

Mirjam Sepesy Maučec*, Gregor Donaj

Metrike za evalvacijo strojnih prevodov

POVZETEK

Spremljanje napredka strojnega prevajanja je neločljivo povezano z ocenjevanjem kakovosti prevodov. Kako oceniti kakovost prevodov, ki jih je ustvaril program? Oceno lahko da usposobljen ocenjevalec. Kljub relativno dobro razviti metodologiji, ki jo v predavanju na kratko predstavimo, pa je tako ocenjevanje še vedno subjektivno, hkrati pa tudi drago, še posebej če želimo s sprotnim ocenjevanjem spremljati napredek strojnega prevajanja. Cenovno ugodnejšo alternativo predstavljajo metrike avtomatskega ocenjevanja kakovosti. V predavanju bomo predstavili nekaj v praksi najbolj uveljavljenih metrik. Posvetili se bomo tudi nekaterim izboljšavam, ki gredo v smeri ocenjevanja prevodov v pregibnem jeziku, med katere sodi tudi slovenščina.

Strojni prevod je le redko uporabljen kot končni produkt, ampak je le vmesna faza prevoda, ki ga prevajalec nato popravi do želene kakovosti. Temu postopku pravimo popravljanje strojnih prevodov. Njegova osnovna ideja je pohitritev in pocenitev prevajalskih postopkov. Metrike avtomatskega ocenjevanja kakovosti prevodov si bomo ogledali tudi v luči popravljanja strojnih prevodov. Predstavili jih bomo kot cenilke napora, potrebnega za popravljanje strojnega prevoda do želene kakovosti.

Strojno prevajanje je/bo orodje sedanjosti/prihodnosti, ki lahko prevajalcu olajša delo, le sprejeti ga mora.

Ključne besede: strojno prevajanje, ročna evalvacija, avtomatska evalvacija, metrika.

1. Uvod

Ljudje v današnjem času ustvarjamo velike količine digitalnih vsebin. Po nekaterih ocenah smo samo v zadnjih nekaj letih ustvarili kar 90 % vseh vsebin, ki so nam na voljo. Črpati informacije iz takšne količine podatkov postane za posameznika hitro neobvladljivo. Potrebuje računalniška orodja, ki so mu pri tem v pomoč [1, 2]. Eno od takih je orodje za strojno prevajanje, ki je še posebej dobrodošlo, če jezika, v katerem je zapisana vsebina, ne poznamo. Po drugi strani pa je lahko uporabno orodje tudi za prevajalca, saj je z njegovo pomočjo v mnogih primerih prevajanje hitrejšo.

Začetki strojnega prevajanja segajo v petdeseta leta preteklega stoletja, ko so v podjetju IBM javno predstavili prvi sistem strojnega prevajanja. Takrat so verjeli, da bo problem strojnega prevajanja rešen v treh do petih letih. To se je kmalu pokazalo za preveč optimistično napoved, saj je več kot pol stoletja zatem strojno prevajanje še vedno intenzivno raziskovano znanstveno področje, ki še ni ponudilo končne rešitve problema, lahko pa sledimo vztrajnemu napredku, ki je pripeljal strojno prevajanje tudi do komercialne uporabe.

2. Evalvacija strojnih prevodov

Razvoj strojnega prevajanja je neločljivo povezan z ocenjevanjem kakovosti prevodov [3]. Kako dober prevod potrebujemo, je odvisno od konteksta uporabe. Če je prevod namenjen končnemu uporabniku, mora praviloma biti brez napak. Če je pa prevod le vmesna faza v nekem postopku obdelave dokumenta, kot na primer v postopku klasifikacije dokumentov po tematiki, je dovolj že informativni približek originala, ki brez škode lahko vsebuje določene slovnične napake.

Zakaj in kdaj ocenjujemo kakovost prevodov? Evalvacija je potrebna, če želimo med seboj primerjati več sistemov za strojno prevajanje in izbrati nam ustrežnejšega. Evalvacijo izvajamo tudi med razvojem novega sistema, da ocenimo izboljšavo ali zaznamo morebiten padec v kakovosti prevajanja.

Pri razmišljanju, o tem kako evalvirati prevode, se srečamo z različnimi problemi. Pravilnih prevodov je več in ali so vsi »enako dobri«? Kako oceniti dva delno pravilna prevoda, če oba vsebujeta napake, ki se med seboj razlikujejo? So ene napake večje od drugih?

2.1 Ročna evalvacija

Zgodovinsko gledano se je najprej uveljavila ročna evalvacija prevodov, ki so ji, predvsem prevajalci, najbolj zaupali. Pri ročni evalvaciji strokovnjaki ocenjujejo natančnost (ang. adequacy) in pravilnost (ang. fluency) prevodov. Pri natančnosti preverjajo, koliko informacije iz izvornika je ohranjene v prevodu. Pri tem ocenjevalec najprej prebere prevod, nato pa izvornik. Ocene so lahko: 1 (kar pomeni vsa informacija), 2 (večina informacije), 3 (delno) 4 (malo) in 5 (nič). Pri ocenjevanju pravilnosti se ocenjuje slovnična pravilnost in jasnost oz. berljivost prevoda. Ocenjevalec presoja,

*Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru, Koroška c. 46, 2000 Maribor
E-pošta: mirjam.sepesy@um.si, gregor.donaj@um.si

kako naraven je prevod. Pri ocenjevanju pravilnosti ocenjevalec običajno nima na voljo izvirnika. Tudi pri tem ocenjevanju se uporablja skala od 1 do 5, pri čemer 1 pomeni tekoč jezik in 1 nerazumljivo zapisana poved.

Pri ročni evalvaciji iste prevode ocenjuje več strokovnjakov, saj se njihove ocene nemalokrat razlikujejo. Ocene so potem statistično obdelane in podana je končna ocena. Tu pa tiči glavni problem. Takšno ocenjevanje je zamudno in drago, zato so se raziskovalci lotili iskanja cenejše alternative za evalvacijo.

2.2 Strojna evalvacija

Strojno evalvacijo opravi računalnik oz. program. Takšna evalvacija je hitra in poceni, od nje pa pričakujemo, da ima pomen, je konsistentna in pravilna, da torej boljšemu sistemu za prevajanje oz. njegovim prevodom da boljšo oceno.

Pri strojni evalvaciji računalniški program strojni prevod primerja z referenčnim (tj. človeškim) prevodom, ki ga imenuje zlati standard, in na različne načine izračuna stopnjo ujemanja ali stopnjo odstopanja [4].

Najprej se je pri razvoju sistemov za prevajanje uveljavilo računanje natančnosti (ang. precision) in priklica (ang. recall). Natančnost je razmerje med številom pravilno prevedenih besed in dolžino prevoda, priklic pa razmerje med številom pravilno prevedenih besed in dolžino referenčnega prevoda. Slabost tega načina ocenjevanja je, da dobi sistem, ki sicer izbira pravilne besede, a v napačnem vrstnem redu, dobre ocene.

Pri strojnem prevajanju se je v preteklosti uveljavil tudi način ocenjevanja z metriko WER (ang. word error rate), ki izvira iz razpoznavne govora. Temelji na Levenshteinovi razdalji med dvema nizoma. Razdalja je enaka minimalnemu številu potrebnih operacij, da en niz spremenimo v drugega. Operacije pa so: ujemanje, zamenjava, vrivanje in brisanje. Končna ocena je razmerje med številom operacij in dolžino referenčnega prevoda. Pri strojnem prevajanju so metriko dopolnili še z operacijo premik in metriko poimenovali TER (ang. translation error rate).

BLEU je metrika, ki temelji na ujemanju besed in zaporedij besed različnih dolžin. Govorimo o n-gramih, na primer 3-gram je zaporedje treh besed. Metrika vpeljuje tudi kazni za prekratke prevode. Končna ocena je uteženo geometrijsko povprečje delnih ocen. Metrika BLEU je bila prva metrika, ki je omogočala primerjavo z več referenčnimi prevodi. V končni oceni je upoštevana ocena glede na „najbližji“ referenčni prevod.

Pri ocenjevanju prevodov v pregibnih jezikih, med katere sodi tudi slovenščina, se je kot najbolj primerna izkazala metrika METEOR. Kot delni oceni uporablja natančnost in priklic in upošteva le nivo unigramov. Tudi METEOR omogoča primerjavo z več referenčnimi prevodi. Njegova največja prednost pa je, da pri ujemanju upošteva tudi ujemanje v korenu besede in sinonime.

3. Popravljanje strojnih prevodov

Strojni prevodi običajno vsebujejo napake. Nekatere med njimi lahko prevajalec enostavno popravi, določene napake pa od prevajalca zahtevajo več kognitivnega napora. Uporaba strojnega prevajanja je smiselna le, če je popravljanje strojnega prevoda (ang. post-editing machine translation, PEMT) manj

zamudno od prevajanja »od začetka«. Kdaj lahko to pričakujemo? Izkazalo se je, da je še najbolj odvisno od jezikovnega para, v katerem prevajamo, pa tudi od smeri prevajanja, saj je za isti jezikovni par prevajanje v eni smeri lahko mnogo boljše kot v drugi. Na primer prevajanje iz slovenščine v angleščino daje neprimerno boljše rezultate kot prevajanje v obratni smeri [5]. Kvaliteta strojnih prevodov je odvisno tudi od učnega gradiva, ki je bilo uporabljeno pri gradnji prevajalnika, saj večina prevajalnikov temelji na statističnih analizah učnih gradiv. Nenazadnje pa ima pri učinkovitosti strojnega prevajanja pomembno vlogo zahtevnost besedila. Manj zahtevno je besedilo, boljši je prevod in potemtakem zahteva manj popravljanja.

4. Strojno prevajanje v prevajalskem poklicu

Prevajalci se danes soočajo z izredno kratkimi roki, v katerih morajo prevesti dokumente, ki so se običajno pripravljali in pilili več tednov, prevod pa potrebujejo »takoj«. Pritisk je velik, odgovornost prav tako. Strojno prevajanje lahko pospeši prevajalske procese. Seveda ne vedno, odvisno je od zahtevnosti vsebine. Če je le-ta dovolj podobna vsebinam, na katerih se je učil prevajalnik, je lahko prevod zelo dober, če pa močno odstopa, je lahko strojni prevod neuporaben.

Zanimalo nas je, kako je strojno prevajanje sprejeto v prevajalskih krogih. S tem namenom smo pobrskali po forumih in našli tako negativne kot pozitivne komentarje. Navedimo najprej nekaj negativnih:

- *»Prevajalec, ki ima rad svoj poklic in z veseljem prevaja besedila, mora zavrniti vsakršno povezavo s stojnim prevajanjem.«*
- *»Strojnemu prevajanju se upiram kolikor je le mogoče, saj je prevajanje moja strast.«*
- *»Zavračam popravljanje strojnih prevodov. "Braincrafted translations only, PEMT nein danke!"«*
- *»Nikakor ne nameravam učiti stroja, kako naj me nadomesti.«*
- *»Dobro bi bilo, ko bi vsi prevajalci ustrezno spoštovali svoj poklic.«*

Iz komentarjev je razvidno, da je eden od vzrokov za negativne odzive strah pred ogroženostjo prevajalskega poklica. Manj je komentarjev, ki bi kazali na neuporabnost strojnega prevajalnika.

Poglejmo še pozitivne komentarje:

- *»Ko imam na voljo malo čas in besedila niso zahtevna, mi je strojno prevajanje v veliko pomoč, saj mi popravljanje strojnih prevodov vzame manj časa.«*
- *»Strojno prevajanje uporabljam že 15 let. Premišljena uporaba mi prihrani veliko tipkanja.«*
- *»Strah, da bi prevajalniki nadomestili človeka, je odveč. Lahko so koristen pripomoček prevajalcu.«*
- *»Iz izkušenj lahko povem, da mi strojno prevajanje razbremeni misli. Posledično delam hitreje in se lahko bolj skoncentriram na zahtevnejše odseke prevoda.«*

Z gotovostjo trdimo, da je strojno prevajanje orodje prihodnosti, s pomočjo katerega bodo prevajalci urejali

prevedene dokumente in jih slogovno ter pomensko prilagajali značilnostim ciljnega jezika. Predvsem zato, da bi lahko prevajalci v prihodnosti ugodili hitrim in vse obsežnejšim zahtevam po kakovostnem prevajanju, bodo morali najti način, kako strojno prevajanje vključiti v svoje delovne procese. Spreminjanje delovnih procesov zahteva čas, nova znanja in predvsem potrpljenje. Šele ko je nov delovni proces utečen, se pokažejo njegove prednosti.

5. Zaključek

Prihodnost prevajalskega poklica je vsekakor prepletena s tehnologijam [6]. Strojni prevajalniki profesionalnih prevajalcev ne bodo izpodrinili. Eden od glavnih razlogov je tudi ta, da prevajalnik ne nosi odgovornosti za kvaliteto in predvsem za vsebinsko ustreznost prevoda. Tudi prevajanje leposlovja je in bo v domeni prevajalcev. Pri prevajanju tehnične dokumentacije, pravnih besedil ipd. pa je strojni prevajalnik lahko koristen pripomoček, ki prevajalca »odreši« suhoparnega in ponavljajočega prevajanja istih oz. podobnih povedi. Nenazadnje je prevajalnik dobrodošlo orodje vsakemu, ki potrebuje informativne približke prevodov iz jezika, ki ga ne razume in jih potrebuje takoj.

Literatura

1. Uszkoreit, Hans: Language Technology. A First Overview. Dostopno na spletnem naslovu: <http://www.dfki.de/~hansu/LT.pdf>, [28.3.2016].
2. Krek, Simon: Slovenski jezik v digitalni dobi, Zbirka bela knjiga, Springer-Verlag Berlin Heidelberg, 2012.
3. Han, A.L.F.; Wong, D.F. Machine Translation Evaluation: A Survey. *arXiv preprint arXiv:1605.04515*, 2016.
4. Verdonik, D.; Sepesy Maučec, M. O avtomatski evalvaciji strojnega prevajanja. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*, 2013, 1(1), 111-133.
5. Sepesy Maučec, Mirjam, Kačič, Zdravko. Statistical machine translation from Slovenian to English. CIT, ISSN 1330-1136, Mar. 2007, vol. 15, no. 1, str. 47-59.
6. Vintar, Š. Sodobne prevajalske tehnologije in prihodnost prevajalskega poklica. *Uporabna informatika*, 2013, 21 (4), 221-227.