Comparative Analysis of Computer Assisted Valuation of Descriptive Answers using WEKA with Different Classification Algorithms

[1] Ruhi Dubey, [2] Rajni Ranjan Singh Makwana
[1]Student, Department of CSE & IT, Madhav Institute of Technology and Science, Gwalior, India
[2]Assistant Professor, Department of CSE & IT, Madhav Institute of Technology and Science, Gwalior, India

Abstract

Most of the exams conducted nowadays are online. Objective exams are the latest trends that are used for most of the competitive examinations. All India objective exams are one of the most successful exam patterns that are followed for different competitive examinations but a university exam are still theoritical type and at most of the places gets evaluated by using manual efforts. Here an comparative analysis is done by using weka tool and it's in built classification algorithms to perform computer assisted valuation. An experiment is carried out in our academic organization to built required dataset. Dataset consist of 530 training samples and 159 test samples that are applied on some predefined algorithms. The algorithms that are used with their respective efficiency without using any threshold value are random forest classification- 61.63%,J48 classification- 51.57%, FT classification- 61.63%,naïve bayes classification - 49.68%, random tree classification - 59.11%, REPtree classification - 53.45%.

Keywords

random forest classification, J48 classification, FT classification, naïve bayes classification, random tree classification, reptree classification automated evaluation.

Abbreviations: FT- Functional tree, REP tree- Reduces Error Pruning, WEKA- Waikato Environment for Knowledge Analysis, ID3- Iterative Dichotomiser 3.

I. INTRODUCTION

Online examination is the easiest method for conducting unbiased exams and performing optimized valuation. Most of competitive exams are performed online and objective type but university exams are performed offline and descriptive type which needs human efforts in valuation and also are biased by some percentage because all the answers do not get evaluated by the same person. Due to which the evaluating pattern and marking styles used in evaluation are different. This is the reason why different students get different marks for the same questions and answers. This leads to biased valuation and hence affects the result of students. Also valuation depends on the mood of valuator at the time of valuation for example if he is in good mood he gives good marks while if he is in a bad mood, he might give bad marks to the student or some times by mistake one gives different marks to someone. Such mistakes leads to the revaluation of answer sheets and which involve a huge amount of money.

According to a survey, every year around 36% students go for revaluation of answer sheets. For saving time and money of revaluations, an approach was proposed in which a classifier based machine learning approach is proposed for the evaluation of descriptive answers. By the same methods and using 5 more classification algorithms, a comparative analysis is performed in this paper to come up with a result that which algorithm gives the best result for the given data set with and without threshold value. In the experiment performed, two rounds are conducted in which one is without threshold value that gives the exact result and one is with +/- 1 threshold value where if the difference in human evaluated marks and system evaluated marks have a difference of +/-1 then it will also considered as a correct valuation and if difference goes beyond the threshold difference then it will considered as error.

II. LITERATURE SURVEY

Followings are the paper published in different journals and conferences which gives a basic idea regarding the topic of this paper.

1) Mohan et al.[5] proposed a Feature Clustering Algorithm for Valuation of Illustrative Type Examination. Their technique utilizing the parts of speech components like Nouns, Pronouns, Verbs, Adverbs and Adjectives as the pre announced clusters. SVM classifier is used to evaluate test samples. The authors claim that method is good for disquisition type answers only and not working accurate for formula based and Mathematical type questions.

2) Kaur et al.[6] proposes an algorithm for the valuation of single sentence illustrative answers. Similarity measurement is performed between student answers and standard answer based on full or partial string match. Work do not contains sufficient examples to validate the system.

3) C. Sunil Kumar et al.[7] presented a significant work which utilizes bagging classifier for the valuation of illustrative answers. Author claim that on an average 76% of accuracy is obtained when tested across 5 datasets using 10 fold cross validation using Naive Bayse, Logistic Regression, Random Forests, Support Vector Machine (SVM), Decision Stump and Decision Trees. Nevertheless, this paper appears to suffer from two drawbacks. The applied dataset consist of student-written essays, it could be better if the author would provide some specific questions and their answers to train the classifiers because essays valuation is totally different from illustrative answer valuation, second instead for 10 folds validation, an unseen test dataset should be supplied to test the system.

4) Mamčenko et al.[8] proposed a Illustrative model to recognize concealed patterns in student's answers using data mining techniques. Clustering techniques is applied so that we can group similar types of object together. Result includes, total time spend, time spent to give incorrect answer and time spend to give correct answers however proposed research is not directly related to the valuation of illustrative answers.

III. DESCRIPTION

A. Random Forest Classification: Random Forest Classification is the combination of a group which predicts the tree model of the data. In this classification each and every tree pivot on the values of a random vector specimen such that they are independent from each other and all each and every tree in the forest have similar allocation. Random Forest Classification Algorithm is an entity learning method i.e. a group of items viewed as a whole rather than individually, for classification, regression and other tasks. Random forests are a way of averaging multiple deep decision trees, trained on different parts of the same training set, with the goal of reducing the variance. This comes at the expense of a small increase in the bias and some loss of interpretability, but generally greatly boosts the performance of the final model[15].

B. J48 Classification: J48 is an extension of ID3. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. In the WEKA data mining tool, J48 is an open source Java implementation of the C4.5 algorithm. The WEKA tool provides a number of options associated with tree pruning. In case of potential over fitting pruning can be used as a tool for précising.[16].

C. FT Classification: FT (Functional Tree) combines a standard univariate decision tree, such as C4.5, with linear functions of the attributes by means of linear regressions. While a univariate DT uses simple value tests on single attributes in a node, FT can use linear combinations of different attributes in a node or in a leaf. In the constructive phase a function is built and mapped to new attributes. A model is built using the constructor function. The constructor function should be a classifier or a regression function depending on the type of the problem[15].

D. Naïve Bayes: The Naive Bayesian classifier is based on Bayes' theorem, it makes use of all the attributes contained in the data, and analyses them individually as though they are equally important and independent of each other. Bayes theorem provides a way of calculating the posterior probability. Naive Bayes classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

$$P\left(\frac{c}{x}\right) = \frac{P\left(\frac{x}{c}\right)P(c)}{P(x)} \quad (1)$$

P(c|x) is the posterior probability of class (target) given predictor (attribute). P(c) is the prior probability of class. P(x|c) is the likelihood which is the probability of predictor given class. P(x) is the prior probability of predictor[4].

***E. Random Tree**:* A random tree is a collection (ensemble) of tree predictors that is called forest. It can deal with both classification and regression problems. The classification works as follows: the random trees classifier takes the input feature vector, classifies it with every tree in the forest, and outputs the class label that received the majority of "votes". In case of a regression, the classifier response is the average of the responses over all the trees in the forest[15].

***F. REPTree**:* Reduces Error Pruning (REP) Tree Classifier is one of the most speedy classifier which are famous for their fast decisiveness output which work on the theory of information gain with entity and minimizing the error arising from variance. REP Tree applies regression tree logic and generates multiple trees in altered iterations. After all the repetitions, it selects the best from all generated trees. This algorithm constructs the regression/decision tree using variance and information gain. Also, this algorithm prunes the tree using reduced-error pruning with back fitting method. At the beginning of the model preparation, it sorts the values of numeric attributes once. As in C4.5 Algorithm, this algorithm also deals the missing values by splitting the corresponding instances into pieces[18].

***1) Dataset Collection*:** To build a required data set, a descriptive test is performed consisting of 8 questions given to 88 students by which we got 704 answers out of that we select 630 answers. Further that answers are divided as 530 answers for training dataset and 159 answers for testing dataset. Division of answers is done on random basis. These answers got evaluated by human expert and marks are allotted manually on the scale of 0 to 3 where 0 is considered as worst and 3 as best.

Test samples are first evaluated by manual evaluator and it also given to the classifiers for automated valuation. Classifier valuation results are compared to the manual valuation results with the objective that classifier valuation and manual valuation will produce the similar results.

2) Experimental details, methods, materials: Experiment is carried out on the system having Ubuntu LTS Linux $_v$14.04 with 1.3 GHz Intel i3 processor and 3 GB Ram. Weka $_V$ 3.6.11 machine learning workbench, developed by university of Waikato is utilized for the classification of illustrative answers.

***3) Experiment steps*:** Figure 1 show the experimental steps performed.

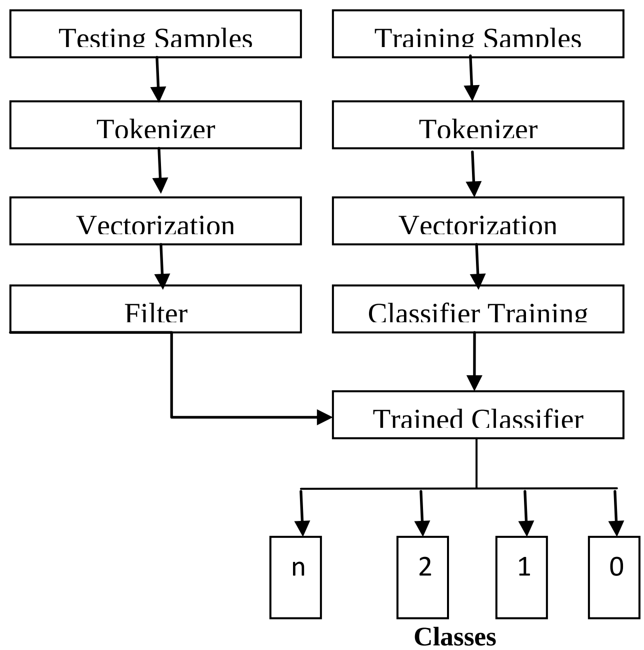

Figure 1: Experimental steps

Following questions are asked by students:

1) What is the history of C language?
2) Who developed java and when?
3) Who was the father of modern computer and when it was developed?
4) What is array? Define different types of arrays?
5) When and where C++ language was developed and by whom?
6) What are the main components of Central Processing Unit?

7) What is operating system? List any five operating systems
8) What do you mean by Loops? Define different types of loops.

***4) Training and Test Samples*:** As over mentioned that 530 answers are selected as training sample and 159 are selected as test samples. Classifiers are trained and tested individually by set of answers for each individual questions (i.e. set of answers of a question are trained individually).

IV. METHODOLOGY

***A) Tokenizer*:** It breaks the incoming answer string down into a stream of terms or tokens. A simple tokenizer divides the string up into terms wherever it encounters whitespace or punctuation. For examples: 4 answers are given

Ans. 1="great work" Ans. 2="excellent work" Ans. 3="worst work ever" Ans. 4="no comment". Identified tokens by tokenizer are: great, work, excellent, worst, ever, no, comments.

***B) Vectorization*:** Tokens obtained from the tokenizer are processed to transform into a column vector. Each vector row is represented by following structure.

$$\mathbf{V = [V_1, V_2, V_3 \ldots V_n, Class]} \quad (2)$$

Example: Identified tokens in the previous example are transformed into column vector, here Class attribute represents the marks (scale 0-3) given by manual evaluator, given bellow:

	Great	work	excellent	worst	ever	no	comment	Class
V1 =	[1,	1,	0,	0,	0,	0,	0,	2]
V2 =	[0,	1,	1,	0,	0,	0,	0,	2]
V3 =	[0,	1,	0,	1,	1,	0,	0,	0]
V4 =	[0,	0,	0,	0,	0,	1,	1,	1]

Figure 2: Vectorization

C) *Training*: Random Forest classifier, J48 classifier, FT classifier, naïve bayes classifier, random tree classifier, reptree classifier are used for training dataset

***D) Filter*:** For successful classification, both training and testing file should have same name, type and equal number of attribute (column vector). However in this work training and testing samples are having unequal column vectors. So it is required to make them compatible, therefore test samples are preprocessed by the arbitrary filter to achieve vector dimension compatibility. The composition of the filter is based solely on the training data and test instances will be processed by the filter without changing their composition.

***E) Classification*:** As mentioned above, random forest classification, J48 classification, FT classification, naïve bayes classification, random tree classification, a reptree classification algorithm are utilized to evaluate the test samples under the scale of 0-3 and produces the result. It is expected that train classifier would evaluate test samples as human evaluates them.

V. RESULTS

***Observations and Discussions*:** The data set is applied to the random forest classifier, J48 classifier, FT classifier, naïve bayes classifier, random tree classifier, reptree classifier and observed result is measured by following factors.

Table 1: Confusion Matrix

		Detected	
		Positive	**Negative**
Actual	**Positive**	A: True Positive	B: False Negative
	Negative	C: False Positive	D: True Negative

***1) True Positive Rate (TP Rate)/ Recall*:** It is the section of cases whose results are positive and that were accurately classified as positive, as calculated by using the following equation:

$$\text{Recall} = A / A + B \quad (3)$$

2) False Positive Rate (FR Rate): It is the section of cases whose results should be negative but that were inaccurately classified as positive, as calculated by using the following equation:

$$\text{FP Rate} = C / C + D \quad (4)$$

***3) Precision*:** It is the portion of the predicted positive cases which were correct and as delibrated using the following given equation:

$$\text{Precision} = A / A + C \quad (5)$$

***4) F-Measure*:** The F-Measure evaluates some mean of all the information retrieval precision and recall metrics.

$$F = 2.\frac{Precision *Recall}{Precision +Recall} \quad (6)$$

***5) ROC Curve*:** This curve represents how excellent, good and worthless experiments are plotted on the same graph. The accuracy of the test pivot on how well the experiment discrete the group being tested. A zone of 1 represents an accurate evaluation; an zone of 0.5 represents a test that has no or very little value.

***6) Kappa Statics*:** Kappa statistic is used to measure the accordance in between predicted and observed categorizations of a dataset, while correcting for an accordance that happens by coincidence. If the results of kappa is 1 then it specifies accurate accordance whereas if the result of kappa is 0 then is specifies accordance equals to chance.

7) Classification %: It is conditional on the number of samples correctly classified. Here t is the number of sample cases correctly classified, and n is the total number of sample cases.

$$Classification\,\% = 100 * \frac{t}{n} \qquad (7)$$

Following are the result and observation tables containing results of the training and testing datasets.

Table 2: Training Dataset Result

Classifier	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Time (Seconds)
J48	0.902	0.065	0.902	0.902	0.901	0.971	10.65
Random Forest	0.987	0.009	0.987	0.987	0.987	1	1.97
Random Tree	0.991	0.008	0.991	0.991	0.991	1	0.34
FT	0.989	0.01	0.989	0.989	0.989	1	36.24
Naïve Bayes	0.642	0.136	0.691	0.642	0.646	0.861	0.5
REP Tree	0.706	0.27	0.734	0.706	0.678	0.844	1.09

Table 3: Testing Dataset Result

Classifier	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Time (Seconds)
J48	0.516	0.316	0.539	0.516	0.515	0.656	07.08
Random Forest	0.616	0.265	0.613	0.616	0.608	0.775	1.7
Random Tree	0.591	0.233	0.603	0.591	0.593	0.677	0.27
FT	0.616	0.259	0.617	0.616	0.611	0.712	29.82
Naïve Bayes	0.497	0.217	0.529	0.497	0.501	0.699	0.5
REP Tree	0.56	0.355	0.593	0.56	0.511	0.623	0.98

Table 4: Result analysis

Classifiers	Without threshold value			With threshold value		
	Correct	Incorrect	Efficiency	Correct	Incorrect	Efficiency
J48	82	77	51.57%	154	5	96.85%
Random Forest	98	61	61.63%	153	6	96.22%
Random Tree	95	65	59.11%	155	4	97.84%
FT	98	61	61.63%	152	7	95.59%
Naïve Bayes	79	80	49.68%	144	15	90.56%
REP Tree	85	74	53.85%	140	19	88.05%

A certain threshold value is applied to the dataset i.e. +/-1. If the difference between human evaluated and system evaluated marks is +/-1 then if will be considered as correct evaluation while if difference goes beyond threshold then it will be considered as error. Table 2 and table 3 shows the different factors of training and testing data set while table 4 shows the final result of all algorithms with and without considering the threshold value. From above experiments, Random Tree and Random Forest gives the most optimized results and also have less error rate with respect to others.

VI. CONCLUSION & FUTURE WORK

Random tree and Random forest gives optimized result with and without threshold value. Still work has to be done in this field for improving grammatical errors and Recognization of such context. Also classifier may get confuse within same type of questions and answers. Natural language processing and artificial intelligence can be used to make system better.

VII. REFERENCES

[1] Andy Liaw, Matthew Wiener, "Classification and Regression by randomForest" ,R News, Vol. 2/3, December 2002, ISSN 1609-3631

[2] Tina R. Patil, Mrs. S. S. Sherekar, "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification", International Journal Of Computer Science And Applications, Vol. 6, No.2, Apr 2013, PP: 256-261

[3] Vasilis Vryniotis, "Machine Learning Tutorial: The Multinomial Logistic Regression" weblink:http://blog.datumbox.com/machine-learning-tutorial-the-multinomial-logistic-regression-softmax-regression/.

[4] Naive Bayesian, web. link: http://www.saedsayad.com/naive_bayesian.htm.

[5] A Krishna Mohan, MHM Krishna Prasad, "A Novel Feature Clustering Algorithm for Valuation of Descriptive Type Examination", International Journal of Computer Applications, Volume 98– No.9, July 2014, PP: 35-41.

[6] Amarjeet Kaur, M Sasikumar, Shikha Nema, Sanjay Pawar, "Algorithm for Automatic Valuation of Single Sentence Descriptive Answer", International Journal of Inventive Engineering and Sciences (IJIES), Volume-1, Issue-9, August 2013, PP:6-9

[7] C. Sunil Kumar, R. J. Rama Sree, "An Attempt to Improve Classification Accuracy through Implementation of Bootstrap Aggregation with Sequential Minimal Optimization during Automated Valuation of Descriptive Answers", Indian Journal of Science and Technology, Vol 7(9) , September 2014,PP: 1370-1375.

[8] J. Mamčenko, I. Šileikienė, J. Lieponienė and R. Kulvietienė, "Evaluating the Data of an e-Examination System using a Descriptive Model in Order to Identify hidden Patterns in Students Answers", The Online Journal on Computer Science and Information Technology (OJCSIT), Vol. (1) – No. (2), PP: 45-49.

[9] S.L. Ting, W.H. Ip, Albert H.C. Tsang "Is Naïve Bayes a Good Classifier for Document Classification", International Journal of Software Engineering and Its Applications, Vol. 5, No. 3, July, 2011, PP: 37-46.

[10] Jehad Ali , Rehanullah Khan , Nasir Ahmad , Imran Maqsood, "Random Forests and Decision Trees", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012 ISSN (Online): 1694-0814.

[11] Performance Measures, http://www.seas.gwu.edu/~bell/csci243/lectures/performance.pdf

[12] Gerard Biau, "Analysis of a Random Forests Model", Journal of Machine Learning Research 13 (2012) 1063-1095.

[13] S. Bharathidason, C. Jothi Venkataeswaran, "Improving Classification Accuracy based on Random Forest Model with Uncorrelated High Performing Trees", International Journal of Computer Applications (0975 – 8887) Volume 101– No.13, September 2014.

[14] Priyanka Sharma, "Comparative Analysis of Various Decision Tree Classification Algorithms using WEKA", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 2 684 – 690

[15] Gaganjot Kaur, Amit Chhabra, " Improved J48 Classification Algorithm for the Prediction of Diabetes", International Journal of Computer Applications (0975 – 8887) Volume 98 – No.22, July 2014

[16] S.K. Jayanthi, S.Sasikala. "REPTree Classifier for indentifying Link Spam in Web Search Engines" IJSC, Volume 3, Issue 2, (Jan 2013), 498 – 505, 2013