Student: Lloyd Shatkin

Title: Automatic Detection of Fake News

Authors: Veronica Perez-Rosas, Bennet Kleinberg, Alexandra Lefevre, Rada Mihalcea

Link: http://aclweb.org/anthology/C18-1287

General Project Description: The intention of the paper I will replicate is to detect Fake News, where fake news consists of news that is realistic but is also not grounded in truth. The term "realistic" is meant to separate a news article that could be easily seen as true, compared to a satirical article that is written with the intention of appearing obviously false. Detecting fake news rather than satirical news presents a more challenging problem, as the obvious indicators of falsity in satire does not present itself in fake news.

The project originally involved creating a dataset with several hundred extractions of news articles, and corresponding fake news articles created by humans to appear similar to the original article but including some false aspects. The dataset was split into two, with one data subset corresponding to news coming from sources of CNN, the New York Times, and others of the like, and other data subset coming from People magazine and other magazines documenting more personal affairs. The authors believed that these types of articles differed in substantial ways, and furthermore that their fake counterparts differed in substantial ways. The full dataset is available online, so I will use the dataset presented rather than re-gathering the data.

The original authors used an SVM model with 5 features. The features are N-Grams, punctuation, psycholinguistic features (provided by the LIWC lexicon), readability and syntax. I will use these same features, trying to best replicate how these features are defined. Using the given features, I will replicate the model built by the authors and display the success metrics displayed by the authors. Further, I will re-create any and all figures presented in the paper.

At this point, I have no direct ideas for further extension other than attempting to add more features. I have found that extensions often present themselves while completing the project, so I plan to continue evaluating and critically thinking about what more I could add to this project. Further, some of the authors of this paper are members of the UM community, and I hope to discuss with them what further research they thought of and would have completed given more time.