Discrimantive Methods for Training Translation Models

Advanced Machine Translation Seminar

Abhaya Agarwal

I am interested in exploring the recent body of work introducing fully discriminative models for Machine Translation. In contrast to the more prevelent approach which uses a handful of sub-models learned in a generative framework (translation lexicons learned using GIZA++, phrase extraction based on the alignments, language models), here everything is placed in a one big translation model which can be trained against different loss functions. Performance of these systems is yet to catch up with the current MERT based aproaches and one of the biggest problems remain that it is hard to scale these methods to large corpora.

Following can be a rough guideline:
- As back ground, brief mention of the maximum entropy approach to MT and introduction to MERT [1,2]
- Series of papers by C Tillmann and T Zhang where they propose a discrimative approach based on block orientation prediction [3]
- Paper by P Liang, A Bouchard-Côté, D Klein, B Taskar in ACL 2006 which explores various update strategies for discrimative training [4]
- Paper by A Ittycheriah, S Roukos where they discuss direct translation models [5]

Going ahead, I would like to look into issues of
- What are the possible ways of solving the scaling up issue?
- Given that the amount of data available is so much, will the hybrid models always beat purely discriminative models?
- Most of the papers employ a loss function based on BLEU. Are there other better and fast enough loss functions?

References:

1. Och, F.J., Ney, H.: Discriminative training and maximum entropy models for statistical machine translation. In: Proc. of the 40th Annual Meeting of the Association for Computational Linguistics (ACL). (2002)
2. Franz Josef Och, Minimum error rate training in statistical machine translation, Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, p.160-167, July 07-12, 2003, Sapporo, Japan
3. Christoph Tillmann , Tong Zhang, A discriminative global training algorithm for statistical MT, Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL, p.721-728, July 17-18, 2006, Sydney, Australia
4. Percy Liang, Alexandre Bouchard-Côté, Dan Klein, Ben Taskar. An end-to-end discriminative approach to machine translation. International Conference on Computational Linguistics and Association for Computational Linguistics (COLING/ACL), 2006.
5. Abraham Ittycheriah and Salim Roukos . Direct Translation Model 2. In proceedings of NAACL 2007, Rochester, 2007