Opinion - Fact Classifiers

Aman Gupta
University of Illinois, Urbana-Champaign

Introduction

In today's 24-hour news cycle, hyper-polorized news pieces and a high level of scrutiny on the news media, it has become more important than ever for consumers of news to understand if they are reading an opinion or fact piece. Additionally, with numerous new outlets parading as legitimate and factual news sources, it is imperative that technology be able to flag which pieces of news are actual fact versus opinion.Being able to consume pieces that are purely factual would have a large positive impact on society.

Additionally, in today's society there is a renewed interest in reading and understanding a large variety of perspectives when understanding a social, political, economic or other issue. It is clear that users that are searching for and consuming news online would benefit from the ability from not only finding factual news articles but also finding and understanding different opinion based pieces. Such clear demarcation would help users better navigate the wealth of news that is produced today. All of these benefits are accentuated by today's socio-political climate - large political events such as the US Election, social movements such as Black Lives Matters, and world-wide events such as the COVID-19 Pandemic have made finding factual news a top priority for most users of the web. In this review I will illustrate the work that has previously been done in fact-opinion classifiers, issues that exist in the work today, and what the future for this field might hold.

Work in classifying opinion and fact based documents can occur at three levels, classifying documents as fact-based or opinion-based, classifying sentences as fact-based or opinion-based, and classifying if a document or a sentence's opinion is positive or negative[1]. The ability to classify documents at these three levels would not only help users, but would also be integral for any NLP or text classification solution that is attempting to segment news information. Work in this field started in the 1990's when Weibe (1999) used a sentence-level Naive Bayes classifier by using the presence or absence of particular parts of speech, sentence position, and punctuation based on 1000+ annotated sentences (fact or opinion)[2]. This was the beginning of sentence - based classification, and was followed by Hatzivassiloglou and Weibe's work in 2000 to showcase that adjectives that are detected and gradable are useful features for similarity[3]. And in 2002, Weibe showed that for document-level subjectivity, using a

k-nearest neighbor algorithm based on total count of opinionated words and phrases in each document.

When it came to classifying the positivity or negativity of a particular opinion, Hatzvassiloglou and McKeown showed in 1997 that an unsupervised learning algorithm concluded with 90% accuracy[4]. More recently however, Yu and Hatzivassiloglou have showcased a new methodology for completing these three tasks in fact-opinion classifiers[1]. For the classification of documents, they used no supervised data, and instead relied completely on content and metadata provided with an article. For sentence level bassifiers, they looked at sentence similarity (similarity of an opinion sentence to another opinion sentence), naive - bayes classifiers, and approaches that utilize multiple classifiers. And for polarity and opinion sentences, they used co-occurrences of words in documents from words in the seed training set, and sentence polarity tagging, looking for the average per word log likelihood. This study represents the state of the art fact-opinion classifiers today, with 97% accuracy at document level classification used a Baysian classifier, 91% accuracy at the sentence level when picking the MAX of the three options, and 90% for polarity of opinionated sentences and documents.

Apart from generic topic and sentence level work, Mullick et. all have also showcased a classifier that detects opinions and facts, not only from news articles but also from social media data, such as tweets and social media comments[5]. They furthermore built a classifier that is able to compare the opinionatedness of sections within a newspaper, and between newspapers as a whole. The approach utilized here was Bagging with Random Forest Trees, which resulted in the highest accuracy for both opinionatedness between sections in a news paper and in opinionatedness of social media comments and hashtags. The implications here are also massive - it is easy to see uses for this technology in flagging highly opinionated news sources on social media, flagging comments by elected officials as opinions not facts, and helping newspapers understand their different sections.

Lastly, Stepinski and Mittal, showcased that UI patterns for helping users parse through opinionated and factual news articles help alleviate user frustration with searching for a variety of news sources[6]. With pre-classified articles, they showcase that a slider allowing a user to pick what ratio of opinion based and fact based articles to view, a user is able to help fine tune classification algorithms and find a mix that suits their needs.

It is clear that the implications and needs for fact-opinion based text are only going to become more important in the future. Future work that needs to be explored here is

facts versus lies, as we have seen the propagation of lies in the news recently. Additionally, there is work to be done in classifying more than just news data, such as social media posts and tweets as a majority of Americans consume news from social media today.

References

1. Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of the 2003 conference on Empirical methods in natural language processing (EMNLP '03). Association for Computational Linguistics, USA, 129–136. DOI:https://doi.org/10.3115/1119355.1119372
2. Rebecca Bruce and Janyce Wiebe. 1999. Recognizing subjectivity: A case study in manual tagging. Natural Language Engineering, 5(2).
3. Vasileios Hatzivassiloglou and Janyce Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In Conference on Computational Linguistics (COLING-2000).
4. Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL, pages 174–181, Madrid, Spain, July. Association for Computational Linguistics.
5. *Ankan Mullick, Shivam Maheshwari, Soumya C., Pawan Goyal, and Niloy Ganguly. 2017. A Generic Opinion-Fact Classifier with Application in Understanding Opinionatedness in Various News Section. In Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 827–828. DOI:https://doi.org/10.1145/3041021.3054270*
6. *Adam Stepinski and Vibhu Mittal. 2007. A fact/opinion classifier for news articles. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '07). Association for Computing Machinery, New York, NY, USA, 807–808. DOI:https://doi.org/10.1145/1277741.1277919*