A PROFILE BASED USER CUSTOMIZABLE PRIVACY PRESERVING WEB SEARCH

M.Malik[1], S.P.Ramya[2], K.Nivetha[3]

[1,2]Department of Information Technology, Periyar Maniammai University, Vallam, Tamil Nadu 613403, India
[3] Department of Software Engineering, Periyar Maniammai University, Vallam, Tamil Nadu 613403, India

[1]malik_m21@yahoo.com
[2]way2ramya@yahoo.com
[3]hanivet.swe@gmail.com

Abstract- **In image retrieval search engine, the images can be retrieved based on the text or content. The existing image retrieval which classifies based on the click through logs may plays a vital role for effective search results. The user satisfaction can be obtained through the user click sequence. From the click sequence, the feedback sessions are calculated and produce effective search results. Thus the results show only the related results, but the users need the related and relevant data results. For this, the novel algorithm may propose to obtain the effective relevant and related search results based on both the click through logs and personalized web search. The personal profile added to the system to get the effective search results that are based on both these two techniques. After that the privacy protection has been implemented to avoid the data leakage when the system use protected details for searching. Our runtime generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the generalized profile. The new system presents two greedy algorithms, namely GreedyDP and GreedyIL, for runtime generalization. The system proposes a Cluster-Based SVM (CB-SVM) method to overcome the problems obtained with the SVM Classifier and also it can be tested with the Big Data Applications. We also provide an online prediction mechanism for deciding whether personalizing a query is beneficial. The experimental results show the effectiveness of the novel algorithm.**

Index Terms- **Image Retrieval, click through logs, Cluster Based SVM Classifier, data leakage.**

I. INTRODUCTION

Use of the digital data is increased due to the digitalization of each sector. Digital data is available in the form of text, audio and video. Mostly knowledge discovery from the data stored in typical alphanumeric database, such as relational database, has been focus of work in data mining. Nowadays, nonstandard databases like images are also used. Mining of these image datasets can be performed to discover the knowledge not explicitly stored in images. Image mining is the concept used to extract useful knowledge from images. It deals with the extracting inherent and embedded knowledge, image data relationship, or other pattern which is not explicitly found in the images [1, 3]. Mining large collection of images, and combined data mining of large collections of images with associated alphanumeric data are the two important themes of image mining [2]. Image mining is still at the experimental stage and growing filed for research. Lack of understanding the research issues of image mining is the obstacle to rapid progress. Image mining is not just the expansion of data mining to image domain. It can be considered to be an efficient hybridization of image processing and data mining concepts to extract the useful knowledge. Various application domains of image mining include natural scene recognition, remote sensing, weather forecasting, criminal investigation, image segmentation, etc.

Keyword-based image search is not only a problem of significant commercial importance but it also raises fundamental research questions at the intersection of computer vision, natural language processing, machine learning, and information retrieval. Our objective is to improve the performance of tail queries in image search engines by leveraging click data aggregated across users and sessions. To this end, we address three limitations of existing search engines in this paper. First, there is no straightforward, fully automated way of going from textual queries to visual features. Image search engines therefore primarily rely on static and textual features for ranking. Visual features are mainly used for secondary tasks such as finding similar images. Second, image rankers are trained on query-image pairs labeled with relevance judgments determined by human experts. Such labels are well known to be noisy due to various factors including ambiguous queries, unknown user intent and subjectivity in human judgments. This leads to learning a sub-optimal ranker. Finally, a static ranker is typically built to handle disparate user queries. The ranker is therefore unable

to adapt its parameters to suit the query at hand which again leads to sub-optimal results.

Here the new search engine for images may propose to obtain the user satisfied image search results. The image search can be done through the click sequences for the feedback sessions and also personalized profile based search. User feedback can provide powerful information for analyzing and optimizing the performance of information retrieval systems. Unfortunately, experience shows that users are only rarely willing to give explicit feedback. To overcome this problem, this paper explores an approach to extracting information from implicit feedback. The user is not required to answer questions, but the system observes the user's behavior and infers implicit preference information automatically.

The particular retrieval setting studied in this paper is web search engines. In this setting, it seems out of question to ask users for relevance judgments about the documents returned. However, it is easy to observe the links the user clicked on. With search engines that receive millions of queries per day, the available quantity of such click-through data is virtually unlimited. This paper shows how it is possible to tap this information source to compare different search engines according to their effectiveness.

Our approach is based on building user profiles based on the user's interactions with a particular search engine. For this purpose, we implemented Wrapper: a wrapper around the search engine, that logs the queries, search results, and clicks on a per user basis. This information was then used to create user profiles and these profiles were used in a controlled study to determine their effectiveness for providing personalized search results.

The study was conducted through three phases: 1. collecting information from users. All searches, for which at least one of the results was clicked, were logged per user. 2. Creation of user profiles. Two different sources of information were identified for this purpose: all queries submitted for which at least one of the results was visited and all snippets visited. Two profiles were created out of either queries or snippets 3. Evaluation: the profiles created were used to calculate a new rank of results browsed by users.

II.   RELATED WORK

Beitzel S.M. presented the approaches[6] for the web query classification. In this, the ambiguous query optimization has been improve the performance and effectiveness of the information retrieval. C-K Huang et. al. presented an effective algorithm[5] for log based approach to find the relevant term

extraction and suggestion. The search engine logs are collected for the similar query sessions for the relevant term suggested.

Zheng Lu et. al proposed the novel approach[7] to infer the user search goals with feedback sessions. The click sequence has to obtain from the feedback sessions and the pseudo documents may generate to obtain the user search goals. The criterion named classified Average Precision may also propose to measure the performance of the information retrieval.

Wang X et. al. presented a approach for effective image retrieval[8 that based both the keyword based and also visual similarity with the query image. The query specific semantic signature based image retrieval may propose to obtain the re-ranked web images.

Shou L et. al. proposed the framework for obtaining user satisfied search goals based on personalized profile data with the online profiling. Thus the framework may also secure the privacy protected personal data of the users without degrading the performance of the user search.

III.   SYSTEM STUDY

The user should search the image results through the keyword based information retrieval. The feedback sessions should be monitored through the click logs or click sequence. From the click sequence, the system should re rank the image results that are obtained in the existing approach. Thus the retrieval should be measure through the classified average precision.
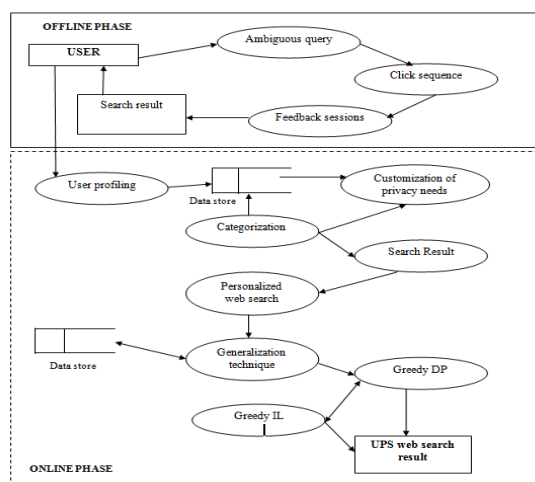

Fig 1. Optimized Image Search

The profile data of the user search goals should be updated in the system. After the retrieval based on click through logs, the image re ranking should be done through the semantic signature specification. The visual features should be identified so that the results should be re-ranked. The GreedyIL and GreedyDP algorithms used to retrieve the optimal solution for the query image.

We propose a Cluster-Based SVM (CB-SVM) method to overcome the problems obtained with the SVM Classifier and also it can be tested with the Big Data Applications. The major contributions of the proposed system as follows:

1. Construct the micro clusters using a CF-Tree. 2. Train an SVM on the centroids of the micro clusters. 3. De-cluster entries near the boundary. 4. Repeat the SVM training with the additional entries. 5. Repeat the above until convergence. At the online stage, images are re-ranked by comparing their semantic signatures obtained from the visual semantic space specified by the query keyword. The new approach significantly improves both the accuracy and efficiency of image re-ranking.

IV.   CONCLUSIONS

Thus the image results should satisfy the user goals by means of both the click through log based and personalized web search with profile data. The security is needed to avoid the eavesdropping attack in personal profile data. The users need to obtain the valid results from the classified average precision. For the generated dataset the results are validated through the user satisfaction. The visual feature similarity should be identified for the query image after that the visually equaled image can be re-ranked to obtain the user satisfaction.

REFERENCES

[1] C. Lakshmi Devasena, T.Sumathi, Dr. M. Hemalatha, "An Experiential Survey on Image Mining Tools, Techniques and Applications" International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 3 Mar 2011.

[2] C. Ordonez and E. Omiecinski, "Image Mining: A New Approach for Data Mining", 1998.

[3] J. Han and M. Kamber, "Data Mining: Concepts and Techniques", 2/E, Morgan Kaufmann Publication, 2006.

[4] C.-K Huang, L.-F Chien, and Y.-J Oyang, "Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query Session Logs," J. Am. Soc. for Information Science and Technology, vol. 54, no. 7, pp. 638-649, 2003.

[5] S. Beitzel, E. Jensen, A. Chowdhury, and O. Frieder, "Varying Approaches to Topical Web Query Classification," Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development (SIGIR '07), pp. 783-784, 2007.

[6] Lu Z, Zha H, Yang X and Zheng Z, "A New Algorithm for Inferring User Search Goals with Feedback Sessions", IEEE. Trans. on Knowledge and Data Engineering, vol. 25, no.3, pp.502-514, 2013.

[7] Wnag X, Liu K and Tang X,"Web Image Re-ranking using query specific semantic signatures",IEEE trans. on Pattern Analysis and Machine Intelligence, vol. 36, no.4, 2014.

[8] Shou L, Bai H, Chen K and Chen G,"Supporting privacy protection in personalized web search", IEEE Trans. on Knowledge and Data Engineering, vol 26, no.2, 2014.