The University of Washington Machine Translation System for IWSL 2009

Mei Yang, Amittai Axelrod, Kevin Duh, Katrin Kirchhoff

Dept. of Electrical Engineering, University of Washington, Seattle, WA, USA

{yangmei, amittai, duh, katrin}@ee.washington.edu

1 Overview

- We participated in two BTEC translation tasks: Chinese-English and Arabic-English
- Our interests include
  - Different preprocessing schemes for Chinese and Arabic
  - Combination of phrase tables based on different alignments
  - Semi-supervised reranking of N-best lists
  - Sentence-type specific part-of-speech (POS) language modeling for rescoring

2 Baseline translation system

- A state-of-the-art two-pass phrase-based SMT system
- Trained within the Moses development and decoding framework
- A 4-gram Language model trained using the SRILM toolkit

3 Preprocessing schemes

- Chinese segmentation and markup
  - The Stanford segmenter for re-segmenting the Chinese data
  - Character-based segmentation for the Chinese data
  - An in-house tool *Decatur* to markup dates and numbers in both the Chinese and English data
  - A simple tool to markup just numbers in both the Chinese and English data
  - Strip off all punctuations in both the Chinese and English data
  - None of the above schemes led to significance improvement over the original segmentation
- Arabic tokenization
  - The Columbia University MADA and TOKEN tools with two schemes:
    - Split off *w+, f+, l+, b+,* and *Al+*
    - TOKAN's D2 scheme, which does not split off *Al+* but instead separates *s+*
  - The first scheme yielded better performance

4 Phrase table combination

- Phrase tables learned from GIZA++ and MTTK alignments respectively
- The two individual tables were combined into a single table
- Additional binary features to indicate which alignment produced each phrase pair entry
- The combined table outperformed the individual tables in the Chinese-English system

5 Semi-supervised reranking

f: ranking function

P_L: labeled data
Pair-wise samples (x^l, y^l) collected from each N-best list of a held out set, such that x^l ranks higher than y^l

P_U: unlabeled data
Pair-wise samples (x^u, y^u) collected from the N-best list of a given test sentence

$$f^* = \underset{f}{argmin} \sum_{P_L} e^{-(f(x^l) - f(y^l))} + \beta \sum_{P_U} e^{-|f(x^u) - f(y^u)|}$$

- The labeled data were produced using smoothed sentence-level BLEU scores
- The ranking function was learned using a modified RankBoost algorithm
  - Maximize the margins of the labeled and unlabeled data jointly
  - Treats the reranking problem as a problem of binary classification on hypothesis pairs
  - Iteratively train a weak ranker and adjust sample weights according to the classification results
  - The final ranking function is a linear combination of the weak rankers from all iterations
- Applied in the second pass for reranking N-best lists
- For IWSLT 2007 Italian-English and Arabic-English data, it achieved substantial improvements
- For this year data, it improved precision based evaluation metrics, such as PER, TER, WER and Precision, but degraded n-gram based metrics, such as BLEU and NIST

6 Sentence-type specific POS language model

- Captures the syntactic differences between questions and statements
- Determine the sentence type using punctuations in the source sentences
- Applied in the second pass for reranking N-best lists
- Led to a small improvement in the Chinese-English system

7 Official evaluation results

	case+punc	no_case+no_punc
BLEU	0.41	0.40
PER	0.42	0.45
Meteor	0.66	0.62
NIST	7.05	7.30

Table 1: the Chinese-English system

	case+punc	no_case+no_punc
BLEU	0.48	0.48
PER	0.35	0.38
Meteor	0.72	0.69
NIST	6.85	6.93

Table 2: the Arabic-English system