Asian Journal of Information Technology 15 (4): 738-742, 2016
ISSN: 1682-3915

Comparative Analysis of Classification Algorithms for Predicting the Advertisements on Webpages

S. Shanmuga Priya and S. Padmavathi
Department of Computer Science and Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham University, Coimbatore, India

Abstract: The fast growth of the Internet has completely changed the way people using computers. In the current scenario, people are more exposed to media and internet has led to the creation of advertisement which can reach users and it has become the ultimate for most business to enhance their profit. More and more ads are being sold on a single-impression basis as opposed to bulk purchases. Identifying whether an image belongs to advertisement or not is of interest to many internet users. This study analyses the performance of probabilistic, tree based and rule based classifier for this classification. Their performances under various conditions are summarized.

Key words: Classifier, image, advertisement, feature, rule, probability, tree, C4.5, Bayes, LAD, decision table

INTRODUCTION

The world wide web has been widely used for the web browsing and continues to grow in advancement towards semantic web. It plays an essential role in all parts of our life and its usage is increasing in an unbelievable manner. Online advertising remains the major source of revenue for most of the service providers on the web which includes search engines, social networks, video sharing websites, blogging sites, etc. Spending on digital advertising as a whole continued to grow at the different rates at different websites. As per the statistics provided by the internet, Google holds about a third of total digital (38%) advertisement revenue but Facebook shows stronger growth than the search giant every year. Facebook's revenue strength lies in display advertising which is a preferred category of digital advertisements of news companies.

Many Internet sites draw income from third-party advertisements usually in the form of images spread across the site's pages (Jushmerick, 1999). There are advantages and disadvantages of these advertisements. From the user's perspective, advertisements on the web pages are preferred based on their needs. The drawback of these advertisements is the increase in actual download time for the images in the web pages. The users, generally, do not prefer these images interfering their work.

El-Deen Ahmeda *et al*. (2015) analysed eleven data mining classification techniques which are comparatively analyzed to find the best classifier fit for consumer online shopping attitudes and behavior. A Classification model is build consisting of 5 phases. The dataset used composed of online ordering log file for three months. The ten-fold cross validation method is used for testing the accuracy of the classification techniques.

Kumar and Arora (2015) uses data mining technique to improve the sales in the departmental store by distribution of coupons among customers visiting the departmental store such that both customers and departmental stores can gain because of increased sales volume. The study was carried out on the profiles of the customers who have visited the departmental store from July 2014 to Dec 2014. From the experimental results, this showed a significant effect in improving sales and hereby achieving targets of departmental store.

Graepel *et al*. (2010) and Shatnawi and Mohamed (2012) created a Bayesian online learning algorithm, adpredictor used for CTR prediction in bing's sponsored search advertising. The study emphasis the importance of this prediction to Sponsored search advertising, impacting the user experience, profitability of advertising and search engine revenue. From the experimental study, the algorithm is superior and makes much more informative predictions than Naive Bayesian.

Kohavi (1995) evaluates the power of decision table as hypothesis space for supervised learning algorithms. The experimental results showed that in cases of discrete valued features, it outperforms the state-of-the-art

Corresponding Author: S. Shanmuga Priya, Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham University, Coimbatore, India

algorithms C4.5.When experimented with feature subset selection from decision tree algorithms, the decision table also represents all combinations of the chosen subset of feature and they might form a good hypothesis space.

In this study, the webpage containing the image is classified as an advertisement or not based on the features of the images such as the image geometry, phrases occurring in the URL, the image's URL, alternate text, the anchor text, words occurring near the anchor text, etc. Among 1558 such features available, a feature selection method using gain ratio is applied to reduce the dimensionality and given to the various classifier algorithms such as decision tree, rule based and Baye's classifier. The performance of these classifiers are analysed on the data set with size of 3279 samples (Sharma and Jain, 2013).

MATERIALS AND METHODS

Classification is a supervised machine learning which uses a set of samples with the class label to develop a model during training. This model is later used to identify the class labels of the data set during testing. In this study, classification algorithms (Nithya *et al*., 2013) that uses probabilistic, rule based, tree based models are considered for analysis. Bayes classifier which is considered as a benchmark for classification algorithms develops probabilistic model to learn from the training data set which are then used during testing. Likewise, rule based model is employed in knowledge based classifier and tree based model in hierarchical classifier. The features selected for classification plays a major role in classifier's performance. Relevance analysis is generally used to identify the best features for classification.

Tree based classifier: Tree based classifier uses divide and conquer strategy to construct a decision tree to classify the dataset into different classes. To construct an efficient tree, a greedy strategy is usually applied on splitting criteria at each level of the tree. During this process, a particular feature is chosen as the splitting criterion. The tree is grown until the termination condition is reached. The splitting criteria, the termination condition and post pruning of the decision tree vary for different algorithms. The C4.5 and LAD tree algorithms are considered in this study.

The C4.5 is widely used tree based classifier which finds the best split using gain ratio of the features (Hssina *et al*., 2014). The termination condition depends on the purity of the node which is determined by homogeneity of class labels. The impurity of the node can be assessed by various factors such as entropy, information gain, etc. For a dataset S with feature vector f and set of classes C, entropy can be calculated as in Eq. 1:

$$\text{Entropy}(P) = -\sum_{i=1}^{2} p_i \log_2 p_i \quad (1)$$

Where:

p_i = The probability of class c

In this study, the number of classes is 2. Information gain also measures the impurities in features as in Eq. 2:

$$\text{Information gain}(f) = \text{Entropy}(P) - \sum_{j=1}^{n}\left(p_j \times \text{Entropy}\left(P_j\right)\right) \quad (2)$$

where, p_j is the set of all possible values of a feature. The best split on the feature is decided by the gain ratio as in Eq. 3:

$$\text{Gain ratio}(f) = \text{Information gain}(f)/\text{splitinfo}(f) \quad (3)$$

Where:

$$\text{Splitinfo}(f) = -\sum_{j=1}^{n} p'(j/p)\log_2 p'(j/p) \quad (4)$$

and p'(j/p) is the set of values of the feature f taking the value of j. Among the features, the one having the highest gain ratio is designated to be the root of the tree and splitting is done based on the values of that feature. By repeating this process recursively on the remaining features, the tree is grown until the nodes are pure. The strength of C4.5 algorithms is attributed by post pruning which removes the irrelevant and redundant nodes of the tree.

LAD (Least Absolute Deviation) tree finds its usefulness especially in two class problems yielding a good accuracy. The tree is built starting with a single node containing the entire feature set. The best splitting criteria is chosen as the one which minimizes the sum of the squared errors. The tree is grow until the total class variance's as calculated in Eq. 5 is less than a threshold:

$$s = \frac{1}{n}\sum C \sum_{i \in c}\left(f_i - m_c\right)^2 \quad (5)$$

where, f, C, n, m_c represents the feature vector, class label, training sample size and class mean as in Eq. 6, respectively:

$$m_c = \frac{1}{n}\sum_{i \in c} f_i \quad (6)$$

Probability based classifier: Probability based classifier uses probabilistic measure to acquire the knowledge from the training dataset for predicting the most probable classes for the test dataset. These algorithms are fast, feature independent, space efficient and perform relatively well when compared with other classification algorithms. Naive Bayesian is most prominent algorithm in this category and is considered in this paper. It uses likelihood of the feature to a class and prior probability of the class to find the posterior probability as in Eq. 7. The posterior gives the probability of a class for the given feature vector. The maximum of which is used to label the class:

$$P(Cj|F) = P(Cj)\Pi_{(P(F|Cj))}, F = f1, f2, \ldots, fn \quad (7)$$

where, P(Cj) prior probability of the class j and P(cj|F) likelihood of feature to class j.

Rule based classifier: Rule based classifier generates rule set for classifying data into different classes. It creates partition such that the matching values of feature vector belong to a class. Decision table which is rule based classifier is considered in this study. This classifier creates decision table cells which are the partitions enclosing the set of samples with the same values for the features and has same class label. These partitions define the rule set for classification which are used as lookup table during testing. The class label is assigned based on the matching features from the lookup table. If exact match is not available the class that matches with the majority features is assigned.

RESULTS AND DISCUSSION

The data set consist of 3279 Instances of which 2821 belong to non-advertisement and 458 belong to advertisement. There are about 1558 features which include 3 continuous features while the rest are binary in nature. Information gain and gain ratio are calculated for the features exempting the height, width, aspect ratio which are continuous in nature and local domain feature. Out of 1554 features defined, 8 were selected based on threshold value. The midpoint of the gain ratio as in Eq. 8 is used as a threshold:

$$\text{Threshold} >= (L+S)/2 \quad (8)$$

Where:

L = The largest gain ratio
S = The smallest gain ratio

Table 1: Evaluation summary of the classifiers with 5 fold cross-validation for the entire feature set

Classifier	True Positive rate	False Positive rate	Precision (%)	Recall (%)	F measure	Accuracy (%)
Bayes	0.959	0.202	95.8	95.9	0.957	95.88
C4.5	0.963	0.176	96.3	96.3	0.962	96.34
LAD tree	0.961	0.191	96.1	96.1	0.960	96.13
Decision table	0.962	0.169	96.1	96.2	0.961	96.16

Table 2: Classifiers accuracy and time complexity with 5, 3 and 2 fold cross-validation

	50-50		70-30		80-20	
Classifier	Accuracy	Time (Sec)	Accuracy	Time (sec)	Accuracy	Time (sec)
Bayes	95.79	0.8900	95.88	0.8600	95.88	1.140
C4.5	95.09	23.4000	95.36	22.9200	96.34	49.030
LAD tree	95.82	78.4100	95.94	80.2000	96.13	85.870
Decision table	95.76	224.8500	95.88	217.3800	96.16	217.180

Table 3: Classifiers accuracy and time complexity with 5 fold cross-validation for the selected feature set

	50-50		70-30		80-20	
Classifier	Accuracy	Time (Sec)	Accuracy	Time (sec)	Accuracy	Time (sec)
Bayes	94.05	0.00	94.05	0.00	94.05	0.00
C4.5	94.05	0.02	94.20	0.03	94.57	0.03
LAD tree	94.27	0.34	94.63	0.34	94.60	0.34
Decision table	94.17	0.13	94.32	0.13	94.35	0.13

A java program is developed and run in netbean IDE and the selected feature are stored in a csv format to be used in weka. The classifier algorithms as discussed in this study are experimented on the data set and their performance are measured as in Eq. 9-11:

$$\text{Precision} = TP/(TP + FP) \quad (9)$$

$$\text{Recall} = TP/(TP + FN) \quad (10)$$

$$\text{FMeasure} = 2(\text{Recall} \times \text{Precision})/(\text{Recall}+\text{Precision}) \quad (11)$$

where, TP, FP, FN represents True Positive, False Positive, False Negative rate of the dataset, respectively. Precision defines the prediction made by the classifier based on the false positive rate. Recall measures the same with respect to the false negative. The F-Measure defines the weighed measure of both.

The classifiers performance metrics for the entire dataset with 5 fold cross validation is summarized in Table 1. The results of 2, 3 and 5 fold cross validation of the classifiers are given in Table 2. After selecting the 8 features, the performance of the classifiers for 2, 3 and 5 fold cross validation are given in Table 3 for comparison. It could be seen that rule based method gives the least false positive while Bayes method gives the maximum false positive. Comparing the overall performance the tree

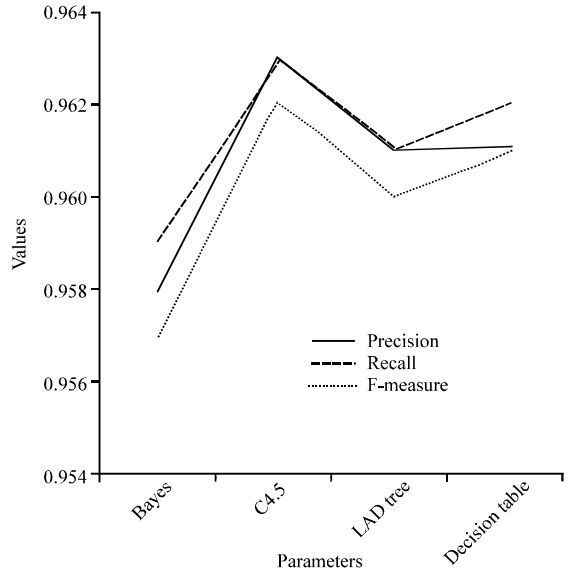

Fig. 1: Performance metrics of the classifiers

based and rule based methods perform better than probabilistic method and C4.5 shows a higher accuracy as shown in Fig. 1.

Table 2 shows that Baye's classifier takes almost a constant time and gives almost a constant accuracy irrespective of the training sample size. Other classifiers show better accuracy with increasing sample size. With regard to the execution time, Bayes classifier consumes least time where as other classifiers take more time to build the model which also increases gradually with the sample size. The decision table method shows the highest time to build the rule based model but the accuracy is close to LAD tree. The LAD Tree has better accuracy than C4.5 when the training sample is less but consumes more time than the later. However, C4.5 outperforms other method when higher training samples are used.

The accuracy of the classifiers with the selected feature set is slightly less when compared with that of the entire feature set where as there is a significant reduction in the time consumed. Baye's classifier and C4.5 takes negligible time in building the model while decision table and LAD tree showed a constant time. With increase in the training sample size, C4.5, decision tree, LAD tree method showed slight increase in the accuracy but Bayes method showed a constant accuracy as in Fig. 2.

When, the percentage of the data set taken is 80% and feature set selection process is repeated by varying the threshold value, a gradual decrease in the accuracy and a constant Time complexity is observed and are shown in the Table 4. It is inferred that when the size of the feature set and the training sample size are still reduced, all the classifiers show a decrease in their performance.

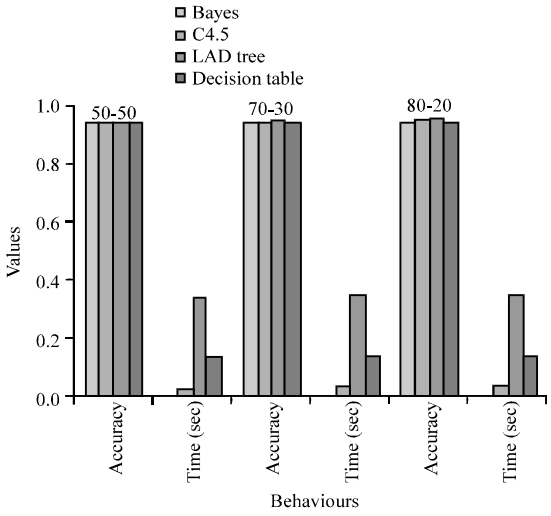

Fig. 2: Accuracy and time complexity of the classifiers

Table 4: Classifiers accuracy and time complexity for reduced data set

	80-20	
Classifier	Accuracy	Time (sec)
Bayes	93.78	0.03
C4.5	94.01	0.02
LAD tree	94.39	0.27
Decision table	94.16	0.12

CONCLUSION

In this study, the image present in the web page is classified as advertisement or not by considering various features. Probabilistic, tree and rule based classification algorithms are experimented on the data set and their performance is summarized under various conditions. Probabilistic method showed constant accuracy for various training sample size. Rule based method consumed more time when compared to tree based methods. Tree based classifier C4.5 gave better accuracy when the entire feature set was considered as it uses an inbuilt feature selection process and it also performs better for continuous attributes. LAD tree method had higher accuracy when the features were selected as it involved binary attributes.

REFERENCES

El-Deen Ahmeda, R.A., M.E. Shehaba, S. Morsya and N. Mekawiea, 2015. Performance study of classification algorithms for consumer online shopping attitudes and behavior using data mining. Proceedings of the 5th International Conference on Communication Systems and Network Technologies, April 4-6, 2015, Gwalior, pp: 1344-1349.

Graepel, T., J.Q. Candela, T. Borchert and R. Herbrich, 2010. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. Proceedings of the 27th International Conference on Machine Learning April 2009, Haifa, pp: 13-20.

Hssina, B., A. Merbou, H. Ezzikouri, M. Erritali, 2014. A comparative study of decision tree ID3 and C4.5. Int. J. Adv. Comput. Sci. Applic., 2104: 13-19.

Jushmerick, N., 1999. Learning to remove internet advertisements. Proceedings of the 3rd Annual Conference on Autonomous Agents, May 1-5, 1999, Seattle, WA., USA., pp: 175-181.

Kohavi, R., 1995. The power of decision tables. Proceedings of the 8th European Conference on Machine Learning Heraclion, April 25-27, 1995, Crete, Greece, pp: 174-189.

Kumar, S. and R.K. Arora, 2015. Analyzing customer behaviour through data mining. Int. J. Comput. Applic. Technol. Res., 4: 884-888.

Nithya, V., S.L. Pandian and R. Regan, 2013. The SQL injection attack detection and prevention by classification and analysis. Asian J. Inform. Technol., 12: 131-139.

Sharma, T.C. and M. Jain, 2013. WEKA approach for comparative study of classification algorithm. Int. J. Adv. Res. Comput. Commun. Eng., 2: 1925-1931.

Shatnawi, M. and N. Mohamed, 2012. Statistical techniques for online personalized advertising: A survey. Proceedings of the 27th Annual ACM Symposium on Applied Computing, March 26-30, 2012, Riva del Garda (Trento), Italy, pp: 680-687.