Karolina Bartos

Association rules in the study of consumer behaviour

Ekonomiczne Problemy Usług nr 105, 279-286

2013

ZESZYTY NAUKOWE UNIWERSYTETU SZCZECIŃSKIEGO
NR 763 EKONOMICZNE PROBLEMY USŁUG NR 105 2013

KAROLINA BARTOS
Uniwersytet Ekonomiczny we Wrocławiu

ASSOCIATION RULES IN THE STUDY OF CONSUMER BEHAVIOUR

Introduction

Consumer behaviour became the subject of research in the 50s and 60s of the 20th century[1]. The growth of household income, which already was sufficient enough for more than just basic needs, contributed to this. Consumers freely disposed of it and could afford to buy much more goods than before. Moreover, the choice between different products of the same kind appeared in the market. Buyers began to demand more about the quality of products and their prices. This led to an increased competition between producers and sellers for customers and to considerations about what influences the decision to buy a particular good[2]. As a further result, more and more effective methods and tools for the analysis of this problem have been developed.

Prerequisite for a study of consumer behaviour is to obtain relevant statistics about it. Data from store receipts are a rich source of information about customers' buying habits. They are increasingly used by retailers for the analysis with association rules, allowing discovery of patterns about buyers' behaviour while making purchases. Knowledge of these patterns is extremely valuable. It allows a better planning of activities aiming at increasing the sales.

---

[1] G. Antonides, W.F. van Raaij: *Zachowanie konsumenta*, Wydawnictwo Naukowe PWN, Warszawa 2003, p. 583.

[2] G. Katona: *The Powerful Consumer*, McGraw-Hill, New York 1960.

1. Analysis of association rules as a data mining method

Methods of data mining, because of their purpose and the types of patterns they discover, can broadly be divided into the following classes[3]:

- classification / regression,
- clustering,
- characteristics exploration,
- discovering sequence,
- timing analysis,
- discovering association rules,
- detect changes and deviations,
- web exploration,
- exploration of text.

Discovering associations belongs to the broadest class of methods, which deals with the discovery of dependencies or correlations of interest (known generally as associations) in large data sets. It is used in many fields, including: medicine[4], education[5] and fraud detection[6]. However, the most common example of its application is called "market basket analysis", where the association rule is generated based on the data from market baskets (transactions). The purpose of this analysis is to find the natural patterns of consumer buying behaviour through the analysis of the products that are most commonly purchased by them.

Association rules take the form: "If *predecessor* then *successor*". Basket analysis result is presented as a set of association rules in the form of the relationship, represented by the following formula:

$$\{(A_1 = 1) \wedge \ldots \wedge (A_k = 1)\} \rightarrow \{(B_1 = 1) \wedge \ldots \wedge (B_k = 1)\}$$

This means that if a customer buys the products A_1, A_2, etc. up to A_k, he or she is likely to also buy the products B_1, B_2, etc. up to B_k. For example, a rule might indicate: "If a consumer buys cheese, sausage and ham he will probably also buy butter, tomatoes and bread."

With each association rule two fundamental measures of characterizing the statistical validity and strength are related:

---

3 T. Morzy: *Studia informatyczne*, http://wazniak.mimuw.edu.pl [21.12.2012].

4 P. Laxminarayan, S.A. Alvarez, C. Ruiz: *Mining statistically significant associations for exploratory analysis of human sleep data*, IEEE Transactions on Information Technology in Biomedicine 2006, Vol. 10, Iss. 3, p. 440-450.

5 D. Radosav, E. Brtka, V. Brtka: *Association Rules from Empirical Data in the Domain of Education*, "International Journal of Computers Communications & Control" 2012, Vol. 7, Iss. 5, 933-944.

6 D. Sanchez, M.A. Vila, L. Cerda: *Association rules applied to credit card fraud detection,* "Expert Systems with Applications" 2009, Vol. 36, Iss. 2, 3630-3640.

- Support,
- Confidence.

The support for an association rule A → B is the percentage of transactions that contain A and B[7]:

$$support = P(A \cap B) = \frac{sum\ of\ transactions\ containing\ A\ and\ B}{sum\ of\ all\ transactions}$$

Confidence for a rule A → B is a metric of the accuracy of the rule, determining what percentage of transactions which contain A also contain B[8]:

$$confidence = \frac{P(A \cap B)}{P(A)} = \frac{sum\ of\ transactions\ containing\ A\ and\ B}{sum\ of\ \ transactions\ containing\ A}$$

Analysts sometimes also use two other metrics, the *correlation* and the *lift,* for characterization. The correlation indicates how the fact that the client has chosen product A increases (positive correlation) or decrease (negative correlation) the probability that he or she will choose the product B as well. The lift is a modification of the correlation. Thereof it also determines the impact of selling product A on the probability of selling the product B too.

Discovering association rules generally takes place in two stages:

1. Find all common sets of events (frequency ≥ ϕ);
2. Based on the collection found, create association rules satisfying the minimum conditions for the level of support and confidence.

The algorithm, most common in use to find rules is *A priori*. But there are other methods, such as generalized rule induction method (GRI) and artificial neural networks, which become more and more popular[9, 10].

2. Application of *A priori* algorithm

Table 1 presents 10 sample transactions (market baskets) featuring 8 types of products purchased by customers in a grocery store.

---

[7] D.T. Larose: *Odkrywanie wiedzy z danych – wprowadzenie do eksploracji danych*, Wydawnictwo Naukowe PWN, Warszawa 2006, p. 188.

[8] *Ibidem*, p. 188-189.

[9] K. Migdał-Najman: *Zastosowanie samouczącej się sieci neuronowej typu SOM w analizie koszykowej*, in: *Taksonomia* z. 17, K. Jajuga, M. Walesiak (eds), Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 107, Wrocław 2010, p. 305-315.

[10] K. Migdał-Najman: *Analiza porównawcza samouczących się sieci neuronowych typu SOM i GNG w poszukiwaniu reguł asocjacyjnych*, in: *Taksonomia* z. 18, K. Jajuga, M. Walesiak (eds), Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 176, Wrocław 2011, p. 272-281.

Table 1

Articles purchased in 10 sample transactions

No. transaction	Dairy	Bread	Vegetables	Fruits	Sweets	Alcohol	Non-alcoholic Beverages	Meat
1	1	1	0	0	1	0	1	1
2	1	0	0	0	0	0	1	0
3	0	1	1	1	0	0	0	1
4	0	1	0	0	0	0	0	1
5	1	1	0	0	0	0	1	1
6	0	0	0	0	1	1	0	0
7	0	0	1	0	0	0	0	0
8	1	1	0	0	0	0	0	1
9	0	0	1	1	0	0	1	0
10	1	1	0	1	0	0	0	0
Sum	5	6	3	3	2	1	4	5

Source: own elaboration.

The first step of the analysis will be finding common item sets. The frequency (ϕ) was established at 3. To single-element sets belong products that are bought at least three times. These are: F_1 = {dairy, bread, vegetables, fruits, non-alcoholic beverages, meat}. Sweets and alcohol have occurred less than 3 times, so they cannot be in a common item set. The property of *A priori* algorithm says: if a set of events is not common, the addition of any article for this set will not cause that it will become common[11]. Therefore, the construction of common two-element sets considers only items that are an element of F_1. Based on the data from table 1 you can see that there are three two-element common sets: F_2 = {{dairy, bread}, {bread, meat}, {dairy, beverages}}, as these products were bought together at least three times. You can also find one three-element common set F_3 = {dairy, bread, meat}}, because three times (in transactions 1, 5 and 8) they were purchased together.

The next step is to find common item sets based on association rules which meet the condition of the specified minimum support and confidence level. The minimum level of support in this example was established at 40% and the confidence level at 70%. Table 2 shows all possible association rules for two-element sets.

[11] D.T. Larose: *Odkrywanie wiedzy z danych – wprowadzenie do eksploracji danych*, Wydawnictwo Naukowe PWN, Warszawa 2006, p. 189.

Table 2

Possible association rules for two-element sets

If *predecessor*, then *successor*	**Support**	**Confidence**
If dairy, then bread	**4/10 = 40%**	**4/5 = 80%**
If bread, then dairy	4/10 = 40%	4/6 = 67%
If bread, then meat	4/10 = 40%	4/6 = 67%
If meat, then bread	**4/10 = 40%**	**4/5 = 80%**
If dairy, then beverages	3/10 = 30%	3/5 = 60%
If beverages, then dairy	3/10 = 30%	3/4 = 75%

Source: own elaboration.

Only two rules satisfy the condition of minimum support and confidence: "If a customer bought a dairy, in 80% of all cases he or she also bought bread", "If a customer bought meat, in 80% of all cases he or she also bought bread". The rules for a three-element set were not selected because they did not meet the specified level of support (was purchased together only three times → support of only 30%).

3. Application of the association rules analysis in the study of consumer behaviour

Knowledge of association rules of customer buying patterns allows a more efficient use of activities aiming at increasing sales. For example, products that are often purchased together, in one basket, you may arrange close to each other to increase their total sales. You can also, on the contrary, place them far apart to force the customers to go around many shelves in the expectation that they will be tempted to buy other products. Knowledge of patterns of consumers shopping behaviour is usually used for:

- Arrangements of store shelves and products,
- Design and conception of advertising folders,
- Organization of promotional campaigns.

Sometimes, dishonest sellers use deception in their promotional campaign. Knowing that the products A and B are purchased together they promote a price reduction of A while in the same time increasing the price of B. In this way, they do not reduce their potential profits.

Analysis of association rules has a much wider application in the study of consumer behaviour than simple exploration baskets of supermarket customers. Instead of products in the rule we can analyze consumer demographics, such as: age, education, income, expenses, and his or her preferences. For example, you can carefully study the behaviour and habits of the internet users by combining the identification number with their visited pages and placed orders[12].

Another interesting example is the identification of the consumer behaviour of different groups of households in Poland[13]. In this study, the rules expressing the relationship between demographic and social characteristics of households and their expenditure on selected services, food and non-food goods have been discovered. In the article of A. Pasztyła[14] association rules analysis was used to search for patterns of behaviour that are likely to characterize the customers of a bank. She determined e.g. what features customers using credit cards usually have.

The use of the analysis of association rules in the field of services increases more and more. It is used, among others for[15]:

- Analysis of services for cross-marketing application,
- Optimization of service packages, tariffs and charges offered,
- Prevention of customer resignation from their chosen services,
- Planning of loyalty programs.

Moreover, managers in the area of services, as well as hypermarket sellers, use the basket analysis for planning promotional campaigns and also for the design and conception of advertising folders.

Conclusions

Discovering association rules from the collected data about customers can generate very useful information that can not be seen without a proper analysis. With them you can better understand the consumer behaviour, preferences and habits. This allows better adjustment to the customer's needs by optimizing the packages of offered services and fees as well as it helps in planning the most beneficial placement of products on store shelves. Moreover the discovered regularities provide the necessary knowledge to maintain customer loyalty to a brand and the use of cross marketing tools.

[12] R. Kita: *Analiza sposobu poruszania się użytkowników po portalu internetowym, Data minig – metody i przykłady*, StatSoft Polska 2002, www.statsoft.pl [21.12.2012].

[13] I. Kurzawa, F. Wysocki: *Wykorzystanie analizy koszykowej do identyfikacji zachowań konsumpcyjnych gospodarstw domowych w Polsce*, w: *Taksonomia* z. 15, K. Jajuga, M. Walesiak (eds), Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 7, Wrocław 2008, p. 527-534.

[14] A. Pasztyła: *Przykład badania wzorców zachowań klientów za pomocą analizy koszykowej*, www.statsoft.pl [21.12.2012].

[15] A. Pasztyła: *Analiza koszykowa danych transakcyjnych – cele i metody*, "Magazyn Systemy IT", p. 51, www.statsoft.pl [21.12.2012].

Literature

1. Antonides G., W.F. van Raaij: *Zachowanie konsumenta*, Wydawnictwo Naukowe PWN, Warszawa 2003.
2. Chen Y.L., Tang K., Shen R.J., Hu Y.H.: *Market basket analysis in a multiple store environment*, Decision Support System 2005, Vol. 40, Iss. 2.
3. Katona G.: *The Powerful Consumer*, McGraw-Hill, New York 1960.
4. Kita R.: *Analiza sposobu poruszania się użytkowników po portalu internetowym, Data minig – metody i przykłady*, StatSoft Polska 2002, www.statsoft.pl
5. Kurzawa I., Wysocki F.: *Wykorzystanie analizy koszykowej do identyfikacji zachowań konsumpcyjnych gospodarstw domowych w Polsce*, w: *Taksonomia* z. 15, K. Jajuga, M. Walesiak (eds), Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 7, Wrocław 2008.
6. Larose D.T.: *Odkrywanie wiedzy z danych – wprowadzenie do eksploracji danych*, Wydawnictwo Naukowe PWN, Warszawa 2006.
7. Laxminarayan P., Alvarez S.A., Ruiz C.: *Mining statistically significant associations for exploratory analysis of human sleep data*, IEEE Transactions on Information Technology in Biomedicine 2006, Vol. 10, Iss. 3.
8. Migdał-Najman K.: *Analiza porównawcza samouczących się sieci neuronowych typu SOM i GNG w poszukiwaniu reguł asocjacyjnych*, w: *Taksonomia* z. 18, K. Jajuga, M. Walesiak (eds), Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 176, Wrocław 2011.
9. Migdał-Najman K. , *Zastosowanie samouczącej się sieci neuronowej typu SOM w analizie koszykowej*, w: *Taksonomia* z. 17, K. Jajuga, M. Walesiak (eds), Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 107, Wrocław 2010.
10. Morzy T.: *Studia informatyczne*, http://wazniak.mimuw.edu.pl
11. Pasztyła A.: *Analiza koszykowa danych transakcyjnych – cele i metody*, "Magazyn Systemy IT, www.statsoft.pl
12. Pasztyła A.: *Przykład badania wzorców zachowań klientów za pomocą analizy koszykowej*, www.statsoft.pl
13. Radosav D., Brtka E., Brtka V.: *Association Rules from Empirical Data in the Domain of Education*, "International Journal of Computers Communications & Control" 2012, Vol. 7, Iss. 5.
14. Sanchez D., Vila M.A., Cerda L.: *Association rules applied to credit card fraud detection*, "Expert Systems with Applications" 2009, Vol. 36, Iss. 2.

REGUŁY ASOCJACYJNE W BADANIU ZACHOWANIA KONSUMENTA

Streszczenie

Artykuł prezentuje jedną z metod eksploracji danych, jaką jest analiza reguł asocjacyjnych. Opisuje podstawowe miary charakteryzujące jej siłę: wsparcie i ufność, oraz skupia się na wykorzystywanym podczas analizy algorytmie *a priori*. Ponadto przedstawia zastosowanie opisanych reguł w badaniach zachowań konsumentów ze szczególnym uwzględnieniem usług i handlu.

Tłumaczenie Karolina Bartos