Building Human Computation Space on the WWW: Labeling Web Contents through Web Browsers

Yoshinari Shirai[1], **Yasue Kishino**[1], **Yutaka Yanagisawa**[1], **Shin Mizutani**[1], **Takayuki Suyama**[1]
[1]NTT Communication Science laboratories, NTT Corporation
2-4, Hikaridai, Seika-cho, Soraku-gun Kyoto, Japan
yoshinari.shirai.wm@hco.ntt.co.jp

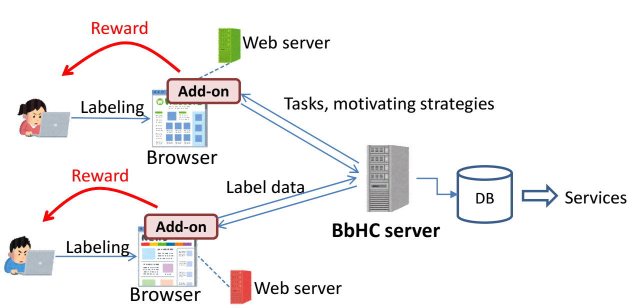

Figure 1: Overview of Browser-based Human Computation

Abstract

Although contents on the WWW are potentially valuable data to achieve automated services, they are difficult to use in their current state. Our proposed Browser-based Human Computation (BbHC) offers a cost-effective way to extract desirable data from web contents. BbHC enables people to label various web contents through the web browsers they normally use for web browsing. We implemented three systems based on the BbHC to explain how BbHC works.

Introduction

The entire WWW, rather than a specific web page or website, is an ideal place for human computation (Sabou et al. 2018) because huge amounts of potentially useful data for various services are buried within it. While such data are already digitalized, services sometimes require not only the data but also appropriate labels for these data. Consequently, we propose a novel human computation system, called Browser-based Human Computation (BbHC), to enable users to label data directly through common web browsers.

In BbHC, browser extensions (add-ons) provide the computation interfaces for labeling tasks on a given web page. Task results are sent to the BbHC server, and the interfaces are updated based on the server's responses (Fig. 1). This mechanism enables users to immediately start working on a labeling tasks as soon as they find web contents that are appropriate for that labeling task while browsing the web for other purposes. To accelerate the labeling of data without the inducement of monetary rewards, extensions that work with servers motivate people to continuously engage in tasks through various human computation techniques (Quinn and Bederson 2011). BbHC server evaluates the results of each human computation task to maintain high standards of label quality.

Presented in the Work in Progress and Demo track, HCOMP 2019.

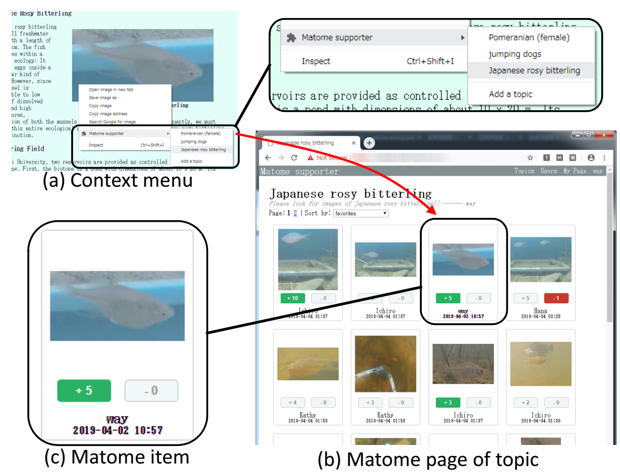

Figure 2: Screenshot of matome supporter

Examples of BbHC system

Here, we show three BbHC systems. We developed chrome extensions to perform labeling functions and implemented a BbHC server (Node.js) with a database (PostgreSQL).

Matome supporter

Matome supporter enables users to build web pages that show a collection of images distributed on various existing web pages. Although it is difficult to manually build these photo album-like web pages, which we call matome pages, matome supporter allows users to easily and coop-

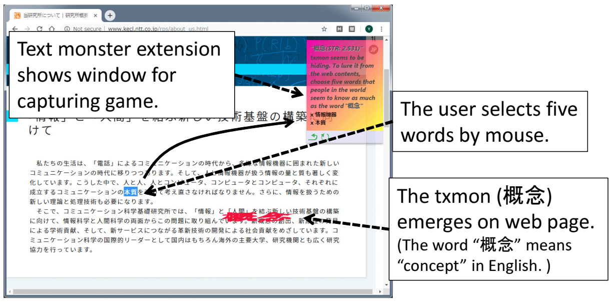

Figure 3: Screenshot of capturing game in text monster

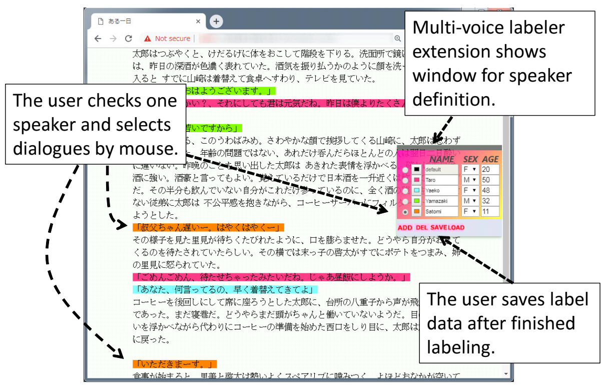

Figure 4: Screenshot of multi-voice labeler

eratively build such a page for each topic. When users of matome supporter find suitable images of a topic on the WWW, they simply select the relevant topic name shown in the right-click context menu (Fig. 2(a)). Topics represented by matome pages can be freely registered by matome supporter users and shared among all users. Accordingly, matome pages can be cooperatively created. Figure 2(b) shows a matome page for "Japanese rosy bitterling," generated by the BbHC server using image URLs registered by users. The user can evaluate each image's accuracy using the + / - buttons shown with each image (Fig. 2(c)) to sort thumbnail images appropriately.

Text monster

Text monster is a game that enables users to scramble among websites using Japanese words. The game personifies Japanese words as monsters, or "txmons." While the basic concept of the game is similar to Pokémon Go(Colley et al. 2017), which is famous as a location-based game (Avouris and Yiannoutsou 2012), text monster evolves and progresses via the player's location on the WWW (i.e. URLs). Users can collect txmons that are hiding in specific web contents, place ('pasture') txmons in the websites they will occupy, and battle with other users' txmons that have been placed in websites in order to rob the other users of those websites. Figure 3 shows how to capture txmons hiding in the web page contents. Users need to respond to these types of instructions to capture txmons:

The "concept" txmon seems to be hiding. To lure it from the web contents, choose five words that people in the world seem to know as much as the word "concept".

In this example, the user can successfully capture the "concept" txmon, if the word familiarity values(Amano, Kondo, and Kato 1999; Amano, Kasahara, and Kondo 2007) of the five words selected by the user are near the word familiarity value of the word "concept."

Multi-voice labeler

Today, we can easily listen to web contents like novels on the WWW using text-to-speech functions on smartphones. However, this is still unnatural, especially when the web contents include conversations. The current text-to-speech functions on smartphones normally continue to read using a default voice, even in a conversation between a young man and an older woman. Multi-voice labeler enables users to annotate speaker information to web contents so that smartphones can appropriately read them using multiple voices. Figure 4 shows the labeling operations. First, the user defines speakers by name, gender, and age on the window displayed by the chrome extension. Then she chooses a speaker and selects sentences that the speaker seems to have spoken. Finally, she presses the save button to send all of the label data (URL, speaker data, and sentences of each speaker) to the BbHC server. A multi-voice app, which we also developed for smartphones, can read web contents with multiple voices using the label data. When reproducing each sentence in the web contents, the app selects the voice with the closest gender and age to the speaker indicated by the speaker data.

Human Computation

These systems are designed to collect labeled data for the following purposes of human computation. Matome supporter collects labeled images to create an image classifier for relatively uncommon categories (e.g. Japanese rosy bittering). Text monster annotates word familiarity values to words that have not yet been scored using the results of a capturing game for updating a word familiarity database (Amano and Kondo 1998). Multi-voice labeler's purpose is to collect writings with speaker information for natural language processing research (He, Barbosa, and Kondrak 2013). Each system motivates users to engage themselves in labeling activities by offering the benefit of using extensions. We have also implemented several quality control techniques for use in evaluating the labeling results.

When requesters request labeling tasks using common crowdsourcing platforms, they have to consider how to prepare samples (e.g. images) which workers label. In our approach, requesters don't need to prepare samples because all contents on the WWW are candidates of samples and workers look for samples appropriate for each labeling purpose.

Conclusion

This paper proposed the Browser-based Human Computation (BbHC) and described three systems based on it. Future work will evaluate the effectiveness of the quality control and incentive supply in our systems and implement privacy protection mechanisms to offer our systems to the public.

References

Amano, S., and Kondo, T. 1998. Estimation of mental lexicon size with word familiarity database. In *The 5th International Conference on Spoken Language Processing*.

Amano, S.; Kasahara, K.; and Kondo, T. 2007. Reliability of familiarity rating of ordinary japanese words for different years and places. *Behavior Research Methods* 39(4):1008–1011.

Amano, S.; Kondo, T.; and Kato, K. 1999. Familiarity effect on spoken word recognition in japanese. In *Proceedings of the 14th ICPhS*, volume 2, 873–876.

Avouris, N. M., and Yiannoutsou, N. 2012. A review of mobile location-based games for learning across physical and virtual spaces. *J. UCS* 18:2120–2142.

Colley, A.; Thebault-Spieker, J.; Lin, A. Y.; Degraen, D.; Fischman, B.; Häkkilä, J.; Kuehl, K.; Nisi, V.; Nunes, N. J.; Wenig, N.; Wenig, D.; Hecht, B.; and Schöning, J. 2017. The geography of pokémon go: Beneficial and problematic effects on places and movement. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, 1179–1192.

He, H.; Barbosa, D.; and Kondrak, G. 2013. Identification of speakers in novels. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, volume 1, 1312–1320.

Quinn, A. J., and Bederson, B. B. 2011. Human computation: A survey and taxonomy of a growing field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, 1403–1412.

Sabou, M.; Aroyo, L.; Bontcheva, K.; Bozzon, A.; and Qarout, R. K. 2018. Semantic web and human computation: The status of an emerging field. *Semantic Web* 9:291–302.