**Cyberbullying Detection based on Semantic-Enhanced Marginalized Denoising Auto-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 is robust and discriminative numerical representation learning of text messages. In this paper, we propose a new representation learning method to tackle 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 word embedding technique. Our proposed method is able to exploit the hidden feature structure of bullying information and learn a robust and discriminative representation of text. Comprehensive experiments on two public cyberbullying corpora (Twitter and MySpace) are conducted, and the results show that our proposed approaches outperform other baseline text representation learning methods.

**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, sentiment features and contextual features to train a support vector machine for online harassment detection.
* Dinakar et.al utilized label specific features to extend the general features, where the label specific features are learned by Linear Discriminative Analysis. In addition, common sense knowledge was also applied.
* Nahar et.al presented a weighted TF-IDF scheme via scaling bullying-like features by a factor of two. Besides content-based information, Maral et.al proposed to apply users’ information, such as gender and history messages, and context information as extra 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 denoising autoencoder (SDA). SDA stacks several denoising autoencoders and concatenates the output of each layer as the learned representation. Each denoising autoencoder 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 denoising autoencoders (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 Denoising Autoencoders (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 Stacked Denoising Autoencoder is able to learn robust features from BoW representation in an efficient and effective way. These robust features are learned by reconstructing original input from corrupted (i.e., missing) ones. The new feature space can improve the performance of cyberbullying detection even with a small labeled training corpus.
* Semantic information is incorporated into the reconstruction process via the designing of semantic dropout noises and imposing sparsity constraints on mapping matrix. In our framework, high-quality semantic information, i.e., bullying words, can be extracted automatically through word embeddings.
* Finally, these specialized modifications make the new feature space more discriminative and this in turn facilitates bullying detection.
* Comprehensive experiments on real-data sets have verified the performance of our proposed model.

**SYSTEM ARCHITECTURE:**



**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 on Semantic-Enhanced Marginalized Denoising Auto-Encoder”, **IEEE Transactions on Affective Computing, 2016**.