Logical Representation of Dependencies of Items and the Complexity of Customer Sets

János Demetrovics
MTA SZTAKI
Lágymányosi u. 11
1111 Budapest, Hungary
demetrovics@sztaki.hu

Hua Nam Son
Budapest Business School
Buzogány u. 11–13
1149 Budapest, Hungary
huanamson@yahoo.com

Ákos Gubán
Budapest Business School
Buzogány u. 11–13
1149 Budapest, Hungary
guban.akos@pszfb.bgf.hu

ABSTRACT

The problem of discovering of frequent market baskets and association rules has been considered widely in literatures of data mining. In this study, by using the algebraic representation of market basket model, we propose a concept of logical constraints of items in an effort to detect the logical relationships hidden among them. Via the relationships of the propositional logics and logical constraints of items we propose also the concept of the complexity of customers. As a result we show that every set of customers can be characterized by a logical constraint and can be divided into different blocks that are characterized by quite simple logical constraints. In the natural way the complexity of a customer set is defined as the number of the blocks it contains.

Keywords

Market basket model, frequent product, propositional logics, lattice

1. INTRODUCTION

Great efforts have been made to discover the information hidden in the sets of market baskets and in the sets of customer transactions. The studies of customer market baskets (MB) and mining the association rules are important in various applications, for example, in decision making and strategy determination of retail economy ([1]). Therefore discovering of *large itemsets* and association rules attracts the interest of researchers (see [8,10]). One can notice that in their studies the researchers deal with the set of items (e.g. bread, milk,....) purchased by customers, but did not consider the quantity of each item. It would be interesting also if we know that 70% of customers buy bread and milk, but only 50% of customers buy 1 kg bread and 2 litter milk, while 1% of customers buy 10 kg bread and 1 litter milk. Similar example can be found for association rules. The reasons for quantitative analysis of transactions are evident.

In the previous studies (see [5, 6]) we have introduced a quantitative analysis of transactions. We are interested not only in the statement "90% of customers who buy bread and milk also buy butter", but in the fact "90% of customers who buy 1 kg bread and 2 litter milk also buy 0,5 kg butter". By dealing with the quantity of items our approach is somehow different of those in other studies (see [1]) and is suitable for a quantitative analysis of transactions.

Similarly to [5, 6], instead of *itemsets* we use *market baskets* or *transactions*. In the following we establish the relationships between the structure of the set of market baskets and propositional logics. In Section 2 we recall the algebraic approach in analysing the structure of the set of market baskets which was defined in [6]. In Section 3 the concept of the constraints of market baskets is defined and we show that every set of market baskets can be characterized by some logical constraint. In Section 4 we introduce the concept of complexity of the sets of market baskets, which we call by the complexity of the customer sets. All these are done via the relationship between the structure of the set of market baskets and propositional logics which are defined in this section.

BCI'12, September 16–20, 2012, Novi Sad, Serbia.

Local Proceedings also appeared in ISBN 978-86-7031-200-5, Faculty of Sciences, University of Novi Sad.

2. A GENERALIZATION OF THE MARKET BASKET MODEL

In this section we recall the concepts and results that are established in [5, 6]. For a finite set of items $P = \{p_1, p_2, ..., p_n\}$ we consider a *market basket* (MB) as a tube $\alpha = (\alpha[1], \alpha[2], ..., \alpha[n])$, where $\alpha[i] \in \mathbb{N}$ is the quantity of p_i in the basket α. The set of all MBs is denoted by Ω. For $\alpha, \beta \in \Omega$ where $\alpha = (\alpha[1], \alpha[2], ..., \alpha[n])$, $\beta = (\beta[1], \beta[2], ..., \beta[n])$ we write $\alpha \leq \beta$ if for all $i = 1, 2, ..., n$ we have $\alpha[i] \leq \beta[i]$. $\langle \Omega, \leq \rangle$ is a lattice with the natural partial order $\leq$ (see [4]). For a set $A \subseteq \Omega$ we denote

$$U(A) = \{\alpha \in \Omega | \forall \beta \in A : \beta \leq \alpha\} \text{ and}$$

$$L(A) = \{\alpha \in \Omega | \forall \beta \in A : \alpha \leq \beta\}.$$

We denote also by $sup(A)$ and $inf(A)$ the smallest, and the biggest MB in $U(A)$ and $L(A)$, respectively.

For a set $A \subseteq \Omega$ and $\alpha \in \Omega$ we denote by

$$supp_A(\alpha) = \frac{|\{\beta \in A | \alpha \leq \beta\}|}{|A|}$$

the *support* of α in A. In word, $supp_A(\alpha)$ denotes the rate of all market baskets that exceeds the given threshold (in the form of a sample market basket) α to A. The support of an market basket is a statistical index by which one can estimate the "vogue" of α in the given group of customers A.

Naturally, the market baskets of more support attract better the attention of the shop managers. Some semantic interpretations are needed to avoid the possible misunderstandings: If p_i (the name of i-th item) is, for example, bread, then α_i is the quantity of bread in the basket (of the customer) α. For $\alpha, \beta \in \Omega$ where $\alpha = (\alpha[1], \alpha[2], ..., \alpha[n])$ and $\beta = (\beta[1], \beta[2], ..., \beta[n])$ we write $\gamma = \alpha \cup \beta$ if $\gamma[i] = max\{\alpha[i], \beta[i]\}$ for all $i = 1, 2, ..., n$. We call $\alpha \longrightarrow \beta$ an *association rule* of β to α. By the *confidence* of $\alpha \longrightarrow \beta$ in a set of MBs A we mean the rate

$$conf_A(\alpha \longrightarrow \beta) = \frac{supp_A(\alpha \cup \beta)}{supp_A(\alpha)}$$

For a set $A \subseteq \Omega$, $\alpha \in \Omega$ and $0 \leq \varepsilon \leq 1$, α is ε-frequent MB, if $supp_A(\alpha) \geq \varepsilon$. The set of all ε-frequent MBs is denoted by Φ_A^ε. In our setting the Apriori Principle is stated as follows:

Apriori Principle: For a set $A \subseteq \Omega$, $\alpha, \beta \in \Omega$ and $0 \leq \varepsilon \leq 1$, if $\alpha \leq \beta$ and β is ε-frequent then α is ε-frequent.
The following example was considered in [5]:

Example 1: Consider a set of items $P = \{a, b, c\}$ and a set of transactions $A = \{\alpha, \beta, \gamma, \delta\}$, where $\alpha = (2, 1, 0)$, $\beta = (1, 1, 1)$, $\gamma = (1, 0, 1)$, $\delta = (2, 2, 0)$. One can see that for $\sigma = (1, 1, 0)$, $\eta = (1, 2, 0)$ we have $supp_A(\sigma) = \dfrac{3}{4}$ and $supp_A(\eta) = \dfrac{1}{4}$. For the threshold $\varepsilon = \dfrac{1}{2}$ the ε-frequent MBs of A are:

$$\Phi_A^{\frac{1}{2}} = \{(2,1,0), (1,0,1), (1,1,0), (2,0,0), (0,0,1), (0,1,0), (1,0,0), (0,0,0)\}.$$

Let us denote
$\Phi_{A,k} = \{\alpha \in \Omega | \exists \alpha_1, \alpha_2, ..., \alpha_k \in A : \alpha \leq \{\alpha_1, \alpha_2, ..., \alpha_k\}\}$

One can remark that if $k \leq l$ then $\Phi_{A,k} \supseteq \Phi_{A,l}$ and $\Phi_A^\varepsilon = \Phi_{A,k}$ where $k = \lceil \varepsilon |A| \rceil$ denotes the smallest integer that is greater or equal to $\varepsilon |A|$. We have Theorem 1 and Theorem 2 (see [5, 6] for the proof):

Theorem 1: For a set of items $P = \{p_1, p_2, ..., p_n\}$, a set of MBs $A \subseteq \Omega$ and a threshold $0 \leq \varepsilon \leq 1$ an MB $\alpha \in \Omega$ is ε-frequent iff there exist $\alpha_1, \alpha_2, ..., \alpha_k \in A$ such that $\alpha \in L(\{\alpha_1, \alpha_2, ..., \alpha_k\})$ where $k = \lceil \varepsilon |A| \rceil$.

By Theorem 1 in [5] we have proposed an algorithm that creates all ε-frequent MBs of a given set of transactions A in $O\left(\binom{|A|}{k} . (m+1)^n\right)$ running time.

Algorithm 1: (Creating all ε-frequent MBs of a given set of transactions A)

Input: Set of items P, Set of MBs $A \subseteq \Omega$ and a threshold $0 \leq \varepsilon \leq 1$.
Output: Φ_A^ε.

Theorem 2: (Explicit representation of *large MBs*) For a set of items $P = \{p_1, p_2, ..., p_n\}$, a set of MBs $A \subseteq \Omega$ and a threshold $0 \leq \varepsilon \leq 1$ there exist $\alpha_1, \alpha_2, ..., \alpha_s \in \Omega$ where $s = \binom{|A|}{\lceil \varepsilon |A| \rceil}$ such that

$$\Phi_A^\varepsilon = \bigcup_{i=1}^{s} L(\alpha_i).$$

We should remark that $\alpha_i \leq \alpha_j$ iff $L(\alpha_i) \subseteq L(\alpha_j)$. For a set of MBs A and a given threshold ε the *basic ε- frequent set of MBs* of A is the set of MBs $\alpha_1, \alpha_2, ..., \alpha_s$ for which

i. $\Phi_A^\varepsilon = \bigcup_{i=1}^{s} L(\alpha_i)$.

ii. $\forall i, j : 0 \leq i, j \leq s$ we have $\alpha_i \nleq \alpha_j$ and $\alpha_j \nleq \alpha_i$

For a given A, ε the basic ε- frequent set of MBs of A is unique, which we denote by S_A^ε. We have:

Theorem 3: For a set of items P, a threshold $0 \leq \varepsilon \leq 1$ every set of MBs $A \subseteq \Omega$ has an unique basic ε- frequent set of MBs S_A^ε.

An algorithm that creates the basic ε- frequent set of MBs $O\left(\binom{|A|}{k} .m.n\right)$ in running time for a given set of MBs $A \subseteq \Omega$ and a given threshold ε is proposed in [5]:

Algorithm 2: (Creating the basic ε- frequent set of MBs S_A^ε)

Input: Set of items P, Set of MBs $A \subseteq \Omega$ and a threshold $0 \leq \varepsilon \leq 1$.
Output:S_A^ε

One can remark that in the case of large amount of transactions A the basic ε - frequent set of MBs S_A^ε can be generated much more quickly than the set of all ε-frequent set of MBs Φ_A^ε.

Example 2: We continue the Example 1. For the set of transactions A Algorithm 2 generates the basic $\frac{1}{2}$ - frequent set of MBs $S_A^{\frac{1}{2}} = \{\rho, \theta\}$ where $\rho = (2, 1, 0)$, $\theta = (1, 0, 1)$. It means that the family of $\frac{1}{2}$- frequent set of MBs of A is $\Phi_A^{\frac{1}{2}} = L(\rho) \cup L(\theta)$.

As shown in [5, 6] we can find all associations with given confidence. For a set of items P, a set of MBs $A \subseteq \Omega$ and a threshold $0 \leq \varepsilon \leq 1$ an association $\alpha \longrightarrow \beta$ is *ε-confident* if $conf_A(\alpha \longrightarrow \beta) \geq \varepsilon$. The set of all ε-confident associations of A is denoted by C_A^ε. We have:

Theorem 4: For a set of products P, a set of MBs $A \subseteq \Omega$ and $0 \leq \varepsilon \leq 1$ an association $\alpha \longrightarrow \beta$ is ε-confident iff $\dfrac{|U(\alpha \cup \beta) \cap A|}{|U(\alpha) \cap A|} \geq \varepsilon$.

A natural question for cross marketing, store layout, ...(see, for example, [1]) is to find all association rules with a given confidence. In our generalized model the following theorem shows in a sense an explicit representation of all association rules. More exactly, we show for a given MB α which set of MBs β may be associated to α with a given threshold of confidence.

For MBs ρ, σ where $\rho \leq \sigma$, let us denote

$$M(\rho,\sigma) = \{\eta \in \Omega | \rho \cup \eta \leq \sigma\}$$

It should be remarked that $M(\rho,\sigma)$ can be represented explicitly. If $\rho = (\rho_1, \rho_2, ..., \rho_s)$, $\sigma = (\sigma_1, \sigma_2, ..., \sigma_s)$ then $\eta = (\eta_1, \eta_2, ..., \eta_s) \in M(\rho,\sigma)$ if and only if $max\,(\rho_i, \eta_i) = \sigma_i$ for all $i = 1, 2, ..., s$, i.e. $\eta_i = \sigma_i$ in the case $\rho_i \lneq \sigma_i$ and $\eta_i \leq \sigma_i$ in the case $\rho_i = \sigma_i$.

Theorem 5: (Explicit representation of association rules) For a set of items $P = \{p_1, p_2, ..., p_n\}$, a set of MBs $A \subseteq \Omega$, an MB $\alpha \in \Omega$ and a threshold $0 \leq \varepsilon \leq 1$ there exist $\alpha_1, \alpha_2, ..., \alpha_k \in \Omega$ such that $\forall \beta \in \Omega : \alpha \longrightarrow \beta$ is ε-confident association rule if and only if $\beta \in \bigcup_{i=1}^{k} M(\alpha, \alpha_i)$.

It was showed in [5] that Theorem 5 in a sense gives an explicit presentation for association rules and by the following algorithm one can find all ε-confident association rules for given left side.

Algorithm 3: (Creating all ε- confident association rules $\alpha \longrightarrow \beta$ for given α)

Input: A set of items P, a set of MBs $A \subseteq \Omega$, a threshold $0 \leq \varepsilon \leq 1$ and an MB α.
Output: $\bigcup_{i=1}^{k} M(\alpha, \alpha_i)$.

Example 3: We continue the Example 1. For the set of MBs A (see Example 1), the MB $\sigma = (1, 1, 0)$ and threshold $\varepsilon = \frac{1}{2}$ we should find all MB η such that $\sigma \longrightarrow \eta$ is ε- confident association rule. We can see $U(\sigma) \cap A = \{(2, 1, 0), (1, 1, 1), (2, 2, 0)\}$ and $s := \lceil \varepsilon |U(\alpha) \cap A| \rceil = 2$. By step 2 in Algorithm 3 we have $k = 4$ and $\alpha_1 = (1, 1, 0)$, $\alpha_2 = (2, 1, 0)$. The set of all MBs η such that $\sigma \longrightarrow \eta$ is $\frac{1}{2}$-confident association rule is

$$M(\sigma, \alpha_1) \cup M(\sigma, \alpha_2) = \{(1, 1, 0), (1, 0, 0), (0, 1, 0), (0, 0, 0), (2, 1, 0), (2, 0, 0)\}$$

As a result we can see that besides the trivial association rules of the form $\sigma \longrightarrow \sigma'$, where $\sigma' \leq \sigma$ we got non-trivial association rules $\sigma \longrightarrow (2, 1, 0)$ and $\sigma \longrightarrow (2, 0, 0)$. In words, in the set of customers A more than 50% of customers that buy a and b also buy 2 a and 1 b items, as well more than 50% of customers who buy a and b also buy 2 a items.

3. LOGICAL CONSTRAINTS OF MARKET BASKETS

In this section we propose a concept of constraints of MBs which we call the *logical constraints of MBs*. The reason for our attempt is clear: The constraint $(\neg\alpha)$ where α means the meat certainly holds with high support for the vegetarian customer groups, while the constraint $(\alpha \wedge \beta) \longrightarrow \gamma$ seemingly gains high support from the householder customers, if α, β and γ means milk, egg and flour respectively. In the same way one can easily see that $\gamma \longrightarrow \alpha \vee \beta$ holds commonly for any customer groups.
Let us construct the logical constraints of MBs. For a set of items $P = \{p_1, p_2, ..., p_n\}$ let Ω be the set of all MBs over P. We define the *logical constraints of MBs* (for short, constraint) as follows:

1. All $\alpha \in \Omega$ are constraints. In this case $\pi(\alpha) = U(\alpha) = \{\beta \in \Omega | \alpha \leq \beta\} \subseteq \Omega$.
2. If α is a constraint then $(\neg\alpha)$ is a constraint and $\pi(\neg\alpha) = (\pi(\alpha))^c$ where by A^c we denote $\Omega \setminus A$ for $A \subseteq \Omega$
3. If α, β are constraints then
   $(\alpha \vee \beta)$ is a constraint and $\pi(\alpha \vee \beta) = \pi(\alpha) \cup \pi(\beta)$.
   $(\alpha \wedge \beta)$ is a constraint and $\pi(\alpha \wedge \beta) = \pi(\alpha) \cap \pi(\beta)$.
4. All constraints are constructed as in 1., 2. and 3.

As usual, the parentheses are omitted where it causes no confusion. We call $\pi(\alpha)$ the *set of supporting market baskets* of α. Two constraints α, β are *equivalent*, noted by $\alpha \equiv \beta$, if $\pi(\alpha) = \pi(\beta)$. A constraint is *tautology* if $\pi(\alpha) = \Omega$. The set of all constraints is denoted by $C(\Omega)$.
The following properties of propositions in propositional calculus hold also for the constraints:

1. If α, β, $\gamma \in GI(\Omega)$ are constraints then
   $\alpha \vee \beta \equiv \beta \vee \alpha$,
   $\alpha \wedge \beta \equiv \beta \wedge \alpha$
   $\alpha \vee (\beta \vee \gamma) \equiv (\alpha \vee \beta) \vee \gamma$,
   $\alpha \wedge (\beta \wedge \gamma) \equiv (\alpha \wedge \beta) \wedge \gamma$
2. If $\alpha \in GI(\Omega)$ is a constraint then $\neg(\neg\alpha) \equiv \alpha$.
3. If α, β are constraints then
   $\neg(\alpha \wedge \beta) \equiv \neg\alpha \vee \neg\beta$ and
   $\neg(\alpha \vee \beta) \equiv \neg\alpha \wedge \neg\beta$
4. For α, $\beta \in GI(\Omega)$ the notation $\alpha \rightarrow \beta$ is used also for $\neg\alpha \vee \beta$.

The above identities are always true. We call these identities the *logical identities*. It is easy to see that for a given A in the same way we can define $\pi_A(\alpha) = \pi(\alpha) \cap A$, which we call the *relative set of supporting MBs* of α. Similarly we say that two constraints α, β are *relatively equivalent* (in A), noted by $\alpha \equiv_A \beta$, if $\pi_A(\alpha) = \pi_A(\beta)$. It is easy to verify the following theorem:

Theorem 6:

1. For any finite set of MBs $A \subseteq \Omega$ there is a constraint $\alpha_A^* \in GI(\Omega)$ such that $\pi(\alpha_A^*) = A$.
2. For all $\beta, \gamma \in GI(\Omega)$, $\beta \equiv_A \gamma$ if and only if $\beta \wedge \alpha_A^* \equiv \gamma \wedge \alpha_A^*$.

Proof:
1) For any finite set of MBs $A \subseteq \Omega$ we find the constraint $\alpha_A^* \in GI(\Omega)$ such that $\pi(\alpha_A^*) = A$. If $P = \{p_1, p_2, ..., p_n\}$, $\rho = (\rho[1], \rho[2], ..., \rho[n]) \in \Omega$ then let

$$\rho_i^+ = (\rho[1], \rho[2], ..., \rho[i] + 1, ..., \rho[n]).$$

One can see that

$$\{\rho\} = \pi(\rho) \setminus \bigcup_{i=1}^{n} \pi(\rho_i^+) = \pi(\rho \wedge \bigwedge_{i=1}^{n} \neg(\rho_i^+)).$$

Let

$$\alpha_A^* = \bigvee_{\rho \in A} [\rho \wedge \bigwedge_{i=1}^{n} \neg(\rho_i^+)]$$

We have $A = \pi(\alpha_A^*)$.
2) The assertion is proved easily by using the definitions. We have $\beta \equiv_A \gamma \Longleftrightarrow \pi_A(\beta) = \pi_A(\gamma) \Longleftrightarrow \pi(\beta) \cap A = \pi(\gamma) \cap A \Longleftrightarrow \beta \wedge \alpha_A^* \equiv \gamma \wedge \alpha_A^*$.
The proof is completed.

One can remark that there are two trivial cases: The first is the case, when α_A^* is tautology. In this case $\equiv_A$ coincides with $\equiv$. This coincidence does not hold in general. We call a set of customers (transactions) *complete* if α_A is tautology. The second one is the case when α_A^* is tautologically false. For $\beta \in GI(\Omega)$ we denote $\beta_A = \beta \wedge \alpha_A^*$.

Example 4: We continue the Example 1. Let $P = \{a, b, c\}$ and a set of transactions $A = \{\alpha, \beta, \gamma, \delta\}$, where $\alpha = (2, 1, 0)$, $\beta = (1, 1, 1)$, $\gamma = (1, 0, 1)$, $\delta = (2, 2, 0)$. If $a =$ "Flour", $b =$ "Egg", $c =$ "Milk", which can be identified by $a = (1, 0, 0)$, $b = (0, 1, 0)$, and $c = (0, 0, 1)$, respectively, then

$\pi(a) = U((1, 0, 0)) = \{(x, y, z)|x \geq 1\}$,
$\pi_A(a) = \{\alpha, \beta, \gamma, \delta\}$
$\pi(b) = U((0, 1, 0)) = \{(x, y, z)|y \geq 1\}$,
$\pi_A(b) = \{\alpha, \beta, \delta\}$
$\pi(c) = U((0, 0, 1)) = \{(x, y, z)|z \geq 1\}$
$\pi_A(c) = \{\beta, \gamma\}$

In this case the constraint $a \wedge b \rightarrow c$ that may be interpreted as $Flour \wedge Egg \rightarrow Milk$, characterises those customers, who if buy Flour and Egg then must buy Milk. It is easy to see that the set of supporting MBs of this constraint is $\pi(a \wedge b \rightarrow c) = \{(x, y, z)|x = 0 \text{ or } y = 0 \text{ or } z \geq 1\}$. One also can see that in this case $\pi_A(a \wedge b \rightarrow c) = \pi(a \wedge b \rightarrow c) \cap A = \{\beta, \gamma\}$, i.e. $(a \wedge b \rightarrow c) \equiv_A c$.

It is easily to see that the properties of propositions in propositional calculus (see [9]) hold also for the constraints in the given set of customers, but the converse is not always true. Although one can verify the following for α, $\beta \in GI(\Omega)$ and an arbitrary set of customers A:

1. $(\alpha \vee \beta)_A \equiv_A \beta_A \vee \alpha_A$.
2. $(\alpha \wedge \beta)_A \equiv_A \beta_A \wedge \alpha_A$.
3. $(\neg\alpha)_A \equiv_A \neg(\alpha_A)$.

One should distinguish $\equiv_A$ and $\equiv$.

4. THE COMPLEXITY OF THE CUSTOMER SETS

In this section we propose a criteria for the complexity of the customer sets. The practical aspect of this attempt is clear: every shop manager want have the answer to the question how complex is their customer set. One can remark that the set of customers that contains only one customer is simple. An other simple customer set is the case when the transactions of the customer in the set (that may be a large mass) are "similar" in some way. In our version the concept of complexity may be understood as following.

Let $P = \{p_1, p_2, ..., p_n\}$ be a a finite set of items and Ω be the set of MBs over P. We recall that $U(\alpha) = \{\beta \in \Omega|\alpha \leq \beta\}$ for $\alpha \in \Omega$. We call a set $B \subseteq \Omega$ a *block of customers* if there are $\alpha_1, \alpha_2, \ldots, \alpha_m \in \Omega$, $\beta_1, \beta_2, \ldots, \beta_n \in \Omega$ such that

$$B = \bigcap_{k=1}^{m} U(\alpha_k) \setminus \bigcup_{k=1}^{n} U(\beta_k)$$

The block is denoted by $[\alpha_1, \alpha_2, \ldots, \alpha_m|\beta_1, \beta_2, \ldots, \beta_n]$. We have the following simple theorem:

Theorem 7: Let $P = \{p_1, p_2, ..., p_n\}$ be a a finite set of items and Ω be the set of all MBs over P.

1. Every $\gamma \in \Omega$ is a block, i.e. there are $\alpha_1, \alpha_2, \ldots, \alpha_m \in \Omega$, $\beta_1, \beta_2, \ldots, \beta_n \in \Omega$ such that $\{\gamma\} = [\alpha_1, \alpha_2, \ldots, \alpha_m| \beta_1, \beta_2, \ldots, \beta_n]$.
2. Every $A \subseteq \Omega$ is union of some blocks, i.e. there are $0 \leq k$, $\alpha_1^k, \alpha_2^k, \ldots, \alpha_{m_k}^k \in \Omega$, $\beta_1^k, \beta_2^k, \ldots, \beta_{n_k}^k \in \Omega$ such that

$$A = \bigcup_{i=1}^{k} [\alpha_1^i, \alpha_2^i, \ldots, \alpha_{m_i}^i|\beta_1^i, \beta_2^i, \ldots, \beta_{n_i}^i]$$

We denote

$$c(A) = min\{k|\exists B_i \text{ blocks}, i = 1, \ldots, k : A = \bigcup_{i=1}^{k} B_i\}$$

We call $c(A)$ the *complexity of* A. If $A = \bigcup_{i=1}^{k} B_i$ where $k = c(A)$ then we say that $A = \bigcup_{i=1}^{k} B_i$ is a *minimal representation* of A by blocks. We should notice that a set $A \subseteq \Omega$ may have different minimal representations, even if we does not take in account of the permutation of blocks. Let us consider an example:

Example 5: (Following the Example 4) As considered in Example 4 let $\alpha = (2, 1, 0)$, $\beta = (1, 1, 1)$, $\gamma = (1, 0, 1)$ and let $\theta = (1, 1, 2)$, $\lambda = (1, 0, 2)$. One can verify

$$\{\gamma\} = U(\gamma) \setminus \{U((2, 0, 1)) \cup U(\beta) \cup U(\lambda)\}$$

and

$$\{\beta, \gamma\} = U(\gamma) \setminus \{U((2, 0, 1)) \cup U((2, 1, 1)) \cup U(\theta) \cup U(\lambda)\}$$

Thus we have $c(\{\beta, \gamma\}) = c(\{\gamma\}) = 1$. One can verify also that $c(\{\alpha, \gamma\}) = 2$.
We have also $c(\{\gamma, \theta, \lambda\}) = 2$ and one can verify that:

$$\begin{aligned}\{\gamma, \theta, \lambda\} &= [\gamma|\beta, \lambda] \cup [\lambda|(1, 2, 2), (1, 0, 3)] \\ &= [\gamma|\beta, (1, 0, 3)] \cup [\theta|(1, 2, 2), (1, 1, 3)]\end{aligned}$$

We use propositional logics in finding the blocks of a given set of MBs. In propositional logics a *full conjunctive clause* is an expression of the form $\bigwedge_{k=1}^{n} x_k$ in which x_k are all variables in either positive or negative form. A *full disjunctive normal form* (full DNF) is a disjunction of full conjunctive clauses. It is well known in propositional logics that all logical formulas can be transformed into full DNF(see, for example, [9]). More exactly, if α is a constraint of items (which is namely a logical formula) then by using simple transformations we can find the full DNF of α:

$$\alpha = \bigvee_{i=1}^{n} [\bigwedge_{k=1}^{m_i} \beta_k^i \wedge \bigwedge_{k=1}^{n_i} (\neg\gamma_k^i)]$$

where for all $\beta \in \Omega$, β appear in every clause of α in either positive or negative form. One can verify that

$$U([\bigwedge_{k=1}^{m_i} \beta_k^i \wedge \bigwedge_{k=1}^{n_i} (\neg\gamma_k^i)]) = [\beta_1^i, \ldots, \beta_{m_i}^i | \gamma_1^i, \ldots, \gamma_{n_i}^i]$$

is a block. By this in fact we have proved the following theorem.

Theorem 8:

1. There is an algorithm by that for any constraint of MBs α we can find the system of full customer blocks of $U(\alpha)$, i.e. we can find

$$\{[\alpha_1^i, \alpha_2^i, \ldots, \alpha_{m_i}^i | \beta_1^i, \beta_2^i, \ldots, \beta_{n_i}^i] | i = 1, 2, \ldots, n\}$$

where $\alpha_1^i, \alpha_2^i, \ldots, \alpha_{m_i}^i, \beta_1^i, \beta_2^i, \ldots, \beta_{n_i}^i$ are all MBs that appear in α, such that

$$U(\alpha) = \bigcup_{i=1}^{k} [\alpha_1^i, \alpha_2^i, \ldots, \alpha_{m_i}^i | \beta_1^i, \beta_2^i, \ldots, \beta_{n_i}^i]$$

2. The decomposition of $U(\alpha)$ into full customer blocks is unique.

3. The minimal representations of $U(\alpha)$ can be obtained by decomposition of $U(\alpha)$ into full customer blocks and by combining some full customer blocks into one to reduce the number of blocks.

4. The complexity of $U(\alpha)$ does not exceed the number of full clauses in the full DNF of α.

Proof:
1. The algorithm that is well known in propositional logics (see [9]) converts a constraint of MBs α into full DNF. By this algorithm we can find the system of full customer blocks of $U(\alpha)$.
2. This is a result in propositional logics (see [9]).
3. If

$$U(\alpha) = \bigcup_{i=1}^{k} [\alpha_1^i, \alpha_2^i, \ldots, \alpha_{m_i}^i | \beta_1^i, \beta_2^i, \ldots, \beta_{n_i}^i]$$

is a minimal representations of $U(\alpha)$ where, for example, some block $[\alpha_1^i, \alpha_2^i, \ldots, \alpha_{m_i}^i | \beta_1^i, \beta_2^i, \ldots, \beta_{n_i}^i]$ is not full. Then using the equivalence $X \equiv (X \wedge a) \vee (X \wedge \neg a)$ we can insert into the block the missing item a. In result we have the decomposition of $U(\alpha)$ into full customer blocks, which, accordingly to 2., is unique. The reverse transformation converts the full DNF of α into the given minimal representation of $U(\alpha)$.
4. The proof is evident by definition of the complexity. The complexity of $U(\alpha)$ is the number of clauses in the minimal representation of α that does not exceed the number of full clauses in the full DNF of α.

Let us consider an example:

Example 6: (Following the Example 4) Let a = "Flour", b = "Egg", c = "Milk", which can be identified by $a = (1, 0, 0)$, $b = (0, 1, 0)$ and $c = (0, 0, 1)$, respectively. The constraint $\alpha = (a \wedge b \rightarrow c)(\neg b \rightarrow (a \vee c))$ characterises the set of all those customers, who if buy flour and egg then buy also milk, and if do not buy egg, then would buy flour or milk. Let us denote this set of customers by A, i.e. $A = U(\alpha)$. By using simple transformations we have the full DNF of α:
$\alpha = (a \wedge b \wedge c) \vee (\neg a \wedge b \wedge c) \vee (\neg a \wedge b \wedge \neg c) \vee (\neg a \wedge \neg b \wedge c)$
$\vee (a \wedge \neg b \wedge c) \vee (a \wedge \neg b \wedge \neg c)$

The full customer blocks of $A = U(\alpha)$ is

$$A = U(\alpha) = [a, b, c] \cup [b, c|a] \cup [a, b|c] \cup [c|a, b] \cup [a, c|b] \cup [a|b, c]$$

One can remark that:

$$\alpha = c \vee (\neg a \wedge b) \vee (a \wedge \neg b)$$

Thus one of the minimal representations of $A = U(\alpha)$ is

$$A = U(\alpha) = [c|] \cup [a|b] \cup [b|a]$$

This means that A can be characterized as the union of three blocks of customers: the first block contains those customers who buy milk, the second block contains all customers who buy flour but do not buy eggs, and the third one is the block of all customers who buy eggs but do not buy flour. One can see that the complexity of A is 3 and the structure of A is clear.

5. CONCLUSION

In this short paper by using the logical structure of the market baskets we have introduced the concept of constraints of market baskets. We showed that every set of market baskets can be characterized by some constraints. The relationships between the structure of the set of market baskets and propositional logics are discovered. Similarly to the well known methods and results in propositional logics we showed that every set of customers can be represented, in different ways, as union of some blocks, and the number of these blocks can be considered as a complexity of the given set of customers.

In fact, the shop managers have to deal with large amounts of itemsets, as well as with the large amounts of market baskets. Finding the efficient algorithms to discover the logical constraints of market baskets is always the problem of practical value.

One should remark that in the last part of this paper we define the complexity for really the *set of market baskets*, not for the *set of customers*, although we call it by the *complexity of the customer sets*. The difference is evident: some customers may buy the same market basket. In this sense the analysis of the customer sets requires more research.

6. REFERENCES

[1] R. Agrawal, R. Srikant (1994) Fast algorithms for mining association rules. VLDB, 487–499.

[2] Benczúr A. , Szabó Gy. I. (2012) Functional Dependencies on Extended Relations Defined by Regular Landuages, Foundations of Information and Knowledge Systems, eds.: T. Lukasiewicz, A. Sali, FoIKS 2012 Kiel, Germany, LNCS 7153, pp. 384–404.

[3] T. Brũggermann, P. Hedstrõm, M. Josefsson (2004) Data mining and Data Based Direct Marketing Activities, Book on Demand GmbH, Norderstedt, Germany.

[4] Davey, B.A., Priestley, H. A. (2002), Introduction to Lattices and Order, Cambridge University Press, ISBN 978-0-521-78451-1.

[5] J. Demetrovics, Hua Nam Son, Akos Guban (2011) An algebraic approach to market basket model: explicit representation of frequent market baskets and associations rules, CSIT 2011. Computer science and information technologies. Proceedings of the conference. Yerevan, pp. 170–173.

[6] J. Demetrovics, Hua Nam Son, Akos Guban (2011) An algebraic representation of frequent market baskets and association rules, Cybernetics and Information Technologies, Vol. 11, No 2, pp. 24–31.

[7] Demetrovics, J., Katona, G.O.H., Dezsö, M., Thalheim, B. (2006) On the number of independent functional dependencies, Lecture Notes in Computer Science, Springer-Verlag, 3861, pp. 83–91.

[8] Heikki Mannila, Hannu Toivonen (1996) Discovering generalized episodes using minimal occurrences. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD' 96), AAAI Press, pp. 146–151.

[9] Mendelson, Elliot (1964), Introduction to Mathematical Logic, D. Van Nostrand Company.

[10] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal (1999) Discovering frequent closed itemsets for association rules. ICDT, 398–416

[11] Ping-Yu Hsu, Yen-Liang Chen, Chun-Ching Ling (2004) Algorithms for mining association rules in bag databases, Information Sciences, Volume 166, Issues 1–4, pp. 31–47.