Learning Segmentations that Balance Latency versus Quality in Spoken Language Translation

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

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 erstmals in ein kunstmuseum ging

▶ Reference Sentence:

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

▶ BLEU Score: Low (15.6)

▶ Segments/Second: High
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

my → I was in

Segmentation

Translation
Segmentation Classifier
* 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.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
### Example Data for Annotation - Feature frequencies

<table>
<thead>
<tr>
<th>Feat</th>
<th>Freq</th>
<th>Feat</th>
<th>Freq</th>
<th>Feat</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-P</td>
<td>6</td>
<td>J-N</td>
<td>3</td>
<td>V-R</td>
<td>1</td>
</tr>
<tr>
<td>P-D</td>
<td>5</td>
<td>N-N</td>
<td>2</td>
<td>P-S</td>
<td>1</td>
</tr>
<tr>
<td>D-N</td>
<td>4</td>
<td>P-N</td>
<td>2</td>
<td>P-J</td>
<td>1</td>
</tr>
<tr>
<td>N-.</td>
<td>3</td>
<td>D-J</td>
<td>2</td>
<td>S-N</td>
<td>1</td>
</tr>
<tr>
<td>N-V</td>
<td>3</td>
<td>R-P</td>
<td>1</td>
<td>A-V</td>
<td>1</td>
</tr>
<tr>
<td>V-D</td>
<td>3</td>
<td>N-A</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Full Segmentation Set Size**: 40
Example Data for Annotation - Feature frequencies

<table>
<thead>
<tr>
<th>Feat</th>
<th>Freq</th>
<th>Feat</th>
<th>Freq</th>
<th>Feat</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-P</td>
<td>6</td>
<td>J-N</td>
<td>3</td>
<td>V-R</td>
<td>1</td>
</tr>
<tr>
<td>P-D</td>
<td>5</td>
<td>N-N</td>
<td>2</td>
<td>P-S</td>
<td>1</td>
</tr>
<tr>
<td>D-N</td>
<td>4</td>
<td>P-N</td>
<td>2</td>
<td>P-J</td>
<td>1</td>
</tr>
<tr>
<td>N-</td>
<td>3</td>
<td>D-J</td>
<td>2</td>
<td>S-N</td>
<td>1</td>
</tr>
<tr>
<td>N-V</td>
<td>3</td>
<td>R-P</td>
<td>1</td>
<td>A-V</td>
<td>1</td>
</tr>
<tr>
<td>V-D</td>
<td>3</td>
<td>N-A</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 ($\mu$)
  $\Rightarrow$ 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** ($\mu$)
  - total number of expected segments ($K$)

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

* Sentence boundaries do not count towards $K$
Greedy Segmentation Strategy - An Example for $\mu = 13$

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

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

I am a contemporary artist with a bit of an unexpected background. 

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Greedy Segmentation Strategy - An Example for $\mu = 8$

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

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

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Greedy Segmentation Strategy - An Example for $\mu = 8$

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

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

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Greedy Segmentation Strategy - An Example for $\mu = 8$

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

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

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Greedy Segmentation Strategy - An Example for $\mu = 8$

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

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

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.

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 \( \frac{\text{BLEU}}{\text{#Segments}} \) per sentence
- The input is the same desired average segment length \( \mu \)
Pareto-Optimal Segmentation - An Example for $\mu = 8$

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

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

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Pareto-Optimal Segmentation - An Example for $\mu = 8$

$$K = 2 = \left\lceil \frac{\text{#Words} = 43}{\mu = 8} \right\rceil - \text{#Sentences} = 3$$

Avg $\{ \frac{\text{BLEU}}{\text{#Segments}} \} / \text{Sentence} = 9.0$, Segs/Sec = 0.956

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
### Sample Data Review

<table>
<thead>
<tr>
<th>Feat</th>
<th>Freq</th>
<th>Feat</th>
<th>Freq</th>
<th>Feat</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-P</td>
<td>6</td>
<td>J-N</td>
<td>3</td>
<td>V-R</td>
<td>1</td>
</tr>
<tr>
<td>P-D</td>
<td>5</td>
<td>N-N</td>
<td>2</td>
<td>P-S</td>
<td>1</td>
</tr>
<tr>
<td>D-N</td>
<td>4</td>
<td>P-N</td>
<td>2</td>
<td>P-J</td>
<td>1</td>
</tr>
<tr>
<td>N-</td>
<td>3</td>
<td>D-J</td>
<td>2</td>
<td>S-N</td>
<td>1</td>
</tr>
<tr>
<td>N-V</td>
<td>3</td>
<td>R-P</td>
<td>1</td>
<td>A-V</td>
<td>1</td>
</tr>
<tr>
<td>V-D</td>
<td>3</td>
<td>N-A</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Full Segmentation Set Size** 40

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Sample Data Review

<table>
<thead>
<tr>
<th>Feat</th>
<th>Freq</th>
<th>Feat</th>
<th>Freq</th>
<th>Feat</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-P</td>
<td>6</td>
<td>J-N</td>
<td>3</td>
<td>V-R</td>
<td>1</td>
</tr>
<tr>
<td>P-D</td>
<td>5</td>
<td>N-N</td>
<td>2</td>
<td>P-S</td>
<td>1</td>
</tr>
<tr>
<td>D-N</td>
<td>4</td>
<td>P-N</td>
<td>2</td>
<td>P-J</td>
<td>1</td>
</tr>
<tr>
<td>N-.</td>
<td>3</td>
<td>D-J</td>
<td>2</td>
<td>S-N</td>
<td>1</td>
</tr>
<tr>
<td>N-V</td>
<td>3</td>
<td>R-P</td>
<td>1</td>
<td>A-V</td>
<td>1</td>
</tr>
<tr>
<td>V-D</td>
<td>3</td>
<td>N-A</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Full Segmentation Set Size 40

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Pareto-Optimal Segmentation - Initiating the Segmentation

---

I am a contemporary artist with a bit of an unexpected background.

\[
\begin{array}{cc}
\text{N} & \text{V} \\
\text{D} & \text{J}
\end{array}
\]

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

\[
\begin{array}{cc}
\text{N} & \text{V} \\
\text{P} & \text{S} \\
\text{N}
\end{array}
\]

I grew up in the middle of nowhere on a dirt road in rural Arkansas.

\[
\begin{array}{cc}
\text{N} & \text{V} \\
\text{R} & \text{P} \\
\text{D} & \text{N} \\
\text{P} & \text{N} \\
\text{P} & \text{D} \\
\text{N} & \text{N} \\
\text{P} & \text{J} \\
\text{N}
\end{array}
\]
Pareto-Optimal Segmentation - Searching for first point

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Pareto-Optimal Segmentation - Searching for first point

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Pareto-Optimal Segmentation - Searching for first point

\[
\begin{array}{cccccccc}
\end{array}
\]

\[
\begin{array}{cccccccc}
N & V & P & S & N & P & N & A \\
\end{array}
\]

\[
\begin{array}{cccccccc}
\end{array}
\]
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
Pareto-Optimal Segmentation - Searching for first point

![Graph showing Pareto-Optimal Segmentation]

- Best Segs/Sec So Far!
- Best Accuracy in First Search!
Pareto-Optimal Segmentation - Searching for second point

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Pareto-Optimal Segmentation - Searching for second point

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.

---

Average BLEU Score

<table>
<thead>
<tr>
<th># Segments per second</th>
<th>Average BLEU Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.2</td>
</tr>
<tr>
<td>0.10</td>
<td>0.4</td>
</tr>
<tr>
<td>0.15</td>
<td>0.6</td>
</tr>
<tr>
<td>0.20</td>
<td>0.8</td>
</tr>
<tr>
<td>0.25</td>
<td>1.0</td>
</tr>
<tr>
<td>0.30</td>
<td>1.2</td>
</tr>
</tbody>
</table>
Pareto-Optimal Segmentation - Searching for second point
Pareto-Optimal Segmentation - Searching for second point

Best in features happening twice!
Pareto-Optimal Segmentation - Searching for second point

![Graph showing average BLEU score versus # segments per second with points labeled P-S, P-N, P-S,N-A.]

- I am a contemporary artist with a bit of an unexpected background.
- I was in my twenties before I ever went to an art museum.
- I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Pareto-Optimal Segmentation - Searching for second point

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Pareto-Optimal Segmentation - Searching for second point

I am a contemporary artist with a bit of an unexpected background.

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

I grew up in the middle of nowhere on a dirt road in rural Arkansas.
Pareto-Optimal Segmentation - Searching for second point

![Graph showing Pareto-Optimal Segmentation](image-url)
Pareto-Optimal Segmentation - Searching for second point

Final Result Set Contains Just One Element !!!

Average BLEU Score

# Segments per second
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

- **PO Segmenter**
- **GDP Segmenter**
- **Heuristic Segmenter**

- **Avg Sentence BLEU Score**
  - 0.15
  - 0.16
  - 0.17
  - 0.18
  - 0.19
  - 0.20
  - 0.21
  - 0.22
  - 0.23

- **Average #segments per second**
  - $\mu = 2$
  - $\mu = 3$
  - $\mu = 4$
  - $\mu = 5$
  - $\mu = 6$
  - $\mu = 7$
  - $\mu = 8$
  - $\mu = 9$
  - $\mu = 10$
  - $\mu = 11$
  - $\mu = 12$
  - $\mu = 13$
  - $\mu = 14$
  - $\mu = 15$
Result comparison for $\mu = 3$ and $\mu = 8$

<table>
<thead>
<tr>
<th></th>
<th>$\mu = 3$</th>
<th></th>
<th>$\mu = 8$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Segs/Sec</td>
<td>BLEU</td>
<td>Segs/Sec</td>
<td>BLEU</td>
</tr>
<tr>
<td>Pareto-Optimal Segmenter</td>
<td>0.474</td>
<td>18.07</td>
<td>0.315</td>
<td>21.77</td>
</tr>
<tr>
<td>Greedy Segmenter</td>
<td>0.424</td>
<td>18.07</td>
<td>0.305</td>
<td>21.63</td>
</tr>
</tbody>
</table>
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
Algorithm 1 Pareto-Optimal Segmentation

1: $S_0^* \leftarrow \emptyset$
2: for $k = 1$ to $K$ do
3:     $S_k^* \leftarrow \arg\text{pareto frontier}_{p \in FSS \land p \notin S_{k-1}^*} \left\{ B_\alpha(S_{k-1}^* \cup \{p\}), \Lambda_\alpha(S_{k-1}^* \cup \{p\}) \right\}$
4: end for
5: return $S_K^*$
Algorithm 2: Computationally Efficient Pareto-Optimal Segmentation

1: $\Phi_0 \leftarrow \emptyset$
2: for $k = 1$ to $K$ do
3:     for $j = 0$ to $k - 1$ do
4:         $\Phi' \leftarrow \{\phi : (\phi \notin \Phi_j) \land \text{count}(\phi; \mathcal{F}) = k - j\}$
5:         $\Phi_{k,j} \leftarrow \Phi_j \cup \left\{ \text{arg pareto frontier}_{\phi \in \Phi'} \left\{ B_\alpha(s(\mathcal{F}, \Phi_j \cup \{\phi\})), \Lambda_\alpha(s(\mathcal{F}, \Phi_j \cup \{\phi\})) \right\} \right\}$
6:     end for
7:     if $k < K$ then
8:         $\Phi_{k,j} \leftarrow \underset{\phi \in \{\Phi_{k,j} : 0 \leq j \leq k\}}{\text{argmax}} B_\alpha(s(\mathcal{F}, \phi))$
9:     end if
10:    $\Phi_k \leftarrow \text{arg pareto frontier}_{\Phi \in \{\Phi_{k,j} : 0 \leq j \leq k\}} \left\{ B_\alpha(s(\mathcal{F}, \Phi)), \Lambda_\alpha(s(\mathcal{F}, \Phi)) \right\}$
11: end for
12: return $s(\mathcal{F}, \Phi_K)$
Pareto-Optimal Segmentation - Formulae

- $K$ and $\mu$ are the same as Greedy Segmentation Strategy
- Accuracy measure

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

- Latency measure

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

- The best set of segmentation strategies

$$S^* = \arg\text{ pareto frontier}_{s \in S_{all}} \{B_\alpha(s), \Lambda_\alpha(s)\}$$
Size of Data used in Experiments

<table>
<thead>
<tr>
<th></th>
<th>Sentences</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT Train</td>
<td>1033491</td>
<td>105267</td>
<td>27948041</td>
</tr>
<tr>
<td>MT Tune</td>
<td>1730</td>
<td>3937</td>
<td>31568</td>
</tr>
<tr>
<td>Seg Train</td>
<td>3669</td>
<td>6773</td>
<td>74883</td>
</tr>
<tr>
<td>Seg Test</td>
<td>1025</td>
<td>3181</td>
<td>22026</td>
</tr>
</tbody>
</table>
Greedy Segmentation Strategy - Formulae

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

\[ K := \max(0, \left\lceil \sum_{f \in F} \frac{|f|}{\mu} \right\rceil - N) \]

- $\mu = \text{the average expected segment length}$

- 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

\[ S^* = \arg\max_{s \in S_{all}} \{B_\alpha(s)\} \]