

An Information-centred Approach to Combat Cyberbullying Among Adolescents

Abstract

Due to the proliferation of social media within our information society, cyberbullying has emerged as a key issue within digital environments across the world, particularly among vulnerable adolescents. Cyberbullying is defined as cruel covert bullying used primarily by young people to harm others using Internet technologies such as social networking sites, chat-rooms, game forums, instant messaging, web forums or image sharing (Kowalski et. al., 2012). A recent Australian study (Sakellariou et. al., 2012) found that victimisation via the Internet was the most common form of cyberbullying with 11.5% of students reporting at least one experience of it during the school year, with junior secondary school students (Years 8 to 10) the most likely to be victimised in this manner, and text-based bullying being the most common form of cyberbullying. Cyberbullying is a group phenomenon most prevalent during the transitional ages between primary and secondary school shows an overlap in roles between 'bully' and 'victim' (Price and Dalgleish, 2012). Despite the serious emotional impacts of cyberbullying, over a quarter of victims did not seek support for various including the risk of losing access to devices or to the Internet if parents find out about it. This highlights the need for bystander education and 'upstander' action.

We are developing an online monitoring tool for schools, parents and public librarians to detect emergent episodes of cyber-bullying occurring within their local networks. To do this, we are using an expert informant approach that uses human expertise (from parents, schoolteachers, and researchers) to train a machine-learning program to automatically detect emerging cyberbullying episodes. This program can be used to train adolescents and their parents in identifying bullying episodes, and can also be automated to alert the user when patterns of emergent bullying episodes are detected. This can lead to early intervention to stop the continuation of such incidents and also to identify those who may need counselling support, be they perpetrators or victims.

This paper discusses the design and development of this open-source application that will be made freely available to schools, libraries, Internet cafes, and other providers. The application allows problem interactions within social media messages to be identified with the least intrusion using existing technologies (Xu et. al., 2012; Nahar et. al, 2013; Dadvar et. al., 2012) that have

been enhanced and 'taught' by human informants to capture traces of language and classify them based on emotion, sentiment, roles (i.e. participants role identification in cyberbullying episodes – perpetrators, victims, and enablers) and provide an understanding of the content being exchanged in cyberbullying episodes through topic modeling. The paper also discusses the design of the user-friendly visual dashboard developed for training the software with human informants.

References

- Dadvar, M., de Jong, F., Ordelman, R., Trieschnigg, R.: (2012). Improved cyberbullying detection using gender information. In: *Twelfth Dutch-Belgian Information Retrieval Workshop, DIR 2012*, 23-24 February 2012, Ghent, Belgium pp. 23-25.
- Kowalski, R.M., Limber, S.P., Agatston, P.W. (2012). *Cyberbullying in the Digital Age*. 2nd Ed., Blackwell Publishing.
- Nahar, V., Li, X., Pang, C. (2013). An effective approach for cyberbullying detection. *Communications in Information Science and Management Engineering*, Vol. 3 Iss. 5, (2013) pp 236-247.
- Price, M., Dalgleish, J. (2010). Experiences, impacts and coping strategies as described by Australian young people. *Youth Studies Australia* 29, pp. 51.
- Sakellariou, T., Carroll, A., Houghton, S. (2012). Rates of cyber victimization and bullying among male Australian primary and high school students. *School Psychology International* 33, pp. 533–549
- Xu, J.M., Jun, K.S., Zhu, X., Bellmore, A. (2012). Learning from bullying traces in social media. In: *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics pp. 656–666