

Proposal of Grade Training Method in Private Crowdsourcing System

Masayuki Ashikawa and Takahiro Kawamura

Corporate Research and Development Center
Toshiba Corporation
Kanagawa 212-8582, Japan

Akihiko Ohsuga

Graduate School of Information Systems
The University of Electro-Communications
Tokyo 182-8585, Japan

Abstract

Current crowdsourcing platforms such as Amazon Mechanical Turk provide an attractive solution for processing of high-volume tasks at low cost. However, problems of quality control remain a major concern. We developed a private crowdsourcing system (PCSS) running in a intranetwork, that allow us to devise for quality control methods. In the present work, we designed a novel task allocation method to improve accuracy of task results in PCSS. PCSS analyzed relations between tasks from workers' behavior using Bayesian network, then created learning tasks according to analyzed relations. PCSS increased quality of task results by allocating learning tasks to workers before processing difficult tasks. PCSS created 8 learning tasks automatically for 2 target task categories and increased accuracy of task results by 10.77 point on average. We found that creating learning tasks according to analyzed relations is a practical method to improve the quality of workers.

Introduction

The numbers of workers who process crowdsourcing tasks are increasing in line with the expansion of the domains in which crowdsourcing is used. Therefore, the way in which work is performed in crowdsourcing is expected to become common practice. However, support for crowdsourcing workers, such as education and improvement of the working environment, is insufficient. This problem is thought to be due to crowdsourcing workers being many and unspecified. Crowdsourcing workers are employed and disemployed easily since they are unspecified. This poor management of workers could lead to declining quality of task results and unjustified termination of workers.

Therefore, the importance of education for crowdsourcing workers is expected to increase to a level comparable with that of education in the traditional working environment.

We proposed a grade-based training method for application in the education of microtask crowdsourcing workers. In this method, workers process appropriate training tasks before processing difficult tasks. Workers' skill is upgraded by processing training tasks. Specifically, our system creates appropriate training tasks by analyzing the relation be-

tween many tasks and workers' behavior using a Bayesian network.

The Grade Training Method in PCSS

Education is based on the assumption that experience in certain tasks has a positive effect on people's behavior in the same or similar tasks. Therefore, in order to obtain good results in any situation, people should have experience in the same or similar situations. The efficacy of the grade-based training method is demonstrated in school education. Therefore, we propose a method of upgrading workers' skill with training tasks allocated in stages. However, this grade-based training method has a cost problem. It is difficult to create the subjects and the curriculum automatically in a crowdsourcing system, since the purposes of tasks vary and there are many unspecified workers. The method that we propose creates training tasks automatically by reusing existing tasks.

Automatic Task Classification

In order to create training tasks by reusing existing tasks, it is necessary to analyze the relation between task and task contents. However, analyzing each task is costly in terms of both time and calculation, because PCSS(Ashikawa 14) has many tasks. Therefore, PCSS classifies task contents to task category groups automatically to reduce cost.

The tasks in PCSS have titles and descriptions written by requesters. PCSS calculates the tf-idf of each task using keywords extracted from the title and description using morphological analysis.

PCSS calculates similarity between tasks using tf-idf of each words and cosine similarity. PCSS calculated 777,924 patterns of the cosine similarity between 882 tasks. Consequently, PCSS classifies tasks with cosine similarity of 0.4 or more to the same category. As a result, PCSS classifies 882 tasks to 50 task categories.

Analyzing Relation Between Task Categories

An intelligent tutoring system (ITS) is a computer system that aims to provide immediate and customized instruction or feedback to learners usually without intervention from a human teacher. ITS is very useful for a grade-based training method(Ueno 2000). Methods for representing learner models in ITS using a predicate logical representation have been

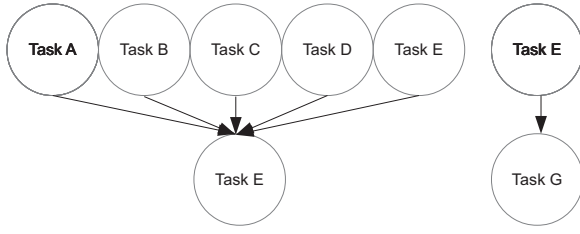


Figure 1: Directed acyclic graph for target task categories

popular for many years. However, a predicate logical representation has some problems, such as difficulty in handling exceptions to rules and inconsistent workers' behavior. For example, workers' careless mistakes and guesswork. However, these cases often happen in microtask crowdsourcing. A stochastic method, such as a Bayesian network can treat such exceptions to rules and inconsistent workers' behavior. A Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies with a directed acyclic graph. PCSS analyzes relations between task categories analyzed in Section 4.1 using a Bayesian network.

$P(A | B)$ means a probability that a worker who can process task B with high accuracy can process task A. $P(A | B)$ can be defined as;

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad (1)$$

, where $P(A)$ is the probability that the worker can process task A with more than 90% accuracy, and then it is regarded as true. If $P(A | B)$ shows a high probability, PCSS can create training tasks for task A by reusing task B.

Discussion

In order to use a Bayesian network in crowdsourcing, PCSS calculates the average accuracy of task results in each task category in Section 4.1. The Bayesian network learns from the calculated average accuracy and creates a directed acyclic graph. In the present work, we use Weka (Waikato Environment for Knowledge Analysis) 3.6.11 (Weka) in order to apply the Bayesian network in crowdsourcing.

As a result, we were able to create directed acyclic graphs for 'determine if two web pages have similar content, or not' and 'cutting a sentence by paragraph'. The created directed acyclic graphs are shown in Figure 1, where Task A is 'select correct accent data from the following options', Task B is 'spell checking', Task C is 'determine if two synthesized speeches have similar sound, or not', Task D is 'put phonetics data to words', Task E is 'checking if phonetics data of word is correct', Task F is 'put accent data and phonetics data', Task G is 'cutting a sentence by paragraph'.

The created directed acyclic graphs show the relations between target tasks and training tasks. For example, in order to find a training task for Task E, PCSS calculates the probability $P(A | B)$ in Eq.1. A is target task Task E and B is the training task for A. $P(A | B)$ is the probability that a worker

who can process task B with high accuracy can process task A with high accuracy.

In order to check the effectiveness of the directed acyclic graphs, we tested as follows:

1. Allocate a target task for which upgrading accuracy of the task result is required to all workers to check the accuracy first.
2. Classify workers who processed the target task to worker group 1 to process training tasks and to worker group 2 to process the same tasks as in 1.
 - (a) Worker group 1 process training tasks.
 - (b) Worker group 2 process the same tasks as in 1.
3. Allocate the same tasks in 1 to workers who processed in 2, and compare accuracy of the task result in 1 and accuracy of the task result in 3. The workers do not perform other tasks between each step.

we tested 'determine if two web pages have similar content, or not' with 5 training tasks and 'cutting a sentence by paragraph' with 1 training task. These training tasks improve 10.77 point on accuracy of target tasks on average.

Conclusion And Future Work

In this paper, we propose a grade-based training method for PCSS. PCSS creates training tasks automatically by reusing other tasks. Workers' skill is upgraded for the target task by processing a training task before processing the target task.

Additionally, in order to check the effectiveness of grade-based training, we created 2 directed acyclic graphs for 2 task categories, and then 8 training tasks from the directed acyclic graphs. These training tasks improve 10.77 point on accuracy of target tasks on average.

On the other hand, there are some task categories for which there are no training tasks created by PCSS. It is speculated that PCSS could not create training tasks for these task categories because of a shortage of workers' behavior data. Development of a training task method that requires little workers' behavior data is a subject for future work.

References

- Ashikawa, M., Kawamura, T., Ohsuga, A., "Speech Synthesis Data Collection for Visually Impaired Person", Human Computation & Crowdsourcing, (2014)
- Ueno, M., "Intelligent tutoring system based on belief networks", pp141-142, (2000).
- Weka (Waikato Environment for Knowledge Analysis), <http://www.cs.waikato.ac.nz/ml/weka/>