NLP 2 - Project 1

April 6, 2017

This project will help you familiarise yourself with word-based models. Word-based models remain at the core of today's SMT systems in the form of alignment models. You will implement the simplest (though still widely used) word-based models, namely, IBM model 1, a lexical translation model, and IBM model 2, which models an impoverished form of word alignments.

In summary, your task is to

- Implement IBM model 1;
- Implement IBM model 2 using a jump distribution as in Vogel et al. (1996);
- Experiment with maximum likelihood estimation and variational Bayes;
- Write a technical report where you present the models and an empirical comparison. Your report should also present learning curves where applicable along with a discussion explaining aspects such as non-convexity, stability and convergence.

1 IBM model 1

1. a) Implement EM training (Brown et al., 1993) for IBM model 1;
   b) Implement variational inference for Bayesian IBM model 1;
   c) All of the tasks below should be performed for both models.
2. Plot the evolution of the log likelihood function (or ELBO) as a function of the iteration. How to compute the Dirichlet par of the ELBO is described here (I'll add the link tonight when I come home);
3. Plot the evolution of alignment error rate (AER) on validation data as a function of the iteration;
4. Experiment with two criteria for model selection (i.e. deciding on number of training iterations): 1) convergence in terms of log likelihood; 2) best AER on validation data;

5. For the selected modes, obtain Viterbi alignments for every sentence pair in a test corpus and compute AER using a gold-standard provided by the assistant;

2 IBM model 2

1. Extend your previous model by implementing a full IBM model 2 (Brown et al., 1993), however using the cheaper parameterisation in terms of jumps;
2. IBM 2 is non-convex, thus you will see that optimising the log-likelihood function is not as trivial as in the case of IBM model 1, particularly, convergence will depend on how you initialise the model parameters, you will try
   - uniform initialisation
   - random initialisation (try 3 different starting points)
   - initialise the lexical parameters using the output of a complete run of model 1
3. Plot the log-likelihood function as a function of the iteration for all these cases
4. Plot validation AER as a function of the iteration for all these cases
5. Select two models: 1) one in terms of log likelihood, 2) another in terms of validation AER;
6. Compare the selected models to IBM model 1 in terms of AER in the test set.

3 Data

You will be provided with parallel data taken from the Canadian Hansards (parliament proceedings). The data consists of preprocessed sentence pairs (please do not further preprocess the data). There are two files, one for the English and one for the French sentences. Sentences with the same line number are translations of each other.

4 Report

You should use latex for your report, here you will find a template: `http://acl2017.org/downloads/acl17-latex.zip` (unlike the template suggests, your submission should not be anonymous).

We expect short reports (4 pages plus references) written in English covering at least the following:

- abstract;

- introduction: discuss the problem and some background where relevant;
- model: technical description of models;
- experiments: details about the data, experimental setup and findings;
- conclusion.

References

Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., and Mercer, R. L. (1993). The mathematics of statistical machine translation: parameter estimation. *Computational Linguistics*, 19(2):263–311.

Vogel, S., Ney, H., and Tillmann, C. (1996). HMM-based word alignment in statistical translation. In *Proceedings of the 16th Conference on Computational Linguistics - Volume 2*, COLING '96, pages 836–841, Stroudsburg, PA, USA. Association for Computational Linguistics.