Skysong Innovations
Advancing the Arizona State University Knowledge Enterprise

1475 N. Scottsdale Road, Suite 200
Scottsdale, AZ 85287-3538
Phone: 480 884 1996 Fax: 480 884 1984

Case ID:M21-049P^
Published: 10/6/2021

Inventors

Lu Cheng
Kai Shu
Siqi Wu
Yasin Silva
Deborah Hall

Contact

Shen Yan
shen.yan@skysonginnovations.com

Unsupervised Cyberbullying Detection Via Time-Informed Gaussian Mixture Model

Background Cyberbullying, defined as "aggressively intentional acts carried out by a group or an individual using electronic forms of contact, repeatedly or over time against victims who cannot easily defend themselves," has been rising at an alarming rate. Previous research has found that nearly 43% of teens in the United States have been victims of cyberbullying. In light of this, efforts aimed at automatically detecting cyberbullying—which seeks to predict whether or not human interactions within a social media session constitute cyberbullying—can play a profound role in addressing this problem. However, detecting cyberbullying on social platforms is particularly challenging given that a social media session often consists of multi-modal information, for instance, an initial post, a sequence of comments, images/videos, and other social content such as the number of likes and shares. Most existing cyberbullying detection methods are supervised and thus have two key drawbacks: (1) The data labeling process is often labor-intensive and time-consuming; (2) Current labeling guidelines may have limited utility for future instances due to evolving language and social networks. Invention Description Researchers at Arizona State University have developed a principled unsupervised learning algorithm—Unsupervised Cyberbullying Detection via Time-Informed Gaussian Mixture Model (UCD). A central feature of UCD is its incorporation of comment inter-arrival times of a social media session, which enables cyberbullying classification to account for full commenting history. UCD consists of two main components: (1) a network which learns the compact multi-modal representations of a session; and (2) a multi-task learning network which predicts whether or not a session contains bullying behaviors while modeling the temporal dynamics of all comments. Specifically, the representation learning network models social media sessions using a Hierarchical Attention Network (HAN) for textual features and a Graph Auto-Encoder (GAE) for user and network features. The multi-task learning network then takes the multi-modal representations (e.g., text, user, and social network) as input to estimate the bullying likelihood using a time-informed Gaussian Mixture Model (GMM). The two UCD components are jointly optimized to boost learning effectiveness. Potential Applications • Social media platforms • Language modeling Benefits and Advantages • Employs an unsupervised learning approach that does not require labeled data, resulting in an efficient algorithm that can maintain utility even as language and networks change over time • Performance studies using real datasets from Instagram and Vine show that UCD outperformed state-of-the-art unsupervised models and achieved results comparable to those of supervised models Related Publication: Unsupervised Cyberbullying Detection via Time-Informed Gaussian Mixture ModelFaculty Profile of Professor Huan LiuFaculty Profile of Professor Yasin SilvaFaculty Profile of Professor Deborah Hall