Classifier-based modelling of source-side context in Machine Translation

The current state of my Ph.D research

Maarten van Gompel

Centre for Language Studies, Radboud University Nijmegen

September 3rd, 2014

Introduction

Translation

1. Faithfully convey the meaning of all words from source to target
2. Fluent target sentence

Phrase-based Statistical Machine Translation

Combination of two models:

1. Translation Model: $P(target|source)$ (bilingual)
2. Language Model: $P(nextword|previouswords)$ (monolingual, this models target-side context)

An MT decoder optimises a log-linear equation to seek the best translation hypothesis amongst a vast pool of possibilities.

Introduction

Translation

1. Faithfully convey the meaning of all words from source to target
2. Fluent target sentence

Phrase-based Statistical Machine Translation

Combination of two models:

1. Translation Model: $P(target|source)$ (bilingual)
2. Language Model: $P(nextword|previouswords)$ (monolingual, this models target-side context)

An MT decoder optimises a log-linear equation to seek the best translation hypothesis amongst a vast pool of possibilities.

Observation

1. SMT does not explicitly model source-side context..
2. ...yet context intuitively seems informative for disambiguation:
   - (English to Dutch): The **bank** went bankrupt → De bank ging failliet.
   - (English to Dutch): The ship departed from the **bank** of the river → Het schip vertrok van de oever van de rivier.
3. Machine Translation overlaps with Word Sense Disambiguation here

Research Hypothesis

1. Does modelling source-side context improve translation?
2. Limited to surface forms, not using linguistic features!

Observation

1. SMT does not explicitly model source-side context..
2. ...yet context intuitively seems informative for disambiguation:
   - (English to Dutch): The **bank** went bankrupt → De bank ging failliet.
   - (English to Dutch): The ship departed from the **bank** of the river → Het schip vertrok van de oever van de rivier.
3. Machine Translation overlaps with Word Sense Disambiguation here

Research Hypothesis

1. Does modelling source-side context improve translation?
2. Limited to surface forms, not using linguistic features!

Two lines of Research

1. Source-side context modelling in full Statistical Machine Translation tasks
2. Source-side context modelling in a narrower scope, closer to WSD: Translation Assistance

Context Modelling in SMT

Training pipeline

1. Given a sentence-aligned parallel corpus:
2. Compute word alignments (GIZA++, uses Expectation Maximisation)
3. Extract probable phrases (grow-diag-final) from word alignments, results in a phrase-translation table
4. Build classifiers mapping all phrases (from the phrasetable) in all of the contexts they occur, to their translations

Classifiers

- Classifiers are memory-based classifiers (k-nn) (proven method in WSD)
- **Classifier Experts:** one classifier per source-side phrase vs. **Monolithic Classifier:** one classifier for all
- Feature vector consists of local context, x words to the left, y to the right.
- Experiments with global keywords also attempted

Test pipeline

1. Given an input sentence to translate:
2. Find all words/phrases in the input sentence for which we built a classifier
3. Classify all fragments given their contexts, store results in an intermediate phrase-table, one per sentence
    - This effectively alters the distribution of possible translations
4. Translate the sentence using the intermediate phrase-table using the Moses SMT Decoder
5. Evaluate output using standard MT metrics (BLEU et al), against human reference translations

Results

- Baseline: non-context informed, standard MT
- No gain above baseline for most general corpora (Europarl, OpenSubtitles, TED talks).
- **But:** Modest gain above baseline in highly formulaic corpora (EMEA, JRC-Acquis)

Observations

- Our edits are quite local and limited
- Reference translations are often too different; translations too free.

Context Modelling: Zooming in

Translation of L1 Fragments in an L2 Context

- Evaluating the translation of fragments as such: closer to Word Sense Disambiguation
- Side-tracking the full complexity of MT: single fragment, no full sentential translation (no long distance word reordering)
- In an interesting new setting aimed at language learning

Problem setting

1. Sometimes you don't know the proper expression in the language you are writing in
2. ... especially if you're not fluent in the language
3. Fall back to your native language! *(code switching)*

Context Modelling: Zooming in

Translation of L1 Fragments in an L2 Context

- Evaluating the translation of fragments as such: closer to Word Sense Disambiguation
- Side-tracking the full complexity of MT: single fragment, no full sentential translation (no long distance word reordering)
- In an interesting new setting aimed at language learning

Problem setting

1. Sometimes you don't know the proper expression in the language you are writing in
2. ... especially if you're not fluent in the language
3. Fall back to your native language! *(code switching)*

Context Modelling: Zooming in

Translation of L1 Fragments in an L2 Context

- Evaluating the translation of fragments as such: closer to Word Sense Disambiguation
- Side-tracking the full complexity of MT: single fragment, no full sentential translation (no long distance word reordering)
- In an interesting new setting aimed at language learning

Problem setting

1. Sometimes you don't know the proper expression in the language you are writing in
2. ... especially if you're not fluent in the language
3. Fall back to your native language! *(code switching)*

Translation Assistance System

1. Given an L1 fragment in an L2 context..
2. ... translate the fragment in its context, to L2

Intended audience

- Language learners (Computer Aided Language Learning)
- ... intermediate level
- ... usage of the language is encouraged as it facilitates learning
- Translation Assistance for professional translators

Translation Assistance System

1. Given an L1 fragment in an L2 context..
2. ... translate the fragment in its context, to L2

Intended audience

- Language learners (Computer Aided Language Learning)
- ... intermediate level
- ... usage of the language is encouraged as it facilitates learning
- Translation Assistance for professional translators

Examples

- Input (L1=English,L2=Spanish): *Hoy vamos a* **the swimming pool***.*
  Desired output: *Hoy vamos a* **la piscina***.*
- Input (L1-English, L2=German): *Das wetter ist wirklich* **abominable***.*
  Desired output: *Das wetter ist wirklich* **ekelhaft***.*
- Input (L1=French,L2=English): *I* **rentre à la maison** *because I am tired.*
  Desired output: *I* **return home** *because I am tired.*
- Input (L1=Dutch, L2=English): *Workers are facing a massive* **aanval op** *their employment and social rights.*
  Desired output: *Workers are facing a massive* **attack on** *their employment and social rights.*

A new Task

1. Participated as a new task in SemEval 2014
2. Six teams participated and wrote a translation assistance system
3. Prior, to this, we did a pilot study with our own system (presented at ACL 2014)
4. Afterwards, we ran our system on our new test set

Test set creation

- Manual collection of test sentences from learner resources and general resources
- Translation of fragments
- Adding of alternative translations
- Verification by natives or professionals

System

System

1. Construct a training set from a parallel corpus and a phrase-table (L1/L2 substitutions)
2. Build classifier experts for each L1 fragment:
   - Features: Local context window (x words left, y words right)
   - Using k-Nearest Neighbours in Timbl
3. On testing, the corresponding classifier is called to translate the L1 fragment
4. An extra L2 Language Model can be included and a simple log linear search seeks the optimum solution:
   $\hat{H} = \arg\max_H \lambda_1 score_T(H) + \lambda_2 score_{lm}(H)$

System

System

1. Construct a training set from a parallel corpus and a phrase-table (L1/L2 substitutions)
2. Build classifier experts for each L1 fragment:
   - Features: Local context window (x words left, y words right)
   - Using k-Nearest Neighbours in Timbl
3. On testing, the corresponding classifier is called to translate the L1 fragment
4. An extra L2 Language Model can be included and a simple log linear search seeks the optimum solution:
   $\hat{H} = \arg\max_H \lambda_1 score_T(H) + \lambda_2 score_{lm}(H)$

System

System

1. Construct a training set from a parallel corpus and a phrase-table (L1/L2 substitutions)
2. Build classifier experts for each L1 fragment:
   - Features: Local context window (x words left, y words right)
   - Using k-Nearest Neighbours in Timbl
3. On testing, the corresponding classifier is called to translate the L1 fragment
4. An extra L2 Language Model can be included and a simple log linear search seeks the optimum solution:
   $\hat{H} = \arg\max_H \lambda_1 score_T(H) + \lambda_2 score_{lm}(H)$

System

System

1. Construct a training set from a parallel corpus and a phrase-table (L1/L2 substitutions)
2. Build classifier experts for each L1 fragment:
   - Features: Local context window (x words left, y words right)
   - Using k-Nearest Neighbours in Timbl
3. On testing, the corresponding classifier is called to translate the L1 fragment
4. An extra L2 Language Model can be included and a simple log linear search seeks the optimum solution:
   $\hat{H} = \arg\max_H \lambda_1 score_T(H) + \lambda_2 score_{lm}(H)$

Evaluation

Baselines

- Most Frequent Translation baseline (non-context informed)
- LM Baseline (L2)

Evaluation metrics

- **Accuracy**: Exact matches
- **Word Accuracy**: Partial matches, credits partial solutions
- **MT Metrics**: BLEU, METEOR, NIST, WER, PER
- **Recall**: Ratio of L1 fragments that yielded *a* solution

Evaluation

Baselines

- Most Frequent Translation baseline (non-context informed)
- LM Baseline (L2)

Evaluation metrics

- **Accuracy**: Exact matches
- **Word Accuracy**: Partial matches, credits partial solutions
- **MT Metrics**: BLEU, METEOR, NIST, WER, PER
- **Recall**: Ratio of L1 fragments that yielded *a* solution

Output sample

Example

Input: Es la **last** vez que me dirijo a esta Cámara .
MLF baseline: Es la **pasado** vez que me dirijo a esta Cámara .
l1r1: Es la **última** vez que me dirijo a esta Cámara .

Conclusions

Conclusions

Conclusion from the pilot study and later application of our system on our SemEval test set:

- Context-aware solutions improve results!
- Language Model and classifiers complement each other: use both
- Small improvement with feature optimisation

. . . or not?

De-conclusion

- Later experiments with better-integrated MT-based solutions again cast doubts on our results: new pure-MT non-context informed baseline wins again

Re-conclusion or Discussion

- Explicit modelling of source-side context provides little or no added value for SMT
- Why? Information is already implicitly available through translation model and (target-side) language model

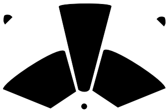
. . . or not?

De-conclusion

- Later experiments with better-integrated MT-based solutions again cast doubts on our results: new pure-MT non-context informed baseline wins again

Re-conclusion or Discussion

- Explicit modelling of source-side context provides little or no added value for SMT
- Why? Information is already implicitly available through translation model and (target-side) language model



Questions?