The Method and Algorithm for Increasing Diversity in Recommendation Systems

Mykhailo Lobur, Mykhaylo Shvarts, Yuriy Stekh

Lviv Polytechnic National University
yuriy.v.stekh@lpnu.ua

INTRODUCTION

Recommendation systems are systems for providing personalized recommendations to users when selecting items. Personalization of recommendations means that the recommendation system offers users the items that most closely match their preferences[1]. Recommended systems arose at the end of the twentieth century[5], [6]. The incentive for the development of recommendation systems was the introduction of a worldwide network INTERNET. E-commerce is a new line of business that arose with the advent of the INTERNET [4]. Electron trade is a new direction of trade relations, which also arose with the introduction of the INTERNET network. E-commerce is a relationship aimed at generating profits arising from the acquisition, modification or termination of civil rights and obligations remotely with the use of information and telecommunications systems, as a result of which the participants in such relations have rights and obligations of property nature. Internet shop is the main subject of e-commerce. The task of e-commerce is the trade in items (goods, objects or services)[3], [4]. The main task of e-commerce is to obtain profit through the maximum satisfaction of the needs of consumers.[3]. Therefore, subjects of e-commerce make significant efforts and resources to research, develop and improve methods and tools for building recommendation systems.[1], [2]. Recommendation systems offer users items that have chosen similar to them consumers. The similarity of users in the method of collaborative filtering is defined as the similarity of user profile vectors. Diversity, novelty, and serendipity remain largely unheeded. In this paper, we propose a method and algorithm for accounting for the dissimilarity of items in the method of collaborative filtering.

THE METHOD AND ALGORITHM FOR INCREASING DIVERSITY IN RECOMMENDATION SYSTEMS

Let there be a set of objects $\mathbf{S} = \{s_1, s_2, \ldots, s_n\}$. This set has a high level of diversity, if there is a big difference between objects in the set. Let the difference between the objects s_i and s_j calculated by the formula $dist(s_i, s_j)$. Then the diversity of the set of objects can be calculated by the equation (1)

$$diversity(\mathbf{S}) = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} dist(s_i, s_j), s_i \in \mathbf{S}, s_j \in \mathbf{S} \tag{1}$$

The recommended system offers the user *Top n* items. Mostly these are ten items with the highest forecasted rating. However, among these items may be those that almost do not differ in their characteristics. The matrix of item - user in the method of collaborative filtering is characterized by high sparsity. The number of non-zero elements does not exceed 6-7% of the total number of elements in the matrix. The weighted sum method is used to predict the ratings of items.

The classical model of collaborative filtering is to predict the unknown product rating for the active user by user-user or product-product model. For a method user-user [1,2]

$$p_{a,i} = \overline{r}_a + \frac{\sum_{u \in \mathbf{U}} (r_{u,i} - \overline{r}_u) w_{a,u}}{\sum_{u \in \mathbf{U}} \left| w_{a,u} \right|} \tag{2}$$

where $p_{a,i}$ – rating *i*-th item for active users;

$\overline{r}_a$ – the average of the rating of the active user;

$\overline{r}_u$ – average rating value of the *u*-th user;

$r_{u,i}$ – rating of the *i* item for *u* user;

$w_{a,u}$ – coefficient of similarity for a rating vector of active user and rating vector of *u* user.

For the method of the item-item [1,2]

$$p_{u,i} = \frac{\sum_{u \in \mathbf{U}} r_{u,n} w_{i,n}}{\sum_{u \in \mathbf{U}} \left| w_{i,n} \right|} \tag{3}$$

where $w_{i,n}$ – coefficient of similarity for a rating vector of *i* item and *n* item.

Accuracy of accounting similarity between vectors $w_{i,n}$ significantly affect the accuracy of prediction of recommendations.

The similarity coefficients for the user profile vectors take into account only the coincident values of the rating estimates. This leads to errors in the prediction recommendations. The predicted items may differ little from each other. Traditionally, recommendation systems do not take into account diversity in forecasting recommendations. The structural scheme of the process of forecasting recommendations taking into account the diversity is shown in Fig.1.

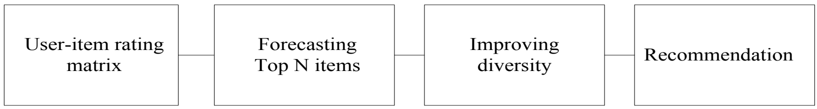

Fig.1 Structural scheme of the process of forecasting recommendations

The following expression is used to calculate the diversity

$$\begin{aligned} FDIV(i, \mathbf{U}, \mathbf{R}) = {} & \alpha F_1(i) + (1-\alpha) F_2(i, \mathbf{R}) * \\ & (\beta F_3(i, \mathbf{U}) + (1-\beta) F_4(i, \mathbf{U})), \end{aligned} \tag{4}$$

where i – an item whose diversification is assess;
$\mathbf{U}$ – active user profile;
$\mathbf{R}$ – set of items that were predicted by the method of collaborative filtering;
$\alpha, \beta \in (0,1]$ – control parameters.

The components of the expression have the following meaning:
$F_1(i)$ – evaluation of the priority of the predicted rating of the i-th item;
$F_2(i, \mathbf{R})$ – an estimate of the diversity of the item i with respect to the set $\mathbf{R}$;
$F_3(i, \mathbf{U})$ – evaluation of the advantages of the item i for the active user;
$F_4(i, \mathbf{U})$ – estimate of the novelty of the object i with respect to the set $\mathbf{R}$.

The following expressions are used to calculate the components of formula (4):

$$F_1(i) = \frac{pr(i)}{MaxRating}, \tag{5}$$

where $pr(i)$ – predicted rating of the i-th item;
$MaxRating$ – highest possible rating;

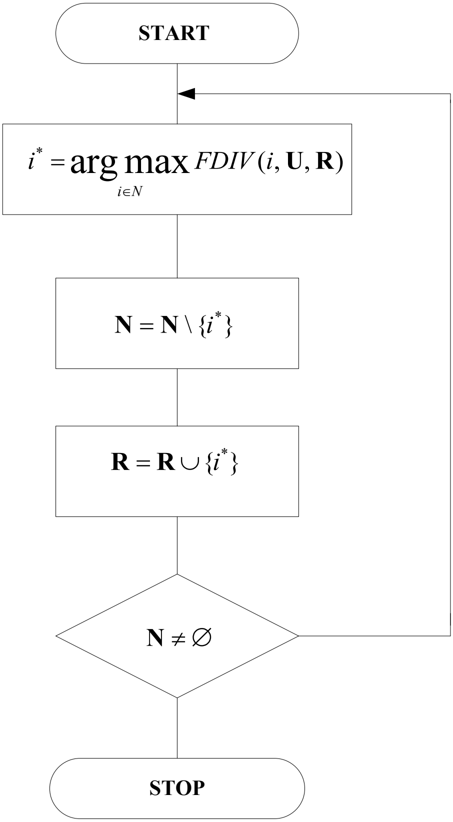

Fig.2. The algorithm for increasing the diversity of items.

$$F_2(i) = \frac{1}{|\mathbf{R}|} \sum_{r \in \mathbf{R}} dist(i, r) , \qquad (6)$$

where $dist(i, r) -$ estimation of the dissimilarity between item i and item r;

$$F_3(i, \mathbf{U}) = \frac{\sum_{u \in \mathbf{U}} sim(i, u)\, r(u, \mathbf{U})}{\sum_{u \in \mathbf{U}} r(u, \mathbf{U})} , \qquad (7)$$

where $sim(i, u) -$ the similarity between item i and item u;

$r(u, \mathbf{U}) -$ rating of the item u in the profile of the active user **U;**

$$F_4(i, \mathbf{U}) = div(i, \{\mathbf{N} \cup \mathbf{U}\}) , \qquad (8)$$

where $\mathbf{N} -$ set of objects from the neighborhood of the active user **U**.

Let the set of recommended items **R**. This set of items iteratively increases to increase diversity due to items from the set **N**. A set is formed as the union of sets of objects from the neighborhood of objects for the active user.

The algorithm for increasing the diversity of items is presented in the Fig.2

References

[1] J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, "Recommender systems survey" // *Knowledge-Based Systems,* vol. 46, pp.109-132, 2013.

[2] L. Lü, M. Medo, C.H.Yeung, Y.-C. Zhang, Z.-K. Zhang, T. Zhou, "Recommender systems" // *Physics Reports,* vol.519, pp.1-49, 2012.

[3] G. Adomavicius, A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions" // *IEEE Transactions on Knowledge and Data Engineering,* vol.17, pp.734-749, 2005.

[4] J.B. Schafer, J. Konstan, J. Riedl, "E-commerce recommendation applications" // *Applications of Data Mining to Electronic Commerce,* pp. 115-153, 2001.

[5] P. Resnick, H.R. Varian, "Recommender systems" // Communications of the ACM, vol.40, pp.56-58, 1997.

[6] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews" In *Proceeding of the ACM Conference on Computer Supported Cooperative Work,* pp. 175–186, 1994.