Lillyan Pan (ldp54)

CS 4740 Critique 2

*The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text*

Authors: Richard Farkas, Veronika Vincze, Gyorgy Mora, Janos Csirik, Gyorgy Szarvas

The following paper gave an analysis of the approaches and methods used in the Uncertainty Detection Task. As the paper described the approaches of over 20 teams, descriptions of each team’s methods were reduced to only the key features of their implementation. For example, the model developed by Zhang et al. was described as “a cascade system which utilized directly the predicted scopes during predicting other scopes in the same sentence.” While this statement does give a high level overview of the group’s implementation, the there is a lack of detail in exactly what feature selection was used, the dataset used, as well as the features employed. However, as this paper aimed to only provide high level overviews– as words such as “typically”, “usually”, “mostly”, were used frequently– it is reasonable to keep implementation details at the minimum and allow readers to look into the specific group’s paper on their CoNLL-2010 implementation for further details.

In terms of limitations, the evaluation metrics for Task 2 were a “scope-level F measure…where true positives were scopes which exactly matched the gold standard cue phrases and gold standard scope boundaries assigned to the cue word.” In terms of applications, this type of evaluation may be overly strict as the goal is to find uncertainty spans, rather than cue words. This may cause lower uncertainty F scores as both span and cue word must be correct, despite the end focus being determining the overall uncertainty of a text.

However, the paper and uncertainty detection methods mentioned bring the possibility for much needed advancement in uncertainty detection and an increase in possible applications. As uncertainty was detected in biological texts, one possible study could look at the relationship between uncertainty scores and correct diagnosis as well as the validity of the biomedical conclusions made in the analyzed paper.

Another interesting study would be to see the extent uncertainty plays in gendered communication. Often women’s speech contains hedges that are used to reduce forcefulness in communication. This stems from common pressure to remain unaggressive to avoid negative perceptions of character. The ability to classify gendered text based off uncertainty may give insight on differences in how aggressiveness is perceived between genders. This understanding may help improve the current gender biases.