МАТЕМАТИКА И КИБЕРНЕТИКА — ПРИКЛАДНЫЕ АСПЕКТЫ

В статті описаний взаємозв'язок між двома методами колаборативної фільтрації: методом найближчих сусідів та методом матричної факторизації, які, зазвичай, представляються як протилежні. В даній роботі показано, що обидва підходи є взаємопов'язаними: процес оцінки рейтингів є схожим і, за певних умов, елементи, що використовуються обома підходами, мають високе значення взаємної кореляції, але не є ідентичними

Ключові слова: колаборативна фільтрація, метод найближчих сусідів, матрична факторизація, інтерпретація латентних характеристик

В статье описана взаимосвязь между двумя методами коллаборативной фильтрации: методом ближайших соседей и методом матричной факторизации, которые, обычно, представляются как противоположные. В данной работе показано, что оба подхода являются взаимосвязанными: процесс оценки рейтингов является похожим и, при определенных условиях, элементы, которые используются обоими подходами, имеют высокое значение взаимной корреляции, но не являются идентичными

Ключевые слова: коллаборативная фильтрация, метод ближайших соседей, матричная факторизация, интерпретация латентных характеристик

COMPARATIVE ANALYSIS OF NEIGHBORHOOD-BASED APPROACHE AND MATRIX FACTORIZATION IN RECOMMENDER SYSTEMS

O. Chertov
Doctor of technical sciences, Head of the Department*
E-mail: chertov@i.ua
A. Brun
PhD, Associate Professor**
E-mail: armelle.brun @loria.fr
A. Boyer
PhD, Professor, Head of the KIWI research team**
E-mail: anne.boyer@loria.fr
M. Aleksandrova*
PhD student**
E-mail: rita.v.aleksandrova@gmail.com
*Applied Mathematics department
National Technical University of Ukraine
"Kyiv Polytechnic Institute"
37, Prospect Peremohy, Kyiv, Ukraine, 03056
**Lorraine Research Laboratory in
Computer Science and its Applications (LORIA)
University of Lorraine
Campus scientifique, BP 239,
Vandoeuvre-lès-Nancy Cedex, France, 54506

1. Introduction

The amount of digital information produced by humanity grows exponentially from year to year [1], which makes the process of useful information search more and more difficult. That is why the development of different approaches and systems that help people navigate digital information available is in demand.

One of the classes of systems that help solve such kind of tasks is the class of recommender systems (RS). Recommender systems aim at recommending users some items that are likely to interest them. They are intensively used in many domains, such as e-commerce, e-tourism, e-learning, etc., and help not only contribute to the satisfaction of the user, but also increase profits of commercial systems. The task of rating prediction by the RS can be considered as a task of filling unknown values of a rating matrix, in which each row represents a user and each column – an item. The intersection of a specific row and column reveals the rating of the current user on the current item.

There are three categories of recommendation algorithms [2]: content-based, collaborative filtering and hybrid approaches. Content-based approaches [3] recommend to the active user those items, which are similar to the items already highly appreciated by him. The main drawback of this kind of methods is that the system cannot follow the change of preferences and tastes of the user. Collaborative filtering (CF) [4] relies on the ratings of other users while estimating unknown user preferences. Hybrid-based approaches [5] use the ideas of both content and collaborative-based recommendation algorithms.

Collaborative filtering is proven to result in accurate recommendations and are widely used, especially in the cases when either no or not sufficient amount of content information (information about the items and their similarity) is provided. Two major approaches are used in CF-based recommender systems: the neighborhood-based approach and the matrix factorization-based approach. The Neighborhood-based approach (NB) [6] relies on the preferences of the user's neighbors (other users with similar preferences)

to estimate his/her preferences. The Matrix Factorization (MF) [7] is a relatively new approach. MF represents the relation between users and items through a set of latent factors (also called features). It forms two low-rank matrices, each representing the relation between users (or items) and this set of features. The multiplication of these two matrices allows estimating users' future preferences. Although Matrix Factorization does not have the same intuitive interpretation as NB-based approaches, it was proven to result in accurate recommendations, especially in the case of sparse input rating matrices [7].

We believe that interpretation of features as real users can reveal the deep ideological interconnection of these two approaches. This can lead to the qualitatively new understanding of the basic Collaborative Filtering algorithms and can open new possibilities for their joined usage.

2. Analysis of Published Works and Problem Statement

MF and NB are usually presented as opposed approaches [8] as they rely on different elements: either neighbors or latent features (the latter do not have specific physical meaning). They have never been compared in other terms than their respective performance, for example in terms of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) [9]. However, the objective of this paper is joint analysis of the ratings estimation processes of Matrix Factorization and Neighborhood-based approaches, which, according to our knowledge, have not been presented in other works.

Recently, we proposed to interpret features of MF-based approach as users [10, 11], referred to as representative users (RU). We assumed that a feature k represents a user if the vector of this user has a canonical form: it is a unitary column with only k one non-zero and equal to 1 element on position . We have shown that Non-negative Matrix Factorization with Multiplicative Update Rules (for the sake of simplicity further referred to as NMF) naturally results in factorization matrices that have vectors with a form close to the one previously described [11]. Other works, dedicated to feature interpretation in MF-based approaches proposed to interpret them as behavioral patterns [12] or groups of users [13]. However, we believe that features interpretation as real users can make a link between so different otherwise NB and MF approaches.

3. Purpose and objectives of the study

The objective of this paper is to propose a connection between NB and MF through the notion of representative users.

In order to fulfill this goal the following tasks were performed:

1. Comparative analysis of MF and NB-based approaches.
2. Proposition of a model for connecting MF and NB through the notion of representative users.
3. Validation of the proposed model.

4. Algorithmic Analysis of NB and MF

To recommend items to a user, called the active user, both NB and MF aim at estimating u_a's ratings on the items that he/she has not rated yet. Let U be the set of users (of size M) and I the set of items (of size N). In order to perform this estimation, both approaches rely on users' ratings, represented as a rating matrix $\mathbf{R}$, where r_{ui} is the rating that a user u assigned to an item i.

4. 1. The MF Approach

Matrix factorization is an unsupervised learning method for latent variable decomposition [14]. It has recently received great popularity, especially since the Netflix Price Competition [7].

MF assumes that a small number of latent factors influences users' ratings. It aims at forming two low rank matrices $\mathbf{W}$ and $\mathbf{V}$, with $\dim(\mathbf{W}) = K \times M$ and $\dim(\mathbf{V}) = K \times N$, where K is the number of features. The product of both matrices approximates the rating matrix: $\mathbf{R} \approx \mathbf{W}^T \mathbf{V}$. $\mathbf{W}$ and $\mathbf{V}$ respectively represent the extent to which users and items are related to these latent factors.

To get the estimated rating of an active user u_a on an item i, MF calculates the dot product of the two vectors in $\mathbf{W}$ and $\mathbf{V}$ that correspond to u_a and i. Features obtained with MF algorithms usually don't have any physical sense.

$$r^*_{u_a,i} = \mathbf{w}^T_{u_a} \mathbf{v}_i = \sum_{k=1}^{K} w_{u_a,k} \cdot v_{k,i} \ . \tag{1}$$

Factor matrices $\mathbf{W}$ and $\mathbf{V}$ correspond to the solutions of an optimization task , which can be obtained with such algorithms as Alternating Least Squires (ALS) [15] and Stochastic Gradient Descend (SGD) [16].

$$\left\| \mathbf{R} - \mathbf{W}^T\mathbf{V} \right\| \to \min \ , \tag{2}$$

where $\|\bullet\|$ denotes Euclidian norm.

Non-negative Matrix Factorization is a variant of MF, which forces the values in both matrices to be non-negative. Non-negative factor matrices can be obtained through posing corresponding conditions on solutions obtained with ALS and SGD methods (first group); or through a special optimization procedure, that ensures non-negativity of matrix elements (second group). One popular approach in the second group is Multiplicative Update Rules [17], which updates factor matrices according to the formulae

$$W_{km} \leftarrow W_{km} \frac{\left(\mathbf{V}\mathbf{R}^T\right)_{km}}{\left(\mathbf{V}\mathbf{V}^T\mathbf{W}\right)_{km}}, V_{kn} \leftarrow V_{kn} \frac{\left(\mathbf{W}\mathbf{R}\right)_{kn}}{\left(\mathbf{W}\mathbf{W}^T\mathbf{V}\right)_{kn}} \ . \tag{3}$$

4. 2. The NB Approach

The NB approach, which has emerged from the beginning of CF [18], assumes that users' preferences are correlated and that similar users rate items similarly. To estimate the rating of an active user u_a on item i, this approach exploits the ratings of a set of users similar to u_a: his/her neighbors. The NB technique defines the neighbors of u_a as the set of his K most similar users who rated item i.

The identification of neighbors thus relies on a similarity measure between users (for this reason a similarity matrix $\mathbf{S}$ ($\dim(\mathbf{S}) = M \times M$) is computed). This measure is generally calculated by the Cosine similarity or the Pearson correlation coefficient [6]. The Cosine similarity, contrary to the Pearson correlation, always results in non-negative values if input vectors are non-negative and is computed by the formula

$$cos\_sim\left(\mathbf{v}_1,\mathbf{v}_2\right)=\frac{\mathbf{v}_1\mathbf{v}_2}{\left\|\mathbf{v}_1\right\|\left\|\mathbf{v}_2\right\|}. \tag{4}$$

The similarity measure is also used to estimate the rating of u_a on item i, as the weight associated to neighbors. Estimated ratings are usually evaluated with equation .

$$r_{u_a,i}^{*}=\sum_{n\in U_{u_a}}sim\left(u_a,n\right)\cdot r_{n,i}, \tag{5}$$

where U_{u_a} is the set of the K nearest neighbors of u_a, who have rated i.

5. Representative Users versus Neighbors

5. 1. Identification of Representative Users

In [11] we have shown that feature of NMF can be associated to a set of real users (representative users), also an algorithm of RU identification was proposed. This algorithm consists of 6 steps presented on fig. 1 and further detailed below.

Step 1. A traditional matrix factorization is performed, resulting in both matrices **W** and **V** with K features.

Step 2. A normalization of each of the M column vectors of the matrix **W** is performed to result in unitary columns. The resulting normalized matrix is denoted by $\mathbf{W}_{norm}$ and the set of normalization coefficients is denoted by **C**.

Step 3. This step is dedicated to the identification of the representative users in the $\mathbf{W}_{norm}$ matrix. As shown in [11] first all users are divided on groups of preimage candidates for each feature according the position of the maximum element in the column-vector $\mathbf{w}_{norm}$: a user w for whom the maximum of the column-vector $\mathbf{w}_{norm}$ is situated on the position k belongs to the preimage group of the k-th feature. After this the quality score q of the each preimage candidate w_m is computed using the formula , and the user with the highest quality score among all candidates is considered as the representative one for the feature k.

$$q\left(w_m\right)=1-\frac{dist\left(\mathbf{f}_k,\mathbf{w}_m\right)}{dist^{max}\left(K\right)}, \tag{6}$$

where $dist\left(\mathbf{v}_1,\mathbf{v}_2\right)\equiv\left\|\mathbf{v}_1-\mathbf{v}_2\right\|$ – is an Euclidian distance between vectors $\mathbf{v}_1$ and $\mathbf{v}_2$; $\mathbf{f}_k$ – k -th canonic column-vector, with one non-zero element situated on the position k; $\mathbf{w}_m$ – column vector of matrix $\mathbf{W}_{norm}$, corresponding to the user w_m; $dist^{max}\left(K\right)$ – the maximum distance between a preimage candidate and a canonic column-vector of dimensionality K. As shown in [11], the maximum distance is computed by a formula

$$dist^{max}\left(K\right)=\sqrt{2\left(1-\frac{1}{\sqrt{K}}\right)}. \tag{7}$$

Once all RU are identified, the matrix $_{norm}$ is modified in the following way: for every column-vector $\vec{\mathbf{w}}_k'$, which corresponds to a representative user of the feature k, all values are set to 0, except the one on the position k, that is set to 1. This transformation performs one-to-one mapping of representative users and corresponding features. The resulting modified matrix is the matrix $\mathbf{W}_{norm}^{mod}$. Fig. 2 presents an

example of such transformation. For the sake of simplicity, all representative users are grouped on the left part of the matrix.
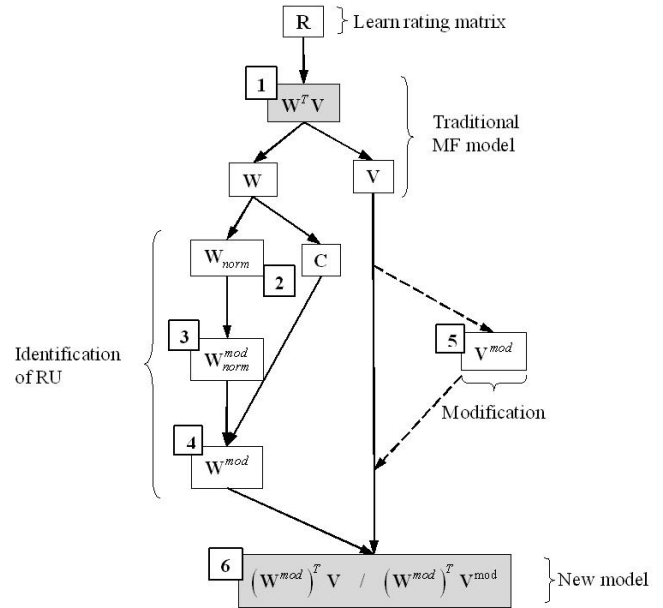

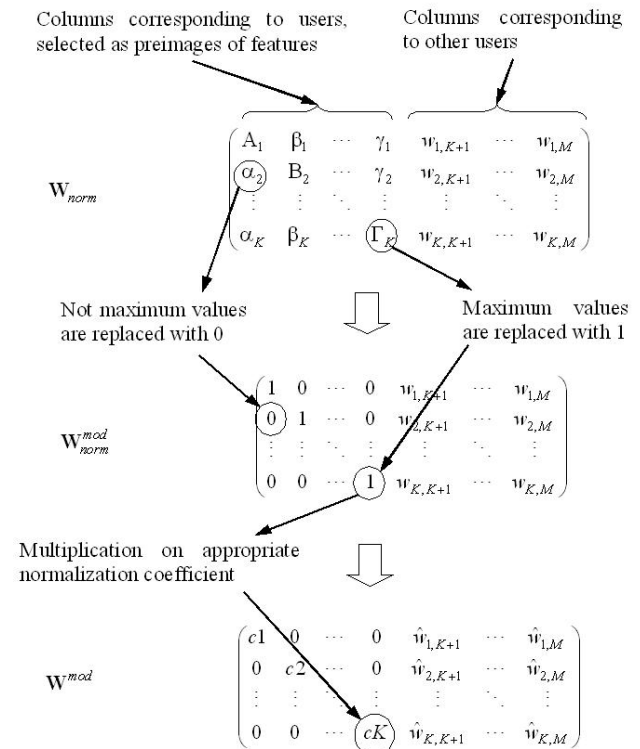Fig. 1. RU identification algorithm

Fig. 2. From $\mathbf{W}_{norm}$ to $\mathbf{W}^{mod}$

In some cases, a feature, say feature k, may have no candidate preimage. In this case, we can either decrease the number of features considered for factorization or search for a vector with the second maximum situated on that specific position.

Step 4. Each column of the matrix $\mathbf{W}_{norm}^{mod}$ is multiplied by the appropriate normalization factor from the set **C**

(Fig. 2). After this, representative users will remain preimages of the features but with scaling coefficients.

Step 5. In order to obtain the best model matrix $\mathbf{V}$ can be modified under the condition of minimal loss. Modification of $\mathbf{V}$ can be performed using optimization methods with the starting value obtained after the first step.

Step 6. The resulting recommendation model is made up of matrices $\mathbf{W}^{\text{mod}}$ and $\mathbf{V}$ (or $\mathbf{W}^{\text{mod}}$ and $\mathbf{V}^{\text{mod}}$).

5. 2. Connection between NB and MF

Let us compare the rating estimation processes of MF and NB .

Both equations perform a sum. In the case of NB, this sum is made of the K nearest neighbors. In the case of MF, it is made of the K features. If the number of neighbors is equal to the number of features, then both sums use the same number of terms. Focusing in details on the terms that are summed up, we can find additional similar points. First, the element $r_{n,i}$ in equation is the rating of the neighbor user n on item i. The element $v_{k,i}$ in the equation represents to what extent item i is related to feature k. In the case features are interpreted as representative users, we rise question (1): does $v_{k,i}$ correspond to the rating of the kth representative user on item i? If yes, both elements $r_{n,i}$ and $v_{k,i}$ can be linked to each other and matrix $\mathbf{V}$ can be considered as an approximation of a rating matrix. Second, the element $\text{sim}(u_a,n)$ in equation represents the similarity between user u_a and his/her neighbor user n. The element $w_{u_a,k}$ in equation represents to what extent user u_a is related to feature k. As this feature is interpreted as a user, $w_{u_a,k}$ may reflect the link between u_a and the kth representative user. We thus raise question (2): does $w_{u_a,k}$ correspond to the similarity between user u_a and the kth representative user? If yes, both elements $\text{sim}(u_a,n)$ and $w_{u_a,k}$ can be linked to each other and matrix $\mathbf{W}$ can be considered as an approximation of the similarity matrix $\mathbf{W}$.

If there is actually a correspondence between these elements, we can conclude that the estimation processes of NB and MF are similar. The questions we raise are schematically presented in Fig. 3.

We have to mention here that there is a big difference between both processes: the set of features (representative users) is unique, whereas the set of neighbors is dependent on each pair (user, item). However, it was shown in [19] that exploiting a unique set of neighbor users in NB, leads to a high quality of recommendations (low MAE). NB and MF may thus be considered as similar.

In the following section, we conduct experiments on a benchmark dataset, to determine if the elements used in the estimation process of NB and of NMF are similar. Because NMF algorithm was used to perform Matrix Factorization, NB with cosine similarity was considered, to ensure non-negativity of both models.
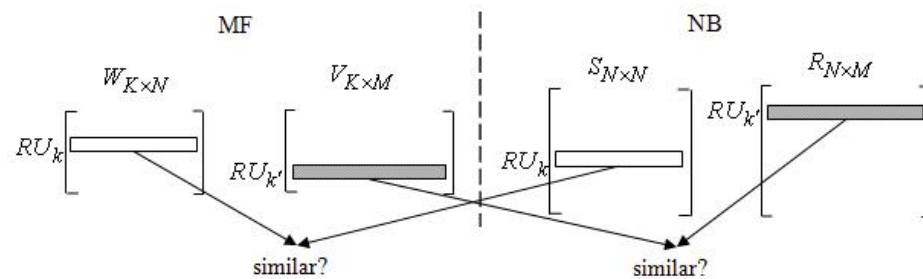
6. Experimental Analysis of NB and MF Rating Estimation Processes

We conduct the experiments on the 100k MovieLens dataset [20], which contains 100 000 ratings, ranging from 1 to 5, assigned by 943 users to 1682 items. 80 % of the ratings are randomly chosen to form the learning set and the 20 % remaining ratings are used for the test set. The number of features used for NMF is $K=10$ (following the experiments in [11], where the best results were obtained with $K=10$), and the number of neighbors used for NB is $K=10$. The accuracy of the models is evaluated with the standard mean absolute error (MAE), computed by formula, where L corresponds to the number of ratings in the test set, r_l represents a rating value from the test set and r_l^* – the corresponding estimated value.

$$\text{MAE} = \frac{1}{L}\sum_{i=1}^{L}\left|r_l - r_l^*\right|. \tag{8}$$

We first aim at answering question (1): can matrix from NMF, be considered as a rating matrix? With NMF, after identifying the users that correspond to the features (the representative users), we study if the values in $\mathbf{V}$ correspond to the ratings of the representative users in matrix $\mathbf{R}$. We calculate the cosine similarity between the corresponding lines in the matrices.

The resulting average similarity is **0.972**, with a standard deviation equal to **0.013**. This shows that matrix $\mathbf{V}$ is highly similar to the lines in $\mathbf{R}$, that correspond to representative users. We can thus answer question (1): matrix $\mathbf{V}$ can be considered as an approximation of the rating matrix of the representative users.

Based on this answer, we can now raise question (2): can matrix $\mathbf{W}$ be considered as a similarity matrix between representative users and all users in the system? As in the previous case, we calculate the cosine similarity between lines of matrix $\mathbf{W}$ and lines of matrix $\mathbf{S}$ that correspond to representative users. The resulting average similarity value is **0.666**, with a standard deviation equal to **0.110**. We can conclude that matrices $\mathbf{W}$ and $\mathbf{S}$ are fairly similar, even if they are less similar than $\mathbf{V}$ and $\mathbf{R}$.

As $\mathbf{V}$ is highly related to $\mathbf{R}$, we perform an additional experiment. We force $\mathbf{V}$ to contain the rating values from $\mathbf{R}$ (those of the representative users). We run one additional iteration of NMF to update $\mathbf{W}$ and we study if the resulting matrix $\mathbf{W}$ is closer to $\mathbf{S}$ or not. First, we assign the value **0** to unknown rating values (model $\text{NMF}_0^{(t+1)}$). In this case the mean and standard deviation values of the similarity between vectors in matrix $\mathbf{W}$ and the corresponding vectors in $\mathbf{S}$ are equal to **0.741** and **0.089**. Second, we assign the values of $\mathbf{V}$ (from NMF) to unknown rating values (model $\text{NMF}_{V^t}^{(t+1)}$). The resulting mean and standard deviation values are **0.671** and **0.110**. Results of similarity analysis of different elements of MF and NB approaches are summarized in Table 1.

We can conclude that filling $\mathbf{V}$ with ratings increases the similarity between $\mathbf{W}$ and $\mathbf{S}$, especially when $\mathbf{V}$ is initialized with the value 0 in the case of unknown ratings.

Fig. 3. Connection between MF and NB

Table 1

Similarity between different elements of MF and NB

cosine similarity	$\mathbf{R}$ and $\mathbf{V}$	$\mathbf{S}$ and $\mathbf{W}$ (NMF)	$\mathbf{S}$ and $\mathbf{W}$ ($\mathrm{NMF}_0^{(t+1)}$)	$\mathbf{S}$ and $\mathbf{W}$ ($\mathrm{NMF}_{V^t}^{(t+1)}$)
mean	0.972	0.666	0.746	0.671
std	0.013	0.110	0.089	0.110

We now focus on the MAE of the models previously considered (Table 2). MAE of NMF and NB are equivalent (respectively **0.802** and **0.801** respectively). The MAE of $\mathrm{NMF}_{V^t}^{(t+1)}$ **(0.833)** remains close to the one of the traditional NMF and NB, but is increased. However, the MAE of $\mathrm{NMF}_0^{(t+1)}$ is more than twice higher than that of NMF **(1.830)**. Thus, filling $\mathbf{V}$ with 0 values, when the ratings are unknown, highly decreases the accuracy of the model.

Table 2

Accuracy for different models

model	NB	NMF	$\mathrm{NMF}_0^{(t+1)}$	$\mathrm{NMF}_{V^t}^{(t+1)}$
MAE	0.801	0.802	1.830	0.833

From these experiments, we can conclude that there is a connection between NB and NMF. Indeed, the elements used by both approaches are highly correlated, thus can be interpreted in the same way, especially since both approaches perform similarly in terms of MAE (for NMF and $\mathrm{NMF}_{V^t}^{(t+1)}$).

7. Conclusion

In this paper, we raised the question whether there exists a connection between the two most popular recommendation approaches: matrix factorization and neighborhood-based approaches, which are usually presented as opposed. First, we have shown that the rating estimation processes are equivalent. Second, based on a series of preliminary experiments, we have shown that NB and NMF can be considered as similar under the condition that features are interpreted as users. After interpreting features from NMF as users (representative users), we have shown that matrix $\mathbf{V}$ from NMF can be considered as an approximation of the rating matrix $\mathbf{R}$ (for these representative users). We have also shown that matrix $\mathbf{W}$ from NMF is an approximation of the user similarity matrix $\mathbf{S}$, traditionally used by NB. Thus, both elements used by NMF (matrices $\mathbf{W}$ and $\mathbf{V}$) correspond to both elements used by NB (matrices $\mathbf{S}$ and $\mathbf{R}$). However, although both approaches have similar MAE, a major difference remains between NB and NMF: the set of representative users and the set of neighbors. NMF results in a unique set of representative users, which is used to predict ratings for all users in the system. At the opposite, NB forms a set of neighbors for each pair (user, item), which makes NB more complex.

In a future work, we would like to perform a similar analysis for other MF techniques (ALS, SGD) and study if each feature could be associated with a set of representative users, not only one user, thus making MF even closer to NB.

References

1. Turner, V. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. IDC iView [Electronic resource] / V. Turner. – Access mode: http://www.emc.com/leadership/digital-universe/2014iview/digital-universe-of-opportunities-vernon-turner.htm

2. Adomavicius, G. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-art [Text] / G. Adomavicius, A. Tuzhilin // IEEE Transactions on Knowledge and Data Engineering. – 2005. – Vol. 17, Issue 6. – P. 734–749. doi: 10.1109/tkde.2005.99

3. Pazzani, M. J. Content-Based Recommendation Systems [Text] / M. J. Pazzani, D. Billsus // The Adaptive Web. Lecture Notes in Computer Science. – 2007. – Vol. 4321 – P. 325–341. doi: 10.1007/978-3-540-72079-9_10

4. Schafer, J. B. Collaborative Filtering Recommender Systems [Text] / J. B. Schafer, D. Frankowski, J. Herlocker, S. Sen // Lecture Notes in Computer Science. – 2007. – Vol. 4321. – P. 291–324. doi: 10.1007/978-3-540-72079-9_9

5. Burke, R. Hybrid Recommender Systems: Survey and Experiments [Text] / R. Burke // User Modeling and User-adapted Interaction. – 2002. – Vol. 12, Issue 4. – P. 331–370.

6. Breese, J. Empirical Analysis of Predictive Algorithms for Collaborative Filtering [Text] / J. Breese, D. Heckerman, C. Kadie // Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI'98), 1998. – P. 43–52.

7. Koren, Y. Matrix Factorization Techniques for Recommender Systems [Text] / Y. Koren, R. Bell, C. Volinsky // Computer. – 2009. – Vol. 42, Issue 8. – P. 30–37. doi: 10.1109/mc.2009.263

8. Takacs, I. Matrix Factorization and Neighbor Based Algorithms for the Netflix Prize Problem [Text] / I. Takacs, I. Pilaszy, B. Nemeth, D. Tikk // Proceedings of the 2008 ACM Conference on Recommender systems, 2008. – P. 267–274. doi: 10.1145/1454008.1454049

9. Shani, G. Evaluating Recommendation Systems [Text] / G. Shani, A. Gunawardana. – Recommender Systems Handbook, 2011. – P. 257–297. doi: 10.1007/978-0-387-85820-3_8

10. Brun, A. Can Latent Features be Interpreted as Users in Matrix Factorization-based Recommender Systems? Vol. 2 [Text] / A. Brun, M. Aleksandrova, A. Boyer // Proceedings of 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014. – P. 226–233. doi: 10.1109/wi-iat.2014.102

11. Aleksandrova, M. Search for User-related Features in Matrix Factorization-based Recommender Systems. Vol. 1 [Text] / M. Aleksandrova, A. Brun, A. Boyer, O. Chertov // European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2014), PhD Session Proceedings, 2014. – P. 1–10.

12. Pessiot, J. F. Factorisation en Matrices Non-negatives pour le Filtrage Collaboratif [Text] / J. F. Pessiot, V. Truong, N. Usunier et al. // Proceedings of 3rd Conference en Recherche d'Information et Applications, 2006. – P. 12.

13. Zhang, S. Learning from Incomplete Ratings Using Non-negative Matrix Factorization [Text] / S. Zhang, W. Wang, J. Ford, F. Makedon // Proceedings of the 6th SIAM Conference on Data Mining, 2006. – P. 548–552. doi: 10.1137/1.9781611972764.58

14. Sarwar, B. Application of Dimensionality Reduction in Recommender System a Case Study [Text]: technical report / B. Sarwar, G. Karypis, J. Konstan, J. Riedl. – Minnesota University Minneapolis Department of Computer Science, 2000. – 15 p.

15. Zhou, Y. Large-scale Parallel Collaborative Filtering for the Netflix Prize [Text] / Y. Zhou, D. Wilkinson, R. Schreiber, R. Pan // Algorithmic Aspects in Information and Management. – 2008. – Vol. 5034. – P. 337–348. doi: 0.1007/978-3-540-68880-8_32

16. Koren, Y. Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model [Text] / Y. Koren // Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008. – P. 426–434. doi: 10.1145/1401890.1401944

17. Lee, D. D. Algorithms for Non-negative Matrix Factorization [Text] / D. D. Lee, H. S. Seung // Proceedings of Advances in Neural Information Processing Systems. 2001. – P. 556–562.

18. Goldberg, D. Using Collaborative Filtering to Weave an Information Tapestry [Text] / D. Goldberg, D. Nichols, B. Oki, D. Terry // Communications of the ACM. – 1992. – Vol. 35, Issue 12. – P. 61–70. doi: 10.1145/138859.138867

19. Boumaza, A. Stochastic Search for Global Neighbors Selection in Collaborative Filtering [Text] / A. Boumaza, A. Brun // Proceedings of the 27th Annual ACM Symposium on Applied Computing, 2012. – P. 232–237. doi: 10.1145/2245276.2245322

20. MovieLens Dataset GroupLens [Electronic resource] / Available at: http://grouplens.org/datasets/movielens/

У статті наведено результати порівняльного аналізу трьох методів метрологічної атестації математичних моделей технологічних елементів газотранспортних систем: методу імітаційного моделювання, методу статистичної лінеаризації, методу речових інтервалів. Показано, що для розглянутих моделей результати метрологічної атестації за трьома розглянутими методами практично збігаються, а найбільш ефективним виявився метод зосереджених інтервалів

Ключові слова: стохастичний модель, лінійна ділянка, метод, лінеаризація, імітаційне моделювання, речові інтервали

В статье приведены результаты сравнительного анализа трех методов метрологической аттестации математических моделей технологических элементов газотранспортных систем: метода имитационного моделирования, метода статистической линеаризации, метода вещественных интервалов. Показано, что для рассмотренных моделей результаты метрологической аттестации по трем рассмотренным методам практически совпадают, а наиболее эффективным оказался метод центрированных интервалов

Ключевые слова: стохастический модель, линейный участок, метод, линеаризация, имитационное моделирование, вещественные интервалы

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МЕТРОЛОГИЧЕСКОЙ АТТЕСТАЦИИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

А. Д. Тевяшев
Доктор технических наук, профессор*
E-mail: tevjshev@kture.kharkov.ua

Ю. С. Асаенко
Аспирант*
E-mail: esset8@gmail.com

А. М. Кобылин
Доцент
Кафедра информационных технологий
Харьковский институт банковского дела Университета банковского дела НБУ
пр. Победы, 55, г. Харьков, Украина, 61202
E-mail: kobilin@khibs.edu.ua
*Кафедра прикладной математики
Харьковский национальный университет радиоэлектроники
пр. Ленина, 14, г. Харьков, Украина, 61166

1. Введение

При выборе математических моделей для решения практических задач в реальном масштабе времени возникает проблема метрологической аттестации моделей, т. е. проблема оценивания степени неопределенности зависимых переменных (результатов вычислений) от степени неопределенности независимых переменных (исходных данных).

В системах реального времени в математическую модель подставляются результаты косвенных измерений технологических параметров, получаемых из SCADA-систем. Любые косвенные измерения содержат определённый уровень неопределенности. Как правило, предполагается, что результаты косвенных измерений (исходные данные) являются случайными величинами, имеющими нормальное распределение с известными статистическими характеристиками

9