Large-scale Discriminative n-gram Language Models for Statistical Machine Translation

Zhifei Li and Sanjeev Khudanpur
Johns Hopkins University
Language Model: Data Mismatch
Language Model: Data Mismatch

- A regular language model is trained on well-formed monolingual corpora (e.g., Gigaword)
Language Model: Data Mismatch

- A regular language model is trained on well-formed monolingual corpora (e.g., Gigaword)
  - it does not require bilingual data
Language Model: Data Mismatch

- A regular language model is trained on well-formed monolingual corpora (e.g., Gigaword)
  - it does not require bilingual data
- During training, the language model does not see the MT outputs
Language Model: Data Mismatch

- A regular language model is trained on well-formed monolingual corpora (e.g., Gigaword)
  - it does not require bilingual data
- During training, the language model does not see the MT outputs
  - But, the LM will be used to rank MT outputs
Language Model: Data Mismatch

- A regular language model is trained on well-formed monolingual corpora (e.g., Gigaword)
  - it does not require bilingual data
- During training, the language model does not see the MT outputs
  - But, the LM will be used to rank MT outputs
- MT outputs differ substantially from Gigaword
Language Model: Data Mismatch

- A regular language model is trained on well-formed monolingual corpora (e.g., Gigaword)
  - it does not require bilingual data

- During training, the language model does not see the MT outputs
  - But, the LM will be used to rank MT outputs
  - MT outputs differ substantially from Gigaword

- Can we make the LM task-specific without losing its big advantage in using enormous monolingual data?
Task: reranking MT outputs
Task: reranking MT outputs

我是最好的翻译。
Task: reranking MT outputs

我是最好的翻译。
Task: reranking MT outputs

我是最好的翻译。 I am the best translation.
Task: reranking MT outputs

<table>
<thead>
<tr>
<th>Hypothesized translation</th>
<th>TM</th>
<th>LM</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am a most best translation .</td>
<td>9.1</td>
<td>10</td>
<td>19.1</td>
</tr>
<tr>
<td>I are the best translation .</td>
<td>9.0</td>
<td>10</td>
<td>19.0</td>
</tr>
<tr>
<td><strong>I am the best translation</strong> .</td>
<td><strong>10</strong></td>
<td><strong>8</strong></td>
<td><strong>18</strong></td>
</tr>
<tr>
<td>I are the good translate .</td>
<td>9</td>
<td>8</td>
<td>17</td>
</tr>
</tbody>
</table>

I am the best translation.
**Task: reranking MT outputs**

I am the best translation.

<table>
<thead>
<tr>
<th>Hypothesized translation</th>
<th>TM</th>
<th>LM</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>i am a most best translation .</td>
<td>9.1</td>
<td>10</td>
<td>19.1</td>
</tr>
<tr>
<td>i are the best translation .</td>
<td>9.0</td>
<td>10</td>
<td>19.0</td>
</tr>
<tr>
<td>i am the best translation .</td>
<td>10</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>i are the good translate .</td>
<td>9</td>
<td>8</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesized translation</th>
<th>TM</th>
<th>LM</th>
<th>Corrective</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>i am the best translation .</td>
<td>10</td>
<td>8</td>
<td>1.0</td>
<td>19</td>
</tr>
<tr>
<td>i am a most best translation .</td>
<td>9.1</td>
<td>10</td>
<td>-0.5</td>
<td>18.6</td>
</tr>
<tr>
<td>i are the best translation .</td>
<td>9.0</td>
<td>10</td>
<td>-0.5</td>
<td>18.5</td>
</tr>
<tr>
<td>i are the good translate .</td>
<td>9</td>
<td>8</td>
<td>-1</td>
<td>16</td>
</tr>
</tbody>
</table>
## Task: reranking MT outputs

I am the best translation.

<table>
<thead>
<tr>
<th>Hypothesized translation</th>
<th>TM</th>
<th>LM</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>i am a most best translation .</td>
<td>9.1</td>
<td>10</td>
<td>19.1</td>
</tr>
<tr>
<td>i are the best translation .</td>
<td>9.0</td>
<td>10</td>
<td>19.0</td>
</tr>
<tr>
<td>i am the best translation .</td>
<td>10</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>i are the good translate .</td>
<td>9</td>
<td>8</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesized translation</th>
<th>TM</th>
<th>LM</th>
<th>Corrective</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>i am the best translation .</td>
<td>10</td>
<td>8</td>
<td>1.0</td>
<td>19</td>
</tr>
<tr>
<td>i am a most best translation .</td>
<td>9.1</td>
<td>10</td>
<td>-0.5</td>
<td>18.6</td>
</tr>
<tr>
<td>i are the best translation .</td>
<td>9.0</td>
<td>10</td>
<td>-0.5</td>
<td>18.5</td>
</tr>
<tr>
<td>i are the good translate .</td>
<td>9</td>
<td>8</td>
<td>-1</td>
<td>16</td>
</tr>
</tbody>
</table>
**Task: reranking MT outputs**

I am the best translation.

<table>
<thead>
<tr>
<th>Hypothesized translation</th>
<th>TM</th>
<th>LM</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am a most best translation .</td>
<td>9.1</td>
<td>10</td>
<td>19.1</td>
</tr>
<tr>
<td>I are the best translation .</td>
<td>9.0</td>
<td>10</td>
<td>19.0</td>
</tr>
<tr>
<td>I am the best translation .</td>
<td>10</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>I are the good translate .</td>
<td>9</td>
<td>8</td>
<td>17</td>
</tr>
</tbody>
</table>

**Corrective**

<table>
<thead>
<tr>
<th>Hypothesized translation</th>
<th>TM</th>
<th>LM</th>
<th>Corrective</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am the best translation .</td>
<td>10</td>
<td>8</td>
<td>1.0</td>
<td>19</td>
</tr>
<tr>
<td>I am a most best translation .</td>
<td>9.1</td>
<td>10</td>
<td>-0.5</td>
<td>18.6</td>
</tr>
<tr>
<td>I are the best translation .</td>
<td>9.0</td>
<td>10</td>
<td>-0.5</td>
<td>18.5</td>
</tr>
<tr>
<td>I are the good translate .</td>
<td>9</td>
<td>8</td>
<td>-1</td>
<td>16</td>
</tr>
</tbody>
</table>
Task: reranking MT outputs

<table>
<thead>
<tr>
<th>Hypothesized translation</th>
<th>TM</th>
<th>LM</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>i am the best translation.</td>
<td>10</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>i am a most best translation.</td>
<td>9.1</td>
<td>10</td>
<td>19.1</td>
</tr>
<tr>
<td>i are the best translation.</td>
<td>9.0</td>
<td>10</td>
<td>19.0</td>
</tr>
<tr>
<td>i are the good translate.</td>
<td>9</td>
<td>8</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesized translation</th>
<th>TM</th>
<th>LM</th>
<th>Corrective</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>i am the best translation.</td>
<td>10</td>
<td>8</td>
<td>1.0</td>
<td>19</td>
</tr>
<tr>
<td>i am a most best translation.</td>
<td>9.1</td>
<td>10</td>
<td>-0.5</td>
<td>18.6</td>
</tr>
<tr>
<td>i are the best translation.</td>
<td>9.0</td>
<td>10</td>
<td>-0.5</td>
<td>18.5</td>
</tr>
<tr>
<td>i are the good translate.</td>
<td>9</td>
<td>8</td>
<td>-1</td>
<td>16</td>
</tr>
</tbody>
</table>
Discriminative LM reranking
Discriminative LM reranking

- A discriminative language model should
Discriminative LM reranking

- A discriminative language model should
  - discover useful n-gram features
Discriminative LM reranking

- A discriminative language model should
  - discover useful n-gram features
  - find optimal weights for these features
A discriminative language model should
- discover useful n-gram features
- find optimal weights for these features

The discriminative LM is trained on
Discriminative LM reranking

- **A discriminative language model should**
  - discover useful n-gram features
  - find optimal weights for these features
- **The discriminative LM is trained on**
  - hypotheses produced by a baseline system
Discriminative LM reranking

- A discriminative language model should
  - discover useful n-gram features
  - find optimal weights for these features
- The discriminative LM is trained on
  - hypotheses produced by a baseline system
  - desired translation
Discriminative Modeling
Discriminative Modeling

- **Global linear model**

\[ s(f, e) = \Phi(f, e) \cdot \bar{\alpha} = \sum_j \Phi_j(f, e) \alpha_j \]
Discriminative Modeling

- **Global linear model**

\[ s(f, e) = \Phi(f, e) \cdot \bar{\alpha} = \sum_j \Phi_j(f, e) \alpha_j \]

- **Training**

\[ \bar{\alpha}^* = \arg \max_{\bar{\alpha}} F(Data, \bar{\alpha}) \]
Discriminative Modeling

- **Global linear model**

  \[ s(f, e) = \Phi(f, e) \cdot \tilde{\alpha} = \sum_j \Phi_j(f, e) \alpha_j \]

- **Training**

  \[ \tilde{\alpha}^* = \arg \max_{\tilde{\alpha}} F(Data, \tilde{\alpha}) \]

*Perceptron*
Discriminative Modeling

- **Global linear model**

  \[
  s(f, e) = \Phi(f, e) \cdot \bar{\alpha} = \sum_j \Phi_j(f, e) \alpha_j
  \]

- **Training**

  \[
  \bar{\alpha}^* = \arg\max_{\bar{\alpha}} F(Data, \bar{\alpha})
  \]
Discriminative Modeling

- **Global linear model**

  \[ s(f, e) = \Phi(f, e) \cdot \tilde{\alpha} = \sum_j \Phi_j(f, e) \alpha_j \]

- **Training**

  \[ \tilde{\alpha}^* = \arg \max_{\tilde{\alpha}} F(Data, \tilde{\alpha}) \]
Discriminative Modeling

- **Global linear model**
  \[ s(f, e) = \Phi(f, e) \cdot \vec{\alpha} = \sum_j \Phi_j(f, e)\alpha_j \]

- **Training**
  \[ \vec{\alpha}^* = \arg \max_{\vec{\alpha}} F(Data, \vec{\alpha}) \]
Discriminative Modeling

- **Global linear model**

\[ s(f, e) = \Phi(f, e) \cdot \vec{\alpha} = \sum_{j} \Phi_j(f, e) \alpha_j \]

- **Training**

\[ \vec{\alpha}^* = \arg \max_{\vec{\alpha}} F(Data, \vec{\alpha}) \]

- **Decision rule**

\[ e^* = \arg \max_{e \in \text{TRANS}(f)} s(f, e) \]
Discriminative Reranking

- Score after reranking

- Features
Discriminative Reranking

- Score after reranking
  \[ s(f, e) = \Phi(f, e) \cdot \tilde{\alpha} \]

- Features
Discriminative Reranking

- **Score after reranking**
  \[ s(f, e) = \Phi(f, e) \cdot \alpha \]
  \[ = \alpha_0 \Phi_0(f, e) + \]

- **Features**
  - *baseline feature*
    \[ \Phi_0(f, e) = \text{Baseline score for translation } e \]
Discriminative Reranking

- Score after reranking
  \[ s(f, e) = \Phi(f, e) \cdot \vec{\alpha} = \alpha_0 \Phi_0(f, e) + \sum_{j \in [1,J]} \alpha_j \Phi_j(f, e) \]

- Features
  - baseline feature
    \[ \Phi_0(f, e) = \text{Baseline score for translation } e \]
  - reranking n-gram features, e.g.,
Discriminative Reranking

- **Score after reranking**

\[
s(f, e) = \Phi(f, e) \cdot \tilde{\alpha} = \alpha_0 \Phi_0(f, e) + \sum_{j \in [1, J]} \alpha_j \Phi_j(f, e)
\]

- **Features**
  - **baseline feature**
    \[
    \Phi_0(f, e) = \text{Baseline score for translation } e
    \]
  - **reranking n-gram features, e.g.,**
    \[
    \Phi_1(f, e) = \text{Count of the bigram “the of” in } e
    \]
Discriminative Reranking

○ **Score after reranking**

\[ s(f, e) = \Phi(f, e) \cdot \tilde{\alpha} \]
\[ = \alpha_0 \Phi_0(f, e) + \sum_{j \in [1, J]} \alpha_j \Phi_j(f, e) \]

○ **Features**

○ *baseline feature*

\[ \Phi_0(f, e) = \text{Baseline score for translation } e \]

○ *reranking n-gram features, e.g.,*

\[ \Phi_1(f, e) = \text{Count of the bigram "the of" in } e \]
Discriminative Reranking

- Score after reranking
  \[ s(f, e) = \Phi(f, e) \cdot \tilde{\alpha} \]
  \[ = \alpha_0 \Phi_0(f, e) + \sum_{j \in [1,J]} \alpha_j \Phi_j(f, e) \]

- Features
  - baseline feature
    \[ \Phi_0(f, e) = \text{Baseline score for translation } e \]
  - reranking n-gram features, e.g.,
    \[ \Phi_1(f, e) = \text{Count of the bigram } \text{“the of” in } e \]
Leave-one-out Baseline Training
Leave-one-out Baseline Training

part-1

part-2

part-n
Leave-one-out Baseline Training

- part-1
- part-2
- part-n
Leave-one-out Baseline Training

part-2

part-1

part-n
Leave-one-out Baseline Training

part-1

part-2

part-n
Leave-one-out Baseline Training

part-2

part-n

Baseline

part-1
Leave-one-out Baseline Training

part-2 → part-1 → Baseline

part-n
Leave-one-out Baseline Training

part-2

part-n

Baseline

part-1
Leave-one-out Baseline Training

part-2

part-n

part-1

Baseline

N-bests
Leave-one-out Baseline Training

part-1

part-2

part-n
Leave-one-out Baseline Training

part-1

part-2

part-n
Leave-one-out Baseline Training

part-1

part-2

part-n
Leave-one-out Baseline Training

part-1

part-n

part-2
Leave-one-out Baseline Training

part-1 → part-2 → Baseline → part-n
Leave-one-out Baseline Training

part-1

part-2

part-n

Baseline
Leave-one-out Baseline Training

part-1

part-n

Baseline

part-2

Tuesday, August 18, 2009
Leave-one-out Baseline Training

part-1

dashed arrow

goes to

part-n

goes to

Baseline

N-bests
Leave-one-out Baseline Training

- part-1
- part-2
- part-n
Leave-one-out Baseline Training

part-1

part-2

part-n
Leave-one-out Baseline Training

part-1

part-2

part-n
Leave-one-out Baseline Training

part-1

part-2

part-n
Leave-one-out Baseline Training

part-1

part-2

part-n

Baseline
Leave-one-out Baseline Training

part-1

part-2

part-n

Baseline
Leave-one-out Baseline Training

part-1

part-2

Baseline

part-n
Leave-one-out Baseline Training

part-1

part-2

Baseline

part-n

N-bests
Data Selection
Data Selection

- Data is very noisy in our MT application
Data Selection

- Data is very noisy in our MT application
  - Human annotation is noisy
Data Selection

- Data is very noisy in our MT application
  - Human annotation is noisy
  - Automatic sentence alignment is noisy
Data Selection

- Data is very noisy in our MT application
  - Human annotation is noisy
  - Automatic sentence alignment is noisy
- We aim to select high-quality training data for discriminative training
Data Selection

- Data is very noisy in our MT application
  - Human annotation is noisy
  - Automatic sentence alignment is noisy

- We aim to select high-quality training data for discriminative training
  - An training example will be selected only if it satisfies certain conditions
G(ref, oracle) > T_1
\[ G(\text{ref, oracle}) > T_1 \]
G(ref, oracle) > T_1 

matched translation
$G(\text{ref, oracle}) > T_1$

$G(\text{ref, oracle}) - G(\text{ref, 1best}) > T_2$
\[ G(\text{ref}, \text{oracle}) > T_1 \]

\[ G(\text{ref}, \text{oracle}) - G(\text{ref}, 1\text{best}) > T_2 \]
G(ref, oracle) > T_1 \quad \Rightarrow \quad \text{matched translation}

G(ref, oracle) − G(ref, 1best) > T_2
$G(\text{ref, oracle}) > T_1$\quad\Rightarrow\quad \text{matched translation}$

$G(\text{ref, oracle}) - G(\text{ref, 1best}) > T_2\quad\Rightarrow\quad \text{profitable}$
\[ G(\text{ref, oracle}) > T_1 \]

\[ G(\text{ref, oracle}) - G(\text{ref, 1best}) > T_2 \]

\[ G(\text{oracle, 1best}) > T_3 \]
\( G(\text{ref, oracle}) > T_1 \)  

\( G(\text{ref, oracle}) - G(\text{ref, 1best}) > T_2 \)

\( G(\text{oracle, 1best}) > T_3 \)
\[ G(\text{ref}, \text{oracle}) > T_1 \quad \rightarrow \quad \text{matched translation} \]

\[ G(\text{ref}, \text{oracle}) - G(\text{ref}, \text{1best}) > T_2 \quad \rightarrow \quad \text{profitable} \]

\[ G(\text{oracle}, \text{1best}) > T_3 \quad \rightarrow \quad \text{correctable} \]
# Experiments: facts

<table>
<thead>
<tr>
<th>Language pair</th>
<th>Chinese to English</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Translation system</strong></td>
<td><strong>Hiero</strong></td>
</tr>
<tr>
<td><strong>Language model data</strong></td>
<td><strong>160 M words</strong></td>
</tr>
<tr>
<td><strong>Translation model data</strong></td>
<td><strong>30M words</strong></td>
</tr>
<tr>
<td><strong>Number of partitions</strong></td>
<td><strong>30</strong></td>
</tr>
<tr>
<td><strong>DEV set for baseline MERT</strong></td>
<td><strong>MT03</strong></td>
</tr>
<tr>
<td><strong>DEV set for reranking</strong></td>
<td><strong>MT04</strong></td>
</tr>
<tr>
<td><strong>Test sets</strong></td>
<td><strong>MT05, MT06</strong></td>
</tr>
<tr>
<td><strong>N-best size</strong></td>
<td><strong>300 unique</strong></td>
</tr>
<tr>
<td><strong>Training algorithm</strong></td>
<td><strong>averaged perceptron</strong></td>
</tr>
</tbody>
</table>
Data Selection: varying $T_1$

Figure 5: BLEU Scores on MT’04 when varying the value of $T_1 \in [0.05, 0.25]$ with a step size 0.01.

$G(\text{ref, oracle}) > T_1$ matched translation
Data Selection: varying $T_2$

Figure 6: BLEU Scores on MT’04 when varying the value of $T_2 \in [0.01, 0.10]$ with a step size 0.01.

$$G(\text{ref, oracle}) - G(\text{ref, 1best}) > T_2$$

-profitable
Data Selection: varying $T_3$

Figure 7: BLEU Scores on MT’04 when varying the value of $T_3 \in [0.20, 0.75]$ with a step size 0.05.

$$G(\text{oracle, 1best}) > T_3$$

\[\text{correctable}\]
Experiments: reranking results

<table>
<thead>
<tr>
<th>Task</th>
<th>Baseline</th>
<th>Reranking</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Full</td>
<td>Selected</td>
<td></td>
</tr>
<tr>
<td>MT’04</td>
<td>0.357</td>
<td><strong>0.365</strong></td>
<td>0.365</td>
<td></td>
</tr>
<tr>
<td>MT’05</td>
<td>0.326</td>
<td>0.332</td>
<td><strong>0.333</strong></td>
<td></td>
</tr>
<tr>
<td>MT’06</td>
<td>0.283</td>
<td>0.292</td>
<td><strong>0.294</strong></td>
<td></td>
</tr>
</tbody>
</table>
## Experiments: reranking results

<table>
<thead>
<tr>
<th>Task</th>
<th>Baseline</th>
<th>Reranking</th>
<th>Reranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Full</td>
<td>Selected</td>
</tr>
<tr>
<td>MT’04</td>
<td>0.357</td>
<td><strong>0.365</strong></td>
<td>0.365</td>
</tr>
<tr>
<td>MT’05</td>
<td>0.326</td>
<td>0.332</td>
<td><strong>0.333</strong></td>
</tr>
<tr>
<td>MT’06</td>
<td>0.283</td>
<td>0.292</td>
<td><strong>0.294</strong></td>
</tr>
</tbody>
</table>

### Selected data

<table>
<thead>
<tr>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_3$</th>
<th>Selected data</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.01</td>
<td>0.25</td>
<td>610K out of 1M</td>
</tr>
</tbody>
</table>
## Experiments: reranking results

<table>
<thead>
<tr>
<th>Task</th>
<th>Baseline</th>
<th>Reranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Full</td>
</tr>
<tr>
<td>MT’04</td>
<td>0.357</td>
<td><strong>0.365</strong></td>
</tr>
<tr>
<td>MT’05</td>
<td>0.326</td>
<td>0.332</td>
</tr>
<tr>
<td>MT’06</td>
<td>0.283</td>
<td>0.292</td>
</tr>
</tbody>
</table>

### n-gram

<table>
<thead>
<tr>
<th></th>
<th>active</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-gram</td>
<td>34k</td>
</tr>
<tr>
<td>2-gram</td>
<td>1908k</td>
</tr>
</tbody>
</table>
Summary

- We have developed a discriminative n-gram LM to rerank MT outputs
Summary

- We have developed a discriminative n-gram LM to rerank MT outputs
- Discriminative LM reranking improves the translation quality over a state of the art system
Summary

- We have developed a discriminative n-gram LM to rerank MT outputs
- Discriminative LM reranking improves the translation quality over a state of the art system
- With data selection, we can train a better/comparable model using less data
Hypergraph-based Discriminative Rescoring
Hypergraph-based Discriminative Rescoring

- generate a hypergraph (instead of an n-best) for each Chinese sentence
Hypergraph-based Discriminative Rescoring

- generate a hypergraph (instead of an n-best) for each Chinese sentence
- identify oracle translations on the hypergraph
Hypergraph-based Discriminative Rescoring

- generate a hypergraph (instead of an n-best) for each Chinese sentence
- identify oracle translations on the hypergraph
- train a model and use it in decoding on a hypergraph
Hypergraph-based Discriminative Rescoring

- generate a hypergraph (instead of an n-best) for each Chinese sentence
- identify oracle translations on the hypergraph
- train a model and use it in decoding on a hypergraph
- the hypergraph is pruned using the posterior pruning
## Hypergraph rescoring results

<table>
<thead>
<tr>
<th>System</th>
<th>MT04</th>
<th>MT05</th>
<th>MT06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chiang’07</td>
<td>34.6</td>
<td>31.8</td>
<td>NA</td>
</tr>
<tr>
<td>Ours</td>
<td>35.7</td>
<td>32.6</td>
<td>28.3</td>
</tr>
<tr>
<td>N-best</td>
<td>36.5</td>
<td>33.3</td>
<td>29.4</td>
</tr>
<tr>
<td>Hypergraph</td>
<td>35.9</td>
<td>33.0</td>
<td>28.2</td>
</tr>
</tbody>
</table>
Joshua: an open-source parsing-based MT decoder

- Team members
  - **JHU**: Zhifei Li, Chris Callison-Burch, Sanjeev Khudanpur, Wren Thornton, Jonathan Weese, and Omar Zaidan
  - **UMD**: Chris Dyer
  - **U of Minnesota**: Lane Schwartz

- Functions
  - Chart-parsing, pruning, language model integration, kbest extraction, distributed and parallel decoding
  - Suffix-array based grammar extraction
  - Minimum error rate training
Thank you!

谢谢！