International Journal of Engineering

Journal Homepage: www.ije.ir

An Ensemble Click Model for Web Document Ranking

D. Bidekani Bakhtiarvand*, S. Farzi

Department of Artificial Intelligence, Faculty of Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran

PAPER INFO

Paper history:
Received 09 February 2020
Received in revised form 31 March 2020
Accepted 30 May 2020

Keywords:
Click Modeling
Document Re-ranking
Modeling Users' Behavior
Search Engine Result Page Enhancement

ABSTRACT

Annually, web search engine providers spend a lot of money on re-ranking documents in search engine result pages (SERP). Click models provide advantageous information for re-ranking documents in SERPs through modeling interactions among users and search engines. Here, three modules are employed to predict users' clicks on SERPs simultaneously, the first module tries to predict users' click behaviors using *Probabilistic Graphical Models*, the second module is a *Time-series Deep Neural Click Model* which predicts users' clicks on documents and finally, the third module is a similarity-based measure which creates a graph of document-query relations and uses *SimRank Algorithm* to predict the similarity. After running these three simultaneous processes, three click probability values are fed to an MLP classifier as inputs. The MLP classifier learns to decide on top of the three preceding modules, then it predicts a probability value which shows how probable a document is to be clicked by a user. The proposed system is evaluated on the Yandex dataset as a standard click log dataset. The results demonstrate the superiority of our model over the well-known click models in terms of perplexity.

doi: *10.5829/ije.2020.33.07a.06*

1. INTRODUCTION

Nowadays, web search engine providers invest huge amounts of money for ranking documents in SERPs according to users' satisfaction criteria [1]. Most users are willing to see search engine result pages (SERPs) which are arranged such that the more relevant documents appear in the higher ranks. Some of them never look at the second result page. Therefore, if the service provider does not want to see the users leaving the service, it should provide them with the most relevant results in the first page. Since sometimes a search engine is unable to arrange a SERP with desired ranking using the traditional methods, e.g. Page Rank, an innovative idea is to employ the collected knowledge from users' behaviors, e.g. how they interact with SERPs, when they stop clicking on documents, and amount of time they spend on documents. In other words, search engines can use the implicit feedback of users to improve the ranking of their documents.

Typically, users click on the documents that they think are more in line with their information needs. Hence, a service provider can find relevant documents more efficiently before arranging a SERP by taking the users' footprints into account.

In recent years, researchers have been encouraged to work on models based on users' behavior such as clicks and mouse movements in order to enrich search engines' qualities [2–4] or effective advertising [5]. Since the click is the most frequent user behavior in web search, most researchers consider click models to provide some useful information for ranking documents. In online advertising markets, knowing about the click possibilities on items leads to changing the priority of displayed items.

In this paper, an Ensemble Click Model, hereafter named *ECM*, is introduced based on combining sophisticated well-known click models, i.e. a Probabilistic Graphical Model based (PGM) click model named User Browsing Model (UBM) [6] and a deep

*Corresponding Author Email: *danial.bidekani@email.kntu.ac.ir* (D. Bidekani Bakhtiarvand)

Please cite this article as: D. Bidekani Bakhtiarvand, S. Farzi, An Ensemble Click Model for Web Document Ranking, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 33, No. 7, (July 2020) 1208-1213

neural network click model named NCM [2] and a structural similarity measure using SimRank [7]. The proposed system relies on the fact that each click model explores separate parts of the hypothesis space, with respect to its assumptions about users' click behaviors. Consequently, an Ensemble click model can scan a larger area of hypothesis space. Our proposed system has employed a Multi Layer Perceptron (MLP) classifier in order to predict the final decision based on UBM and NCM models outputs as Learning models, and SimRank as similarity measure.

In order to evaluate the ECM, a variety of experiments have been performed over a standard dataset called Yandex relevance dataset. The Yandex relevance dataset contains 30,717,251 unique queries and 117,093,258 documents. We followed two scenarios to evaluate the ECM in terms of effectiveness. The purpose of the first scenario is to analyze and set suitable parameters for the ECM and the second experiment attempts to compare the ECM with other well-known click models i.e. UBM, DBN (stands for Dynamic Bayesian Network) and NCM. The results imply that ECM has superiority over the well-known click models in terms of perplexity as quality metric.

The rest of the paper is organized as follows, Section 2 reviews all realted works, the proposed system is defined in Section 3. It is subdivided to four parts which are explained respectively. The experiments have been drawn and explained in Section 4 in which two research questions are discussed, and the conclusion is presented in Section 5. References are all listed in Section 6, respectively.

2. RELATED WORKS

PGM-based models have probabilistic backgrounds and try to model users' behaviors as a sequence of events. These events include attractiveness, examination, satisfaction, etc [8]. UBM and DBN are the state-of-the-art PGM-based click models [2]. The user browsing model is a well-known click model introduced by Dupret and Piwowarski [6]. The user browsing model considers the distance between last clicked document and current document examination to predict user's behavior. DBN is an extension of the Cascade Model [9] which considers a parameter for users satisfaction. The experiments show that the UBM outperforms DBN [3].

To the best of our knowledge, the first model which solved the problem of click modeling by the Neural Network approach was the Neural Click Model defined by Brisov et al. [2]. They defined several representations and also used LSTM and RNN models to train their proposed model. They reported better quality of their model over all PGM-based models and also defined another Neural Network based model using Encoder-Decoder architecture which was a little better than their previous model [10].

Researchers have applied the Convolutional Neural Network in the click modeling problem [11, 12]. There have been accomplished more specific researches on click modeling, specifically in mobile search [13] and sponsored search [14].

3. PROPOSED METHOD

We proposed an Ensemble Click Model which takes the advantages of both PGM-based and Neural network-based click models and a structural similarity-based algorithm called SimRank [7]. In particular, we paid attention to the idea of Ensemble Learning which deals to the concept that if a group of base learners attempts to learn the same problem, they can do a better classification by aggregating their viewpoints [15]. Therefore, base learners feed their decision output (a probability of how a document is likely to be clicked) to the combiner on one hand and SimRank predicts the similarity on the other hand, then the final decision will compute through either polling methods, e.g. majority voting, averaging and borda count, or applying a new classifier constructing an Ensemble Model. In contrast to most Ensemble Methods which use different datasets for every base learner, here, a single training and testing datasets have been employed. It provides us an outstanding achievement that the Ensemble Model may explore a larger subspace of the hypothesis space under the base learners' assumptions. The overall model architecture is depicted in Figure 1. The proposed model should pass four phases in order to train. According to Figure 1, in the undersampling phase, 50000 search sessions are selected from Yandex relevance dataset. In the Encoding phase, data is prepared in a way that it is suitable to base learners. Then in the base learners' training phase, each base learner will learn a hypothesis through its assumptions. Before executing last phase, the training set will be tested by each base learners. Then they return a probability of clicking on documents. These probabilities are fed to an MLP classifier as its attributes.

3. 1. Undersampling Phase In 2011, Yandex published a dataset of its search engine which contained users' clicks history. Because of the computation limitation and unavailability of powerful resources, we had to undersample a set of sessions from Yandex relevance dataset by uniform random sampling. Yandex dataset contains ten documents for each query, however we consider the first six documents in this paper. Table 1 shows the general information of the undersampled dataset. Thus, in the first place, it needs to be shown that the undersampled dataset is a good representative of the whole Yandex dataset. In this regard, Table 2 represents

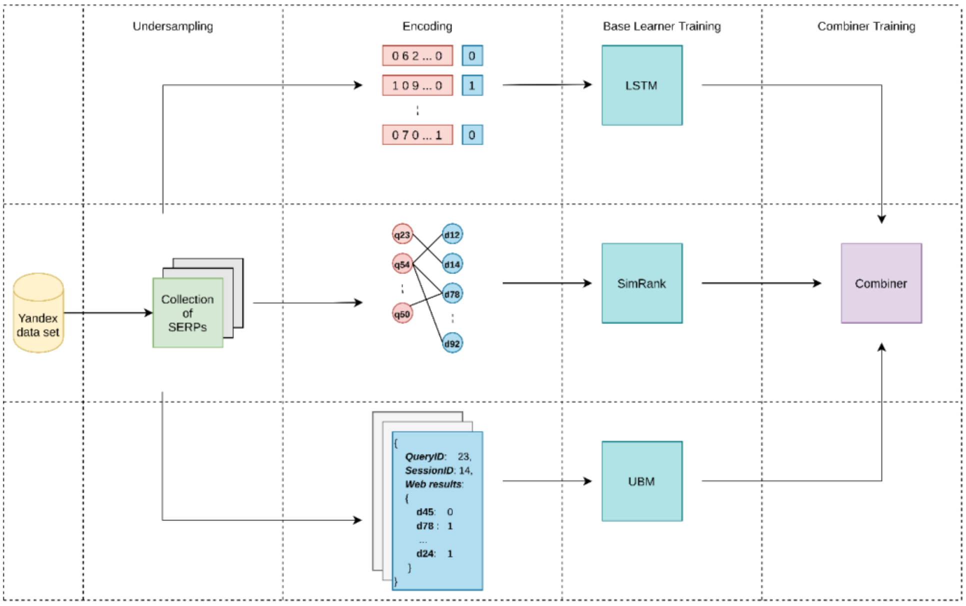

Figure 1. Ensemble click model training architecture

the click frequency ordered by documents ranking. It shows the normal behavior of users intuitively, because of the numerous clicks at the top ranks is the most seen users' behavior and the number of clicks has a descending order which illustrates that the probability of clicking on a document reduces by increasing its rank [16]. In other words undersampled dataset shows users' intention to click on the top documents in SERP which is the normal

behavior of search engine users. Figure 2 shows the users' intent on the sampled dataset. Furthermore, Table 3 shows the number of clicks in each query session of undersampled dataset. From the table, it can be understood that the session which contains less clicks on the results pages are in the majority.

3. 2. Encoding Phase Before each click model learns its hypothesis space, it is required to preprocess dataset in a suitable form according to the intended model. As observed in Figure 1 the encoding phase contains NCM transformation, UBM transformation and SimRank transformation.

Before each click model learns its hypothesis space, it is required to preprocess dataset in a suitable form.

The first representation should be in a vector form to feed in the NCM model. In order to transform dataset to the vector forms, a vector of 2^{SERP} is considered for every <query, document, rank> triplet. Here, the given SERP size is six, so every <query, document, rank> triplet creates a vector of 2^6, because in a six document SERP, there exist 2^6 different click patterns. A click pattern is a binary vector that its value shows the click or skip on the documents. To transform dataset to the vector forms, a vector of 2^6 is created for <query, document, rank> triplet and the search is began for SERPs which include the same <query, document, rank> triplet. Every SERP has a click pattern, it turns the click pattern binary value to an integer, and at the end it adds one unit to the corrosponding vector index. To make a more informative vector, the user interaction will be added at the end of the vectors, e.g. if the user clicks on the previous document, 1 will be appended to the vector representation otherwise 0. To create query

TABLE 1. Sampled dataset information

Item	Value
Search session size	50000
Query session size	166149
Train set query sessions size	124611
Test set query sessions size	41538
Unique query of train set query sessions size	62702
Unique query of test set query sessions size	22383

TABLE 2. Clicks and skips frequency in dataset

Rank	#Clicks	#Skips
1	76693	89456
2	32434	133715
3	21345	144804
4	16104	150045
5	12582	153567
6	10228	155921
Sum	169386	827508

representation, it is only needed to aggregate all vectors which have the same query [2].

This representation of UBM model includes query ID, session ID and the user history on each SERP whether it is a click or a skip.

In order to apply SimRank algorithm as a similarity measure of the ECM, data should be transformed into a graph form. This transformation creates a bipartite graph which contains two different node types; query node and document node. An edge is generated whenever at least there exist a click on a query and a document. The represented bipartite graph is weighted by the Click Through Rate (CTR). CTR for a query document pair is defined as the number of clicks when *q* as query and *d* as document appear together and receive a click event over the times that *q* and *d* appears together despite the event type (click or skip).

3. 3. Base Learner Training Phase

3. 3. 1. NCM Model One of the base learners has the LSTM structure. It is shown that LSTM is an effective model to learn the sequences [17]. Since the click modeling is a sequential problem, it is a good intuition to learn a model based on the LSTM structure [2] called *NCM*. It considers query vector at first, then it predicts whether the user will click on the first document or not. Following the user's interactions with the first document, the model predicts on the second document, the considering query representation at first, and first document representation as next. This process continues until the last document in SERP.

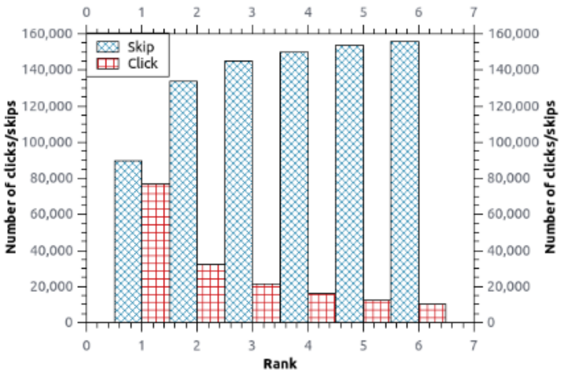

Figure 2. Click and Skip frequency @Rank

TABLE 3. Clicks per session frequency in the dataset

#Session	#Clicks
0	55498
1	77058
2	18804
3	8152
4	3883
5	1792
6	962
Sum	166149

3. 3. 2. UBM Model The second learner is the UBM model which is a well-known PGM-based click model. UBM is an extension of the PBM model that considers the last clicked document in its assumption. It depends on the current document and the last clicked document ranks. This model defines two parameters, *examination parameter* which depends on the document rank and on the previously clicked document rank, and *Attractiveness parameter* which depends on a query and a document. Both examination and attractiveness parameters are learned by the Expectation Maximization (EM) algorithm [6].

3. 3. 3. Similarity-based Model The third part of the ECM is different from the others. The similarity-based model tries to find the similarity between a query node and a document. It considers the SimRank algorithm as a similarity measure to predict a link between a query and a document. The idea behind SimRank is that the two objects are similar if they are referenced by similar objects. SimRank formula is as follows:

$$sim(q,d) = \frac{c}{|N(q)|.|N(d)|} \sum_{q \in N(q)} \sum_{d \in N(d)} sim(q,d) \qquad (1)$$

where *sim(q, d)* denotes the similarity between a document and a query which is initialized by the CTR measure. An object has the most similarity to itself, therefore *sim(q, q)* = 1 and *sim(d, d)* = 1.

3. 4. Combiner Training Phase After the third phase, the base learner training phase, it is necessary to design a model which takes the base learners' output and learns to decide based on the base learners' opinions. MLP is a supervised learning algorithm which can learn a hypothesis spase based on gradient descent algorithm. As shown in Figure 1, MLP acts as a combiner learning in the presented model to aggregate the results of UBM, NCM and SimRank. The combiner takes the output from base learners as its input. Base learners output are similar in the context of domain. The domain is limited to a value in [0, 1]. After training the combiner model, when the model is required to predict on a <query, documen, rank> triplet, first the model transforms the input in the three different representations type as we discussed in Section 3.2, then each representation is fed to base learners as input, and three outputs comes from the base learners training phase. Next, these outputs are fed to the MLP classifier as input and in the end of the process, combiner decides on the base learners decision and predicts that the user finally clicks on the document or not.

4. EXPERIMENTS

In this section, we have provided two experiments to evaluate the model quality. Then we discuss the validity of the results and the superiority of ECM over the well-known click models.

In order to measure the quality of each click model, we used perplexity metric [6] to compare the accuracy of each click models. Perplexity is the best known metric in the context of click modeling. The perplexity measure for the model M on a set of sessions S in the rank r is calculated by Formula (2). The total perplexity is calculated by averaging over all ranks [2].

$$p_r(Model) = 2^{-\frac{1}{|S|}\sum_{s\in S}(c_r^{(s)}\log_2 q_r^{(s)}+(1-c_r^{(s)})\log_2(1-q_r^{(s)}))} \quad (2)$$

Before addressing the results, it is a good idea to have an overall view to see how experiments are done. Because one of the three base learners is based on a Deep Neural Network, the results will be somewhat different for each execution. To make certain that the results make sense, we executed the NCM model over the training dataset 10 times. The results of 10 runs are included in Table 4. Because the execution of NCM model is required to execute the combiner model, we executed the combiner model 10 times as the same way to be sure of the results. The results for the combiner model executions are shown in Table 5.

4. 1. Experiment 1

After running the MLP classifier as the combiner model with different neural network structures, we came up with a two-layer neural network containing 20×5 neurons. The results for the combiner are depicted in Table 5. As it is shown in the table, the standard deviation is not noticeable which conforms the validity of the results.

TABLE 4. Results of 10 times execution of NCM

Run	@1	@2	@3	@4	@5	@6
1	1.789	1.519	1.423	1.384	1.289	1.249
2	1.811	1.545	1.462	1.335	1.299	1.242
3	1.800	1.539	1.390	1.329	1.272	1.236
4	1.719	1.560	1.468	1.360	1.290	1.208
5	1.729	1.514	1.368	1.320	1.298	1.231
6	1.834	1.512	1.405	1.317	1.286	1.237
7	1.688	1.537	1.426	1.329	1.298	1.231
8	1.703	1.510	1.384	1.355	1.298	1.245
9	1.703	1.542	1.448	1.323	1.285	1.244
10	1.673	1.493	1.399	1.313	1.272	1.246
AVG	1.745	1.527	1.419	1.338	1.290	1.239
STD DEV	0.057	0.020	0.033	0.022	0.010	0.011

TABLE 5. Results of 10 times execution of the Combiner

Run	@1	@2	@3	@4	@5	@6
1	1.578	1.052	1.027	1.011	1.008	1.004
2	1.495	1.035	1.025	1.013	1.007	1.004
3	1.421	1.026	1.031	1.012	1.008	1.003
4	1.519	1.057	1.031	1.014	1.009	1.005
5	1.505	1.053	1.029	1.012	1.007	1.004
6	1.420	1.034	1.022	1.011	1.006	1.004
7	1.420	1.036	1.025	1.010	1.005	1.003
8	1.501	1.031	1.032	1.014	1.007	1.006
9	1.544	1.051	1.030	1.013	1.008	1.003
10	1.530	1.036	1.026	1.011	1.006	1.003
AVG	1.493	1.041	1.028	1.012	1.007	1.004
STD DEV	0.055	0.011	0.003	0.001	0.001	0.000

4. 2. Experiment 2

After finding the best structure for the MLP classifier as the combiner of the proposed model, the model perplexity was measured and the results are depicted in Figure 3.

In order to compare the ECM with previous models, it is a good idea to draw the average perplexity of NCM, ECM and UBM models altogether. As it can be understood from Figure 3, the ECM has succeeded in all ranks and average perplexity. To ensure that these results are meaningful, the t-test experiment has been taken and the ECM has superiority over UBM (p-vlaue = 0.00010249401468) and NCM (p-value = 3.01806922588773E-05).

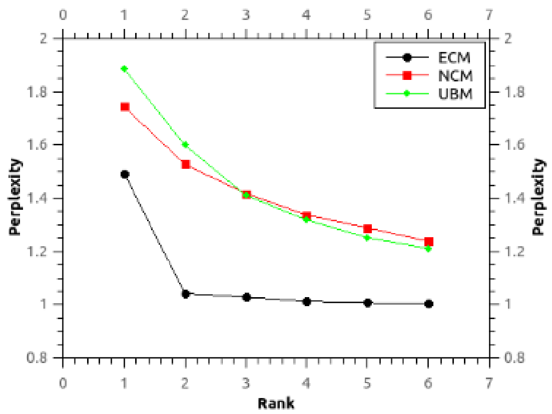

Figure 3. Comparison of the Click Models in terms of perplexity

5. CONCLUSION

In this paper, a new click model called ECM was introduced and it was shown that it has better performance than the well-known click models. The proposed model consists of a PGM-based click model called UBM and a Neural network based click model called NCM. An MLP classier is employed to decide based on UBM and NCM click models output. The superiority of ECM has been shown with experiments based on perplexity measure.

6. REFERENCES

1. Ghose, A., Ipeirotis, P. G., and Li, B., "Examining the impact of ranking on consumer behavior and search engine revenue", ***Management Science***, Vol. 60, No. 7, (2014), 1632–1654. https://doi.org/10.1287/mnsc.2013.1828
2. Borisov, A., Markov, I., De Rijke, M. and Serdyukov, P., "A neural click model for web search", In Proceedings of the 25th International Conference on World Wide Web, (2016), 531–541. https://doi.org/10.1145/2872427.2883033
3. Grotov, A., Chuklin, A., Markov, I., Stout, L., Xumara, F. and de Rijke, M., "A comparative study of click models for web search", In International Conference of the Cross-Language Evaluation Forum for European Languages, (2015), 78–90. https://doi.org/10.1007/978-3-319-24027-5_7
4. Liu, Z., Mao, J., Wang, C., Ai, Q., Liu, Y. and Nie, J.Y., "Enhancing click models with mouse movement information", ***Information Retrieval Journal***, Vol. 20, No. 1, (2017), 53–80. https://doi.org/10.1007/s10791-016-9292-4
5. Aouad, A., Feldman, J., Segev, D. and Zhang, D., "Click-based MNL: Algorithmic frameworks for modeling click data in assortment optimization", Available at SSRN 3340620, (2019). http://dx.doi.org/10.2139/ssrn.3340620
6. Dupret, G. E. and Piwowarski, B., "A user browsing model to predict search engine click data from past observations", In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, (2008), 331–338. https://doi.org/10.1145/1390334.1390392
7. Jeh, G. and Widom, J., "SimRank: a measure of structural-context similarity", In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, (2002), 538–543. https://doi.org/10.1145/775047.775126
8. Guo, F., Liu, C., and Wang, Y. M., "Efficient multiple-click models in web search", In Proceedings of the second acm international conference on web search and data mining, (2009), 124–131. https://doi.org/10.1145/1498759.1498818
9. Chapelle, O. and Zhang, Y., "A dynamic bayesian network click model for web search ranking", In Proceedings of the 18th international conference on World wide web, (2009), 1–10. https://doi.org/10.1145/1526709.1526711
10. Borisov, A., Wardenaar, M., Markov, I. and de Rijke, M., "A click sequence model for web search", In the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, (2018), 45–54. https://doi.org/10.1145/3209978.3210004
11. Liu, Q., Yu, F., Wu, S. and Wang, L., "A convolutional click prediction model", In Proceedings of the 24th ACM international on conference on information and knowledge management, (2015), 1743–1746. https://doi.org/10.1145/2806416.2806603
12. Ni, Z., Ma, X., Sun, X. and Bian, L., "A Click Prediction Model Based on Residual Unit with Inception Module", In Pacific Rim International Conference on Artificial Intelligence, (2019), 393–403. https://doi.org/10.1007/978-3-030-29911-8_30
13. Zheng, Y., Mao, J., Liu, Y., Luo, C., Zhang, M. and Ma, S., "Constructing Click Model for Mobile Search with Viewport Time", ***ACM Transactions on Information Systems (TOIS)***, Vol. 37, No. 4, (2019), 1–34. https://doi.org/10.1145/3360486
14. Zhang, Y., Dai, H., Xu, C., Feng, J., Wang, T., Bian, J., Wang, B. and Liu, T.Y., "Sequential click prediction for sponsored search with recurrent neural networks", In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, (2014). https://arxiv.org/abs/1404.5772
15. Dietterich, T. G., "Ensemble methods in machine learning", In International workshop on multiple classifier systems (pp. 1-15). Springer, Berlin, Heidelberg, (2000), 1–15. https://doi.org/10.1007/3-540-45014-9_1
16. Zhang, Y., Chen, W., Wang, D. and Yang, Q., "User-click modeling for understanding and predicting search-behavior", In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, (2011), 1388–1396. https://doi.org/10.1145/2020408.2020613
17. Sak, H., Senior, A. W., and Beaufays, F., "Long short-term memory recurrent neural network architectures for large scale acoustic modeling", In 15th Annual Conference of the International Speech Communication Association, Singapore, (2014), 338-342. https://www.isca-speech.org/archive/interspeech_2014/i14_0338.html

Persian Abstract

چکیده

به طور سالانه، شرکت‌های فراهم‌کننده‌ی سرویس موتور جست‌وجو، برای رتبه‌بندی مجدد اسناد در صفحه‌های جست‌وجو، مبالغ زیادی هزینه می‌کنند. کلیک‌مدل‌ها از طریق تعاملات کاربران با موتورهای جست‌وجو اطلاعات سودمندی برای رتبه‌بندی مجدد اسناد در صفحه‌های نتایج جست‌وجو فراهم می‌کنند. در این مقاله، به منظور پیش‌بینی کلیک کاربران بر روی صفحه‌های نتایج جست‌وجو، از سه ماژول به طور همزمان استفاده شده است، نخستین ماژول سعی در پیش‌بینی کلیک‌های کاربران، با استفاده از مدل‌های گرافی احتمالی دارد، ماژول دوم بر اساس شبکه‌های عصبی عمیق مختص سری‌های زمانی کلیک کاربران روی اسناد را پیش‌بینی می‌کند و در آخر، ماژول سوم که از یک معیار شباهت به نام SimRank بر روی گرافی از روابط کلیک–سند استفاده می‌کند. پس از اجرای همزمان این سه ماژول، سه مقدار احتمالاتی به دست آمده به عنوان ورودی‌های یک شبکه‌ی عصبی پرسپترون چندلایه استفاده می‌شود. شبکه‌ی عصبی پرسپترون چندلایه آموزش می‌بیند که روی تصمیم سه ماژول پایه، تصمیم بگیرد، سپس یک مقدار احتمالاتی به عنوان احتمال کلیک‌خوردن یک سند توسط کاربر پیش‌بینی کند. مدل ارائه‌شده با استفاده از مجموعه‌داده‌ی موتور جست‌وجوی Yandex ارزیابی شده است و نتایج حاکی از برتری این مدل نسبت به مدل‌های شناخته‌شده‌ی گذشته دارد.