Transductive Non-linear Learning for Chinese Hypernym Prediction

Chengyu Wang\textsuperscript{1}, Junchi Yan\textsuperscript{1,2}, Aoying Zhou\textsuperscript{3}, Xiaofeng He\textsuperscript{4}
\textsuperscript{1} Shanghai Key Laboratory of Trustworthy Computing, East China Normal University\
\textsuperscript{2} IBM Research — China

Introduction

- Learning hypernymy relations is essential for taxonomy construction, fine-grained entity categorization, knowledge base population, etc.
- Extracting hypernyms for entities is still challenging for Chinese due to flexible language expressions.
- Our work maps Chinese hyponyms to hypernyms in the embedding space by transductive non-linear learning.

General Framework

- **Initial Stage**
  - Train two linear projection models to capture the semantics of is-a and not-is-a relations based on the training set (\(D^+ \) and \(D^-\)).
  - Estimate the prediction and confidence scores for unlabeled data \(D_U\).
- **Transductive Learning Stage**
  - Learn the final prediction score for each pair \((x_i, y) \in D_U\) based on initial prediction, linguistic rules and non-linear regularization.

Initial Model Training

- Train skip-gram models to map each word or concept with multiple words \(x_i\) to its embedding vector \(v_i\).
- Train two linear projection models with Tikhonov regularizers based on word embeddings. One for is-a relations, the other for not-is-a relations.

\[
J(M^*) = \frac{1}{2} \sum_{(x,y) \in D^+} ||M^+ x_i - y||^2 + \lambda ||M^+||^2
\]

\[
J(M^-) = \frac{1}{2} \sum_{(x,y) \in D^-} ||M^- x_i - y||^2 + \lambda ||M^-||^2
\]

- Estimate the prediction score \(\text{score}(x_i, y)\) and the confidence score \(\text{conf}(x_i, y)\) for each \((x_i, y) \in D_U\).

\[
\text{score}(x_i, y) = \tanh(||M^+ x_i - y|| - ||M^- x_i - y||)
\]

\[
\text{conf}(x_i, y) = \max\{||M^+ x_i - y||, ||M^- x_i - y||\}
\]

High prediction score: large probability of is-a relation between \(x_i\) and \(y\).
High confidence score: large probability that the models can predict the existence of is-a relations correctly.

Transductive Non-linear Learning (I)

- Initialize the \(m\)-dimensional final prediction vector \(F\) where \(m = |D^+| + |D^-| + |D_U|\).

\[
F_i = \begin{cases} 
1 & (x_i, y) \in D^+ \\
-1 & (x_i, y) \in D^- \\
0 & (x_i, y) \in D_U, u_i \sim \text{Uniform}(-1, 1)
\end{cases}
\]

- Define the objective considering results of initial prediction: \(O_u = ||W(F - S)||^2\).
  \(S\) is the initial prediction vector. \(W\) is set as follows:

\[
W_{ij} = \begin{cases} 
\text{conf}(x_i, y) & i = j, (x_i, y) \in D_U \\
1 & i = j, (x_i, y) \in D^+ \cup D^- \\
0 & \text{Otherwise}
\end{cases}
\]

Transductive Non-linear Learning (II)

- Define the objective considering linguistic rules: \(O_l = ||F - R||^2\).
- Compute the TP/TN rate \(\gamma\) for each positive/negative rule.
- If \((x_i, y)\) matches a collection of positive rules \(C_{(x_i, y)}\), define \(R_i = \max\{F_i, \max_{x \in C(x_i, y)} \gamma\}\).
- If \((x_i, y)\) matches a collection of negative rules \(C_{(x_i, y)}\), define \(R_i = -\max\{-F_i, \max_{x \in C(x_i, y)} \gamma\}\).
- Define the non-linear regularizer based on the TransLP framework: \(O_u = F^T \Sigma \Sigma F\).

\[
\Sigma(i,j) = \begin{cases} 
\cos(x_i, x_j) & y_i = y_j \\
0 & \text{Otherwise}
\end{cases}
\]

It assumes \(F_i\) and \(F_j\) w.r.t. \((x_i, y)\) and \((x_j, y)\) is similar if the candidate hypernyms \(y_i\) and \(y_j\) are the same and the candidate hyponyms \(x_i\) and \(x_j\) are similar in semantics.
- Optimize the combined objective function via blockwise gradient descent.

\[
J(F) = O_u + O_l + \mu_1 ||O_u|| + \mu_2 ||F||^2
\]

- Predict \(y_i\) is a hypernym of \(x_i\) if \(F_i > 0\) \(\theta \in (-1, 1))\).

Experiments

- **Datasets**: Two public Chinese datasets (i.e., FD and BK), consisting of Chinese entity pairs with labeled positive/negative is-a relations.
- **Metrics**: Precision, Recall and F-Measure.
- **Results**: Our approach outperforms all baselines for Chinese.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>FD</th>
<th>BK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>\text{Taxonomy Matching}</td>
<td>54.3</td>
</tr>
<tr>
<td></td>
<td>Linear Projection</td>
<td>64.1</td>
</tr>
<tr>
<td></td>
<td>Piecewise Linear Projection</td>
<td>66.4</td>
</tr>
<tr>
<td></td>
<td>Iterative Linear Projection</td>
<td>69.3</td>
</tr>
<tr>
<td></td>
<td>Vector Linearization Model</td>
<td>57.7</td>
</tr>
<tr>
<td></td>
<td>Vector Addition Model</td>
<td>65.3</td>
</tr>
<tr>
<td></td>
<td>Vector Subtraction Model</td>
<td>71.9</td>
</tr>
<tr>
<td>Ours (Initial)</td>
<td>70.7</td>
<td>69.2</td>
</tr>
<tr>
<td>Ours</td>
<td>72.8</td>
<td>70.5</td>
</tr>
</tbody>
</table>

- **Examples of model prediction.**

<table>
<thead>
<tr>
<th>Candidate Hyponym</th>
<th></th>
<th>Candidate Hypernym</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity (Stefanie Sun)</td>
<td></td>
<td>Entity (Chemical Product)</td>
</tr>
<tr>
<td>Chemical</td>
<td>\checkmark</td>
<td>(Stefanie Sun)</td>
</tr>
<tr>
<td>Organic</td>
<td>\checkmark</td>
<td>(Chemical Product)</td>
</tr>
<tr>
<td>Organic</td>
<td>\checkmark</td>
<td>(Stefanie Sun)</td>
</tr>
<tr>
<td>Chemical Product</td>
<td>\checkmark</td>
<td>(Stefanie Sun)</td>
</tr>
</tbody>
</table>

Supplementary experiments: Our approach is comparable to many existing methods in the English environment. (please refer to the paper for details).

Conclusion and Future Work

- We propose a transductive non-linear learning approach for Chinese hypernym prediction. It has high accuracy and does not require parsing Chinese texts and training deep classification models.
- Future work: constructing a complete taxonomy from texts in Chinese.

References