Abstract

Association rule mining has been typically used in transaction databases to understand correlation between various items and/or events. This paper describes a work done in the education sector in India to understand association of various dimensions of behavior of school children and represents a case which has actually been executed successfully.

1. Introduction

Education institutes like schools are the nursery for future minds of a nation. Knowledge represents the “intangible assets of the organization” (Buckman, 2004; Rylatt, 2003). In a school, it is readily observed that knowledge is both the process and product of the organization. Knowledge and its management is the business of schools. It would be difficult to find disagreement with this assertion. However it has been found schools are notoriously poor with their own management processes. Majority of the school management faces common set of problems in their schools, which they would like to address. However, they lack understanding of advancement in the use of Information Systems for better decision making maintaining the status quo of these problems. For example, the school management would like to understand performance patterns of a particular student. Performance patterns would help in understanding a student’s strong and weak areas. Performance pattern study would help in taking necessary preventive steps,. School teachers and respective authorities frequently find various event correlations as a part of their daily experience. For example, a student who is good at languages will be good at literary activities like essay-writing, students who are good at mathematics may be good at arts or creativity, students who are good at sports may not good at studies etc. However, if there are inverse performance patterns in the same field under different situations by students, this would call for attention. For example, if a student is seen to be good in languages and not good at literary events, then she should be given extra care so that she can pick up in literary events. Students also have problems in deciding their career path based on their individual aptitude.

These problems were highlighted while discussing the use and applicability of integrated information systems with a day-boarding school named Anand Niketan at Ahmedabad, India. The need was getting clearly defined (as discussed later in this paper), we tried using a couple of data mining techniques for better understanding of few patterns being exhibited by the students. Extensive experiments are being conducted to come up with better ideas on the need and suitability of such systems in Indian schools. In this paper, we discuss the case on usage and suitability of these techniques at Anand Niketan school. One such application of Information System techniques has been adopted by the school authority as a part of their routine exercise. We conducted a survey among teachers, and school management to understand the suitability and usefulness of data mining techniques in school. We also present the result of these surveys in this paper. The rest of the paper is organized as follows: in section 2, we present the case of Anand Niketan and in section 3 glimpses of the technique used. Although we have used and continue to address a number of issues in the school using data mining, but in this paper, we limit discussing only one such application (Question #1 in table 1). In section 4, we discuss the results of experiment and present the findings of stakeholders survey in section 5. Section 6 presents the previous work done in this domain using the set of tools discussed in the paper and concluding with the road ahead.

2. Anand Niketan – A Day Boarding School in Ahmedabad, India

Anand Niketan, if translated in English, would mean an abode of joy. It is a day-boarding school at Ahmedabad and started in 1996 by an NGO group with only two classes and twenty students. The school management believes in the principle that, “children will start learning the moment they feel excited about it”. Hence, from the days of incubation, the philosophy of the school has been learning and not teaching. The school management believed in a simple philosophy that every child has the potential to grow into wonderful human beings, provided that they are given correct guidance. Through this, children can discover their own path and create destinies. With the practice of this philosophy, by the end of year 2005,, they had more than
800 children studying in classes from Nursery to Standard 12.

However, school management’s difficulties aggravated with rise in number of students and staff, along with increased competition from other schools in the locality. The school management thought of improving some of their core processes to enable better understanding of their students’ performance and providing individual attention to students. However, the teachers were in dilemma regarding monitoring students’ performance in non-academic activities, which remain ungraded. However, they considered these non-academic activities crucial for students overall growth. With these dilemmas in mind, the school management wanted to answer the following questions (as represented in table 1). It was exploring the possible usage of IT in solving these problems.

1. How fast can we understand the behavior of each student better than before?
2. How can we ensure quality of delivery of education to every child while teachers keep on changing?
3. How can we convert information about individual students into actionable knowledge?
4. How can IT be used for answering these questions as has been used in the corporate and scientific communities?
5. How can these analyses be done in minutes as compared to weeks and days?

Table 1 Questions being asked by the school management of Anand Niketan

Keeping these questions in mind, we approached the management of Anand Niketan about the data availability pertaining to each student’s behavioral aspects as well as their performance records to be made accessible on a computer. A system was developed by an external software services firm to routinely capture these data for six months. Thereafter, we decided to use a couple of data mining techniques to look at the impact on the problems which were tried to be solved. In this paper, we will discuss about the usage of association rule mining for students behavioral analysis.

3. Glimpses of the Data Mining Techniques Used

Although many detailed definitions of data mining exist in academic sources, an easy to understand definition of data mining is “the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories and by using pattern recognition technologies as well as statistical and mathematical techniques”. Data Mining is a multi-disciplinary field with many relevant techniques.

3.1 Association Rule Mining

Association rules are one of the techniques used in data mining, and particularly useful with e-commerce transactional information. Association rules describe events that tend to occur together. They are formal statements in the form of \( X \Rightarrow Y \), where if \( X \) happens, \( Y \) is likely to happen. Such a rule reveals that the transactions in the database, containing items in \( X \) tend to contain items in \( Y \), and the probability, measured as the fractions of the transactions containing \( X \) also containing \( Y \), is called the confidence of the rule. The support of the rule, is the fraction of the total transactions that contain all items both in \( X \) and \( Y \) (Han & Kamber, 2001). Out of the many techniques available, we have used the Apriori algorithm (Agarwal, Imielinski, & Swami, 1993) due to its simplicity and ease of implementation.

4. Data Mining Applications in Anand Niketan: Behavioral Analysis Using Patterns

At Anand Niketan, Apriori algorithm was used to conduct experiments for analyzing students behavioural aspects. It is worth mentioning at this point that 70 per cent of the teachers were unaware of data mining and its applicability in their context. In the following sub-sections, we will describe the applications in brief.

4.1. Behavioral Analysis using Association Rule Mining

One of the major concerns of the school management and the teachers was about getting a rough idea of the students behavioral patterns. This posed a problem as they are not captured in academic examination, albeit the school has a strong inclination towards students holistic development. The teachers had a guess of students behavioural aspects, but were non-confident due to lack of data. They lacked quantitative measure and no correlations existed to provide them with sufficient inputs on concerned students’ behavioral aspects. To get some clues towards the solution of this problem, we asked the school management and teachers to list types of behavioral aspects they think actually worth to look at. The teachers and concerned authorities of the school (with the help of an external behavioral psychoanalyst) came up with such a list of fifteen attributes, like attentiveness, cleanliness, class participation, discipline, cooperation etc. (a complete list of these factors is as provided in table 2, attributes have been ordered alphabetically and not on relevance, significance or importance).
Table 2 List of Behavioral Attributes that the School wanted to have a look at

Once the list of these attributes was finalized, we took some sample classes for experimentation. The teachers were asked to mark corresponding attributes for each student in the sampled classes for a particular term. After data was collected, we found attributes (in a particular class) where most of the students were strong (as shown in figure 1) and attributes in which the students are weak (as shown in figure 2). We applied Apriori algorithm to find correlation between these attributes to give the teachers and decision makers an idea about the correlation between parameters. The support and confidence threshold was varied to see the impact on the results. However, later it was decided that the support threshold be kept at 60% cut-off. 60% cut-off would give the decision makers insight about the patterns occurring at least in 60% of students, or majority of the class. The confidence threshold was kept at 40%. At the given thresholds, we found majority of the students of a particular class (with 28 students) were showing strength in behavioral attributes like class participation, cleanliness, cooperation, discipline, emotional stability and doing home assignments (as shown in figure 1). Whereas, majority of the students were weak in attributes like innovation, leadership, performance, self-confidence and presentation skills (as shown in figure 2). The correlation between these attributes in forms of association rules were also found and shown to the teachers. To keep things simpler for teachers, we had shown them the patterns of length two only. For example, we found that students, who are attentive in class, are also good in class participation and cooperation (as shown in figure 3). Similarly, rules for behavioral attributes where students were showing weakness were also found (figure 4).

These results were then easily mapped for individual student which helped the teachers to discuss about the next steps to overcome weakness areas. For example, the above figure indicates areas that needed attention were increasing innovativeness, leadership etc.. The school administration immediately took requisite steps to introduce necessary bridge-up courses and exercises so that these behavioral attributes may be attended. In areas of strength of a student as well this had significant importance. For e.g. if student is more attentive then there are high chances of that she must like cleanliness and like to participate in class activities. Whenever a teacher sees that a student is attentive but is not enough participative in class activities, with some simple remedial steps, the teacher can bring in the behavioral change..

Moreover, with the bar graph, related attributes can be easily referred by the teachers, which was of great help to them. In addition to this, the teachers actually cross checked these behaviors and said it to be correct (which is reflected in the survey presented in the following section). As supposed by the Management a regular exercise would strike two birds with one stone. First, it would help to gauge any improvements in students weak areas and secondly if teachers are giving enough attention to individual students to build up all aspects of a particular student.
5. Understanding Impact of Association Rules for Behavioral Attributes

To understand the impact and correctness of this exercise, we conducted a survey by administering a questionnaire to the school teachers and the principal at Anand Niketan. To have it cross checked, we also administered the questionnaire to the external consultant psychoanalyst. To save space, we have indicated the questions and their responses in Appendix only.

6. Conclusion & Future Work

Use of data mining techniques in schools and education institutes is increasing interest (Petrides & Guiney, 2002). The emerging interest in the application of data mining and related techniques in schools has been prominently stated in education literature as well (Kongshem, 1999; Streifer, 1999). Additionally, school districts are implementing technology applications at the administrative level as a means to execute data-based decision making that affects the organization and its students (Bushweller, 2000). Moreover, association rule mining has also been reported to be used for better marketing to target right school students (Ma, Liu, Wong, & Yu, 2000). (Luan, 2004) in his executive report also discusses the application of data mining techniques, for higher education, typically for the purposes of understanding student registration and forecasting credit hours etc. However, we have not found any references of using association rule mining techniques for analyzing behavioral patterns in schools. Experiments on using Data Mining techniques are also being carried out in Anand Niketan. For example, we are exploring the idea of using clustering techniques for coming up with an intelligent grading system. This will form clusters based
on total marks of students as well as subject wise for understanding and analysis of each student’s performance. The research is also planned for testing suitability of these techniques in appraisal of teachers, analyzing activities of students and expectations of parents, alumni interaction etc. We also plan to study the acceptability of the technology concerned by the end-users, their reactions and managing that transformation efficiently. Future work will also deal with replicating the study to rural and semi-urban schools. To study usage of these techniques and having an integrated school information system may promise to provide better education to students of these schools, otherwise poor in infrastructure.

8. Reference


7. Acknowledgement

The author is grateful to the teachers and the management of Anand Niketan Day Boarding School for all the support and help for this research. Special thanks to Mr. Fathepuria, Principal of Anand Niketan. Thanks are also due to Mr. Abhay Panjiar of Ceon Solutions for helping in the data collection.

Figure 4 Finding length two attribute-patterns where most of the Students (who are already weak and above support level) are also weaker (above confidence level)
## Appendix

<table>
<thead>
<tr>
<th>Question</th>
<th>No Reply (%)</th>
<th>Not at all (%)</th>
<th>Little (%)</th>
<th>Moderate (%)</th>
<th>High (%)</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 How much are you aware about the Association Rule Analysis to understand the Trend in Student's Behavior, after the experiments are conducted?</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>35</td>
<td>53</td>
<td>100</td>
</tr>
<tr>
<td>Q2 How, according to you, does Association Rule Analysis helps in finding Attitude and Trend of Every student in various attributes?</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>59</td>
<td>100</td>
</tr>
<tr>
<td>Q3 Does Association Rule Analysis help in understanding how many students are weak in various Attributes?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>88</td>
<td>100</td>
</tr>
<tr>
<td>Q4 Does Association Rule Analysis helps in understanding trends in Students' Behavior and finding Correlation between them?</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>18</td>
<td>65</td>
<td>100</td>
</tr>
<tr>
<td>Q5 Has Association Rule Analysis helps in identify Attributes which indicates that students are away from trend?</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>65</td>
<td>100</td>
</tr>
<tr>
<td>Q6 How much Datamining help in Education sector to quickly understand the trend in Performance and guide students on various aspects?</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>88</td>
<td>100</td>
</tr>
</tbody>
</table>