Opinion Word Expansion and Target Extraction through Double Propagation

Srijan R. Shetty
11727
Indian Institute of Technology, Kanpur

Opinion Word and Target extraction

There are three possible relations between the targets and words:
- Opinion-Target relation (OT - Rel)
- Target-Target Relation (TT - Rel)
- Opinion-Opinion Relation (OO - Rel)

Each relation is expressed as a four tuple <POS(w), D, R, POS(v)>, where D is the dependency type in the dependency graph, R is the syntactic relation.

The relations OT - Rel, OO - Rel and TT - Rel can be characterized by four rules R1 to R4 given in table 1 which can be used to extract one knowing the other.

<table>
<thead>
<tr>
<th>Rule ID</th>
<th>Observations</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>O - O Dep - T s.t. O ∈ (O), O Dep ∈ t - T [MR], POS(O) ∈ [NW]</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>O - O Dep - H - T Dep - T s.t. O ∈ t - T [O], OT Dep ∈ [MR], POS(T) ∈ [NN]</td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>O - O Dep - H - T Dep - T s.t. O ∈ t - T [O], OT Dep ∈ [MR], POS(O) ∈ [NN]</td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>T Dep - T Dep - T Dep - T s.t. T ∈ t - T</td>
<td></td>
</tr>
</tbody>
</table>

After the extraction of words, a polarity is assigned to targets and words one the basis of the following rules:
- Heterogeneous rule: Opinion words and targets which are extracted from known targets and words respectively have the same polarity as that of known ones.
- Homogenous rule: Opinion words and targets extracted from known words and targets respectively have the same polarity as the known ones.
- Intra-review rule: Across reviews targets and opinion words have no relation. In such a case, the polarity of the review is used to assign a polarity to the word.
- Polarity assignment may lead to multiple polarities being assigned to a word or target, a total count of positive and negative polarities is used to disambiguate the polarity in such a case.

Noisy targets are removed by using frequency based global pruning. After the extraction of words, a polarity is assigned to targets and words one the basis of the following rules:

Method

A dependency parser is used to create a dependency graph of the sentence, the nodes of the graph are then labelled by Stanford POS tagger.

The extraction happens through Double Propagation; it is an iterative procedure, where we employ rules R1 to R4 at each step to new obtain new words and targets using the older ones.

Noisy targets are removed by using clause pruning.

References
