Chapter 4
Web Mining–Based Method for Cyberbullying Detection

ABSTRACT

In this chapter, the authors present a method for automatic detection of cyberbullying entries based on a Web mining technique, in particular, on an extended SO-PMI-IR method calculating relevance of new input documents with training documents. The method uses seed words from three categories to calculate semantic orientation score and then maximizes the relevance of categories. The method outperformed previously proposed Web-mining-based methods in both laboratory and real-world conditions. The developed system is deployed and tested in practice. After a year of testing, the authors notice an over 30% point drop in its performance. They hypothesize on the reasons for the drop. To regain the lost performance and sustain it in the future, the authors propose additional improvements including automatic acquisition and filtering of seed words. Experimentally selected optimal improvements regained much of the lost performance.

INTRODUCTION

Web-mining-based methods have been widely acknowledged in document classification, particularly in sentiment and affect analysis (Turney, 2002; Pang & Lee, 2008; Ptaszynski et al., 2009), even in such novel fields as machine ethics (Komuda et al., 2010). Their advantage lies mostly in their high effectiveness and, most of all, efficiency. Web-mining methods usually

require minimal training data to perform a mass-level search on available Web data, such as the Internet (Ptaszynski et al., 2009), or large scale local corpora (Komuda, Rzepka & Araki, 2013) due to minimal human effort needed to achieve satisfactory results.

Here, we present our attempt of applying a Web mining method to the detection of cyberbullying. The reason for this attempt was backed by the main goal of our research, namely, to provide help to the members of Internet Patrol, which have been carried out mostly manually. It takes much time and effort to find harmful entries (entries that contain information and expressions aimed to harm other users) in a large amount of contents appearing on countless numbers of electronic Bulletin Board (BBS) pages. Moreover, the task comes with a great psychological burden on mental health to the net-patrol members.

To solve the above problem and decrease the burden of net-patrol members, we firstly proposed a preliminary method to detect harmful entries automatically (Matsuba et al., 2011). In this method we extended the method of relevance calculation SO-PMI-IR, developed by Turney (2002) to the relevance of a document with harmful contents. With the use of a small number of seed words we were able to detect large numbers of document candidates for harmful entries with satisfying effectiveness.

The initial method was demonstrated to determine harmful entries with an accuracy of 83% on test data for which about a half contained harmful entries. However, it was not verified how well the method would perform in real life conditions, where cyberbullying entries and normal contents do not appear in equal numbers.

In the research described in this chapter, initially introduced by Nitta et al. (2013) in 2013, we firstly proposed an optimization of the original method by maximizing the relevance scores. In the method we divided the seed words into multiple categories and calculated maximal relevance values for each seed word from each category. By calculating the maximized score, representing semantic orientation of harmfulness of an Internet entry, the method detected harmful entries more effectively than in the previous attempts. Moreover, we additionally evaluated our method on data sets with real-life ratios of harmful contents to verify the usability of the method in the most realistic way.

After confirming the performance, the method was deployed and tested "in the field" for over one year. Unfortunately, after re-evaluating the method after that time we found out that the performance significantly dropped by losing over 30-percentage-points when compared to its original performance measured in 2013. In the original evaluation process, applied also in this research and previously in Matsuba et al. (2011), Nitta et al. (2013) assigned

24 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/web-mining-based-method-for-cyberbullying-detection/217352

Related Content

Feeling Connected: A Sense of Belonging and Social Presence in an Online Community of Learners
Sara Jane Archard (2014). *International Journal of Cyber Ethics in Education (pp. 16-28).*
www.irma-international.org/article/feeling-connected/123980

Research Commentary: "CyberEthics"?
Wanbil Lee (2011). *International Journal of Cyber Ethics in Education (pp. 58-59).*
www.irma-international.org/article/research-commentary-cyberethics/52100

Adolescents, Third-Person Perception, and Facebook
John Chapin (2019). *Analyzing Human Behavior in Cyberspace (pp. 28-38).*
www.irma-international.org/chapter/adolescents-third-person-perception-and-facebook/211042

Discursive Practice and the Nigerian Identity in Personal Emails
Innocent Chiluwa (2010). *Handbook of Research on Discourse Behavior and Digital Communication: Language Structures and Social Interaction (pp. 112-129).*
www.irma-international.org/chapter/discursive-practice-nigerian-identity-personal/42775

Transparent Classrooms: How the Mobile Phone is Changing Educational Settings
Carla Ganito (2011). *International Journal of Cyber Ethics in Education (pp. 59-69).*
www.irma-international.org/article/transparent-classrooms-mobile-phone-changing/56109