Adviosor Search Methodology For Fine Grained Knowledge Sharing

Suvarna Hiwale[1], Prof. Smita Ponde [2]
Department of Computer Science & Engineering
[1, 2]Deogiri Institute of Engineering & Management StudiesAurangabad, India

Abstract-*In the era of collaborative environment members may try to get access of similar type of information related to same domain. For example in research lab one member may want to solve the problemusing graphic modeling which is not familiar to him so he can resort to another person who has studied sametopic before for efficiently studying that topic than studying by itself. We have fine grainedknowledge sharing framework which has two steps: 1) web surfing data is clustered into tasks by a nonparametric generative model; (2) a new infinite Hidden Markov Model 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*

Keywords-Fine grained knowledge sharing; Advisor Search, Infinite Hidden Markov Model, Collaborative environment, nonparametric generative model.

I. INTRODUCTION

In a collaborative environment, it is common thing that members try to acquire similar information on the web in order to gain specific knowledge in same domain. We propose here a new method to identify, how to enable such knowledge sharing mechanism by analyzing user data. For example, Alice starts to surf the web and wants to learn how to develop a Java multithreading program, which has already been studied by Bob. In this case, it might be a good idea to consult Bob, rather than studying by her. Such recommendations are provided for the ease of user with this methodology by analyzing surfing activities automatically. In this example, not necessarily Bob is an expert in every aspect of Java programming; however, due to his significant surfing activities in Java multithreading, it becomes easier to assume that he has gained enough knowledge in this area so that he can help Alice.

This method is from the customary expert search difficulty in that expert search goals to find domain experts based on their documents in an project storehouse, while the goal of this project work is to find proper "advisors" who are most likely having the desired piece of fine grained mined knowledge based on their web surfing activities. In order to get the knowledge gained by web users, new method is proposed to monitor and study user's web surfing data. User's interactions with the web can be divided into different "tasks". Documentary contents of a task are usually structured. This paper defines a session as a aggregation of consecutively surfed web contents of a user that belong to the same domain. Sessions are the atomic units in our examination. A task can be further divided into fine-grained aspects (called micro-aspects). A micro-aspect could be more cohesive subset of sessions in a task. In above example, learning java can be a task while java I/O and java multithreading can be micro aspects. This is a two step framework. The first one is to formulate task from sessions. In this we have used Infinite Gaussian Mixture model, based on dirichlet process to cluster sessions. The second one is to extract micro aspects from sessions in each task. By using these two steps it is going to mine knowledge. Finally, a language model based expert search method is applied over the mined micro aspects for advisor search.

II. LITERATURE SURVEY

A. T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," It demonstrate the use of a deviation of the nonparametric Bayesian (NPB) forward-backward (FB) method for case state sequences of hidden Markov models, when the continuous-valued observations follow autoregressive (AR) processes. It shows that by roughly integrating out some parameters of the model, one can ease this problem considerably. Drawbacks-A Bayesian analysis approach is fruitful in many ways but it has rather been unsuccessful in terms of non-parametric problems.

B. In year 2003 D. M. Blei, A. Y. Ng, and M. I. Jordan [1] 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.

C. H. Deng, I. King, and M. R. Lyu, "Formal models for expert finding." Finding relevant experts in a specific field is often crucial for, both in industry and in academia. The aim of this paper was to address the expert-finding task in a real world academic field. Evaluation results show that the hybrid model outperforms other models in few metrics. Disavow:-Evaluation results show that the hybrid model outperforms other models only in particular types of metrics but failed to work in all.

D. In year 2005 X. Liu, W. B. Croft, and M. Koll [2] 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.

E. In year 2011 A. Kotov, P. Bennett, R. White, S. Dumais, and J. Teevan [3] 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.

III. PROPOSED SYSTEM

The proposed system is based on advisor search problem which is different from traditional expert search. (1) Advisor search is keen to retrieving people who are most likely having the desired piece of fine-grained knowledge, while traditional expert search does not overtly take this goal. (2) The critical difference lies in the data. This paper is about mine micro aspects and show the advantage of our search scheme over the simple idea of applying traditional expert search methods on session data directly. A person typically generates multiple sessions for micro aspects of a task. In example stated above, person could spend many sessions in learning about java multithreading skills. This line of research tries to recover tasks from peoples search behavior. Web surfing data provides more comprehensive information about knowledge gaining activities of users.

Recently researchers have focused on detecting, modeling and analyzing user search tasks from query logs. The search tasks are interleaved and used classifiers to segment the sequence of user queries into tasks. First, we consider general web surfing contents (including search), rather than search engine query logs. Query logs do not record the subsequent surfing activity after the user clicked a relevant search result. Web surfing data provides more comprehensive information about the knowledge gaining activities of users. Even if various methods were projected for finding search process in logs, these methods cannot be applied in our setting since they exploit query log specific properties[4]. Second, none of the above works tried to mine fine-grained aspects for each task. When studying, people could spend some effort on one fine-grained aspect of a task and generate multiple contents. Summarizing fine-grained aspects can provide a fine- grained description of the knowledge gained by a person. Finally, none of existing works which analyze user online behaviors address advisor search by exploiting the data generated from users' past online behaviors.

We refer to a member in a collaborative environment as a "candidate". In our problem, we have a group of h candidates {e1, e2,….eh} Where each candidate generate sequence of sessions. It has following three sub problems

1. Partition sessions into set of clusters, where each cluster represents a task.
2. Partition sessions in each cluster into a set of micro aspect where micro aspect is subset of cohesive sessions.
3. Compute association weight between candidate and session.

The aim of our system is find a person who has the desired piece of knowledge rather than domain experts. The proposed method provides technique to find proper "advisors" who are most likely having 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. First the user entered web surfing data including queries and name is analyzed and extracted. Following diagram shows the architecture of system.

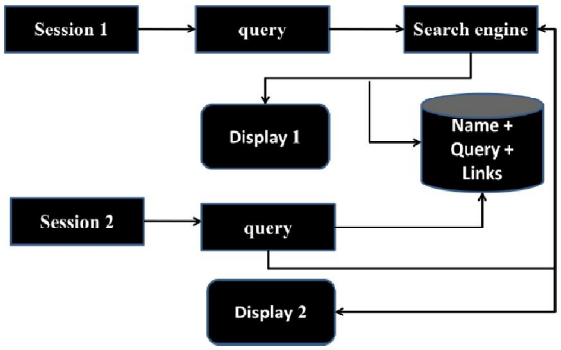

Fig: 1 Architecture diagram of system

The working can be elaborate as follows:

1. First, User will start the session 1 with his /her name and enters the query in search engine.
2. Search engine then gives the information and as a background task, database is created with name, query entered and related searched links.
3. When next user comes with next session 2 the query is first compared with queries in our database.
4. If it is equivalent then the stored data is suggested and if not then new display is provided with required topic.

IV. CONCLUSION AND FUTURE SCOPE

We introduced the system "Advisor Search methodology for Knowledge Sharing", which is desirable in practice. We have mined knowledge reflected by people from web surfing data. This system finds the url to related query which is hitter maximum number of time and the advisor who has searched that url more number of time. 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. There are open issues for this problem. (1) The fine grained knowledge could have a hierarchical structure. For example, "Java IO" can contain "File IO" and "Network IO"as sub-knowledge. We could iteratively apply d-iHMM on the learned micro-aspects to derive a hierarchy, but how to search over this hierarchy is not a trivial problem.

ACKNOWLEDGMENT

Thanks to the analysts and also distributors for making their resources accessible to us. We would like to thank our internal guide professor Smita Ponde for helping us and for her guidance and support we needed. Their valuable suggestions were very helpful.

REFERENCES

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

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

[3] 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.

[4] 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.