

## 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](http://www.igi-global.com/chapter/traditional-classifiers-vs-deep-learning-for-cyberbullying-detection/217354)

## Related Content

---

### A Digital Forensic Analysis of Advance Fee Fraud (419 Scams)

Christine I. Ofulue (2010). *Handbook of Research on Discourse Behavior and Digital Communication: Language Structures and Social Interaction* (pp. 296-317).

[www.irma-international.org/chapter/digital-forensic-analysis-advance-fee/42787](http://www.irma-international.org/chapter/digital-forensic-analysis-advance-fee/42787)

### Cyberbullying on Chinese Social Media: Language Features and Influence Factors in Comments on Key Opinion Leaders' Posts

Jin Liu, Caiwei Wang, Yinyu Chen, Yuantao Peng, Jingyi Guo and Michael Prieler (2023). *Handbook of Research on Bullying in Media and Beyond* (pp. 116-131).

[www.irma-international.org/chapter/cyberbullying-on-chinese-social-media/309854](http://www.irma-international.org/chapter/cyberbullying-on-chinese-social-media/309854)

### Assertiveness and Anxiety Effects in Traditional and Online Interactions

Amy E. Baker and Debora Jeske (2015). *International Journal of Cyber Behavior, Psychology and Learning* (pp. 30-46).

[www.irma-international.org/article/assertiveness-and-anxiety-effects-in-traditional-and-online-interactions/134388](http://www.irma-international.org/article/assertiveness-and-anxiety-effects-in-traditional-and-online-interactions/134388)

### An Empirical Analysis of Receiver's Psychological Characteristics in eWOM Engagement

Anshu Rani and Shivaprasad H. N. (2022). *International Journal of Cyber Behavior, Psychology and Learning* (pp. 1-19).

[www.irma-international.org/article/an-empirical-analysis-of-receivers-psychological-characteristics-in-ewom-engagement/298686](http://www.irma-international.org/article/an-empirical-analysis-of-receivers-psychological-characteristics-in-ewom-engagement/298686)

### Reply Timing as Emotional Strategy in Mobile Text Communications of Japanese Young People: Focusing on Perceptual Gaps between Senders and Recipients

Yuuki Kato, Shogo Kato and Kunihiro Chida (2014). *User Behavior in Ubiquitous Online Environments* (pp. 1-18).

[www.irma-international.org/chapter/reply-timing-as-emotional-strategy-in-mobile-text-communications-of-japanese-young-people/81166](http://www.irma-international.org/chapter/reply-timing-as-emotional-strategy-in-mobile-text-communications-of-japanese-young-people/81166)