Karlsruhe Institute of Technology

Punctuation Insertion for Real-time Spoken Language Translation

IWSLT 2015

Eunah Cho, Jan Niehues, Kevin Kilgour, Alex Waibel | December 4, 2015

INSTITUTE FOR ANTHROPOMATICS AND ROBOTICS

Let's eat, grandpa!

Let's eat grandpa

Punctuation insertion

OK but then after a while I realized this is my life this is six months of my life and this...

OK. But then, after a while, I realized this is my life. This is six months of my life. And this...

- Readability
- Performance of MT

Punctuation and MT

- Impact of segmentation and punctuation marks on MT performance
- Oracle experiments
- German to English

	BLEU
ASR output	20.70
+ Segmentation	21.42
+ Full stop	22.18
+ All punctuations	22.48
Transcript	27.99

Conventional strategies...

LM and prosody based model

- Incorporates acoustic features
- Low latency

Monolingual translation system

- MT that translates non-punctuated language into punctuated one
- Improve MT performance
- Overlapping window for segment insertion, longer context $\rightarrow$ latency

LM and prosody based model

- Consider two words prior and two after the possible punctuation mark
- 4-gram LM trained on punctuated text
  - Score without an inserted punctuation mark $P(w_{i-1}, w_i, w_{i+1}, w_{i+2})$
  - Score with a comma $P(w_{i-1}, w_i, \texttt{@COMMA}, w_{i+1}, w_{i+2})$
  - Score with a fullstop $P(w_{i-1}, w_i, \texttt{@STOP}, w_{i+1}, w_{i+2})$
- Dynamic scaling factor to prevent very short or long sentences
- Pauses over θs then force terminate segment

Monolingual MT: Random segmentation for training data

- **Original corpus**
  this is Bill Lange.
  I am Dave Gallo.
  and we are going to tell you some stories from the sea here in video.
  we have got some of the most incredible video of Titanic that is. . .

- **Randomly segmented corpus**
  this is Bill Lange. I am
  Dave Gallo. and we are going to tell you
  some stories from the sea here in video. we have got
  some of the most incredible video of Titanic that is. . .

Monolingual MT: Sliding window for testing

- Sliding window to observe words in longer, various contexts
- Example segments: *. . . der bildet die sogenannte konjunktive Normalform wir haben gesehen dass wir diese . . .*

der	bildet	die	sogenannte	konjunktive	Normalform	wir	haben
bildet	die	sogenannte	konjunktive	Normalform	wir	haben	gesehen
die	sogenannte	konjunktive	Normalform	wir	haben	gesehen	dass
sogenannte	konjunktive	Normalform	wir	haben	gesehen	dass	wir
konjunktive	Normalform	wir	haben	gesehen	dass	wir	diese
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Output of the Monolingual MT

- Punctuated words in bold letter
- An empirically chosen threshold

der	bildet	die	sogenannte	konjunktive	**Normalform.**	Wir	haben
bildet	die	sogenannte	konjunktive	**Normalform.**	Wir	haben	**gesehen,**
die	sogenannte	konjunktive	**Normalform.**	Wir	haben	**gesehen,**	dass
sogenannte	konjunktive	**Normalform.**	Wir	haben	**gesehen,**	dass	wir
konjunktive	**Normalform.**	Wir	haben	**gesehen,**	dass	wir	diese
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- *. . . der bildet die sogenannte konjunktive Normalform. Wir haben gesehen, dass wir diese. . .*

Efficient punctuation insertion scheme

- For real-time SLT system
- Based on the monolingual translation system
- Designed to decrease the latency
- No loss in performance

Real-time spoken language translation

- Latency: the time between when a word is spoken and when its transcript and translation are displayed to the user
- Each component adds to the latency
  - Computation time $\Rightarrow$ fast servers with multiple cores, parallelized computations, smaller faster models
  - Communication time $\Rightarrow$ fast connection, low overhead between components
  - Required (future) context, ...

Resending of ASR

- Reduce the apparent latency
- ASR continually outputs its current best hypothesis e.g., once a second
- Updated by newer, possibly better, hypothesis
- Higher user acceptance than waiting for a complete, stable hypothesis

Example: resending of ASR

. . . in this planet you would have to prove . . .

. . . in this planet you would have to provide 36 million translation . . .

. . . many dialects it is of course a dog . . .

. . . many dialects it is of course a daunting challenge . . .

Segmentation for streamed input

- MT system requires the ASR output to be segmented
- The MT/ASR components constantly update its current hypothesis
  - The punctuation prediction component has to deal with possibly changing, frequently sent inputs fast
  - The MT component requires sentence boundary info as soon as possible

- Monolingual translation system with overlapping window
  - Long shifting window of 10 words, requiring a long future context $\Rightarrow$ latency!

Model

- Finalized ASR output $w = \{w_l, \ldots, w_m\}$
- Flexible ASR output $v = \{v_{m+1}, \ldots, v_n\}$

. . . would not exist in one hundred years ***one***

. . . hundred years why because they look at the ***curb*** *its*

. . . why because they look at the ***curve*** *and say if*

. . . the curve and say if the population keeps growing at ***these***

. . . keeps growing at ***this*** *rate*

- Another stack for recognized words before w for punctuation insertion setup:
- History stack $h = \{h_{l-c}, \ldots, h_{l-1}\}$, with context length c

Model

- Input
  city of New York would not exist in one hundred years one
- SegOutput
  city of New York would not exist in one hundred years. One

- Input
  not exist in one hundred years why because they look at the curb its
- SegOutput
  not exist in one hundred years. Why? Because they look at the curb its

History stack and punctuation output

- History stack in yellow box

Input	OK but then after a while
Output	OK. But then, after a while,
Input	then, after a while, I realized this is
Output	then, after a while, I realized this is
Input	I realized this is my life this is six months of
Output	I realized this is my life. This is six months of
Input	is six months of my life and
Output	is six months of my life. And
Input	of my life. And this . . .
Output	of my life. And this . . .

What about overlapping window?

and I said, “OK, it ’s the huge file. OK,
I said, “OK, it ’s the huge file. OK, but
said, “OK, it ’s the huge file. OK, but then
OK. it ’s the huge file. OK, but then, after
it ’s the huge file. OK, but then, after a
’s the huge file. OK, but then, after a while,
the huge file. OK, but then, after a while, I
huge file. OK, but then, after a while, I realised
file. OK, but then, after a while, I realised this
OK , but then, after a while, I realized. this is
but then, after a while, I realized. this is my
then , after a while, I realized. this is my life.
after a while, I realized. this is my life. this
a while, I realized. this is my life. this is
while I realized. this is my life. this is six
I realized. this is my life. this is six months
realized . this is my life. this is six months of
this is my life. this is six months of my
is my life. this is six months of my life
my life. this is six months of my life, and
life . this is six months of my life, and this
this is six months of my life. and this fire
is six months of my life. and this fire. so
six months of my life. and this fire. so, I
months of my life. and this fire. so I was
of my life. and this fire. so I was a
my life. and this fire. so I was a little
life , and this fire. so I was a little bit
and this fire. so I was a little bit skeptical
this fire. so I was a little bit skeptical of

- With 10 words of overlapping window, a much longer future context is required

Experimental setup

- Punctuate English tst13
  - LM and prosody based segmentation
  - Monolingual translation system with an overlapping window 10
  - Monolingual translation system with streaming input*
- Translate it into German
  - Online setup
  - Not possible to prepare 100% fit phrase table for each test data
  - Phrase table is built based on training data vocabulary, filtered

Monolingual translation system

- Built on EPPS, NC, TED, and cleaned crawl corpus of English, 10.1 m words
- Randomized sentence boundaries, so that punctuation marks can be observed in all possible locations
- Removed .,?! from the corpus and lowercased all letters for the source side
- Phrase-based translation model

Results

Punctuation	ASR Output	Manual Transcript
LM, Prosody	9.74	-
Baseline	11.18	19.57
StreamingInput	**11.55**	**19.41**

- Scores in BLEU
- Baseline: monolingual translation system with an overlapping window 10
- 1.8 BLEU points of improvements

Segmentation Comparison - 1

- LM and prosody based

I also ask myself does not really work can they really store all.
this information about us and every time I use my mobile phone.
so I ask my phone company Deutsche Telekom which was at that time the largest phone company.
in Germany and they ask them please send me all the information you have started.
about me.
and there is some one thousand against and I got no real. . .

Segmentation Comparison - 1

- Streaming input segment

I also ask myself, does not really work?
can they really store.
all this information about us.
and every time I use my mobile phone.
so I ask my phone company, Deutsche Telekom, which was at that time the largest phone company in Germany, and they ask them, please send me all the information you have started about me.
and there is some one thousand against, and I got no real, . . .

Segmentation Comparison - 2

- LM and prosody based

  . . . the city of New York would.
  not exist in one hundred years.
  why because they look at the curve and say if the population keeps growing at.
  this rate to move the population of New York year round they would have needed.

Segmentation Comparison - 2

- Streaming input segment

  the city of New York would not exist in one hundred years.
  why?
  because they look at the curve, and say, if the population keeps growing at this rate to move the population of New York year round.
  they would have needed . . .

Conclusion

- Latency issue of real-time spoken language translation systems
- Performance of punctuation prediction

- Monolingual translation system: good performance but latency due to input in sliding window

- Using streamed input
- Based on monolingual translation system
- comparable performance

Thank you!

Segmentation Comparison, using Sliding Window

- Example 1

  I also ask myself does not really work.
  can they really store all this information about us.
  and every time I use my mobile phone.
  so I ask my phone company, Deutsche Telekom, which was at that time the largest phone company in Germany, and they ask them, please send me all the information you have started about me and there some one thousand against, and I got no real and . . .

Segmentation Comparison, using Sliding Window

- Example 2

  the city of New York would not exist in one hundred years.
  why?
  because they look at the curve, and say, if the population keeps growing at this rate to move the population of New York year round.
  they would have needed . . .