Approach to Search Result Rankings Based on User Ratings

Viacheslav Zosimov [a], Oleksandra Bulgakova [a]

[a] *Taras Shevchenko National University of Kyiv, Bohdan Hawrylyshyn str. 24, Kyiv, 04116, Ukraine*

Abstract

This paper proposed an approach to rank search results based on user ratings, with a subjective approach to the ranking process. Expert groups unique to each user are created to achieve this effect. When no common ratings are available, the expert group is formed based on a model built on the user's social profile, utilizing inductive algorithms such as a polynomial neural network with active neurons.

The approach of ranking search results provides a unique order of elements for each user's list of web resources. This effect is achieved by using evaluations from an expert group that is unique to each user, as well as by incorporating each evaluation into the model for calculating final rankings of web resources with its unique weight, calculated based on their previous activity in the system. The resulting ranking model is much more difficult to falsify because it is based on subjective factors. The ranking model will be unique for each user. Falsifying it would require reproducing the preferences of each user, which is much more difficult than acquiring links to one's web resource from authoritative sources.

Keywords 1

Ranking, search results, user ratings, expert groups, social profile, inductive algorithms, polynomial neural network, active neurons

1. Introduction

Currently, online advertising is the most effective way of promoting a business. This has resulted in search engines no longer being a tool for obtaining information, but rather becoming advertising platforms that display to the user not the pages that are most relevant to their informational needs, but rather those for which more money has been invested in their promotion. Search engines benefit from artificially lowering the quality of organic search results, as contextual advertising, which is often irrelevant to the search query, appears more appropriate against that backdrop.

The ranking algorithms of search engines take into account a large number of factors, but the main weight is given to the page ranking, which is calculated based on the analysis of the quantity and quality of external links to the page. Such evaluation methods are objective, but they can be easily falsified with a certain advertising budget by purchasing the necessary amount of high-quality links from external sources. This implies that they are oriented towards satisfying the needs of advertisers rather than users.

In [1], described learning-based ranking model is proposed to improve recommendation systems with implicit user feedback. In [2], a ranking model is described that uses adaptive learning to improve content-based recommendation systems. In [3] develops a hybrid ranking model for scientific articles that combines content-based and citation-based methods. The ranking model using neural networks and weak supervision is described in [4-5]. It can work with incomplete data, making it more universal and convenient to use. However, the drawback is the complexity of understanding the principles of neural network operation. [6] is dedicated to using BERT (Bidirectional Encoder Representations from Transformers) for ranking in search engines. Research results have shown that

BERT provides high accuracy compared to traditional ranking methods. However, its implementation requires significant computational resources. [7] proposed a new approach to ranking that uses reinforcement learning to aggregate different page ratings. Results have shown improvement in ranking accuracy. However, the drawback is the need for significant computational resources and implementation complexity.

In [8] authors discuss sentiment analysis methods, including rule-based and machine learning-based methods, and provide examples of their applications in various domains, such as social media monitoring and product review analysis. In conclusion, the authors highlight the importance of integrating recommender systems and sentiment analysis for creating more effective and personalized recommendation systems.

The development describes in [9] is designed to improve the relevance of search within organizations and enterprises by automatically identifying and capturing the knowledge and expertise of their employees. This can provide more accurate and relevant search results by linking search queries to employees who possess the necessary expertise.

However, one potential drawback of this system is its potential inefficiency in using contextual information. It's important to consider not only the keywords but also the context in which they are used and other related factors when conducting searches. If the system doesn't account for context and can't adapt to changing user requirements, it can lead to irrelevant or incomplete results.

Brytsov R.A. researches this problem. In his work [10], he presented a theoretical ranking model based on web resource visit statistics and document viewing time. However, it does not take into account the opinions of users about the web resources, as well as the level of agreement between users and experts. Ranking methods based on user ratings are widely used for ranking products in online stores. However, they are also based solely on the number of ratings and their numerical values.

Overall, recent research has shown that the use of neural networks and reinforcement learning can improve ranking accuracy in search engines. However, these approaches require significant computational resources.

Information retrieval is the process of searching for unstructured documentary information that satisfies the user's information needs [11]. Given that user information needs are individual and subjective, this definition suggests that search result ranking algorithms should be based on subjective factors specific to each user.

Therefore, the development of methods for personalizing search engine results by providing users with search management tools and the development of new ranking models based on subjective user information needs is a pressing task.

In this paper proposed an approach to ranking search results based on user ratings. The main difference of this method is the subjective approach to the ranking process. This effect is achieved by creating expert groups unique to each user. In the absence of any common ratings, the expert group is formed based on a model built on the user's social profile using inductive algorithms, namely a polynomial neural network with active neurons.

2. Formation of expert groups

In the real task of ranking search results, user ratings are used as input data for which their status as an expert is not defined. Obviously, accepting the opinions of all users who have rated a web resource as expert would be incorrect. It is also evident that the ratings and personal data provided during registration are not sufficient to unambiguously identify a user as an objective expert in the subject area. However, this data is sufficient to determine subjective expert groups for each user based on the criterion of proximity of the user's ratings to those of each expert.

In this work, for the search query of the current user, the ranks of web resources from the search output are calculated as the weighted harmonic mean of all ratings from the members of the expert group. Expert groups for each network user are unique and are formed in the background mode of the system using three methods, depending on the presence of common ratings on a certain set of web resources between the current user and a potential expert. For convenience of presentation, experts are divided into three levels according to the method of calculating their weight. The division into levels

is only of a logical nature. The weight of experts from all levels is equivalent and is taken into account when ranking search results without additional coefficients.

To establish clarity in the definitions, the following terms are proposed within the framework of the research:

Current user U_0 - the user for whom the group of experts is formed and for whom the ranking of search results is carried out.

Expert weight - a measure of the agreement between the expert's opinion and the current user's opinion, calculated based on the similarity of their assessments for a certain set of web resources.

First-level potential expert U_i - a user who shares ratings with the current user, but whose level of agreement has not yet been calculated.

Second-level potential expert $\hat{U}_j$ - a user who does not share ratings with the current user, but shares ratings with a first-level expert.

Third-level potential expert $\tilde{U}_k$ - a user who does not share ratings with the current user or first-level experts.

2.1. Calculating the weight of first-level experts

Direct calculation can be applied in case the current user shares ratings with a set of potential experts X for some set of web resources. It allows calculating the similarity of ratings for each pair of "current user - potential expert" $d(U_0, U_i)$ separately. User ratings have values from 1 to 10, where 10 represents the most acceptable option. The expert's weight value is determined as the arithmetic mean of the absolute differences between each pair of the user's and potential expert's ratings:

$$d(U_0, U_i) = \sum_{j=1}^{m} \frac{\left|U_{0j} - U_{ij}\right|}{m} \tag{1}$$

where m is the number of common ratings that the current user has with the i-th potential expert from the set X, and j is the ordinal number of the rating. The quality of the connection is evaluated using the Shewhart chart. Candidates with high and very high connection strength are selected as experts. The weight of expert U_i relative to the current user U_0 is expedient to consider as a metric $d(U_0, U_i)$ in the metric space (X, d), where $X[U_0, \ldots, U_n]$ is the set of all system users. The function d satisfies the axioms of identity, symmetry, and triangle, but it cannot be defined for every pair of elements in the set X, since not all users have common ratings. Therefore, in the context of this problem, it is incorrect to use the term metric space, so in the future, we will refer to the function $d(U_0, U_i)$ for determining the weight of the expert as an analogue of a metric. Expert weight values are numbers within the interval $d(U_0, U_i) \in [0, \ldots, 0.9]$, therefore, to determine a qualitative assessment, the result of a normalizing function is used, rather than the value itself:

$$W(d) = 1 - \left(1,1 \cdot \frac{d(U_0, U_i)}{10}\right) \tag{2}$$

2.2. Calculating the weight of second-level experts

Forming expert groups only from those users who have common ratings with the current user significantly narrows down the circle of potential experts. To solve this problem, a method for calculating the weight of experts in the absence of common ratings with the current user has been proposed. It assumes the presence of common ratings with first-level experts and determines the overall weight of a potential second-level expert with respect to the current user $d(U_0, \hat{U}_j)$ taking into account their weight relative to the first-level expert $d(U_i, \hat{U}_j)$ and the weight of the first-level expert relative to the current user $d(U_i, \hat{U}_j)$.

The weight of second-level experts relative to the current user is expedient to calculate as the product of the weight of the first-level expert relative to the second-level potential expert who has common ratings and the weight of the first-level expert relative to the current user:

$$W(d(U_0,\hat{U}_j)) = W(d(U_i,\hat{U}_j)) \cdot W(d(U_0,U_i))$$ (3)

The expert group includes second-level potential experts whose weight relative to the current user is $W(d(U_0,\hat{U}_j)) > 0.7$.

2.3. Calculating the weight of third-level experts

The proposed method can be applied when there are a low number or absence of shared ratings, based on a model for calculating the weight of potential experts constructed from personal social profiles. The calculation of the weight of potential experts based on their past activity always yields the most accurate results. However, during the implementation and early stages of system operation, there will inevitably be situations where there is insufficient data to apply this method. Accumulating a certain base of ratings, i.e., the presence of shared ratings in the database in sufficient quantities to form expert groups for most system users, is necessary for its application. Therefore, to ensure correct system operation at an early stage of implementation, a method for determining the weight of users who have no shared ratings was developed. The user's social profile is formed based on the information they provide during registration in the system [12]. To build a model for calculating the weight of experts, a number of subjective features x_m were selected, which may directly or indirectly affect the visitor's evaluation.

For modeling the degree of consensus among experts, a generalized iterative algorithm (GIA) of inductive modeling was chosen - a polynomial network with active neurons [13].

Formally, in the general case, the GIAI for a layer r is defined:

1. The input matrix is $X_{r+1} = (y_1^r,...,y_F^r,x_1,...,x_m)$.

2. Transition operators of the form $y_l^{r+1} = f(y_i^r,y_j^r)$, $l=1,2,...,C_F^2$, $i,j=\overline{1,F}$ and $y_l^{r+1} = f(y_i^r,x_j), l=1,2,...,Fm, i=\overline{1,F}, j=\overline{1,m}$ with a quadratic partial description are applied.

3. For each description, the optimal structure $f(u,v) = a_0 d_1 + a_1 d_2 u + a_2 d_3 v + a_3 d_4 uv + a_4 d_5 u^2 + a_5 d_6 v^2$, $d_{opt} = \arg\min_{l=1,q} CR_l$, $q = 2^p - 1$, $d_k = \{0,1\}$, $f_{opt}(u,v) = f(u,v,d_{opt})$ is found;

4. The algorithm stops when $CR_{\min}^{r+1} \geq CR_{\min}^r$ and the optimal model corresponds to the value $CR_{\min}^r$ for the r-th layer. Usually, the criterion of regularity is applied:

$$AR_{B|A} = \|y_B - \hat{y}_B\|^2 = \|y_B - X_B\hat{\theta}_A\|^2,$$ (4)

based on dividing the data sample into two non-overlapping sub-samples - training set A and validation set B, $W^T = (A^T B^T)$ where $\hat{y}_B$ is the estimate of the output on B by a model, $\hat{\theta}_A$ vector of parameters of which is calculated on A.

Figure 1 represents a system that takes an input X and uses it to generate different models on next layer r. The generated models are passed to an active neuron block that optimizes their complexity using combinatorial optimization and selects F best models. The system then computes the error for each model, selects F best models again, and repeats the process until only one model is left. The final model is selected based on having the lowest error. Finally, the system outputs the value of y.

Figure 1: GIA scheme

3. Results

For comparing the results of the coherence of experts' opinions using the methods described above, a data sample was used in which 20 users evaluated 20 web resources based on the quality of presented information and ease of use. The evaluation scale ranged from 1 to 10, where 10 is the best score. Expert #0 is the current user, and a group of experts is selected based on their ratings. For ease of understanding, the web resources will be denoted by capital Latin letters with a digital index equal to the number of the data sample, and the user numbers will be represented by digits.

For each potential expert paired with user #0, a measure of consensus of opinions was calculated (Table 1).

Table 1
Measures of concordance of user #0's opinions with each of the potential experts

User	The weight value of the potential expert | correlation strength >0.7
1	0.9285 | +
2	0.8735 | +
3	0.9010 | +
4	0.9120 | +
5	0.6370 | -
6	0.9120 | +

7	0.5765 | -
8	0.8955 | +
9	0.8955 | +
10	0.9010 | +
11	0.9285 | +
12	0.9175 | +
13	0.6975 | -
14	0.8570 | +
15	0.6315 | -
16	0.5600 | -
17	0.5270 | -
18	0.7580 | +
19	0.7195 | +
20	0.7525 | +

To develop a methodology for considering the weight of second-level experts in ranking, it is necessary to choose criteria for assessing the ratio of the weight of second-level experts to the current user, expressed through the weight of first-level experts. Fig. 2-3 shows the modules of the differences in weight of second-level experts and their weights relative to the current user.

The diagram in Fig. 2 shows the dependence of the average difference in weight of second-level experts relative to the weight of first-level experts. The graph shows that the weight deviation significantly increases when the weight values of first-level experts are less than 0.8.

Potential second-level experts whose weight relative to the current user is $W(d(U_0, \hat{U}_j)) > 0.7$ are selected for the expert group.
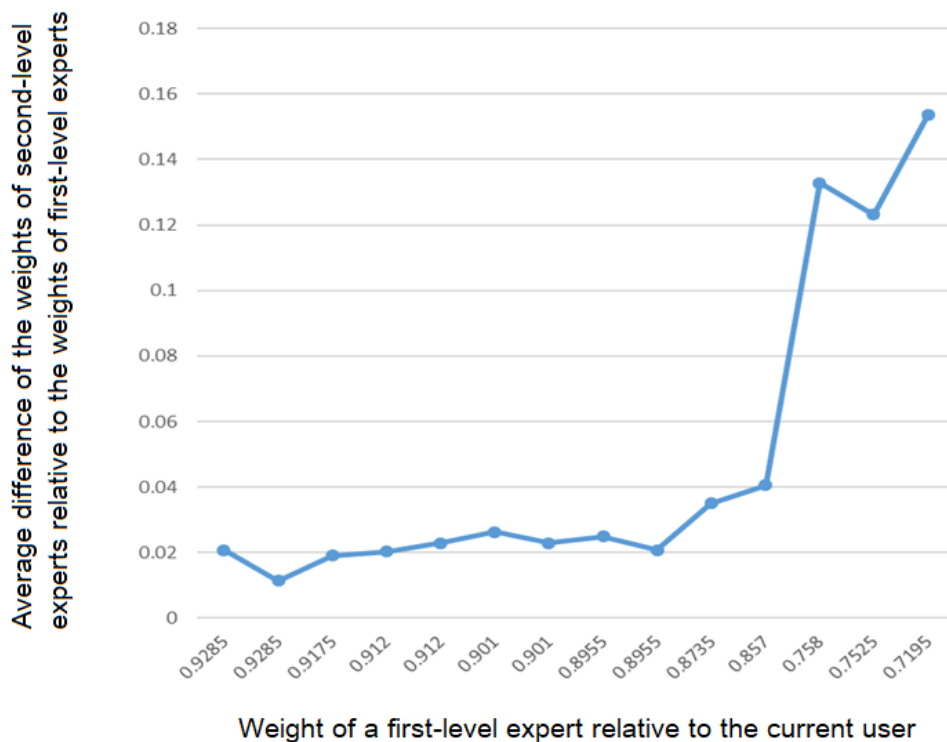

Figure 2: The dependence of the average difference in weight of a second-level user on the weight of a first-level expert
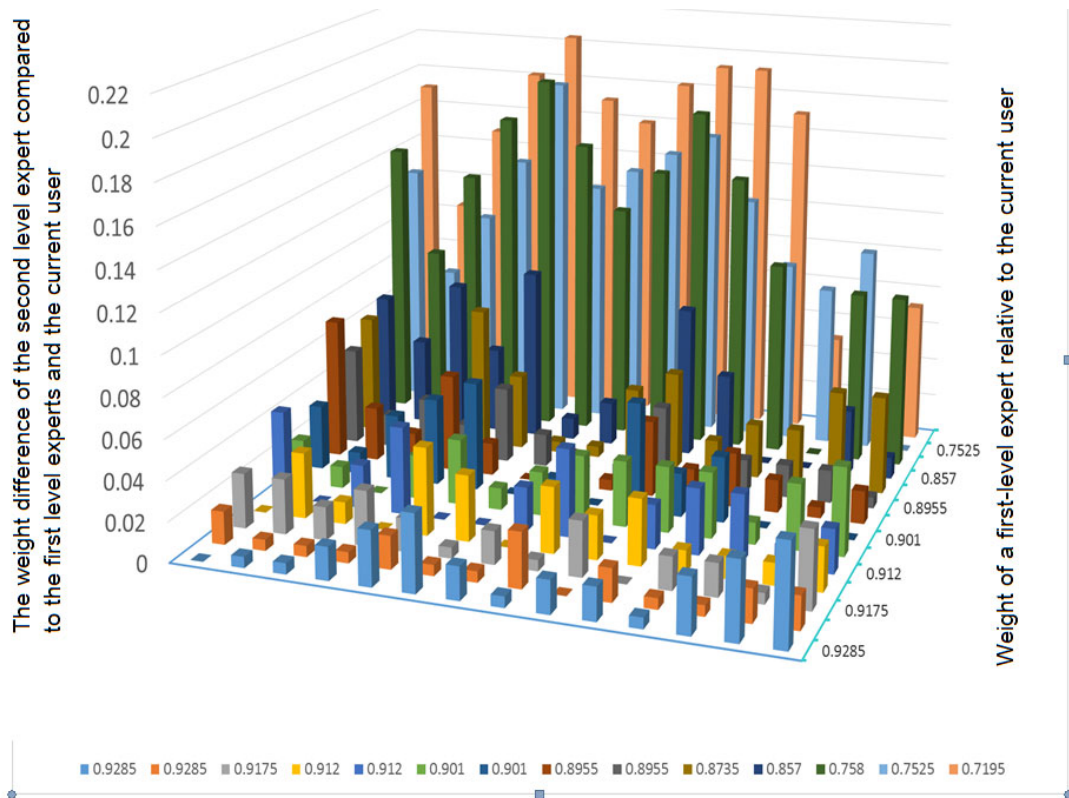
Figure 3: The weight differences in of second level users

The calculation of the weight of potential experts based on the analysis of their previous activities always gives the most accurate results. However, during the system's introduction and the period from the beginning of its operation until a certain database of ratings has accumulated, situations will inevitably arise where there is not enough data for its application. By accumulating a certain database of ratings, it is meant having a sufficient number of shared ratings in the database to form expert groups for most users of the system. Therefore, to ensure the proper functioning of the system at an early stage of its operation, the GIA algorithm is used at the third level to determine the weight of users who have no shared ratings at all. The generalized iterative algorithm is a set of iterative and iterative-combinatorial algorithms defined by the components of a vector of three index sets: DM (Dialogue Mode), IC (Iterative-Combinatorial), MR (Multilayered-Relaxative), where any iterative algorithm is defined as a specific case of the generalized UIA = {DM, IC, MR}. DM can take three values: 1 - standard automatic mode; 2 - planned automatic mode; 3 - interactive mode; IC can take two values: 1 - iterative and 2 - iterative-combinatorial algorithms; MR can take three values: 1 - classical multilayered, 2 - relaxative, 3 - combined algorithms [13].

The input data consisted of social profiles of users (U_0-U_{60}) who participated in previous experiments. Since their weight relative to user U_0 had already been calculated on the rating data samples, the same expert groups were used for building the model.

To determine in terms of users who have no common ratings with both the current user and the members of the expert group $U_{0,exp}$, we will refer to them as potential third-level experts. It is necessary to build a model of the influence of socio-personal factors of Internet users on the degree of agreement of opinions between the potential third-level expert and the current user. The sample was randomly divided into two sub-samples in the following proportions: 70% - testing, 30% - validation. The obtained model is an intermediate stage for solving the given task under time constraints. Table 2 shows the values of actual and modeled results on samples A and B for GIA. In Figure 4 the generalized iterative algorithm reaches a minimum on the 7th layer.

The proposed methodology is applied to the task of ranking search results in the search engine Google.com.ua. At first glance, this task seems very laborious, considering the large number of results, which in most cases reaches hundreds of thousands, or even several million. However, upon

closer analysis of the algorithms used by search engines, it turns out that the computational complexity of the task is several orders of magnitude lower [14]. The actual number of search results that a user has access to be much smaller than what the system claims when performing a search query.

Table 2
The values of actual and modeled results on samples A and B for GIA

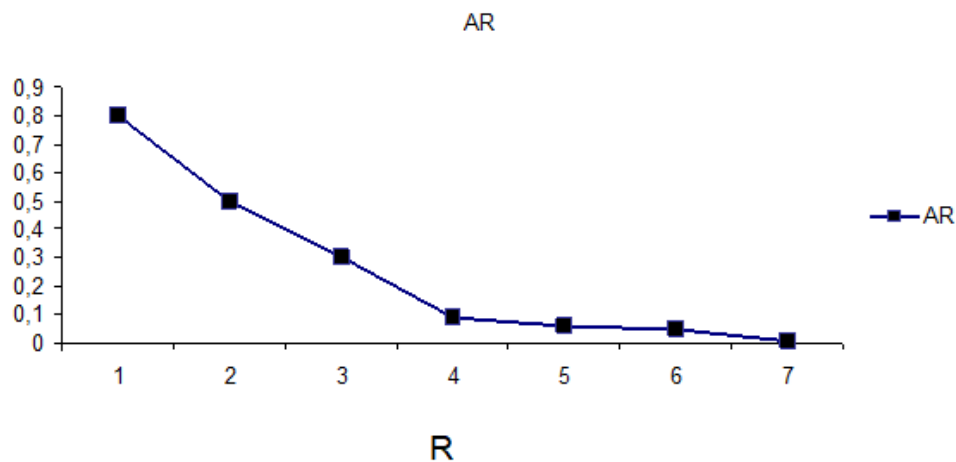| Sub-samples	Real data, y	Model data, $\hat{y}$
	1	0,904
	0,9285	0,895
	0,8735	0,752
	0,901	0,882
	0,912	0,810
	0,637	0,628
	0,912	0,865
A	0,5765	0,637
	0,8955	0,818
	0,8955	0,961
	0,901	0,911
	0,9285	0,886
	0,9175	0,928
	0,6975	0,737
	0,857	0,856
	0,6315	0,730
	0,56	0,680
B	0,527	0,574
	0,758	0,788
	0,7195	0,814
	0,7525	0,750

Figure 4: The values of the evaluation criterion (AR criterion) for the layers R

To verify the effectiveness and correctness of the methods presented above for solving the problem of ranking web resources based on user ratings, four experiments were conducted. Four data samples were formed, one for each experiment (Table 1-2). In each sample, 20 users rated 20 web resources based on the quality of the presented information and ease of use. A total of 60 users were involved in forming the four samples. Some users participated in rating web resources in more than one sample.

However, the same users were denoted with the same numbers in all four experiments, for example, user #1 in each experiment is the same person. Each experiment involved several experts from the previous experiment and new users. This was done to be able to trace the weight values of the same experts in different data samples. Before starting the ranking of web resources according to the chosen methods, it is necessary to obtain a reference ranking against which the results will be compared. Simple sorting of web resources in decreasing order of user 0's ratings only provides approximate results. Such a method only orders a group of web resources with the same ratings in decreasing order. The true order of web resources within each group remains unknown. Therefore, to construct a reference ranking, user 0 assigned a rank from 1 to 20 to each web resource, where 1 is the highest rank.

Manual ranking was used to create the reference ranking for the current user. The next step is to construct a ranking based on median values and assign new ranks. The new rank for a web resource with a unique value is equal to its position in the ordered list. In case several web resources have the same ranks based on median values, their new ranks are the arithmetic means of their positions in the ordered list. As some web resources received the same score, they are considered equivalent and are grouped into an equivalence class. To evaluate the efficiency of the aforementioned methods for calculating average scores, the average deviation from the reference ranking is used. It is calculated as the arithmetic mean of the differences between the positions of web resources.

The summarized results of the conducted experiments are presented in Table 3.

Table 3
Measures of concordance of user #0's opinions with each of the potential experts

Experiment number	Weighted harmonic mean
#1 The sum of deviations	18
Average value of deviation:	0,9
#2 The sum of deviations	24
Average value of deviation:	1.2
#3 The sum of deviations	22
Average value of deviation:	1.1
#4 The sum of deviations	20
Average value of deviation:	1
Best results	4

The last row of the Table 3 shows the number of times the determined method provided the best ranking results. If more than one method showed the best results during the experiment, they are all considered the best.

The methodology for building a personalized web ranking model based on user ratings has shown high effectiveness, indicating the potential for developing this direction further.

The developed method for ranking search results generates a unique ordering of web resources for each user. This is achieved by using ratings from expert groups unique to each user, and by assigning each rating a unique weight calculated based on the user's previous activity in the system.

The resulting ranking model is much more difficult to falsify because it is based on subjective factors. Each user will have a unique ranking model, and falsifying it would require reproducing each user's preferences, which is much more difficult than acquiring links to their website from authoritative sources.

4. Conclusions

An approach to ranking search results based on user ratings is proposed. The main difference of this method is the subjective approach to the ranking process. This effect is achieved by forming expert groups unique to each user. Experts are selected based on the degree of agreement with the current user, calculated based on the common ratings for a certain set of web resources. Selection of

users for the expert group is based on their weight relative to the current user, which is a measure of their agreement.

A methodology for forming unique expert groups for each user is proposed, which includes three levels depending on the presence of common ratings for a certain set of web resources between the current user and potential experts. The first level involves selecting potential experts based on their common ratings with the current user for a certain set of web resources. The second level involves selecting potential experts based on their agreement with the current user on a certain set of web resources. The third level involves selecting potential experts based on their weight relative to the current user, which is a measure of their agreement.

The proposed method has several advantages over traditional ranking methods. First, it takes into account the subjective nature of user preferences. Second, it allows for the formation of unique expert groups for each user, which can improve the accuracy of the ranking results. Third, it provides users with tools to manage their search results.

The next step in the development of this approach could be to construct ranking models based on a large number of factors, similar to modern search engines. By incorporating additional factors such as click-through rates, time spent on a page, and other user behavior data, it may be possible to further improve the accuracy of the ranking results.

5. References

[1] R. Yu, Q. Liu, Y. Ye, M. Cheng, E. Chen and J. Ma, "Collaborative List-and-Pairwise Filtering From Implicit Feedback". IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 6, pp. 2667-2680, 1 June (2022), doi: 10.1109/TKDE.2020.3016732.

[2] N. S. Raj and R. V G, "A Rule-Based Approach for Adaptive Content Recommendation in a Personalized Learning Environment: An Experimental Analysis". IEEE Tenth International Conference on Technology for Education (T4E), (2019), pp. 138-141, doi: 10.1109/T4E.2019.00033.

[3] Li Li, Yujie Zhang, Shuai Zhang, and Xinyu Li. A Hybrid Approach for Ranking Research Articles. Information Processing & Management (2017), vol. 53, no. 6, pp. 1407-1419.

[4] Zhaopeng Meng, Jianxun Lian, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng, "Personalized Ranking Framework via Multilayer Perceptron Learning". IEEE Transactions on Knowledge and Data Engineering (2017), vol. 29, no. 7, pp. 1509-1521.

[5] Shuai Shen, Zhiyuan Liu, Maosong Sun, and Yang Liu, "Neural Ranking Models with Weak Supervision." in Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (2019), pp. 111-120. doi:10.48550/arXiv.1704.08803

[6] M. Makarov, E. Tutubalina, P. Braslavski, "BERT for Ad-hoc Retrieval: Baselines and Analysis", in Advances in Information Retrieval - 42nd European Conference on IR Research, ECIR 2020, (2020), pp. 516-530.

[7] Yongqiang Zhang, Min Zhang, Yiqun Liu, Shaoping Ma, and Jintao Li, "A Reinforcement Learning Approach for Rank Aggregation in Information Retrieval", IEEE Access (2021), vol. 9.

[8] S.Mohammed AL-Ghuribi Shahrul, A. Mohd Noah. A Comprehensive Overview of Recommender System and Sentiment Analysis. URL: https://arxiv.org/ftp/arxiv/papers/2109/2109.08794.pdf

[9] S. Brave, R. Bradshaw, J. Jia, C. Minson. Method and apparatus for identifying, extracting, capturing, and leveraging expertise and knowledge. URL: https://patents.google.com/patent/US7698270

[10] R. Britsov Ranking of information based on user ratings and behavior. T-Comm, Telecommunications and Transport. (2016) Vol. 10. No. 1. P. 62.

[11] Christopher D. Manning. Introduction to Information Retrieval. Cambridge University Press, (2008) 520 p. URL: https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

[12] V. Zosimov, O. Bulgakova, V. Pozdeev. Semantic Profile of Corporate Web Resources. CEUR Workshop Proceedings (2021), 3179, pp. 389–397. URL: https://ceur-ws.org/Vol-3179/Short_13.pdf

[13] Bulgakova, O., Stepashko, V., Zosimov, V. "Numerical study of the generalized iterative algorithm GIA GMDH with active neurons" in Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2017, (2017), 1, pp. 496–500. doi: 10.1109/STC-CSIT.2017.8098836.

[14] Markus Spies (Ed.) Database and Expert Systems Applications. Inductive Building of Search Results Ranking Models to Enhance the Relevance of Text Information Retrieval. Los Alamitos: IEEE Computer Society, (2015) P. 291-295. doi: 10.1109/DEXA.2015.70