Interestingness Measures for Association Rules in Educational Data

Agathe Merceron[1] and Kalina Yacef[2]
merceron@tfh-berlin.de, kalina@it.usyd..edu.au,
[1]Computer Science and Media Department,
University of Applied Sciences TFH-Berlin, Germany
[2]School of Information Technologies, University of Sydney, Australia

Abstract. Educational data differs from traditional knowledge discovery 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 deduce which the best is. It is therefore essential to use techniques and measurements that are fairly intuitive and easy to interpret. Extracting the most interesting association rules can be quite tricky. One of the difficulties is that many measures of interestingness do not work effectively for all datasets and are hard to understand intuitively by the teachers. We argue in this paper that cosine and added value (or equivalently lift) are well suited to educational data, and that teachers can interpret their results easily. We argue that interestingness should be checked with cosine first, and then with lift if cosine rates the rule as non-interesting. If both measures disagree, teachers should use the intuition behind the measures to decide whether or not to dismiss the association rule. We provide a case study with data from a LMS.

1 Introduction

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 hand-written 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 [1], 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 hundreds 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 [8, 9, 12]. However, measuring the interestingness of a rule can be problematic, as explained in [2]. 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 $X \rightarrow Y$ has been extracted, it is wise to double check how much X and Y are related. About 20 measures have been proposed in the literature to do so. Unfortunately, no measure is better than all the others in all situations, though measures tend to agree when support is high [11].

In this paper, we revisit measures of interestingness for association rules and argue that two of them, cosine and added value, are particularly suited for educational data because their meaning is fairly intuititive even to non data mining experts and one of them, cosine, does not depend on the data size. We present a case study using our usage data from the Learning Management System Moodle [10].

In the next section we present the underlying concepts: the association rules, the cosine, the added value and the lift, with typical values for each of these measures followed by a discussion. We then present a case study, corresponding to a common situation of using a LMS to put additional learning material for students. Supporting teachers in this task is giving them some way to be informed of the use of this extra material by students, and of its impact on the marks. Association rules are used to refine the findings made while exploring data. Cosine and lift agree on most rules found, but not on all. As we will see, the intuition behind the two measures helps to understand this information.

2 Association Rules, cosine, Added Value and Lift

Association rules come from market basket analysis and capture information such as "if customers buy book X, they also buy book Y". This can be written as $X \rightarrow Y$. Two measures characterize an association rule: support and confidence.

2.1 Association rules

Let $I = \{I_1, I_2, ..., I_p\}$ be a set of p items and $T = \{t_1, t_2, ..., t_n\}$ be a set of n transactions, with each t_i being a subset of I.

An *association rule* is a rule of the form $X \rightarrow Y$, 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: $sup(X \rightarrow Y) = |X, Y| / n$. This is also called $P(X, Y)$, the probability that a transaction contains both X and Y. Note that the support is symmetric: $sup(X \rightarrow Y) = sup(Y \rightarrow X)$.

Let us denote by $|X|$ the number of transactions that contain X. The confidence of a rule $X \rightarrow Y$ is the proportion of transactions that contain Y among the transactions that contain X: $conf(X \rightarrow Y) = |X, Y| / |X|$. An equivalent definition is : $conf(X \rightarrow Y) = P(X,Y) / P(X)$, 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.

2.2 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 pairwise 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*: $x=(x_{1,} \dots , x_n), y=(y_{1,} \dots , y_n)$.

Then $cosine(x, y)=\frac{(x\cdot y)}{(\|x\|.\|y\|)}$, where . above indicates the vector dot product $x\cdot y=\sum_{k=1}^{n} x_k y_k$ 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 $cosine(x, y)=\frac{P(X,Y)}{\sqrt{(P(X).P(Y))}}=cosine(X\rightarrow Y)$, 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 $cosine(X\rightarrow Y)=\frac{|X, Y|}{\sqrt{|X|.|Y|}}$. 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 a symmetric measure.

2.3 Added value and lift

The added value of the rule *X→Y* is denoted by $AV(X\rightarrow Y)$ 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*.

$AV(X\rightarrow Y)=P(Y/X)-P(Y)=conf(X\rightarrow Y)-P(Y)$. 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.

$$lift(X\rightarrow Y)=\frac{P(X,Y)}{P(X).P(Y)}=\frac{conf(X\rightarrow Y)}{P(Y)}.$$

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 $AV(X \rightarrow Y)$ tends towards *1* when $lift(X \rightarrow Y)$ tends towards infinity, and $AV(X \rightarrow Y)$ tends towards *-1* when $lift(X \rightarrow Y)$ tends towards *0.*

Note that $lift(X \rightarrow Y) = \frac{|X, Y| . n}{|X| . |Y|}$ so the result is proportional to *n*, the total number of transactions. As opposed to cosine, lift does not hold the null-invariant property.

2.4 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 $m \leq n$, and that $n - m$ transactions contain neither *X* nor *Y*.

First consider the case where all *m* transactions contain both *X* and *Y*. Then: $cosine(X \rightarrow Y) = 1$. Conversely, it is easy to show that $cosine(X \rightarrow Y) = 1$ implies that all *m* transactions contain both *X* and *Y*. As for the lift, $lift(X \rightarrow Y) = (m.n)/(m.m) = n/m$. So if $m = n$, $lift(X \rightarrow Y)=1$. If $m = ½ . n$, $lift(X \rightarrow Y)=2$ and so on.

Consider now the case where 90% of the *m* transactions contain both *X* and *Y*, and 10% of the rest contain *X* but not *Y*. Then:

$$cosine(X \rightarrow Y) = \frac{\frac{90}{100} . m}{\sqrt{m . \frac{90.m}{100}}} = \frac{90}{100} . \sqrt{\frac{100}{90}} = 0.949$$

$$lift(X \rightarrow Y) = (\frac{90}{100} . m . n)/(m . \frac{90}{100} . m) = \frac{n}{m}$$

Now consider again the case where 90% of the *m* transactions contain both *X* and *Y*, but 5% of the rest contain *X* and not *Y*, and the other 5% contain *Y* and not *X*. In other words *X* and *Y* are evenly spread among the transactions containing either *X* or *Y* but not both Then:

$$cosine(X \rightarrow Y) = \frac{\frac{90}{100} . m}{\sqrt{\frac{95.m}{100} . \frac{95.m}{100}}} = \frac{90}{100} . \frac{100}{95} = 0.947$$

$$lift(X \rightarrow Y) = (\frac{90}{100} . m . n)/(\frac{95}{100} m . \frac{95}{100} . m) = 0.99 \frac{n}{m} .$$

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.

Discussion: 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, as pointed out in [7]. 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 same conclusion as [3]: 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)

3 Improving Teacher Support: Case Study

The present case study describes a standard use of a Learning Management System (LMS) for providing additional resources to students in a face-to-face teaching context. Teachers want to figure out whether students use these resources and possibly whether their use has any (positive) impact on marks.

The LMS Moodle [10] was used in the context of the course *Formal Basics of Computer Science* for first semester students enrolled in the degree “Computer Science and Media” at the University of Applied Sciences TFH Berlin during Winter Semester 2007/08. The cohort of 84 students enrolled in that course is divided into two groups. Students had a 3-hour weekly lecture. It includes formal teaching where concepts are explained, paper/pencil exercises to apply these concepts, and exercises discussed on the spot. To pass this course students take two exams. The first one takes place about 8 weeks after the beginning of the semester and the second one at the end of the semester. The present case study uses data gathered till the first exam.

Moodle is used for posting lecture slides and accessing the following extra resources:

- *Book*: a link to the homepage of the text book "Introduction to Automata Theory, Languages and Computation" used for this course [4]. From this homepage students could access a set of exercises with solutions.
- *DP*: extra reading "Design Patterns for finite automata" [6].
- *Jflap* [5], a software to practice automata construction.
- *Ex1, Ex2 ... Ex7* : a set of seven extra self-evaluation exercises. One exercise is published in Moodle each week right after the lecture. The last exercise *Ex7* was put 2 weeks before the exam.
- *TrEx01* and *TrEx02* : two sample exams, published 3 weeks before the exam.
- *TrEx01S* and *TrEx02S*, the solutions to the sample exams, published 10 days before the exam.

The use of Moodle, its additional resources and its self-evaluation exercises were not compulsory though strongly encouraged. Therefore for the teacher it is quite important to know: what do students do with those extra resources? What do they view? Is there any relationship between their use of these resources and their result in the exam? To answer these questions we have used solely the log data available in Moodle. Log data gives, for each resource and each student login, when the resource was accessed. It also gives, for each exercise and each student login, whether the exercise has been attempted, and whether the first trial was a success or not.

3.1 Exploring Data

From the 84 students enrolled in the course, 81 were enrolled in Moodle. The case study considers only those 81 students. From them, 52 passed the exam, 8 failed and 21 did not come. From the 60 who took the exam, statistics on their marks is given in the first line "General" of Table 4.

Did students do the exercises? Table 2 summarizes the figures. Lines should be read as follows. For example line 2 means that 46 students did not attempt exercise 1, 21 students gave a correct answer on their first trial and 14 gave a wrong answer on their first trial. One notices that as time goes there is always less students attempting exercises.

Table 2. Exploring exercises among all students.

Exercise	**Ex1**	**Ex2**	**Ex3**	**Ex4**	**Ex5**	**Ex6**	**Ex7**
No attempt	46	53	63	65	70	70	71
Success	21	19	11	8	6	9	8
Fail	14	9	7	8	5	2	2

Table 3. Viewing resources.

TrEx01	TrEx01S	TrEx02	TrEx02S	JFlap	DP	Book	AtLeast1Ex
59	52	53	47	39	36	23	38

Did they access other resources? Table 3 summarizes the figures. The first column says that 59 students have viewed the first sample exam, the second column says that 52 students have viewed the solution of the first sample exam, and so on. One extra column has been added. *AtLeast1Ex* says that 38 students have attempted at least 1 exercise.

What are the results in the exam for each group of Table 3? Table 4 summarizes the results. Two extra lines have been added. *NoEx* shows the results for students who have never attempted any exercise. *AtLeast1Ex* shows the results for the students who have attempted at least 1 exercise. Table 3 and Table 4 suggest that the standard preparation for the exam is to look at sample exams and/or their solutions. Students who invest some more time with extra material tend to have better marks. The biggest positive impact on the marks is given by *DP*.

Table 4. Viewing resources and marks in the exam.

Resource	minimum	maximum	average	s.deviation
General	11	50	36.45	10.85
TrEx01	14	50	36.86	10.57
TrEx01S	14	50	36.61	10.71
TrEx02	14	50	37.12	10.59
TrEx02S	14	50	36.73	10.79
JFlap	14	50	37.90	10.52
DP	14	50	44.08	8.83
Book	14	50	40.22	9.37
NoEx	11	50	33	10.91
AtLeast1Ex	14	50	39.09	10.18

Table 3 and 4 confirm the expected outcome. Table 4 also shows something that was not known before: students tend to access a sample exam more that its solution.

This first exploration gives also directions for more investigation: If students attempt exercise 2, do they also attempt exercise 1? If they look at the solution of a sample exam, do they also look at the sample exam itself? This kind of questions can be investigated with association rules.

3.2 Association Rules

We begin with association rules tackling sample exams. The following rules again confirm the expected finding. If students look at the solution of a sample exam, they look also at the sample exam itself. Further, if they view the second exam, then they also view the first one. The other way round does also hold, but with a slightly lower confidence.

Table 5. Association rules for sample exams

rule	sup.	conf.	cos.	lift
TrEx01S → TrEx01	0.59	0.92	0.87	1.27
TrEx01 → TrEx01S	0.59	0.81	0.87	1.27
TrEx02S → TrEx02	0.56	0.96	0.90	1.46

TrEx02 → TrEx02S	0.56	0.85	0.90	1.46
TrEx01S → TrEx01	*0.72*	*0.96*	*0.90*	*1.11*
TrEx01 → TrEx01S	*0.72*	*0.84*	*0.90*	*1.11*
TrEx02 → TrEx01	0.64	0.98	0.93	1.35
TrEx01 → TrEx02	0.64	0.88	0.93	1.35

Results are similar when rules are mined restricting the population to the students who came to the exam as shown for the first sample exam in the lines in italic of Table 5. Notice however that the lift diminishes as the rules become stronger [7].

Table 2 gives a direction for further rules to investigate: Is there any association between attempting exercise i and exercise j? One expects that many students enthusiastically have begun with exercise 1 at the beginning of the semester and then slowly have stopped doing them, till exercise 4 where a bunch of students just keep doing them. The rules we have obtained confirm this interpretation.

We have mined these rules restricting the data to students who have attempted at least 1 exercise, which means 38 transactions. Rules with a high confidence relate attempting exercise 2 and exercise 3, exercises 4 to 7, as well as exercise 1, and not attempting exercises 2 to 3, or not attempting exercises 4 to 7. Table 6 presents a sample of the extracted associations. Note that *!Ex2* means that exercise 2 has not been attempted. So the first line says if students don't attempt exercise 2, then they don't attempt exercise 3.

Table 6. Association rules for attempted exercises

rule	**sup.**	**conf.**	**cos.**	**lift**
!Ex2 → !Ex3	0.26	1	0.71	1.9
Ex3 → Ex2	0.47	1	0.80	1.36
!Ex4, !Ex5 → !Ex6	0.56	1	0.93	1.41
!Ex6, !Ex7 → !Ex5	0.63	0.96	0.92	1.35
Ex1, !Ex4 → !Ex5, !Ex6	0.55	1	0.90	1.46
Ex1, Ex7 → Ex5, Ex6	0.21	1	0.89	3.8
!Ex5, !Ex6 → Ex1	0.63	1	0.80	1.0

For all these rules, except the last one, cosine and lift rate associations the same way. The drop between attempting exercises 1 to 3 and attempting the others has led us to investigate the marks of this population. Surprisingly, their average mark is smaller than for all students who have attempted at least 1 exercise.

Table 7. Attempting exercises 4 to 7 and marks in the exam.

Resource	**minimum**	**maximum**	**average**	**s.deviation**
Ex 4 to 7	16	50	38.76	10.35

As for the other resources, were they consulted by the same students? We have looked at associations between *DP*, *Jflap*, *Book* and *AtLeast1Ex* considering the full population and show two rules found in Table 8. Here lift does not confirm the non-interesting rating given by cosine. As before, *!DP* means that the resource *DP* has not been viewed.

Table 8. Association rules for the other resources

rule	sup.	conf.	cos.	lift
!DP, !AtLeast1Ex → !Book	0.32	0.92	0.62	1.29
Book, DP, AtLeast1Ex → Jflap	0.12	1	0.51	2.08

Keeping in mind the meaning of measures can help deciding what to do with an association.

Let us consider the last rule of Table 6. Cosine indicates that, among the students who have not done *Ex5* nor *Ex6*, over 60% had done Ex1 (consult the typical figures in Table 1), while lift indicates that the proportion of students who have not done *Ex5* and *Ex6* is not larger in students who did *Ex1* (which represented 43% of the students, according to table 2) than in all students. However, from a pedagogical point of view, the case of students who did not attempt *Ex1* is not relevant for this analysis. Therefore the teacher would probably find it useful to keep this rule, hence following cosine, though lift gives here an interesting complementary information.

Let us now consider the second rule of Table 8. Cosine gives us the following information: among the students who consulted the *Book* web site, the extra material on Design Patterns (*DP*) and done at least one exercise, less than 40% used *Jflap* (refer to the typical values in Table 1). The lift gives us the following information: the proportion of students who looked at *Jflap* is higher among the students who looked at the *Book* web site, the *DP* material and done at least one exercise than in the whole student population. Given the very small number of students (as we can see from the support, there are 10 students who satisfied these three criteria on the left hand side of the rule), it is prudent to follow cosine and reject the rule. It is interesting to note though that if the cosine and lift had given similar values, but with a higher number of students satisfying the three criteria in the left side of the rule, it would then have been advisable to follow the lift and retain the rule.

Conclusion

Association rules are useful in Educational Data Mining for analysing learning data. This technique requires not only that adequate thresholds be chosen for the two standard parameters of support and confidence, but also that appropriate measures of interestingness be considered to retain meaningful rules and filter uninteresting ones out. In this paper we revisited and gave an interpretation for two interestingness measures: cosine and added value (which we saw is closely related to the lift). We presented typical values for these measures. An association rule is rated uninteresting by cosine if its value is around or smaller than 0.65, whereas it is rated uninteresting by the lift if its value is around or under 1. We came to a similar conclusion as in [3]: the interestingness of a rule should be first measured by the cosine, then with lift if cosine rated it as uninteresting. In case of conflict between the two measures, the user needs to take into account the intuitive information provided by each measure and decide upon it. The case study presented in the paper depicts a standard situation: a LMS provides additional resources for students in complement to the face-to-face teaching context. Teachers want to figure out whether students use these resources and whether these have any (positive) impact on marks. Few association rules (without being strong symmetric ones) came out with a contradictory result for cosine and lift. Keeping in mind the intuition behind cosine and lift helped to decide whether to discard these rules.

Another conclusion of this work is that common LMS are far from being data mining friendly. Log data concerning access to resources and test data are not stored the same way for example. Complex data manipulation is needed to get all data consolidated in a useful form. LMS present statistics, however these are very limited. LMS should be enhanced with a special module with good facilities for exploring data. Data mining tools for LMS should have an association rules module with good facilities to choose the attributes to derive association rules for and with the two interestingness measures cosine and lift.

References

[1] 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 .

[2] 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.

[3] Han J., Kamber M.,: Concepts and Techniques, Morgan Kaufmann Publishers, 2001.

[4] Introduction to Automata Theory, Languages, and Computation, http: //infolab.stanford.edu/~ullman/ialc.html, last consulted March 01 2008.

[5] Jflap http://www.jflap.org/ last consulted March 01 2008.

[6] Merceron, A. Unterstützung des Erlernens von endlichen Automaten mit Hilfe von Mustern. *Proceedings Wokshop für Modellierung in Lehre und Weiterbildung,* Desel, J. & Gliz, M. (Eds.), Modellierung 2008, Berlin, Germany.

[7] Merceron, A., Yacef, K. Revisiting interestingness of strong symmetric association rules in educational data. *Proceedings of the International Workshop on Applying Data Mining in e-learning (ADML'07),* Crete, Greece 2007, p. 3-12.

[8] Merceron, A., Yacef , K. Educational Data Mining: a Case Study. *Proceedings of Artificial Intelligence in Education (AIED2005),* Amsterdam, The Netherlands, IOS Press, 2005.

[9] Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer, G., Punch, W.F. 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,

[10] Moodle http://moodle.org/ last consulted March. 01 2008.

[11] Tan P.N., Kumar V., Srivastava J. 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, 2001.

[12] Wang, F. 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, 2006.