Student Frustration Level Estimation using Machine Learning

1. Komal Kalbhor, computer dept, KJCOEMR, Pune
2. Poorva Kurhade, computer dept, KJCOEMR, Pune
3. Rupali Mohite, computer dept, KJCOEMR, Pune
4. Pooja Talathi, computer dept, KJCOEMR, Pune
5. Prof. R.R.Patil ,computer dept, KJCOEMR, Pune

Abstract

Most of the time students often get frustrated due to long class hours and hectic schedule in the academic process. To maintain their enthusiasm up it always needs to motivate them in the right direction, otherwise the student may end up failing in the exams and in turn it may yield in discontinuation of the current education . There are no particular methodologies are existed to measure their frustration level so that right action can be taken to boost them up. Some finger counting methodologies are existed which are used to measure the frustration level by using some data mining techniques. So to boost the level of precision proposed model uses the student's behavior by providing some tests. The scores of these tests are used in the machine learning process using Decision tree to evaluate the frustration level of the student. Based on this frustration level, proposed model displays some motivational suggestions to the student to come out of the current frustration level.

Keywords: *K-Nearest Neighbor, Entropy Estimation, colloborative filtering, Decision Tree.*

I. INTRODUCTION

Feature Extraction is one of the most important aspects of machine learning. Features of a certain dataset need to be extracted efficiently for optimum performance of the machine learning algorithm. The features that are extracted from the samples are extremely useful and allow for a greater degree of precision in the calculations. The values that are extracted from the datasets have to be completely distinct.

As the distinction itself that provides a very accurate means of classification that can be achieved from the dataset. Therefore, the value that is considered to be the defining feature of the entire data, should be definitely very unique as the algorithm that works should be able to discern the originality of the content being segregated.

This is one of the most important phases of the pre-processing that is done on the data that is supposed to be fed to a machine learning application. And according to how well the data is segregated, has its effects on the actual processing of the data by the machine learning algorithm. a poorly pre-processed data would be practically useless as it would increase the space and time complexity of the system as a whole.

If the data that is extracted is contaminated with unwanted data then it adds to the space in the clusters that are being generated. This increase in size can result in a lot of failed clusters with the non-optimal data that is usually expensive to be processed on the already bottlenecked computational power of the machine. As limited computational power cannot be wasted on useless data frivolously. Therefore, feature extraction is one of the most important fields in the realm of Machine Learning pro-processing. A decision tree is one of the most widely used tools for the prediction or segregation of data. A decision tree is also really simple and intuitive to use and apply. It provides a greater degree of precision over most components that rely on various use cases that exist around most decisions. The Decision Tree follows a flowchart-like structure which determines various possibilities that can occur.

The Decision Tree is one of the most powerful and versatile tools for the purpose of selection and filtering of the desired elements. The desired elements need to be extracted through the process of eliminating the undesirable elements and processing them for every possible parameter that is essential. The decision tree appears like branches of a tree hence the name that is given. The source set is divided into two subsets based on certain parameters that are relevant to the task. The desiredoutcomes are then made to go through further iterations on the parameters are performed recursively to get even smaller and smaller subsets. These subsets correspond to the desired outcomes that are being filtered through these parameters. This is a very powerful process that can filter the undesirable data quite efficiently.

Collaborative Filtering is one of the widely used methods for the filtering of data or prediction of various data items. Collaborative Filtering is a highly flexible and versatile technique for the automatic estimation of various values. It is commonly used for the purpose of a recommendation engine. The main criteria for the Collaborative Filtering is the recommendation of various items to the user based oncertain data about the user, such as spending habits, taste etc.

Collaborative Filtering also utilises the tastes and selections of various users before the current user with similar tastes for the recommendation system. As the data from various users with a somewhat similar preferences are pooled in to collect and analyse the data. This data is then used further to recommend various items that can be very precise as the precision depends upon the quality of the data. Collaborative Filtering has been user predominantly almost everywhere, such as e-commerce websites, mineral exploration etc.

This research paper dedicates section 2 for analysis of past work as literature survey, section 3 deeply elaborates the proposed technique and whereas section 4 evaluates the performance of the system and finally section 5 concludes the paper with traces of future enhancement.

II. RELATED WORKS

R. Rajendran [1] ITS is an Intelligent Tutoring System which is basically based on the students who are interested in a self-Learning system which provides the learning material to the student which they need. This paper presents the student frustration level based on the data mining approach. By using the theoretical data and by understanding the feedback of data the predictors can predict the frustration level of the students. Once the frustration is detected the motivational video or the messaged is conveyed or shown to students to reduce the frustration level.

R. Rajendran [2] There has been vast research made on the data mining technique in the last decades, but there is some important research done by students. Thus, one of the researches is student Frustration detection, which is detected by gathering information from the student side. By using the data mining approach, the accuracy rate of the frustration is very high and the accurate compared to the other researches. This model has tested on the different peoples the accuracy and the frustration level is detected by clear interpretation.

E. Ibsen [3] This paper presents the detection of the frustration level of the students due to the workload and two assists them and to support the action of the students. The main aim of this paper is to reduce the workload and help them in learning the concepts. With reduced the student load it also finds why students are affected or getting frustrated and after finding this they find the way to help them to reduce frustration by showing the motivational videos or by giving the motivational speech and also analysis the students' performance.

S.Ahmad [4] The proposed methodology the given presents three different techniques improve the student performance there are CS, i.e. cognitive skills, frustration, CS Related Human Factors (CSRF) There are four steps include improving the performance first is to prepare the primary range of CS for the accurate division of student skills. Second is dividing the frustration by using the layers of the umbrella. Thirdly the divide the SCRF in two factors such as job such as services, health, teacher, and home and second is rural and urban. And the last is to test this on two data sets.

S. Warren [5] There is the student who isfrustrating from the c programming languages due to the lack of mathematical concepts and the main part, and the important part is the linear system to solve this difficulty Kansas

State University (KSU) Electrical & Computer Engineering (ECE) Department recently added a required course, ECE 540 – Applied Scientific Computing to understand the c programming. Thus, this brings greater confidence in the field of development.

J. Laser [6] Authors proposed the frustration level of the students and the workplace user due to loss of time. When there crash in the network or in computers or because of the poor interfaces, there are consequences for the individuals and for the organizations. Thus, this paper studies the frustration level on the daily interchange of the user and how much time it has taken to fix the problem and it also checks on what level the frustration can increase. The frustration cannot be of personal it also effects of disrupting the workplace.

L. Aik [7] There are various traffic systems which gives the details of traffic and analysis the traffic. There are models which study the traffic behavior and the network traffic situation. to study the traffic behaviors, there are various things to be known such as acceleration rule, slow down probability, traffic density, maximal and minimal speeds are the main effects by studying and by knowing this method we can prepare for the determine driving behavior. There is an additional factor to know how much time it will take to reach the destination.

P. Kurtiker [8] In this paper the black-box technique is used to detect the frustration level among the learners. Error Quotient (EQ) is calculated of each student which is entered in the laboratory the student who is having the highest EQ has the higher frustration rate and the student who has the lowest EQ will have low frustration there is threshold value set that is 0.5 the learner having the value higher than 0.5 is frustrated and the learner having the value less than 0.5 then learned not frustrated. The result shows that there were 13 to 0 students frustrated in OOP lab.

J. Octavia [9] There are the users suffering from the frustration during interaction in virtual environments in this paper they have developed a module that will continuously capture the measure of the user and detect the frustration in real time. When a user is frustrated the feedback is taken by visual, haptic, and multimodal. This paper results in the frustration level based on the pressure released for the 3D input device. It also presents the different pattern observed by the user frustration states

F. Leong [10] Detecting the frustration level of the student is very important nowadays otherwise the student will lose interest in studies so it is very important to detect the frustration. In these paper contextual and keystroke featuresare used in the Java tutoring system to detect the frustration there are some psychological sensors is used to detect the contextual and keystroke logs. The logistic regression with lasso regularization is used to prevent over-fitting. The result of this module the accuracy rate level is 0.67. This paper plays an important role to detect the frustration level of the student.

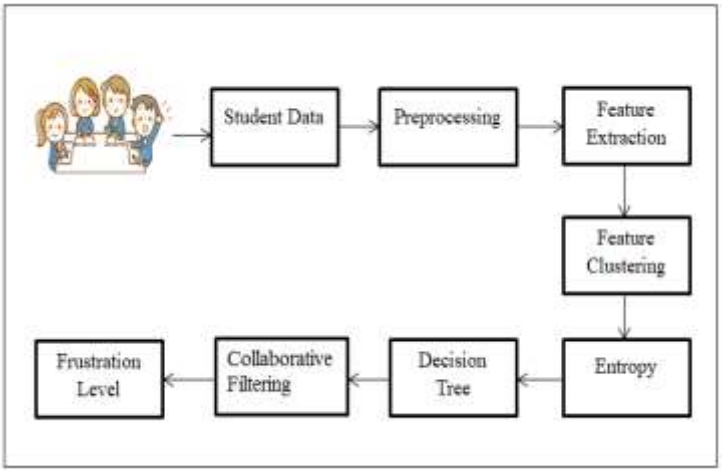
Haemal Naik [11] In this paper there is research made on Industrial Augmented Reality in this industry, there are modifications made on application in the industrial workflows In this poster, we propose a resourceful solution to achieve end-to-end workflow integration with minimum effort from the user end. PPF approach is preferred because it provides reasonable initialization to target specific regions of the object for closed loop pose refinement. This is a first attempt to understand an existing concept of mechanical alignment and replicate them to the optical tracking domain.

A.Alabdulkarim [12] There are has been rapid growth in the last decade in the field of technology to understand the emotion of the user to see whether he is happy or sad for the better solutions of the user. Thus, in this paper, they have used a hand gesture to find whether the user is frustrated or not. They can also assume whether it is positive or negative instances, thus to have the most accurate results. Leading to better interactive systems, that has the ability to understand the user's emotional state.

M. Brecht [13] Understanding the students problem is one of the important parts of nowadays if they are frustrated they may lose interest in their academics and they may be higher dropouts rate in coming time, so it is very important to detect the frustrate level of the student and if there frustrate them to give the solution on this by giving a motivational speech or by showing the motivational videos to reduce their frustration Thus frustration level can be check by using the data mining approach.

III PROPOSED METHODOLOGY

Figure 1: Overview of the Student frustration Level measurement system



The proposed system of student's frustration level estimation is described with the below mentioned steps.

Step 1: Student Data collection - An interactive user interface is created for the student to give the 4 types of tests like Personality Test, Academic Test, General knowledge Test and cognitive ability Test .

To provide these tests, model is equipped with the 100 questions in the database. Whenever a student is ready to give the test, then proposed model provides 10 random questions along with their optional answers to the student. Once a student is given all the tests, then the scores of these tests are stored in the database for the further evaluation of the frustration level.

Step 2:Preprocssing and Feature extraction - Once these tests are finished, then all the data from the database is retrieved to get the important features like scores of Personality Test, Academic Test, General knowledge Test and cognitive ability Test into a preprocessed list, that will be used further for clustering process.

Step 3:Feature Clustering - To cluster the preprocessed list K-nearest neighbor algorithm is used. Here each row of the preprocessed list is considered where it contains the scores of the 4 tests. Then for each of this row a Euclidean distance is created with respect to all other rows of the preprocessed list. The average of all of this row Euclidean distance yields a Euclidean distance of the complete preprocessed list called as P_{ED}.

All the rows of the preprocessed list are appended with their respective row distances. Based on these row distances the preprocessed list is sorted in ascending order. Random data points are allocated based on the requirement of the number of clusters. Then centroids are formed based on the data points of the sorted list. These centroids are used to form the boundaries of the cluster using the $P_{ED.}$ The boundaries are used to form the clusters which contain the nearest data.

Step 4: Entropy Estimation - Once the clusters are formed, these clusters are used to analyze the entropy of the current test of the student. To evaluate the entropy Each and every row is matched for its 4 scores with the current scores of the student to make a count.

Then this count is used to evaluate the entropy or distribution factor for each of the rows of a cluster using the equation 1. Here entropy of a row lies in between 0 and 1. Any row which yields the entropy more than 0.5 is selected to form the entropy clusters which is used for the next level of the process .

$$E = \frac{A}{C} log \frac{A}{C} \frac{B}{C} log \frac{B}{C} \quad \text{_____(1)}$$

Where

C= 4 (As there are 4 tests are there)
A= matched score count
B= C-A
E = Entropy Gain

Step 5 : Decision Tree - Each and every row of the entropy clusters are appended with the entropy of the respective rows. These entropy of all the rows are summed to get the mean Entropy of the particular cluster. So the cluster, which is having least mean is decided to select for the collaborative filtering process. As the cluster with the smallest entropy is having the highest resemblance for the current test scores.

Step 6 : Collaborative Filtering - Here the current score set is tend to select the most resembled scores for the individual tests from the selected cluster of the decision tree step. This process can be shown in algorithm 1.

Algorithm 1: Collaborative Filtering

// Input : Current Score Set $C_{SET} = \{P_S , C_S, A_S, G_S \}$
[P_S : Personality Test Score, C_S : Cognitive Test Score, A_S : Academic Test Score , G_S : General Knowledge Test Score]
// DT_C ; Decision Tree Cluster
// Output : Recommended Score Set R_{SET}
Function : cf($C_{SET,}$ DT_C)
Step 0: Start
Step 1: ***for*** i=0 **to** size of C_{SET}
Step 2: MIN_{SCORE} =10
Step 3: $S_G = C_{SETi}$
Step 4: ***for*** j=0 **to** size of DT_C
Step 5: ROW= $DT_{C\,J}$
Step 6: DIST= ROW_i - S_G
Step 7: **If**(DIST < MIN_{SCORE}), THEN
Step 8: $MIN_{SCORE=}$DIST
Step 9: **End *for***
Step 10: $R_{SET=}$ R_{SET+} MIN_{SCORE}
Step 11: **End *for***
Step 12: return R_{SET}
Step 13: Stop

The obtained current scores are matched with the past score of the student to evaluate his/ her frustration level to provide the valid suggestions based on the table 1.

Test Type / Range	0 to 3	4 to 5	6 to 7	8 to 10
Personality Test	You can improve in communication skills.	Be a better listener and read more and expand your interests.	Have a positive outlook and attitude	Meet New People , Be Yourself and Be Fun and see the humorous side of life.
Cognitive Ability	Exercise daily and follow meditation	Need to work on analytical skills. Practice atleast 100 questions daily from RS Aggarwal	Play the brain game Dual N-Back. Do this 20 minutes a day. Solve crosswords and game puzzles.	Practice puzzles from TOKETA
Academics Test	Clear your basics of C, Java, DBMS, OS. Read Let Us C for basics. Read Balguruswamy to improve basics of C++ and Java	Improve C programming skills by practicing atleast 20 programs daily.	You are good at C. You need to clear your basics of DBMS. Read Navathe to improve your database skills.	You are very good at C, Java and DBMS. You need to work more on OS. Read TENANBAUM. To improve on Os.
GK	Study all the books from class 6–10 NCERT sincerely	Read "Times of India" and "The Hindu" daily	Read Lucent General Knowledge	Refer Affair Cloud and GradeUp

Table 1: Suggestion Table

IV RESULT AND DISCUSSIONS

The proposed methodology of Frustration level estimation using machine learning is deployed using the Java programming language. And the model uses the Netbeans 8.0 as development IDE and Mysql as a database server.To measure the effectiveness of the system some experiments are carried out as described below.

RMSE (Root mean square Error) is a measurement protocol between the expected outcomes and obtained outcomes. This is used to measure the expected improvement in frustration level and obtained improvement in frustration level. This can be represented by the following equation.

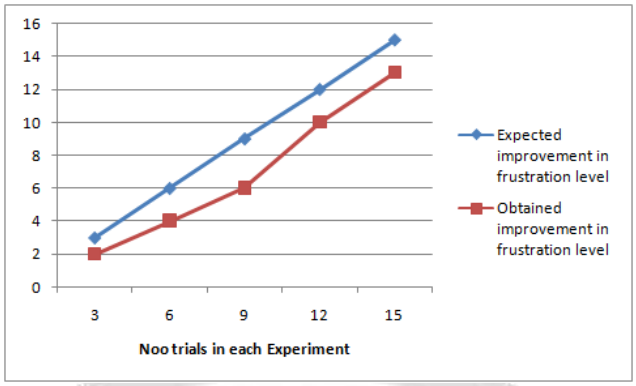
$$\mathrm{RMSE}_{fo} = [\sum_{i=1}^{N} (z_{f_i} - z_{o_i})^2 / N]^{1/2}$$

Where
∑ - Summation
$(Zfi - Z_{oi})^2$ - Differences Squared
N - Number of samples or Trails

Experiment No	No of trails	Expected improvement in frustration level	Obtained improvement in frustration level	MSE
1	3	3	2	1
2	6	6	4	4
3	9	9	6	9
4	12	12	10	4
5	15	15	13	4

Table 1: Mean Square Error of the Conducted Experiement for RMSE

Figure 2: Plot for Expected outcome and



Obtained Outcome

The experiment is conducted to measure the RMSE and obtained results are tabulated in the table 1. The Average MSE of the of Table 1 is 4.4 and RMSE is estimated as 2.097618. Any RMSE of this Range is always good for any kind of measurement models. And hence this experiment indicates that, the proposed model yields much effective result in the very first attempt regarding this research and it is very good sign of the Machine learning models.

IV CONCLUSION AND FUTURE SCOPE

This research article has dealt with various issues that are faced by the student in their respective academic lives. Because of this, students often get frustrated and that leads to the bad effect on their academic career. So this model conducts some tests for the student like GK, Congnitive Ability, Academic and Personality Tests. Based on the present and past test score proposed model learns about the ability of the student to handle the things using the K Nearest neighbor algorithm. Which eventually brings the all possible past values of the student to estimate the Frustration level. And the Decision tree and Collaborative Filtering helps to make a proper decision about the student frustration which is measured using RMSE. The proposed system yields a good RMSE of 2.097618, Which is a good sign of the Machine Learning Model.

In the future this system can be enhanced to work with all possible parameters of the student like Financial health, Family background, Residing premise and early education. And this type of the system can be developed as a secure mobile application too.

REFERENCES

[1] Ramkumar Rajendran," Student Model in an Intelligent Tutoring System.", IEEE Transactions on Learning Technologies, Submitted on Jan 2013.

[2] Ramkumar Rajendran, Sridhar Iyer, Sahana Murthy, Campbell Wilson, Judithe Sheard, "A Theory-Driven Approach to Predict Frustration in an ITS," IEEE transactions on learning technologies, vol. xx, no. y, month year

[3] Edécio Fernando Iepsen, Magda Bercht," Detection and Assistance to Students Who Show Frustration in Learning of Algorithms" 978-1-4673-5261-1/13/$31.00 ©2013 IEEE

[4] SADIQUE AHMAD, KAN LI, ADNAN AMIN AND SALABAT KHAN," A Novel Technique for the Evaluation of Posterior Probabilities of Student Cognitive Skills" Digital Object Identifier 10.1109/ACCESS.2017.DOI

[5] Steve Warren and Dwight Day," The Impact of a Scientific Computing Prerequisite on Student Performance in a Linear Systems", 978-1-4673-1352-0/12/$31.00 ©2012 IEEE

[6] Jonathan Lazara, Adam Jonesa, Mary Hackleya, Ben Shneiderman," Severity and impact of computer user frustration: A comparison of student and workplace users" Interacting with Computers 18 (2006) 187–207

[7] Lim Eng Aik Tan Wee Choon," Effects of Traffic Speed Level on Frustration Index in Cellular Automata Traffic Simulation", Journal of Advanced Transportation, Vol. 39, pp. 323-349, 2004.

[8] Prathmi Pingu Kurtiker, and Ramrao S. Wagh," Monitoring Student Frustration Level using Blackbox Logs" 2018 IEEE 18th International Conference on Advanced Learning Technologies 2161-377X/

[9] Johanna Renny, Octavia Karin,and Coninx Kris Luyten," Squeeze Me and I'll Change: An Exploration of Frustration-triggered Adaptation for Multimodal Interaction" IEEE Symposium on 3D User Interfaces 2011

[10] Fwa Hua Leong," Automatic Detection of Frustration of NoviceProgrammers from Contextual And", The 10th International Conference on Computer Science & Education (ICCSE 2015) July 22-24, 2015. Fitzwilliam College, Cambridge University, UK.

[11] Hemal Naik, MahmoudBahaa andFederico Tombari," Frustration-Free Pose Computation for Spatial AR Devices in Industrial Scenario", IEEE International Symposium on Mixed and Augmented Reality Ad, 2016

[12] Arwa Alabdulkarim," Towards Hand-Gesture Frustration Detection in Interactive Systems", 2014 3rd International Conference on User Science and Engineering (i-USEr)

[13] Edécio Fernando Iepsen Magda Bercht," Detection and Assistance to Students Who ShowFrustration in Learning", 978-1-4673-5261-1/13/$31.00 ©2013 IEEE

*****