Hetesi, E. – Kürtösi, Zs. (eds) 2011: The diversity of research at the Szeged Institute of Business Studies. JATEPress, Szeged, pp. 59–71.

The role of collaborative filtering in marketing

Zoltán Majó – Balázs Révész

In the past few years, there have been many lectures and articles about products that sell well via the Internet. Books, CD's, videos, computers and home electronic devices are among the items found on current bestseller lists and on the shelves of the largest online stores. There are tools available, as outlined in this paper, that can aid the efficient online marketing of these products and the sales of these can be supported using recommender systems. The collaborative filtering technique, also described in this research paper, has been employed by market leading US dotcom companies with great success in recent years. Because of said marketing success, the appearance of European recommendation systems can be anticipated very soon.

Keywords: collaborative filtering, recommendation systems, CRM, online marketing

1. Introduction

The theoretical foundations of recommender systems have been known for a long time in everyday business practice. An example of a common recommender system is "The Chef's Favourites" on a restaurant menu or an expert's article on services provided by car washes. It is customary to find recommendations on the covers of books while browsing through the shelves of a bookshop. In the literary works business, famous people (reviewers, analysts, journalists, editors, etc.) are often called upon to review a book, thereby helping or influencing customers in their choice. This recommender technique has been a generally applied marketing tool in the business world for a long time.

In addition to individual and expert recommendation techniques, community recommendation tools are also known in marketing practice (Nikolaeva–Sriram 2006). To be able to isolate and use these tools, a vast number of consumer preference data are quantified, systematized and then put into an easy-to-communicate form. Commonly speaking, this is the method of top lists or popularity ratings which are present in all channels of the media (films, books, sales and hit lists in music). These recommender techniques, either individual or community-based ones, are common tools of mass marketing. However, the question is how to formulate individual offers from individual preferences and whether it is possible to develop recommender techniques into recommender systems.

Perhaps one of the best known among the community sites using a web-based collaborative filtering process is MovieLens (www.movielens.org), a personal

movie recommender, which is a good illustration of how collaborative filtering works.

To be able to use this system, online registration is needed in which users have to rate films that they have previously seen (in this case a 5-grade scale is used). Then the system compares the ratings (i.e. a user's profile) with those of other users and tries to contrast the profiles. Basically, by using the movie ratings, the system looks for the closest neighbour who is most similar to that particular user's profile. Of course, for a comparison of profiles, a reliable scale is needed (Cosley et al 2003) in addition to a program that makes multivariable statistical calculations (correlation calculus, a cluster analysis with groups) (Goldberg et al 2001). In the case of several thousand users, significant computer resources are required to satisfy the needs of masses (Cho–Kim 2004), since it is necessary to calculate the coefficient of correlation in a large calculation capacity. In a collaborative filtering system, the data are compared in two dimensions. In this context, overlapping refers to the number of films that contrasting users evaluated, and correlation indicates how similarly those particular users rated those films that both had seen.

Ideally, if a particular system has a sufficient number of users, there will be user groups within which both correlation and overlapping are high. A high correlation value (~1) suggests that the same opinion is held. A high overlapping, on the other hand, means that same opinions are not incidental, considering that they are based on a large number of data.

2. Collaborative Filtering Solutions

Algorithms in collaborative filtering can be coupled with various IT solutions (McNee et al 2002). There are manual collaborative filtering systems, in which people make or ask for such recommendations, but the majority of commercial applications work as automated systems that collect, store and analyse customer preference data; they look for customers with a similar taste and recommend certain products using the obtained data.

Collaborative filtering systems depend on a single thing: customer preferences (Riedl et al 2002, Huang et al 2007). These customer preferences not only reflect the taste of individual customers but – indirectly – they also create the set of data needed to determine the closest neighbours. These data are transaction data gathered during shopping; besides basic sales data (like what is purchased or when or at what price), online sales are also concerned with how much time users spend on a webpage, what they view, what they print out, what they save or even how they rate certain products.

Once ratings and/or preferences for a particular consumer group have been obtained, collaborative filtering techniques can begin being used. Going back to the above example, the system can also give recommendations as to whether a given

film is worth being seen or not, but there also is choice for result in which recommendations can be obtained from the system without any intervention.

Taking these different approaches into consideration, John Riedl and Eric Vrooman identified three large groups of recommender systems, and based on their functions, they distinguished three different collaborative filtering techniques: pull-active CF, push-active CF and automated CF. Since the three systems generate different inputs, sometimes all three might be used at the same organisation or on the same website.

2.1. Pull-active CF

With pull-active collaborative filtering applications, the user is an active participant of the process in which the system creates recommendations (based on queries) according to his or her request. This recommender works by knowing other users' preferences or personal interests within a community, and while looking for a solution to a problem or task, a search is made for other people's suggestions and remarks.

The first, widely-known computerized collaborative filtering system was Tapestry. The primary goal of Tapestry, developed by Xerox PARC as a research project, was to set up workgroups to help find out which articles (mainly on electronic bulletin boards) are worth reading. Users of Tapestry could add remarks to an article and other users could make the system search for those texts that fulfilled certain criteria, e.g. key words of an article (using information retrieval and filtering), other people's remarks or what others did with a particular article.

2.2. Push-active CF

In typical business life, emails are often forwarded with nothing other than an abbreviation of FYI; to others that one might think would take an interest. Frequent internet users have undoubtedly been added to chain emails in which jokes are forwarded to friends and acquaintances that ideally have the same sense of humour as they do. PUSH-ACTIVE CF follows this principle, that with the help of a programme, users can simply recommend (push) the information to others if they find it interesting or useful. The first CF prototype of this kind was developed by David Maltz and Kate Ehrlich at Lotus Research.

2.3. Automated CF

The major difference between automated CF and PUSH and PULL active CF is that while the latter are applications requiring human intervention, automated CF gathers information on user preferences; it compares and analyses them and then sends them to users (Ahn 2006). A pioneer in automated CF technology was GroupLens.

3. Collaborative Filtering on the Internet

The Internet has opened up new possibilities for scientists, economists, politicians and any other users who can access the worldwide web. In this non-hierarchical network, more and more information is accumulated. Today, it is better not to type in the words "marketing conference" into the search field of a search engine because it might take more than a week to look through all the links (at the time this article was being written, Google showed about 5,040,000 links). Thus, there are two factors that must be coped with these days when using the Internet, the abundance and the reliability of information. Some years ago, the top management of a company was still able to read through a brochure compiled by a press-monitoring agency every day. Today it is a daunting task to try to process the contents on the Internet, no matter whether it is by keyword search or just product information. Even when time and energy are not at issue, there is still the problem of the quality or reliability of the information that is found on the web which is rather difficult, and sometimes even impossible, to control. For example, there is an industrial company employing several hundred people, on the website of which there has been a job announcement for a sales manager's position for several months. In this case, it is fairly difficult to decide whether they have not yet managed to fill this position or they have just forgotten to update their site. What can we do to solve this problem of information filtering and retrieval? Where can we find help?

Collaborative filtering techniques might be useful in order to use not only keywords, but also topics, quality, taste or fields of interest to determine the theme of a website. One of the major advantages in doing business on the Internet is that there is a two-way interaction between the visitor and the website owner during every visit. While browsing the information published by a company, visitors constantly give information about themselves (Vandermerwe 2000), though not for the general public, but for the owner of the site. Information is obtained when a visitor asks questions, places orders, fills in a registration form or a questionnaire, or writes to the company or a discussion group about his or her opinion of a product or a particular situation. Additional information is obtained by making statistical analyses of visits (e.g. analysing the log file) which primarily helps form a logical structure of the website and measure its popularity.

To sum up, automated collaborative filtering is based on data obtained from former interactions between an individual and the system. In their simplest form, automated collaborative filtering systems keep track of every item that a user rated, including the fact of how much he or she liked it. Subsequently, based on the similarities in consumer tastes, the system calculates which consumers could "predict" others' taste. Eventually, it tries to recommend new products using these predictors.

4. The Role of Collaborative Filtering in Marketing

With the help of group-based, information-filtering and recommender systems described in this article, the sale of particular products can be transferred to the Internet, not to mention the potential increase in sales volume through cross selling. Recommender systems can make sure (primarily if they are based on the technique of group-based information filtering) that visitors and customers not only consider buying products they originally planned to buy, but also have the option to shop for other - either closely or less closely related, or sometimes even unrelated - articles as well (Winoto–Tang 2008).

In the past few years, online sales and e-commerce have begun to develop very rapidly. Everybody interested knows that almost all fast moving consumer goods (FMCG) can be ordered on the Internet today. However, there are only a few business-to-consumer (B2C) solutions that care for a professional presentation of their products and services. Very often, photos are missing in web catalogues or the pictures of the articles are placed in a separate "picture gallery". At the present level of development, it is difficult to find an online shop where, apart from an "Add to shopping cart" function, there are additional features that help customers in their shopping (e.g. detailed product descriptions, branding, warranty conditions, etc.).

For online marketing to become more efficient, it is important to make experts aware that the inputs on a webpage coming via the Internet are not just a set of data consisting of the binary combination of 0 and 1. They are not even just statistics or log files that are only valuable to IT experts and systems administrators, but rather an imprint of visitors' and customers' behaviour (Mild–Reutterer 2003). There are humans sitting at the other end of the net who have individual preferences and expectations. The conscious information supplying and gathering behaviour of visitors, as well as their self-expressing conduct, are interesting to us because by knowing these, a selection of the target group and positioning takes place and is a precondition to successful marketing activity. Information on the target group such as implicit inputs obtained from visitors' and customers' behaviour (e.g. statistics, log files) and explicit inputs (customer ratings, recommendations) is indispensable to formulate offers and to use various tools of promotion.

Yet, it is not only mass media offers that the Internet makes possible. Individual visitors can be targeted with personalized recommendations if there is sufficient data on their personal characteristics, purchase history, preferences and records of the behaviour of users who are similar to them.

The first step that separated sales and marketing as a company activity was the appearance of mass marketing. With such production and sales volumes, the mass media became the primary tools of promotion by persuading customers with advertisements to purchase a product at a later point of time. Due to the media explosion that occurred at the end of the 20th century, the tools of mass marketing could become more sophisticated as well as the economic cost/benefit principle

applied in marketing. Target group selection and segmentation according to demographic features (income level, age, sex, nationality, religion, place of residence, educational level) has become localized. Now an advertisement can be made for young women and it can appear in a magazine that they read, and a separate message can be sent to highly qualified men or large families can be reached through a film channel for the whole family. Segmentation has reached mailing lists on the Internet as well; different messages or offers can be sent to smaller groups. Instead of an initially generic message, people get messages that are suited to them better and closer to their actual preferences.

However, marketing based on demographical data has its limits. In reality, people cannot be clearly assigned to such simple categories. Almost simultaneously with the advance of technology, two things happened. Customer relationship management (CRM) software and computer databases made it possible for the tools of one-to-one marketing to evolve (Allen et al 2001, Kandikó 2009). This kind of marketing model was first presented in Peppers and Rogers' The One to One Future in 1993. In this book, they tried to treat customers individually by tracking and remembering their preferences and then, based on purchase transactions (see mass customisation), a company's range of products and its special offers, new recommendations can be made (Peppers–Rogers 1993, Peppers et al 1999). At the same time, in the second half of the 90s, the worldwide web and the advance of printing technology brought about cheaper delivery mechanisms. Unlike shops, the Internet is now able to present every user with a personalised interface and offer them customized products. Being that writing software and information technology itself involve relatively high fixed and low variable costs, it means that a new visitor to a website entails practically no extra costs. Using effective custom printing, every customer can be sent a more or less individualized catalogue, newsletter, voucher booklet or offer. Today's level of technology has been enabled to retrieve customer data (data mining, CRM, business intelligence) and to open up a customized channel of communication to users through the Internet (Newell 2000), so the only need is to know what specifically to recommend.

One-to-one marketing relies on data about individual customers found in databases and on human processing of product information. In order to define what offerings or products should show up to a customer, especially if there is a broad clientele, additional resources are required.

This is where automated recommender systems prove really useful as they can help attain the objectives of one-to-one marketing with their precisely customized methods (Kandikó 2009). Using recommender systems, a customized marketing campaign can be organized for even a customer base consisting of several hundred thousand people; no matter whether it is for web-based sales, cross selling by phone, an e-mail or a mailing campaign.

5. Inputs and Outputs in Collaborative Filtering

In their book, John Riedl and Eric Vrooman distinguished three input types (explicit, implicit and community) and four output types (suggestion, prediction, rating and review) of collaborative filtering. The model-like use of this classification and business intelligence systems can be a basis for a marketing management process.

5.1 Inputs types

Explicit and implicit

In this context, an input means that customers express their preferences. These inputs can be either explicit (inputs that require buyers' active co-operation) or implicit (inputs received from the natural behaviour of a customer as observed on a webpage). The most common explicit inputs are ratings, i.e. quantified or symbolic evaluations of a product; keywords/attributes, which refer to the expressed personal interest of a customer; and profiles i.e. personal data supplied by customers during registration. The most frequent inputs are purchase history and browsing data. From purchase history we learn which products a customer found valuable, whereas browsing data (including products and information viewed as well as goods put into the shopping cart) help identify the current interests of a visitor.

Community

The other inputs reflect the community. These include purchase history, the average and statistical indices of individual ratings, as well as reviews written by others. Products are often classified based on sales lists and ratings that are characteristic of the entire community (e.g. films or books according to genre). Popularity indices, such as ticket sales or bestseller lists, help customers see what a community appreciates.

5.2 Output types

Suggestion

The simplest output type is suggestion which involves mentioning or presenting a product without stressing that it actually is a sort of recommendation.

Prediction

In addition to simple suggestions, certain systems can predict in a quantitative or symbolic manner how much a customer will like a product (e.g. the movie recommender above).

Rating and review

Some systems even allow customers to view the ratings and reviews of other shoppers thus community inputs can turn into evaluations and reviews. This is especially common in such cases when several products have to be rated.

Amazon.com, for instance, encourages its customers to rate books and write reviews (explicit input). Then this information is made available to other customers as well. EBay asks sellers and buyers to critically evaluate each other (explicit input). Finally, these evaluations and reviews are summed up for those who wish to do business.

6. The Spread of Collaborative Filtering in the US and Europe

Among the flagships of online electronic retail, the American companies, eBay.com and Amazon.com, have used recommender techniques and algorithms of collaborative filtering for years. Both content providers have now become global players on the Internet, owning not just national and multilingual homepages – thereby making their services distinct – but have professionalized one of the crucial and fundamental ideas of collaborative filtering, the involvement of communities (turn communities into content) as well. We do not find such a 'blockbuster' in European business practice, although a number of companies and organisations have started employing collaborative filtering techniques (GUS, Deutsche Telecom, etc.).

Studying business practice in addition to the scientific and research-specific aspects of collaborative filtering on the Internet, conferences and studies dealing with this have been found as early as at the end of the 90s. From the point of view of recommender systems, this period is of historical importance since the first public academic application (GroupLens), also found in the book of the two American authors, and was officially launched in 1996 while the first business solution (GUS) was born in 1999. Simultaneously with US research at Berkeley and the University of Minnesota, development programmes were started in the European Union as well as financed by the 4th Development Framework Programme for Research and Technology of the European Union. In the form of international co-operation, the framework programme financed the development of complex, web-based software systems called Web4Groups which aided the work of teams. As a result of this, an experimental online voting system and a programme for evaluating online documents called SELECT were created.

The programme called SELECT, an article recommender system, developed that was similar to GroupLens. While with the development of GroupLens, recommending news and articles was in focus, the development of SELECT concentrated on evaluating webpages. At the beginning, GroupLens users could follow the ratings given by readers of current scientific articles. After a particular user had evaluated several articles, GroupLens was able to recommend him or her articles that he or she most likely would be interested in. Instead of scientific articles, SELECT, which is still available on the web, specialised in documents published on homepages and websites. Whenever a new site is visited, its contents can be rated on a 5 grade scale; other people's ratings can be looked at and remarks

can be stored on a given web-based document. As a matter of fact, this system is the core of an online recommender system.

The method of GroupLens was used very soon on non-academic sites in the US like MovieLens, which became popular in the States and was followed by several sites, e.g. Zagat's restaurant guide. According to our research, no such direct business relationships can be derived from Select.

6.1. Online recommender systems in the consumer electronics branch

Examining the spread of the commercial use of online recommender systems, it seems an obvious solution to look for examples among the companies trading in consumer electronics products. Investigations have proved that the areas of recommender systems and collaborative filtering are rather varied. Through research, US and European sites are compared in e-commerce carrying similar product lines and analysed to the extent to which they use collaborative filtering. On these sites, it was primarily examined what input (implicit, explicit and community) and output (suggestion, prediction, ratings and reviews) items appear on particular pages. The subject of these examinations was leading companies with an extensive network.

One of the best known e-commerce sites of the world is www.ebay.com, and www.bestbuy.com is one of America's click-and-mortar store chains. With its chain stores (in the Czech Republic, Denmark, Finland, France, Greece, Hungary, Great Britain, Norway, Italy, Spain and Sweden), www.dixons.co.uk is the site of Europe's leading consumer electronics retailer; www.nitro-shopping.uk operates (in France, England, Germany, Italy, Norway, Spain and Sweden) as a dotcom company and www.fotexnet.hu is Hungary's leading e-commerce site.

While analysing the availability of recommender systems, we especially focused on the following aspects:

- Are there any special offers displayed to visitors on the entry page of the online store that function as recommenders for those customers who look for immediate discounts?
- Is there a New Products section on the main page serving as a recommender for those buyers who look for novelties?
- Is there a TOP section on the homepage (e.g. TOP electronics, TOP DVD's, TOP music) serving as the basis for collaborative filtering?
- Is there a rating scale for quantifying customer opinions?
- In addition to popularity indices, is there a section (e.g. "voice your opinion") for customers of certain products where they can write their opinion of these products?
- Do we find any services on the site that visitors can use to notify their friends about a product by e-mail so that visitors can become involved in selling the product?

– Are there any links to other brands or further offers which could help visitors in their choice or to compare products? Does the system support cross selling?

If we analyse homepages in terms of recommender systems, the results seem rather varied. All dotcom companies in our survey use recommender systems but in the case of click-and-mortar companies, functions suggesting the use of recommender systems are minimal. The following table shows the availability of those elements that are important from the point of view of recommender systems.

Table 1. Inputs and Outputs of collaborative filtering at popular online stores

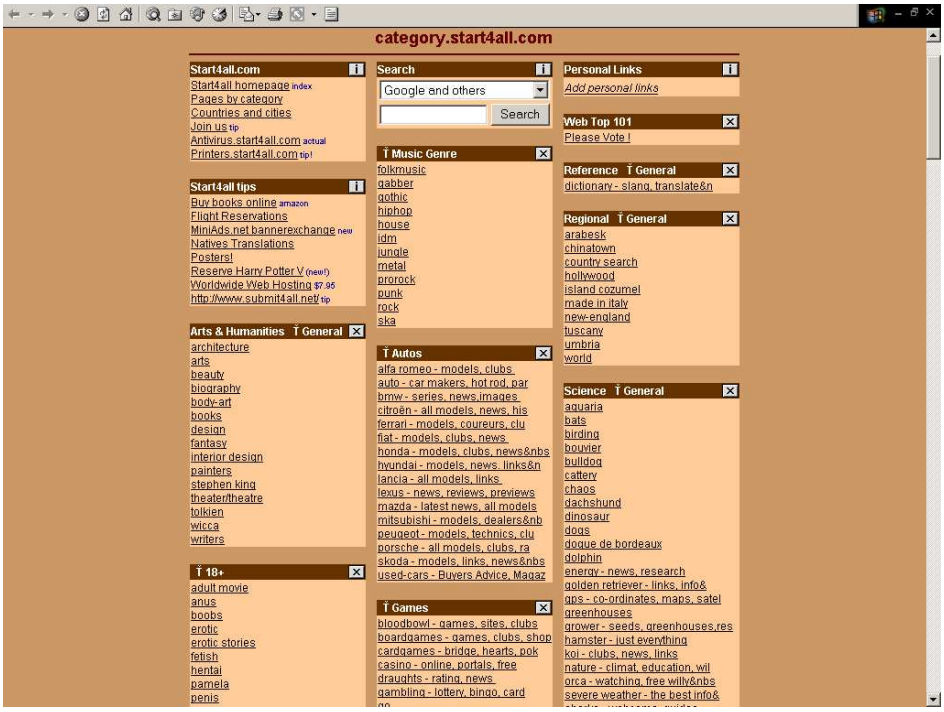
Store	Bestbuy (USA)	Dixons (EU)	eBay (USA)	Nitro shopping (EU)	Fotexnet (Hu)
Inputs					
Implicit*	Registration, history of orders	Registration, history of orders	Registration, history of orders	Registration, history of orders	Registration, history of orders
Explicit	None	None	Positive and negative and neutral rating, short comments	Quick rate 1 to 10, consumer reviews	Quick rate 1 to 5, consumer comments
Community*	No data	No data	Feedback score	Compare prices	Number of visits
Outputs					
Suggestion	Hot offers, accessories of products, e-mail to friends	Top sellers, deal of the day, deal of the week, e-mail to friends	Featured items, spotlights on, e-mail to friends	Most popular, e-mail to friends	Best-selling articles of the past 7 days, popular products
Prediction	None	None	Most-wanted-item notifications	Estate agent	Average popularity index
Evaluations and reviews	None	None	Seller and buyer information, feedback information	Rating, reviews	Customer opinions

Note: *Conclusions could only be reach on applied implicit and community inputs from observed outputs.
Source: own construction

7. Power of Communities – Added Value in Collaborative Filtering

As already described in this paper, collaborative filtering involves processing the information (interactions, ratings and reviews) received from users, finding the closest neighbours and generating recommendations from the above. Subsequently, these user interactions might even develop into independent contents and the Internet users themselves may become "algorithms" of collaborative filtering. In this way the basis for a new business opportunity is created which is illustrated by the initiative of a Dutch site called start4all.com.

Figure 1. Start4all.com – online knowledge base and recommendation system



Source: http://category.start4all.com
Downloaded: 2 Dec. 2007.

If there is desire to buy a second hand car and the words "second hand car" or "car dealer" are typed into Google, the search engine shows several hundred thousand hits. In such a case, it seems more hopeful to go and visit the dealers than to visit their websites. This is where digital catalogues can help in the retrieving and filtering of information. Therefore, obviously these catalogues have rapidly gained lots of users. Although these start pages do not have independent contents, they have become major starting pages in a number of countries worldwide.

The system itself is made up of simple boxes: the start page usually consists of a single page containing several boxes. Each box contains 5 to 15 links in a given topic. From the News box you get to the most important news pages while in the Telephone Directory box you find online phone book editors. In the Timetables box both local and international bus, coach and railway timetables are available. Each and every topic is located in a separate box.

Today start4all.com is like a global online editorial office with start pages coming to life in more and more countries of the world. In Europe, 23 countries have their own start page. Among them, Hungary's page has more than a thousand topics operated by a thousand "amateur" page editors. Some of them are fanatics (u2.start4all.com), and many consider this their hobby (modelcar.start4all.com) while others hope to earn some money in this way (business.start4all.com) or just want to tell the world about their job (antivirus.start4all.com). Although motivating factors are different, one thing is for certain; they intend to tell the world about their community or field of interests and show their own contribution to the worldwide web. By now, start4all.com's system has grown into a full-fledged online knowledge base with an editorial staff, editing principles and sales policy.

8. Summary

Using recommender techniques in online marketing can only bring success in selling top books, CD's, videos and home electronics goods online. This marketing technique could be a successful model for every web-based store including such products as wine, chocolate or clothing. Recommender techniques make it possible for click-and-mortar companies to formulate community recommendations (top lists such as Wine of the Week) using the purchase data at their disposal, and after analysing user profiles, they can design personal offers (using the method of closest neighbours) for their customers.

Collaborative filtering techniques do have their own problems. Quantifying consumer preferences, designing suitable algorithms and privacy concerns raise a number of questions where marketing still needs to find answers.

References:

Ahn, H. 2006: Utilizing popularity characteristics for product recommendation. *International Journal of Electronic Commerce,* Vol. 11., No. 2., pp. 59–80.

Allen, C. – Kania, D. – Yaeckel, B. 2001: *One-to-One Web Marketing*. New York, John Wiley & Sons, Inc.

Cho, Y. H. – Kim, Y. K. 2004: Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce. *Expert Systems with Applications,* Vol. 26., Issue 2., pp. 233–246.

Cosley, D. – Lam, S. K. – Albert, I. – Konstan, J. A. – Riedl, J. 2003: *Is seeing believing? How recommender interfaces affect users' opinions*. CHI 2003, April 5–10., Lauderdale, Florida.

Goldberg, K. – Roeder, T. – Gupta, D. – Perkins, C. 2001: Eigentaste: A constant time collaborative filternig algorithm. *Information Retrieval Journal*, Vol. 4., No. 2., pp. 133–151.

Huang, Z. – Zeng, D. D. – Chen, H. 2007: Analysing consumer – product graphs: empirical findings and applications in recommender systems. *Management Science*, Vol. 53., No. 7., pp. 1146–1164.

Kandikó, J. 2009: Kódolt vásárló – behálózott fogyasztói. *Magyar Tudomány*, Vol. 170., No. 6., pp. 691–696.

McNee, S. M. – Albert, I. – Cosley, D. – Gopalkrishnan, P. – Lam, S. K. – Rashid, A. M. – Konstan, J. A. – Riedl, J. 2002: *On the recommending of citations for research paper*. CSCW'02, November 16–20, New Orleans, Louisiana, USA.

Mild, A. – Reutterer, T. 2003: An improved collaborative filternig approach for predicting cross-category purcheses based on binary market basket data. *Journal of Retailing and Consumer Services*, Vol. 10., Issue 3., pp. 123–133.

Newell, F. 2000: *Loyalty.com, Customer Relationship Management in the New Era of Internet Marketing*. McGraw-Hill, USA.

Nikolaeva, R. – Sriram, S. 2006: The moderating role of consumer and product characteristics on the value of customized on-line recommendations. *International Journal of Electronic Commerce*, Vol. 11., No. 2., pp. 101–123.

Peppers, D. – Rogers, M. 1993: *The one to one future*. Doubleday, New York.

Peppers, D. – Rogers, M. – Dorf, B. 1999: Is your company ready for one-to-one marketing? *Harvard Business Review*, 77/1999, pp. 151–156.

Riedl, J. – Konstan, J. – Vrooman, E. 2002: *Word of Mouse, The Marketing Power of Collaborative Filtering*. Warner Books, New York.

Vandermerwe, S. 2000: How increasing value to costumers improves business results. *Sloan Management Review*, Vol. 42., pp. 27–37.

Winoto, P. – Tang, T. 2008: If you like the Devil wears Prada the book, will you also enjoy the Devil wears Prada the movie? A study of cross-domain recommendations. *New Generation Computing*, Vol. 26., pp. 209–225.