Master's Thesis Review

Author: Fahim A. Salim
Title: Combining outputs from machine translation systems
Opponent: RNDr. Ondřej Bojar, Ph.D.

The thesis presented by Fahim A. Salim aims at combining outputs of several MT systems. Despite the original assignment of English-to-Czech translation, the author reverses the translation direction. In my opinion, this significantly changes the specifics of the task.

The overall structure of the text is better than in the version submitted previously. Still, a lot of space is given to the overall motivation for black-box MT system combination while the experiments selected are not motivated enough. For instance, why the author has evaluated Cn_noskel_12_34_56 and not e.g. Cn_noskel_13_25_46 in Section 6.3.3, or why 2, 3 and 6 systems were combined in Section 6.4.3 and not 3, 4 and 6 as in the previous sections?

Despite the complaints in my review of the previous submitted version, there are still many details missing in the description so that the experiments are not reproducible:

- Section 5.4.1: "...feature scores are scaled and normalized in order to make them more consistent with each other". How exactly is this "scaling and normalization" performed (a formula is needed), and how is the "scaling" different from the "normalization"? What does the author mean by "more consistent with each other"? What weights did the author eventually use?
- Section 6.3.1: "...based on those alignments the Confusion Network was created." How exactly? Specifically, how unaligned words of the skeleton and unaligned words of the hypothesis are treated? An illustration is necessary for the default technique as well as for the technique called "_withempty". (There is some description of "Cn_6_withempty" but it is rather incomprehensible.)
- Section 6.3.3: Insufficient description of how the (mutually aligned) pairs of hypotheses are combined to a single CN. The author says only: "...CN was created by joining those learnt alignments...". Again, an illustration is needed.

I have reservations towards the two critical parts of the text: the abstract and the conclusion. The abstract provides a long motivation but very little of what the thesis is actually about and no summary of findings at all. Admittedly, the results need not be presented in the abstract. The conclusion is overly optimistic with respect to sentence-based techniques. (Only very few setups and with a very restricted set of systems succeeded. Moreover, the author does not use any method of confidence checking to test the significance of the improvements.) On the other hand, "general purpose tools" for CN combination are criticized while they arguably work for other researchers, as obvious from the cited literature. Perhaps the author just failed to apply them successfully...

The style of the thesis remains rather poor in many respects, despite some improvement. There are still many typos, errors in the grammar (most notably subject-predicate agreement in number), and some sentences or whole paragraphs are incomprehensible (e.g. "A word lattice is defined as a acyclic directed graph with a single starting point and edges labeled with a word or node and it may also has a weight associated with it." on p. 39).

Conclusion

The presented thesis unfortunately remains below the standard of M.Sc. theses at Charles University. I acknowledge that the author has proposed and evaluated new methods for system combination but too often he fails to present his proposals and findings clearly enough (detail of documentation, language and style).

With reservations, I recommend the thesis to be accepted, provided that the author explains and illustrates the three procedures listed above (normalization and the two techniques of CN construction) during or after his defense presentation.

Prague, January 27, 2011.

RNDr. Ondřej Bojar, Ph.D.
Charles University in Prague, ÚFAL