A Hybrid Neural Model for Type Classification of Entity Mentions

Li Dong†*  Furu Wei‡  Hong Sun§  Ming Zhou‡  Ke Xu†
†State Key Lab of Software Development Environment, Beihang University, Beijing, China
‡Microsoft Research, Beijing, China
§Microsoft Corporation, Beijing, China
donglixp@gmail.com  {fuwei, hosu, mingzhou}@microsoft.com  kexu@nlsde.buaa.edu.cn
Motivation

- Types group entities to categories
- Entity types are important for various NLP tasks
- Our task: predict an entity mention’s type
  - Input
    - \([c_{-S} \ldots c_{-1}]\) left context
    - \([w_1 \ldots w_n]\) mention
    - \([c_1 \ldots c_S]\) right context
  - Output
    - Type
    - [an initiative sponsored by][Bill & Melinda Gates Foundation][to fight HIV infection]
- **Mention**
  - Bill & Melinda Gates Foundation (*Organization*)
  - Bill, Melinda, Gates -> {Person Name}
  - {Person Name} + Foundation -> Organization

- **Context**
  - [The greater part of ] [Gates] [ ' population is in Marion County .]  (*Location*)
  - [Gates] [ was a baseball player .]  (*Person*)
Architecture

\[
y = \frac{1}{\sum_{j=1}^{C} e^{\theta_j x}} \begin{bmatrix} e^{\theta_1 x} \\ \vdots \\ e^{\theta_C x} \end{bmatrix}
\]
RNN-based Mention Model

- Learn composition patterns for entity mention
  - \{Name\} + Foundation / University -> (Organization)
  - \{Body Region\} + \{Disease\} -> (Disease)

- **Recurrent Neural Networks** (Elman Networks)
  - Use a global composition matrix to compute representation recurrently
  - A natural way to learn composition patterns

\[
v_2 = \tanh(W \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + b^m) \quad v_3 = \tanh(W \begin{bmatrix} v_2 \\ w_3 \end{bmatrix} + b^m) \quad \ldots \quad v_5 = \tanh(W \begin{bmatrix} v_4 \\ w_5 \end{bmatrix} + b^m)
\]
MLP-based Context Model

- Use context to disambiguate
  - [The greater part of ] *Gates*' population is in Marion County.
    *(Location)*
  - *Gates* was a baseball player.
    *(Person)*

- MultiLayer Perceptrons
  - Location-aware, jointly trained

![Diagram](image-url)
Model Training

- Objective function
  \[
  \min_{\theta} - \sum_i \sum_j t^i_j \log y^i_j + \frac{\lambda_\theta}{2} \| \theta \|_2^2
  \]
  - cross entropy loss
  - regularization

- Back-propagation algorithm
  - Back-propagate errors of softmax classifier to other layers

- Optimization
  - Mini-batched AdaGrad
Automatically Generating Training Data

Wikipedia Article

The later stages of his career, Gates has pursued a number of philanthropic endeavors, donating large amounts of money to various charitable organizations and supporting research programs through the **Bill & Melinda Gates Foundation** established in 2000.

Gates stepped down as chief executive officer of Microsoft in January 2000. He remained on created the position of chief software architect for himself. In June 2000, Gates announced that he would be transitioning

**Wikipedia ID** ➔ **DBpedia** ➔ **DBpedia Entity**

**Anchor link**

**Context**

**Mention**
Automatically Generating Training Data

- DBpedia ontology
  - 22 top-level types
    - Organisation, MeanOfTransportation, Holiday, Work, Food, Award, AnatomicalStructure, Device, Colour, Language, TopicalConcept, EthnicGroup, Currency, Disease, Drug, Person, Place, Activity, CelestialBody, Event, Species, BioChemSubstance

- Wiki-22
  - #Train: 2 million
  - #Dev: 0.1 million
  - #Test: 0.28 million
Evaluation on Wiki-22

- micro-F1 / macro-F1 score
- Baseline methods
  - Support Vector Machine (SVM)
  - Multinomial Naive Bayes (MNB)
  - Sum word vectors (ADD)
    - Use a softmax classifier
- *-mention
  - Only use mention
- *-context
  - Only use context
- *-joint
  - Use both mention and context

<table>
<thead>
<tr>
<th>Method</th>
<th>Micro-F1</th>
<th>Macro-F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM-mention</td>
<td>90.2</td>
<td>89.7</td>
</tr>
<tr>
<td>MNB-mention</td>
<td>87.0</td>
<td>87.6</td>
</tr>
<tr>
<td>ADD-mention</td>
<td>90.1</td>
<td>90.7</td>
</tr>
<tr>
<td>HNM-mention</td>
<td>93.4</td>
<td>93.6</td>
</tr>
<tr>
<td>SVM-context</td>
<td>76.3</td>
<td>73.3</td>
</tr>
<tr>
<td>MNB-context</td>
<td>72.8</td>
<td>70.0</td>
</tr>
<tr>
<td>ADD-context</td>
<td>75.4</td>
<td>73.1</td>
</tr>
<tr>
<td>HNM-context</td>
<td>81.1</td>
<td>78.3</td>
</tr>
<tr>
<td>SVM-joint</td>
<td>93.5</td>
<td>93.4</td>
</tr>
<tr>
<td>MNB-joint</td>
<td>85.9</td>
<td>82.8</td>
</tr>
<tr>
<td>ADD-joint</td>
<td>94.1</td>
<td>93.9</td>
</tr>
<tr>
<td>HNM-joint (our)</td>
<td>96.8</td>
<td>96.5</td>
</tr>
</tbody>
</table>
Comparison with Previous Systems

- HYENA [Yosef et al., 2012]
  - Support Vector Machine
  - unigrams, bigrams, and trigrams of mentions, surrounding sentences, mention paragraphs, part-of-speech tags of context words, gazetteer dictionary

- FIGER [Ling and Weld, 2012]
  - Perceptron
  - unigrams, word shapes, part-of-speech tags, length, Brown clusters, head words, dependency structures, ReVerb patterns

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>Micro-F1</th>
<th>Macro-F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiki-5</td>
<td>HYENA</td>
<td>95.2</td>
<td>91.9</td>
</tr>
<tr>
<td></td>
<td>HNM-joint</td>
<td>95.0</td>
<td>93.6</td>
</tr>
<tr>
<td>News</td>
<td>FIGER</td>
<td>72.6</td>
<td>80.1</td>
</tr>
<tr>
<td></td>
<td>HNM-joint</td>
<td>75.1</td>
<td>80.6</td>
</tr>
</tbody>
</table>
Evaluation on Unseen Mentions

- Evaluate on unseen mentions (length > 2)
  - Mentions which do not appear in the train set

<table>
<thead>
<tr>
<th>Method</th>
<th>Micro-F1</th>
<th>Macro-F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM-mention</td>
<td>75.8</td>
<td>68.8</td>
</tr>
<tr>
<td>MNB-mention</td>
<td>75.5</td>
<td>69.0</td>
</tr>
<tr>
<td>ADD-mention</td>
<td>76.1</td>
<td>69.3</td>
</tr>
<tr>
<td>HNM-mention</td>
<td>82.5</td>
<td>75.6</td>
</tr>
</tbody>
</table>

- Help us deal with uncommon or unseen mentions
  - RNN-based mention model utilizes the compositional nature of mentions
Examples: Compositionality of Mentions

- Query similar mention examples
  - cosine similarity of mentions' vector representations

<table>
<thead>
<tr>
<th>English civil war</th>
<th>Spanish civil war / Greek civil war / Nigerian civil war / Angolan civil war</th>
</tr>
</thead>
<tbody>
<tr>
<td>Columbia University School of Law</td>
<td>Northwestern University School of Law / West Virginia University College of Law / University of Iowa College of Law / Golden Gate University School of Law</td>
</tr>
<tr>
<td>Subdural Hematoma</td>
<td>Intracranial Haemorrhage / Cardiac Arrhythmia / Duodenal Ulcer / Arterial Thrombosis</td>
</tr>
<tr>
<td>Joseph Jefferson Award</td>
<td>Margaret A. Edwards Award / Marian Engel Award / Doug Wright Award / Timothy Findley Award</td>
</tr>
<tr>
<td>Red-bellied Lemur</td>
<td>Oriental White-eye / Red-legged Honeycreeper / Black-crowned White-eye / Snowy Egrets</td>
</tr>
</tbody>
</table>

- Mentions that are of similar patterns are closer
Evaluation in Question Answering (QA)

- Web-based QA system [Cucerzan and Agichtein, 2005; Lin, 2007]
  - Add Q&A type interaction feature template

Q: who is the ceo of microsoft?

Search Engine (Bing)

Candidates Extracted from Titles and Snippets

Ranker

Answer Types

Classifier (18 types)

Person

[left context] [Satya Nadella] [right context]

[left context] [Xbox] [right context]

Feature Template: \{Type(Q) | Type(A)\}

\{Person | Person\} – positive weight

\{Person | Device\} – negative weight
Evaluation in Question Answering (QA)

- WebQuestions dataset [Berant et al., 2013]
  - Manually annotated question-answer pairs

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc@1</th>
<th>Acc@3</th>
<th>Acc@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/oTYPE</td>
<td>29.2</td>
<td>50.8</td>
<td>61.2</td>
</tr>
<tr>
<td>w/TYPE</td>
<td>33.5</td>
<td>55.6</td>
<td>64.4</td>
</tr>
</tbody>
</table>

Table 6: Evaluation results on the QA task. Type information obtained by our approach improves the accuracy. w/oTYPE: Without using type features in the answer ranking model. w/TYPE: Using type features in the answer ranking model.

- Our type classifier improves the accuracy of QA systems
Conclusion

- Recurrent Neural Networks are good at learning soft patterns
  - Compositional nature of entity mentions
  - Generalize for Unseen or uncommon mentions
- Automatically generate training data instead of annotating manually
- Type information is important for many NLP tasks

Future work

- Fine-grained type classification
  - Person -> doctor, actor, etc.
- Utilize hierarchical taxonomy
- Multi-label
- Utilize global information (e.g., document topic)
- ...
THANKS!