International Journal of Pure and Applied Mathematics
Volume 118 No. 24 2018
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue

http://www.acadpubl.eu/hub/

Subjective Answer Evaluation Using Machine Learning

Piyush Patil [1],Sachin Patil [2] ,
Vaibhav Miniyar[3],Amol Bandal[4]
[1,2,3,4]Department of Computer Engineering,
Sinhgad Institute of Technology, Lonavala,India
piyushpatil666@gmail.com[1],
sachin.patil521@gmail.com [2],
vaibhavminiyar@gmail.com[3],
amolbd1987@gmail.com[4]

May 23, 2018

Abstract

The current way of checking subjective paper is adverse. Evaluating the Subjective Answers is a critical task to perform. When human being evaluates anything, the quality of evaluation may vary along with the emotions of Person. In Machine Learning, all result is only based on the input data provided by the user. Our proposed system uses machine learning and NLP to solve this problem. Our Algorithm performs a task like Tokenizing words and sentences, Part of Speech tagging, Chunking, Chinking, Lemmatizing words and Wordnetting to evaluate the subjective answer. Along with it, our proposed algorithm provides the semantic meaning of the context. Our System is divided into two modules. The first one is extracting the data from the scanned images and organizing it in the proper manner and the second is applying ML and NLP to the text retrieved from the above step and giving marks to them.

***Key Words*:**Nave bayes, Cosine Similarity, Classifier, Semantic Checking, Machine Learning

1 Introduction

The manual system for evaluation of Subjective Answers for technical subjects involves a lot of time and effort of the evaluator. Subjective answers have various parameters upon which they can be evaluated such as the question specific content and writing style. Evaluating subjective answers is a critical task to Perform. When human being evaluates anything, the quality of evaluation may vary along with the emotions of the person. Performing evaluation through computers using intelligent techniques ensures uniformity in marking as the same inference mechanism is used for all the students. In Machine Learning, all result is only based on the input data provided by the user. Our Proposed System uses machine learning and NLP to solve this problem. Our Algorithm performs a task like Tokenizing words and sentences, Part of Speech tagging, Chunking, chinking, Lemmatizing words and Wordnetting to evaluate the subjective answer. Along with it, our proposed algorithm provides the semantic meaning of the context. Our System is divided into two modules, Extracting the data from the scanned images and organizing it in the proper manner and Applying ML and NLP to the text retrieved from the above step and giving marks to them. The software will take a scanned copy of the answer as an input and then after the preprocessing step, it will extract the test of the answer. This text will again go through processing to build a model of keywords and feature sets. Model answer sets and keywords categorized as mentioned will be the input as well. The classifier will then, based on the training will give marks to the answers. Marks to the answer will be the final output. The need for online examination aroused mainly to overcome the drawbacks of the existing system. The main aim of the project is to ensure user-friendly and more interactive software to the user. The online evaluation is a much faster and clear method to define all the relevant marking schemes. It brings much transparency to the present method of answer checking The answers to all the questions after the extraction would be stored in a database. The database is designed as such that it is very easily accessible. Automating repetitive tasks has been the main aim of the industrial and technological revolution. The work of checking hundreds of answer sheets which more or less contains the same answer can be quite a boring task for the teachers. This system can be used instead in order to reduce their burden. It will save a lot of effort and time on teachers part. The human efforts applied in this repetitive task can be saved and spent more in other academic endeavors. The obvious human mistakes can be reduced to obtain an unbiased result. The system calculates the score and provides results fairly quickly. This system can be widely used in academic institutions such as schools, colleges, coaching and institutes for checking answer sheets. It can also be implemented in different organizations which conduct competitive examinations.

The software will take scanned copy of the answer as an input and then after the preprocessing step it will extract the test of the answer. This text will again go through processing to build a model of keywords and feature sets. Model answer sets and keywords categorized as mentioned will be the input as well. Classifier will then, based on the training will give marks to the answers. Marks to the answer will be the final output.

The paper is organized as follows: Section II contains the review of related work. Section III gives brief idea about working of system. Section IV contains Experimental Analysis and section V contain the conclusions of this research work.

2 LITERATURE SURVEY

Evaluation of subject answer checking isnt a new thought. It has been in the works since a decade and a half. A large number of techniques where experimented with to solve the problem efficiently. Natural Language processing, Latent Semantic Analysis, Generalized Latent Semantic Analysis, Bayes theorem, K- nearest neighbor, etc. In general they can be categorized as follows : Clustering techniques, classification techniques and natural language processing techniques.Intelligent Essay evaluator developed by Landauer[3],[4-7] in 2003 using a technique known as Latent Semantic Analysis. It gives results in the accuracy range of 60-90 %. A slightly better version of using the probabilistic LSA technique[8-10] used to develop automatic essay evaluator tool by Kakkonen. Generalized LSA[11] technique extends the LSA approach by working on vectors(n-gram, bag of vectors) instead of the dual document-term representation.

It gave a better accuracy upto 96% .

Along with the above clustering methods, classification methods such as Bayes theorem[12], K-nearest neighbour[13], Maximum entropy[14], etc were also experimented with. Bayes theorem used by Rudner in 2002 has an accuracy of 80%. The clustering technique of K nearest neighbour is based on random selection of cluster heads and then carving out clusters based on their distances from those heads. It produced results with 76% accuracy. A tool named as C-rater which uses the Maximum Entropy technique for evaluation of short answers. It gives an 80% accuracy with the score assigned by a human-grader. One major natural language processing technique we ought to look is BLEU (bilingual evaluation understudy)[19-20] is a basically an algorithm for evaluation of the text quality which has been translated with the help of a machine from one language to another. Though it is built to mimick human evaluations at a corpus level, it has a bad performance if it is used to evaluate the quality of individual sentences, which explains the poor accuracy of 50%.

3 WORKING

This system can be widely used in academic institutions such as schools, colleges, coaching and institutes for checking answer sheets. It can also be implemented in different organizations which conduct competitive examinations. Our Algorithm performs a task like Tokenizing words and sentences, Part of Speech tagging, Chunking, chinking, Lemmatizing words and Wordnetting to evaluate the subjective answer. Along with it, our proposed algorithm provides the semantic meaning of the context.

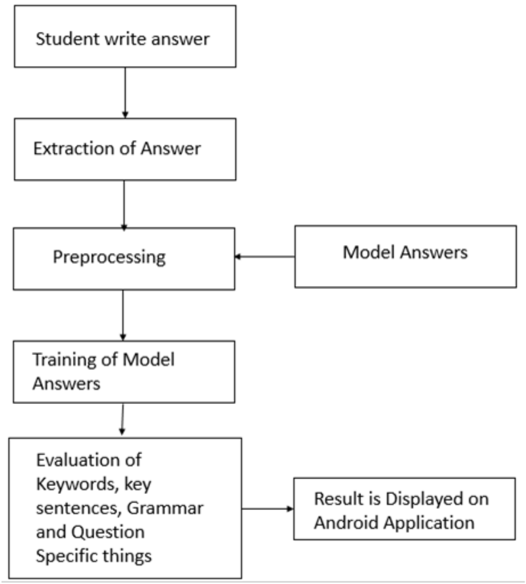

Student write answer
Extraction of Answer
Preprocessing
Model Answers
Training of Model Answers
Evaluation of Keywords, key sentences, Grammar and Question Specific things
Result is Displayed on Android Application

Fig. 1-Workflow diagram

Student writes answer on answer-sheet.The system will take scanned copy of the answer as an input and then afterthe preprocessing step it will extract the text of the answer. Model answer sets will be provided by the moderator/evaluator. This model answers will be then trained. The evaluator also provides with the keywords and question specific things(QST). Model answer sets and keywords categorized as mentioned will be the input aswell. Now the answer text extracted from the student ie. the user answer will be searched for the presence of keywords, QST, grammar and will be categorized and named internally as provided in table1. Grammar will be checked with an api and internalized as given in table 2. Nave bayes is used as a based classifier in our system.Nave bayes is based on three parameters i.e. Keywords, Grammar and Question Specific things.

Mathematical model:

P(Class — Keyword,Grammar,QST) = P(Keywords — Class) *P(grammar — class)*P(QST — class)*P(class)

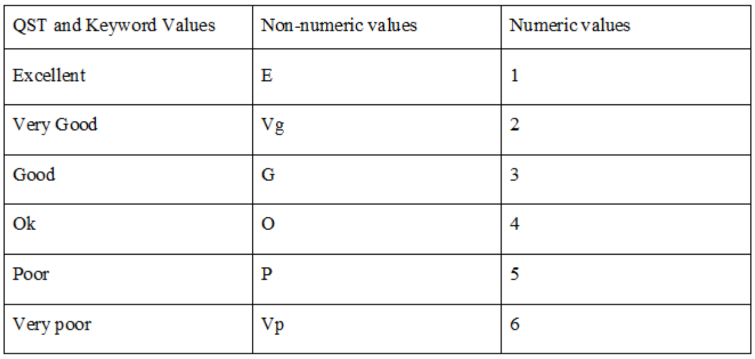
For example If we have the values of Keywords,grammar,QST as 2,0,2 respectively. Then we can evaluate the class. For the input values given, it is evaluated against all the classes and then the class having the maximum probability is the class to which the given input will belong.

Each Answer will be given biased value in between 1-10. Actual marks will be evaluated on that basis. For example, If after evaluation 6 is the value we are getting and question is of 5 Marks then the actual marks calculated from this using -

Marks obtained for the answer out of 10 = 6 * 5 /10 = 3 i.e. generically we can define the formula as

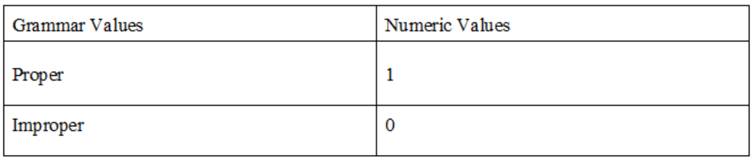
Actual Marks=(Biased Value After Evaluation) * (Max marks that can be obtained for the answer/ 10)

As the dataset in ML approaches works well with numeric dataset we have mapped our six values as follows:-



QST and Keyword Values	Non-numeric values	Numeric values
Excellent	E	1
Very Good	Vg	2
Good	G	3
Ok	O	4
Poor	P	5
Very poor	Vp	6

Table. 1-Handling Non-Numeric Values of Keywords and key sentences



Grammar Values	Numeric Values
Proper	1
Improper	0

Table. 2-Handling Non-Numeric Values of Grammer

These 3 values i.e. Keywords, Grammar and Question Specific things is passed to Nave bayes classifier as a input. Naive bayes classifier is probabilistic classifier which is based on the maximum probability to which the given input belongs.

Now inorder to evaluate these 3 parameters we are using following strategy :-

i]Keywords and key sentences: (e, vg, g, o, p, vp)

In Cosine Similarity first we will make the vector of both model answer and users answer.We transfer the answer into vector form using cosine method.Lesser the Angle greater the similarity and greater the value of cosθ.

It is calculated using two components numearator(num) and denominator(den)

Num = $\sum$(vec1[x] * vec2[x])

Where vec1[x] andvec2[x] are answer vectors and model answers vector respectively.

Den= $\frac{sum1}{den} * \sqrt{sum2}$

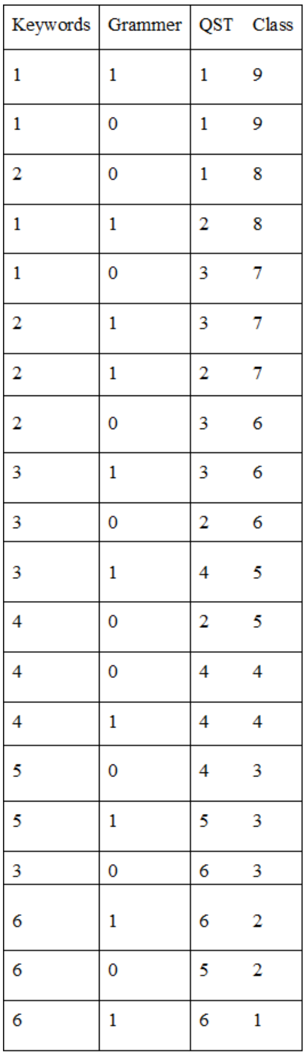
Where sum1,sum2 are keys obtained from model answer and user answer.

ii]Grammar: (y, e) - API which gives number of errors in the answer. This is evaluated only if above phase has some value. For Improper Grammar: 0, For Correct Grammar: 1

iii]Question Specific things: (e, vg, g, o, p, vp).Here we are using Fuzzy wuzzy - using multiple ratio functions available in Fuzzy Logic.The Fuzzywuzzy library analyses the text using degrees/features of text instead of the rigid Boolean values of 0 or 1.

After having some observations we have 21 of them as our training dataset.We will get any one of 1-9 value as the output from this Classifier. NB classifier gives the output with evidence/ probability value. Further Depending on that value we can increase the accuracy

Example: if output is 6 and the probability that given query belongs to 6 is 70% then we can increase this 6as $6 + 0.7 = 6.7$. So the total marks evaluated on the basis of 6.7.



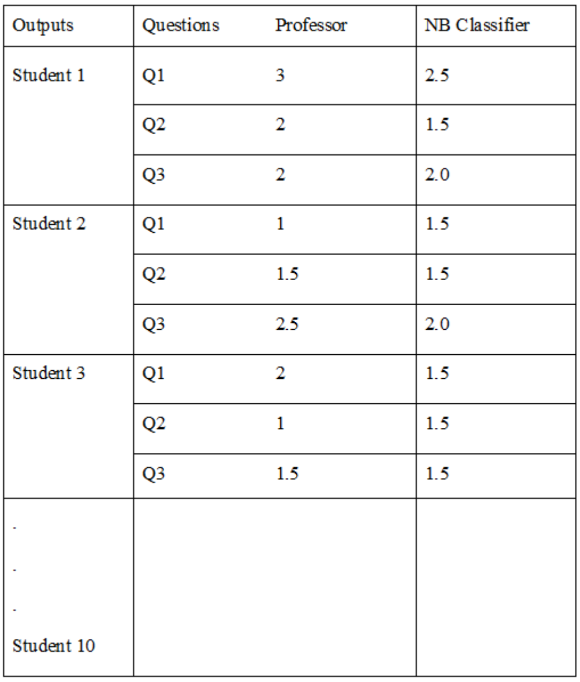
Keywords	Grammer	QST	Class
1	1	1	9
1	0	1	9
2	0	1	8
1	1	2	8
1	0	3	7
2	1	3	7
2	1	2	7
2	0	3	6
3	1	3	6
3	0	2	6
3	1	4	5
4	0	2	5
4	0	4	4
4	1	4	4
5	0	4	3
5	1	5	3
3	0	6	3
6	1	6	2
6	0	5	2
6	1	6	1

Table. 3-Dataset Marking Scheme

We have trained our model using above dataset. The values that we have defined in the yable are set according to the requirement of the answer. The evalautor/moderator/teacher of the answersheet can define these values for themselves to suit their needs.

4 EXPERIMENTAL ANALYSIS

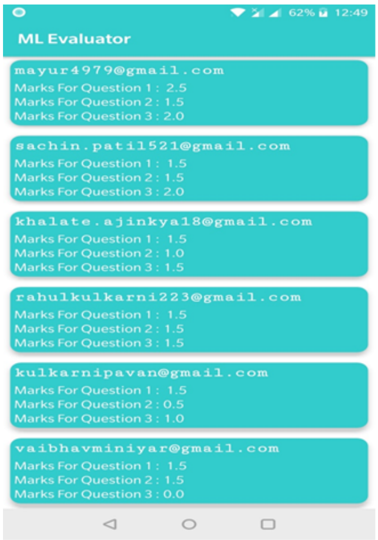
We have given 3 questions to each student. Total number of student was 20.Each question carries 5 Marks .All answers are evaluated firstly by 10 Professors then our algorithm will evaluate them .Then the similarity betweenProfessorEvaluation and our algorithmevaluation is taken into consideration .we have found :-



Outputs	Questions	Professor	NB Classifier
Student 1	Q1	3	2.5
	Q2	2	1.5
	Q3	2	2.0
Student 2	Q1	1	1.5
	Q2	1.5	1.5
	Q3	2.5	2.0
Student 3	Q1	2	1.5
	Q2	1	1.5
	Q3	1.5	1.5
. . . Student 10			

Table. 4- Comparative Evaluation Result Table

We have made python flask web app for experiment purpose, where students will write the subjective question answersand we also have made an android application to show the results.



62% 12:49

ML Evaluator

mayur4979@gmail.com
Marks For Question 1 : 2.5
Marks For Question 2 : 1.5
Marks For Question 3 : 2.0

sachin.patil521@gmail.com
Marks For Question 1 : 1.5
Marks For Question 2 : 1.5
Marks For Question 3 : 2.0

khalate.ajinkya18@gmail.com
Marks For Question 1 : 1.5
Marks For Question 2 : 1.0
Marks For Question 3 : 1.5

rahulkulkarni223@gmail.com
Marks For Question 1 : 1.5
Marks For Question 2 : 1.5
Marks For Question 3 : 1.5

kulkarnipavan@gmail.com
Marks For Question 1 : 1.5
Marks For Question 2 : 0.5
Marks For Question 3 : 1.0

vaibhavminiyar@gmail.com
Marks For Question 1 : 1.5
Marks For Question 2 : 1.5
Marks For Question 3 : 0.0

Fig. 2- Students Mark Evaluation Application Screenshot

5 CONCLUSION

The techniques discussed and implemented in this project should have a high agreement (up to 90 percent) with Human Performance. The project works with the same factors which an actual human being considers while evaluation such as length of the answer, presence of keywords, and context of key-words. Use of Natural Language Processing coupled with robust classification techniques, checks for not only keywords but also the question specific things. Students will have certain degree of freedom while writing the answer as the system checks for the presence of keywords, synonyms, right word context and coverage of all concepts. It is concluded that using ML techniques will give satisfactory results due to holistic evaluation. The accuracy of the evaluation can be increased by feeding it a huge and accurate training dataset. As the technicality of the subject matter changes different classifiers can be employed. Further improvement by taking feedback from all the stakeholders such as students and teachers can improve the system meticulously.

References

[1] B. Rujiang and L. Junhua, Improving documents classification with semantic features, 2nd Int.Symp. Electron. Commer. Secur. ISECS 2009, vol. 1, pp. 640643, 2009.

[2] P. D. Turney and P. Pantel, From frequency to meaning: Vector space models of semantics, J.Artif. Intell. Res., vol. 37, pp. 141188, 2010.

[3] T. K. Landauer, Automatic Essay Assessment, Assess. Educ. Princ. Policy Pract., vol. 10, no. 3, pp.295308, 2003.

[4] T. K. Landauer, P. W. Foltz, and D. Laham, An introduction to latent semantic analysis, DiscourseProcess., vol. 25, no. 23, pp. 259284, 1998.

[5] T. K. Landauer and P. W. Foltz, An introduction to latent semantic analysis, Discourse Process.,no. April 2012, pp. 3741, 2012.

[6] P. W. Foltz, W. Kintsch, and T. K. Landauer, The measurement of textual coherence with latent semantic analysis, Discourse Process., vol. 25, no. 23, pp. 285307, 1998.

[7] P. W. Foltz, Latent Semantic Analysis for Text-Based, Behav. Res. Methods, Instruments Comput.,vol. 28, no. 2, pp. 197202, 1996.

[8] T. Kakkonen, N. Myller, E. Sutinen, and J. Timonen, Comparison of dimension reduction methodsfor automated essay grading, Educ. Technol. Soc., vol. 11, no. 3, pp. 275288, 2008.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, Latent Dirichlet Allocation, J. Mach. Learn. Res., vol. 3,no. 45, pp. 9931022, 2012.

[10] T. Hofmann, Probabilistic latent semantic indexing, Sigir, pp. 5057, 1999.

[11] M. Islam, Automated Essay Scoring Using Generalized, in Proceesings of 13th InternationalConference on Computer and Information Technology (ICCIT 2010), 2010.

[12] L. Rudner and T. Liang, Automated essay scoring using Bayes theorem, J. Technol. Learn. ,vol. 1, no. 2, 2002.

[13] L. Bin, L. Jun, Y. Jian-Min, and Z. Qiao-Ming, Automated essay scoring using the KNN algorithm,Proc. - Int. Conf. Comput. Sci. Softw. Eng. CSSE 2008, vol. 1, pp. 735738, 2008.

[14] C. Leacock and M. Chodorow, C-rater: Automated scoring of short-answer questions, Comput.Hum., vol. 37, no. 4, pp. 389405, 2003.

[15] J. Z. Sukkarieh, Using a MaxEnt classifier for the automatic content scoring of free-text responses,AIP Conf. Proc., vol. 1305, pp. 4148, 2010.

[16] J. Sukkarieh and S. Stoyanchev, Automating Model Building in c-rater, Proc. 2009 Work. , no.August, pp. 6169, 2009.

[17] J. Burstein, K. Kukich, S. Wolff, C. Lu, M. Chodorow, L. Braden-Harder, and M. D. Harris, Automated scoring using a hybrid feature identification technique, Proc. 17th Int. Conf. Comput.Linguist. -, vol. 1, p. 206, 1998.

[18] D. Callear, J. Jerrams-Smith, and V. Soh, Bridging gaps in computerised assessment of texts, Proc.- IEEE Int. Conf. Adv. Learn. Technol. ICALT 2001, pp. 139140, 2001.

[19] P. Diana, A. Gliozzo, C. Strapparava, E. Alfonseca, P. Rodr, and B. Magnini, Automatic Assessment of Students free-text Answers underpinned by the Combination of a B LEU inspired algorithm and Latent Semantic Analysis, Mach. Transl., 2005.

[20] F. Noorbehbahani and a. a. Kardan, The automatic assessment of free text answers using a modified BLEU algorithm, Comput. Educ., vol. 56, no. 2, pp. 337345, 2011.

[21] W. Wang and B. Yu, Text categorization based on combination of modified back propagation neural network and latent semantic analysis, Neural Comput. Appl., vol. 18, no. 8, pp. 875881, 2009.

[22] C. A. Kumar, M. Radvansky, and J. Annapurna, Analysis of a Vector Space Model , Latent Semantic Indexing and Formal Concept Analysis for Information Retrieval, vol. 12, no. 1, pp. 3448, 2012.