

Evaluation of Mgr. Nathan Green's doctoral thesis:

"Improvements to Syntax-based Machine Translation using Ensemble Dependency Parsers"

(referee: Dr. phil. Eckhard Bick)

Research topics, methodology and results

The thesis belongs in the subject areas of Dependency parsing, Treebanks, Machine Learning (ML) and Machine Translation (MT), to all of which the author makes innovative contributions. Specifically, Nathan Green uses ML techniques to achieve improved dependency trees for Deep Transfer MT. That Dependency parsing algorithms, tree depth and part-of-speech (POS) error distribution are evaluated not just as such, but through the optics of indirect MT improvements, in a modular MT chain setup, is one of the methodological innovations presented in the thesis. The author's direct contribution to Dependency parsing is three-fold: First, he presents a method to achieve improved training data by better conversion of constituent trees to dependency trees. Second, he shows that parsers differ considerably more with regard to the POS and attachment errors they make, than their overall performance differences would suggest. The third, and arguably most important, contribution builds upon the second, and concerns a technique called Ensemble parsing, where different dependency parsers (or different parameter settings of only one or two parsers) are combined into a compound system weighted according to error types and gold standard recall. Nathan Green has previously published in this area, with a focus on under-resourced languages (Tamil and Indonesian), but in his thesis, standard English training and test suites from ML joint task conferences are also used, making his findings comparable to the state of the art. The best results were achieved with parser weighting based on UAS values raised to the 10th power. This is an interesting empirical finding, and will probably prompt other researchers in the field to find corresponding optimal exponent values for various languages and genre-specific data sets. Another creative contribution within Ensemble parsing, demonstrated for Tamil, is an SVM-based meta-classifier that solely makes use of model agreement as a feature, making the system virtually language-independent and, hence, valuable for under-resourced languages. This sparse-data advantage is further compounded by successful experiments with self-training (for Indonesian), departing from as little as 100 hand-annotated sentences as initial training data. Finally, and also in connection with SVM Ensemble parsing, the author shows that contrary to what one might expect, more training data for the base parsers - though reducing the risk of out-of-vocabulary words - is not always a good idea, since it may leave too little tuning data

for the SVM.

Form

As an empirical study, the text has to reconcile the duality of data presentation on the one hand, and theoretical-analytical discussions on the other. This is achieved by a clear structuring of the text. where a presentation of state-of-the-art tools and methods is used as an introduction, followed by the author's own innovations, and finally a description of the experiments and evaluation of the results. The thesis is written in clear and idiomatic English, and generally error-free, though there are some minor errors that could easily have been spotted by proof-reading, such as "manor"/"manner" (p2), unfinished sentences "from different. (p4), "similarly as" (p22), and other small oddities "an complete/a incomplete" (p35), "seems to of had" (p.67)- While the latter do not disturb the reading process, there are some inconsistencies between tables and text, that are probably due to repeated results updates but can potentially confuse the reader. For instance, p.38 states that the 5-parser UAS10 combination gives the 2nd best rating, while it comes out on top in table 3.3, with a value that is also higher than what is declared the top score (on p.49) for fuzzy clustering. Similarly, text (p.41) and table (3.5) disagree about the average relative error rate (8.64% vs. 7.79%), or the cluster weighting for CC on p.48. However, these are minor problems compared to the scale of the work, and the overall impression is quite positive, and there is no doubt that the experiments have been carefully prepared, described and executed. Thus, it is at all times possible for the reader to understand how the author arrives at his conclusions.

Methodology

The methodology used is empirical (corpus-based) and statistical, and in a strict sense results are thus "true" or "false" only at given level of statistical significance and relational strength. That Nathan Green is well-versed in this methodology, can be seen from a number of factors - for instance, he uses large numbers of random samples, and he consistently provides (Wilcoxon) statistical significance values, or even evaluates parameters directly in terms of significance thresholds (table 3.18). He is also aware of the importance of data set sizes for this method, and proposes methods, such as self-training and his language-independent model agreement classifier, to offset sparse data problems. On the other hand, Nathan Green is aware of non-statistical issues, too, not least the importance of linguistic structure. This becomes apparent where he discusses improving np tree depth, or looks for carry-over effects (on MT quality) from improved individual tags and dependencies, independently of average/statistical tagging performance.

Research contextualisation

In each chapter, the author provides the necessary introduction to the topic's research context, and literature references are provided in satisfactory detail, with 7 pages of bibliography, which constitutes a relatively large section considering the total size of the thesis. Throughout, the use and placement of such references, as well as general background information such as explanations and definitions, documents that the author is sufficiently oriented in his field of research.

Conclusion

All in all, I have a clearly positive impression. It is my opinion that Nathan Green's thesis is an interesting, well-researched and methodologically sound piece of work. In particular, it proves that Nathan Green is capable of creative scientific work, and also of meaningfully relating and integrating neighbouring fields such as treebanking, dependency parsing and machine translation, in a unified methodology. Both methods and findings are presented in sufficient detail for other researchers to verify and build upon Nathan Green's work.

Aarhus, August 8th, 2013

A handwritten signature in black ink, appearing to read 'E. Bick', written in a cursive style.

Dr. phil. Eckhard Bick
Institute of Language and Communication
University of Southern Denmark