ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 5, Issue 4, April 2016

IMPLEMENTING ADVISOR SEARCH WITH KNOWLEDGE SHARING

Nilofar N. Pathan,
Computer Science and Engineering,
H.V.P.M's C.O.E.T,
Amravati, India

Prof.Ranjit R. Keole
Information Technology
H.V.P.M's C.O.E.T
Amravati, India

Abstract— In order to analyze knowledge acquired by web users analysis of users web surfing data is very useful. Using web surfing data it is possible to find advisor who most likely possesses the desired piece of fine grained knowledge related with given query. In this dissertation work investigation is done on fine-grained knowledge sharing in collaborative environments. In this work a methodology is proposed to analyze member's web surfing data to summarize the fine-grained knowledge acquired by them and keep a record. A two-step framework is proposed for mining fine-grained knowledge: (1) web surfing data is clustered into tasks by a nonparametric generative model; (2) a novel discriminative infinite Hidden Markov Model (d-iHMM) is developed to mine fine-grained aspects in each task. Finally, the classic expert search method is applied to the mined results to find proper advisor for knowledge sharing.

Index: Web surfing data, Clustering, knowledge sharing, infinite Hidden Markov Model, fine-grained knowledge, nonparametric generative model.

I. INTRODUCTION

There is no existing method which considers the micro-aspects based mining. The task which is further divided into fine grained aspects is known as micro-aspects in web surfing data. Mining these micro-aspects is critical as it can provide a detailed description of the knowledge gained by a person, which is the basis for advisor search. In collaborative environment it is usual that members try to acquire similar information on the web in order to gain specific knowledge. For example, in software development company, where members are focused on projects which require similar background knowledge, it may happen that the developer tries to solve a particular problem using one method which he/ she is not familiar with but has been studied by another developer. In this case finding the right individual who already knows something in that field is always superior than studying by oneself. Because people can provide live interaction, better communication and give more insights on that particular problem and also gives what were the problems faced by him/her. It will be productive to get them connected and share learned knowledge.

This dissertation presents a novel method to identify, how to enable such knowledge sharing mechanism by analyzing user data. For example, Alice wants to download a dictionary and she starts to surf the web and she doesn't get the direct download link which has already been accessed by Bob. In this case, it might be a good idea to consult Bob or go through the links accessed by Bob, rather than searching by herself and wasting time on useless links redirecting to the another page. Such recommendations are provided with this methodology by analyzing surfing activities automatically. Also if any person wants to keep his name secret about any topic then the privacy can also be provided. Without showing his name the search made by him can be suggested to another person. This can give a proper direction to ones search.

The goal of this proposed work is to find proper "advisors" who are most likely possessing the desired piece of fine-grained knowledge based on their web surfing activities. This work departs from the traditional expert search method in that expert search aims to find domain experts based on their associated documents in an enterprise repository. In order to analyze the knowledge acquired by web users, new method is proposed to log and analyze user's web surfing data. User's interactions with the web can be segmented into different "tasks", e.g., "web mining" and "shopping". Textual contents of a task are usually cohesive. This dissertation defines a session as an aggregation of consecutively browsed web contents of a user that belong to the same task. A task can be further decomposed into fine-grained aspects (called micro-aspects). A micro-aspect could be roughly defined as a significantly more cohesive

subset of sessions in a task. For example, the task "web mining" might contain "usage mining" and "content mining". To this end, a novel discriminative infinite Hidden Markov Model (d-iHMM) is proposed to mine micro-aspects in each task. Finally, a language model based expert search method is applied over the mined micro aspects for advisor search.

II. LITERATURE REVIEW & RELATED WORK

In year 2001 M. J. Beal, Z. Ghahramani, and C. E. Rasmussen [1] had shown that it is possible to extend hidden Markov models to have a countably infinite number of hidden states. By using the theory of Dirichlet processes they implicitly integrated out the infinitely many transition parameters.

In year 2003 D. M. Blei, A. Y. Ng, and M. I. Jordan [2] had analyzed topic modeling. Topic modeling is a popular tool for analyzing topics in a document collection. The most prevalent topic modeling method is Latent Dirichlet Allocation (LDA). It is a generative probabilistic model for collections of discrete data. Topic modeling decomposes a document into topics.

But it doesn't recover the semantic structures of people's online learning activities from their web surfing data, i.e. identifying groups of sessions representing tasks (e.g. learning "Java") and micro-aspects (e.g. learning "Java multithreading"). After applying topic modeling methods on session data, it is still difficult to find the right advisor by using the mined topics.

In year 2005 X. Liu, W. B. Croft, and M. Koll [3] has also been studied expert retrieval in other scenarios, e.g. online question answering communities. People using such services are like a community – anyone can ask, anyone can answer, and everyone can share, since all of the questions and answers are public and searchable immediately.

But there are hundreds of questions asked each day but some portion of them may not be answered or there may be a lag between the time when a question is asked and when it is answered. Also the answers may not be satisfactory.

In year 2006 K. Balog, L. Azzopardi, and M. de Rijke [4] proposed a language model framework for expert search. Expert search aims at retrieving people who have expertise on the given query topic. Their Model 2 is a document-centric approach which first computes the relevance of documents to a query and then accumulates for each candidate the relevance scores of the documents that are associated with the candidate. It locates documents on topic, and then finds the associated expert. Balog showed that Model 2 performed better.

But the nature of these methods is still accumulating relevance scores of associated documents to candidates. Traditional expert search does not explicitly retrieving people who are most likely possessing the desired piece of fine-grained knowledge it focused on finding experts only rather than to mine fine-grained aspects for each task.

In year 2008 R. Jones and K. Klinkner [5] found that search tasks are interleaved and used classifiers to segment the sequence of user queries into tasks. They studied real sessions manually labeled into hierarchical task. They proposed and evaluated a method for the automated segmentation of users' query streams into hierarchical units.

But it considers search engine query logs only, rather than general web surfing contents (including search). Query logs do not record the subsequent surfing activity after the user clicked a relevant search result. Also it dint try to address advisor search by exploiting the data generated from users' past online behaviors.

In year 2010 David M.Ble, Thomas L.Griffiths and Michael I.Jordan [6] presented the nested Chinese restaurant process (nCRP), that assigns probability distributions to infinitely deep branching trees. Proposed Bayesian nonparametric model is based on representations that are allowed to grow structurally as more data are observed.

But analyzing the richly structured data requires extending this approach.

In year 2011 A. Kotov, P. Bennett, R. White, S. Dumais, and J. Teevan [7] designed classifiers to identify same-task queries for a given query and to predict whether a user will resume a task. They introduced and addressed the two problems in the context of analysis of cross-session search tasks: (i) identifying queries from earlier sessions on the same task, and (ii) predicting whether a user will return to the same task during a later session.

But it doesn't provided richer prediction models and alternative feature sets, exploring new prediction and classification problems in the context of cross session information needs. It also didn't tried to mine fine-grained aspects for each task. Summarizing fine-grained aspects can provide a fine-grained description of the knowledge gained by a person.

In year 2015 Ziyu Guan, Shengqi Yang, Huan Sun, Mudhakar Srivatsa, and Xifeng Yan [8] suggested a fine-grained knowledge sharing in collaborative environments. They proposed a method to find proper "advisors" who are most likely possessing the desired piece of fine-grained knowledge based on their web surfing activities.

But the fine-grained knowledge could have a hierarchical structure. And how to search over this hierarchy is not a trivial problem. Also this work creates an issue of privacy.

III. PROBLEM STATEMENT

Working in isolation doesn't helps in discovery of learning and gaining new information in collaborative environment. The creation of knowledge is needed to be aligned with sharing activities. There is a need of sharing knowledge and working together to accomplish stated goals and objectives. Approaching to a right person can be far more efficient than studying by oneself, since people can provide digested information, insights and live interactions, compared to the web.

The problems are:

- Many times people don't get direct links or proper direction to absorb the actual data they want.

- Many people pointed out the problem of finding a right advisor due to the variety of information needs.
- Repeating efforts occurs when people try to acquire similar information on the web in order to gain specific knowledge in one domain.
- These repeating efforts cause waste of time and energy.
- Users are not getting information sharing environment which can enhance their knowledge.
- Studying on the web is not as efficient as learning from an advisor and from the accessed results.

IV. AIM OF PROPOSED METHOD

To address advisor search by exploiting the data generated from user's past online behaviors and also to provide sharing of previously accessed data by another users.

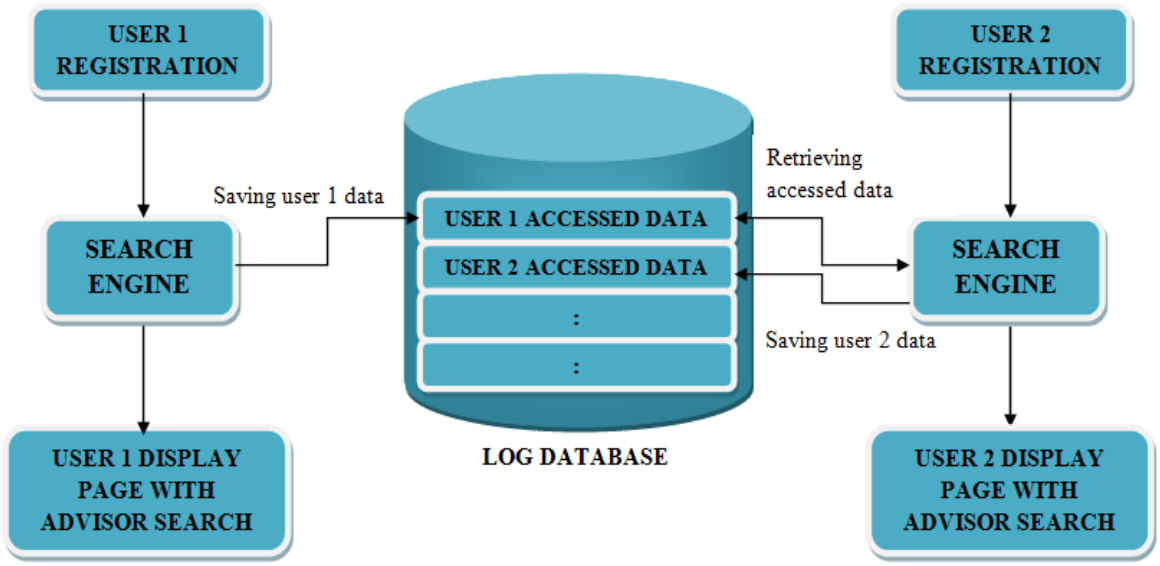
V. THE PROPOSED OBJECTIVES

- To retrieve people who are most likely possessing the desired piece of fine-grained knowledge.
- To summarize fine-grained aspects which can provide a fine grained description of the knowledge gained by a person.
- To provide a healthy sharing environment which can enhance their knowledge
- To analyze user online behaviors to mine the tasks and provide advisor search.
- To provide ease of access of desired information and save time of repetitive efforts.

VI. DESCRIPTION OF PROPOSED METHODOLOGY

This work provides the advisor and the accessed data of that advisor to the user to make the search operation more productive and easy. The goal of this method is not finding domain experts but a person who has the desired piece of knowledge. This methodology provides a way to find proper "advisor" who are most likely possessing the desired piece of fine-grained knowledge based on their web surfing activities. This work proposes the fine-grained knowledge sharing in collaborative environments. This method is proposed to solve the problems by first summarizing web surfing data into fine grained aspects, and then search over these aspects. To this end, a novel discriminative infinite Hidden Markov Model (d-iHMM) is proposed to mine micro-aspects in each task. Also if any person wants to keep his name secret about any topic then the privacy can also be provided. Without showing his name the search made by him can be suggested to another person. This can give a proper direction to ones search. A real user-generated web surfing data is collected to test the feasibility of this method. Then the mined micro-aspects of each task is obtained, advisor search can then be implemented on the collection of learned micro-aspects. Each member is first registered and provided a unique ID which helps to keep record of member's search data and helps in advisor search.

VII. ARCHITECTURE OF PROPOSED METHOD



The architecture is explained step by step as follows:

Step 1: New user is registered first and enters the query to search engine.

Step 2: Here user is asked for privacy whether to click it or not depends on user.

Step 3: User entered web surfing data including queries and user detail is analyzed and saved.

Step 4: This web surfing data is clustered into tasks by a nonparametric generative model.

Step 5: These tasks can be further decomposed into fine-grained aspects (called micro-aspects).

Step 6: Then novel discriminative infinite Hidden Markov Model (d-iHMM)is developed to mine fine-grained aspects in each task and to employ comparison among same searches.

Step 7: Finally, a language model based expert search method is applied over the mined micro aspects for advisor search.

Step 8: If the query matches to the search done previously then the advisor search is provided to the end user.

VIII. DESIRED IMPLICATIONS

- Time consumed in repeating efforts is reduced.
- This method can be implied best in a collaborative environment, where it is common that members try to acquire similar information on the web in order to gain specific knowledge in one domain.
- It provides a good sharing environment in working area along with privacy.
- It can be used for relevant search in collaborative environment.

IX. CONCLUSION

This methodology provides a healthy sharing environment which can enhance people knowledge. It provides an easy way to retrieve people who are most likely possessing the desired piece of fine-grained knowledge by addressing advisor search by exploiting the data generated from user's past online behaviours. It provides access to the previously accessed search of users to make data easily available. It provides ease of access of desired information and save time of repetitive efforts. Also identified digging out fine-grained knowledge reflected by people's interactions with the outside world as the key to solving the problems. This method proposed a two-step framework to mine fine-grained knowledge and integrated it with the classic expert search method for finding right advisors.

ACKNOWLEDGMENT

My thanks to the Guide, Prof. R.R.Keole and Principal Dr.A.B.Marathe, who provided me constructive and positive feedback during the preparation of this paper.

REFERENCES

[1] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The infinite hidden Markov model," in Proc. Adv. Neural Inf. Process. Syst., 2001, pp. 577–584.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., vol. 3, pp. 993–1022, 2003.

[3] X. Liu, W. B. Croft, and M. Koll, "Finding experts in communitybased question-answering services," in Proc. 14th ACM Int. Conf. Inf. Knowl. Manage., 2005, pp. 315–316.

[4] K. Balog, L. Azzopardi, and M. de Rijke, "Formal models for expert finding in enterprise corpora," in Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2006, pp. 43 50.

[5] R. Jones and K. Klinkner, "Beyond the session timeout: Automatic hierarchical segmentation of search topics in query logs," in Proc. 17th ACM Conf. Inf. Knowl. Manage., 2008, pp. 699–708.

[6] David M.Blei, Thomas L.Griffiths,Michael I.Jordan," The Nested Chinese Restaurant Process and Bayesian Nonparametric Inference of Topic Hierarchies" in Journal of the ACM, Vol.57, No.2, Article7, January 2010.

[7] A. Kotov, P. Bennett, R. White, S. Dumais, and J. Teevan, "Modeling and analysis of cross-session search tasks," in Proc. 34th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2011, pp. 5–14.

[8] Ziyu Guan; Shengqi Yang; Huan Sun; Srivatsa, M.; Xifeng Yan, "Fine-Grained Knowledge Sharing in Collaborative Environments," in Knowledge and Data Engineering, IEEE Transactions on , vol.27, no.8, pp.2163-2174, Aug. 1 2015

AUTHOR PROFILE

Nilofar N. Pathan, received the B.E.degree in Information Technology from H.V.P.M's College Of Engineering And Technology, Amravati in 2014. She is currently persuing Master's Degree in Computer Science and Engineering from H.V.P.M's College of Engineering And Technology, Amravati.

Prof. Ranjit R.Keole, received the B.E.and M.E degree in Computer Science from Prof. Ram Meghe Institute of Technology, Badnera in 1992 and 2008, respectively. His field of specialisation is web Mining. He is currently working as Associate Professor at H.V.P.M's college of Engineering and Technology,Amravati.