

Suitability of various machine learning approaches for recognition of antisocial behaviour on social networks

Kristína Machová¹, Tomáš Tomčík¹

¹Department of Cybernetics and Artificial Intelligence, Technical University of Košice, Slovakia.

Abstract

Nowadays, social networks allow web users to express publicly agreement or disagreement with other people and express freely their opinions. This freedom is often abused and that is why we can see social networks that are full of offensive comments. The increase in textual data on the Internet has stimulated the emergence of new scientific fields as web mining that examine short texts in the online space and look for hate or offensive speech, and that try to analyze textual data in online space. Our paper is focused on a special type of analysis concentrated on detection of some forms of antisocial behaviour, particularly on hate speech, offensive posts, and cyberbullying recognition in the online space. The main goal of the work was to find out which of the machine learning strategies - classic, deep or ensemble - are the most effective in detecting of these forms of antisocial behaviour on social networks. We have compared models generated by the following methods: deep learning of neural networks (LSTM, and GRU), classical methods (SVM, NB, and DT), and ensemble learning (RF, AdaBoost). We have tested those methods on three datasets created from posts of various volume to find how the volume of data available for training affects the results of machine learning models. The best result on the smallest Hate Speech Dataset were achieved by ensemble learning using AdaBoost (Accuracy=0,904). On the other hand, the best result on the largest Offensive Speech Dataset was achieved by deep learning using GRU (Accuracy=0.964).

Keywords: Machine learning; deep learning; ensemble learning; social web mining; detection of antisocial behaviour.
