ACL Paper Summary – Prachi Patel and Kim Ngo

The paper we chose to look at is called, “How Did This Get Funded?! Automatically Identifying Quirky Scientific Achievements”. This paper was written by Chen Shani (a PHD student), Nadav Borenstein, and Dafna Shahaf. Shani and Borenstein had equal contribution to this paper and all three authors are affiliated with The Hebrew University of Jerusalem.

The authors choose to address the issue of identifying humor. In particular, the authors chose to try and identify humor in funny and unusual scientific papers as inspired by the Ig Nobel Prize – awarded to celebrate funny scientific achievements. In this case the humor involved, requires common sense and specialized knowledge and understanding of scientific culture. The funniness of a paper can be discovered from the title a lot of the times, so this was a relatively clean way to explore automatic learning of humorous scientific papers.

The authors relied on a lot of prior work to get them started with this problem. They broke down the prior work, the research is based upon into two categories: Humor in Humanities and Humor Detection in AI. Humor in Humanities is regards to the fact that a lot of work on theoretical humor comes from linguistics and psychology. Humor in this sense is divided into three categories: incongruity, sexual, and nonsense, which they discovered from Ruch’s work. They chose to neglect the nonsense category, because published scientific papers and articles would not have any nonsense, and thus focused on incongruity and sexual humor. Using the work of prior researchers, they discovered that incongruous humor results from the “violation of the expectations”. They also found sex-related humor to be common in funny scientific papers. The next category of prior work involved Humor Detection in AI. They discovered that most computational humor detection involved supervised or semi-supervised methods and focused on specific, narrow, types of jokes or humor. These narrow types of humor they found based on domains created by researchers prior included knock-knock jokes, funny tweets, one-liners, funny product reviews, and even “That’s what she said” jokes. They also saw in the prior work of other researchers that irony, sarcasm, and satire have also been explored.

These authors created unique contributions to the community. They formulated a new humor recognition task in the scientific domain by trying to identify funny scientific papers and as a result they also created a dataset that contains thousands of funny scientific papers that could be used for future research. The authors Shani, Borenstein, and Shahaf combined existing findings from psychology and linguistics with recent NLP advances to develop many classifiers. Using a corpus of more than 0.6 M

papers, they evaluated these classifiers on their dataset and in a real-world setting to identify potential Ig Nobel Prize candidates. They created a unique, rigorous, and data-driven way to aggregate crowd worker's annotations for subjective questions regarding the funniness of a paper and its title. These are some of the ways these authors have created a unique contribution to the researching community.

To evaluate their research, the authors took two approaches. One was to evaluate the labeled dataset on the five models (Iggy(the one they created), Bert, Bertf, SciBert, and SciBertf)(f at the end indicates feature based). For a naïve baseline, to compare, they added two Bag-of-Words models: random forest and logistic regression. After splitting the data into a 80/10/10 split for the train/development/test sets respectively while using the development set to tune hyper-parameters, they found accuracy, precision and recall for each model. They noticed that the BOW models fell behind while the other five models gave high accuracies. The SciBertf model outperformed all the models across accuracy, precision, and recall. To see if the models were able to retrieve and accurately predict the Ig Nobel winners, the authors took 1,496 humorous papers found on the web and the 211 Ig Nobel winners, with the test set containing the Ig Nobel winners and 211 negative titles. The idea was to see if the models trained on just web-based funny papers are good predictors for Ig Nobel winners. The SciBertf and the Bertf performed the best showing that the pretrained models with literature-based features ended up being better. They concluded that the features actually were informative for the Ig Nobel-worthy papers detection task. The second way that they evaluated performance was testing the models in a more realistic setting by computing the funniness of the 630k articles from Semantic Scholars and checking the accuracy by crowdsourcing on Amazon Mechanical Turk to ask people to rate the funniness of the title and the paper from 1-5. They also compared the crowdsourced people's results with the results of a scientific expert who had been trained on the problem. Further comparison was done working with labeled data by asking the crowd workers to rate papers that they knew was funny or not. They found that their model Iggy had better precision with k between 0-50 and that SciBert and Bertf did better with larger k values from 50-300. Using these two evaluative methods, the authors chose to test the models' accuracy and precision in predicting funny scientific articles.

Chen Shani who is a PHD student has received 87 total citations. Looking at her work, I can see influences of psychiatry, and much work involving language modeling. Nadav Borenstein of all three authors has the least number of citations with only 1 citation. Dafna Shahaf who is an associate professor has by far the most citations of the three authors with 2,044 citations and 1,180 of those citations being since 2017. In her work, Shahaf has experience in dealing with humor detection when

she also did research in identifying humorous cartoon captions. In our opinion, their work is important because it shows that their humor identification model, that they created based on literature Iggy, which is relatively simple, can still perform well compared to complex BERT based models. It also shows that pretrained models combined with features that are representative of a problem can give good accuracy in predicting humor.

References

Chen Shani, Nadav Borenstein, and Dafna Shahaf. 2021. How Did This Get Funded?! Automatically Identifying Quirky Scientific Achievements. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 14–28, Online. Association for Computational Linguistics.