УДК 004.89

Maydanyuk D. V., Team Leader (Company Segater)

MATHEMATICAL METHODS IN DESIGNING OF THE UNIQUE RECOMMENDER SYSTEMS

Recommender system methodology and principles of work overview is shown. These systems are able to generate unique person oriented recommendations and help to effective understanding of Web visitor's needs.

***Keywords*: recommender system, collaborative filtering, content method, hybrid recommender, similarity metrics.**

I. Introduction

Nowadays, mathematical methods in designing Web-oriented systems become more relevant because they enable to identify exactly user needs and, thus, to give a unique competitive advantage to the system. The use of various recommender services (RS) implies a possibility to form unique person oriented recommendations for a Web-resource user basing on the analysis of his activity and preferences; it is widely spread in business (Amazon, NetFlix and similar services).

The use of collaborative filtering methodology in its pure form drives to relevant and satisfactory results. Practice shows that it is necessary to develop so-called hybrid systems, which allow combining collaborative and different statistical methods for granting person-oriented recommendations. Besides, most part of the created systems are based on formalized heuristics and needs further improvements for more efficient work.

II. Definition of the problem

Recommender systems are mainly based on the analysis of ratings, which users gave to the product. Generally, the task of getting a recommendation comes down to the definition of some rating of the recommended product. This rating is estimated according to the preceding recorded user activity and any other available information about him (location, sex, age, etc.). Only products, which got the highest ratings, are further recommended.

The task can be formalized as follows; suppose that:

C – set of all users,

S – set of offered products.

U – utility function, describing the utility of a product from S for C.

Then, for each consumer from $\in$ C we would like to choose such a product from s`$\in$ S, which most closely corresponds to the utility for the user:

$$s` = \arg\max u(c,s), \quad \forall c \in C .$$

Product utility can be expressed in scores, which show how much the user liked the particular product. One of the major problems in RS is that the product utility u (c, s) is not defined for the totality C×S. As a rule, utility is defined only for products, which user have evaluated. Any user action with the product (preview, purchase, recommendation to friends, etc.) can be seen as a rating. The matrix, showing the interaction between user and product, is presented in table. In this case, $u(c_i,s_j)=r_{ij}$, where r_{ij} – a rating of i-th user, given to j-th product.

Thus, RS must be able to "predict" rating of the products, not directly evaluated by the user, and give recommendations based on these predictions.

Also, it is necessary to take into account that a user can be described using an individual profile, which contains different information about him such as age, sex, income level, geographical location; and a product can be described using the information about price and different consumer categories.

Table

Matrix «Users – Products»

Users	Products				
	s_1	**s_2**	**s_3**	**…**	**s_n**
c_1	r_{11}	∅	r_{13}	…	r_{1n}
c_2	∅	r_{22}	r_{23}	…	r_{2n}
…	…	…	…	…	…
c_m	r_{m1}	r_{m2}	∅	…	∅

III. Results

Desired ratings of the products, not evaluated by the users, can be obtained by using different methods of multivariate approximation, classification, clusterization and heuristic methods. Now there is a provisional classification of task-solving procedures and they are divided in 3 groups: content, collaborative and mixed (hybrid).

Content methods base on the definition of similarities between the user and the product, which means that the user will get recommendations about products he chose earlier.

Collaborative methods base on the definition of similarities between the users, which means that the user will be offered products that were chosen in the past by the people with similar tastes and interests.

Hybrid methods combine content and collaborative methods.

The advantages, disadvantages and peculiarities of these methods are listed below. Under content methods the utility function u (c, s) is mostly defined heuristically as a similarity function between content profiles of the product and user [2]:

$$u(c,s) = f(Content[c], Content[s]).$$

Where $Content[c], Content[s]$ – weight vectors, determining characteristics and properties of user's and product profile. As a product category, by which this weight is determined, different key words can be chosen. They should give more complete description of product characteristics and properties, for example:

$$Content[s_i] = (w_{i1}, w_{i2}, w_{i3}, \ldots, w_{in}).$$

w_{i1} is understood to be keyword weight in profile i. In order to calculate weight we can use information retrieval, which are methods using frequency of occurrence (methods of direct/reverse frequency) [1]. For users the content profile is formed according to his activity history on the Web resource, previewed and estimated products, and consequently, is a certain reflection of user's preferences in product properties. Utility function is defined as a line similarity measure between product and user's profile.

Limitations of content methods are directly connected with properties of subjects of recommendations. For consistent performance, the product presentation should have a shape, available for computer analysis, which means a set of strictly formalized properties. There are following existing disadvantages of content methods:

1. The system will not differentiate two different products with alike set of properties;
2. RS will not be able to recommend products that are substantially different from those familiar to the user;
3. RS will not be able to give recommendations to a new user who made few or no ratings at all.

Under collaborative methods, the product utility for the user results from rating given earlier by other users with similar interests and preferences. Utility function $u(c,s)$ of a product s for a user c is calculated basing on utilities $u(c_j, s)$ of the product s given by users c_j similar to the user c.

There are two algorithm classes: anamnestic (predicts user ratings basing on previous ratings done by this user earlier) and model (builds a model by different ratings, this model will be a base for predictions about future ratings).

The value of unknown rating $r(c,s)$ for a user *c* and a product *s* is calculated as total ratings of the product *s* given by other users [4]:

$$r(c,s) = aggr\, r(c`, s), c` \in C`$$

where $C`$ – total of N users with the most similar preferences to the user c and who rated the product *s*. $r(c,s)$ is often a function showing the similarity degree between users in the set S_{xy} of all products rated both by the user *x* and the user *y*:

$$S_{xy} = \{s \in S \,|\, r(x,s) \neq \varnothing \,\&\, r(y,s) \neq \varnothing\}.$$

Correlation method and method of line similarity are the most often used to identify similarity between users.

Collaborative methods also have some substantial drawbacks. There is so called problem of a new user; that means RS should first study user's preferences basing on the ratings given by him. This is exactly the same with the product, RS cannot recommend any product until it gets enough quantity of ratings. Tasks of collaborative filtration are, as a rule, sparse, i.e. quantity of ratings, which are needed to be predicted, usually exceed by far the quantity of ratings got from users.

Hybrid methods, which combine collaborative and content approaches, try to eliminate the shortcomings of the methods listed above. This allows avoiding limitations common to each method. There are several variants of method uniting [2, 3]:

1. uniting solutions, in other words realization of collaborative and content algorithm goes separately, but their recommendations are united;
2. inlining content rules to collaborative method;
3. inlining collaborative rules to content method.

Despite these processes, hybrid methods just partially improve results gained through basic methods.

IV. Conclusions

Last years proposed us content, collaborative and hybrid algorithms for gaining the recommendations, which give us unique advantages. Many systems are successfully applied in business (Amazon, MovieLens, NetFlix and others).

Nevertheless, most part of the systems bases on formalized heuristics and for more efficient work requires further improvements, among which are:

- RS «understanding» of user needs and product properties;
- Dimensionality of recommendations, their orientation towards different user groups;
- Obtrusiveness, system ability to restructure.

RS keep developing and problem solutions are made by involving new methods of data analysis and computer learning.

1. B. Sarwar, G. Karypis, J. Konstan and J. Riedl (2000) Application of Dimensionality Reduction in Recommender Systems – A Case Study Proc. ACM Web KDD Workshop. **2.** Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions Gediminas Adomavicius **/** Alexander Tuzhilin **//** IEEE Transactions on Knowledge and Data Engineering. – Vol. 17, No 5. **3.** J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining Collaborative Filtering Recommendations, Proc. ACM Conf. Computer Supported Cooperative Work, 2000. **4.** B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," Proc. 10thInt'l WWW Conf., 2001.

Рецензент: д.е.н., професор Сазонець І. Л. (НУВГП)

Майданюк Д. В., керівник групи (Компанія Segater)

МАТЕМАТИЧНІ МЕТОДИ В РОЗРОБЦІ УНІКАЛЬНИХ РЕКОМЕНДУЮЧИХ СИСТЕМ

В роботі подано огляд методології створення і принципів функціонування рекомендуючих систем, що здатні генерувати унікальні адресні пропозиції і приводити таким чином до більш ефективного розуміння потреб користувачів Web ресурсів.
***Ключові слова*: рекомендаційна система, коллаборативна фільтрація, контентний метод, гібридна система, метрики схожості.**

Майданюк Д. В., руководитель группы (Компания Segater)

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В РАЗРАБОТКЕ УНИКАЛЬНЫХ РЕКОМЕНДУЮЩИХ СИСТЕМ

В работе дается обзор методологии создания и принципов функционирования рекомендательных систем, способных выдавать уни-

кальные адресные предложения и вести, таким образом, к эффективному пониманию потребностей пользователей Web-ресурса.
***Ключевые слова*: рекомендательная система, коллаборативная фильтрация, контентный метод, гибридный метод, метрики схожести.**