Score Prediction by Using Machine Learning Technique

1Akshay Ozarkar, 2Sumit Pawar, 3Swapnil Sathe, 4Suraj Tapkir, 5Prof. N.D. Kale

Abstract: In the field of online learning, the main part is an online test. Sometimes it becomes difficult to assess the scoring of these tests, quizzes or complex answers. Also, it becomes difficult for a teacher to analyze all these answers for huge class. The proposed system should be able to predict scores for descriptive answers and approach makes use of data mining techniques of the k-nearest neighbors (KNN) algorithm to predict the score. This paper presents text classification for subjective scoring using k-nearest neighbors in order to automatically score the subjective answer. The product can be used for evaluating scores of descriptive answers by using TF-IDF and machine learning classifier. The proposed similarity is based on word matching and word ordering. If the pair of text match words/phrases and have the same order, the similarity is high. The results suggest the new approach is able to predict the subjective score and achieve the highest predictive accuracy than the standard classifiers. To reduce human efforts to analyze descriptive answers, we are developing a system that will automatically evaluate the answers and predict score using TF-IDF and KNN. As a result, the proposed method can be a useful tool for instructors. This product can provide accurate score than manual analysis by teachers. It can be a useful tool that helps the instructor to score the subjective answer.

Keywords: KNN Algorithm, Keyword Extraction, Learning Theory, Similarity.

Introduction:
Now a day’s E-learning is becoming popular in educational field. E-learning adapts the use of computer and network to improve quality of education. The motive of E-learning is to allow student to learn independently at any place at any time. In order examine skills and knowledge of student, test is important. As there are many number of students and their different answers, teacher may need long time for evaluation of these answers. Sometimes teacher may give unfair score for corresponding answer. In order to evaluate skills, knowledge, or attitudes of students, testing is required. Different type of testing depends on the purpose of testing, for example, the objective that helps Learners measure their own progress may use frequent short tests, for motivate learning may use informal pre-tests, for instances. The questions of testing can be classified to objective and subjective. Objective test questions or closed-response questions are the question that provides many choices. A Student must select a correct answer. The objective question is quickly and easily scored by a computer. In contrast, subjective test questions or open-response questions require human to evaluate the answer. The answer is complex, includes many words. The teacher must use their skill and time to evaluate the answer. In a large class, high number of students, high number of answers, lots of different answers, a teacher must take a long time to evaluate and may unfair score the answer. In order to solve this problem, this paper proposed an automatic scoring for subjective answers based on k-nearest neighbor’s algorithm. It uses synonyms, pronominal reference in order to improve the predictive accuracy score of short answer. This paper introduces text classification for subjective scoring using k-nearest neighbour’s algorithm (KNN). The main process of the proposed algorithm includes three main steps of: generates supplied test set, split the answer text to the list of words/phrases, and finally apply TF-IDF for keyword extraction and k-nearest neighbor classifier for score prediction with the proposed similarity measure. This study defines the available similarity measures for KNN algorithms based on word matching and word ordering. The algorithm can apply to short and long subject answers.

Related work:

A. Text Matching

In the text matching the main task in natural language processing. It uses words, phrases, or sentences to produce the similarity score between the pair of texts. Text matching can start form word matching. It finds the same word between the pair of texts. In a high number of words, phrase matching, it finds the same phase using n-gram. In a high number of phrase, sentence matching, it matches the sentence between the pair of texts. It requires word and phrase matching in order to produce the matching score. In this study, text (subjective answer) can be written in Thai and English languages, it is hard to use phrase matching and sentence matching. In order to find the similarity score, this study uses the basic word matching.

B. K-Nearest Neighbors Algorithm (KNN):

KNN algorithm is one of the simplest classification algorithm and it is one of the most used learning algorithms. KNN is in the top 10 algorithms in data mining. K-Nearest Neighbors is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection. The algorithm finds the answer of a sample using the majority vote of its neighbor’s samples.

TF-IDF Technique:

To compute the TF-IDF value for a one document, we take that document and calculate the TF-IDF score for each unique word without stop words. For each unique word, or term,

1. Term frequency: This is used for measuring how frequently particular term appears in documents.
TF(term, document) = \(\frac{\text{Number of times the term appears in doc}}{\text{Total number of words in a doc}}\) ........eq1

2. Inverse Document Frequency: This is used for measuring how frequently particular term appears in all documents.

\[
\text{IDF}(\text{term}) = \begin{cases} 
0 & \text{Total number of docs} \\
\ln \left( \frac{\text{Number of docs containing the term}}{\text{Total number of docs}} \right) & \text{otherwise}
\end{cases}
\] 

...eq2

3. On equation 1 and 2 calculate overall TFIDF of word,

TF-IDF = TF * IDF

**Motivation:**

It was difficult and time consuming task for teachers to analyze descriptive answers. To reduce human efforts to analyze descriptive answers, we are developing a system that will automatically evaluate the answers and predict score. This system will provide more reliable approach as it uses two methods. Teachers will create a test. Students will appear for that test. Answers given by students will be compared to answers set by teachers using KNN and TF-IDF method. Based on similarity score will be predicted. This system will provide the new approach is able to predict the subjective score and achieve the highest predictive accuracy than the standard classifiers. As a result, the proposed method can be a useful tool for instructors.

**System Architecture:**

It can be a useful tool that helps the instructor to score the subjective answer. However, in a high number of target classes, high number of difference score, the predictive accuracy may be low. In our System Teachers will create a test with descriptive answer and this answer save in database. Students will Attempt the test. Answers given by students will be compared to answers set by teachers and keyword extraction using TF-IDF method. Based on similarity, score will be predicted using KNN.

**Fig. system overview**

**Conclusion:**

This paper presents text classification for subjective scoring using k-nearest neighbor’s algorithm. As there are many numbers of students and their different answers, teacher may need long time for evaluation of these answers. However, in a high number of target classes, high number of difference score, the predictive accuracy may be low. The lower number of samples with high number of target classes causes the classification model did not generalize. For resolving this issue, we are going to propose an automatic Score Prediction System for subjective answers based on TF-IDF and k-nearest neighbor’s algorithm. It can be a useful tool that helps the instructor to score the subjective answer. In this paper, we will be developing a system that will minimize the effort of teacher to predict student’s score.

**Reference:**


