Chapter 6

Traditional Classifiers vs. Deep Learning for Cyberbullying Detection

ABSTRACT

In this chapter, the authors present their approach to cyberbullying detection with the use of various traditional classifiers, including a deep learning approach. Research has tackled the problem of cyberbullying detection during recent years. However, due to complexity of language used in cyberbullying, the results obtained with traditional classifiers has remained only mildly satisfying. In this chapter, the authors apply a number of traditional classifiers, used also in previous research, to obtain an objective view on to what extent each of them is suitable to the task. They also propose a novel method to automatic cyberbullying detection based on convolutional neural networks and increased feature density. The experiments performed on actual cyberbullying data showed a major advantage of the presented approach to all previous methods, including the two best performing methods so far based on SO-PMI-IR and brute-force search algorithm, presented in previous two chapters.

INTRODUCTION

Along with Ptaszynski et al.'s (2010) first attempt, there have been a number of research attempting to develop methods for automatic detection of cyberbullying (Dinakar et al., 2012; Basave et al., 2013; Ptaszynski et al., 2010, 2016). Unfortunately, even with multiple improvements, the results remained only partially satisfying. This has been caused by a multitude of language ambiguities and styles used in CB.

To improve the effectiveness of automatic cyberbullying detection, in this chapter we perform an in-depth analysis of the performance of all classifiers previously used in CB detection, and above that, propose a novel, Deep Learning approach based on Convolutional Neural Networks (CNN). Moreover, based on the analysis of the characteristics of CNN and the initial results, we propose an optimization of CNN by increasing Feature Density of training data.

The rest of the chapter is organized in the following way. We describe all applied methods, including traditional classifiers as well as methods described in previous Chapters, and the final proposed method based on Convolutional Neural Networks. We also present the ways of dataset preprocessing used in this research, which to some extent overlap with previous Chapter. We explain the evaluation settings for the experiment and present a thorough analysis of experiment results and discussion.

PROPOSED METHODS

Below we describe the details of the applied methods. Firstly, we describe basics of data preprocessing and feature extraction. Next, we shortly explain all classifiers with their settings and modification applied in the experiments, including the proposed model based on CNN.

Data Preprocessing

The sentences from the original dataset used in this (Ptaszynski et al., 2010, 2015a, 2015b, 2016; Nitta et al., 2013) were preprocessed in the following ways:

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/traditional-classifiers-vs-deep-learning-for-cyberbullying-detection/217354

Related Content

Comparing Online Personality of Americans and Chinese
Xingyun Liuand Tingshao Zhu (2016). *International Journal of Cyber Behavior, Psychology and Learning (pp. 88-98).*
www.irma-international.org/article/comparing-online-personality-of-americans-and-chinese/149173

Workplace Cyberbullying in Organizations With Criminal Intent and Subtle Means: A Very Pleasant and Safe Organization Culture Brings in a Better Work Culture in an Organization
Karthikeyan C. (2022). *Handbook of Research on Digital Violence and Discrimination Studies (pp. 568-588).*
www.irma-international.org/chapter/workplace-cyberbullying-in-organizations-with-criminal-intent-and-subtle-means/296101

Gender Effects on Managerial Communication and Work Performance
Rita S. Mano (2013). *International Journal of Cyber Behavior, Psychology and Learning (pp. 34-46).*
www.irma-international.org/article/gender-effects-managerial-communication-work/78280

Bullying, Cyberbullying, and Interventions in Schools
Elena Bianchini (2022). *Research Anthology on Combating Cyber-Aggression and Online Negativity (pp. 844-861).*
www.irma-international.org/chapter/bullying-cyberbullying-and-interventions-in-schools/301671

Community Programs: Local School Boards and Anti-Bullying Programs
(2018). *Cyber Harassment and Policy Reform in the Digital Age: Emerging Research and Opportunities (pp. 79-99).*
www.irma-international.org/chapter/community-programs/201678