William Diedrick

CS 467 – Machine Learning Project

Wikipedia Vandalism Detection

This topic includes the use of machine learning to detect vandalism on Wikipedia servers. Wikipedia is an open source project so obtaining the code and resources necessary to complete this project is possible. The project targets specific Wikipedia pages to remove vandalism that may have been placed on it. By utilizing the data Wikipedia stores and machine learning algorithms, vandalism can correctly be identified and removed. This project provides an opportunity to help a company improve their current vandalism identification through the use of an open source project.

I will utilize a Random Forest learning algorithm to locate specific cases of vandalism on the servers. The first step of this project is to email the Wikipedia moderators to obtain the code necessary to access the Wikipedia data. I will develop a machine learning algorithm based around test cases that can be utilized for multiple web pages. Using a Random Forest learning algorithm I can hopefully identify a significant amount of vandalism that occurs and remove it. This kind of algorithm relies on the identification of specific words and analyzing them to see if they are candidates for vandalism.

Citations

Adler, Thomas. “Combining natural language, metadata, and reputation features.” *Wikipedia vandalism detection*, Springer-Verlag, dl.acm.org/citation.cfm?id=1964776.

Gotze, Paul. “Advanced Vandalism Detection on Wikipedia.” 3 Sept. 2014. http://www.uni-weimar.de/medien/webis/teaching/theses/goetze\_2014.pdf