ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

A Highly Enhanced Deep Learning Method to Deal Cyber browbeat

Latha R[1], Praicy Joseph[2], Archana D[3], Nanditha C M[4], Jayasheela B N[5]

8th semester B. E student, Dept. Of Information Science and Engineering, Jnanavikas Institute of Technology, Bidadi, Karnataka, India[1,2,3,4]

Asst. Professor, Dept. of Information Science and Engineering, Jnanavikas Institute of Technology, Bidadi, Karnataka, India[5]

ABSTRACT: As we can see the drastic evolution of social media, often cyberbrowbeat has also become a major issue affecting the children, adolescents and young adults. Machine learning techniques help in automatic recognition of bullying messages or posts in possible online social media, and this might help in constructing a healthy and safe social network environment. In this meaningful research area, one critical issue is robust and discriminative numerical representation learning of text messages. In this paper, we propose a new representation learning method to overcome this problem. Our method named Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA) is developed via semantic extension of the popular deep learning model stacked denoising autoencoder. The semantic extension consists of semantic dropout noise and sparsity constraints, where the semantic dropout noise is designed based on domain knowledge and the term embedding technique. Our proposed method is able to venture the hidden feature structure of bullying information and learn a robust and discriminative representation of text. Comprehensive experiments on two public cyberbrowbeat corpora (Twitter and MySpace) are conducted, and the results show that our proposed approaches outperform other baseline text representation learning methods.

KEYWORDS: Cyberbrowbeat detection, machine learning techniques, stacked denoising autoencoder, text mining, representation learning

I. INTRODUCTION

The term “Cyberbrowbeat” can be defined as a rude, purposeful and offensive action performed by an individual or a group of people via online media by sending messages/posts/comments against a victim. Where bullies are free to hurt the victim, the victims are easily exposed to harassment. And this is too much nowadays as youths are constantly connected to the internet or social media. As reported in [1], Cyberbrowbeat victimization rate ranges from 10% to 40%. In the United States, approximately 43% of teenagers were bullied on social media [2].

One way to approach this problem is by automatically detecting and promptly reporting bullying messages in order to take proper measures to prevent some tragedies. The previous works on computational studies of bullying have shown that the natural language processing and machine learning are powerful tools to study bullying [3], [4]. Three kinds of information including text, user demography and social network features are most commonly used in Cyberbrowbeat detection [5]. Firstly, a classifier is trained by humans, and then the trained classifier is used to recognize a bullying message

In text-based Cyberbrowbeat detection, the first as well as the critical step is the numerical representation learning for text messages. Bag-of-words (BoW) model is one commonly used model where each dimension corresponds to a term. Latent Semantic Analysis (LSA) and topic models are other popular text representation models, where both are also based on BoW models. In cyberbrowbeat detection, as messages on social networks may often contain short forms and also usage of informal languages used with misspellings, robust representations for these messages are required to reduce their ambiguity.

One of the major issues being very challenging is data sparsity. Firstly, labeling data is labor intensive and time consuming. Secondly, Cyberbrowbeat is tough to describe and judge from a third-eye due to its intrinsic ambiguities. Thirdly, protection of internet users and privacy issues, partial messages are posted on internet and bullying messages

DOI: 10.15680/IJIRCCE.2017. 0505127

are deleted. The goal of this present study is to develop methods that can learn robust and discriminative representations to tackle the above problems in Cyberbrowbeat detection.

II. RELATED WORK

This work aims to learn a robust and discriminative text representation for cyberbrowbeat detection. Hence, text representation and automatic cyberbrowbeat detection are both related to our work.

A. Text representation

As we know, in text mining, information retrieval and natural language processing are very essential, thereby effective numerical representation of linguistic units is a key issue. The Bag-of-words (BoW) model is the most classical text representation and the cornerstone of some states-of-arts models including Latent Semantic Analysis (LSA) [6]. BoW model represents a document in a textual corpus using a vector of real numbers indicating the occurrence of words in the document. Although BoW model has proven to be efficient and effective, there is sometimes sparsity in the representation.

However, our approach has some distinct merits.

Firstly, the multi-layers and non-linearity of our model can ensure a deep learning architecture for text representation, which has been proven to be effective for learning high-level features [7]. Second, the applied dropout noise can make the learned representation more robust. Third, specific to cyberbrowbeat detection, our method employs the semantic information, including bullying words and sparsity constraint imposed on mapping matrix in each layer and this will in turn produce more distinctive representation.

B. Cyberbrowbeat Detection

Browbeat is everywhere. Now, with the advent of social media it has become an uncontrollable situation equally affecting toddlers to the young and old. Even dead people are not left offended with criticism. Many efforts to facilitate our understanding for cyberbrowbeat the psychological science approach based on personal surveys is very time-consuming and may not be suitable for automatic detection of cyberbrowbeat.

Since machine learning is gaining increased popularity in recent years, the computational study of cyberbrowbeat has attracted the interest of researchers. Several research areas including topic detection and affective analysis are closely related to cyberbrowbeat detection. Owing to their efforts, automatic cyberbrowbeat detection is becoming possible.

As an introductory work, they did not develop specialized models for cyberbrowbeat detection. Yin et.al proposed to combine Bow features, sentiment feature and contextual features to train a classifier for detecting possible harassing posts [8]. The introduction of the sentiment and contextual features has been proven to be effective.

Dinakar et.al used Linear Discriminative Analysis to learn label specific features and combine them with BoW features to train a classifier [9]. The performance of label-specific features largely depends on the size of training corpus. Huang et.al also considered social network features to learn the features for cyberbrowbeat detection. The shared deficiency among these fore mentioned approaches are constructed text features are still from BoW representation, which has been criticized for its inherent over-sparsity and failure to capture semantic structure[10], [11].

III. PROPOSED SYSTEM

In this paper, we investigate one deep learning method named Stacked Denoising Autoencoder (SDA) [12]. SDA stacks several denoising autoencoder as well concatenates the output of each layer as the learned representation. Each denoising autoencoder in SDA is trained in order to recover the input data from a corrupted version of it. The input is corrupted by randomly setting some of the input to zero, which is called as dropout noise.

Here, we develop a new text representation model based on a variant of SDA: marginalized stacked denoising autoencoder (mSDA), which adopts linear instead of nonlinear projection to accelerate training and marginalizes infinite noise distribution in order to learn more robust representations.

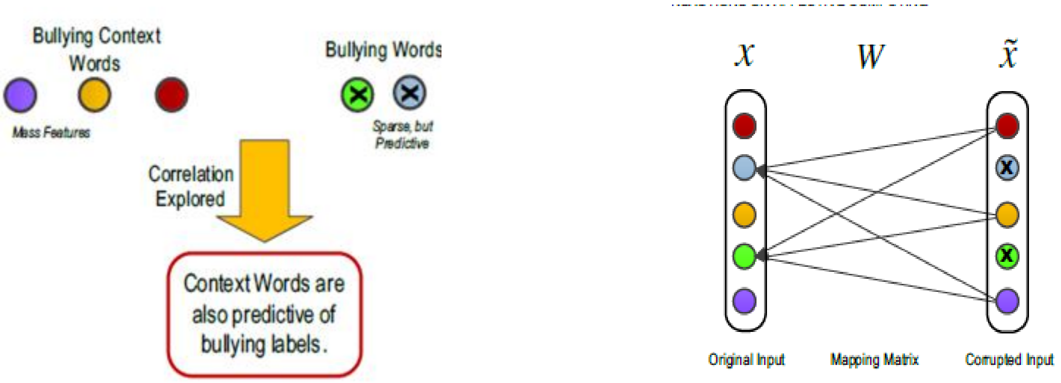
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

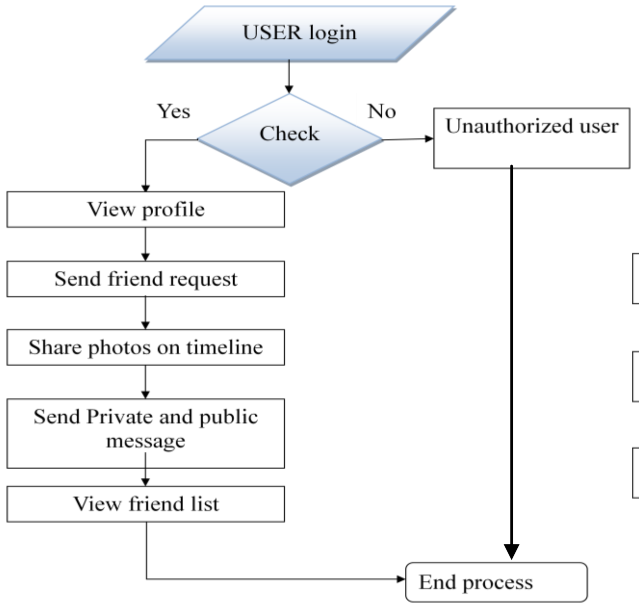


(a) (b)

Fig.1(a). System Implementation Fig.1(b). Illustration of smSDA, the cross symbol denotes that its corresponding feature is corrupted.

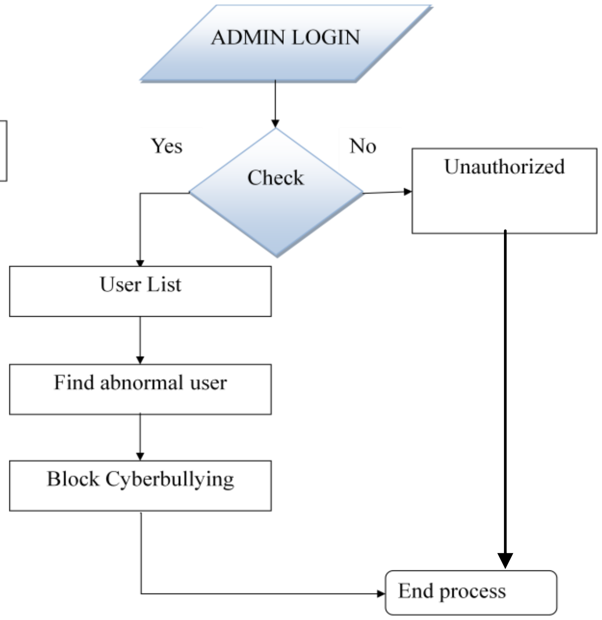
From Figure 1(a) we can realise the correlation between words and detect if it as a bullying entity and later label them accordingly. A very simple yet intuitive example, "Leave him alone, he is after all a pest". This indeed is a bullying in terms of human emotion. Hence the proposed method uses a classifier to bifurcate between a bullying and non-bullying conditions. Later in Fig1 (b) a denoising auto encoder is trained to reconstruct the removed feature values such that the learned mapping matrix W is able to capture the correlation between the removed features and the other features.

We utilize semantic information to expand mSDA and develop Semantic-enhanced Marginalized Stacked Denoising Autoencoder (smSDA). The semantic information consists of bullying words. An automatic extraction of bullying words based on word embeddings is proposed so that the involved human labour can be reduced. During training of smSDA, we attempt to reconstruct bullying features from other normal words by discovering the latent structure, i.e. correlation, between bullying and normal words. The correlation information discovered by smSDA helps to reconstruct bullying features from normal words, and this in turn facilitates detection of bullying messages without containing bullying words.



(a)

Fig 2(a). Workflow diagram for User



(b)

Fig 2(b). Workflow diagram for admin

In the Figure 2(a), we can see the user's perspective with respect to a social media where the user can view profile, send friend requests and so on. On the other hand we have the admin perspective of usage in a social media which is seen Figure 2(b).

Every user or admin, has to go through an authentication process, during which user has to give valid mail-id and also password, which will be given during the time of registration. As a user, he is given access to send/receive friend requests, send/receive public messages, view friend list, view friend's profile. As a admin, he given authority to check for the user details, find the abnormal users trying to access any account, or usage of bullying words. Admin can view all the details of any user, anytime by simply logging in.

IV. PROPOSED ALGORITHM

a. Construction of OSN

```
Procedure Construction of OSN (User and Admin)
Input: n user and admin
Output: basic OSN
            Begin
Step 1: Create the web application like a facebook with basic features.
Step 2: User should be able to register, login and send/accept friend requests.
Step 3: User can be able to post on their walls and on friend's wall too.
Step 4: Store all the user information.
            End
```

b. Construction of Bullying Features

```
Procedure Construction of Bullying Features (BoW, dict)
Input: bag of words (BOW), and dict
Output: Construction of bulling features.
    Begin
Step1: Add the bulling words to our own corpus with word embedding technique
Step2: Assign the weights to the words.
Step3: find the correlation between words.
    i)  if co-related words add into bulling features
    ii) else repeat step2.
    End
```

c. SMSDA

```
Procedure SMSDA (Corrupted words, Bullying features)
Input: Corrupted words, bullying features
Output: Reconstruction of corrupted words and block the bullying messages
    Begin
Step 1: Read corrupted words
Step 2: Find the co-relation between the corrupted words with normal words
Step 3: Re-construct the corrupted words
Step 4: Detect if the reconstructed word is bullying word or normal
Step 5: if it is bullying word
            Block
      else
            Post the message on user wall
    End
```

```
Algorithm
lch(word1, word2)
Construct T1 which is co related words to the word1
Construct T2 which is co related words to the word2
Lowest Common Subsumer(s) = argmin(length(subsumer(T1,T2)))
        = { subsumer(T1[1], T2[2]), subsumer(T1[1], T2[3]) } = { word1 }

Length( word1 ) = 12
Max Depth( n ) = 20 (from parameter)
Score = -log( length(LCS) / (2 * max_depth(LCS.pos) ) = -log( 12 / ( 2 * 20 ) )
return score;
```

V. PSEUDOCODE

```
Pseudocode for SMSDA working
SMSDA
Input: User Post p, and Dictionary d.
Output: Proper words.
    BEGIN
        Read User Post p,
        for(post:each word) then,
            if(Misspelled words) then
                Analysis the next word and previous word co-relation with dictionary
                Predict the suitable word.
            end if
        endfor
    End
```

VI. SIMULATION RESULTS

In this context, we have presented a comparison of our proposed smSDA method with six benchmark approaches in Twitter and MySpace Datasets. Our approaches, especially smSDA, gains a significant performance improvement compared to sBoW. This is because bullying features only account for a small portion of all features used. It is difficult to learn robust features for small training data by intensifying each bullying features' amplitude. Our approach aims to find the correlation between normal features and bullying features by reconstructing corrupted data so as to yield robust features. The details of the dataset are shown in Table 1.

TABLE 1
Statistical properties of two datasets

Statistics	Twitter	MySpace
Feature No.	4413	4240
Sample No.	7321	1539
Bullying Instances	2102	398

In Table 1, we analyze the statistical data of Twitter and MySpace with prominence to three features namely Feature no, Sample number and the number of bullying instances.

We also compare our methods with two stat-of-arts text representation learning methods LSA and LDA. These two methods do not produce good performance on all datasets. This may be because that both methods belong to dimensionality reduction techniques, which are performed on the document-word occurrence matrix. Although the two methods try to minimize the reconstruction error as our approach does, the optimization in LSA and LDA is conducted after dimensionality reduction. The reduced dimension is a key parameter to determine the quality of learned feature space. Here, we fix the dimension of latent space to 100.

TABLE 2
Accuracy (%), and F1 scores (%) for compared methods onn twitter and MySpace Datasets. The mean values are given, respectively. Bold face indicates Best Performance

Dataset	Measures	BWM	BoW	sBoW	LSA	LDA	mSDA	smSDA
Twitter	Accuracies	69.3	82.6	82.7	81.6	81.1	84.1	**84.9**
	F1 Scores	16.1	68.1	68.3	65.8	66.1	70.4	**71.9**
MySpace	Accuracies	34.2	80.1	80.1	77.7	77.8	87.8	**89.7**
	F1 Score	36.4	41.2	42.5	45.0	43.1	76.1	**77.6**

In Table 2, the average results, for these two datasets, based on the parameters namely, Accuracy and F1 score. It is quite clear that our approaches outperform the other approaches in these two Twitter and MySpace corpora.

Based on mSDA, our proposed smSDA utilizes semantic dropout noise and sparsity constraints on mapping matrix, in which the efficiency of training can be kept. This extension leads to a stable performance improvement on cyberbrowbeat detection.

VII. CONCLUSION AND FUTURE WORK

This paper focuses on the text-based cyberbrowbeat 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 cyberbrowbeat detection. In addition, word embeddings have been used to automatically expand and refine bullying word lists that are initialized by domain knowledge. The performance of our approaches has been experimentally verified through two cyberbrowbeat 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. R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner, “Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth”, Psychological bulletin, Vol. 140(4), pp.1073-1137, 2014.
2. 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.
3. A.Kontostathis, L. Edwards, and A. Leatherman, “Text mining and cybercrime,” Text Mining: Applications and Theory. John Wiley & Sons, Ltd, Chichester, UK, pp.1-14, 2010.
4. J.-M. Xu, K.-S. Jun, X. Zhu, and A. Bellmore, “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, 2012.
5. Q. Huang, V. K. Singh, and P. K. Atrey, “Cyber bullying detection using social and textual analysis,” in Proceedings of the 3rd International Workshop on Socially-Aware Multimedia. pp.3-6,2014
6. T. K. Landauer, P. W. Foltz, and D. Laham, “An introduction to latent semantic analysis,” Discourse processes, vol. 25, no. 2-3, pp. 259-284, 1998
7. Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” Pattern analysis and Machine Intelligence, IEEE Transitions on, vol.35, no. 8, pp.1798-1828, 2013
8. 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.
9. K. Dinakar, R. Reichart, and H. Lieberman, “Modelling the detection of textual cyberbullying.” In The Social Mobile WEB, vol. 2, issue 3, 2012.
10. T. L. Griffiths and M. Steyvers, “Finding scientific topics,” Proceedings of the National academy of Sciences of the United States of America, vol. 101, no. Suppl 1, pp.5228-5235, 2004.
11. D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” the Journal of machine Learning research, vol. 3, pp.993-1022, 2003.
12. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, “Stacked denoisingautoencoders: Learning useful representations in a deep network with a local denoising criterion,” The Journal of Machine Learning Research, vol. 11, pp. 3371–3408, 2010.