Hybrid Adaptation of Named Entity Recognition for Statistical Machine Translation

Vassilina Nikoulina, Ágnes Sándor, Marc Dymetman

Second ML4HMT Workshop, Coling 2012

xerox

Outline

- Introduction
- Approach
  - NER integration within SMT
  - NER adaptation for SMT
  - NER prediction for SMT
- Experiments
- Conclusion

Introduction

- Incorrect NE translation can seriously harm quality

- Main problems caused by NEs in standard PBMT:
  - Ambiguity:
    - *Grant wonderful in Bridget Jones's Diary / Grant obtained by French university*
      - Detecting the NE is crucial for producing the right translation
  - Sparsity:
    - Some named entities can be very sparse (eg. DATEs, UNITs, NAMEs), although they are often used in similar contexts
      - Standard SMT training does not cope well with this situation

Our Approach

- Adaptation of NER for better integration within SMT
  - Additional rules on top of generic NER rule-based model
- NE generalization:
  - Replace NE with a place-holder (specific to the NE type)
- NE translation:
  - Translation of NE with specific NE-translator
- NE-replacement predictor, for choosing between:
  1. Replacing NE by place-holder and using NE-translator
  2. Leaving NE as is and using SMT baseline translation

xerox

Example of proposed framework

- <u>Src:</u> The Author , F. Mellozzini , carries out an in-depth analysis of the objectives of agricultural policy which have arisen during a meeting held in Rome by the Confederation of Agricultural Workers on 18 - 19 October .

- <u>Reduced Src:</u> The Author , PERSON , carries out an in - depth analysis of the objectives of agricultural policy which have arisen during a meeting held in Rome by the ORGANIZATION on DATE .

- <u>Reduced Translation:</u> L'auteur, PERSON, exerce une analyse approfondie des objectifs de la politique agricole qui ont ainsi présentée au cours de la réunion tenue à Rome par la ORGANIZATION en DATE.

Example of proposed framework

- Reduced Translation: L'auteur, PERSON, exerce une analyse approfondie des objectifs de la politique agricole qui ont ainsi présentée au cours de la réunion tenue à Rome par la ORGANIZATION en DATE.

- NE Translation: can be rule-based, dictionary-based, specific for different NE types etc.
  - F. Mellozzini = F. Mellozzini
  - Confederation of Agricultural Workers = Confédération des travailleurs agricoles
  - 18 - 19 October = 18 - 19 octobre

- Final Translation: L'auteur, F. Mellozzini, exerce une analyse approfondie des objectifs de la politique agricole qui ont ainsi présentée au cours de la réunion tenue à Rome par la Confédération des travailleurs agricoles en 18 - 19 octobre .

NER adaptation and prediction

- NER errors may lead to decrease in translation quality
- The internal structure of NE's should be adapted for SMT (different from structure required for IE)
  - We propose post-processing rules on top of our baseline NER system
- Not all NEs should be replaced by the place-holder:
  - e.g. If the NE is frequent in the bilingual training data, then the baseline SMT may perform well in translating it
    - We propose to learn a predictor for making the choice

Adaptation of NER for SMT

- Many existing NER systems are created for Information Extraction (IE)
- Translation works better with a minimal pattern:

NER4IE	NER4SMT
[Queen Elisabeth]	[Elisabeth]
[on July 15th]	[July 15th]

- Modification of NER system so that it does not extract
  - common nouns
  - function words
- Advantages:
  - Simplifies NE translation model
  - Reduces sparsity in phrase extraction

Prediction model for NE replacement

- Prediction model: 0/1 classifier deciding whether NE replacement is beneficial for final translation quality

- Some features :
  - NE type
  - NE frequency in training data
  - NE context in source
  - Confidence in NE translation

- In order to learn this classifier, we need to create some training data…

Creating a training set for the prediction model

For each sentence *s* in a dev-set :

- Translate *s* with the baseline SMT model : SMT(s)
- For each *ne* found by NER in *s* :
  - Replace *ne* with place-holder: $s|_{ne}$
  - Translate $s|_{ne}$ with the placeholder-enabled SMT model : SMT_NE ($s|_{ne}$)
  - Compare SMT(s) and SMT_NE ($s|_{ne}$) relative to the reference translation (BLEU or TER)
  - Label *ne* **positive** if the comparison strongly in favor of SMT_NE ($s|_{ne}$), **negative** in the opposite case, **neutral** if difference is small

Train the classifier on the positive/neutral/negative labels

Note: This model can be generalized to a multiple-class classication problem, when different NE translators are available.

Overall Training of the NE-aware SMT system

- Create reduced parallel corpus:
  - Use NER on the source side of our bilingual corpus
  - Project source NEs on the target (through word-alignment)
  - Replace aligned NEs with a place-holder
    - This replacement is done only with probability *alpha,* so as to keep a proportion of NEs in their original form
- Train reduced SMT model:
  - This model will be able to deal not only with the place-holders, but also with the original form of frequent Named Entities
- Train Prediction model for reduced SMT model

Experimental settings

- English-French translation task
- Data: titles and abstracts of scientific publications in Agricultural domain (European Project Organic.Lingua)
- Baseline SMT: Moses with standard settings trained on ~150K in-domain parallel sentences
- Baseline NER: Xerox Incremental Parser, rule-based
- NE prediction model: SVM 3-class classifier (libsvm)

  1 : replace with a place-holder; 0/-1 : do not replace
- NE-specific translation model: a combination of two techniques
  - Bilingual dictionary extracted by projection from the bilingual corpus
  - When not found in this dictionary, baseline SMT system, but tuned on a set of parallel NE's

Experimental results

	Titles		Abstracts	
	BLEU	TER	BLEU	TER
Baseline SMT	0.3135	0.6566	0.1148	0.8935
NE-aware SMT, *baseline* NER	0.3213	0.6636	0.1211	0.9064
NE-aware SMT, *adapted* NER	0.3258	0.6605	0.1257	0.8968
NE-aware SMT, *baseline* NER + NE prediction model	0.3371	0.6523	0.1228	0.9050
NE-aware SMT, *adapted* NER + NE prediction model	**0.3421**	**0.6443**	**0.1341**	**0.8935**

Conclusions and future work

- Proposed framework for NE integration within SMT addressing sparsity issues
- Adaptation of standard NER + Prediction of NE-replacement are beneficial for final translation quality
- Future work: replace pipeline architecture with confusion network

Questions ?

Also can be addressed to: *Vassilina.Nikoulina@m4x.org*