International Journal of Innovations & Research Analysis (IJIRA)
ISSN : 2583-0295, Impact Factor: 5.449, Volume 02, No. 03(II), July- September, 2022, pp 168-173

PERSONALIZED INFORMATION LEARNING AND CLASSIFICATION: A REVIEW

Dr. Vandna Bhalla[*]
Dr. Rinki Sharma[**]

ABSTRACT

Today's user wants to browse efficiently through huge personal collections made possible due to the easy and economical availability of digital cameras and immense storage at lower prices. The proliferation of low cost multimedia devices has resulted in an unprecedented growth of personal data which may or may not be interrelated, is heterogeneous and unstructured. Personal collections can be in various forms and the variety and the amount of personal information that an individual deals with everyday is growing constantly. Full use and benefits of such collections are meaningless if the retrieval and access methods are limited and ineffective. It is labour as well as time intensive to select specific data for sharing or for personal use given the sheer size of these collections. We present a review of personalized information learning tools.

Keywords: *Information, Personalized, Retrieval, Small Data Set, Training, Samples.*

Introduction

We are entering the era of owning huge personal data. The tools for managing these often cluttered collections are gaining significance. The variety and the amount of personal information is constantly growing and the current tools for managing these are inadequate. Personalized search is very different from search in personal data collections or albums. Personalized search are the tailored experiences on the web according to the user's interest beyond the query. Mechanisms for retrieving from personal databases require different techniques than those used for general web search. Personal data retrieval is an important arena for research but has many challenges. It is difficult to design a generalized framework for retrieval as individuals use different tools to create a varied mix of personal data. Evaluation strategies for personal data retrieval mechanism are costly as these require long term studies of the users and sometimes their active participation too. Further privacy rights can prevent sharing of such personal data amongst researchers. In this work we review the existing techniques for personalized information retrieval and their inadequacies for retrieving from personal data. We review the current techniques for search in personal collections. The personal datasets have limited number of training samples and therefore have their own challenges. The personal data is now being increasingly stored on different social media which presents different challenges as the data is now fragmented since it is scattered across many devices (e.g. phone) and services (e.g. Google photos).

Personalized Information Retrieval

Information Retrieval Systems assist users to find relevant information from a gigantic sea of information. Personalization in Information Retrieval systems goes a little beyond traditional methods and tries to satisfy the user's specific and personal information requirements by enhancing the search results such that they are of particular relevance to the user who queried the information. In a conventional Web search system, a user expresses short usually ambiguous query in a textual format using a limited

[*] Department of Electronics, Sri Aurobindo College, Delhi University, New Delhi, India.
[**] Department of Electronics, Sri Aurobindo College, Delhi University, New Delhi, India.

number of keywords. Mere keyword queries are actually inadequate to accurately describe the user's current requirement intent. Different people can have different intention of the same keywords. 'Cricket' could be an insect or the game and the results returned by search engine are mixed irrespective of the preferences of the user. 'Jaguar' for a car lover should ideally return various car models whereas the same query by an animal enthusiast should return different pictures of the canine cat, Fig. 1.

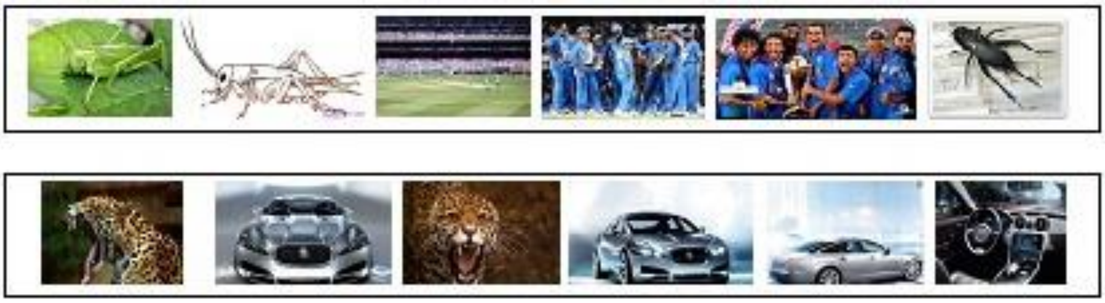

Fig. 1: Non personalised search result for 'cricket' and 'jaguar'

The existing information retrieval systems like Yahoo, Google etc. are unable to provide personalized search results as they do not have the personal information of the user. The in-formation thus returned is rarely exact or precise and never personalized. Google relies only on keywords to search and uses PageRank technology which returns results as per the popular public demand and not user specific. It generally returns results of poor quality when handling multimedia content as it lacks semantic information. 'One size fits all' paradigm is insufficient in today's scenario. It is highly implausible that hundreds of billion minds are so alike in their interests and thought processes that same perspective and techniques to browsing and searching fits every individual's requirements. The concept of personalization is undergoing a paradigm change. An ideal personalized response should be as shown in Fig. 2.

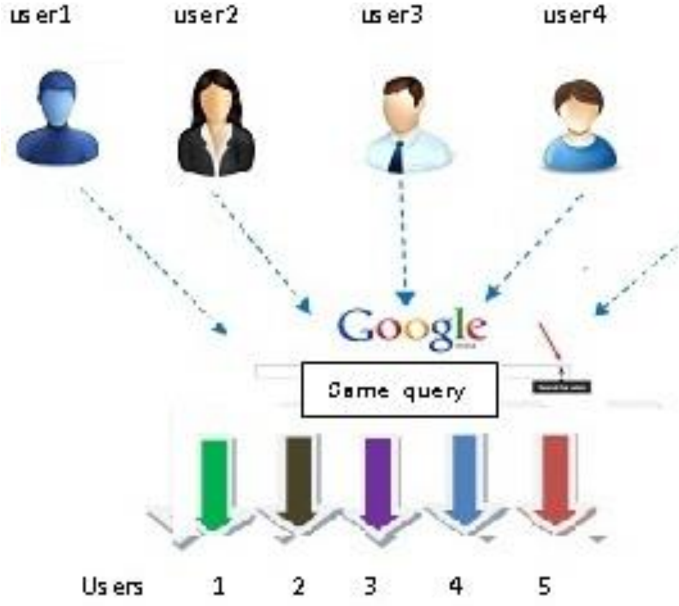

Fig. 2: Ideal Personalized Query Handling

A Review

Personalized Search modifies the results returned conforming to the user's intent and hence greatly improves the experience of web search[1]. Users' personal preferences, interactions and interests have to be taken into consideration and monitored over time to retrieve results that would be of relevance to the user at the time of query[2]. Relevance for the user also may undergo temporal shifts which have to be mapped constantly and patterns generated need to be examined. Typically, the user preferences and interests go a long way in helping model search personalization. A user's interaction with the system and past search behaviour can be tracked and aggregated to help build a personalized information retrieval model. Personalizing search based on context requires built in intelligence to match the requirements with various learning contexts. Semantic knowledge about the domain being explored and the user query play a critical role in personalized web information access. Re ranking search results according to the user context information brings more relevance to search results. A lot of past work is on ontology based personalized search approaches [3] [4]. Based on semantics, a framework was presented in 2008 which builds a semantic domain using known information. It generates the learner's profile, clusters the documents and learns more refined sub concepts, re ranks the search results based on matching concepts in the content and user profile and provides with semantic recommendations [5]. This study is very domain specific and focuses on e-learning environment.

Zhang, Zhu, Zhao and Xu in 2008 [6] suggested personalized retrieval based on multi model cross mutual knowledge of the user. Their work provides personalization based entirely on the user interaction, implicit as well as explicit. Initially the user interest model is built based on information provided by the user. This is updated regularly based on the user action and the feedback so generated. The system calculates the high-level semantic level similarity function along with the low-level perception features using vector clustering. In this model the critical part is the user interest model and its updation. There is a lot of user's time involved to build the initial interest model and periodically to update this model by providing constant feedback information of query results. Personalised image retrieval based on user interest model started a new trend. The low-level visual features and the high-level semantic concepts help build up the short-term interests. The long-term interests are inferred from the accumulated short-term interests. Weights are assigned to the semantic concepts indicating the user interest in each of them. The question arises as to how the users' interests or semantic meanings can be best captured.

User information [7] can be acquired in two ways (1) Explicit: users input their information including their personal interests into the system. This requires the user to willingly provide their personal information and puts extra demand on the user time. Further users may not be very accurate about their data or profile may become inaccurate over time as user's interests undergo transitions. (2) Implicit: user information is collected by the system automatically without user intervention. The disadvantage is that this model collect only positive information. Browsing histories typically provide information on user's interests but are generally shared with one particular website which means that only that website can provide personalized services. Moreover, it usually collects information from a single computer. This method requires larger investment in relevant software deployment.

The user preferences are constantly changing which is difficult for the slow learning processes to track. An adaptive user profile method considering the change in user's preferences by Jeon, Kim and Choi address these problems and improve profiles using Genetic Algorithms [8]. A user's profile is updated through feedbacks sent by the user. They have used a user profile approach and collaborative filtering approach to account for changing user preferences. The whole approach is based on user feedback. The weight value increases with more feedback which puts extra burden on the user since it requires explicit user interaction. Explicit models in which the users offer information on their own initiative [9] are not as popular as mostly people are hesitant to express their preferences. There is no data to certify the performance of this model. The year 2008 saw maximum work so far on personalized retrieval based on user inputs. A personalized retrieval approach using implicit user information and interest quotient was proposed in 2009 [10]. A multidimensional user model based on implicit interaction with the user using navigation to measure the interest level in a given entity is used to build the user profile. Both semantic and spatial user contexts are considered to define the similarity measures. A multidimensional user model is built iteratively using user information. User's navigation information provides estimations which iteratively build the model. Three measures i.e number of interesting objects, spatial distance and semantic distance are applied to improve each consecutive iteration. How the various user models interact and whether similarity measurements between user models can enhance personalization is still unexplored. More or less all these PIR systems work on the common principle of re ranking the results of a general query based on user profiles. In 2011 Bennet, Radlinski, White, and Yilmaz [11] personalized the web search by using Location Metadata. For every relevant website, generalized Gaussian Expectation Maximization is used to learn about the user location. They first create a model consisting of all likely locations of interest for each website. This estimation of geographic distribution for each website is a cumbersome task.

Summary of Popular Techniques

Figure 2.3 gives a diagrammatic summarization of popular current methodologies of personalized information retrieval.

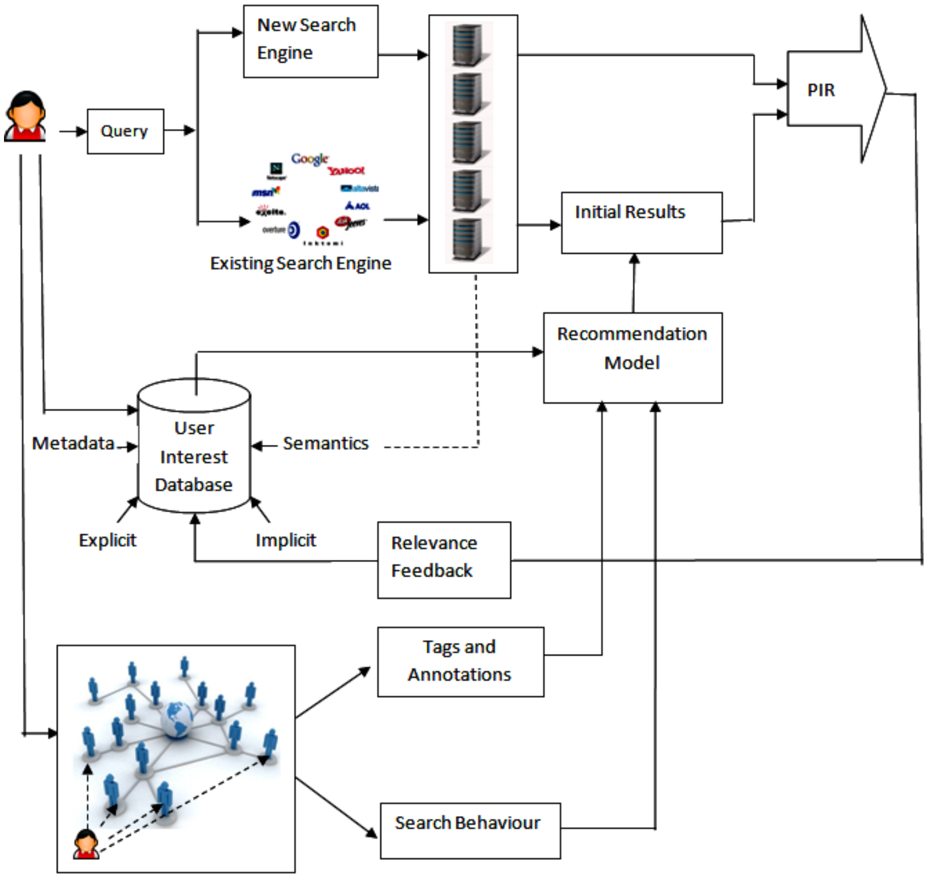

Fig. 3: Summarization

Keyword based search has been so far the most popular search paradigm though its performance is not very satisfactory. Over 50 percent of the time it does not yield satisfactory results. The reasons could be as follows:

- Queries are short and nonspecific.
- Intentions of the users are not captured optimally by words.

Most personalized systems built so far require explicit user intervention [12] and are in text related domains. Though tedious, wearisome and time-consuming, yet a lot of researches were dedicated to enhance and ease the annotation tools [13]. Some of these techniques perform well for other data scope as per the literature review but their performance is ordinary in personal data collections. Most of the work done for personalized retrieval has been for text-based documents. Today the net is used for putting up so much more than text. A personal repository typically contains a huge amount of information in the form of text, images and videos and its volume is growing rapidly. With the increasing popularity of digital multimedia such as images and videos and the advent of the cloud computing paradigm, a fast-growing amount of private and sensitive multimedia data are being stored and managed over the network cloud as personal data by users. Personalization has the potential to vastly benefit the searching experience by taking into consideration the exact intentions of the user queries and rerank the searched results. The search processes of today do not take into account the users' interest, the diversity and the continual increase in the content of the search. Hence results returned by existing search methods do not give the information as required by a particular user. It also contains an assumption that user query is static.

Future: Intelligence for Search in Personal Collections

We are on the threshold of taking this personalization to the next level, where a user is offered personalized service for the constantly increasing collections of personal data. Now modern PIR architectures with intelligent techniques are required to perform personalised services for current realms and these should help store/ retrieve data efficiently, providing personalization at an individual level particularly for personal databases. Photos by far are the biggest class of personal data today. Most of the current photo management systems are based on text/ keywords-based annotations [14] which are intuitive but the user needs to acquaint themselves with concepts like class, property relations and instances etc. Early systems, like FotoFile [15] and Photo finder [16], annotate content with keywords and names and use archives to generate tags and annotations. Though the search performance is good but the tedious process of manually annotating each photo by the user is a requirement. Actually, most personalized systems built so far require specific and precise user intervention [12]. All the early systems like Fotofile[15],

Photo finder [16] and Photo Mesa [17] too need the tedious inputs of the user to either annotate or arrange photos personally. Some models use the metadata to help identify the relevant categories like for instance Namaan [18] and Orii [19]. But these have not shown satisfactory results in the domain of personal photo collections. If the data is placed in the correct category then retrieval will be more efficient [20]. Quite a few Manual, Semi Automated and Automated annotation techniques are available in literature and Kustanowitz [21] proposes a frame work highlighting the weakness and strengths. Another technique is to use the capability of knowledge inference where the preceding commented data is used for future data. This technique is inadequate to yield inference for the prospective photos of future. PhotoMesa [17] maximizes the screen-space usage, but here also the photos have to be arranged personally by the user. The lexically motivated keyword-based approach does not resolve multiplicity and semantic relevance remains a issue. Naaman et al [18] automatically generate additional metadata from each photo and based on a user study and survey identified the useful and relevant categories of contextual meta data for retrieval. Extracting information from the meta data like location, time of the day, light status, weather status and temperature and maybe additional categories or any similar additional information to generate related contextual information and use it to recall relevant photographs is not sufficient today. The fast visual scanning mechanism like pan and zoom do not scale to manage the large personal photo collections. Y.Orii et al. [19] shows that congregating on the basis of time stamp makes very little difference for unfamiliar photo collections and in their subsequent work say that it helps browsing experience. This is useful when you want to cluster contiguous photos and not personal photo collections. Kai-En Tsay et al [22] did develop an organizer but their results are case studies and not statistically significant analysis. Also, very broad categories like for instance people/non-people and indoor/outdoor were chosen. Today people are in need of systems wherein they can store their images/videos over a long period of time, which they can access privately or give access to, on a selective basis, to individuals or a group of individuals. A personalized framework is needed which is capable of retrieving data for a user irrespective of size of training data, inter and intra class diversity, number of classes with discerning automation built in.

References

1. Jitao, S., Changsheng, X., & Dongyuan, L. (2012). Learn to personalized image search from the photo sharing websites. *IEEE Trans. Multimedia, 14*(4), 963–974.
2. Jaime, T., Susan, T. D., & Eric, H. (2005). Personalizing search via automated analysis of interests and activities. *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (New York, NY, USA), SIGIR '05, ACM,* 449–456.
3. Susan, G., Jason, C., & Alaxander, P. (2003). Ontology-based personalized search and browsing. *Web Intelli. and Agent Sys, 1*(3-4), 219–234.
4. Joana, T., & Susan, G. (2004). Improving ontology-based user profiles.
5. Leyla, Z., & Olfa, N. (2008). Personalized cluster-based semantically enriched web search for e-learning. *Proceedings of the 2Nd International Workshop on Ontologies and Information Systems for the Semantic Web (New York, NY, USA), ONISW '08, ACM,* 105–112.
6. Zhaowen, Q., Haiyan, C., & Haiyi, Z. (2015). Personalized web image retrieval based on user interest model. *Springer Berlin Heidelberg, Berlin, Heidelberg,* 188–195.

7. Fikadu, G., Liu, T., & Zhang, Y. (2010). A framework for personalized information retrieval model, Computer and Network Technology. *International Conference on 00* 500–505.

8. Hochul, J., Taehwan, K., & Joonqmin, C. (2008). Adaptive user profiling for personalized information retrieval. *Third International Conference on Convergence and Hybrid Information Technology,* 2, 836–841.

9. Liqiang, G., & Howard, J. H. (2006). Interestingness measures for data mining: A Survey. *ACM Comput. Surv. 38* (3).

10. Azza, H., Sahbi, S., Malek, G., & Henda, B. G. (2010). How to improve information retrieval using user's profile: Survey and open issue. *Proceedings of the 2010 International Conference on Information & Knowledge Engineering, IKE 2010, July 12-15, 2010, Las Vegas Nevada, USA,* 445–451.

11. Paul, N. B., Filip, R., Ryen, W. W., & Emine, Y. (2011). Inferring and using location metadata to personalize web search. *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (New York, NY, USA), SIGIR '11, ACM,* 135–144.

12. Esin, G., Thomas, O., Else, L., & Moncef, G. (2014). Instance based personalized multi-form image browsing and retrieval. *Multimedia tools and applications, 71*(3), 1087–1104.

13. Brendan, E., & Z. Meral, O. (2007). A comparison of methods for semantic photo annotation suggestion., *22nd international symposium on Computer and information sciences, IEEE,* 1–6.

14. Yanmei, C., Tian, X., Jianming, Z., & Haifeng, Li. (2010). Intelligent digital photo management system using ontology and swrl. *International Conference on Computational Intelligence and Security (CIS),* 18–22.

15. Allan, K., Celine, P., Michael, L. C., Dennis, F., Bill, S., & Jacek, G. (1999). Fotofile: a consumer multimedia organization and retrieval system. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM,* 496–503.

16. Hyunmo, K., & Ben, S. (2000). Visualization methods for personal photo collections: Browsing and searching in the photofinder. https://www.cs.umd.edu/~ben/papers/Kang2000Visualization.pdf

17. Benjamin, B. B. (2001). Photomesa: a zoomable image browser using quantum treemaps and bubblemaps. *Proceedings of the 14th annual ACM symposium on User interface software and technology, ACM,* 71–80.

18. [18] Mor, N., Susumu, H., Qian, Y. W., Hector, G. M., & Andreas, P. (2004). Context data in geo-referenced digital photo collections. *Proceedings of the 12th Annual ACM International Conference on Multimedia (New York, NY, USA), MULTIMEDIA '04, ACM,* 196–203.

19. Yuki, O., Takayuki, N., & Toshiyuki, K. (2008). Web-based intelligent photo browser for flood of personal digital photographs. *WI-IAT'08. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 3*, 127–130.

20. Rong, Y., Apostol, N., & Murray, C. (2007). An efficient manual image annotation approach based on tagging and browsing. *Workshop on Multimedia Information Retrieval on The Many Faces of Multimedia Semantics (New York, NY, USA), MS '07, ACM,* 13–20.

21. Kustanowitz, J., & Ben, S (2004). Motivating annotation for digital photographs: Lowering barriers while raising incentives.

22. Kai-En, T., Yi-Leh, W., Maw-Kae, H., & Cheng-Yuan, T. (2009). Personal photo organizer based on automated annotation framework., *5th International Conference on Intelligent Information Hiding and Multimedia Signal Proceedings IEEE*.

□❍□