Learning to Match for Natural Language Processing and Information Retrieval

Hang Li
Huawei Technologies

* Work was done at Microsoft Research, with former colleagues and interns
Language Understanding is Difficult for Computer, If Not Impossible

- 真热！
- 枯藤老树昏鸦 小桥流水人家
- 韩寒 方舟子
Language Processing without Language Understanding

- Transformation
- Transform one string to another string
- Applications
  - Machine Translation

- Matching
- Match between two strings
- Applications
  - Search
  - Question Answering
Talk Outline

• Introduction
• Regularized Latent Semantic Indexing
• Matching in Latent Space
• String Rewriting Kernel
• Conclusion
Same Search Intent, Different Query Representations

Example = “Distance between Sun and Earth”

• "how far" earth sun
• "how far" sun
• "how far" sun earth
• average distance earth sun
• average distance from earth to sun
• average distance from the earth to the sun
• distance between earth & sun
• distance between earth and sun
• distance between earth and the sun
• distance from earth to the sun
• distance from sun to earth
• distance from sun to the earth
• distance from the earth to the sun
• distance from the sun to earth
• distance from the sun to the earth
• distance of earth from sun
• distance between earth sun
• how far away is the sun from earth
• how far away is the sun from the earth
• how far earth from sun
• how far earth from the sun
• how far from earth is the sun
• how far from earth to sun
• how far from the earth to the sun
• distance between sun and earth
Same Search Intent, Different Query Representations
Example = “Youtube”

- yutube
- ytube
- youtubo
- youtube om
- youtube
- youtub com
- youtub
- you tube
- you tube videos
- www youtube
- yotube
- ww youtube com
- utube videos
- u tube com
- u tube
- outube
Semantic Matching Project: Solving Document Mismatch in Web Search
Matching at Different Levels

- **Term**
  - NY → NY
  - youtube → youtube
- **Phrase**
  - hot dog → hot dog
- **Word Sense**
  - utube → youtube
  - NY → New York
- **Topic**
  - Microsoft Office → ... Microsoft ... PowerPoint, Word, Excel...
- **Structure**
  - how far is sun from earth → ... distance between sun and earth

The pyramid illustrates the matching process at different levels of semantics, from terms to phrases, word senses, topics, and structures.
Query Understanding

- Structure Identification
- Topic Identification
- Similar Query Finding
- Phrase Identification
- Spelling Error Correction

michael jordan: **main phrase**
michael jordan berkely: **machine learning**
michael l. jordan
michael jordan

[michael jordan] berkeley
michael jordan berkeley

michael jordan berkele
Homepage of Michael Jordan

Michael Jordan is Professor in the Department of Electrical Engineering
Online Matching

Matching is conducted at different levels
Related Work

• Studied in long history of IR
• Query expansion, pseudo relevance feedback
• ... ... 

• New problem setting
  – Large amount of data available
  – New machine learning techniques
# Matching vs Ranking

In search, first matching and then ranking

<table>
<thead>
<tr>
<th></th>
<th>Matching</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prediction</strong></td>
<td>Matching degree between query and document</td>
<td>Ranking list of documents</td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>$f(q, d)$</td>
<td>$f(q,d1), f(q,d2), \ldots, f(q,d_n)$</td>
</tr>
<tr>
<td><strong>Challenge</strong></td>
<td>Mismatch</td>
<td>Correct ranking on top</td>
</tr>
</tbody>
</table>
Matching between Heterogeneous Data is Everywhere

• Matching between user and product (collaborative filtering)
• Matching between text and image (image annotation)
• Matching between people (dating)
• Matching between languages (machine translation)
• Matching between receptor and ligand (drug design)
Regularized Latent Semantic Indexing

Joint Work with Quan Wang, Jun Xu, and Nick Craswell

SIGIR 2011
Regularized Latent Semantic Indexing

• Motivation
  – Matching between query and document at topic level
  – Scale up to large datasets (vs. existing methods)

• Approach
  – Matrix Factorization
  – Regularization on topics and documents (vs. Sparse Coding)
  – Learning problem can be easily decomposed

• Results
  – $l_1$ on topics leads to sparse topics and $l_2$ on documents leads to accurate matching
  – Comparable with existing methods in topic discovery and search relevance
  – But can easily scale up to large document sets
Query and Document Matching in Topic Space

Document Space

Topic Space

q
d1
d2
dn

q
d1
d2
dn
Regularized Latent Semantic Indexing

\[ \min_{U, \{v_n\}} \sum_{n=1}^{N} \|d_n - Uv_n\|_2^2 + \lambda_1 \sum_{k=1}^{K} \|u_k\|_1 + \lambda_2 \sum_{n=1}^{N} \|v_n\|_2^2 \]

documents are smooth

topics are sparse
Optimization Strategy

\[
\min \sum_{n=1}^{N} \left( \|d_n - Uv_n\|_2^2 + \lambda_1 \|u_k\|_1 + \lambda_2 \sum_{n=1}^{N} \|v_n\|_2^2 \right)
\]

Coordinate Decent

Update \(U\)

Update \(V\)

\[
\min \sum_{m=1}^{M} \left( \|d_m - V^T \tilde{u}_m\|_2^2 + \lambda_1 \|\tilde{u}_m\|_1 \right)
\]

for \(m = 1, \ldots, M\)

\[
\min \|d_m - V^T \tilde{u}_m\|_2^2 + \lambda_1 \|\tilde{u}_m\|_1
\]

\[
\min \sum_{n=1}^{N} \left( \|d_n - Uv_n\|_2^2 + \lambda_2 \|v_n\|_2^2 \right)
\]

for \(n = 1, \ldots, N\)

\[
\min \|d_n - Uv_n\|_2^2 + \lambda_2 \|v_n\|_2^2
\]

\[
u_n^* = \left( U^T U + \lambda_2 I \right)^{-1} U^T d_n
\]

Analytic Solution

\[
u_n = \begin{cases} 
\frac{(r_{mk} - \sum_{l\neq k} s_{kl} u_{ml}) - \frac{1}{2} \lambda_1}{s_{kk}}, & \text{if } u_{mk} > 0 \\
\frac{s_{kk}}{(r_{mk} - \sum_{l\neq k} s_{kl} u_{ml}) + \frac{1}{2} \lambda_1}, & \text{if } u_{mk} < 0
\end{cases}
\]
RLSI Algorithm

- Single machine multi core version
- Multiple machine version (MapReduce and MPI)

**Algorithm 2 Update U**

Require: $D \in \mathbb{R}^{M \times N}$, $V \in \mathbb{R}^{K \times N}$

1: $S \leftarrow VV^T$
2: $R \leftarrow DV^T$
3: for $m = 1 : M$
4: \hspace{1em} $\tilde{u}_m \leftarrow 0$
5: \hspace{1em} repeat
6: \hspace{2em} for $k = 1 : K$
7: \hspace{3em} $w_{mk} \leftarrow r_{mk} - \sum_{l \neq k} s_{kl} u_{ml}$
8: \hspace{3em} $u_{mk} \leftarrow \left( |w_{mk}| - \frac{1}{2} AN \right) \cdot \text{sign}(w_{mk}) / s_{kk}$
9: \hspace{2em} end for
10: \hspace{1em} until convergence
11: end for
12: return $U^{(T)}$, $V^{(T)}$

**Algorithm 3 Update V**

Require: $D \in \mathbb{R}^{M \times N}$, $U \in \mathbb{R}^{M \times K}$

1: $\Sigma \leftarrow \left(U^T U + \theta I\right)^{-1}$
2: $\Phi \leftarrow U^T D$
3: for $n = 1 : N$
4: \hspace{1em} $v_n \leftarrow \Sigma \phi_n$, where $\phi_n$ is the $n^{th}$ column
5: \hspace{1em} end for
6: return $V$
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Max Dataset Applied (#docs; #words)</th>
<th># Topics</th>
<th># Processors Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLDA and PLDA+ (by Google)</td>
<td>Wiki-200T(2,112,618; 200,000)</td>
<td>1000</td>
<td>2,048</td>
</tr>
<tr>
<td>AD-LDA</td>
<td>NY Times (300,000; 102,660)</td>
<td>200</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>PubMed (8,200,000; 141,043)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wiki-All (3,239,884; 6,043,069)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B01 (1,562,807; 7,014,881)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bing News (940,702; 500,033)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wiki-All (3,239,884; 6,043,069)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSWeb Data (2,635,158; 2,371,146)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLSI</td>
<td>B01 (1,562,807; 7,014,881)</td>
<td>500 ~ 1000</td>
<td>single machine, 24 cores</td>
</tr>
<tr>
<td></td>
<td>Bing News (940,702; 500,033)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wiki-All (3,239,884; 6,043,069)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSWeb Data (2,635,158; 2,371,146)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experimental Results on Topic Discovery

Topics discovered by RLSI are equally readable compared with LDA, PLSI, LSI

<p>| Table 8: Topics discovered by RLSI, LDA, PLSI, and LSI from AP dataset. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|</p>
<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
<th>Topic 5</th>
<th>Topic 6</th>
<th>Topic 7</th>
<th>Topic 8</th>
<th>Topic 9</th>
<th>Topic 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLSI</td>
<td>opec</td>
<td>africa</td>
<td>aid</td>
<td>school</td>
<td>noriega</td>
<td>percent</td>
<td>plane</td>
<td>israeli</td>
<td>nuclear</td>
</tr>
<tr>
<td>AvgComp = 0.0075</td>
<td>oil</td>
<td>south</td>
<td>virus</td>
<td>student</td>
<td>panama</td>
<td>billion</td>
<td>crash</td>
<td>palestinian</td>
<td>soviet</td>
</tr>
<tr>
<td></td>
<td>cent</td>
<td>african</td>
<td>infect</td>
<td>teacher</td>
<td>panamanian</td>
<td>rate</td>
<td>flight</td>
<td>israel</td>
<td>treaty</td>
</tr>
<tr>
<td></td>
<td>barrel</td>
<td>angola</td>
<td>test</td>
<td>educate</td>
<td>delval</td>
<td>0</td>
<td>air</td>
<td>arab</td>
<td>missile</td>
</tr>
<tr>
<td></td>
<td>price</td>
<td>apartheid</td>
<td>patient</td>
<td>college</td>
<td>canal</td>
<td>trade</td>
<td>airline</td>
<td>plo</td>
<td>weapon</td>
</tr>
<tr>
<td>LDA</td>
<td>soviet</td>
<td>school</td>
<td>dukakis</td>
<td>party</td>
<td>water</td>
<td>price</td>
<td>court</td>
<td>air</td>
<td>iran</td>
</tr>
<tr>
<td>AvgComp = 1</td>
<td>nuclear</td>
<td>democrat</td>
<td>year</td>
<td>year</td>
<td>year</td>
<td>year</td>
<td>year</td>
<td>charge</td>
<td>plane</td>
</tr>
<tr>
<td></td>
<td>union</td>
<td>campaign</td>
<td>minister</td>
<td>television</td>
<td>animal</td>
<td>trade</td>
<td>judge</td>
<td>case</td>
<td>flight</td>
</tr>
<tr>
<td></td>
<td>state</td>
<td>educate</td>
<td>bush</td>
<td>elect</td>
<td>film</td>
<td>0</td>
<td>percent</td>
<td>attorney</td>
<td>crash</td>
</tr>
<tr>
<td></td>
<td>treaty</td>
<td>university</td>
<td>jackson</td>
<td>nation</td>
<td>nation</td>
<td>nation</td>
<td>nation</td>
<td>airliner</td>
<td>navy</td>
</tr>
<tr>
<td>PLSI</td>
<td>company</td>
<td>israeli</td>
<td>year</td>
<td>bush</td>
<td>court</td>
<td>soviet</td>
<td>year</td>
<td>plane</td>
<td>state</td>
</tr>
<tr>
<td>AvgComp = 0.9534</td>
<td>million</td>
<td>iran</td>
<td>state</td>
<td>charge</td>
<td>treaty</td>
<td>state</td>
<td>flight</td>
<td>air</td>
<td>new</td>
</tr>
<tr>
<td></td>
<td>share</td>
<td>israel</td>
<td>state</td>
<td>democrat</td>
<td>missile</td>
<td>new</td>
<td>air</td>
<td>people</td>
<td>nation</td>
</tr>
<tr>
<td></td>
<td>billion</td>
<td>palestinian</td>
<td>new</td>
<td>campaign</td>
<td>judge</td>
<td>nuclear</td>
<td>atom</td>
<td>crash</td>
<td>people</td>
</tr>
<tr>
<td></td>
<td>stock</td>
<td>arab</td>
<td>nation</td>
<td>republican</td>
<td>trial</td>
<td>gorbachev</td>
<td>govern</td>
<td>air</td>
<td>nation</td>
</tr>
<tr>
<td>LSI</td>
<td>soviet</td>
<td>567</td>
<td>earthquake</td>
<td>drug</td>
<td>0</td>
<td>israel</td>
<td>yen</td>
<td>urgent</td>
<td>student</td>
</tr>
<tr>
<td>AvgComp = 1</td>
<td>percent</td>
<td>234</td>
<td>quake</td>
<td>school</td>
<td>dukakis</td>
<td>oil</td>
<td>oil</td>
<td>school</td>
<td>noriega</td>
</tr>
<tr>
<td></td>
<td>police</td>
<td>0</td>
<td>richter</td>
<td>test</td>
<td>israel</td>
<td>dukakis</td>
<td>bush</td>
<td>student</td>
<td>noriega</td>
</tr>
<tr>
<td></td>
<td>govern</td>
<td>percent</td>
<td>dollar</td>
<td>court</td>
<td>jackson</td>
<td>dollar</td>
<td>dollar</td>
<td>dukakis</td>
<td>panama</td>
</tr>
<tr>
<td></td>
<td>state</td>
<td>12</td>
<td>scale</td>
<td>dukakis</td>
<td>dem</td>
<td>jackson</td>
<td>dukakis</td>
<td>cent</td>
<td>teacher</td>
</tr>
</tbody>
</table>
Experimental Results on Web Search

- Reducing vocabulary hurts ranking accuracy.
- RLSI can help improve search relevance.

![Graph showing MAP and NDCG values for different metrics and models.](chart.png)
Matching in Latent Space

Joint Work with Wei Wu, Zhengdong Lv
Under review
Matching in Latent Space

• Motivation
  – Matching between query and document in latent space

• Assumption
  – Queries have similarity
  – Document have similarity
  – Click-through data represent “similarity” relations between queries and documents

• Approach
  – Projection to latent space
  – Regularization or constraints

• Results
  – Significantly enhance accuracy of query document matching
Matching in Latent Space

Query Space

Document Space

Latent Space

$q_1 \rightarrow q_m \rightarrow d_1 \rightarrow d_2 \rightarrow q_2$

$L_q$

$L_d$
Example: Projecting Keywords and Images into Latent Space
Partial Least Square (PLS)

- **Setting**
  - Two spaces: \( \mathcal{X} \subset \mathbb{R}^m \) and \( \mathcal{Y} \subset \mathbb{R}^n \).

- **Input**
  - Training data: \( \{(x_i, y_i, r_i)\}_{1 \leq i \leq N}, \ r_i \in \{+1, -1\} \) (or \( r_i \in \mathbb{R} \))

- **Output**
  - Similarity function \( f(x, y) \)

- **Assumption**
  - Two linear (and orthonormal) transformations \( L_x \) and \( L_y \)
  - Dot product as similarity function \( \langle L_x^T x, L_y^T y \rangle = x^T L_x L_y^T y \)

- **Optimization**
  
  \[
  \arg \max_{L_x, L_y} \sum_{r_i=+1} x_i^T L_x \ L_y^T \ y_i - \sum_{r_i=-1} x_i^T L_x \ L_y^T \ y_i
  \]

  subject to \( L_x^T L_x = I_{k \times k}, L_y^T L_y = I_{k \times k} \)
Solution of Partial Least Square

- Non-convex optimization
- Can prove that global optimal solution exists
- Global optimal can be found by solving SVD (Singular Value Decomposition)
- SVD of Matrix \( M_S - M_D = U \Sigma V^T \)
Regularized Mapping to Latent Space (RMLLS)

• Setting
  – Two spaces: \( \mathcal{X} \subset \mathbb{R}^m \) and \( \mathcal{Y} \subset \mathbb{R}^n \).

• Input
  – Training data: \( \{(x_i, y_i, r_i)\}_{1 \leq i \leq N}, \ r_i \in \{+1, -1\} \) (or \( r_i \in \mathbb{R} \))

• Output
  – Similarity function \( f(x, y) \)

• Assumption
  – L1 and L2 regularization on \( L_X \) and \( L_Y \) (sparse transformations)
  – Dot product as similarity function \( \langle L_X^T x, \ L_Y^T y \rangle = x^T L_X \ L_Y^T y \)

• Optimization

\[
\text{argmax}_{L_X, L_Y} \ \sum_{r_i=+1} x_i^T L_X \ L_Y^T y_i - \sum_{r_i=-1} x_i^T L_X \ L_Y^T y_i
\]

subject to \( |lx| \leq \vartheta x, \ |ly| \leq \vartheta y, \ \|lx\| \leq \lambda x, \ \|ly\| \leq \lambda y, \)
Solution of Regularized Mapping to Latent Space

• Coordinate Descent
• Repeat
  – Fix $L_x$, update $L_y$
  – Fix $L_y$, update $L_x$
• Update can be parallelized by rows
## Comparison

<table>
<thead>
<tr>
<th></th>
<th>PLS</th>
<th>RMLS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assumption</strong></td>
<td>Orthogonal</td>
<td>L1 and L2 Regularization</td>
</tr>
<tr>
<td><strong>Optimization Method</strong></td>
<td>Singular Value Decomposition</td>
<td>Coordinate Descent</td>
</tr>
<tr>
<td><strong>Optimality</strong></td>
<td>Global optimum</td>
<td>Local optimum</td>
</tr>
<tr>
<td><strong>Efficiency</strong></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>
Experimental Results

<table>
<thead>
<tr>
<th>Enterprise Search</th>
<th>Web Search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NDCG@1</td>
</tr>
<tr>
<td>MPLS&lt;sub&gt;Com&lt;/sub&gt;</td>
<td>0.715</td>
</tr>
<tr>
<td>MPLS&lt;sub&gt;Conca&lt;/sub&gt;</td>
<td>0.700</td>
</tr>
<tr>
<td>MPLS&lt;sub&gt;Word&lt;/sub&gt;</td>
<td>0.688</td>
</tr>
<tr>
<td>MPLS&lt;sub&gt;Bipar&lt;/sub&gt;</td>
<td>0.659</td>
</tr>
<tr>
<td>BM25</td>
<td>0.653</td>
</tr>
<tr>
<td>RW</td>
<td>0.654</td>
</tr>
<tr>
<td>RW+BM25</td>
<td>0.664</td>
</tr>
<tr>
<td>LSI</td>
<td>0.656</td>
</tr>
<tr>
<td>LSI+BM25</td>
<td>0.692</td>
</tr>
<tr>
<td></td>
<td>0.681</td>
</tr>
<tr>
<td>MPLS&lt;sub&gt;Com&lt;/sub&gt;</td>
<td>0.676</td>
</tr>
<tr>
<td>MPLS&lt;sub&gt;Conca&lt;/sub&gt;</td>
<td>0.674</td>
</tr>
<tr>
<td>MPLS&lt;sub&gt;Word&lt;/sub&gt;</td>
<td>0.612</td>
</tr>
<tr>
<td>MPLS&lt;sub&gt;Bipar&lt;/sub&gt;</td>
<td>0.637</td>
</tr>
<tr>
<td>BM25</td>
<td>0.655</td>
</tr>
<tr>
<td>RW</td>
<td>0.671</td>
</tr>
<tr>
<td>RW+BM25</td>
<td>0.588</td>
</tr>
<tr>
<td>LSI</td>
<td>0.649</td>
</tr>
</tbody>
</table>

- RMLS and PLS work better than BM25, Random Walk, Latent Semantic Indexing
- RMLS works equally well as PLS, with higher learning efficiency and scalability
String Re-writing Kernel

Joint work with Fan Bu and Xiaoyan Zhu
ACL 2012
Learning with String Re-wring Kernel

- ((distance between sun and earth), (how far sun earth), +1)
- ((distance between beijing and shanghai), (how far is beijing from shanghai), +1)
- … ….
- ((distance between moon and earth), (how far sun earth), -1)

```
Learning System
```

```
Prediction System
```

```
Model
```

+1/-1?

```
((distance between moon and earth), (how far is earth from moon))
```
Problem Formulation

• Training data

\[((s_1, t_1), y_1) \cdots ((s_n, t_n), y_n)\]

• Model

\[y = \text{sign} \left( \sum_{i=1}^{n} \alpha_i y_i K((s_i, t_i), (s, t)) \right)\]

• String Re-writing Kernel

\[K((s_i, t_i), (s, t))\]
Re-Writing Rule

• Measure similarity between two pairs of strings using re-writing rule

Shakespeare wrote Hamlet

Hamlet was written by Shakespeare

Cao Xueqin wrote Dream of the Red Chamber

Dream of the Red Chamber was written by Cao Xueqin

Re-writing Rule

Cao Xueqin wrote Dream of the Red Chamber

Hamlet was written by Shakespeare
Formulation of String Re-writing Kernel (SRK)

\[ K((s_1,t_1),(s_2,t_2)) = \langle \Phi(s_1,t_1), \Phi(s_2,t_2) \rangle \]

\[ \Phi(s,t) = (\phi_r(s,t))_{r \in R} \]

\[ \phi_r(s,t) = n\lambda^i \quad \lambda \in (0,1] \]
String Re-writing Kernel

• Advantage: Matching between informally written sentences such as long queries in search can be effectively performed

• Challenge
  • Number of re-writing rules is infinite
  • Number of matched rules increase exponentially when length of sentence increases

• Our Approach
  • Sub-class: kb-SRK
Definition of kb-SRK

• Special class of SRK
• Re-writing rules in kb-SRK
  – String patterns in rule are of length $k$
  – Wildcard $?\text{ only substitutes a single character}$
  – Alignment between string patterns is bijective

? wrote ?
? described ?
Formulation of kb-SRK

\[ K_k ((s_1, t_1), (s_2, t_2)) = \sum \sum \bar{K}_k ((\alpha_{s_1}, \alpha_{t_1}), (\alpha_{s_2}, \alpha_{t_2})) \]

\[ \bar{K}_k = \sum_{r \in R} \phi_r (\alpha_{s_1}, \alpha_{t_1}) \phi_r (\alpha_{s_2}, \alpha_{t_2}) \]
Experiment: Paraphrase Identification

- Comparison with state-of-the-arts methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang and Patrick (2005)</td>
<td>71.9</td>
</tr>
<tr>
<td>Lintean and Rus (2011)</td>
<td>73.6</td>
</tr>
<tr>
<td>Heilman and Smith (2010)</td>
<td>73.2</td>
</tr>
<tr>
<td>Qiu et al. (2006)</td>
<td>72.0</td>
</tr>
<tr>
<td>Wan et al. (2006)</td>
<td>75.6</td>
</tr>
<tr>
<td>Das and Smith (2009)</td>
<td>73.9</td>
</tr>
<tr>
<td>Das and Smith (2009)(PoE)</td>
<td>76.1</td>
</tr>
<tr>
<td>Our baseline (PR)</td>
<td>73.6</td>
</tr>
<tr>
<td>Our method (ps-SRK)</td>
<td>75.6</td>
</tr>
<tr>
<td>Our method (pw-SRK)</td>
<td>75.0</td>
</tr>
<tr>
<td>Our method (kb-SRK)</td>
<td>76.3</td>
</tr>
</tbody>
</table>
Experiment: Recognizing Textual Entailment

- Comparison with state-of-the-arts methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harmeling (2007)</td>
<td>59.5</td>
</tr>
<tr>
<td>de Marneffe et al. (2006)</td>
<td>60.5</td>
</tr>
<tr>
<td>M&amp;M, (2007) (NL)</td>
<td>59.4</td>
</tr>
<tr>
<td>M&amp;M, (2007) (Hybrid)</td>
<td>64.3</td>
</tr>
<tr>
<td>Zanzotto et al. (2007)</td>
<td>65.75</td>
</tr>
<tr>
<td>Heilman and Smith (2010)</td>
<td>62.8</td>
</tr>
<tr>
<td>Our baseline (PR)</td>
<td>62.0</td>
</tr>
<tr>
<td>Our method (ps-SRK)</td>
<td>64.6</td>
</tr>
<tr>
<td>Our method (pw-SRK)</td>
<td>63.8</td>
</tr>
<tr>
<td>Our method (kb-SRK)</td>
<td>65.1</td>
</tr>
</tbody>
</table>

Lexical-based
Conclusion
Conclusion

• Transformation and matching are two fundamental problems in natural language processing
• Query document mismatch is greatest challenge in search
• Learning to match can deal with mismatch
  – Topic Modeling
  – Latent Space
  – String Re-writing Kernel
Publications of the Project

Thank You!

Contact: hangli.hl@huawei.com