Learning Segmentations that Balance Latency versus Quality in Spoken Language Translation

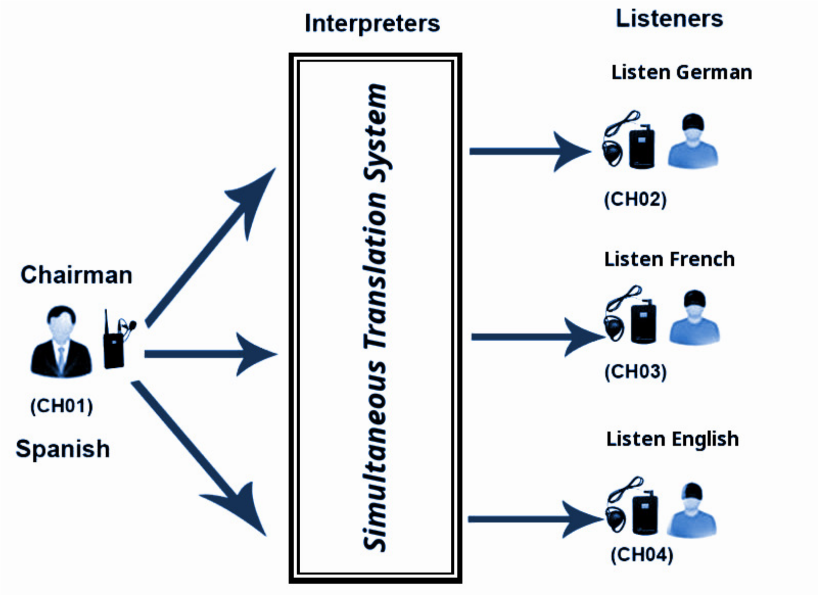
Hassan S. Shavarani Maryam Siahbani Ramtin M. Seraj
Anoop Sarkar

SFU NatLangLab
Natural Language Processing Lab (Natlang)
School of Computing Science
Simon Fraser University

IWSLT 2015

Introduction

Simultaneous Translation (Interpretation)



Simultaneous Translation - Extreme Strategies

- First Translation Strategy:

I was in my twenties before I ever went to an art museum

↓

Ich war in meinen zwanzig bevor ich in ein kunstmuseum ging

- Reference Sentence:

Ich war in meinen zwanzigern bevor ich erstmals in ein kunstmuseum ging

- BLEU Score: High (57.6)
- Segments/Second: Low

Simultaneous Translation - Extreme Strategies

- Second Translation Strategy:

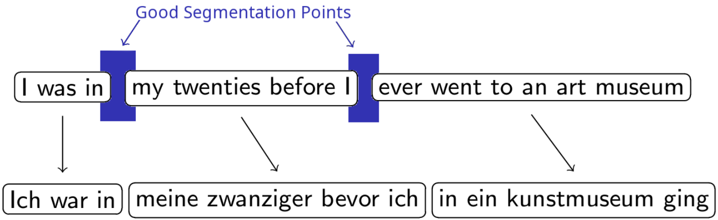
I	was	in	my	twenties	before	I	ever	went	to	an	art	museum
Ich	war	in	meine	zwanziger jahre	bevor	ich	je	ging	zu	ein	kunst	museum

- Reference Sentence:

  Ich war in meinen zwanzigern bevor ich erstmals in ein kunstmuseum ging

- BLEU Score: Low (15.6)
- Segments/Second: High

Segmentation - A Trade-off between Extremes



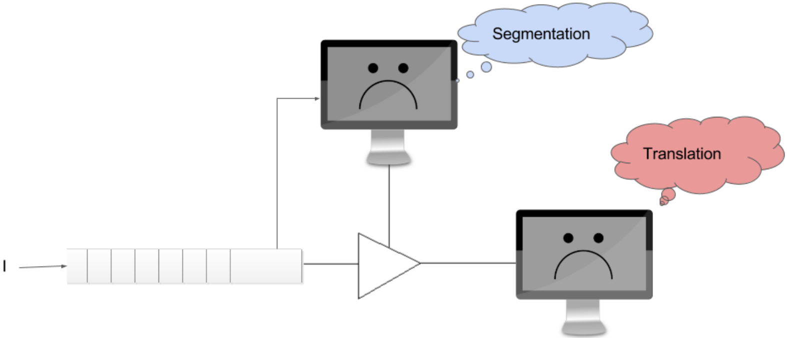
- Reference Sentence:

  Ich war in meinen zwanzigern bevor ich erstmals in ein kunstmuseum ging

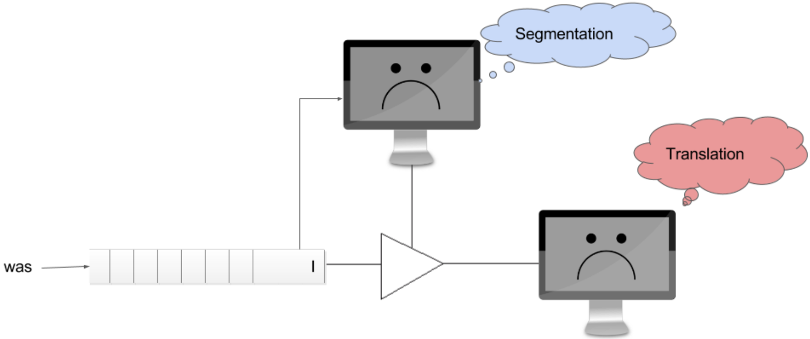
- BLEU Score: Acceptable (38.2)
- Segments/Second: Acceptable

Segmentation Classifier

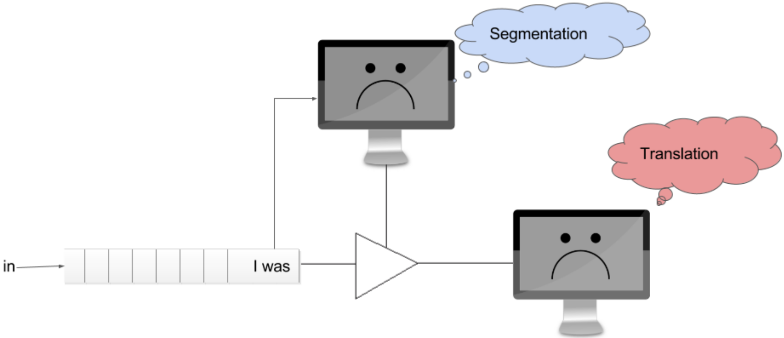
Segmentation Classifier



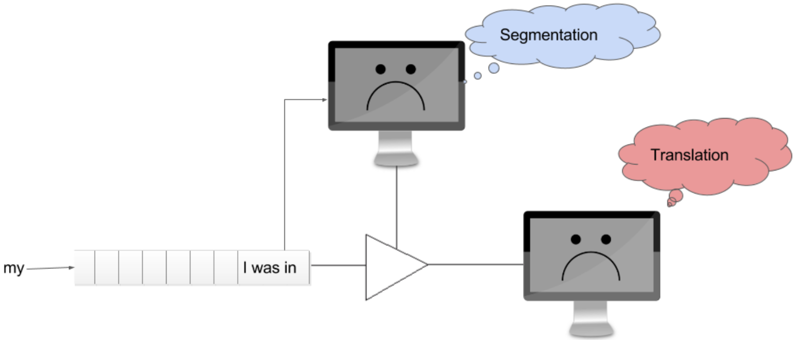
Segmentation Classifier



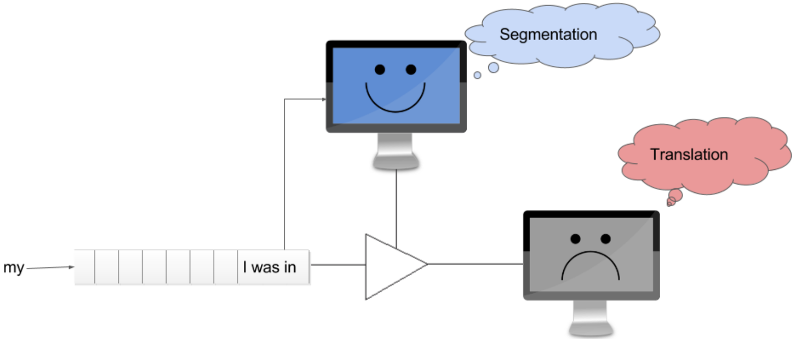
Segmentation Classifier



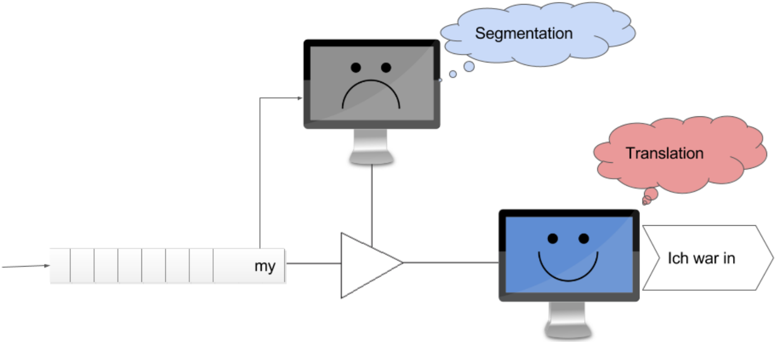
Segmentation Classifier



Segmentation Classifier



Segmentation Classifier



Classifier Data Annotation

Training Classifier Needs Annotated Data

* We are going to provide a method that
will create this annotated data

Classifier Data Annotation - An Example

- Task: English-German
- Features: Bigram part-of-speech tags
- Only source side is shown here !

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	.

N[noun], V[verb], D[determiner], J[adjective], P[preposition], S[possessive pronoun], A[adverb], R[particle], .[dot]

Example Data for Annotation - Feature frequencies

Feat	Freq	Feat	Freq	Feat	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	

Example Data for Annotation - Feature frequencies

Feat	Freq	Feat	Freq	Feat	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		

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	.

Greedy Segmentation Strategy

[Oda et al. 2014]

Greedy Segmentation Strategy

- Greedily maximize the sum of BLEU Scores of Sentences
  - Decoding is done Sentence by Sentence

Greedy Segmentation Strategy

- Greedily maximize the sum of BLEU Scores of Sentences
  - Decoding is done Sentence by Sentence

- Input: the desired average segment length (μ)

  $\Rightarrow$ total number of expected segments (K)

Greedy Segmentation Strategy

- Greedily maximize the sum of BLEU Scores of Sentences
  - Decoding is done Sentence by Sentence

- Input: the desired average segment length (μ)

  $\Rightarrow$ total number of expected segments (K)

$$\boxed{K = \left\lfloor \frac{\# Words}{\mu} \right\rfloor - [\# Sentences]}$$

* Sentence boundaries do not count towards K

Greedy Segmentation Strategy - An Example for $\mu = 13$

$$\boxed{K = 0 = \left\lfloor \frac{[\#Words{=}43]}{[\mu{=}13]} \right\rfloor - [\#Sentences = 3]}$$

Sum of BLEU Scores [of the 3 sentences] = 57.6

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	.

Greedy Segmentation Strategy - An Example for $\mu = 8$

$$\boxed{K = 2 = \left\lfloor \frac{[\#Words{=}43]}{[\mu{=}8]} \right\rfloor - [\#Sentences = 3]}$$

Sum of BLEU Scores [of the 3 sentences] $= 13.8$

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
N	V	P	S	N	P	N	A	V	P	D	N

museum	.
N	.

I	grew	up	in	the	middle	of	nowhere	on	a	dirt
N	V	R	P	D	N	P	N	P	D	N

road	in	rural	Arkansas	.
N	P	J	N	.

Greedy Segmentation Strategy - An Example for $\mu = 8$

$$\boxed{K = 2 = \left\lfloor \frac{[\#Words{=}43]}{[\mu{=}8]} \right\rfloor - [\#Sentences = 3]}$$

Sum of BLEU Scores [of the 3 sentences] = 27.2

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	.

Greedy Segmentation Strategy - An Example for $\mu = 8$

$$\boxed{K = 2 = \left\lfloor \frac{[\#Words{=}43]}{[\mu{=}8]} \right\rfloor - [\#Sentences = 3]}$$

Sum of BLEU Scores [of the 3 sentences] = 38.2

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	.

Greedy Segmentation Strategy - An Example for $\mu = 8$

$$\boxed{K = 2 = \left\lfloor \frac{[\#Words{=}43]}{[\mu{=}8]} \right\rfloor - [\#Sentences = 3]}$$

Sum of BLEU Scores [of the 3 sentences] = 38.2

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	.

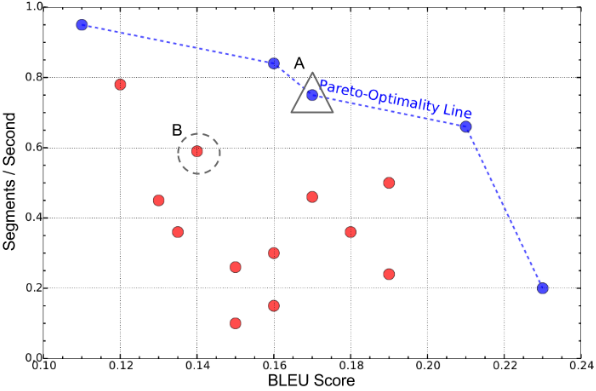
I	grew	up	in	the	middle	of	nowhere	on	a	dirt	road	in	rural	Arkansas	.
N	V	R	P	D	N	P	N	P	D	N	N	P	J	N	.

Only maximizes the BLEU score

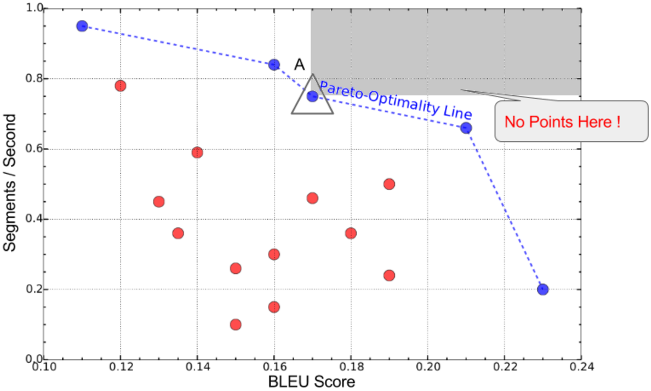
Tends to oversegment fewer sentences

Pareto-Optimal Segmentation Strategy

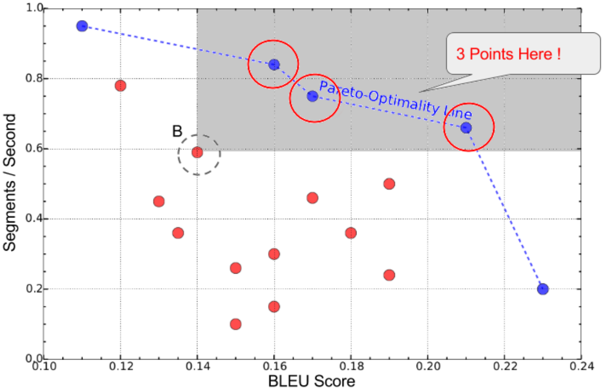
Pareto-Optimality



Pareto-Optimality



Pareto-Optimality



Pareto-Optimal Segmentation

- Tries to find the best segmentation points regarding both Accuracy and Segs/Sec
  - Our measure of accuracy is the average of $\{\frac{\textsc{Bleu}}{\#\mathit{Segments}}\}$ per sentence
- The input is the same desired average segment length μ

Pareto-Optimal Segmentation - An Example for $\mu = 8$

$$\boxed{K = 2 = \left\lfloor \frac{[\#Words=43]}{[\mu=8]} \right\rfloor - [\#Sentences = 3]}$$

Avg $\{\frac{\text{BLEU}}{\#Segments}\}$/ Sentence = 12.7, Segs/Sec = 0.560

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	.

Pareto-Optimal Segmentation - An Example for $\mu = 8$

$$\boxed{K = 2 = \left\lfloor \frac{[\#Words=43]}{[\mu=8]} \right\rfloor - [\#Sentences = 3]}$$

Avg $\left\{\frac{\textsc{Bleu}}{\#Segments}\right\}$/ Sentence $= 9.0$, Segs/Sec $= 0.956$

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	.

Sample Data Review

Feat	Freq	Feat	Freq	Feat	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	

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	.

Sample Data Review

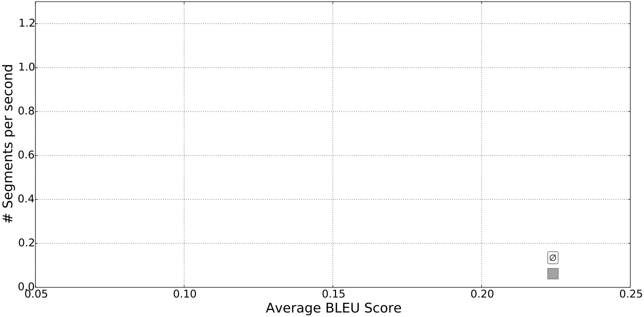
Feat	Freq	Feat	Freq	Feat	Freq
N-P	6	J-N	3	V-R	1
P-D	5	N-N	2	P-S	1
D-N	4	P-N	2	P-J	1
N-.	3	D-J	2	S-N	1
N-V	3	R-P	1	A-V	1
V-D	3	N-A	1		
Full Segmentation Set Size				40	

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	.

Pareto-Optimal Segmentation - Initiating the Segmentation

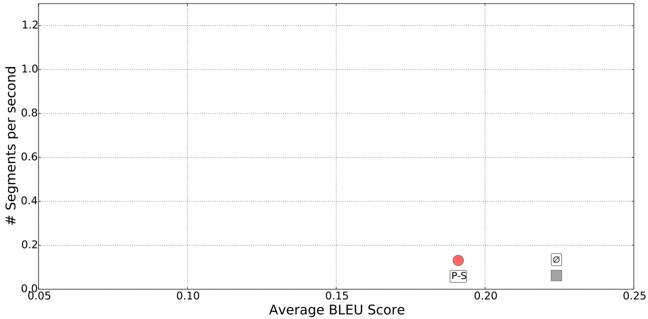


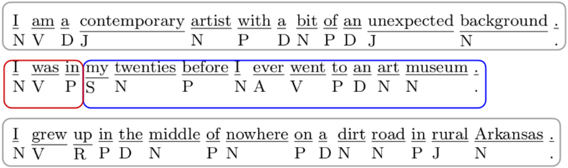
I am a contemporary artist with a bit of an unexpected background .
N V D J N P D N P D J N .

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 .

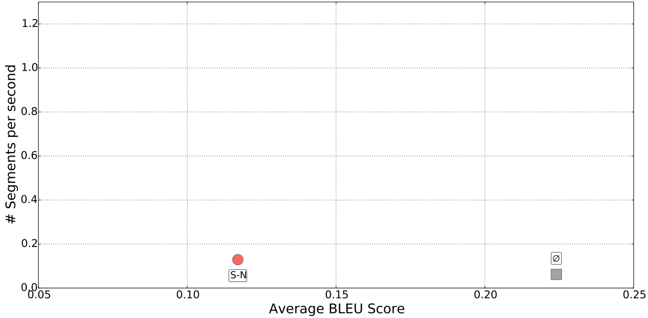
I grew up in the middle of nowhere on a dirt road in rural Arkansas .
N V R P D N P N P D N N P J N .

Pareto-Optimal Segmentation - Searching for first point





Pareto-Optimal Segmentation - Searching for first point

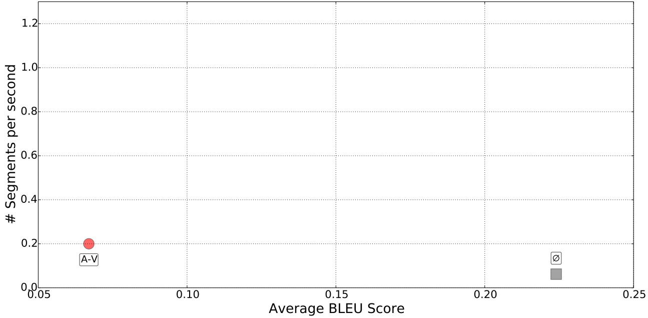


I am a contemporary artist with a bit of an unexpected background .
N V D J N P D N P D J N .

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 .

Pareto-Optimal Segmentation - Searching for first point

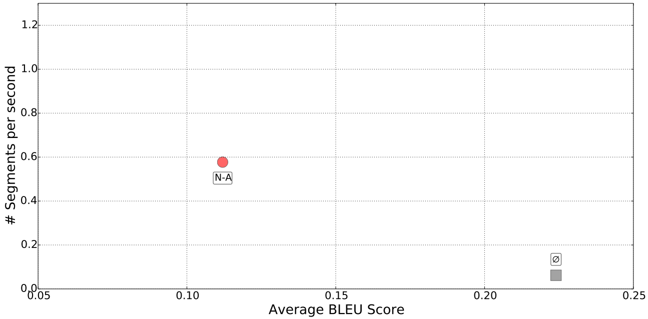


I am a contemporary artist with a bit of an unexpected background .
N V D J N P D N P D J N .

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 .

Pareto-Optimal Segmentation - Searching for first point



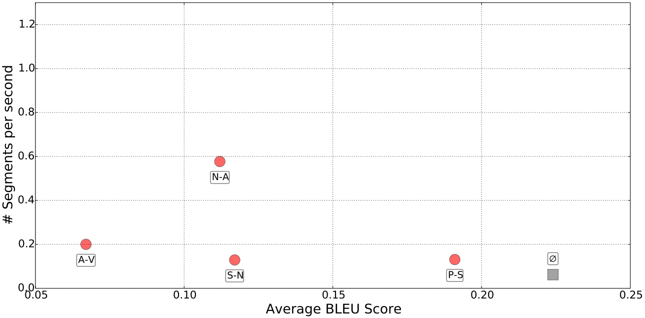
I	am	a	contemporary	artist	with	a	bit	of	an	unexpected	background	.
N	V	D	J	N	P	D	N	P	D	J	N	.

I	was	in	my	twenties	before	I
N	V	P	S	N	P	N

ever	went	to	an	art	museum	.
A	V	P	D	N	N	.

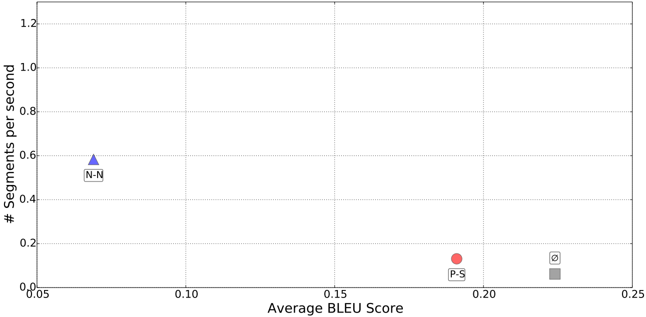
I	grew	up	in	the	middle	of	nowhere	on	a	dirt	road	in	rural	Arkansas	.
N	V	R	P	D	N	P	N	P	D	N	N	P	J	N	.

Pareto-Optimal Segmentation - Searching for first point



Pareto-Optimal Segmentation - Searching for first point

Pareto-Optimal Segmentation - Searching for second point



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

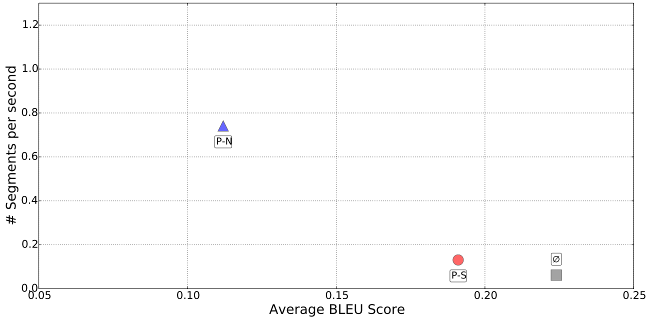
I	was	in	my	twenties	before	I	ever	went	to	an	art
N	V	P	S	N	P	N	A	V	P	D	N

museum	.
N	.

I	grew	up	in	the	middle	of	nowhere	on	a	dirt
N	V	R	P	D	N	P	N	P	D	N

road	in	rural	Arkansas	.
N	P	J	N	.

Pareto-Optimal Segmentation - Searching for second point



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

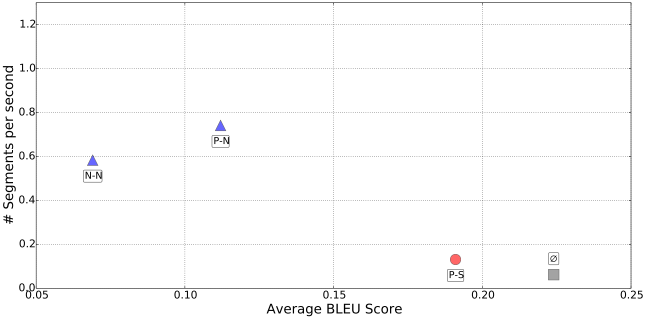
I	was	in	my	twenties	before
N	V	P	S	N	P

I	ever	went	to	an	art	museum	.
N	A	V	P	D	N	N	.

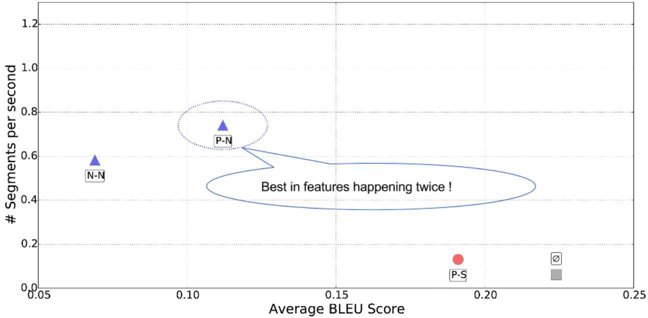
I	grew	up	in	the	middle	of
N	V	R	P	D	N	P

nowhere	on	a	dirt	road	in	rural	Arkansas	.
N	P	D	N	N	P	J	N	.

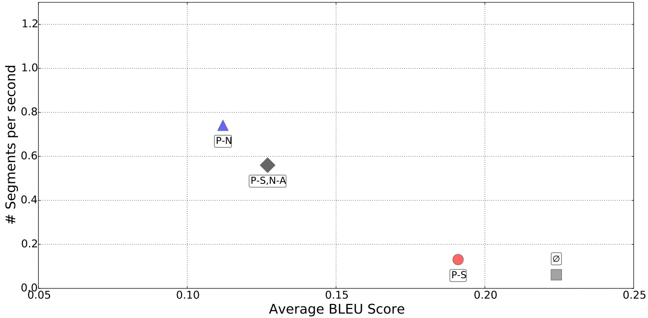
Pareto-Optimal Segmentation - Searching for second point



Pareto-Optimal Segmentation - Searching for second point



Pareto-Optimal Segmentation - Searching for second point

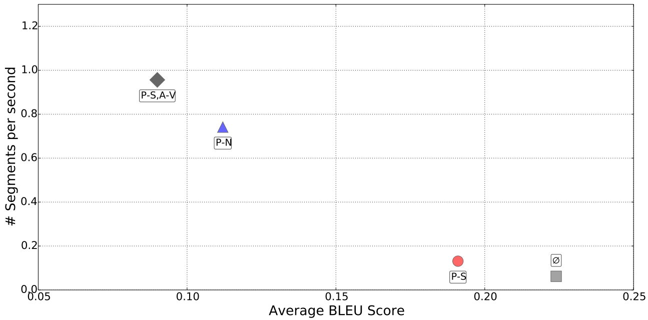


I am a contemporary artist with a bit of an unexpected background .
N V D J N P D N P D J N .

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 .

Pareto-Optimal Segmentation - Searching for second point

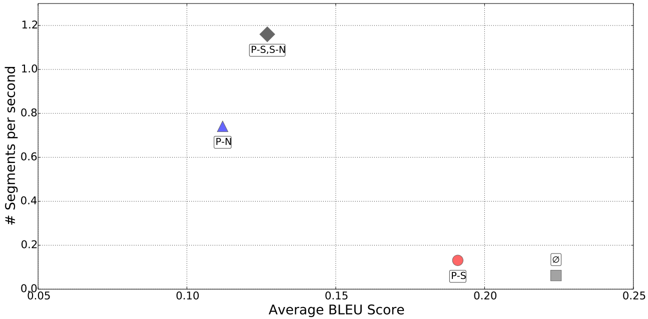


I	am	a	contemporary	artist	with	a	bit	of	an	unexpected	background	.
N	V	D	J	N	P	D	N	P	D	J	N	.

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	.

Pareto-Optimal Segmentation - Searching for second point

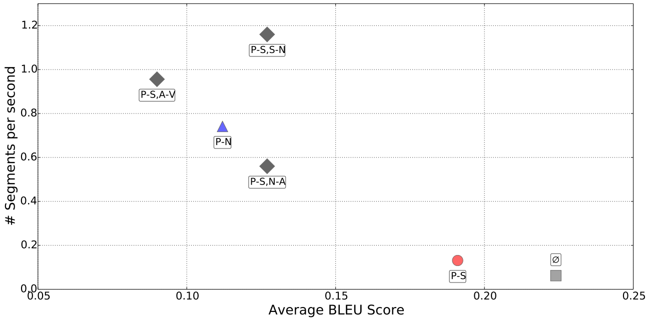


I am a contemporary artist with a bit of an unexpected background .
N V D J N P D N P D J N .

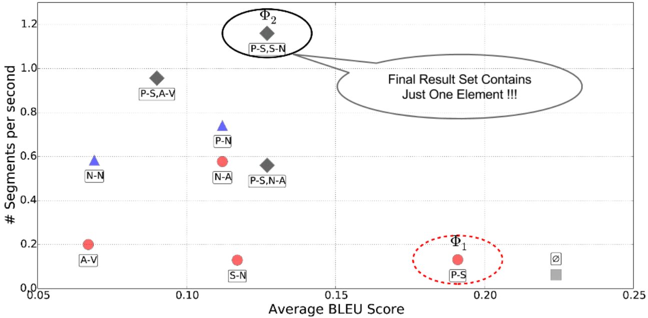
I was in | my | twenties before I ever went to an art museum .
N V P | S | N P N A V P D N N .

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 .

Pareto-Optimal Segmentation - Searching for second point



Pareto-Optimal Segmentation - Searching for second point



Experiments and Results

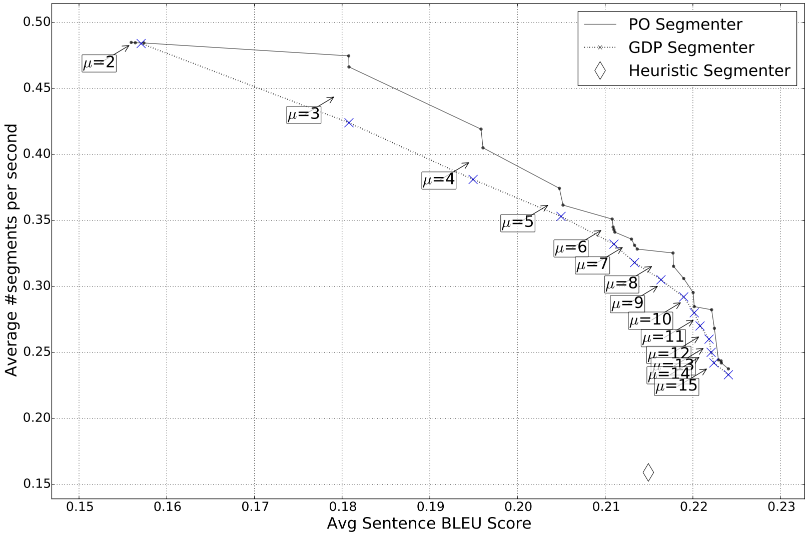
Experimental Setup

- Task: English-German TED speech translation
- MT System Training Data: IWSLT 2013 Train data + half of the Europarl data [Koehn 2005]
- MT System Tuning Data: IWSLT Test 2012
- German Language Model Data: monolingual data from WMT 2013 Shared Task
- Segmenter Training Data: IWSLT Dev 2010 and 2012 and Test 2010
- Segmenter Test Data: IWSLT Test 2013
- Segmentation Train Size: 3669
- Segmentation Test Size: 1025

Accuracy vs. Latency-Accuracy Evaluation Experiment

- We compared
  - the state-of-the-art heuristic speech segmenter [Rangarajan et al. 2013]
  - Greedy Segmentation Approach [Oda et al. 2014]
  - Pareto-Optimal Segmentation Approach

Results on the Test Data



Result comparison for $\mu = 3$ and $\mu = 8$

	$\mu = 3$		$\mu = 8$	
	Segs/Sec	BLEU	Segs/Sec	BLEU
Pareto-Optimal Segmenter	**0.474**	18.07	**0.315**	**21.77**
Greedy Segmenter	0.424	18.07	0.305	21.63

Summary

In this work we:

- Concentrated on the problem of data annotation for training the segmentation classifier
- Presented a multi-metric optimization algorithm over both latency and accuracy to solve the problem
- Showed that our algorithm performs better than the state-of-the-art methods
  - While we managed to keep the same translation quality of the state-of-the-art

We Aim To:

- Extend this work with a larger variety of features
- Use the annotated data to fine-tune the simultaneous translation system
  - Which results in pushing "the knee of the plot" further

Thank You!

contact: sshavara@sfu.ca

Pareto-Optimal Segmentation - Algorithm

Algorithm 1 Pareto-Optimal Segmentation

1: $\mathcal{S}_0^* \leftarrow \emptyset$
2: **for** $k = 1$ to K **do**
3:
$$\mathcal{S}_k^* \leftarrow \underset{p \in FSS \wedge p \notin \mathcal{S}_{k-1}^*}{\text{arg pareto frontier}} \left\{ \begin{array}{c} B_\alpha(\mathcal{S}_{k-1}^* \cup \{p\}), \\ \Lambda_\alpha(\mathcal{S}_{k-1}^* \cup \{p\}) \end{array} \right\}$$
4: **end for**
5: **return** $\mathcal{S}_K^*$

Pareto-Optimal Segmentation - Efficient Algorithm

Algorithm 2 Computationally Efficient 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 $s(\mathcal{F}, \Phi_K)$

Pareto-Optimal Segmentation - Formulae

- K and μ are the same as Greedy Segmentation Strategy
- Accuracy measure

$$B_\alpha(s) = \sum_{j=1}^{N} \frac{\beta(\mathcal{D}(f_j,s_j),e_j)}{|s_j|} - \alpha|\Phi|$$

- Latency measure

$$\Lambda_\alpha(s) = \frac{|s|}{\sum_{j=1}^{N} \gamma(\mathcal{D}(f_j,s))} - \alpha|\Phi|$$

- The best set of segmentation strategies

$$\mathcal{S}^* = \text{arg pareto frontier}_{s\in\mathcal{S}_{all}} \{B_\alpha(s), \Lambda_\alpha(s)\}$$

Size of Data used in Experiments

	Sentences	Types	Tokens
MT Train	1033491	105267	27948041
MT Tune	1730	3937	31568
Seg Train	3669	6773	74883
Seg Test	1025	3181	22026

Greedy Segmentation Strategy - Formulae

- total number of expected segments in the corpus (K)

$$K := max(0, \left\lfloor \frac{\sum_{f \in F} |f|}{\mu} \right\rfloor - N)$$

- $\mu =$ the average expected segment length
- Accuracy measure

$$B_\alpha(s) = \sum_{j=1}^{N} \beta(\mathcal{D}(f_j, s), e_j) - \alpha|\Phi|$$

- The best set of segmentation strategy

$$\mathcal{S}^* = \text{argmax}_{s \in \mathcal{S}_{all}} \{B_\alpha(s)\}$$