INFORMATION CONTENT OF ASSOCIATION RULES

Arnost Vesely[*]

Abstract: Database records can be often interpreted as state descriptions of some world, system or generic object, states of which occur independently and are described by binary properties. If records do not contain missing values, then there exists close relationship between association rules and propositions about state properties. In data mining we usually get a lot of association rules with large confidence and large support. Since their interpretation is often cumbersome, some quantitative measure of their "informativeness" would be very helpful.

The main aim of the paper is to define a measure of the amount of information contained in an association rule. For this purpose we make use of the tight correspondence between association rules and logical implications. At first a quantitative measure of information content of logical formulas is introduced and studied. Information content of an association rule is then defined as information content of the corresponding logical implication in the situation when no knowledge about dependence among properties of world states is at our disposal. The intuitive meaning of the defined measure is that the association rule that allows more appropriate correction of the distribution of world states, acquired under unfair assumption of independence of state properties, contains also larger amount of information. The more appropriate correction here means a correction of the current probability distribution of states that leads to the distribution that is closer to the true distribution in the sense of Kullback-Leibler divergence measure.

Key words: *Association rules, information divergence, information content of logical formulas, information content of association rules*

Received: February 28, 2013 **DOI:** 10.14311/NNW.2014.24.014
Revised and accepted: May 12, 2014

1. Introduction

In data mining different methods for gaining knowledge are studied and used. The mining of association rules is one of the most successful. The purpose of mining association rules is to discover the associations among data in large databases or data sets, i.e. to find items that imply the presence of other items in the same database records or transactions. Association rules were firstly introduced by Agrawal [1] and then successfully applied by many authors.

[*]Arnost Vesely, Czech University of Life Sciences, Kamycka 129, Prague, Czech Republic, Tel.: +420 22438 2215, E-mail: vesely@pef.czu.cz

Having association rules, the main problem is to interpret them in some useful way. One of the most appealing way of interpretation is to interpret them as logical formulas, because logic provide main means for scientific reasoning. In paragraphs 2-4 the interpretation of association rules as logical formulas is systematically pursued. We consider databases that contain only records with binary items that can be interpreted as truth values of predicates and we assume that these predicates denote properties describing independently realized states of some world, system or generic object. In paragraph 2 we prove that if the database does not contain missing data, then knowing association table (contingency table) of an association rule $\alpha \Rightarrow \gamma$ brings in a possibility to interpret the association rule as an arbitrary logical formula built up from the antecedent α and consequent γ. Moreover, the probability that the formula is valid can be estimated from the standard parameters of the association rule, namely from its confidence cf, support sup and consequent coverage cov_Y. Moreover, if the association rule is interpreted as implication, then knowing only parameters cf and sup is sufficient (see also [2]).

During data mining process we usually get a lot of association rules with large confidence $cf \cong 1$ and support $sup > sup_{min}$. To simplify their interpretation, we should select only those of them that are "informative" in some way. So we need some quantitative measure of "informativeness" of an association rule. A lot of measures based on contingency table have been devised, see e.g. Tan [3] or Blanchard [4]. In this article we propose a quite different and original approach. We establish a quantitative information measure for logical formulas and then we utilize the close relationship between association rules and logical implications.

Quantitative measure for evaluating information content of logical formulas based on Shannon entropy was firstly proposed in [5] and [6]. To behave efficiently in the surrounding world means to have a good knowledge of probabilities of the states observed in this world, or of the events occurring there. Therefore the formula enabling to specify more precisely an unknown true distribution of these events or states should be considered more "informative" than a formula leading to a less precise specification. Assume that we measure the distance between a priori distribution q of the world states and their true distribution p. The distance between distributions is measured in the standard information-theoretic manner, namely by the Leibler-Kullback divergence and is denoted $D(p||q)$. After we find out that some formula β, expressing logical dependence among the world states is valid, the a priori distribution q can be transformed into a posteriori distribution $\tilde{q}$ as it is specified in §5. The main result of §5 asserts that the distance between a posteriori distribution $\tilde{q}$ and the true distribution p (denoted $D(p||\tilde{q})$) is less than $D(p||q)$. The information content of formula β is then defined

$$I(\beta, q) = D(p||q) - D(p||\tilde{q}) \geq 0.$$

In §5 we prove that for evaluation of $I(\beta,\, q)$ the knowledge of true distribution p is not necessary and that the value $I(\beta, q)$ depends only on the probability that formula β is valid (denoted $[\beta]$) and on the a priori distribution q. The soundness of the defined measure stems from the fact that the more informative formula enables to get more precise estimate $\tilde{q}$ of the true probability distribution p.

In data mining the association rules are used to support expert's decision-making. As we show in paragraph 4, the most important way, how association rules could be interpreted, is to interpret them as logical implications. Since association rules can be in natural way interpreted as logical implications and we have at our disposal a quantitative measure of information content of logical formulas, we may apply this information measure, originally devised for logical formulas, also to association rules.

Association rules reflect dependence among properties of world states. The more this dependence is reflected, the more "informative" the association rule is supposed to be. Before mining of association rules we may assume to be in situation, in which we have no knowledge of the dependence that among properties exists. Assume for a while their mutual independence. Under this assumption the probability distribution of world states q^* can be easily evaluated from the database (see paragraph 6). Assume further that association rule $\alpha \Rightarrow \gamma$ with confidence $cf \cong 1$ and support $sup > sup_{\min}$ has been mined. The association rule can be interpreted in natural way as logical formula $\alpha \supset \gamma$ and the probability $[\alpha \supset \gamma]$ that the formula $\alpha \supset \gamma$ is valid in the given world can be estimated as $[\alpha \supset \gamma] \cong 1 + sup - sup/cf$ (see paragraph 2). The larger the information content $I(\alpha \supset \gamma, q^*)$ of the formula $\alpha \supset \gamma$ is, the more precise correction of the distribution q^*, obtained under likely false assumption of independence of properties of world states, might be carried out (see paragraph 5). Thus the natural measure of information content of an association rule is

$$I(\alpha \Rightarrow \gamma) = I(\alpha \supset \gamma, q^*).$$

2. Association Rules and Logical Formulas

At the beginning of this paragraph we introduce association rules and their parameters confidence (cf), support (sup), antecedent coverage (cov_X) and consequent coverage (cov_Y) in the standard way following the seminal work [Agrawal 1994]. Then we show how association rules can be interpreted as logical formulas. Our thorough analysis results in proving that an association rule $\alpha \Rightarrow \gamma$ can be interpreted as an arbitrary logical formula that is build up from α and γ. Moreover, we prove that probability of its validity can be estimated from parameters cf, sup, and cov_Y, provided the transaction set is complete.

Assume that $I = \{i_1, i_2, \ldots, i_m\}$ is a set of labels, called items. Let $\mathbf{T}$ be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. We say that a transaction T contains X, a set of some items in I, if $X \subseteq T$.

An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$.

The rule $X \Rightarrow Y$ holds in the transaction set $\mathbf{T}$ with confidence cf if $(cf \times 100)\%$ of transactions in $\mathbf{T}$ that contain X also contain Y.

The rule $X \Rightarrow Y$ has support sup in the transaction set $\mathbf{T}$ if $(sup \times 100)\%$ of transactions in $\mathbf{T}$ contain $X \cup Y$.

The rule $X \Rightarrow Y$ has antecedent coverage cov_X in the transaction set $\mathbf{T}$ if $(cov_X \times 100)\%$ of transactions in $\mathbf{T}$ contain X.

The rule $X \Rightarrow Y$ has consequent coverage cov_Y in the transaction set $\mathbf{T}$ if $(cov_Y \times 100)\%$ of transactions in $\mathbf{T}$ contain Y.

Assume an world Ω and $P = \{P_1, \dots, P_m\}$ a set of properties describing world states $\omega \in \Omega$. Assume that world states realize in random and that their realizations are mutually independent. Then world Ω can be in one of $n = 2^m$ states $\omega_1, \dots, \omega_n$ and each state ω_i can be unequivocally described with the conjunction

$$\alpha_i = L_1 \wedge \dots \wedge L_m,$$

where for all $j = 1, \dots, m, L_j = P_j$ if the world state ω_i has the property P_j or $L_j = \neg P_j$ otherwise.

According to the above definitions $I = \{P_1, \neg P_1, \dots, P_m, \neg P_m\}$ constitutes an itemset and $T \subset I, T = \{L_1, \dots, L_m\}$, where $L_i = P_i$ or $L_i = \neg P_i$, $i = 1, \dots, m$ are transactions of a special type. We will call them complete transactions. A transaction set consisting of complete transactions will be called complete transaction set.

An arbitrary formula φ created from literals $L_1, L_2, \dots, L_m$ by means of propositional operators $\wedge$, $\vee$, $\supset$, $\neg$, $\equiv$ is or is not valid in a transaction T. Assume a complete transaction set $\mathbf{T}$ that consists of N transactions. If φ is valid in m transactions of $\mathbf{T}$, then

$$[\phi]_T = \frac{m}{N}$$

is validity of φ in the transaction set $\mathbf{T}$. Formula φ can be interpreted as an assertion about world Ω. The probability that φ is valid in randomly chosen state of world will be denoted $[\varphi]$. Clearly, for large N the value $[\varphi]_T$ can be considered being an estimate of the probability $[\varphi]$.

Definition 1 Let $X \Rightarrow Y, X = \{K_1, \dots K_m\}, Y = \{M_1, \dots M_p\}$ be an association rule on a complete transaction set $\mathbf{T} = \{T_1, \dots, T_N\}$. Denote $\alpha = K_1 \wedge \dots \wedge K_m$, $\gamma = M_1 \wedge \dots \wedge M_p$. Association table $A_{X \Rightarrow Y}$ of the association rule $X \Rightarrow Y$ are four real numbers (a, b, c, d), where

$$a = \frac{a'}{N},\ b = \frac{b'}{N},\ c = \frac{c'}{N},\ d = \frac{d'}{N} \text{ and}$$

a' is the number of world states in which both α and γ are valid,
b' is the number of world states in which α is valid and γ is not valid,
c' is the number of world states in which α is not valid and γ is valid,
d' is the number of world states in which neither α nor γ are valid.

Association table is often given in the form of the table:

	γ	$\neg\gamma$
α	a	b
$\neg\alpha$	c	d

Obviously equation $a + b + c + d = 1$ holds. Association rule $X \Rightarrow Y$ will be also written as $\alpha \Rightarrow \gamma$.

Lemma 1 Let $X \Rightarrow Y$ be an association rule on a complete transaction set $\mathbf{T} = \{T_1, \dots, T_N\}$ and let $A_{X \Rightarrow Y} = (a, b, c, d)$ be its association table. Let *sup*,

cf, and cov_Y be support, confidence and consequent coverage of the rule $X \Rightarrow Y$respectively. Then if $sup \neq 0$, the following holds:

$$\begin{aligned} &\text{a)} \quad & sup &= a, & (1)\\ &\text{b)} \quad & cf &= a/(a+b), & (2)\\ &\text{c)} \quad & cov_Y &= a+c. & (3) \end{aligned}$$

Proof

a) According to Definition 1 α and γ are both valid in $a' = aN$ transactions from $\mathbf{T}$. It means that aN transactions from $\mathbf{T}$ contain $X \cup Y$. Therefore support of the rule $X \Rightarrow Y$ equals to a.

b) According to Definition 1 α is valid in $a'+b' = (a+b)N$ transitions but only in aN of them also γ is valid. In other words only $Na/(a+b)$ transactions from $\mathbf{T}$ have the property that with X also contain Y. Therefore $cf = a/(a+b)$.

c) According to Definition 1 γ is valid in $a' + c' = N(a+c)$ transactions. Therefore $cov_Y = a + c$.

Notice that if $sup \neq 0$, then also $cf \neq 0$. ■

Lemma 2 Let $X \Rightarrow Y$ be an association rule on a complete transaction set $\mathbf{T}$ and let $A_{X \Rightarrow Y} = (a, b, c, d)$ be its association table. Let sup, cf, and cov_Y be support, confidence and consequent coverage of the rule $X \Rightarrow Y$ respectively. Then if $sup \neq 0$, the following holds:

$$\begin{aligned} &\text{a)} \quad & a &= sup, & (4)\\ &\text{b)} \quad & b &= sup(1-cf)/cf, & (5)\\ &\text{c)} \quad & c &= cov_Y - sup, & (6)\\ &\text{d)} \quad & d &= 1 - cov_Y - sup(1-cf)/cf, & (7) \end{aligned}$$

Proof

a) See Lemma 1.a.

b) According to Lemma 1.b

$$cf = \frac{a}{a+b}.$$

Therefore

$$b = \frac{a(1-cf)}{cf} = sup\frac{(1-cf)}{cf}.$$

c) $c = cov_Y - a = cov_Y - sup$ (from (3) and (1)).

d) $d = 1 - a - b - c = 1 - cov_Y - sup(1-cf)/cf$. ■

Lemma 3 Assume association rule $X \Rightarrow Y$, $X = \{K_1, \dots K_m\}$, $Y = \{M_1, \dots M_p\}$ and denote $\alpha = K_1 \wedge \dots \wedge K_m$, $\gamma = M_1 \wedge \dots \wedge M_p$. If $sup \neq 0$, then

$$\begin{aligned} &\text{a)} \quad [\alpha \supset \gamma]_T = 1 + sup - sup/cf, &(8)\\ &\text{b)} \quad [\alpha]_T = sup/cf, &(9)\\ &\text{c)} \quad [\alpha \wedge \gamma]_T = sup. &(10)\end{aligned}$$

Proof

a) According to Def. 1 we have $[\alpha \supset \gamma]_T = a + c + d = 1 - b$. Hence, $b = 1 - [\alpha - \gamma]_T$.

According to (5)

$$\frac{b}{sup} = \frac{(1 - cf)}{cf}.$$

Therefore

$$\frac{1 - [\alpha \supset \gamma]_T}{sup} = \frac{(1 - cf)}{cf}$$

and

$$[\alpha \supset \gamma]_T = 1 + sup - \frac{sup}{cf}.$$

Proofs of b) and c) are obvious. ■

Assume an association rule $X \Rightarrow Y$ with association table $A_{X \Rightarrow Y}$, where $X = \{K_1, \dots K_m\}$, $Y = \{M_1, \dots, M_p\}$, $\alpha = K_1 \wedge \dots \wedge K_m$, $\gamma = M_1 \wedge \dots \wedge M_p$, defined on a complete transaction set **T**. Obviously, using logical connectives indefinite number of formulas can be built up from α and γ. From Tab. I one can easily see that all these formulas are members of 16 classes of equivalence such that any two formulas in the same class are logically equivalent and any two formulas from different classes are not. Any formula inside the class can represent the class and its validity in **T** can be easily computed from parameters a, b, c, d of the association

α	γ	$\alpha \wedge \neg\alpha$	$\alpha \wedge \gamma$	$\alpha \wedge \neg\gamma$	α	$\neg\alpha \wedge \gamma$	γ	$\alpha \not\equiv \gamma$	$\alpha \vee \gamma$	$\neg\alpha \wedge \neg\gamma$	$\alpha \equiv \gamma$	$\neg\gamma$	$\alpha \vee \neg\gamma$	$\neg\alpha$	$\alpha \supset \gamma$	$\neg\alpha \vee \neg\gamma$	$\alpha \vee \neg\alpha$
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
validity		0	a	b	$a+b$	c	$a+c$	$b+c$	$a+b+c$	d	$a+d$	$b+d$	$a+b+d$	$c+d$	$a+c+d$	$b+c+d$	1

Tab. I *Classes of equivalence of formulas built up from α and γ of the association rule $\alpha \Rightarrow \gamma$.*

table $A_{X\Rightarrow\ Y}$. For example all formulas of the class represented by formula $\alpha \supset \gamma$ have validity equal to $[\alpha \supset \gamma]_T = a + c + d$. The sufficient and necessary condition for a formula from this class to be valid in whole transition set $\mathbf{T}$ is $a + c + d = 1$. Obviously, the equivalent condition to the previous one is $b = 0$.

Theorem 1 Assume an association rule $X \Rightarrow Y$, where $X = \{K_1, \ldots, K_m\}$, $Y = \{M_1, \ldots, M_p\}$, $\alpha = K_1 \wedge \cdots \wedge K_m$, $\gamma = M_1 \wedge \cdots \wedge M_p$, defined on a complete transaction set $\mathbf{T}$. Let $X \Rightarrow Y$has confidence cf, support sup and consequent coverage cov_Y. Then validity in $\mathbf{T}$ of any formula built up from α and γ by means of logical connectives can be evaluated using sup, cf and cov_Y.

Proof The validity of any formula that has been built up from α and γ can be evaluated from parameters a, b, c, d of the association rule $X \Rightarrow Y$ according to Tab. I. Values of a, b, c, d can be evaluated from sup, cf and cov_Y according to Lemma 2. ■

Example 1 Assume a set of patients described with 7 properties *HD*, *LD*, *SBP*, *DBP*, *OLD*, *FAT*, *IM* with the following meaning.

HD... Patient has high level of high density cholesterol.
LD... Patient has high level of low density cholesterol.
SBP... Patient has high value of systolic blood pressure.
DBP... Patient has high value of diastolic blood pressure.
OLD... Patient is older than the given age (e.g. 60 years).
FAT... Patient's body mass index BMI is greater than 30.
IM... Patient suffered from myocardial infarction in the next 2-years period after examination.

We can consider the patients being realizations of a generic patient. Generic patient here constitutes world Ω and particular patients constitute realizations of world states. Suppose we have at our disposal 10 patients with descriptions given in Tab. II.

Patient	Patient's description						
T_1	HD	LD	SBP	$\neg DBP$	OLD	FAT	IM
T_2	$\neg HD$	LD	SBP	DBP	OLD	FAT	IM
T_3	$\neg HD$	LD	SBP	$\neg DBP$	OLD	FAT	IM
T_4	HD	LD	$\neg SBP$	$\neg DBP$	OLD	FAT	$\neg IM$
T_5	$\neg HD$	$\neg LD$	SBP	$\neg DBP$	$\neg OLD$	$\neg FAT$	$\neg IM$
T_6	$\neg HD$	LD	SBP	$\neg DBP$	OLD	FAT	$\neg IM$
T_7	$\neg HD$	LD	SBP	DBP	$\neg OLD$	FAT	IM
T_8	$\neg HD$	LD	SBP	DBP	OLD	$\neg FAT$	IM
T_9	$\neg HD$	$\neg LD$	SBP	$\neg DBP$	$\neg OLD$	$\neg FAT$	$\neg IM$
T_{10}	$\neg HD$	LD	SBP	DBP	OLD	FAT	$\neg IM$

Tab. II *Small patient database that constitutes complete transaction set.*

From the point of view of association rule theory the properties constitute itemset $I = \{HD, \neg HD, LD, \neg LD, SBP, \neg SBP, DBP, \neg DBP, OLD, \neg OLD, FAT, \neg FAT, IM, \neg IM\}$ and patient's descriptions $T_1, \ldots, T_{10}$ complete transaction set $\mathbf{T}$ that contain 10 transactions.

Formula $LD \wedge OLD \wedge FAT \supset SBP$ is not valid only in T_4 and therefore $[LD \wedge OLD \wedge FAT \supset SBP]_T = 0.9$. From Tab. II we see that association rule $X \Rightarrow Y$, $X = \{LD, OLD, FAT\}, Y = \{SBP\}$ has support $sup = 0.5$ and confidence $cf = 5/6 = 0.83$. According to Lemma 3 the probability estimate of validity of the formula $LD \wedge OLD \wedge FAT \supset SBP$ is

$$[LD \wedge OLD \wedge FAT \supset SBP]_T = 1 + sup - \frac{sup}{cf} = 0.9.$$

On the base of the Tab. II also association table $A_{X \Rightarrow Y} = (a, b, c, d)$ can be established and we are getting $A_{X \Rightarrow Y} = (0.5, 0.1, 0.4, 0)$. From Tab. I we will get the same value of the formula probability estimate

$$[LD \wedge OLD \wedge FAT \supset SBP]_T = a + c + d = 0.9.$$

As another example let us take formula $LD \wedge \neg HD \wedge SBP \supset IM$. It is not valid only in the states T_6 and T_{10} and therefore $[LD \wedge \neg HD \wedge SBP \supset IM]_T = 0.8$. The association rule $X \Rightarrow Y, X = \{LD, \neg HD, SBP\}$, $Y = \{IM\}$ has support $sup = 0.4$ and confidence $cf = 2/3 = 0.66$ (see Tab. II). Therefore

$$[LD \wedge \neg HD \wedge SBP \supset IM]_T = 1 + sup - \frac{sup}{cf} = 0.8.$$

The association table of this rule is $A_{X \Rightarrow Y} = (0.4, 0.2, 0.1, 0.3)$ and therefore according to Tab. I

$$[LD \wedge \neg HD \wedge SBP \supset IM]_T = a + c + d = 0.8.$$

3. Confidence Intervals

Assume that $[\alpha]_T$ is validity of a formula α in a complete transaction set $\mathbf{T}$. Suppose that transaction set $\mathbf{T}$ has N transactions that are interpreted as states of world Ω. If we take $[\alpha]_T$ as an estimate of the probability that formula α is valid in randomly chosen world state $\omega \in \Omega$, how accurate this estimate can be?

For answering this question we shall use the standard statistical technique of confidence intervals. Assume $[\alpha]$ be the probability that α is valid in the world Ω and let r be the number of transactions in which α is valid. The random variable X that takes r as its value follows Binomial distribution $\mathrm{Bi}(N, [\alpha])$ with mean

$$\mu_X = N[\alpha]$$

and standard deviation

$$\sigma_X = \sqrt{N[\alpha](1 - [\alpha])}.$$

If $N[\alpha](1 - [\alpha]) \geq 5$ the Binomial distribution is very well approximated with Normal distribution and the confidence interval for X is $\mu_X \pm z_c \sigma_X$, where z_c

is the constant that defines the width of the smallest interval about the mean that includes $100c\%$ of the total probability mass under the bell-shaped Normal distribution (e.g. for $c = .95$ the value is $z_c = 1.96$). If we take $[\alpha]_T$ as an estimator of $[\alpha]$ we are getting

$$\tilde{\mu}_X = N[\alpha]_T,$$

$$\tilde{\sigma}_X = \sqrt{N[\alpha]_T(1 - [\alpha]_T)}.$$

The variable X takes r as its value and its confidence interval is

$$N[\alpha]_T \pm z_c\sqrt{N[\alpha]_T(1 - [\alpha]_T)}$$

Obviously, the confidence interval of the random variable that takes as its value $[\alpha] = r/N$ is

$$[\alpha]_T \pm z_c\sqrt{\frac{[\alpha]_T(1 - [\alpha]_T)}{N}}. \tag{11}$$

In Example 1 the transaction set $\mathbf{T}$ is too small so that the confidence intervals of considered formulas might be evaluated. E.g. for formula $LD \wedge OLD \wedge FAT \supset SBP$ we have

$$N[\alpha]_T(1 - [\alpha]_T) \cong 0.9 \ll 5.$$

4. Associations Rules Interpreted as Implications

Association rules are used in data mining to support expert decision-making. Today's models of expert decision-making used in the field of artificial intelligence suppose that the expert knowledge can be decomposed into small peaces or chunks, which can be expressed as IF-THEN rules [7]. IF-THEN rules are then interpreted as implications, i.e. from validity of the assertion in the IF part of the rule the validity of the assertion in the THEN part of the rule is deduced. Therefore in data mining the interpretation of an association rule as implication $\alpha \supset \gamma$ is the mostly appealing.

If we assume sufficiently large N and $cf \cong 1$, then from (8) and (9) we are getting

$$[\alpha \supset \gamma] \cong [\alpha \supset \gamma]_T = 1 + sup - \frac{sup}{cf} \cong 1,$$

$$[\alpha] \cong [\alpha]_T = \frac{sup}{cf} \cong sup.$$

Thus association rule $\alpha \Rightarrow \gamma$ that has $cf \cong 1$ and large support sup can be in natural way interpreted as implication $\alpha \supset \gamma$ that is valid with probability $[\alpha \supset \gamma] \cong 1$ and that can be often used for deducing consequent. Often here means with probability $[\alpha] \cong sup$.

To search for acceptable association rules using systematic passing through all possible states of the world Ω would be, for large number of literals contained in the association rule, too time-consuming. Nevertheless, we can use efficient Agrawal's algorithms [1] for searching of association rules with great support. These algorithms retrieve association rules with $sup \geq sup_{min}$ and afterwards they check their confidence cf and retain only those with $cf \geq cf_{min}$. According to (8) and

(11) found association rules can be interpreted as implications $\alpha \supset \gamma$ that are valid in the world Ω with probability

$$[\alpha \supset \gamma]_T \pm z_c \sqrt{\frac{[\alpha \supset \gamma]_T (1 - [\alpha \supset \gamma]_T)}{N}}, \text{ where } [\alpha \supset \gamma]_T = 1 + sup - \frac{sup}{cf} \quad (12)$$

and according to (9) and (11) their antecedents are valid in Ω with probability

$$[\alpha]_T \pm z_c \sqrt{\frac{[\alpha]_T (1 - [\alpha]_T)}{N}}, \text{ where } [\alpha]_T = \frac{sup}{cf}. \quad (13)$$

5. Information Content of a Logical Formula

Assume a world Ω and $P = \{P_1, P_2, \ldots, P_m\}$ a set of properties describing world states $\omega \in \Omega$. Thus world Ω can be in one of $n = 2^m$ states $\omega_1, \ldots, \omega_n$. Denote $p = \{p_i\}$ the true probability distribution of world states. Clearly, each state ω_i can be unequivocally described with the conjunction

$$\alpha_i = L_1 \wedge \cdots \wedge L_m, \quad (14)$$

where for all $j = 1, \ldots, m$ $L_j = P_j$ if the world state ω_i has the property P_j or $L_j = \neg P_j$ otherwise.

For an arbitrary formula β that consists of predicates from P we define its spectrum

$$I_\beta = \{i \in <1, 2^m>: \beta \text{ is valid in the world state } \omega_i\}. \quad (15)$$

Obviously,

$$\beta \equiv \bigvee_{i \in I_\beta} \alpha_i. \quad (16)$$

Theorem 2 Let Ω be a world with true probability distribution $p = \{p_i\}$ of its world states ω_i. Let the present knowledge about world Ω can be described by a probability distribution $\{q_i\}$, $q_i > 0$ if $p_i > 0$. Let the following formula

$$\beta \equiv \bigvee_{i \in I_\beta} \alpha_i, \text{ where } I_\beta \text{ is spectrum of } \beta,$$

be valid in Ω with probability $0 < [\beta] < 1$. Assume that the distribution $\{q_i\}$ is updated to $\{\tilde{q}_i\}$ according to the following equations

$$\tilde{q}_i = \frac{[\beta]}{\sum\limits_{i \in I_\beta} q_i} q_i \quad \text{if} \quad i \in I_\beta, \quad (17)$$

$$\tilde{q}_i = \frac{(1 - [\beta])}{\sum\limits_{i \notin I_\beta} q_i} q_i \quad \text{if} \quad i \notin I_\beta. \quad (18)$$

Then the following assertions are valid:

1. $\{\tilde{q}_i\}$ is a probability distribution. Moreover $\tilde{q}_i < 1$ for all i and $\tilde{q}_i > 0$ if $p_i > 0$.
2. $D(p\|q) - D(p\|\tilde{q}) = D(r\|s)$, where r and s are probability distributions

$$r = \{r_1, r_2\}, s = \{s_1, s_2\}, r_1 = [\beta], r_2 = (1 - [\beta]), s_1 = A, s_2 = (1 - A), A = \sum_{i \in I_\beta} q_i. \tag{19}$$

Here $D(p\|q)$ measures the difference between probability distributions $\{p_i\}$ and $\{q_i\}$ and is known as information (or Leibler-Kullback) divergence

$$D(p\|q) = \sum_i p_i \log \frac{p_i}{q_i}, \tag{20}$$

where $p \log p/q = 0$ if $p = 0$, $q \geq 0$ and $p \log p/q = \infty$ if $p > 0$, $q = 0$ (see for example Cover [8]).

Proof

1. If $[\beta] > 0$, then $\sum\limits_{i \in I_\beta} p_i > 0$. Hence, for some $j \in I_\beta$ must be $p_j > 0$ and consequently $q_j > 0$. Thus $\sum_{i \in I_\beta} q_i > 0$ holds. If $[\beta] < 1$, then $\sum_{i \notin I_\beta} p_i > 0$. Hence, for some $j \notin I_\beta$ must be $p_j > 0$ and $q_j > 0$. Thus also $\sum_{i \notin I_\beta} q_i > 0$ holds. From (17) and (18) we see that $\tilde{q}_i$ is defined for all i and that for all i $0 \leq \tilde{q}_i < 1$ must hold.

   Denote $A = \sum_{i \in I_\beta} q_i$ and $(1 - A) = \sum\limits_{i \notin I_\beta} q_i$. Then

   $$\sum_i \tilde{q}_i = \sum_{i \in I_\beta} \frac{[\beta]}{A} q_i + \sum_{i \notin I_\beta} \frac{(1 - [\beta])}{(1 - A)} q_i = [\beta] + (1 - [\beta]) = 1.$$

   Thus $\{\tilde{q}_i\}$ is a probability distribution. Moreover, theorem assumption that $q_j > 0$ if $p_j > 0$ together with (17), (18) yields that $\tilde{q}_i > 0$ if $p_i > 0$.

2. Let us put $A = \sum\limits_{i \in I_\beta} q_i$ Then

   $$\begin{aligned} D(p\|\tilde{q}) &= \sum_i p_i \log \frac{p_i}{q_i} = \sum_{i \in I_\beta} p_i \log \frac{A p_i}{[\beta] q_i} + \sum_{i \notin I_\beta} p_i \log \frac{(1 - A) p_i}{(1 - [\beta]) q_i} = \\ &= \sum_{i \in I_\beta} p_i \log \frac{A}{[\beta]} + \sum_{i \in I_\beta} p_i \log \frac{p_i}{q_i} + \sum_{i \notin I_\beta} p_i \log \frac{p_i}{q_i} + \sum_{i \notin I_\beta} p_i \log \frac{(1 - A)}{1 - [\beta]}. \end{aligned}$$

   Since $\sum\limits_{i \in I_\beta} p_i = [\beta]$ and $\sum\limits_{i \notin I_\beta} p_i = 1 - [\beta]$ hold, we are getting

   $$\begin{aligned} D(p\|\tilde{q}) &= D(p\|q) + [\beta] \log \frac{A}{[\beta]} + (1 - [\beta]) \log \frac{1 - A}{1 - [\beta]} = \\ &= D(p\|q) - ([\beta] \log \frac{[\beta]}{A} + (1 - [\beta]) \log \frac{1 - [\beta]}{1 - A}) = D(p\|q) - D(r\|s), \end{aligned}$$

   where r, s are probability distributions (19). ■

Theorem 3 Assume Ω be a world with true probability distribution of its states $p = \{p_i\}$. Let the present knowledge about world Ω be described by a probability distribution $q = \{q_i\}$, $q_i > 0$ if $p_i > 0$. Let a formula β be valid in Ω with probability $[\beta] = 1$ and let $A = \sum\limits_{i \in I_\beta} q_i$. Assume that the distribution $\{q_i\}$ is updated to $\{\tilde{q}_i\}$ according to the following equations

$$\tilde{q}_i = \frac{q_i}{A} \quad \text{if} \quad A < 1 \text{ and } i \in I_\beta, \tag{21}$$

$$\tilde{q}_i = 0 \quad \text{if} \quad A < 1 \text{ and } i \notin I_\beta. \tag{22}$$

Then the following assertions are valid:

1. $\tilde{q} = \{\tilde{q}_i\}$ is a probability distribution and $\tilde{q}_i > 0$ if $p_i > 0$.
2. $D(p\|\tilde{q}) = D(p\|q) + \log A$

If $A = 1$, then the knowledge of $[\beta]$ cannot be used for refinement of the current probability distribution q. In this case we put $\tilde{q} = q$.

Proof

1. Assume $A < 1$. Due to the assumption $[\beta] = 1$, there exists $j \in I_\beta$ such that $p_j > 0$. For this j $q_j > 0$ must hold according to the Theorem's assumptions. Hence, $A = \sum\limits_{i \in I_\beta} q_i > 0$ and $\tilde{q}_i$ is defined for all i. Moreover, $0 \le \tilde{q}_i \le 1$ for all i and $\sum\limits_{i=1}^{n} \tilde{q}_i = \sum\limits_{i \in I_\beta} \frac{q_i}{A} = 1$. Hence, $\tilde{q} = \{\tilde{q}_i\}$ is a probability distribution. If $p_i > 0$, then $q_i > 0$ and $i \in I_\beta$ holds due to $[\beta] = 1$. Then the value $\tilde{q}_i$ is evaluated according to (21) and therefore also $\tilde{q}_i > 0$.

2. $$D(p\|\tilde{q}) = \sum_{i \in I_\beta} p_i \log \frac{Ap_i}{q_i} + \sum_{i \notin I_\beta} p_i \log \frac{p_i}{\tilde{q}_i}.$$

   Since $[\beta] = 1$, $p_i = 0$ must hold for all $i \notin I_\beta$. Consequently, the second sum in the preceding equation must be zero and we are getting

   $$D(p\|\tilde{q}) = \sum_{i \in I_\beta} p_i \log \frac{Ap_i}{q_i} = \sum_{i \in I_\beta} p_i \log \frac{p_i}{q_i} + \sum_{i \in I_\beta} p_i \log A = D(p\|q) + \log A.$$

■

The amount of information inherent in a formula β may be defined as follows.

Definition 2 Assume that Ω is a world with the true probability distribution of its world states $p = \{p_i\}$. Let the present state of knowledge be $q = \{q_i\}$, $q_i > 0$ if $p_i > 0$ and let a formula β be valid in Ω with probability $[\beta] > 0$. Then the amount of information contained in β is defined as

$$I(\beta, q) = D(p\|q) - D(p\|\tilde{q}), \tag{23}$$

where the distribution update $\tilde{q}$ is defined according to the Theorem 2 if $[\beta] < 1$ or according Theorem 3 if $[\beta] = 1$.

Theorem 4 Assume a world Ω with the true probability distribution $p = \{p_i\}$ of its world states ω_i. Assume further that the current knowledge about Ω is given by means of probability distribution $q = \{q_i\}$, $q_i > 0$ if $p_i > 0$. Assume that β and γ are two formulas valid in Ω with probabilities $[\beta]$ and $[\gamma]$. Let q_β, q_γ be two probability updates of distribution q based on the known probabilities $[\beta]$ and $[\gamma]$ respectively. Then the following assertions hold.

1. $I(\beta, q) = ([\beta]\log\frac{[\beta]}{A} + (1-[\beta])\log\frac{1-[\beta]}{1-A}),\ \text{if } 0 < [\beta] < 1. \quad (24)$

2. $I(\beta, q) = -\log A,\ \text{if } [\beta] = 1, A < 1. \quad (25)$

3. $I(\beta, q) = 0,\ \text{if } [\beta] = 1, A = 1. \quad (26)$

4. $I(\beta, q)$ is non negative.

5. If $I(\beta, q) \geq I(\gamma, q)$, then $D(p\|q_\beta) \leq D(p\|q_\gamma)$, i.e. the q_β update is closer to the true probability distribution p than q_γ.

The symbol A that occurs in theorem items 1-3 is defined $A = \sum\limits_{i \in I_\beta} q_i$.

Proof Assertions 1–3 simply follow from the Theorem 2 and 3. Assertion 4 follows from Theorem 2 and 3 and from the fact that both expressions $-\log A$ and $D(r, s)$ are nonnegative. To prove Assertion 5 assume $I(\beta, q) \geq I(\gamma, q)$. According to Definition 2 we have

$$I(\beta, q) = D(p\|q) - D(p\|q_\beta),$$

$$I(\gamma, q) = D(p\|q) - D(p\|q_\gamma).$$

Therefore

$$D(p\|q_\gamma) - D(p\|q_\beta) = I(\beta, q) - I(\gamma, q) \geq 0.$$

■

Notice that $I(\beta, q)$ can be computed using only $[\beta]$ and q. The soundness of the just defined measure stems from the fact that according to the Theorem 4 the more informative formula enables to get more precise estimate of the true probability distribution p of the world states.

If β consists of only some predicates from $P = \{P_1, \ldots, P_m\}$, i.e. if β consists of $P_{j1}, \ldots, P_{jp} \in P, p < m$, we can instead of world Ω consider the restrict world Ω', the states ω_i' of which are described only with predicates $P_{j1}, \ldots, P_{jp}$. Then the spectrum of a formula β is

$$I_\beta' = \{i \in \langle 1, 2^p\rangle : \beta \text{ is valid in the world state } \omega_i'\}.$$

Obviously probability distribution $\{q_i'\}$ is a marginal distribution of the probability distribution $\{q_i\}$ and the following holds

$$A' = \sum_{i \in I_\beta'} q_i{}' = \sum_{i \in I_\beta} q_i = A.$$

Example 2 Assume that world states ω_i of a world Ω are characterized with properties $P = \{P_1, P_2, P_3, P_4\}$. Let $\beta = (P_1 \wedge (P_3 \vee \neg P_4))$ be a formula. Let a current knowledge about world be given by probability distribution $q = \{q_i\}$ defined in the Tab. III.

Then from the Tab. III we can see that

$$\begin{aligned} A &= \sum_{i \in I_\beta} q_i = q_9 + q_{11} + q_{12} + q_{13} + q_{15} + q_{16} = \\ &= 0.504 + 0.2016 + 0.0864 + 0.0336 + 0.1344 + 0.0576 = 0.564 \end{aligned}$$

α_i	P_1	P_2	P_3	P_4	q_i	$P_1 \wedge (P_3 \vee \neg P_4)$
α_1	0	0	0	0	.0336	0
α_2	0	0	0	1	.0144	0
α_3	0	0	1	0	.1344	0
α_4	0	0	1	1	.0576	0
α_5	0	1	0	0	.0224	0
α_6	0	1	0	1	.0096	0
α_7	0	1	1	0	.0896	0
α_8	0	1	1	1	.0384	0
α_9	1	0	0	0	.0504	1
α_{10}	1	0	0	1	.0216	0
α_{11}	1	0	1	0	.2016	1
α_{12}	1	0	1	1	.0864	1
α_{13}	1	1	0	0	.0336	1
α_{14}	1	1	0	1	.0144	0
α_{15}	1	1	1	0	.1344	1
α_{16}	1	1	1	1	.0576	1

Tab. III *States of the world Ω.*

Let us restrict the initial world Ω to world Ω' with predicates P_1, P_3, P_4 only. Then the probability distribution $\{q'_i\}$ of states of the restricted world Ω' is given in the Tab. IV. Clearly,

$$\begin{aligned} q'_5 &= q_9 + q_{13} = 0.084 \\ q'_7 &= q_{11} + q_{15} = 0.336 \\ q'_8 &= q_{12} + q_{16} = 0.144 \end{aligned}$$

$$A' = \sum_{i \in I'_\beta} q_i{}' = q'_5 + q'_7 + q'_8 = 0.084 + 0.336 + 0.144 = 0.564$$

6. Information Content of an Association Rule

Assume as in paragraph 2 that transactions $T \in \mathbf{T}$ consist of items describing states ω of some world Ω. More precisely, assume that each transaction $T = \{L_1, \ldots, L_m\}$

α'_i	P_1	P_3	P_4	q'_i	$P_1 \wedge (P_3 \vee \neg P_4)$
α'_1	0	0	0	.056	0
α'_2	0	0	1	.024	0
α'_3	0	1	0	.224	0
α'_4	0	1	1	.096	0
α'_5	1	0	0	.084	1
α'_6	1	0	1	.036	0
α'_7	1	1	0	.336	1
α'_8	1	1	1	.144	1

Tab. IV *States of the restricted world Ω'.*

of a complete transaction set $\mathbf{T}$ determines an world state $\omega \in \Omega$ that can be unequivocally described with formula $\alpha = L_1 \wedge \cdots \wedge L_m$. Assume further that $\alpha \Rightarrow \gamma$ is a association rule defined on transaction set $\mathbf{T}$. Then we define the information content of an association rule $\alpha \Rightarrow \gamma$ as follows.

Definition 3 The information content of an association rule $\alpha \Rightarrow \gamma$ is defined as

$$I(\alpha \Rightarrow \gamma) = I(\alpha \supset \gamma, q^*), \tag{27}$$

where $q^* = \{q_i^*\}$, $q_i^* > 0$ is the probability distribution of world states that is derived from the true probability distribution $p = \{p_i\}$ of world states under assumption of statistical independence of world properties (i.e. properties that describe world states). It means that if the state ω_i has description $\alpha_i = L_1 \wedge \cdots \wedge L_r$, then

$$q_i^* = [L_1] \dots [L_r], \tag{28}$$

where $[L_j], j = 1, \dots, r$ is the marginal probability of L_j, induced by the true probability distribution p.

The soundness of just defined information measure stems from the fact that more informative association rule enables, provided it is interpreted in a natural way as logical implication, to get more precise correction of the initial probability distribution of world states, which has been obtained under likely false assumption of statistical independence of world properties.

Marginal probabilities L_j can be easily estimated from the transaction set as well as support *sup* and confidence cf of the association rule. Probability distribution $q^* = \{q_i^*\}$ can be determined using (28). In paragraph 2 we have showed that from support *sup* and confidence cf the estimate $[\alpha \supset \gamma]_T$ of the probability that formula $\alpha \supset \gamma$ is valid in randomly chosen world state can be evaluated using (8). Eventually, information content of the association rule can be computed according to the equations (24), (25), (26) respectively.

Example 3 Assume association rule $LD \wedge OLD \wedge FAT \Rightarrow SBP$ from Example 1. From the Tab. II we see that the rule has support $sup = 0.5$, confidence $cf = 5/6$ and that estimates of marginal probabilities of predicates are $[LD]_T = 0.8$, $[OLD]_T = 0.7$, $[FAT]_T = 0.7$ and $[SBP]_T = 0.9$. The probability that formula $LD \wedge OLD \wedge$

$FAT \supset SBP$ is valid can be estimated from its confidence cf and support sup

$$[\alpha \supset \gamma]_T = 1 + sup - \frac{sup}{cf} = 0.9$$

Our task is to evaluate the information content of the association rule according to the Definition 3. Denote $A = \sum_{i \in I_{\alpha \supset \gamma}} q_i^*$. Since $\neg(\alpha \supset \gamma) \equiv \alpha \wedge \neg\gamma$, we are getting

$$(1 - A) = \sum_{i \notin I_{\alpha \supset \gamma}} q_i^* = \sum_{i \in I_{\neg(\alpha \supset \gamma)}} q_i^* = \sum_{i \in I_{\alpha \wedge \neg\gamma}} q_i^*.$$

Formula $\alpha \wedge \neg\gamma \equiv LD \wedge OLD \wedge FAT \wedge \neg SBD$ is valid only in the one state, which we denote ω_k and which has description $\alpha_k = LD \wedge OLD \wedge FAT \wedge \neg SBD$. The probability q_k^* of this state is estimated from the transaction set

$$q_k^* = [LD]_T [OLD]_T [FAT]_T (1 - [SBP]_T).$$

Therefore

$$1 - A = q_k^* = 0.8 \times 0.7 \times 0.7 \times 0.1 = 0.0392.$$

The information content of the association rule $LD \wedge OLD \wedge FAT \Rightarrow SBP$ is then evaluated according to (24)

$$I(LD \wedge OLD \wedge FAT \Rightarrow SBD) = I(LD \wedge OLD \wedge FAT \supset SBP, q^*) = \\ = 0.9 \log \frac{0.9}{0.9608} + 0.1 \log \frac{0.1}{0.0392} = 0.0254.$$

In a similar way we may calculate the information content of the association rule $LD \wedge \neg HD \wedge SBP \Rightarrow IM$. The support and the confidence of this rule are $sup = 0.4$ and $cf = 2/3$ and marginal probabilities estimates are $[LD]_T = 0.8$, $[HD]_T = 0.2$, $[SBP]_T = 0.9$ and $[IM]_T = 0.5$ (see the Tab. II). The estimate of probability that formula $LD \wedge \neg HD \wedge SBP \wedge \neg IM$ is valid can be evaluated from sup and cf according to (8) and it equals to 0.8.

Consequently, the value $1 - A$ and the information content of the association rule are

$$1 - A = 0.8 \times 0.8 \times 0.9 \times 0.5 = 0.288,$$

$$I(LD \wedge \neg HD \wedge SBP \supset IM, q^*) = 0.8 \log \frac{0.8}{0.712} + 0.2 \log \frac{0.2}{0.288} = 0.0203.$$

Thus the information content of association rule $LD \wedge OLD \wedge FAT \Rightarrow SBP$ is about 25% greater than the information content of the association rule

$$LD \wedge \neg HD \wedge SBP \Rightarrow IM.$$

7. Conclusion

In the paper a quantitative measure of "informativeness" of association rules has been introduced. The measure can be used if the underlying database model fulfills the following conditions.

1. The records in the database describe independently occurring states of a world or a system.
2. The components of the records are values of binary properties of the world states.
3. The records have fixed length and do not contain any missing values.

For example, database can be a database of inhabitants of a town district, a database of clients of a bank, a database of medical patient records of some hospital department etc. Components of records are supposed to be binary properties. Of course, if some record component takes more values than two or if it takes real values, then it can be converted into several binary components using binary coding. If the database contains records with missing values, then those with missing values must be discarded.

For evaluation of the introduced information measure $I(\alpha \Rightarrow \gamma)$ of an association rule $\alpha \Rightarrow \gamma$ that contains r properties we need to know the following.

1. Probabilities q_i^* of all world states, in which the implication $\alpha \supset \gamma$ is valid. There are at maximum 2^r such probabilities. They are to be evaluated under assumption of independence of world properties from r marginal property probabilities. The estimates of property probabilities can be easily obtained from database. Having q_i^*, we can simply compute constant A from Theorem 4.
2. Probability $[\alpha \supset \gamma]$ that a formula $\alpha \supset \gamma$ is valid in the underlying world. This probability can be easily estimated using confidence and support of the association rule $\alpha \Rightarrow \gamma$.

From constant A and from probability $[\alpha \supset \gamma]$ the information content of association rule $\alpha \Rightarrow \gamma$ can be evaluated according to the Theorem 4. To be able to interpret association rules in some plausible way, we are usually interested only in association rules with less then several tens of properties. In this case the effective evaluation of the constant A and then that of the information measure is quite feasible.

The information content of an association rule has been defined under assumption that no knowledge about dependence among properties describing world states is at our disposal. However, it is not always the case. Sometimes some a priori knowledge about the true distribution of world states might be available. For example, consider a database of patient records. Then if in some record the property "male" has value true, then in the same record the property "suffered from complicated child delivery" must have value false. A priori knowledge of this kind often reflects logic dependence that may exist among properties. This dependence may be taken into consideration and it may be included into the definition of information content of an association rule. To do it, we represent known logical dependence with logical formula and then refine the initial probability distribution q^*, obtained under assumption of independency, to distribution $\tilde{q}$ by means of equations (21) and (22). According to Theorem 4, $\tilde{q}$ is a probability distribution and moreover, it is a real refinement, since the Kullback-Leibler divergence between $\tilde{q}$ and the true probability distribution of states p is less then that between distributions q^* and p.

For example, assume that logical dependence among properties may be expressed with implication $\alpha \supset \gamma$. Assume that we have taken this piece of knowledge into consideration and refined initial distribution q^* into $\tilde{q}$. Since implication $\alpha \supset \gamma$ is valid, confidence of the association rule $\alpha \Rightarrow \gamma$ must equal to 1. If its support were sufficiently large, it might be mined by means of Agrawal's algorithm. However, its information content would be according to (26) $I(\alpha \Rightarrow \gamma) = I(\alpha \supset \gamma, \tilde{q}) = 0$. We see that if we interpret an association rule in natural way as logical implication and if the validity of this implication is a priori known and considered, then information content of the association rule equals to zero. Putting it in another way, we are getting intuitively reasonable result that in this case the association rule does not yield any new information.

Introduced measure of "informativeness" of association rules enables us to focus our attention on those association rules that mostly reflect regularities inherent in the data. The interpretation of these association rules gives the largest hope to obtain new interesting knowledge. However, the practical usefulness of the introduced information measure may be proved only by means of its successful applications.

References

[1] Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules, VLDB, Sep. 12-15, Chile, ISBN 1-55860-153-8, 1994, pp. 487–99.

[2] Vesely, A.: On logical interpretation of association rules. In Proceedings of the 13th IASTED International Conference on Artificial Intelligence and Soft Computing, Palma de Mallorca, 2009.

[3] Tan, P., Kumar V., Srivastava, J.: Selecting the right objective measure for association analysis, Information Systems, 29, 2004, pp. 293–313.

[4] Blanchard, J., Guillet, F., Gras, R., Briand, H.: Using Information-theoretic Measures to Assess Association Rule Interestingness, The Fifth IEEE Conference on Data Mining, 2005.

[5] Vajda, I., Vesely, A., Zvarova, J.: On the amount of information resulting from empirical and theoretical knowledge, Revista Mathematica Complutense, 18, 2, 2005, pp. 275–283.

[6] Vesely, A.: On the information content of knowledge. In Proceedings of the 11th IASTED International Conference on Artificial Intelligence and Soft Computing, Palma de Mallorca, 2007.

[7] Anderson, J. R., Lebiere, C.: Atomic components of thought, Hillsdale, 1998.

[8] Cover, T. M., Thomas J. A.: Elements of Information Theory, John Wiley & Sons, New York, 2006.