Cyberbullying Detection Based on Category Relevance Maximization

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

†Department of Computer Science, Kitami Institute of Technology, 090-8507, Japan, nitta@ialab.cs.kitami-it.ac.jp,{masui,ptaszynski}@cs.kitami-it.ac.jp
‡Department of Information and Management Science, Otaru University of Commerce, 047-8501, Japan, kimura@res.otaru-uc.ac.jp
§Graduate School of Information Science and Technology, Hokkaido University, 060-0814, Japan, {kabura,araki}@media.eng.hokudai.ac.jp

1 Introduction

"Cyberbullying", or slandering other people on the Internet, leads to depression or even suicides of victims. To solve the problem, in Japan members of Parent-Teacher Association (PTA) perform Web site monitoring. Unfortunately, searching through the Web manually costs time and fatigue of PTA members. We propose a method for detecting harmful information automatically.

2 System Description

In the proposed method, we extend the semantic orientation determination method [1], to calculate the degree of relevance between a document and harmful words. We divide harmful seed-words into three categories (obscene, violent and abusive) and calculate the relevance between documents and each category. The documents with the highest overall harmfulness score are considered as malicious. The method consists of three steps: (1) *phrase extraction*, where we extract word-pairs, such as N-N ("monkey face"), N-V ("kill him") or N-Adj ("he [is] stupid"); (2) *harmful word detection and categorization*, where we use a set of 255 words divided into the three categories to extract harmful contents; and (3) *maximization of relevance score*, where we perform maximization of harmfulness score calculated for each phrase with all seed words from each of the three categories. Maximization prevents harmful words used in non harmful context to achieve high scores. The experiment showed that the proposed method performed better than the baseline [2] achieving 90% of Precission (P) at 10% Recall (R) window and keeping up high P (80-70%) at R close to 50%.

References

1. Turney, P.D.F Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, Proc. of ACL2002, pp.417–424, (2002)
2. Ptaszynski, M., Dybala, P., Matsuba, T., Masui, F., Rzepka, R., Araki, K., Momouchi, Y.: In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis, Int. J. Comp. Ling. Research 1(3), 135–154, (2010)