Engineering Universe for Scientific Research and Management
(International Journal)

EUSRM A Universe of Research ISSN(ONLINE) : 2319-3069

Vol. 7 Issue 6
June 2015

The Hybrid Approach to Classify Student's learning Experience using Fuzzy-Apriori & CART

Abhay Narayan Singh[1], Anurag Jain[2], Himanshu Yadav[3]
Computer Science & Engineering
Radharaman Institute of Technology & Science, Bhopal
narayan.upc@gmail.com[1], anurag.akjain@gmail.com[2], himanshuyadav86@gmail.com[3]

Abstract

There are various techniques implemented that are used for the mining of Online Social Media are discussed and their applications such as Community detection and Message Wall Filtering and Student are learning experience and their opinions. Since Online Social Media seems to be advantageous and its usage by various users over internet may create some harm to the other users. Here a new and efficient technique is implemented which is a combinatorial method of applying Fuzzy-Apriori algorithm and then applying CART algorithm for the classification of student's learning experience.
The methodology adopted here for the classification of student's learning experience on Online Social Network such as Facebook and twitter provides efficient results and better classification as compared to existing technique implemented.
The proposed methodology implemented here works in two stages, in the first stage Fuzzy-Apriori is applied in the social media dataset for the generation of candidate sets and association rules and these rules are then classified using Classification and Regression Tree.
The experimental results are performed on facebook dataset and it provides better accuracy of classification.

Keywords: *Online Social Media, Facebook, Apriori, Frequent Sets, Association rules, CART, Classification.*

1. Introduction

Many educational technology researchers leverage social media data to answer questions about trends, collaboration or learning networks. Educational data mining is emerging as a research area with a suite of computational and psychological methods and research approaches for understanding how students learn.

New computer-supported interactive learning methods and tools—intelligent teaching methods, models, and games—have released up occasions to collect and evaluate student data, to determine models and developments in those data, and to formulate new detection and analysis hypotheses about how students be trained. Data accumulated from online learning methods can be aggregated over large numbers of students and can contain many variables that data mining algorithms can explore for model structure. Just as with early on attempts to recognize online performances, before time endeavors at learning data mining involved mining website log data [1], but now more integrated, mechanism, and complicated online learning methods provide more varieties of data. Educational data mining commonly underlines reducing learning into small components that can be analyzed and then influenced by software that adapts to the student [2].

Student learning data collected by online learning systems are being explored to develop predictive models by applying educational data mining methods that classify data or discover connections. These representations contributed a means responsibility in structure adaptive learning systems in which adaptations or interventions based on the model's predictions can be used to change what students experience next or even to recommend outside academic services to support their learning. An important and unique feature of educational data is that they are hierarchical. Data at the keystroke stage, the response stage, the session stage, the student stage, the classroom stage, the teacher level, and the school level are nested surrounded by one another [3], [4].

Other important characteristics are time, progression, and circumstance. Time is significant role to confine data, for example length of perform gatherings or time to become skilled at series characterizes how ideas put together on one another and how practice and tutoring should be categorized. Context is significant for give explanation consequences and knowing where a model may or may not effort. Techniques for hierarchical data mining and longitudinal data modeling have been important developments in mining educational data.

2. Theoretical Concept

Over the last few decades the growth of Information and Communication Technologies ICT has evolved rapidly and brought forth a plethora of new services. However, the theme of data delivery using the client-server models has not changed. The most popular means of data release, at a distance from e-mail has been the World Wide Web (WWW). With the arrival of Web 2.0, content presentation has become richer and brought about the genesis of a large range of tools that enable content generation, adaptation as well as delivery [5]. Several tools, both general purpose (databases, blogs, Wikis, etc.) as well as specific (Moodle, Blackboard, etc.), are used in the context of e-learning. Mutually, as communications, they make available the overall functionality of e-learning [6]. Each of these tools generates a log of user activities as a part of their operation.

These logs are cumulative and it is possible to correlate activities of a single user across these functional blocks to assess the utilization or extract resource right of entry developments [5]. Ahead of this, the opportunity of removing information to provide a feedback to the overall educational process to make it dynamic and adaptive is exciting. It is in this context of learning that two new but similar streams of research have emerged in the recent past, Educational Data Mining (EDM) [7] and learning analytics [8]. This article aims to investigate the role played by data mining techniques in e-learning particularly focusing on Educational Data Mining methods.

3. Literature Survey

In this paper the proposed [9] method was well again put in ordered; as it appended two research questions to the direct learning and a more concrete idealistic environment. One of the most significant transforms done in the method was to enhanced the complete structure it by have access to the straight forward structure of information competencies, with the intention of enhanced make a distinction literacy proficiencies the students may be using per period of the learning interferences. In this way, it is waiting for that the responsibility of student’s literacy’s in this type of learning would materialize and would be probable to improve investigated in it.

The research endeavors were communicated to influential: important concerns, confronts and opening promising from the combination of social networks in a higher education learning situation; how students understanding learning under such circumstances; to resolve if their literacy’s have an effect on the way they features this type of learning understanding; and how social networks and learning knowledge’s can be equally profile one another. The contestants of this learning were comprehensive profits of worldwide master lessons, divided in four teams. All students were enthusiastically involved in all the behavior and blocked up the investigation and the surveys; the applicant’s discussion was four team leaders, selected by their own teams. The mechanisms used for the collected works of data were: a diagnostic examination, online and offline discussions, student’s reports, feedback forms and semi structured conferences. The technique for investigating data was substance analysis and so a structure of grouping could be generated to make available, categorize and present the data.

Even though expression with a significant constraint [9], as it is the loss of the support the learning has in the period of its pilot revise, it is unmoving enduring. This necessitated a reorganizing of the method, on how to accomplish the learning on the minimum amount of time potential and at the same time, to assemble significant and significant results. This limitation distinguished a face up to moreover for the reason that improve has to be accomplished in disconnection of the tracks the students is attractive. On the other hand and optimistically, the students would be able to connect the dots and see the association between what they are knowledge in their routes and inside this knowledge, with the objective of enhanced make offered the investigator with some having an important effect approaching into the convenience of this instrument in higher education.

Students' informal conversations on social media (e.g., Twitter, Facebook) shed light into their educational knowledge’s-estimations, reactions, and apprehensions on the subject of the education development. Data from such un-instrumented situations can provide valuable knowledge to inform student learning. Analyzing such data, however, can be difficult. The difficulty of students' understandings reflected from social media content requires human interpretation. However, the growing scale of data demands automatic data analysis techniques. In this paper [10], we developed a workflow to integrate both qualitative analysis and large-scale data mining techniques. Here they focused on engineering student’s Twitter posts to understand issues and problems in their educational understandings. We earliest accomplished a

qualitative investigation on samples taken from concerning 25,000 tweets communicated to engineering student's college life. We found engineering students encounter problems such as heavy study load, lack of social engagement, and sleep deprivation. Based on these results, we implemented a multi-label classification algorithm to classify tweets reflecting students' problems. We then used the algorithm to train a detector of student problems from about 35,000 tweets streamed at the geo-location of Purdue University. This work, for the first time, presents a methodology and results that show how informal social media data can provide insights into students' experiences.
In this learning methodology [10] is helpful to researchers in learning analytics enlightening data mining and learning knowledge. It make available a workflow for investigating social media data for educational principles that defeats the most important constraints of both instruction manual qualitative analysis and big scale computational analysis of user created textual substance. Our learning can notify educational administrator's practioners and other significant pronouncement manufacturers to gain additional sympathetic of engineering student's college knowledge's.

In this paper [11], author here to show one potential approach of analyzing social media alteration data with the intention of enhanced appreciate customers. Finally, our objective is to evaluate customer performance as it is communicated in free-form discussions and take out from it profitable precious information about the customer. In this study, the main focus on author has deliberated on using statistical methods for analyzing this unstructured data at two levels: 1) at the level of the words used in the discussion and 2) by mapping those words to abstract conceptions. The objective of such a statistical analysis is twofold. At initially, the numerically important expressions used by the users and the ideas connected with them provide approaching on a user's concentrations that money-making examinations can use, for example with the purpose of objective announcements. Additionally, be acquainted with the development of a customer's concentrations and leisure pursuit can be developed money-making by put on the markets, media and diversion companies, telecommunications companies, and lot more companies. In this paper, author describes a common structure for the investigation of social media data and, consecutively the application of the structure to the statistical analysis of the verbal communication of tweets.

This paper shows that [12], within this exacting illustration, academic Facebook activity and the frequency of scrupulous academic-related topics vary at certain points in the academic semester on an surveillance of 70 university students' use of their personal social network site (SNS), Facebook, over a 22-week university study stage. The study required to decide the amount that university students use their personal SNSs to sustain learning by searching frequencies of academic-related substance and topics being argued. As technology is increasingly ever-present, it shows that some students may be leaving suggestions of their university experience in their personal online freedoms and that these traces may present valuable approaching into understanding students' learning processes, their knowledge's and interactions in social gaps.

The discovering recommended that many students nowadays may be leaving traces of their academic journey online and that academics should be responsive that these interfaces may also continue living in their own students' online public gaps. Technologies [12], and particularly SNSs, obtain a negative description in the media for being disorderly to learning, which brings us to ask whether students' use increases in occurrence because students are adjourned during study or whether students enhance use of SNSs during these eras to support learning. The decisions from this study call for a need to additional scrutinize how students use SNSs, mainly around significant points in the academic study era, and how SNS supports or entertains students from learning, in addition to the amount to which universities should or can connect SNSs to get better the student knowledge practices robust in relative to other communication techniques and the student experience can control SNSs.

4. Proposed Methodology

Here the proposed methodology is based on the combinatorial method of rules generation and classification. The proposed methodology works in the following phases:

1. Take an input dataset of facebook or twitter
2. Input Support & Confidence for the Apriori to generate Candidate Sets and rules.
3. For ($k=1; L_k \neq \emptyset ; k{+}{+}$) do begin
4. C_{k+1}=candidate generated from L_k
5. For each transaction t in database do
6. Increment the count of all candidates in C_{k+1} that are contained in t
7. L_{k+1}=candidate in C_{k+1} with min_support
8. End

9. Return U_kL_k
10. Traverse with each of the candidate sets and rules generated from Apriori.
11. For each item set L_k to P in Apriori.
12. If it is frequent (based on Count[L_k]) over the whole dataset E
13. Output (L_k)
14. Remove it
15. For each remaining item sets L_k
16. Identify constituent singletons

$$s1, s2, \ldots \ldots sm\ of\ Lk\ \forall\ Lk = s1 \cap s2 \ldots \ldots \cap sm$$

17. Calculate activation function for Lk using td[Lk]
18. Count[L_k] + = meu
19. If no item sets remain to be enumerated
20. Exit
21. Start at the root node of the candidate item sets
22. For each of the ordered variable X, convert it to an unordered variable X' by grouping its values in the node into a small number of intervals.
23. If X is unordered, set X'=X
24. Find the split set {$X^* \in S^*$} that minimizes the sum of small index and split the node into two child nodes.
25. Prune the tree with CART method

4.1 Dataset used

Here two types of dataset are used Twitter and Facebook dataset taken from UCI repository. Here the dataset is in ARFF (Attribute Relation File Format) version so that the input is directly proceeding on the algorithm. Since the base work is for infrequent item sets, here we take both frequent and infrequent and weighted and un-weighted datasets.

4.2 Apriori Algorithm

Given a transaction database DB and a minimum support threshold, the problem of finding the complete set of frequent patterns is called the frequent pattern mining problem.

Step 1: Build a compact data structure called the Apriori.

- Built using 2 passes over the data-set.

Step 2: Extracts frequent itemsets directly from the Apriori

Pass 1:

- Scan data and find support for each item.
- Discard infrequent items.
- Sort frequent items in decreasing order based on their support.
- Use this order when building the Apriori, so common prefixes can be shared.

Pass 2:

Nodes correspond to items and have a counter

1. Apriori reads 1 transaction at a time and maps it to a path
2. Fixed order is used, so paths can overlap when transactions share items (when they have the same prefix).

- In this case, counters are incremented

3. Pointers are maintained between nodes containing the same item, creating singly linked lists (dotted lines)

- The more paths that overlap, the higher the compression. Apriori may fit in memory.
- Frequent item sets extracted from the Apriori.

4.3 Fuzzy Over Apriori Algorithm

After the generation of rules from Apriori, Fuzzy is applied over these rules to generate minimum rules.

Fuzzy sets for quantitative attributes

It is composed of three steps:

Step 1: Transform the original database into positive integer

Step 2: For each attribute

- Cluster values of the attribute i^{th} into k medoids
- Classify the attribute i^{th} into k fuzzy sets
- Generate membership functions for each fuzzy set

End for

Step 3: Transform the database based on fuzzy sets

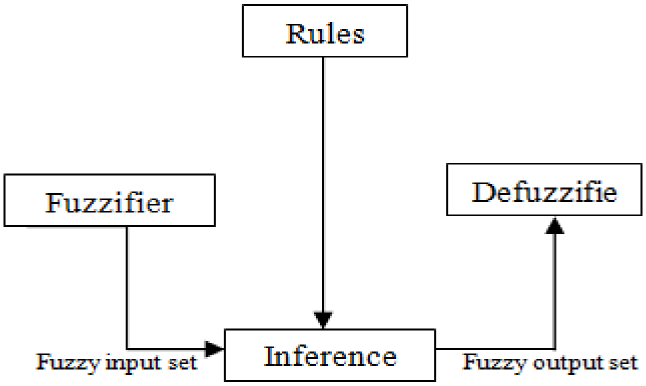

Figure 1. A Fuzzy Logic System

4.4 CART Algorithm

The algorithm is based on Classification and Regression Trees by Breiman et al (1984). A CART tree is a binary decision tree that is constructed by splitting a node into two child nodes repeatedly, beginning with the root node that contains the whole learning sample.

1. Find each predictor's best split.
2. Find the node's best split.
3. Among the best splits found in step 1, choose the one that maximizes the splitting criterion.
4. Split the node using its best split found in step 2 if the stopping rules are not satisfied.

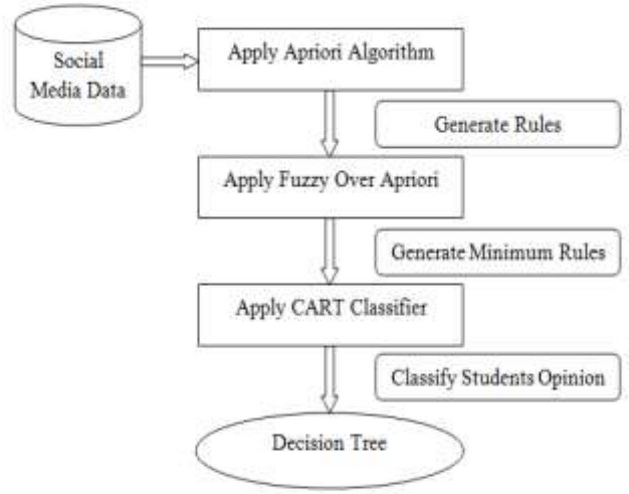

Figure 2. Flow chart of the methodology

5. Result Analysis

The table shown below is the analysis and comparison of existing and proposed work on various parameters such as True positive and true negative, precision and recall. The proposed methodology shows the performance over existing work.

Table 1. Analysis and comparison of Various Parameters

Measures	**Existing Work (Naïve Bayes Classifier)**	**Proposed Work (CART Classifier)**
True Positive	0.92	0.97
False Positive	0.45	0.23
True Negative	0.89	0.92
False Negative	0.37	0.23
Precision	0.93	0.96
Recall	0.94	0.97
F-Measure	0.934	0.965

$$\text{Precision} = \frac{\text{No. of Correctly Classified Instances}}{\text{Total No. of Instances Fetched}} \quad (1)$$

$$\text{Recall} = \frac{\text{No. of Correctly Classified Instances}}{\text{Total No. of Instances in the Dataset}} \quad (2)$$

$$\text{F} - \text{Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

TP – True positive is the number of instances correctly classified from the dataset.

TN – True Negative is the number of instances classified from the dataset that are not true.

FP – False Positive is the condition in which an instance detected is misclassified.

FN – False Negative is the condition in which an instance detected as negative is actually correctly classified

The Figure shown below is the analysis and comparison of existing and proposed work on various parameters such as True positive and true negative, precision and recall and F-

Measure. The proposed methodology shows the performance over existing work.

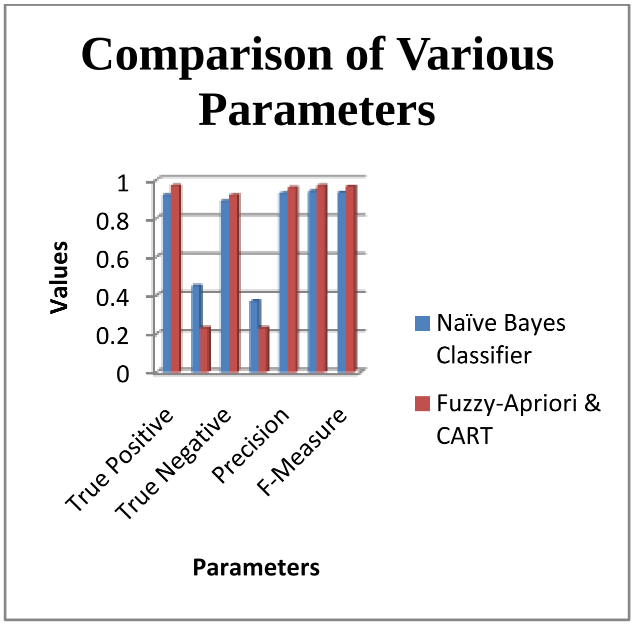

Figure 3. Performance Comparison of Various Parameters

The figure shown below is the analysis and comparison of Rules Generated from the existing and the proposed work.

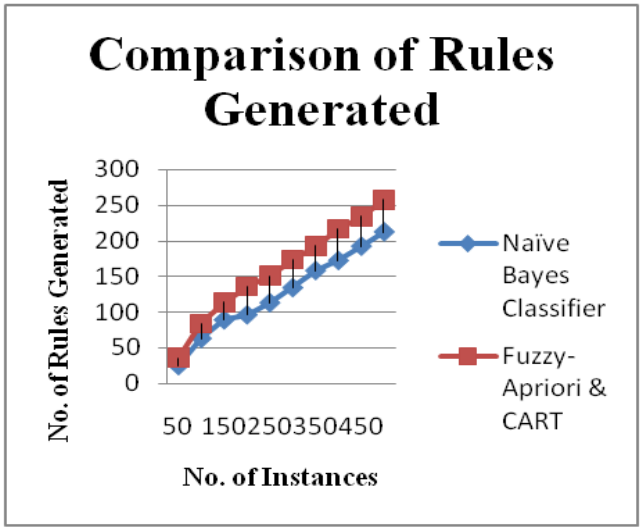

Figure 4. Comparison of Rules Generated

5. Conclusion

The proposed methodology implemented here for the classification of student's opinion using Decision tree provides efficient results as compared to the existing technique implemented for the student's opinion in Social media. The proposed methodology implemented here provides better classification as well provides reduces the chances of fake opinion of the student's.

The proposed methodology increases the accuracy of classifying student's learning experience in social media datasets.

Although the technique implemented here for the Classification of Student's learning experience using hybrid combination of Fuzzy-Apriori and CART is efficient but further enhancements can be done for the improvement of classification of other keywords.

The Methodology implemented here for the student's learning experience is efficient and provides better classification of student's views, but there are further enhancements which can be done for the improvement of classification using some supervised machine learning approach such that any type of data can be classified.

References

[1] Baker, R. S. J. D., and K. Yacef. 2009. "The State of Educational Data Mining in 2009: A Review and Future Visions." Journal of Educational Data Mining 1 (1): 3–17.

[2] Siemens, G., and R. S. J. d. Baker. 2012. "Learning Analytics and Educational Data Mining: Towards Communication and Collaboration." In Proceedings of LAK12: 2nd International Conference on Learning Analytics & Knowledge, New York, NY: Association for Computing Machinery, 252–254.

[3] Baker, R. S. J. d. 2011. "Data Mining for Education." In International Encyclopedia of Education, 3rd ed., edited by B. McGaw, P. Peterson, and E. Baker. Oxford, UK: Elsevier.

[4] Romero, C. R., and S. Ventura. 2010. "Educational Data Mining: A Review of the State of the Art." IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews 40 (6): 601–618.

[5] Clark, R. C. & Mayer, R. E., (2011). e-learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning, 3rd ed. Pfeiffer, San Francisco

[6] Osimo, D., (2008). Web 2.0 in Government: Why and How? JRC Scientific and Technical Reports, EUR 23358.

[7] [online], IEDM, (2012). International Educational Data Mining Society, Available at http://www.educationaldatamining.org

[8] [Online], Society for Learning Analytics Research,(2012). http://www.solaresearch.org/ .

[9] Juan Daniel Machin Mastromatteo, “The Mutual Shaping of Social Networks, Learning Experiences, and Literacies: The Methods Revisited” Qualitative and Quantitative Methods in Libraries (QQML) 2:195–205, 2013.

[10] Xin Chen. Vorvoreanu, M., Madhavan, Mining Social Media Data for Understanding Students’ Learning Experiences Learning Technologies, IEEE Transactions on Volume: 7, Issue: 3, 2014.

[11] Konopnicki, D. Shmueli-Scheuer, M., “A statistical approach to mining customers' conversational data from social media IBM Journal of Research and Development (Volume: 57, Issue: 3/4) May-July 2013.

[12] Rebecca Vivian, Alam Barnes, “The academic journey of university students on Facebook: an analysis of informal academic-related activity over a semester” Vol 22 - 2014.