Translation Assistance

Translating L1 Fragments in an L2 Context

Maarten van Gompel & Antal van den Bosch

Centre for Language Studies, Radboud University Nijmegen

June 24th, 2014

Introduction

Introduction

1. Pilot study: a novel application for Machine Translation
   Cross-lingual translation assistance
2. Close to (Cross-Lingual) Word Sense Disambiguation

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)*

Introduction

Introduction

1. Pilot study: a novel application for Machine Translation
   Cross-lingual translation assistance
2. Close to (Cross-Lingual) Word Sense Disambiguation

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)*

Introduction

Introduction

1. Pilot study: a novel application for Machine Translation
   Cross-lingual translation assistance
2. Close to (Cross-Lingual) Word Sense Disambiguation

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)*

Introduction

Introduction

1. Pilot study: a novel application for Machine Translation
   Cross-lingual translation assistance
2. Close to (Cross-Lingual) Word Sense Disambiguation

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)*

Introduction

Introduction

1. Pilot study: a novel application for Machine Translation
   Cross-lingual translation assistance
2. Close to (Cross-Lingual) Word Sense Disambiguation

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)*

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.*

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

Characteristics

- Replacing and going beyond dictionaries...
  - Fragments can be any word or phrase: no need to be a fixed expression or a linguistic constituent
  - Context-sensitivity
- Our solution is fully data-driven, using techniques from..
  - Machine Translation
  - Word Sense Disambiguation
  - Machine Learning

Characteristics

- Replacing and going beyond dictionaries...
  - Fragments can be any word or phrase: no need to be a fixed expression or a linguistic constituent
  - Context-sensitivity
- Our solution is fully data-driven, using techniques from..
  - Machine Translation
  - Word Sense Disambiguation
  - Machine Learning

This study

1. **Hypothesis:** L2 context can be used to better translate an L1 fragment
2. Memory-based classifier techniques are a means to achieve this
3. **Issue:** Generating artificial training/test data

Creating training and test data (1/2)

- **Ideally:** Corpus of code-switching by language learners
- ....such data does not exist
- **Instead:** Quick alternative: auto-generated on the basis of parallel corpora:
  1. Word alignments trained using GIZA++
  2. Phrase-translation table trained using Moses tools
  3. Iterate through all sentence pairs in the word-aligned parallel corpus
  4. for each sentence pairs, find all matching fragments in the phrase table (above a certain translation probability threshold)
  5. Insert the L1 translation in the L2 sentence
- **Better alternative:** Manually selected sentences from cloze tests (i.e. gap exercises) or learner corpora; with manual L1 insertions

Creating training and test data (1/2)

- **Ideally:** Corpus of code-switching by language learners
- ....such data does not exist
- **Instead:** Quick alternative: auto-generated on the basis of parallel corpora:
  1. Word alignments trained using GIZA++
  2. Phrase-translation table trained using Moses tools
  3. Iterate through all sentence pairs in the word-aligned parallel corpus
  4. for each sentence pairs, find all matching fragments in the phrase table (above a certain translation probability threshold)
  5. Insert the L1 translation in the L2 sentence
- **Better alternative:** Manually selected sentences from cloze tests (i.e. gap exercises) or learner corpora; with manual L1 insertions

Creating training and test data (1/2)

- **Ideally:** Corpus of code-switching by language learners
- ....such data does not exist
- **Instead:** Quick alternative: auto-generated on the basis of parallel corpora:
  1. Word alignments trained using GIZA++
  2. Phrase-translation table trained using Moses tools
  3. Iterate through all sentence pairs in the word-aligned parallel corpus
  4. for each sentence pairs, find all matching fragments in the phrase table (above a certain translation probability threshold)
  5. Insert the L1 translation in the L2 sentence
- **Better alternative:** Manually selected sentences from cloze tests (i.e. gap exercises) or learner corpora; with manual L1 insertions

Creating training and test data (2/2)

- Training and test data are constructed on two equally-sized but completely independent splits of a parallel corpus
- Test data has more stringent thresholds and is sampled down

System

System

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

System

System

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

System

System

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

System

System

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

Experiments

Data

- Europarl (200k train, 200k test)
  - L2=Spanish, L1=English
  - L2=French, L1=English
  - L2=Dutch, L1=English
- IWSLT 2012 Data set
  - L2=Chinese, L1=English
  - L2=English, L1=Chinese

Experiments

- Uniform configuration: same feature vector configuration for all classifier experts
- **Feature optimisation**: Find the best feature configuration per classifier (cross-validation/leave-one-out)
- L2 Language Model
- Asymmetry: Left context only (interesting for as-you-type applications)

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

Europarl results. L1 = English L2 = Spanish. Recall = 0.9422:

Configuration	Accuracy	Word Accuracy	BLEU	WER	PER
MLF baseline	0.6164	0.6662	0.972	1.4465	1.4209
+LM baseline	0.7158	0.7434	0.9785	1.1735	1.1574
l1r1	0.7588	0.7824	0.9801	1.1625	1.1444
l2r2	0.7574	0.7801	0.9800	1.1750	1.1569
l3r3	0.7514	0.7742	0.9796	1.1946	1.1780
l1r1+LM	**0.7810**	**0.7973**	**0.9816**	**1.0946**	**1.077**
auto	0.7626	0.7850	0.9803	1.1594	1.1424
auto+LM	0.7796	0.7966	0.9815	1.1021	1.0845
l1r0	0.6924	0.7223	0.9757	1.3415	1.3249
l2r0	0.6960	0.7245	0.9759	1.3364	1.3193
l2r1	0.7624	0.7849	0.9803	1.1554	1.1378

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

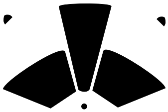
Conclusions

- Context-aware solutions improve results
- Language Model and classifiers complement each other: use both
- Small improvement with feature optimisation
- Right context adds much strength, asymmetric solutions with left-context only are much weaker

New Research since the pilot study

New Research since the pilot study

- Experiments on a more representative data-set.
- Task has run for **SemEval 2014**, with manually constructed data set (Maarten van Gompel, Antal van den Bosch, Iris Hendrickx, Els Lefever, Veronique Hoste)
- Improving recall: breaking down large untranslatable fragments using standard MT techniques



Questions?