World Wide Journal of Multidisciplinary Research and Development

WWJMRD 2017; 3(8): 1-6
www.wwjmrd.com
International Journal
Peer Reviewed Journal
Refereed Journal
Indexed Journal
UGC Approved Journal
Impact Factor MJIF: 4.25
e-ISSN: 2454-6615

M. Sornamageswar
PG&Research Department of Computer Science, Govt Arts College, Coimbatore. India

S. Geetharani
Assistant Professor
Department of Computer Science, PSG College Of Arts and Science, Coimbatore. India

A Design and Implementation of Ranking Search Results Based on LC2R and LBRM

M. Sornamageswar, S. Geetharani

Abstract
Personalized search is an essential research area that has main goal to determine the uncertainty of query terms. In order to enhance the relevance of search results, personalized search engines form user profiles which is captures the users' personal preferences and by using those preference find out the actual goal of the input query. By using User profile we can rank the documents in a search engine according to the query which is submitted by user. A better user profiling strategy is an important and primary component in search engine personalization. Many researches are developed for the personalized search based on the user profile in recent years. Our early work, proposed a novel technique i.e., Link-Click-Concept based Ranking Algorithm. In this work, The Link-Click-Concept user profiling strategy extracts user's conceptual preferences from user's click through data resulted from the web search. These preferences are used to rank the pages in a search engine. The LC2R Method is robust, and generic but the query results vary from location to location. Then this research adopts LC2R method in the area of location. The user location is a major factor for determining search results relevancy. So a Location-based Ranking Method (LBRM) is proposed to know the location effects of the queries. It ranks the search results by considering the similarity between the location and retrieved pages. In addition, users' locations (positioned by GPS) are used to complement the location concepts in LBRM.

Keywords: Personalized search, User profile, Ranking Algorithm, LBRM and Search engine

Introduction
The Link-Click-Concept user profiling strategy extracts user's conceptual preferences from user's click through data resulted from the web search. These preferences are used to rank the pages in a search engine. The LC2R Method is robust and generic but the query results vary from location to location. Then this research adopts LC2R method in the area of location. The user location is a major factor for determining search results relevancy. So a Location-based Ranking Method (LBRM) is proposed to know the location effects of the queries. It ranks the search results by considering the similarity between the location and retrieved pages. In addition, users' locations (positioned by GPS) are used to complement the location concepts in LBRM.
The similarity value between the locations and retrieved pages is computed and derived via two databases Page-location Database and Location-page Database for similarity identification. The frequent retrieval patterns are retrieved by calculating the support value. The support value represents the frequent occurrence of retrieved pages from the particular location. Based on the frequent occurrences the weighted score is estimated for the patterns and the search patterns are ranked based on the highest weighted score. Hence the proposed LBRM rank the pages based on the location effect of the queries which is more significant for improving the performance of search engine.

Related Work
To rank the search more efficiently based on location effects in the search engine Location-based Ranking Method (LBRM) is proposed. Users submitted their queries in different location and they retrieve different results. In this research work location of user is also considered to rank the search results. This research work consists of three modules. Initially

Correspondence:
M. Sornamageswar
PG & Research Department of Computer Science, Govt Arts College, Coimbatore. India

user submitted their queries from a particular location and the user locations are identified by the geographic information and obtain the user locations. Then the similarity is determined between the user location and retrieved pages. For the identification of similarity two databases namely Page Location Database (PLD) and Location Page Database (LPD). The computation of support values retrieves the frequent retrieval patterns. This support values represent the frequent occurrences of the retrieved pages from a particular location. Based on the support values the weight of web page is assigned and the web pages are ranked according to the highest weighted score.
The main aim of the similarity identification process is to compute the similarity between the location and retrieved pages automatically from the database. The database consists of various information including for a particular page which location is retrieved this page and for a given location, identified which pages are retrieved. This information is utilized to determine which location and pages are similar. The geographic information is used to find the particular location of a user. A similarity score is assigned to every pair of pages or locations.

A. LBRM User Profiling Strategy
Input:
Page Set$P = \{p_1, p_2, \dots p_k\}$, Query set $Q = \{q_1, q_2, \dots q_n\}$, Location set of the users $L = \{l_1, l_2, \dots l_m\}$

Output:
Ranking of search results
Step 1: Using the geographic information the location is determined for the queries

Step 2: Calculate the similarity among the retrieved pages and location

Step 3: Calculate the location similarity

$$sim(l_1, l_2) = \frac{\sum_{\rho \in \Gamma_{l_1}} MaxSim(\rho, \Gamma_{l_1}) + \sum_{\varrho \in \Gamma_{l_2}} MaxSim(\varrho, \Gamma_{l_2})}{|\Gamma_{l_1}| + |\Gamma_{l_2}|}$$

Γ_{l_1}andΓ_{l_2}represents the two page sets, $MaxSim(e, E) = Max_{e' \in E} sim(e, e')$ indicates the maximum similarity among e and element in E.
Step 4: Calculate the page similarity

$$sim(p_1, p_2) = 1 - \frac{\sum_{\beta \in \Omega_{p_1}} MinSim(\beta, \Omega_{p_2}) + \sum_{\omega \in \Omega_{p_2}} MinSim(\omega, \Omega_{p_1})}{|\Omega_{p_1}| + |\Omega_{p_2}|}$$

Ω_{p_1}andΩ_{p_2} denotes the set of pages, $MinSim(e, E) = Min_{e' \in E} sim(e, e')$represents the minimum similarity between e and E.
Step 5: Create the patterns that includes retrieved pages$P = \{p_1, p_2, \dots p_k\}$, $Q = \{q_1, q_2, \dots q_n\}$ and user locations$L = \{l_1, l_2, \dots l_m\}$
Step 6: Calculate the support value to find the frequent access pattern

$$Sup(P) = \frac{\sigma(l_i, p_j)}{|G|}$$

G denotes the set pair with the pages and locations. The frequency of occurrence of the search of a particular location is denoted by $\sigma(l_i, p_j)$. The frequent access pattern is determined by comparing the patterns with the minimum support value. The minimum support value is greater than the pattern then is termed as frequent access pattern.

Step 7: Calculate the pattern similarity

$$ps(P_u, P_v) = \begin{cases} 0, & if\ u < 1\ or v < 1 \\ sim(l_1, l_2) + sim(p_1, p_2) & otherwise \end{cases}$$

The similarity among the locations l_1 and $l_2$is indicated as $sim(l_1, l_2)$ and the similarity among the pages $p_1$and $p_2$is denoted by $sim(p_1, p_2)$.
Step 8: Calculate the weighted score for the pattern

$$WS(P_u, P_v) = \sum_{j=0}^{\max(m,n)-1} ps(P_{u_{m-j}}, P_{v_{n-j}}) \times w_{\max(m,n)-j} \times \sup(P_u)$$

$ps(P_u, P_v)$denotes the pattern similarity and W represents the weight of the pattern. Based on the weight score the pages are ranked.

Experimental Results
In this section the performance of the LC2R methods is evaluated and compared with the Location Based Ranking method (LBRM) with different user profiling strategies. The LC2R is compared with the LBRM in terms of precision, recall, precision-recall, F-Measure,ΔSimilarity, Average Relevant Rank and Exponential Ration.

Precision
Precision is briefly explained in Chapter 3. The comparison of the precision and recall values LC2R and LBRM methods with different profiling strategies are given in Table 3.

Table 3.1: Comparison of Precision and Recall

No.of queries	Precision						Recall					
	LC2R			LBRM			LC2R			LBRM		
	Joachisms	mJoachisms	SpyNB	Joachisms	mJoachisms	SpyNB	Joachisms	mJoachisms	SpyNB	Joachisms	mJoachisms	SpyNB
5	0.8953	0.8860	0.9073	0.9478	0.9271	0.9629	0.8897	0.8804	0.8919	0.9359	0.9267	0.9568
10	0.8728	0.8558	0.8841	0.9186	0.9003	0.9458	0.8634	0.8539	0.8689	0.9112	0.9009	0.9356
15	0.8410	0.8346	0.8600	0.8980	0.8733	0.9156	0.8362	0.8321	0.8509	0.8855	0.8766	0.9085
20	0.8171	0.8063	0.8339	0.8749	0.8457	0.8930	0.8119	0.8024	0.8322	0.8709	0.8433	0.8899
25	0.8091	0.7921	0.8168	0.8691	0.8347	0.8810	0.7875	0.7740	0.8084	0.8455	0.8232	0.8698

Figure 3.1 shows that the comparison of retrieval precision and recall values made between LC2R and LBRM with different profiling strategies.

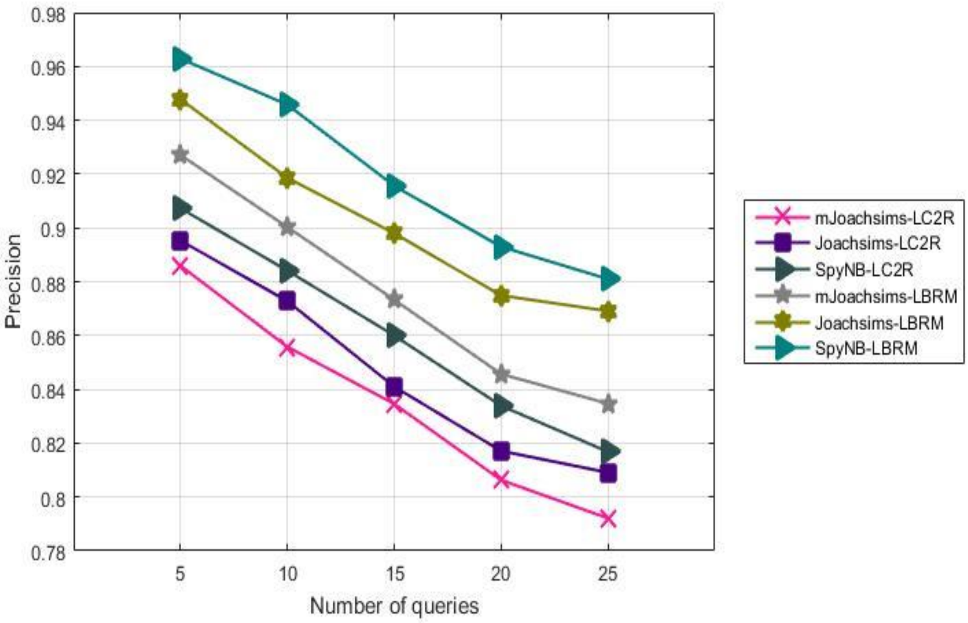

Fig. 3.1: Comparison of Precision

Figure 3.1 shows that comparison of precision between LC2Rand LBRM with different profiling strategies. X axis be elected by the number of queries and Y axis be elected by precision value. In figure 3.1, if the number of queries is 5, the retrieval precision of LC2R with SpyNB ranking approach is 0.9073 and LBRM with SpyNBis 0.9629. While the number of queries is 10 the corresponding retrieval precision of LC2RSpyNB is 0.8841 and LBRM withSpyNB is 0.9458. The retrieval precision of LC2R with SpyNB is 0.8600 andLBRM with SpyNB is 0.9156 while number of queries is 15. If number of queries is 20, then the respective retrieval precision of LC2R with SpyNB is 0.8339 andLBRM with SpyNB is 0.8930. The retrieval precision of LC2R with SpyNB is 0.8168 andLBRM with SpyNB is 0.8810 when subset size is 25. This result illustrates that the LBRM with SpyNBhas high retrieval precision.

Recall

Figure 3.2 shows that the comparison of retrieval recall values made between LC2R and LBRM with different profiling strategies.

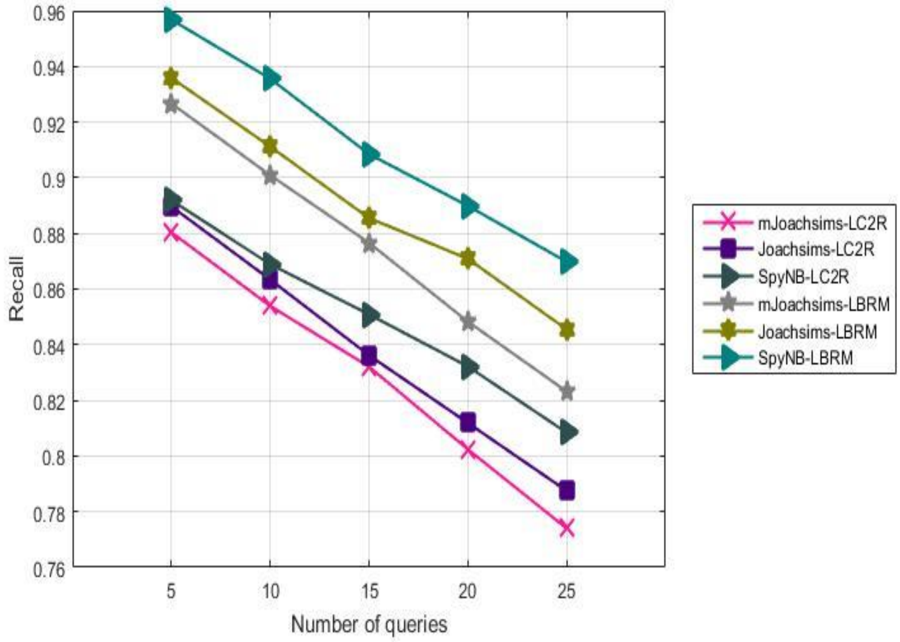

Fig. 3.2: Comparison of Recall

Figure 3.2 shows that comparison of recall between LC2R Ranking and LBRMwith different profiling strategies. X axis be elected by the number of queries and Y axis be elected by recall value. In figure 3.2, if the number of queries is 5, the retrieval recall of LC2R with SpyNB ranking approach is 0.8919 and LBRM with SpyNB is 0.9568. While the number of queries is 10 the corresponding retrieval recall of LC2R with SpyNB is 0.8689 and LBRM with SpyNB is 0.9356. The retrieval recall of LC2R with SpyNBis 0.8509 andLBRM with SpyNB is 0.9085 while number of queries is 15. If number of queries is 20, then the respective retrieval recall of LC2R with SpyNB is 0.8322 andLBRM with SpyNB is 0.8899. The retrieval recall of LC2R with SpyNB is 0.8084 andLBRM with SpyNB is 0.8698 when subset size is 25. This result illustrates that the LBRM with SpyNBhas high retrieval recall.

Precision- Recall

The connection among the precision and recall may be located. It is in briefly explained in chapter 3.The contrast of the precision-recall values for LC2R and LBRM with different profiling strategies are given in Table 3.3.

Table 3.3: Comparison of Precision-Recall

Recall	Precision					
	LC2R			LBRM		
	Joachisms	mJoachisms	SpyNB	Joachisms	mJoachisms	SpyNB
0.2	0.8923	0.8805	0.9083	0.9486	0.9202	0.9608
0.4	0.8778	0.8639	0.8873	0.9294	0.9043	0.9520
0.6	0.8535	0.8472	0.8652	0.9066	0.8856	0.9306
0.8	0.8359	0.8290	0.8450	0.8957	0.8661	0.9128
1	0.8189	0.8007	0.8206	0.8783	0.8444	0.8939

Figure 3.3 shows that the comparison of retrieval precision-recall values made between LC2R and LBRM with different profiling strategies.

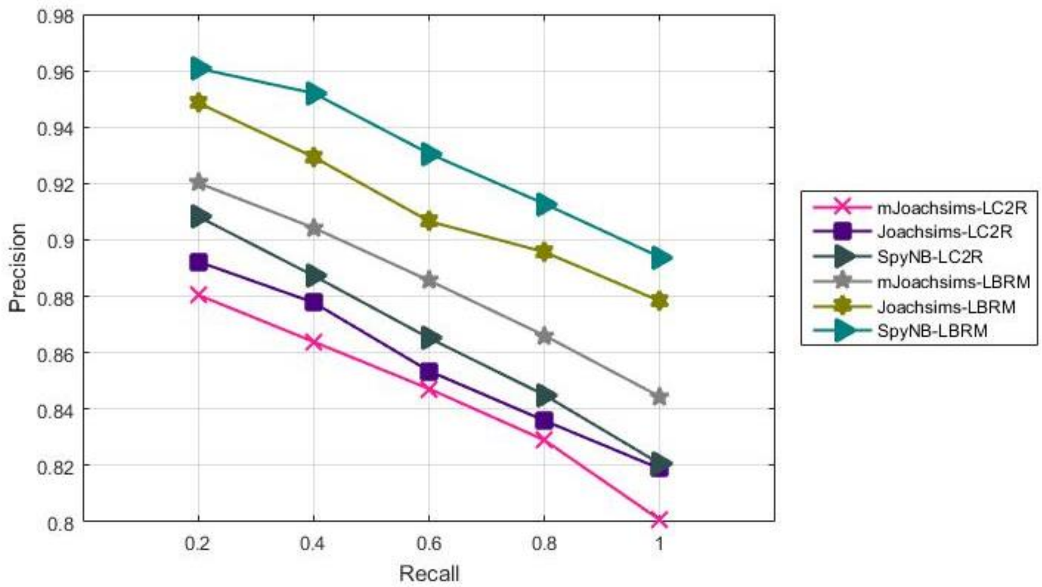

Fig. 3.3: Comparison of Precision-Recall

Figure 3.3 shows that comparison of precision-recall between LC2R and LBRM with different profiling strategies. X axis represents the precision and Y axis represents recall value. In figure 4.3, if the precision is 0.2, the retrieval recall of LC2R with SpyNB ranking strategy is 0.9083 and LBRM with SpyNB is 0.9608. When the precision is 0.4 the corresponding retrieval recall of LC2R with SpyNB is 0.8873 and LBRM with SpyNB is 0.9520. The retrieval recall of LC2R with SpyNB is 0.8652 andLBRM with SpyNB is 0.9306 while precision is 0.6. If precision is 0.8, then the respective retrieval recall of LC2R with SpyNB is 0.8450 andLBRM with SpyNB is 0.9128. The retrieval recall of LC2R with SpyNB is 0.8206 andLBRM with SpyNB is 0.8939 when precision is 1. This result illustrates that the LBRMwith SpyNBhas high retrieval precision-recall.

Conclusion

In this chapter the location of the user is considered for ranking the search results. The LBRM user profiling strategy is fast, robust, and generic but the query results vary from one location to another location. The user's location is a major factor for determining search results relevancy. So in addition to the LC2R user profiling strategy, Location-based Ranking Method (LBRM) is proposed to know the location effects of the queries. It ranks the search results by considering the similarity between the location and retrieved pages. Hence the proposed LBRM performs better than the LC2R method. The experimental results prove that the proposed LBRM has high precision, recall, precision-recall, F-Measure, Δ similarity, Average Relevant Rank and Exponential ratio.

References

1. Abou-Assaleh, T., Das, T., GAO, W., Miao, Y., O'Brien, P., & Zhen, Z. (2007, May). A Link-based ranking scheme for focused search. In Proceedings of the 16th international conference on World Wide Web (pp. 1125-1126). ACM.
2. Akhlaghian, F., Arzanian, B., &Moradi, P. (2010, February). A personalized search engine using ontology-based fuzzy concept networks. In Data Storage and Data Engineering (DSDE), 2010 International Conference on (pp. 137-141). IEEE.
3. Alhenshiri, A. A. (2010). Web information retrieval and search engines techniques. Al-Satil Journal, 55-92.
4. Arens, R. J. (2009). Learning to rank documents with support vector machines via active learning.
5. Arora, N., &Govilkar, S. (2016). Efficient Hybrid Ranking Algorithm for Search Engine. In International Journal of Computer Science and Information Technologies, 7(3), 1624-1629, 2016.
6. Azimzadeh, M., Badie, R., &Esnaashari, M. M. (2016, April). A review on web search engines' automatic evaluation methods and how to select the evaluation method. In Web Research (ICWR), 2016 Second International Conference on (pp. 78-83). IEEE.
7. Bouramoul, A., Kholladi, M. K., & Doan, B. L. (2012, May). An ontology-based approach for semantics ranking of the web search engines results. In Multimedia Computing and Systems (ICMCS), 2012 International Conference on (pp. 797-802). IEEE.
8. Brindha, S., &Sabarinathan, P. (2014). an effective search on web log from most popular downloaded content. International Journal of Distributed and Parallel Systems, 5(1-3), 51.
9. Campos, R., Dias, G., Jorge, A., &Nunes, C. (2016). GTE-Rank: A time-aware search engine to answer time-sensitive queries. Information Processing & Management, 52(2), 273-298.
10. Chaudhuri, S., Theocharous, G., &Ghavamzadeh, M. (2016). Personalized Advertisement Recommendation: A Ranking Approach to Address the Ubiquitous Click Sparsity Problem. arXiv preprint arXiv:1603.01870.
11. Chawla, S. (2016). A novel approach of cluster based optimal ranking of clicked URLs using genetic algorithm for effective personalized web search. Applied Soft Computing, 46, 90-103.
12. Das, N. N., & Kumar, E. (2013). Approaches of Page Ranking Algorithms: Review. International Journal of Computer Applications, 82(2).
13. Deepa, C., &Prathiba, S. (2013). Modified Approach of User Profiling from Search Engine Logs. In International Journal of Computer Science and Mobile Computing, 2(4), 555-558, 2013.

14. Delbru, R., Campinas, S., &Tummarello, G. (2012). Searching web data: An entity retrieval and high-performance indexing model. Web Semantics: Science, Services and Agents on the World Wide Web, 10, 33-58.
15. Delbru, R., Toupikov, N., Catasta, M., Tummarello, G., & Decker, S. (2010, May). Hierarchical Link analysis for ranking web data. In Extended Semantic Web Conference (pp. 225-239). Springer Berlin Heidelberg.
16. Du, Y., &Hai, Y. (2013). Semantic ranking of web pages based on formal concept analysis. Journal of Systems and Software, 86(1), 187s-197.