COMP 90049: Project 2

Identifying trolls on Twitter by supervised machine learning

1. Introduction

With the raise of social media, people can share and discuss ideas, information, news in various forms such as blogs, videos. However, social media platforms and new technologies give the power of anonymity to the users who could consistently use trolls, write fake posts and comments in public forums. This phenomenon already threatens to millions of regular individual users, governments and organizations, and to the safety and harmonious development of the social media platform. Twitter is one of the biggest social media platform and there are more than half a billion tweets are posted every day in average by millions of users on Twitter [1]. Therefore, detecting the existence of "Twitter trolls" is an important task to maintain the integrity, security and performance of Twitter and to protect regular users.

The aim of this report is to identify troll users and classify tweet text to three labels(left trolls, right trolls and other) based on Twitter dataset and to gain knowledge by implementing KNN, Naive Bayes and Random Forest. This is an attempt to understand those supervised machine learning methods and how those methods deal with trolls.

2. Related Literature

There are a lot of related studies about detecting trolls. Kumar et al. [2] investigated TIA (the fastest algorithm) on troll detection on Slashdot Zoo to classify troll users. Mihaylov [3] made experiments to train a classifier to find opinion manipulation trolls and to find troll's comments in news community forums. Jain et al. [4] categorized twitter users on the basis of their interests using Hadoop/Mahout platform and compared K-means with fuzzy K-means algorithm. Manikandan et al. [5] applied IBK, Naïve Bayes and J48 Decision Tree to build classification model and concluded that Naïve Bayes had the best performance among those classification techniques. Sobran et al. [6] worked on Naïve Bayes on imbalanced tree and made improvements on Naïve Bayes to solve imbalanced tree problems. Santos et al. [7] presented a methodology to detect and associate fake profiles on Twitter social network and accompanied with a successful real life use case to detect and stop a cyberbullying situation in elementary school. (S. Vijayarani & M. Muthulakshmi, 2013)Vijayarani et al. [8] presented a comparative analysis of Bayes and Lazy classification algorithms and concluded that lazy classification is much more efficient than Bayes.

3. Data Set

In this report, there are two dataset are applied. One is mostXXX dataset and the

other is bestXXX dataset. For mostXXX, term frequency for the top XX terms are recorded according to document frequency. For bestXXX, term frequency for the terms with the greatest Mutual Information and Chi-Square values are recorded. XXX represents small, medium or large which is the size of the dataset. BestXXX and mostXXX dataset are both constructed of 175 randomly chosen users and a total of 223K tweets with 56194 instances. Each tweet text instances can be classified as leftTroll, rightTroll or other. Table 1 below shows the number of attributes of two dataset.

Attribute	Small	Medium	Large
Best	28	112	430
Most	13	53	203

Table 1: number of attributes of two dataset

4. Methodology

Throughout this report, IBK, Naïve Bayes and Random Forest classification algorithms are implemented in Weka. We analyze and compare the results based on two dataset. Also, the best algorithm for predicting the class of trolls will be discussed.

1). IBK

It is a K-nearest neighbors classifier. The number of nearest neighbors can be specified explicitly in the object editor or determined automatically using leave-one-out cross-validation focus to an upper limit given by the specified value[9]. Each tweet text instance is classified by a majority vote of its neighbors, with each instance being assigned to the class most common among its 2 nearest neighbors. The most common distance function used is the Euclidian distance. In this report, we choose K=2.

2). Naïve Bayes

It implements a probabilistic classifier based on conditional Bayes probabilities with strong naïve independence assumptions. There are many different distribution techniques such as Bernoulli, Multinomial used to find conditional probability, but Gaussian Distribution is used in this report. For each feature, conditional probability is calculated and then the product of all probabilities is calculated in order to compare. The big value is a determinant for the class.

3). Random Forest

Random Forests are an ensemble learning method for classification (and regression) that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees[7]. RF is a collection of CART-like trees using bagging mechanism. Each node is split according to the best among the subset of predicators in random forest.

5. Experiments

A. Evaluation

For evaluation metrics in this report, we measure the accuracy(correctly predicted label) and error rate(falsely predicted label) of three classifiers and they are shown in the following tables.

Before applying KNN, Naïve Bayes and Random Forest classifiers, we remove the tweet-id attributes since it does not have any effect on the accuracy for those three classifiers. Attributes like user-ID cannot be removed since single user-id could have multiple tweet text which will influence the classification results.

B. Results

a. mostXXX dataset:

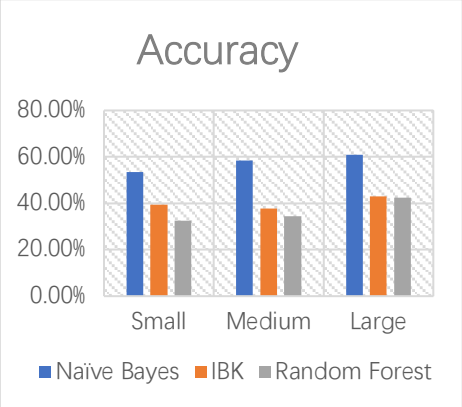

Figure 1: accuracy measure

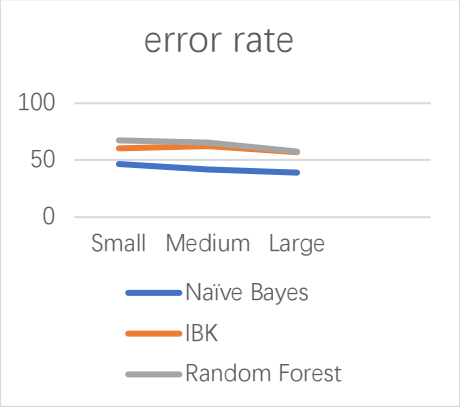

Figure 2: error rate measure

From the Figure 1 above, for fixed data size, we can see that Naïve Bayes has the highest accuracy and the Random Forest has the lowest accuracy among three classifiers. Also, the error rate of Naïve Bayes is the lowest among three classifiers no matter what training size is implemented.

In the aspect of data size, for fixed techniques, we can see that the accuracy of Naïve Bayes and Random Forest increase as the data size become larger. However, the accuracy of IBK is irrelevant with training size since the accuracy decrease from 39.3672 % to 37.625%.

b. bestXXX dataset:

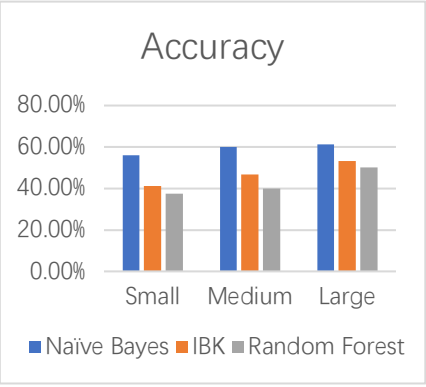

Figure 3: accuracy measure

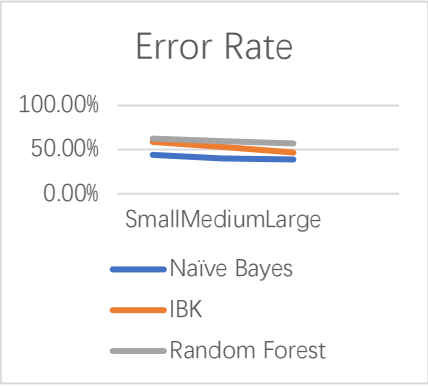

Figure 4: error rate measure

From the Figure 3 and Figure 4, we can conclude that Naïve Bayes has the highest accuracy(also lowest error rate) among three classifiers for fixed training size. Furthermore, IBK has the biggest growth with accuracy among three techniques as the training size increase. However, Naïve Bayes only increase 5.23% as the increase of training size from small to large.

C. Analysis

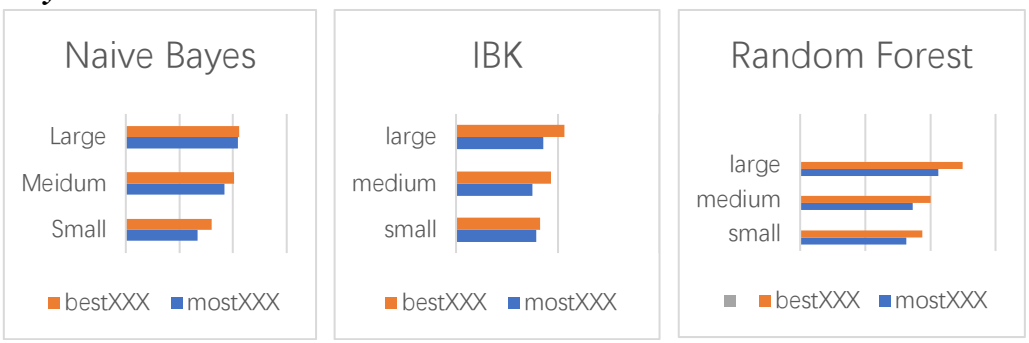

Figure 5: accuracy measurement based on different dataset

1. Best dataset

From the perspective of dataset, bestXXX is more suitable for those three classifiers since it has better fitting and higher accuracy for Naïve Bayes, IBK and Random Forest. The reason of bestXXX is a better training dataset is that bestXXX has much more attributes ranther than mostXXX. Therefore, more feature weight will be added to build models.

2. Best Classifier

In section B, the results show that Naïve Bayes has the highest accuracy no matter on bestXXX dataset or mostXXX dataset. Therefore, Naïve Bayes is the best classifier to build model to classify trolls.

3. Training size

From the discussion in section B, we can see that the larger the training size is, the higher accuracy Naïve Bayes and Random Forest have. IBK does not follow this rule, adjusting to an appropriate parameter or changing the combination of attributes may increase the accuracy under medium training size.

4. Attribute

Improvements on removing and recombining the attributes for each training dataset can be made to improve the accuracy. For example, removing 'the' attribute on the most-Medium dataset can increase the accuracy of Naïve Bayes from 58.4173% to 59.3302%. The reason is that 'the' does not have actual meaning in English grammar. It's a definite article used to specify a noun that previously considered.

6. Conclusion

In this paper, we analyzed that tweet text can help us to identify trolls on Twitter and Naïve Bayes is a better classification technique rather than KNN and Random Forest. Also, bestXXX is a better training set than mostXXX dataset since more attributes adds on the bestXXX dataset. For training size, larger training size results

in a higher accuracy generally.

In the future, the study can be extended to improve the accuracy by adjusting parameter and attributes and build model to determine every kind of trolls more precisely.

Reference

[1] Gal´an-Garc´ıaPatxi, Carlos Laorden G´omez, Jos´e Gaviria de la Puerta, SantosIgor, &. Garc´ıa BringasPablo . (2013). Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Application to a Real Case of Cyberbullying. DeustoTech Computing, University of Deusto.

[2] JainKS , & JainEeti. (9th 2014). Categorizing Twitter users on the basis of their interests using Hadoop/Mahout platform. International Conference on Industrial and Information Systems (ICIIS).

[3] KavithaB., KarthikeyanS., & chitraB. (2010). Efficient Intrusion Detection with Reduced Dimension Using Data Mining Classification Methods and Their Performance Comparison. Information Processing and Management Communications in Computer and Information Science , vol. 70, pp. 96-101.

[4] Kumar, S., Spezzano, F., & Subrahmanian, V. (August, 2014). Accurately Detecting Trolls in Slashdot Zoo via Decluttering. *IEEE/ACM Internatonal Conference* , 188-195.

[5] ManikandanP, & RamyachitraD, (July-September,2014),. Naïve Bayes Classification Technique for Analysis of Ecoli Imbalance Dataset. International Journal of Computational Intelligence and Informatics Vol. 4: No. 2.

[6] MihaylovBorisovTodor.(July, 2015). Finding Public Opinion Manipulation Trolls in Bulgarian Online News Media. St. Kliment Ohridski: Department Software Technologies, Sofia University.

[7]. MutluBusra, MutluMerve, OztoprakKasim, & DogduErdogan. (2016). Identifying Trolls and Determining Terror Awareness Level in Social Networks Using a Scalable Framework. IEEE International Conference on Big Data . Konya, Turkey 42020: KTO Karatay Unviersity.

[8] S. Vijayarani, & M. Muthulakshmi. (August, 2013). Comparative Analysis of Bayes and Lazy Classification Algorithms. International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2.8: 3118-3124.

[9] SobranMohdNur Maisarah, AhmadArfah, & IbrahimZuwairie. CLASSIFICATION OF IMBALANCED DATASET USING CONVENTIONAL NAÏVE BAYES CLASSIFIER. Universiti Teknikal Malaysia Melaka.