UMBC Team: Event Nugget Detection and Coreference Classification of Event Nugget using Context Features and Semantic Similarity

Taneeya Satyapanich
University of Maryland, Baltimore County
Baltimore, MD, 21250, USA
taneeya1@umbc.edu

Tim Finin
University of Maryland, Baltimore County
Baltimore, MD, 21250, USA
finin@umbc.edu

1 System Description

We considered the event nugget detection problem as a document classification problem. We thought that the document that contained the similar event will contain the similar group of words. We did not use the entire document to build the classification model. We filtered only the context of the event nuggets such as a phrase, a sentence to be used as input data. Then we built features vectors using: word n-gram, character n-gram, prefix and suffix, word shape, the semantic similarity between the event nugget and its event subtypes. For the similarity values, we used our UMBC STS system (Han, 2013). We used the Stanford Classifier tool (Manning and Klein., 2003) to build the classification model of each event subtypes.

2 Training data

We combined the LDC2015E69 and LDC2014E121 corpus. The total is 351 files. Then we divided them into training data (80%) and development data (20%). We experiment to find the best parameters for classification model. We can reach 0.80 of F1-measure when testing with development data.

3 Testing

The testing step was tricky different from when we build the training data. Because we do not know where the event nugget is in the document. We cannot find the context of the event nugget to be our input data. So we have to scan through every sentence and then divided them into chunks as noun phrase or verb phrase followed theirs part of speech tags. Then we build features vectors from these chunks instead. After we put these features through the classifier, we selected only the answer of the classifier that has confidence score higher that our thresholds. These thresholds were defined by observing the confidence score of development data. About the coreference resolution of the event nugget, we introduced the Stanford CoreNLP (Manning, et.al., 2014) to find the coreference in the document. We got the coreference sets from Stanford CoreNLP and we have the event nugget sets from our system that include the context of each nugget. We compared these two sets and selected the event as the coreference if they appeared in both sets.

References

Lushan Han, Abhay L. Kashyap, Tim Finin, James Mayfield, and Johnathan Weese. 2013. *UMBC EBIQUITY-CORE: Semantic Textual Similarity Systems*,. In Second Joint Conf. on Lexical and Computational Semantics. Association for Computational Linguistics.

Christopher Manning and Dan Klein.. 2003. *Optimization, Maxent Models, and Conditional Estimation without Magic*. Tutorial at HLT-NAACL 2003 and ACL 2003.

Manning, Christopher D., Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., and McClosky, David. 2014. *The Stanford CoreNLP Natural Language Processing Toolkit*. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.