

A proposed system for machine translation by crowdsourcing

Kantaro Mera
Graduate School of Engineering
Doshisha University
Kyoto, Japan

Taichi Sugimura
School of Engineering
Doshisha University
Kyoto, Japan

Takahiro Koita
School of Engineering
Doshisha University
Kyoto, Japan

Abstract—To enjoy video content, the hearing impaired and viewers of different languages need video captioning. Existing automatic captioning systems have already been employed on commercial platforms such as YouTube. These systems generate captions from the audio data of the videos. The voice data handled by existing systems is a synthesis of multiple sounds, making it difficult to caption. In addition, the insufficient accuracy of machine-translated captions cannot correctly express the meaning of the original conversation. In this paper, we propose a new captioning system using crowdsourcing to improve the accuracy of foreign language captions in conversations with multiple people. The proposed system includes a step to create machine translation adaptive sentences to increase accuracy. We compared and evaluated the accuracy of captions created by the proposed system and YouTube's existing system. In this paper, we demonstrate that our proposed system has higher accuracy than the existing system.

Keywords—crowdsourcing, machine translation, automatic captioning system, multilingual communication

I. INTRODUCTION

In recent years, the spread of smartphones has increased the opportunity to enjoy videos. We not only watch, but also create, share, and comment. Captions are indispensable for the hearing impaired or viewers of different languages to enjoy these videos. While it appears that all smartphone users are enjoying videos, not all videos have captions. In fact, there are few videos with foreign language captions. Currently, the accuracy of these captions varies depending on the video content. This large gap in the accuracy of captions occurs because making captions is not easy; it requires skill, time, and a devoted worker. Currently, in order to eliminate the hassle of creating captions for users who create and share videos, video sharing platforms have implemented automatic caption creation systems and a function to translate the created captions into foreign language captions by machine translation. According to the existing research, using a crowdsourcing caption creation method makes it possible to create captions that are about 10 times more accurate than the captions currently made by an automatic caption creation system implemented on a platform [1].

Yoshino proposed using a "machine translation adaptive sentence" to create a machine translated sentence that makes it easier for a monolingual speaker to interpret the machine translated sentence, and conducted an adaptive sentence creation experiment to evaluate its success [2]. A machine translation adaptive sentence is a sentence that has a high degree of agreement with the back-translated sentence. As a result of the experiments, this method shows a higher degree of agreement with the back-translated sentence than the original example for all examples. When Yoshino conducted a parallel translation creation experiment using machine translation adaptive sentences and verified the effect of these

sentences, he was able to create machine translation sentences with an accuracy of 97% [3].

In this paper, we propose a new foreign language subtitling method using crowdsourcing and machine translation adaptive sentences, and compare the conventional method with the proposed method.

II. PROPOSED SYSTEM

The proposed system is composed of four steps. The difference between this system and YouTube's existing system is that this system uses crowdsourcing for computers to create the best translation. It is difficult for a computer to create captions from uploaded videos and to understand dialects. Computers cannot distinguish the sounds of speakers, but people can. Computers are also only programmed with general languages, so the dialects of specialized speakers are easily misunderstood. However, this is a simple task for people to accomplish. For these reasons, our system uses crowdsourcing to create captions. Following bulleted points show steps of proposed system in Figure 1.

- (1) selecting the video
- (2) making captions by crowdsourcing
- (3) making machine translation adaptive sentences by crowdsourcing
- (4) translating by a machine translation system

III. EVALUATION

In this paper, we evaluate whether the proposed system is more useful than existing systems. As our existing system, we used the system implemented on YouTube for creating English captions. We used the English captions of a Japanese comedy video created via three methods. This video consists of 179 sentences. In the first method, we asked 23 subjects to make captions and asked 14 subjects to translate using CrowdWorks (CW). CW is a Japanese crowdsourcing platform. The captions were made in the same way as in [1]. The cost of making captions was \$0.55 and the cost of translating was \$9. In the second method, we asked 23 subjects to create captions in CW and translate using Google Translate. The third method is our proposed system; we asked 23 subjects to make captions and 14 subjects to make machine translation adaptive sentences in CW. The price of making machine translation adaptive sentences was \$7. We then translated using Google Translate. The steps for making the captions was the same in all methods. After making the English captions, we evaluated them line by line. The rates of

the number of correct sentences is defined as the accuracy. We then compared the accuracy.

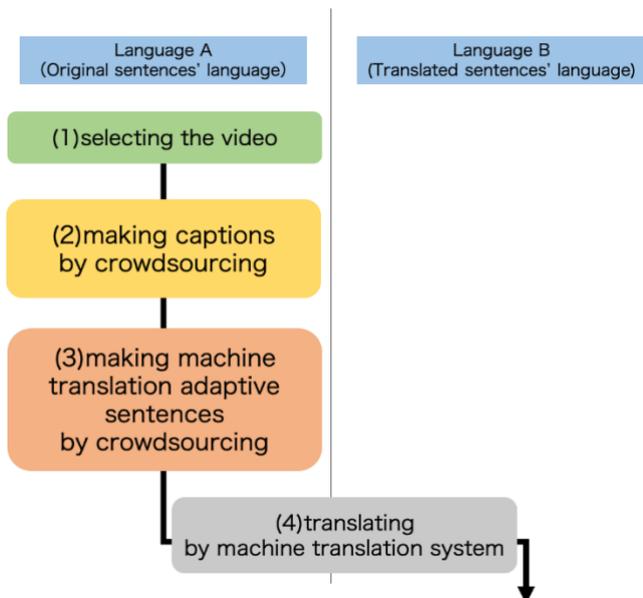


Figure 1: Overview of the proposed system

IV. RESULTS

Table 1 shows the classification of the number of correct sentences. For the first method, the number of correct sentences is 158 and the accuracy is 88.4%. In the second method, the number of correct sentences is 108 and the accuracy is 60.3%. For our proposed, third method, the number of correct sentences is 140 and the accuracy is 78.2%.

Table 1: Classification of the number of correct sentences

	# of correct sentences	Accuracy (%)
First	158	88.4
Second	108	60.3
Third	140	78.2

V. DISCUSSION

We discuss our results from two viewpoints: Accuracy and utility.

The third method's accuracy is 88.4%. Accuracy was not 100% because the subjects could not always understand the speakers' dialects, backchannels, and English words that correspond to various expressions. As a result, they could not make accurate captions and machine translation adaptive sentences.

Next, we approach the results from two sides to verify the utility of the proposed system: Accuracy and cost. Table 2 shows the result of this multilateral consideration. We used the accuracy of English captions in YouTube as the existing method. The third method demonstrates significant accuracy

improvement over the existing method (YouTube). However, if we consider the result from the view of both accuracy and cost, we cannot say that the proposed system is always useful. The third method's accuracy is about only 1.3 times higher than the second method; in contrast, the third method's cost is about 13.7 times higher than the second.

Table 2: Result of multilateral consideration

	Accuracy (%)	Cost (\$)
First	88.4	16
Second	60.3	0.55
Third	78.2	7.55
YouTube	0	0

VI. CONCLUSION

In this paper, we proposed a new system for creating English captions. The proposed system demonstrated a 78.2% improvement in the accuracy of the English captions, but, due to its cost, we cannot say that the proposed system is always useful.

In our future work, we will study new approaches to dialects, backchannels, and English words that correspond to various expressions in Japanese to increase the accuracy of the proposed system and improve the cost performance, by taking more effective way to assign crowdworkers.

REFERENCES

- [1] Kantaro Mera, Takahiro Koita, The Method for Making Captions by Crowdsourcing, IEICE, 2020
- [2] Takashi Yoshino, Proposal of Using Back Translation for a Method of Bilingual Translation by a Monolingual Speaker on Crowdsourcing [translated from Japanese], GN Workshop 2015
- [3] Takashi Yoshino, Proposal of Parallel Texts Creation Method Using Multiple Machine-translation by Monolingual on Crowdsourcing Workers, Journal of Information Processing Society of Japan, 2018