Sentiment Analysis of Tweets using SVM

Article *in* International Journal of Computer Applications · November 2017

DOI: 10.5120/ijca2017915758

3 authors:


Munir Ahmad
National College of Business Administration & Economics
46 PUBLICATIONS **398** CITATIONS
SEE PROFILE


Shabib Aftab
Virtual University of Pakistan
46 PUBLICATIONS **690** CITATIONS
SEE PROFILE


Iftikhar Ali
Virtual University of Pakistan
15 PUBLICATIONS **398** CITATIONS
SEE PROFILE

Sentiment Analysis of Tweets using SVM

Munir Ahmad
Department of Computer Science
Virtual University of Pakistan

Shabib Aftab
Department of Computer Science
Virtual University of Pakistan

Iftikhar Ali
Department of Computer Science
Virtual University of Pakistan

ABSTRACT
Community's view and feedback have always proved to be the most essential and valuable resource for companies and organizations. With social media being the emerging trend among everyone, it paves way for unprecedented analysis and evaluation of various aspects for which organizations had to rely on unconventional, time consuming and error prone methods earlier. This technique of analysis directly falls under the domain of "sentiment analysis". Sentiment analysis encompasses the vast field of effective classification of user generated text under defined polarities. There are several tools and algorithms available to perform sentiment detection and analysis including supervised machine learning algorithms that perform classification on the target corpus, after getting trained with training data. Lexical techniques which performs classification on the basis of dictionary based annotated corpus and Hybrid tools which are combination of machine learning and lexicon based algorithms. In this paper we have used Support Vector Machine (SVM) for sentiment analysis in Weka. SVM is one of the widely used supervised machine learning algorithms for textual polarity detection. To analyze the performance of SVM, two pre classified datasets of tweets are used and for comparative analysis, three measures are used: Precision, Recall and F-Measure. Results are shown in the form of tables and graphs.

Keywords
Polarity Detection, Sentiment Analysis, Opinion Mining, Data Mining, Data Classification, Machine Learning, Support Vector Machine, SVM

1. INTRODUCTION
Need for effective and efficient text mining tools and techniques is increasing now days due to the staggering amount of textual data. This data is increasing day by day due to social networking websites (Facebook and Twitter etc). The organizations can get unlimited benefits from mining the sentiments and polarity of this massive amount of information and reviews. With the implementation of sentiment analysis, organizations can take effective measures in order to maintain and improve their place in the market by assessing which products or services require improvement, from which price allocations the majority is unsatisfied with and what type of new features the community wants etc. Mostly three techniques have been discussed in the literature for sentiment analysis which are Lexicon based, Machine learning based, and their Hybrid [1], [9], [10], [18]. Lexicon based approach comprises of a predefined dictionary which includes weightage of words and their sentiment orientation to determine the sentiment inclination of textual data. It effectively classifies text using its set dictionary as explained by [11]. Some well-known lexicon based tools are SentiStrength 3.0, SentiWrodNet, WordNet, Linguistic Inquiry Word Count (LIWC), Affective Norms for English Words (ANEW) and SenticNet as discussed in [12]. Now if we talk about the supervised versions of machine learning techniques then it is necessary to mention here that they need a training dataset to get themselves trained for the real input. In this technique, some of the dataset with the pre identified output class is given to the algorithm in order to make the rules and then the real input data (test data) is given. Some well-known machine learning techniques include Maximum Entropy, Stochastic Gradient Descent (SGD), Random Forest (RF), SailAil Sentiment Analyzer (SASA), Multi-Layer Perceptron (MLP), Naïve Bayes (NB), Multinomial Naïve Bayes (MNB) and Support Vector Machine (SVM) as discussed in detail by [13]. A hybrid platform is one which combines both the techniques elaborated above. It uses the lexicon classification through a predefined dictionary and classifies that data using machine learning methods. Most commonly used hybrid techniques include pSenti [14], SAIL [15], NILC_USP [16] and Alchemy API [17] as discussed in detail by [18]. In this research, Support Vector Machine (SVM) is selected for sentiment analysis of two pre classified sets of tweet. SVM is formally introduced by [19] and proved to be one of the widely used supervised machine learning algorithms for the purpose of classification. It is a prevalent method which has proved to be very effective at various fronts of text categorization and has outperformed Naïve Bayes classifiers on many occasions as pointed out by [20]. For the purpose of performance evaluation of SVM Precision, Recall and F-Measure are used for both datasets.

Further organization of this paper is as follows. Section 2 describes the related work. Section 3 elaborates materials and methods. Section 4 is about classification. Section 5 discusses the results and finally section 6 concludes the paper.

2. RELATED WORK
Sentiment analysis of the textual data is one of the hot topics today. Many researchers are working on the automated techniques of extraction and analysis of huge amount of user generated data, which is available in social networking websites. In [21], the authors proposed a way to get the pre labeled data from twitter which can be used to train SVM classifier. They used the twitter hash tags to judge the polarity of tweet. To analyze the accuracy of proposed technique, a test study on the classifier was conducted which showed the result with the accuracy of 85%. In [22], the authors analyzed the performance of J48 and MLP for classification of five different datasets. Parameters to measure the accuracy in the study were TP rate, FP rate, Precision, Recall, F-measure and ROC Area. MLP performed better on each dataset. Moreover the results showed that the Neural Network also has the better learning capability and can be a good option for classification problems. The authors in [23] introduced a new technique to classify the sentiment of tweets as positive or negative. They presented and discussed the results of machine learning algorithms for twitter sentiment analysis by using distant supervision. Training data, the authors used consisted of tweets with emotions which were used as noisy labels.

According to authors, the machine learning algorithms such as Naive Bayes, Maximum Entropy and SVM when trained with emotion tweets can have accuracy more than 80%. The study also highlighted the steps used in preprocessing stage of classification for high accuracy. [24] Presented an application of Arabic sentiment analysis on twitter data. They analyzed 1000 tweets for polarity detection by using machine learning techniques, NB and SVM. In the proposed approach feature vectors were applied to machine learning classifiers for higher accuracy. The authors also pointed out some problem areas in training data such as multiple occurrences of tweets, opinion spamming and dual opinion tweets. These issues could put the question mark for the level of achieved accuracy. In [25], the authors have used three different machine learning algorithms Naïve Bayes, Decision Trees and Support Vector Machine for sentiment classification of Arabic dataset which was obtained from twitter. This research has followed a framework for Arabic tweets classification in which two special sub-tasks were performed in pre-processing, Term Frequency-Inverse Document Frequency (TF-IDF) and Arabic stemming. They have used one dataset with three algorithms and performance has been evaluated on the basis three different information retrieval metrics precision, recall, and f-measure. In [26], the authors have proposed an efficient feature vector technique by dividing the feature extraction process in two steps after the preprocessing. In first step, those features are extracted which are twitter specific and then added to feature vector. After that these features are removed from the tweets and then again the feature extraction process is done just like the case with normal text. These extracted features are also added to the feature vector. The accuracy of the proposed feature vector technique is same for Nave Bayes, SVM, Maximum Entropy and Ensemble classifiers. However this technique performed well for the domain of electronic products. [27] Proposed adaptive multiclass SVM model which works with topic adaptive sentiment classifier. The authors focused on non-text features to handle the sparsity of tweets. An iterative algorithm is proposed, consisted of three steps: optimization, unlabeled data selection and adaptive feature expansion. With 6 topic tweets, the proposed algorithm achieved promising high accuracy as compared to other well-known supervised and semi supervised classifiers. The authors in [28] focused on the polarity of hashtags as a classification feature of tweets in political domain. They proposed the rules for automatic dataset labeling based on the positive and negative hashtags, and finally proposed a method to enrich terms in the tweet by hashtag term extraction. The authors highlighted that use of positive and negative hashtags for dataset labeling and sentiment classification has accuracy of more than 95%. Moreover this hashtag feature outperforms the unigram feature when combined with Naïve Bayes, SVM or Logistic Regression algorithms, but the accuracy decreases when combined with Random Forest algorithm based on computing time to build the model. In [29], three data mining techniques are used to predict and analyze students' academic performance. The authors have used Decision tree (C4.5), Multilayer Perception and Naïve Bayes. All these techniques were applied on student's data which was collected from 2 undergraduate courses in two semesters. According to results, Naïve Bayes showed the prediction accuracy of 86% which was higher among other MLP and Decision tree. With this type of prediction it would be easy for teachers to detect those students early, who are expected to get F grade in the course. So ultimately, with the teacher's special care to those students, the academic performance can be improved.

3. MATERIALS AND METHODS

This paper aims to analyze the performance of Support Vector Machine (SVM) for polarity detection (positive, negative and neutral) of textual data. Two Pre-labeled twitter datasets are considered for this analysis. The reason of choosing the pre-labeled tweets as test data is to analyze the performance and accuracy of SVM. The output polarity for each tweet from this algorithm will be compared to the pre-labeled class and then the difference will be calculated by Weka. The performance will be measured in terms of precision, recall and f measure [1], [2], [3], [8].

3.1 Weka

In this study, we have used Weka [4], [7] for classification and performance analysis of SVM. It is one of the widely used tools to analyze the working of data mining and machine learning algorithms. Weka is developed in Java language at the University of Waikato, New Zealand. It is widely accepted due to its easy to use GUI interface. It is very famous tool due to its portability and General Public License.

3.2 Datasets

Two pre-labeled datasets of tweets are used in this research. First dataset contains the tweets about self-driving cars [5]. It contains 110 very negative, 685 slightly negative, 4245 neutral, 1444 slightly positive, 459 very positive and 213 irrelevant tweets.

Table 1. Twitter dataset for self-driving cars

Class	Tweets
Very Negative	110
Slightly Negative	685
Neutral	4245
Slightly Positive	1444
Very Positive	459
Irrelevant	213
Total	7156

Second dataset [6] contains tweets about apple products (iphone, iPod etc). This dataset consists of 1218 negative, 2162 neutral, 423 positive and 81 irrelevant tweets.

Table 2. Twitter dataset for Apple products

Class	Tweets
Negative	1218
Neutral	2162
Positive	423
Irrelevant	81
Total	3884

The dataset or input phase of our classification approach includes the downloading of relevant datasets and transformation of this data into CSV/ARFF format to use in WEKA Workbench [4], [7]. We have used simple CLI to convert text files into ARFF format by using "weka.core.converters.TextDirectoryLoader" function as shown in Figure 1.

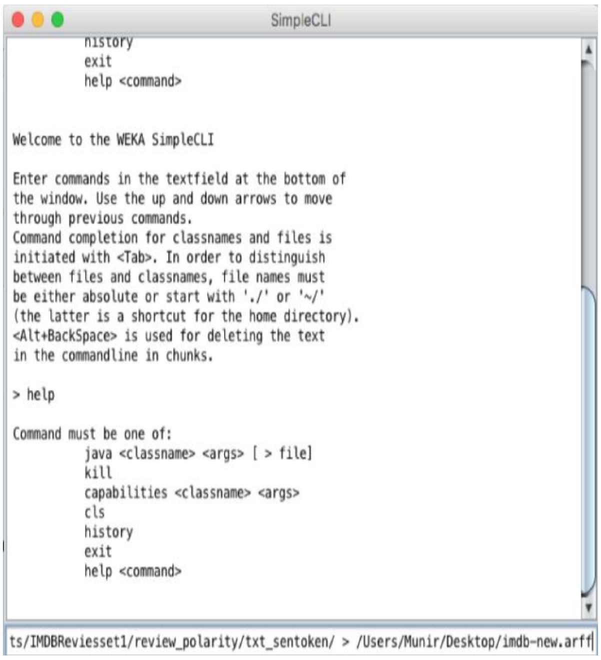

Fig 1: Simple CLI in Weka

3.3 Pre-processing

Pre processing of the input data is very important stage of classification procedure. In this stage the dataset get normalized and prepared for the classification algorithm so that the particular algorithm can run smoothly and bring effective results in minimum time [8]. According to many researches, parameters for pre-processing includes TF-IDF, Stemmer, stopwords Handler and tokenizer etc [1], [25], [30]. In this study we have used the default parameters for preprocessing as shown in Figure 2.

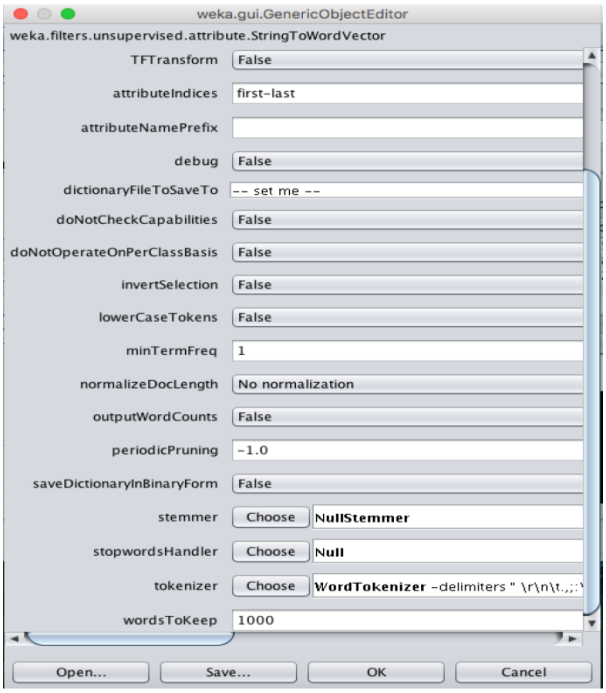

Fig 2: Default parameter selection in Weka

4. CLASSIFICATION

This is the phase in which SVM runs on the normalized data for classification and gives the results. Performance analysis of any supervised machine learning algorithm can be performed by providing the pre classified data as test data and comparing the output polarities with the pre classified polarities. We have used two datasets of pre-label tweets as input data. The results are measured in terms of precision, recall and f measure.

5. RESULTS

This section focuses on the results and comparative analysis of SVM in different measures for both datasets. For comparison, three evaluation parameters are used in this study: Precision, Recall and F Measure.

The precision can be calculated using TP and FP rate as shown below:

$$Precision = \frac{TP}{(TP + FP)}$$

TP is used for sentences, which are correctly classified, and FP is for those sentences, which are wrongly classified.

Recall can be calculated as shown below:

$$Recall = \frac{TP}{(TP + FN)}$$

FN is used for non-classified sentences and TP is for correctly classified sentences (as explained above).

F-measure can be computed as bellow:

$$\mathrm{F-measure} = \frac{Precision * Recall * 2}{(Precision + Recall)}$$

First dataset is taken from [5] and contains the tweets regarding self driving cars. According to results, the average Precision, Recall and F-Measure is 55.8%, 59.9% and 57.2% respectively.

These results are arranged in Table 3 and class wise result in each measure is shown with graph (Figure 3).

Table 3. Class wise Precision, Recall and F-Measure for First Dataset

Class	Precision	Recall	F-Measure
Very Negative	0.224	0.1	0.138
Slightly Negative	0.247	0.184	0.211
Neutral	0.708	0.841	0.769
Slightly Positive	0.428	0.305	0.356
Very Positive	0.278	0.237	0.256
Irrelevant	0.225	0.136	0.17
Average	0.558	0.599	0.572

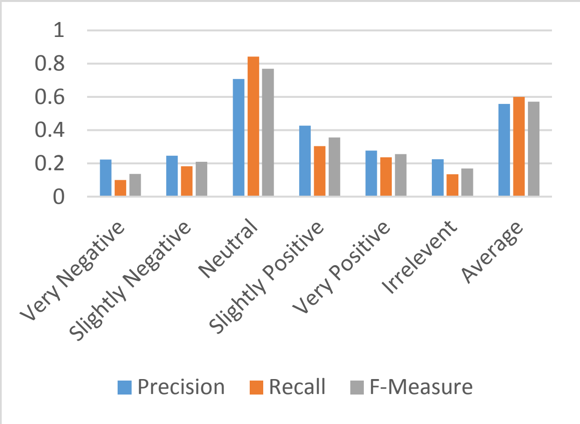

Fig 3: Twitter dataset for self-driving cars

The Score for neutral class in Precision, Recall and F-Measure is 70.8%, 84.1% and 76.9% respectively, which is higher than other classes.

Second dataset is taken from [6] and contains the tweets regarding 'apple' products. According to results, the average Precision, Recall and F-Measure is 70.2%, 71.2% and 69.9% respectively.

Complete results are arranged in Table 4 and class wise result in each measure is shown with graph (Figure 4).

Table 4. Class wise Precision, Recall and F-Measure for Second Dataset

Class	Precision	Recall	F-Measure
Negative	0.732	0.602	0.661
Neutral	0.729	0.859	0.789
Positive	0.548	0.376	0.446
Irrelative	0.318	0.173	0.224
Average	0.702	0.712	0.699

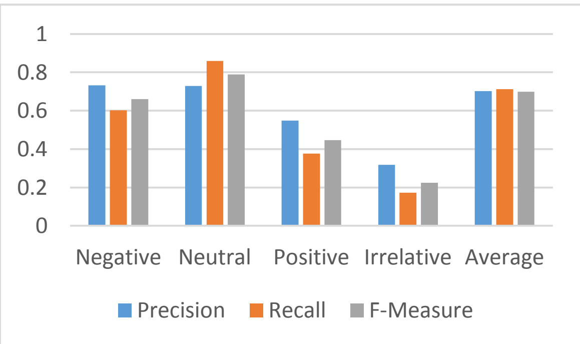

Fig 4: Twitter dataset for apple

According to results the Precision is high in Negative class (73.2%) however Recall and F-Measure both are high in Neutral class (85.9% and 78.9% respectively).

According to Weka the accuracy of SVM during sentiment classification is different in both datasets. For first dataset of self-driving cars, it is 59.91% and for second dataset of apple, it is 71.2%.

Table 5. SVM Accuracy

Datasets	Accuracy %
Self-Driving Cars	59.91%
Apple	71.2%

6. CONCLUSION

In this paper, we have analyzed the performance of Support Vector Machine (SVM) for sentiment analysis. For performance analysis of SVM, we have used two pre classified datasets of tweets, first dataset consisted of tweets regarding self driving cars and second dataset was about the apple products. Weka tool is used for performance analysis and comparison. Results are measured in terms of precision, recall and f-measure. According to results, for first dataset the average precision, recall and f-measure is 55.8%, 59.9% and 57.2% respectively. For second dataset the average Precision, Recall and F-Measure is 70.2%, 71.2% and 69.9% respectively. Complete results are shown in tabular and in graphical forms. The results clearly show the dependency of SVM performance upon input dataset. The performance dependency of SVM and other machine learning techniques should be explored further by using large and different datasets. For comparative analysis the results of this paper can be used as baseline. Moreover it should also be investigated that for classification purpose, which machine learning algorithm performs better on which type of dataset and what might be the reasons? This can lead the researchers to the improved versions of machine learning algorithms for classification purpose.

7. REFERENCES

[1] Ahmad, M., & Aftab, S. (2017). Analyzing the Performance of SVM for Polarity Detection with Different Datasets. International Journal of Modern Education and Computer Science (IJMECS), 9(10), 29-36.

[2] Sharma, A., & Dey, S. (2013, October). A boosted SVM based sentiment analysis approach for online opinionated text. In Proceedings of the 2013 Research in Adaptive and Convergent Systems (pp. 28-34). ACM.

[3] Singh, V. K., Piryani, R., Uddin, A., & Waila, P. (2013, January). Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches. In Knowledge and Smart Technology (KST), 2013 5th International Conference on (pp. 122-127). IEEE.

[4] Holmes, G., Donkin, A., & Witten, I. H. (1994, December). Weka: A machine learning workbench. In Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on (pp. 357-361). IEEE.

[5] Crowdflower.com. (2017). [online] Available at: https://www.crowdflower.com/wp-content/uploads/2016/03/Twitter-sentiment-self-drive-DFE.csv [Accessed 15 Aug. 2017].

[6] Crowdflower.com. (2017). [online] Available at: https://www.crowdflower.com/wp-content/uploads/2016/03/ Apple-Twitter-Sentiment-DFE.csv [Accessed 15 Aug. 2017].

[7] Weka: http://www.cs.waikato.ac.nz/~ml/weka/

[8] Zainudin, S., Jasim, D. S., & Bakar, A. A. (2016). Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction. International Journal on Advanced Science, Engineering and Information Technology, 6(6), 1148-1153.

[9] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1-135.

[10] Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. Information Processing & Management, 52(1), 5-19.

[11] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167

[12] Ahmad, M., Aftab, S., Muhammad, S. S., & Waheed, U. (2017). Tools and Techniques for Lexicon Driven Sentiment Analysis: A Review. Int. J. Multidiscip. Sci. Eng, 8(1), 17-23.

[13] Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2017). Machine Learning Techniques for Sentiment Analysis: A Review. Int. J. Multidiscip. Sci. Eng, 8(3), 27-32.

[14] Mudinas, A., Zhang, D., & Levene, M. (2012, August). Combining lexicon and learning based approaches for concept-level sentiment analysis. In Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining(p. 5). ACM.

[15] Malandrakis, N., Kazemzadeh, A., Potamianos, A., & Narayanan, S. (2013, June). SAIL: A hybrid approach to sentiment analysis. In SemEval@ NAACL-HLT (pp. 438-442).

[16] Balage Filho, P., & Pardo, T. (2013, June). NILC_USP: A Hybrid System for Sentiment Analysis in Twitter Messages. In SemEval@ NAACL-HLT (pp. 568-572).

[17] "AlchemyAPI." [Online]. Available: https://www.ibm.com/watson/alchemy-api.html.

[18] Ahmad, M., Aftab, S., Ali, I., & Hameed, N. (2017). Hybrid Tools and Techniques for Sentiment Analysis: A Review. Int. J. Multidiscip. Sci. Eng, 8(3)

[19] Cortes, C., & Vapnik, V. (1995). Support vector machine. Machine learning, 20(3), 273-297

[20] Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79-86). Association for Computational Linguistics

[21] Zgheib, W. A., & Barbar, A. M. A Study using Support Vector Machines to Classify the Sentiments of Tweets.

[22] Arora, R. (2012). Comparative analysis of classification algorithms on different datasets using WEKA. International Journal of Computer Applications, 54(13).

[23] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(2009), 12.

[24] Shoukry, A., & Rafea, A. (2012, May). Sentence-level Arabic sentiment analysis. In Collaboration Technologies and Systems (CTS), 2012 International Conference on (pp. 546-550). IEEE.

[25] Altawaier, M. M., & Tiun, S. (2016). Comparison of Machine Learning Approaches on Arabic Twitter Sentiment Analysis. International Journal on Advanced Science, Engineering and Information Technology, 6(6), 1067-1073.

[26] Neethu, M. S., & Rajasree, R. (2013, July). Sentiment analysis in twitter using machine learning techniques. In Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on (pp. 1-5). IEEE.

[27] Liu, S., Li, F., Li, F., Cheng, X., & Shen, H. (2013, October). Adaptive co-training SVM for sentiment classification on tweets. In Proceedings of the 22nd ACM international conference on Information & Knowledge Management (pp. 2079-2088). ACM.

[28] Alfina, I., Sigmawaty, D., Nurhidayati, F., & Hidayanto, A. N. (2017, February). Utilizing Hashtags for Sentiment Analysis of Tweets in The Political Domain. In Proceedings of the 9th International Conference on Machine Learning and Computing (pp. 43-47). ACM.

[29] Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. International Journal of Modern Education and Computer Science, 8(11), 36.

[30] Isa, D., Lee, L. H., Kallimani, V. P., & Rajkumar, R. (2008). Text document preprocessing with the Bayes formula for classification using the support vector machine. IEEE Transactions on Knowledge and Data engineering, 20(9), 1264-1272.