Finding Knowledgeable Groups in Enterprise Corpora

Liang, S.; de Rijke, M.

Published in:
SIGIR '13: the proceedings of the 36th International ACM SIGIR Conference on Research & Development in Information Retrieval: July 28-August 1, 2013, Dublin, Ireland

DOI:
10.1145/2484028.2484109

Citation for published version (APA):
https://doi.org/10.1145/2484028.2484109
ABSTRACT

The task of finding groups is a natural extension of search tasks aimed at retrieving individual entities. We introduce a group finding task: given a query topic, find knowledgeable groups that have expertise on that topic. We present four general strategies to this task. The models are formalized using generative language models. Two of the models aggregate expertise scores of the experts in the same group for the task, one locates documents associated with experts in the group and then determines how closely the documents are associated with the topic, whilst the remaining model directly estimates the degree to which a group is a knowledgeable group for a given topic. We construct a test collections based on the TREC 2005 and 2006 Enterprise collections. We find significant differences between different ways of estimating the association between a topic and a group. Experiments show that our knowledgeable group finding models achieve high absolute scores.

Categories and Subject Descriptors
H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing

Keywords
Group finding, expertise retrieval, language modeling

1. INTRODUCTION

A major challenge within any organization is managing the expertise within the organization such that groups with expertise in a particular area can be identified [2]. Rather than finding knowledgeable individuals, sometimes locating a group with appropriate skills and knowledge in an organization is of great importance to the success of a project being undertaken [6].

Traditional approaches to finding knowledge, whether in individuals or in groups within an organization, often include two main steps. For a given task the expertise of the experts in each group is recorded and then the expertise of a group is computed by aggregating the expertise values of all group members. Both steps are traditionally done manually and require considerable effort. In addition, this approach is usually restricted to a fixed set of expertise areas [7]. To reduce the effort of recording and evaluating the expertise of people from their representations, many automatic approaches have been proposed. There has been an increasing move to automatically extract such representations for evaluating expertise from heterogeneous document collections [2]. To compute the expertise values of a group, in principle, many aggregation operators are available, e.g., sum or average. These can be employed to combine individual experts’ expertise values. There are at least 90 families of aggregation operators [11], which have been put to use in a range of applications. But the problem of how to aggregate expertise values of experts within a group so that the expertise scores of different groups can be easily compared and ranked is unknown.

We treat the problem of finding a knowledgeable group differently. Four distinct models are proposed. Our models are based on probabilistic language modeling techniques. Each model ranks groups according to the probability of the group being a knowledgeable group for a query topic, but the models differ in how this is performed. Three types of variable play a key role in our estimations: groups (G), queries (Q) and documents (D). The order in which we estimate these is reflected in our naming conventions. E.g., the model named GDQ proceeds by first collecting evidence of whether a group is knowledgeable on the topic via the experts in the group (G), and then determining whether each expert in the group has expertise on the topic via documents (D), and finally whether a document is talking about the given query (Q) topic.

2. RELATED WORK

Significant research effort has been invested in locating a group of individuals in an organization. Yang et al. [10] try to find a group of attendees familiar with a given activity initiator, and ensure each attendee in the group to have tight social relations with most of the members in the group. Sozio and Gionis [9] study a query-dependent variant of the community-detection problem: given a graph, and a set of nodes in the graph as their input query, find a subgraph that contains the input query nodes and is densely connected. Lappas et al. [6] study the problem of given a task, a pool of individuals χ with different skills and a social network that captures the compatibility among them, finding a subset of χ, who together have the skills to complete a task with minimal communication costs. Kargar and An [5] design communication cost functions for two types of communication structures.

The problem we deal with is different. We introduce a new group finding task: given a topic query, determine a list of knowledgeable groups within which the experts have expertise on the topic. Our group finding problem includes two sub-problems. The first is to answer questions such as “Which groups are knowledgeable groups on topic T?” whilst the second is to answer the question “What does group G know?” We focus on the first sub-problem.
3. MODELING GROUP FINDING

In our modeling of the knowledgeable group finding task, groups, documents and queries are considered in different orders. Groups are ranked according to how likely they have expertise on the given query according to the estimated language model.

Problem definition and context. We address the following problem: what is the probability of a group \( g \) being a knowledgeable group given query topic \( q \)? We have to estimate the probability of a group \( g \) given a query \( q \) and then rank groups according to this probability. The top \( k \) groups will be considered to be the most knowledgeable groups for the given query topic. Instead of computing this probability directly, we apply Bayes’ Theorem, and obtain \( p(g|q) = p(q|g)p(g)\), where \( p(q|g) \) is the probability of a query and \( p(g) \) the probability of a group, both of which can be assumed to be uniform for a query and a group, considering that \( q \) is the same during retrieval and there is no group that is more likely to be relevant. Hence, ranking groups according to the estimated language model.

We address the following probability given query topic \( q \) and term \( t \) given document \( d \), and \( p(d|ex) \) is the probability of \( d \) given expert \( ex \). Now we can obtain the probability of a group given a query, i.e., our GQD model:

\[
p(g|q) = \text{rank} \sum_{ex \in g} \left\{ \prod_{t \in q} p(t|d) p(d|ex)^{\alpha_{s(ex,g)}} \right\}^{a_{s(ex,g)}}
\]

The DGQ Model. We can compute the probability of a group \( g \) given an expert \( ex \) in a different way. By taking the sum over all documents \( d \), \( p(q|ex) \) can be obtained. Formally, this can be expressed as: \( p(q|ex) = \sum_d p(q|d)p(d|ex) \), where \( p(q|d) \) and \( p(d|ex) \) are the probability of query \( q \) given document \( d \) and of document \( d \) given query \( q \), respectively. Based on this, we obtain our second aggregation model, i.e., our DGQ model:

\[
p(g|q) \overset{\text{rank}}{=} \sum_{ex \in g} \left\{ \prod_{t \in q} p(t|d) p(d|ex)^{\alpha_{s(ex,g)}} \right\}^{a_{s(ex,g)}}
\]

The DGQ model. Next we consider a document model. Instead of aggregating expertise scores of all the experts within a group as in our aggregation models, as the key terms DGQ in this model’s name suggests, the probability \( p(g|q) \) can be computed directly via the documents \( D \). For each we compute how likely the group \( G \) is associated with it, and how likely it is talking about the given query \( Q \), such that: \( p(g|q) = \sum_d p(g|d)p(d|q) \), where \( p(g|d) \) and \( p(d|q) \) are the probability of group \( g \) given document \( d \) and the probability of \( d \) given query \( q \), respectively. This, then, is how \( p(g|q) \) can be represented, i.e., our DGQ model:

\[
p(g|q) \overset{\text{rank}}{=} \sum_d \left\{ \prod_{ex \in g} p(d|ex)^{a_{s(ex,g)}} \right\}^{a_{s(ex,g)}}
\]

The DGQ model. Finally, we present a query model for group finding. We use “query model” not in the sense of building rich representations of a query but to indicate that our estimations of a group finding model start with the query. As the name DGQ indicates, it first considers how likely a group knows about a query \( Q \) topic. DGQ computes this via documents \( D \) and then determines how likely each expert in the group \( G \) is associated with each document. We collect evidence of how knowledgeable group \( g \) is via all documents in the collection and obtain \( p(t|g) = \sum_d p(t|d)p(d|g) \), where \( p(d|g) \) is the probability of document \( d \) given group \( g \). The final version of \( p(g|q) \) can then be represented as:

\[
p(g|q) \overset{\text{rank}}{=} \sum_{ex \in g} \left\{ \prod_{t \in q} p(t|d) \prod_{ex \in g} p(d|ex)^{a_{s(ex,g)}} \right\}^{a_{s(ex,g)}}
\]

And this is our DGQ model. It first computes the probability of how likely a group is talking about a query topic; it collects evidence of how knowledgeable the group is for a given query via all documents in the collection. For each expert within a group, we determine how likely the expert is associated with the documents.

4. ASSOCIATIONS AND SMOOTHING

Expert-document associations. For all models described in the previous section, we need to be able to estimate the probability of an expert \( ex \) in group \( g \) being associated with document \( d \). In recent years, this problem has attracted considerable attention [2]. Following Balog et al. [1], to define this probability, we assume that associations \( a(d,ex) \) between experts \( ex \) and documents \( d \) have been calculated and define

\[
p(d|ex) = \frac{a(d,ex)}{\sum_{d \in D} a(d,ex)}
\]

where \( D \) is the set of documents in the collection, and \( a(d,ex) \) is simply defined as to be 1 if the full name or email address of expert \( ex \) (exactly) appears in document \( d \), otherwise \( a(d,ex) = 0 \).
Group-expert associations. For all of the group finding models described in the previous section, we also need to be able to estimate the strength of the association between expert $ex$ and group $g$ to which the expert belongs. We define the following group expert association $as(ex, g) = |g|^{-1}$, where $|g|$ is the total number of experts within the group to which they belong.

Smoothing strategies. In our four models, the term $p(g|q)$ may contain zero probabilities due to data sparsity. E.g., in our aggregation models, GQD and DGQ, $p(g|q)$ will contain zero probabilities if there exist experts who have no expertise on the given query. Hence, we have to infer a group model $θ_g$, such that the probability of a group given a query model is $p(θ_g|q)$. We employ Jelinek-Mercer smoothing [4] to estimate $p(θ_g|q)$: we consider two types.

To facilitate comparisons and for the sake of uniformity, instead of estimating $p(g|q)$ directly, we can easily infer a document model $θ_d$ such that the probability of term $t$ given a document $d$ model is $p(t|θ_d)$, and infer an expert model $θ_ex$ such that the probability of a document $d$ given an expert $ex$ is $p(d|θ_ex)$. The document model, then, is a linear interpolation of the background model $p(t)$ and the smoothed estimate: $p(t|θ_d) = (1 - α)p(t|d) + αp(t)$, where $α$ is a smoothing parameter ($0 < α < 1$). The expert model is a linear interpolation of the background model $p(d)$ and the smoothed estimate: $p(d|θ_ex) = (1 - β)p(d|ex) + βp(d)$, where $β$ is a smoothing parameter ($0 < β < 1$). Let $θ(α, t, d)$ be short for $p(t|θ_d) = (1 - α)p(t|d) + αp(t)$, and $θ(β, d, ex)$ be short for $p(d|θ_ex) = (1 - β)p(d|ex) + βp(d)$. Then, the group finding model GQD can be smoothed and estimated as

$$p(g|q)_{rank} = \prod_{ex \in g} \left\{ \prod_{t \in q} \left( θ(α, t, d) \cdot θ(β, d, ex) \right)^{n_t(q)} \right\}^{as'},$$

where $as'$ abbreviates $as(ex, g)$. The other group finding models, GDQ, DGQ, and QDG, can be smoothed and estimated in an analogous manner. As to the GDQ model:

$$p(g|q)_{rank} = \prod_{ex \in g} \left\{ \sum_d \left( θ(α, t, d) \cdot θ(β, d, ex) \right)^{n_t(q)} \right\}^{as'},$$

with $as'$ as before. For the DGQ model we have

$$p(g|q)_{rank} = \sum_d \left\{ \prod_{ex \in g} θ(β, d, ex)^{as(ex, g)} \prod_{t \in q} θ(α, t, d)^{n_t(q)} \right\},$$

and the QDG model can be smoothed and estimated as

$$p(g|q)_{rank} = \prod_{t \in q} \left( \sum_d θ(α, t, d) \prod_{ex \in g} θ(β, d, ex)^{as(ex, g)} \right)^{n_t(q)}.$$

For the GDQ model, we also consider a second type of smoothing strategy with one parameter: $p(g|q)_{rank} = \prod_{ex \in g} \left( (1 - λ) \sum_d p(t|d)p(d|ex) + λp(t) \right)^{n_t(q)}^{as'}$.

5. EXPERIMENTAL SETUP

Next, we describe the experimental setup for testing our knowledgeable group finding methods. We specify our research questions, describe our data set, and detail our ground truth.

Research questions. We consider the following questions. How do different group finding models perform compared against each other, under different ground truths or different evaluation metrics? Are some queries harder than others for the same model? Finally: Are the models different from each other?

Experimental collection. For evaluation purposes we use data made available for the expert finding task at the TREC 2005 and 2006 Enterprise tracks [3, 8]. The document collection used is a crawl of the World Wide Web Consortium (W3C; 330K documents, 5.7GB). The expert finding qrels for the two years differ: in 2005, 50 working group titles were the test topics for the expert finding task, resulting in 1509 expert-group pairs, with 2 to 391 experts in the same group and approximately 30 experts per group on average; names and e-mail addresses of 1092 expert candidates (W3C members) are given as part of the collection. For the TREC 2006 expert finding task, 55 queries queries were created, but only 49 are provided with expert finding ground truth.

Three types of ground truth. We use the qrels of the TREC 2006 expert finding task and propose three types of ground truth: binary, graded and number.

Binary A working group $g$ is considered relevant for topic $q$ if there is at least one expert $ex$ who is a member of $g$ (according to the TREC 2005 expert finding qrels) and has expertise on the topic $q$ (according to the TREC 2006 expert finding qrels).

Grade A slightly more sophisticated definition of group relevance uses grades: the level of relevance of group $g$ for query $q$ is defined based on the fraction of the experts in the group. We distinguish between $|L|$ different levels of relevance, i.e., $\{0, 1, 2, \ldots, |L| - 1\}$. The relevance grade of group $g$ for topic $q$ is defined as follows: let $f(q, g) = \frac{|\{ex \in g : rel(ex, q) > 1\}|}{|g|}$, where $\{ex \in g : rel(ex, q) = 1\}$ is the set of experts in $g$ with expertise on topic $q$ according to the TREC 2006 expert finding qrels and $|g|$ is the total number of experts in group $g$. If $f(q, g) < \frac{1}{|L|}$, the level grade for this group is $l$. In this paper, we set $|L| = 10$.

Number Here, the level of relevance of group $g$ for query $q$ is defined based on the number of experts in the group. For instance, if there are 15 experts who have expertise on the given query topic, then the level of the relevance for this group is 15. The level of relevance ranges from 0 to 30 with a majority smaller than 4.

Runs. We run our experiments with all documents in the collection for our four group finding models. We perform a grid search to find optimal settings of the smoothing parameters (with 0.1 increments).

We generate runs on the full collection and on subsets defined by taking the top $n$ documents returned by a standard document retrieval run when using the topic as query. Evaluation measures used are MAP, precision@5, 10, nDCG and nDCG@5, 10 against our three types of ground truth. Evaluation was done using the trec_eval program (available from http://trec.nist.gov).

6. RESULTS

We start by comparing the results of the optimized models with two smoothing parameters. Next, we present the results of query differences. Finally, we test whether the models smoothed by two or one parameters are statistically significantly different.

Model comparison. How do our knowledgeable group finding models perform compared to each other? For each specific performance evaluation metric, we compare the models using optimal smoothing parameters. We use two parameters $α$ and $β$ to smooth the proposed four knowledgeable group finding models, i.e., GQD, GDQ, DGQ and QDG.

Table 1 lists the scores for the various metrics. Clearly, DGQ outperforms the other models on all metrics using the binary and graded ground truth, but GDQ outperforms DGQ on all metrics using the number ground truth. QDG model is the worst performing model for all metrics and against all types of ground truth. The table also shows that GQD, GDQ and DGQ have a similar performance for all metrics against all types of ground truth. (The MAP and p@N scores against the number ground truth are the same as those against the binary ground truth, and are therefore omitted.)

Query differences. Our aim here is to find out whether some queries are harder than others for the same model against the metrics. We turn to a topic-level analysis of the MAP performance for
Table 1: Evaluation results for all optimal models with two smoothing parameters, using the binary, graded and number ground truths. For each metric, we report the evaluation results, followed by the optimal smoothing parameters $\alpha$ and $\beta$.

<table>
<thead>
<tr>
<th>Ground truth</th>
<th>Model NDCG@ $\alpha$ $\beta$ 5 $\alpha$ $\beta$ MAP $\alpha$ $\beta$ 5 $\alpha$ $\beta$ 10 $\alpha$ $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>binary</td>
<td>GQD .8861 1.2 .8165 .1 1.2 .7850 .1 1.3 .7127 .1 .7804 .1 .9 .7306 .1 .3</td>
</tr>
<tr>
<td></td>
<td>GDQ .9099 1.9 .8291 .3 1.5 .7936 .2 1.5 .7552 .1 1.9 .8122 .1 .6 .7408 .2 .8</td>
</tr>
<tr>
<td></td>
<td>DGQ .9133 1.9 .8680 .1 1.9 .8420 .1 1.9 .7772 .1 .9 .8571 .1 .9 .7918 .1 .9</td>
</tr>
<tr>
<td></td>
<td>QDG .7623 1.2 .5694 .1 1.1 .5568 .1 1.2 .4882 .1 .4 .5992 .1 .1 .5265 .1 .2</td>
</tr>
<tr>
<td>graded</td>
<td>GQD .8245 2.3 .7237 2.3 .7595 1.2 .7403 .1 .3 .6367 .1 .1 .5224 .1 .2</td>
</tr>
<tr>
<td></td>
<td>GDQ .8457 .4 .7675 .3 .7945 1.4 .7673 .1 .4 .6776 .1 .3 .5510 .1 .4</td>
</tr>
<tr>
<td></td>
<td>DGQ .8631 2.9 .7991 .5 1.9 .8160 .7 1.8 .8092 6.8 .7102 5.6 .5331 .7 .9</td>
</tr>
<tr>
<td></td>
<td>QDG .3916 1.8 .1012 .1 .1 .1282 .1 .2 .2182 .1 .4 .1714 .1 .1 .1673 .1 .1</td>
</tr>
<tr>
<td>number</td>
<td>GQD .7964 1.9 .6210 1.8 .6730 .6 .8</td>
</tr>
<tr>
<td></td>
<td>GDQ .8160 1.9 .6496 .1 .9 .6905 .1 1.9</td>
</tr>
<tr>
<td></td>
<td>DGQ .7907 1.9 .6062 .1 1.9 .6758 .1 .9</td>
</tr>
<tr>
<td></td>
<td>QDG .6222 1.2 .3328 .1 .1 .4258 .1 .1</td>
</tr>
</tbody>
</table>

Figure 1: Topic-level differences from the mean scores for GQD using the binary/number and graded ground truths.

(a) GQD, binary  
(b) GQD, graded

We introduced a group finding task. We proposed four models, GQD, GDQ, DGQ and QDG. We also constructed an experimental collection by using the TREC 2005 and 2006 Enterprise collections. We introduced three kinds of ground truth and evaluated our models along many dimensions. Directly collecting expertise evidence from documents is the most effective way to find knowledgeable groups when using the binary or graded ground truths, and aggregating the expertise of each experts in the same group can also be a good way to find the groups. Our models are not very sensitive to changes of the parameters when using a two parameter smoothing strategy. We found statistically significant differences between the models when using MAP scores based on multiple types of ground truth.

Statistical significance. Finally, we determine whether the observed differences between our group finding approaches with two smoothing parameters strategy are statistically significant. We use a two-tailed paired t-test between two models on NDCG and MAP data and test for significance differences at the 0.95 confidence level. Table 2 indicates that when using NDCG as a metric the differences between GQD and GDQ against graded ground truth are not significant. This is also true for the differences between GQD and DGQ, and between GDQ and DGQ against the graded and number ground truth. When using MAP, the differences between all models are statistically significant except for those between GQD and GDQ. We also test differences between the optimal GQD model with smoothing with two parameters and with a single smoothing parameter based on NDCG and MAP against three types of ground truth; there are no statistically significant differences at the 0.95 confidence level except based on NDCG against binary ground truth where the p value is 0.0270. Hence, we cannot really distinguish between GQD with smoothing with two parameters and GQD with smoothing with one parameter in our experiments.

7. CONCLUSIONS

8. REFERENCES