

# Cyberbullying Detection based on Semantic-Enhanced Marginalized DE noising Auto-Encoder

Technofist,

YES Complex, 19/3&4, 2<sup>nd</sup> Floor, Dinnur Main Road, R.T.Nagar, Bangalore-560032

Ph: 080-40969981, Website: [www.technofist.com](http://www.technofist.com). E-mail: [technofist.projects@gmail.com](mailto:technofist.projects@gmail.com)

**ABSTRACT:**

As a side effect of increasingly popular social media, cyber bullying 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 auto encoder. 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 cyber bullying corpora (Twitter and MySpace) are conducted, and the results show that our proposed approaches outperform other baseline text representation learning methods.

Technofist,

YES Complex, 19/3&4, 2<sup>nd</sup> Floor, Dinnur Main Road, R.T.Nagar, Bangalore-560032

Ph: 080-40969981, Website: [www.technofist.com](http://www.technofist.com). E-mail: [technofist.projects@gmail.com](mailto:technofist.projects@gmail.com)

## Introduction

- Three kinds of information including text, user demography, and social network features are often used in cyber bullying detection. Since the text content is the most reliable, our work here focuses on text-based cyber bullying detection. In this paper, we investigate one deep learning method named stacked denoising auto encoder (SDA). SDA stacks several denoising auto encoders and concatenates the output of each layer as the learned representation. Each denoising auto encoder 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 auto encoders to learn robust representation. In addition, each auto encoder 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 auto encoders (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.

Technofist,

YES Complex, 19/3&4, 2<sup>nd</sup> Floor, Dinnur Main Road, R.T.Nagar, Bangalore-560032

Ph: 080-40969981, Website: [www.technofist.com](http://www.technofist.com). E-mail: [technofist.projects@gmail.com](mailto:technofist.projects@gmail.com)

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.

TECHNOFIST

Technofist,

YES Complex, 19/3&4, 2<sup>nd</sup> Floor, Dinnur Main Road, R.T.Nagar, Bangalore-560032

Ph:080-40969981, Website: [www.technofist.com](http://www.technofist.com). E-mail: [technofist.projects@gmail.com](mailto:technofist.projects@gmail.com)