Citizen + X: Volunteer-Based Crowdsourcing in Science, Public Health, and Government:
Papers from the 2014 HCOMP Workshop

Speech Synthesis Data Collection for Visually Impaired Person

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

Crowdsourcing platforms provide attractive solutions for collecting speech synthesis data for visually impaired person. However, quality control problems remain because of low-quality volunteer workers. In this paper, we propose the design of a crowdsourcing system that allows us to devise quality control methods. We introduce four worker selection methods; preprocessing filtering, real-time filtering, post-processing filtering, and guess-processing filtering. These methods include a novel approach that utilizes a collaborative filtering technique in addition to a basic approach involving initial training or use of gold-standard data. These quality control methods improved the quality of collected speech synthesis data. Moreover, we have already collected 140,000 Japanese words from 500 million web data for speech synthesis data.

Introduction

Text-to-speech(TTS) synthesis is preferable for many visually impaired person to support their work. However, acquiring a lot of speech synthesis data, such as reading data, class data and accent data to improve the TTS synthesis is quite costly. Therefore, volunteer-based crowdsourcing can be considered as a solution for the problem. However, volunteer-based crowdsourcing services (VCS) suffer from undesirable (volunteer) worker practices, including low-quality contributions. In this paper, we propose quality control methods for the TTS synthesis.

There are many proposals of quality control methods for crowdsourcing in the literature. The most popular quality control method is change of task design, such as lowering of task difficulty and simplification of task description. However, controlling all task design by the administrator is impossible, because task design depends on the requester's purpose. Therefore, the worker selection method is adopted in VCS. Proposals of quality control methods for workers in the literature include: excluding spam workers on the basis of learning data from task results (Halpin 12); making conjectures on worker's skills from worker's behaviour in unrelated tasks (Kilian 12); excluding spam workers and poorly qualified workers by ranking workers (Raykar 11); and excluding spam workers and poorly qualified workers by calculating the threshold between highly qualified workers and poorly qualified ones (Donmez et al., 2009). These proposals of quality control methods are mainly excluding low-quality workers. However, decreasing workers lowers performance in crowdsourcing system.

VCS

VCS is shown in Figure 1. VCS has been in operation since 2011. The total number of workers is 1,568, and 5.7 million tasks have been processed.We have collected 140,000 Japanese words from 500 million web pages for speech synthesis data. The words are extracted from web text by using morphological analysis. In order to use these words as speech synthesis data, the noise rejection task, read data task, class data task, and accent data task are designed and processed in VCS.

Quality Control Mechanism

For quality control, VCS implements four worker selection methods: 1) preprocessing filtering, 2) real-time filtering, 3) post-processing filtering, and 4) guess-processing filtering. These filtering methods are applied in descending order of the number of workers.

Preprocessing Filtering

This filtering method is applied in the recruitment of workers. A questionnaire survey is conducted and, based on the results, workers who are clearly of low quality are excluded. The questionnaire ascertains 'time available for work', 'ambitions concerning work', 'educational background', and 'ICT experience'. Through this filtering, 1,630 workers were recruited.

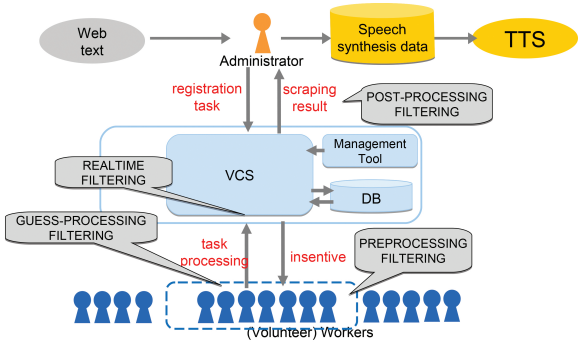

Figure 1: Volunteer-based crowdsourcing service (VCS)

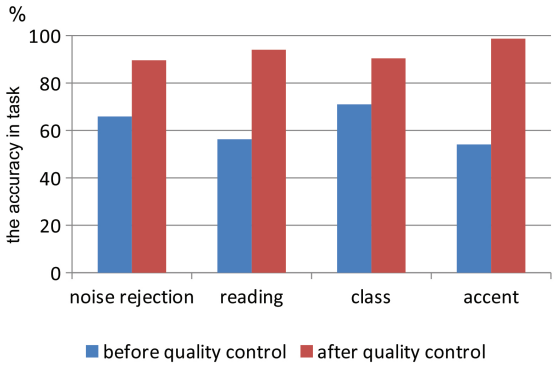

Figure 2: Result of data accuracy

Real-Time Filtering

This filtering method is applied in processing tasks performed by workers. VCS calculates the accuracy rate of all task as follows: the number of correct task results / the number of tasks processed per workers'. VCS excludes any worker whose accuracy rate for all tasks is below 70%. This filtering method excluded 62 workers considered to be spam workers. In addition, VCS calculates the accuracy rate for each task category as follows: the number of correct task results for each task category/ the number of tasks processed per task category per workers'. VCS does not allocate tasks to any worker, whose task category accuracy rate is below 70% for the task category in question.

Post-Processing Filtering

Workers tend to have strong task categories and weak task categories. VCS analyses a worker's strong task categories and weak task categories based on the results of the tasks that the worker performed, and controls task allocation accordingly. In addition, a requester can analyse worker's strong task categories and weak task categories from the task results. The requester collects workers whose strengths correspond to the requester's task categories. This information is registered in VCS and shared by all requesters as 'Skill'. Examples are Determination of word class from word (31 workers), Determination of correct or incorrect accent of word (207 workers), Selection of correct accent from candidates (142 workers), Writing of correct accent from word (53 workers), and Low quality for accent task (242 workers).

Guess-Processing Filtering

This filtering method controls the allocation of tasks to workers based on the degree of similarity between workers. Specifically, in the case that worker A is similar to worker B, it is assumed that worker A's strong task categories and weak task categories are similar to worker B's. Similarity score between workers is calculated with the pearson correlation coefficient. Similarity $S_{u,v}$ between worker u and worker v can be written as

$$S_{u,v} = \frac{\sum_{i \in T}(r_{u,i} - \bar{r_u})(r_{v,i} - \bar{r_v})}{\sqrt{\sum_{u \in W}(r_{u,i} - \bar{r_u})^2}\sqrt{\sum_{v \in W}(r_{v,i} - \bar{r_v})^2}} \quad (1)$$

where W is the set of workers, u, v are the elements of W, T is the set of task categories, i, j are the elements of T, $r_{u,i}$ is the accuracy of task i result by worker u, and $\bar{r_u}$ is the average accuracy of all tasks by worker u. The conjectured accuracy $P_{u,i}$ that is the estimated accuracy of task i not processed by worker u can be written as

$$P_{u,i} = \bar{r_u} + \frac{\sum_{v \in W}(r_{v,i} - \bar{r_v})S_{u,v}}{\sum_{v \in W} \mid S_{u,v} \mid} \quad (2)$$

To verify the validity of the calculated conjectured accuracy, we compared the conjectured accuracy and the actual accuracy.The number of workers, whose conjectured accuracy in the task of adding reading data (the pronunciation of kanji) is above 90%, is 219, whereas the number of workers whose the actual accuracy in the same task is above 90%, is 194. The number of workers, whose conjectured accuracy in the task of adding word class is above 90%, is 23, and the number of workers, whose the actual accuracy in the same task is above 90%, is also 23. The number of workers, whose conjectured accuracy in the task of adding word accent is above 90%, is 138, whereas the number of workers whose the actual accuracy in the same task is above 90%, is 121.

As the results, we found that the guess-processing filtering is useful for extracting high-quality volunteer worker without a test task.

Evaluation System

To verify the quality control methods employed in VCS, we compared accuracy in each task without the quality control and accuracy in each task with the quality control. The results are shown in Figure 2. This figure indicates that the quality control methods improved the accuracy of collected speech synthesis data.

Conclusion

VCS has collected a large amount of speech synthesis data for visually impaired person. Additionally, the worker control method in VCS improves the quality of data to a level suitable for speech synthesis. A future work is the further

improvement in the quality of collected speech synthesis data based on feedbacks from the impaired person.

References

Donmez, P. et al., Efficiently learning the accuracy of labeling sources for selective sampling, In proc. of KDD, 2009.

Halpin, H., Blanco, R.,"Machine-Learning for Spammer Detection in Crowd-Sourcing", HCOMP, (2012)

Kilian, N., Krause, M., Runge, N., Smeddinck, J., "Predicting Crowd-Based Translation Quality with Language-Independent Feature Vectors", HCOMP, (2012)

Raykar, V., Yu, S., "Ranking annotators for crowdsourced labeling tasks", NIPS, (2011).