The International Scientific Conference INPROFORUM 2017, *November 9, 2017, České Budějovice*, 327-332, ISBN 978-80-7394-667-8.

Modelling E-Commerce Processes in the Presence of a Price Comparison Site

Ladislav Beránek, Radim Remeš

Abstract: *Simultaneously with the development of e-business, related business applications have been developed. Price comparison sites that allow comparison of prices in different online stores belong to the most successful. These sites are also known as price comparisons, shopbots, or internet buying agents. Online buyers use them to obtain information about the price or the relevant stores. They reduce buyer search costs and help them in their decision making by providing comparison of price of for example some product in various e-shops. Such information is rarely found in the context of physical retail purchases. Compare pricing pages were mainly surveyed in terms of the impact of price comparison pages on the price of products and services and the sensitivity of online shoppers to the price. In this paper, however, we want to look at online shopping processes with network analysis optics. In our model, we use weighted network-based inference as a basic, and we suggest further a model of relationships in the presence of price comparators sites. The goal of our model is to predict what e-shop a customer chooses when buying a particular product.*

Key words: E-commerce · Network-based Inference · Model · Price Comparison Site

JEL Classification: D85 · L81

1 Introduction

A price comparison website is one of successful e-commerce applications. They offer to users (e.g., Heureka.cz or Zbozi.cz) a possibility to find some e-shop that sells a particular product or service at a good price compared to other e-shops. These price comparison sites help users to search for some products, provide them with insight into prices, and influence consumers' perceptions of the risks associated with online shopping.

Simultaneously with the development of e-business, related business applications have been developed and many new approaches have been widely used as for example recommendation systems or other. Price comparisons that allow comparison of prices in different online stores are the most successful. These sites are also known as price comparisons, shopbots, or internet buying agents. Buyers online use them to obtain information about the price or the relevant stores. They reduce buyer search costs and help them make decisions by providing price comparison information that is rarely found in the context of physical retail purchases (Brynjolfsson & Smith, 2000). Compare pricing pages were mainly surveyed in terms of the impact of price comparison pages on the price of products and services and the sensitivity of online shoppers to the price (Takayuki, Tsutomu, 2013; Pathak, 2012). Degeratu et al. (2001) states that the existence of price comparison sites increases price competition and buyer price sensitivity.

We can look at e-commerce also as a network. Social and business networks are increasingly important areas of research in many disciplines (Borgatti & Halgin, 2011; Easley & Kleinberg, 2010; Lamanauskas et al., 2013). However, stable equilibrium and models have focused primarily on this, while their dynamics and productivity have gained limited attention on research. One of the main tasks is to better understand, anticipate and control their dynamics, including how they form, develop and shape their behaviors and performances (Rauch, 2010; Schweitzer, Fagiolo & Sornette, 2009). Enough progress has been made in e-commerce applications so far, and e-commerce plays a very important role in the economy. A large number of buyers and sellers cooperate with each other via web site transactions (Beranek & Nydl, 2013; Beranek, Tlusty & Remes). These interactions support development and shape complex e-commerce market structures. Getting a deep look at e-commerce research has a deep and lasting significance.

Our paper presents a model that would outline the basic processes of online shopping if the user uses one of the price comparators. The aim of our model is to predict customers' behavior and their preferences when using price comparison sites. This means customers' product selection and e-shop selection. The product is recommended based on a recommendation system (or mostly based on peers recommendation or other), and e-shop based on customer recommendations and a price of some product which are presented on the respective price comparison site. We use data

doc. Ing. Ladislav Beranek, CSc., University of South Bohemia in České Budějovice, Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, České Budějovice 370 05, e-mail: beranek@ef.jcu.cz

Mgr. Radim Remeš, University of South Bohemia in České Budějovice, Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, České Budějovice 370 05, e-mail: inrem@ef.jcu.cz

collected from performed experiment to verify our model. It follows from a comparison that the model has good predictive abilities.

The aim of our research is to develop a model for exploring online shopping processes with the existence of a comparison page. We look at the relationship between shoppers and e-shops as a tripartite network. Then we use belief function theory to express user preferences. The objectives of this study are therefore twofold:

1. Creating a basic theoretical model to explain how buyers use price comparison sites pages when purchasing.

2. Performing empirical research to investigate validity and explaining the strength of this model.

For this purpose, we use simulation data and compare results with data from other sources.

The article is structured as follows: Section 2 briefly reviews related work on recommended systems. Section 3 defines some definitions that we use throughout this article. Algorithmic details can be found in Chapter 4. Section 5 describes the experimental procedure and results, followed by further discussion in Chapter 6. Finally, conclusions are made in Section 7.

2 Methods

In this section, we present a network analysis framework for e-commerce market. Nodes and links between these nodes (edges) constitute the network. Nodes represent the elements in a complex system, and the edges represent the interaction between system elements. E-commerce market is a complex network, and its complexity reflects in the following areas:

- The system has a huge number of nodes, and its network structure is complex and presents a variety of different characteristics. There are the generation and disappearance of nodes or edges. The emergence and demise of elements that have the life-cycle characteristic is very common. The relationships among elements are also dynamic changed.
- Networks are no governance structures. There is not one dominant organization to control and organize the other members in the network. Rather, the network concept is a way to visualize and understand the way firms and organizations are interconnected directly and indirectly through relationships. Networks are not under the control of individual nodes.
- Networks are formed in a self-organizing way through the actions and interactions of actions and interactions of actors involved, as they occur over time. They are continually being made and remade (or not) through ongoing structuring and restructuring processes. The multiple interactions and feedback effects continually taking place in networks lead to a complexity that makes it very difficult to control and predict for any individual actor.
- Time plays a central role in explaining and understanding exchange. Business relations develop over time and they are path-dependent. Buyers and sellers actively take into account what has happened before and they form plans and have expectations of what is likely to happen in the future, both of which affect their decisions in the present. The state of the network subject (person or organization) changes over time.

From a mathematical point of view, network can be described by set N={V, E} composed with nodes set V={v_1, v_2, …, v_n} and edges set E={e_1, e_2, …, e_m}. Network analysis framework for e-commerce market is as the following Figure. 1. From the graph we can see that, there are three steps for analyzing on e-commerce market from a network point of view—definition of network, analysis of network topology and analysis of network environment.

In the following paragraph, we give a brief introduction to the basic notions of the Dempster-Shafer theory (frequently called theory of belief functions or theory of evidence) which we use for the customers' preference prediction.

Considering a finite set referred to as *the frame of discernment* Ω, a *basic belief assignment (BBA)* is a function
m: $2^{\Omega} \rightarrow [0,1]$ so that

$$\sum_{A \subseteq \Omega} m(A) = 1 , \tag{2}$$

where $m(\emptyset) = 0$, see (Shafer, 1976). The subsets of 2^{Ω} that are associated with non-zero values of *m* are known as *focal elements* and the union of the focal elements is called *the core*. The value of *m(A)* expresses the proportion of all relevant and available evidence that supports the claim that a particular element of Ω belongs to the set *A* but not to a particular subset of *A*. This value pertains only to the set *A* and makes no additional claims about any subsets of *A*. We denote this value also as a *degree of belief* (or *basic belief mass* - *BBM*).

Shafer further defined the concepts of *belief* and *plausibility* [24] as two measures over the subsets of Ω as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B), \qquad Pl(A) = \sum_{B \cap A = \phi} m(B) \tag{3}$$

A *bba* can also be viewed as determining a set of probability distributions P over Ω so that $Bel(A) \le P(A) \le Pl(A)$. It can be easily seen that these two measures are related to each other as $Pl(A) = 1 - Bel(\neg A)$. *Moreover, both of them are equivalent to m.* Thus, one needs to know only one of the three functions m, Bel, or Pl to derive the other two. Hence, we can speak about belief function using corresponding *bba*s in fact.

Dempster's rule of combination can be used for pooling evidence represented by two belief functions Bel_1 and Bel_2 over the same frame of discernment coming from independent sources of information. The Dempster's rule of combination for combining two belief functions Bel_1 and Bel_2 defined by (equivalent to) *bba*s m_1 and m_2 is defined as follows (the symbol ⊕ is used to denote this operation):

$$(m_1 \oplus m_2)(A) = \frac{1}{1-k} \sum_{B \cap C = A} m_1(B) \cdot m_2(C), \tag{4}$$

where

$$k = \sum_{B \cap C = \varnothing} m_1(B) \cdot m_2(C).$$

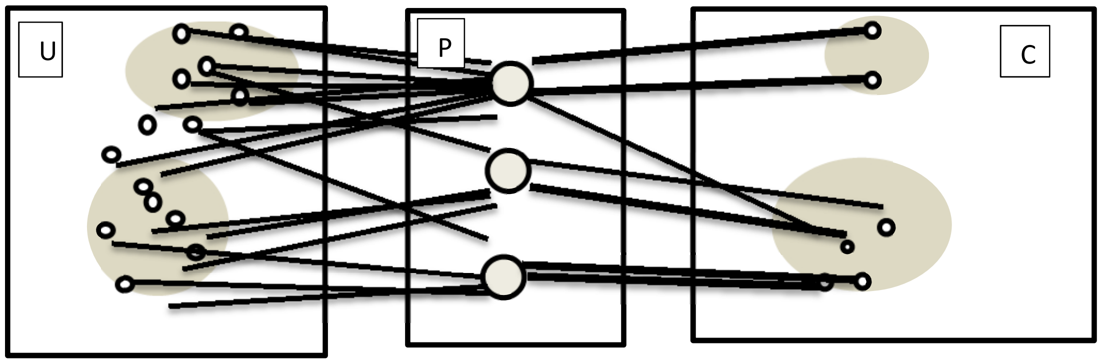
Here k is frequently considered a *conflict measure* between two belief functions m_1 and m_2 or a measure of conflict between m_1 and m_2 (Shafer, 1976). Unfortunately this interpretation of k is not correct, as it includes also internal conflict of individual belief functions m_1 and m_2 (Daniel, 2013). Demspter's rule is not defined when $k = 1$, i.e. when cores of m_1 and m_2 are disjoint. This rule is commutative and associative; as the rule serves for the cumulating of beliefs, it is not idempotent.
We will express belief functions for the price parameter and for the reputation of e-shop.

2.1 The design of our model

The problem description is as follows. The customer is looking for a specific product (such as a robotic vacuum cleaner) he has heard from acquaintances, from social networking friends, etc. A specific product can also be recommended to him/her by a recommendation system on a website or on a specific e-shop site. The customer gets a certain idea of the products he has at his disposal. He/she may also have a preference of a certain brand, for example iRobot when he/she buy robotic vacuum cleaner. He/she wants to buy this product at the lowest price. He therefore clicks on the website of a price comparison site (Heureka.cz). He/she selects chosen product and finds a range of e-shops selling the product. Offers of e-shops are sorted by price, lowest to highest. For each e-shop, ratings of previous customers are displayed. The customer chooses now e-shop according the price and rating of this e-shop and then buys the chosen product in this e-shop. We can model the before mentioned e-commerce processes with the existence price comparison sites as a tripartite network. In general, tripartite networks are the networks whose vertices are composed of three disjoint sets. Nodes in these tripartite networks are different in each layer (see Figure 1). Here, the set U are buyers, the set P are products and nodes in the set C present e-shops presented by price comparison site. However, in this paper, we will describe the process of e-shop selection, i.e., the relationship of U and C set, we suppose that user have chosen some concrete product. The reason we limit our model in this manner is to simplify the problems. Then, we can use bipartite network for the description of U and C relationship.

Suppose there are some categories of products (e.g., robotic vacuum cleaners that are manufactured by different companies and may have different functions). Let's denote the set of users $U = \{U_1, U_2, ..., U_m\}$, we suppose that these users have some referrals from friends, from recommendation systems from various information sources, or they prefer different brands of seeking item (robotic vacuum cleaner).
We further expect that this product be sold in various e-shops (the number of e-shops is m) at a price p_i. The price of delivery may also be important for users. It may vary depending on the price of the product. We include the price of delivery to the price of a product. Customers rate each e-shop after performed previous transactions with a certain number of points. This rating is available for other users to decide whether they will buy products in certain e-shop or not. We can express these rating (preferences of e-shops) using a matrix $R = (r_{ij})_{n \times m}$ where r_{ij} represents user' (j) preferences (expressed by a certain number of points, for example in range 1 - 5) given to this e-shop (j).

Figure 1 Example of a tripartite network as a model for e-commerce process in the presence of price comparison site



Source: authors

The total number of point each for each of m e-shops is:

$$d_j = \sum_{i=1}^{n} r_{ij} \tag{1}$$

We have two parameters influencing the shopping behavior: the price of product and the reputation of e-shop expressed by the number of points obtained by various customers after previous transaction. We will combine these two parameters with the help of Dempster-Shafer theory (Shafer, 1976).

Price parameter

This parameter shows how the shopping behavior of customers is influenced by the price. The respective belief functions have the following form:

$$m_p(\{buy_i\}) = \alpha \frac{p_{\min}}{p_i}$$

$$m_p(\{\neg buy_i\}) = 0 \tag{5}$$

$$m_p(\Theta_i) = 1 - \alpha \frac{p_{\min}}{p_i},$$

where α is the weight of this evidence. We can intuitively read this weight as a reliability of this evidence; p_i is the price in the *i-th* e-shop; $p_{\min}$ is the minimal price at which the product is sold in respective e-shops. With this equation, we have expressed the tendency to purchase the product based on the price. Usually, the lower the price the higher the temptation to buy the product in the *i-th* e-shop. At the same time, we assume that the equation is reflecting the temptation to buy the product, it does not show the reluctance to buy goods from the *i-th* e-shop, i.e. $m_p(\{\neg buy_i\}) = 0$.

Reputation parameter

This parameter shows how is the affection of customers to buy goods in the *i-th* e-shop is influenced by the reputation of respective e-shop. The belief functions have the following form:

$$m_R(\{buy_i\}) = \beta \frac{d_i}{d_{\max}}$$

$$m_R(\{\neg buy_i\}) = 0 \tag{6}$$

$$m_R(\Theta_i) = 1 - \beta \frac{d_i}{d_{\max}},$$

where d_i is the total number of point each for each of m e-shops, $d_{\max}$ is the maximum pints assigned to some e-shop, and β is the weight of this evidence. With this equation, we have expressed the tendency to buy the product in a particular e-shop based on the good reputation of this e-shop. Usually, the better the reputation of the e-shop expressed by the number of points assigned to this e-shop by customers according their satisfaction with previous transactions. At the same time, we assume that the equation is reflecting the tendency to buy the product in certain e-shop. It does not show some reluctance to the buying goods from the *i-th* e-shop, i.e. $m_R(\{\neg buy_i\}) = 0$.

The selection of e-shops

Once we obtain the belief functions, we combine them in a consistent manner to get a more complete assessment of what the whole group of signs indicates. The combination of belief functions is done with the help of the Dempster's combination rule, see equation (4). We express the assumption that customer will buy certain product chosen by him/her in previous stage belief functions $m(\{buy_i\})$ in some e-shop *i*. We calculate the value $m(\{buy_i\})$ using the combination (see equation 4) of single belief functions expressing appropriate evidence for every of *i* e-shop:

$$m(\{buy_i\})=(m_p \oplus m_R)(\{buy_i\}) \tag{7}$$

We obtained the vector of $m(\{buy_i\})$ values as recommendation for user *U* to by certain product in *m* e-shops. In order to evaluate the accuracy and efficiency of recommendation algorithm we defined an evaluation criterion as follows:

$$AC = \frac{1}{n}\sum_{i=1}^{n}\frac{N}{L}, \tag{8}$$

where *n* is the number of customers, *L* is the length of list *LE* of e-shops taking into consideration in experiments (less than the total number of e-shops available) and *N* is the number of correctly determined e-shops in the list *LE*.

3 Research results

3.1 Data set description

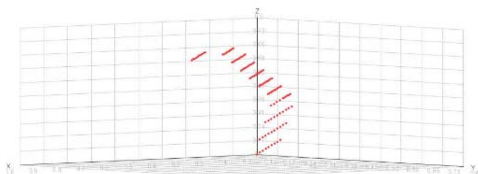
In this paper, we used data obtained by experiments with the group of 20 students. They were told to buy certain product with the use of price comparison site Heureka.cz for this purchase. They had to proceed like in real situation, to consider price and reputation of e-shops. Then, they had to written a list of e-shops in the order in which they would wanted to buy certain product from these e-shops.

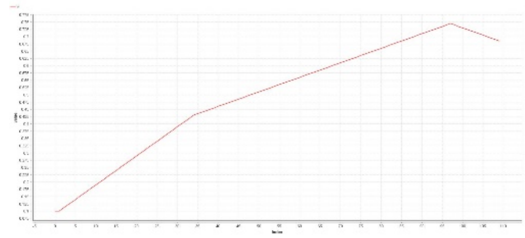
On Heureka.cz, users grade e-shops using integers from one to five, which one stands for "extremely don't like" and five stands for "extremely like." We randomly divided the data set into two parts: 80% is the training set, and 20% is the testing set. To ensure the experimental results are accurate, all experiments were carried out with a fivefold cross validation test.

3.2 Experiment results

Taking reality into consideration, we could also assume the length L of the list of e-shops to be 5. We used the training set and the testing set to measure *α* and *β* by varying them from 0.0 to 1.0 step by 0.1 to see the values of *AC*. The aim was to determine optimal values of *α* and *β*.

Figure 2 Value of *AC* criterion by varied *α* and *β*





Source: Own processing

We determined the best values *α* and *β* as following: *α* = 0.98, *β* = 0.89 (see Figure 2).

4 Conclusions

We have suggested a model of e-commerce processes in the presence of a price comparison site. Experimental results on the group of students show accuracies of the proposed method. The accuracy of our model is sensitive to both amount of data and the length of list of e-shops. Our model can be useful at prediction of shopping behavior of customers in the presence of a price comparison site. In this paper, we focused mainly on the last part of the modeling process, where we used the belief function theory. However, it is clear that the graph approach brings advantages over other methods as it also allows the expression of relationships between individual actors in the field of e-commerce. Therefore, we want to focus on the use of network methods in the presented model in our next work.

A significant challenge for inference with networks is the available information is only an approximation of people's relationships and preferences. For use of social networks in e-commerce, the network information could be incomplete and out of date, that is, noisy. Thus, in practice, evaluating the usefulness of network-based inference for e-commerce requires understanding the consequence of errors in the data. Fortunately, mechanisms relying on aggregated information from social networks are somewhat robust: performance degrades gradually rather than abruptly with noise. In such cases, estimates of consumer interests based on approximate network information is beneficial compared to not using the information at all. Evaluating the amount of noise in online networks and its effects on mechanisms relying on those networks is an important direction for future work.

A further challenge arises from the using real networks data about e-commerce. While available online networks can include thousands or millions of users, and thus give strong statistical correlations, detailed information on why users form links is usually lacking. Thus, it is difficult to distinguish links arising from prior similarity from influence of linked individuals creating similar preferences. In our future work, we want to explore further insight into the network of interest and make a more dynamic analysis of the network possible.

Acknowledgement

This study was supported by project grant No. IGS21C1, Faculty of Economics, University of South Bohemia Czech Republic.

References

Beranek, L., & Nydl, V. (2013). The Use of Belief Functions for the Detection of Internet Auction Fraud. In *Proceedings of the 31st International Conference Mathematical Methods in Economics 2013* (pp. 31-36). Jihlava: College of Polytechnics Jihlava.

Beranek, L., Tlusty, P., & Remes, R. (2010). The Usage of Belief Functions for an Online Auction Reputation Model. In *Proceedings of the 28th Conference on Mathematical Method in Economy 2010* (pp. 49-54). Economical faculty. University of South Bohemia, Ceské Budejovice, Czech Republic, September 2010.

Borgatti, S. P., & Halgin, D. S. (2011). Analyzing a affiliation networks. In Carrington, P. and Scott, J., eds. *The Sage Handbook of Social Network Analysis* (pp. 417-433). Thousand Oaks: Sage Publications.

Brynjolfsson, E. & Smith, M. D. (2000). Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science,* 46, 563-585.

Daniel, M. (2013) Properties of Plausibility Conflict of Belief Functions. In Rutkowski, L; Korytkowski, M; Scherer, R; et al. *12th International Conference on Artificial Intelligence and Soft Computing, ICAISC 2013* (pp. 235-246), Zakopane, POLAND, Jun 09-13, 2013, PIEEE Computat Intelligence Soc, Poland Chapter, Lecture Notes in Artificial Intelligence, Vol. 7894.

Degeratu, A.M., Rangaswamy, A., & Wu, J. (2001). Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of Research in Marketing*, 17(1), 55-78.

Easley, D., & Kleinberg, J. (2010). *Networks, Crowds, and Markets*. Cambridge: Cambridge University Press.

Lamanauskas, V., Slekiene, V., Balog, A., & Pribeanu, C. (2013). Exploring the Usefulness of Social Networking Websites: a Multidimensional Model. *Studies in Informatics and Control,* 22(2), 175-184.

Pathak, B. K. (2012) Comparison Shopping Agents and Online Price Dispersion: A Search Cost based Explanation. *Journal of Theoretical and Applied Electronic Commerce Research,* 7(1), 64-76.

Rauch, J. E. (2010). Does Network Theory Connect to the Rest of Us? A Review of Matthew O. Jackson's Social and Economic Networks. *Journal of Economic Literature*, 2010, 980-986.

Schweitzer, F., Fagiolo, G., & Sornette, D. (2009). Economic Networks: The New Challenges. *Science,* 325, 422.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

Takayuki, M. & Tsutomu, W. (2013). Why Are Product Prices in Online Markets Not Converging? *Plos One*. 8(8), 1-7.