IJASRM

International Journal of Advanced Scientific Research and Management, Volume 3 Issue 9, Sept 2018

www.ijasrm.com

ISSN 2455-6378

A survey on web mining techniques for capturing user intention for personalized websites

D.K Vijayalakshmi[1] , D.Sridhar, [2]

[1]M.Phil Scholar, Department of Computer Science, Dr.SNS Rajalakshmi College Of Arts and Science, Coimbatore, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science, Dr.SNS Rajalakshmi College Of Arts and Science, Coimbatore, Tamil Nadu, India

Abstract

With the rapid growth of internet are useful source of information for almost every activity huge and more complex data are provide by internet, obtaining valuable information and capture User Intention has become a major and difficult challenge, especially in personalized websites. The field of ,web mining has involved and adopt all data mining techniques specifically Recommendation systems, artificial intelligence, and so on to the web data traces user's visiting behaviors and extracts their user intention using patterns. The previous studies have used several features to retrieve information according to the user interest and preference. The main drawbacks of the previous studies are that need better and effective data cleaning, feature extraction and ranking model. Data mining is an effective way to solve such problems in the personalized sites. This paper surveys various techniques and methods used to capture user intention for the personalized websites.

Keywords: *User Intention, personalization, Web mining, personalized Websites, Intelligent retrieval.*

1. Introduction

Capture user intention is the biggest research topic in nowadays. CPSSs (Cyber-physical-social systems), includes cyber, physical, social world, can provide effective and high quality personalized services for users. The emergence of CPSSs has resulted in the explosive growth of information. To address the explosive large growth of information, the development of effective information retrieval technique is more urgent. Various data mining techniques such as keywords-based, vertical search, multi-keywords queries, and ranking model have been extensively employed to retrieve massive amounts of Internet information. The keywords-based technique uses retrieves information based on keywords provided by the user. Vertical search algorithm is used to provide the retrieve information from web data, but result has contained some noise data. The ranking model method is used to rank the search result but has not been effectively addressed.

However, the retrieve result contain substantial of amount of unnecessary information, some required results can be hidden in the back of the webpage, some of the result it might be non-relevant for users so that user need to spend more and more lot of time to spend finding the user relevant and user intention result. This remains difficult to retrieve more accurate and more user intention related information. These will not provide satisfactory results to the user. Analyzing, capture the user's intention and personalized retrieval still important research topic in data mining. In order to capture user intention extract the interests and preferences accurately for personalized site need machine learning algorithms and with the help of computer technology.

1.1 Web Mining

The intention of Web mining is to detect useful information or knowledge from the Web page content, hyperlink structure along with usage data. Various data mining techniques used by Web mining include supervised learning, unsupervised learning, association rule mining and sequential pattern mining.

1.2 Personalized System

The personalized system (PS) is to predict user's interest and intention on information among the

tremendous amount of available data by using the data mining techniques and prediction algorithms. For example, it analyzing my previous behavior and promoting products that it thinks I might want, this details are analyzed, which means to see right on the homepage in the website. PS have the potential to help and improve the quality of the decisions consumers make while searching for and selecting products online.PS has introduced the need for information filtering techniques that are use to help users by filter out information in which they are interested in. PS changed the way as the websites communicate with their users. Instead of providing a static feel for the users, in shopping items searching, this provides potential suggestions which increases communication to provide a higher experience.PS is recognize intention on individual users based on past browsing history, profiles, click through data feedback and location and other behavior too. Capture the intentions of user mainly based on following factor these important factors are list out and explain here.

1.2.1 Implicit and explicit feedback

Users interactions is large in ecommerce site extracted user intention is more precious task, so that This implicit and explicit feedback system mostly helps to analysis and capture intention in personalized sites, implicit feedback which means system passively track different sorts of user behavior, such as purchase history, watching habits and browsing activity, and click information in order to model user intention. Explicit feedback refer user provide a score based on their interest, such as a rating or a like forget effective user intention.

1.2.2 Keywords extraction

This is a statistical model, initially it which calculates the frequency of words in the text. After that extract all possible words, phrases, term that can potentially be keywords. A score or probability threshold, is then used to select the final set of keywords and identify the relationship among the words

1.2.3 Page Ranking

Page Rank Model is used to rank user search query result. Page Ranking (PR) assigns a numerical weighting that is measuring the value of page to each element of a hyperlinked set of documents. This model is to count page rank that is how many number of users to access the website (or) a web page which this means and it hold with the purpose of measuring its relative importance. When the user performs multiple clicks, the ranking will change due to the influence of the personalized Page Rank value. The WebPages with more clicks will be ranked in higher positions. Based on page ranking values helps to order to model user intention.

2. Literature Review

This paper (Qingyao, 2017) proposed a hierarchical embedding model for personalized product search. In this hierarchical embedding model queries, users and items are represented with their associated text data, then jointly learn embeddings for words, queries, users and items. These text data with this hierarchical structure by directly maximizing the likelihood of observed query-user-item triples. If take this model, that is this model is apply or tested, so this concept is moved in to experimental way, then this Experiments show that our hierarchical embedding model significantly outperforms existing product search baselines on multiple datasets.

This paper (seung-taek park, 2012) author has presented how to personalize recommendation on dynamic content using predictive bilinear models. Dynamic contents such as news articles, E-commerce, etc. present a feature-based machine learning approach for personalization that is takes information about the interested content based by the user. It maintains the profiles of content of interest. This concept is moved on is general and flexible for any other personalized tasks, obviously this model is to provide accurate personalized recommendations for both existing and new users.

This paper (Taher H, 2013) authors had presented Page Rank algorithm for improving the ranking of search-query results. RIV Ranking mode used to improve the ranking of results in response to a query. In this propose computing a set of Page Rank vectors, biased using a set of representative topics. Then, this algorithm is to capture more accurately the notion of importance with respect to a particular topic. For ordinary keyword search queries compute multiple importance scores for each page; we compute a set of scores of the importance of a page with respect to various topics. If the query time that is user recommendation time, these importance scores (counted value) are combined. The combined scores are based on the topics of the query. The Query or requirements is to form a composite PageRank score for those pages matching the query. This score can be used in conjunction with other IR-based scoring schemes to produce a final effective rank for the result pages with respect to the query. In this experiments

result based on PageRank algorithm and this experiment result shows can generate more accurate rankings than with a single, generic PageRank vector.

Authors in this paper (Legenstein.R, 2013) proposed RS (Recommender System) these systems are personalized so it generates user specific recommendations accurately and more efficiently. The concept of personalization is based on User profile and rating. This approach Integrates user profile similarity and user rating similarity to build the user based Recommendation model. These Recommender systems are to more help users find items that they deem of interest to them.

This paper (Xu. C. Z, 2014) authors had presented multi-keyword ranked search approach which makes the query results more consistent with the user's requirements. In this approach secure search protocol is used to enable along with keyword transformation and stemming algorithm. With these techniques, this approach is able to efficiently handle more misspelling mistake. This takes the keyword weight into consideration during ranking process. The approach produces many unrelated results, which lead to a massive waste of computational resources.

In this paper (Lee. W. C, 2016), authors proposed click-through model captured the user's interests and preferences. Effective ranking function is employed to present the search results in some proper order according to the user preferences. The first step in this method is preference mining, which discovers user's preferences of search results from click through data. The second step is the ranking function optimization, which optimizes the ranking according to the user's preferences.

Authors in paper (Dik Lun Lee, 2010) intended that making personalization approach based on query clustering. Initially extract concepts from the web-snippets of the search result returned from a query and use the concepts to identify related queries for that query. A new two phase personalized agglomerative clustering algorithm and fuzzy clustering algorithm, which is able to generate personalized query clusters. This paper Experimental results show that our approach has better precision and recall than the existing query clustering methods.

This paper (Weiyi Meng, 2012) author has presented the concept of personalized web search by mapping user queries to categories. Personalization of web search is carried out the user intention's that is interests. In this paper, propose a novel technique and to map a user query to a set of categories that is grouped, so easily find the user decisions. This personalization differently finds out the user intension, steps are followed by take the user decision is, a user profile learned from the user's history. Then, general profile learned from the user's history. After category classification respectively. So this following steps are finally combine and then give the results is to map a user query into a set of categories.

This paper (ChengXiang Zhai, 2015) author presented personalized search for implicit user modeling. This paper decision theoretic framework used for optimizing interactive information retrieval based on individual users. This approach shows how to infer a user's interest from the user's search context and use the inferred implicit user model for personalized search. Experiments search result show that our search agent can improve search accuracy over the popular search engine.

Authors in (Lau, F.C.M, 2011) presented a personalized webpage re-ranking algorithm through mining dwell times of a user. Dwell times derived from a user's previously online reading or browsing activities. This paper introduces a quantitative model to derive concept word level user dwell times from the observed document level user dwell time. In this re-ranking algorithm is to predict user dwell time allows us to carry out the personalized webpage re-ranking. The concept of
personalization use the re-ranking algorithm, and then, to explore the effectiveness of our algorithm, measured the performance of our algorithm under two conditions its effectively generating personalized webpage rankings to satisfy a user's personal preference.

Authors in (Jia-Xuan Wei, 2014) presented general-purpose and practical meta-algorithmic for a personalized recommendation system which combining user profile, intrapersonal and interpersonal interest similarity along with the interpersonal influence. The thing of user individual interest makes associations between user and product with hidden features. This RS utilizes interpersonal and intra personal similarities from the user log. Creates a memory based collaborative filtering for online shopping product recommendation. Identifying user's usage patterns and how many users accepted the recommendations.

Authors in this paper (Mei, T, 2014) present an image search re-ranking algorithm, called click-based relevance feedback retrieve more accurate results.

Table 1.0. Comparison table

Paper No	Technique	Advantages	Disadvantage
1	hierarchical embedding model	This model significantly outperforms product search.	Learning the semantic representation for multiple entities takes time consuming process.
2	predictive bilinear models	accurate personalized recommendations for Dynamic contents	offline model with light computational overhead
3	RIV Ranking mode	Produce a accurate effective rank for the result pages with respect to the query.	Sink problem occurs because of infinite network. This is not too fat.
4	Recommender System	Generates user specific recommendations more accurately.	Need a lot of dataset in order to effectively make recommendations.
5	Multi-keyword ranked search Model	This model offers and efficiently handle more misspelling mistake. Reduce the maintenance overhead during the Keyword dictionary expansion.	The approach produces many unrelated results, which lead to a massive waste of computational resources.
6	Click-through model	This model Accurately Captured the user's interests and preferences. .	Capture the diverse click behavior patterns more difficult and time consuming process.
7	fuzzy clustering algorithm	Generate effective personalized query clusters. Better precision and recall.	Difficult to predict the number of clusters (K-Value), The order of the data has an impact on the final results.
8	Ranking algorithm	Generates user product interests with the help of accurate mapping strategies.	Matching user Search queries with product categories are tedious task.
9	Implicit user modeling	Improve 90% search accuracy	This is not suitable for dealing with huge quantities of data.
10	webpage re-ranking algorithm	User friendly and more effectively satisfy user's personal preference.	Find semantic similarities between the result page and the query to re-rank the results tedious process.
11	Collaborative filtering	This model can be a powerful way of recommending items based on user history.	Complexity and expense and it create cold start problem.
11	Image search re-ranking algorithm	This model offers more accurate results. Improve classification accuracy.	It does not show the relationship Between the image similarity. This needs more attention.

From the survey analysis Click-through data and leverage multiple kernel learning simultaneously to boost image search performance. Algorithm, which transforms image re-ranking into a classification problem. It leverages the clicked images as positive data and images from other queries as negative data to improve classification accuracy and can automatically

learn the fusion weight of each modality for different queries at the feature level.

Table 1.0. Depicts the working methodologies of various data mining techniques which can be used to achieve personalization and capture user intention.

3. Conclusion

Capture User Intention in personalized sites is one of the major problems in nowadays which leads to dissatisfaction results. Predicting and capture intention is possible only by the consideration of some of the important user's factors. This factors analyzing can be achieved by the inclusion of data mining techniques. Data mining methodologies embraces methods such as ranking, Recommendation systems, clustering mechanisms, classification, and etc.Further implementation has to be done in order to predict user intention for personalized websites.

References

[1] Ai, Qingyao, et al. "Learning a hierarchical embedding model for personalized product search." Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, (2017).

[2] Chu, Wei, and Seung-Taek Park. "Personalized recommendation on dynamic content using predictive bilinear models." Proceedings of the 18th international conference on World Wide Web. ACM, (2012).

[3] Haveliwala, Taher H. "Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search." IEEE transactions on knowledge and data engineering 15.4: 784-796, (2013).

[4] Jahrer.M, Toscher.A, and Legenstein.R. "Combining predictions for accurate recommender systems". KDD'10, pp. 693-702, (2013).

[5] Li, R., Xu, Z., Kang, W., Yow, K. C., & Xu, C. Z. Efficient multi-keywords ranked query over encrypted data in cloud computing. Future Generation Computer Systems, 30(1), 179-190, (2014).

[6] Leung, W. T., Lee, D. L., & Lee, W. C. Personalized Web search with location preferences. IEEE, International Conference on Data Engineering Vol.41(pp):701-712, (2016).

[7] Leung, Kenneth Wai-Ting, Wilfred Ng, and Dik Lun Lee. "Personalized concept-based clustering of search engine queries." IEEE transactions on knowledge and data engineering 20.11: 1505-1518, (2010).

[8] Liu, Fang, Clement Yu, and Weiyi Meng. "Personalized web search by mapping user queries to categories." Proceedings of the eleventh international conference knowledge management. ACM,(2012).

[9] Shen, Xuehua, Bin Tan, and ChengXiang Zhai. "Implicit user modeling for personalized search." Proceedings of the 17th ACM international conference on Information and knowledge management. ACM, (2015).

[10] u, S., Jiang, H., & Lau, F.C.M.. Mining user dwell time for personalized web search reranking. In Proceedings of the 20th International Joint Conference on Artificial Intelligence.Palo Alto, CA: AAAI Press(pp. 23672372). (2011).

[11] Zhang, Ruisheng, Qi-dong Liu, and Jia-Xuan Wei. "Collaborative filtering for recommender systems." 2014 Second International Conference on Advanced Cloud and Big Data (CBD). IEEE, 2014.

[12] Zhang, Y., Yang, X., & Mei, T.. Image search reranking with query-dependent click-based relevance feedback.IEEE Transactions on Image Processing IEEE Signal Processing Society, 23(10), 4448 , (2014).