Semantic Annotation with RescoredESA: Rescoring Concept Features Generated From Explicit Semantic Analysis

Zhuoren Jiang
Dalian Maritime University

Miao Chen*
Indiana University Bloomington

Xiaozhong Liu
Indiana University Bloomington

INTRODUCTION
Annotating text with a set of concepts reveals semantically important entities and thus is an important step in semantic annotation. This study we try to extend a semantic annotation approach, Explicit Semantic Analysis (ESA), by enriching it with context and knowledge base.

ESA formula: \( v_E = v_d \cdot M^T \)

RescoredESA:

\[
e_i + \alpha \left( \frac{\sum_{j=1}^{k} LS(C_{E_i}, C_{I_j})}{k} \right)
\]

\[
r_i = \frac{1 + \alpha}{1 + \alpha}
\]

THE RESCORED ESA APPROACH

In-text Wikipedia Concepts

<table>
<thead>
<tr>
<th>String</th>
<th>Matched Wikipedia concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radio</td>
<td>[radio]</td>
</tr>
<tr>
<td>Tape</td>
<td>[Cassette deck]</td>
</tr>
<tr>
<td>Toyota Tercel</td>
<td>[Toyota Tercel]</td>
</tr>
</tbody>
</table>

Wikipedia Concepts from ESA

[Intim Van Ostrand] [Ostrander, Minnesota]
[USS Tercel (AM-386)] [Toyota Paseo]
[Toyota Tercel] [Ostrander, Ohio]
[John Ostrander] [Toyota Origin] [Toyota Raum]
[T. C. Ostrander]

Text

“...I’m looking for a replacement radio/tape player for a 1984 Toyota Tercel...”

RESULTS
The results show that for the three text categorization methods other than BalanceWinnow, the general trend is that: Concat (BOW + RescoredESA) > BOW > InText > RescoredESA > ESA. For BalancedWinnow, the order is similar but with ESA and RescoredESA reversed: Concat (BOW + RescoredESA) > BOW > InText > ESA > RescoredESA.

EXPERIMENT
In order to examine performance of RescoredESA, we use the 20 newsgroups data, a widely used news data set for text classification. RescoredESA, along with other comparable approaches, are evaluated using their performance on classifying 20 newsgroups - that is, first representing text with these approaches and then perform text classification based on the representation results.

Parameter \( \alpha \)

Parameter \( \gamma \)