LExL: A Learning Approach for Local Expert Discovery on Twitter

Wei Niu, Zhijiao Liu, James Caverlee

Department of Computer Science and Engineering, Texas A&M University
College Station, Texas, USA
{wei,lzj,caverlee}@cse.tamu.edu

Abstract. In this paper, we explore a geo-spatial learning-to-rank framework for identifying local experts. Three of the key features of the proposed approach are: (i) a learning-based framework for integrating multiple factors impacting local expertise that leverages the fine-grained GPS coordinates of millions of social media users; (ii) a location-sensitive random walk that propagates crowd knowledge of a candidate's expertise; and (iii) a comprehensive controlled study over AMT-labeled local experts on eight topics and in four cities. We find significant improvements of local expert finding versus two state-of-the-art alternatives.

1 Introduction

Identifying *experts* is a critical component for many important tasks. For example, the quality of movie recommenders can be improved by biasing the underlying models toward the opinions of experts [1]. Making sense of information streams – like the Facebook newsfeed and the Twitter stream – can be improved by focusing on content contributed by experts. Along these lines, companies like Google and Yelp are actively soliciting *expert reviewers* to improve the coverage and reliability of their services [7].

Indeed, there has been considerable effort toward expert finding and recommendation, e.g., [2, 3, 6, 10, 11]. These efforts have typically sought to identify *general topic experts* – like the best Java programmer on github – often by mining information sharing platforms like blogs, email networks, or social media. However, there is a research gap in our understanding of *local experts*. Local experts, in contrast to general topic experts, have specialized knowledge focused around a particular location. Note that a local expert in one location may not be knowledgeable about a different location. To illustrate, consider the following two local experts:

• A "health and nutrition" local expert in Chicago is someone who may be knowledgeable about Chicago-based pharmacies, local health providers, local health insurance options, and markets offering specialized nutritional supplements or restricted diet options (e.g., for gluten allergies or strictly vegan diets).

• An "emergency response" local expert in Seattle is someone who could connect users to trustworthy information in the aftermath of a Seattle-based disaster, including evacuation routes and the locations of temporary shelters.

Identifying local experts can improve location-based search and recommendation, and create the foundation for new crowd-powered systems that connect people to knowledgeable locals. Compared to general topic expert finding, however, there has been little research in uncovering these local experts or on the factors impacting local expertise.

Hence, we focus on developing robust models of *local expertise*. Concretely, we propose and evaluate a geo-spatial learning-to-rank framework called **LExL** for identifying local experts that leverages the fine-grained GPS coordinates of millions of Twitter users and their relationships in Twitter lists, a form of crowd-sourced knowledge. The framework investigates multiple classes of features that impact local expertise including: (i) user-based features; (ii) list-based features; (iii) local authority features; and (iv) features based on a location-sensitive random walk that propagates crowd knowledge of a candidate's expertise.

Through a controlled study over Amazon Mechanical Turk, we find that the proposed local expert learning approach results in a large and significant improvement in Precision@10, NDCG@10, and in the average quality of local experts discovered versus two state-of-the-art alternatives. Our findings indicate that careful consideration of the relationships between the location of the query, the location of the crowd, and the locations of expert candidates can lead to powerful indicators of local expertise. We also find that high-quality local expert models can be built with fairly compact features.

2 Learning Approach to Local Expert Finding

In this section, we introduce the learning approach framework for finding local experts – **LExL**: **L**ocal **Ex**pert **L**earning. Given a query, composed of a topic and a location, the goal of LExL is to identify high-quality local experts. We assume there is a pool of local expert candidates $V = \{v_1, v_2, ..., v_n\}$, each candidate is described by a matrix of topic-location expertise scores (e.g., column i is Seattle, while row j is "web development"), and that each matrix element indicates to what extent the candidate is an expert on the corresponding topic in the corresponding location. Given a query q that includes both a topic t and a location l, our goal is to find the set of k candidates with the highest local expertise in query topic t and location l. For example, find the top experts on t_q = "web development" in l_q = Seattle, WA.

Learning Approach. We propose to address the local expert ranking problem with a supervised learning-to-rank framework that can combine any number of local expertise features, using a tool such as LambdaMART [12]. The basic idea of LambdaMART is to train an ensemble of weak models and to linearly combine the prediction of them into a final model which is more accurate. But what features should we investigate?

We propose four classes of features that potentially contribute to local topic expertise of a user. Compared to much existing work on expert finding that is content based (e.g., [9]), we focus on features that are independent of what a candidate has posted, instead relying on activity and network based features.

We anticipate integrating content-based features in our future work. In total, we focus on 25 features. Here we briefly introduce the features we used.

User-based Features. This group of features capture user-oriented aspects that are independent of the query topic and query location. Three aspects of information are considered. user's network (for example, number of follower), user's activity (for example, number of status), and longevity (for example, how long the user has joined).

List-based Features. We extract expertise evidence directly from the Twitter list, but ignoring the geo-spatial features of the lists (those aspects are part of the following two groups of features). Twitter lists have been recognized as a strong feature of expertise in previous work [6]. In particular, lists can shed light on a candidate from two perspectives: appearing on list and maintaining the list. We also defined a feature to characterise the quality of the list.

Local Autority Features. These features focus on the local authority of candidates revealed through the geo-located Twitter lists. The main idea is to capture the "localness" of these lists. Intuitively, a candidate who is well-recognized near a query location is considered more locally authoritative. We measure the local authority of a candidate in multiple ways, like Haversine distances among candidate, labeler and query location. We also adopt Candidate Proximity and Spread-Based Proximity as two features [4].

Distance-Biased Random Walk Features. We introduce a set of features that incorporate additional network context beyond these one-hop relationships in local authority features. Concretely, we explore features based on a random walk model that directly incorporates the query location, the location of a candidate expert, and the location of external evidence of a candidate's expertise (e.g., in the case of the Twitter lists, the location of the list labeler). The main intuition is to bias a random walker according to the distances between these different components (the query location, the labeler, the candidate) for propagating local expertise scores. In this way, each candidate can be enriched by the network formed around them via Twitter lists. We have a total of 7 features by considering distance among candidate, labeler and query location.

3 Evaluation

In this section, we present the experimental setup, including the collection of ground truth data via AMT, alternative local expert ranking methods, and metrics for comparing these methods, followed by experimental results and analysis.

3.1 Experimental Setup

Our experiments rely on the dataset described in [4], totaling 15 million list relationships in which the coordinates of labeler and candidate are known.

Queries. We adopt a collection of eight topics and four locations that reflect real information needs. The topics are divided into broader local expertise topics – "food", "sports", "business", and "health" – and into more specialized local expertise topics which correspond to each of the broader topics – "chefs", "football", "entrepreneur", and "healthcare". The locations are New York City, San Francisco, Houston and Chicago, which all have relatively dense coverage in the dataset for testing purposes. For each method tested below, we retrieve a set of candidates for ranking based on topics derived from list names.

Proposed Method: Local Expert Learning (LExL). There are a wide variety of learning-to-rank approaches possible; in this paper we evaluate four popular learning-to-rank strategies: Ranknet, MART, Random Forest and LambdaMART. We use an open source implementation of these methods in the RankLib toolkit. For each topic, we randomly partition the collected candidates together with their four categories of features into two equal-sized groups for training and testing. We use four-fold cross validation for reporting the results. We compare our proposed approach with two state of the art approaches for finding local experts: Cognos+ [6] and LocalRank [4].

Ground Truth. Since there is no publicly-available data that directly specifies a user's local expertise given a query (location + topic), we employ human raters (turkers) on Amazon Mechanical Turk to rate the level of local expertise for candidates via human intelligent tasks (HITs). In total, we collect 16k judgments about user's local topic expertise in a scale of 5 (0-4) across the eight topics and four locations. To explore the validity of turker judgments, we measure the kappa statistic [5], where a value of 0.46 on average means "moderate agreement."

Evaluaton Metrics. To evaluate the quality of local expertise approaches, we adopt two well-known metrics Precision@k, NDCG@k. We also use Rating@k for a query pair to measure the average local expertise rating by the turkers for the top-k experts output by each approach, defined as: $R@k = \sum_{i=1}^{k} rating(c_i, q)/k$, where c is candidate and q is the query pair. In our scenario, we set k=10.

3.2 Results

Comparison versus Baselines. We begin by comparing the proposed learning method (LExL) versus the two baselines. Figure 1 shows the Precision@10, Recall@10, and NDCG@10 of each method averaged over all queries.[1] We consider the LambdaMART version of LExL, in addition to methods using Ranknet, MART and Random Forest. First, we observe that three versions of LExL clearly outperform all alternatives, resulting in a Precision@10 of around 0.78, an average Rating@10 of more than 3, and an NDCG of around 0.8.

Cognos has been shown to be effective at identifying topic experts. However, we see even a modified version to include distance factor is not compatible with local expert finding. For example, Cognos may identify a group of "healthcare" experts known nationwide, but it has difficulty uncovering local experts.

[1] Note that the results reported here for LocalRank differ from the results in [4] as the experimental setups are different. First, our rating has 5 scales, which is intended to capture more detailed expertise level. Second, [4] only considers ideal ranking order for the top 10 results from LocalRank when calculating maximum possible (ideal) DCG@10, while we consider a much larger corpus.

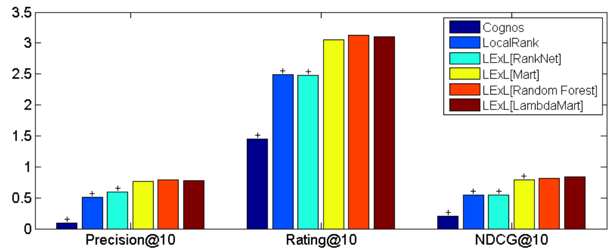

Fig. 1: Evaluating the proposed learning-based local expertise approach versus two alternatives. '+' marks statistical significant difference with LExL[LambdaMART] according to paired t-test at significance level 0.05.

LocalRank has a much better Precision@10 of around 0.5 compared to Cognos+, which indicates that 50 percent of the candidates it identifies have at least “some local expertise” for the query. The average Rating@10 is 2.49, which means the candidates are generally rated between “a little expertise” and “some expertise”. Since LocalRank explicitly builds on both topical and local signals (by exploiting the distance between a candidate's labelers and the query location), it performs much better than Cognos+. However, LocalRank is only a linear combination of these two factors, and so does not exploit either additional factors (like the random walk presented in this paper) nor take advantage of a learning approach for optimizing the weighting of these factors.

For the four LExL approaches, Ranknet performs comparably to LocalRank, but the remaining three all result in significantly better performance, with both Random Forest and LambaMART achieving comparably good results. These two methods have a Rating@10 of around 3.1, indicating that the local experts discovered have from “some local expertise” to “extensive local expertise”. The Precision@10 and NDCG@10 also support the conclusion that these learning-based methods result in high-quality local experts. Since LambdaMART is significantly less computationally expensive (∼1/6 of the computing time of Random Forest), we adopt it for the remainder of the paper.

Effectiveness Across Topics and Locations. We next turn to comparing the effectiveness of LExL with LamdaMART across topics.

We observe in Table 1 that NDCG@10 is consistently high for the four general topics, with an average value of 0.8074. Precision@10 and Rating@10 are also consistent for general topics except for the topic of “health” which has relatively low values. We attribute this poor showing due to data sparsity: through manual inspection we find that there are inherently only a limited number of candidates with high local expertise for the “health” topic in the training and testing datasets. However, since the learning framework is effective at identifying even those few local experts in “health”, we see a high NDCG@10.

We observe comparable results for the four narrower topics. The Precision@10 is lower than for the general topics (0.74 versus 0.82), but the NDCG@10 is

Table 1: Quality of local expert rankings across topics

Topics	P@10	R@10	N@10	Topics	P@10	R@10	N@10
food	0.8250	3.125	0.7004	chefs	0.8250	3.163	0.8554
sports	0.9375	3.225	0.8913	football	0.7220	2.925	0.8820
business	0.9250	3.131	0.8810	entrepreneurs	0.7333	3.040	0.7768
health	0.5750	3.059	0.7570	healthcare	0.7125	3.125	0.9423
General topic AVG	0.8156	3.135	0.8074	Subtopic AVG	0.7482	3.063	0.8641

higher (0.86 versus 0.81). Part of the higher NDCG results may be attributed to the decrease in the denominator of NDCG for these narrower topics (the Ideal DCG), so the ranking method need only identify some of a pool of moderate local experts rather than identify a few superstar local experts.

4 Conclusion

In this paper, we have proposed and evaluated a geo-spatial learning-to-rank framework for identifying local experts that leverages the fine-grained GPS coordinates of millions of Twitter user and carefully curated Twitter list data. We introduced four categories of features for learning model, including a group of location-sensitive graph random walk features that captures both the dynamics of expertise propagation and physical distances. Through extensive experimental investigation, we find the proposed learning framework produces significant improvement compared to previous methods.

References

1. Amatriain, X., Lathia, N., Pujol, J.M., et al.: The wisdom of the few: A collaborative filtering approach based on expert opinions from the web. In: SIGIR (2009)
2. Balog, K., Azzopardi, L., De Rijke, M.: Formal models for expert finding in enterprise corpora. In: SIGIR (2006)
3. Campbell, C.S., Maglio, P.P., et al.: Expertise identification using email communications. In: CIKM (2003)
4. Cheng, Z., Caverlee, J., Barthwal, H., Bachani, V.: Who is the barbecue king of texas?: A geo-spatial approach to finding local experts on twitter. In: SIGIR (2014)
5. Fleiss, J.L., et al.: Large sample standard errors of kappa and weighted kappa. Psychological Bulletin (1969)
6. Ghosh, S., Sharma, N., et al.: Cognos: crowdsourcing search for topic experts in microblogs. In: SIGIR (2012)
7. Google: Overview of local guides (2015), `https://goo.gl/NFS0Yz`
8. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. JACM (1999)
9. Lappas, T., Liu, K., Terzi, E.: A survey of algorithms and systems for expert location in social networks. In: Social Network Data Analytics. Springer (2011)
10. Pal, A., Counts, S.: Identifying topical authorities in microblogs. In: WSDM (2011)
11. Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: WSDM (2010)
12. Wu, Q., Burges, C.J., Svore, K.M., Gao, J.: Ranking, boosting, and model adaptation. Tecnical Report, MSR-TR-2008-109 (2008)