

Student's Performance Analysis Using Machine Learning Tools

L.NAVEEN , PG Scholar,ASCET,Gudur

Email: naveenleburu@gmail.com

Dr.P.CHANDRAKANTH Assoc Prof Dept of CSE,ASCET,Gudur

Email: chandrakanthc4u@gmail.com

Technology in today's world has reached to extent that it can be used to do various task in day to day life easily with less effort and time. There are lot of research work going to enhance the Learning Management system. World has realized importance of education in one's life which has led to revolution in field of education like universities and colleges. Nowadays colleges have loads of task to be completed in given timeline. In today's scenario colleges needs to analyse student results manually before entering the students into the university exam so that the performance in the university exams may not get affected. This process takes lots of time and effort by faculties who need to analyse the results of the students individually. Hence in order to simplify this tasks using big data analysis system is introduced which can perform student result analysis. This system takes file of student results obtained in university exam in data sheet format as an input. The data is loaded and trained by machine using algorithm. Based on the previous year data the new data will be predicted. The future results of analysis will then be displayed according to student performance. This proposed system shows a typology of learning management system.

Keywords— LMS, Classification, Association rule.

I. INTRODUCTION

There has been an exponential increase in the amount of data stored in a data warehouse and real-time data which spawn the phenomenon of Big Data over the past decade. Big data as anticipated by can be defined by these three key concepts which are volume, velocity and variety. Moreover, the knowledge of big data is not just about the size, the speed and the variation but a concept in which it can provide new insights as well as future guidelines to overcome challenges through analysis methods under in big data creation. Data creation in the field of education is emerging as a new direction to explore and developed the current learning environment especially in eLearning where data can be collected digitally to understand students's performance. The eLearning systems can collect an immense amount of data over a period of time from students activities involving background studies, online tasks, various logs, and course performance which are valuable for evaluating the effectiveness of the courses offered and student's overall learning ability. The analyzed data can help the development of eLearning overall and increase its effectiveness and provide a competent online learning atmosphere. Analyzing the eLearning data can be quite difficult due to its complexity and finding the relationship between different information. Therefore, applying the right technique such as association

rule can reveal certain patterns. All things considered, the importance of big data analytic is to realize the hidden patterns, the co-occurrence between data attributes. In spite of the fact that the infrequent item sets can be considered as outliers or uninteresting, at times they are valuable especially those that resemble in negative associations, for example, in the field of healthcare, security and education. In this research work, the association rule applying on eLearning data such as Apriori and Predictive Apriori algorithms are employed to evaluate their performances, namely time consumption, rules generation and accuracy. As well as a number of experiments with different configurations will be conducted to compare their interestingness. The results from each association rules generate will be inspected to generalize the behavior of students enrolled in eLearning courses for future improvement.

II. LITERATURE SURVEY

Course Recommender System in E-Learning

Now a day E-learning is becoming popular. E-learning is essentially the computer and network-enabled transfer of knowledge & skill. E-learning is also referred to as web-based education and e-teaching. Some examples of commercial systems are Blackboard, WebCT and Top-Class while some examples of free systems are Moodle, Ilias and Claroline. In this paper, we propose the architecture for Course Recommender System & how the data flows through this system. This system predicts the best combination of subjects i.e. the subjects in which students are more interested. Here we use the learning management system such as Moodle to collect the data from student regarding their course choices.

Apriori Based: Mining Infrequent and Non-Present Item Sets from Transactional Data Bases

Item set mining has been an active area of research due to its successful application in various data mining scenarios including finding association rules. Though most of the past work has been on finding frequent item sets, infrequent item set mining has demonstrated its utility in web mining, bioinformatics and other fields. In this paper, we propose a new method based on Apriori algorithm to find infrequent item sets and non-present item sets. Finally, we analyze the behavior of our proposed method by considering a transactional data base.

Mining Association Rules between Sets Of Items In Large Databases, SIGMOD Conference

We are given a large database of customer transactions. Each transaction consists of items purchased by a customer in a visit. We present an efficient algorithm that generates all significant association rules between items in the database. The algorithm incorporates buffer management and novel estimation and pruning techniques. We also present results of applying this algorithm to sales data obtained from a large retailing company, which shows the effectiveness of the algorithm.

Finding Association Rules that Trade Support Optimally Against Confidence

When evaluating association rules, rules that differ in both support and confidence have to compare; a larger support has to be traded against a higher confidence. The solution which we propose for this problem is to maximize the expected accuracy that the association rule will have for future data. In a Bayesian framework, we determine the contributions of confidence and support to the expected accuracy on future data. We present a fast algorithm that finds the n best rules which maximize the resulting criterion. The algorithm dynamically prunes redundant rules and parts of the hypothesis space that cannot contain better solutions than the best ones found so far. We evaluate the performance of the algorithm (relative to the Apriori algorithm) on realistic knowledge discovery problems.

Association Rule Mining on Big Data – A Survey

Frequent pattern mining is the key concept in Association Rule Mining task. Main aim of frequent pattern mining is to find the recurrent patterns occurring in a dataset. Finding patterns identify the relationship between items in an item domain; these relationships are useful for strategic decision making. Data is flooded in a day to day life, called -Big Data, because massive amount of data is produced everywhere. Mining frequent patterns from the huge volumes of data has many challenges due to memory requirement, multiple data dimensions, heterogeneity of data and so on. The complexities related to mining frequent itemsets from a Big Data can be minimized by parallelizing the mining task with Map Reduce framework in Hadoop Cluster. In this paper, an introduction to Big Data, Association Rule Mining, concepts and basic methods for frequent pattern mining are given. The various methods proposed by different authors to mine frequent patterns from enormous dataset effectively are also discussed.

A Survey on Association Rule Mining

Apriori calculation has been basic calculation in association rule mining. Principle proposition of this calculation is to discover valuable examples between various arrangements of information. It is the least complex calculation yet having numerous downsides. Numerous specialists have been accomplished for the improvement of this calculation. This paper does a study on couple of good improved methodologies of Apriori calculation. This will be truly exceptionally supportive for the up and coming specialists to locate some new thoughts of this methodology.

Comparing Rule Measures for Predictive Association Rules," in Machine Learning

We study the predictive ability of some association rule measures typically used to assess descriptive interest. Such measures, namely conviction, lift and χ^2 are compared with confidence, Laplace, mutual information, cosine, Jaccard and ϕ -coefficient. As prediction models, we use sets of association rules. Classification is done by selecting the best rule, or by weighted voting. We performed an evaluation on 17 datasets with different characteristics and conclude that conviction is on average the best predictive measure to use in this setting. We also provide some meta-analysis insights for explaining the results.

Evaluating the performance of apriori and predictive apriori algorithm to find new association rules based on the statistical measures of datasets.

Recently, various advancements has emerged in the field of data mining. One of the hottest topic in this area is mining for association rules from the existing massive collection of datasets. The pattern obtained from these databases are used in various fields like super market sales-prediction, fraud detection and weather forecasting etc. So it is necessary that only strong rules are mined by using appropriate algorithm. In this paper, out of the various existing algorithms of association rule mining, two most important algorithm i.e. apriori and predictive apriori algorithm are chosen for experiment. Their performance is compared based on the interesting measures using weka3.7.5 which is a java based machine learning tool. After that, various statistical measures are calculated of different datasets and then based on the comparison of algorithms and statistical measures of data, new rules are generated using see5 tool.

Bayesian Estimation of Disclosure Risks for Multiply Imputed, Synthetic Data

Agencies seeking to disseminate public use microdata, i.e., data on individual records, can replace confidential values with multiple draws from statistical models estimated with the collected data. We present a framework for evaluating disclosure risks inherent in releasing multiply-imputed, synthetic data. The basic idea is to mimic an intruder who computes posterior distributions of confidential values given the released synthetic data and prior knowledge. We illustrate the methodology with artificial fully synthetic data and with partial synthesis of the Survey of Youth in Custody.

Implementation of Apriori Algorithm using WEKA

Knowledge exploration from the large set of data, generated as a result of the various data processing activities due to data mining only. Frequent Pattern Mining is a very important undertaking in data mining. Apriori approach applied to generate frequent item set generally espouse candidate generation and pruning techniques for the satisfaction of the desired objective. This paper shows how the different approaches achieve the objective of frequent mining along with the complexities required to perform the job. This paper demonstrates the use of WEKA tool for association rule mining using Apriori algorithm.

IV. PROBLEM DESCRIPTION

Existing method

In Existing process the students results are analysed manually by the faculties of the corresponding institution. Application of association rule after collecting student’s data for years from Learning Management System (LMS) has been investigated. Interestingness metrics and other relevant student preferences are captured after considering several years of data of the students such as marks obtained. Some visualized presentations of rules and their relevant results will be demonstrated in terms of performance metrics and their suitability in eLearning environments.

Demerits

The use of linear search in file handling might increase time complexity and no advanced feature benefit can be used in this application. Also updating of data is very difficult as all data needs to revise again. Predicting the results are very difficult. Staff work load will be increased.

Proposed method

Student Result Analysis System is a student driven system that is designed in order predict student results by use of machine learning algorithm. This proposed system takes file of student results obtained by universities in data sheet format as an input and the machine is trained with the previous year records and predict the forthcoming results. This system will manage Student result and keep track the large dataset in the data sheet format. The data set is loaded into the machine and the machine is trained with the large set of student records. The loaded data is pre-processed with the necessary details and the unwanted details are eliminated and then system is predicted the forthcoming results. The system is covered under all validation checks for authentication. Student data has been analysed based on various aspects like cycle test marks and model exam marks. This system which is helpful for the college authorities to predict the overall performance of the students. In this application we can analysis and predict the student result according to their semester exam and cycle test marks. The machine is trained with the previous year records and present records to predict the results by using machine learning algorithm. The data set used in this system is from the Learning Management System database, with the loaded data, the system will pre-process the data. System will be designed in such a way that it can be later updated to accept different format. For that all data will be stored in temporary storage first, and then formatted before inserting in database. The result of analysis will then be displayed in sorted manner using machine learning according to rank of students.

Merits

This system identify the gaps in existing prediction methods and the variables used in analyzing students’ performance. This system will Study the existing prediction methods for predicting students’ performance. Smart management of student’s dataset. Avid student performance. Streamline and simplify everyday tasks.

IV. METHODOLOGY

Sign-In Module

This module will provide user id and password for authentication. This module is to prevent data from hackers or data stealers. Admin and user module is allowed admin can update data set by providing valid admin id and password. After registration user will provided with user id and password to view and analyse result.

Forget password

If the user forget the password, the user can enter through this module by giving user id, name, old password and new password and code.

New registration

If the user is new to the system, he/she can enter through this module by providing user name, code and password.

Load analysis list

This module can load the student’s data from the dataset. The user can view the students details like name and marks obtained in previous years in the form of data sheet. The dataset are stored in data sheet

Pre processing

In this module the dataset from load analysis list is processed. The absence field of a particular student is marked by the average value of the overall students.

Feature selection

Feature selection is made in this module. The necessary field for calculation of marks are considered and unwanted field are discarded

Prediction

The data set is getting ready for the working process of j48 algorithm. The output of this module is not loaded, but the data is given to the j48 module. Which also used as another dataset for machine learning.

J48 algorithm

Result will be analysed in this module and make it as a tree structure. Tree will display the students predicted result. Which also used as another dataset for machine learning.

Clear dataset

The tree structure will be cleared, if we click this option all the data are cleared.no more testing can be performed. To perform testing we have to use j48 module for creating data set.

Multiple testing

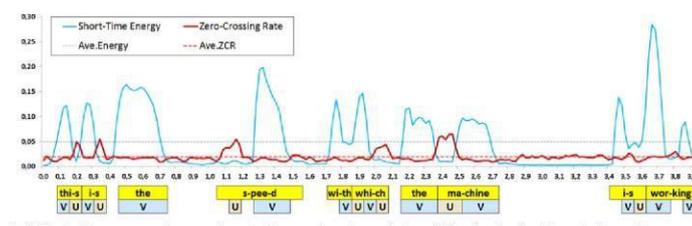
The overall data is loaded from the multiple data set by clicking the load option. After loading we can check the results of the students of the current year.

close

After performing all task in the panel, we can close the application by clicking this module.

V. EXPERIMENTAL RESULTS

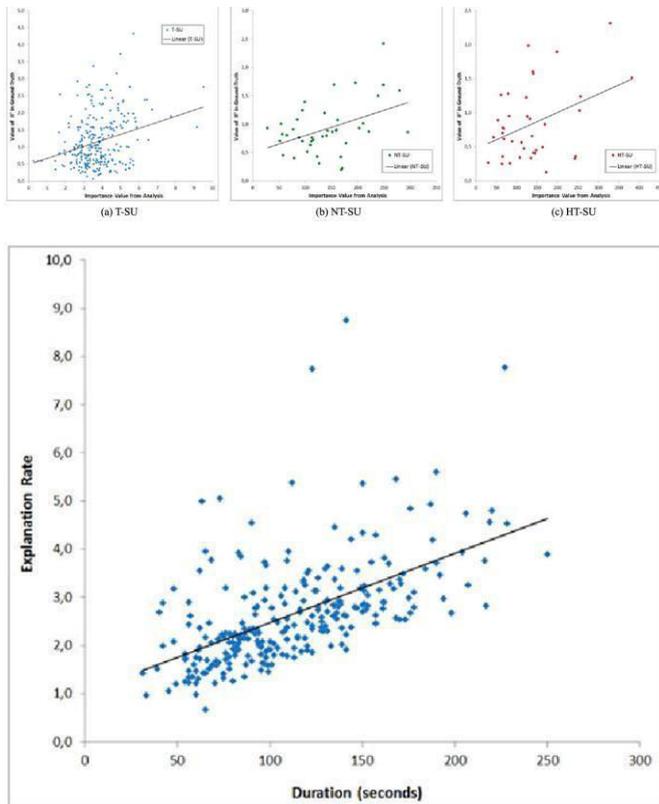
Method	All Sentences		Highlighted Sentences			
	Num	Ave	Num	Ave	Hit	Precision
ToneAnalyzer	124	0.86	22	0.86	8	36.4%
Vokaturi	124	0.86	22	1.11	12	54.5%
Proposed	124	0.86	22	1.14	14	63.6%



Type	All Segments		Selected Key Segments (<i>Top 1/6</i>)			
	Num	A-IF'	Num	A-IF'	Correct	Accuracy
T-SU	268	1.15	44	1.58	31	70.5%
NT-SU	42	0.82	7	1.40	5	71.4%
HT-SU	38	0.83	6	1.13	4	66.7%
All	348	-	57	-	40	70.2%

REFERENCES

1. Big Data Analytic Using Association Rule In ELearning, 2018- IEEE 3rd international conference on big data analytics by Korn Poonsirivomg, Albert Lawrence school of communication Thailand, Chantorn Jittawiriyakoon, graduate school of eLearning Thailand
2. D. Laney, "3D Data Management: Controlling Data Volume, Velocity, and Variety," META Group2001.
3. S. M. D. Aher and L. L.M.R.J., "COURSE RECOMMENDER SYSTEM IN E-LEARNING," International Journal of Computer Science and Communication, vol. 3, no. 1, pp. 159-164, 2012.
4. S. Kamepalli, R. Sekhara, R. Kurra, and S. Krishna, "Apriori Based: Mining Infrequent and Non-Present Item Sets from Transactional Data Bases," International Journal of Electrical & Computer Science IJECS-IJENS, vol. 14, no. 3, pp. 31-35, 2014.
5. R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," in SIGMOD '93 Proceedings of the 1993 ACM SIGMOD international conference on Management of data, Washington DC., USA, 1993, vol. 22, no. 2, pp.207-216.
6. T. Scheffer, "Finding Association Rules that Trade Support Optimally Against Confidence," in In: 5th European Conference on Principles of Data Mining and Knowledge Discovery PKDD2001, Freiburg, Germany, 2001, vol. 2168, pp. 424-435: Springer, Berlin, Heidelberg.
7. R. Nedunchezian and K. Geethanandhini, "Association Rule Mining on Big Data – A Survey," International Journal of Engineering Research & Technology (IJERT), vol. 5, no. 5, pp. 42-46, 2016.
8. T. Karthikeyan and N. Ravikumar, "A Survey on Association Rule Mining," International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, no. 1, 2014.
9. P. J. Azevedo and A. M. Jorge, "Comparing Rule Measures for Predictive Association Rules," in Machine Learning: ECML 2007, 18th European Conference on Machine Learning, Warsaw, Poland, 2007, pp. 510 - 517
10. J. P. Reiter, Q. Wang, and B. Zhang, "Bayesian Estimation of Disclosure Risks for Multiply Imputed, Synthetic Data," Journal of Privacy and Confidentiality, vol. 6, no. 1, pp. 17-33, 2014.
- A. K. Shrivastava and R. N. Panda, "Implementation of Apriori Algorithm using WEKA," KIET International Journal of Intelligent Computing and Informatics, vol. 1, no. 1, pp. 12-15, 2014.



VI. CONCLUSION AND FUTURE WORK

Result Analysis system is an integrated application to automate the result predicting procedure of an educational institution. It ensures that all the processes are carried out in time with no scope of human induced error. In this research the exploration of machine learning algorithm on e-learning data has been proposed as a useful method to compare student's performance as well as supporting certain prediction. The performance evaluation of machine learning algorithm is also compared to determine the effectiveness for the task. This prediction can benefit teacher and course planner learns how to increase student's results in the university exam. This paper has reviewed previous studies on predicting students _performance with various analytical methods. In this research I have used cycle test mark, model mark as data sets. While for prediction techniques, the classification method is frequently used in educational data mining area. Under the classification techniques, Neural Network and Decision Tree are the two methods highly used by the researchers for predicting students' performance. There is a future scope of this system that this system can be implemented as a web based application and android application so that the many of the institution may get benefits by using this system and many more features can be added like school related data.