Paraphrase Identification as Probabilistic Quasi-Synchronous Recognition

Dipanjan Das and Noah A. Smith
Language Technologies Institute
School of Computer Science
Carnegie Mellon University
{dipanjan, nasmith}@cs.cmu.edu
Outline

- Introduction
- Quasi-Synchronous Grammars
- Probabilistic Model
- Data
- Experiments and Results
- Conclusion and Future Work
Outline

• Introduction
• Quasi-Synchronous Grammars
• Probabilistic Model
• Data
• Experiments and Results
• Conclusion and Future Work
Task

• Paraphrase identification
  • do two given sentences convey the same meaning?
  • bidirectional entailment
  • a binary classification task
Examples

Example of a sentence level paraphrase:

The decision was among the most significant steps toward deregulation undertaken during the Bush administration.

The decision is among the far-reaching deregulatory actions made during the Bush administration.

(from the Microsoft Research Paraphrase Corpus)
Examples

Example of a sentence level paraphrase:

The decision was among the most significant steps toward deregulation undertaken during the Bush administration.

The decision is among the far-reaching deregulatory actions made during the Bush administration.

significant lexical overlap
Examples

Another example:

Revenue in the first quarter of the year dropped 15 percent from the same period earlier.

With the scandal hanging over Stewart’s company, revenue in the first quarter of the year dropped 15 percent from the same period earlier.

significant lexical overlap
Previous Work

- Various approaches:
  - predicate-argument annotations (Qiu et al., 2006)
  - inversion transduction grammars (Wu, 2005)
Previous Work

• Various approaches:
  • predicate-argument annotations (Qiu et al., 2006)
  • inversion transduction grammars (Wu, 2005)
  • statistical classifiers leveraging shallow lexical, syntactic or semantic overlap (Finch et al., 2005; Wan et al., 2006; Corley and Mihalcea, 2005)
Various approaches:

- predicate-argument annotations (Qiu et al., 2006)
- inversion transduction grammars (Wu, 2005)
- statistical classifiers leveraging shallow lexical, syntactic or semantic overlap (Finch et al., 2005; Wan et al., 2006; Corley and Mihalcea, 2005)

a lot can be achieved using ‘overlap’ based classifiers
Baselines

Results on the Microsoft Research Paraphrase Corpus
## Baselines

Results on the Microsoft Research Paraphrase Corpus

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

all p  Wan et al. SVM
Baselines

Results on the Microsoft Research Paraphrase Corpus

Accuracy

inter-annotator agreement = 83%

all p  Wan et al. SVM
Baselines

Results on the Microsoft Research Paraphrase Corpus

Accuracy

inter-annotator agreement = 83%
Baselines

Results on the Microsoft Research Paraphrase Corpus

inter-annotator agreement = 83%

Accuracy

ACL-IJCNLP 2009 • D. Das and N.A. Smith • Paraphrase Identification
Baselines

Results on the Microsoft Research Paraphrase Corpus

Accuracy

inter-annotator agreement = 83%

Features:
- unigram overlap
- BLEU score
- difference in sentence length
- syntactic dependency overlap

ACL-IJCNLP 2009 • D. Das and N.A. Smith • Paraphrase Identification
Baselines

Furthermore:

inter-annotator agreement = 83%

Features:
Just unigram, bigram and trigram overlap!

Accuracy

ACL-IJCNLP 2009 • D. Das and N.A. Smith • Paraphrase Identification
Baselines

Furthermore:

- inter-annotator agreement = 83%

Features: Just unigram, bigram and trigram overlap!

Accuracy

- all p: 66.49
- Wan et al. SVM: 75.42
- Logistic Regression: 75.36

indistinguishable
Examples

Another example:

About 120 potential jurors were being asked to complete a lengthy questionnaire.

The jurors were taken into the court room in groups of 40 and asked to fill out a questionnaire.

less lexical overlap: hard for a naive classifier to get
Examples

Another example:

About 120 potential jurors were being asked to complete a lengthy questionnaire.

The jurors were taken into the court room in groups of 40 and asked to fill out a questionnaire.

aligning similar lexical units may help
A false paraphrase:

"There were a number of bureaucratic and administrative missed signals -- there's not one person who's responsible here," Gehman said.

In turning down the NIMA offer, Gehman said, "there were a number of bureaucratic and administrative missed signals here.

(from the Microsoft Research Paraphrase Corpus)
Examples

A false paraphrase:

"There were a number of bureaucratic and administrative missed signals -- there's not one person who's responsible here," Gehman said.

In turning down the NIMA offer, Gehman said, "there were a number of bureaucratic and administrative missed signals here."

(from the Microsoft Research Paraphrase Corpus)
Examples

A false paraphrase:

"There were a number of bureaucratic and administrative missed signals -- there's not one person who's responsible here," Gehman said.

**makes the focus of the sentence different!**

In turning down the NIMA offer, Gehman said, "there were a number of bureaucratic and administrative missed signals here.

(from the Microsoft Research Paraphrase Corpus)

difficult for an ‘overlap’ based classifier to get
In a Nutshell

• What’s achieved in the current work:
  • *not* outperforming the state-of-the-art significantly
In a Nutshell

What’s achieved in the current work:

- not outperforming the state-of-the-art significantly
- a model of loose syntactic transformation coupled with lexical semantics to recognize paraphrases
  - monolingual translation
In a Nutshell

• What’s achieved in the current work:
  • *not* outperforming the state-of-the-art significantly
  • a model of loose syntactic transformation coupled with lexical semantics to recognize paraphrases
    • monolingual *translation*
  • discovery of *hidden* alignment between structures of the two sentences without supervision
In a Nutshell

• What’s achieved in the current work:
  • *not* outperforming the state-of-the-art significantly
  • a model of loose syntactic transformation coupled with lexical semantics to recognize paraphrases
    • monolingual *translation*
  • discovery of *hidden* alignment between structures of the two sentences without supervision
  • a probabilistic model that is easily extensible with more features
    • e.g. semantic structure
In a Nutshell

• Our approach:
  
  • a quasi-synchronous dependency grammar for paraphrases
  
  • models correspondences between words in the two sentences, in *loose syntactic* terms
  
  • employs *lexical semantics* in conjunction with syntax
In a Nutshell

• Smith and Eisner (2006) proposed quasi-synchronous grammars for machine translation

• Various applications:
  • QA: Wang, Smith, Mitamura (2007)
  • MT: Gimpel and Smith (2009)
  • Parser adaptation: Smith and Eisner (2009)

• We model **paraphrase identification** with competitive results
Outline

- Introduction
- Quasi-Synchronous Grammars
- Probabilistic Model
- Data
- Experiments and Results
- Conclusion and Future Work
Dependency Grammar

A simple dependency grammar

\[ t \rightarrow t \ t' \]

or

\[ t \rightarrow t' \ t \]
Dependency Grammar

A simple dependency grammar

\[ t \rightarrow t \ t' \]

or

words in a sentence

\[ t \rightarrow t' \ t \]
A simple dependency grammar

\[ t \rightarrow t \ t' \]

or

\[ t \rightarrow t' \ t \]

or

\[ t \rightarrow t' \ t' \]
Quasi-Synchronous Grammar

From dependency grammars to quasi-synchronous dependency grammars

Source sentence = \langle s_1, s_2, \ldots, s_m \rangle

\langle t, l \rangle \rightarrow \langle t, l \rangle \langle t', k \rangle

or

\langle t, l \rangle \rightarrow \langle t', k \rangle \langle t, l \rangle

indices in the source sentence
Quasi-Synchronous Grammar

From dependency grammars to quasi-synchronous dependency grammars

Source sentence = \( \langle s_1, s_2, \ldots, s_m \rangle \)

\[ \langle t, l \rangle \rightarrow \langle t, l \rangle \langle t', k \rangle \]

or

\[ \langle t, l \rangle \rightarrow \langle t', k \rangle \langle t, l \rangle \]
Quasi-Synchronous Grammar

From dependency grammars to quasi-synchronous dependency grammars

Source sentence = \( \langle S_1, S_2, \ldots, S_m \rangle \)

\[
\langle t, l \rangle \rightarrow \langle t, l \rangle \langle t', k \rangle
\]

or

\[
\langle t, l \rangle \rightarrow \langle t', k \rangle \langle t, l \rangle
\]
Quasi-Synchronous Grammar

From dependency grammars to quasi-synchronous dependency grammars

Source sentence = \( \langle S_1, S_2, \ldots, S_m \rangle \)

\( \langle t, l \rangle \rightarrow \langle t, l \rangle \langle t', k \rangle \)

or

\( \langle t, l \rangle \rightarrow \langle t', k \rangle \langle t, l \rangle \)
Quasi-Synchronous Grammar

From dependency grammars to quasi-synchronous dependency grammars

Source sentence = \( \langle s_1, s_2, \ldots, s_m \rangle \)

\( \langle t, l \rangle \rightarrow \langle t, l \rangle \langle t', k \rangle \)

\( \langle t', l \rangle \rightarrow \langle t', k \rangle \langle t, l \rangle \)

\( \text{or} \)

\( t \)

\( s_l \)

\( t' \)

\( s_j \)

\( s_k \)

\( \text{grandparent-child} \)

\( \text{so forth} \)
Quasi-Synchronous Grammar

An example

$ About 120 potential jurors were being asked to complete a lengthy questionnaire.

$ The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.
Quasi-Synchronous Grammar

An example

$ About 120 potential jurors were being asked to complete a lengthy questionnaire.

$ The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.
Quasi-Synchronous Grammar

An example

$\begin{align*}
\text{About 120 potential jurors were being asked to complete a lengthy questionnaire.} \\
\text{The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.}
\end{align*}$
About 120 potential jurors were being asked to complete a lengthy questionnaire.

The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.

An example assuming that $ in source aligns to $ in target.

The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.
About 120 potential jurors were being asked to complete a lengthy questionnaire.

moreover, $ collects all unaligned words

The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.
Quasi-Synchronous Grammar

An example

$ About 120 potential jurors were being asked to complete a lengthy questionnaire.

$ The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.

$\langle $, 0$\rangle \rightarrow $\langle $, 0$\rangle$ \langle were, 5$\rangle$

parent-child
Quasi-Synchronous Grammar

An example

$\text{About 120 potential jurors were being asked to complete a lengthy questionnaire.}$

$\langle\text{were, 5}\rangle \rightarrow \langle\text{jurors, 4}\rangle \langle\text{were, 5}\rangle$

parent-child

$\text{The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.}$
Quasi-Synchronous Grammar

An example

$ About 120 potential jurors were being asked to complete a lengthy questionnaire.

\[ \langle \text{were,5} \rangle \rightarrow \langle \text{were,5} \rangle \langle \text{taken,0} \rangle \]

child-null

$ The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.

ACL-IJCNLP 2009 • D. Das and N.A. Smith • Paraphrase Identification
Quasi-Synchronous Grammar

An example

$ About 120 potential jurors were being asked to complete a lengthy questionnaire.

$ The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.

\langle \text{were,5} \rangle \rightarrow \langle \text{were,5} \rangle \langle \text{and,0} \rangle$

child-null
Quasi-Synchronous Grammar

An example

$ About 120 potential jurors were being asked to complete a lengthy questionnaire.

$ The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.
About 120 potential jurors were being asked to complete a lengthy questionnaire.

The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.
Quasi-Synchronous Grammar

An example

$\text{About 120 potential jurors were being asked to complete a lengthy questionnaire.}$

$\text{The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.}$
About 120 potential jurors were being asked to complete a lengthy questionnaire.

The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.
About 120 potential jurors were being asked to complete a lengthy questionnaire.

The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.
Quasi-Synchronous Grammar

An example

$\text{About 120 potential jurors were being asked to complete a lengthy questionnaire.}$

$\text{The jurors were taken into the courtroom in groups of 40 and asked to fill out a questionnaire.}$
Quasi-Synchronous Grammar

An example

$\text{Dozens were injured across Afghanistan yesterday.}$

$\text{Sixty-one people were killed and dozens wounded.}$
Dozens were injured across Afghanistan yesterday.

Sixty-one people were killed and dozens wounded.

Quasi-Synchronous Grammar

More configurations

$\text{Dozens were injured across Afghanistan yesterday.}$

$\text{Sixty-one people were killed and dozens wounded.}$
Quasi-Synchronous Grammar

More configurations

$ EDS$ reported a first-quarter loss of 6 million dollars.

$ EDS$ declared a loss of six million dollars in the first quarter.
Quasi-Synchronous Grammar

More configurations

$ EDS reported a first-quarter loss of 6 million dollars .

$ EDS declared a loss of six million dollars in the first quarter .

\langle \text{quarter,4} \rangle \rightarrow \langle \text{first,4} \rangle \langle \text{quarter,4} \rangle$

same-node

$ EDS declared a loss of six million dollars in the first quarter .
Quasi-Synchronous Grammar

A generative model for dependency parsing

\[ P(\tau^i \mid G_0) \]

*probability of a tree rooted at word i given a base grammar*
Quasi-Synchronous Grammar

A generative model for dependency parsing

\[ P(\tau^i \mid G_0) = \prod_{j \in i's \ kids} P(\tau^j \mid G_0) \times p_{kid}(t_j \mid t_i, G_0) \]

a product term for each kid
Quasi-Synchronous Grammar

A generative model for dependency parsing

\[ P(\tau^i \mid G_0) = \prod_{j \in i's \ kids} P(\tau^j \mid G_0) \times p_{kid}(t_j \mid t_i, G_0) \]

probability of generating a kid
Quasi-Synchronous Grammar

A generative model for dependency parsing

\[ P(\tau^i \mid G_0) = \prod_{j \in i's \ kids} P(\tau^j \mid G_0) \times p_{kid}(t_j \mid t_i, G_0) \]

Note the recursion
Quasi-Synchronous Grammar

A generative model for dependency parsing

\[
P(\tau^i \mid G_0) = \prod_{j \in i's \ kids} P(\tau^j \mid G_0) \times p_{kid}(t_j \mid t_i, G_0)
\]

Parsing can be done using dynamic programming!
Quasi-Synchronous Grammar

A generative model for **quasi-synchronous** parsing

\[ P(\tau^i_t, x \mid \tau_s, G_{QG}(s)) \]

*Probability of a subtree rooted at word i and the alignment, given the source tree and the QG based on it*
Quasi-Synchronous Grammar

A generative model for \textit{quasi-synchronous} parsing

\[
P(\tau^i_t, \mathbf{x} \mid \tau_s, G_{QG}(s)) = \prod_{j \in i's \ kids} P(\tau^j_s, \mathbf{x} \mid \tau_s, G_{QG}(s)) \\ \times p_{kid}(t_j, \mathbf{x} \mid \tau_s, t_i, G_{QG}(s))
\]

\textit{a product term for each child}
Quasi-Synchronous Grammar

A generative model for quasi-synchronous parsing

\[ P(\tau_t^i, x \mid \tau_s, G_{QG}(s)) = \prod_{j \in i's \ kids} P(\tau_s^j, x \mid \tau_s, G_{QG}(s)) \times p_{kid}(t_j, x \mid \tau_s, t_i, G_{QG}(s)) \]

Probability of generating each kid and the alignments
Quasi-Synchronous Grammar

A generative model for quasi-synchronous parsing

\[
P(\tau^i_t, x \mid \tau_s, G_{QG}(s)) = \prod_{j \in i's \ kids} P(\tau^j_s, x \mid \tau_s, G_{QG}(s)) \times p_{kid}(t_j, x \mid \tau_s, t_i, G_{QG}(s))
\]

Probability of generating each kid and the alignments

depends on the WordNet relationships between \(t_i, t_j\), their POS and NE tags, and the syntactic configuration created at the alignment site
Quasi-Synchronous Grammar

A generative model for \textit{quasi-synchronous} parsing

\[
P(\tau_i^t, x | \tau_s, G_{QG}(s)) = \prod_{j \in i's \ kids} P(\tau_j^s, x | \tau_s, G_{QG}(s)) 
\times p_{kid}(t_j, x | \tau_s, t_i, G_{QG}(s))
\]

\text{defines the QG!}

Probability of generating each kid and the alignments
Quasi-Synchronous Grammar

A generative model for \textit{quasi-synchronous} parsing

\[ P(\tau_t^i, x | \tau_s, G_{QG}(s)) = \prod_{j \in i's \ kids} P(\tau_s^j, x | \tau_s, G_{QG}(s)) \times p_{kid}(t_j, x | \tau_s, t_i, G_{QG}(s)) \]

\textit{recursion}
Quasi-Synchronous Grammar

A generative model for quasi-synchronous parsing

\[ P(\tau^i_t, x \mid \tau_s, G_{QG}(s)) = \prod_{j \in i's \ kids} P(\tau^j_s, x \mid \tau_s, G_{QG}(s)) \times p_{kid}(t_j, x \mid \tau_s, t_i, G_{QG}(s)) \]

Decoding can be solved using dynamic programming, with some constraints on the alignments.
Outline

• Introduction
• Quasi-Synchronous Grammars
• Probabilistic Model
• Data
• Experiments and Results
• Conclusion and Future Work
Probabilistic Model

Given: sentences $s_1$ and $s_2$
Probabilistic Model

Given: sentences $s_1$ and $s_2$

\[ p(\text{paraphrase} \mid s_1, s_2) = \]
Probabilistic Model

Given: sentences $s_1$ and $s_2$

\[
p(\text{paraphrase} \mid s_1, s_2) = \frac{p(s_1, s_2 \mid \text{paraphrase}) \cdot p(\text{paraphrase})}{p(s_1, s_2 \mid \text{paraphrase}) \cdot p(\text{paraphrase}) + p(s_1, s_2 \mid \text{non-paraphrase}) \cdot p(\text{non-paraphrase})}
\]
Probabilistic Model

**Given:** sentences $s_1$ and $s_2$

$$p(s_1, s_2 | \text{paraphrase})$$
Probabilistic Model

Given: sentences $s_1$ and $s_2$

$$p(s_1, s_2 | \text{paraphrase})$$
Probabilistic Model

Given: sentences \( s_1 \) and \( s_2 \)

\[
p(s_1, s_2 | \text{paraphrase}) = p(s_1 | G_0) \cdot p(s_2 | G_p(s_1))
\]
Probabilistic Model

Given: sentences $s_1$ and $s_2$

$$p(s_1, s_2|\text{paraphrase}) = p(s_1 | G_0) \cdot p(s_2 | G_p(s_1))$$

probability of $s_1$
under base grammar $G_0$
Probabilistic Model

Given: sentences $s_1$ and $s_2$

$$p(s_1, s_2 \mid \text{paraphrase}) = p(s_1 \mid G_0) \cdot p(s_2 \mid G_p(s_1))$$

probability of $s_2$
under quasi-synchronous
grammar $G_p$ based on $s_1$
Probabilistic Model

Given: sentences $s_1$ and $s_2$

$$p(s_1, s_2 | \text{paraphrase}) = p(s_1 | G_0) \cdot p(s_2 | G_p(s_1))$$

models only one direction of paraphrasing
Probabilistic Model

Given: sentences $s_1$ and $s_2$

$$p(s_1, s_2 | \text{paraphrase}) = 0.5 \cdot p(s_1 | G_0) \cdot p(s_2 | G_p(s_1))$$
$$+ 0.5 \cdot p(s_2 | G_0) \cdot p(s_1 | G_p(s_2))$$
Probabilistic Model

Given: sentences $s_1$ and $s_2$

$$p(\text{paraphrase} \mid s_1, s_2) = \frac{p(s_1, s_2 \mid \text{paraphrase}) \cdot p(\text{paraphrase})}{p(s_1, s_2 \mid \text{paraphrase}) \cdot p(\text{paraphrase}) + p(s_1, s_2 \mid \text{non-paraphrase}) \cdot p(\text{non-paraphrase})}$$
Probabilistic Model

\[ p(s_1, s_2 | \text{non-paraphrase}) \]
Probabilistic Model

\[
p(s_1, s_2 | \text{non-paraphrase}) = 0.5 \cdot p(s_1 | G_0) \cdot p(s_2 | G_n (s_1)) \\
+ 0.5 \cdot p(s_2 | G_0) \cdot p(s_1 | G_n (s_2))
\]
Probabilistic Model

\[ p(t \mid G_P(s)) \] \quad \text{and} \quad \[ p(t \mid G_N(s)) \]

parameterized in the same way
Probabilistic Model

\[ p(t \mid G_p(s)) \]
Probabilistic Model

\[ p(t \mid G_p(s)) \]

\[ = p(\tau_t \mid G_p(\tau_s)) \quad \text{trees assumed to be observed} \]
Probabilistic Model

\[ p(t \mid G_p(s)) \]

\[ = p(\tau_t \mid G_p(\tau_s)) \]

\[ = \sum_{x} p(\tau_t, x \mid G_p(\tau_s)) \]

Trees assumed to be observed
Probabilistic Model

\[ p(t \mid G_p(s)) \]

\[ = p(\tau_t \mid G_p(\tau_s)) \quad \text{trees assumed to be observed} \]

\[ = \sum_{x} p(\tau_t, x \mid G_p(\tau_s)) \]

Alignments not visible!
\[ p(t \mid G_p(s)) = p(\tau_t \mid G_p(\tau_s)) \]
\[ = \sum_{x} p(\tau_t, x \mid G_p(\tau_s)) \]

*Marginalizing all possible alignments*
Probabilistic Model

\[ p(t \mid G_p(s)) \]

\[ = p(\tau_t \mid G_p(\tau_s)) \]  
   trees assumed to be observed

\[ = \sum_{x} p(\tau_t, x \mid G_p(\tau_s)) \]

*Can be done efficiently using dynamic programming in \(O(m^2n)\) time and \(O(mn)\) space*
Inference

\[ c^* = \arg \max_{c \in \{\text{●, ●}\}} p(c \mid s_1, s_2) \]
Training

\[
\max_{\Theta} \sum_{i=1}^{N} \log p(c^{(i)} | s_1^{(i)}, s_2^{(i)}, \Theta) - C \| \Theta \|_2^2
\]

We directly optimize this using L-BFGS
Outline

• Introduction
• Quasi-Synchronous Grammars
• Probabilistic Model
• Data
• Experiments and Results
• Conclusion and Future Work
Data

• Microsoft Research Paraphrase Corpus
  (Dolan et al., 2004; Quirk et al., 2004)
  • 5801 sentence pairs with human annotations
  • 3900 “true paraphrases”
  • training set: 4076 sentence pairs (2753/1323)
  • test set: 1725 sentence pairs (1147/578)
  • 83% inter-annotator agreement
  • κ=0.62
Outline

- Introduction
- Quasi-Synchronous Grammars
- Probabilistic Model
- Data
- Experiments and Results
- Conclusion and Future Work
Experiments

- Baselines:
  - all paraphrase
  - a system by Wan et al. (2006)
    - SVM trained on “overlap” features
    - replication with our preprocessing tools
  - a logistic regression (LR) model with n-gram overlap features
Results
## Results

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>all p</th>
<th>Wan et al. SVM</th>
<th>LR</th>
<th>QG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

inter-annotator agreement = 83%

Accuracy | Precision | Recall
--- | --- | ---
all p | 66.49 | 66.49 | 100.00
Wan et al. SVM | 75.42 | 76.88 | 90.14
LR | 75.36 | 78.12 | 87.44
QG |  |  |  

ACL-IJCNLP 2009 • D. Das and N.A. Smith • Paraphrase Identification
Results

inter-annotator agreement = 83%
Alignments

- fill → complete
- questionnaire → questionnaire
- dozens → wounded
- dozens → injured
- quarter → first-quarter
- first

- parent-child
- child-parent
- same-node
Alignments

grandparent-grandchild
Alignments

- signatures
- necessary

- collected
- approaching
twice

- signatures

- 897,158
the

- needed

- c-command
Experiments

- The QG cannot capture long lexical overlaps
- To include such overlap features in a grammar is intractable
- We combine the LR model with the QG
  - Product of Experts (Hinton, 99)
Results

inter-annotator agreement = 83%
Results

inter-annotator agreement = 83%
Outline

- Introduction
- Quasi-Synchronous Grammars
- Probabilistic Model
- Data
- Experiments and Results
- Conclusion and Future Work
Conclusion

- A novel approach for **paraphrase identification**
  - use of an elegant probabilistic grammar model
- Models and discovers alignments using latent variables
- Probabilistic combination of the QG model with a lexical overlap model giving state-of-the-art results
Future Work

- Incorporate more semantic structure in the QG
- Extend the model to more sentence-pair tasks
- Investigate more features at the QG’s alignment sites
Questions?