Chapter 39
Spam Detection on Social Media Using Semantic Convolutional Neural Network

Gauri Jain
Banasthali University, India

Manisha Sharma
Banasthali University, India

Basant Agarwal
Swami Keshvanand Institute of Technology (SKIT), India

ABSTRACT

This article describes how spam detection in the social media text is becoming increasing important because of the exponential increase in the spam volume over the network. It is challenging, especially in case of text within the limited number of characters. Effective spam detection requires more number of efficient features to be learned. In the current article, the use of a deep learning technology known as a convolutional neural network (CNN) is proposed for spam detection with an added semantic layer on the top of it. The resultant model is known as a semantic convolutional neural network (SCNN). A semantic layer is composed of training the random word vectors with the help of Word2vec to get the semantically enriched word embedding. WordNet and ConceptNet are used to find the word similar to a given word, in case it is missing in the word2vec. The architecture is evaluated on two corpora: SMS Spam dataset (UCI repository) and Twitter dataset (Tweets scrapped from public live tweets). The authors' approach outperforms the-state-of-the-art results with 98.65% accuracy on SMS spam dataset and 94.40% accuracy on Twitter dataset.

DOI: 10.4018/978-1-7998-0414-7.ch039

INTRODUCTION

Social Media (Jain & Sharma, 2016a) is the way of communication today and as a result, spammers have turned wholeheartedly to this platform and mobile networks, which has resulted in an exponential surge in the volume of spams. Spam started spreading with spam mails, also known as unsolicited bulk mail (UBE) or unsolicited commercial email (UCE) (Cranor & LaMacchia, 1998). This type of mail occupies a chunk of bandwidth and storage space.

To fight against the spams many techniques are proposed. Automatic spam detection using traditional machine learning classifiers like Naïve Bayes (Zhang, 2004), Support Vector Machine (Tresch & Luniewski, 1995), KNN (Healy, 2005), Artificial Neural Network (Jain, 1996) and Random Forest (Breiman, 2001) have already shown good results. These classifiers depend on the handcrafted features which are extracted on the basis of some pre – defined techniques like BoW, information gain (IG), chi-square, maximum redundancy maximum relevancy (nRMR) etc. These techniques are discussed by Agarwal and Mittal (2016a). Common sense knowledge with the help of ConceptNet is also used for feature selection (Agarwal & Mittal, 2016b). These methods along with the traditional classifiers were used to identify spam messages. Joe and Shim (2010) used SVM, along with a thesaurus for SMS detection. Wang (2010) used NB classifier along with content-based features to classify Twitter spam. Longzhen, An, and Longjun (2009) proposed a multi-step filtration process using various steps which include blacklisting, whitelisting, content based rough sets and KNN. Silva (2010) compared different types of ANN and concluded that some of them have high potential. Ensemble methods have proven their capability as a classifier in the field of spam detection. One of the highly efficient classifier is random forest. Mccord and Chuah (2011) integrated the user-based features with the traditional classifiers to conclude that Random Forest gave the best accuracy. A comprehensive comparison of content-based techniques is given by Almeida and Yamakami (2010), Karami and Zhou (2014), Behjat, Mustapha, Nezamabadi-pour, Sulaiman and Mustapha, (2012) and Zhang, Wang, Phillips and Ji (2014).

This paper proposes a novel architecture known as Semantic Convolutional Neural Network (SCNN) for automatic spam detection. The model uses a new technique known as deep learning (Bengio, 2009). A popular deep learning architecture known as Convolutional Neural Network (CNN) is used for classification of spams. For processing the text data, it needs to be converted into numeric vector form, also known as word vectors or word embedding. Word vectors are the set of real numbers having random initial values and these are fine – tuned during the learning process. In our research, the vectors are pre – trained using word2vec which converts the textual word into a set of real numbers. It also binds semantic meaning to the text by mapping related or similar words close to each other in multi–dimensional space. A word not present in the word2vec, is looked up in the WordNet or ConceptNet for finding a similar word. The word2vec is used to find the word vector for the similar word. The better initialization of word vectors has helped in increased performance of CNN for spam detection especially when the text size is small. The CNN architecture uses these vectors for learning the features and the classification of the text as spam or ham.

The performance comparison between SCNN and the baseline machine learning techniques: Support Vector Machine (SVM), Naïve Bayes (NB), Artificial Neural Network (ANN), k-Nearest Neighbor (KNN) and Random Forest (RF) are presented. The metrics used to measure and analyze the performance of the classifier are accuracy and F1 measure. The results show that SCNN model outperforms the baseline methods.

Related Content