From CEOs to Presidents: Transfer Learning for Relation Extraction

Danushka Bollegala
Graduate School of Information Science & Technology
Short Bio

Name
Danushka Bollegala

Education
2001~2005: BS (Eng.), The University of Tokyo
2005~2007: MS (Eng.), The University of Tokyo
2007~2009: PhD (Info. Sci.), The University of Tokyo

Work & Fellowships
2007~2010: JSPS (Japan Society for Promotion of Science) Research Fellow
2010 Winter: University of Sussex, Visiting Research Fellow
2011 Summer: University of Cambridge, Visiting Research Fellow
2010~Present: Assistant Professor, The University of Tokyo
Research Background

Natural Language Processing, Web Intelligence.

Multi-Document Summarization

Discourse Analysis

Entity Disambiguation, Alias Detection

Latent Relational Web Search Engines

Collaborative Exploratory Web Search Engines

Attributional & Relational Similarity

Evolutionary Computation-based Rank Learning for IR

Domain Adaptation Cross-Domain Sentiment Classification

The University of Tokyo

BS  MS  PhD  JSPS PD  Assistant Professor

2005  2006  2007  2008  2009  2010  2011  2012
The Web as a network of documents connected by links.

The Web as a network of entities connected by relations.
Applications of Relation Extraction

- **Information Retrieval**
  - For user queries on entities, we can return search results for related entities to improve user satisfaction
  - Relations can be used to navigate the search results

- **Semantic Web**
  - By automatically extracting relational tuples, we can overcome the metadata sparseness problem

- **Social Network Extraction**
  - Given a set of personal names, automatically extract the social network among those people from the Web
  - *spysee.jp*
浅田真央

トップ画像の変更

彼女の画像検索結果

浅田真央

1990年

2006年

TBS系「学校に行こうMAX」放映（「ニッポン…

YouTube

Twitter

浅田 真央 - Wikipediaより引用

フィギュアスケート選手

愛知県名古屋市

1090年生

女性

身長153cm

中京大学体育部

冬はスケート

バブル時代は五輪

五輪メダリスト

タグを編集

つながりの強いひと

関連人物をもっと詳しく見る

※不適切なページとして報告する
Relational Search (or $XYZ$ search)

If $X$ is to $Y$, then $Z$ is to?

If $USA$ is to $Lady Gaga$, then $Japan$ is to?

(USA, Lady Gaga), (Japan, ?)

Challenges in Relation Extraction from the Web

- Both structured and unstructured texts exist on the Web
- Large amounts of noisy data exist on the Web
- Novel entities and relations constantly appear on the Web
- Relations between existing entities vary with time
- Lack of labeled data for supervised relation extraction
Entity Disambiguation Problem

Multiple entities can have the same name on the Web (namesakes) Bollegala et al., Disambiguating Personal Names on the Web using Automatically Extracted Keyphrases, ECAI 2006.

Jim Clark (Founder of Netscape)   Jim Clark (F1 Champion)
Name Alias Detection Problem

Will Smith
Fresh prince

A single entity can have multiple names on the Web (aliases) Bollegala et al., Automatic Discovery of Personal Name Aliases on the Web, IEEE Trans. on Knowledge & Data Eng., 2011.
Relation Extraction – Problem Definition

- Given a crawled corpus of Web text, identify all the different semantic relations that exist between entities mentioned in the corpus.

**Approaches**

- **Supervised Relation Extraction**
  - Requires labeled data for each relation type to be extracted!
  - Number and types of relations must be specified in advance.

- **Unsupervised Relation Extraction (Open IE)**
  - Clustering entity pairs considering their semantic relations
  - Might not always extract relation types of interest.

- **Relation Adaptation**
  - Learn novel relation types using labeled data for existing relation types.
  - Minimum supervision
Relation Extraction – Problem Definition

- Given a crawled corpus of Web text, identify all the different semantic relations that exist between entities mentioned in the corpus.

Approaches

- **Supervised Relation Extraction**
  - Requires labeled data for each relation type to be extracted!
  - Number and types of relations must be specified in advance.

- **Unsupervised Relation Extraction (Open IE)**
  - Clustering entity pairs considering their semantic relations
  - Might not always extract relation types of interest.

- **Relation Adaptation**
  - Learn novel relation types using labeled data for existing relation types.
  - Minimum supervision
Relational Duality

ACQUISITION

(Microsoft, Powerset)

(Google, YouTube)

\[ \vdots \]

\[ \vdots \]

\[ X \text{ acquires } Y \]

\[ X \text{ buys } Y \text{ for } $ \]

Extensional definition

DUALITY

Intensional definition

Unsupervised Relation Extraction

Web -> crawler -> Text Corpus -> Sentence splitter -> POS Tagger -> NP chunker -> Pattern extractor

Lexical patterns
Syntactic patterns

Entity pair clusters
Lexico-syntactic pattern clusters

Cluster labeler (L1 regularized multi-class logistic regression)

Sequential Co-clustering Algorithm

Entity pairs vs. Patterns Matrix

X acquires Y
X buys Y
(Google, YouTube)
(Microsoft, Powerset)
...
Lexical-Syntactic Pattern Extraction

- Replace the two entities in a sentence by \( X \) and \( Y \)
- Generate subsequences (over tokens and POS tags)
  - A subsequence must contain both \( X \) and \( Y \)
  - The maximum length of a subsequence must be \( L \) tokens
  - A skip should not exceed \( g \) tokens
  - Total number of tokens skipped must not exceed \( G \)
  - Negation contractions are expanded and are not skipped

**Example**

- \( \ldots \text{merger/NN is/VBZ software/NN maker/NN [Adobe/NNP System/NN]} \text{ acquisition/NN of/IN [Macromedia/NNP]} \)
- \( X \) acquisition of \( Y \), software maker \( X \) acquisition of \( Y \)
- \( X \text{ NN IN Y, NN NN X NN IN Y} \)
Entity pairs vs. lexico-syntactic pattern matrix

- Select the most frequent entity pairs and patterns, and create an entity-pair vs. pattern matrix.

Entity pairs vs. Patterns Matrix

- (Google, YouTube)
- (Microsoft, Powerset)
- X acquires Y
- X buys Y
Relation Extraction – Problem Definition

- Given a crawled corpus of Web text, identify all the different semantic relations that exist between entities mentioned in the corpus.

**Approaches**

- **Supervised Relation Extraction**
  - Requires labeled data for each relation type to be extracted!
  - Number and types of relations must be specified in advance.

- **Unsupervised Relation Extraction (Open IE)**
  - Clustering entity pairs considering their semantic relations
  - Might not always extract relation types of interest.

- **Relation Adaptation**
  - Learn novel relation types using labeled data for existing relation types.
  - Minimum supervision
Domain Adaptation is the problem of generalizing a model trained on one domain to another domain.
Relation Adaptation

- Given training instances for some source relations $S_1, ..., S_k$ and some seed instances for a target relation $T$, learn a classifier to extract target relation.

Characteristics of relation adaptation

- Multiple source relation types
- Many training instances for the source relations
- Only a few (seeds) for the target relation type
- We are only interested in obtaining good performance on the target relation type
Challenges in Relation Adaptation

- **Challenge One**
  - features (e.g. lexical-syntactic patterns) that occur for source relations might not occur for the target relation.
  - **Proposed Solution**
    - Learn a low-dimensional mapping between source and target relation types.

- **Challenge Two**
  - Number of training instances for the target relation type is far less compared to that for the source relation types.
  - **Proposed Solution**
    - Perform one-sided undersampling to select a subset of source relation training instances to train a multi-class classifier.
# Relational Mapping

<table>
<thead>
<tr>
<th>leaderOf (source relation)</th>
<th>ceoOf (target relation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>President George Bush directed U.S. to an unnecessary war against Iraq.</td>
<td>Tim Cook personally directs Apple and makes final decisions on various UI designs.</td>
</tr>
<tr>
<td>[X direct Y]</td>
<td>[X direct Y]</td>
</tr>
</tbody>
</table>

| U.S. president George Bush attended the G8 summit last month. | Tim Cook became the CEO of Apple in August 2011. |
| [Y president X] | [X ceo Y] |

---

Recognizing Relation Independent Patterns

- Entropy of a pattern as a measure of independence
  - Hypothesis
    - If a pattern co-occurs with numerous entity pairs that have different relation types, then that pattern is relation independent.

\[
H(\rho) = \sum_{R \in \Omega} \sum_{(A,B) \in R} p(\rho, A, B) \log_2 p(\rho, A, B).
\]
Relational Mapping Algorithm

Input: An edge-weight matrix, $M \in \mathbb{R}^{(n-l) \times l}$ of a bipartite graph $G(V_{RS} \cup V_{RI}, E)$, and the number of clusters (latent dimensions) $k$.

Output: A projection matrix, $U \in \mathbb{R}^{n \times k}$.

1. Compute the affinity matrix, $A \in \mathbb{R}^{n \times n}$, of the bipartite graph $G$ as
   \[
   A = \begin{bmatrix}
   0 & M \\
   M^\top & 0
   \end{bmatrix}.
   \]

2. Compute the Laplacian, $L$, of the bipartite graph $G$ as $L = I - D^{-1}A$, where the diagonal matrix $D$ has elements $D_{ii} = \sum_j A_{ij}$, and $I \in \mathbb{R}^{n \times n}$ is the unit matrix.

3. Find the eigenvectors corresponding to the $k$ smallest eigenvalues of $L$, $u_1, \ldots, u_k$, and arrange them in columns to form the projection matrix $U = [u_1, \ldots, u_k] \in \mathbb{R}^{n \times k}$.

4. return $U$
Unbalanced Source vs. Target Instances

- The amount of source domain labeled training data is much larger than that for the target domain.
  - Supervised training from an unbalanced dataset leads to features in the minority class being *washed out.*
One-sided Undersampling

- Source Relation $S_1$
- Source Relation $S_2$

Initial Sample
All target relation instances + One randomly selected instance from each source relation

- Use the current sample to perform 1-nearest neighbor classification on source relation instances
- Append each misclassified instance to the current sample
Experiments

- **Dataset**
  - 20 relation types from the YAGO Ontology
    - Manually confirmed accuracy of 95%
  - 100 entity pairs (relation instances) for each relation type
  - source:60, target 20, test:20
    - source vs. target = 1140 (19 x 60) vs. 20

- **Relation Types**
  - ceoOf, isMarriedTo, actedIn, directed, isLeaderOf, livesIn, acquiredBy, hasChild, isCitizenOf, hasWonPrize, hasCurrency, hasOfficialLanguage, participatedIn, hasPredecessor, discovered, politicianOf, hasCapital, graduatedFrom, diedIn, worksAt

  - #of relation independent patterns = 1000, dimensions = 1000
  - Trained a multi-class logistic regression classifier with L2 regularization
  - Evaluate macro-averaged precision/recall/F-score for the target relation type
## Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>7.25</td>
<td>7.33</td>
<td>7.24</td>
</tr>
<tr>
<td>Relation Specific (RS) patterns</td>
<td>77.08</td>
<td>29.99</td>
<td>41.41</td>
</tr>
<tr>
<td>Relation Independent (RI) patterns</td>
<td>81.38</td>
<td>40.22</td>
<td>51.40</td>
</tr>
<tr>
<td>All patterns</td>
<td>79.34</td>
<td>37.11</td>
<td>47.94</td>
</tr>
<tr>
<td>Projected patterns</td>
<td>78.48</td>
<td>33.56</td>
<td>44.86</td>
</tr>
<tr>
<td>Combined (All patterns + Projected)</td>
<td>84.21</td>
<td>45.11</td>
<td>56.99</td>
</tr>
<tr>
<td>RS patterns + Sampling</td>
<td>83.58</td>
<td>38.44</td>
<td>49.78</td>
</tr>
<tr>
<td>RI patterns + Sampling</td>
<td>75.60</td>
<td>45.11</td>
<td>54.83</td>
</tr>
<tr>
<td>All patterns + Sampling</td>
<td>80.94</td>
<td>47.111</td>
<td>57.62</td>
</tr>
<tr>
<td>Projected + Sampling</td>
<td>72.07</td>
<td>37.33</td>
<td>47.61</td>
</tr>
<tr>
<td>Jiang ACL 2009</td>
<td>81.06</td>
<td>44.89</td>
<td>55.62</td>
</tr>
<tr>
<td>Combined + Sampling (PROPOSED)</td>
<td><strong>86.47</strong></td>
<td><strong>51.78</strong></td>
<td><strong>62.77</strong></td>
</tr>
</tbody>
</table>
Conclusions

- Extracting semantic relations between entities is an important problem that has numerous applications.
- Two approaches presented in this talk
  - **Unsupervised Approach**
    - Use duality between definitions of semantic relations in a sequential co-clustering algorithm
    - Open Information Extraction
  - **Relation Adaptation Approach**
    - Use labeled data for a set of source relation types to learn a classifier for a target relation type
    - Spectral clustering on the bi-partite graph reduces feature mismatch
    - One-sided undersampling reduces the imbalance of training data between source and target relation types
Open Issues..

- How to avoid *negative transfer*?
- How to handle temporal variations in semantic relations?
- How to verify that a particular relation actually holds?
- How to avoid expensive batch processing of corpora to reflect novel relations and entities?
- How to merge relation extractions from different snapshots of the corpus?
- What type of pre-compiled knowledge bases do we need/must build to answer relational queries on the Web?
Thank You

Danushka Bollegala
www.iba.t.u-tokyo.ac.jp/~danushka
danushka@iba.t.u-tokyo.ac.jp
@Bollegala