ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 7, July 2016

Deep Learning Model for Cyber Bullying Detection in the Text Representation for Social Media

Tintu Jose
UG student, Department of IT, Panimalar Institute of Technology,
Anna University, Chennai, India.
tintjose@gmail.com

***Abstract---* The social media plays major role in introducing the innovative learning and teaching using the social media for making the clear understanding of the theme in both the beginning stage of research on web 2.0 technologies, which represented by wikis and blogs The most of the research on social networking sites such as Twitter and Face book are utilized for making the analysis of clear theme of the web. In this study, one of the important issues is discriminative and robust in the text representation of understanding in the messages. In this paper, we propose a deep representation learning mechanism to handle such issues in text representation. Our scheme, is termed as Semantic-Enhanced Marginalized De-noising Auto-Encoder (SMSDA) is enhanced from the familiar deep learning model stacked of the de-noising auto encoder. In our proposed method to detect the cyber bulling in the text representation, we include sparsity constraints and semantic dropout noise, where it improves to reconstruct the original data from the domain knowledge with the help of word embedding method. Our proposed scheme is capacity to feat the hidden feature structure of bullying data and learns a more discriminative and robust in the text representation.**

***Index Terms*- Social Media, Cyber bulling detection, Text Representation and Semantic-Enhanced Marginalized De-noising Auto-Encoder**

I. INTRODUCTION

In the research domain of machine learning, data sets are acquiring higher in both diversity and volume. The bigger and faster growing structure of the hidden data in those data sets needs both the improvement of advanced machine learning strategies and interdisciplinary studies [1]. The traditional parametric proficiencies is well known difficulties in managing huge class of artificial and natural data; therefore neural network (2) concept has been broadly utilized in various fields because of their ability to resolve the complex non linear mappings directly from the input forms. According to the Huang et al had introduced the extreme learning machine (ELM) method [1] that has obtained the greater scalability and minimum computational complexity than back propagation of gradient descent based algorithm in the neutral network. Extreme learning machine (ELM) was improved particularly for the single-hidden-layer feed-forward neural networks (SLFNs) at the starting stage, and later on it is utilized for “generalized”, which is one of the type in the single-hidden-layer feed-forward neural networks that may not be similar as neuron

The usage of the recent technologies, particularly in the social media is getting ubiquitous in students day to day life. Inexpensive or free apps are shared freely from the app store have made increase to a social media to concentrate on the culture which is figuring out the way of communication, teaching and learning [2]. Nevertheless, the rapid changes in the culture and social environment does not appears to contribute the similar changes in the schools because of the various factors such as, significant technology integration, hardware availability, rigid school networking strategies [3]. As per Tess et al (3)has resolved that empirical evidences is lagging in asserting the argument for incorporating social media as significant educational tools. Majority of the attention is provided to the industries culture, resources, professional improvement for the teachers in accepting the recent technologies for learning and teaching, it is very essential to look at the students affordances and perspectives that may determine the design, analysis, development and execution of the efficient instructional schemes [2]. This is particularly true, while utilizing social media to encourage learning due to the perceived difficulties in incorporating its emerging fluid storms and meanings into extremely structured learning environments (Lewis & Rosen) [4].

In the related study, the term " social media" has been utilized interchangeably with Web 2.0 tools and the social networking software. In this analysis, social media are determined as advanced technologies and applications that use the internet and Web 2.0 technologies and permits the users to introduce and enter in several communities through the important function such as managing, collaborating, sharing, publishing, interacting , sharing and communicating [1].

Social media can be classified into the following groups [3]:

- Social networking elements such as instant messengers (Face book, what sup, Skype and etc)
- Social sharing or publishing tools consists of blogs, Twitter, Glogster and social bookmarking or tagging tools like a Flickr, Picasa, You Tube Google spreadsheets and docs and so on
- Social and content management tools involving Edmodo or Moodle; Internet- based tools such as survey, calendars and Polls
- Gaming and virtual world based surroundings including club penguin, WEE World and Playstation

1.1 Social Media In Education

The social media plays major role in introducing the innovative learning and teaching using the social media for making the clear understanding of the theme in both the beginning stage of research on web 2.0 technologies, which represented by wikis and blogs The most of the research on social networking sites such as Twitter and Face book are utilized for making the analysis of clear theme of the web [4].Web 2.0 technologies was to enhance the student engagement, college experiences, and pedagogical practices, and has been advocating innovations and changes to remain the present with changing education market in the world [3].

In the text- related cyber bullying detection, the first and also important process step is the representing ion the numerical for the deep learning for text messages. In general, text representation based learning is broadly studied in natural language processing (NLP), text mining, and data retrieval. Bag-of-words (BoW) model is most frequently utilized model that in which each dimension represents to a term. Latent Semantic Analysis (LSA) are most familiar mole for the text representation models, that are both related on BoW models. By analyzing the text data as units into fixed-length vectors, the representation for text learning can be classify progressed for numerous language processing functions. Hence, the usage of learning representation should identity the exact meaning behind text units. While evaluating cyber bullying detection, the numerical representation for data which is available internet that messages should be discriminative and robust. As the messages available social media are frequently very precise and comprise a higher number of informal language and misspellings, robust learning representations for these data are essential to minimize their ambiguity. In some worse case , the lack of high-quality and insufficient training datasets , i.e., called as data sparsity make the problem more tough. At the beginning, data label is very intensive in laboring and consuming lot of time [6]. Later on, cyber bullying is difficult to judge and describe froma third view because of its intrinsic ambiguities. Finally, as to protection of Internet users and privacy related problem, only a particular portion of messages or data are remains in the internet, and almost bullying posts are removed. As a result, the trained classifier may not normalize well on testing messages that comprisenonactivated but it has discriminative features in it. The aim of this current study is to enhance methods that can learn robust and discriminative representations of the text learning to handle the common problems in cyber bullying detection [4].

In this paper, we analyze one deep learning mechanism named stacked de-noising auto encoder (SDA) to introduce the a new text representation paradigm based for the constant marginalized stacked de-noising auto encoders (mSDA), which includes nonlinear projection to perform training and marginalizes infinite noise distribution in order to obtain the learn more robust representations. Our scheme, is termed as Semantic-Enhanced Marginalized De-noising Auto-Encoder (SMSDA) is enhanced from the familiar deep learning model stacked of the de-noising auto encoder. In our proposed method to detect the cyber bulling in the text representation, we include sparsely constraints and semantic dropout noise, where it improves to reconstruct the original data from the domain knowledge with the help of word embedding method. Our proposed scheme is capacity to feat the hidden feature structure of bullying data and learn a more discriminative and robust in the text representation.

II. LITERATURE SURVEY

This study aims to make detailed learning about a discriminative and robust text representation for cyber bullying detection. Text representation and automatic cyber bullying detection are both associated in this work.

In this analysis, it makes study on importance and benefits of Social Networking Sites (SNS), according to the

Mason (2006) (6) has reviewed that concluded that Social Networking Sites could be utilizedas educational platforms with a efficient potential capacity of the students to cultivate critical thinking among students. As per Piriyasilpa (2010), in a her paper on the learning of new language among university students in Thailand has concluded thatFace book was indeed a encouraging supporting tool in developing the students' learning capcity and their experience. According to the English and Duncan-Howell (2008) has demonstrated the usage of social networking sites such as Twitter and Face book is regarded as the support tool for students in the business sector for undertaking teaching practicum. As per Ziegler (2007), on the other hand, illustrated that social networking tools have the capacity to change students from being passive learners to becoming intentional learners and active presence, which is defined as the most requirement for making the student-centred learning. These findings agrees with those of Cress and Kim merle(2008), Collins and Halverson (2010), Minorca and Schneider (2010), and Wodzicki, Schwämmlein and Moskaliuk (2012) who discovered that the usage of social software in higher education has given a collaborative methodology for teaching and learning and teaching, permitting increased peer interaction as well as interaction between the students and teachers.

According to Murphy, and Simonds (2007) has concluded that classroom climate can develop the teachers and students relationship through social networking sites platform. whereas, Ventura and Quero (2013) demonstrated that utilizing social networks in aiding learning and teaching in business Economics and Studies, guided the students to enhance themselves in the set of competences. McCarthy (2010) has described that Face book as an special host for a blended learning environment as it was identified to improve peer relationships as students comprehended the interactive discussions that particularly occurs in the virtual learning platform [8]. The Face book activity logs also represented the development in learner engagement in the course, particularly with an assessment process. Current proof of (Ellison, Stein field and Lampe,2007; Abidin, 2010; Ng and Wong, 2013) have represented that Face book is one of the significant process to improve communication, motivate, provides a more postive learning attitude, encourage students to learn new things, support them totake their learning tasks more usefully and ideveloping their virtual interactions via social capital.

Data mining in cross-enterprise manufacturing context is to obtain efficient data and their useful information and knowledge for involving and capability management [4] . Most of previous research work has concentrated on the data mining from analyzing the structured context database. For instance, Hui and Jha [7] has combined neural network, rule-based and case-based analyzing scheme to obtain knowledge from the database to encouragefault diagnosis and service decision.

Agard and Kusiak [8] has utilizedassociation rule-based clustering scheme for customer functional essential segmentation in the development of customer response information and product families. Therefore, nowadays, in the protype of social and cloud manufacturing, huge amount of context data from the social networking sites were collected in un-structured plain text form instead of predetermined exchanging model or framework [7] .

Text-based context data is the important knowledge source from social interaction context and knowledge discovery [8,9]. Several text mining analysis have been performed in text classification, hypothesis generation, relationship extraction, terminology extraction opinion mining and entity recognition. Two important schemes have been analyzed and reviewed in the literature for these text mining applications. The first scheme deals with two important process in the text mining. It introduce the pre-processing section for the elimination of noise in the text, and later on it further analysis on equated and cleaned form of data like as ontology . The second schemes progress with the noisy data text directly utilizing machine learning strategy which would analyze patterns and learn from the underlying text . The adopted schemes generally based on the on the huge data extraction process. Their study concentrates on the manufacturing relationship extraction from the text-based context data, and includes the second scheme to handle with the big-data and high sparsity nature of the MSIC. It is represented in the fig.1

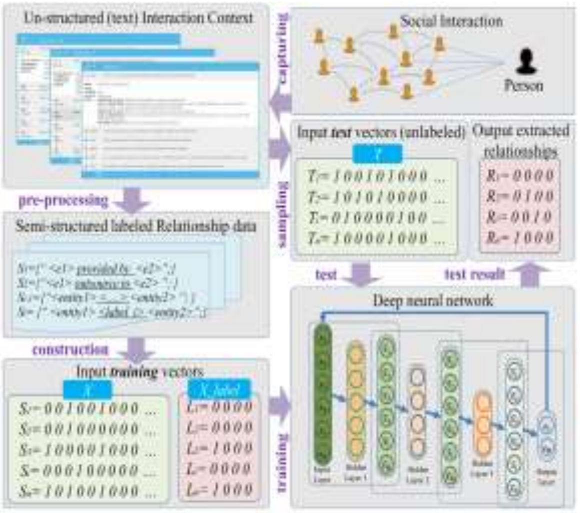
Un-structured (text) Interaction Context
capturing
Social Interaction
Person
pre-processing
Semi-structured labeled Relationship data
sampling
Input test vectors (unlabeled)
Output extracted relationships
T
T₁= 1 0 0 1 0 1 0 0 0 ...
T₂= 1 0 1 0 1 0 0 0 0 ...
T₃= 0 1 0 0 0 0 1 0 0 ...
Tₙ= 1 0 0 0 0 1 0 0 0 ...
R₁= 0 0 0 0
R₂= 0 1 0 0
R₃= 0 0 1 0
Rₙ= 1 0 0 0
test
test result
construction
Input training vectors
X
X_label
S₁= 0 0 1 0 0 1 0 0 0 ...
S₂= 0 0 1 0 0 0 0 0 0 ...
S₃= 1 0 0 0 0 1 0 0 0 ...
S₄= 0 0 0 1 0 0 0 0 0 ...
Sₙ= 1 0 1 0 0 1 0 0 0 ...
L₁= 0 0 0 0
L₂= 0 0 0 0
L₃= 1 0 0 0
L₄= 0 0 0 0
Lₙ= 1 0 0 0
training
Deep neural network
Input Layer
Hidden Layer 1
Hidden Layer 2
Hidden Layer 3
Output Layer

Fig .1 Deep learning model

Ptaszynski et.al (10) had evaluated sophisticated patterns in a brute-force way, in which it describes about the extracted data pattern and their weights are required to be determineddepend on annotated training corpus, and hence the performance would not be assured if the training corpus has a restrained with size. Apart content-based information, Maral et.al were also discussed about the employ users' data, like as gender, age and historymessages, and context data as an additionally features.

According to the Huang et.al (11)also believed that social network features to study the features for cyber bullying detection. There have shared lack of techniques in the aforementioned schemes, which is developed for text features are yet from BoW representation, which has been seriously criticized for its inherent over-sparsity and failure to capture semantic structure. Apart from these mechanisms, our proposed study can analyze and laearn more robust features by reconstructing the original information from corrupted information and create semantic corruption noise and sparsity mapping matrix to develop the feature structure which are prognosticative of the existence of bullying so that the learned representation can be robust and discriminative.

III. PROPOSED SYSTEM

3.1 SEMANTIC-ENHANCED MARGINALIZED STACKED DE-NOISING AUTO-ENCODER

At first, we describe about the notations utilized in this paper. Let $D = \{w_1, ..., w_d\}$ be the represented as the dictionary extending all the words previously in the text corpus. We demonstrate each message utilizing a BoW vector $X \in □^{d}$. After, the whole corpus can be indicated as a matrix: $X = [x_1, ..., x_n] \in □^{d \times n}$, where n is the number of posts which is available.We describe detailed manner about the proposed semantic marginalized stacked de-noisingauto-encoder

.

3.1 Auto Encoder

An auto-encoder is a from the neutral network with an input layer, hidden layer and output layer, it is represented in the below diagram.

$$E = \frac{1}{2}(\bar{x} - f(Vf(Wx + b^{(ih)}) + b^{(ho)})) \qquad (1)$$

Where

x -represnts orginal data,

f -represnts activation funations

W-denotes matrix from input to hidden layer

V- denotes matrix from hidden to ouput layer

$b^{(ih)}$ -bias vector from input to hidden layer

$b^{(ho)}$ - bias vector from hidden layer

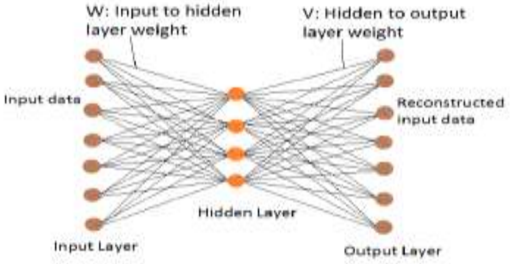

Fig.2 graphical representation of auto encoder

The weights W and V should be selected in a such manner $V = W^T$, where T represents matrix transpose. The dependency V and W is referred as tied weights. The tied weights is utilized to keep the autoencoder from the identify function of the learning representation.

3.2 Marginalized De-noising Auto-encoder:

In this paradigm,de-noising auto-encoder tries to rebuilds original information or data by utilizing linear projection for the corrupted data. The projection matrix for the learning is represented as

$$W = \arg\underset{W}{win}\frac{1}{2n}tr[(X - W\tilde{x})^T(X - W\tilde{x})] \quad (2)$$

Whereas $W \in \square^{d\times d}$, $\tilde{X} = [\tilde{x}_1, ..., \tilde{x}_n]$is presents the corrupted version of the data or information which is described in the above equation , it is also termed as ordinary least square problem with closed-form solution.

In this proposed system, this corruption value would be termed as the marginalized value over the noise distribution [17]. When there is huge corruptions of the data is taking in the scenario, we consider in using de-noising auto-encoder, which is more robust transformation can be learned for the text representation. Hence, the best preference is applying the infinite versions of corrupted information. In the case, the corrupted data corpus is corrupted infinite times, then it uses the matrix such as P and Q are meet to their corresponding expectation in the process.

$$W = E[P]E[Q]^{-1} \quad (3)$$

The expected results for the matrices would be calculated by the noise distribution. In [11], dropout noise is included to corrupt data sets by forming a feature to zero valuewith the help of the probability distribution p. We consider the scatter matrix for the reconstructing the original data sets is represented as S = XXT, the expectedmatrices can be estimated as

$$E[P]_{i,j} = (1-p)S_{i,j} \quad (4)$$

Whereas i and j represented as the indices of features sets. It can be demonstrated that it is very significant to estimate Wby using the marginalizing dropout noise in de-noising auto-encoder. Later on , it applies the mapping weights W, which is equated with non linear squashing function for performing the tangent function, can be used to obtain the final results of the marginalized de-noising auto-encoder.

3.3Semantic Enhancement for MSDA

The major benefits of corrupting the original input data in the Marginalized semantic de-noising auto-encoder mSDAcan be described by using the feature co-occurrence statistics in the computing the matrix value. The cooccurrencedata is present to obatin a robust feature and discriminative representation for text mining under an unsupervised learning mechanism, andthis also supports various text based feature learning schemes such as Latent Semantic Analysis and topic models. It is demonstrated in the in Figure 3. (a), a de-noising autoencoderis trained to rebuild these damaged featuresvalues from the rest uncorrupted ones. Hence, the learnedmapping matrix W is proposed to collect the correlation betweenthese damaged features and other features of the text mining. It is demonstrated thatthe learning based text representation is discriminative and robust and can be considered asa greater level concept feature of the data as the correlation data is constant to domain-specific vocabularies. It is discussed in the below section how to improve the mSDA for cyber bullying detection.The most of the correlation consists of the include sparse mapping constraints and semantic droupout noise.

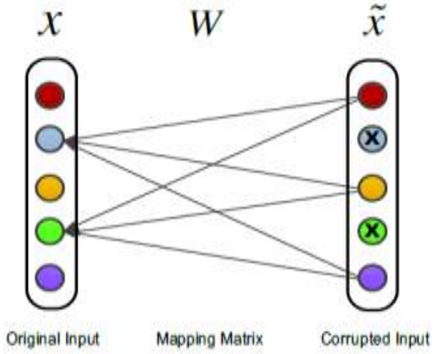

Fig 3.(a) Illustration of SMSDA

As demonstrated in Figure 3. (b), the correlation between normal features sets can connect other typical words to detect bullying labels. Assuming a normal but intuitive instance, ”Leave him alone, he is just a chink”1, which is actually referred a bullying message. Moreover, the classifier set will distinguish the weight of the discriminative word ”chink” to zero, if the small sized training corpus does not cover it.

Our proposed SMSDA canhandlesuch issues by using text based learning a robust feature representation, which is a greater standard of concept in text representation. In the learned based text representation, the word ”chink” are rebuild by context words co occurring with the particular word called (”chink”) and the context words may be distributed by other bullying words present in training corpus. Hence, the correlation is relabeled by this auto-encoder structure ensures the subsequent classifier to distinguish the discriminative word and enhance the performance in the classification.

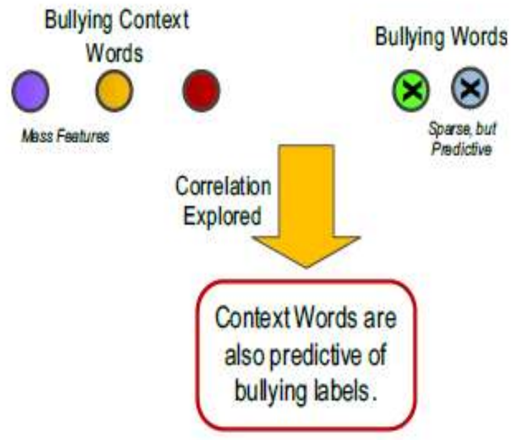

Fig 3.(b) cross symbol denotes that its corresponding feature is corrupted

3.4 SMSDA For Cyber bullying Detection

We propose the Semantic-enhanced Marginalized Stacked De-noising Auto-encoder (SMSDA). In this part, we analyze how to leverage it for cyber bullying detection. SMSDA provides discriminative and robust text based learning representations The learned numerical representations can then given into Support Vector Machine (SVM). In the new space, because of the collected semantic data and feature correlation in the Support Vector Machine, even trained datasets in a small size of training corpus, is capable to obtain a better result son testing documents. Dependon prior knowledge, we develop a pre-defined bullying wordlist and equate it with the original word vocabulary of the whole corpus X. The words occurring in both the vocabulary and the bullying wordlist are chosen as insulting seeds. The insulting seeds are then extended and redefined as automatically through word embeddings scheme, which describes the bullying features ZB for layer one.

Fig.4 Word Cloud Visualization of the Bullying Features

Furth more, Bullying Word Matching (BWM), as a easily and intuitive scheme of applying semantic information, provides the worst performance. In older BWM, the existence of bullying words are determined as rules sets for word classification. It represents that only an elaborated usage of such type of bullying words rather of a using normal one can support cyber bullying detection. In Deep learning methods including MSDA and SMSDA actually obtains the best performances while comparing to the other standard mechanisms. This trend is special prominent in F1 measure due to the cyber bullying detection issues are class-imbalance. The huge development on F1 score confirm the performance of the proposed methods, moreover, it uses the sparsity constraints and semantic dropoutnoiseon mapping matrix, in which the significantly used as the training datasets can be obtained. This developments causes to a outperform in enhancing on cyber bullying detection and the detailed analysis has been described in the following table.

Table 1: Word Reconstruction

	Reconstruct words	
Bullying words	mSDA	SMSDA
Bitch	Shut Friend tell	HTTPLINK Fuck up shut
Shift	Some Big With lol	Abuse This shit Shift lol big

It is demonstrated in the table 1, that how it reconstructs the very approximate than the existing methods. This represents that SMSDA can learn more thewords' and its correlations which may be the effectively reconstructs of bullyingsemantics, and hence the learned text presentation are more discrimative and robust features boostthe effective performance on cyber bullying detection.

IV CONCLUSION

This paper addresses the text-based cyber bullying 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 de-noising auto encoder as a specialized representation learning model for cyber bullying detection.

In addition, word embeddings have been used to automatically expand and refine bullying word lists that is initialized by domain knowledge. 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 thechallenges and opportunities of social media," Business horizons,vol. 53, no. 1, pp. 59–68, 2010.

[2] R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R.Lattanner, "Bullying in the digital age: A critical review and metaanalysisof cyber bullying research among youth." 2014.

[3] M. Ybarra, "Trends in technology-based sexual and non-sexualaggression 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 inthe anxiety–depression link: Test of a mediation model," Anxiety,Stress, & Coping, vol. 23, no. 4, pp. 431–447, 2010.

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

[6] G. Gini and T. Pozzoli, "Association between bullying and psychosomaticproblems: A meta-analysis," Pediatrics, vol. 123, no. 3,pp. 1059–1065, 2009.

[7] A. Kontostathis, L. Edwards, and A. Leatherman, "Text miningand 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 frombullying traces in social media," in Proceedings of the 2012 conferenceof the North American chapter of the association for computational linguistics: Human language technologies. Association for ComputationalLinguistics, 2012, pp. 656–666.

[9] Q. Huang, V. K. Singh, and P. K. Atrey, "Cyber bullying detectionusing social and textual analysis," in Proceedings of the 3rd Internation InternationalWorkshop 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, vol. 2, pp. 1–7, 2009.

[11] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detection of textual cyber bullying." in The Social Mobile Web, 2011.

[12] V. Nahar, X. Li, and C. Pang, "An effective approach for cyber bullying detection," Communications in Information Science and Management Engineering, 2012.

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