

Chornous, G., Nikolskyi, I., Wyszyński, M., Kharlamova, G., & Stolarczyk, P. (2021). A hybrid user-item-based collaborative filtering model for e-commerce recommendations. *Journal of International Studies*, 14(4), 157-173. doi:10.14254/2071-8330.2021/14-4/11

Journal
of International
Studies

© Foundation
of International
Studies, 2021
© CSR, 2021

Scientific Papers

A hybrid user-item-based collaborative filtering model for e-commerce recommendations

Galyna Chornous

Department of Economic Cybernetics, Taras Shevchenko National University of Kyiv
Ukraine
chornous@univ.kiev.ua
ORCID 0000-0003-4889-1247

Ihor Nikolskyi

Department of Business Automation, BIZZABO LTD
Ukraine
ibor.nikolskyi@gmail.com

Mateusz Wyszyński

Faculty of Economics and Management, Łazarski University, Warsaw, Poland
mateusz.wyszynski@lazarski.pl
ORCID 0000-0003-2969-634X

Ganna Kharlamova

Department of Economic Cybernetics, Taras Shevchenko National University of Kyiv
Ukraine
akharlamova@ukr.net
ORCID 0000-0003-3614-712X

Piotr Stolarczyk

Faculty of Economics and Management, Łazarski University, Warsaw, Poland
piotrstolarczyk@o2.pl
ORCID 0000-0003-0089-7252

Abstract. The COVID-19 pandemic deepened understanding of e-commerce as an extremely promising sphere. Nowadays, even small businesses are widely using e-shops and e-markets. Thus, small and medium-sized e-commerce companies need powerful, flexible recommender systems, which do not require significant

Received:
February, 2021
1st Revision:
October, 2021
Accepted:
December, 2021

computing and financial resources. Currently, the main vector of such systems developing is targeted at hybridization, i.e. a combination of well-known effective methods. The paper proposes a hybrid model (bagging) to achieve high-quality e-commerce recommendations, which are based on the effective combination of collaborative filtering techniques. The model consists of the following components: User-based collaborative filtering (UBCF, classical variant and method involving text comments to compute the rating matrix), and Item-based collaborative filtering (IBCF) using the attributes of the predicted objects to calculate their similarity. The proposed model can become a methodological basis to introduce the recommender system for the medium-scale e-commerce platforms that are not able to afford deep learning. The recommender system based on this model does not require real-time updates and is easy for website integration. Besides, Root Mean Squared Error (RMSE) of the proposed method is significantly lower than in IBCF and UBCF models. Due to the significant improvement of recommendations accuracy, e-commerce companies will have a chance to positively affect customer loyalty without considerable investment.

DOI:
10.14254/2071-
8330.2021/14-4/11

Keywords: recommender system, hybrid model, hybridization, collaborative filtering, rating matrix, item-based technique, user-based technique, e-commerce.

JEL Classification: C49, L81, M31

1. INTRODUCTION

Due to the rapid growth of e-commerce, entrepreneurs faced an urgent need to stimulate consumers through giving a valuable advice. The impetuous growth and variety of information available on the web and the speedy development of new e-commerce services (buying products, product comparison, different auctions, etc.) frequently overburden the customers and result in their poor decisions (Vynogradova et al., 2020; Yasynska et al., 2019). This negatively affects profitability.

The solution is seen in the implementation of recommender systems (RSs). Over recent years they have proven to be effective in decreasing the level of information pressure on the customers. The key idea of the RSs is to direct a consumer to a new object (product, service, etc.) that would be new for him/her and at the same time relevant to his/her current goals. This is a crucial factor in substantiating the relevance of the RSs' concept development.

Ricci, Rokach, and Shapira (2015) define RS as “software and techniques providing suggestions for items to be of use to a user.” Initially the RSs were developed in the early 1990s (Resnik et al., 1994; Schafer et al., 1999). In those days, a large number of algorithms and approaches to the creation of such systems appeared (Ricci et al., 2015; Aggarwal, 2016). Since that time, the RSs have had great potential for application in e-commerce. In this case, products, manufacturers or suppliers, product sets or service packages turned into the objects for recommendations.

Schafer, Konstan, and Riedl (1999, 2001) outlined the features of the RSs usage in e-commerce and proved its capacity to overcome the information overload and to achieve the effect of mass customization. In their opinion, RSs improve e-commerce sales in three ways: converting "browsers" into buyers, cross-sell increasing, and loyalty building.

Ricci et al. (2015) listed some other reasons for the companies to use RSs technology on their sites, they are as follows: to increase the number of objects sold, to sell more diverse objects, to increase customer

satisfaction, to increase customer fidelity, to better understand what the customer wants. Ultimately, all the above results in the business profit increase and raising convenience for the client.

A large number of approaches have been developed for non-personalized and personalized RSs such as product association approach (Poriya et al., 2014), content-based, information filtering (Hooda et al., 2014; Woldan et al., 2020), collaborative filtering (memory-based - user-based (Resnik et al., 1994; Li & Li, 2020) and item-based (Sarwar et al., 2001; Ujkani et al., 2020; Ajaegbu, 2021), model-based (Kane, 2018)), knowledge-based (Aggarwal, 2016), multi-criteria recommender systems (Shambour, 2021; Smirnov & Ponomarev, 2020), etc.

However, the wide spreading of the RSs is restrained by several problems.

First of all, as a rule, only a few enterprises have the necessary human and capital resources to develop, implement and maintain a powerful RS that would be able to be updated in real-time. This is why the urgent question is to find the ways to optimize classical methods that do not require additional specific conditions.

On the other hand, there are a lot of other problems: dealing with the RS development, namely, the lack of data, the problem of security, privacy and fraud, etc. (Abid et al., 2020). For useful work of the RS, a reasonable content of information is required. This is not only data of the ratings obtained directly from the consumer (like or dislike, on a scale of 1 to 5, etc.) or indirect ratings reflecting the customer's opinion based on his/her behaviour (data generated in the customer purchase history that can show tastes and patterns) (Dehuri et al., 2013). Also, the quality of the recommendations is influenced by manipulations (interference of subjects (people or robots) that factitiously over- or underestimate the object rating) (Khusro et al., 2016).

The study of consumers' feedbacks (comments, reviews) is one of the tools to overcome some of the obstacles dealing with the above problems. Researchers quite actively explore this area. Chen et al. (2015) divided those studies into two main categories - based on the feedback from the modelling of a user profile or a product profile. Among the developed methods, the improvement of the traditional collaborative filtering method is valuable. It works by the forming of one-dimensional virtual ratings based on the results of the feedback tonality classification (Kane, 2018).

However, the effectiveness of these methods is also limited due to a high level of general data sparsity (Chen et al., 2015). This confirms the urgency in finding the ways of the feedback exploiting while making the recommendations. On the other hand, the usage of single reviews as a source of recommendations is not always sufficient. The reason is that, at least, this approach is not suitable for all types of recommendations and all e-resources.

The e-commerce software offers a wide range of RSs applications. However, it should be emphasized that the issue of hybridization is currently determining the main vector of RSs development and is the main source for further improvements of the approach.

Today, the bulk of the RSs in e-commerce uses only one type of data that customers create when purchasing (e.g., direct ratings, comments, or historical data). While the idea of hybridization involves the use of all available information at the same time. It is combined in an integrated intelligence information system. That helps to form recommendations based on a much more complete portrait of the consumer.

The COVID-19 pandemic demonstrated that e-commerce is an extremely promising area. Nowadays, even small businesses are widely using e-shops and e-markets. The competition is becoming more fierce. Now, a lot of researchers try to explain online purchase intention, examine the factors that affect customer loyalty to e-commerce and online shopping (Dakduk et al., 2017; Shafiee et al., 2018; Zatonatska et al., 2019; Abdallah et al., 2018). The results indicate that the use of RSs influences e-service quality positively and directly. They have a positive impact on e-loyalty, which in turn significantly affects online purchases and repurchases. Thus, the implementation of RSs adds e-commerce companies more competitive advantages that confirms the need for quite powerful but affordable systems.

The discussed problems determine the objective absence of a universal and accessible (in the case of limited resources of enterprises) tool for communication between the seller and the buyer. To improve widespread approaches and search for new ones, the individual characteristics of consumers should be used to for effective combination of the existing algorithms, as well as for the development of new ones. Consequently, this determines the relevance of the research of the RSs.

The purpose of this study is to develop a hybrid model to achieve more accurate and high-quality e-commerce recommendations based on the efficient combination of collaborative filtering techniques that do not require significant computing power and easy-to-integrate into the website's architecture. Such RSs will be useful for medium-sized commerce companies that are gaining ground in the electronic market.

The paper is organized as follows. The next section reviews the related literature. The description of the main collaborative filtering techniques is given in the third section. The proposed hybrid model of collaborative filtering is presented in the same section. In the fourth section, the main experimental results are shown. Finally, conclusions and opportunities for further research are considered in the last section.

2. LITERATURE REVIEW

The section of the literature review presents the related studies that demonstrate the broad acceptance of a hybrid approach in the development of RSs and confirm the variety of hybridization methods.

2.1. Popularity of the methods combination

Comparison of different recommender algorithms on the accuracy, the complexity of preparation and optimization time, presented in (Hnot, 2016), suggests that using separate methods demonstrates a direct connection between the complexity of the RS and its effectiveness. Today, the only hybridization animates the usage of simple classical methods for more accurate predictions.

Burke classified the main models of hybrid RSs as follows: weighted hybridization, switching hybridization, mixed hybridization, feature combination hybridization, cascade hybridization, feature augmentation hybridization, meta-level hybridization (Burke, 2007). These models differ in architecture and organization of the usage of the results for each involved algorithm in the hybridization process.

Besides, it is common to combine collaborative filtering techniques with another method or traditional measurement metrics to avoid a cold-start problem. Feng et al. (2021) considered the extreme case of sparse data (the new user cold-start problem) and proposed the collaborative filtering ranking model, which combines rating-oriented and pairwise ranking-oriented approaches, precisely the Probabilistic Matrix Factorization and a Bayesian Personalized Ranking. Ajaegbu (2021) proposes to balance the three current traditional measurement metrics such as: Cosine-based similarity, Pearson correlation similarity and Adjusted cosine similarity, in the direction of cold-start situations. This algorithm complements the strength of the traditional measurement metrics with evidence shown when measured with Mean Absolute Error.

The widespread use of collaborative filtering-based hybridization is confirmed by an analysis of research over the past two years.

Walek and Fojtik (2020) proposed a monolithic hybrid RS, which combines a collaborative filtering system, a content-based system, and a fuzzy expert system for use in the area of movie recommendation. Wu, Sang, and Cui (2021) suggested a semi-supervised ensemble filtering method to improve the recommendation performance by combining three popular collaborative filtering techniques. First, three weak predictors with labelled data are independently initialized by three different algorithms. Predictors are generated by neighbourhood methods and by latent factors model too. The final prediction is achieved by mixing the results from the three predictors enhanced with unlabeled data. Yassine et al. (2021) presented a new intelligent RS that combines collaborative filtering with the popular clustering algorithm K-means. The

authors involved user demographic attributes (gender and age) to create segmented user profiles. All of the proposed RSs improve the performance and time response speed of the traditional collaborative filtering and the content-based techniques.

2.2. Deep learning and other approaches

Most state-of-the-art studies on RS development refer to this use of deep learning methods.

Martins, Papa, and Adeli (2020) conducted a systematic review focusing on deep learning techniques applied in collaborative filtering recommendations. The authors considered the diverse non-linear modelling strategies to deal with rating data, the combination of deep learning techniques with traditional collaborative filtering-based linear methods. Farashah et al. (2021) propose the original method, which consists of the four phases: 1) clustering and classification of new users using Deep Neural Network algorithm; 2) using collaborative filtering RS based on hybrid similarity criterion which based on users features, such as age, gender, and occupation; 3) running improved Friendlink algorithm to count the similarity between users; 4) combining results of collaborative filtering RS and improved Friendlink algorithm. Molaei et al. (2021) developed a complete deep learning conceptual model by learning latent social features to embed in a collaborative filtering approach. First, representation learning applies to the rating matrix to extract the latent social features. Then, an advanced deep learning approach based on cascade tree forest is involved in the recommendation process.

However, RSs based on deep learning are not available to most small and medium-sized e-commerce companies. Such companies need powerful, flexible tools that do not require significant computing and financial resources. The following systems meet this request in some way.

Leng et al. (2021) proposed a novel collaborative filtering approach based on multiple attribute decision making. The method determines the weight of each item and computes the preference similarities between the active and inactive users. According to the preference similarities, the candidate neighbourhood of the active user is determined. The authors used the most frequent item recommendation method to provide top-N recommendations to the active user. Li and Li (2020) developed a collaborative filtering recommendation algorithm based on user characteristics and interests. The similarity calculation based on user score similarity, user attribute feature similarity and user interest similarity was presented. The algorithm uses the number of user evaluations of attributes to determine the interest of users in different attributes. The similarity calculation formula is used to estimate the interest similarity score between users. The final similarity for recommendation is a combination of the user attribute feature similarity, user interest similarity, and the user rating similarity. However, these studies do not take into account the user review content.

2.3. Analyzing User Review Content

The RSs, which involve such elements as overall product rating, and users' comments, are much more appropriate resources to support consumer decision-making in e-commerce. In this context, research on hybrid algorithms using text comment analysis is noteworthy.

Fidan (2020) demonstrated that the weighting methods such as Term Frequency and Term Frequency-Inverse Document Frequency as well as commonly used classification algorithms such as Naïve Bayes and SVM have some inadequacies in short text analysis. The author developed a grey relational classification model based on the Vector Space Model and Bag of Words. Yang et al. (2021) proposed a hybrid personalized recommendation model that extracts user preferences by analyzing user review content in different sentiment polarity at the sentence level, based on jointly applying user-item score matrices and dimension reduction methods. The authors demonstrate that fine-grained emotion recognition has good

adaptability to a sparse rating matrix with a reasonable and good performance. Visa and Patel (2021) presented the approach, which initially generates the text score based on users' reviews with the help of opinion mining. Then the rating corresponding to the text estimates is submitted to the input of Convolutional Neural Network which results in better predictions in a product RS.

Having in mind the existing research gap, in this paper, we propose a new approach to the RS development, which includes a modification of the evaluating the similarity in the context of the collaborative filtering model and the use of users text comments as an additional source for a recommendation.

3. METHODOLOGY AND RESEARCH MODEL

The methodology section describes in detail the collaborative filtering approach to RS development. Then we introduce a new hybrid (bagging) model, which allows circumventing some limitations of the collaborative filtering methods.

3.1. Main collaborative filtering techniques

To achieve study purposes, firstly, we introduce the most common algorithmic approaches for generating personalized buying proposals. Jannach et al. (2011) proposed to divide these approaches into two main groups: collaborative filtering and content-based, information filtering. Naturally, it is possible to apply a hybrid approach, meaning to use different techniques within the RSs.

Content-based filtering provides a personalized approach to each user, taking into account the contextual features of both the user and the recommended objects.

Collaborative filtering (CF) techniques use given rating data for many objects by many users as the basis for missing rating forecasting and a top-N recommendation list generating for a given user who acts on the system as an active user.

Formally, we have a set of users (in terms of e-commerce called customers) $U = \{u_1, u_2, \dots, u_m\}$ and a set of objects (for example, products) $I = \{i_1, i_2, \dots, i_n\}$. Customer's preferences are stored in a $m \times n$ customer-object rating matrix $R = (r_{jk})$ where each row represents the customer u_j ($j = \overline{1, m}$) and columns represent the object i_k ($k = \overline{1, n}$).

CF aims to create recommendations for a customer called the active user $u_a \in U$. We define the set of objects unknown to this customer as $I_a = I \setminus \{i_t \in I \mid r_{at} = 1\}$. The two basic tasks are to predict ratings for all objects in I_a or to create a list of the best N recommendations to u_a .

There are two main methods of CF: the nearest neighbour (NN) methods and the latent factor methods. The NN methods are based on the principle that consumers who have preferred similar objects in the past tend to prefer similar objects in the future. These methods can be user-based (UBCF) or item-based (IBCF).

The UBCF approach, also known as k-NN CF, comes under a class of memory-based algorithms. Originally, the methodology was presented in (Resnik et al., 1994). In the UBCF, the objects recommended to a consumer are those that other ones, with similar preferences, have chosen previously. A very important decision is the choice of similarity function. The literature (Ricci et al., 2015; Jannach et al., 2011; Herlocker et al., 2004) presents several different approaches for determining the similarity function - Pearson correlation coefficient, Spearman rank correlation coefficient, cosine similarity, etc. However, empirical

analyses show that for UBCF the Pearson coefficient outperforms other measures of users comparing (Ricci et al., 2015; Herlocker et al., 2004).

Although UBCF is an effective method, there are some scalability issues when the set of customers grows. To expand CF capabilities for processing with large databases and facilitate the deployment of e-commerce sites, a more scalable algorithm, IBCF, was proposed. The main idea of the IBCF is to form proposals based on the relationship between objects inferred from the rating matrix. The assumption behind this approach is that users will prefer objects that are similar to other objects they like.

The IBCF finds the most similar objects (k-NN is used) to generate predictions and recommendations. In this case, the advantage of IBCF is the ability to calculate the similarity matrix in advance. Relating to the choice of the similarity function, the IBCF examines cosine similarity and Pearson correlation coefficient, as well as the UBCF, does. However, it has been emphasized in (Hahsler, 2013), the cosine similarity measure consistently outperforms the Pearson correlation metric for the later-described item-based recommendation techniques.

In RSs with a sufficiently high degree of consumer-object correspondence, one consumer adding or changing a rating is unlikely that significantly alter the similarity between the two objects, especially if the object has many ratings. This allows substantiating the expediency of a preliminary calculation of object similarity in the similarity matrix. Rows of this matrix can even be shortened to store only the most similar objects. When consumers change their ratings, these data will become a bit outdated, but they are likely to still get good recommendations, and the data can be fully updated by converting the similarity matrix during a low system load.

The first step in preparing a model for a recommendation is to divide a complete sample into training and test ones. We can use several approaches (Hahsler, 2013): splitting, bootstrap sampling and k-fold cross-validation. The last approach suggests splitting the total sample into k sets (called folds) of approximately the same size, then we evaluate k times. It produces more robust results and error estimates.

Evaluation of different techniques for developing RSs is an important topic and its reviews were presented in (Herlocker et al., 2004; Gunawardana & Shani, 2009; Silveira et al., 2019). The evaluation metrics concerning CF scenarios were studied in (Martins et al., 2020).

A typical way to evaluate a prediction is to compute its deviation from the true value. This is the basis for the Mean Average Error (MAE) and the Root Mean Square Error (RMSE). The average RMSE and MAE have corrected error metrics for unbalanced test sets. RMSE penalizes larger errors stronger than MAE and thus is suitable for situations where small prediction errors are not very important.

Not all RSs work on the principle of predicting such a parameter as the rating of an object. They recommend to consumers those objects that potentially can be of interest to them. To evaluate the quality of such a recommendation, a confusion matrix is constructed (Hahsler, 2013). It shows how many objects are recommended correctly, where negative recommendations are truly negative and positive ones are true positive. In reliance on this matrix, some basic metrics are calculated for the analysis of the recommendation effectiveness, for instance, the accuracy of the prediction. Precision and recall are the two best-known classification metrics; they are also used for measuring the quality of information retrieval tasks in general.

In addition to the above methods, the following user-centric quality metrics are used to evaluate the perceived quality of the RSs: perceived accuracy (also called relevance), novelty, overall users' satisfaction, etc. (Cremonesi et al., 2011).

Traditional neighbourhood-based CF algorithms are widely used in the RSs field. However, they may present such problems as cold-start (the system is unable to make reliable recommendations due to the lack of initial ratings) and the data sparseness (the number of ratings available is generally very small compared to the number of ratings that need to be provided).

To overcome these and other problems of CF, as well as to improve the quality of rating predictions, we can use hybridization - a combination of different approaches of collaborative and content-based filtering. The hybrid approach is most popular when RSs are developing for commercial sites.

3.2. Hybrid model of collaborative filtering

To achieve study purposes, firstly, we introduce the most common current algorithmic approaches for generating personalized buying proposals. Jannach et al. (2011) proposed to divide these approaches into two main groups: collaborative filtering and content-based, information filtering. Naturally, it is possible to apply a hybrid approach, meaning to use different techniques within the RSs.

A new hybrid (bagging) model ST-UICBF, which allows circumventing the limitations of the methods of IBCF and UBCF, is composed of such components as UBCF (classical method), T-UBCF (UBCF involving text comments to compute the rating matrix), and S-IBCF (IBCF using the properties of predicted objects to calculate their similarity). Thus, the model can use multiple information sources, including text.

To realize this model for the recommendation, the following problems are solved:

- development of the algorithm to the optimal model;
 - preliminary processing and analysis of input data;
 - transformation of text comments into ratings;
 - experimenting to find the optimal RS in different configurations;
 - confirmation or refutation of the hypotheses:
1. Hybrid RS demonstrates more accurate results than individual algorithms (without the use of text comments);
 2. The use of text comments as the basis for the consumer rating matrix demonstrates higher-precision results in hybrid models;
 3. Using the content component to determine the similarity of the objects improves the accuracy of the IBCF model.

The structure of the proposed hybrid model is shown in Figure1, consists of four main parts. The first part is to generate the first rating (predicted rating No. 1) by the improved IBCF approach based on the principle of combining two types of objects similarity. The second part is responsible for creating the predicted rating No. 2, No. 3 by the UBCF approach, where a part is a regular forecast by the ratings, and the second part is the prediction of convertible set into the numerical indicator of text comments. The three rating types are combined to construct an objective function f in part 4, which is an integrated optimal model. It is configurable by the optimization parameters that are determined during the learning process. Parts 1-3 form the foundation for finalizing the appearance of the hybrid model in part 4.

The efficiency of the RS based on the CF models depends on the calculation of the similarity of the objects. This similarity is determined by internal and external factors. The classical IBCF approach involves only external factors to define similarity by the reckoning of ratings. In fact, the similarity between objects also depends on internal factors, such as the properties of objects that reflect the gained semantic charge (Xu & Raahemi, 2016). In other words, the similarity of the objects depends on both internal and external factors. In this work, internal factors denote the properties of the object, which are used to characterize and depend on object specificity. For instance, if an object is a product, properties can be trade dress, colour, functionality, price, quality, etc.

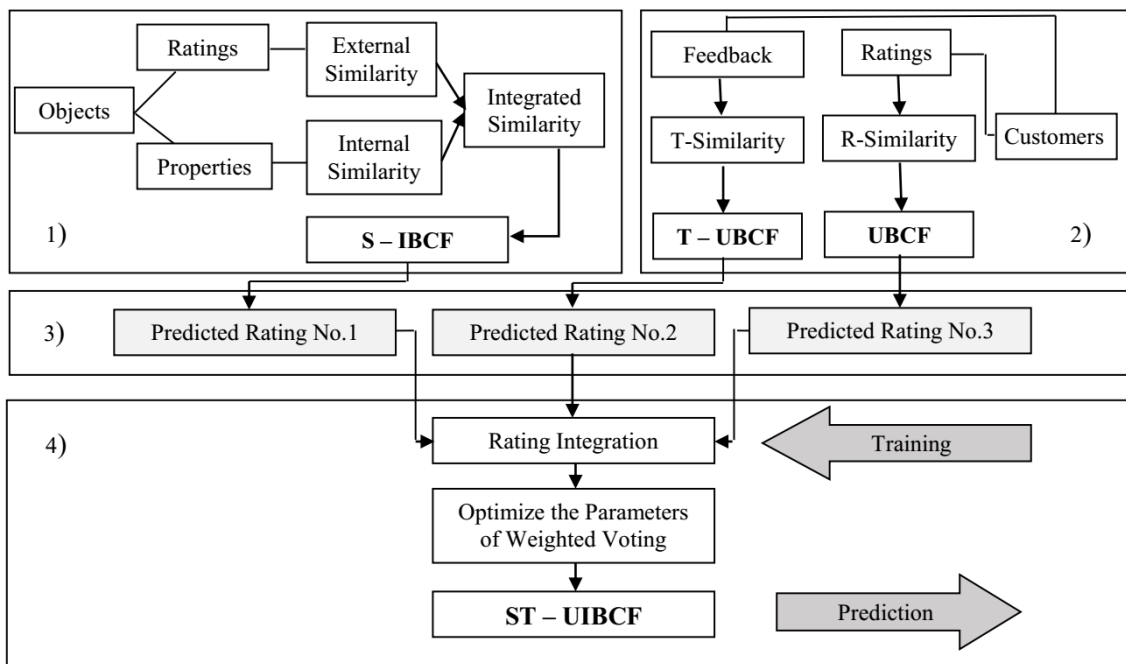


Figure 1. Conceptual operation scheme of the hybrid model ST-UICBF

Source: Authors' results

The paper proposes to use the Pearson correlation coefficient to determine the external similarity of objects i and j ($i, j \in I$):

$$sim_{out}(i, j) = \frac{\sum_{k \in K} (r_{k,i} - \bar{r}_i)(r_{k,j} - \bar{r}_j)}{\sqrt{\sum_{k \in K} (r_{k,i} - \bar{r}_i)^2} \sqrt{\sum_{k \in K} (r_{k,j} - \bar{r}_j)^2}}, \quad (1)$$

where $K = U_i \cap U_j$ - the set of customers that rated both objects i and j . The symbols \bar{r}_i and \bar{r}_j correspond to the average rating of the objects.

The internal similarity of the objects $i, j \in I$ can be rated as:

$$sim_{in}(i, j) = \sum_{p \in P} \varphi(p) sim(\Theta(p), i, j), \quad (2)$$

where Θ - the set of characteristics $p \in P$, $sim(\Theta(p), i, j)$ is a measure of the similarity of the object i to the object j according to the characteristic p of the set Θ , and $\varphi(p)$ - the weight of the characteristic p (Cheng et al., 2017).

A combination of similarity functions is planned to be carried out based on the model of a switching hybridization based on the indicator of the rating matrix for each object.

The sparsity function for the set of consumers who have rated i (U_i^I) and j (U_j^I) is defined as follows:

$$SP_{i,j}^I = \frac{2 * |U_i^I \cup U_j^I| - |U_i^I \cup U_j^I|}{2 * |U_i^I \cup U_j^I|}. \quad (3)$$

According to the analysis, it is advisable to combine the internal and external similarities of objects. Data sparsity coefficient can be used to balance external and internal similarities. To do this, we determine the weight function between the internal and external similarity, which includes an element of local similarity to the sigmoid function in the following way:

$$f(SP_{i,j}^I) = \begin{cases} \frac{1}{1+e^{-SP^I}}, & 0 \leq SP_{i,j}^I \leq 1 \\ 1, & SP_{i,j}^I = 1 \end{cases} . \quad (4)$$

Due to this approach, the value of the function is $f(SP_{i,j}^I) \in [0,5;1]$, which means that internal similarity will always be used since it has already proven its effectiveness (Resnik et al., 1994). Summarizing all the above concerning the integration of similarity functions, the function ultimately gets the form:

$$sim(i, j) = f(SP_{i,j}^I) \cdot sim_{in}(i, j) + (1 - f(SP_{i,j}^I)) \cdot sim_{out}(i, j) . \quad (5)$$

Another important aspect of the model is the rating prediction based on text comments. First, we need to choose the law that is supposed to be distributed data. After that, according to the given examples, the parameters of this distribution are calculated for future classification.

The target class, in this case, is the positive feedback class. As a result, the probability of a comment hit is converted into a discrete five-point rating system by the following rule:

$$r_{i,j}^T = \begin{cases} 1, & P_{i,j}(c|d) < 0,2 \\ 2, & 0,2 \leq P_{i,j}(c|d) < 0,4 \\ 3, & 0,4 \leq P_{i,j}(c|d) < 0,6 \\ 4, & 0,6 \leq P_{i,j}(c|d) < 0,8 \\ 5, & 0,8 \leq P_{i,j}(c|d) \leq 1 \end{cases} . \quad (6)$$

The final component - the integration of ratings - is realized based on the principle of weighted voting, where, as a result of modelling the preferences of the consumer u , three predicted ratings $\hat{r}_{u,i}^{(1)}$, $\hat{r}_{u,i}^{(2)}$, $\hat{r}_{u,i}^{(3)}$ for the object i are obtained. To get more accurate predictions, the proposed model combines predicted ratings in this way:

$$\hat{r}_{u,i} = \hat{r}_{u,i}^{(1)} * w_1 + \hat{r}_{u,i}^{(2)} * w_2 + \hat{r}_{u,i}^{(3)} * w_3 = \vec{\hat{r}}_{u,i} \vec{w}^T, \quad (7)$$

where the vector $\vec{w} = (w_1, w_2, w_3)$ consists of three components to determine the effect of the algorithms S-IBCF, T-UBCF and UBCF.

In this way, we have the optimization problem:

$$\min_w (\ell(r_{u,i}, \hat{r}_{u,i}) + \mathbb{R}(\vec{w})), \quad (8)$$

Here is $\ell(r_{u,i}, \hat{r}_{u,i})$ - the penalty function, in our case, the sum of squared errors of prediction (SSE):

$$\ell(r_{u,i}, \hat{r}_{u,i}) = \frac{1}{2} \sum_{(u,i)} (r_{u,i} - \hat{r}_{u,i})^2, \quad (9)$$

which specifies the difference between the actual and predicted values, and $\mathbb{R}(\vec{w})$ - the Frobenius regularization function, which further penalizes the model to avoid overfitting problems, and is defined as:

$$\mathbb{R}(\vec{w}) = \frac{1}{2} \lambda \|\vec{w}\|^2, \quad (10)$$

The impact of the regularization is controlled by constant λ . The constant λ controls the extent of regularization and is usually determined by cross-validation. $\|\cdot\|$ denotes the Frobenius norm.

The stochastic gradient descent (SGD) algorithm solves the optimization equation (8).

Consequently, we propose a new approach to the RS development, which includes a modification of the evaluating the similarity in the context of the IBCF model and the use of text comments as an additional source for the formation of the rating matrix. In the final configuration, the model can be called Similarity and Text improved – User-Item Based Collaborative Filtering (ST-UICBF).

For the convenient experimental results following, we present the main acronyms of models in Table 1.

Table 1

Collection of acronyms

Acronym	Model Name
UBCF	User-based Collaborative Filtering
IBCF	Item-based Collaborative Filtering
T-UBCF	UBCF involving text comments to compute the rating matrix
S-IBCF	IBCF using the properties of predicted objects to calculate their similarity
ST-UICBF	Similarity and Text improved – User-Item Based Collaborative Filtering

Source: Authors' results

4. RESULTS

In this section, the main experimental results based on the Amazon review data are presented. Also, a comparison of the other research results obtained on the same dataset is provided in the Discussion part.

During the 4 series of experiments using 5-fold cross-validation, the following hypotheses are confirmed:

1. the hybrid system exhibits greater accuracy of predictions due to the mutual compensation of the errors of each ensemble component;
2. CF model can be constructed exclusively on text comments without, at least, loss of accuracy,
3. usage of the content component in determining the similarity of the objects greatly improves the accuracy of the IBCF model.

To test the proposed hybrid model ST-UICBF, we choose the Amazon Fine Food Reviews dataset, which consists of 568,454 food reviews that Amazon users left up to October 2018 as our dataset (McAuley & Leskovec, 2013). It includes food information from the Amazon Web Store and customers' ratings and text reviews.

The original database is cleared from the unnecessary objects, customers without ratings, duplicate comments, but it still required additional manipulations to implement the ST-UICBF approach.

First, it is necessary to prepare the basis for improving the IBCF model in the context of using the semantic features of objects while calculating similarity coefficients. To do this, it is necessary to establish a correspondence between different objects as demonstrated in (Cosley et al., 2003).

The second step is to develop the UBCF algorithm based on the transformation of text comments into the numerical rating. For the experiment, the Linear SVM (Support Vector Machine) algorithm is chosen as the main transformation algorithm, which proved its high efficiency in working with the classification of text data in unbalanced samples with a certain class dominance that can be assumed from the initial rating

distribution (Press et al., 2007). Assuming that there is a relationship between the numerical ratings and the text reviews, and given the fact that the real rating distribution is significantly shifted towards positive ratings (5), it is more appropriate to use the binary classification, rather than to determine the 5 classes. The distribution of the classification results turned out to be 23% negative and 77% positive ratings. Classification accuracy is 0,89.

During the experiment, four groups of tests with different sizes of samplings are performed: 40%, 60%, 80% and 100% of the basic dataset. Table 2 shows comparison between such methods accuracy: Popular - the recommendation of only popular objects (baseline); classical IBCF; classical UBCF; T-UBCF - UBCF based on text comments; S-IBCF - IBCF with improved similarity mechanism; ST-UIBCF - hybrid model S-IBCF + UBCF + T-UBCF.

Table 2

Accuracy comparison (RMSE) of recommended methods

Methods	Sampling percentage			
	40%	60%	80%	100%
Popular	1,368	1,390	1,385	1,396
IBCF	1,310	1,239	1,204	1,157
UBCF	1,273	1,260	1,225	1,136
T-UBCF	1,019	0,981	0,925	0,892
S-IBCF	0,849	0,843	0,835	0,819
ST-UIBCF	0,801	0,779	0,786	0,768

Source: Authors' calculations.

The results identify ST-UIBCF as the best of the proposed methods. Algorithms T-UBCF and S-IBCF are much more accurate than their classical versions, but they are slightly inferior to ST-UIBCF.

As can be seen in Table 2, the testing results are greatly improving with the increase of the involved dataset up to 80% and 100%. This can be explained in this way. As the percentage of sample coverage increases, the dataset is replenished by new consumers, new objects and new ratings that qualitatively supplement the complete dataset. This allows making higher-precision predictions about the similarity of consumers and objects.

The last step is to analyse the structure of ST-UIBCF in each series of experiments. Every group of experiments is completed by the formation of 5 optimal vectors, which determine the weight of each component in the structure of the hybrid method. The average values of the optimal vectors $\bar{w} = (w_1, w_2, w_3)$ are shown in Figure 2.

A similar situation is observed in all series of experiments. S-IBCF is a dominant algorithm in a composition that plays a decisive role in forecasting ratings, especially in terms of the increasing dataset. The impact of UBCF and T-UBCF decreased with an increase in the sampling. Most probably, this stems from the fact that more consumers, objects, and ratings provided more opportunities to uncover the potential of the S-IBCF, which increased its impact from 62.4% to 76% while using complete sampling.

Consequently, the experiments carried out to confirm all the introduced hypotheses and, in fact, open the potential to continue the research in the direction of optimization and development of classical CF methods that could compete in efficiency with state-of-art algorithms of machine learning.

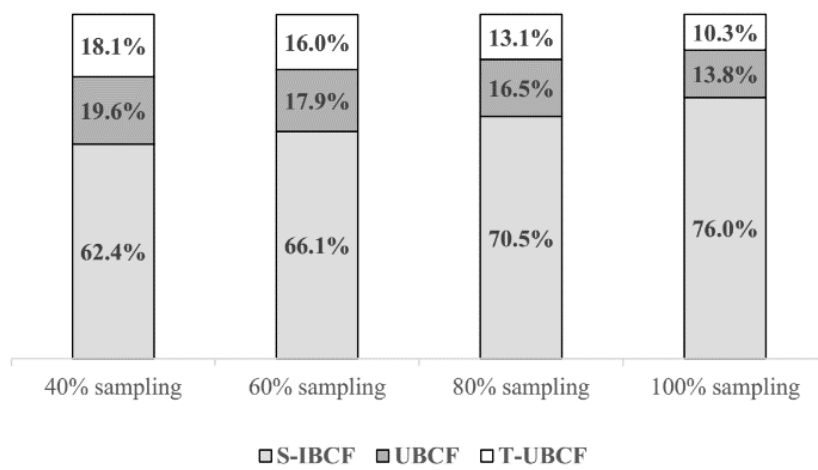


Figure 2. ST-UICBF structure in the experiments

Source: Authors' results

5. CONCLUSIONS AND DISCUSSION

In this section, we conclude our main findings and discuss the possible directions for future research.

5.1. Discussion: Other research results based on the Amazon reviews data

Since the publication of Amazon Fine Food Reviews datasets in 2013 (McAuley & Leskovec, 2013), the collection has been used widely for evaluating recommendation techniques. According to Google Scholar, there are currently 397 citations (last checked: March 19, 2021) to the original paper (McAuley & Leskovec, 2013), where the authors use the collection to evaluate their method that models consumers personal evolution or experience for recommendations. To understand how different experience levels affect consumer's taste, they introduced the latent factor model. Chu et al. (2015) use the sets to focus on the inner connection and relationship among consumers and objects by letting it learn consumer's habits and what they purchased. Their models are designed for predicting the rating of different objects based on the data given, such as the time when the review was written, the number of helpfulness voting, as well as the length of the review. The best model they have found is the latent factor model (RMSE=0.78). Zheng et al. (2016) implemented several kinds of regressions with the feature involving text content and the regressions only focus on nontext features. The most powerful regression with the text feature is the mixture of unigrams and bigrams representation (RMSE=1.06).

Most studies chose the latent factor model as the best one to deal with a recommender problem. In the researchers' opinion, this approach has a really good performance on data that has a chaotic and unpredictable feature so far. Later examples using the datasets include Shao et al. (2017) that compared different methods to establish more reasonable and reliable RSs based on latent factor models. As a result, the researchers implemented Matrix Factorization, SVD, Deep Learning, Random Forest and Times Series. For the involved dataset, the Matrix Factorization gives a very good prediction: the train set RMSE is 0.184 and the test set RMSE is 0.188. We accentuate that a powerful pre-processing of data was implemented and some Deep Learning methods were used to build an RS for Amazon users. RMSE value of the ST-UICBF model is higher, but appropriate RS is less complicated and readily available for the medium-scale Internet

portal. The neighbourhood-based CF algorithms are widely used in the RS field for their interpretability and operability.

The conclusions on performance comparisons between different models, stated in the literature, may change when tested on cleaned data. This effect is apparently stronger for more complex methods. It was essential to recall a stronger susceptibility of these methods to overfitting. Thus, it is needful to be cautious in using data for evaluation and competitive comparison of methods (Basaran et al., 2017).

5.2. Conclusions

The experimental results identified ST-UIBCF as the best of the proposed methods. They exceeded other algorithms by RMSE value in all four series of experiments (Androniceanu et al., 2020).

The practical value of the research results is a combination of the typical methods of CF that allows to achieve higher accuracy of prediction than while using the individual methods. The proposed approach is efficient to reduce the complexity of optimization, compared with the basic methods of machine learning for solving the problem of recommendation (Androniceanu et al., 2019). The proposed hybrid model is capable to put in place more information sources than individual models. In addition to the ratings, the semantic component and customer feedback are considered to be the factors to achieve accurate and qualitative recommendations and improve classical methods of CF. Moreover, the significant advantages of this hybrid method are no requirements for additional computing power and easiness to integrate into the website architecture in the absence of necessity to update the recommendations in real-time.

According to modern realities, Internet access is a prerequisite for the survival and attraction of new clients, especially for media-oriented enterprises (movie portals, news portals, etc.), regardless of the format of doing business (Mičík & Mičudová, 2018; Saksiriruthai, 2018). The proposed model can become a methodological basis to introduce RSs to the medium-scale Internet portal, such as an e-shop, a movie portal, etc. As a rule, these companies are not able to afford the RSs that require a real-time update, which means the proposed model can give them a chance to positively affect the loyalty and engagement of their customers, to increase the average cost of the consumer basket due to recommendations' accuracy growth without heavy capital commitment.

5.3. Future research

The opportunities for hybridization are inexhaustible, and the ST-UIBCF method can be expanded and improved. The next steps in this direction might be:

- development and inclusion into the ensemble of the S-UBCF model by analogy to the S-IBCF model. S-UBCF will determine the similarity of consumers based on their socio-demographic attributes;
- improvement of the S-IBCF model to the SH-IBCF (Similarity-History Improved Item Based Collaborative Filtering). This approach provides for examination whether the customer purchase history can be an additional effective source of information for accuracy increase of identifying goods similarity;
- implementation of a more complex approach to the text sentiment analysis and conversion of the results into a scale from 1 to 5;
- rearrangement of the model architecture. Testing various hybrid model configurations to obtain a higher synergy effect.

REFERENCES

- Abdallah, G. M. A., Abou-Moghli, A. A., & Al-Thani, A. H. (2018) An examination of the e-commerce technology drivers in the real estate industry. *Problems and Perspectives in Management*, 16(4), 468-481. [https://doi.org/10.21511/ppm.16\(4\).2018.39](https://doi.org/10.21511/ppm.16(4).2018.39)
- Abid, U., Ashraf, M.U., Butt, M.U. (2020) *A Critical Survey on Privacy Prevelling in Collaborative FiltringRecomender System: Challenges, State-Of-The-Art Methods and Future Directions*. 2020 International Conference on Engineering and Emerging Technologies, ICEET 2020, Lahore, Pakistan, 1-7. <https://doi.org/10.1109/ICEET48479.2020.9048206>.
- Aggarwal, C.C. (2016) *Recommender Systems: The Textbook*. Springer: New York, USA.
- Ajaegbu, C. (2021) An optimized item-based collaborative filtering algorithm. *Journal of Ambient Intelligence and Humanized Computing*. <https://doi.org/10.1007/s12652-020-02876-1>
- Androniceanu, A. (2019) *Collaborative Digital Economy, Main Attributes, and Challenges*. In: Peters M. (eds) Encyclopedia of Educational Philosophy and Theory. Springer, Singapore. <https://doi.org/10.1007/978-981-287-532-7>
- Androniceanu, A., Kinnunen, J., & Georgescu, I. (2020) E-Government clusters in the EU based on the Gaussian Mixture Models. *Administratiiesi Management Public*, 35, 6-20. <https://doi.org/10.24818/amp/2020.35-01>
- Basaran, D., Ntoutsis, E., & Zimek, A. (2017) *Redundancies in Data and their Effect on the Evaluation of Recommendation Systems: A Case Study on the Amazon Reviews Datasets*. Proceedings of the 2017 SIAM International Conference on Data Mining, Houston, USA, April 27-29. 390-398. <https://doi.org/10.1137/1.9781611974973.44>
- Burke, R. (2007) Hybrid web recommender systems. In *The Adaptive Web*; P. Brusilovsky, A. Kobsa, W. Nejdl (eds.); Springer-Verlag: Berlin, Germany. 377-408. https://doi.org/10.1007/978-3-540-72079-9_12
- Chen, L., Chen, G., & Wang, F. (2015) Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction*, 25, 99-154. <https://doi.org/10.1007/s11257-015-9155-5>
- Cheng, S., Zhang, B., & Zou, G. (2017) A rating-based integrated recommendation framework with improved collaborative filtering approaches. *International Journal of Computers, Communications & Control*, 12, 307-322. <https://doi.org/10.15837/ijccc.2017.3.2692>
- Chu, Y., Pham, Y., Duong, N., & Wu, Y. (2015) *Predicting Rating of Amazon Fine Food from Reviews*. CSE 190 Assignment 2. Available online: <http://jmcauley.ucsd.edu/cse190/projects/fa15/021.pdf> (accessed on 20/03/2021).
- Cosley, D., Lam, S., Albert, I., & Konstant, J. (2003) *Is seeing believing? how recommender system interfaces affect users' opinions*. Proceedings of the CHI 2003 Conference on Human factors in Computing Systems, Ft. Lauderdale, Florida, USA, April 5-10. 585-592. <https://doi.org/10.1145/642611.642713>
- Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A.V., & Turrin R. (2011) Looking for “Good” Recommendations: A Comparative Evaluation of Recommender Systems. In *Human-Computer Interaction – INTERACT 2011. INTERACT 2011. Lecture Notes in Computer Science*; Campos P., Graham N., Jorge J., Nunes N., Palanque P., Winckler M. (Eds); Springer: Berlin, Heidelberg, 6948, 152-168. https://doi.org/10.1007/978-3-642-23765-2_11
- Dakduk, S., Horst, E., Santalla, Z., Molina, G., & Malavé J. (2017) Customer Behavior in Electronic Commerce: A Bayesian Approach. *Journal of Theoretical and Applied Electronic Commerce Research*, 12, 1-20. <https://doi.org/10.4067/S0718-18762017000200002>
- Dehuri, S., Patra, M.R., Misra, B.B., & Jagadev, A.K. (2013) *Intelligent Techniques in Recommendation Systems: Contextual Advancements and New Methods*. IGI Global: Hershey.
- Farashah, M.V., Etebarian, A., Azmi, R., & Dastjerdi, E. (2021) A hybrid recommender system based on link prediction for movie baskets analysis. *Journal of Big Data*, 8, 32. <https://doi.org/10.1186/s40537-021-00422-0>
- Feng, J., Xia, Z., Feng, X., & Peng, J. (2021) RBPR: A hybrid model for the new user cold-start problem in recommender systems. *Knowledge-Based Systems*, 214, 106732. <https://doi.org/10.1016/j.knosys.2020.106732>
- Fidan, H. (2020) Grey Relational Classification of Consumers' Textual Evaluations in E-Commerce. *Journal of Theoretical and Applied Electronic Commerce Research*, 15, 48-65. <https://doi.org/10.4067/S0718-18762020000100105>
- Gunawardana, A., Shani, G. (2009) A survey of accuracy evaluation metrics of recommendation tasks. *Journal of Machine Learning Research*, 10, 2935-2962. <https://doi.org/10.1145/1577069.1755883>

- Hahsler, M. (2013) *Recommenderlab: A Framework for Developing and Testing Recommendation Algorithms*. Available online: <https://cran.r-project.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf> (accessed on 20/03/2021).
- Herlocker, J.L., Konstan, J.A., Terveen, L.G., & Riedl, J.T. (2004) Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22, 5-53. <https://doi.org/10.1145/963770.963772>
- Hnot, T. (2016) *Recommender Systems Comparison*. Available online: <https://www.softserveinc.com/uk-ua/blog/recommender-comparison-algorithm> (accessed on 20/03/2021).
- Hooda, R., Singh, K., & Dhawan, S. (2014) A Study of Recommender Systems on Social Networks and Content-based Web Systems. *International Journal of Computer Applications*, 97, 23-28. <https://doi.org/10.5120/16996-7128>
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2011) *Recommender Systems. An Introduction*; Cambridge University: New York, USA.
- Kane, F. (2018) *Building Recommender Systems with Machine Learning and AI: Help people discover new products and content with deep learning, neural networks, and machine learning recommendations*; Sundog Education: Winter Springs, Florida.
- Khusro, S., Ali, Z., & Ullah, I. (2016) Recommender Systems: Issues, Challenges, and Research Opportunities. In *Information Science and Applications (ICISA), Lecture Notes in Electrical Engineering*, Kim K., Joukov N. (Eds); Springer: Singapore, 376, 1179-1189. https://doi.org/10.1007/978-981-10-0557-2_112
- Leng, Y.-J., Wu, Z.-Y., Lu, Q., & Zhao, S. (2021) Collaborative filtering based on multiple attribute decision making. *Journal of Experimental & Theoretical Artificial Intelligence*. <https://doi.org/10.1080/0952813X.2021.1882000>
- Li, S., Li, X. (2020) Collaborative filtering recommendation algorithm based on user characteristics and user interests. *Journal of Physics: Conference Series*, 1616, 012032. <https://doi.org/10.1088/1742-6596/1616/1/012032>
- Martins, G.B., Papa, J.P., & Adeli, H. (2013) Deep learning techniques for recommender systems based on collaborative filtering. *Expert Systems*, 37, e12647. <https://doi.org/10.1111/exsy.12647>
- McAuley, J., Leskovec, J. (2013) *From amateurs to connoisseurs: modelling the evolution of user expertise through online reviews*. Proceedings of the 22nd international conference on World Wide WebMay, Rio de Janeiro, Brazil, May, 897-908. <https://doi.org/10.1145/2488388.2488466>
- Mičík, M., Mičudová, K. (2018) Employer Brand Building: Using Social Media and Career Websites to Attract Generation Y. *Economics and Sociology*, 11(3), 171- 189. doi:10.14254/2071-789X.2018/11-3/11
- Molaei S., Havvaei, A., Zare H., & Jalili M. (2021) Collaborative Deep Forest Learning for Recommender Systems. *IEEE*, 9, 22053-22061. <https://doi.org/10.1109/ACCESS.2021.3054818>
- Poriya, A., Bhagat, T., Patel, N., & Sharma, R. (2014) Non-Personalized Recommender Systems and User-based Collaborative Recommender Systems. *International Journal of Applied Information Systems*, 6, 22-27. <https://doi.org/10.5120/ijais14-451122>
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P. (2007) *Numerical Recipes: The Art of Scientific Computing*, 3rd ed.; Cambridge University Press: New York, USA.
- Resnik, P., Iacovou, N., Suchak, M., & Bergstrom, P. (1994) *GroupLens: an open architecture for collaborative filtering of netnews*. Proceedings of the 1994 ACM conference on Computer supported cooperative work, North Carolina, USA, October 22-26, 175-181. <https://doi.org/10.1145/192844.192905>
- Ricci, F., Rokach, L., & Shapira, B. (2018) *Recommender Systems. Handbook*, 2-nd ed.; Springer: New York, USA.
- Saksiriruthai, S. (2018) Impact of media technology on wage changes: The case of Thailand. *Journal of International Studies*, 11(4), 69-78. <https://doi.org/10.14254/2071-8330.2018/11-4/5>
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001) Item-Based Collaborative Filtering Recommendation Algorithms. Proceedings of the 10th International Conference on World Wide Web, Hong Kong, May 1-5. 285-289. <http://doi.acm.org/10.1145/371920.372071>
- Schafer, J.B., Konstan, J.A., & Riedl, J. (2001) E-Commerce Recommendation Applications. *Journal of Data Mining and Knowledge Discovery*, 5, 115-153. <https://doi.org/10.1023/A:1009804230409>
- Schafer, J.B., Konstan, J.A., Riedl, J. (1999) *Recommender systems in e-commerce*. Proceedings of the 1st ACM conference on Electronic commerce, Denver, Colorado, USA, November 03-05, 158-166. <https://doi.org/10.1145/336992.337035>

- Shafiee, M.M., Bazargan, N.A. (2018) Behavioural Customer Loyalty in Online Shopping: The Role of E-Service Quality and E-Recovery. *Journal of Theoretical and Applied Electronic Commerce Research*, 13, 26-38. <https://doi.org/10.4067/S0718-18762018000100103>
- Shambour, Q. (2020) A deep learning-based algorithm for multi-criteria recommender systems. *Knowledge-Based Systems*, 211, 106545. <https://doi.org/10.1016/j.knosys.2020.106545>
- Shao, D., Tian, X., Du, S., & Zhang, T. (2017) Amazon Fine Food Recommendation System. Final Report. Available online: https://github.com/Neocryan/Amazon-Fine-Food-Recommendation/blob/master/Amazon_Fine_Food_Review.pdf (accessed on 20/03/2021).
- Silveira, T., Zhang, M., Lin, X., Liu, Y., & Ma, S. (2019) How good your recommender system is? A survey on evaluations in the recommendation. *International Journal of Machine Learning and Cybernetics*, 10, 813-831. <https://doi.org/10.1007/s13042-017-0762-9>
- Smirnov, A. V., Ponomarev, A. V. (2020) Multicriteria Context-Driven Recommender Systems: Model and Method. *Scientific and Technical Information Processing*, 47, 298-303. <https://doi.org/10.3103/S014768822005007X>
- Ujkani, B., Minkovska, D., & Stoyanova, L. (2020) A recommender system for WordPress themes using item-based collaborative filtering technique. Proceedings of the 29th International Scientific Conference Electronics, ET 2020, Sozopol, Bulgaria, 1-3. <https://doi.org/10.1109/ET50336.2020.9238305>
- Visa, M.R., Patel, D.B. (2021) A Deep Learning Approach of Collaborative Filtering to Recommender System with Opinion Mining. In *Rising Threats in Expert Applications and Solutions. Advances in Intelligent Systems and Computing*, Rathore V.S., Dey N., Piuri V., Babo R., Polkowski Z., Tavares J.M.R.S. (Eds); Springer: Singapore, 1187, 119-131. https://doi.org/10.1007/978-981-15-6014-9_15
- Vynogradova, O., Drokina, N., Yevtushenko, N., Darchuk, V., & Irtlach, M. (2020) Theoretical approaches to the definition of Internet marketing: Ukrainian dimension. *Innovative Marketing*, 16(1), 89-103. [https://doi.org/10.21511/im.16\(1\).2020.09](https://doi.org/10.21511/im.16(1).2020.09)
- Walek, B., Fojtik, V. (2020) A hybrid recommender system for recommending relevant movies using an expert system. *Expert Systems With Applications*, 158, 113452. <https://doi.org/10.1016/j.eswa.2020.113452>
- Woldan, P., Duda, P., & Hayashi, Y. (2020) Visual Hybrid Recommendation Systems Based on the Content-Based Filtering. In *Artificial Intelligence and Soft Computing. ICAISC 2020. Lecture Notes in Computer Science*, Rutkowski L., Scherer R., Korytkowski M., Pedrycz W., Tadeusiewicz R., Zurada J.M. (eds); Springer: Cham, Switzerland, 12416, 455-465. https://doi.org/10.1007/978-3-030-61534-5_41
- Wu, J., & Sang, X.; Cui, W. (2021) Semi-supervised collaborative filtering ensemble. *World Wide Web*, 24, 657-673. <https://doi.org/10.1007/s11280-021-00866-7>
- Xu, S., Raahemi, B. (2016) A semantic-based service discovery framework for collaborative environments. *International Journal of Simulation Modelling (IJSIMM)*, 15, 83-96. [https://doi.org/10.2507/IJSIMM15\(1\)7.326](https://doi.org/10.2507/IJSIMM15(1)7.326)
- Yang, C., Chen, X., Liu, L., & Sweetser, P. (2021) Leveraging semantic features for recommendation: Sentence-level emotion analysis. *Information Processing and Management*, 58, 102543. <https://doi.org/10.1016/j.ipm.2021.102543>
- Yassine, A., Mohamed, L., & Al Achhab, M. (2021) Intelligent recommender system based on unsupervised machine learning and demographic attributes. *Simulation Modelling Practice and Theory*, 107, 102198. <https://doi.org/10.1016/j.simpat.2020.102198>
- Yasynska, N., Fomichenko, I., Voloshyna, O.; Byvsheva, L., & Krikunenko, E. (2019) Assessment of the level of business readiness for digitalization using marketing and neural network technologies. *Innovative Marketing*, 15(3), 42-59. [https://doi.org/10.21511/im.15\(3\).2019.04](https://doi.org/10.21511/im.15(3).2019.04)
- Zatonatska, T., Dluhopolskyi, O., Chyrak, I., & Kotys, N. (2019) The internet and e-commerce diffusion in European countries (modelling at the example of Austria, Poland and Ukraine). *Innovative Marketing*, 15(1), 66-75. [https://doi.org/10.21511/im.15\(1\).2019.06](https://doi.org/10.21511/im.15(1).2019.06)
- Zheng, C., Zhang, Y., & Huang, Y. (2016) Rating prediction on Amazon Fine Foods Reviews. Available online: <http://www.quincyh.com/AmazonFood.pdf> (accessed on 20/03/2021).