Relation Extraction From Text

Purvanshi Mehta

Objective-Support Vector Machine based relation extraction.

Studies Done-

1.**Distant Supervision for Relation Extraction using Tree Kernels**
Goal-
Improving the Distant Supervision algorithm for automatically creating labeled training data.

Dataset-
The New York Times (NYT) corpus to train the system.
Training set:4,700 relations
Test set:1900 relations

Data Processing-
- Both lexical and syntactic features were used in the feature vector.
- Named entities were tagged using Standford named entity recognizer
- Lexical features were extracted from the OpenNLP tool.
- Constituency Parse trees were obtained by Charniak parser.
- Learning model-SVM-Light- TK3
- Overlapping conditions were handled by collecting classifier decisions of one vs all classifier.

2. **End-to-End Relation Extraction Using Distant Supervision from External Semantic Repositories**
Goal-
Relation extraction on Wikipedia by considering relations defined on external repositories(YAGO and subset of Wikipedia documents)

Dataset-
Training Set-1.YAGO, a large knowledge base of entities and relations
2.Freebase, a collection of Wikipedia articles.

Testing set-200 articles from Freebase.

Method-

- Relation extractions were obtained from YAGO. For each pair of entities in some YAGO relation, all the sentences of the Freebase documents that contain such entities were retrieved.
- 18 regular expressions to extract dates and other 25to extract numerical attributes were designed.
- All of Wikipedia articles in Freebase were scanned for its NEs.
- Learning Model-Two parsing paradigms were used $(\mathrm{K_P}$, and $\mathrm{TK})$ by adding the kernel above with six sequence kernels.
- The category labels in the nodes of the trees and sequences had to be removed to apply the kernel to Wikipedia documents.
- The candidate relations were generated by iterating all pairs of entity mentions in the same sentence.
- The One vs. Rest strategy is employed by selecting the instance with largest margin as the final answer.

3. Distant supervision for relation extraction without labeled data
Goal-

Relation extraction using unlabelled corpora, avoiding the domain dependence of ACEstyle algorithms, and allowing the use of corpora of any size.

Dataset-

 Freebase

Method-

- Freebase was used to give training set of relations and entity pairs (Using entity tagger).
- FEATURE EXTRACTION-lexical and syntactic(Minipar parser) features were used .Named entity tagging was done using Stanford four-class named entity tagger.Rather than using all features only conjugated features were used.
- DATA PROCESSING-

1.The full text of all Wikipedia articles was sentence tokenized by Metaweb Technologies.
2.Each sentence of this unstructured text is dependency parsed by MINIPAR to produce a dependency graph.
3.Consecutive words with name entity tagged are chunked.
4.Multi-class logistic classifier optimized using L-BFGS with Gaussian regularization.

Results
1.**Distant Supervision for Relation Extraction using Tree Kernels**-
The precision and recall obtained is rather low as compared to the previous work done,but is obtained with much less effort.Moreover SVM classifier has been has been used to handle overlapping

relation problem.Several methods are being tried to decrease the number of false positives generated by DS.

2. **End-to-End Relation Extraction Using Distant Supervision from External Semantic Repositories-**

In traditional DS was implemented by the Infobox, which is connected to the sentences by a proximity relation but here the use of DB of relations such as YAGO and any document of Freebase was allowed. Gave a high a high F1 of 74.29% on extracting 52 YAGO relations from any Wikipedia document. This result improves the previous work by 13.29 absolute percentage points.

3. **Distant supervision for relation extraction without labeled data-**

The classifier takes as input an entity pair and a feature vector ,returns a relation name and confidence score .The combination of syntactic and lexical features provides better performance than either feature set on its own.

References

1. Truc-Vien T. Nguyen and Alessandro Moschitti End-to-End Relation Extraction Using Distant Supervision
2. from External Semantic Repositories Department of Information Engineering and Computer Science
3. University of Trento
4. Mike Mintz, Steven Bills, Rion Snow, Dan Jurafsky Distant supervision for relation extraction without labeled data Stanford University
5. Azad Abad and Alessandro Moschitti Distant Supervision for Relation Extraction using Tree Kernels DISI, University of Trento, Italy Qatar Computing Research Institute, Qatar
6. Culotta, A., Sorensen, J.:Dependency tree kernels for relation extraction.: 42nd An-nual Meeting on Association for Computational Linguistics. Association for Com-putational Linguistics (2004)
7. Eugene Agichtein and Luis Gravano. 2000. Snowball:Extracting relations from large plain-text collections.In Proceedings of the 5th ACM International Conference on Digital Libraries.
8. Michele Banko, Michael J. Cafarella, Stephen Soderland,Matthew Broadhead, and Oren Etzioni. 2007. Open information extraction from the web. In Proceedings of IJCAI, pages 2670–2676.