Probabilistic Local Expert Retrieval

Wen Li[1(✉)], Carsten Eickhoff[2], and Arjen P. de Vries[3]

[1] University College London, WC1E 6BT, London, UK
wen.li@ucl.ac.uk
[2] Department of Computer Science, ETH Zurich, Zurich, Switzerland
c.eickhoff@acm.org
[3] Faculty of Science, Radboud University, 6525 EC Nijmegen, The Netherlands
arjen@acm.org

Abstract. This paper proposes a range of probabilistic models of local expertise based on geo-tagged social network streams. We assume that frequent visits result in greater familiarity with the location in question. To capture this notion, we rely on spatio-temporal information from users' online check-in profiles. We evaluate the proposed models on a large-scale sample of geo-tagged and manually annotated Twitter streams. Our experiments show that the proposed methods outperform both intuitive baselines as well as established models such as the iterative inference scheme.

Keywords: Domain expertise · Geo-tagging · Twitter

1 Introduction

When visiting unfamiliar cities for the first time, visitors are confronted with a number of challenges related to finding the right spots to go, sights to see or even the most appropriate range of cuisine to sample. While residents quite naturally familiarize themselves with their surroundings, strangers often face difficulties in efficiently selecting the best location to suit their preferences. We refer to such location-specific, and frequently sought-after [11], knowledge as local expertise.

Local expertise can be acquired with the help of online resources such as review sites (e.g., yelp.com or tripadvisor.com) that rely on both paid professionals as well as user recommendations. General-purpose Web search engines, especially in the form of entertainment or food verticals, provide valuable information. However, these services merely return basic information and results are not specifically tailored to the individual. Ideally, a more effective way of solving this task would be to ask someone who is local and/or has the knowledge about the area in question. Seeking out this kind of people is an example of expert retrieval, and, more specifically, of local expert retrieval.

Location-based social networks allow users to post messages and document their whereabouts. When a user checks in at a given location, the action of *check-in* is not merely a user-place tuple. The physical attendance at the location also suggests that the user, at least to some extent, gets familiar with the

location and its environment. The more frequently such evidence is observed, the more accurate our insights into the user's interests and expertise become. This paper introduces two novel contributions over the state of the art in local expert retrieval. **(1)** We propose a range of probabilistic models for estimating users' local expertise on the basis of geo-tagged social network streams. **(2)** In a large-scale evaluation effort, we demonstrate the merit of the methods on real-world data sampled from the popular microblogging service Twitter.

The remainder of this paper is structured as follows: Sect. 2 gives an overview of local expert retrieval methods as well as social question answering platforms. Section 3 derives a range of probabilistic models for local expert retrieval that are being evaluated on a concrete retrieval task in Sect. 4. Finally, Sect. 5 concludes with a brief discussion of our main findings as well as an outlook on future work.

2   Related Work

The task of expertise retrieval has first been addressed in the domain of enterprises managing and optimizing human resources (detailed in [1,17]). Early expertise retrieval systems required experts to manually fill out questionnaires about their areas of expertise to create the so-called expert profiles. Later, automated systems were employed for building and updating such profiles and probabilistic models were introduced for estimating a candidate's expertise based on the documents they authored, *e.g.*, [1,4,6,7,15]. These works inspire us to use probabilistic models for estimating local expertise. Since location information is not always presented in textual format, we approach the problem by building models based on candidates' check-in profiles. Li et al. [11] proposed the problem of local expert retrieval (using the term "geo-expertise") and investigated the main intuitions that could naturally support an automatic approach which considers user check-in profiles as evidence of having knowledge regarding locations they had visited. A preliminary empirical evaluation demonstrated its feasibility using three heuristic methods for automating local expert retrieval, however, without giving a formal derivation that would underline the soundness of these methods. In this paper, we follow up upon this line of research and investigate the probabilistic reasoning behind the methods proposed in the previous work. Cheng et al. [5] also proposed finding local experts as a retrieval problem, for which they combined models of local and topical authority to rank candidates based on data collected from Twitter. The authors rely on textual queries accompanied by a location to specify spatial constraints. In our setting, queries are phrased in terms of locations. This can for example be either a specific restaurant or a type of restaurants to which users are interested in paying a visit. The main difference between this study and the aforementioned one is that we focus on evidence of location knowledge in geo-spatial movement profiles while Cheng et al. introduce location constraints in text-based expertise retrieval.

Another domain closely related to expertise retrieval is found in the context of community question answering (CQA) platforms, such as Quora (http://www.quora.com), Stackoverflow (http://stackoverflow.com), or, Yahoo Answers (https://answers.yahoo.com). These services rely on routing questions to the most suitable potential answerers. These platforms also provide researchers with an opportunity to access large-scale topical expertise profiles. In particular, they provide data that can be used directly for evaluation, *i.e.*, whether the top-ranked candidates give satisfactory answers to the questions they have been retrieved for. Based on this kind of data, Liu et al. [13] proposed to use language models to profile candidates, Zhang et al. [18] used heuristic features from asker-answerer networks to rank candidates, and Horowitz et al. [9] relied on probabilistic models similar to Balog et al. [1] in their social search engine Aardvark. Aardvark is a CQA-platform-based instant messaging system including a location-sensitive classifier to decide whether a given question requires local expertise. Although the providers do not detail the algorithm of their classifier [9], their paper gives examples rendering the problem as a place entity recognition task. Studies based on CQA data focus on textual and social network features but do not currently explore candidates' movement history. We consider this type of information a crucial factor in local expert retrieval tasks, *i.e.*, modelling candidate's knowledge about locations.

There are several domain-specific studies of user expertise modelling. Bar-Haim et al. [3] aimed to identify stock experts on Twitter by testing candidates' predictions of stock prices (*e.g.*, buying or selling a stock at the right time) in their tweets. Whiting et al. [16] suggested using changes of Wikipedia pages as evidence for developing events. The authors retrieved tweets containing relevant terms and considered the authors of these tweets to be influencers for this topic. In a closely related effort, Bao et al. [2] proposed a method to finding local experts for location categories. They applied a hyperlink inference topic search algorithm on the user-location matrix. We have included their method as a performance baseline in this study.

3   Models of Local Expertise

In this paper, we define local expertise as knowledge regarding given (categories of) places of interest (POI). The POI information and POI categories are provided respectively by Twitter and Foursquare via their APIs. Consequently, a topic in local expert retrieval can be either a specific POI or a category of POIs within a geographical scope, *e.g.*, the [Blue Ribbon Fried Chicken] in New York or [Chinese Restaurants] in Los Angeles. The former is referred to as *POI topics*, which describe knowledge regarding a single location, such as opening times, or admittance fees. The latter is referred to as *category topics*, which describe knowledge regarding all locations in a specific category, such as different themes or decoration of locations in the category. High-ranking candidates should be able to answer questions about the locations or the category of locations in the topic. For simplicity, both types of topics are also referred to as locations in the rest

of paper. In [11], check-ins are considered to be links between candidates' visits to a location and knowledge they may have about the location. They proposed three methods characterizing check-in profiles in three different aspects, *i.e.*, visiting *frequency*, *diversity* and *recency*. While the previous work captured these notions in a heuristic manner, the following sections pursue a formal derivation of these empirically proven notions. That is we develop a probabilistic model for each of these heuristic methods.

3.1   Within-Topic Activity (WTA)

The first approach we propose considers only the candidates' check-in frequency. This method focuses on knowledge about a single location or a single type of locations. We take a co-occurrence modelling approach, inspired by expert finding via text documents [6]. To be specific, we rank a candidate u by their probability of having local expertise in a given topic q, *i.e.*, $P(u|q)$. We estimate the conditional probability by aggregating over the user's check-ins at all locations (l), that is

$$P(u|q) = \sum_l P(u|l, q)P(l|q).$$

Assuming conditional independence of the candidate u and the query topic q given the location l, *i.e.*, $P(u|l, q) = P(u|l)$, we obtain

$$P(u|q) = \sum_l P(u|l)P(l|q).$$

As for $P(u|l)$, we apply Bayes' Rule which gives

$$P(u|l) \propto P(l|u)P(u)$$

where we assume a uniform prior for $P(l)$.
Putting these together, we obtain

$$P(u|q) \xlongequal{\text{rank}} \sum_l P(l|q)P(l|u)P(u). \qquad (1)$$

$P(u)$ is the query-independent confidence in the user model estimated by the number of check-ins observed for this user, *i.e.*,

$$P(u) = \frac{N_u}{N}$$

where N_u and N, respectively, are the number of check-ins posted by candidate u and the overall number of check-ins. Intuitively, the more data is available for a given user, the more we trust the model built from his/her check-in profiles to be accurate. The conditional probability $P(l|q)$ captures the user's query intent, *i.e.*, the possible range of locations users may be interested in. In our setting of local expertise, the query is actually a location or a type of locations. The final conditional probability can be estimated by

$$P(l|q) = \begin{cases} \frac{1}{|L_q|} & \text{if } l \in L_q, \\ 0 & \text{otherwise,} \end{cases}$$

where L_q is the set of locations matching the query. To estimate $P(l|u)$, we use

$$P(l|u) = \frac{N_{l,u}}{N_u},$$

where $N_{l,u}$ is the number of check-ins candidate u made at location l. The scoring function can be derived from simplifying Eq. 1, that is

$$S_n(u,q) = \frac{1}{|L_q| \cdot N} \sum_{l \in L_q} N_{l,u} \stackrel{\text{rank}}{=\!=\!=} \sum_{l \in L_q} N_{l,u}.$$

Intuitively, the more check-ins a candidate has at the queried location(s) in L_q, the more likely they are to be interested in those locations and knows about them.

3.2   Within-Topic Diversity (WTD)

Our second method uses the language model referred to as *Model 1* in [1], that is

$$P(\theta_u|q) \propto P(q|\theta_u)P(\theta_u).$$

where θ_u is a language model of a candidate based on his/her check-in profile. To estimate $P(q|\theta_u)$, we assume independence between individual locations. That is

$$P(q|\theta_u) = \prod_{l \in L_q} P(l|\theta_u) = \prod_{l \in L_q} \frac{N_{l,u}}{N_u}. \tag{2}$$

For the prior $P(\theta_u)$, we use

$$P(u) = \frac{N_u^{|L_q|}}{\sum_{u' \in U} N_{u'}^{|L_q|}},$$

so it will simplify the scoring function, that is

$$P(q|u) = \frac{N_u^{|L_q|}}{\sum_{u' \in U} N_{u'}^{|L_q|}} \prod_{l \in L_q} \frac{N_{l,u}}{N_u}.$$

By applying the logarithm (to avoid underflow in computation), we obtain

$$S_d(u,q) = \log \frac{1}{\sum_{u' \in U} N_{u'}^{|L_q|}} + \sum_{l \in L_q} \log N_{l,u} \stackrel{\text{rank}}{=\!=\!=} \sum_{l \in L_q} \log N_{l,u}.$$

The following smoothing function is used to differentiate the profiles containing visits to different numbers of locations but each location has been visited only once.

$$S_d(u,q) = \sum_{l \in L_q} \log(N_{l,u} + 1).$$

The above scoring function indicates that check-ins at multiple distinct locations (within the queried location set) should increase the score more than repeatedly checking in at the same location. This means that a candidate will gain a high local expertise rating if he/she makes check-ins at a variety of relevant locations. This fits the intuition that candidates with experience at a variety of locations may know more about the essence of the topic rather than mere specifics of a single place within that category. For example, if we seek advice about Italian restaurants, individuals who have been to many Italian restaurants in town will be more suitable candidates than those who have been to the same restaurant a lot.

The prior $P(u)$ is selected for the scoring function so that the candidate-dependent denominator in the conditional probability $P(l|u)$ will be cancelled when combined with the prior. This accounts for the fact that language models represent users' topical focus rather than their knowledge, $i.e.$, they are biased towards shorter profiles, when two profiles have the same amount of relevant check-ins. Since check-ins are positive evidence of candidates knowing about a location, additionally knowing about other types of locations should not negatively affect the local expertise score. For example, if a candidate has visited two place categories A and B each for n times while another candidate only has been to A for n times, it is not reasonable to assume that the latter candidate should have more knowledge about A than the former candidate, even if the latter has focused on A more.

3.3   Within-Topic Recency (WTR)

Experts are humans and as such they rely on their memories to support their expertise. Therefore, we should take into account the fact that (1) people forget the knowledge they once gained and have not refreshed for a while, and (2) the world is changing as time goes by, $e.g.$, restaurants may have new chefs, and old buildings may have been replaced. The more time has passed since the creation of the memory, the more likely it will be forgotten or outdated. To incorporate such effects, we explicitly model the candidates' memory by $P(c|u)$, which indicates the probability that candidate u can recall his/her visit represented by the check-in c. As suggested in the domain of psychology, human memory can be assumed to decay exponentially [14]. Consequentially, we use an exponential decay function to represent the retention of individual check-ins, by which we obtain:

$$P_t(c|u) = \frac{e^{-\lambda(t-t_c)}}{\sum_{c \in C_u} e^{-\lambda(t-t_c)}},$$

where t is the time of query and t_c is the time when the user posted the check-in. Similarly, we define a prior for each candidate as follows

$$P_t(u) = \frac{\sum_{c \in C_u} e^{-\lambda(t-t_c)}}{\sum_{c \in C} e^{-\lambda(t-t_c)}}$$

The decay of the weight on check-ins models our belief on how up-to-date the information is, while the prior reflects the average recency of knowledge borne by the whole community on the social network. Then, for estimating the candidate's expertise, we weight each check-in according to its recency, *i.e.*, we marginalize the user's old check-ins.

$$P_t(l|u) = \sum_{c \in C_u} P(l|c)P_t(c|u) = \sum_{c \in C_u} \frac{\mathbf{1}(l_c = l)e^{-\lambda(t-t_c)}}{e^{-\lambda(t-t_c)}}, \tag{3}$$

where $\mathbf{1}(\cdot)$ is an indicator function, which equals 1 *if and only if* the condition in the parentheses evaluates to true. Given these two estimations, we have

$$\mathcal{S}_r(u,q) = P_t(u|q) = \sum_{l \in L} P_t(u|l)P(l|q) \stackrel{\text{rank}}{=\!=\!=} \sum_{l \in L} P_t(l|u)P_t(u)P(l|q)$$

$$\stackrel{\text{rank}}{=\!=\!=} \sum_{l \in L_q} \sum_{c \in C_u, l_c = l} e^{-\lambda(t-t_c)}.$$

As can be seen, $\mathcal{S}_r$ down-weights older check-ins' contribution to a candidate's expertise due to the fact that they may be vaguely memorized and become unreliable. The decay parameter λ is fixed to $\frac{1}{150}$ at a granularity of days and we leave the fine-tuning of this parameter for future work.

3.4   Combining Recency and Diversity (WTRD)

Diversity and recency of check-ins can both be important factors in estimating a candidate's local expertise. Thus, we propose a combination of the two features introduced in Sects. 3.2 and 3.3. In Eq. 2, the conditional probability can be transformed into

$$P(q|u) = \sum_{c \in C_u} P(l|c,u)P(c|u) = \sum_{c \in C_u} P(l|c)P(c|u) = \sum_{c \in C_u} \mathbf{1}(l_c = l)P(c|u)$$

in which we assume that candidate and location are conditionally independent given a check-in (*i.e.*, from the first equation to the second equation). Then, we estimate the conditional probability $P(c|u)$ with Eq. 3.

$$P(c|u) = \frac{e^{-\lambda(t-t_c)}}{\sum_{c \in C_u} e^{-\lambda(t-t_c)}}$$

Thus, we have

$$P(q|u) = \prod_{l \in L_q} \sum_{c \in C_u} \frac{\mathbf{1}(l_c = l)e^{-\lambda(t-t_c)}}{\sum_{c \in C_u} e^{-\lambda(t-t_c)}}$$

Similar to the prior probability used in the diversity method,

$$P(u) = \frac{(\sum_{c \in C_u} e^{(-\lambda(t-t_c))})^{|L_q|}}{(\sum_{c \in C} e^{-\lambda(t-t_u)})^{|L_q|}}$$

By replacing the counterparts with these into Eq. 2 and applying the logarithm on both sides of the equation, we obtain

$$\mathcal{S}_d(u,q) = \log \frac{(\sum_{c \in C_u} e^{-\lambda(t-t_c)})^{|L_q|}}{(\sum_{c \in C} e^{-\lambda(t-t_u)})^{|L_q|}} \prod_{l \in L_q} \frac{\sum_{c \in C_u} \mathbf{1}(l_u = l)e^{-\lambda(t-t_c)}}{\sum_{c \in C_u} e^{-\lambda(t-t_c)}}$$

$$= \log \frac{1}{(\sum_{c \in C} e^{-\lambda(t-t_c)})^{|L_q|}} + \sum_{l \in L_q} \log \sum_{c \in C_u} \mathbf{1}(l = l_c)e^{-\lambda(t-t_c)}$$

$$\overset{\text{rank}}{=\!=\!=} \sum_{l \in L_q} \log \sum_{c \in C_u} \mathbf{1}(l = l_c)e^{-\lambda(t-t_c)}$$

The decay parameter λ is set to the same value as that in the WTR method.

3.5   Iterative Inference Model (HITS)

Bao et al. [2] propose a model for estimating local expertise based on the Hyperlink-Induced Topic Search (HITS) algorithm, an approach originally designed for link analysis of Web pages [10]. The model defines two properties for users and locations respectively, *i.e.*, hub scores for users and authority scores for locations. The hub score indicates how well a user can serve as an information source about a place and the authority score presents how popular a place is. We implement a normalized version of the algorithm and focus on hub scores for users (candidates) which are used as estimates of local expertise and are calculated as $\mathcal{S}_h(u,l) = \boldsymbol{h}_{u,l}^{(n+1)}$, where

$$\boldsymbol{h}^{(n+1)} = \frac{\mathbf{M}^T \mathbf{M} \cdot \boldsymbol{h}^{(n)}}{\| \mathbf{M}^T \mathbf{M} \cdot \boldsymbol{h}^{(n)} \|}.$$

3.6   Candidate Profiling

As a mainstream location-based social network, Foursquare attempts to increase user engagement by encouraging users, through gamification elements, to check in at a location far more times than they actually need to (*i.e.*, the user never left the location but checked in again in order to collect rewards). To mitigate the effect of this twisted relation between check-ins and actual visits, we define a different type of candidate profile, *i.e.*, the Active-Day Profile (referred to as +A while +C is used to refer to original check-in profiles). It is a subset of a user's check-ins which is defined as: $\{c | c \in C_u, \nexists c' \in C_u : l_c = l_{c'}, t_c < t_{c'} < \lceil t_c \rceil_D\}$ where $\lceil \cdot \rceil_D$ is a ceiling function towards midnights. Informally, the Active-Day profile contains only the last check-in at each place within each day, reducing the influence of multiple check-ins at the same place to at most one per day.

4   Evaluation

To evaluate the various discussed methods and profile types, we implement a configurable local expert retrieval system. The system accepts a topic which is

composed of a scope of city and a (type of) location(s) and returns a list of related candidate experts. The dataset used in this study is an extended version of the collection of POI-tagged tweets proposed by Li et al. [12]. It comprises 1.3M check-ins from 8 K distinct users from New York, Chicago, Los Angeles, and San Francisco. The data collection process was set up such that each user's full check-in profile would be included. To filter out accounts that are solely used for branding and advertisement purposes (*e.g.*, by companies) we remove all users having a "*speed*" between consecutive check-ins higher than 700 kph (which corresponds approximately to the speed of a passenger aircraft). Similarly, users showing less than 5 geo-tagged tweets were excluded as well. As a consequence of this thresholding approach, Fig. 1, shows that the check-in distribution over users does not follow a complete power law which was observed in the previous dataset.
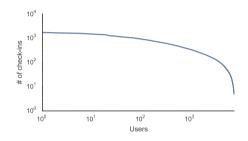

Fig. 1. Distribution of check-in frequencies across users.

To prepare a set of topics for evaluation, we rely on stratified sampling to identify a seed set of location categories and POIs, based on their popularity among the users. Two strata are composed respectively for popular POIs (top 10 %) and less popular POIs (remaining 90 %). Locations are selected randomly using a uniform distribution per stratum, and the number of samples is in accordance with the size of stratum. As for category topics, we include all 9 top categories from Foursquare's category hierarchy (*e.g.*, Food) and apply 10 %:90 % stratified sampling to the categories at the lower levels (*e.g.*, Chinese Food, Mexican Food). This results in a seed set totalling 275 topics for all 4 cities. To reduce the work load for human annotators, we eliminate entities which have less than 5 visitors and whose names are obscure (Building, Home – Private, Field, Professional & Other, Residence). As a result, we obtain 95 topics from 4 cities in total, among which there are 71 category topics and 24 POI topics. The top 5 returned candidates from each method are then pooled and annotated. This process resulted in a pool of 1588 distinct topic-candidate pairs across all methods. With the annotated topic-candidate pairs, we measure the performance of the proposed methods in terms of P@1, P@5 and MAP. A random selection approach and our implementation of [2]'s method are included as baselines.

4.1 Annotation

Topic-candidate pairs are annotated by human judges that assess each candidate's level of expertise about the given topics. To facilitate this process, an interactive annotation interface for displaying the candidate's historical check-ins has been designed and can be accessed at https://geo-expertise.appspot.com. For each topic-candidate pair, an annotator is asked to assign a value from 1 to 5 indicating their assessment of the candidate's knowledge about the topic, where "5" means the candidate knows the topic very well and "1" indicates the candidate knows barely anything about the topic. For greater reliability, we recruit assessors from different channels. **(1)** The first run has been carried out on the crowdsourcing platform CrowdFlower where each participant is paid 0.5 USD per task (each containing 10 topic-candidate pairs to annotate). Additionally, we invited students and staff from Delft University of Technology to contribute their assessments. Via Cohen's Kappa [8], we found that annotators are inclined to agree ($\kappa > 0.4$) whenever they have strong opinions on whether a candidate is a local expert on a give topic.

4.2 Quantitative Evaluation

We carry out separate evaluations on two runs of annotation, *i.e.*, one from the recruited annotators on CrowdFlower and one from the university staff and students. Annotations are converted into binary labels, in which topic-candidate pairs assigned with scores 4 or 5 are considered relevant (local experts) and those with scores 3 or below are considered as irrelevant (non-experts). Based on the binary relevance annotation, `trec_eval` (http://trec.nist.gov/trec_eval/) is used for measuring the performance of the proposed methods. We test the statistical significance of differences between the performance scores using a Wilcoxon Signed Rank Test ($\alpha < 0.05$).

According to the crowdsourced annotation shown in Table 1, the WTD method with Active-Day profiles performs the best under P@1 and P@5, while WTA with check-in profiles performs the best under MAP. All proposed methods with both types of profiles significantly outperform the random baseline. The HITS baseline performs significantly better than the random selection approach but is outperformed by the proposed movement profile methods even though this difference was not found to be significant.

The two types of profiles (+A and +C) were designed for comparing the potential influence of check-in gamification by Foursquare that might encourage users to check in as often as possible. In our evaluation, however, we do not observe a clear preference for either of the two profile types. This may suggest that the two types of profiles do not diverge much and check-in gamification does not have an observable influence on assessing candidates' local expertise.

Turning towards university annotations, we note that the annotators' relevance assessment is also in favour of configurations with WTD for P@1 and P@5 and WTA for MAP, although no significant differences between these two methods are observed. The proposed methods and the HITS method are all

significantly better than the random baseline on this set of annotations. Different from the crowdsourced annotation, here, we observe a significant preference of all the proposed methods (WTD, WTA, or WTRD) over HITS. At the same time, we have not observed any significant differences between the three methods WTD, WTA and WTRD configured with either profile types.

Table 1. Local expert retrieval results.

Method	Profile	Crowdsourced annotation			University annotation		
		MAP	P@1	P@5	MAP	P@1	P@5
WTA	+A	0.2750	0.4211	0.3979	**0.3218**	0.5579	0.4463
	+C	**0.2771**	0.4316	0.3895	0.3147	0.5579	0.4337
WTD	+A	0.2340	**0.4789**	**0.4197**	0.2878	0.5775	0.4817
	+C	0.2280	0.4507	0.4169	0.2908	**0.5915**	**0.4845**
WTR	+A	0.2442	0.3789	0.3747	0.2814	0.5263	0.4042
	+C	0.2508	0.4211	0.3768	0.2824	0.5368	0.4063
WTRD	+A	0.2434	0.4316	0.3684	0.2862	0.5368	0.4042
	+C	0.2491	0.4421	0.3726	0.2919	0.5368	0.4168
HITS	+A	0.2327	0.4507	0.4113	0.2041	0.5211	0.3859
	+C	0.2363	0.4366	0.4028	0.2045	0.5493	0.3831
Rand	−	0.1343	0.2316	0.2063	0.1041	0.1579	0.1600

5   Conclusion

In this paper, we presented a range of probabilistic-model-based approaches to the task of local expert retrieval. Based on the existing theoretical work in expertise retrieval, we designed three models to capture the candidate's check-in profiles. We further designed a method for distilling users' check-in profile to test whether the gamification of online location-based social networks would affect the accuracy of geo-expertise estimation. To evaluate the proposed methods, we collected a large volume of check-ins via Twitter's and Foursquare's public APIs, for which we finally collected judgements from both online recruited annotators and university annotators. Our evaluation shows that the proposed methods do capture local expertise better than both random as well as refined baselines. During our experiments, we did not observe a significant difference between Active-Day profiles and the raw check-in profiles in the evaluation.

In the future, we propose to carry out this evaluation task *in-vivo* by building a dedicated local expert retrieval system. Such a system can access the Twitter/Foursquare APIs for users who authorize the application to analyse their check-in profiles as well as their friends' geo-tagged media streams to find the friend that is assumed to know most about the user's desired location. In consequence, it can be assumed to produce much more reliable expertise annotations

since it would allow us to observe which recommendations are being followed-up on in practice, without the need of external assessment by judges. Additionally, it would be interesting to investigate the social ties between potential answerers and local expertise seekers, to ensure engagement of both parties and allow for greater personalization of answers.

References

1. Balog, K., Fang, Y., de Rijke, M., Serdyukov, P., Si, L.: Expertise retrieval. Found. Trends Inf. Retrieval **6**(2–3), 127–256 (2012)
2. Bao, J., Zheng, Y., Mokbel, M.F.: Location-based and preference-aware recommendation using sparse geo-social networking data. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems - SIGSPATIAL 2012, pp. 199–208 (2012)
3. Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., Goldstein, G.: Identifying and following expert investors in stock microblogs. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing - EMNLP 2011, pp. 1310–1319 (2011)
4. Campbell, C.S., Maglio, P.P., Cozzi, A., Dom, B.: Expertise identification using email communications. In: Proceedings of the 12th International Conference on Information and Knowledge Management - CIKM 2003, pp. 528–531 (2003)
5. Cheng, Z., Caverlee, J., Barthwal, H., Bachani, V.: Who is the barbecue king of Texas? In: Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR 2014, pp. 335–344 (2014)
6. Fang, H., Zhai, C.X.: Probabilistic models for expert finding. In: Amati, G., Carpineto, C., Romano, G. (eds.) ECiR 2007. LNCS, vol. 4425, pp. 418–430. Springer, Heidelberg (2007)
7. Fang, Y., Si, L., Mathur, A.P.: Discriminative models of integrating document evidence and document-candidate associations for expert search. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR 2010, p. 683 (2010)
8. Fleiss, J.L.: Measuring nominal scale agreement among many raters. Psychol. Bull. **76**(5), 378–382 (1971)
9. Horowitz, D., Kamvar, S.D.: The anatomy of a large-scale social search engine. In: Proceedings of the 19th International Conference on World Wide Web - WWW 2010, pp. 431–440 (2010)
10. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM **46**(5), 604–632 (1999)
11. Li, W., Eickhoff, C., de Vries, A.P.: Geo-spatial domain expertise in microblogs. In: de Rijke, M., Kenter, T., de Vries, A.P., Zhai, C.X., de Jong, F., Radinsky, K., Hofmann, K. (eds.) ECIR 2014. LNCS, vol. 8416, pp. 487–492. Springer, Heidelberg (2014)
12. Li, W., Serdyukov, P., de Vries, A.P., Eickhoff, C., Larson, M.: The where in the tweet. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management - CIKM 2011, pp. 2473–2476 (2011)
13. Liu, X., Croft, W.B., Koll, M.: Finding experts in community-based question-answering services. In: Proceedings of the 14th ACM International Conference on Information and Knowledge Management - CIKM 2005, pp. 315–316 (2005)

14. Loftus, G.R.: Evaluating forgetting curves. J. Exp. Psychol. Learn. Mem. Cogn. **11**(2), 397–406 (1985)
15. Wagner, C., Liao, V., Pirolli, P., Nelson, L., Strohmaier, M.: It's not in their tweets: modeling topical expertise of twitter users. In: SocialCom/PASSAT 2012, pp. 91–100 (2012)
16. Whiting, S., Zhou, K., Jose, J., Alonso, O., Leelanupab, T.: CrowdTiles: presenting crowd-based information for event-driven information needs. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management - CIKM 2012, pp. 2698–2700 (2012)
17. Yimam-Seid, D., Kobsa, A.: Expert-finding systems for organizations: problem and domain analysis and the DEMOIR approach. J. Organ. Comput. Electron. Commer. **13**(1), 1–24 (2003)
18. Zhang, J., Ackerman, M.S., Adamic, L.: Expertise networks in online communities: structure and algorithms. In: Proceedings of the 16th International Conference on World Wide Web - WWW 2007, pp. 221–230 (2007)