EXPLORING CONSUMER BEHAVIOR: USE OF ASSOCIATION RULES

Pavel Turčínek[1], Jana Turčínková[2]

[1] Department of Informatics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
[2] Department of Marketing and Trade, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

Abstract

TURČÍNEK PAVEL, TURČÍNKOVÁ JANA. 2015. Exploring Consumer Behavior: Use of Association Rules. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis,* 63(3): 1031–1042.

This paper focuses on problematic of use of association rules in exploring consumer behavior and presents selected results of applied data analyses on data collected via questionnaire survey on a sample of 1127 Czech respondents with structure close to representative sample of population the Czech Republic. The questionnaire survey deals with problematic of shopping for meat products. The objective was to explore possibilities of less frequently used data-mining techniques in processing of customer preference.
For the data analyses, two methods for generating association rules are used: Apriori algorithm and FP-grow algorithm. Both of them were executed in Weka software. The Apriori algorithm seemed to be a better tool, because it has provided finer data, due to the fact that FP-growth algorithm needed reduction of preference scale to only two extreme values, because the input data must be binary. For consumer preferences we also calculated their means.
This paper explores the different preferences and expectations of what customers' favorite outlet should provide, and offer. Customers based on the type of their outlet loyalty were divided into five segments and further explored in more detail. Some of the found best association rules suggest similar patterns across the whole sample, e.g. the results suggest that the respondents for whom a quality of merchandise is a very important factor typically also base their outlet selection on freshness of products. This finding applies to all types of retail loyalty categores. Other rules seem to indicate a behavior more specific for a particular segment of customers. The results suggest that application of association rules in customer research can provide more insight and can be a good supplementary analysis for consumer data exploration when Likert scales were used.

Keywords: knowledge discovery, association rules, Apriori, consumer behavior, marketing research, meat products

INTRODUCTION

Perception of information and communication technology is gradually transformed from something rather unique, bringing a competitive advantage in the market, to the necessity of conditioning the existence or not existence of business between the competitive business organizations (Chalupová, Motyčka, 2008). To imagine data exploration of any type without use of information technologies is in these days unthinkable.

The issue of consumer behavior is explored in the field of customer relationship management. Customer Relationship Management (CRM) is seen as a holistic framework for interaction of organizations with their customers (Dařena, 2008). CRM uses marketing research as a strong tool to explore consumer behavior. Marketing research, a process of collecting and using information for marketing decision making (Boone, Kurt, 2013), plays an essential role in customer relationship management. Tools to facilitate individual steps of marketing research, particularly collection of data and their analysis can be more effective through increased use of databases and data mining techniques (Bradly, 2007; Kříž, Dostál, 2010). As a part of a Marketing Information System (Dařena, 2007) such tools provide decision makers with a continuous flow of information relevant to their area of responsibility (Boone, Kurt, 2013).

Every organization should focus on optimizing the workflows while ensuring compliance with regulation and dynamically responding to the market situation and customer requirements (Rábová, 2012). Use of Business Intelligence (Kříž, Klčová, Sodomka, 2011) and other modern approaches (Knížek *et al.*, 2011) are absolutely inherent in managerial decision-making in these days.

Each consumer behaves in his individual way. There might be various motives for shopping. In order to make marketer's work more efficient, it is recommended to use market segmentation which is based on the need to aggregate consumers into groups that combine certain characteristics they have in common. If this is done properly, consumers within a segment should react similarly to same market stimuli (Bárta *et al.*, 2009).

The segmentation can be approached: a. intuitively, b. based on experience, c. systematically and deliberately. These approaches can therefore be divided into two groups:

1. A priori segmentation – when marketer decides for one of the criteria which then determines belonging to a particular segment.
2. Post hoc segmentation – the number of segments is not known in advance, but through data processing relationships between major criteria are uncovered, and then subsequently selected such criteria that allow for the segment to meet requirements of homogeneity and heterogeneity (Bártová *et al.*, 2005; Suarez *et al.*, 2004).

Numerous research studies (starting in 70's of the 20th century) have focused on preferences for retail channels, choice of retail format and patronage behavior, depending on both external and internal factors to the consumer, mostly with emphasis on understanding the influence of store attributes, situational factors, households demographics, shopping patterns, price levels, attitudes toward stores (e.g. Monroe, Guiltinan, 1975; Arnold, Ma, Tiger, 1978; Mason, Durand, Taylor, 1983; Keng, Ehrenberg, 1984; Louvier, Geath, 1987; Spiggle, Sewall, 1987; Dawson, Bloch, Ridgway, 1990; Burke *et al.*, 1992; Woodside and Trappey, 1992; Medina, Ward, 1999; Outi, 2001; Sinha, Banerjee, 2004; Carpenter, Moore, 2006).

The most popular basis for segmenting customer groups is demographic segmentation, based on age, gender, family size, family life cycle, income, occupation, education, religion, race or nationality (Kotler, 2007), however, needs, aspirations and desires of consumers are changing frequently in close association with these variables; therefore, it is necessary to segment on a different basis. The choice of retail mix elements is for vendors crucial (Milosavljevic *et al.*, 2011). Numerous studies have identified a great variety of factors influencing the store choice of various consumers. For example Morschett *et al.* (2005) points out a product quality, assortment, price, one stop purchase, speed and quality of service and the atmosphere of the in-store experience as the most important ones. Briesch *et al.* (2009) states it is the overall comfort of customers during the purchase process that matters. Netopil and Kalábová (2013) proved an influence of information posted in flyers (leaflets) regularly delivered by retail chain to household on consumers' choice of their outlet, with higher impact on seniors than young customers.

The share of expenditure on food of Czech households is declining, and getting close to the European average. Competitive environment, with the influence of past economic crisis certainly has a significant impact on all markets including the food one. In the past decade, the number of hypermarkets in Czech Republic more than tripled. Results of Shopping monitor by Incoma (2010) describe hypermarkets, supermarkets and discount stores as the main place of realizing grocery purchases in Czech Republic, with as supermarkets, hypermarkets and small local shops the most popular. Only a limited number of customers have their one favorite store or chain store which is visited regularly and rarely buy elsewhere (Netopil, 2012).

It is now more than ever crucial to get to know our customer. Therefore, the question is how food retailers can maintain their regular clients to continue shopping in their stores? Customer loyalty is understood as both subjective and behavioral tendency to prefer one brand (product) over all others. Customer orientation is a priority on a market, where an increasing number of firms are competing for every customer (Lošťáková, 2009). The higher the quality of the products, however, does not automatically mean the higher customer satisfaction (Zamazalová, 2008). Customer satisfaction leads not only to repeated purchases, but also extends customer loyalty to a specific firm (Horáková, 1992). Faed and Forbes (2011) reported an increase in profitability of the business by 25–85% already at 5% increase in customer retention.

The aim of this paper is to summarize results of the research project "Application of modern methods to data processing in the field of marketing research" which was solved at the Department of Informatics, Faculty of Business and Economics of Mendel University in Brno. Most of the results were presented at international conferences.

METHODS AND RESOURCES

Association Rules

Let $I = I_1, I_2, \ldots, I_m$ be a set of binary attributes, called items. Let T be database of transactions. Each transaction is represented as a binary vector, with $t[k] = 1$ if t bought the item I_k, and $t[k] = 0$ otherwise. There is one tuple in database for each transaction. Let X be a set of some items in I. Agrawal *et al.* (1993) say that a transaction t satisfies X if for all items I_k in X, $t[k] = 1$. By an association rule, Agrawal *et al.* (1993) mean an implication of form $X \Rightarrow I_j$, where X is a set of some items in I, and I_j is a single item in I

that is not presented in X. The rule $X \Rightarrow I_j$ is satisfied in the set of transactions T with confidence factor $0 \leq c \leq 1$ if at least c% of transaction in T that satisfy X also satisfy I_j. It is given notation $X \Rightarrow I_j | c$ to specify that the rule $X \Rightarrow I_j$ has confidence factor c.

In other words if Han and Kamber (2006) think of the universe as a set of items at the store, then each item has a Boolean variable representing the presence of that item. Each basket can then be represented by Boolean vector of values assigned to these variables. The Boolean vector can be analyzed for buying patterns that reflect items that are frequently associated or purchased together. These patterns can be represented in form of association rules. Example of buying computers and antivirus together is shown below:

$$Computer \Rightarrow antivirus \: [support = 2\%, confidence = 60\%].$$

As you can see, there are two metrics how to measure rule interestingness. A **support** of 2% for Association Rule means that 2% of all transaction under analysis shows that computer and antivirus are purchased together. A **confidence** of 60% means that 60% of customers who purchased a computer also bought the antivirus software.

Apriori

The essential idea is to iteratively generate the set of candidate patterns of length $(k + 1)$ from the set of frequent patterns of length k (for $k \geq 1$), and checks their corresponding occurrence frequencies (Han *et al*, 2004). More about Apriori algorithm can be found in e.g. Han, Kamber (2006), Han *et al.* (2004), or Zaki (2000).

Frequent Pattern Mining

Frequent pattern mining plays an essential role in mining association rules. Most of the studies adopt an *Apriori*-like approach. However, candidate set generation is still costly, especially when prolific patterns and/or long patterns exist there (Han, Pei, Yin, 2000).

An interesting method is **frequent-pattern growth** (**FP-growth**), which adopts a *divide-and-conquer* strategy (Han, Kamber, 2006). First, the compact data structure, called frequent pattern tree (FP-tree), is constructed, which is an extended prefix-tree structure storing crucial, quantitative information about frequent patterns. Only frequent length-1 items will have nodes in the tree.

Second, an FP-tree-based pattern fragment growth mining method, is developed, which starts from a frequent length-1 pattern, examines only its conditional pattern base, constructs its (conditional) FP-tree, and performs mining recursively with such a tree. The pattern growth is achieved via concatenation of the suffix pattern with the new ones generated from a conditional FP-tree.

Third, the search technique employed in mining is a partitioning-based, divide-and-conquer method rather than Apriori-like bottom-up generation of frequent item sets combinations. This dramatically reduces the size of conditional pattern base generated at the subsequent level of search as well as the size of its corresponding conditional FP-tree (Han, Pei, Yin, Mao, 2004).

Data Source

In this paper are presented selected results of applied data analyses on data collected via questionnaire survey on a sample of 1127 Czech respondents with structure close to representative sample of population the Czech Republic (data collection took place in 2011). Customers have different preferences and expectations of what their favorite outlet should provide and offer. A list which included following 20 items was set up and respondents were asked to mark how important they were for them when selecting a particular outlet for shopping for meat products:
- Current offers in leaflets;
- A convenient store location;
- A possibility to park next to the store;
- Proximity to public transport stops;
- Low prices;
- Friendly and courteous store staff;
- Fast service and care;
- Short waiting times at cash registers;
- Easy handling of claims and complaints;
- Cleanliness and store layout;
- Easy orientation inside the store (easy to find merchandise);
- Correct indication of products (most of all the prices);
- Wide assortment;
- Sufficient stock of merchandise currently in a special offer;
- Fresh products;
- Quality of merchandise;
- The store offers Czech and local products;
- Possibility to pay with lunch tickets;
- Bags free of charge;
- Habit.

For respondents a Likert scale from 1 to 10 was used, where 1 = completely unimportant and 10 = very important, however, for better results of our analyses when using Apriori algorithm it was decided to reduce the scale from 1–10 to 1–5 in our data processing, using the rule where original scores 9 and 10 were replaced with 5, then 7 and 8 with 4, etc. It is believed that the higher level of satisfaction a customer experiences, the higher his loyalty to the outlet would be. The FP-growth method in Weka is suited just for binary inputs so we had to reduce scale for this algorithm from 1–10 to 1–2. We used the same principle of replacing as in previous (Apriori) method just 10–6 become 2 and 5–1 become 1.

Our aim was to find a way of how to classify respondents according to their answer to a question:

"How many outlets (stores) do you usually use for your shopping for these products?", thus, what represents their store loyalty, based on above given 20 factors which characterize their customer behavior. The possible answers were:

- I have one favorite store where I do my regular shopping. I go elsewhere only exceptionally.
- I have one favorite retail chain where I do my regular shopping. I go elsewhere only exceptionally.
- I have several stores I prefer, and I make my choice of a particular outlet regardless their current special offers and sales (I base my choice on other attributes).
- I switch among the outlets based on their current special offers advertised in leaflets.
- I chose the outlets randomly, based on their proximity at the moment.

Our aim was to analyze with use of various methods (see also Turčínková *et al.*, 2014) whether there are any significant characteristics differentiating respondents based on their loyalty type. The total sample based on the loyalty type had a following structure as indicated in Fig. 1.

For the individual outlet loyalty category and for the whole sample we calculated mean levels of importance that our respondents assigned to individual factors according to their importance on respondents' choice of outlet when shopping for meat products. Statistical software Statistica v. 10 was used.

RESULTS

For finding association rules were used methods implemented in software Weka (2013). Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization.

In our research the focus was on association rules. Two different methods were used: *Apriori* and *FPGrowth*. The questions in survey we focused on were evaluated on scale of 1 to 10 (as mentioned earlier, the original Likert scale 1–10 was reduced to 1–5 for processing and then numbers were replaced with letters, so the final values use in our analyses were: 1 = a, 2 = b, 3 = c, 4 = d and 5 = e). *FPGrowth* algorithm in Weka, however, accepts only binary inputs, so the data had to be transform into acceptable form. Also data pro calculation by *Apriori* algorithm had to be transformed. Both methods have parameters among others support and confidence which were used to adjust the results.

Respondents were divided into five groups according their answer to question: ***"How many outlets (stores) do you usually use for your shopping for these products[1]?"***

1. I have one favorite store where I do my regular shopping. I go elsewhere only exceptionally (273 respondents).
2. I have one favorite retail chain where I do my regular shopping. I go elsewhere only exceptionally (155 respondents).
3. I have several stores I prefer, and I make my choice of a particular outlet regardless their current special offers and sales (I base my choice on other attributes) (272 respondents).
4. I switch among the outlets based on their current special offers advertised in leaflets (154 respondents).
5. I chose the outlets randomly, based on their proximity at the moment (269 respondents).

The Fig. 1 depicts relative shares of individual outlet loyalty categories in the whole sample.
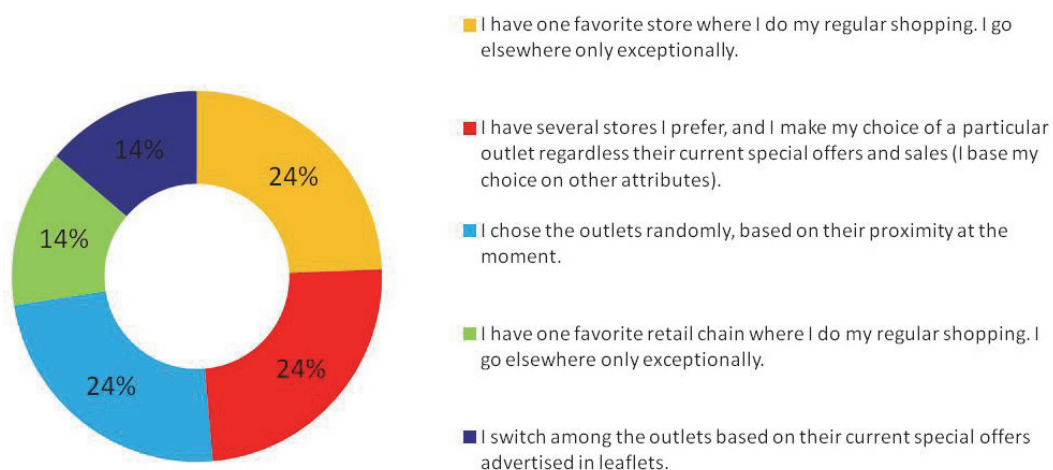
1: *How many outlets (stores) do you usually use for your shopping for these products? Select the option that most reflects your shopping habits.*
Source: Questionnaire survey, MENDELU, 2011, n = 1127

---

1 i.e. groceries

I: *Mean levels of importance of factors influencing outlet choice when shopping for meat products (Likert scale 1 to 10, 1 = unimportant)*

Mean	Respondents' loyalty category					
	outlet loyal[1]	retail chain loyal[2]	several favorite stores[3]	special offers[4]	proximity[5]	all
Fresh products	9.03	9.03	9.34	8.96	8.87	9.06
Quality of merchandise	8.83	9.00	9.28	8.60	8.65	8.89
Correct indication of products (most of all the prices)	8.35	8.36	8.38	7.93	7.84	8.18
Cleanliness and store layout	8.30	8.17	8.56	7.74	7.77	8.15
Easy orientation inside the store (easy to find merchandise)	7.94	7.78	7.93	7.57	7.52	7.77
Wide assortment	7.44	7.94	7.98	7.65	7.64	7.72
Short waiting times at cash registers	7.60	7.69	7.66	7.54	7.90	7.69
Fast service and care	7.56	7.57	7.64	7.52	7.81	7.64
A convenient store location	7.43	7.50	6.85	7.48	8.33	7.52
Low prices	7.09	7.18	6.66	8.31	7.00	7.14
The store offers Czech and local products	7.35	7.04	7.79	6.73	6.08	7.02
Friendly and courteous store staff	7.17	7.10	7.30	7.05	6.37	6.98
Sufficient stock of merchandise currently in a special offer	6.61	7.58	6.64	8.10	6.25	6.87
Easy handling of claims and complaints	6.44	6.80	6.48	6.82	5.44	6.32
Habit	6.98	6.67	6.08	5.64	5.24	6.12
Current offers in leaflets	5.45	6.34	5.05	7.95	4.06	5.48
A possibility to park next to the store	4.73	5.83	5.86	5.05	4.54	5.15
Proximity to public transport stops	4.80	4.71	4.17	5.66	6.00	5.04
Bags free of charge	3.95	4.30	3.30	4.46	3.75	3.87
Possibility to pay with lunch tickets	3.25	4.12	4.09	3.99	3.60	3.76

Note: The headings are short versions of following statements:
[1] I have one favorite store where I do my regular shopping. I go elsewhere only exceptionally.
[2] I have one favorite retail chain where I do my regular shopping. I go elsewhere only exceptionally.
[3] I have several stores I prefer and I make my choice of a particular outlet regardless their current special offers and sales (I base my choice on other attributes).
[4] I switch among the outlets based on their current special offers advertised in leaflets.
[5] I chose the outlets randomly, based on their proximity at the moment.
Source: Questionnaire survey, MENDELU, 2011, n = 1127

The calculated means suggested that there are differences among customers according to their outlet loyalty category based on their choice of answer how many outlets (stores) they usually use for their shopping for meat products.

The Tab. I shows that the highest levels of importance for majority of the factors were assigned by customers who have several preferred stores and their choice of the outlet is based on other attributes than current special offers. They seem to be the most demanding and expecting a high level of performance.

Customers that base their outlet choice on current special offers have a different profile. Not that a quality of merchandise would not matter to them at all, but when compared to other categories their level of importance is the lowest on average. They are also the least demanding on the cleanliness and store layout. Their answers show that it really are the price cues (low prices, and sufficient stock of merchandise currently in a special offer) that influence them the most of all outlet loyalty categories. With some exaggeration they could perhaps be called bargain shoppers.

For customers who stress the importance of outlet proximity the results show high levels of importance of convenient location, fast services and care, short waiting times at cash registers and proximity to public transport stops. These customers seem to value their time and are willing to reduce their expectations on other attributes if a particular outlet can help them eliminate waste of time.

Both algorithms in Weka were used to generate association rules. For each group parameters of support and confidence were modified to get reasonable number of rules. In the following text, we choose the most interesting ones. However, the results of Apriori method provided finer results than the FP-growth method. This is caused by reduction of the scale because FP-growth algorithm allows just binary input.

II: *Best rules found with use of Apriori method*

Best rules found (and score) Min. support: 0.5 Min. metric <confidence>: 0.1	Respondents' loyalty category				
	outlet loyal[1]	retail chain loyal[2]	several favorite stores[3]	special offers[4]	proximity[5]
Fresh products (e)	75.46	78.71	84.19	73.55	69.26
Quality of merchandise	70.33	77.42	81.62	62.81	62.59
Fresh products (e) + Quality of merchandise (e)	67.03	74.19	79.78	59.35	58.15
Correct indication of products (most of all the prices) (e)	56.41	56.13	55.51		
Correct indication of products (most of all the prices) (e) + Fresh products (e)		52.26	52.57		
Correct indication of products (most of all the prices) (e) + Quality of merchandise (e)			51.84		
Correct indication of products (most of all the prices) (e) + Fresh products (e) + Quality of merchandise (e)			51.47		
Cleanliness and store layout (e)	55.68	54.84	61.40		
Cleanliness and store layout (e) + Fresh products (e)	51.28	50.97	58.46		
Cleanliness and store layout (e) + Fresh products (e) + Quality of merchandise (e)			56.61		
Cleanliness and store layout (e) + Quality of merchandise (e)			56.99		
Low prices (e)				56.77	
Current offers in leaflets (e)				50.32	
A convenient store location (e)					56.30
Possibility to pay with lunch tickets (a)	54.95				
Bags free of charge (a)			54.78		

Note: The headings are short versions of statements as indicated below Tab. I.
Source: Questionnaire survey, MENDELU, 2011, n = 1127

Tab. II presents results that suggest again that customers with several favorite outlets seem to be the most demanding, however this time with even more detail.

Let us explore the customer categories in more detail individually.

I Have One Favorite Store Where I Do My Regular Shopping. I Go Elsewhere Only Exceptionally.

In general, customers who picked this option have a lot in common with customers who have one favorite retail chain; however, they seem to be less demanding as we can see in Tab. I. It may be important to note that their share in sample is higher than of those who are retail chain loyal. What really matters to them is an easy orientation inside the store and easy location of merchandise. A habit plays a great role in their shopping behavior. As the found best rules confirm, also for this category of customers provided quality of merchandise and freshness of products play an important role when choosing their outlet for a purchase of meat products, their level is very close to the mean of the whole sample. The analysis has not provided any category specific rule, but the results suggest that the respondents for whom the quality of merchandise is important also stress freshness of products, they also expect correct indication of products, cleanliness and store layout, because the assigned highest levels of importance of these factors on their choice of outlet where they would buy meat products.

To find reasonable amount of rules the metrics of minimum support were set on 0.4 (109 instances) and minimum confidence to 0.8.

I Have One Favorite Retail Chain Where I Do My Regular Shopping. I Go Elsewhere Only Exceptionally.

Even though these customers do not stand out in any level of importance, but the possibility to pay with lunch tickets (they exhibited the highest level of importance of this option when selecting an outlet for their meat product purchases), they are mostly more demanding on store performance than the outlet loyal customers. That is possible to see also from the results of Apriori method. They expect wide assortment, but about a half of them insists it must be of the highest quality merchandise and fresh. Short waiting times are of greater importance to them.

To find reasonable amount of rules the metrics of minimum support were set on 0.4 (62 instances) and minimum confidence to 0.9.

III: *Best rules found for outlet loyal customers*

	Best rules found	conf.
1	Correct indication of products (most of all the prices)=e Quality of merchandise=e 128 ⇒ Fresh products=e 124	0.97
2	Cleanliness and store layout=e Quality of merchandise=e 136 ⇒ Fresh products=e 131	0.96
3	Quality of merchandise=e 192 ⇒ Fresh products=e 183	0.95
4	Cleanliness and store layout=e Fresh products=e 140 ⇒ Quality of merchandise=e 131	0.94
5	Cleanliness and store layout=e 152 ⇒ Fresh products=e 140	0.92
6	Correct indication of products (most of all the prices)=e Fresh products=e 135 ⇒ Quality of merchandise=e 124	0.92
7	Cleanliness and store layout=e 152 ⇒ Quality of merchandise=e 136	0.89
8	Fresh products=e 206 ⇒ Quality of merchandise=e 183	0.89
9	Easy orientation inside the store (easy to find merchandise)=e 131 ⇒ Fresh products=e 116	0.89
10	Correct indication of products (most of all the prices)=e 154 ⇒ Fresh products=e 135	0.88
11	Easy orientation inside the store (easy to find merchandise)=e 131 ⇒ Correct indication of products (most of all the prices)=e 113	0.86
12	Cleanliness and store layout=e 152⇒ Fresh products=e Quality of merchandise=e 131	0.86
13	Correct indication of products (most of all the prices)=e 154 ⇒ Quality of merchandise=e 128	0.83
14	Correct indication of products (most of all the prices)=e 154 ⇒ Fresh products=e Quality of merchandise=e 124	0.81

Source: Generated in Weka, based on questionnaire survey, MENDELU, 2011, n = 273

IV: *Best rules found for retail chain loyal customers*

	Best rules found	conf.
1	Cleanliness and store layout=e Quality of merchandise=e 77 ⇒ Fresh products=e 77	1
2	Correct indication of products (most of all the prices)=e Quality of merchandise=e 76 ⇒ Fresh products=e 76	1
3	Wide assortment=e Quality of merchandise=e 63 ⇒ Fresh products=e 63	1
4	Easy orientation inside the store (easy to find merchandise)=e Quality of merchandise=e 62 ⇒ Fresh products=e 62	1
5	Cleanliness and store layout=e Fresh products=e 79 ⇒ Quality of merchandise=e 77	0.97
6	Wide assortment=e 68 ⇒ Fresh products=e 66	0.97
7	Easy orientation inside the store (easy to find merchandise)=e Fresh products=e 64 ⇒ Quality of merchandise=e 62	0.97
8	Quality of merchandise=e 120 ⇒ Fresh products=e 115	0.96
9	Easy orientation inside the store (easy to find merchandise)=e 67 ⇒ Fresh products=e 64	0.96
10	Short waiting times at cash registers=e Quality of merchandise=e 67 ⇒ Fresh products=e 64	0.96
11	Wide assortment=e Fresh products=e 66 ⇒ Quality of merchandise=e 63	0.95
12	Short waiting times at cash registers=e 72 ⇒ Fresh products=e 68	0.94
13	Fresh products=e 122 ⇒ Quality of merchandise=e 115	0.94
14	Short waiting times at cash registers=e Fresh products=e 68 ⇒ Quality of merchandise=e 64	0.94
15	Correct indication of products (most of all the prices)=e Fresh products=e 81 ⇒ Quality of merchandise=e 76	0.94
16	Correct indication of products (most of all the prices)=e 87 ⇒ Fresh products=e 81	0.93
17	Short waiting times at cash registers=e 72 ⇒ Quality of merchandise=e 67	0.93
18	Cleanliness and store layout=e 85 ⇒ Fresh products=e 79	0.93
19	Wide assortment=e 68 ⇒ Quality of merchandise=e 63	0.93
20	Wide assortment=e 68 ⇒ Fresh products=e Quality of merchandise=e 63	0.93
21	Easy orientation inside the store (easy to find merchandise)=e 67 ⇒ Quality of merchandise=e 62	0.93
22	Easy orientation inside the store (easy to find merchandise)=e 67 ⇒ Fresh products=e Quality of merchandise=e 62	0.93
23	Cleanliness and store layout=e 85 ⇒ Quality of merchandise=e 77	0.91

Source: Generated in Weka, based on questionnaire survey, MENDELU, 2011, n = 155

I Have Several Stores I Prefer, and I Make My Choice of a Particular Outlet Regardless Their Current Special Offers and Sales (I Base My Choice on Other Attributes).

Customers who have several favorite stores have the highest expectations on product quality, freshness, cleanliness and layout of stores, correct indication of products width of assortment, availability of Czech and local products, but they are the least sensitive to price level and offer of bags free of charge. They also value friendly and courteous store staff. Results of Apriori analysis show strong

V: *Best rules found for customers with several preferred outlets*

	Best rules found	conf.
1	Quality of merchandise=e The store offers Czech and local products=e 118 ⇒ Fresh products=e 118	1
2	Fresh products=e The store offers Czech and local products=e 118 ⇒ Quality of merchandise=e 118	1
3	Wide assortment=e Fresh products=e 114 ⇒ Quality of merchandise=e 114	1
4	Cleanliness and store layout=e Correct indication of products (most of all the prices)=e Quality of merchandise=e 114 ⇒ Fresh products=e 114	1
5	Cleanliness and store layout=e Quality of merchandise=e 155 ⇒ Fresh products=e 154	0.99
6	Correct indication of products (most of all the prices)=e Quality of merchandise=e 141 ⇒ Fresh products=e 140	0.99
7	Cleanliness and store layout=e Correct indication of products (most of all the prices)=e 117 ⇒ Fresh products=e 116	0.99
8	Easy orientation inside the store (easy to find merchandise)=e Quality of merchandise=e 117 ⇒ Fresh products=e 116	0.99
9	The store offers Czech and local products=e 120 ⇒ Fresh products=e 118	0.98
10	The store offers Czech and local products=e 120 ⇒ Quality of merchandise=e 118	0.98
11	The store offers Czech and local products=e 120 ⇒ Fresh products=e Quality of merchandise=e 118	0.98
12	Wide assortment=e Quality of merchandise=e 116 ⇒ Fresh products=e 114	0.98
13	Cleanliness and store layout=e Correct indication of products (most of all the prices)=e Fresh products=e 116 ⇒ Quality of merchandise=e 114	0.98
14	Correct indication of products (most of all the prices)=e Fresh products=e 143 ⇒ Quality of merchandise=e 140	0.98
15	Quality of merchandise=e 222 ⇒ Fresh products=e 217	0.98
16	Quality of merchandise=e Bags free of charge=a 132 ⇒ Fresh products=e 129	0.98
17	Fresh products=e Bags free of charge=a 132 ⇒ Quality of merchandise=e 129	0.98
18	Cleanliness and store layout=e Correct indication of products (most of all the prices)=e 117 ⇒ Quality of#merchandise=e 114	0.97
19	Cleanliness and store layout=e Correct indication of products (most of all the prices)=e 117 ⇒ Fresh products=e Quality of merchandise=e 114	0.97
20	Cleanliness and store layout=e Fresh products=e 159 ⇒ Quality of merchandise=e 154	0.97
21	Easy orientation inside the store (easy to find merchandise)=e Fresh products=e 121 ⇒ Quality of merchandise=e 116	0.96
22	Cleanliness and store layout=e 167 ⇒ Fresh products=e 159	0.95
23	Fresh products=e 229 ⇒ Quality of merchandise=e 217	0.95
24	Correct indication of products (most of all the prices)=e 151 ⇒ Fresh products=e 143	0.95
25	Easy orientation inside the store (easy to find merchandise)=e 128 ⇒ Fresh products=e 121	0.95
26	Wide assortment=e 124 ⇒ Quality of merchandise=e 116	0.94
27	Correct indication of products (most of all the prices)=e 151 ⇒ Quality of merchandise=e 141	0.93
28	Cleanliness and store layout=e 167 ⇒ Quality of merchandise=e 155	0.93
29	Correct indication of products (most of all the prices)=e 151 ⇒ Fresh products=e Quality of merchandise=e 140	0.93
30	Cleanliness and store layout=e 167 ⇒ Fresh products=e Quality of merchandise=e 154	0.92
31	Wide assortment=e 124 ⇒ Fresh products=e 114	0.92
32	Wide assortment=e 124 ⇒ Fresh products=e Quality of merchandise=e 114	0.92
33	Easy orientation inside the store (easy to find merchandise)=e 128 ⇒ Quality of merchandise=e 117	0.91
34	Easy orientation inside the store (easy to find merchandise)=e 128 ⇒ Fresh products=e Quality of merchandise=e 116	0.91

Source: Generated in Weka, based on questionnaire survey, MENDELU, 2011, n = 272

connection of expectation level among quality of merchandise, freshness of products, correct indication of products, but also the offer of Czech and local product.

To find reasonable amount of rules the metrics of minimum support were set on 0.4 (109 instances) and minimum confidence to 0.9.

I Switch Among the Outlets Based on Their Current Special Offers Advertised in Leaflets.

Customers who base their outlet selection on current special offers are relatively the least demanding in quality of merchandise expectations and cleanliness and store layout when compared with the other four groups. On the other hand, as we could expect, they base their choice on price level and sufficient stock of merchandise currently in a special offer the most out of all. However, more than anybody else they care for bags free of charge.

What the best rules found show is interesting too. The rules show that also customers who expect low prices and sufficient stock of merchandise currently in a special offer also expect that the merchandise will be of high quality of merchandise and the products will be fresh. To find reasonable amount of rules the metrics of minimum support were set on 0.35 (54 instances) and minimum confidence to 0.8.

VI: *Best rules found for customers basing choice of outlet on current special offers*

	Best rules found	conf.
1	Correct indication of products (most of all the prices)=e Quality of merchandise=e 55 ⇒ Fresh products=e 54	0.98
2	Sufficient stock of merchandise currently in a special offer=e Quality of merchandise=e 56 ⇒ Fresh products=e 54	0.96
3	Wide assortment=e 60 ⇒ Fresh products=e 57	0.95
4	Quality of merchandise=e 97 ⇒ Fresh products=e 92	0.95
5	Cleanliness and store layout=e 61 ⇒ Fresh products=e 57	0.93
6	Low prices=e Quality of merchandise=e 60 ⇒ Fresh products=e 56	0.93
7	Easy orientation inside the store (easy to find merchandise)=e 63 ⇒ Fresh products=e 55	0.87
8	Easy orientation inside the store (easy to find merchandise)=e 63 ⇒ Correct indication of products (most of all the prices)=e 54	0.85
9	Correct indication of products (most of all the prices)=e 75 ⇒ Fresh products=e 64	0.85
10	Correct indication of products (most of all the prices)=e Fresh products=e 64 ⇒ Quality of merchandise=e 54	0.84
11	Sufficient stock of merchandise currently in a special offer=e Fresh products=e 64 ⇒ Quality of merchandise=e 54	0.84
12	Sufficient stock of merchandise currently in a special offer=e 76 ⇒ Fresh products=e 64	0.84
13	Low prices=e Fresh products=e 68 ⇒ Quality of merchandise=e 56	0.82
14	Fresh products=e 114 ⇒ Quality of merchandise=e 92	0.81

Source: Generated in Weka, based on questionnaire survey, MENDELU, 2011, n = 154

VII: *Best rules found for customers who value outlet proximity*

	Best rules found	conf.
1	Cleanliness and store layout=e Quality of merchandise=e 95 ⇒ Fresh products=e 91	0.96
2	Correct indication of products (most of all the prices)=e Quality of merchandise=e 101 ⇒ Fresh products=e 96	0.95
3	Quality of merchandise=e 169 ⇒ Fresh products=e 157	0.93
4	A convenient store location=e Quality of merchandise=e 105 ⇒ Fresh products=e 96	0.91
5	Correct indication of products (most of all the prices)=e Fresh products=e 106 ⇒ Quality of merchandise=e 96	0.91
6	Quality of merchandise=e Possibility to pay with lunch tickets=a 93 ⇒ Fresh products=e 84	0.90
7	Cleanliness and store layout=e 114 ⇒ Fresh products=e 102	0.89
8	Cleanliness and store layout=e Fresh products=e 102 ⇒ Quality of merchandise=e 91	0.89
9	A convenient store location=e Fresh products=e 109 ⇒ Quality of merchandise=e 96	0.89
10	Correct indication of products (most of all the prices)=e 124 ⇒ Fresh products=e 106	0.85
11	Fast service and care=e 109 ⇒ Short waiting times at cash registers=e 92	0.84
12	Fresh products=e 187 ⇒ Quality of merchandise=e 157	0.84
13	Cleanliness and store layout=e 114 ⇒ Quality of merchandise=e 95	0.83
14	Correct indication of products (most of all the prices)=e 124 ⇒ Quality of merchandise=e 101	0.81
15	Fast service and care=e 109 ⇒ Fresh products=e 88	0.81

Source: Generated in Weka, based on questionnaire survey, MENDELU, 2011, n = 269

I Chose the Outlets Randomly, Based on Their Proximity at the Moment.

Customers who seek proximity most probably do so in order to avoid waste of time. As the mean values for individual factors influencing outlet choice show, they have the highest expectations on short waiting times at cash registers, fast service and care and, of course, convenient store location or at least proximity to public transport stop. For most of other factors the mean values are the lowest relatively to other groups.

To find reasonable amount of rules the metrics of minimum support were set on 0.3 (81 instances) and minimum confidence to 0.8. The application of Apriori method has not, however, provided any unexpected results (see Tab. VII).

DISCUSSION

In this paper the focus was on possibilities of use of association rules in exploring consumer behavior. It presents selected results of applied data analyses on data collected via questionnaire survey on a sample of 1127 Czech respondents with structure close to representative sample of population the Czech Republic. The questionnaire survey deals with problematic of shopping for meat products. Two methods for generating association rules were used: Apriori algorithm and FP-grow algorithm for analyzing the data. Both of them were executed in Weka (2013) software which is data mining tool. This paper explores the different preferences and expectations of what customers' favorite outlet should provide, and offer. The Apriori algorithm seemed to be a better tool, because it has provided finer data, because FP-growth algorithm needed reduction of preference scale to only two extreme values, because the input data must be binary.

The advantage of use of association rules is their simplicity and easy understanding for anyone. Association rules were originally used to analyze consumer basket. Thus, it was sufficient information whether the product is in the basket or not (binary approach as FP-growth method). While Apriori method allows researchers to work with multiple values within a single criterion. These various values are then treated as all equally different. If the range is used, both the differences between values 1 and 5 and between values 1 and 2 is treated as the same. This leads to a loss of a certain part of the information. Improvements could be implemented by use of optimized methods for association rules, as proposed by Brin, Rastogi and Shim (2003).

There was a number of association rules that were common for all customer segments no matter what their outlet loyalty type was which helped to confirm that customers in general share certain common traits, e.g. those who have high expectations on quality of merchandise also demand freshness of products. Because there were more variations in preferences in the middle of the Likert scale, the best association were found mostly only for the customer choices with the highest scores (typically 5). Therefore, we could see that application of associations rules in customer research can provide more insight and can be a good supplementary analysis for consumer data exploration when Likert scales were used.

REFERENCES

AGRAWAL, R., IMIELINSKI, T. and SWAMI, A. 1993. Mining association rules between sets of items in large databases. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 22(2): 207–216.

ARNOLD, S. J., MA; S. and TIGERT, D. J. 1978. A comparative analysis of determinant attributes in retail store selection. In: HUNT, H. K. (ed.), *Advances in Consumer Research*, 5: 663–667.

BÁRTA, V., PÁTÍK, L. and POSTLER, M. 2009. *Retail marketing*. 1st ed. Praha: Management Press.

BÁRTOVÁ, H., BÁRTA, V. and KOUDELKA, J. 2005. *Chování spotřebitele a výzkum trhu* [*Consumer behavior and marketing research*]. 2. vyd. Praha: Oeconomika.

BOONE, L. E. and KURT, D. L. 2013. *Contemporary Marketing*. Mason: Cengage Learning.

BRADLEY, N. 2007. *Marketing Research: Tools and Techniques*. New York: Oxford University Press.

BRIESCH, R., CHINTAGUNTA, P. and FOX, E. 2009. How does assortment affect grocery store choice? *Journal of Marketing research*, 46: 176–189.

BRIN, S., RASTOGI, R. and SHIM, K. 2003. Mining optimized gain rules for numeric attributes. *IEEE Transactions on Knowledge and Data Engineering*, 15(2): 324–338.

BURKE, R., BARI, R., HARLAM, A. et al. 1992. Comparing dynamic consumer choice in real and computer-simulated environments. *Journal of Consumer Research*, 19(1): 71–82.

CARPENTER, J. and MOORE, M. 2006. Consumer demographics, store attributes, and retail format choice in the US grocery market. *International Journal of Retail & Distribution Management*, 34(6): 434–452.

CHALUPOVÁ, N. and MOTYČKA, A. 2008. Situation and trends in trade-supporting information technologies. *Acta Univ. Agric. Silvic. Mendelianae Brun.*, 56(6): 25–36.

DAŘENA, F. 2007. Global architecture of marketing information systems. *Agricultural Economics*, 52(9): 432–440.

DAŘENA, F. 2008. A research on CRM systems in the Czech Republic. *Acta Univ. Agric. Silvic. Mendelianae Brun.*, 56(3): 29–34.

DAWSON, S., BLOCH, P. H. and RIDWAY, N. M. 1990. Shopping motives, emotional states and retail outcomes. *Journal of Retailing*, 66(4): 408–427.

FAED, A. and FORBES, D. 2011. Impact of Customer Management System in Improving Customer Retention: Optimization of Negative Customer Feedback. *World Academy of Science, Engineering & Technology* 72: 171–175. EBSCO. [online]. Accessible at: <http://ehis.ebscohost.com/eds/detail?vid=4&hid=120&sid=7f441ac1-9b4e-4676-b0130215eeced880%40sessionmgr13-&bdata=Jmxhbmc9Y3Mmc2l0ZT1l ZHMtbGl2ZQ%3d%3d#db=a9h&AN=60784078&anchor=GoToAllQVI>. [2012-02-01].

HAN, J. and KAMBER, M. 2006. *Data Mining: Concepts and Techniques*. 2nd ed. San Francisco: Morgen Kaufmann Publishers.

HAN, J., PEI, J. and YIN, Y. 2000. Mining frequent patterns without candidate generation. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 29(2): 1–12.

HAN, J., PEI, J., YIN, Y. et al. 2004. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8(1): 53–87.

HORÁKOVÁ, H. 2003: *Strategic marketing* [*Strategický marketing*]. 1. vyd. Praha: Grada Publishing.

INCOMA GFK, s. r. o. 2012. Incoma.cz Shopper typology 2010. [online]. Accessible at: <http://www.incoma.cz/cz/ols/reader.aspx?msg=1136&>. [2012-04-22].

KENG, K. A., EHRENBERG, A. S. C. 1984. Patterns of store choice. *Journal of Marketing Research*, 21(4): 399–409.

KOTLER, P. 2007: *Marketing management*. 1st ed. Praha: Grada.

KŘÍŽ, J. and DOSTÁL, P. 2010. Database System and Soft Computing. *Systémová integrace*, 17(4): 17–26.

KŘÍŽ, J., KLČOVÁ, H. and SODOMKA, P. 2011. The Use of Business Intelligence Tools for Prediction and Decisionmaking Processes in the Academic Environment: A Case Study. Innoviation and Knowledge Management – A Global Competitive Advantage. In: *Proceedings of The 16th International Business Information Management Association Conference*. IBIMA. Kuala Lumpur, Malaysia. International Business Information Management Association (IBIMA), 1259–1265.

KNÍŽEK, J., ŠINDELÁŘ, J., VOJTĚŠEK, B. et al. 2011. Using markers to aid decision making in diagnostics. *International Journal of Tomography and Statistics*, 16(11): 41–55.

LOŠŤÁKOVÁ, H. 2009. *Diferencované řízení vztahů se zákazníky* [*Differenciated customer relationship management*]. Praha: Grada Publishing.

LOUVIERE, J. J. and GEATH, G. J. 1987. Decomposing the determinants of retail facility choice using the method of hierarchical and international comparisons. *Journal of Marketing Research*, 63(1): 149–157.

MASON, J. B., DURAND, R. M. and TAYLOR, J. L. 1983. Retail patronage: A causal analysis of antecedent factor. In: DARDEN, W. and LUSCH, R. (eds.), *Patronage behavior and retail management*. pp. 339–352. New York: North-Holland.

MEDINA, S. and WARD, R. W. 1999. A model of Retail Outlet Selection for Beef. *International Food and Agribusiness Management Review*, 2(2): 195–219.

MILOSAVLJEVIC, M., KOCH, C. and RANGEL, A. 2011. Consumers can make decisions in as a little as a third of a second. *Judgment and Decision making*, 6: 520–530.

MONROE, K. B. and GUILTINAN, J. P. 1975. A path-analytic exploration of retail patronage influences. *Journal of Consumer Research*, 2(1): 19–28.

MORSCHETT, D., SWOBODA, B. and SCHRAMM-KLEIN, H. 2005. Perception of store attributes and overall attitude towards grocery retailers: The role of shopping motives. *The international review of Retail, Distribution and Consumer Research*, 15: 423–447.

NETOPIL, T. and KALÁBOVÁ, J. 2013. Marketing communication: loyalty to grocery stores built on flyers. *Journal of International Scientific Publications: Media and Mass Communication*, 2(2): 347–356.

NETOPIL, T. 2012. Consumer behaviour of customers on the food market in Czech Republic. [CD-ROM]. In: *PEFnet 2012*. 1–11.

UUSITALO, O. 2001. Consumer perception of grocery retail formats and brands. *International Journal of Retail and Distribution Management*, 29(5): 214–225.

SINHA, P. K. and BANERJEE, A. 2004. Store choice behavior in an evolving market. *International Journal of Retail and Distribution management*, 32(10): 482–494.

SPIGGLE, S. and SEWALL, M. A. 1987. A choice set model of retail selection. *Journal of Marketing*, 51(2): 97–111.

SUAREZ, A., DEL BOSQUE, I. R., RODRIGUEZ-POO, J. M. et al. 2004. Accounting for heterogeneity in shopping centre choice models. *Journal of Retailing and Consumer Services*, 11(2): 119–129.

TURČÍNKOVÁ, J., TURČÍNEK, P., MOTYČKA, A. et al. 2014. Exploring How Customers Shop for Meat Products. In: *Recent Advances in Economics, Management and Marketing*. 1st ed. Cambridge, MA, USA: WSEAS Press.

WEKA, 2013: *Weka 3 – Data Mining with Open Source Machine Learning Software in Java*. [online]. HTML Document. 2013. Accessible at: http://www.cs.waikato.ac.nz/ml/weka/. [2013-11-28].

WOODSIDE, A. G. and TRAPPEY, R. J. 1992. Finding out why customers shop your store and buy your brand. *Journal of Advertising Research*, 32(6): 59–78.

ZAKI, M. J. 2000. Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, 12(3): 372–390.

ZAMAZALOVÁ, M. 2009. *Marketing of a corporate business* [*Marketing obchodní firmy*]. 1. vyd. Praha: Grada Publishing.

Contact information

Pavel Turčínek: pavel.turcinek@mendelu.cz
Jana Turčínková: jana.turcinkova@mendelu.cz