Semi-automated Knowledge Elicitation for Modelling Plant Defence Response

Dragana Miljkovic
Department of Knowledge Technologies, Jožef Stefan Institute, Jamova cesta 39, Ljubljana, Slovenia
E-mail: Dragana.Miljkovic@ijs.si

Thesis Summary

This article represents a summary of the doctoral dissertation of the author on the topic of knowledge elicitation for modelling plant defence response.

Povzetek: Članek predstavlja povzetek doktorske disertacije, ki obravnava temo zajemanja znanja za modeliranje obrambnega odziva rastlin.

1 Introduction

In systems biology, the growth of experimental data is not uniform for different types of biological mechanisms, hence some biological mechanisms still have few datasets available, like plant defence against pathogen attack. Pathogens are a serious threat to living organisms and can lead to fitness costs, physiological damage or even death [1]. In particular, plants have specially evolved sophisticated mechanisms that can effectively fight-off infections with various pathogens. Upon pathogen recognition, plants trigger a complex signalling network, referred as plant defence response or plant defence signalling (PDS). Biologists have been investigating plant defence response to virus infections; however a comprehensive mathematical model of this complex process has not been developed. One challenge in developing a dynamic model, useful for simulation, are scarce experimental data from which the model parameters could be determined.

2 Methods and results

The thesis [2] describes a novel methodology for the construction of biological models by eliciting the relevant knowledge from literature and domain experts. The methodology has been applied to build the PDS model, and can be used to construct models of other biological mechanisms. The thesis also presents a PDS model.

Most of the plant-pathogen interaction studies are focused on individual interactions or subsets of the whole PDS mechanism. The models that are commonly used are structural models with no information on their dynamics [3]. Several dynamical models of plant defence have been developed. However, they are either simple [4], or do not contain sufficiently detailed information on the pathways of interest in this dissertation [5], or focusing only on one pathway [6].

In order to build the PDS model the standard approach to the construction of dynamic models is enhanced with the following methods: a method for model structure revision by means of natural language processing techniques, a method for incremental model structure revision, and a method for automatic optimisation of model parameters guided by the expert knowledge in the form of constraints.

The initial model structure was first constructed manually by defining the representation formalism, encoding the information from public databases and literature, and composing a pathway diagram. To complement the model structure with additional relations, a new approach to information extraction from texts was developed. This approach, named Bio3graph [7], allows for automated extraction of biological relations in the form of triplets followed by the construction of a graph structure which can be visualised, compared to the manually constructed model structure, and examined by the experts. Using a PDS vocabulary of components and reaction types, Bio3graph was applied to a set of 9,586 relevant full text articles, resulting in 137 newly detected relations. The resulting PDS pathway diagram represents a valuable source for further computational modelling and interpretation of omics data.

An incremental variant of the Bio3graph tool was developed to enable easy and periodic updating of a given model structure with new relations from recent scientific literature. The incremental approach was demonstrated on two use cases. In the first use case, a simple PDS network with 37 components and 49 relations, created manually, was extended in two incremental steps yielding the network with 183 relations. In the second use case, a complex PDS model structure in *Arabidopsis thaliana*, consisting of 175 nodes and 524 relations [7], was incrementally updated with relations from recently published articles, resulting in

an enhanced network with 628 relations. The results show that by using the incremental approach it is possible to follow the development of knowledge of specific biological relations in recent literature.

One obstacle in developing simulation models, are scarce kinetic data from which the model parameters could be determined. This problem was addressed by proposing a method for iterative improvement of model parameters until the simulation results meet the expectations of the biology experts. These expectations were formulated as constraints to be satisfied by model simulations. To estimate the parameters of the salicylic acid pathway, the most important PDS pathway, three iterative steps were performed. The method enabled us to optimise model parameters which provide a deeper insight into the observed biological system. As a result, the constraint-driven optimisation approach allows for efficient exploration of the dynamic behaviour of biological models and, at the same time, increases their reliability.

3 Conclusion

The main results of this thesis are: a new methodology for constructing biological models using the expert knowledge and literature and a PDS model, which was built by applying this methodology. Most notably, the standard approach to constructing dynamic models was upgraded with the following methods: a method for model structure revision by means of natural language processing techniques, a method for incremental development of biological model structures and a method for constraint-driven parameter optimisation.

The thesis also contributes to publicly available biological models and scientific software. The PDS model structure of *Arabidopsis thaliana* in the form of directed graphs is publicly available. Also, the Bio3graph approach is implemented and provided as a publicly accessible scientific workflow.

References

[1] Z. Zhao, J. Xia, O. Tastan, I. Singh, M. Kshirsagar, et al. (2011) Virus interactions with human signal transduction pathways, *International Journal of Computational Biology and Drug Design*, 4: 83 - 105.

[2] D. Miljkovic (2014) *Semi-automated knowledge elicitation for modelling plant defence response*, PhD Thesis, IPS, Jožef Stefan, Ljubljana, Slovenia.

[3] P. E. Staswick (2008) Jazing up jasmonate signaling. *Trends in Plant Science*, 13, 66-71.

[4] T. Genoud, M. B. T. Santa Cruz, J. P. Métraux (2001) Numeric simulation of plant signaling networks, *Plant Physiology*, 126, 1430-1437.

[5] M. Naseem, N. Philippi, A. Hussain, G. Wangorsch, N. Ahmed, et al. (2012) Integrated systems view on networking by hormones in arabidopsis immunity reveals multiple crosstalk for cytokinin, *Plant Cell*, 24, 1793-1814.

[6] A. Devoto, Turner, J. G. (2005) Jasmonate-regulated arabidopsis stress signalling network, *Physiologia Plantarum*, 123, 161-172.

[7] D. Miljkovic, T. Stare, I. Mozetič, V. Podpečan, M. Petek, et al. (2012) Signalling Network Construction for Modelling Plant Defence Response, *PLoS ONE*, 7(12):e51822.