氏　　名	TRIEU HAI LONG			
学位の種類	博士(情報科学)			
学位記番号	博情第 372 号			
学位授与年月日	平成 29年 9 月 22 日			
論文題目	A Study on Machine Translation for Low-Resource Languages （低いリソースの言語のための機械翻訳についての研究）			
論文審査委員	主査	Nguyen Minh Le	北陸先端科学技術大学院大学	准教授
		東条　敏	同	教授
		飯田　弘之	同	教授
		白井　清昭	同	准教授
		Steven Gordon	CQU	上級講師
		Ashwin Itoo	Liege University	准教授

論文の内容の要旨

Current state-of-the-art machine translation methods are neural machine translation and statistical machine translation, which based on translated texts (bilingual corpora) to learn translation rules automatically. Nevertheless, large bilingual corpora are unavailable for most languages in the world, called low-resource languages, that cause a bottleneck for machine translation (MT). Therefore, improving MT on low-resource languages becomes one of the essential tasks in MT currently.

In this dissertation, I present my proposed methods to improve MT on low-resource languages by two strategies: building bilingual corpora to enlarge training data for MT systems and exploiting existing bilingual corpora by using pivot methods. For the first strategy, I proposed a method to improve sentence alignment based on word similarity learnt from monolingual data to build bilingual corpora. Then, a multilingual parallel corpus was built using the proposed method to improve MT on several Southeast Asian low-resource languages. Experimental results showed the effectiveness of the proposed alignment method to improve sentence alignment and the contribution of the extracted corpus to improve MT performance. For the second strategy, I proposed two methods based on semantic similarity and using grammatical and morphological knowledge to improve conventional pivot methods, which generate source-target phrase translation using pivot language(s) as the bridge from source-pivot and pivot-target bilingual corpora. I conducted experiments on low-resource language pairs such as the translation from Japanese, Malay, Indonesian, and Filipino to Vietnamese and achieved promising results and improvement. Additionally, a hybrid model was introduced that combines the two strategies to further exploit additional data to improve MT performance. Experiments were conducted on several language pairs: Japanese-Vietnamese, Indonesian-Vietnamese, Malay-Vietnamese, and Turkish-English, and achieved a significant improvement. In addition, I utilized and investigated neural machine translation (NMT), the state-of-the-art method in machine translation that has been proposed currently,

for low-resource languages. I compared NMT with phrase-based methods on low-resource settings, and investigated how the low-resource data affects the two methods. The results are useful for further development of NMT on low-resource languages. I conclude with how my work contributes to current MT research especially for low-resource languages and enhances the development of MT on such languages in the future.

Keywords: machine translation, phrase-based machine translation, neural-based machine translation, low-resource languages, bilingual corpora, pivot translation, sentence alignment

論文審査の結果の要旨

This thesis presents a study for dealing with machine translation on low-resource language in which statistical machine translation and neural translation models are explored. The candidate has also proposed a method to improve sentence alignment based on word similarity learned from monolingual data in building bilingual corpora. A semantic similarity is combined with the use of linguistic issues (such as syntactic parsing and morphological knowledge) for enhancing the accuracy of conventional pivot machine translation methods. Experimental results show that the proposed method is effectively working on low-resource languages including Japanese and Vietnamese language. Another major contribution of the candidate is that he automatically creates a large scale of parallel corporal for Southeast Asian languages from Wikipedia. This resource is used for a study of machine translation. The presentation of the thesis is also nice. The publication of the thesis consisting of two journals and seven international conferences.

In conclusion, the thesis proposes a novel study for machine translation with low-resource language. All committee members have agreed that the work is appropriate for granting the candidate a doctoral degree.

We would like to conclude that this is an excellent dissertation and we approve awarding a doctoral degree to Mr. Trieu Hai Long.