**Cyberbullying Detection based onSemantic-Enhanced Marginalized DenoisingAuto-Encoder**

**ABSTRACT:**

As a side effect of increasingly popular social media, cyberbullying has emerged as a serious problem afflicting children,adolescents and young adults. Machine learning techniques make automatic detection of bullying messages in social media possible,and this could help to construct a healthy and safe social media environment. In this meaningful research area, one critical issue isrobust and discriminative numerical representation learning of text messages. In this paper, we propose a new representation learningmethod to tackle this problem. Our method named Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA) is developedvia semantic extension of the popular deep learning model stacked denoisingautoencoder. The semantic extension consists ofsemantic dropout noise and sparsity constraints, where the semantic dropout noise is designed based on domain knowledge and theword embedding technique. Our proposed method is able to exploit the hidden feature structure of bullying information and learn arobust and discriminative representation of text. Comprehensive experiments on two public cyberbullying corpora (Twitter andMySpace) are conducted, and the results show that our proposed approaches outperform other baseline text representation learningmethods.

**EXISTING SYSTEM:**

* Previous works on computational studies of bullying have shown that natural language processing and machine learning are powerful tools to study bullying.
* Cyberbullying detection can be formulated as a supervised learning problem. A classifier is first trained on a cyberbullying corpus labeled by humans, and the learned classifier is then used to recognize a bullying message.
* Yin et.al proposed to combine BoW features, sentimentfeatures and contextual features to train a support vectormachine for online harassment detection.
* Dinakaret.al utilized label specific features to extend the generalfeatures, where the label specific features are learned byLinear Discriminative Analysis. In addition, commonsense knowledge was also applied.
* Nahar et.al presented aweighted TF-IDF scheme via scaling bullying-like featuresby a factor of two. Besides content-based information,Maral et.al proposed to apply users’ information, such asgender and history messages, and context information asextra features

**DISADVANTAGES 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.

**PROPOSED SYSTEM:**

* Three kinds of information including text, user demography, and social network features are often used in cyberbullying detection. Since the text content is the most reliable, our work here focuses on text-based cyberbullying detection.
* In this paper, we investigate one deep learning method named stacked denoisingautoencoder (SDA). SDA stacks several denoisingautoencoders and concatenates the output of each layer as the learned representation. Each denoisingautoencoder in SDA is trained 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 dropout noise. This denoising process helps the autoencoders to learn robust representation.
* In addition, each autoencoder layer is intended to learn an increasingly abstract representation of the input.
* In this paper, we develop a new text representation model based on a variant of SDA: marginalized stacked denoisingautoencoders (mSDA), which adopts linear instead of nonlinear projection to accelerate training and marginalizes infinite noise distribution in order to learn more robust representations.
* We utilize semantic information to expand mSDA and develop Semantic-enhanced Marginalized Stacked DenoisingAutoencoders (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 labor 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 intuition behind this idea is that some bullying messages do not contain bullying 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.

**ADVANTAGES OF PROPOSED SYSTEM:**

* Our proposed Semantic-enhanced Marginalized StackedDenoisingAutoencoder is able to learn robustfeatures from BoW representation in an efficientand effective way. These robust features arelearned by reconstructing original input from corrupted(i.e., missing) ones. The new feature spacecan improve the performance of cyberbullying detectioneven with a small labeled training corpus.
* Semantic information is incorporated into the reconstructionprocess via the designing of semanticdropout noises and imposing sparsity constraintson mapping matrix. In our framework, high-qualitysemantic information, i.e., bullying words, can beextracted automatically through word embeddings.
* Finally, these specialized modifications make thenew feature space more discriminative and this inturn facilitates bullying detection.
* Comprehensive experiments on real-data sets haveverified the performance of our proposed model.

**SYSTEM ARCHITECTURE:**



**MODULES:**

* OSN System Construction Module
* Construction of Bullying Feature Set
* Cyberbullying Detection.
* Semantic-Enhanced Marginalized Denoising Auto-Encoder.

**MODULES DESCSRIPTION:**

**OSN System Construction Module**

* In the first module, we develop the Online Social Networking (OSN) system module. We build up the system with the feature of Online Social Networking. Where, this module is used for new user registrations and after registrations the users can login with their authentication.
* Where after the existing users can send messages to privately and publicly, options are built. Users can also share post with others. The user can able to search the other user profiles and public posts. In this module users can also accept and send friend requests.
* With all the basic feature of Online Social Networking System modules is build up in the initial module, to prove and evaluate our system features.

**Construction of Bullying Feature Set:**

* The bullying features play an importantrole and should be chosen properly. In the following, thesteps for constructing bullying feature set Zb are given, inwhich the first layer and the other layers are addressedseparately.
* For the first layer, expert knowledge and wordembeddings are used. For the other layers, discriminativefeature selection is conducted.
* In this module firstly, we build a list of words with negativeaffective, including swear words and dirty words. Then, wecompare the word list with the BoW features of our owncorpus, and regard the intersections as bullying features.
* Finally, the constructed bullying features are used totrain the first layer in our proposed smSDA. It includes twoparts: one is the original insulting seeds based on domainknowledge and the other is the extended bullying wordsvia word embeddings
* Observe Attentively Over A Period Of Time.

**Cyberbullying Detection:**

* In this module we propose the Semantic-enhanced Marginalized Stacked Denoising Auto-encoder (smSDA). In this module, we describe how to leverage it for cyberbullying detection. smSDA provides robust and discriminative representations The learned numerical representations canthen be fed into our system.
* In the new space, due to the captured feature correlation and semantic information, even trained in a small size of training corpus, is able to achieve a good performanceon testing documents.
* Based on word embeddings, bullying features canbe extracted automatically. In addition, the possiblelimitation of expert knowledge can be alleviated bythe use of word embedding
* BLOCK THE ACCOUNTS:
* Abnormal user.
* Cyber- Crime user.

**Semantic-Enhanced Marginalized Denoising Auto-Encoder:**

* An automatic extraction of bullying words based on word embeddings is proposed so that the involved human labor 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 intuition behind this idea is that some bullying messages do not contain bullying 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. For example, there is a strong correlation between bullying word fuck and normal word off since they often occur together.
* If bullying messages do not contain such obvious bullying features, such as fuck is often misspelled as fck, the correlation may help to reconstruct the bullying features from normal ones so that the bullying message can be detected. It should be noted that introducing dropout noise has the effects of enlarging the size of the dataset, including training data size, which helps alleviate the data sparsity problem.

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium Dual Core.
* Hard Disk : 120 GB.
* Monitor : 15’’LED
* Input Devices : Keyboard, Mouse
* Ram : 1GB.

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 7.
* Coding Language : JAVA/J2EE
* Tool : Netbeans 7.2.1
* Database : MYSQL

**REFERENCE:**

Rui Zhao and Kezhi Mao, “Cyberbullying Detection based onSemantic-Enhanced Marginalized DenoisingAuto-Encoder”, **IEEETransactions on Affective Computing, 2016**.