

Statistical Machine Translation without Parallel Corpora

Person
University of Awesome
Location
person@email

Person
University of Awesome
Location
person@email

Person
University of Awesome
Location
person@email

Abstract

1 Introduction

2 Related work

SMT state of the art: phrase-based (?), hierarchical (?), etc. Will review the phrase-based pipeline in greater detail in Section 3, and describe our extensions in Section 4. For a recent thorough review of statistical machine translation we refer the reader to (?).

A lot of related work in lexicon induction (e.g. (?; ?; ?; ?; ?)). Although often motivated by the resource constrained machine translation, it was never used in MT. In this work, we make direct use of these methods in the machine translation pipeline.

3 Background

Review the MT pipeline, emphasizing: phrase scoring and ordering.

4 Reducing the Parallel Data requirement / Estimating Parameters from Monolingual Data

4.1 Phrase extraction

If we get it to work.

4.2 Phrase scoring

Describe how we compute monolingual similarity instead of estimating scores from alignments.

4.3 Reordering

TODO: Reordering figure

Describe the algorithm for estimating orientation probabilities. Talk about the issue of too much weight on out-of-order orientation.

5 Experiments

5.1 Data

Describe data we use in the experiments: Europarl (?), Gigaword, our own crawls¹.

5.2 Single language

1. *Phrase features.* (a) Augment phrase scores with mono features. If we see better performance, reduce the amount of parallel data until it matches the performance of the original system. Make the tradeoff argument. (b) (**lesion experiments**) See how well we do with mono features alone.
2. *Orientation features.* Use mono orientation features.
3. *Induce phrase table.*
4. *Put everything together.* Run the entire pipeline.

5.3 Big experiment

Now, run the entire pipeline on a handful of languages extracting monolingual features from the Gigaword and our crawls.

6 Discussion

7 Conclusions and Future Work

First to make use of plentiful monolingual data to reduce the dependence on expensive parallel data. In particular:

- Showed that augmenting standard pipeline with monolingual features helps.
- Demonstrated that monolingual features are informative enough on their own for a competitive system.

¹Promise to distribute after publication.

- Proposed an algorithm for estimating orientation probabilities from monolingual data alone.
- Build complete systems for X low-resource languages.