Expert Finding in A Social Network

Jing Zhang, Jie Tang, and Juanzi Li

Department of Computer Science and Technology, Tsinghua University
10-201, East Main Building, Tsinghua University, Beijing, China, 100084
{zhangjing, tangjie, ljz}@keg.cs.tsinghua.edu.cn

Abstract. This paper addresses the issue of expert finding in a social network. The task of expert finding, as one of the most important research issues in social networks, is aimed at identifying persons with relevant expertise or experience for a given topic. In this paper, we propose a propagation-based approach that takes into consideration of both person local information and network information (e.g. relationships between persons). Experimental results show that our approach can outperform the baseline approach.

1. Introduction

Expert finding is one of the most important subjects for mining from (web-based) social networks. The task of expert finding is aimed at identifying persons with relevant expertise or experience for a given topic.

Several research efforts have been made for expert finding. However previous works usually studied the person local information and relationships separately and combined them in an ad-hoc approach. For example, Campbell et al. [3] have investigated the issue of expert finding in an email network. They utilize the link between authors and receivers of emails to improve the expert finding result. Link structure-based algorithms, for example PageRank [7] and HITS [6], can be used to analyze the relationships in a social network, which might improve the performance of expert finding. However, both PageRank and HITS have a common problem: topic drift, which makes the most in-links in the network tend to dominate [8].

Many approaches focus on finding the experts from the Web. For example, Balog et al. view the task of expert finding as that of information retrieval [1]. They propose using two language models to search experts on the Web. TREC'2005 and TREC'2006 have provided a common platform for researchers to empirically assess methods and techniques devised for expert finding. However the number of candidate experts is limited, for example only 1092 in TREC'2005, while in a network everyone can be the candidate, thus the number of candidate experts is up to millions. In addition, in the Web only unstructured data is available, the candidates' names and the topics are presented as keywords in the plain texts, while a social network contains not only person local information but also complex relationships.

In this paper, our focus is how to make use of person local information and relationships between persons in a unified approach. We proposed a propagation-based approach for finding expert in a social network. The approach consists of two steps. In the first step, we make use of person local information to estimate an initial expert score for each person and select the top ranked persons as candidates. The selected persons are used to construct a sub-graph. In the second step, we propose a propagation-based approach, which propagates one's expert score to the persons with whom he/she has relationships.

We use an academic researcher network as the experimental data, which is built automatically using information extraction approaches [9]. In the network, a person can have different types of information: person profile, contact information, and publications/documents. The relationship in the network is coauthor. In total, we gathered 448,289 persons and created 2,413,208 coauthor relationships between them. We also collected 725,655 publications of the researchers into the network.

2. Expert Finding in A Social Network

Expert finding addresses the task of finding the right persons with the appropriate skills and knowledge: "Who are the experts on topic X?"

Formally, a social network can be defined as a graph $G = (V, E)$, where $v \in V$ represents a person in the social network and $e^t_{ij} \in E$ represents a relationship with type t between persons v_i and v_j. (t can be, for example, coauthor or colleague) The task of expert finding is defined as: given a query topic q, it is to *find* a subset of the persons from the social network and return them in a *ranked* list.

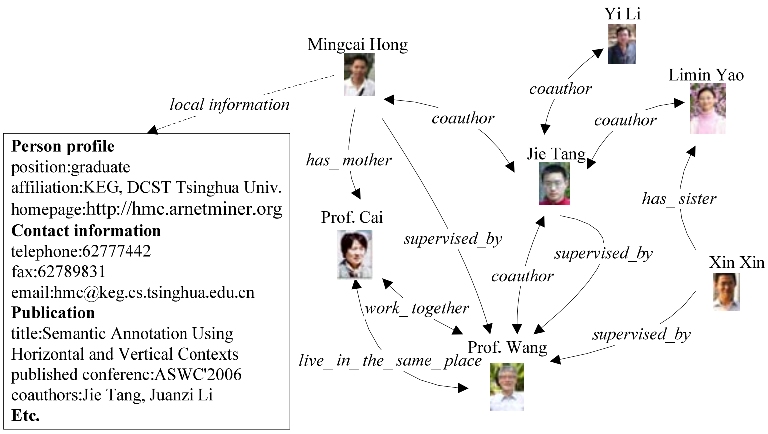

Fig. 1. An example of academic researcher network

Figure 1 shows a snippet of the academic researcher network. In the network, each person has several types of *local information*, for example, personal profile, contact information, and publications. Two persons can have relationships with each other. The relationship can be directional or bi-directional. In Figure 1, "Jie Tang" has one out-relationship (i.e. a *supervised_by* relationship with "Prof. Wang") and four bi-directional relationships (e.g. a *coauthor* relationship with "Mingcai Hong"). Two persons in the social network may have more than one relationship, for example, "Jie Tang" and "Prof. Wang" have two relationships, *supervised_by* and *coauthor*.

In this paper, we propose a new approach to expert finding in a social network which takes into consideration of not only person local information but also relationships between persons. It consists of two steps, Initialization and Propagation.

In Initialization, we use the person local information to calculate an initial expert score for each person. The basic idea in this stage is that if a person has authored many documents on a topic or if the person's name co-occurs in many times with the topic, then it is likely that he/she is a candidate expert on the topic. Our strategy for calculating the initial expert scores is based on the probabilistic information retrieval model. For a person, we first create a 'document' *d* by combining all his/her person local information. We estimate a probabilistic model for each 'document' and use the model to calculate the relevance score of the 'document' to a topic. The score is then viewed as the initial expert score of the person.

In Propagation, we make use of relationships between persons to improve the accuracy of expert finding. The basic idea here is that if a person knows many experts on a topic or if the person's name co-occurs in many times with another expert, then it is likely that he/she is an expert on the topic. Based on the propagation theory [5], we propose a propagation-base approach.

We view the social network as a graph. In the graph, we assign a weight on each edge to indicate how well the expert score of a person propagates to its neighbors and back. These so-called propagation coefficients range from 0 to 1 inclusively and can be computed in many different ways.

In general, the expert score $s(v_i)^{i+1}$ is computed from $s(v_i)^i$ as follows (normalization is omitted for clarity):

$$s(v_i)^{n+1} = s(v_i)^n + \sum_{v_j \in U} \sum_{e \in R_{ji}} w((v_j, v_i), e) s(v_j)^n \tag{1}$$

where $w((v_j, v_i), e)$ represents the propagation coefficient and $e \in R_{ji}$ is one kind of relationship from the person v_j to v_i; U stands for all neighboring nodes to v_i in the graph and R_{ji} stands for all relationships from the person v_j to v_i.

The approach runs in iterations. After propagation in each iteration, all expert scores are normalized, i.e., divided by the maximal expert score of current iteration. So far, we define the terminal conditions as: if the maximal change of the expert score is below a predefined threshold ε for some n>0 or the iteration times exceed a predefined number (defined as 100 in our experiments), then stop the propagation.

The approach has good convergence property. We omitted the proof due to the space limitation.

3. Experimental Results

To evaluate our proposed approach, we collected 13 test sets, each of which consists of a topic and a list of experts (http://keg.cs.tsinghua.edu.cn/project/PSN/dataset.html). We defined a baseline approach (called Baseline hereafter) for expert finding using only person local information.

We carried out the experiments as follows. We first created a 'document' for each person by combining all of his/her person local information. Given a topic, we use the traditional probabilistic IR method to estimate the relevance score of a 'document' to the topic as the initial expert scores for the second stage of our approach, at the same time as the final score in Baseline. Next, we conducted propagation in the graph to update the expert score of each person based on the equation (1) (the convergence threshold ε is set as 0.025). Finally, we outputted the ranked persons as results.

We evaluated the found results in terms of the Precision@5, Precision@10, Precision@20, Precision@30, R- Prec, mean average precision (MAP), and bpref [2] [4]. Table 2 shows the experimental results on the 13 topics.

Table 1. The average expert search result of 13 topics with the runs from two approaches (%)

Approach	Precision@5	Precision@10	Precision@20	Precision@30	R-prec	*MAP*	*bpref*
Baseline	46.15	38.00	35.80	32.82	34.60	9.73	13.64
Our Approach	61.54	48.00	40.40	36.15	37.82	11.03	16.11

From the evaluation results, we can see that our approach significantly outperforms Baseline in terms of all evaluation measures, which indicates that the relationships can be very useful for finding experts in a social network.

4. Conclusions

In this paper, we proposed a propagation-based approach to expert finding in a social network. The approach takes into consideration of both person local information and relationships. Experimental results show that the proposed approach performs better than the baseline.

References

[1] K. Balog, L. Azzopardi, and M. de Rijke. Formal Models for Expert Finding in Enterprise Corpora. In Proc. of SIGIR ’06. 2006, pp.43-55

[2] C. Buckley, E.M. Voorhees. Retrieval Evaluation with Incomplete Information. In Proc. of SIGIR ’04. 2004, pp.25-32

[3] C. S. Campbell, P. P. Maglio, A. Cozzi, B. Dom. Expertise Identification Using Email Communications. In Proc. of CIKM’03. 2003, pp.528-531

[4] N. Craswell, A. de Vries, and I. Soboroff. Overview of the Trec-2005 Enterprise Track. TREC 2005 Conference Notebook. 2005, pp.199-205

[5] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient Belief Propagation for Early Vision. International Journal of Computer Vision, Vol. 70, No. 1, October 2006.

[6] J. M. Kleinberg . Authoritative Sources in a Hyperlinked Environment. In Proc. of the 9th SODA, 1998. pp.668-677.

[7] L. Page, S. Brin, R. Motwani and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Stanford Digital Library working paper SIDL-WP-1999-0120, Nov.1999

[8] M. Richardson and P. Domingos. The Intelligent Surfer: Probabilistic Combination of Link and Content Information in PageRank. In Proc. of NIPS’01. pp.1441-1448

[9] J. Tang, M. Hong, J. Zhang, B. Liang, and J. Li. A New Approach to Personal Network Search based on Information Extraction. Demo paper. In Proc. of ASWC’2006.