensitivity of medical information, close consideration has to be paid to the security framework offered by the CORBA platform. Implemented in the form of the CORBAservices, OMG among other things provides additional capabilities for security routines. This framework includes features like identification and authentication of users, authorization and access control, auditing, secure communication, non-repudiation and administration of various security policies. Undoubtedly, these functionalities are of great importance to performance sensitive applications such as EHCR management. However, they do not offer a complete solution in our case. Services provided by the standard DOC platforms cannot fully satisfy the legal and ethical rights of the patient and his/her medical data, and some additional measures have to be taken in order to meet those requirements.

In connection to middleware layer shown in Figure 1, framework for NGN communication architecture also includes some common functions such as registration, profile management, usage recording etc, which are also not CORBA's core functionalities. Those modules can be separated based on their orientation towards network transport layer or service layer, and together comprise a fully functional communication system for NGN architecture. Research in this area has been a subject of separate efforts conducted in our laboratory.[11]

Open R&D issues of standard DOC communication platforms

Employment of CORBA middleware communication architecture in medical information systems provides the application developers with number of advantages, especially in reference to classical client/server communication architectures. However, it still does not provide a complete solution for large-scale distributed applications. There are still some very important open R&D issues that are the subject of research in many laboratories and interest groups around the globe, two of which are of high importance for medical systems.

Communication overhead – traditionally, performance results of communication between CORBA objects and the application based on the ORB core were inferior to time requirements for client/server communication in two-tier network architecture.[12,13] Caused by the substantial progress in the standard middleware platforms, the performance results have significantly improved over the last couple of years,[14] however they still do not outperform classical client/server applications.

QoS management – first-generation DOC middleware was not targeted for performance sensitive applications with stringent QoS requirements. Not surprisingly, its efficiency, predictability, scalability and dependability was problematic. Over the last couple of years, however, the use of CORBA-based DOC

middleware has increased significantly in high performance distributed systems with real-time QoS requirements. Advancements of CORBA architecture model include Massaging[15] and Real-time[16] specifications that provide the control of many end-to-end ORB QoS policies such as timeouts or priority queuing, and implement standard interfaces for managing ORB processing, communication and memory resources. As a result, some of the focus of overhead, non-determinism, and priority inversion problems has shifted to the commercial operating systems and networks, which are once again responsible for the majority of end-to-end latency and jitter.[17] However, in respect to quality solutions introduced by these specifications, there are additional requirements set for QoS management framework like dynamic resource management, portable network QoS APIs, and multiple QoS property, which together comprise guidelines for future research and development efforts.

Beside common QoS parameters like predictable latency and jitter control it is important to keep in mind that the term QoS also includes a wide range of system properties like scalability, security and dependability. Common middleware solutions still do not provide a complete solution for those issues, and the improvements in this area has been the subject of work for number of research teams around the world, some of which have achieved high level of performance and usability.[18,19]

EHCR management system design

Following the principles and the requirements summarized in the previous sections, we have designed an experimental laboratory EHCR management system based on the communication architecture framework shown in Figure 1. The focus of attention has been paid to the middleware services that are targeted to satisfy very strict demands set for EHCR system implementation. Figure 2 depicts the laboratory system schema.

The module that requires special developers attention is preserving security of patients' personal and medical data. Like stated before, one of the basic requirements for EHCR management system is to ensure privacy, medico-legal and ethical needs of all persons known to the system.[7] The basic principle used in medicine is that the access to patient's information is granted to a very limited group of people, which posses the legal right to retrieve and edit patient's data.[20] In most of the cases that means that only general practitioner (GP) chosen by the patient is allowed to edit the medical record of the particular patient. In every other case the medical staff, other GPs or specialists have to acquire explicit patient's permit to access his/her information. Also, it is a common practice that all the data used for scientific research has to be used anonymously, except when the process itself requires personal information. Again, in that case the project has to acquire the permit of the subjects used in the study.[21]

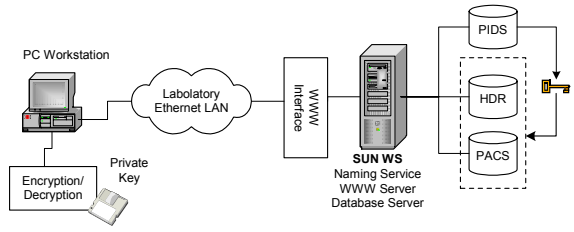


Figure 2 Laboratory System Infrastructure

Third part of the ENV 13606 recommendations called "Distribution Rules" addresses the problems of legal rights to view, edit and transfer a part of or a complete patient's medical record.[5] By adopting these proposals it is possible to define very detailed conditions when an access to patients' data can be granted, and what operations are permitted for different cases. Again, this represents only a mechanism to implement access rules at the data level, defined by the local users and national guidelines, and does not address logical and semantic security problems.

The solution implemented here provides the developers with the possibility to adopt all these specifications, but also introduces additional security measures against possible illegal and unauthorized access to the data repository. The

basic concept of the EHCR management system is strict separation of patient's personal and medical data.[22] More precisely, the EHCR system consists of two completely separated databases: Person Identification Service (PIDS), which contains personal and demographic information, and Healthcare Database Repository (HDR), which contains medical data only (Figure 3). The connection between those two databases is achieved through Master Patient Index (MPI), which is stored in the PIDS database in encrypted form, opposite to HDR where it is kept as plain text. The encryption and decryption processes are following the concepts of the asymmetric cryptography,[23] in which one key, called "public" is used for data encryption and the other, called "private" for decryption process. When a new account for a physician is opened in the system, the administrator creates at least two pairs of key (one pair for encryption and decryption of data, and the other for digital signatures), and digitally signs them. Public keys are added to the public keyring and published on the key server that in general could use LDAP service to manage public keys, but this is not mandatory . Private keys are transferred on floppy discs or CD-ROMs that also have a copy of the fingerprint of the administrators' signing key. These portable discs act as smart cards, without which one is theoretically unable to use the service. When encrypting data, every key first has to be checked for the administrator's signature. If the fingerprint on the physician's public key matches the one on the floppy disc, key is used. Otherwise, the key is treated as untrustworthy, revocation certificate is published on the key server, and the key is no longer used.

The administration of the patients goes as follows: when the patient is introduced to the system for the first time, his/her personal information including MPI is entered in the PIDS database. At the same time the patient chouses the GP, and the administrator selects corresponding public key for the first entry in the HDR database. Moreover, if the patient wants to enable more than one physician with the access rights to his/her medical information, theoretically there is no limit of public keys that can be used in encryption process. Also, if the patient during lifetime changes the GP, which usually happens a couple of times, the soul procedure that needs to take place is to replace the old encrypted MPI in the PIDS database with the new one. In this sense the EHCR framework completely follows and fully meets the demand that the consumer, in this case the patient, is the legal owner of the health record content.[1] In combination with the selected physician's public key the encryption process automatically also uses Master public key. The purpose of this key is a "safety net mechanism", which is used in special situations like the loss of a private key, i.e. when there is no other way to decouple the connection between PIDS and HDR archives. Since Master private key is able to unlock all the records, the size of these keys should be much bigger than standard key size and therefore harder to break. It is also very important that the private key of the pair is kept in a high security location like a safe .

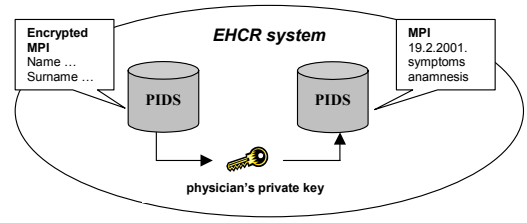


Figure 3 Configuration and logical architecture of EHCR system

With this model we have accomplished some very important features. First, because of the fact that all the relevant data is stored as plain text within both modules, PIDS module can serve as the primary interface to all the other modules defined on the system. It can be opened for access not only to physicians, but also to other groups of users like nurses, hospital administration staff etc. Furthermore, by storing medical data as plain text in the HDR data repository, this archive can easily be used for education or research purposes. The logical structure of the HDR database system is not limited by any means and can adopt specifications proposed in ENV 13606 document,

including the distribution rules that comply with the generic standard.

Developed security module resides on client side of the application (Figure 2). This partially steps out of the classical middleware networks framework, where the clients don't require special additions in order to be able to use the service. It is insecure to encrypt data on the server side, since the automation takes away all the control about the keys that are used in the process. Possible intruder could compromise some of the keys in such way that the server logic is unable to locate the problem. If the encryption is connected to the client side, the client has the ability to autonomously check every public key that is used in the process. Comparing the fingerprints of the public key used in the current process and the value stored on the floppy disc, the user is certain whether the public key is trustworthy or not. Decryption is even more insecure if it would be placed on the server side. In such case the server would need to have some kind of an access to physicians' private keys, which is inherently a security leak.

EHCR management system is based on CORBA's middleware architecture, which is responsible for object localization, naming service and communication between server and client components. By default clients access the application through their WWW browsers, i.e. using HTTP or HTTP/SSL communication protocols (Figure 2). System that adopts this type of communication architecture is straightforward and can be mapped to a wide variety of network access, which makes the application transparent to the details and characteristics of the client terminals. Especially if the system is being accessed from within the hospital LAN, clients can theoretically use raw CORBA's IIOP communication protocol, but its' efficiency from the performance point of view is rather questionable.

Finally, the question that is quite expected is how secure actually are we? Unfortunately, the answer to this problem is not completely unambiguous. Implementing CORBA security services and additionally employing SSL communication protocol, all the information is transferred in the encrypted form and therefore hidden from eavesdroppers. Local hospital networks are usually protected by the firewalls, and even if an attacker breaks into the system, without the possession of the private keys he is unable to compromise the medical data repository. Regular database backups as well as the use of digital signatures for data integrity check would easily diagnose possible misuse of the system. The last potential security gap are the keys used in the encryption/decryption process. In the recent years there has been a lot of discussion and research in this area that tried to find the answer to what level of security Public Key Infrastructure (PKI) offers. The results of empirical studies have shown that PKI can provide extremely high level of security. The size of used keys directly influence the complexity of the possible break. Private keys use additional security measures in such way that they are kept in the encrypted form on the floppy discs and protected by the passfrase. Bottom line, the consensus of developers and researchers is that this type of security infrastructure is more profound, sophisticated and qualitative than the standard measures used in paper health record management.

Results

Laboratory prototype performance results

To support the system architecture illustrated in Figure 2, we have built a laboratory prototype of the encryption/decryption module. Also, in order to simulate a real situation environment, we have designed an interface to the image management system that was developed during our previous work,[24] which among other characteristics featured a diagnostic images database repository. These images complied with the DICOMv3 standard and were taken using different image modalities. Prior to the integration the system needed some changes, because its' data archive contained parts

of patients personal information. That information was directly deleted from the database, and replaced with the newly created MPIs.

Details of laboratory devices, servers and terminals are as follows:

- Client terminal is a PC with a Pentium II processor and 256 MB RAM, and private keys are stored on floppy discs.
- WWW, CORBA and database servers are implemented on a single SUN Ultra 5 WS running on 400MHz RISC processor and 256 MB RAM.
- LAN is based on Ethernet 100BaseT technology.
- Asymmetric and symmetric keys used in encryption and decryption process are 1024 and 128 bits of size respectively.
- CORBA and database modules were implemented using Java™ Programming Language.

The encryption/decryption module is fully implemented using C/C++ Programming Languages. Encryption algorithms were provided by the GnuPG[25] application, and using provided APIs and Java Native Interface we have programmed a Dynamic Loadable Library (DLL) that interfaces Java GUI in WWW browser and controls the use of GnuPG application (Figure 4). Dynamic library also performs various security routines such as validation of the public/private keys and integrity check of the GnuPG application and Java Security File.

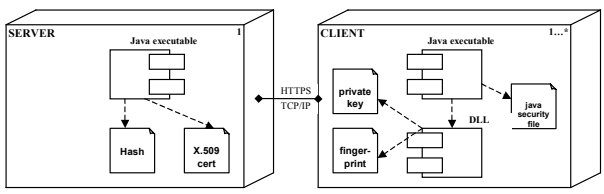


Figure 4 UML Component/deployment diagram of application modules

In order to gain a prospective about the performance characteristics we have conducted some basic measurements of time needed for different processes. The idea behind this work is to find out which module or process is most time consuming and in future could cause bad performance results. Our hypothesis was the encryption/decryption module could be rather sensitive procedures, since both processes need to access data stored on the floppy disc. Table 1 illustrates the measurements results.

Table 1 Security module performance results

Procedure	Time interval (sec)
1 thread ORB context initialisation	0.317
100 thread ORB context initialisation	2.173
Decryption	0.091
Encryption	0.364
US image retrieval	1.624
MRI image retrieval	1.861
CT image retrieval	2.381

ORB initialisation threads are used to simulate more than one connection at the same time, which is usually the case in a real situation. Second column depicts average time for 20 iterations.

Time needed to initialise the context or to retrieve and transfer images is highly dependant about the network conditions and number of clients accessing the service, whereas that does not influence the neither encryption nor decryption, since those are client-side processes. Image retrieval was measured from the time the user sent a request for an image and until that image appeared on the screen. Every image before being sent to client terminal must be additionally processed, since common WWW browsers like Netscape Communicator or Internet Explorer do not support DICOM image format. In this particular case images were formatted to PNG (Portable Network Graphics) format, which is supported by all of the standard browsers. Furthermore, at this point the laboratory prototype supports rendering only one image at the time, and therefore no image compression was used.

The comparison of the results shows that our assumption the encryption/decryption of patient IDs will strongly influence the time performance of the system was not confirmed. The use of GnuPG's implementation of crypto algorithms and DLL control has shown no disadvantages apart from being platform dependant. It is especially useful because the control of modules is distributed, i.e. Java module controls DLL and GnuPG and vice versa. This makes the application robust, since possible attacker would need to compromise both modules in order to do the damage. Table 1 also illustrates the complexity of QoS problem for image transfer. Namely, studies within imaging departments have shown that clinicians find it acceptable for studies to appear at workstations within 2 seconds of the images being requested,[26] which in our case the results for CT image retrieval do not satisfy this boundary.

Conclusion

Medical information systems require an increasingly broad range of features, which impose number of research questions on the computer and telecommunication scientists. Large-scale integrated EHCR systems are faced with many very strict requirements such as security and privacy, sensitivity and diversity of data and media types that need to be processed, support of various QoS aspects etc. Our goal was to make the EHCR management system secure from the unauthorized access from both outside and inside local hospital network, and at the same time to meet the demand of legal patients' ownership of their own medical data. The EHCR management system presented here successfully copes with the requirements and provides the developers with key performance factors such as flexibility, modularity and scalability. It also solves the necessity of controlling very strict access rights to patients' medical data, and fully respects the recommendations and proposals of CEN standardization body.

Our plans for further development include the design and research of the system modules according to the framework shown in Figure 1. We are planning to further investigate the performance issues of IIOP opposite HTTP/SSL communication protocol stack, based also on the results of some other research teams that show clear advantage of server-based applications and latter type of access. Parallel to that another research team in our laboratory is working on communication architecture for NGN. We are especially interested in the standard and healthcare specific middleware components, which would introduce important new features like personalization, profile management, terminology services etc. All of these address different issues of the QoS properties, which is of paramount importance to EHCR management system.

References

1. T. Beale, "Health Information Standards Manifesto", rev 2.5, 2001.
2. Torleif Olhede, PhL: "Archiving of Care Related Information in XML-format", Proceedings of MIE2000, pp. 1146-1150, Hannover, Germany.
3. F.H. Roger France, Cl. Beguin, R. van Breugel, et.al. "Long Term Preservation of Electronic Health Records. Recommendations in a large teaching hospital in Belgium", Proceedings of MIE2000, pp. 1146-1150, Hannover, Germany.
4. National Health Records Task Force, Australia, "The Health Information Network for Australia", 2000.
5. CEN/TC 251 Health Informatics, ENV 13606-3:1999, "Health Informatics – Electronic healthcare record communication – Part 3: Distribution Rules".
6. European Standards in Health Informatics official URL: http://www.centc251.org.
7. P. Schloeffel, T. Beale, S. Heard, et.al. "Background and overview of the Good Electronic Health Record", openEHCR Fundation, 2001.
8. I. Klapan et.al. "Our Tele-3D C-FESS remote surgical procedures: real time transfer of live video image in parallel with volume rendered models of surgical field", SoftCom 2001 proceedings, pp. 967-979, 2001.
9. Irfan Pyarali, Timothy H. Harrison, Douglas C. Schmidt: "Design and Performance of an Object-

Oriented Framework for High-Speed Electronic Medical Imaging", Computing Systems Journal, USENIX, Vol. 9, No. 3.
10. Jeremy Rosenberger: "Teach Yourself CORBA In 14 Days", SAMS Publishing, Macmillan Computer Publishing.
11. "Mobile Multimedia Portal – Role In UMTS Services", D. Šimić et.al. SoftCom2001 proceedings, Vol. II, pp. 529-535.
12. A. Gokhale and D. C. Schmidt, "Measuring the Performance of Communication Middleware on High-Speed Networks," SIGCOMM'96 proceedings, (Stanford, CA), ACM, 1996.
13. A. Gokhale and D. C. Schmidt, "Performance of the CORBA Dynamic Invocation Interface and Internet Inter-ORB Protocol over High-Speed ATM Networks," GLOBECOM'96, (London, England), IEEE, 1996.
14. G. Coulson, S. Baichoo, "Implementing the CORBA GIOP in a High-Performance Object Request Broker Environment", ACM Distributed Computing Journal, vol. 14, 2001.
15. CORBA Messaging Specification, OMG Document orbos/98-05-05 ed., 1998.
16. Realtime CORBA Joint Revised Submission, OMG Document orbos/99-02-12 ed., 1999.
17. D.C. Schmidt, M. Deshpande, C. O'Ryan, "Operating System Performance in support of Real-time Middleware", in Proceedings of WORDS'02, San Diego, 2002.
18. Yamuna Krishnamurthy, Vishal Kachroo, David A. Karr, et.al. "Integration of QoS-enabled Distributed Object Computing Middleware for Developing Next-generation Distributed Applications", ACM SIGPLAN Workshop on Optimization of Middleware and Distributed Systems (OM 2001), Snowbird, Utah, 2001.
19. D.C. Schmidt, D.L. Levine, S. Mungee, "The Design and Performance of Real-Time Object Request Brokers", Computer Communications, vol. 21, pp. 294-324, 1998.
20. Reglementation applicable aux banques de donnes medicales automatisees, Recommandation No. R(81)1 adopte par le Comite des Ministres du Conseil de l'Europe, Strasbourg, 1981.
21. Recommendation on the Protection of Medical Data R(96) of the Council of Europe, Strasbourg 1996.
22. Hans Schuell, Volker Schmidt: "MedStage – Platform for Information and Communication In Healthcare", Proceedings of MIE2000, pp. 1101-1106, Hannover, Germany, 2000.
23. Phil Zimmerman: "Introduction to Cryptography", copyright© 1990-1999 by Network Associates, Inc.
24. M. Končar and S. Lončarić: "WWW-based image management system", MIE2000, Hannover, Germany, 2000.
25. Gnu Privacy Guard Official WWW Site, http://www.gnupg.org.
26. CEN/TC 251 Health Informatics, "Quality of service requirements for health information interchange", Technical Report, 2002.

Research Paper ■

Testing the Suitability and the Limitations of Agent Technology to Support Integrated Assessment of Health and Social Care Needs of Older People

Haralambos Mouratidis, Gordon Manson, Ian Philp

Abstract. This paper explores the potential and the limitations of agent technology to support delivery of integrated information systems for the health and social care sector. In doing so, it points out the similarities and the mutual characteristics (such as distribution of expertise) of integrated health and social care information systems and agent technology. On the other hand, it identifies an important limitation of agent technology in the development of health and social care systems, which is the lack of a complete and mature analysis and design methodology that will provide guidance in the analysis and design of complex computer-based systems for health and social care. The Single Assessment Process (SAP) [http://www.doh.gov.uk/scg/sap/], an integrated assessment of health and social care needs of older people is used as an example of an integrated health and social care information system throughout the paper.

■ **Infor Med Slov** 2003; 8(1): 38-46

Author's institution: Department of Computer Science, University of Sheffield (H.M, G.M), Sheffield Institute for Studies on Ageing, University of Sheffield (I.P).

Contact person: H. Mouratidis, Department of Computer Science, 211 Portobello Street, Sheffield, S1 4DP, England. Email: H. Mouratidis@dcs.shef.ac.uk.

Introduction

In a distributed health care setting different health care professionals, such as general practitioners and nurses, must cooperate together in order to provide patients with appropriate care and must also work closely with social care professionals, such as social workers, because health and social care needs quite often overlap. National policy in England is to promote the Single Assessment Process (SAP), an integrated health assessment of health and social care needs of older people. The Single Assessment Process aims to create closer working for providing primary health and social care for older people and other groups.

Computerising this process will help to automate some of the administration tasks (such as the management of the health and social care teams, appointments procedures and secure and anonymised sharing of medical records) of the health and social care professionals and thus leave the professionals with more time for the actual care of the older person.

Nevertheless, computerising this process is not an easy task. Only about 1% (in UK) of some health and social care professionals are using computer systems. Apart from the complexity of such a system because of the integration, the security concerns and the mobility, there are other concerns related to the domain of providing health care to older people, which must be taken into account. Thus, the fact that most of the time professionals have to use the system whilst dealing with older people is a concern that must be taken into consideration. Thus, one of the most important decisions in computerising such a process is the choice of the technology that must be used.

The number of information systems based on Agent Technology has increased in the last few years; however, this is not the case in the development of health and social care information systems. More traditional technologies, such as web and database technologies, are still mainly used in the development of such systems. Although, each of those technologies provide advantages, they fail to adequately provide the main reason of using computer systems in the health and social care information systems, which is to reduce the workload and make procedures easier and quicker for health and social care professionals and to improve quality of care for the older people.

We believe that Agent Technology looks very promising to fulfil the requirements of an integrated health and social care computerised system. The scenario of distributed health and social care suits well to an agent-based system since there is distribution of data (each professional owns their data about the patient), cooperation between the different professionals (exchange of information about the older person), and different expertise areas between the professionals.

We are developing the electronic Single Assessment Process (eSAP), an electronic system to deliver an integrated assessment of health and social care needs of older people. The project is run jointly between the Computer Science Department of the University of Sheffield and the Sheffield Institute for Studies on Ageing (SISA), and it is funded by the RANK Foundation. Analysing and designing such a system is not an easy task. Apart from the complexity of the system itself, another important factor is the lack of an existing system, either electronic or "human". Thus, apart from trying to understand the functionality of the system, an understanding of the environment of the system is essential.

This paper, tests the suitability and (some of) the limitations of agent technology in the development of health and social care information systems. It focuses on the lack of a complete and mature agent oriented software engineering methodology, by comparing and evaluate four state of the art agent oriented software engineering methodologies. By performing this evaluation, this paper, originally identifies features and concepts, such as security and mobility, necessary to the development of health and social

care information systems that current agent oriented methodologies fail to capture.

The next section of the paper introduces the Single Assessment Process, an integrated health assessment of health and social care needs of older people. The Single Assessment Process is used as a case study throughout this paper, to identify suitability and limitations of the agent technology for the development of integrated health and social care information systems. The paper then describes the mutual characteristics between agent technology and integrated information systems for the health and social care sector, and also outlines the main limitations of the agent technology in the development of integrated health and social care computerised systems. Especially it describes the limitations of current analysis and design methodologies for health and social care agent-based information systems. Finally, some concluding remarks and directions for further research work are presented.

The Single Assessment Process case

Older people often have a complex mixture of health and social care needs. Several different health and social care professionals such as General Practitioners (GPs), nurses, and social workers are mostly involved in the care of older people with assessments often undertaken in the older person's home. Currently these professionals might belong to different organisations such as GP offices, community services, and social services. The current situation very often results in duplication of assessments, lack of awareness of key concepts of need and fragmentation of care.

National policy in England is to promote the Single Assessment Process (SAP), an integrated assessment of health and social care needs of older people. The Single Assessment Process aims to create closer working for providing primary health and social care for older people and other groups. With closer working, professionals will work in teams that will be responsible for the health and social care of the older person. Each team will consists of many different (health and social care) professionals that they will cooperate and share information between them. In addition each of those professionals will posses some expertise. The teams will promote person-centred care by allowing the patient to actively involved in their care.

The single assessment process will also provide the older person and their carer with a personal copy of their care plan to support person-centred care. The single assessment process involves three main stages; the initial contact, the overall assessment and the follow-up action. During the first stage, contact assessment will provide basic information. In the second phase an overall assessment using a validated assessment instrument, such as Easy-Care,[1] will take place. The third stage will provide older people with more care in particular problems (might be different problems for each individual such as housing and loneliness problems) and with more detailed assessment if appropriate. The selection of the problems is determined by the results obtained from the Easy-Care assessment instrument.

Computerising this process will help to automate some of the administration tasks, such as the appointments set up between the health and social care professionals, and the management of the health and social care teams, and thus leave the professionals with more time for the actual care of the older person. Furthermore, it will help older people to be actively involved in their health and social care, since they will have access to the system.

Suitability of Agent Technology

In a Multi-Agent System a software agent is considered as a problem-solving entity. In this type of system a complex task is accomplished by combining different software agents that possess

different skills (expertise). All the different software agents that co-exist in the system possess their own expertise, which can be related to the other agents of the system but it is distinct, and they use this expertise at different stages of the solving process in order to accomplish the system's aim.

A reason that makes this kind of Multi-Agent System very attractive to researchers is because these systems can be viewed as human societies in which the roles of the human beings are played by the software agents, or there is a mix between human beings and software agents that cooperate and communicate in order to solve a complex problem. Thus, a software agent in the system can be viewed either as an entity that acts on behalf of a human, or as an autonomous entity that possesses some expertise and it is able to cooperate and communicate.

Table 1 Mutual characteristics of the single assessment process and an agent-based system

Single Assessment Process	Agent-Based System
Professionals	Software Agents
Cooperate	Cooperate
Expertise	Expertise
Distribution	Distribution

It is concluded from the above that the single assessment process can be modelled into a multi agent system that consists of many different software agents (and humans), which can cooperate with each other, share information (distribute information) and each of them possesses some expertise. A summary of the mutual characteristics of the Single Assessment Process and a multi-agent system is given in Table 1.

In our system, the software agents will act on behalf of professionals. Each professional will have their "own" software agent, which they will customise according to their needs. The agent will have enough information about the professional, such as personal information and professional commitments, and it will be intelligent enough (capable of analysing the information and take decisions) to enable it to act on their behalf, and also negotiate for the interest of the professional.

The system will have the following characteristics:

1. The system will consists of software agents as well as human professionals.
2. Each professional will have his/her software agent.
3. The professionals must able to customize their software agents through an easy-to-use interface.
4. The system must be developed with mobility in mind since many of the professionals will use it whilst in the older person's house
5. The system must be secure.
6. The software agent will be capable of analysing information and taking decisions. Also, it will have information about the professional (personal and professional) that will be able to act on his/her behalf.
7. Software Agents in the system will be able to communicate between themselves as well as with the human professionals.

Agent technology is suitable to fulfil all the characteristics required by such a system. However, in developing such a complex system the first step is to analyse and design the system. In our case we have to model the single assessment process into a multi-agent system in which software agents and humans cooperate together to deliver better health.

Limitations of Current Agent Based Methodologies

Before starting the implementation of such a system, it is necessary to fully analyse and design it. Designing and analysing a system before it is actually implemented helps to better understand the system requirements, the user needs and thus the whole system. Starting to implement a system without a design might work for small systems that only require a few lines of code and one developer. However, trying to implement complex distributed systems (like health care systems), that require hundreds or thousands of lines of code, and large teams of developers, without a design proves to be a nightmare.

To help with analysing and designing systems, partial methodologies have been created. But as Kinny et al argue "If multi-agent systems are to become widely accepted as a basis for large scale applications, adequate agent-oriented methodologies and modelling techniques will be essential[2] ". This is not just to ensure that systems are reliable, maintainable, and conformant, but to allow their design, implementation, and maintenance to be carried out by software analysts and engineers rather than researchers. Thus, it is recognised amongst the agent research society[3,4,5,6,7] that there is a need for a complete analysis and design methodology for multi-agent systems.

The main role of such a methodology will be to help in all the phases of the development of an agent-oriented system. There are plenty of issues that must be considered when analysing and designing such systems, such as coordination, cooperation and communication between the agents[7] . In addition, analysing and designing health care information systems, such as the Single Assessment Process, involves integration and sharing of information and introduces some extra requirements. First of all, security is a major concern in such a system. The system will contain personal and medical information and thus must be very secure. Also, the methodology must give developers the flexibility to use any kind of software agents that might be needed. In the Single Assessment Process computerised system, mobile agents are most likely to be used. Mobile agents are software agents that can be moved to other computers in the network to obtain some information that the user requires. Another important point is the interface that will be used to enable the health and social care professionals to interact with their personal agents. Designing and analysing such interfaces will help to make the system easier for the users (professionals). Thus, the methodology must support the analysis and design of the characteristics that a health and social care system introduces.

Testing Current Agent-Oriented Methodologies

Although, there are methodologies for analysing and designing agent-based systems, these are failing when designing agent-based information systems for the health and social care. Four leading methodologies (GAIA, MaSE, MESSAGE and MASB) for agent-based systems were reviewed and evaluated. These methodologies were chosen since they represent the state-of –the-art in agent-oriented software engineering.

- GAIA methodology is specifically tailored to the analysis and design of agent-based systems. The methodology deals with both the societal (macro) level and the agent (micro) level aspects of the design. The developers of the methodology argue that GAIA allows a designer to systematically go from requirements to a detailed design that can be implemented. GAIA methodology separates the analysis and the design phases explicitly and are considered (both analysis and design) as a process of developing models of the system under development.[8]

- The Multi-agent Systems Engineering Methodology (MaSE) is similar to the GAIA methodology but more specialised for its use in the distributed agent

paradigm and goes further by providing support for generating code using the MaSE code generation tool. One of the main differences between this methodology and other agent-based methodologies is that in the MaSE methodology the general components of the system are designed before the system itself is actually defined .[9]

- The Methodology for Engineering Systems of Software Agents (MESSAGE) focuses on the analysis and design steps. It provides a set of analysis and design models suitable for analysing and designing Multi-Agent Systems and gives some recommendations on how these models can be built. The methodology starts by using UML and extends it by adding entity and relationship concepts required for the analysis and design of Multi-Agent Systems. Thus, it provides additional "knowledge level" concepts and uses the methodology's meta-model in order to define these concepts.[7]

- Multi-Agent Scenario-Based (MASB) method is a multi-agent system design approach based on the analysis and design of scenarios involving human and artificial agents. The methodology can be used to design multi-agent systems in the area of cooperative work. The method is divided into two phases, the analysis and the design. In the scenario description step, the designers give a description of a scenario, using natural language, which emphasises the roles played by the humans and the agents. Thus, the typical information exchanges between the agents and the humans are described along with the events and the actions performed by the agents.[10]

Eleven (11) evaluation points were identified for the evaluation of these methodologies. These points are enough to obtain adequate results and conclusions about the methodologies, since they evaluate the two main aspects necessary for every agent-oriented methodology. That is, the modelling of the agents of the system and some software engineering concerns, such as flexibility that every methodology should provide the engineer. The evaluation points are as follows: (1) Identification of the agents of the system - can the methodology identify the correct number of software agents of the system; (2) Communication - can the methodology capture all the available communications that happen in the system; (3) Intelligence - can the methodology capture the intelligence of the agents of the system; (4) Agent Expertise - can the methodology capture the expertise that the agents of the system must posses; (5) Interfaces - can the methodology model the interfaces used from the humans to communicate with their software agents in the system; (6) System Environment - can the methodology capture adequately other aspects of the system environment, such as the borders of the system; (7) Mobility Aspects - can the methodology capture the mobile agents of the system; (8) Security - can the methodology capture the initial requirements for the security of the system; (9) Easy-to-Use - is the methodology easy-to-use and understand when employed by not very experienced software engineers; (10) Consistency - does the methodology provide rules to test the consistency between the analysis and the design phases; (11) Flexibility - does the methodology allow for design flexibility.

In order to evaluate the methodologies an appointment system was partially analysed and designed using each of the above methodologies. In this system, each professional has a personal software agent and the software agents have enough information about the professionals so that they (agents) can book appointments on their behalf. The decision of employing an appointment system took place since such a system provides all the functionality necessary to evaluate a methodology according to the above-mentioned evaluation points. The results obtained from the evaluation are shown in table 2. The "√" symbol means the methodology provides means to model

adequately the characteristic. "~" means that improvements are needed in order to fully capture the characteristic, while "X" means the methodology cannot capture the characteristic.

The approach of analysing and designing the appointment system with each of the above-mentioned methodologies proved to be quite useful for many reasons. First of all, valuable knowledge was obtained for each of the methodologies. By obtaining this knowledge and having a common ground for comparison (the appointment system) the final evaluation came natural. Also the approach of using different methodologies in the analysis and design of the system provided us with very useful feedback for the system itself since we obtained different "views" of the system from different perspectives.

Table 2 identifies important limitations of the methodologies under evaluation. The most important of these limitations, especially in the design of health and social care computer systems, is the lack of modelling the security aspects of the system under development. The need for security is a major concern, especially in health information systems, since revealing a medical history could have serious consequences for particular individuals. Although security aspects are taken into consideration after the design of the system has been finished and during the implementation stage, this is not the best approach. Security must be concerned from the start, since there is not much hope to design a secure system by making changes and additions to an insecure system.

In addition, the capturing of mobile agents is another important limitation of the current methodologies. Mobile agents are a crucial part in most agent-based systems and the lack of a model to capture them restricts the usefulness of the existing methodologies. Especially in the health and social care information systems, mobile agents can play a major role. For example a mobile agent can migrate off a mobile device (from a health care professional visiting an older person in their house), and roam the Internet to gather information. Since it is not in continuous contact with the device (it can transfer itself in the requested network location), even if the professional disconnects (to visit another patient) the mobile agent is not affected. Thus, when the professional re-connects the mobile agent will deliver the requested information to the mobile device.

Table 2 Comparison of the methodologies

Characteristic / Methodology	GAIA	MaSE	MESSAGE	MASB
1	√	√	√	√
2	~	√	√	√
3	√	√	~	~
4	~	~	~	~
5	X	X	X	X
6	~	~	~	~
7	X	X	X	X
8	X	X	X	X
9	√	√	~	~
10	√	√	√	~
11	√	√	√	√

Finally, the modelling of the interfaces between the humans and the software agents involved in the system is an important characteristic that is not captured by any of the existing methodologies. Although there is rich research on the Human-Computer Interfaces (HCI) area, it is important to identify how the design of agent-based systems is depended on the interfaces between the humans and the software agents of a system.

Conclusions and future work

This paper argues that agent technology has the potential to support the development of health and social care information systems. Nevertheless, although there are some attempts of employing agent technology in health and social care information systems, [11,12] these are the exemption rather than the rule. Mostly, developers of health and social care information systems are reluctant to use agent technology. One of the reasons is the lack of a mature and complete analysis and design

methodology for agent-based health and social care information systems, which will help developers throughout the development of such a system. Thus, in order for the agent technology to be widely accepted in the development of computer based medical systems, it is necessary to develop a complete and mature analysis and design methodology to support all the stages of the development of agent-based medical systems.

The results presented in this paper are the initial exploration of the suitability of agent technology in the development of health and social care information systems. Much remains to be done for further research. Future work is directed towards extensions to support the integration of security and the modelling of mobile agents during the development of our system. Thus, we are extending Tropos [13,14] agent-oriented methodology to support integration of security aspects during the analysis and design stages. In doing so, we aim to provide a set of concepts and notations customised to security modelling, and a clear process that will guide the developer during the development stages. Future work on mobile agents modelling involves the identification of a process that will help developers to correctly identify the kind of agents that are suitable for their system, and then provide concepts and notations to capture them.

By doing so, we hope that agent technology will advance and will be easier to use in the development of health and social care information systems providing all the advantages that are described in this paper, such as problem-solving capabilities, information sharing, encapsulation of expertise, and even more.[11,12]

Acknowledgements

The first Author is grateful to the RANK Foundation for the funding of his research project, in which this work was carried out.

References

1. Philp I: Can a medical and social assessment be combined?. Journal of the Royal Society of Medicine 1997, 90(32), 11-13.
2. Kinny D, Georgeff M, Rao A: A Methodology and Modelling Technique for Systems of BDI Agents. Lecture Notes in Artificial Intelligence 1996, Springer-Verlag, Vol 1038.
3. Jennings N. R: An agent-based approach for building complex software systems. Communications of the ACM 2001, 44 (4).
4. Wooldridge M, Ciancarini P: Agent-Oriented Software Engineering: The state of the art. Lecture Notes in Artificial Intelligence 2001, Springer-Verlag, Vol. 1957.
5. Wooldridge M: Agent-Based Software Engineering. IEEE Proceedings on Software Engineering 1997, 144 (1), 26-37.
6. Wooldridge M, Jennings N.R: Software Engineering with Agents: Pitfalls and Pratfalls. IEEE Internet Computing May/June1999.
7. Evans R, Kearney P, Stark J et al: MESSAGE: Methodology for Engineering Systems of Software Agents, AgentLink Publications, September 2001.
8. Wooldridge M, Jennings N.R, Kinny D: The GAIA methodology for agent-oriented analysis and design. Journal of Autonomous Agents and Multi Agent Systems 2000, 3(3), 285-312.
9. Wood M.F, DeLoach S.A: An Overview of the Multiagent Systems Engineering Methodology (MaSE), Proceedings of the First International Workshop on Agent-Oriented Software Engineering (AOSE), June 2000, Ireland.
10. Moulin B: Collaborative Work Based On Multi Agent Architectures: A Methodological Perspective. Soft-Computing – Fuzzy Logic and Neural Networks 1994.
11. Huang J, Jennings N.R, Fox J: An Agent Architecture for Distributed Medical Care. Lecture Notes in Artificial Intelligence 1995, Springer-Verlag, 219-232.
12. Moreno A, Valls A, Bocio J: Management of hospital teams for organ transplants using multi-agent systems. In the Proceedings of the 8th European Conference on Artificial Intelligence in Medicine, Portugal, July 2001.
13. Castro J, Kolp M, Mylopoulos J: A Requirements-Driven Development Methodology. In Proceedings of the 13th International Conference on Advanced

Information Systems Engineering (CaiSE'01), Switzerland, June 2001.
14. Mouratidis H, Giorgini P, Manson G, Philp I: Using Tropos Methodology to Model an Integrated Health Assessment System. Proceedings of the 4th International Bi-Conference Workshop on Agent-Oriented Information Systems (AOIS'02), Canada, May 2002.

Research Paper ■

Non-Invasive Methods for Children's Cholesterol Level Determination

Petra Povalej, Peter Kokol, Jernej Završnik

Abstract. Today, there is a controversy about the role of cholesterol in infants and the measurement and management of blood cholesterol in children. Several scientific evidences are supporting relationship between elevated blood cholesterol in children and high cholesterol in adults and development of adult arteriosclerotic diseases such as cardiovascular and cerebrovascular disease. Therefore managing and measuring the level of blood cholesterol in children is very important for the health of the whole population. Non-invasive methods are much more convenient for the children because of their anxieties about blood examinations. In this paper we will present a new attempt to find non-invasive methods for determining the level of blood cholesterol in children with the use of intelligent systems.

■ **Infor Med Slov** 2003; 8(1): 47-55

Author's institution: University of Maribor, Faculty of Electrical Engineering and Computer Sciences (PP, PK), Adolf Drolc Health Centre (JZ).

Contact person: Petra Povalej, University of Maribor, Faculty of Electrical Engineering and Computer Sciences, Smetanova 17, 2000 Maribor, email: petra.povalej@uni-mb.si.

Introduction

For most of the children every medical examination causes anxiety. Therefore finding some painless, timesaving methods as a substitute for expensive and time-consuming medical examinations would be a great success.

In this paper we will present current results of our efforts to find a non-invasive method for determining the level of blood cholesterol in children. This research was performed in cooperation with the Adolf Drolc Health Centre in Maribor, where 729 five-year-old children were examined. For each child the following data was gathered: gender, height, weight, head circumference, chest circumference, upper arm circumference, skinfold thickness, pulse, systolic and diastolic blood pressure, blood cholesterol level, LDL and HDL cholesterol level and triglycerides. Our aim was to find some relations among these attributes and the blood cholesterol level in order to predict (with a considerate degree of accuracy) whether the child has elevated blood cholesterol without blood testing.

What are Lipids and Cholesterol?

Lipids are fats in the bloodstream and in all of the body's cells. Among the components of the lipids in blood are triglycerides, which come from the fats eaten or being made by our body from other things eaten, including carbohydrates. If the calories consumed are not used immediately for energy, they are stored as triglycerides in the fat cells. They are released when the body needs energy, such as between meals. Having too many triglycerides has been linked to coronary artery disease.

A certain amount of cholesterol is important to the healthy function of our body. It is an oily substance that is used to build cell walls and form some hormones and tissues. High level of cholesterol in the blood, known as hypercholesterolemia, is a major risk factor for heart disease and can lead to a heart attack.

We accumulate cholesterol in two ways. The liver produces about 1,000 milligrams of cholesterol a day. Another 150 to 250 milligrams comes from the foods we eat.

Cholesterol and triglycerides are carried in the bloodstream by lipoproteins. Two kinds - low-density lipoproteins (LDL) and high-density lipoproteins (HDL) - are the most important.

Low-density lipoproteins, sometimes called "bad" cholesterol, are the primary cholesterol carrier. If there is too much LDL in the bloodstream, it can build up on the walls of the arteries that lead to the heart and the brain. This buildup forms plaque, a thick, hard substance that can block arteries. If a blood clot forms and gets jammed in a clogged artery leading to the heart or the brain, you could have a heart attack or a stroke.

The remainder of body's cholesterol, about one-third to one-fourth of it, is moved through the blood by high-density lipoproteins, or HDL. These are sometimes known as "good" cholesterol because they carry the cholesterol away from the arteries and back to the liver, where it is passed from the body.

High levels of LDL cholesterol (the bad cholesterol) are related to a risk for heart disease and stroke, whereas high levels of HDL cholesterol (the good cholesterol) can protect against these. Conversely, low levels of LDL cholesterol are good and low levels of HDL cholesterol are bad.

High levels of triglycerides and cholesterol are a major risk factor for coronary artery disease, also known as atherosclerosis. An adult's heart attack or stroke has its origins in the development of atherosclerosis, which begins in earnest during the late teen years. Paying attention to cholesterol levels in children can lead to proper diet and medical treatment throughout life, slowing the progress of atherosclerosis, and either delaying or preventing heart attacks and stroke.

Childhood cholesterol levels were not tracked until recently, and some experts think that high cholesterol in kids is a major underreported public health problem. The health risks associated with high cholesterol - heart disease and stroke, for example - generally don't show up for years, even decades, so making the connection between kids and cholesterol is difficult for many people. Children with elevated cholesterol levels may be a precursor, some doctors believe, to a generation of teenagers with cardiovascular disease.

Decision trees

Decision trees have been successfully used for years in many decision-making applications[6]. One of the main advantages of using decision trees, in compare with other methods of machine learning, is very simple and clear representation of the path to acquired decision. Inducing a decision tree is a form of machine learning, where we extract knowledge from a set of examples (objects) and present it in a 2-dimensional form of a decision tree[1].

A decision tree is inducted on a training set, which consists of training objects. Every training object is completely described by a set of attributes (object properties) and class (decision, outcome). Attributes can be numeric or discrete, but numeric attributes are not suitable for learning a tree. Therefore they must be mapped into a discrete space.

There are two types of nodes in a decision tree: internal and external nodes. Each internal node (non-terminal node) contains a test of a specific attribute value. External nodes (terminal nodes, decision nodes, leaves) are labeled with a class, which represents a decision. Nodes are connected with edges (links). Edges are labeled with different outcomes of a test performed on an attribute in a source node.

For testing a decision tree a testing set is used. Testing set consists of testing objects described with the same attributes as training objects except that testing objects are not included in training set.

The results are described with specificity, sensitivity and total accuracy. Specificity is defined as the number of correctly classified children with normal cholesterol level divided by the number of all children with normal cholesterol level. Sensitivity is the number of correctly classified children with abnormal cholesterol level divided by the number of all children with abnormal cholesterol level. The overall quality of a decision tree is described with total accuracy.

The tool we used is called MtDecit 3.0[2]. It basically follows the same principles as many other decision tree building tools, but additionally implements different ways of numeric attribute's discretization. [3]

Genetically induced decision trees

Disadvantages of classic decision tree induction such as sensitivity to noise (missing or corrupted data),[8] encouraged us to try another method of machine learning that combines two methods: decision tree induction and genetic algorithms. This hybrid method merges all advantages of both methods and therefore usually gives better results.

Genetic algorithms are based on the evolutionary ideas of natural selection and genetic processes of biological organisms.[4,8] They are often capable of finding optimal solutions even in most complex search spaces or at least they offer significant benefits over other search and optimization techniques.

The first phase of genetic process is generation of initial population. Enough individuals have to be constructed to fulfill the whole population. Every individual in this method is represented as a decision tree.

Second phase of genetic process is evolution of population with the use of three genetic operators. First genetic operator is selection when individuals are evaluated on the basis of fitness function and the best ones (parents) are chosen for creating new individuals (children) with a second genetic operator - crossover. This way a new population is created.

After the new individual is constructed by crossover, a genetic operator of mutation is applied with certain (low) probability. Mutation serves as a random change of individuals with intention to find an optimal solution to the given problem faster and more reliably. The tool we used in research is Vedec.[5]

Data collection

The database used in the study included 729 objects described with 14 properties (gender, height, weight, head circumference, chest circumference, upper arm circumference, skinfold thickness, pulse, systolic and diastolic blood pressure blood cholesterol level, LDL cholesterol level, HDL cholesterol level and triglycerides). Since we were focused on the problem of determining cholesterol levels, we defined the last four attributes as outcomes (class attributes). All other properties were used as attributes.

Because of the nature of decision trees, numeric class attribute had to be mapped into two discrete values (normal/abnormal). For discretization we used standard recommended values presented in table 1.

All objects in the database with more than 90% of missing parameters and all objects with unknown class attribute were deleted, so the new database has reamining 712 objects.

Training and testing sets

For the training purposes we had to build a training set from the filtered database. While examining our database we found out that the percentage of objects with abnormal levels of blood cholesterol, HDL and LDL cholesterol and triglycerides were substantially lower than 50% (see table 2). Such distribution could influence decision trees in such a way that they would learn more about normal blood lipids than abnormal.

Table 1 Normal levels for blood lipids in children

Lipids	Normal level
Blood cholesterol level	< 5 mmol/l
HDL cholesterol level	> 1 mmol/l
LDL cholesterol level	< 3 mmol/l
Triglycerides	< 1,4 mmol/l

Table 2 The number of objects with normal / abnormal levels of blood lipids

Class	Triglycerides	Blood chol	HDL chol	LDL chol
NORMAL	680	539	565	436
ABNORMAL	32	173	147	276

Therefore we were very careful in building training sets in such a way that all classes of an outcome attribute were represented equally.

For a purpose of assessing decision trees constructed from the training set we used a test set which included all objects from initial database that were not used for training purposes.

Experiment No. 1

In our first experiment we were trying to predict the level of blood cholesterol, HDL cholesterol and LDL cholesterol in children. Therefore we defined all three attributes as an outcome with 8 different classes. We built training and testing sets and applied them to both tools.

We used different types of discretization for classic decision tree induction, the dynamic discretization on the whole data set gave better results, but both methods resulted in poor accuracy. The best decision tree in fact was genetically induced and it

classified testing objects with the total accuracy of 46%. The method of classic tree induction produced even worse results.

Experiment No. 2

Based on the poor results of the first experiment we concluded that bad results were a consequence of to many different classes in outcome so we combines some of the classes following the evaluation of physicians

From the medical point of view it is very important to compare the levels of LDL and HDL cholesterol because they together form complete information about the level of cholesterol in blood. For that reason we defined an output attribute "degree". We evaluated different combinations of blood cholesterol level, LDL and HDL levels with a degree form 1 to 4, where 1 represents the worst possible combination of cholesterol levels and 4 represents the best (see table 3) from the medical point of view

Table 3 Evaluation of comparison among cholesterol levels with a degree from 1 to 4, where 1 represents the worst state of child's blood lipids an 4 represents the best

Blood chol level	HDL chol level	LDL chol level	Degree	Number of objects
abnormal	abnormal	abnormal	1	129
abnormal	abnormal	normal	2	0
abnormal	normal	abnormal	3	
abnormal	normal	normal	3	
normal	normal	abnormal	3	485
normal	abnormal	normal	3	
normal	abnormal	abnormal	3	
normal	normal	normal	4	98

Once again we built a training set with 60 randomly chosen objects from each of the outcome classes. All other objects were used in testing set. The results were not much different than in previous experiments. The decision trees induced with both methods had highest total accuracies 50%.

Experiment No. 3

In our third experiment we reduced the number of classes in outcome attribute even more. We combined all three cholesterol levels (blood cholesterol level, HDL and LDL cholesterol level) in one outcome – "lipids". We defined "lipids" as discrete attribute with two possible values:

- lipids = "normal" if an object has normal all three cholesterol levels or
- lipids = "abnormal" otherwise.

We got approximately the same proportion of object with normal (349) and abnormal lipids (363) in the database.

Training set included 178 objects with normal lipids and the same number of objects with abnormal lipids. All other objects were included in a testing set. The best results are presented in table 4. We can see that all decision trees had very low accuracies.

Table 4 Results of the comparison of two methods applied on training and test sets for determining the level of lipids in blood

Method	Specificity	Sensitivity to abnormal lipids.	Total accuracy
Classic	75%/**53,8%**	73,5%/**53%**	73%/ **53,4%**
Genetic	73,5%/**50,8%**	70,7%/**53,5%**	72,2%/**52,3%**

Poor results stimulated us to add a new attribute to objects in the data set – Roher index (ROI), which is similar to body mass index (BMI), but often used with children because it correlates less with height than BMI but equally well with skinfold thickness. It is calculated as follows:

$$ROI = \frac{weight}{height^3} .$$

Children with ROI > 1,5 are obese and those with ROI < 1,1 are lean. On this bases we defined the following values for the new attribute ROI: obese, normal and lean. Then we again used both decision tree induction methods on new data sets (training and testing set).

Nonetheless the accuracy of induced decision trees was not much higher than without ROI (see table 5).

Table 5 Results of the comparison of two methods applied on training and test sets with added attribute ROI for determining the level of lipids in blood

Method	Specificity	Sensitivity to abnormal lipids	Total accuracy
Classic	73,3%/ **55%**	70,4% / **53%**	72%/ **54%**
Genetic	71,3%/**58,4%**	70,2%/ **51,8%**	70,8%/ **55%**

Our next attempt to improve results was distribution of the initial database in three data sets according to Roher index. First data set included objects with ROI=obese (175 objects), second data set included objects with ROI=normal (530 objects) and third with ROI=lean (7 objects). The third data set was too small for experimenting, so we used only the first two. This way we tried to induce two specialized decision trees for determining level of lipids: one for 'obese' children and one for 'normal' children on the basis of Roher index.

We tried both methods of decision tree induction separately for objects with ROI=obese and for objects with normal ROI.

The training set for objects with ROI=obese included 100 objects with the same proportion of object with normal and abnormal lipids. All other objects were included in a testing set (normal lipids: 21 objects, abnormal lipids: 54 objects). Once again the results were bad and the classic decision tree induction was just a little bit better than genetic algorithms (see table 6).

Table 6 Results of the comparison of two methods applied on data sets where only objects with ROI=obese are included

Method	Specificity	Sensitivity to abnormal lipids	Total accuracy
Classic	78%/ **57,1%**	75% / **53,7%**	76% / **54,7%**
Genetic	88%/ **52,3%**	78% / **51,8%**	83% / **52%**

Objects with ROI=normal were also divided into training and testing data sets. 200 objects were included in the training dataset with the same proportion of normal and abnormal level of lipids. The testing set included all other objects. In the table 7 we can see that there was no significant difference between the accuracies for objects with ROI=obese and objects with ROI=normal.

Table 7 Results of the comparison of two methods applied on data sets where only objects with ROI=normal are included

Method	Specificity	Sensitivity to abnormal lipids	Total accuracy
Classic	78%/**57,1%**	70% / **53,7%**	75% / **54,7%**
Genetic	69%/**52,5%**	80% / **52,5%**	74,5% / **52,4%**

At the end of the third experiment we concluded that combining all cholesterol levels in one output leads to bad results. Both methods of decision tree induction gave us similar accuracies.

Experiment No.4

In our last experiment we decided to limit our research. Therefore we restricted the outcome on blood cholesterol level only. We eliminated LDL and HDL cholesterol and also triglycerides from our database. In order to improve the power of classifiers (see table 2) we reduced the number of objects with normal blood cholesterol level in the training set so that both values were represented in approximately the same proportion (149 classified as 'normal' and 108 classified as 'abnormal'). All other objects were used for testing purposes.

The best decision tree classified test objects with 72,6% total accuracy, but the sensitivity to abnormal cholesterol level was only 37,5%. That shows that our decision tree specialized for classifying objects with normal cholesterol level (specificity 74,3%) and therefore it is not applicable for practical usage.

Consequently we tried to achieve better results with a hybrid method of genetically induced decision trees. The problem of classifying objects with abnormal cholesterol level was not solved either.

As in previous experiment we added Roher index as a new attribute in object's description. The most interesting results are shown in the table 8.

Table 8 Results of the comparison of two methods applied on data sets with new attribute (Roher index) added to the description of objects. Accuracies of induced decision trees are represented for training set / testing set

Method	Specificity	Sensitivity to abnormal chol.	Total accuracy
Classic	74,5%/**60,3%**	41,7%/**63,2%**	60,7%/**60,5%**
Genetic	77,1%/**47,3%**	75,9%/**52,6%**	76,7%/**47,6%**

The results were (as expected) better than before. Sensitivity to abnormal cases of blood cholesterol level was much higher in both methods, but genetic induction of decision trees was less accurate on the testing set than classic decision tree induction.

Since height and weight were already included in the calculation of Roher index, we presumed that they themselves might not be regarded as an influencing factor for blood cholesterol level determination. Therefore we excluded them from object's description. In the table 9 you can see best results of inducing decision trees using classic and genetic method.

Table 9 Results of the comparison of two methods applied on data sets with added attribute (Roher index) and attributes height and weight excluded from an object's description

Method	Specificity	Sensitivity to abnormal chol.	Total accuracy
Classic	91,9%/**62,8%**	74,1%/**52,6%**	84,5%/**62,3%**
Genetic	87,2%/**65,3%**	43,5%/**42,1%**	68,9%/**63,8%**

In comparison with previous results we can see that new decision trees were less accurate, and therefore we can establish that height and weight have some influence on blood cholesterol level.

Further we divided the database in two data sets on the basis of ROI. After examining both data sets we established that within obese children 75,5% had normal cholesterol level and 24,5% had abnormal cholesterol level. Among children with normal Roher index 79,6% had normal cholesterol level and 20,4% had abnormal.

Each data set was first divided into training and testing sets considering the ratio of objects with normal cholesterol level and object with abnormal cholesterol level in the training set (see table 10).

Table 10 The Number of objects in training and testing sets for two data sets: first with obese children and second with normal children according to Roher index

	ROI = obese		ROI = normal	
	Nor. chol	Abnor. chol	Nor. chol	Abnor. chol
Training set	55	40	130	110
Testing set	70	10	274	16

First we used the training set for obese children with both tools. The results are described in table 11 where you can see that genetically induced decision tree has higher accuracy than classic decision tree. The decision tree induced with genetic algorithms classified test objects with accuracy over 70%, which is high enough to be used for medical purposes.

Table 11 Results of the comparison of two methods applied on data sets where only objects with ROI=obese are included

Method	Specificity	Sensitivity to abnormal chol.	Total accuracy
Classic	84,6%/**61,6%**	85,7%/**71,4%**	85,1%/**62,5%**
Genetic	86,8%/**78%**	65,5%/**71,4%**	77,6%/**77,5%**

After encouraging results we were inquisitive if the results would change for the better in case of excluding weight and height from the list of attributes in object's description (see table 12). As it was established before the results are better when the height and weight are included in the induction of decision trees.

Table 12 Results of the comparison of two methods applied on data sets where only objects with ROI=obese are included and attributes weight and height are excluded

Method	Specificity	Sensitivity to abnormal chol.	Total accuracy
Classic	100%/**69,9%**	82,1%/**71,4%**	92,5%/**70%**
Genetic	86,8%/**72,6%**	68,9%/**71,4%**	79,1%/**72,6%**

Similar to experiments with the obese children data set we tried to obtain some good results with the data set that included only children with normal Roher index. You can see the results in tables 13 and 14.

Table 13 Results of the comparison of two methods applied on data sets where only objects with ROI=normal are included

Method	Specificity	Sensitivity to abnormal chol.	Total accuracy
Classic	94,5%/**61,6%**	70,5%/**71,4%**	84,6%/**62,5%**
Genetic	88,1%/**49,1%**	73%/**72,7%**	81,9%/**50,2%**

Surprisingly the decision trees for children with normal Roher index were less accurate than decision trees for obese children. We can also notice that accuracies on training sets were much higher than accuracies on testing set. Therefore those decision trees cannot be useful in practice.

The results of this experiment show that Roher index has an influence on determining blood cholesterol level. The decision tree induction on the dataset with obese children by Roher index was more successful than with children classified as normal according to Roher index. If we compare the methods used for decision tree induction we can see that genetic algorithms usually gave us decision trees with higher accuracy than classic decision tree induction.

Table 14 Results of the comparison of two methods applied on data sets where only objects with ROI=normal are included and attributes weight and height are excluded

Method	Specificity	Sensitivity to abnormal chol.	Total accuracy
Classic	91,8%/**62,9%**	73,1%/**54,5%**	84%/**62,5%**
Genetic	88,1%/**52,9%**	60,2%/**54,5%**	76,6%/**53%**

Discussion

After examining the database we expected to gain better results with the use of genetic algorithms because of their insensibility to missing and corrupted data.

First we tried to classify objects according to the level of blood cholesterol, HDL and LDL cholesterol, but there were too many classes for outcome attribute and therefore the results were poor. Reducing the number of classes in outcome attribute lead us to a little bit better results.

Higher results were gained when we reduced our classification on blood cholesterol level only. We divided the initial database to three data sets according to Roher index: first for obese children, second for normal children and third for lean children. The accuracy we achieved during our experiments on obese children was substantially higher than the accuracy achieved on children classified as normal according to Roher index. The only reason for that which arises at the moment is that with obese children the influencing factors for determining blood cholesterol level are clearer.

From the results gained with all four experiments we can conclude that the attributes in our database are not significant enough for determining cholesterol levels in children with non- invasive methods.

Experiments for determining cholesterol level with the use of machine learning have not yet been performed on the children so young. That is why this research is very interesting from medical point of view.

Conclusion

In this paper we compared two methods of machine learning (classic decision tree induction and decision tree induction with genetic algorithms) on the problem of children's cholesterol level determination. Presented results show that with the use of genetics the induction of decision trees was in most cases more successful than classic decision tree induction. We tried many different experiments in order to determine cholesterol levels in children, but the results were not very incentive. The best results were obtained in our last experiment where we tried to determine blood cholesterol level on the data set with only obese children included (according to Roher index). These results show that the obesity has an influence on the level of blood cholesterol.

To summarize we have reached the conclusion that attributes used in the database are not enough for determining cholesterol level in children. For that purpose we should obtain some other attributes that are more significantly linked with the cholesterol level.

In the future we will try some other new methods of machine learning such as rough sets on the present database and we are expecting some interesting results.

References

1. Quinlan JR: C4.5: Programs for machine learning. Morgan Kaufmann publishers, San Mateo, CA, 1993.
2. Zorman M, Hleb Š, Šprogar M: Advanced tool for building decision trees MtDecit 2.0. In: Kokol Peter (ed.), Welzer-Družovec Tatjana (ed.), Arabnia Hamid R (ed.). International conference on artificial intelligence, June 28 – July 1, 1999, Las Vegas, Nevada, USA. Las Vegas: CSREA, (1999), book. 1: 315-318.
3. Zorman M, Kokol P: Dynamic discretization of continuous attributes for building decision trees. In: Fyfe C. (ed.). Proceedings of the second ICSC symposium on engineering of intelligent systems, June 27-30, 2000, University of Paisley, Scotland, U.K.: EIS 2000. Wetaskiwin; Zürich: ICSC Academic Press, (2000): 252-257.
4. Baeck T: Evolutionary Algorithms in Theory and Practice, Oxford University Press, Inc., 1996.
5. Sprogar M, Kokol P, Zorman M, Podgorelec V, Yamamoto R, Masuda G, Sakamoto N: Supporting Medical Decisions with Vector Decision Trees In: V: PATEL, V. L. (ed.), ROGERS, R. (ed.), HAUX, R. (ed.). 10th World Congress on Medical Informatics MEDINFO, London, 2001. *MEDINFO 2001:* proceedings of the 10th World congress on medical informatics, London, UK, 2-5 September, 2001, Amsterdam: IOS Press: Ohmsha, (2001): 5.
6. Kokol P, Završnik J, Zorman M, Malčić I, Kancler K: Participative Design, Decision Trees, Automatic Learning and Medical Decision Making. In: Brender J. (ed.). Medical Informatics Europe '96, (Studies In Health Technology And Informatics, Vol.34). Amsterdam [Etc.]: Ios Press; Tokyo: Ohmsha, (1996): 501-505.
7. Kerry SJ., Brown CS, Hickey CM, McFarland LD, Weinhofer JJ, Gottlieb SH: Physical fitness, physical activity, and fatness in relation to blood pressure and lipids in pre-adolescent children: Results from the FRESH Study. Journal of Cardiopulmonary Rehabilitation, (1995): 122-129.
8. Podgorelec V, Kokol P: Induction of medical decision trees with genetic algorithms. In: Proceedings of the international ICSC congress on Computational intelligence methods and applications, June 22-25, 1999, Rochester, N.Y. USA. [s. l.]: ICSC - International Computer Science Conventions, (1999): 7.

Research Paper ■

Objectifying Researches on Traditional Chinese Pulse Diagnosis

Lisheng Xu, Kuanquan Wang, David Zhang*, Yingshun Li, Zhijie Wan**, Jing Wang****

Abstract. This paper describes our recent researches on objectifying of Traditional Chinese Pulse Diagnosis (TCPD) by means of some modern signal processing methods. In order to demystify TCPD and prove its efficiency, its significance, theory and features are briefed firstly. Secondly, a survey of recent developments in the researches of TCPD is provided. Thirdly, our researches on baseline removal, monitoring of the pulse and the feature extraction of the pulse are introduced. Furthermore, our pulse acquisition diagnosis system is presented. Finally, the prosperities and future works are also pointed out.

■ **Infor Med Slov** 2003; 8(1): 56-63

Author's institution: Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin, China.

*Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong.

**Hospital of Harbin Institute of Technology, Harbin Institute of Technology, Harbin, China.

Contact person: Lisheng Xu, Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin, China, email: lishengxu@hotmail.com.

Introduction

TCPD has been proven to be worthwhile and clinically valid over 5000 years of the Chinese medicine history recorded. However, due to the difficulty to master it, many people still take it as a mystery. Thus, it is extremely necessary to introduce TCPD and let more and more people understand it. Many kinds of apparatus and systems that can automatically detect pulse from patients demonstrate that the researches of TCPD are significant and successful, but the modern research of TCPD has slowed down for a long time due to pulse's complexity and variation. [1] Nevertheless, the development in medical, sensor, pattern recognition, signal processing, database and other relative fields accelerate the research of TCPD forward recently.

This paper aims to employ some modern feature extraction and signal processing technologies to the objectifying researches on TCPD and point out its brighter future. First, the background, significance and the features of TCPD is stated. Then, an overview of recent achievements of TCPD is presented and our research on TCPD is introduced. Finally, we point out future tasks, emphases and restrictions of modern research on TCPD. For the clarity of understanding, some Chinese explanations corresponding to the English terms of TCPD are given in the round brackets together.

Traditional Chinese Pulse Diagnosis

TCPD, one of the four diagnostic methods of TCM, is to judge disease by means of fingertips palpating patient' pulse image shown in the superficial arteries. Many western people may consider that pulse waveform is just the same as electrocardiogram (ECG) and the patient's ECG analysis is enough. The signal of ECG acquired through several electrodes only reflects the bioelectrical information of body. Having analyzed the pressure fluctuation signal of pulse, doctors can detect and predict some symptoms that ECG cannot. TCPD can not only deduce the positions and degree of pathological changes, but also is a convenient, inexpensive, painless, bloodless, noninvasive and non-side effect method promoteded by U.N. [2]

The substance of pulse is the blood and the power of pulse is the heart. The heart pumps and blood into all parts of the body through vessels and then the blood enter viscera inward and reach limbs & skin outwards incessantly. Besides, the blood circulation also depends on other viscera, which coordinates the heart. The lung meets all vessels and the blood circulation all over the body should converge into the lung; the liver stores blood and is in charge of its conducting; the kidney stores essence. Thus through the vessels, all visceral state and disease condition can be understood by means of pulse diagnosis. [3,4] Pulse diagnosis is to palpate pulses with fingertips and then to understand and judge the disease condition through the process of diagnostician's comprehension. It also named pulse-palpating, pulse-feeling, pulse-touching, pulse-reading, pulse examination or pulse taking. Pulse taking is the common word. To sum up the ancient Chinese Medicine, the significances of TCPD research today are as follows:

1. The physical examinations for the people of special careers such as students, pilots, athletes and some others, especially for the workers in chemical plants;

2. The researches of drug' s functions and effects on blood vessels & heart;

3. The monitoring of patients, pregnant women, fetus and so on;

4. The important reference for the doctors to recognize the exterior and interior of disease, to judge the deficiency and excess, to ascertain nature of disease, to identify the cause of disease, to predict the prognosis and to inspect the disease mechanism;

5. The medical education and training for medicos;

6. The researches on the circulation system, nerve system, body fluid regulation, the emotions and so on; [5]

7. The researches on fitness and exercise (checking the effects and revising the exercise plan);

8. The surveying of psychology and the detecting of liar; [6-9]

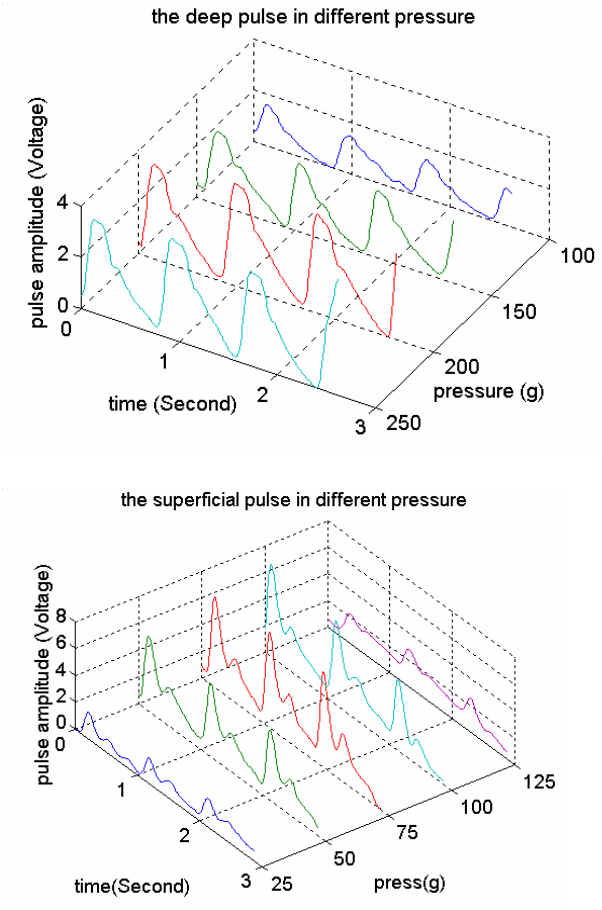


Figure 1 (a) Deep pulse (Cheng Mai) images, (b) Superficial pulse images

Since ancient times, doctors have been paying great attention to pulse taking and have accumulated rich experiences. Taking pulse in TCPD, involving counting the number of beats and identifying its form & pattern, does not just mean the identifying of pulse waveform as the researcher of modern medicine did. [10] It should be borne in mind that each of the types doesn't represent just one aspect of a given pulse. For example, floating and sinking describe the depth; slow and rapid describe the rate, whereas surging and fine describe the size of the pulse. Actually, these parameters occur in combinations. In most cases, a patient's pulse is described with a composite term such as floating, slippery, and rapid, or sinking, wiry, and thin.

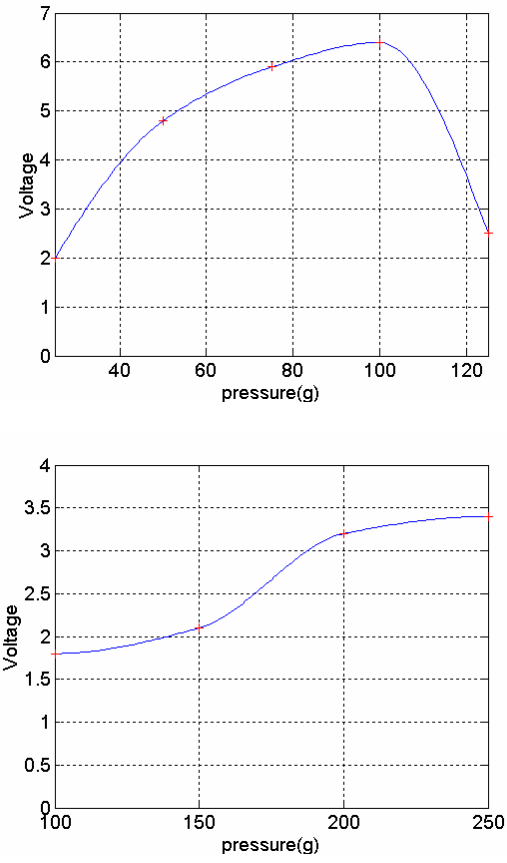


Figure 2 (a) Trend of superficial pulse images, (b) Trend of deep pulse images

According to the theory of TCPD, we use different pressure to acquire the pulse image and then judge the pulse whether floating or sinking, whether excess or deficiency and so on. Pulse shape varies with pressures. When the pulse waveform amplitude is the highest among those pulse

waveforms, it is named optimal pulse waveform. If the pressure acquired the optimal pulse waveform is smaller than 100g, this kind of pulse must be superficial pulse. When the pressure acquired the optimal pulse waveform is more than 200g, this kind of pulse must be deep pulse. Normal pulse's pressure acquired the optimal pulse waveform is smaller than 200g and more than 100g. Figure 1(a) and (b) are the superficial pulse and deep pulse we acquired. Their trends are illustrated in Figure 2. The deep pulse, defined only by its deep position, is often described as deficient on light pressure and excess by heavy pressure. Only when the pressure is more than 100g, the deep pulse can be felt. The best shape of the pulse is at the pressure of 200g or so. When the pressure is bigger than 250g, the pulse shape is still clear. The superficial pulse, defined only by its superficial position, is often described as excess on light pressure and deficient by heavy pressure. When the pressure is 25 g or so, the superficial pulse can be felt with ease, but when the pressure is bigger than 125g the pulse is not so clear. The best shape of the superficial pulse is at the pressure of 75g or so.

Thus, the researches on TCPD are more than the studies on pulse waveform. It just means the multi-dimension information. According to TCPD, we name those pulse waveforms as Pulse Image. What's more, new disease and new problems associated with our modern civilization have begun to show consistencies in TCPD. For example, the "ceiling dripping" scattered pulse of AIDS and a kind of knotted pulse related to cancer are among the few recently identified syndromes which seem to have characteristic pulse images. [11]

Developments of Researches on TCPD

To reveal the scientific essence of pulse diagnosis, a lot of researches have been made in the fields of TCM, western medicine, medical engineering and their related fields from 1950's. But some of them did not base on the theory of TCPD. Beside the researchers in China, some researchers in Japan,[12] Korean,[13,14] German, [15] Canada and US got interested in this research of TCPD. [16] In order to objectify pulse, engineers have designed many kinds of pulse sensors to acquire pulse. Of all these kinds of pulse sensors, the pressure sensors can reflect the information just as pulse feeling based on TCPD better. The HMX pulse sensor made by Shanghai Medical Instrument Company has better reproducibility in operation. [17] According to the theory of elastic cavity, McDonald, [18] Liu Zhaorong studied the circulation system. [19] But the cardiovascular system is so complicated that it cannot be modeled accurately. It is meaningful but it still needs the further systematic research.

Our Researches on TCPD

At present, the parameters' extraction is mainly carried out by time-domain signal processing method such as computing the amplitude, slope, area and so on. Among these parameters, the ratio between some of them also can be used to justify vascular elasticity and peripheral resistance. Due to some limits of related fields, the researches in TCPD make less progress. With the application of modern signal processing methods and technologies, some biometrics technologies such as speech recognition and signature recognition have made rapid progress. Thus, the research of TCPD should combine with the modern signal processing too.

In this section, all the pulse data are acquired by our pulse diagnosis system, which comprises a set of pulse sensor, adapter, amplifier, and computer. The sensor, named HMX-4, was made by Shanghai Medical Instrument Company. It is a hyperbolic contact-terminal type of the strain cantilever beam transducer, which is not the same as the previous sensors for studying the western medicine. Our sensor's probe is a trapezoid whose area is 29.4 mm^2, that makes the probe's little deviation do not influence its repeatability. Thus the impersonal, stable, high-precision pulse waveform is ensured. The following is our works

on baseline drift removal, monitoring and features extraction of pulse.

Baseline Drift Removal of Pulse Waveform

Pulse waveform can easily be influenced by many factors such as respiration, body temperature, muscle's dithering, body's movement and so on. The whole pulse goes down when exhaling and goes up when inhaling. Holding the breath may make pulse more stable. But these restricts not only make the patient uncomfortable and inconvenient, but also prevent us from acquiring the long period of stable pulse. Thus, we developed an algorithm for baseline removal.[20] The pulse, with its baseline being adjusted, is ***signal3*** in Figure 3, and its original pulse curve is ***signal1***. ***Signal2*** is the baseline drift.

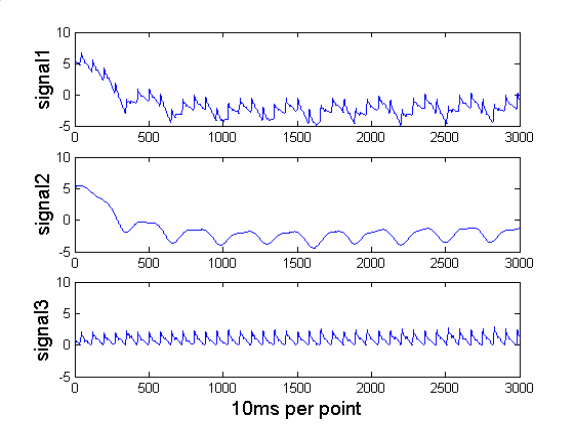


Figure 3 Actual pulse and its results filtered by wavelet

Monitoring of Pulse

TCPD has been researched worldwide. Some success has achieved. The means of acquiring the pulse information and the performance of pulse sensor are satisfied. But the research of pulse's monitor is reported seldom because of baseline drift and noise interferences. Having combined some modern signal processing technology, we extracted the baseline drift and noise interference. Thus, the monitoring of pulse can be realized. This does have the pathological and physiology meaning. Figure4 illustrates a period of the monitored pulse data. The signal in the upper is the contaminated pulse; the signal in the lower panel is the baseline being extracted; the signal in the middle is the real pulse.

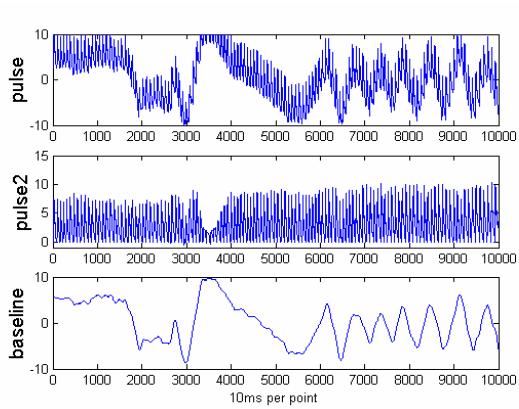


Figure 4 The monitoring of pulse

During the process of monitoring, we can study the pulse rate's variation too. As far as pulse' rhythm concerned, it often changes even to a healthy person. Figure5 shows us this phenomenon. The period's mean value is 0.8 second and its standard deviation is 0.0667. The knotted pulse **(Jie Mai)**, scattered pulse **(San Mai)** and intermittent pulse **(Dai Mai)**, all have their distinctive characters of the rhythm respectively.

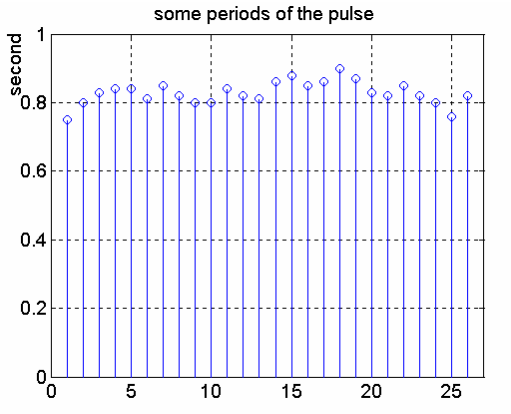


Figure 5 The pulse's period fluctuation

Table 1 Comparison of three different pulse's RPA

RPA \ pulse name	Taut pulse	Normal pulse	Smooth pulse
RPA (Ratio of pulse area) Maximum	0.65	0.40	0.35
RPA (Ratio of pulse area) Minimum	0.40	0.35	0.25

Feature Extraction

During the time domain analysis of the pulse, we find that the ratio of pulse area is a feature for differentiating the typical pulses. As Figure6 shown, the pulse area of every period (S_{pulse}) and its rectangle area (S_{rect}) can be calculated. The rectangle area (S_{rect}) equals to the value of pulse's main peak multiple its period. The ratio of pulse area (RPA) to the rectangle area is defined as follows.

$$RPA = S_{pulse} / S_{rect} \qquad (1)$$

Illustrated in Table1, the RPA of taut pulse ***(Xian Mai)*** is bigger than 0.4, while the RPA of normal pulse ***(Ping Mai)*** is bigger than 0.35 and less than 0.4, the RPA of smooth pulse ***(Hua Mai)*** is more than 0.25 and less than 0.35.

About the extraction of pulse features, this article promotes two kinds of area grade analysis methods, namely X-axis area analysis and Y-axis area analysis. Applying the X-axis area analysis method, the systolic area and diastolic area and some related parameters are calculated. What's more, by means of the Y-axis area analysis method, the main peak's width and the variation of pulse waveform shape characters can be got. As Figure7 illustrated, we gets the main peak's value P_m at first, then draw a line y=Pm. Then draw the equispaced lines parallel with the X-axis such as y=0.99* P_m, y=0.98* P_m,…, y=0.02* P_m and y=0.01* P_m. Next, the pulse waveform intersects with these lines and the areas of these intersects can be calculated. According to these areas trend, we can classify the various pulse images. If we combine these two kinds of area analysis method, the classification will be more satisfied.

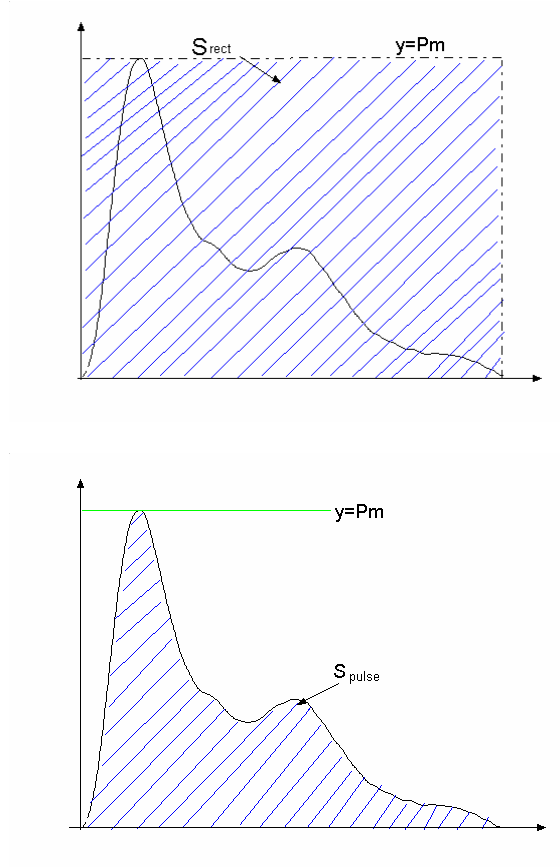


Figure 6 The scheme of areas computing, (a) The calculation of Srect, (b) The calculation of Spulse

Ling Y Wei found that the energy rate of pulse power spectrum did have some relation with the disease. This illustrates that the frequency analysis of pulse image is significant. From the power spectrum analysis, we can find that the ratio of the spectral peaks is very important in analyzing people's physical condition. Table 2 lists some of the comparisons. A_0, A_1, A_2, A_3, A_4,

A_5, ……, stands for the direct current component, first harmonic, second harmonic, third harmonic, fourth harmonic, fifth harmonic of the pulse and so on respectively. Applying method of power spectrum analysis, we can also analyze slow pulse ***(Man Mai)***, rapid pulse ***(Kuai Mai)***, moderate pulse ***(Huan Mai)***, scatter pulse ***(San Mai)***, knotted pulse ***(Jie Mai)***, running pulse ***(Cu Mai)***, intermittent pulse ***(Dai Mai)*** and so on.[21-25]

The choice on the method of the pulse's analysis is significant. Due to the variability of pulse mentioned above, some statistical approaches need to be used. Statistics alone do not help all the time, however. There is also a need for some signal processing algorithms, which are robust to this variability. Although several modern signal-processing algorithms have been developed for the research of pulse-taking, some technologies such as wavelet, STFT (Short Time Fourier Transform), Higher order spectrum, AR-spectrum array, neural network and so on were successfully applied in the research of heart's sound can be applied to study the TCPD too.[26,27] The methods of speech processing also can be used for reference. Based on our ever-growing database of pulse, our lab is on the way to improve the pulse image's efficiency of signal processing and recognition.

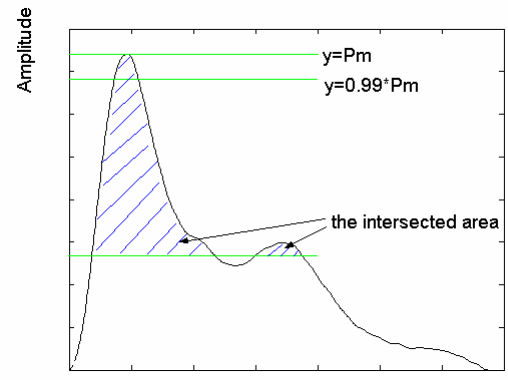


Figure 7 Schematic figure of Y-area analysis method

Table 2 Comparison of three different pulse's harmonics

Harmonic Amplitude	A0	A1	A2	A3	A 4	A5	A6
Taut pulse	0.47±0.05	0.30±0.05	0.12±0.02	0.04±0.01	0.02±0.01	0.02±0.01	0.02±0.01
Normal pulse	0.34±0.05	0.29±0.05	0.15±0.02	0.09±0.02	0.06±0.01	0.04±0.01	0.03±0.01
Smooth pulse	0.46±0.05	0.28±0.05	0.18±0.02	0.05±0.01	0.02±0.01	0.01±0.01	0.0

Conclusion

For the purpose of probing the mechanism of manifestations of the pulse of Traditional Chinese Medicine (TCM), this article has made lots of researches on pulse image by using signal processing methods. What's more, the monitoring of pulse is researched for the first time. In time domain, a brand-new area analysis method is proposed. In frequency domain, harmonic features are also extracted. With these comments, we end our discussion of TCPD by stating some of its developing directions.

1. Unifying the instrument for acquiring pulse pressure, the analysis methods and the normalization of TCPD;
2. Combining the integral and dynamic researches on TCPD with clinic;
3. Applying some modern signal processing & technology and looking for the new breakthrough of TCPD;
4. Combining with some other diagnosis methods such as tongue diagnosis, ECG, EEG and heart sound.

Acknowledgements

The research was partly supported by the Multidiscipline Scientific Research Foundation of Harbin Institute of Technology (HIT.MD2001.36).

References

1. Maruya, Tokinori, "Apparatus and system for pulse diagnosis and the method", *European patent 97307209.3*, 1998.
2. Feng Yuanzhen, *Chinese Journal of biomedical engineering*, 1(1), 1983.
3. Li Shizhen, *The lakeside master's study of the pulse*, Blue Poppu Press.
4. Li Shizhen, *Pulse diagnosis*, Paradigm publication, Syndey, Australia, 1985.
5. Fei Zhaofu, Takehito Fujino, "Relation between radial sphygmogram and autonomic nerve function", *Research and application on TCME*, pp.337, 1992.
6. Shou Xiaoyun, *Clinic and pulse diagnosis research of Shou Xiaoyun,* Chinese medicine press, 1998.
7. Shou Xiaoyun, "Clinical recognition of psychological pulse image", *Journal of Beijing University of TCM*, Vol.20, No.3, 1997.
8. http://www.catacheater.com/HandyTruster.htm
9. http://science.163.com/tm/000929/000929_37312.html.
10. Stockman, G. Kanel, L.N., Kyle, M.C., "Structural pattern recognition of carotid pulse waves using a general waveform parsing system", *Commun. ACM*, pp.688, 19(12), 1976.
11. Hammer, "Contemporary pulse diagnosis: introduction to an evolving method for learning an ancient art-part one", *American Journal of Acupuncture*, Vol.21, No. 2, 1993.
12. Seng He et al, "Objectifying of pulse-taking", *Journal of Japanese Eastern Medicine Society*, 27(4):7, 1977.
13. Paik Hee Soo, Hee Soo "Type electronic pulse diagnosis", *Abstract of papers of the 7-th world acupuncture conference*, Sriklanka, 1981.
14. Young-Zoon Yoon, Myeong-Hua Lee, Kwang-Sup Soh,Michael, "Pulse type classification by varying contact pressure", *IEEE engineering in medicine and biology,* pp.106-110, 2000.
15. Von. D. J. Yoo, "Elektronishe Puisographie-eine hene diegno-Stishe Moghichkeit auf chrobobiorhythmischer", *Grundlage Akupunktur Theorie und Paris*, Vol.3, pp. 90, 1980.
16. Broffman, Michael MeCulloch, "Instrument-Assisted pulse evaluation in the Acupuncture practice", *American Journal of Acupuncture*, Vol. 14, No. 3, pp.255-259, 1986.
17. Li Jingtang, "The possibility of objectifying and detection traditional Chinese pulse image and the design of its figures", *International conference on Traditional Chinese Medical*, Shanghai, China, 1987.
18. McDonald D.A., *Blood Flow in Arteries*, Oxford University Press, 1978.
19. Liu Zhaorong, Li Xixi, "The variation of the pressure waves of the radial artery with some physiological parameters", *Journal of Mechanics*, pp.244-250, 1982.
20. Xu Lisheng, Wang Kuanquan, David Zhang, "Adaptive Baseline Wander Removal in the Pulse Waveform", *IEEE proceeding of CBMS2002 international conference*, 2002.
21. Liu Guanjun, *Chinese Pulse Diagnosis*, Chinese Medicine and Drugs Press, 2002.
22. Hong Zhipin, Zhang Shixu, "The spectral characters of taut pulse and smooth pulse", *Liaoning Journal of TCM*, pp.147-148, Vol.22, No.4, 1995.
23. Zhang Jinren, Yang Tianquan, "Spectral Analysis of healthy people's pulse", *Liaoning Journal of TCM*, pp.435-436, Vol.22, No.10, 1995.
24. Wang WK, Hsu TL, Chang Y, et al, "Study on Pulse Spectrum change before deep sleep and its possible relation to EEG", *Chinese J Med Eng*, pp.107, 1992.
25. Ling Y Wei, "Frequency distribution of human pulse spectra", *IEEE Trans. Biomed Eng*, Vol.32, pp.245, 1985.
26. Wu Yanjun, Xu Jingping, Zhao Yan, "A study of using three time-frequency methods in the analysis of heart sound signals", *Journal of Chinese Medical Instruments*, Vol.20, No.1 , 1996.
27. Hu Jiangning, Yan Shuchi, Wang Xiuzhang, et al. "An Intelligent traditional Chinese Medicine pulse analysis system model based on artificial neural network", *Journal of Chinese Medical University,* pp.134-137, Vol.26, No.2, 1997.

Strokovno-znanstveni prispevek ■

Uporaba računalniških inteligentnih sistemov v kardiologiji

Milojka Molan Štiglic, Peter Kokol

Izvleček. V članku opisujemo inteligentne sisteme in vlogo zdravnika pri njihovi uporabi. Podajamo uspešno aplikacijo iz področja otroške kardiologije, katere rezultat je odkritje novega medicinsko znanja.

The Application of Intelligent Systems in Cardiology

Abstract. The role of the physician in developing and using intelligent systems is presented in this paper. A success story resulting in new medical knowledge is briefly described.

■ **Infor Med Slov** 2003; 8(1): 64-68

Institucija avtorja: Splošna bolnišnica Maribor.

Kontaktna oseba: Milojka Molan Štiglic, Splošna bolnišnica Maribor, Ljubljanska 5, 2000 Maribor. email: molan.stiglic@sb-mb.si.

Uvod

V zadnjih letih človekova možnost za shranjevanje podatkov vse bolj presega njegove možnosti za njihovo analizo. Pričenjamo govoriti o t.i. podatkovnih grobnicah – podatke shranjujemo, in jih nato pustimo, da počivajo v miru. Vendarle pa ti podatki skrivajo mnoga neodkrita, dragocena znanja, in vse bolj se zavedemo, da je potrebno razviti nove avtomatske tehnike za njihovo odkrivanje.

V članku bomo na kratko opisali inteligentne sisteme in njihovo uporabo pri analizi medicinskih podatkov. Podali bomo primer iz kardiologije in razmišljene o koristnosti izkopavanja podatkov z medicinskega stališča.

Inteligentni sistemi

Besedica inteligenca je že dolgo srž mnogih diskusij in raziskav. Žal še dandanes nimamo enotne definicije, zato je toliko težje definirati kaj je računalniška inteligenca. Ena izmed najbolj uporabljanih definicij je naslednja:

Niz (mentalnih akcij) je inteligenten, če doseže »nekaj«, kar bi imenovali inteligentno, če bi to »nekaj« dosegel človek

Znameniti matematik in eden pionirjev računalništva je kot definicijo inteligence podal naslednji test:

V neki sobi je skrit pameten stroj ali človek. Izpraševalec ne ve kaj je v sobi, in tej entiteti postavlja vprašanja. Če iz odgovorov ne moremo ugotoviti, ali se pogovarja s človekom ali strojem, in če je v sobi stroj, potem je ta stroj inteligenten.

Bolj laično lahko inteligentne sisteme opišemo z naslednjo postavko:

Podobno kot mehanski stroji povečajo naše fizične sposobnosti, optični instrumenti (npr. mikroskop) naše čutne sposobnosti inteligentni sistemi povečujejo / podpirajo naše intelektualne sposobnosti.

Inteligentni sistemi morajo posedovati naslednje lastnosti:

- izražajo adaptivno ciljno usmerjeno obnašanje;
- se učijo iz izkušenj;
- uporabljajo velike »količine« znanja;
- izražajo samozavedanje;
- komunicirajo z ljudmi z uporabo jezika in govora;
- tolerirajo napake pri komunikacijah;
- odgovarjajo v realnem času.

Podajmo še nekaj razlik.

Ekspertni sistem : inteligentni sistem

- pri ekspertnem sistemu znanje dobimo v obliki pravil od ekspertov na področju
- pri inteligentnem sistemu znanje dobimo s pomočjo strojnega učenja na podlagi rešenih primerov

Statistika : inteligentna analiza

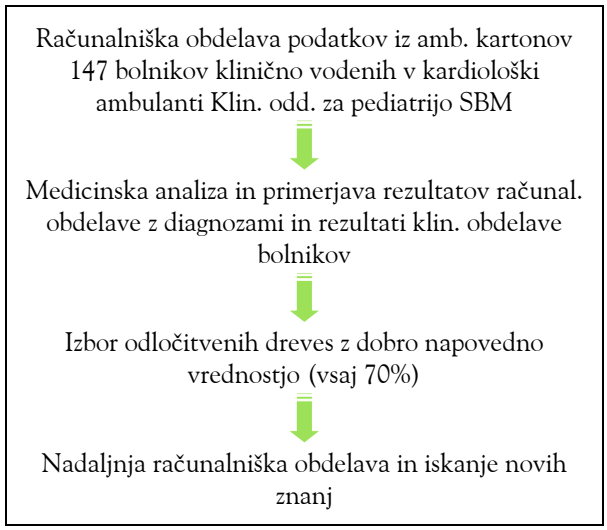
- pri statistiki potrjujemo znane relacije med podatki
- pri inteligentni analizi inteligentni sistem sam išče neznane relacije med podatki (temu procesu rečemo tudi podatkovno rudarjenje ali izkopavanje)

Odločitvena drevesa

Odločitvena drevesa so tipičen predstavnik strojnega učenja, kjer so učni primeri predstavljeni kot par (*lastnosti*, *odločitev*). Lastnosti so opisane kot zbirka oziroma vektor več atributov, ki naj bi na najboljši možen način predstavljale posamezni objekt. Izbira atributov je odvisna od snovalcev množice učnih primerov, od okoliščin in od

zmožnosti opravljanja meritev. Odločitev je tista lastnost, ki je znana pri objektih v učni množici, ne pa tudi pri objektih, o katerih bomo kasneje s pomočjo odločitvenega drevesa sprejemali odločitve. Običajno gre pri odločitvi za lastnost, ki se ne da izmeriti (npr. nek dogodek, ki se bo zgodil v prihodnosti) oziroma je njena meritev povezana z velikimi stroški, časovno zahtevnostjo ali zahtevami po zapletenih postopkih. Z uporabo odločitvenih dreves lahko zato poizkušamo:

- napovedati dogodek v prihodnosti,
- poiskati alternativne možnosti za dosego cilja, ki bodo skrajšale čas, zmanjšale stroške ali celo omogočile doseganje želenih rezultatov.

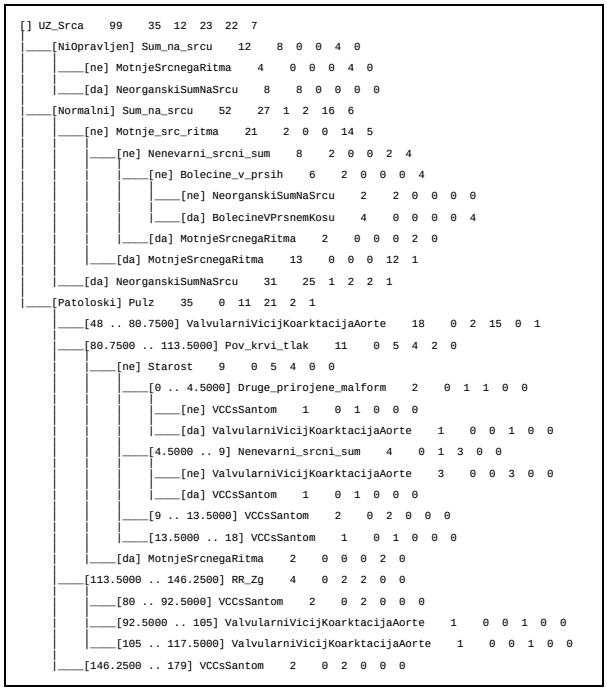


Slika 1 Metode dela

Mesto zdravnika pri izgradnji in uporabi inteligentnih računalniških sistemov v medicini?

Zdravnik in ostalo zdravstveno osebje na inteligentne sisteme ponavadi gledajo z določeno mero dvoma in nezaupanje. Vendarle lahko zdravnika umestimo na naslednji način:

- Zdravnik ocenjuje skladnost (pravilnost) oz. neskladnost računalniške odločitve z dognanji medicinske znanosti
- Pravilnim odločitvam daje medicinsko razlago na temelju klasičnega medicinskega znanja
- Išče najbolj racionalne algoritme diagnostičnih postopkov oz. potrebne baze podatkov za nadaljnje raziskave s pomočjo modelov umetne inteligence
- Išče nova znanja oz. zaenkrat še nepotrjene povezave, ki so lahko osnova novih znanstvenih raziskav



Slika 2 Primer odločitvenega drevesa

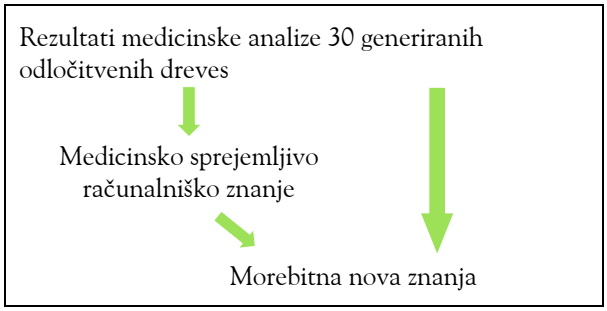
Primer študije v SBM

Metoda dela je podana na sliki 1.

Preiskovance smo razdelili v naslednje skupine:

- neorganski šum na srcu
- prirojene srčne napake z L-D šantom
- bolezni srčnih zaklopk
- motnje srčnega ritma
- bolečine v prsnem košu

in generirali 30 odločitvenih dreves, eno izmed zanimivejših je podano na sliki 2. Njihovo analizo shematično podajamo na sliki 3, rezultate analize v Tabeli 1 in 2. in drevo, ki vsebuje novo znanje na sliki 4.



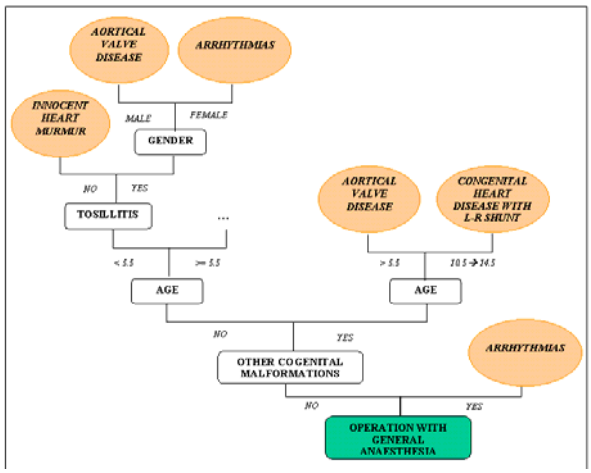
Slika 3 Shematski prikaz analize generiranih odločitvenih dreves

Tabela 1 Odkrito znanje

Medicinsko sprejemljivo računalniško znanja	Morebitna nova znanja
Povezava neorganskega šuma na srcu in tahikardije ob vročinskih stanjih Povezava težke okvare aortne zaklopke in sinkope oz. kolapsa ob naporu	Povezava oz. večja pogostost srčnih aritmij pri otrocih, ki so bili operirani v splošni anesteziji
Povezava večje verjetnosti za spremembe v EKG-ju, okvare zaklopk ali aritmije po številnih prebolelih anginah Povezava priroj. srč. napak in drugih priroj. malformacij	Povezava okvar aortne zaklopke in srčnih aritmij

Tabela 2 Zanimivo znanje

Medicinsko že znano dejstvo	Morebitno novo znanje
Možnost pojava aritmij med samim aktom operacije (vpliv anestetikov, sprememb AB ravnotežja, oksigenacije)	Nagnjenje k srčnim aritmijam še mesece po operaciji v spl. anesteziji!



Slika 4 Drevo z novim znanjem

Zaključek

Kot zanimivost navajamo, da smo gornjo metodo in rezultate pomladi predstavili na kongresu pediatrične kardiologije v Portu, in kar na istem kongresu dobili potrditev našega dognanja v »Lecture of the art« predstavitvi Prof. Marie C. Seghaye.

Literatura

1. Seghaye MC et al: Lecture of the Art: Impact of the inflamatory reaction on organ disfunction after susrgery. *Cardiology in the Young*. Association for European Paediatric Cardiology, XXXVII Annual General Meeting, Porto 2002.
2. Kokol P, Zorman M, Molan Štiglic M: Intelligent system for cardiac diseases decision making in the young. *Cardiol. young*, pp. 47.
3. Zorman M, Hleb Š, Šprogar M: Advanced tool for building decision trees MtDecit 2.0. In: Kokol P (ed.), Welzer-Družovec T (ed.), Arabnia Hamid R (ed.). International conference on artificial intelligence, June 28 – July 1, 1999, Las Vegas, Nevada, USA. Las Vegas: CSREA, (1999), book. 1: 315-318.
4. Šprogar M, Kokol P, Zorman M, Podgorelec V, Yamamoto R, Masuda G, Sakamoto N: Supporting Medical Decisions with Vector Decision Trees In: V: Patel, V. L. (ur.), Rogers, R. (ur.), Haux, R. (ur.). 10th World Congress on Medical Informatics MEDINFO, London, 2001. MEDINFO 2001: proceedings of the 10th World

congress on medical informatics, London, UK, 2-5 September, 2001, Amsterdam: IOS Press: Ohmsha, (2001): 5.
5. Mitchell T: *Machine Learning*, Addison Wesley, MA, 1997.
6. Zorman M, Kokol P: Dynamic discretization of continuous attributes for building decision trees. In: Fyfe C. (ed.). Proceedings of the second ICSC symposium on engineering of intelligent systems, June 27-30, 2000, University of Paisley, Scotland, U.K.: EIS 2000. Wetaskiwin; Zürich: ICSC Academic Press, 2000: 252-257.
7. Baeck T: *Evolutionary Algorithms in Theory and Practice*, Oxford University Press, Inc., 1996.
8. Kokol P, Završnik J, Zorman M, Malčić I, Kancler Kurt: Participative Design, Decision Trees, Automatic Learning and Medical Decision Making. In: Brender J. (ed.). Medical Informatics Europe '96, (Studies In Health Technology And Informatics, Vol.34). Amsterdam [Etc.]: Ios Press; Tokyo: Ohmsha, (1996): 501-505.

Znanstveno-raziskovalni prispevek ■

Razvrščanje profilov izražanja genov z metodami strojnega učenja

Tomaž Curk, Blaž Zupan, Gaj Vidmar

Izvleček. Nedavno razvita tehnologija mikromrež DNK omogoča opazovanje časovne aktivnosti (profilov izražanja) večjega števila genov, kar nam lahko pomaga pri določanju funkcije genov. V primeru, da za določen nabor genov njihovo funkcijo že poznamo, lahko iz podatkov gradimo modele, ki napovejo funkcijo genov na podlagi njihovih časovnih aktivnosti. Na dveh zbirkah podatkov smo pokazali, da so metode strojnega učenja primerne za indukcijo napovednih modelov funkcij genov. Navkljub splošnemu prepričanju na področju bioinformatike, da je za obravnavo tovrstnih podatkov najprimernejša metoda podpornih vektorjev, smo pokazali, da dosti bolj preprosta in časovno veliko bolj učinkovita metoda naivnega Bayesa dosega podobne oziroma celo boljše rezultate. Razvili smo tudi novo metodo kvalitativnega modeliranja profilov izražanja genov, ki se je v napovedih izkazala za manj točno, lahko pa uspešno služi za vizualizacijo aktivnosti genov v času.

Classification of gene expression profiles with machine learning

Abstract. The recently developed DNA microarray technology provides a way to measure expression profiles of a large number of genes and assign functions to genes. Given prior knowledge on gene functions and the microarray data, one can build models that predict functions of genes based on their expression profiles. We demonstrate on two genetic data sets that machine learning methods are suitable for induction of such prediction models. Surprisingly, naive Bayesian method proved at least as accurate but much faster than the currently prevailing support vector machines. We also present a new method for qualitative modelling of gene expression profiles, which makes less accurate predictions but it may be very useful for visualization of gene expression profiles.

■ **Infor Med Slov** 2003; 8(1): 69-80

Instituciji avtorjev: Fakulteta za računalništvo in informatiko, Univerza v Ljubljani (TC, BZ), Department of Human and Molecular Genetics, Baylor College of Medicine (BZ), Inštitut za biomedicinsko informatiko, Medicinska fakulteta, Univerza v Ljubljani (GV).

Kontaktna oseba: Tomaž Curk, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Tržaška 25, 1000 Ljubljana. email: tomaz.curk@fri.uni-lj.si.

Uvod

V zadnjih nekaj letih je na področju molekularne biologije in genetike prišlo do tehnološkega preobrata s pojavom in razvojem tehnologije mikromrež DNK, ki omogoča merjenje nivojev aktivnosti oziroma izražanja več deset tisoč genov hkrati. Analiza tovrstnih podatkov je eden izmed pristopov, ki ga genetiki na področju funkcijske genomike (ang. *functional genomics*) lahko uporabljajo za sklepanje o funkciji genov. Vedenje, kdaj in kje je nek gen izražen, nam namreč lahko olajša nalogo določanja njegove funkcije.[1]

Profil izražanja gena je časovno zaporedje nivoja aktivnosti gena in pove, kako se aktivnost gena spreminja s časom. Profile izražanja genov dobimo tako, da v nekem biološkem eksperimentu opravimo več merjenj izražanja genov ob različnih časih. Slika 1 prikazuje izraze triindvajsetih genov amebe *D. discoideum* v trinajstih točkah časa (razmik med dvema točkama je dve uri).

Velike količine tako zbranih eksperimentalnih podatkov je možno in smotrno obdelati le z uporabo računalnikov in specializiranih metod. Od slednjih so za to področje še posebej zanimive metode za odkrivanje znanja iz podatkov. Med njimi glede na preliminarne študije največ obetajo metode za strojno učenje, ki pa jih je zaradi specifičnosti področja potrebno ustrezno prilagoditi. Strojno učenje se torej ponuja kot ena od tehnik obdelave tovrstnih podatkov in razvoja napovednih modelov, njegova uporaba na tem področju pa je še v povojih. Potrebno je še podrobno preučiti uporabnost posameznih metod strojnega učenja in razviti specializirane metode za reševanje problema določanja funkcije gena iz njegovega profila izražanja.

V pričujočem prispevku smo preučili uporabnost različnih metod strojnega učenja za gradnjo modelov, ki lahko določen gen na osnovi njegovega profila izražanja razvrstijo v neko funkcijsko skupino. Med sabo smo primerjali različne metode strojnega učenja in preučevali njihovo uspešnost na dveh različnih naborih podatkov. Navkljub splošnemu prepričanju, da je za reševanje tovrstnih problemov najbolj primerna metoda podpornih vektorjev, smo pokazali, da preprostejša in časovno veliko bolj učinkovita metoda naivnega Bayesa daje vsaj enako dobre rezultate. V prispevku predstavljamo tudi novo metodo, ki temelji na kvalitativni obravnavi profilov izražanja genov; ta sicer daje slabše rezultate z vidika napovedi, a je zanimiva in uporabna pri iskanju trendov genskih izrazov in vizualizaciji.

Uporabljeni genetski podatki

Za študijo smo uporabili dve zbirki podatkov, ki opisujeta izražanje genov v različnih poskusnih pogojih. Uporabili smo podatke o kvasovki *S. cerevisiae*, ki so jih pripravili Brown in soavtorji,[2,3] in podatke o amebi *D. discoideum*, ki jih je zbrala skupina Gad Shaulsky-ja iz Baylor College of Medicine v Houstonu, Texas. Nivo izražanja gena je navadno definiran kot logaritem razmerja med nivojem izražanja gena v testnem (poskusnem) tkivu in nivojem izražanja istega gena v kontrolnem (normalnem) tkivu.

Podatke o izražanju genov lahko predstavimo s tabelo (primer je tabela 1), kjer so vsi podatki o enem genu zapisani v eni vrstici. Vsak gen, torej vsaka vrstica v tabeli tako predstavlja en učni primer za nadzorovano strojno učenje, kjer so atributi nivoji izražanja gena v različnih točkah časa (profil izražanja), razred pa je funkcijska skupina, v katero je gen razvrščen. Druga možnost predstavitve profilov izražanja genov je graf, kjer abscisna os predstavlja čas meritev, ordinata pa izražanje gena. Uporabna je predvsem za lažji vpogled v podatke in jo genetiki precej uporabljajo pri določanju funkcije genov. Primera takšnih grafov sta sliki 1 in 3.

Podatki o kvasovki *S. cerevisiae*

Podatki o *S. cerevisiae* obsegajo 2467 učnih primerov z 79 atributi (meritvami) iz osmih bioloških poskusov: prehod iz anaerobnega v aerobno dihanje (ang. *diauxic shift*, 7 meritev),

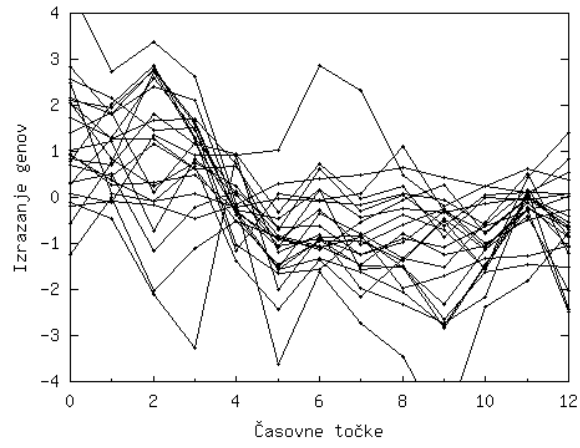
mitoza (ang. *mitotic cell division cycle*, 18+14+15 meritev), sporulacija (ang. *sporulation*, 11 meritev), temperaturni šok (ang. *temperature shock*, 4+6 meritev) in shujševalni šok (ang. *reducing shock*, 4 meritve). Gene so izbrali Eisen in soavtorji[5] glede na razpoložljive in natančne funkcijske oznake (ang. *functional annotations*) in jih razvrstili v šest funkcijskih skupin na podlagi oznak Munich Information Center for Protein Sequences Yeast Genome Database (MYGD): TCA - Tricarboxylic-acid pathway (17 primerov), Resp - Respiration chain complexes (30 primerov), Ribo - Cytoplasmic ribosomal proteins (121 primerov), Proteas – Proteasome (35 primerov), Hist - Histones (11 primerov), HTH - Helix-turn-helix (16 primerov), other – ostalo (2240 primerov). Podatki so bili že v obliki, ki nam je omogočila hitro in enostavno pripravo. Združiti smo morali le dve tabeli. V eni tabeli so bile meritve izražanja genov in imena genov, v drugi pa je bila vsakemu genu določena funkcijska skupina.

Podatki o amebi *D. discoideum*

Podatki o *D. discoideum* obsegajo 126 učnih primerov s 13 atributi (meritvami) iz enega poskusa, kjer je bil organizem podvržen stradanju. Izvirne meritve, dobljene neposredno iz mikromrež DNK, so vsebovale podatke o izražanju 3031 različnih genov, kjer so bili nekateri geni izmerjeni večkrat, in kjer so bile nekatere meritve nepopolne oziroma neuspele. S predhodno obdelavo smo takšne meritve izločili. Gene smo izbrali in razvrstili v skupine tako, da smo združili razpoložljive in natančne funkcijske oznake, ki so dostopne na medmrežju (http://dicty.sdsc.edu/annotationdicty.html), in delne funkcijske oznake, ki nam jih je posredoval Gad Shaulsky. Pri tem smo na njegov predlog uporabili prag *Identity* ≥ 50, ki določa verjetnost pravilnosti funkcijske oznake. Tako smo dobili 126 učnih primerov (genov), ki so razvrščeni v dve funkcijski skupini: sinteza beljakovin (ang. *protein synthesis*, 23 primerov) in ostalo (103 primeri).

Gradnja napovednih modelov s strojnim učenjem

Podatki, kjer so geni eksperimentalno opisani z genskimi izrazi ter označeni s pripadajočo funkcijo, so primerni za obravnavo s strojnim učenjem. Tu je cilj strojnega učenja gradnja modelov, ki lahko za določen gen na podlagi njegovega profila izražanja napovejo pripadnost funkcijski skupini. Na podlagi podatkov o označenih genih iz tabele 1 (prvih deset genov v tabeli) tako lahko zgradimo model, ki za določen profil izražanja gena napove verjetnost pripadnosti funkcijskima skupinama Resp in Ribo. Tak model, zgrajen z metodo *naivnega Bayesa* ($m = 10$), napove, da enajsti (nerazvrščeni) gen iz tabele 1 pripada skupini Ribo z verjetnostjo 0.93.



Slika 1 Izražanje genov funkcijske skupine sinteza beljakovin iz podatkov o D. discoideum.

Tabela 1 Izrazi desetih genov kvasovke S. cerevisiae pri prehodu iz anaerobnega v aerobno dihanje.

gen	X1	X2	X3	X4	X5	X6	X7	funkcijska skupina
YGR207C	0.04	-0.12	0.38	0.14	0.15	0.90	-0.04	Resp
YNL052W	-0.23	-0.23	-0.09	0.08	0.64	1.80	2.30	Resp
YGL187C	0.12	0.29	0.51	0.62	0.94	2.03	2.18	Resp
YGL191W	0.04	-0.07	0.03	0.03	0.89	2.49	2.35	Resp
YLR395C	0.08	0.03	0.42	0.33	0.86	1.32	1.66	Resp
YDL184C	0.19	0.28	0.58	0.30	0.30	-0.62	-1.40	Ribo
YBR048W	0.11	0.20	0.08	-0.42	-1.00	-1.51	-2.47	Ribo
YDR450W	0.44	0.10	0.11	-0.36	-0.09	-1.32	-2.00	Ribo
YHL033C	0.34	0.23	0.19	-0.18	-0.51	-1.56	-2.47	Ribo
YBR189W	-0.03	-0.30	0.03	-0.27	-0.56	-2.00	-2.64	Ribo
YPL198W	0.11	-0.20	0.01	-0.60	-0.64	-1.79	-2.18	?

V ilustracijo problema, s katerim se ukvarja članek, enajsti gen (YPL198W) ni razvrščen: cilj strojnega učenja je zgraditi napovedni model iz podatkov o prvih desetih genih ter za nerazvrščeni gen napovedati, kateri funkcijski skupini pripada.

Uporabnost tovrstnih modelov je potencialno velika, saj je za veliko večino organizmov, na katerih genetiki preučujejo osnovne genetske in biološke zakonitosti, funkcija večine genov še neznana. Modele za razvrščanje genov bi torej gradili iz množice funkcijsko določenih genov, rezultati razvrstitve nerazvrščenih genov pa bi genetikom nudili osnovo za razmišljanje in morebitno načrtovanje dodatnih poskusov, na osnovi katerih bi nedvoumno določili funkcije preostalih genov.

Cilj prispevka je bil raziskati uporabnost metod strojnega učenja na omenjenem področju. Izhajali smo iz zbirk genetskih podatkov, kjer so bili vsi geni razvrščeni v eno od funkcijskih skupin, uspešnost učenja pa smo preverili preko metod, ki so množico genov razbile na učno, t.j. množico za gradnjo modelov, in testno, t.j. množico genov, na kateri smo ovrednotili dobljene modele.

Metode strojnega učenja

Poleg *metode podpornih vektorjev* (SVM, ang. *Support Vector Machines*) smo uporabili še metodo *naivnega Bayesa*, *k-najbližjih sosedov*, *odločitveno drevo* in preizkusili več različic novo razvite metode *QMP*. Metode smo implementirali v sistemu za strojno učenje Orange.[4]

Naivni Bayesov klasifikator

Naivni Bayesov klasifikator predpostavlja pogojno neodvisnost vrednosti različnih atributov pri danem razredu. Osnovna formula Bayesovega pravila je:[8]

$$P(r_k \mid V) = P(r_k)\prod_{i=0}^{a}\frac{P(r_k \mid v_i)}{P(r_k)} \qquad (1)$$

Naloga učnega algoritma je s pomočjo učne množice podatkov oceniti apriorne verjetnosti razredov $P(r_k), k = 1...n_0$ in pogojne verjetnosti razredov $r_k, k = 1...n_0$ pri dani vrednosti v_i atributa $A_i, i = 1...a : P(r_k \mid v_i)$. Za ocenjevanje apriornih verjetnosti se navadno uporablja Laplaceov zakon zaporednosti,[8] za ocenjevanje pogojnih verjetnosti se uporablja *m*-ocena.[8] Vrednost parametra *m*-ocene smo določili z notranjim prečnim preverjanjem.

Za uporabo metode *naivnega Bayesa* smo morali najprej diskretizirati zvezne atribute (meritve izražanja genov v različnih točkah časa). Uporabili smo diskretizacijo Fayyada in Iranija[7], ki s pristopom od zgoraj navzdol (ang. *top-down*) razdeli začetni obseg vrednosti atributa v manjše intervale. Delitev lokalno poveča informacijsko vsebino diskretiziranega atributa (ang.

informativity) in se ustavi, ko je dolžina opisa (ang. *description length*) večja od pridobljene informacije.

Odločitvena drevesa

Odločitveno drevo predstavlja klasifikacijsko funkcijo, ki je hkrati simbolični opis in povzetek zakonitosti v dani problemski domeni. Sestavljeno je iz notranjih vozlišč, ki ustrezajo atributom, vej, ki ustrezajo podmnožicam vrednosti atributov, in listov, ki ustrezajo razredom. Pot v drevesu od korena do lista ustreza enemu odločitvenemu pravilu za klasifikacijo novega primera. Pri tem so pogoji (pari atribut - podmnožica vrednosti), ki jih srečamo na poti, konjunktivno povezani.[8]

Za izbor "najboljšega atributa" smo pri gradnji drevesa uporabili mero razmerje informacijskega prispevka (ang. *gain ratio*). Ker je zanesljivost ocene kvalitete atributa odvisna od števila učnih primerov, ni dobro, da se učna množica prehitro razdeli na majhne podmnožice.[8] Zato smo gradili binarno odločitveno drevo.

k-najbližjih sosedov

Učenje z algoritmom *k*-najbližjih sosedov temelji na shranjenih vseh učnih primerih. Ko želimo napovedati razred r_x novemu primeru u_x, poiščemo med učnimi primeri *k* najbližjih $u_1,...,u_k$ in pri klasifikaciji napovemo večinski razred, t.j. razred, ki mu pripada največ izmed k najbližjih sosedov. Učenja pri tej metodi skorajda ni. Glavnina procesiranja je potrebna pri klasifikaciji novega primera in je zato cena klasifikacije precej večja kot pri drugih metodah učenja.[8] Za računanje razdalj med novim in učnimi primeri smo uporabili evklidsko razdaljo.

Parameter *k* je običajno liho število. S povečevanjem parametra k povprečimo napovedi več bližnjih učnih primerov in s tem zmanjšamo verjetnost, da je vseh k učnih primerov napačnih. Po drugi strani pa z večanjem števila k povečujemo možnost, da h klasifikaciji prispevajo tudi tisti učni primeri, ki niso dovolj podobni novemu primeru. Zato je potrebno za vsak problem posebej eksperimentalno določiti optimalni k.[8]

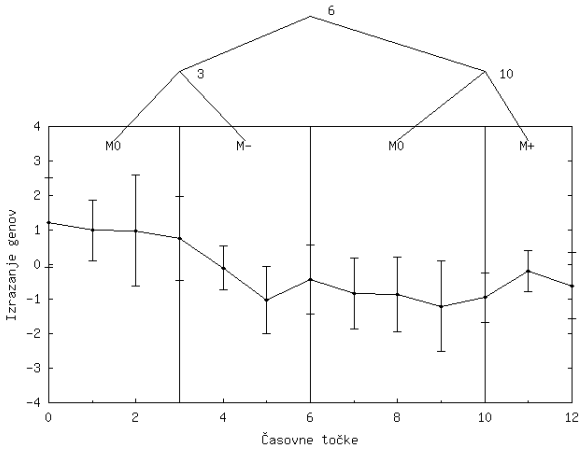
Metoda podpornih vektorjev

Profil izražanja gena si lahko predstavljamo kot točko v m-dimenzionalnem prostoru, kjer je m število meritev v profilu gena. Teoretično bi lahko za vsak razred zgradili binarni klasifikator tako, da bi v prostoru primerov določili hiperravnino, ki bi uspešno ločevala pozitivne od negativnih učnih primerov.

Večina realnih problemov pa vsebuje neločljive podatke (ang. *nonseparable data*), za katere takšna hiperravnina ne obstaja. Problem lahko rešimo z uporabo tako imenovane jedrne funkcije, ki preslika prostor atributov vhodnih podatkov v prostor atributov višje dimenzije (ang. *feature space*), kjer lahko poiščemo ločitveno hiperravnino. Umetno ločevanje učnih primerov na ta način izpostavlja učni sistem nevarnosti generiranja trivialnih rešitev in prevelikemu prileganju podatkom. Metoda podpornih vektorjev (ang. *support vector machines*, SVM)[2,3,10] elegantno rešuje vse te težave. Prevelikemu prileganju podatkom se izogne tako, da vedno izbere hiperravnino z največjo razdaljo do najbližjega učnega primera. Hiperravnino predstavimo kot linearno kombinacijo le tistih učnih točk, ki so ji dovolj blizu. Takšne točke so tipično majhna podmnožica vseh učnih točk, kar dela klasifikacijo učinkovito. Učne točke imenujemo tudi podporni vektorji, ker podpirajo oziroma določajo ločitveno hiperravnino.

Izbira pravega jedra je pomembna, saj določa mero podobnosti dveh vektorjev oziroma točk in tako izraža neko predhodno znanje o pojavu, ki ga modeliramo. Uporabljali smo dve jedrni funkciji: polinomsko jedro $K(\vec{X},\vec{Y}) = (\vec{X}\cdot\vec{Y}+1)^d$, kjer je d stopnja polinoma, in jedro z radialno bazno funkcijo (RBF) $K(\vec{X},\vec{Y}) = \exp\left(\frac{-\left\|\vec{X}\cdot\vec{Y}\right\|^2}{2\sigma^2}\right)$, kjer je

σ širinski parameter Gaussove funkcije. V poskusih Browna in soavtorjev in tudi v naših poskusih σ ustreza mediani vseh evklidskih razdalj v učni množici med vsakim pozitivnim in njemu najbližjim negativnim učnim primerom.[3]



Slika 2 Povprečni profil genov funkcijske skupine sinteza beljakovin amebe in za ta profil zgrajeno drevo kvalitativnih omejitev. Navpične črte so meje med intervali, ki so zapisane v notranjih vozliščih drevesa.

QMP - kvalitativno modeliranje profilov

Poleg uporabe navedenih metod smo razvili in uporabili tudi novo metodo, poimenovano *QMP* (ang. *qualitative modelling of gene profiles*), ki časovne aktivnosti genov modelira kvalitativno. Motivacija za razvoj metode *QMP* je način, na katerega genetiki opisujejo profile izražanja genov. Navadno namreč opisno navajajo časovne intervale, v katerih se izražanje genov ne spreminja, narašča ali pada. Iz takšnih opisov potem sklepajo o funkciji genov. Pri učenju smo želeli modelirati prav te lastnosti profilov in tako upoštevati časovnost, ki je ostale metode ne upoštevajo, saj posamezno meritev (točko na profilu) obravnavajo neodvisno od ostalih. Določiti smo želeli intervale naraščanja, padanja in konstantnega izražanja genov posamezne funkcijske skupine. Tako naučeni model omogoča lažje razumljivo razlago odločanja pri klasifikaciji novih primerov.

Učni algoritem QMP

Osnova metode *QMP* je algoritem *ep-QUIN*,[6] ki se iz numeričnih podatkov nauči binarna drevesa kvalitativnih omejitev. Metoda *QMP* je sestavljena iz *učnega* in *izvajalnega* algoritma. Učni algoritem iz množice podatkov in predznanja tvori model. Izvajalni algoritem pa uporabi naučeni model za reševanje novih problemov.[8] Razvili in preizkusili smo tri različice učnega in tri različice izvajalnega algoritma z dvema možnima vrednostima notranjega parametra (stopnja značilnosti pri statističnih testih). Preizkušali smo torej osemnajst različic ($3\times3\times2$).

Učni algoritem *QMP*(*S*)
Vhod: *S* je profil izražanja funkcijske skupine X. Je polje trojic (X_t, s_{Xt}, n_{Xt}) za vsako točko časa *t*.
Izhod: Drevo kvalitativnih omejitev.

1. naredi vozlišče za Koren drevesa
2. g := **mostConsistentQCF**(*S*)
3. $E(g)$:= cena QCF g ; $E_{best} := E(g)$
4. **for** vsak indeks t polja S **do begin**
5. razdeli profil izražanja gena S v dva dela S_{leq} in S_{grt} glede na *indeks* $\le t$
6. E_{left} := cena **mostConsistentQCF**(S_{leq})
7. E_{right} := cena **mostConsistentQCF**(S_{grt})
8. E_{tree} := cena drevesa (enačba 2)
9. **if** $E_{tree} < E_{best}$ **then begin**
10. $E_{best} := E_{tree}$; $t_{best} := t$
11. **end**
12. **end**
13. **if** $E_{best} < E(g)$ **then begin**
14. v Koren drevesa vstavi indeks t_{best}
15. razdeli S v S_{leq} in S_{grt} glede na *indeks* $\le t_{best}$
16. pod Koren vstavi poddrevesi dobljeni z **QMP**(S_{leq}) in **QMP** (S_{grt})
17. **end else** Koren je list z QCF g
18. **return** Koren

Slika 3 Učni algoritem QMP za učenje drevesa kvalitativnih omejitev.

Vhod v metodo so povprečni profili izražanja genov (vrednosti X_t) vsake funkcijske skupine in njihov standardni odklon (vrednosti s_{Xt}). Učni algoritem uporablja požrešno metodo (podobno

kot algoritmi za učenje odločitvenih dreves) in za vsako funkcijsko skupino zgradi eno drevo kvalitativnih omejitev. Naučeno drevo predstavlja delitev profila izražanja funkcijske skupine na podintervale z enakim kvalitativnim obnašanjem. Notranja vozlišča drevesa so razmejitve profila glede na indeks delitve profila, v listih pa so kvalitativno omejene funkcije QCF, ki opisujejo izražanje na podintervalu v odvisnosti od časa: konstantno (M^0), naraščajoče (M^+) ali padajoče (M^-). Algoritem je prikazan na sliki 3.

Algoritem pri izbiri najprimernejše kvalitativno omejene funkcije (QCF) v listu drevesa uporablja napako, ki temelji na najkrajši dolžini kodiranja.[6] Pri določitvi delitve v notranjem vozlišču izbere vrednost indeksa delitve, ki minimizira napako obeh poddreves. Delitev izvaja, dokler je napaka delitve manjša od napake, kjer za celoten podinterval uporabimo samo eno, najprimernejšo kvalitativno omejeno funkcijo (spremenljivka *g* v zapisu na sliki 3).

$$E_{tree} = E_{left} + E_{right} + SplitCost$$
$$SplitCost = \log_2(Splits_i) \qquad (2)$$

Cena napake (E_{tree}) v vozlišču drevesa (enačba 2) je vsota napak obeh poddreves (E_{left} in E_{right}) in cene kodiranja indeksa delitve (*SplitCost*).[6] $Splits_{\text{i}}$ je število vseh možnih delitev na dva podintervala. Cena napake v listu, za podano kvalitativno omejeno funkcijo, je dolžina kodiranja tistih parov indeksov (pod)intervala, katerih vektor spremembe ne ustreza kvalitativno omejeni funkciji.

Vektor spremembe je smer spremembe vrednosti izražanja med dvema točkama na intervalu, ki sta urejeni po naraščajoči vrednosti indeksa. Možne so tri smeri spremembe, ki sovpadajo s kvalitativno omejenimi funkcijami: pozitivna, negativna in konstantna (brez spremembe).

Razvili smo tri načine izbiranja najprimernejše kvalitativno omejene funkcije (mostConsistentQCF(*S*)) na intervalu *S*: *majority*, *votingTH* in *variance*. Metoda *majority* izbere tisto kvalitativno omejeno funkcijo, katere vektorji sprememb na intervalu so najbolj številčni. Metoda *votingTH* dela podobno kot majority. Med funkcijama M^+ in M^- izbira le, če sta podprti z dovolj velikim deležem vektorjev sprememb, sicer izbere funkcijo M^0. Metoda *variance* s statističnim testom enakosti varianc najprej ugotovi, ali interval ustreza modelu M^0. Sicer izbere med funkcijama M^+ in M^- tisto, ki je podprta z večjim številom vektorjev sprememb.

Statistično značilen vektor spremembe $v_s(i,j)$ med točkama *i* in *j* v intervalu določimo odvisno od izbranega načina računanja najprimernejše QCF. Pri metodah *majority* in *votingTH* uporabimo t-test, pri metodi *variance* pa uporabimo t-test z *Bonferronijevim popravkom*, ki je izrazito konzervativen. Notranji parameter je tako stopnja značilnosti $\alpha \in \{0.05, 0.01\}$, ki naj se uporabi za statistični test.

$$v_s(i,j) = \begin{cases} poz., \text{razlika znacilna in } X_i < X_j \\ neg., \text{razlika znacilna in } X_j > X_i \\ konst., \text{sicer} \end{cases} \qquad (3)$$

Razvrščanje z modelom QMP

Vhod v izvajalni algoritem je profil izražanja gena z neznano funkcijsko skupino. Izvajalni algoritem izračuna napako modela posamezne funkcijske skupine na vhodnem profilu. Gen klasificira v funkcijsko skupino z najmanjšo napako modela. Zaupanje (verjetnost) v odločitev za posamezni razred je nasprotno premosorazmerna z normalizirano vrednostjo izračunane napake.

Izračun napake modela oziroma kvalitativnega drevesa na vhodnem profilu poteka podobno kot v učnem algoritmu. Rekurzivno razdeli profil na podintervale glede na indekse v notranjih vozliščih drevesa, dokler ne pride do listov, kjer so zapisane kvalitativno omejene funkcije. Za vsak list izračuna ceno napake, ki jih nato skozi vozlišča sešteva proti korenu drevesa, ki predstavlja napako celotnega drevesa. Težava nastopi pri računanju napake v listih. Na prvi pogled ne moremo uporabiti statističnih testov za izračun vektorjev sprememb kot smo to počeli v učnem algoritmu

(enačba 3), ker imamo samo en, vhodni profil izražanja gena. Težavo smo poskusili odpraviti na tri načine: *THzero*, *meanDiffKeep* in *meanDiffZero*. Metoda *THzero* uporabi od uporabnika podano minimalno spremembo vrednosti T_{zero}, ki se še obravnava kot konstantna (enačba 4). Privzeta vrednost T_{zero} je 1% razlike med maksimalno in minimalno vrednostjo[6] izražanja profilov genov učne množice.

$$v_s(i,j)=\begin{cases} poz., \text{razlika znacilna in } \mathrm{X_i} + T_{zero} < X_j \\ neg., \text{razlika znacilna in } \mathrm{X_i} - T_{zero} > X_j \\ konst., \text{sicer} \end{cases} \quad (4)$$

Drugi dve metodi, *meanDiffKeep* in *meanDiffZero*, pa uporabita statistični *z*-test za določitev vektorja spremembe med točkama *i* in *j* v intervalu (3). Metoda *meanDiffZero* v primeru, da testiramo model M^0, privzame ničelno razliko med točkama. Povprečja populacij in standardni odkloni pri *z*-testu so vrednosti v točkah profila funkcijske skupine, ki smo jih izračunali v učnem algoritmu.

Obravnavanje podatkov o več poskusih

Kadar podatki o genu vsebujejo meritve iz več neodvisnih poskusov, kot je to v našem primeru pri organizmu *S. cerevisiae*, jih je potrebno obravnavati ločeno. V takšnem primeru za vsak poskus zgradimo en model. Pri razvrščanju novega gena je potrebno upoštevati napovedi vseh modelov. To smo storili na dva načina. Izbrali smo razred, ki je prejel največjo verjetnost pri enem ali več modelih, ali pa smo za vsak razred med seboj zmnožili izračunane verjetnosti modelov in izbrali razred z najvišjo tako izračunano verjetnostjo.

Ocenjevanje uspešnosti učenja

Kvaliteto dobljenih modelov oziroma uspešnost metod strojnega učenja smo ocenjevali z 10-kratnim prečnim preverjanjem, s katerim smo množico razpoložljivih učnih primerov razdelili na 10 približno enako močnih podmnožic. Učenje in ocenjevanje smo tako izvajali desetkrat. V *i*-tem izvajanju smo za ocenjevanje vzeli *i*-to podmnožico, za učenje pa preostalih devet. Za vse metode strojnega učenja smo uporabili isto razbitje na učno in testno množico.

Poleg klasifikacijske točnosti (v rezultatih označena s KT) smo uspešnost učenja merili še s ceno napake oziroma prihrankom cene (označena s prihranek) in površino pod krivuljo ROC (ang. *Receiver Operating Characteristics*, označena z AUC).[9] Statistično značilnost razlike med klasifikatorji smo določili z McNemarovim testom s stopnjo značilnosti $\alpha = 0.005$.

Za prihranek cene napak smo vzeli funkcijo, ki so jo predlagali Brown in soavtorji.[2] Cena uporabe dobljenega modela (klasifikatorja) z metodo strojnega učenja M je definirana kot $C(M) = fp(M) + 2\,fn(M)$, kjer je $fp(M)$ število negativnih primerov, ki jih klasifikator napačno klasificira za pozitivne (ang. *false positives*), in $fn(M)$ število pozitivnih primerov, ki jih napačno klasificira za negativne (ang. *false negatives*). Ceno klasifikatorja metode *M* primerjajo s ceno klasifikatorja ničelne metode strojnega učenja *N*, ki vse primere klasificira za negativne. Prihranek cene (ang. *cost savings*) pri uporabi metode M je tako $S(M) = C(N) - C(M)$. Večji kot je prihranek cene $S(M)$, bolj je metoda M uspešna.

Poskusi in rezultati

Poskuse smo izvajali na obeh zbirkah podatkov in primerjali uspešnost metod strojnega učenja. Ocenili smo metode *večinski klasifikator* (majority), *odločitveno drevo* (TDIDT), k-najbližjih sosedov (*k-NN*), *naivni Bayes* (nb), štiri različice *SVM* in osemnajst različic *QMP*. Gradili smo klasifikator *k-NN* z vrednostjo parametra $k = 11$, en klasifikator *SVM* z jedrom RBF (SVM RBF) in tri s polinomskimi jedri stopenj $d = 1,2,3$ (SVM p1, SMV p2 in SVM p3). Najprej smo izmerili vpliv vrednosti parametra *m* na uspešnost učenja metode *naivnega Bayesa*. Nato smo med seboj primerjali osemnajst različic metode *QMP* in za končno primerjavo izbrali najboljšo. Nazadnje smo

izvedli primerjavo med izbranimi in vsemi ostalimi metodami strojnega učenja.

Večinski klasifikator smo pri primerjavah uporabili kot referenco za najslabši in najenostavnejši možni model, ki določa spodnjo mejo uspešnosti učenja. Klasifikator prešteje učne primere, ki pripadajo posameznemu razredu, s čimer določi verjetnost vsakega razreda. Rezultat učenja je klasifikator, ki vedno vrača najbolj verjetni (večinski) razred in v fazi učenja ocenjene verjetnosti vseh razredov.[4]

Parameter *m* metode *naivnega Bayesa* smo določili z notranjim prečnim preverjanjem. V okviru (glavnega) 10-kratnega prečnega preverjanja smo z (notranjim) 5-kratnim prečnim preverjanjem na učni množici izbrali vrednost parametra $m \in \{0.1,0.2,0.5,1.0,2.0,5.0,...,10000.0\}$, pri kateri je uspešnost učenja največja (za vse tri mere). To vrednost smo uporabili za učenje na celotni učni (pod)množici trenutnega koraka glavnega prečnega preverjanja. Z dvosmernim t-testom s stopnjo značilnosti $\alpha = 0.05$ smo testirali, ali je uspešnost s tako izbranimi vrednostmi parametra statistično značilno različna od uspešnosti pri vrednosti $m = 0$. V končno primerjavo smo vključili boljšega.

Uporabljena imena klasifikatorjev *QMP* (npr. QMP.majority.THzero.0.05) so sestavljena iz štirih, s piko ločenih delov, ki opisujejo značilnosti posameznega klasifikatorja. Ime se vedno začne s *QMP*. Drugi del opisuje uporabljeno metodo pri gradnji modela, tretji del opisuje uporabljeno metodo pri klasifikaciji novih primerov, četrti del pa je lahko 0.05 ali 0.01 in opisuje uporabljeno stopnjo značilnosti za določanje statistično značilnih razlik med gradnjo modela in klasifikacijo novih primerov. Deli imen so zaradi preglednosti v tabelah smiselno okrajšani. Posamezni deli so podrobno obrazloženi v začetnem razdelku opisa metode QMP .

Kvasovka S. cerevisiae

Merili smo klasifikacijsko točnost in prihranek cene napak. Ker v podatkih nastopa več razredov, nismo merili površine pod krivuljo ROC.

Najprej smo izmerili vpliv vrednosti parametra *m* metode *naivnega Bayesa* na uspešnost učenja tako, da smo vrednost parametra *m* izbrali vnaprej in ta isti *m* uporabili pri vseh delih učenja v prečnem preverjanju. To smo ponavljali za različne vrednosti *m* in tako dobili odvisnost uspešnosti učenja od vrednosti parametra *m*. Uspešnost za obe meri je bila največja pri vrednosti $m = 2000$.

Nato smo za obe meri uspešnosti določili povprečje in standardni odklon vrednosti parametra *m* metode *naivnega Bayesa*, pri kateri je bila uspešnost učenja z uporabo notranjega prečnega preverjanja največja (tabela 2).

Tabela 2 Najboljše vrednosti parametra m.

Mera uspešnosti učenja	$\overline{m}$	s_m
Klasifikacijska točnost	1375.000	176.777
prihranek cene napak	1375.000	176.777

Povprečna vrednost in standardni odklon najboljše vrednosti parametra *m*, določene z notranjim prečnim preverjanjem za podatke o *S. cerevisiae*.

Notranje prečno preverjanje uspešno določi vrednost parametra *m* ($m = 1375.0$), saj je ta blizu točke ($m = 2000.0$), kjer sta klasifikacijska točnost in prihranek cene napak največji. Kljub temu pa razlika uspešnosti v primerjavi z uspešnostjo pri vrednosti $m = 0$ ni statistično značilna za obe meri.

Primerjava različic QMP

Primerjali smo osemnajst različic *QMP*. Vsi klasifikatorji *QMP* so slabši od *večinskega klasifikatorja*. Za obe meri dobimo enaki lestvici. McNemarov test označi vse klasifikatorje *QMP* za statistično značilno slabše od *večinskega klasifikatorja*. Najboljši klasifikator *QMP.variance.meanDiffKeep.0.05*, ki smo ga tudi izbrali za primerjavo z ostalimi metodami, je statistično značilno boljši od vseh klasifikatorjev

QMP razen dveh (*QMP.majority.meanDiffKeep.0.05* in *QMP.votingTH.meanDiffKeep.0.05*).

Primerjava metod strojnega učenja

Tabela 3 prikazuje uspešnost učenja klasifikatorjev glede na obe meri uspešnosti. Za obe meri uspešnosti dobimo isto lestvico najboljših klasifikatorjev. Statistično značilno različni pari, ki jih določi McNemarov test, so prikazani v tabeli 4. Tabela je simetrična preko diagonale. Metode so razvrščene po naraščajoči uspešnosti učenja od leve proti desni v vrstici oziroma od zgoraj navzdol v stolpcu. Najboljši klasifikator *k- NN* je statistično značilno boljši od vseh klasifikatorjev *SVM* razen dveh.

Tabela 3 Uspešnost metod strojnega učenja na podatkih o S. cerevisiae.

Mera uspešnosti	$\bar{x}_{prihranek}$	$s_{prihranek}$	$\bar{x}_{KT}$	s_{KT}
majority	425.0	0.0	0.907	0.000
SVM p1	436.7	12.6	0.923	0.017
SVM p2	464.9	9.2	0.961	0.012
SVM p3	464.6	6.8	0.960	0.009
SVM RBF	459.8	8.4	0.954	0.011
nb prihranek	459.8	8.5	0.954	0.011
k-NN	**468.5**	4.7	**0.966**	0.006
TDIDT	425.0	0.0	0.907	0.000
QMP*	384.5	39.7	0.852	0.054

*QMP.var.mDK.0.05

Ameba D. discoideum

Merili smo klasifikacijsko točnost in prihranek cene napak. Ker v podatkih nastopata samo dva razreda, smo lahko merili tudi površino pod krivuljo ROC.

Tudi tu smo na enak način kot pri kvasovki najprej izmerili vpliv vrednosti parametra *m* metode *naivnega Bayesa* na uspešnost učenja in dobili primerljive rezultate. Tabela 5 prikazuje povprečje in standardni odklon vrednosti parametra *m*, pri kateri je bila uspešnost učenja z uporabo notranjega prečnega preverjanja največja.

Tudi tu notranje prečno preverjanje uspešno določi vrednost parametra *m*. Razlika uspešnosti v primerjavi z uspešnostjo pri vrednosti $m = 0$ ni statistično značilna za vse tri mere. Uspešnost, merjena s površino pod krivuljo ROC, je skorajda konstantna in ne kaže vpliva vrednosti *m* na uspešnost učenja.

Tabela 4 Najboljše vrednosti parametra m.

Mera uspešnosti učenja	$\overline{m}$	s_m
klasifikacijska točnost	9.280	15.610
prihranek cene napak	4.290	6.414
površina pod krivuljo ROC	2.600	3.459

Povprečna vrednost in standardni odklon najboljše vrednosti parametra *m*, določene z notranjim prečnim preverjanjem za podatke o D. discoideum.

Primerjava različič QMP

Tudi tu so vsi klasifikatorji *QMP* statistično značilno slabši od *večinskega klasifikatorja*, če jih primerjamo na podlagi klasifikacijske točnosti in prihranka cene napak. Za obe meri dobimo enaki lestvici. Najboljši klasifikator *QMP* je *QMP.votingTH.THzero.0.01*. Lestvica najbolj uspešnih klasifikatorjev se spremeni, če uporabimo za oceno uspešnosti učenja površino pod krivuljo ROC. *Večinski klasifikator* v tem primeru pade na predzadnje mesto, najboljši klasifikator pa je *QMP.majority.meanDiffKeep.0.01*. Za primerjavo z ostalimi metodami smo izbrali *QMP.votingTH.THzero.0.01*.

Primerjava metod strojnega učenja

Tabela 6 prikazuje uspešnost učenja klasifikatorjev za vse tri mere uspešnosti. McNemarov test pokaže statistično značilne razlike le med klasifikatorjem *QMP* in ostalimi. Za meri klasifikacijske točnosti in cene napak je *naivni Bayes* najboljši, sledijo mu klasifikatorji *SVM* in *k-NN*.

Primerjava na podlagi površine pod krivuljo ROC nam da drugačne rezultate. *Naivni Bayes* je še vedno najboljši, sledijo mu pa *k-NN*, *SVM* in *QMP*. Metodi *TDIDT* in *majority* si delita zadnje mesto.

Zaključek

Strojno učenje se je izkazalo za uporabno na obravnavani problemski domeni. Kot bolj uspešni od metode podpornih vektorjev, čeprav ne statistično značilno, sta se izkazali metodi *naivnega Bayesa* in *k-NN*. Metoda *QMP* je manj uspešna, nam pa namesto kompleksnih matematičnih modelov daje zelo razumljive simbolne modele, ki dobro opisujejo profile izražanja genov.

Poleg razvoja nove metode *QMP*, ki lahko v prikazani inačici služi predvsem za pomoč pri vizualizaciji časovnih odzivov genov, je osnovni prispevek pričujočega dela ugotovitev, da se enostavna metoda *naivnega Bayesa* pri napovedi funkcije genov obnese vsaj tako dobro kot metoda *SVM*. Ta ugotovitev je presenetljiva, saj velja *SVM* za *de-facto* standard za tovrstno obdelavo podatkov, in je v nasprotju z nekaterimi objavljenimi rezultati.[2] Poleg preprostosti sta prednosti *naivnega Bayesa* tudi časovna učinkovitost (gradnja napovednega modela za podatke o *S. cerevisiae* traja približno 4 sekunde namesto 5 minut, kolikor jih potrebuje SVM) in možnost razlage odločitve.

Zahvala

Dr. Gad Shauslky in dr. Chad Shaw iz Baylor College of Medicine sta nam posredovala genetske podatke o *D. discoideum* ter gene iz podatkov v namene opisane analize razvrstila v funkcionalne skupine. Hvala tudi dr. Dorianu Šucu iz Fakultete za računalništvo in informatiko Univerze v Ljubljani za nasvete pri razvoju metode QMP. Delo je nastalo v okviru MŠZŠ projekta Metode odkrivanja znanj za funkcionalno genomiko (J2-3387, BZ in TC).

Tabela 5 Statistično načilne razlike med klasifikatorji na podatkih o S. cerevisiae.

	QMP	majority	TDIDT	SVM p1	SVM RBF	nb prih.	SVM p3	SVM p2	*k*-NN
QMP	*	*	*	*	*	*	*	*	*
majority	*				*	*	*	*	*
TDIDT	*				*	*	*	*	*
SVM p1	*				*	*	*	*	*
SVM RBF	*	*	*	*					*
nb prih.	*	*	*	*					*
SVM p3	*	*	*	*					
SVM p2	*	*	*	*					
k-NN	*	*	*	*	*	*			

Statistično značilne razlike (označene z zvezdico) med klasifikatorji, določene z McNemarovim testom, na podatkih o *S. cerevisiae*. Klasifikator *QMP.variance.meanDiffKeep.0.05* je poimenovan *QMP*.

Tabela 6 Uspešnost metod strojnega učenja na podatkih o D. discoideum.

Mera uspešnosti učenja	$\bar{x}_{prih.}$	$s_{prih.}$	$\bar{x}_{KT}$	s_{KT}	$\bar{x}_{AUC}$	s_{AUC}
majority	18.3	1.3	0.818	0.034	0.500	0.000
SVM p1	19.2	5.3	0.842	0.139	0.763	0.231
SVM p2	19.5	4.9	0.850	0.128	0.760	0.191
SVM p3	19.2	5.4	0.843	0.146	0.695	0.227
SVM RBF	19.8	4.5	0.859	0.124	0.810	0.205
nb prihranek	**20.7**	3.9	**0.882**	0.104	**0.858**	0.198
k-NN	19.5	4.9	0.849	0.123	0.831	0.202
TDIDT	18.3	1.3	0.818	0.034	0.500	0.000
QMP.var.mDK.0.05	-3.3	7.6	0.249	0.208	0.702	0.257

Literatura

1. Dennis C, Gallagher R (urednika): The Human Genome. Nature Publishing Group, 2001.
2. Brown MPS, Grundy WN, Lin D, Cristianini N, et al.: Knowledge-based Analysis of Microarray Gene Expression Data Using Support Vector Machines, PNAS, vol. 97(1), pp. 262-267, 2000.
3. Brown MPS, Grundy WN, Lin D, et al.: Support Vector Machine Classification of Microarray Gene Expression Data, University of California, Santa Cruz and University of Bristol, Tec. Rep. UCSC-CRL-99-09, June 1999.
4. Demšar J, Zupan B: Programski paket za strojno učenje Orange. http://magix.fri.uni-lj.si/orange, 2002.
5. Eisen M, Spellman P, Brown P, Botstein D: Cluster analysis and display of genome-wide expression patterns. PNAS, vol. 95, pp. 14863-14868, Dec. 1998.
6. Šuc D: Machine reconstruction of human control strategies, Doktorska disertacija, Fakulteta za računalništvo in informatiko, Ljubljana, 2001.
7. Fayyad UM, Irani KB: The Attribute Selection Problem in Decision Tree Generation. Proc. of AAAI-92, pp. 104-110, San Jose, CA, 1992.
8. Kononenko I: Strojno učenje, Založba FE in FRI, Ljubljana, 1997.
9. Provost F, Fawcett T:, Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions, Proc. of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97), 1977.
10. Jakulin A: Knjižica orngExtn-1.0. http://ai.fri.uni-lj.si/~aleks, 2002.

Strokovno-znanstveni prispevek ■

Avtomatiziran video nadzor Centra za preprečevanje in zdravljenje odvisnosti od prepovedanih drog v Trbovljah

Branko Ikica, Andrej Ikica, Uroš Prelesnik, Aleksandar M. Caran

Izvleček. Naraščanje nasilja in kriminala se vse pogosteje širi tudi med stene javnih ustanov, predvsem med centre, ki se ukvarjajo s problemi zdravljenja odvisnosti od mamil. Ker so delavci podobnih centrov stalno izpostavljeni nevarnosti izbruha nasilja, je zanje ključnega pomena dober nadzorni sistem. V članku podajamo opis sistema za avtomatiziran video nadzor Centra za preprečevanje in zdravljenje odvisnosti od prepovedanih drog v Trbovljah. Ta med drugim omogoča popolnoma avtomatiziran video nadzor centra izven delovnega časa ter možnost hitrega vpogleda v dogajanje v čakalnici centra.

Automated video surveillance of "Centre for drugs and drug addiction" in Trbovlje

Abstract. Among many public institutions, violence and criminal rate are also rapidly spreading to those dealing with a drug abuse and drug addiction problems. Working staff in such centres is quite often exposed to violence, so a good video surveilance system is of main concern to them. In this article we present an automated video surveillance system installed in "Centre for drugs and drug addiction" in Trbovlje (Slovenia). Among many other options it allows a fully automated video surveilance of centre and quick overview of centre's headquaters.

■ **Infor Med Slov** 2003; 8(1): 81-85

Institucije avtorjev: Iks d.o.o., računalniški inženering (BI, AI), Zdravstveni dom Trbovlje (UP), Center za preprečevanje in zdravljenje odvisnosti od prepovedanih drog Trbovlje (AMC).

Kontaktna oseba: Branko Ikica, Iks d.o.o., računalniški inženering, Obrtniška cesta 11A, 1420 Trbovlje. email: info@iks-doo.com.

Uvod

Centri za preprečevanje in zdravljenje odvisnosti od prepovedanih drog (CPZOD) so ves čas izpostavljeni visokemu tveganju ropa in izbruha nasilja. Do slednjega lahko pride med samimi pacienti kot tudi nad osebjem centra.

CPZOD Trbovlje se je razvil iz ambulante za zdravljenje odvisnosti, ki je od leta 1995 delovala v ZD Trbovlje še pod okriljem ljubljanskega CPZOD.

V trboveljskem centru lahko dobijo informacije o drogah, preprečevanju in zdravljenju zasvojenosti in zmanjševanju škode zaradi zlorabe prepovedanih drog vsi zainteresirani s širšega področja Zasavja. Pri tem se obvezno upošteva tudi njihova želja po anonimnosti.

Za tiste, ki se odločijo za zdravljenje, je na voljo več programov, v katere se pacienti lahko vključijo na osnovi lastne želje in po posvetu oz. triaži z zdravnikom.

Pacienti, ki so odvisni od prepovedanih drog, se lahko odločajo med programom psihosocialne podpore in vodenja, substitucijsko terapijo, programom za rehabilitacijo in socialno reintegracijo, zdravljenjem in pomočjo pri recidivu bolezni ter svetovanjem in pripravo na vključevanje v terapevtske skupnosti in komune.

Vsi pacienti se obravnavajo individualno in/ali skupinsko. Obstaja tudi program skupinskega svetovanja s starši in partnerji zasvojencev.

V centru so na voljo: dva zdravnika, dva psihiatra, psihologinja, socialna delavka in dve medicinski sestri. Specifično delo z zasvojenci je organizirano tako, da je nekdo od osebja vedno dosegljiv.

Sistem *Security Pro*[1] razvit v podjetju Iks d.o.o. ter prirejen potrebam CPZOD Trbovlje igra pomembno vlogo pri delu v centru. Sistem omogoča neprekinjen nadzor nad celotnim prostorom centra med delovnim časom in izven njega. Le-to pa zagotavlja varnejše počutje in bolj sproščeno delo celotnega osebja.

Zahteve CPZOD Trbovlje

Pri zasnovi avtomatiziranega video nadzornega sistema za CPZOD Trbovlje smo izhajali iz programskega paketa *Security Pro* podjetja Iks d.o.o, ki smo ga nadgradili v skladu s specifičnimi zahtevami centra:

1. Avtomatičen video nadzor izven delovnega časa (nočni čas), ki vse sumljive dogodke hrani v podatkovno bazo. Zdravnik mora imeti vsako jutro možnost vpogleda v dogajanje prejšnjega dne in v primeru ropa tudi pregled arhiva slik.

2. Nadzor med delovnim časom. Tu je bistvenega pomena nadzor čakalnice. Sestra oz. zdravnik morata imeti v vsakem trenutku možnost takojšnjega pregleda nad dogajanjem v čakalnici in v primeru, da je le-to sumljivo (prekupčevanje, pretep ipd.), možnost ročnega snemanja scene.

3. Nemoteno opravljanje ostalega dela na računalniku. Programski paket Security Pro omogoča delovanje v preprostem načinu (delovanje v ozadju, minimizirano okno, možnost hitrega snemanja).

Opis strukture video nadzornega sistema

Video nadzorni sistem v CPZOD Trbovlje vsebuje:

1. barvno video kamero tipa JVC (nameščena v čakalnici)

2. programski paket *Security Pro* prirejen zahtevam CPZOD Trbovlje

[1] Več informacij o sistemu dobite na strani www.iks-doo.com.

Slika 1 Kamera nameščena v čakalnici CPZOD

Programski paket Security Pro za potrebe CPZOD Trbovlje

Programski paket *Security Pro* je v celoti razvit v podjetju Iks d.o.o. Osnovna različica paketa omogoča neprekinjen, 24-urni avtomatiziran video nadzor objektov. V primeru, ko nadzorni sistem detektira gibanje v sceni, se odloči ali gre za regularen dogodek ali le senco oz. nepomembno spremembo v osvetlitvi scene in informacijo o dogodku hrani v podatkovno bazo, do katere ima uporabnik dostop.

V poglavju 2 smo že izpostavili dodatne zahteve CPZOD. Te so: možnost ročnega video nadzora ter možnost nemotenega opravljanja ostalega dela na računalniku, medtem ko aplikacija teče.

Opis strukture in mehanizmov sistema je podan v naslednjih podpoglavjih.

Struktura programskega paketa

Strukturo programskega paketa prikazuje slika 2. Le-tega sestavljajo *vmesnik za zajem žive slike iz video kamere (1)*, *modul z vgrajenim mehanizmom za detektiranje oseb (3)*, ki se gibljejo po prostoru ter *podatkovno bazo (4)* za hranjenje ključnih podatkov.

Programski vmesnik za zajem slikovnega toka (videa) iz video kamere je zasnovan na podlagi knjižnic VFW (Video For Windows), kar omogoča namestitev poljubne digitalne kamere na sistem. K dodatni fleksibilnosti sistema pripomore poseben pretvornik A/D, ki nudi možnost priključitve starejših analognih kamer na sistem.

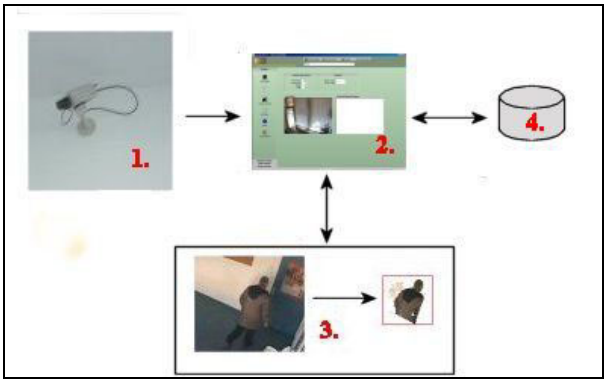
Zajeti slikovni tok v obdelavo dobi modul za detekcijo oseb, ki temelji na prirejenih algoritmih detekcije gibanja. Osnovni algoritem obdeluje vsako sličico posebej in na njej detektira premikajoče se regije. Le-te grupira in na podlagi izločevalnega mehanizma izloča elemente, ki ne ustrezajo kriteriju premikajočega se objekta - spremembe v osvetlitvi, sence, neregularni objekti. Ko je v sceni detektirana premikajoča se oseba, pravimo da pride do nastopa dogodka.

Ob nastopu dogodka sistem informacijo o tem shrani v podatkovno bazo. Informacija vsebuje uro in datum nastopa dogodka, zaporedno številko scene (kadra) ter samo sliko dogodka.

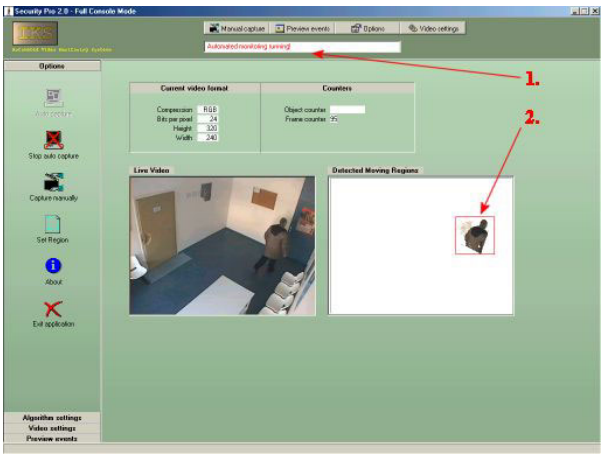
Avtomatičen nadzor

Že osnovna različica programskega paketa vsebuje možnost avtomatičnega video nadzora, ki v CPZOD Trbovlje pride v poštev izven delovnega časa. Ob koncu delovnega dne zdravnik oz. sestra vkluči avtomatični način delovanja sistema, ki avtonomno deluje čez noč. V primeru kakršnegakoli dogodka sistem shrani informacijo v podatkovno bazo, tako da je ta vsako jutro dostopna zdravniku.

Na to, da sistem deluje avtonomno, nas opozarja indikator (1), ki je označen na sliki 3. V primeru, ko pride do dogodka, sistem v desnem oknu (2) pokaže detektirani objekt ter istočano shrani informacijo o dogodku v podatkovno bazo.



Slika 2 Struktura programskega paketa Security Pro

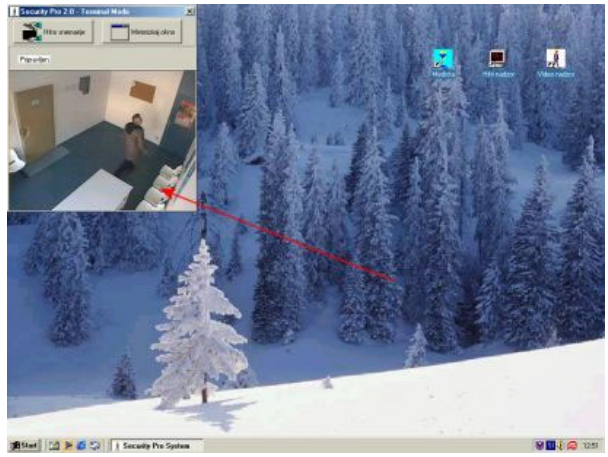


Slika 3 1 - indikator, da sistem teče avtomatično, 2 - detektirana oseba

Ročni nadzor

Dodatna zahteva CPZOD Trbovlje je možnost nadzora med delovnim časom. V tem primeru avtomatičen nadzor ni potreben, saj lahko sestra spremlja dogajanje v čakalnici v živo preko računalniškega zaslona.

V ta namen ponuja programski paket možnost izvajanja v preprostem načinu. V tem primeru aplikacija teče v minimiziranem oknu v levem zgornjem kotu računalniškega zaslona. S klikom na minimizirano okno sestra v trenutku poveča okno na velikost prikazano na sliki 4. Če sestra opazi sumljivo dogajanje – pretep, prekupčevanje v čakalnici, lahko takoj prične s snemanje videa, ki se evidentira v podakovni bazi in predstavlja pomembno dokazno gradivo.



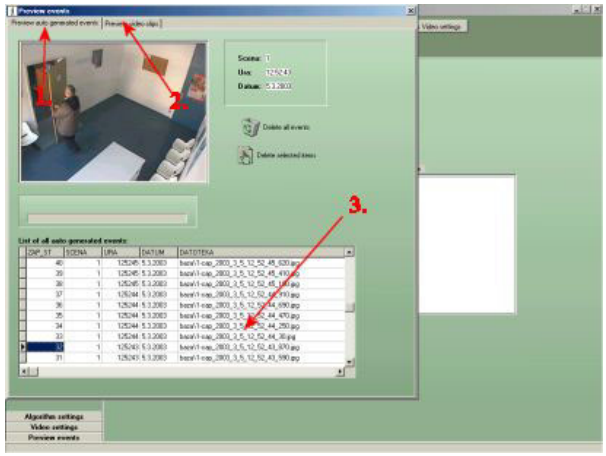
Slika 4 Preprosi način dela (nemenjen sestri)

Pregled zgodovine in ostale nastavitve

Programskemu paketu je dodana podatkovna baza, ki hrani celotno zgodovino dogajanja v CPZOD Trbovlje. Slike dogodkov so shranjene v podatkovni bazi v formatu JPEG.

Zaradi »pametnega mehanizma« detekcije ni potrebno shranjevati 24-urnega slikovnega gradiva, temveč le ključne informacije o dogodkih, kar zmanjša potrebo po diskovnem prostoru tudi do 99%!

Z navigacijo po podatkovni bazi se ustrezno spreminja shranjena slika (v oknu levo zgoraj), kar pomeni hiter in učinkovit pregled aktivnosti (slika 5).



Slika 5 1 - pregled dogodkov, 2 - pregled video posnetkov, 3 - podatkovna baza

Primerjava s podobnimi sistemi

Za video nadzor se večinoma uporablja analogna oprema. Kamere zajemajo sliko in jo preko analognega multiplekserja pošiljajo na videorekorder, ki celotno dogajanje snema na trak. Slabost tega sistema je količina shranjenih podatkov, ki zajema ure in ure slikovnega gradiva, iskanje po njem pa zahteva dolgotrajno pregledovanje in previjanje kupa trakov. V svetu so se pred kratkim začeli pojavljati tudi digitalni sistemi, ki namesto traku uporabljajo trdi disk. Cena teh sistemov sega zelo visoko, tudi v miljonone slovensikh tolarjev. Poleg tega ti sistemi večinoma snemajo celotno 24-urno dogajanje tudi tedaj, ko v sceni ni premikov.

Sistem Security Pro zgoraj opisane pomankljivosti učinkovito rešuje. Detekcija gibanja omogoča shranjevanje le tistih kadrov, v katerih pride do premika. Prav tako je iskanje kadrov enostavno, saj sistem vsak kader evidentira z enolično identifikacijsko številko, z datumom in časom nastopa kadra ter s sliko, ki kader označuje. S premikanjem po podatkovni bazi lahko uporabnik hitro pregleduje zgodovino dogajanja.

Trenutna slabost sitema Security Pro predstavlja shranjevanje ročno zajetih kadrov v obliki datotek AVI, ki zajemajo precej prostora na disku. V avtomatskem načinu dela naš sistem še vedno deluje optimalno, saj zajema posamezne sličice v formatu JPEG.

Zaključek

Sistem smo instalirali v CPZOD marca 2003. Sistem se je izkazal kot zelo uporaben in praktičen pri vsakdanjem delu, saj ga osebje centra neprestano uporablja pri kontroli dogajanja v čakalnici. Ker sta čakalnica in ordinacija ločeni z vrati, je pregled nad dogajanjem v čakalnici brez sistema za video nadzor nemogoč. Sistem Security Pro sestri omogoča preverjanje oseb, preden so le-te sprejete v ordinacijo.

V prihodnosti načrtujemo inplementacijo sistema z uporabo metod Microsoft DirectShow, ki omogočajo hitrejše procesiranje slik in uporabo algoritmov za kompresiranje kadrov (MPEG). Pokazala se je tudi potreba po postavitvi dodatne kamere v hodnik pred čakalnico CPZOD. Tako bomo varnostni sistem centra še dodatno okrepili.

Z uporabo ustrezne kamere z mikrofonom lahko izvajamo tudi avdio nadzor. Načrtujemo implementacijo algoritmov za zajem in analizo zvočnih podatkov, kar bo omogočilo shranjevanje videa v primeru hrupa, kadar sestra ni prisotna.

Literatura

1. Microsoft Developer Network (MSDN), Platform SDK: Windows Multimedia.
2. Marco Cantu – Mastering Delphi 5, Sybex, 1999.
3. Domača stran CMU, računalniški vid, http://www-2.cs.cmu.edu/afs/cs/project/cil/ftp/html/vision.html

Poročilo o obisku ■

Poročilo o obisku Informacijskega oddelka univerzitetne bolnišnice v Tokiju

V sklopu štipendije JSPS (Japan Society for Promotion of Science) sem imel čast obiskati tudi Informacijski oddelek univerzitetne bolnišnice v Tokiju in v tem kratkem poročilu bom podal nekaj zanimivosti. Programsko opremo je razvila firma Fujitsu, na katerih računalnikih oprema tudi teče. Enak program uporablja še nekaj drugih večjih japonski bolnišnic, drugače pa podobno kot v svetu še ni standardov in medsebojno bolnišnice še vedno ne izmenjujejo podatkov. Sam informacijski sistem pa je vendarle zelo napreden in omogoča zdravstvenemu osebju vpisovanje in spremljanje zgodovine pacienta, naročanje pregledov, doziranje in predpisovanje zdravil terapij, pregled kliničnih pravil, ipd. Sistem je zelo integriran in enostaven za uporabo, zdravstveno osebje pa ga v omejeni obliki lahko uporablja tudi preko spleta (samo pregled zgodovine). Varnost je zagotovljena tako z biometriko kot gesli, zdravniki in sestre pa lahko dobijo vse podatke o vseh pacientih v bolnišnici – zanesejo se predvsem na etiko uslužbencev. Poleg računalniških terminalov uvajajo še zelo robustne osebne dlančnike opremljene s čitalci kode in brezžičnim mrežnim dostopom. Vsak pacient ima zapestnico s črtno kodo tako, da lahko s pomočjo omenjenega dlančnika kjerkoli v bolnišnici preverijo njegove osnovne podatke, kot krvna skupina, diagnoza, predpisana terapija, ipd.

Posebna zanimivost je tudi vključenost pacientov v sam informacijski sistem. S pomočjo svoje zapestnice lahko vstopijo v sistem in dobijo nekaj osnovnih podatkov o svojem stanju in terapiji ter naročajo različne vrste obrokov in ostalih uslug. Vsaka bolniška soba, v kateri sta največ dva bolnika je opremljen z računalniškim terminalom, ki je hkrati tudi televizija, ter poleg že omenjenih storitev bolniku omogoča tudi dostop do interneta.

Peter Kokol

■ **Infor Med Slov** 2003; 8(1): 86

Poročilo s simpozija ■

AMIA 2002, 9. do 13. november 2002, San Antonio, ZDA

Tokraten simpozij AMIA (American Association for Medical Informatics) je bil organiziran od 9 do 13 novembra 2002 v Nenry B. Gonzales centru v 9 največjem ameriškem mestu San Antonio, ki slovi predvsem po Alamu (znamenita bitka med Američani in Mehičani, ki so jo upodobili v znamenitem vesternu Alamu, kjer je glavno vlogo igral John Wyne) in River Walku, sprehodu ob rečici, predel z mostički, restavracijami, majhnimi parki in trgovinami spominja na miniaturne Benetke.

Teme simpozija: "Biomedical Informatics: One Discipline"

Glavnino simpozija predstavljajo tutoriali (organiziranih kar 38), delavnice (16) in paneli (27). Poudarek znanstvenih sekcij je bil predvsem na bionformatiki (plenarni trak) in nato še na izobraževanju, uporabnikovih potrebah, podpori odločanju in iskanju informacij. Poudariti velja, da je bila konkurenca kljub relativno velikemu številu predstavljenih člankov (147) in postrov (300) zelo huda, saj je bil delež sprejetih prispevkov zelo nizek, le 37%.

Na konferenci so kot avtorji ali kot sodelovali tri Slovenci: Ankica Babič, Blaž Zupan in Peter Kokol. Vsi trije so predstavljali posterje, Blaž pa je sodeloval tudi kot predavatelj pri dveh delavnicah s področja odkrivanja znanj iz podatkov.

Nekaj vtisov

Iz uvodnega plenarnega predavanja Dr, Botsteina, ki je govoril o funkcionalni genetiki, bi povzel predvsem dve, sicer že dobro znani dejstvi, a vendar še vedno premalo poudarjeni:

medicinci in biologi so slabi s številkami,

ZATO JE

naloga je naloga informatikov, da za njih izdelajo primerna orodja, ki jim bodo numerične rezultate prikazali na njim primeren način.

V primerjav z MIE konferencami je AMIA simpozij vsaj po mojem mnenju manj raziskovalno usmerjen, v glavnem so predstavljene že tekoče, dobro evalvirane in testirane aplikacije in ne toliko najbolj sveže raziskave. Mnogo več je tudi sponzorjev in konkurenca članov je veliko večja, kakor tudi udeležba.

Peter Kokol

■ **Infor Med Slov** 2003; 8(1): 87

Poročilo o sestanku ■

Poročilo o udeležbi na sestanku IMIA – NI Sig, 12. november 2002, San Antonio, ZDA

V okviru konference AMIA 2002 Symposium (AMIA – American Medical Informatics Association) je bil organiziran tudi sestanek edine specialne interesne grupe (SIG), ki deluje v okviru IMIA (International Medical Informatics Association), ki sem se je udeležil kot slovenski opazovalec.

Uvodne besede

Prispevek naj bo napisan v slovenščini, dodatno naj še vsebuje angleški naslov in povzetek.

Predsednica SIG Dr. Virginia Saba je predstavila novo definicijo informatike v zdravstveni negi (Nursing informatics) in poudarila, da se le ta razlikuje od ameriške, je pa naslednja:

Informatika v zdravstveni negi je integracija zdravstvene nege in menedžmenta informacije s procesiranjem informacije in komunikacijskimi tehnologijami v smislu podpiranja zdravja ljudi po celem svetu.

Nato je predstavila delovna telesa (working groups):

- Preprezentacija konceptov (Virginia Saba)
- Raziskovanje (Heather Stratchan)
- Menedžment (Robyn Carr)
- Izobraževanje (Margaret Ehborfs)
- Zgodovina (Marianne Talberg)
- Podatkovni standardi (Kathlen McCormik)
- Praksa na podlagi evidence (C Weaver)
- Uporabniško zdravstvo – Consumer health (Betty Chang)
- Telematika v zdravstveni negi (Paula Procter)

Poročila delovnih skupin

Prisotne vodje delovnih skupin, so podale kratka poročila:

Izobraževanje

Skupina želi ustanoviti mednarodni certifikat znanja iz informatike v zdravstveni negi, ki bi ga podeljevala IMIA na podlagi opravljenih izpitov. Zapletlo se je pri imenu certifikata, ker ima ta beseda v različnih državah različne pomene. Evropskih predstavniki smo predlagali naj bo podeljeni naziv imenuje diploma, s čimer so se bolj ali manj strinjali tudi ostali, vendar celotna akcija še ni končana, bo pa verjetno stvar dogovorjena na kongresu NI 2003 v Riu de Jainero.

Uporabniško zdravstvo

Skupina je bila ustanovljena 2001. Do sedaj je izdelala brošuro za ocenjevanje zdravstvenih mrežnih strani, v izvedbi pa je pregled (survey) o tem kaj »uporabniško zdravstvo« sploh je. Obljubil sem, da bo v pregledu sodelovala tudi Slovenija.

Minimalni nabor podatkov

Projekt traja že več let, izgleda pa da bo v naslednjem letu ali dveh končno dosežen konsenz mednarodne skupnosti in bo baziral na ICNP.

Standardi

Delovna skupina razvija standardno terminologijo v zdravstveni negi pod okriljem ISO/TC 215 WG3. TC 215 se ukvarja z terminologijo v zdravstvu nasplošno, WG3 pa s terminologijo v zdravstveni negi. Sodeluje 28 držav, konsenz je skoraj dosežen, edini, ki še niso dali pristanka so Švedi.

Sugestije

1. Glede slovenskega predstavnika v IMIA NI SIG vlada majhna zmeda, saj se ne ve točno kdo ta predstavnik je. Govori se o več imenih, najbolje bo, da SDMI na naslednji seji imenuje svojega predstavnika in ga uradno prijavi – predlagam vodjo slovenske sekcije za informatiko v zdravstveni negi.
2. Prav tako bi morali čimprej imenovati svoje predstavnike tudi v delovne skupine.
3. Razmisliti bi morali morda tudi o organizaciji Nursing informatics 2009, ki bo organiziran v Evropi.

Peter Kokol

■ **Infor Med Slov** 2003; 8(1): 88-89