Knowledge Sharing in Collaborative Enviornment Using Advisor Search

Suvarna Hiwale[1], Prof. Smita Ponde[2]
[1, 2] Dept of computer science & Engineering
[1, 2] Deogiri Institute of Engineering & Management Studies
Aurangabad, India

Abstract- *In the collaborative environment members may try to get access of information on internet which is of similar type related to same domain. Therefore knowledge sharing becomes the important aspect. For example in an organization several department want ro buy business software for that they search on the internet then it will be beneficial for them to be connected and share a network for their studies. It reduces their time and efforts to get to know about the software. We have a framework with knowledge sharing in collaborative environment with the 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, d-iHMM model, Dirichlet process.

I. INTRODUCTION

In a collaborative environment, it is regular practice that members are trying to acquire similar information on the web in order to gain specific information in one domain. We present a new method how to identify, how to enable such knowledge sharing. A two-step methodology is proposed for mining knowledge: (1) Web surfing data is clustered into tasks (Domain) 2) Mine micro aspects (Sub domain) in each task. This method is different from the traditional expert search in that expert search aims to find domain experts based on their associated documents in an enterprise repository, while the goal of this proposed work is to find proper “advisors” who are most likely possessing the desired piece of knowledge based on their web surfing activities. 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., “Java Studying” and “shopping”. Textual contents of a task are usually cohesive. This system has session as an aggregation of consecutively browsed web contents of a user that belong to the same task. Sessions are atomic units in our analysis. A task is further decomposed into micro-aspects (Sub Domain). A micro-aspect could be roughly defined as a significantly more cohesive subset of sessions in a task. For example, the task “Learning java” might contain “Java I/O” and “Java multithreading”. For example, User2 starts to surf the web and searching for advantages, which has already been studied by User3. In this case, it might be a good idea to consult User3, rather than studying by herself. Such recommendations are provided with this methodology by analyzing surfing activities automatically. In this example, not necessarily User3 is an expert in every aspect of Java multithreading; however, due to his significant surfing activities in Operating system advantages, it is fair to assume that he has gained enough knowledge in this area so that he can help User2. In this case, resorting to a right person could be far more efficient than studying by oneself, since people can provide digested information, insights and live interactions, compared to the Web.

II. LITERATURE SURVEY

The agenda of literature review is to do the thoroughly study of the topic, its meaning, selected relevant documents. The basic starting point of survey is the review of any research process regarding topic itself. We get the idea instantly by checking review from other people about project topic.

Management of knowledge resources has turn out to be a strong demand for development. Discovering the valuable knowledge has also significant approach for management and decision making. In information retrieval research, retrieval can be seen as the main task in interacting with an information resource, not browsing. The capability to tailor retrieval by means of obtaining user response to retrieved items has been implemented in several information retrieval systems through retrieval clustering (Cutting, et al., 1992) and through relevance feedback.

Expert search aims at retrieving people who have expertise on the given query topic. Early approaches involve building a knowledge base which contains the descriptions of people's skills within an organization [1]. Expert search became a hot research area since the start of the TREC enterprise track [2] in 2005. Balog et al. proposed a language model framework for expert search. 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.

Ziyu Guan, Shengqi Yang, Huan Sun, Mudhakar Srivatsa, and Xifeng Yan [3] have proposed "Fine-Grained Knowledge Sharing in Collaborative Environments". In paper they explain new method of expert finding and implement it. They find out expert using web surfing and browsing contents. Web surfing data gives more accurate results than traditional document based method. Gaussian Dirichlet process mixture model is used for clustering session in each task. In order to implement to mine micro aspects in each task discriminative infinite Hidden Markov Model is used. Krisztian Balog [4] and Group have presented paper on "Formal models for expert finding in enterprise Corpora". Two general strategies have been presented to expert searching given a collection of document. The first directly models an expert's knowledge based on the documents that they are associated with and while the second locates documents on the queried topic and then finds the associated expert. Carl Edward Rasussen [5] explained "The Infinite Gaussian Mixture Model ". In which author show that how infinite mixture model has several advantages over finite mixture model. Jurgen Van Gael , Yunus Saatci ,Yee Whye The , Zoubin Ghahramani given paper on "Beam Sampling for the Infinite Hidden Markov Model ".Author proposes Dynamic Expertise Modeling from Organizational Information Resources Server approach to formulate expert finding system that meets the said requirements. Finding experts in collaborative environment to share information has lots of advantages; early approach to advisor search is that data about skills and knowledge of each individual in the organization is collected. This data is stored in the database manually, such an approach requires considerable effort to set up and maintain the data [6]. More recent techniques judge the expert automatically. The task of expert search received significant importance since it has been implemented in the TREC enterprise track..In expert identification query is passed related to the topic X and the relevant documents are generated as output. In the output the relevance of each document is found and those documents have the highest relevance will be ranked high. From the highest rank document it will be easy to recognize advisor. Yi Fang, Luo Si and Aditya P. Mathur had published article on "Discriminative models of integrating document evidence and document-candidate associations for expert search". The proposed research can naturally integrate various documental evidence and candidate-document associations into a single model without extra modeling assumptions or effort.

III. PROPOSED SYSTEM

This research investigates a framework for finding proper advisor while surfing in internet by using micro aspect mining. The Gaussian mixture model is used on Dischrete process for clustering the sessions then expert search is applied over it to find out the proper advisor based on their surfing activity. Moreover this search motivated the hidden morkov model which is applied on the sessions for their clusters for finding more relevant data from sessions. These sessions are then clustered into tasks for the further procurement. These are then mined by non parametric generative model. Finally the classic search method is applied to find proper advisor for relevant knowledge sharing. We propose a two-step framework for mining finegrained knowledge (micro-aspects): (1) In the first step, we formulate tasks from sessions. We design an infinite Gaussian mixture model based on Dirichlet Process (DP) to cluster sessions. We adopt a nonparametric scheme since the number of tasks is difficult to predict. (2) We then extract micro-aspects from sessions in each task. The challenges are: the number of micro-aspects in a task is unknown; sessions for different micro-aspects of a task are textually similar; sessions for a micro-aspect might not be consecutive. To this end, a novel discriminative infinite Hidden Markov Model (d-iHMM) is proposed to mine micro-aspects and evolution patterns (if any) in each task.

- **Mining fine grained knowledge**

The major challenge of mining micro-aspects is that the micro-aspects in a task are already similar with one another. If we model each component (i.e. micro-aspect) independently (as most traditional models do), it is likely that we mess up sessions from different micro-aspects, i.e. leading to bad discrimination. \To this end, we extends the infinite Hidden Markov Model (iHMM) and propose a novel discriminative infinite Hidden Markov Model to mine micro-aspects and possible evolution patterns in a task. The graphical representation is shown in fig. The observed variables of the model is a sequence of T sessions $\{W_t\}_{t=1}^{T}$ belonging to a cluster outputted by the GDP. In practice, a cluster can contain multiple sequences from different people.

- **Solving d-iHMM by Beam Sampling**

Beam sampling adopts the slice sampling idea to limit the number of states considered at each time step to a finite number, so that dynamic programming can be used to sample whole state trajectories efficiently. Specifically, we first sample an auxiliary variable µt for each time step t in the sequence with conditional distribution µt ~(Uniform $\Pi_{t,}$ Π_{t-1}). Given µt, the sequence S_t is resample, considering only the values of st_1 that satisfy Π_t> ut. Hence, a truncation of the infinite number of states and make the number of trajectories with positive probabilities finite, so that the whole sequence can be sampled holistically. Each sampling iteration samples {μ_t}, {S_t}, {Z_{ti}},{Θ_k} and ß in turn. We first sample {μ_t} as described above and create more states if the maximum unassigned probabilities in {Π_t} (the last element of each Π_t) is greater than the minimum of μ_t. Then we perform a forward sweep of S_t Where for the t^{th} step we compute

$$\begin{aligned} & p(s_t|w_{1:t}, u_{1:t}, z_t) \\ & \quad \propto p(s_t, u_t, w_t|w_{1:t-1}, u_{1:t-1}, z_t) \\ & \quad = p(w_t|s_t, z_t) \\ & \qquad \times \sum_{s_{t-1}} \mathbb{I}(u_t < \pi_{s_{t-1}s_t}) p(s_{t-1}|w_{1:t-1}, u_{1:t-1}, z_{t-1}), \end{aligned}$$

Recently researchers have focused on detecting, modeling and analyzing user search.

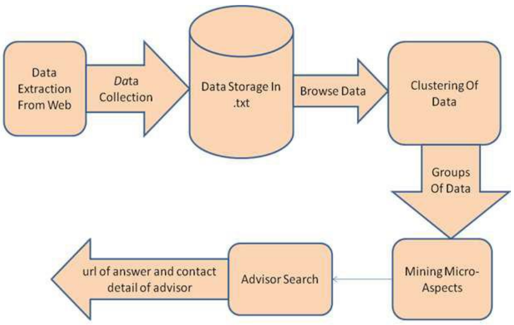


Fig: 1 Architecture diagram of system

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.

Expert search aims at retrieving people who have expertise on the given query topic. Early approaches involve building a knowledge base which contains the descriptions of people"s skills within an organization. Expert search became a hot research area since the start of the TREC Balog et al. proposed a language model framework for expert search.

The experimentation is based on calculating the processing time required to searching the advisor from the complete search. The evaluation methodology is as follows: we generate 20 queries for each dataset Queries are shown in Table 1 as some examples. The ground truth labels of a query are obtained by showing candidates their top relevant sessions assessed by the language model method and asking them to assign a relevance score to themselves for the data collection period on a scale from 0 to 2: (1) score= 0 means "irrelevant"; (2) score = 1 means "partially relevant", i.e. he/she has background knowledge for the query; (3) score= 2 means "very relevant". Normalized discount cumulative gain (NDCG) is used as the evaluation metric:

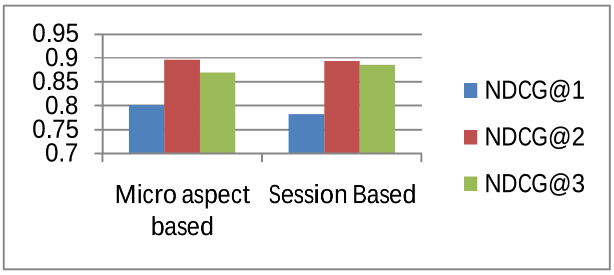
$$\mathrm{NDCG@n} = Z_n \sum_{i=1}^{n} \frac{2^{r_i} - 1}{\log_2(i+1)}$$

Where ri is the relevance score of the candidate at rank i and Zn is anormalization term to let the perfect ranking have a NDCG value of 1.

The results are shown in Tables 2 and 3.We can see that the micro-aspect-based scheme outperforms the other two schemes.

Method	NDCG@1	NDCG@2	NDCG@3
Micro aspect based	0.801	0.896	0.87
Session Based	0.783	0.893	0.885

Table: Above table shows the NDCG rate of n, where n is the number of words in query on both micro aspect and session based method



Based on the above observation, we have represented the readings in the graphical format as above based on both aspects.

IV. CONCLUSION AND FUTURE SCOPE

We introduced the system "Knowledge Sharing in collaborative enviornmentUsing Advisor Search", 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.

V. 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] P. R. Carlile, "Working knowledge: How organizations manage what they know," Human Resource Planning, vol. 21, no. 4, pp. 58– 60, 1998.
[2] N. Craswell, A. P. de Vries, and I. Soboroff, "Overview of the TREC 2005 enterprise track," in Proc. 14th Text REtrieval Conf., 2005, pp. 199–205
[3] Ziyu Guan, Shengqi Yang, Huan Sun, Mudhakar Srivatsa, and Xifeng Yan "Fine-Grained Knowledge Sharing in Collaborative Environments". IEEE transactions on Knowledge and data Engineering, vol.27,No. 8,August 2015, pp.2163-2174 17th ACM Conf. Inf. Knowl. Manage., 2008, pp. 699–708.
[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.
[5] J. Van Gael, Y. Saatci, Y. Teh, and Z. Ghahramani, "Beam sampling for the infinite hidden Markov model," in Proc. Int. Conf. Mach. Learn., 2008, pp. 1088–1095