Modeling Punctuation Prediction as Machine Translation

Stephan Peitz, Markus Freitag, Arne Mauser and Hermann Ney
Human Language Technology and Pattern Recognition, RWTH Aachen University

RWTH AACHEN UNIVERSITY

Motivation

- ASR output does not contain punctuation marks
- MT systems are trained on text data with punctuation
- prediction errors affect translation quality
  - loss of up to 4 BLEU points if punctuation marks need to be predicted, compared to correct punctuation in the input

system	MT dev		MT test	
`+s2t TED trip.`	BLEU	TER	BLEU	TER
correct punctuation	27.5	57.0	30.8	50.9
restored punctuation	24.0	61.7	26.6	55.9

Introduction

- in this work, we consider all kinds of punctuation
  - sentence-end punctuation marks
  - commas
  - parentheses and quotation marks
- punctuation prediction is performed via
  - tool from the SRI LM toolkit [Stolcke, ICSLP 2002]
  - statistical machine translation [Hassan et al., IWSLT 2007] [Ma et al., IWSLT 2008]
- comparison and combination of different methods
- applied in the IWSLT 2011 evaluation campaign

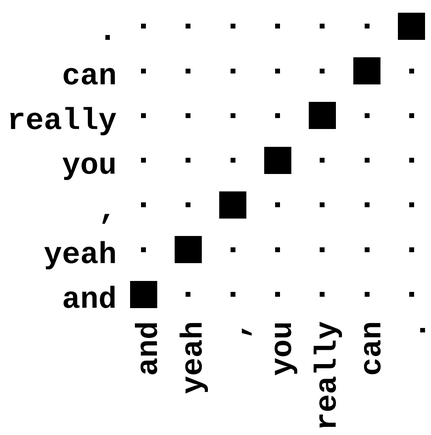
Strategies [Matusov et al., IWSLT 2006]

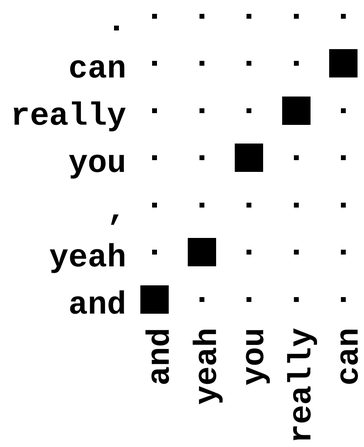
three different stages at which prediction is done

- before translation in the source language (**FULLPUNCT**)
  - no modification to the training data or the translation system
  - prediction errors can affect the translation
- during translation implicitly (**IMPLICIT**)
  - removing all punctuation marks from the source language data
  - re-extracting phrase and word lexicon models
  - prediction and translation are not separate
- after translation in the target language (**NOPUNCT**)
  - all punctuation marks are removed from the training data as well as from the development and test sets
  - translation model and target language model have to rebuilt
  - translation produces errors, make the punctuation prediction less accurate

Punctuation Prediction

- with `hidden-ngram` tool from the SRI LM toolkit (**H-NGRAM**)
  - standard setting with 9-gram language model
- with statistical machine translation (**PPMT**)
  - based on phrase-based MT system
  - additional features besides the language model
  - translate from unpunctuated to punctuated text
  - system is tuned with standard MERT on BLEU





Comparison of the Translation Quality

- IWSLT 2011 English-to-French speech translation of talks [Federico et al., IWSLT 2011]
- uses `+s2t TED trip.` from English-French MT for all punctuation prediction strategies [Wuebker et al., IWSLT 2011]
- system combination [Matusov et al, EACL 2006]
  - combine translation output from multiple punctuation prediction schemes

system	SLT dev		SLT test	
	BLEU	TER	BLEU	TER
IMPLICIT	18.0	69.5	21.8	62.5
FULLPUNCT (H-NGRAM)	18.2	69.3	21.1	62.9
FULLPUNCT (PPMT)	18.3	69.2	21.9	62.2
NOPUNCT (H-NGRAM)	17.3	67.9	20.4	62.8
NOPUNCT (PPMT)	17.8	69.0	21.2	62.2
system comb.	18.5	68.3	22.3	61.6

Comparison of the Punctuation Prediction Accuracy

- remove all punctuation from test set of the correct manual transcription (pseudo ASR output)
- restore the punctuation marks with H-NGRAM and PPMT
- use the original test set as reference
- measure the accuracy regarding three different classes of punctuation marks:
  - class 1: `.,?` and `!`
  - class 1.1: `.`
  - class 1.2: `?`
  - class 2: `,`
  - class 3: `"`, `'`, `;`, `(` and `)`

tool	class 1			class 1.1			class 1.2		
	Prec.	Rec.	F_1	Prec.	Rec.	F_1	Prec.	Rec.	F_1
H-NGRAM	87.9	85.0	86.4	88.9	90.7	89.8	59.7	23.0	33.2
PPMT	88.2	81.7	84.8	89.0	87.5	88.2	63.4	17.6	27.5

tool	class 2			class 3			all punct.		
	Prec.	Rec.	F_1	Prec.	Rec.	F_1	Prec.	Rec.	F_1
H-NGRAM	83.5	44.8	58.3	18.3	6.7	9.8	81.5	57.3	67.3
PPMT	80.6	59.3	68.3	47.2	22.7	30.7	80.7	64.2	71.5

Example

system	tool	
pseudo ASR output	-	they say The plants talk to us
reference	-	they say , ” The plants talk to us . ”
FULLPUNCT	H-NGRAM	they say The plants , talk to us .
FULLPUNCT	PPMT	they say , ” The plants talk to us .

Conclusion

- compared different approaches for predicting punctuation in a speech translation setting
- PPMT outperformed H-NGRAM
- FULLPUNCT (PPMT) slightly better than the implicit method
- main advantage of FULLPUNCT: no modification to the translation system
- system combination improved translation quality further
- future work:
  - investigate special features for parentheses or quotes
  - try different optimization criteria, e.g. F-measure or WER

Lehrstuhl für Informatik 6 – Computer Science Department – RWTH Aachen University – Aachen, Germany
e-mail: `<surname>@cs.rwth-aachen.de`
www: `http://www-i6.informatik.rwth-aachen.de`