Information Sciences 460–461 (2018) 346–363

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Context-aware result inference in crowdsourcing

Yili Fang[a,b], Hailong Sun[a,b,*], Guoliang Li[c], Richong Zhang[a,b], Jingpeng Huai[a,b]

[a] *SKLSDE Lab, School of Computer Science and Engineering, Beihang University, Beijing 100191, China*
[b] *Beijing Advanced Innovation Center for Big Data and Brain Computing, Beijing 100191, China*
[c] *Department of Computer Science, Tsinghua University, Beijing, China*

A R T I C L E I N F O

Article history:
Received 24 January 2017
Revised 5 May 2018
Accepted 25 May 2018
Available online 26 May 2018

Keywords:
Crowdsourcing
Human computation
Quality control
Context-sensitive tasks

A B S T R A C T

Many result inference methods have been proposed to address the quality-control problem in crowdsourcing. However, existing methods are ineffective for context-sensitive tasks (CSTs), e.g., handwriting recognition, translation, speech transcription, where context correlation within a task cannot be ignored for two reasons. Firstly, it is ineffective to crowdsource a whole CST (e.g., recognizing handwritten texts) and use task-level inference methods to infer the answer, because it is rather hard to correctly complete a whole complicated task. Secondly, although a CST is composed of a set of atomic subtasks (e.g., recognizing a handwritten word), it is unsuitable to split it into multiple subtasks and adopt a subtask-level inference algorithm to infer the result, because this will lose the context correlation (e.g., phrases) among subtasks and increase the difficulty to complete a task. Thus it calls for a new approach to handling CSTs. In this work, we study the result inference problem for CSTs and propose a context-aware inference algorithm. We design an inference algorithm by incorporating the context information. Furthermore, we introduce an iterative method to improve the quality. The results of experiments on real-world CSTs demonstrated the superiority of our approach compared with the state-of-the-art methods.

1. Introduction

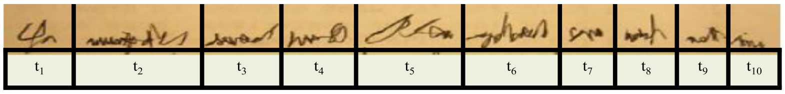
Crowdsourcing aims at leveraging the wisdom of crowds to deal with the problems that are difficult for computers to solve [2]. Its success has been witnessed in numerous applications especially in the area of data management [25], ranging from simple tasks (e.g., image labeling [10,21] and entity resolution [38]) to complex ones (e.g., text editing [4] and software development [19]).

As workers may return noisy results, a core issue of crowdsourcing is to guarantee the result quality [25]. A widely-adopted method to control the quality is *result inference*, which first assigns each task to multiple workers, then uses an inference algorithm to aggregate the results from the assigned workers. Take image labelling as an example. An image is assigned to multiple workers who will provide tags describing the contents of the image. The final result is obtained by choosing a subset of tags from all the collected tags through voting [40] or other inference methods [41].

In crowdsourcing, there is an important category of tasks that are composed of a set of contextually correlated subtasks, e.g., handwriting recognition [7], translation [47], route planning [45], and audio transcription [34]. In this work, we call such crowdsourcing tasks context-sensitive tasks (i.e., CSTs). For instance, Fig. 1(a) shows an example for *handwriting recog-*

* Corresponding author.
E-mail addresses: fangyili@act.buaa.edu.cn (Y. Fang), sunhl@buaa.edu.cn (H. Sun), liguoliang@tsinghua.edu.cn (G. Li), zhangrc@act.buaa.edu.cn (R. Zhang), huaijp@buaa.edu.cn (J. Huai).

https://doi.org/10.1016/j.ins.2018.05.050
0020-0255/



(a)

$$\begin{bmatrix} \text{You} & \text{misspelled} & \perp & \text{words,} & \text{Plan} & \text{spellcheck} & \perp & \text{work} & \perp & \text{tip.} \\ \text{You} & \text{misspelled} & \text{several} & \text{words,} & \text{Please} & \perp & \text{your} & \perp & \text{next} & \text{time.} \\ \perp & \text{misspelled} & \text{several} & \text{works,} & \text{Play} & \perp & \text{your} & \text{work} & \perp & \text{time.} \end{bmatrix}$$

(b)

Fig. 1. A CST task; (a) a handwriting recognition task and (b) answers provided by three workers.

nition, where workers are asked to recognize a handwritten sentence consisting of multiple words. The words in a sentence are closely correlated within the same semantic context determined by the corresponding sentence. Thus *handwriting recognition* is a typical kind of CSTs. However, as for result inference, existing methods are not effective for context-sensitive tasks (CSTs). On the one hand, CSTs are rather difficult and each worker cannot correctly answer the whole task (e.g., correctly recognizing all words in a sentence for the handwriting recognition task). Fig. 1(b) illustrates the answers given by three workers, each of which is a fully or partially recognized sentence. And any worker cannot correctly recognize the sentence. Thus for the task-level inference algorithm [18,38,48], which assigns each complete task to different workers and aggregates the answers of each task, it is hard to obtain high-quality answers from individual workers. On the other hand, each task has internal context correlation among its subtasks. For example, the words in a sentence are not independent, and they are contextually correlated with each other. Obviously the context cannot be ignored in result inference. For example, the second answer to words 3–5 is “misspelled several words” while the third answer is “misspelled several works”. Apparently the second answer is more reasonable because “misspelled” is more closely correlated with “several words” than “several works”. In addition, suppose another worker recognizes words 3–5 as “misspelled several works”. If the context correlation is ignored, the voting-based inference may consider “misspelled several works” as the final result, which is apparently wrong. Thus the subtask-level inference algorithm [4,38,44], which splits each task into multiple subtasks and aggregates the answers of each subtask from workers, is not effective either. Thus, neither of the two approaches is suitable for CSTs because they do not utilize the context correlation to jointly consider the recognized answers in result inference.

There are two main challenges to handle CSTs. First, it is non-trivial to capture the context correlation as a CST usually does not contain enough information. This brings a challenge of modeling the context correlation, solving the model with sparse information, and inferring high-quality answers. Second, CSTs are relatively more complex than simple tasks, and one iteration usually cannot obtain high-quality results. To address these two challenges, we study two problems: *Result Inference* and *Iterative Decision*. The former aims at inferring the best result from the answers submitted by workers and the latter mainly checks whether the inference result can be further improved: if the result is good enough, the iterative process will be terminated; otherwise a new iteration will be started.

In this work we propose a context-aware inference method for improving the quality of CSTs. First, we use the hidden Markov model (HMM) [15,33] to depict the context correlation and design context-aware inference algorithms (*Context-Inf*). Particularly, we incorporate external knowledge to address the challenges of the lack of context correlation information. Then, we propose to support iterative improvement of crowdsourcing with *Context-Inf*. To summarize, we make the following contributions:

- We identify an important category of crowdsourcing tasks, namely context-sensitive tasks (CSTs). To the best of our knowledge, this is the first attempt to study context-aware inference.
- We establish a probabilistic model to describe the crowdsourcing process of CSTs and propose *Context-Inf* to infer results by incorporating HMM into MLE and EM algorithms with external knowledge.
- For complex CSTs, we devise an iterative decision model based on the POMDP to improve the quality.
- We have conducted an extensive set of experiments with two types of CSTs: *handwriting recognition* and *audio transcription*. Experimental results showed the superiority of our approach.

The remainder of this paper is organized as follows. We formalize our problem in Section 2 and present an overview of our approach in Section 3. In Section 4, we respectively describe our probability model and inference algorithms. Section 5 gives our iterative decision model with POMDP and corresponding algorithms. Experimental setup and results are described in Section 6. Section 7 discusses related work and we conclude this work in Section 8.