Diploma Thesis Review

Thesis title: Combining Outputs from Machine Translation Systems
Thesis author: Fahim A. Salim

Thesis description

The aim of the work under review was to design, implement and evaluate a new system for combining sentence translations resulting from diverse Machine Translation (MT) systems. This aim is based on the hypothesis that different MT systems make different types of errors and thus there should be a chance to reach a better translation by concatenating the most promising pieces of sentences produced by the individual systems.

The thesis is structured as follows. Chapter 1 gives a motivation for combining MT outputs and presents a brief classification of current MT systems. Chapter 2 explains the distinction of the so called glass box and black box approaches to MT combination. Chapter 3 lists the previously existing software tools employed in the thesis. Chapter 4 mentions the data used for experiments (especially for training language models) and the MT systems whose outputs are used. Chapter 5 focuses on sentence-level output combinations, while Chapter 6 uses confusion network of aligned sentence pieces, so that shorter sentence fragments can be utilized too. Conclusions follow in Chapter 7. The total number of pages is x+56. A CD-ROM containing source codes and experimental data is attached.

Comments

As it is clear from the thesis text, the author understands the main principles of the contemporary approaches to the task of MT combination (aligning the translation hypotheses, scoring word subsequences, search for the optimal path through the network). He was able to construct an experimental combination systems mostly by utilizing several existing standard statistical NLP tools and gluing them together using Perl and bash. He designed and evaluated a set of experiments and drew some conclusions from the results.

The author evaluated a number of possible combination configurations, differing in the main strategy (sentence selection vs. confusion network decoding), in the set of employed features (language model score, number of common trigrams, number of words...), and in corpus sets used for training language models.

Two of the experimental configurations outperformed the natural baseline, which is choosing always the output of the system which is the best in average. This achievement is the main positive result of the thesis.

There are numerous language mistakes that remained in the text, however, the overall quality of the thesis has been significantly improved compared to the version submitted for defense in autumn 2010.

Conclusion

I recommend to accept the thesis for the defense.

In Prague, 24th January, 2011

Zdeněk Žabokrtský
Institute of Formal and Applied Linguistics
Charles University in Prague