Combining Multi-Engine Machine Translation and Online Learning through Dynamic Phrase Tables

Rico Sennrich

University of Zurich
Institute of Computational Linguistics

30.05.2011
Overview

Multi-Engine Machine Translation
- Combine output of multiple translation systems
  - Motivation
  - Implementation
  - Results

Online Learning
- In post-editing environment: (partially) retrain system on corrected translation
- Similar implementation as multi-engine MT; results and combination with multi-engine MT
Text+Berg Corpus

- Collection of Alpine texts (publication of the Swiss Alpine Club since 1864)
- Since 1957: parallel edition DE–FR → parallel corpus of 4 million tokens.
- Research project: domain-specific SMT

| System                  | BLEU | METEOR |
|-------------------------|------|--------|
| in-domain SMT system    | 17.18| 38.28  |
| Personal Translator 14  | 13.29| 35.68  |
| Google Translate        | 12.94| 34.36  |

Table: MT performance DE–FR.
## Domain-specific translations

| DE       | Text+Berg                                      | Europarl                               |
|----------|-----------------------------------------------|----------------------------------------|
| Angriff  | tentative ([climbing] attempt)                | attaque (attack)                       |
| Führer   | guide (guide)                                 | dirigeant (leader)                     |
| Pass     | col (mountain pass)                           | passeport (passport)                   |
| Spitze   | pointe (peak)                                 | tête (head [of an organisation])       |
| Vorsprung| ressaut (ledge)                               | avance (lead)                          |
Do we need a full-fledged SMT system for system combination?

- In WMT system combination tasks, approaches that do not consider source text still work well.
- Target side alignment; confusion network decoding with LM
- Examples: MANY [Bar10], MEMT [HL10]

Let’s see if it helps...

- Our observations:
  - In-domain system suffers from data-sparseness (high OOV rate).
  - Out-of-domain and rule-based systems are worse than in-domain system, but have greater lexical coverage.

- Our conclusions:
  - Promising strategy: prefer in-domain system for phrases it knows, and choose other systems otherwise.
  - We hope to profit from source-side information and source-target alignment.
Architecture

- Moses framework
- Primary system trained on in-domain training data
- Translation hypotheses are integrated through additional phrase table (alternative translation path during decoding)
- Optimization with MERT
This architecture is similar to [CEF+07].

image source: Chen et al. (2007): Multi-Engine Machine Translation with an Open-Source (SMT) Decoder. In Proceedings of the Second Workshop on Statistical Machine Translation.
Training secondary phrase table

- Trained on translation hypotheses for sentences to be translated → dynamic (re-)training for any number of sentences
- Word alignment with MGIZA++ (using existing model from primary system)
- Phrase extraction with Moses heuristics
- Features in phrase table: $p(t|s)$; $p(s|t)$, lexical weights $lex(t|s)$; $lex(s|t)$ (and constant phrase penalty)
- Two different scoring methods to obtain feature values: vanilla and modified
vanilla scoring

- Scoring of phrase pairs as implemented in Moses
- Calculations based on Maximum-Likelihood Estimation (MLE)
- Problem: MLE is unreliable if frequencies are low ($\frac{1}{1}, \frac{1}{2}$)

modified scoring

- Add frequencies of primary and secondary corpus
- Secondary corpus has little effect if phrase is frequent in primary corpus: 
  \[ \frac{500}{1000} = 0.5 \text{ vs. } \frac{500+2}{1000+2} = 0.501 \]
- Secondary corpus has large effect if phrase is rare in primary corpus: 
  \[ \frac{1}{3} = 0.333 \text{ vs. } \frac{1+2}{3+2} = 0.6 \]
- Fits our strategy of preferring primary corpus where possible, and considering external hypotheses for rare/unknown words
Systems

- Software from WMT 2010 system combination shared task. Dominant paradigm: output alignment and confusion network decoding
  - MANY (Loïc Barrault) [Bar10]
  - MEMT (Kenneth Heafield) [BL05]
- Concatenation of parallel training corpus and translation hypotheses → slow
- Dynamic - vanilla scoring
- Dynamic - modified (re-)scoring
| Combination System                | BLEU  | METEOR |
|----------------------------------|-------|--------|
| Personal Translator 14           | 13.29 | 35.68  |
| Google Translate                 | 12.94 | 34.36  |
| in-domain SMT system             | 17.18 | 38.28  |
| MANY                             | 18.23 | 39.68  |
| MEMT                             | 18.39 | 39.01  |
| Concat                           | 19.11 | 39.45  |
| Dynamic (vanilla)                | 19.33 | 40.00  |
| Dynamic (modified)               | 20.06 | 40.59  |

*Table*: SMT performance DE–FR for multiple system combination approaches.
Results:
Performance with Varying Phrase Table Size

Figure: SMT performance DE–FR as a function of dynamic phrase table size. Comparison of vanilla scoring and modified scoring.
Multi-Engine MT

- Multi-engine MT gives large performance boost (2.9 BLEU points over best individual system)
- Re-scoring with frequencies from primary corpus is effective:
  - Performance gain over vanilla scoring (0.7 BLEU points)
  - Performance does not degrade if secondary corpus is small
| Source                  | Er ist ein Konditionswunder.  
|                        | He is in miraculous shape.    |
| Reference              | C’est un miracle de condition physique. |
| System 1 (Moses)       | C’est un Konditionswunder.    |
| System 2 (PT 14)       | C’est un miracle de condition. |
| System 3 (Google Translate) | Il est un miracle de remise en forme. |
| Multi-Engine (vanilla) | C’est un miracle de condition. |
| Multi-Engine (modified)| C’est un miracle de condition. |
| Source | Wir konnten das Aussehen der Pässe nur ahnen.  
We could only guess at the look of the mountain passes. |
|---|---|
| Reference | Nous ne pouvions que deviner l’aspect des cols. |
| System 1 (Moses) | nous ne pouvions seulement deviner l’aspect des cols. |
| System 2 (PT 14) | Nous ne pouvions que nous douter de l’air des passeports. |
| System 3 (Google Transl.) | Nous ne pouvions imaginer l’aspect de la passe. |
| Multi-Engine (vanilla) | nous ne pouvions de l’air des cols de la passe. |
| Multi-Engine (modified) | nous ne pouvions l’aspect des cols que deviner. |
Learning from Previous Translations

- In post-editing environment, how can we use previous, corrected translations to improve SMT quality?
- Hardt and Elming [HE10] propose incremental re-training of secondary phrase table.
- → same principle that we used for multi-engine MT.

Implementation

- We simulate approach with reference translations instead of actual post-editing.
- Alignment/scoring as for multi-engine MT - but with previous reference translations instead of translation hypotheses.
- Phrase table is dynamically rebuilt after each sentence.
- No new MERT; instead, both phrase tables use baseline weights.
### Table: SMT performance DE–FR with online learning system.

| System              | BLEU  | METEOR |
|---------------------|-------|--------|
| baseline            | 17.18 | 38.28  |
| vanilla scoring     | 16.81 | 37.61  |
| modified scoring    | 17.57 | 38.60  |
Combination of Multi-Engine MT and Online Learning

| System             | BLEU  | METEOR |
|--------------------|-------|--------|
| baseline           | 17.18 | 38.28  |
| online learning    | 17.57 | 38.60  |
| multi-engine MT    | 19.93 | 40.52  |
| combined           | 20.05 | 40.61  |

Table: SMT performance DE–FR with system combining multi-engine MT and online learning.
Online Learning & Combination

- Online learning led to relatively small performance gain.
- Incremental re-training more effective for texts with high text-internal repetition (Hardt and Elming [HE10], clinical trial protocols: 4 BLEU points increase).
- Combination of multi-engine MT and online learning possible, but no performance gain in this evaluation.
Multi-engine MT simple to implement, and promising for people/companies with little training data.

In-domain system is more than Yet Another Hypothesis

Approach has strong dependence on primary corpus: your mileage may vary

Online learning experiments (and combination of both) were below expectations – not necessary failure of technique, but applied to wrong corpus.
Multi-engine MT simple to implement, and promising for people/companies with little training data.

In-domain system is more than Yet Another Hypothesis

Approach has strong dependence on primary corpus: your mileage may vary

Online learning experiments (and combination of both) were below expectations – not necessary failure of technique, but applied to wrong corpus.

Thank you for your attention!
Barrault, Loïc: MANY: Open source MT system combination at WMT'10.
In Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR, pages 277–281, Uppsala, Sweden, July 2010. Association for Computational Linguistics.
http://www.aclweb.org/anthology/W10-1740.

Banerjee, Satanjeev and Alon Lavie: METEOR: An automatic metric for MT evaluation with improved correlation with human judgments.
In Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, pages 65–72, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.
http://www.aclweb.org/anthology/W/W05/W05-0909.

Chen, Yu, Andreas Eisele, Christian Federmann, Eva Hasler, Michael Jellinghaus, and Silke Theison: Multi-engine machine translation with an open-source decoder for statistical machine translation.
In Proceedings of the Second Workshop on Statistical Machine Translation, StatMT '07, pages 193–196, Morristown, NJ, USA, 2007. Association for Computational Linguistics.
http://portal.acm.org/citation.cfm?id=1626355.1626381.

Hardt, Daniel and Jakob Elming: Incremental re-training for post-editing SMT.
In Conference of the Association for Machine Translation in the Americas 2010 (AMTA 2010), Denver, CO, USA, 2010.

Heafield, Kenneth and Alon Lavie: CMU multi-engine machine translation for WMT 2010.
In Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR, WMT '10, pages 301–306, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics, ISBN 978-1-932432-71-8.
http://portal.acm.org/citation.cfm?id=1868850.1868894.