Effects of Expertise Assessment on the Quality of Task Routing in Human Computation

Umair ul Hassan
Digital Enterprise Research Institute
National University of Ireland, Galway
umair.ul.hassan@deri.org

Sean O'Riain
Digital Enterprise Research Institute
National University of Ireland, Galway
sean.oriain@deri.org

Edward Curry
Digital Enterprise Research Institute
National University of Ireland, Galway
edward.curry@deri.org

ABSTRACT

Human computation systems are characterized by the use of human workers to solve computationally difficult problems. *Expertise profiling* involves assessment and representation of a worker's expertise, in order to route human computation tasks to appropriate workers. This paper studies the relationship between the assessment workload on workers and the quality of task routing. Three expertise assessment approaches were compared with the help of a user study, using two different groups of human workers. The first approach requests workers to provide self-assessment of their knowledge. The second approach measures the knowledge of workers through their performance against tasks with known responses. We propose a third approach based on a combination of self-assessment and task-assessment. The results suggest that the self-assessment approach requires minimum assessment workload from workers during expertise profiling. By comparison, the task-assessment approach achieved the highest response rate and accuracy. The proposed approach requires less assessment workload, while achieving the response rate and accuracy similar to the task-assessment approach.

Author Keywords

Human computation, expertise assessment, task routing

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Collaborative Computing, H.3.4 Systems and Software: User profiles and alert services

General Terms

Human Factors; Experimentation

INTRODUCTION

Human computation research focuses on the design of algorithms that leverage humans for solving computationally hard problems [5]. Recent applications have successfully utilized humans for tasks such as image tagging [1], query processing [8], and travel planning [14]. Human computation systems have *explicit control* of the crowd, as compared to general crowdsourcing [5]. In other words, algorithms are responsible for definition, assignment and execution of tasks that are computed by human workers. Assignment of tasks to appropriate humans, also known as *task routing*, is one of the major aspects of human computation. Current human computation research defines two methods of task routing

- *Pull routing* lets humans actively select tasks by using the search and browse capabilities of the platform.

- *Push routing* takes active control of the routing decision, as humans receive tasks assigned to them.

This work addresses the problem of operationalizing *push routing* in human computation systems. Effective push routing requires an understanding of the *expertise* of human workers, for the purpose of matching tasks with appropriate workers. Research questions related with push routing include:

- How to define the expertise requirements of a task? Simple tasks can be performed easily by all human workers. Knowledge intensive tasks may require a higher degree of knowledge or skills, therefore underlining the necessity of expertise.

- How to profile the expertise of human workers, via suitable assessment methods and representation models?

- How to design push routing techniques that effectively leverage the expertise of workers, while considering the requirements of tasks?

We focus on the first two challenges of push routing, while using simple task routing techniques. In the rest of paper, we refer to humans and workers interchangeably. The rest of this section defines the *expertise profiling* problem in human computation and highlights our contributions.

Problem Definition

Human expertise has been studied in other fields such as operations research and cognitive science, to model the knowledge and skills of humans in specific areas [7,9]. Information retrieval techniques have been applied to gather evidence of a person's expertise within large corpuses of digital documents [2]. However, the study of human expertise for routing tasks within human computation is still underexplored. In this paper, we examine the effects of expertise assessment on the quality of task routing. The measurement of worker's expertise

becomes critical in an operational system; specifically when worker expertise changes over time or workers are limited in numbers and expertise.

Modeling the *expertise requirements* of a task is the first challenge of push routing. We model the expertise requirements of a task in terms of concepts related to the task at hand [10]. For example, if a task requires workers to verify a fact about the "Die Hard" movie, then the genre "Action films" of the movie serves as the related concept. Given a set of human workers, the second challenge of push routing involves *expertise profiling* of workers against concepts. Therefore, the *expertise assessment* problem involves finding out the expertise level of a worker against a given concept.

Proposed Solution

We compare three approaches of expertise assessment for building profiles of workers: 1) the *self-assessment* of a worker's knowledge against concepts, 2) the *task-assessment* by observing worker's performance on test tasks (i.e. tasks with known responses), and 3) a proposed combination of 1 and 2 that filters test tasks according to self-assessed conceptual knowledge. The *self-assessment* approach is suitable for the cases where no other source of information about worker's expertise is feasible. By contrast, the *task-assessment* approach is synonymous to the measurement of empirical accuracy of workers. We do not consider other methods of expertise profiling, such as expertise retrieval and social network analysis, due to their use of external information sources.

We have conducted a user study to compare the assessment approaches along two dimensions: *cost* and *quality*. The study employs human computation for improving the data quality of a web-based knowledgebase, using real workers. The *cost* of assessment approaches is defined in terms of the workload (i.e. number of decisions) on workers during assessment. The *quality* of resulting expertise profiles, in supporting task routing, is measured in terms of the response rate and accuracy. Results suggest that expertise profiles generated through task-assessment approach performed the best. However, profiles generated by using combined approach were effective in reducing the workload of assessment while achieving the quality similar to the task-assessment approach.

Contributions

The specific contributions of this paper are

- A combined self-assessment and task-assessment approach for building expertise profiles of workers. The approach when compared to baseline techniques shows between 20% to 50% reduction in worker's workload of expertise assessment.

- The quality of the resulting expertise profiles, in terms of supporting task routing, is comparable to the baseline techniques. The variation in response rate and accuracy, on routed tasks, is within 8% and 5% respectively.

The rest of this paper is organized as follows. Section 2 provides the review of existing research work in closely related areas. In Section 3 we provide details of the knowledge intensive tasks used for the user study, followed by the specifics of the experimental evaluation. Section 4 presents the results of experiment and their implications are discussed in Section 5. Finally we conclude in Section 6 and suggest some directions for the future research.

RELATED WORK

We build upon three main areas of related work: expertise assessment, expertise profiling, and task routing.

Expertise Assessment

The operations research and cognitive science communities have been active in *expertise assessment* research [7,9,12]. In this context, the goal of assessment is to fit a mathematical model on observed data of individuals. For instance, David et al. [12] argued that consistency and discrimination are the fundamental characteristics of expertise, they developed the *Cochran-Weiss-Shanteau* (CWS) index to distinguish between experts and non-experts. However, the CWS index is suited for qualitative comparisons instead of quantitative modeling of expertise.

Lee et al. [7] developed cognitive models for measurement of expertise using the differences between responses of workers. The approach is limited by the dependency on large number of workers, required for the effective assessment via cross examination of responses to the tasks. These approaches have limited applicability, due to their modeling of expertise for specific task types, specifically in situations where the type and skills of tasks are not defined beforehand. By comparison, we use similar assessment approaches for expertise profiling around concepts associated with the knowledge domain of tasks.

Expertise Profiling

Informational retrieval approaches take a user-centric view of expertise inference and modeling. It is assumed that a user is interested in searching for evidence of expertise within a corpus of documents such as emails, publications, webpages, etc. In this context, *expertise profiling* is the process of inferring the competence level of an individual on particular concepts or topics [2,10].

Current expertise profiling approaches associate concepts with persons by analyzing the evidence of association within given corpora. The expertise is modeled in terms of a matrix having concepts and experts as rows and columns respectively [2]. While effective for searching well documented experts these approaches fail to account for the performance of experts on specific tasks. Furthermore these approaches cannot be used in the cases where textual information is limited or not available. By comparison, we measure the performance of workers on test tasks and calculate expertise level for concepts related to the tasks.

Dataset Characteristics	Movies Dataset	Actors Dataset
Total entities	724	14
Total concepts	42	14
Total tasks	230	120
Avg. tasks per concept	9	8.6
Avg. concepts per task	1.64	1

Table 1: Characteristics of the Movies and Actors datasets describing entities describing movies and actors in DBpedia. Fact verification tasks are based on attribute values of the entities. The concepts associated with each task were based on genres for Movies dataset and actor names for Actors dataset.
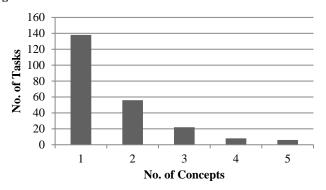

Figure 1: Distribution of number of tasks versus number of concepts in Movie Tasks dataset.

Task Routing

Matching tasks with workers in a crowd or community has been an active area of research in recent years. For instance, Law et al. [6] studied the effects of self-rated expertise, interests, confidence and understanding, on pull based task selection by crowd workers. The study remained limited to relevance judgment tasks only. Zhang et al. [13] proposed *peer routing*, a rules-based incentivisation method to support people in jointly contributing to task solution and routing decisions. Peer routing relies on assessment of neighbors' expertise in a social network, as opposed to worker specific methods discussed in is this paper.

The task routing problem has also been studied in the context of online communities. Zhou et al. combined three approaches to profiling users, based on information available in online question answering systems, for actively pushing question to appropriate users [15]. The value of intelligent task routing in community maintained knowledge system has been demonstrated in recent studies [3]. By contrast, we attempt to study the expertise assessment for human computation instead of explaining the human behavior in a specific system.

METHOD

We have conducted a user study to analyze the workload of expertise assessment on workers, as well as to study the effects of expertise assessment on the quality of push-based task routing.

Knowledge Intensive Tasks

We consider the problem of data quality as an application area of human computation. For this purpose, the DBpedia[1] project serves as an appropriate use case. DBpedia aims at creating a database of facts about real world entities such as cities, actors, books, games, etc. However, DBpedia suffers from data quality issues such as incorrect values, incorrect mappings, and missing values [4]. Consequently, applications using DBpedia need to review the data with the help of humans or experts [11].

A set of knowledge intensive tasks was created from entities in DBpedia. Each task simply required human verification of a fact in DBpedia. Tasks were created for two types of entities in DBpedia; Movies and Actors. For example, following task verifies birthplace of an actor

> **Fact:** *Tom Hanks was born in Concord, California.*
> **Do you:** *Agree, Don't Know, Disagree*

Some of the concepts associated with this fact that could be used for expertise profiling include *famous actors*, *Oscar winners*, *California*, etc. Generally speaking the concepts can be any topic or keyword describing a task and workers knowledge. In this context, the task-related concepts are based on Movies and Actors classification schemes used within DBpedia.

Datasets

This section details the two datasets used in user study, as summarized in the Table 1. The tasks in each dataset were related to films. The choice of creating datasets from films has two advantages; it is relatively easy to recruit people for the study and people have varying degree of knowledge about film depending on various factors such as genre, language, actors, etc.

The Movies dataset was created by selecting Academy Award winning, Indian FilmFare Award winning, and top 100 grossing movies (from both Hollywood and Bollywood). DBpedia provides variety of concept schemes for entities. We chose the 42 film genres associated with the selected movies to serve as concepts. Each task consisted of a fact about a movie entity, where the genres of the movie served as the concepts related to the task. The distribution of tasks against the number of concept is shown in Figure 1.

The Actors dataset was also generated manually by selecting popular persons 10 from Hollywood and 4 from Bollywood. In this case, the names of the movie stars served as concepts thus providing close relationship with their associated tasks. The objective of this selection was to facilitate easy association of conceptual knowledge with the task response, during assessment. Similar to the Movies dataset, a task required feedback on a fact about actor entity.
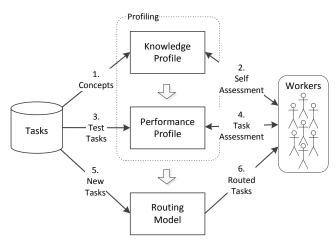
---

[1] www.dbpedia.org

Figure 2: A workflow of experimental prototype used for data collection from workers. The workflow shows, from top to bottom, the two phases of profiling and routing. The profiling phase has steps of self-assessment and task-assessment, which can be performed in different combinations.

Profiling, Assessment, & Routing

Figure 2 illustrates the workflow of web-based prototype developed for the user study. The prototype employed push-based task routing approach that was supported by expertise profiles of workers. First, the worker provides rating of her own knowledge level for each concept. Second, the worker performs test tasks. Finally, the routing model exploits the profiles, generated in previous steps, for assigning tasks to appropriate workers. This workflow shows one particular realization of the steps of the profiling and task routing process; however other variations are also possible. For example, the routing model can directly utilize knowledge profile while ignoring the performance based profile.

We follow a two phased process of *profiling* and *routing* for our experiments [5]. Tasks in each dataset were divided into two mutually exclusive sets of *test tasks* and *routed tasks*, to be used in each phase respectively. The two phases of the experiments as described as

- The *profiling* phase builds worker profiles according to the selected assessment approach. During this phase the workers were required to rate their knowledge about concepts and/or provide responses to test tasks.

- The *routing* phase uses the profiles built earlier for assigning tasks to the appropriate workers. The responses to *routed tasks* were used for calculating the accuracy of responses gathered during this phase.

Next we discuss the modeling of expertise profiles and the assessment approaches used for generating those profiles. Furthermore, we describe the routing strategies that leverage the expertise profiles for assigning tasks to the appropriate workers.

Concept	w_1	w_2	w_3
c_1: Tom Hanks	0.6	0.2	0.2
c_2: Johnny Depp	0.6	0.2	0.6
c_3: Tom Cruise	0.8	0.4	0.4
c_4: Amir Khan	0.8	0.6	0.6

Table 2: Example of a matrix representing knowledge profiles of three workers, for some concepts taken form the Actors dataset for the purpose of illustration.

Expertise Profiling

A profile is defined in terms of the knowledge level of a person for a given concept. Let W be the set of worker, C be the set of all concepts associated with the set of all tasks T, and $A \subset T$ be the set of test tasks with known responses used for task-assessment. Given that that $|W| = m$ and $|C| = n$, the knowledge profiles are created through self-assessment and defined as a matrix K, where

$$K(C,W) = \begin{bmatrix} k_{1,1} & \cdots & k_{1,m} \\ \vdots & \ddots & \vdots \\ k_{n,1} & \cdots & k_{n,m} \end{bmatrix}$$

Therefore, the knowledge profiling involves calculating $k_{i,j}$ from the rating of worker's knowledge provided by her. The normalized $k_{i,j}$ is a real value between 0 and 1, that quantifies knowledge level of worker j for the concept i. For example Table 2 shows knowledge profiles of 3 workers for 4 concepts related to movies.

Similarly performance profiles of worker are defined as a matrix E, which is generated during task-assessment

$$E(C,W) = \begin{bmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{bmatrix}$$

where $e_{i,j}$ quantifies the expertise level of worker j for the concept i, associated with test tasks. In the next section, we detail the three assessment approaches used for populating the knowledge and performance profiles.

Expertise Assessment

We compare three approaches of expertise assessment

- *Self-Assessment*: The knowledge profile is generated by asking workers to provide self-assessment of their knowledge for each concept. We used a simple ordered belief scale for rating knowledge level, to help workers provide their self-assessment. We used a 5 level belief scale for conceptual knowledge rating; with ordered level of *none*, *poor*, *fair*, *good*, and *excellent* respectively. The selected knowledge level is converted to a normalized value, to be used in knowledge profiles.

- *Task Assessment*: The performance profile is calculated from the worker's responses to test tasks. For each concept the expertise level is recorded as the percentage of correct responses to relevant test tasks. For example, if

a worker provides 3 correctly response out of 4 test tasks associated with concept "gang films", then the expertise level is considered to be 0.75

- *Combined Assessment*: The proposed approach in which test tasks are filtered based on self-assessed knowledge of concepts. The worker is asked to rate their conceptual knowledge, followed by task assessment on a subset of test tasks (filtered according to knowledge level of concepts).

Task Routing

The profiles, described in previous section, are exploited by the push routing model. Given that, T is the set of all tasks, C is the set of all concepts, and W is the set of all workers. The routing model matches tasks with appropriate workers. Assuming that c_t denotes the set of concepts associated with a task t, such that $c_t \subseteq C$. We define the task routing problem as ranking the workers for assignment of task t, according to expertise of workers for concepts in c_t. For this purpose, we employ four strategies of calculating the ranking score for an individual worker $w \in W$

- *Random* (*RND*) approach assigns a random value sampled from uniform distribution U

$$score(w) = U(0,1)$$

- *Self-Assessment* (*SA*) defines ranking score as the average of the values in knowledge profile K, only for the worker w and concepts in c_t

$$score(w) = avg\ (K(c_t, w))$$

- *Task Assessment* (*TA*) assigns the average of the values in performance profile E, only for worker w and concepts in c_t, as ranking score

$$score(w) = avg\ (E(c_t, w))$$

- *Combined Assessment* (*CA*) combines knowledge and performance profiles for routing. The ranking score is defined as the average of pair-wise multiplication of knowledge and performance profiles.

$$score(w) = avg(\ K(c_t, w) \times E(c_t, w))$$

A discussion on effects of different levels of filtering on quality and workload of knowledge workers is provided later in results section.

Knowledge Workers

The participants of the user study were recruited through an open call in a research institute. Separate calls were made for Movies and Actors datasets. The resulting two groups of workers consisted of participants coming from countries in Asia, Europe and America. Since, some workers were from South Asian countries, they possessed higher level of knowledge about concepts and tasks related to Bollywood films. Table 3 summarizes the number of workers and division of tasks for each dataset. During the data collection exercise each worker was asked to perform both self-assessment and task-assessment, through the prototype described earlier. Additionally workers had to provide

	Movies Dataset	Actors Dataset
No. of knowledge workers	11	26
No. of knowledge concepts	42	14
No. of test tasks (profiling phase)	100	56
No. of new tasks (routing phase)	130	64

Table 3: Number of volunteer knowledge workers recruited for data collection for the Movies and Actors datasets. Also for both datasets, number of tasks used for the performance assessment during *profiling phase* and number of tasks routed to appropriate workers during *routing phase* for evaluation.
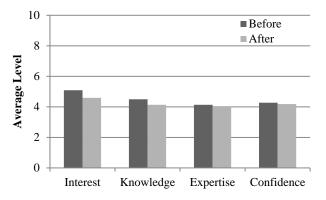

Figure 3: Comparison of the survey performed before and after the experiment, for the Actors dataset. The bars show average level of interest, knowledge, expertise and confidence, as indicated by 22 workers participating in the experiment.

responses to the *routed tasks* assigned to them after profiling phase. The workers were asked to respond quickly and truthfully without looking up answers on the Web.

To measure the effects of the experiment on participants, we performed a pre and post survey, from the group of workers for the Actors dataset. The survey asked the participants to indicate, on 10 point belief scale, their level of 1) *interest* in information about actors, 2) *knowledge* about actors, 3) *expertise* in answering question about actors, and 4) c*onfidence* in answering question about actors.

Figure 3 shows the comparison of results for pre and post surveys. A paired t-test was performed to determine if the belief level of workers, for each question, changed after the experiment. The average loss in *interest* (mean=0.5, standard deviation=1.06, count=22) was significantly greater than zero, where $t(21)=2.22$ and $p=0.04$, providing the evidence that the experiment resulted in decreased *interest* of workers, in information about actors. A 95% confidence interval for average loss in *interest* is (0.03, 0.97). The average difference in the level of *knowledge*, *expertise*, and *confidence* is not statistically significant. Some participants indicated that they lost interest in the experiment due to large number of tasks.

Evaluation

We evaluate the assessment approaches through four routing strategies, discussed in previous section. Each routing strategy was employed to assign the *new tasks* to workers, similar to the situation of an operational human computation system. For each of the new task, the response provided by the assigned worker is compared with known response.

Metrics

We use following metrics to evaluate the quality of routing during *routing phase* and the workload of assessment approaches during *profiling phase*.

- *Response Rate* is the percentage of tasks with either "agree" or "disagree" response, out of all routed tasks during routing phase.

- *Accuracy* is the percentage of correctly responded tasks, out of all routed tasks during routing phase.

- *Workload* is the cognitive load on an individual worker during assessment, in terms of decisions made by her. A decision is either self-rating of the knowledge for a concept, or providing response to a test task.

The *RND* strategy serves as the baseline in terms of workload since it does not leverage any expertise. *TA* strategy achieves maximum possible accuracy and response rate. Objective of the user study was to validate following:

- **Hypothesis:** The routing quality of *CA* strategy approaches the quality of *TA* strategy during routing phase, while requiring comparatively less assessment workload during profiling phase.

RESULTS

In this section we discuss the results of the experiments. We calculate the quality of each routing strategy according to the metrics described earlier. During the routing phase each *new task* was assigned to only one worker, by selecting the top-1 candidate from the ranked (according to the active routing strategy) list of workers.

Routing Quality versus Assessment Workload

Figure 5 shows the comparative quality of task routing strategies, in terms of response rate and accuracy, for both datasets. For each task routing strategy, a two-sample t-test between percentages was performed to determine whether there was a significant difference of quality between datasets. As expected the response rate and accuracy of *RND* were minimum and *TA* strategies were maximum, with no significant difference between datasets. However both metrics had statistically significant different values between datasets for the *SA* strategy. The semantic relationship of concepts (movie genres) and tasks (movie facts) was not strong for the Movies dataset; therefore workers claiming high conceptual knowledge were unable to respond to the assigned tasks. Despite this observation, the quality of *CA* strategy was similar to the quality of *TA* profiles, with no significant difference between datasets.
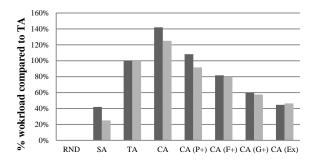

Figure 4: Average workload required from workers during the *profiling phase* for both Movies and Actors datasets. *CA* refers to the routing strategy using combined approach of self-assessment and task-assessment without any filtering, therefore representing the highest level of workload during assessment for expertise profiling.
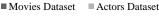
The workload is quantified in terms of total number of decisions made by worker during assessment. For example, the expertise profiles used in *SA* strategy required a worker to makes 42 decisions of concepts ratings during assessment in Movies dataset. The *TA* strategy used profiles generated with 100 decisions of providing responses to the test tasks. Therefore the workload required for profiles used with *CA* includes 42 concept rating decisions and 100 task responses. Clearly, there is an overhead associated with combined approach of assessment, as highlighted by the maximum workload attributed to the *CA* task routing strategy in Figure 4.

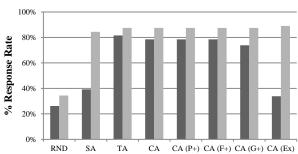**Effects of Filtering on Assessment Workload**

To compensate for the extra workload, due to test tasks, filters were applied according to various levels of concept knowledge in combined approach. As a result, workers were not given test tasks with concepts below a certain knowledge level. Following four level of filtering were applied
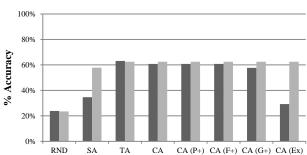
- *CA (P+)*: select test tasks for concepts with minimum *poor* knowledge level on belief scale

- *CA (F+)*: select test tasks for concepts with minimum *fair* knowledge level on belief scale

- *CA (G+)*: select test tasks for concepts with minimum *good* knowledge level on belief scale

- *CA (Ex)*: select test tasks for concepts with only *excellent* knowledge level on belief scale

As shown in Figure 5, the quality of filtering strategy is similar to TA, except for the *CA (Ex)*. The graphs in Figure 6 compare the workload per worker for each strategy (in terms of percentage as compared to *CA*) with the response rate, for both the datasets. The filtered *CA* assessment approaches required comparatively less workload while achieving the levels of quality close to *TA*. The response rate on *new tasks* suffers for *CA (F+)* and *CA (G+)*
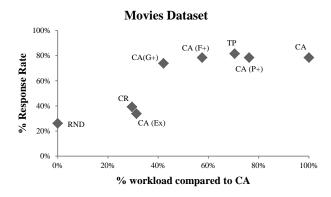
Figure 5: Graphs for the comparison of response rate and accuracy of all routing strategies used for Movies and Actors datasets
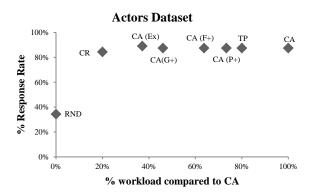

Figure 6: Comparison of assessment approaches for average workload per worker against response rate and accuracy

filtering. Nonetheless, the comparative decrease in quality is less as compared to the rate of decrease in the average workload. However, in case of Movies Dataset the response rates drops below original *SA* approach if filtering is too restrictive i.e. *CA (Ex)* based profiles. Similar results were observed for accuracy versus average workload per worker.

DISCUSSION

In terms of reducing workload while maintaining a high-level of task routing quality, the results demonstrated the effectiveness of filtering within the combined approach of self-assessment and task-assessment. We expect our results to generalize to other memory based tasks like tagging bird types in images, as opposed to observation based tasks such as comparing images. Now we revisit the design of the study and discuss how it may in general inform the design of operationalizing worker performance in human computation systems.

Unified conceptual expertise models for tasks and workers
Compared to the previous works on task routing for human computation, our approach is distinguished in its use of concepts for assessment, representation and exploitation of workers knowledge. This unified approach provides a common framework for representing expertise requirements of a task and expertise profiles of workers, thus allowing effective task routing based on concept matching. Therefore this approach is more suitable, for routing knowledge intensive tasks based on semantics, rather than approaches

where routing decisions are simply based on single measure of empirical accuracy of a worker.

Minimize assessment workload
The self-assessment of conceptual knowledge allows workers to indicate their preferences of tasks to be assigned to them. In our study we found that the sequential process of responding to tasks for task-assessment can be tedious for workers. Therefore limiting the number of test tasks based on self-assessment of knowledge is an effective strategy. The routing decisions based on the resulting knowledge profiles have similar response rate and accuracy; however the cost of building the expertise profile is much lower. Therefore application domains, such as scientific data management, with high diversity of knowledge among workers and across tasks can benefit from our approach.

Relationship between concepts and tasks
The choice of concepts to be used for profiling and routing affects the quality of routing to some extent. Although the general patterns of accuracy and response rate was same for various levels of filtering, there was sharp decline for Movies dataset with very restrictive filters i.e. *CA (Ex)*. In case of Movies dataset the concepts were broader than the concepts for Actors dataset. For example, in Movies dataset the task was related to missing value of a Film entity and a concept was the genre of the same Film entity. In contrast, the task was related to missing value of an Actor entity and

the concept was the same Actor entity. Therefore while some workers felt confident about their excellent level of knowledge for some film genres but were unable to response to specific questions about films from those genres.

Scaling for large number of concepts

While this study used relatively small number of concepts, it would be interesting to study the scalability proposed approach for large number of concepts. Although out of scope of this paper, we suggest some strategies; using concept hierarchies, or applying clustering techniques to group concepts, or using distribution of tasks for ranking important concepts.

CONCLUSION

In this paper, we studied the effects of assessment, in terms of conceptual knowledge, on the quality push-based task routing for human computation. The expertise requirements of tasks and expertise of human workers were defined in terms of concepts. It was observed that expertise profiling of human worker through self-assessment is beneficial for supporting simple task routing. Furthermore, the workload of expertise profiling with task-assessment is reduced by filtering tasks according to self-assessed conceptual knowledge, without sacrificing the quality of task routing significantly.

Enrichment of expertise profiles with information gathered from external sources is part of future work. Analysis of trade-off between assessment and exploitation is also a promising direction for further research. Although the discussion here is limited to using expertise profiling for push-based routing, extending these techniques to pull-based routing would not be difficult.

REFERENCES

1. Von Ahn, L. and Dabbish, L. Labeling images with a computer game. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2004), 319–326.

2. Balog, K. Expertise Retrieval. *Foundations and Trends® in Information Retrieval 6*, 2-3 (2012), 127–256.

3. Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. SuggestBot: using intelligent task routing to help people find work in wikipedia. *Proceedings of the 12th International Conference on Intelligent User Interfaces*, ACM Press (2007), 32–41.

4. Heath, T. and Bizer, C. Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, (2011).

5. Law, E. and Ahn, L. von. Human Computation. In *Synthesis Lectures on Artificial Intelligence and Machine Learning*. 2011, 1–121.

6. Law, E., Bennett, P.N., and Horvitz, E. The effects of choice in routing relevance judgments. *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information in Information Retrieval*, ACM Press (2011), 1127–1128.

7. Lee, M.D., Steyvers, M., De Young, M., and Miller, B. Inferring expertise in knowledge and prediction ranking tasks. *Topics in cognitive science 4*, 1 (2012), 151–63.

8. Marcus, A., Wu, E., Karger, D.R., Madden, S., and Miller, R.C. Crowdsourced Databases: Query Processing with People. *Fifth Biennial Conference on Innovative Data Systems Research Systems Research*, (2011), 211–214.

9. Shanteau, J., Weiss, D.J., Thomas, R.P., and Pounds, J.C. Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research 136*, 2 (2002), 253–263.

10. Ul Hassan, U., O'Riain, S., and Curry, E. Towards Expertise Modelling for Routing Data Cleaning Tasks within a Community of Knowledge Workers. *Proceedings of the 17th International Conference on Information Quality*, (2012).

11. Ul Hassan, U., O'Riain, S., and Curry, E. Leveraging Matching Dependencies for Guided User Feedback in Linked Data Applications. *Proceedings of the Ninth International Workshop on Information Integration on the Web*, ACM Press (2012), 1–6.

12. Weiss, D.J. and Shanteau, J. Empirical Assessment of Expertise. *Human Factors: The Journal of the Human Factors and Ergonomics Society 45*, 1 (2003), 104–116.

13. Zhang, H., Horvitz, E., Chen, Y., and Parkes, D.C. Task Routing for Prediction Tasks. *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, International Foundation for Autonomous Agents and Multiagent Systems (2012), 889–896.

14. Zhang, H., Law, E., Miller, R.C., Gajos, K.Z., Parkes, D.C., and Horvitz, E. Human computation tasks with global constraints. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems*, ACM Press (2012), 217–226.

15. Zhou, Y., Cong, G., Cui, B., Jensen, C.S., and Yao, J. Routing Questions to the Right Users in Online Communities. *Proceedings of the 25th IEEE International Conference on Data Engineering*, Ieee (2009), 700–711.