Alanis Mattia
MIS 2502
Extra Credit
11 April 2021

Cyberbullying Prevention Through Natural Language Processing

Natural language processing (NLP) is a branch of artificial intelligence (AI) that combines computational linguistics with statistical, machine learning, and deep learning models to give computers the ability to process textual and verbal context (IBM Cloud Education, 2020). The combination of these technologies allows computers to understand human language and its intent, similar to the ability of a human. It is crucial that programmers teach computers how to understand human language irregularities—such as metaphors, sarcasm, and idioms—for the applications to be considered useful (IBM Cloud Education, 2020). There are many uses of NPL in current applications such as spam detection, machine translation (such as Google Translate), virtual assistants and chatbots, social media sentiment analysis, and text summarization. One interesting, and highly relevant, current application of NPL is the detection and prevention of cyberbullying. With the rapid growth of technology and social media use among children and teens, cyberbullying has become prevalent, resulting in high rates of depression, feelings of isolation and loneliness, and in extreme cases, suicide (Wu, 2019).

There are two critical steps involved in successful machine learning-based cyberbullying detection: (1) Representation learning for Internet messages and (2) Classification (Zhao, Zhou, & Mao, 2016). To carry out the first step, many programmers utilize the Bag of Words (BoW) model. This technique essentially pre-processes text by converting it into simple lower-case text with no punctuation or non-word characters, all while keeping a count of the most frequently used words (Zhao, 2019). Feature extraction techniques are then used to discard words that are infrequently used or not classified as insulting words. Word embeddings work in conjunction with the BoW model to identify semantic

information behind words. This information is then utilized to provide a robust text analysis, in which the computer can then classify words as harmful (Zhao, Zhou, & Mao, 2016).

NPL has been used to detect cyberbullying on a number of platforms such as Facebook and Instagram. In 2016, Facebook introduced DeepText—a learning-based text engine that can understand textual content (Wu, 2019). The company combines this form of AI with human knowledge through a content moderating team to identify harassment and remove fake or harmful accounts (Wu, 2019). Though Instagram initially used the tool to detect spam, it recently began applying the technique to eliminate online trolls and identify cases of harassment. Other online platforms, including Twitter, Google, Jigsaw, and Youtube, have also begun utilizing AI and NLP to detect and eliminate bullying and harassment. Though the addition of human intervention is currently recommended to elicit the best results, the applications of NPL have been extremely beneficial in combating one of the worst forms of bullying plaguing the communication platforms of today's youth.

Throughout this course, we reviewed a number of topics regarding the information architectures of organizations, as well as the data entry, extraction, and analysis processes. When learning about NoSQL, we briefly reviewed big data, which are essentially data sets that contain too high of a volume, velocity, and variety for traditional relational databases to process. Natural language processing is extremely beneficial in data analytics, facilitating data processing. NLP is specifically useful in managing big data (Ridgers). Through NLP, users have access to more data than ever before. Computers have the ability to find, analyze, and summarize data from tens of thousands of articles and webpages (Ridgers). NLP is crucial in the research process, as data analysts can program these computers to fully understand and effectively answer users' questions. The important relationship between natural language processing and processing big data demonstrates the strong connection between NLP and the course material.

References

IBM Cloud Processing. Big data analytics. (n.d.). Retrieved from https://www.ibm.com/analytics/hadoop/big-data-analytics

IBM Cloud Processing. (2020, July 2). What is natural language processing? Retrieved from https://www.ibm.com/cloud/learn/natural-language-processing

Ridgers, M. (n.d.). How natural language processing is changing data analytics. Retrieved from https://www.kdnuggets.com/2020/08/natural-language-processing-changing-data-analytics.html

Wu, J. (2019, July 10). AI, cyberbullying, and social media. Retrieved from https://towardsdatascience.com/ai-cyberbullying-and-social-media-321d91d5b4ba

Zhao, R., Zhou, A., & Mao, K. (2016, January). Automatic detection of cyberbullying on social networks based on bullying features. Retrieved from https://www.researchgate.net/profile/Rui-Zhao-33/publication/310768726_Automatic_Detection_of_Cyberbullying_on_Social_Networks_based_on_Bullying_Features/links/5c6cbd1392851c1c9dee9d9d/Automatic-Detection-of-Cyberbullying-on-Social-Networks-based-on-Bullying-Features.pdf

Zhou, V. (2019, December 11). A simple explanation of the Bag-of-Words model. Retrieved from https://towardsdatascience.com/a-simple-explanation-of-the-bag-of-words-model-b88fc4f4971