ASSOCIATION RULE AND THE EDUCATIONAL DATA

*Monika Srivastava
*Vijay Chaudhary
*Mohd. Iliyas Khan

Abstract

With the growing amount of data a new field called data mining is emerging zxtremely quickly. Data mining tools which perform data analysis may uncover important data patterns, contributing greatly to business strategies, knowledge bases, and scientific, educational and medical research. Association rules mining is one of the most well studied data mining tasks. It discovers relationships among attributes in databases, producing if-then statements concerning attribute values.

There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, concerns with developing methods that discover knowledge from data come from educational environments. This paper shows how data mining can be used to come up with interesting knowledge from student database. We identify the measure of interestingness and implement them on student database and come up with interesting rules that can help teachers and related people to deal with students and understand their behavior and activities.

Keywords data mining, educational data, association rule, support, confidence.

1. Introduction: The abundance of data generates the appearance of a new field named data mining. Data collected in large databases become raw material for these knowledge discovery techniques and mining tools for gold were necessary. The current expert system technologies, which typically rely on users or domain experts to manually, input knowledge into knowledge bases. This procedure contains errors, and it is extremely timeconsuming and costly. The definition from Gartner Group seems to be most comprehensive, as they define data mining as the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of datastored in repositories and by using pattern recognition technologies as well as statistical and mathematical techniques. Data mining is a field of increasing interest combining databases, artificial intelligence and machine learning .Data mining or knowledge discovery in databases (KDD) is the automatic extraction of implicit and interesting patterns from large data collections. Data mining or knowledge discovery in databases (KDD) is the automatic extraction of implicit and interesting patterns from large data collections. Data mining represents the automatic process to discover patterns and relations between data stored in large databases, the final product of this process being the knowledge, meaning the significant information provided by the unknown elements. Several data mining tasks are association, clustering, classification etc.

*Department of Information Technology, Technocrats Institute of Technology, Bhopal,

Using these methods many kinds of knowledge can be discovered such as association rules, classifications and clustering. The discovered knowledge can be used to better understand students' behavior, to assist instructors, to improve teaching, to evaluate and improve e-learning systems, to improve curriculums and many other benefits.

There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, concerns with developing methods that discover knowledge from data come from educational environments. The data can be collected form historical and operational data reside in the databases of educational institutes. The student data can be personal or academic. Also it can be collected from e-learning systems which have a vast amount of information used by most institutes. Educational data mining used many techniques such as decision trees, neural networks, k-nearest Neighbour, Naive Bayes, support vector machines and many others. In the modern world, information and communication technologies become to be fully involved in the educational processes. As the result, large amount of data can be generated as a side effect. The data can be either directly related to the educational process (e.g. testing of students, learning management systems LMS) or can be generated from other related activities (e.g. leisure time activities, library entries). Educational data may contain many interesting information and potential knowledge about students and their learning habits. However, such knowledge is hidden and their extraction is not a trivial task. Many applications try to describe and then effectively apply the knowledge hidden in the databases or in large data warehouses. Knowledge acquiring methods can be used to extract non-trivial, previously unknown and potentially effective and useful information (knowledge) from the data. The principle of these methods is mainly based on the application of analytic methods. They typically use specially preprocessed data as an input and return knowledge information as an output. This branch of science is also called Data mining, Information harvesting or Knowledge discovery in databases. Educational data mining differs from knowledge discovery in other domains in several ways. One of them is the fact that it is difficult, or even impossible, to compare different methods or measures a posteriori and decide which is the best. Take the example of building a system to transform handwritten documents into printed documents. This system has to discover the printed letters behind the hand-written ones. It is possible to try several sets of measures or parameters and experiment what works best. Such an experimentation phase is difficult in the educational field because the data is very dynamic, can vary a lot between samples and teachers just cannot afford the time and access to the expertise to do these tests on each sample, especially in real time. Therefore, as argued in, one should care about the intuition of the measures, parameters or methods used in educational data mining. Another difference is the size of the data: while tremendous amounts of data are collected about students' work, the size of the data on one sample is usually small. Typically in a classroom there are at best a few hundred students enrolled. Students may not all do the same exercises or activities. Collecting several years of data is certainly an option but there are instances where one wants to analyse the data as early as the first year. Besides, there are often changes between offerings of a course that have an impact on the common attributes of the data (for example not exactly the same topics/exercises/resources are offered from one year to the next). Therefore one should also be careful to avoid measures, parameters and methods where sample size has a predominant effect on the result. Association rules are increasingly used in educational data mining. However, measuring the interestingness of a rule can be problematic. Two measures, support and confidence, are commonly used to extract association rules. However it is well known that even rules with a strong support and confidence may in fact be uninteresting. This is why, once the association rule *X Y* has been extracted, it is wise to double check how much *X* and *Y* are related.

2. The Process Of Implementation Of Association Rule In Educational Database: The general data mining process follows following steps: collecting data, preprocessing, applying the actual data mining tasks and post-processing. We particularize these steps for association rule mining in the Educational database.

3. Collecting data: Most of the current educational databases do not store logs as text files. Instead, they normally use a relational database that stores all the systems information, personal information of the users (profile), academic results, the user s interaction data, etc.. Databases are more powerful, flexible and bug-prone than the typically textual log files for gathering detailed access and high level usage information from all the services available in the educational databases. The educational databases keep detailed logs of all activities that students perform. Not only every click that students make for navigational purposes (low level information) is stored, but also test scores, elapsed time, etc. (high level information).

4. Data pre-processing: Most of the traditional data reprocessing tasks, such as data cleaning, user identification, session identification, transaction identification, data transformation and enrichment, data integration and data reduction are not necessary in educational databases. Data preprocessing of educational databases data is simpler due to the fact that most educational databases store the data for analysis purposes, in contrast to the typically observational datasets in data mining, that were generated to support the operational setting and not for analysis in the first place. Educational databases also employ a database and user authentication (password protection) which allows identifying the users in the
logs. Some typical tasks of the data preparation phase are: data discretization (numerical values are transformed to categorical values), derivation of new attributes and selection of attributes (new attributes are created from the existed ones and only a subset of relevant attributes are chosen), creating summarization tables (these tables integrate all the desired information to be mined at an appropriate level, e.g. student), transforming the data format (to format required by the used data mining algorithms r frameworks).

5. Applying the mining algorithms: In this step it is necessary: 1) to choose the specific association rule mining algorithm and implementation; 2) to configure the parameters of the algorithm, such as support and confidence threshold and others; 3) to identify which table or data file will be used for the mining; 4) and to specify some other restrictions, such as the maximum number of items and what specific attributes can be present in the antecedent or consequent of the discovered rules.

6. Data post-processing: The obtained results or rules are interpreted, evaluated and used by the teacher for further actions. The final objective is to putting the results into use. Teachers use the discovered information (in form of if-then rules) for making decisions about the students and the educational databases activities of the course in order to improve the students learning. So, data mining algorithms have to express the output in a comprehensible format by e.g., using standardized e-learning metadata.

7. Association Rule: Association rules mining is one of the most well studied data mining tasks. It discovers relationships among attributes in databases, producing if-then statements concerning attribute-values. An association rule X ⇒Y expresses that in those transactions in the database where X occurs; there is a high probability of having Y as well. X and Y are called respectively the antecedent and consequent of the rule. The strength of such a rule is measured by its support and confidence. The confidence of the rule is the percentage of transactions with X in the database that contain the consequent Y also. The support of the rule is the percentage of transactions in the database that contain both the antecedent and the consequent.

Let *I = {I1, , I2, ...,Ip}* be a set of p items and *T = {t1, , t2, ...,tn}* be a set of *n* transactions, with each *ti* being a subset of *I*. An *association rule* is a rule of the form *X Y*, where *X* and *Y* are disjoint subsets of *I* having a support and a confidence above a minimum threshold. Let us denote by *X , Y* the number of transactions that contain both *X* and *Y*. The support of that rule is the proportion of transactions that contain both *X* and *Y*: *sup(X Y) = X , Y / n*. This is also called *P*(*X, Y*), the probability that a transaction contains both *X* and *Y*. Note that the support is symmetric: *sup(X Y) = sup(Y X)*. Let us denote by *X* the number of transactions that contain *X*. The confidence of a rule *X→Y* is the proportion of transactions that contain *Y* among the transactions that contain *X: conf* (X →Y) = / *X* , *Y* / *X*/ . An equivalent definition is : *conf* (*X*→ *Y*) = *P(X,Y) / P(X)*, with *P(X) = X / n* This is also written *P(Y/X)*, the probability that a transaction contains *Y* knowing that it contains *X* already. Note that confidence is not symmetric, usually *conf(X →Y)* is different from *conf(Y X)*, and gives its direction to an association rule.

A. Apriori Alogorithm: The Apriori algorithm finds the frequent sets L in Database D.

- Find frequent set Lk -1.
- Join Step.
  - Ck is generated by joining Lk -1with itself
- Prune Step.
  - Any (k 1) item set that is not frequent cannot be a subset of a frequent k item set, hence should be removed.

Where

(*Ck*: Candidate item set of size *k*)
(*Lk*: frequent item set of size *k*)

B. Cosine

Consider two vectors *x* and *y* and the angle they form when they are placed so that their tails coincide. When this angle nears 0°, then cosine nears 1, i.e. the two vectors are very similar: all their coordinates are pair wise the same (or proportional). When this angle is

90°, the two vectors are perpendicular, the most dissimilar, and cosine is 0.
Let x and y be two vectors of length n: x=$x = (x_1, \ldots x_n)$, y= $(y_1, \ldots y_n)$.
Then $cosine(x\ ,\ y)=(\ x\cdot y)/\ (\|x\|.\|y\|)$,
where above indicates the vector dot product

$$\sum_{n}^{k=1} xkyk$$

and $\|x\|$ is the length of vector x, $\|x\|= \sqrt{\sum_{k=1}^{n} x_k^2}$

Borrowing this idea, it is easy to associate two vectors *x* and *y* to the rule *X Y*. Let us interpret x_k as being *1* if transaction t_k contains *X* and *0* otherwise, and similarly for y_k and *Y*. Then it is immediate that the equation for cosine can be rewritten as *cosine(x , y)=cosine (X→ Y)* , the usual form that is given for cosine of an association rule *X Y*. The closer *cosine(X Y)* is to *1*, the more transactions containing item *X* also contain item *Y*, and vice versa. On the contrary, the closer *cosine(X Y)* is to *0*, the more transactions contain item *X* without containing item *Y*, and vice versa. Simplifying with *n* gives *cosine* *(X →Y)=/X,Y/*/$\sqrt{|X|.|Y|}$ This equality shows that transactions not containing neither item *X* nor item *Y* have no influence on the result of *cosine(X Y)*. This is known as the null-invariant property. Note also that cosine is asymmetric measure.

C. Added value and lift: The added value of the rule *X →Y* is denoted by *AV (X→Y)* and measures whether the proportion of transactions containing *Y* among the transactions containing *X* is greater than the proportion of transactions containing *Y* among all transactions. Then, only if the probability of finding item *Y* when item *X* has been found is greater than the probability of finding item *Y* at all can we say that *X* and *Y* are associated and that *X* implies *Y*. *AV(X→ Y)=P(Y / X) P(Y) =conf(X →Y) P(Y).* A positive number indicates that *X* and *Y* are related, while a negative number means that the occurrence of *X* prevents *Y* from occurring. Added Value is closely related to another well-known measure of interest, the lift. *Lift(X→Y)=P(X ,Y)/P(X). P(Y)=conf(X→ Y) /P(Y).* Note that if *P(X, Y) = P(X) . P(Y)* the lift is *1*. In terms of probability, this means that the occurrence of *X* and the occurrence of *Y* in the same transaction are independent events, hence *X* and *Y* are not correlated. It is easy to show that the lift is *1* exactly when added value is *0*, the lift is greater than 1 exactly when added value is positive and the lift is below *1* exactly when added value is negative. Further *AV(X Y)* tends towards *1* when *lift(X Y)* tends towards infinity, and *AV(X Y)* tends towards *-1* when *lift(X Y)* tends towards *0*. Note that *lift (X→Y)= X ,Y . n/ X . Y* so the result is proportional to *n*, the total number of transactions. As opposed to cosine, lift does not hold the null-invariant property.

D. Typical values for cosine and lift*:* To fix ideas let us look at typical values for these measures Suppose that among *n* transactions, *m* contain either *X* or *Y* or both, with *m n*, and that *n - m* transactions contain neither *X* nor *Y*. First consider the case where all *m* transactions contain both *X* and *Y*. Then: *cosine(X→Y)= 1*. Conversely, it is easy to show that *cosine (X→Y) = 1* implies that all *m* transactions contain both *X* and *Y*. As for the lift, *lift (X→Y)=* (*m.n)*/(*m.m)=n /m* . So if *m = n,lift(X→Y)*=1. If *m = ½ . n, lift(X→Y)*=2 and so on. Table 1 summarizes further results. Lines should be read as follows: (*a*,*b*,*c*) means that *a %* of the *m* transactions contain both *X* and *Y*, *b%* contain *X* and *c%* contain *Y*. Therefore (75, 100, 75) means that 75% of the *m* transactions contain both *X* and *Y* and that the remaining 25% contain *X* but not *Y* (*X* is present in 100% of the transactions and *Y* in 75% of them), while (75, 87.5, 87.5) means that *X* or *Y* are evenly spread among the 25% of the remaining transactions. Let us now explain all this. In the case of strong symmetric association rules, which means that *X* , *Y* and *X* ,*Y* are all big numbers close to *n*, cosine and lift do not rate rules the same way. In this case, cosine performs better than lift. Added value and lift rely on probabilities, which make more sense when the number of observations is large. Further we see also that lift and added value, unlike cosine, depend on the number of transactions that contain neither X nor Y. In the educational field it is not clear that these null-transactions should play a role. We come to the conclusion as double check the interestingness of association rules with cosine first, then with lift if cosine is not conclusive. Table 1 suggests that a value around or below 0.65 is rejected by cosine as we can see 0.66 corresponds to the lowest threshold with 50% of common values (50,75, 75). In case of contradictory results then decide using the information that these two measures represent. Table 1. Typical values for cosine and lift, where the 3 figures of the first column show the percentage of transactions containing X and Y, X and Y

% transactions (X and Y, X, Y)	**cosine(X→Y)**	**lift(X→Y)**
(100, 100, 100)	1	n/m
(90, 100, 90)	0.949	n/m
(90, 95, 95)	0.947	0.997.(n/m)
(75, 100, 75)	0.87	n/m
(75, 87.5, 87.5)	0.86	0.98(n/m)
(60, 100, 60)	0.77	n/m
(60, 80, 80)	0.75	0.94(n/m)
(50, 100, 50)	0.707	n/m
(50, 75, 75)	0.66	0.88(n/m)
(40, 100, 40)	0.63	n/m
(40, 70, 70)	0.57	0.82(n/m)
(30, 100, 30)	0.55	n/m
(30, 65, 65)	0.46	0.71(n/m)

8. Conclusion: We thus see that implementation of association rule helps a lot in generating interesting rules. But the interestingness of the rules entirely depend on the correct choice of the threshold values of two most important factors the support and the confidence. These values and the resultant rules can thus be used successfully to generate knowledge for educational data too. Thus in this way association rule is implemented. The drawback of this paper is that after the rules are generated it is not necessary that these rules will be of human interest all the time. After this it relies entirely on the human intelligence upon how much the knowledge is useful, meaningful and interesting. This drawback is also a drawback of data mining.

9. References:

1. Arun K Pujari Data Mining Techniques Univerrsies Press Hyderabad Vol. 13, 2207,pp. 71-81
2. Jiawei Han Micheline Kamber, Data Mininhg Concepts and Techniques Morgan Kaufmann Publishers, Edition -3, pp. 5 33.
3. Using Association Rule For Behavioral Analysis of Schhol Students : A case From India, Dass, R. , System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference
4. Beck J. Difficulties in inferring student knowledge from observations (and why you should care). Educational Data Mining workshop, in conjunction with 13th International Conference of Artificial Intelligence in Education, Marina del Rey, CA. USA. July 2007, pp.21-30.
5. Garcia J., Romero C., Ventura S., Calders T. Drawbacks and solutions of applying association rules mining in learning management systems. Proceedings of the International Workshop on Applying Data Mining in e-learning (ADML'07), Crete, Greece 2007, p. 13-22.
6. Jflap http://www.jflap.org/ last consulted March 01 2008K.
7. Merceron, A. (2008) Unterstützung des Erlernens von endlichen Automaten mit Hilfe von Mustern.*Proceedings Wokshop für Modellierung in Lehre und Weiterbildung,* Desel, J. & Gliz, M.(Eds.), Modellierung, Berlin, Germany.
8. Merceron, A., Yacef, K. (2007) Revisiting interestingness of strong symmetric association rules in educational data. *Proceedings of the International Workshop on Applying Data Mining in elearning (ADML'07),* Crete, Greece 2007, p. 3-12.
9. Merceron, A., Yacef , K. (2005) Educational Data Mining: a Case Study. *Proceedings of Artificial Intelligence in Education (AIED2005),* Amsterdam, The Netherlands, IOS Press, 2005.
10. Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer, G., Punch, W.F. (2003), "Predicting student performance: an application of data mining methods with the educational web-based system LON-CAPA. *ASEE/IEEE Frontiers in Education Conference*. 2003. Boulder, CO: IEEE.
11. Moodle http://moodle.org/ last consulted March. 01 2008.
12. Tan P.N., Kumar V., Srivastava J. (2001) Selecting the Right Interestingness Measure for Association Patterns. *Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* San Francisco, USA, p. 67-76.

13. Wang, F., (2006) On using Data Mining for browsing log analysis in learning environments. *Data Mining in E-Learning. Series: Advances in Management Information*,, Romero, C., Ventura, S.,Editors, WIT press. p. 57-75.