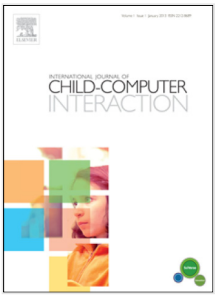
Accepted Manuscript



Sustainable cyberbullying detection with category-maximized relevance of harmful phrases and double-filtered automatic optimization

Michal Ptaszynski, Fumito Masui, Taisei Nitta, Suzuha Hatakeyama, Yasutomo Kimura, Rafal Rzepka, Kenji Araki

PII: S2212-8689(16)30054-X
DOI: http://dx.doi.org/10.1016/j.ijcci.2016.07.002
Reference: IJCCI 44

To appear in: *International Journal of Child-Computer Interaction*

Received date: 1 May 2015
Revised date: 30 June 2016
Accepted date: 27 July 2016

Please cite this article as: M. Ptaszynski, F. Masui, T. Nitta, S. Hatakeyama, Y. Kimura, R. Rzepka, K. Araki, Sustainable cyberbullying detection with category-maximized relevance of harmful phrases and double-filtered automatic optimization, *International Journal of Child-Computer Interaction* (2016), http://dx.doi.org/10.1016/j.ijcci.2016.07.002

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

*Manuscript
Click here to view linked References

ACCEPTED MANUSCRIPT

Sustainable cyberbullying detection with category-maximized relevance of harmful phrases and double-filtered automatic optimization

Michal Ptaszynski †, Fumito Masui †, Taisei Nitta †, Suzuha Hatakeyama †, Yasutomo Kimura ‡, Rafal Rzepka §, Kenji Araki§

†Department of Computer Science, Kitami Institute of Technology
‡Department of Information and Management Science, Otaru University of Commerce
§Graduate School of Information Science and Technology, Hokkaido University

Abstract

We develop a supporting solution for "cyberbullying" prevention based on recent discoveries in Artificial Intelligence and Natural Language Processing. Cyberbullying, defined as using Internet to humiliate and slander other people has become a burning problem. In Japan members of Parent-Teacher Association perform manual Web monitoring to stop cyberbullying activities. Unfortunately, reading through the whole Web manually is an uphill task. Although the complexity of cyberbullying makes it a problem unsolvable solely with the help of technology, we demonstrate that technology could make cyberbullying prevention more efficient. We propose a novel method of automatic detection of cyberbullying entries on the Internet. In the method we use seed words from three categories to calculate semantic orientation score and then maximize the relevance of categories. The proposed method outperformed baseline settings in both laboratory and real world conditions. The developed system is deployed and tested in practice. After a year of testing we notice a major over 30 percent-point-drop in its performance. We hypothesize on the reasons for the drop. To regain the lost performance and retain it in the future we propose additional improvements including automatic acquisition and filtering of seed words. Experimentally selected optimal improvements regained much of the lost performance.

Keywords: Cyberbullying, Natural language processing, Semantic orientation, Optimization

Download English Version:

https://daneshyari.com/en/article/4931296

Download Persian Version:

https://daneshyari.com/article/4931296

Daneshyari.com