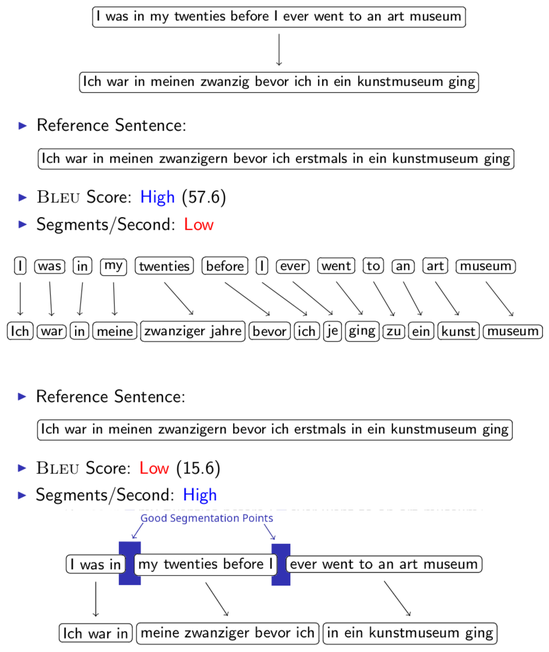
Learning Segmentations that Balance Latency versus Quality in Spoken Language Translation

Hassan S. Shavarani, Maryam Siahbani, Ramtin M. Seraj, Anoop Sarkar
Simon Fraser University, 8888 University Dr., Burnaby, BC, Canada
{sshavara,msiahban,rmehdiza,anoop}@sfu.ca

Contributions

- We provide a method that will create annotated training data for segmentation classifier, considering both Latency and Accuracy
- Our method extends (Oda et al., 2014)'s greedy approach [2]
- Our method explores all potential segmentation points anywhere in the corpus to find the optimal set for data annotation (using dynamic programming)
- We provide experiments that show this method works better than the state-of-the-art methods

Segmentation Alternatives



Example Corpus

I am a contemporary artist with a bit of an unexpected background .
N V D J N P D N P D J N .

I was in | my twenties before I ever | went to an art museum .
N V P | S N P N A | V P D N N .

I grew up in the middle of nowhere on a dirt road in rural Arkansas .
N V R P D N P N P D N N P J N .

Point	Freq	Point	Freq	Point	Freq
N-P	6	J-N	3	V-R	1
P-D	5	N-N	2	P-S	1
D-N	4	P-N	2	P-J	1
N-.	3	D-J	2	S-N	1
N-V	3	R-P	1	A-V	1
V-D	3	N-A	1		
Full Segmentation Set Size				40	

Pareto-Optimal Segmentation Algorithm

Algorithm Pareto-Optimal Segmentation

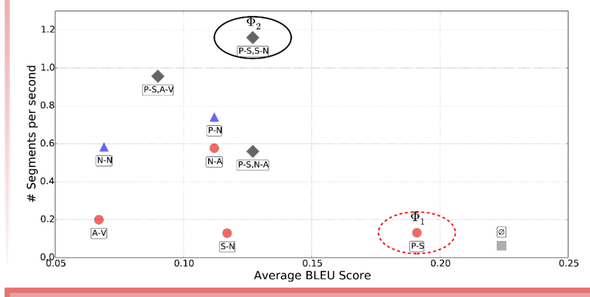
$\Phi_0 \leftarrow \emptyset$
for $k = 1$ to K **do**
  for $j = 0$ to $k-1$ **do**
    $\Phi' \leftarrow \{\phi : (\phi \notin \Phi_j) \wedge (count(\phi; \mathcal{F}) = k - j)\}$
    $\Phi_{k,j} \leftarrow \Phi_j \cup \left\{ \arg \text{pareto frontier}_{\phi \in \Phi'} \{B_\alpha(s(\mathcal{F}, \Phi_j \cup \{\phi\})), \Lambda_\alpha(s(\mathcal{F}, \Phi_j \cup \{\phi\}))\} \right\}$
  end for
  if $k < K$ **then**
    $\Phi_{k,j} \leftarrow \text{argmax}_{\phi \in \{\Phi_{k,j} : 0 \le j \le k\}} B_\alpha(s(\mathcal{F}, \phi))$
  end if
  $\Phi_k \leftarrow \arg \text{pareto frontier}_{\Phi \in \{\Phi_{k,j} : 0 \le j \le k\}} \{B_\alpha(s(\mathcal{F}, \Phi)), \Lambda_\alpha(s(\mathcal{F}, \Phi))\}$
end for
return $s(\mathcal{F}, \Phi_K)$

Running Example

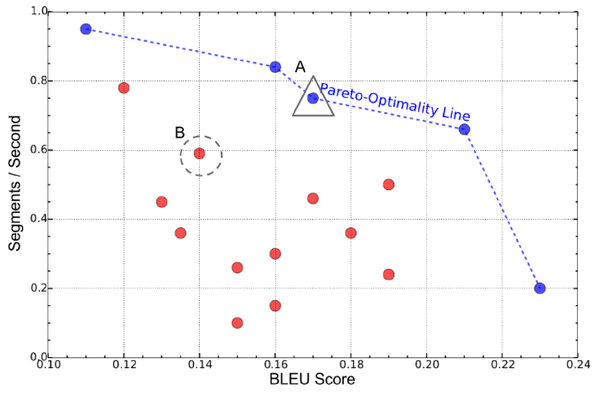
- For $K = 2$, We can have different segmentation choices for $K = 2$

  [N-N] happening twice OR [P-S,S-N] each happening once OR ...

- the run of algorithm over the example data will produce the following plot.



Pareto-Optimality



References

[1] H. S. Shavarani, M. Siahbani, R. M. Seraj, and A. Sarkar Learning Segmentations that Balance Latency versus Quality in Spoken Language Translation In *Proc. of IWSLT 2015*.
[2] Y. Oda, G. Neubig, S. Sakti, T. Toda, and S. Nakamura Optimizing segmentation strategies for simultaneous speech translation In *Proc. of ACL 2014*.
[3] V. K. Rangarajan Sridhar, J. Chen, S. Bangalore, A. Ljolje, and R. Chengalvarayan Segmentation strategies for streaming speech translation In *NAACL, 2013*.

Pareto-Optimal Segmentation

- Assumption: Sentence boundaries are predefined in the corpus!
- Greedily chooses the best potential seg. point and adds it to the previous selected points (as [2] does).
- Accuracy Measure: avg. $\{\frac{\text{BLEUp1}}{\#Segments}\}$
- Latency Measure: avg. $\{\frac{\#Segments}{Translation\ Time}\}$

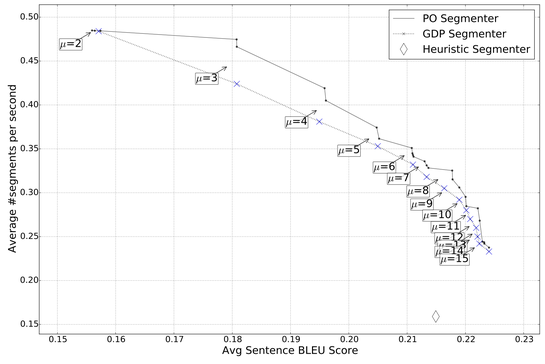
Input: the desired avg. segment length (μ)
$\Rightarrow$ total number of expected segments (K)

$$K = *\frac{\#Words}{\mu} - [\#Sentences]$$

- Sentence boundaries do not count towards K

Results

- **Task:** English-German TED speech translation.
- Training/tuning the MT system data: IWSLT Train 2012-2013 + half of Europarl
- Segmenter Train/Test/Held-out data: IWSLT Dev/Test 2010,2011,2013
- **Methods** to be compared:
  - The state-of-the-art heuristic speech segmentation approach [3]
  - Greedy segmentation approach [2]
  - Pareto-optimal segmentation approach [1]



	$\mu = 3$		$\mu = 8$	
	Segs/Sec	Bleu	Segs/Sec	Bleu
PO	**0.474**	18.07	**0.315**	**21.77**
Greedy	0.424	18.07	0.305	21.63