Large-scale Reordering Models for Statistical Machine Translation

Abdullah Alrajeh

PhD Viva
May 12, 2015
University of Southampton

Outline

Objective

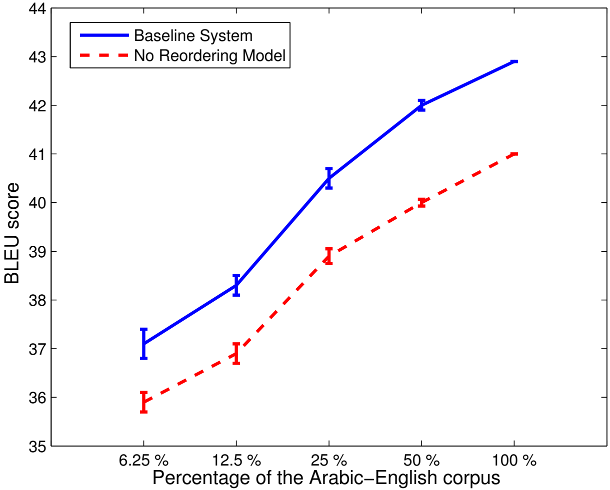
Language model is not sufficient:

- It has a bias towards short translations (Koehn, 2010).
- It does not consider the dependence of the target sentence on the source sentence.

Efficient Reordering Models

A large-scale parallel corpus is important for improving such reordering model, this improvement comes at a price of computational complexity. It is particularly pronounced when discriminative models are considered such as maximum entropy-based model.

Objective



Baseline System
No Reordering Model
BLEU score
44
43
42
41
40
39
38
37
36
35
6.25 %
12.5 %
25 %
50 %
100 %
Percentage of the Arabic–English corpus

Contributions

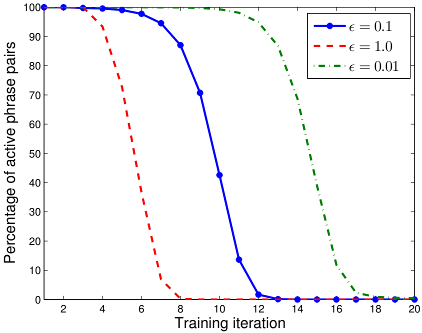
Bayesian Reordering Model

I explore a generative learning approach to phrase reordering namely naive Bayes. My Bayesian model using a Dirichlet prior is shown to be superior to the lexicalized model of estimating probabilities as relative frequencies of phrase movements. The training time of my model is as fast as the lexicalized model and its storage requirement is many times smaller.

Reordering Model	Disk Storage	Accuracy	BLEU Score
Lexicalised	5.9 GB	75.9	43.8
Bayes-MAP estimate	**0.1 GB**	**77.4**	**44.0**
Bayes-Baysien inference	**0.1 GB**	77.3	**44.0**

Dual Multinomial Logistic Regression

I proposed training a reordering model using the dual MLR with a shrinking method. It is a couble of times faster than MaxEnt and more memory-efficient.



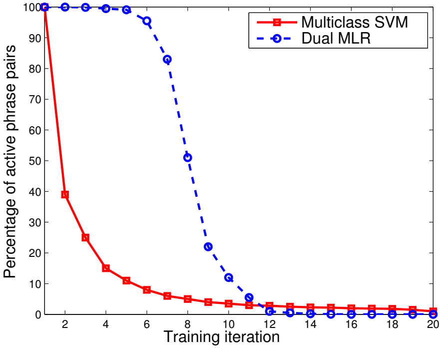
Reordering Model	Train Time	Accuracy
L-BFGS MaxEnt	12.2 h	81.3
Dual MLR ϵ:0.1	**3.4 h**	81.2

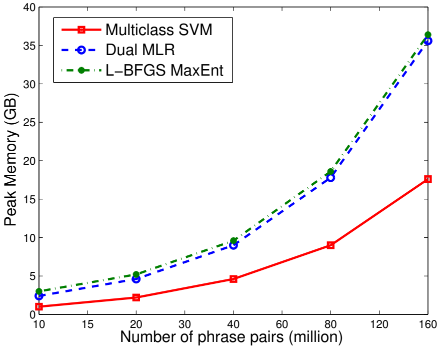
Reordering Model	Disk Storage	BLEU Score
Lexicalised	5.9 GB	43.8
L-BFGS MaxEnt	**0.8 GB**	**44.5**
Dual MLR ϵ:0.1	**0.8 GB**	**44.5**

Contributions

Multiclass SVM

Using dual coordinate descent methods for learning, I provide a mechanism to shrink the amount of training data required for each iteration (**4-fold speed, 50% memory reduction**).





Contributions

Efficient solver for linear SVM

My proposed algorithm is faster than LIBLINEAR, a state-of-the-art solver. My shrinking heuristic could reduce up to 99% of some data sets from the first iteration. It is very useful when the training data cannot fit in memory.

Word Alignment Features

I found that a reordering model based on alignments features is more compact (80% reduction) than using boundaries features without losing translation quality.

Contributions

Ni's Approach is not Necessary

Ni et al. (2011) proposed to break down the learning complexity into small sub-models. I found that having one reordering model is more beneficial to a machine translation system.

Reordering Model	Disk Storage	Accuracy	BLEU Score
Lexicalised	5.9 GB	75.9	43.8
Multiclass SVM (S7)	**0.1 GB**	**79.3**	**44.3**
Perceptron sub-models (S7)	1.1 GB	77.7	43.2

Contributions

Ordinal Regression

I found ordinal regression empirically not useful. The classifier is **restrictive** due to the requirement of parallel discriminant hyperplanes. This leads to a **low accuracy** compared to SVM and Multinomial Logistic Regression.

Voted Spheres

Its **non-parametric** nature is attractive. I found that the classifier has high accuracy. However, it is **very slow during prediction** phase (i.e. 100 times or more slower than competing classifiers) which makes it not practical for machine translation systems.

Future Research

Structured Learning

Feng et al. (2012) propose a reordering model based on sequence labelling techniques. For a sentence pair, they assign each source word an orientation label. Then they let an algorithm such as CRFs or RNN to do the learning task. The computational costs are high (around 16 hours and 120G RAM).

Sparse Reordering Features

Cherry (2013) proposed using sparse features optimise BLEU with the decoder instead of training a classifier independently. Recently, Auli et al. (2014) train a discriminative reordering model with millions of sparse features on a large- scale corpus.

Publications and Software

Article

Scalable reordering models for SMT based on multiclass SVM. The Prague Bulletin of Mathematical Linguistics, 2015.

Conference Papers

Large-scale reordering model for statistical machine translation using dual multinomial logistic regression. In, EMNLP 2014.

Bayesian reordering model with feature selection. In, ACL 2014: WMT.

Memory-efficient large-scale linear support vector machine. In, ICMV 2014.

Generative and discriminative reordering models for statistical machine translation. In 8th SSC, Imperial College Press.

Software

Large-scale Classification Tool (LCT)

Scalable Reordering Models (SRM)