International Workshop on Spoken Language Translation
Kyoto, Japan
November 27-28, 2006

Automatic Sentence Segmentation and Punctuation Prediction for Spoken Language Translation

Evgeny Matusov, Arne Mauser, and Hermann Ney

Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University
D-52056 Aachen

Content

1. motivation
2. related work
3. handling speech input in phrase-based MT
4. methods for predicting punctuation
5. automatic sentence segmentation algorithm
6. experimental results
7. summary

Motivation

ASR systems perform only acoustic segmentation

segments may be too short/long for machine translation systems

⇒ a novel approach to sentence segmentation

- **using a log-linear combination of lexical and prosodic features**
- **using an explicit sentence length model and length contraints**

MT users expect to see sentences with proper punctuation

⇒ three different strategies for punctuation prediction:

- **in the source language (before translation)**
- **in the target language (after translation)**
- **implicitly (using translation models)**

Related work

- annotation of the ASR output as the end product delivered to the user:
  - combine lexical and prosodic cues to improve sentence boundary prediction (Liu et al., 2004)
  - maximum entropy model (Huang and Zweig, 2002)
    decision trees (Kim and Woodland, 2001)
- comma restoration for ASR output translation (Lee et al., 2006)
  - segments are already known and are assumed to end with a period
  - thresholds for bigram/trigram probabilities for commas

Phrase-based MT for Speech Translation

- translation of sentence units
- word and phrase reordering limit the maximum length of a SU to about 50-60 words
- punctuation is treated as words when training the MT system
- models: phrase-based and word-based lexica, language model, etc.
- loglinear model combination

MT Setups for Punctuation:

three different possible setups:

- **MT without punctuation, prediction in the target language**
- **MT with punctuation, prediction in the source language**
- **implicit generation of punctuation marks in MT**

MT without punctuation marks

∗ **remove punctuation from corpus if present**

∗ **insert punctuation marks into the MT output based**

\+ **improved phrase coverage**

– **errors in translation hypotheses reduce prediction quality**

– **only language information available**

Various Punctuation Prediction Strategies

MT with punctuation marks

∗ **keep all punctuation marks in training**

\+ **prosodic features can be used for prediction**

− **incorrect prediction affects MT**

− **difficult to used with ASR lattice translation**

− **punctuation is different in languages**

Various Punctuation Prediction Strategies

Implicit prediction of punctuation in the MT process

- **remove punctuation marks only from the source**
- **punctuation marks will be "inserted" through phrasal translations**
- **example of alternative phrasal translations:**

```
                        did you say
sagten Sie =>         , did you say
                        did you say ,
                      , did you say ,
```

\+ **SMT features help to select best translation and punctuation variant**

\+ **MET of MT system also optimizes punctuation prediction**

\+ **easy to use, applicable to translation of word lattices**

Automatic Sentence Segmentation: Idea

Motivation:

- limit the maximum/minimum length of automatically determined SUs
- maximum: most MT algorithms work well up to 40-50 words
- minimum: context information is lost with too short SUs

Idea:

- utilize lexical and prosodic features (language model, pause duration)
- explicit optimization over segment length
- score of hypothesized segment boundary based on optimal previous segment boundary
- minimum/maximum SU length can be parametrized

Automatic Sentence Segmentation: Theory

- given: unsegmented word sequence $w_1^N := w_1, w_2, \ldots, w_N$
- find optimal segmentation $i_1^K := (i_1, i_2, \ldots, i_K = N)$
- i.e. segmentation with the highest posterior probability

$$\hat{i}_1^{\hat{K}} = \operatorname*{argmax}_{K, i_1^K} \left\{ Pr(i_1^K | w_1^N) \right\} \tag{1}$$

- posterior probability modeled as log-linear combination of several features:

$$Pr(i_1^K | w_1^N) = \frac{\exp\left(\sum_{m=1}^M \lambda_m h_m(i_1^K, w_1^N)\right)}{\sum\limits_{K', i'^{K'}_1} \exp\left(\sum_{m=1}^M \lambda_m h_m(i'^{K'}_1, w_1^N)\right)} \tag{2}$$

- decision rule

$$\hat{i}_1^{\hat{K}} = \operatorname*{argmax}_{K, i_1^K} \left\{ \sum_{m=1}^M \lambda_m h_m(i_1^K, w_1^N) \right\} \tag{3}$$

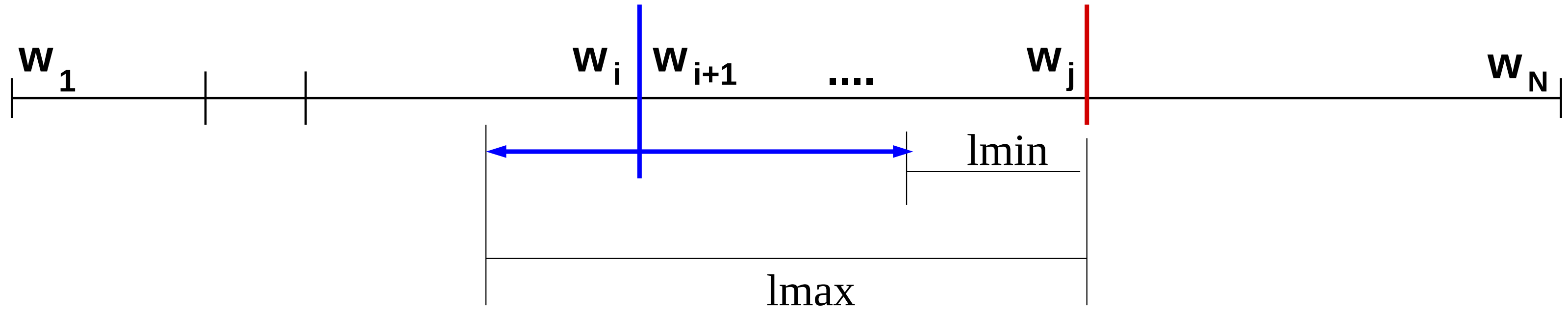
Automatic Sentence Segmentation: Models Used - 1

features used in log-linear combination

- n**-gram language model probability incl. hidden event for boundary**
- **explicit sentence length probability** $p(j - i)$ **(lognormal distribution)**
- **normalized pause duration between** w_i **and** w_{i+1}
- **segment penalty**
- **possible extensions: prosodic features form as (pitch, energy, ...)**

scaling factors tuned manually for precision/recall on development data

Automatic Sentence Segmentation: Search



process the word sequence w_1^N from left to right

recursion

$$p(w_1^j) = \max_{j-lmax \leq i \leq j-lmin} p(w_1^i) \cdot p(w_{i+1}^j)$$

$p(w_1^i)$ was determined in the previous recursion step

processing time: less than 2 seconds for segmentation of 30 000 words

Experimental results

Evaluation Methodology

sentence segmentation:

- **precision and recall compared to manual segment boundaries**

machine translation:

- **objective measures WER, PER, BLEU, NIST with human references**

evaluation of automatically segmented output:

- **MT output automatically aligned to reference (Matusov et al. 2005)**

Experiments

sentence segmentation:

- **use language model and length model features only (IWSLT task)**

+ **add pause duration feature (TC-STAR task)**

machine translation:

- **test the influence of automatic segmentation**
- **three strategies for punctuation prediction**

Quality of Sentence Segmentation (IWSLT 2006 task)

Compare with the standard approach: SRI hidden-ngram tool

corpus	RWTH tool P	RWTH tool R	hidden-ngram P	hidden-ngram R
IWSLT test 2005	84.2	84.1	84.1	85.5
IWSLT dev 2006	59.5	64.6	57.0	62.4
IWSLT test 2006	56.4	61.0	54.9	57.6
IWSLT test 2006 (ASR)	56.0	55.2	55.4	52.6

- precision (P) and recall (R) in %
- 4-gram LM used in both tools
- minimum sentence length 3 words, maximum sentence length 30 words
- Chinese transcriptions
  - 2005 test set: 3208 words and 506 segments
  - 2006 test set: 5550 words and 500 segments (15.2% character error rate of the ASR output)
- 2005 test set is very similar to the language model training data

Quality of Sentence Segmentation (TC-STAR task)

	Development		Test	
	P	R	P	R
baseline (4-gram LM only)	54.2	52.1	54.0	50.4
+ length model	54.7	52.5	55.3	51.7
+ pause model	68.8	68.4	70.5	69.7
baseline + pause model	68.1	68.3	69.9	70.3

- precision (P) and recall (R) in %
- minimum sentence length 3 words, maximum sentence length 50 words
- English ASR output, word error rate of 6.9%, 28K words, 1155 reference segments
- pause duration feature is very important for good segmentation performance

Translation Quality: IWSLT 2006 Chinese-to-English Task

transcription	segmentation	punctuation	BLEU [%]	WER [%]	PER [%]	NIST
DEV 2006						
correct	correct	source	19.8	70.5	54.3	5.99
		implicit	22.0	71.0	53.0	5.86
		target	18.9	70.7	55.2	6.03
	automatic	source	17.3	66.1	54.9	5.34
		implicit	20.7	62.1	52.0	5.41
		target	17.5	67.2	55.9	5.49
automatic	correct	source	15.9	73.9	58.5	5.28
		implicit	19.0	69.1	56.7	5.18
		target	15.4	73.2	58.2	5.37
	automatic	source	14.4	68.4	58.2	4.51
		implicit	17.1	64.8	55.2	4.62
		target	13.8	69.0	59.1	4.60

Translation Quality on the IWSLT Task: Discussion

- errors introduced by ASR have a higher impact than errors introduced by automatic segmentation
- implicit prediction of PM in translation process performs best
  - \+ phrasal translations are not being “broken” by incorrectly inserted PM on the source side
  - \+ especially important for small vocabulary tasks, with limited phrase coverage

Translation Quality: TC-STAR English-to-Spanish Task

transcription	segmentation	punctuation prediction	BLEU [%]	WER [%]	PER [
correct	correct	manual (source)	45.2	43.3	32.2
automatic	correct (aligned)	source	37.8	50.6	37.6
	automatic	source	36.7	51.2	38.1
		implicit	36.1	51.5	38.6
		target	36.3	51.3	38.4
		full stop only (source)	35.8	50.2	38.6

- **only a small degradation in MT quality when automatic segmentation is used**
- **low recognition WER on this corpus makes punctuation prediction in the source language sufficiently reliable to serve as input to the MT system**

Conclusions

- **methods for automatic segmentation and punctuation prediction**
- **improved interface between ASR and MT**
- **sentence segmentation:**
  - **performs at least as well as existing state-of-the-art algorithm**
  - **sentence length can be explicitly adjusted**
  - **robust also in terms of machine translation quality**
- **three different approaches for punctuation prediction**
- **the three methods have advantages and disadvantages**
- **for small vocabulary tasks: implicit generation punctuation leads to higher translation quality**

References

- **Huang, J., and Zweig, G. "Maximum entropy model for punctuation annotation from speech". In *Proc. of ICSLP*, pp. 917-920, 2002.**
- **Kim, J., and Woodland, P. "The use of prosody in a combined system for punctuation generation and speech recognition". In *Proc. of Eurospeech*, pp. 2757-2760, 2001.**
- **Lee, Y., Al-Onaizan, Y., Papineni, K., and Roukos, S. "IBM Spoken Language Translation System". In *Proc. TC-STAR Workshop on Speech-to-Speech Translation*, pp. 13-18, Barcelona, Spain, June 2006.**
- **Liu, Y., Shriberg, E., Stolcke, A., Hillard, D., Ostendorf, M., Peskin, B., and Harper, M. "The ICSI-SRI-UW Metadata Extraction System," *ICSLP 2004, International Conf. on Spoken Language Processing*, Korea, 2004.**
- **Matusov, E., Leusch, G., Bender, O., and Ney, H. "Evaluating Machine Translation Output with Automatic Sentence Segmentation". In *Proc. of IWSLT 2005*, pp. 148-154, Pittsburgh, PA, October 2005.**