Finding Local Experts on Twitter

Zhiyuan Cheng, James Caverlee, Himanshu Barthwal, and Vandana Bachani
Department of Computer Science and Engineering
Texas A&M University
College Station, TX, USA
{zcheng, caverlee, barthwal, vbachani}@cse.tamu.edu

ABSTRACT

We address the problem of identifying local experts on Twitter. Specifically, we propose a local expertise framework that integrates both users' topical expertise and their local authority by leveraging over 15 million geo-tagged Twitter lists. We evaluate the proposed approach across 16 queries coupled with over 2,000 individual judgments from Amazon Mechanical Turk. Our initial experiments find significant improvement over a naive local expert finding approach, suggesting the promise of exploiting geo-tagged Twitter lists for local expert finding.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data Mining

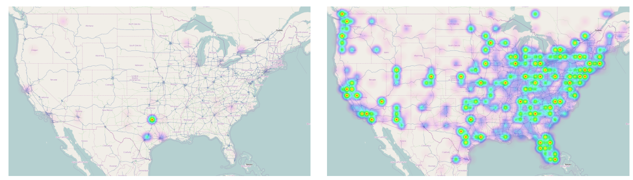
Keywords

Twitter, expert finding, local expert, social tagging

1. INTRODUCTION

We tackle the problem of finding *local experts* in social media systems like Twitter. Local experts bring specialized knowledge about a particular location and can provide insights that are typically unavailable to more general topic experts. A recent Yahoo! Research survey found that 43% of participants would like to directly contact local experts for advice and recommendations online, while 39% would not mind being contacted by others [1]. And yet finding local experts is challenging. Traditional expert finding has focused on either small-scale, difficult-to-scale curation of experts (e.g., a magazine's list of the "Top 100 Lawyers in Houston") or on automated methods that can mine large-scale information sharing platforms. These approaches, however, have typically focused on finding general topic experts, rather than local experts.

We present here our initial framework for local expert finding – LocalRank – that integrates both a person's topical expertise and local authority. Our approach is motivated by the widespread adoption of GPS-enabled tagging of social media content via smartphones and social media services (e.g., Facebook, Twitter, Foursquare) that provide a geo-social overlay of the physical environment. This massive scale geo-social resource provides unprecedented opportunities to study the connection between people's expertise and locations. Concretely, LocalRank views a local expert as *someone who is well recognized by the local community*, where we estimate this local recognition via a novel spatial proximity expertise approach that leverages over 15 million geo-tagged Twitter lists. To illustrate, Figure 1(a) shows a heatmap of the locations of Twitter users who have labeled Daniel Vaughn (@BBQsnob) on Twitter. As one of the foremost barbecue experts in Texas, Vaughn's expertise is recognized regionally in Texas, and more specifically by local barbecue centers in Austin and Dallas. In contrast, late-night host Jimmy Fallon's heatmap suggests he is recognized nationally, but without a strong local community. Intuitively, Daniel Vaughn is recognized as a *local expert* in Austin in the area of Barbecue; Jimmy Fallon is certainly an expert (of comedy and entertainment), but his expertise is diffused nationally.



(a) @BBQsnob (b) @JimmyFallon

Figure 1: Heatmap of the location of Twitter users who have listed @BBQsnob or @JimmyFallon

2. PROBLEM STATEMENT AND SOLUTION

We are interested to find local experts with particular expertise in a specific location. We assume there is a pool of expert candidates $V = \{v_1, v_2, ..., v_n\}$, that each candidate v_i has an associated location $l(v_i)$ and a set of areas of expertise described by a feature vector $\vec{v_i}$. Each element in the vector is associated with an expertise topic word t_w (e.g., "technology"), and the element value indicates to what extent the candidate is an expert in the corresponding topic. We define the **Local Expert Finding** problem as:

DEFINITION 1. ***(Local Expert Finding)*** *Given a query q that includes a query topic $t(q)$, and a query location $l(q)$, find the set of k candidates with the highest local expertise in query topic $t(q)$ and location $l(q)$.*

Topical vs. Local Authority: Identifying a local expert requires that we can accurately estimate not only the candidate's expertise on a topic of interest (e.g., how much does this candidate know about barbecue), but also that we can identify the candidate's local authority (e.g., how well does the local community recognize this candidate's expertise). Hence, we propose to decompose the local expertise for a candidate v_i into two related dimensions: (i) **Topical Authority**: which captures the candidate's expertise on the topic area $t(q)$; and (ii) **Local Authority**: which captures the candidate's local authority in query location $l(q)$. The local experts we are trying to identify should have both great topical authority and local authority: e.g., Daniel Vaughn (@bbqsnob) is an example of an expert with high topical authority (on barbecue), as well as high local authority (in Texas).

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577354.

Local Expert Finding with LocalRank: We formally define candidate v_i's **LocalRank** (LR) $s(v_i, q)$ in query q as:

$$s(v_i, q) = s_l(l(v_i), l(q)) * s_t(\vec{v_i}, t(q))$$

where $s_l(l(v_i), l(q))$ denotes the *Local Authority* of v_i in query location $l(q)$, and $s_t(\vec{v_i}, t(q))$ denotes the *Topical Authority* of v_i in query topic $t(q)$ that is estimated using the candidate's expertise vector $\vec{v_i}$.

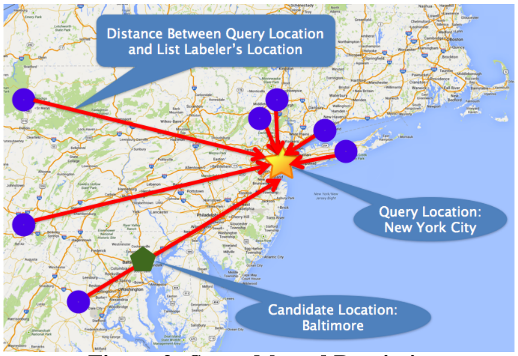


Figure 2: Spread-based Proximity

Estimating Local Authority: The first approach to estimate candidate v_i's local authority for query q is to use the distance between candidate v_i's location $l(v_i)$ and the query location $l(q)$. We define this *Candidate Proximity* ($s_{l_{CP}}$) as:

$$s_{l_{CP}}(l(v_i), l(q)) = \left(\frac{d_{min}}{d(l(v_i), l(q)) + d_{min}} \right)^{\alpha}$$

where $d(l(v_i), l(q))$ denotes the distance between $l(v_i)$, and $l(q)$. This first local authority approach captures the intuition that closer candidates are more locally authoritative.

The second approach – *Spread-based Proximity* ($s_{l_{SP}}$) – measures the "spread" of the locations of a candidate's core audience (by mining Twitter lists to determine which users have labeled a candidate expert) compared to the query location:

$$s_{l_{SP}}(L(V_l(v_i)), l(q)) = \frac{\sum\limits_{v_{l_j} \in V_l(v_i)} s_{l_{CP}}(l(v_{l_j}), l(q))}{|V_l(v_i)|}$$

where v_{l_j} denotes one of the core audience $V_l(v_i)$ of candidate v_i. Basically, the "spread" it measures considers how far candidate v_i's core audience are from the query location $l(q)$ on average. If the core audience of a candidate is close to a query location on average, the candidate gets a high score of $s_{l_{SP}}$. For example, in Figure 2, the green pentagon and the gold star represent the expert candidate's location and the query location, respectively. And the candidate's local authority in New York City is estimated by the distances between the locations of the candidate's labelers (blue dots) and the query location.

Estimating Topical Authority: We adapt the user-centric model (addressed as Directly Labeled Expertise (DLE) method in this paper) that Balog et al. proposed in [2] to estimate the *Topical Authority Score* $s_t(\vec{v_i}, t(q))$ of v_i with respect to the query topic $t(q)$.

3. EVALUATION

We evaluate the proposed local expert finding framework. We seek answers to the following questions:

- What impact does the choice of local authority have on the quality of local expert finding in LocalRank?
- How well does LocalRank perform compared to alternative local expert finding approaches?

Queries: In total, we evaluate four general query topics coupled with four locations, totaling 16 topic-location queries. Specifically, we look for local experts in the areas of "technology", "entertainment", "food", and "travel" in New York City, Houston, San Francisco, and Chicago.

Alternative Approach for Finding Local Experts. We consider one alternative baseline that combines simple versions of topical and local authority:

- *Most Popular in Town by Listed Count on Topic* (*MP (on-topic)*): Rank candidates from the query location by the number of on-topic lists the candidate appears on.

We compare the baseline with the proposed LocalRank approach. For LocalRank, we investigate the two approaches for estimating local authority – by Candidate Proximity (CP), and Spread-based Proximity (SP) – and the Directly Labeled Expertise (DLE) approaches for estimating topical authority. When applying both the *Candidate Proximity*, and *Spread-based Proximity*, we preset the d_{min} to be 100 (miles), and $alpha$ to be 2.0. We calculate the local expertise score using the normalized topical authority score and the normalized local authority score.

Gathering Ground Truth: We gather ground truth by employing human raters on Amazon Mechanical Turk. For each local expert candidate, we have 5 human raters label to what extent the candidate has local expertise in query topic and query location using a 4 Likert scale. In total, we collect over 2,000 individual judgments.

Metrics: To evaluate each local expert finding approach, we measure the *Precision@k*, and *NDCG@k* across all queries in our testbed. For the following experiments, we consider all the 0 and -1 ratings as 0s.

Evaluating Local Expert Finding Approaches: In Table 1, we present the results for the alternative approach, as well as two LocalRank approaches Candidate Proximity + Directly Labeled Expertise (CP + DLE), and Spread Proximity + Directly Labeled Expertise (SP + DLE). In terms of the comparison between the two local authority metrics, we observe that SP significantly improves the performance of local expert finding in comparison with CP. Using CP, the LocalRank approach only identifies true local expert 55% of the time on average among the top 10 candidates. Similarly, we see comparatively low values of $NDCG$@10 as 0.685. In contrast, SP reaches $Precision@k$ of almost 85%, and $NDCG$@10 of 0.90. This indicates the core audience for an expert candidate is crucial to estimating a candidate's local authority. In addition, we observe that the baseline approach (MP on-topic) performs quite effectively, even better than the LocalRank approach with CP. However, we also observe that the LocalRank approach coupled with Spread Proximity performs much better comparing to the baseline in terms of both $Precision$@10 and $NCDG$@10.

Table 1: Comparing LocalRanks with Baseline Approach

Approach	$Precision$@10	$NDCG$@10
MP (on-topic)	0.628	0.750
LR: CP + DLE	0.553	0.685
LR: SP + DLE	**0.842**	**0.896**

4. CONCLUSION

We have proposed and evaluated the LocalRank framework for finding local experts, by integrating both a candidate's local authority and topical authority. We see a significant improvement in performance (35% improvement in $Precision$@10 and around 18% in $NDCG$@10) over the alternative approach. These results demonstrate the viability of mining fine-grained geo-social signals for expertise finding, and highlight the potential of future geo-social systems that facilitate information flow between local experts and the local community.

5. REFERENCES

[1] J. Antin, M. de Sa, and E. F. Churchill. Local experts and online review sites. In *CSCW 2012*.

[2] K. Balog, L. Azzopardi, and M. De Rijke. Formal models for expert finding in enterprise corpora. In *SIGIR 2006*.