IFERP
connecting engineers... developing research

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 8 , August 2017

Automated Tormenting Recognition in light of Semantic-Enhanced Marginalized Stacked Denoisng Auto-Encoder

[1] Megha Rani Raigonda, [2] Bhavani Namdar
[1] Assistant Professor [2] Post-Graduate Student
[1][2] Department Of Studies in Computer Applications
Visvesvaraya Technological University Centre for Pg Studies, Kalaburagi

***Abstract*— Social media now a day is a major communication bridge among peoples almost around the world which helps them to share their day to day activities to the one who is in contact with them in that network; hence it acts as an intermediary among them. As social media have many excellent features for better communication, it does have the many drawbacks which may hurt the intensions of the peoples sometimes due to some weird actions of the users on social media, such as teasing or posting something which hurts the user intensions by using bullying or tormenting words. The major criterion to deal this problem is learning robust representations of texts through text mining concepts in machine learning and NLP. So to prevent this kind of activities on social networks, we are proposing a technique named Semantic-Enhancement for Marginalized Denoising Auto- Encoder(smSDA), which is an extension of the admired machine learning technique Stacked Denoising Auto-Encoder (SDA).**

***Keywords*— Tormenting Recognition, Machine Learning, Social Network, Text mining, Semantic-Enhancement**

1. INTRODUCTION

Social Media from the past years, ruling the human beings with its excellent features such as communication among peoples of different countries by sharing their feelings, day to day activities in textual, audio or video format with the help of these active networks called social network. It is providing précised and valuable set of features for its users that the users of the network are getting huge day by day; along with many benefits it is having major critical issues which may ruin the life of any human being when some abnormal activities or any crimes are performed on such networks. These kinds of activities being performed on social networks are mainly labelled as Cyber crime, which majorly affects the youths and teenagers as they are the most apparent users of social networks. From a recent report [1], it is stated that 43% of teenagers in the United States were the victims for this crime and the discrimination rate ranges from 10% to 40% relatively [2]. As similar to the normal tormenting (bullying), the cyber tormenting also hurts the user feelings and it is to be considered as a negative impact on users of the social network.

To avoid this situation, the best remedy over here is automated recognition of tormenting words being posted on social networks so that we can prevent harmful and tragic things happening through these kinds of cyber crimes. According to some previous records, natural language processing (NLP) and machine learning are the efficient tools for handling the problem of cyber tormenting [3]. Since the tormenting recognition is devised as a supervised learning problem, a classifier is used for recognition of tormenting words. For the recognition of tormenting words, some particulars are considered which includes Text, User details and the features of the social networks [4]. Fig (1) represents the general model for supervised learning in textual representations.

While recognition of tormenting related texts, exceedingly crucial phase is learning of numerical representations for each of the text message. For representation learning of text messages, the concepts such as Text mining, data reclamation and Natural Language Processing (NLP) are studied in concise.

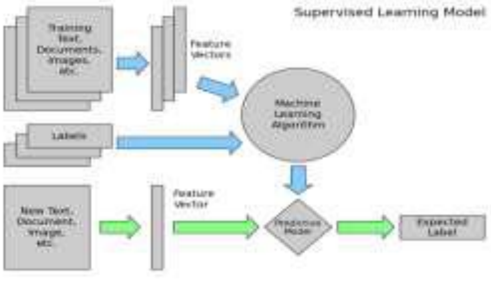
Supervised Learning Model
Training Text, Documents, Images, etc.
Feature Vectors
Labels
Machine Learning Algorithm
New Text, Document, Image, etc.
Feature Vector
Predictive Model
Expected Label

Fig (1) Supervised Learning Model

The more often used techniques for these kinds of issues are Latent Semantic Analysis (LSA) and topic models, but these couldn't achieve more robust and discriminative representation of texts as they both rely on BoW models. The robust and discriminative representation for text messages is very necessary as the messages used in social networks are almost short texts with most unceremonious speech and misused words. As the social networks are wide spreading, the problems associated with those and the solutions for such problems are also determined in an efficient manner. Several approaches have been already anticipated to deal with the problems happening on social networks with the help of feature learning. Yin et.al trained a support vector machine by merging BoW features, relative features and sentiment features for the recognition of online harassment [5]. By using the Linear Discriminative Analysis, Dinakar et.al operated the label précised features to extend common features [6]. Nahar et.al proposed a biased TF-IDF layout through scaling tormenting related features by aspect of two, which was based on common sense awareness [7]. Maral et.al tried to apply content based information of users like gender of user and message history of particular user till date as extra features [8]. All of these approaches had a drawback of not so robust representation of the texts since they use BoW assumptions as a baseline. For more robustness one needs to apply extensive domain knowledge as the quality of text messages can't be predicted easily.

2. PROPOSED WORK

In this paper, to overcome with the existing drawbacks and to bring more robust representation we are examining one of the most efficient deep learning models known as Stacked Denoising Autoencoder (SDA) [9]. The major role of SDA is to stack numerous denoising autoencoders and to concatenate the results produced in each layer for better understanding of text representation. For recovering the input data in corrupted format, each of the denoising autoencoders in SDA are trained with some functions. By casually setting few of inputs to zero, the input is corrupted and is knows as Dropout Noise. With the help of Denoising Process, autoencoders are able to learn robust representation very effectively and all autoencoder layers must have to learn abstract representation in an increasing manner [10]. In this paper, we are building a model by using a deviation on SDA, known as marginalized SDA (mSDA) for the implementation of nonlinear projection to marginalize vast noise supply along with robust representation [11]. Semantic information is exploited for the extension of mSDA to present a developed model Semantic-Enhanced Marginalized Stacked Denoising Autoencoders (smSDA). The semantic information contains tormenting word lists. For dropping the individual work of a human being, an approach for mining of tormenting words is introduced which follows the concept of word embeddings. While preparing SVM for smSDA, we need to differentiate among tormenting words and the normal words with their features for ascertaining the association between them. This association determined by smSDA assists for reformation of tormenting features and thus ease the tormenting recognition. Let's take an example, there is a strong ascertain among the tormenting word jerk and normal word you as they always crop up mutually. Since the users of social networks always use misspelled or short words such as jerk may be written as jrk then the association among those words assists in reconstruction of tormenting word with normal word and hence tormenting words are recognized. The major role of this work is concise as follows.

- The proposed smSDA approach assists in learning the vigorous features as of the BoW illustration in a very effectual manner and while the reformation of original input from damaged ones, these vigorous features are learned. This can even progress the recognition of tormenting words.
- While the scheming of semantic dropout noise and commanding of the sparsity restraints on mapping matrix, the semantic data is integrated into the process of reformation as the tormenting words are mined involuntarily using word embeddings. Hence these focused variations bring the new aspect space further discriminative for tormenting recognition.

For learning the robust and discriminative depiction of texts for automated tormenting recognition, some related works are commenced and are explained briefly as follows.

2.1 Text Depiction Learning

A major concern in text mining, data reclamation and natural language processing (NLP) is the efficient arithmetical depiction of linguistic elements. A very conventional and foundation approach for text depiction is the BoW model which consists of Latent Semantic

Analysis (LSA) [12] and topic models [13]. Even though BoW model is demonstrated to be competent but is frequently much sparse. To deal with this crisis, LSA uses Singular Value Decomposition (SVD) along with this topic models uses Probabilistic Latent Semantic Analysis [14] and Latent Dirichlet Allocation [15] were proposed. The essential thought behind topic model is that word selection in a record will be inclined by the subject of the record probabilistically. Topic models attempt to classify the invention procedure of each word emerged in a report. As alike the methods stated, our proposed method uses the BoW depiction as the input and it is having several diverse merits such as multi-layer and non-linearity support for depiction learning, more robust depiction with dropout noise, and usage of semantic information makes discriminative depiction of the tormenting words.

2.2 Tormenting Recognition

Tormenting recognition in machine learning has two major issues: firstly, each social media messages or posts need to be transformed to the arithmetic vector which makes learning of text depiction more difficult and secondly, training a classifier to adopt this approach in all ways. The approaches which are already proposed were failed in one or the other requirement such as some were not so robust and discriminative, and others were unable to capture the semantic structure but our proposed method smSDA proves to be very efficient as it performs Dropout noise, builds sparsity mapping matrix, and more necessarily makes text depiction as discriminative.

3. MODULES DESCRIPTION

The proposed method Semantic-Enhanced Marginalized Stacked Denoising Auto-Encoder (smSDA) can be illustrated as phase by phase for acquiring the robust depiction of texts over social networks. Fig (2) represents the overall mechanism for the recognition of tormenting words. The major phases of this approach are explained as follows.

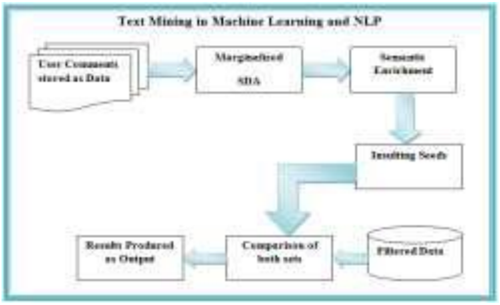
Text Mining in Machine Learning and NLP
User Comments stored as Data
Marginalized SDA
Semantic Enrichment
Insulting Seeds
Filtered Data
Comparison of both sets
Results Produced as Output

Fig (1) Applying Text Mining in Machine Learning and NLP

3.1 Marginalized Stacked Denoising Auto-Encoder

A transformed version of Stacked Denoising Auto-Encoder was proposed by Chen et.al which utilizes nonlinear projection as a substitute of linear projection for attaining a closed-form elucidation for reforming the original input as of the damaged one for achieving robust depiction [11]. In this module, mSDA reforms original input data using the damaged data with the use of linear projection matrix. Chen et.al [11] also projected the stacking structures on mSDA, where output of previous layer is taken as the input for next layer.

3.2 Semantic Enhancement for mSDA

The benefit of damaging the original input in mSDA is able to describe feature co-occurrence information. The co occurrence information is capable of achieving a robust feature depiction beneath an unsupervised learning structure, and it as well inspires other state-of-the-art text feature learning schemes such as Latent Semantic Analysis and topic models [12], [13]. An mSDA is trained to reform these isolated feature standards from the remaining uncorrupted ones. Accordingly, the learned mapping matrix is capable of capturing the association among these isolated features and other features. It signifies that the illustration is robust. Now we explain the extension of mSDA for tormenting recognition. The foremost modifications consist of semantic droupout noise and sparse mapping constraints.

3.2.1 Semantic Dropout Noise

In tormenting recognition, most tormenting posts include tormenting words such as vulgarity words and foul languages. These tormenting words are exceedingly analytical of the survival of cyber tormenting. However, a straight utilize of these tormenting features might not accomplish high-quality presentation since these words only report for a tiny segment of the entire glossary and these offensive words are only one kind of discriminative features for tormenting [5], [16]. We are able to investigate these cyber tormenting words with the help of diverse dropout noise that features comparable to tormenting words comprise a greater probability of corruption than other features. The obligatory huge possibility on tormenting words stresses the association among tormenting features and normal features. This sort of dropout noise is referred as semantic dropout noise, since semantic data is used to propose dropout structure and the proposed structure utilizes the association among tormenting features and normal features enhanced and therefore assists cyber tormenting recognition easily.

3.2.2 Sparsity Constraints

In the smSDA, the sparsity constriction is recognized with the integration of L1 regularization expression into the objective function as mentioned in the Lasso problem [17]. For the mapping matrix a regularization parameter controls the sparsity. Here we also apply an Iterated Ridge Regression technique for solving this problem since it is demonstrated as very efficient method [18].

3.3 Formation of Tormenting Feature Set

In this phase, proper selection of tormenting words with features plays a very vital role hence the formation of tormenting feature set is built upon several layers. For the preparation of first layer, professional skills are requires and the concept of word embeddings are employed. For the further layers, discriminative feature assortment is accomplished. For the first layer, we make a list of words with the negative sentimental, including terrible words and vulgar words. Then, we evaluate this word record with the BoW features of the social media entries record, and consider those associations as tormenting features. Even we inflate the record of predefined insulting words known as insulting seeds, based on word embeddings which use real valued vectors for signifying the semantics of records. For every insulting seed, associated words are extracted if they both have cosine similarities with exceeding threshold. Hence these created tormenting feature set are trained on a first layer of smSDA classifier. The subsequent layers of smSDA are constructed using Fisher score for selecting the tormenting features as the Fr (Fisher Score) is a non-varying metric which reflects discriminative strength of a feature [19].

3.4 smSDA for Tormenting Recognition

Here we proposed the Semantic enhanced Marginalized Stacked Denoising Auto-encoder (smSDA) for the Tormenting Recognition. The learned numerical illustration can then be sustained into Support Vector Machine. In the latest space, due to the detained feature association and semantic information of wordings, the SVM, even trained in a small extent of training corpus, is competent to attain a superior presentation on testing records. On the basis of predefined dropout possibilities for tormenting features and other features the matrices of each is computed if the semantic dropout is implemented. In case of unbiased dropout noise inverse of semantic dropout noise computation will be followed.

4. CONCLUSION AND FUTURE WORK

For Automated Tormenting Recognition, we have achieved the robust and discriminative representation for the text messages and posts on social networks which was the major problem now a day. With the design of semantic dropout noise and implementing sparsity we have successfully executes Semantic Enhanced Marginalized Denoising Auto-Encoder (smSDA) as a focused d the epiction learning for the automatic recognition of tormenting words. The use of word embeddings allows usual expansion and filtering of tormenting word lists which are initialized by expert domain knowledge. While talking about future works we can further enhance the robustness towards the learned depiction by taking into consideration of word orders in the messages posted by users. In extensive part, the disclosure of the abusing words will be able to make out involuntarily while any user is typing such words as of which anticipation can be attained with admiration to vanish such sorts of offenses.

REFERENCES

[1] 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.

[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] 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, 2012, pp. 656–666.

[4] 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. ACM, 2014, pp. 3–6.

[5] 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.

[6] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detection of textual cyberbullying." In The Social Mobile Web, 2011.

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

[8] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, "Improving cyberbullying detection with user context," inAdvancesin Information Retrieval. Springer, 2013, pp. 693–696.

[9] 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, vol. 11, pp. 3371–3408, 2010.

[10] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," Unsupervised and Transfer Learning Challenges in Machine Learning, Volume 7, p. 43, 2012.

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

[12] 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.

[13] 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.

[14] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," Machine learning, vol. 42, no. 1-2, pp. 177–196, 2001.

[15] D.M.Blei,A.Y.Ng, and M.I.Jordan ,"Latent dirchlent allocation," the Journal of machine Learning research, vol. 3, pp. 993–1022, 2003.

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

[17] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B (Methodological), pp. 267–288, 1996.

[18] J.FanandR.Li,"Variable selection via non concave penalized likelihood and its oracle properties," Journal of the American statistical Association, vol. 96, no. 456, pp. 1348–1360, 2001.

[19] T.H.DatandC.Guan,"Feature selection based on fisher ratio and mutual information analyses for robust brain computer interface," in Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, vol. 1. IEEE, 2007, pp. I–337.