Asia-pacific Journal of Convergent Research Interchange
Vol.2, No.4, December 31 (2016), pp. 9-16
http://dx.doi.org/10.21742/apjcri.2016.12.02

Recognition using Cyber bullying in view of Semantic-Enhanced Minimized Auto-Encoder

Jung Hyun Kim[1)], Kirti Raj Bhatele[2)]

Abstract

As a reaction of progressively famous online networking, cyber bullying has risen as a significant issue distressing kids, teenagers and youthful grown-ups. Machine learning strategies make programmed identification of harassing messages in online networking conceivable, and this could develop a solid and safe web-based social networking environment. In this significant research zone, one basic issue is strong and discriminative numerical representation learning of instant messages. In this paper, we propose another representation learning strategy to handle this issue. Our strategy named Semantic-Enhanced Marginalized Demising Auto-Encoder (smSDA) is produced by means of semantic expansion of the famous profound learning model stacked demising auto-encoder. The semantic expansion comprises of semantic dropout clamor and sparsely limitations, where the semantic dropout commotion is planned in view of area learning and the word installing procedure. Our proposed strategy can abuse the concealed element structure of tormenting data and take in a powerful and discriminative representation of content. Exhaustive tests on two open cyber bullying corpora (Twitter and MySpace) are led, and the outcomes demonstrate that our proposed approaches beat other gauge content representation learning techniques.

Keywords : cyber bullying detection, text mining, representation learning, stacked demising auto-encoders, word embedding.

1. Introduction

Social Media, as defined in [1], is ''a group of Internet based applications that expand on the ideological and innovative establishments of Web 2.0, and that permit the creation and trade of client produced content.'' Via web-based social networking, individuals can appreciate huge data, advantageous correspondence experience et cetera. Be that as it may, online networking may have some reactions, for example, cyberbullying, which may impactsly affect the life of individuals, particularly youngsters and adolescents.

Cyberbullying can be characterized as forceful, deliberate activities performed by an

Received(August 30, 2016), Review Result(1st: September 19, 2016, 2nd: October 14, 2016), Accepted(December 10, 2016)

[1]Conversing Technology, Hoseo Graduate School of Venture, Seoul
email: hyun2@hanmail.net
[2](Corresponding Author) Department of Computer Science and Engineering, KL University
email: kirtirajbhatele8@gmail.com

ISSN: 2508-9080 APJCRI

individual or a gathering of individuals through advanced specialized techniques, for example, sending messages and posting remarks against a casualty. Not quite the same as conventional harassing that normally happens at school amid vis-à-vis correspondence, cyberbullying via web-based networking media can occur anyplace whenever. For spooks, they are allowed to offend their peers since they don't have to face somebody and can take cover behind the Internet. For casualties, they are effectively presented to provocation since every one of us, particularly youth, are always associated with Internet or web-based social networking. As reported in [2], cyberbullying exploitation rate ranges from 10% to 40%. In the United States, around 43% of youngsters were ever harassed via web-based networking media [3]. The same as traditional bullying, cyberbullying has negative, insidious and sweeping impacts on children [4-6]. The results for casualties under cyberbullying may even be awful, for example, the event of self-harmful conduct or suicides.

2. Proposed system

2.1 Related works

This work expects to take in a strong and discriminative content representation for cyberbullying discovery. Content representation and programmed cyberbullying identification are both identified with our work. In the accompanying, we quickly survey the past work in these two areas[7-17].

2.2 Text Representation Learning

In text mining, data recovery and common dialect handling, powerful numerical representation of etymological units is a key issue. The Bag-of-words (BoW) model is the most traditional content representation and the foundation of a few conditions of-expressions models including Latent Semantic Analysis (LSA) [18] and subject models [19], [20]. BoW display speaks to a report in a literary corpus utilizing a vector of genuine numbers demonstrating the event of words in the record. In spite of the fact that BoW display has ended up being proficient and powerful, the representation is frequently exceptionally inadequate. To address this issue, LSA applies Singular Value Decomposition (SVD) on the word-report framework for BoW model to determine a low-rank estimate. Each new component is a direct mix of every unique element to reduce the sparsity issue. Subject models, including Probabilistic Latent

Semantic Analysis [21] and Latent Dirichlet Allocation [20], are likewise proposed. The essential thought behind subject models is that word decision in an archive will be affected by the point of the record probabilistically. Point models attempt to characterize the era procedure of every word happened in a report[21-26].

2.3 Cyberbullying Detection

With the increasing popularity of social media in recent years, cyberbullying has developed as a significant issue afflicting children and young adults. Past investigations of cyberbullying concentrated on broad studies and its mental impacts on casualties, and were for the most part directed by social researchers and clinicians [6]. In spite of the fact that these endeavors encourage our comprehension for cyberbullying, the mental science approach in view of individual reviews is exceptionally tedious and may not be reasonable for programmed location of cyberbullying. Since

machine learning is increasing expanded notoriety as of late, the computational investigation of cyberbullying has pulled in light of a legitimate concern for specialists. A few research zones including subject location and emotional examination are firmly identified with cyberbullying identification. Attributable to their endeavors, programmed cyberbullying recognition is getting to be distinctly conceivable. In machine learning-based cyberbullying recognition, there are two issues: 1) content representation figuring out how to change every post/message into a numerical vector and 2) classifier preparing. Xu et.al exhibited a few off-the-rack NLP arrangements including BoW models, LSA and LDA for representation figuring out how to catch harassing signals in online networking [8]. As an initial work, they didn't create specific models for cyberbullying identification. Yin et.al proposed to join BoW highlights, feeling highlight and logical elements to prepare a classifier for recognizing conceivable badgering posts [10]. The presentation of the notion and relevant elements has been turned out to be compelling. Dinakar et.al [11] utilized Linear Discriminative Analysis to learn name particular components and consolidate them with BoW elements to prepare a classifier [11]. The execution of name particular elements to a great extent relies on upon the measure of preparing corpus. In addition, they need to construct a bullyspace knowledge base to boost the performance of natural language processing methods.

2.4 Existing system

Previous works on computational investigations of bullying have demonstrated that common dialect preparing and machine learning are effective apparatuses to study bullying.

Cyberbullying detection can be planned as a regulated learning issue. A classifier is initially prepared on a cyberbullying corpus named by people, and the scholarly classifier is then used to perceive a bullying message.

Yin et.al [10] proposed to consolidate BoW highlights, assessment highlights and relevant elements to prepare a bolster vector machine for online provocation recognition.

Dinakar et.al [11] used mark particular components to amplify the general elements, where the name particular elements are found out by Linear Discriminative Analysis. What's more, judgment skills information was additionally connected.

Nahar et.al [12] introduced a weighted TF-IDF plot by means of scaling harassing like elements by an element of two. Other than substance based data, Maral et.al proposed to apply clients' data, for example, sex and history messages, and setting data as additional features

2.4 Disadvantage of existing system

The first and also critical step is the numerical representation learning for text messages.

Secondly, cyberbullying is hard to describe and judge from a third view due to its intrinsic ambiguities.

Thirdly, due to protection of Internet users and privacy issues, only a small portion of messages are left on the Internet, and most bullying posts are deleted.

2.5 Proposed system

Three sorts of data including content, client demography, and social network features are regularly utilized as a part of cyberbullying location. Since the content substance is the most solid, our work here spotlights on content based cyberbullying detection.

In this paper, we explore one profound learning strategy named stacked denoising autoencoder (SDA). SDA stacks a few denoising autoencoders and links the yield of every layer as the scholarly representation. Each denoising autoencoder in SDA is prepared to recoup the info information from an undermined adaptation of it. The information is undermined by

haphazardly setting a portion of the contribution to zero, which is called dropout commotion. This denoising procedure helps the autoencoders to learn robust representation.

In addition, each autoencoder layer is expected to take in an inexorably dynamic representation of the input.

In this paper, we build up another content representation display in view of a variation of SDA: minimized stacked denoising autoencoders (mSDA), which embraces direct rather than nonlinear projection to quicken preparing and underestimates interminable commotion dissemination so as to take in more robust representations.

We use semantic data to grow mSDA and create Semantic-enhanced Marginalized Stacked Denoising Autoencoders (smSDA). The semantic data comprises of harassing words. A programmed extraction of tormenting words in view of word embeddings is proposed so that the included human work can be lessened. Amid preparing of smSDA, we endeavor to remake harassing highlights from other ordinary words by finding the dormant structure, i.e. relationship, amongst tormenting and ordinary words. The instinct behind this thought is that some tormenting messages don't contain harassing words. The correlation information discovered by smSDA recreates tormenting highlights from ordinary words, and this thus encourages identification of bullying messages without containing bullying words.

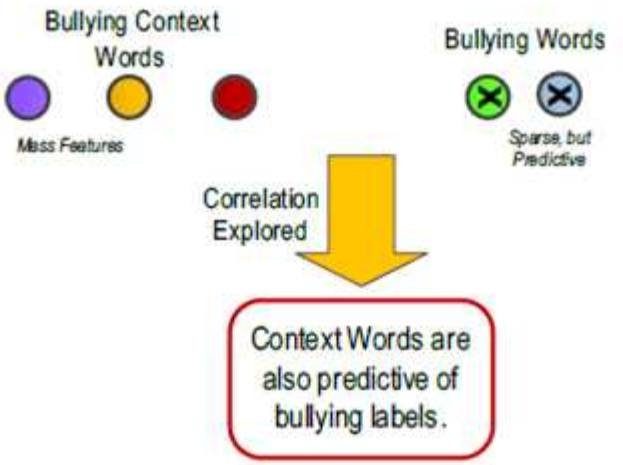
2.6 Advantages of proposed system

Our proposed Semantic-enhanced Marginalized Stacked Denoising Autoencoder can take in vigorous elements from BoW representation in a productive and compelling way. These vigorous components are found out by reproducing unique contribution from defiled (i.e., missing) ones. The new component space can enhance the execution of cyberbullying identification even with a little named preparing corpus.

Semantic data is joined into the recreation procedure by means of the outlining of semantic dropout clamors and forcing sparsity limitations on mapping lattice. In our system, top notch semantic data, i.e., harassing words, can be extricated naturally through word embeddings.

At long last, these specific alterations make the new component space more discriminative and this thusly encourages tormenting location.

Comprehensive experiments on real-data sets have verified the execution of our proposed model.



[Fig 1] System architecture

3. Conclusion

This paper addresses the text-based cyberbullying detection problem, where robust and discriminative representations of messages are critical for an effective detection system. By designing semantic dropout noise and enforcing sparsity, we have developed semantic-enhanced marginalized denoising autoencoder as a specialized representation learning model for cyberbullying detection. In addition, word embeddings have been used to automatically expand and refine bullying word lists that is initialized by domain knowledge. The performance of our approaches has been experimentally verified through two cyberbullying corpora from social medias: Twitter and MySpace. As a next step we are planning to further improve the robustness of the learned representation by considering word order in messages.

References

[1] A. M. Kaplan and M. Haenlein, Users of the world, unite! The challenges and opportunities of social media, Business horizons, **(2010)**, Vol.53, No.1, pp.59-68.

[2] R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner, Bullying in the digital age: A

critical review and metaanalysis of cyberbullying research among youth, **(2014)**.

[3] M. Ybarra, Trends in technology-based sexual and non-sexual aggression over time and linkages to nontechnology aggression, National Summit on Interpersonal Violence and Abuse Across the Lifespan: Forging a Shared Agenda, **(2010)**.

[4] B. K. Biggs, J. M. Nelson, and M. L. Sampilo, Peer relations in the anxiety‐depression link: Test of a mediation model, Anxiety, Stress, & Coping, **(2010)**, Vol.23, No.4, pp.431-447.

[5] S. R. Jimerson, S. M. Swearer, and D. L. Espelage, Handbook of bullying in schools: An international perspective. Routledge/Taylor & Francis Group, **(2010)**.

[6] G. Gini and T. Pozzoli, Association between bullying and psychosomatic problems: A meta-analysis, Pediatrics, **(2009)**, Vol.123, No.3, pp.1059-1065.

[7] A. Kontostathis, L. Edwards, and A. Leatherman, Text mining and cybercrime, Text Mining: Applications and Theory, John Wiley & Sons, Ltd, Chichester, UK, **(2010)**.

[8] J. M. Xu, K. S. Jun, X. Zhu, and A. Bellmore, Learning from bullying traces in social media, Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies. Association for Computational Linguistics, **(2012)**, pp.656-666.

[9] Q. Huang, V. K. Singh, and P. K. Atrey, Cyber bullying detection using social and textual analysis, Proceedings of the 3rd International Workshop on Socially-Aware Multimedia. ACM, **(2014)**, pp.3-6.

[10] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, Detection of harassment on web 2.0, Proceedings of the Content Analysis in the WEB, **(2009)**, Vol.2, pp.1-7.

[11] K. Dinakar, R. Reichart, and H. Lieberman, Modeling the detection of textual cyberbullying, The Social Mobile Web, **(2011)**.

[12] V. Nahar, X. Li, and C. Pang, An effective approach for cyberbullying detection, Communications in Information Science and Management Engineering, **(2012)**.

[13] M. Dadvar, F. de Jong, R. Ordelman, and R. Trieschnigg, Improved cyberbullying detection using gender information, Proceedings of the 12th -Dutch-Belgian Information Retrieval Workshop (DIR2012). Ghent, Belgium: ACM, **(2012)**.

[14] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, Improving cyberbullying detection with user context, Advances in Information Retrieval. Springer, **(2013)**, pp.693-696.

[15] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, The Journal of Machine Learning Research, **(2010)**, Vol.11, pp.3371-3408.

[16] P. Baldi, Autoencoders, unsupervised learning, and deep architectures, Unsupervised and Transfer Learning Challenges in Machine Learning, **(2012)**, Vol.7, p.43.

[17] M. Chen, Z. Xu, K. Weinberger, and F. Sha, Marginalized denoising autoencoders for domain adaptation, arXiv preprint arXiv: 1206.4683, **(2012)**.

[18] T. K. Landauer, P. W. Foltz, and D. Laham, An introduction to latent semantic analysis, Discourse processes, **(1998)**, Vol.25, No.2-3, pp.259-284.

[19] T. L. Griffiths and M. Steyvers, Finding scientific topics, Proceedings of the National academy of Sciences of the United States of America, **(2004)**, Vol.101, No.1, pp.5228-5235.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan, Latent dirichlet allocation, the Journal of machine Learning research, **(2003)**, Vol.3, pp.993-1022.

[21] T. Hofmann, Unsupervised learning by probabilistic latent semantic analysis, Machine learning, **(2001)**, Vol.42, No.1-2, pp.177-196.

[22] Y. Bengio, A. Courville, and P. Vincent, Representation learning: A review and new perspectives, Pattern Analysis and Machine Intelligence, IEEE Transactions on, **(2013)**, Vol.35, No.8, pp.1798-1828.

[23] B. L. McLaughlin, A. A. Braga, C. V. Petrie, M. H. Moore et al., Deadly Lessons:: Understanding Lethal School Violence. National Academies Press, **(2002)**.

[24] J. Juvonen and E. F. Gross, Extending the school grounds? bullying experiences in cyberspace, Journal of School health, **(2008)**, Vol.78, No.9, pp.496-505.

[25] M. Fekkes, F. I. Pijpers, A. M. Fredriks, T. Vogels, and S. P. Verloove-Vanhorick, Do bullied children get ill, or do ill children get bullied? a prospective cohort study on the relationship between bullying and health-related symptoms, Pediatrics, **(2006)**, Vol.117, No.5, pp.1568-1574.

[26] M. Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, and K. Araki, Brute force works best against bullying, Proceedings of IJCAI 2015 Joint Workshop on Constraints and Preferences for Configuration and Recommendation and Intelligent Techniques for Web Personalization, ACM, **(2015)**.