Relation Extraction, Neural Network, and Matrix Factorization

Presenter: Haw-Shiuan Chang
UMass CS585 guest lecture on 2016 Nov. 17
Relation Extraction
Knowledge Graph
Entities

- Common Types
  - NIL
  - Person
  - Organization
  - Location
    - country
    - state
    - city
  - Numbers/Date
- You can have different ontology
  - more fine-grain types
  - Common nouns (common sense)
- Classifiers is called “Named Entities Recognizer” (NER)
- Features
  - POS tagging
  - Lexicon dictionary (gazetteer)
  - word embedding or topic models
Relations

- Brendan born in USA
- Brendan traveled to Pittsburgh
- Brendan likes Amherst
- Brendan thinks midterm is hard

Brendan, birthplace, USA

Relation Extraction
Event Extraction
Sentiment Extraction
Belief Extraction
January 15, 2000

Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. Gates will now focus on the charitable foundation he runs with his wife Melinda Gates. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant.

Gates will now focus on the charitable foundation he runs with his wife Melinda Gates.

Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant.

Gates will now focus on the charitable foundation he runs with his wife Melinda Gates.

Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.

Edges = {
  spouse,
  located in,
  lives in
}
Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant.

Gates will now focus on the charitable foundation he runs with his wife Melinda Gates.

Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
Edges =
{  
  spouse,  
  located in,  
  lives in  
  founded  
}
Edges =

{ 
  spouse, 
  located in, 
  lives in 
  founded 
}
Edges = 
{
  spouse,
  located in,
  lives in,
  founded
}
is married to,
became the CEO of,
is headquartered in,
helped to start,
resides in the city of,
....
....
}
Research Questions

• How to extract relations?
  • Relation Extraction

• How to perform reasoning on relations?
  • Knowledge Base Completion

• What can be qualified as a relation?
  • Schema/Ontology Definition
Challenges in Relation Extraction

- The problem could be very general
  - Many different relations
  - Many ways to expressing one relation in different context
- Hard to get annotated data
  - Hard to get training data
  - Hard to evaluate the results completely
- Unlabeled data: Missing positive or False?
Challenges in Relation Extraction

• Different ways of defining what is correct
  • Correct in reality, correct in author’s belief, probably correct?

• Error propagation at different stages
  • Entity detection: F1 0.8~0.9
  • Entity Linking: F1 0.5~0.6
  • Relation Extraction: F1 0.2~0.3
Neural Networks
Or any other tags

Naive Bayes

chairs

HMM

end with s

comfortable

der end with s

Logistic Regression

chairs

end with s

CRF

comfortable

der end with s
Discrete Model

- Input/output:
  - sequence/sequence
- The feature is discrete
  - \( f(x): [0, \ldots, 1, 0, \ldots] \)
  - \( x: [\ldots, \text{chairs}, \ldots] \)
- The state is discrete
Neural Networks

Noun
Adj
V

word embedding

chairs

LSTM
GRU
CNN

comfortable
Continuous State Transition

- We can extend the size of the state to model longer dependency

Words with similar meaning

Traverse the state space similarly
# Applications of Sequence Learning

## Tags
- POS tag
- Dependency tags
- Entity types
- Event types
- Words
- Other language
- ...

## Applications
- POS tagging
- Dependency parsing
- Entity detection
- Event detection
- Language modeling
- Translation
- ...

...
How Discriminative you want?

- Model flexibility:
  - HMM < CRF < LSTM

- Number of parameters:
  - HMM < CRF < LSTM

- Performance (give sufficient data):
  - HMM < CRF < LSTM

- Interpretability:
  - HMM > CRF > LSTM

- Optimization easiness
  - HMM > CRF > LSTM

- Robustness against Overfitting
  - HMM > CRF > LSTM
Constraints and Problems

Prior knowledge of the problem

Training Data
External Resources

Model
Feature Engineering

Prediction

Trying many models?
Getting more data?
Describing precisely?
Matrix Factorization
Interpretations of Word Embedding

Unsupervised learning

- Word similarity
  - documents -> short context, semantic -> syntax
- Document similarity
- Topic modeling (LSA)
Structure on Interactions

- words
- words
- key words
- users
- entities
- entity pairs

near word (context)
documents
documents
documents
near entities
relations
## Applications

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Natural Language Processing</th>
<th>Application Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>words</td>
<td>word embedding</td>
<td>near word (context)</td>
</tr>
<tr>
<td>words</td>
<td>topic modeling</td>
<td>documents</td>
</tr>
<tr>
<td>key words</td>
<td>summarization</td>
<td>documents</td>
</tr>
<tr>
<td>users</td>
<td>recommendation</td>
<td>documents</td>
</tr>
<tr>
<td>entities</td>
<td>coreference</td>
<td>near entities</td>
</tr>
<tr>
<td>entity pairs</td>
<td>relation extraction</td>
<td>relations</td>
</tr>
</tbody>
</table>
Matrix Factorization plus Neural Networks

- Users
- Entity Pairs
- Sentence
- Words

Sentence Embedding
LSTM
GRU
CNN

word embedd
An example
Universal Schema

• How to extract relations?
  • Neural Network (LSTM)

• How to perform reasoning on relations?
  • Matrix Factorization

• What can be qualified as a relation?
  • Everything could be a schema
Wei Li studies at Xinghua U. Her 2008 publications include W. Li. "Scalable NLP" ACL, 2008.
Wei Li studies at Xinghua U. Her 2008 publications include "Scalable NLP" ACL, 2008.

Relation Extraction

Entity Extraction

Entity Mentions

Resolution (Coref)

Entities, Relations

Relation Extraction

Relation Mentions

query

KB

answer

Member(Wei Li, Xinghua U.)
Compositional Universal Schema

January 15, 2000

Tech pioneer **Bill Gates** stepped down today as chief executive officer of **Microsoft**, the Seattle-headquartered software giant. His long-time friend, **Steve Balmer**, will take over as CEO of **Microsoft**. Gates will now focus on the charitable foundation he runs with his wife **Melinda French Gates**. Bill and Melinda were married in a ceremony in **Hawaii**, rather than her hometown of **Dallas**. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside **Seattle**.
January 15, 2000

Tech pioneer **Bill Gates** stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.

Relation Types

<table>
<thead>
<tr>
<th>Structured</th>
<th>Textual</th>
</tr>
</thead>
<tbody>
<tr>
<td>worker</td>
<td>in</td>
</tr>
<tr>
<td>in</td>
<td>member</td>
</tr>
<tr>
<td>member</td>
<td>state</td>
</tr>
</tbody>
</table>

\[ \mathbb{P}((s, r, o)) = \sigma(u_{s,o}^Tv_r) \]
January 15, 2000

Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.

Bayesian Personalized Ranking (BPR)

\[ u_{s,o}^T v_r > u_{s,o}^T v_r \]
January 15, 2000

Tech pioneer **Bill Gates** stepped down today as chief executive officer of **Microsoft**, the Seattle-headquartered software giant. His long-time friend, **Steve Balmer**, will take over as CEO of **Microsoft**. **Gates** will now focus on the charitable foundation he runs with his wife **Melinda French Gates**. **Bill** and **Melinda** were married in a ceremony in **Hawaii**, rather than her hometown of **Dallas**. **Gates** moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside **Seattle**.
January 15, 2000

Tech pioneer **Bill Gates** stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife **Melinda French Gates**. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
January 15, 2000

Tech pioneer **Bill Gates** stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, **Steve Balmer**, will take over as CEO of Microsoft. **Gates** will now focus on the charitable foundation he runs with his wife **Melinda French Gates**. Bill and Melinda were married in a ceremony in **Hawaii**, rather than her hometown of **Dallas**. **Gates** moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside **Seattle**.
query: **spouse**

Joe NewEntity worked with his wife Jane NewEntity on relation extraction.
Joe NewEntity worked with his wife Jane NewEntity on relation extraction.
Joe NewEntity worked with his wife Jane NewEntity on relation extraction.
Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.

Input:
[per:spouse]
[María Múnera está casado con Juan M Santos]
Conclusions

• Relation extraction is one step further toward bridging the gap between text and knowledge

• Neural networks are powerful, but data is usually the bottleneck in relation extraction

• Matrix factorization can alleviate the requirements of large amount of annotations

• Universal Schema is an example of combining complex methods to solve complex problem
Multilingual Relation Extraction
Multilingual Relation Extraction

English

Spanish

Steve, Microsoft
Bill, Steve
Bill, Microsoft
Melinda, Bill
Microsoft, Seattle
Melinda, Dallas
Obama, US
Obama, Michelle

Bill, Microsoft
Melinda, Dallas
Obama, Michelle
Bill, Seattle
Multilingual Universal Schema

Relation Types

- **structured**
- **English**
  - textual
- **Spanish**
  - textual

Steve, Microsoft
Bill, Steve
Bill, Microsoft
Melinda, Bill
Microsoft, Seattle
Melinda, Dallas
Obama, US
Obama, Michelle
Bill, Seattle
Entity pair  Relation

- **spouse**
- **married to**
- **esposa**

**LSTM**

**Input:**
- Steve, Microsoft
- Bill, Steve
- Bill, Microsoft
- Melinda, Bill
- Microsoft, Seattle
- Melinda, Dallas
- Obama, US
- Obama, Michelle
- Bill, Seattle

**Output:**
- vector
- embedding
- distribute
- of
- semantically learned

**Relations:**
- CEO, chairman, president
- leader-of, head-of
- married-to, HQ-in
- spouse, presidente, vive en
- residente, fundador

**Entity Pairs:**
- Melinda, Bill
- Michelle, Obama

**Diagram Components:**
- LSTM network
- Input vectors
- Output vectors
- Relations represented by arrows
<table>
<thead>
<tr>
<th>Steve, Microsoft</th>
<th>Bill, Steve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steve, Microsoft</td>
<td>Bill, Microsoft</td>
</tr>
<tr>
<td>Melinda, Bill</td>
<td>Barack Obama, US</td>
</tr>
<tr>
<td>Microsoft, Seattle</td>
<td>Barack Obama, US</td>
</tr>
<tr>
<td>Melinda, Dallas</td>
<td>Barack Obama, US</td>
</tr>
<tr>
<td>Bill, Seattle</td>
<td>Barack Obama, US</td>
</tr>
</tbody>
</table>

**LSTM**

- **vector**
- **embedding**
- **distribute**
- **representational**
- **semantically learned**
- **of**
- **~~~~~~~**

**Words**
- **esposa**
- **casado**
- **gato**
- **married**
- **spouse**
- **presidente**
- **resident**
- **CEO**
- **chairman**
- **president**
- **leader of**
- **head of**
- **married to**
- **HQ in**

**Relationships**
- **spouse**
- **born in**
- **friend**
- **co-worker in**
- **lives in**
- **top member of**
- **member of**

**Diagram**

- **LSTM**
- **x**
- **y**
<table>
<thead>
<tr>
<th>Steve, Microsoft</th>
<th>Bill, Microsoft</th>
<th>Bill, Steve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obama, Michelle</td>
<td>Obama, US</td>
<td>learned</td>
</tr>
<tr>
<td>Microsoft, Seattle, Bill</td>
<td>Melinda, Bill</td>
<td>of</td>
</tr>
<tr>
<td>distributed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **LSTM**
- **vector**
- **embeddings**
- **distribute**
- **representative**
- **learned**
- **parameters**
- **~x**
- **~y**

- **spouse**
- **born-in**
- **friend**
- **per:co-worker**
- **per:lives-in**
- **org:top-members**
- **org:member**
- **CEO**
- **chairman**
- **president**
- **leader-of**
- **head-of**
- **married-to**
- **HQ-in**
- **esposa**
- **presidente**
- **vive en**
- **resident**
- **fundador**

- **cat**
- **gato**
- **resident**
- **spouse**
- **married**
- **casado**
- **esposa**
- **presidente**
LSTM

\( \vec{x} \Rightarrow \text{vector} \Rightarrow \text{embeddin} \Rightarrow \text{distribute} \Rightarrow \text{cat} \Rightarrow \text{married} \Rightarrow \text{casado} \)

\( \vec{y} \Rightarrow \text{spouse} \Rightarrow \text{born-in} \Rightarrow \text{friend} \Rightarrow \text{per:co-worker} \Rightarrow \text{per:top-members} \Rightarrow \text{org:member} \Rightarrow \text{CEO} \Rightarrow \text{chairman} \Rightarrow \text{president} \Rightarrow \text{leader-of} \Rightarrow \text{head-of} \Rightarrow \text{married-to} \Rightarrow \text{HQ-in} \Rightarrow \text{esposa} \Rightarrow \text{resident} \Rightarrow \text{vive en} \Rightarrow \text{residente} \Rightarrow \text{esposa} \Rightarrow \text{presidente} \Rightarrow \text{fundador} \Rightarrow \text{cat} \Rightarrow \text{married} \Rightarrow \text{casado} \)

\( \text{Steve, Microsoft} \Rightarrow \text{Bill, Steve} \Rightarrow \text{Bill, Microsoft} \Rightarrow \text{Melinda, Bill} \Rightarrow \text{Microsoft, Seattle} \Rightarrow \text{Melinda, Dallas} \Rightarrow \text{Obama, US} \Rightarrow \text{Obama, Michelle} \Rightarrow \text{Bill, Seattle} \Rightarrow \text{vector} \Rightarrow \text{embeddin} \Rightarrow \text{distribute} \Rightarrow \text{cat} \Rightarrow \text{married} \Rightarrow \text{casado} \)