Generalizing Association Rules to Ordinal Rules

Sylvie Guillaume[1] **Ali Khenchaf**[2] **Henri Briand**[1]

[1]IRIN - Université de Nantes
Ecole Polytechnique de l'université de Nantes
2, Rue de la Houssinière – BP 92208
44322 Nantes Cedex 3 – France

sguillau@ireste.fr

[2]Lab. IRCCyN, UMR 6597 CNRS, Div. SETRA
Ecole polytechnique de l'université de Nantes
Rue C. Pauc, La Chantrerie - BP 50609
44306 Nantes cedex 3 – France

akhencha@ireste.fr

Abstract

The development of good measures of interestingness of the discovered rules is one of the important problems in data mining. Such measures of interestingness are divided into objective measures : – those that depend only on the structure of a rule and the underlying data used in the discovery process, and the subjective measures – those that depend on the class of users who examine the rule. However, most objective measures are suitable for binary attributes and require an appropriate transformation of the initial set of attributes into binary attributes for all unsupervised usual algorithms for the discovery of association rules. As a result, the complexity of these algorithms increases exponentially with the number of attributes, and this transformation can lead us, on the one hand to a combinatorial explosion, and on the other hand to a prohibitive number of weakly significant rules with many redundancies. Moreover, the few measures suitable for numeric attributes, like for example correlation coefficient, are not selective. In this paper, we propose a new objective measure, called *ordinal intensity of implication*, which generalizes intensity of implication suitable for binary attributes and which evaluates whether the number of transactions not clearly verifying rule $X \rightarrow Y$ (i.e., the number of transactions containing a high value for attribute X and a low value for attribute Y) is significantly small as compared to a random draw. We finish the study with an evaluation on banking data and show some discovered ordinal rules, and connection between data / information and quality.

Keywords : Measures of interestingness, objective measures, intensity of implication, statistical analysis, numeric attribute.

1 Introduction

The aim of data mining is to facilitate the understanding of large amounts of data by discovering useful or interesting rules for the user [Piatetsky-Shapiro & Frawley 1991], [Fayyad *et al.* 1996]. However, a discovery system can generate a glut of rules, most of which are of no interest to the user [Frawley *et al.* 1991], [Piatetsky-Shapiro & Matheus, 1994], [Piatetsky-Shapiro *et al.* 1994], [Liu & Hsu 1996]. The presence of the huge number of rules makes it difficult for the user to analyze them and to identify those that are of interest to him/her.

Identifying interesting rules is not a simple task because a rule could be interesting to one user but uninteresting to another. In general, the evaluation of the interestingness of discovered rules has both an objective[1] [Silberschatz & Tuzhilin 1996], [Major & Mangano 1993] and a

[1] data-driven

subjective[2] aspect [Piatesky-Shapiro *et al.* 1994], [Klemettinen *et al.* 1994], [Silberschatz & Tuzhilin 1996], [Liu & Hsu 1996].

The objective approach depends only on the structure of a rule and the underlying data used in the discovery process, whereas the subjective approach depends on the class of users who examine the rule.

In practice, both objective and subjective approaches should be used to select interesting rules. The objective approach can be used as a kind of filter to select potentially interesting rules, among the many rules discovered by a data mining algorithm. The subjective approach can be used as a kind of filter to reduce further the number of potentially interesting rules; the user will judge the ultimate interestingness of the remaining rules.

There are many objective measures :

(1) entropy-based measures : [Goodman & Smyth 1989] argue that an interesting rule is a high-power predictive and general rule, and use a measure called the J-measure.

(2) probability-based measures [Silberschatz & Tuzhilin 1996]. [Schektman *et al.* 1992] keep rules X→Y that have a certain user-specified minimum confidence[3]. [Pavillon 1992] keeps rules X→Y whose confidence Pr(Y/X) is different from a priori probability Pr(Y). [Ganascia 1987], [Ganascia *et al.* 1990] propose the use of 2Pr(Y/X)-1 while [Piatetsky-Shapiro 1991] proposes the use of Pr(X)[Pr(Y/X)-Pr(Y)]. Selecting association rules[4] [Agrawal *et al.* 1993] is based on the support-confidence framework [Agrawal *et al.* 1996].

(3) statistical measures associated with hypothesis tests. [Brin *et al.* 1997a] use the chi-squared test associated first, with a measure of interest and then, with a measure called conviction measure [Brin *et al.* 1997b]. [Fleury *et al.* 1995], [Guillaume *et al.* 1998], [Suzuki & Kodratoff 1998] use intensity of implication [Gras 1979] in knowledge discovery systems.

However, these measures are essentially suitable for binary attributes[5] and require a transformation of attributes for numeric and categorical attributes which can disturb the relevance of results in the case of numeric attributes. Moreover, given that the complexity of unsupervised usual algorithms for the discovery of association rules [Agrawal *et al.* 1996], [Houtsma & Swami 1995], [Mannila *et al.* 1994], [Park *et al.* 1995] increases exponentially with the number of attributes, a transformation of initial attributes into binary attributes can lead us, on the hand to a combinatorial explosion, and on the other hand to a prohibitive number of weakly significant rules with many redundancies.

Moreover, measures suitable for numeric attributes such as correlation coefficient or Lerman's measure [Lerman 1981] are not selective for large databases and require another measure for finding the direction of implication (X→Y or Y→X). [Lagrange 1997] uses a selective measure, called intensity of propensity, but only suitable for numeric attributes with values in interval [0,1].

To fill the gap, we propose in this study a new objective rule-interest measure called ordinal intensity of implication which is computable over numeric attributes, and with an appropriate coding over ordinal categorical attributes. This measure, selective for very large databases, discards the step of transformation of attributes into binary attributes and makes it possible to

[2] user-driven

[3] the confidence of the rule is the conditional probability Pr(Y/X).

[4] an association rule is an implication of the form X→Y, where X and Y are conjunctions of attributes and $X \cap Y = \emptyset$

[5] or variables.

obtain the global behavior of the population since numeric attributes of discovered rules are not split into intervals.
The remainder of the paper is organized as follows. In section 2 we present measures suitable for numeric attributes and explain why they are not selective for large databases and in section 3 we present our new rule-interest measure called *ordinal intensity of implication*. In section 4 we explain the meaning of rules extracted with this measure, ordinal rules, and apply in section 5 this new measure to some banking data before ending up with a set of conclusions in the final section.

2 Numeric Measures

In this section, we present three numeric measures : correlation coefficient with statistical test, Lerman's similarity measure and intensity of propensity. And next, we explain why these three measures are not suitable for large databases.

2.1 Correlation Coefficient with statistical test

For a population whose size is n superior to 100, the random variable R, whose correlation coefficient r is an observed value, is approximated by the normal distribution with a mean 0 and a variance $\frac{1}{\sqrt{n-1}}$ [Saporta 1990]. The larger the size of the population, the smaller the variance. Therefore, the cumulative distribution for large populations has only two values : 0 for negative values and 1 for positive values; and then this measure also has two values : 0 and 1. So for large databases, this measure is not selective.

2.2 Lerman's Similarity Measure

Let X and Y be two numeric attributes, and r be the correlation coefficient. The similarity measure S(X,Y) [Lerman 1981] is defined by :

$$S(X,Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{n-1}\times r} e^{-\frac{1}{2}t^2} dt$$

When r is negative (positive) and the size of population n is large, the value of $\sqrt{n-1}\times r$ tends towards the negative infinite (positive infinite), and then the value of S(X,Y) tends towards 0 (1). To finish, when r is equal to 0, then the value of S(X,Y) is equal to 0.5.
Therefore, for large databases, this measure is not selective either because it has only three values : 0, 0.5 and 1.

2.3 Intensity of propensity

Let X and Y be two numeric attributes with values in interval [0,1]. Let n be the size of the sample E of the population Ω. Let x_i and y_i be respectively values for the transaction[6] i of the sample E, i=1,..,n. Let m_X and m_Y be respectively the arithmetic means of attributes X and Y and, v_X and v_Y be respectively the variances of attributes X and Y.

[6] or observations or records.

Intensity of propensity [Lagrange 1997] generalizes intensity of implication[7] defined by [Gras 1979] and used in knowledge discovery systems : FIABLE [Fleury *et al.* 1995] and PEDRE [Suzuki & Kodratoff 1998]. Intensity of propensity generalizes it for numeric attributes whose values are in interval [0,1]. This measure P(X→Y) is defined by :

$$P(X \to Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} e^{-\frac{1}{2}t^2}\, dt \quad \text{where} \quad s = \frac{\dfrac{\sum_{i=1}^{n} x_i(1-y_i)}{n} - m_X(1-m_Y)}{\sqrt{\dfrac{(v_X + m_X^2)(v_Y + (1-m_Y)^2)}{n}}}\,.$$

Unlike the two previous measures, intensity of propensity is a selective and implicative measure but suitable for numeric attributes having values in interval [0,1]. We can use this measure by adapting numeric attributes (i.e. with the following operation for all numeric attributes : $(X-x_{min})/(x_{max}-x_{min})$ with x_{min} and x_{max} representing respectively the minimum and maximum values for attribute X); but we *do* want to evaluate implications over the set of initial attributes because such a transformation step can be time-consuming for large databases. This is why we have adapted this measure for numeric attributes whose values are in any interval $[x_{min}, x_{max}]$.

3 Ordinal Intensity of Implication

In this section, we present ordinal intensity of implication which generalizes intensity of implication [Guillaume & Kenchaf 2000] for numeric attributes which have values in any interval and for any ordinal categorical attributes after an appropriate coding into numeric attributes.

3.1 Principle and notations

Let X and Y be two numeric attributes taking values respectively in $[x_{min}, x_{max}]$ and $[y_{min}, y_{max}]$.
Ordinal intensity of implication evaluates whether the number of transactions not strongly verifying the rule X→Y (i.e., the number of transactions verifying a high value for X and a low value for Y or the number of "negative transactions") is significantly small compared to the expected number of transactions under the assumption that X and Y are independent.
Therefore, we say that X→Y if the number of negative transactions is significantly small compared to the expected number of negative transactions under the assumption that X and Y are independent.
Let T be a database consisting of n transactions t_i (i=1, …,n) described by p attributes X_j (j=1, …,p) (see table 1).

Row ID	X_1	…	X_j	…	X	Y	…	X_p
t_1			x_{j1}		x_1	y_1		
…			…		…	…		
t_i			x_{ji}		x_i	y_i		
…			…		…	…		
t_n			x_{jn}		x_n	y_n		

Table 1 : Database T.

x_{ji} represents the value of attribute X_j for transaction t_i in set T.

[7] Intensity of implication is a measure suitable for binary attributes which evaluates whether the number of transactions not verifying the rule (i.e. transactions containing the attribute X and not containing the attribute Y or "negative transactions") is significantly small compared to a random draw.

3.2 Raw ordinal implication measure

Based on Lagrange's work [Lagrange 1997] and Gras' work [Gras *et al.* 1996], the proposed raw ordinal implication measure is defined by the following number of negative transactions $n_{X\overline{Y}}$:

$$\boxed{n_{X\overline{Y}} = \sum_{i=1}^{n}(x_i\text{-}x_{min})(y_{max} - y_i)}$$

with x_i and y_i representing respectively values of attributes X and Y for transactions t_i in set T.
We note that this number $n_{X\overline{Y}}$ is called raw ordinal implication measure and not raw numeric implication measure because all the attributes which have a countable and completely ordered number of values can use it. Then, if the values of X or Y are not numerical, we can replace these with an appropriate coding in R. Then, we can also use this measure with ordinal categorical attributes.

Remark

If X and Y take only two values {0,1}, we find the number of negative transactions for intensity of implication, that is to say X and not Y.

3.3 Random model

We must compare this number of negative transactions $n_{X\overline{Y}}$ with the expected number under an assumption of independence. Then, we have to determine the cumulative distribution F(x) = Pr(M≤x) of a random variable M where $n_{X\overline{Y}}$ is an observed value.

For this purpose, let us define a very large population Ω and a sample E of the population of size n. Let U and W be two independent random variables where u_i and v_i are respectively values of variables U and W in set E for i=1,…,n. In order to compare them with attributes X and Y, the random variables U and W must have the same arithmetic means and the same variances as respectively the means and the variances of attributes X and Y. Let m_U, m_W, m_X and m_Y be respectively the arithmetic means of U, W, X and Y; and v_U, v_W, v_X and v_Y be respectively the variances of U, W, X and Y. Then, we have $m_U=m_X$, $m_W=m_Y$, $v_U=v_X$ and $v_W=v_Y$.
Let u_{min} be the minimum value of U in set E and w_{max} be the maximum value of W in set E.

Therefore, $M = \sum_{i=1}^{n}(u_i - u_{min})(w_{max} - w_i)$ is a random variable whose $n_{X\overline{Y}}$ is an observed value. The random variable M can be approximated asymptotically by normal distribution with a mean $m=n\,(m_X–x_{min})(y_{max}\text{-}m_Y)$ and a variance $v=n(v_X+(m_X\text{-}x_{min})^2)(v_Y+(y_{max}\text{-}m_Y)^2)$ (see in appendix A for the demonstration).
We must evaluate whether the number of these negative transactions is high as compared to the random variable M. If this observed number $n_{X\overline{Y}}$ is unusually small compared to the expected number under the assumption of independence, then we can reject the assumption of independence and accept the rule X→Y.

3.4 Statistical test

Let H_0 be the null hypothesis of independence between X and Y and H_1 be the alternative hypothesis. Let α be the significance level[8] (i.e. probability to reject H_0 when it is true).

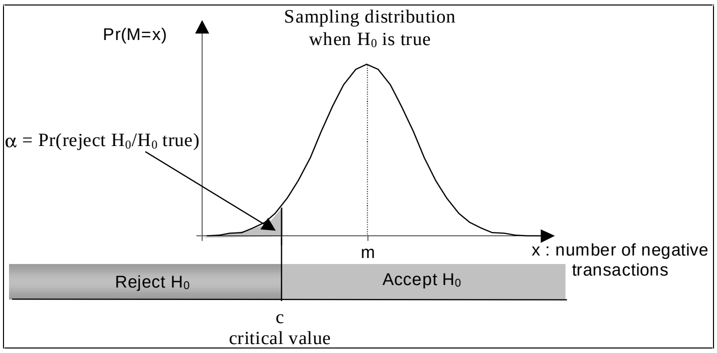

Pr(M=x)
Sampling distribution when H₀ is true
α = Pr(reject H₀/H₀ true)
m
x : number of negative transactions
Reject H₀
Accept H₀
c
critical value

Figure 1 : Statistical test.

Decision rule :

Accept H_0 if $F(n_{X\bar{Y}}) = \Pr(M \le n_{X\bar{Y}}) > \alpha$

Reject H_0 if $F(n_{X\bar{Y}}) = \Pr(M \le n_{X\bar{Y}}) \le \alpha$

It is the observation of the "smallness of $n_{X\bar{Y}}$ compared to the expected number" which is taken as a basic feature of the implication X→Y. If the quantity $\Pr(M \le n_{X\bar{Y}})$ is small, it means that under an assumption of independence we are unlikely to obtain so few negative transactions as compared with a random draw, then this implication is relevant and is evaluated by $1-\Pr(M \le n_{X\bar{Y}}) = 1-F(n_{X\bar{Y}})$.

3.5 Ordinal implication measure

The ordinal implication measure retained, $\varphi(X \to Y)$, between numeric attributes X and Y is, under an assumption of independence, the probability that the random attribute M has a number of negative transactions bigger than the observed number $n_{X\bar{Y}}$.

$$\varphi(X \to Y) = \Pr(M > n_{X\bar{Y}}) = 1 - \Pr(M \le n_{X\bar{Y}}) = 1 - F(n_{X\bar{Y}})$$

Then, the implication X→Y can be admitted at a level of confidence (1-α) if and only if :

$$\varphi(X \to Y) = \Pr(M > n_{X\bar{Y}}) \ge 1-\alpha$$

The ordinal implication measure is :

$$\boxed{\varphi(X \to Y) = \frac{1}{\sqrt{2\pi v}} \int_{n_{X\bar{Y}}}^{+\infty} e^{-\frac{(t-m)^2}{2v}} \, dt}$$

4 Ordinal Rules

In this section, we present ordinal rules. First, we explain what they mean physically and give an example. Next, we show that we can specify these rules in order to capture the behavior of sub-

[8] or the type I error

populations[9]. To finish, we give examples of discovered ordinal rules from simplified banking data.

4.1 Meaning of ordinal rules

Figure 2 shows a two-dimensional scatter diagram for variables X and Y from table 1.

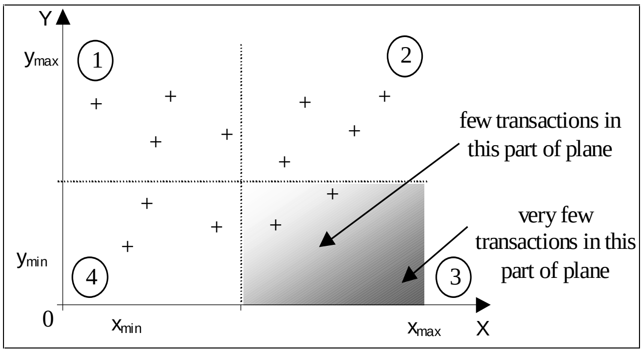

Y
ymax
1
2
few transactions in this part of plane
very few transactions in this part of plane
ymin
4
3
0
xmin
xmax
X

Figure 2 : Scatter diagram for variables X and Y.

Crosses represent transactions of set T as defined in section 3.

Ordinal intensity of implication evaluates, for the rule X→Y, whether the number of transactions in zone 3 is statistically small. However, not all the transactions in this zone have the same importance. Figure 3 shows the graph of $g(x,y)=(x-x_{min})(y_{max}-y)$ where the number $(x_i-x_{min})(y_{max}-y_i)$ for transaction t_i corresponds to $g(x_i,y_i)$. In this figure, we assume that X takes values in [10, 70] and Y in [0, 30].

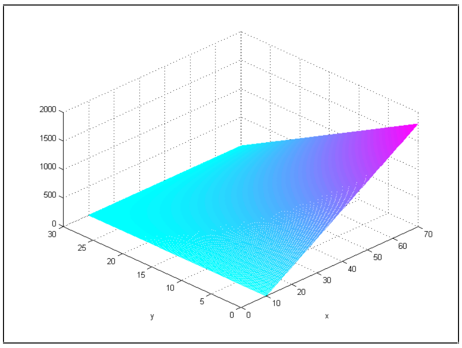


Figure 3 : Graph of g(x,y) showing that not all transactions have the same importance.

We can see that values of g(x,y) are high with high values for x and small values for y; and g(x,y) is maximum with $x=x_{max}$ and $y=y_{min}$ (max{g(x,y)}=g(70,0) for this example).

We could have defined the raw ordinal implication measure as follows $n_{X\overline{Y}} = \sum_{i=1}^{n} x_i(y_{max} - y_i)$, but we would have given too much importance to transactions which have the minimum value x_{min} for X.

In order to obtain a significant ordinal rule X→Y, we must find few transactions in zone 3 and very few transactions with a high value for X and a small value for Y (see dark part in figure 2).

In conclusion, this measure guarantees that if we have a high value for X then we also have statistically a high value for Y, and particularly with very high values for X.

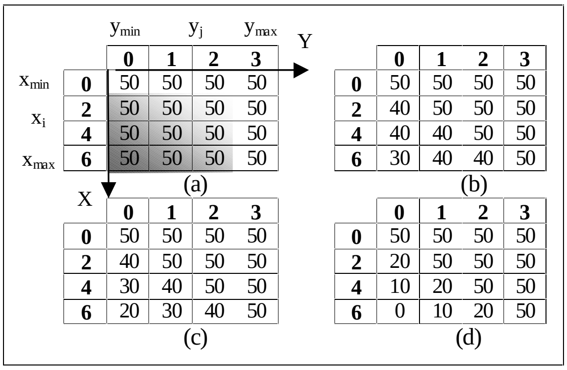
[9] or a sub-set of T.

Then the probability of having a high value for Y when the value for X is high, is significant : Pr(Y being high/ X being high) is significant and particularly with very high values for X.

The discovery of ordinal rules allows us to capture the overall behavior of the population and to obtain a synthesis easily since we do not have a lot of rules.

Examples

We can verify what we said with an example. In this example, attribute X takes values in {0,2,4,6} and Y takes values in {0,1,2,3}.



y_{min}
y_j
y_{max}
Y
x_{min}
x_i
x_{max}
X

(a)

	0	1	2	3
0	50	50	50	50
2	50	50	50	50
4	50	50	50	50
6	50	50	50	50

(b)

	0	1	2	3
0	50	50	50	50
2	40	50	50	50
4	40	40	50	50
6	30	40	40	50

(c)

	0	1	2	3
0	50	50	50	50
2	40	50	50	50
4	30	40	50	50
6	20	30	40	50

(d)

	0	1	2	3
0	50	50	50	50
2	20	50	50	50
4	10	20	50	50
6	0	10	20	50

Table 2 : Four examples where the number of negative transactions decreases (from (a) to (d)).

The intersection between rows x_i and columns y_j represents the number of transactions containing x_i and y_j.
When transactions are uniformly distributed for all values of X and Y (case a) we obtain φ(X→Y)=0.5. The smaller the number of negative transactions, the bigger the ordinal intensity of implication : φ(X→Y)=0.73 (case b), φ(X→Y)=0.89 (case c) and φ(X→Y)=1 (case d).

Refinement of ordinal rules

It is possible to evaluate the "smallness" of the number of transactions in the others zones : the ordinal measure of Y→X will evaluate the number of transactions in zone 1 (see figure 2); the ordinal measure of X→ $\overline{Y}$ will evaluate the number in zone 2; and the measure of $\overline{Y}$ → X will evaluate the number in zone 4. This is graphically represented in the left part of figure 4.
Then, if we have X→Y and Y→X simultaneously, we know that these two attributes are inclined to go in the same direction : to a small value for one of these two attributes corresponds a small value for the other, and to a high value for one of these two attributes corresponds a high value for the other (see left part of figure 4). On the other hand, if we have X→ $\overline{Y}$ and $\overline{Y}$ → X simultaneously, we know that these two attributes are inclined to go in opposite directions (see right part of figure 4).

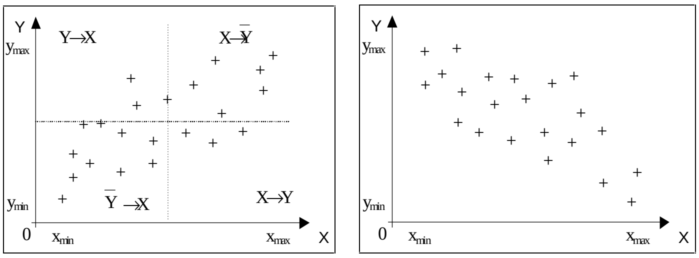


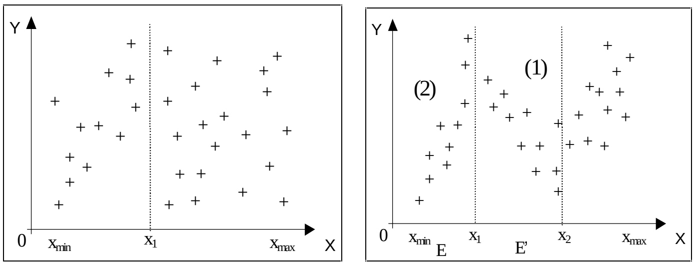
Figure 4 : Meaning of equivalences.

The discovery of such equivalences can reduce the number of attributes in database T, and therefore reduce the complexity of association rules discovery algorithms, since we can keep only one of these two attributes.

4.2 Specific ordinal rules

The discovery of ordinal rules allows us to capture the overall behavior of the population. However, this general behavior may not exist. In this case, there is likely to be a pattern for a sub-set of the population. Then, it could be interesting to find specific ordinal rules. Figure 5 (case (a)) shows such an example : $X \rightarrow Y$ is not significant but $X=[x_{min}, x_1] \rightarrow Y$ is significant. It means that $X \rightarrow Y$ is significant for transactions which have values x_i between x_{min} and x_1, thus for the sub-set $E=\{t_i \in T/x_i \in [x_{min},x_1]\}$.

There is another case where the discovery of specific ordinal rules could be interesting. Though an overall behavior $X \rightarrow Y$ has been discovered for database T, we can have either (1) an opposite behavior for a sub-set of T, or (2) a behavior stronger for a sub-set than for the whole of T. Figure 5 (case (b)) shows these two possibilities (1) and (2). The overall rule $X \rightarrow Y$ is significant but there is an ordinal rule more significant for $E=\{ t_i \in T/x_i \in [x_{min},x_1]\}$ because $\varphi(X=[x_{min},x_1] \rightarrow Y) > \varphi(X \rightarrow Y)$. Moreover, this behavior is opposite for $E'= \{t_i \in T/x_i \in [x_1,x_2]\}$ where $X=[x_1,x_2] \rightarrow \overline{Y}$ and $\overline{Y} \rightarrow X=[x_1,x_2]$ are significant.



(a) (b)

Figure 5 : Examples of interesting specific rules.

It is very important, therefore, to discover these specific rules to capture the complex behavior of the population.

4.3 Example

Consider the simplified banking data in table 3 where there are only ten customers t_1, t_2, …, t_{10} and six attributes : "Age", "Married", "Category", "Years", "Student" and "Income". "Category" determines category of customer : "C_1" represents bad customers, "C_3" good customers and "C_2" average customers. "Years" represents number of years with the bank, "Student" represents the number of student loans and "income" represents total income.

Row ID	Age	Married	Category	Years	Student	income
t_1	20	0	C_1	1	2	2000
t_2	25	0	C_1	2	1	3500
t_3	30	1	C_2	4	2	4000
t_4	35	1	C_2	8	1	7500
t_5	40	1	C_3	11	0	9500
t_6	45	1	C_3	7	0	13500
t_7	50	1	C_3	9	0	15000
t_8	55	1	C_2	12	0	12000
t_9	60	0	C_1	14	0	11000
t_{10}	65	0	C_1	11	0	10500

Table 3 : A simplified banking dataset.

First, we are going to map the ordinal categorical attribute "Category" into numeric attribute : "-1" represents category "C_1", "0" represents category "C_2" and "1" represents category "C_3".

In order to obtain significant rules, we consider that each row represents 10 customers. Implications will be admitted at a 95% level of confidence.

Table 4 shows some discovered overall ordinal rules and table 5 shows some discovered specific rules.

Rules	**$\varphi(X \to Y)$**	**Rules**	**$\varphi(X \to Y)$**
Age→Years	0.997	Years→Age	0.993
Age→$\overline{\text{Student}}$	1	$\overline{\text{Student}}$ →Age	0.986
Age→Income	0.993		

Table 4 : Some discovered global ordinal rules from simplified dataset.

Rules	**$\varphi(X \to Y)$**
Age=[20,50]→Income	0.994
Age=[50,65]→ $\overline{\text{Income}}$	0.995
Age=[20,50]→Category	0.999
Category→Age=[20,50]	0.995
Age=[40,65]→ $\overline{\text{Category}}$	0.993
$\overline{\text{Category}}$ → Age=[40,65]	0.998
Age=[20,55]→Married	0.999

Table 5 : Some discovered specific rules from simplified database.

Equivalence between "Age" and "Years" shows that the older the customer, the longer he will have been with the bank. This relationship indicates loyalty of customers.

Equivalence between "Age" and "$\overline{\text{Student}}$" indicates that this kind of loan is taken by young customers.

Generally, income increases with age, particularly between 20 and 50 years, and decreases a little after 50 years.

The older the customer, the better the customer and this is true for customers between 20 and 50 years. This trend is reversed for customers between 40 and 65 years.

5 Evaluation on Banking Data

In this section, we present experimental tests on a banking database. First, in section 6.1, we describe the banking database and then, in section 5.2, we give results.

5.1 Banking Data

The banking database consists of 47,112 transactions described by 52 attributes, 96% of which are numeric attributes.

Attributes can be broken down into three categories :

- information about customers (age, number of years with bank, …),
- information about various accounts opened with the bank (bonds, mortgages, savings accounts, …) and
- statistics about various accounts (rate of indebtedness, total income, …).

Information about various accounts can also be broken down into two categories :

- attributes representing total of balances of various accounts of a given customer and
- attributes representing the number of accounts opened by the customer for each financial service proposed by the bank.

5.2 Results

First, we discuss results of the discovery of general ordinal rules and give some rules and some strong relationships. Next, we do the same thing with specific ordinal rules and discuss the problem of the number of discovered specific rules.

General ordinal rules

208 relevant ordinal rules at a level of α=5% have been discovered with the ordinal intensity of implication.

This experiment has allowed us to discover financial services likely to interest customers who have a given type of account. For example, we have discovered that customers who have a savings account are potentially interested in the following services : house purchase savings plan with an ordinal intensity equal to 1, house purchase savings account with an ordinal intensity equal to 1, permanent overdraft facility with an ordinal intensity equal to 0,95, credit card (0,96) and the individual savings plan (1).

This kind of ordinal implications have been discovered for all financial services, thus 76 rules have been extracted. We have also extracted seven general rules of the kind $X \rightarrow \overline{Y}$ showing services not likely to interest customers who have a given type of account. For example, we have discovered that customers who have bonds or stocks are not potentially interested in borrowing money with respective ordinal intensities of implication equal to 0.96 and 0.98.

Other examples of discovered general rules : the higher the number of stock market investment savings account, the longer the customer will have been with the bank with an ordinal measure equal to 0.98; the higher the number of individual savings plans (PEP), the older the customer with a measure equal to 0.95.

Now, we shall present the strong relationships that we have discovered (equivalences) between attributes.

- An equivalence between customer's age and number of years with the bank. This relation means that in general the older the customer, the longer he will have been with the bank. ($\varphi(\text{Age}\rightarrow\text{Years})=1$ and $\varphi(\text{Years}\rightarrow\text{Age})=1$).
- Moreover, many equivalences have been discovered between the total of the balances of various accounts for a given service X (balancesX) and the number of accounts opened for this service X (numberX). This kind of relationship has been discovered for 50% of the

financial services offered by the bank. For example, this is true for savings plans and individual savings plans.

These relationships are very interesting because they can reduce the complexity of the problem : we can proceed further with a single attribute : balancesX or numberX.

- To finish, a very strong relationship has been discovered between the house purchase savings plan (PEL) and the house purchase savings account (CEL) as can be seen in the following figure :

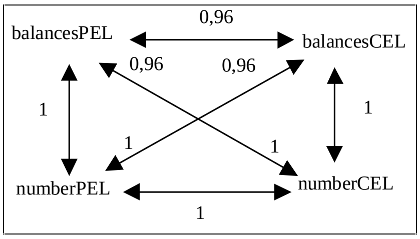


Figure 6 : Strong relationship between the PEL and the CEL.

This relationship proves that in general a customer subscribes to these two services.

Specific Ordinal Rules

We have found interesting rules where the general rule is not significant like, for example, customers who have between 0 and 2 investment trust accounts "SICAV FCP SCPI" are potentially interested in bonds; the maximum value for the number of investment trust accounts is equal to 8.

6 Conclusion and Further Work

We have found a selective measure of implication for ordinal attributes in large databases : the ordinal intensity of implication. This measure allows us to extract some information quality and to discard the transformation step of initial attributes into binary attributes for discovering association rules. Moreover, it allows us to discover a new kind of rules : ordinal rules which reveal the joint evolution of attributes in the same direction or in opposite directions. Discovery of ordinal general rules reveals the overall behavior of the population as well as a synthesis of its behavior more easily than would be possible with rules whose attributes are split into intervals (we have at the most 2p(2p-1) rules, where p is the number of attributes).

However, it is not sufficient and we need to know the behavior of sub-sets of the database. A first answer has been given with specific ordinal rules. This study has to be extended with, for example, ordinal association rules, i.e. ordinal rules $X \rightarrow Y$ where X and Y are conjunctions of either attributes or intervals of attributes and $X \cap Y = \emptyset$.

Appendix A : Proof of random model

Let U' and W' be two independent random variables with values (u_i' i=1…n) and (w_i' i=1…n) in interval [0,1].

[Lagrange, 1997] proved that the random variable given by $Z'=\frac{\sum_i U'_i(1-W'_i)}{n}$ where U'_i and W'_i have the same distribution as U' and W' respectively, can be approximated by normal distribution with a mean $m'=E[(U'(1-W'))]$[10] and a variance $v'=\frac{E[(U'(1-W'))^2]}{n}$.

Let U and W be two new independent random variables with respective values in interval $[u_{min}, u_{max}]$ and $[w_{min}, w_{max}]$.

We have the following relationships between random variables : $U'=U/u_{max}$ and $W'=W/w_{max}$.

Therefore, $Z'=\frac{\sum_i \frac{U_i}{u_{max}}(1-\frac{W_i}{w_{max}})}{n}=\frac{\sum_i U_i(W_{max}-W_i)}{n\times u_{max}w_{max}}$ can be approximated by normal distribution with a mean $\frac{E[U(w_{max}-W)]}{u_{max}w_{max}}$ and a variance $\frac{E[(U(w_{max}-W))^2]}{n\times u_{max}^2 w_{max}^2}$

As U and W are independent, we have :

$E[U(w_{max}-W)]=E(U)(w_{max}-E(W))$ and

$E[(U(w_{max}-W))^2]=E(U^2)E((w_{max}-W)^2)$

$=[Var(U)$[11]$+E(U)^2][Var(W)+(w_{max}-E(W))^2]$

(König-Huyghens[12]' formula)

Then, the random variable $Z=n\times u_{max}w_{max}\times Z'=\sum_i U_i(w_{max}-W_i)$ can be approximated by normal distribution with a mean $nE(U)(w_{max}-E(W))$ and a variance $n(Var(U)+E(U)^2)(Var(W)+(w_{max}-E(W))^2)$.

In conclusion, the random variable $\sum_{i=1}^{n}(U_i - u_{min})(w_{max}-W_i)$ can be approximated by normal distribution with a mean $n[E(U)-u_{min})(w_{max}-E(W)]$ and a variance $n[Var(U)+(E(U)-u_{min})^2][Var(W)+(w_{max}-E(W))^2]$.

References

Agrawal, R.; Imielinski, T.; and Swami A. 1993. Mining Association Rules between Sets of Items in Large Databases. In *Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data*, 207-216, May.

Agrawal, R.; Mannila, H.; Srikant, R.; Toivonen H.; and Verkamo. 1996. A.I. Fast Discovery of Association Rules. In Fayyad U.M., Piatetsky-Shapiro G., Smyth P. and Uthurusamy R. eds., Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press. Chapter 12, 307-328.

Brin, S.; Motwani, R.; and Silverstein, C. 1997a. Beyond Market Baskets : Generalizing Association Rules to Correlations. In *Proc. Of the 1997 ACM SIGMOD Conference*, 265-276, Tucson, Arizona, USA, May.

[10] E(U) is another notation for the arithmetic mean of the random variable U.

[11] Var(U) is another notation for the variance of the random variable U.

[12] Let U be a random variable and let "a" be a constant, we have $Var(U) = E[(U-a)^2] - [E(U)-a]^2$

Brin, S.; Motwani, R.; Ullman, J. D.; and Tsur, S. 1997b. Dynamic Itemset Counting and Implications Rules for Market Basket Data. In *Proc. Of the 1997 ACM SIGMOD Conference*, 255-264, Tucson, Arizona, USA, May.

Fleury, L.; Briand, H.; Philippe, J.; and Djeraba, C. 1995. Rules Evaluations for Knowledge Discovery in Database. In *Proc. of 6th International Conference and Workshop on Database and Expert Systems Applications*, DEXA, London.

Fayyad, U.; Piatetsky-Shapiro, G.; and Smyth, P. 1996. From Data Mining to Knowledge Discovery in Databases. AI Magazine, 37-54.

Frawley, W.J.; Piatetsky-Shapiro, G.; and Matheus, C.J. 1991. Knowledge Discovery in Databases : an Overview. In Piatetsky-Shapiro G. et Frawley W.J., editors, Knowledge Discovery in Databases, 1-27. AAAI Press / The MIT Press, Menlo Park, CA.

Ganascia, J.G. 1987. Agape et Charade : Deux Techniques d'Apprentissage Symbolique Appliquées à la Construction de Bases de Connaissances, Thèse d'Etat, Université Paris-Sud, Mai.

Ganascia, J.G.; Puget, J.F.; and Helft, N. 1990. Comportement des Systèmes d'Apprentissage, *3ème journées nationales PRC-GDR Intelligence artificielle*, 2-7 Mars.

Goodman, R.M.; and Smyth, P. 1989. The Induction of Probabilistic Rules Set, the ITRULE Algorithm. In *Proceedings of Sixth International Workshop on Machine Learning*, Spatz, B., ed, 129-132, San Mateo, CA : Morgan Kaufmann.

Gras, R. 1979. Contribution à l'Etude Expérimentale et à l'Analyse de certaines Acquisitions Cognitives et de certains Objectifs Didactiques en Mathématiques, Thèse d'Etat, Université de Rennes I, October.

Gras, R.; Almouloud, S.A.; Bailleul, M.; Larher, A.; Polo, M.; Ratsimba-Rajohn, H. and Totohasina, **A**. 1996. L'implication Statistique, Ouvrage de 320 pages dans la Collection Associée à "Recherches en Didactique des Mathématiques", La Pensée Sauvage, Grenoble.

Guillaume, S.; Guillet, F.; and Philippé, J. 1998. Improving the Discovery of Association Rules with Intensity of Implication, In *Second European Symposium on Principles of Data Mining and Knowledge Discovery* (PKDD'98), 318-327, Nantes, France.

Guillaume S., Khenchaf A., 2000 Generalizing Association Rules with Ordinal Intensity of Implication, In proceedings of *Information Ressources Management Association International Conference*, IRMA2000, Anchorage, Alaska, USA, May 2000.

Houtsma, M.; and Swami, A. 1995. Set-oriented mining of association rules. In *Int'l Conference on Data Engineering*, Taipei, Taiwan, March.

Klemettinen, M.; Mannila, H.; Ronkainen, P.; Toivonen, H.; and Verkamo, A.I. 1994. Finding Interesting Rules from Large Sets of Discovered Association Rules. *Proceedings of the Third International Conference on Information and Knowledge Management*, 401-407.

Lagrange, J.B. 1997. Analyse Implicative d'un Ensemble de Variables Numériques ; Application au Traitement d'un Questionnaire à Réponses Modales Ordonnées. Prépublication 97-32 de l'Institut de Recherche Mathématiques de Rennes, 1-27, Décembre.

Lerman, I.C. 1981. Classification et analyse ordinale des données, Dunod.

Liu, B.; and Hsu W. 1996. Post-analysis of Learned Rules. AAAI-96, 828-834.

Major, J.; and Mangano J. 1993. Selecting among Rules Induced from a Hurricane Database. KDD-93, 28-41.

Mannila, H.; Toivonen, H.; and Verkamo, A.I. 1994. Efficient algorithms for Discovering Association Rules. *Usama M. Fayyad et Ramasamy Uthurusamy, éditeurs, AAAI Workshop on Knowledge Discovery in Databases*, 181-192, Seattle, Washington, July.

Park, J.S.; Chen, M.S.; and Yu, P.S. 1995. An Effective Hash-Based Algorithm for Mining Association Rules. In *Proc. 1995 ACM-SIGMOD Int. Conf. Management of Data*, San Jose, CA, May.

Pavillon, G. 1992. Un Algorithme d'Apprentissage des Relations de Dépendance et des Relations Causales, rapport MASI RR92/96, Décembre.

Piatetsky-Shapiro, G.; and Frawley, W.J. 1991. editors. Knowledge Discovery in Databases. AAAI Press / The MIT Press, Menlo Park, CA.

Piatesky-Shapiro, G.; and Matheus, C.J. 1994. The Interestingness of Deviations. In Usama M. Fayyad and Ramasamy Uthurusamy, editors, AAAI Workshop on Knowledge Discovery in Databases, 25-36, Seattle, Washington, July.

Piatetsky-Shapiro, G. 1991. Discovery, Analysis, and Presentation of Strong Rules. In Piatetsky-Shapiro G. et Frawley W.J., editors, Knowledge Discovery in Databases, 229-248. AAAI Press / The MIT Press, Menlo Park, CA.

Piatesky-Shapiro, G.; Matheus, C.; Smyth, P.; and Utharusamy, R. 1994. KDD-93 : Progress and Challenge …, AI Magazine, Fall, 77-87.

Saporta, G. 1990. *Probabilités, Analyse des Données et Statistique,* Editions Technip.

Schektman, Y.; Trejos, J.; and Troupe, M. 1992. Un générateur de règles floues à partir des bases de données volumineuses, 3èmes journées symboliques-numériques, 121-143, Paris, France, May.

Silberschatz, A.; and Tuzhilin, A. 1996. What makes Patterns interesting in Knowledge Discovery Systems. IEEE Trans. On Know. And Data Eng. 8(6) :, 970-974.

Suzuki, E.; and Kodratoff, Y. 1998. Discovery of Surprising Exception Rules Based on Intensity of Implication, In *Second European Symposium on Principles of Data Mining and Knowledge Discovery* (PKDD’98), 10-18, Nantes, France.