VIETNAMESE NATURAL LANGUAGE PROCESSING FOR INTERACTION BETWEEN HUMAN AND ROBOT

Luong Huu Thanh Nam, Nguyen Dao Xuan Hai,
Tuong Phuoc Tho, Nguyen Truong Thinh
Ho Chi Minh University of Technology and Education, Vietnam

Received 8/10/2018, Peer-reviewed 15/10/2018, Accepted for publication 2/11/2018

ABSTRACT

Service robot has been designed and developed for different objectives and requirements. Since the robotic revolution, Vietnam is one of the most influenced among South-east Asia countries in developing artificial intelligence. People believe that robot will replace all blue-collars in almost company, even hospital and school in the next few decades, therefore developers are trying to improve natural language interaction between human and robot, especially helping new vision for Vietnamese robotics. We have built a robot application that capable of understanding Vietnamese natural language, there are four tasks which were mentioned solving the problems, even more AI method. Deep learning, on the other hand, is a sub-field of machine learning. In this paper, with these algorithms, artificial intelligent supported is much complex, people can communicate naturally with robot, not only English but also Vietnamese as well. In this paper, we will introduce the specification and intelligent interaction processing in naturally Vietnamese.

Keywords: *Service robot; Artificial intelligence; word tokenization; blue-collars; deep learning; machine learning.*

1. INTRODUCTION

The field of service robot has shown its potential in its near future, currently attracting more researchers and technology companies more than ever. Many governments around the world also realize the potential of the field. The Vietnamese government has funded many programs that their main objective is to encourage the use of the robot in an educational environment. It is expected that when children encountering more with robot, they will show curious and grow their interest in the technology field. Not only education benefits from a service robot. In many other places in the world, service robots are also put into operation in boring, repeated task or dangerous environment. Although the topic is controversial, it is certain that more and more service will be in operation in the near future. However, the interaction between human and robot pose some problem. The main issue is that humans hesitate when interacting with robot. From the psychological point of view, the main cause for this is that most people don't feel comfortable interacting with a robot, they feel that most a robot is not friendly toward them. The main key here is, besides its main functionality, service robots also have to be seen as a friendly, human-like being in order to attract their users. This also is the main objective of me and other teams working on the same project. With the use of artificial intelligence, my goal to create a robot application that capable of understanding natural language and further contribute to the development of Vietnamese robotics.

2. ARTIFICIAL INTELLIGENCE

Artificial Intelligence is an area of computer science and the broadest concept of machine learning and deep learning. The aim of artificial intelligence is to infer information based on the signal it's received from the external sources to solve a different problem. Artificial intelligence takes advantage of many mathematics tools for solving a complex problem that is impossible with the conventional algorithm. Since its early time in the 50s, artificial intelligence is

applied in many different aspects of modern life, from autonomous car to recommendation tool on an online shopping website. Machine learning is a method to achieve artificial intelligence. For most of the time, it is about learning a mathematical model of a particular dataset without being explicitly programmed [1]. Machine learning is not about creating a system and that solve any given problem, rather it's about an optimized approach to a specific problem. In machine learning, three type of different machine learning is defined:

Supervised learning is the task of training on a labeled data, consists of set training examples as a pair of input and the desired output value. Supervised learning algorithm then analyzes the training data and produce an inferred function. The above figure is a sample of the VLSP (Vietnamese Language Processing) data set from training a part of speech tagging (which will be explained in the next section). The training data have 30.000 of part of speech labeled natural language in Vietnamese and is generated from Vietnamese newspaper article. Unsupervised learning is contrary to supervised learning, it's the learning task performing on unlabeled training data, therefor there is no accuracy evaluation on this type of learning. Usually, unsupervised learning is use for recognizing data pattern such as clustering.

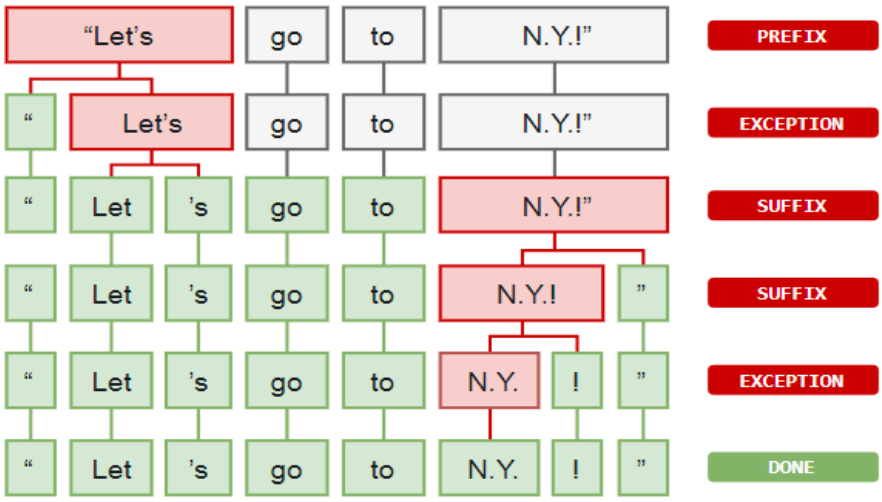

Fig. 1. *The tokenization process of the sample sentence.*

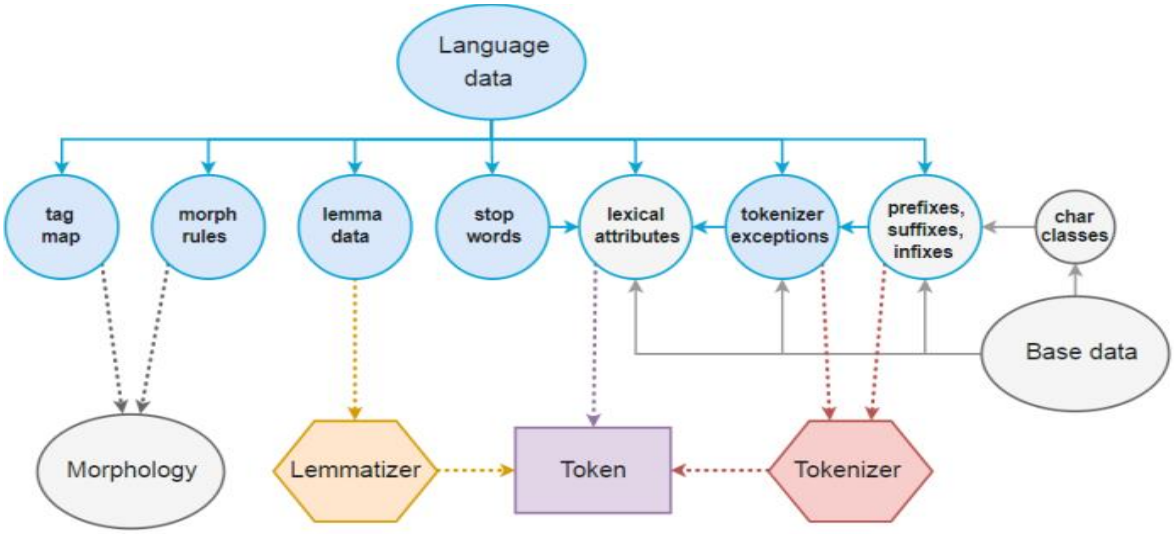

Fig. 2. *Language data for word tokenization*

Deep learning, on the other hand, is a sub-field of machine learning. The field in involving create every large and deep multi-layers of the neural network and must be capable of processing very large data (big data), giving precise prediction according to its learning data. This processing resembles the human learning process. Giving of a set of training data includes many – some time thousands – of input pictures of kittens, adult cats, golden retrievers, dachshunds and raccoons, their features are extracted through many hidden layers. Upon the learning process is done, the system should be able to conclude that input picture whether it is a picture a cat or a raccoon.

3. NATURAL LANGUAGE PROCESSING

Natural language processing (NLP) is a field which consists of the different separate subject such as computer science, artificial intelligence and computational linguistics together with big data algorithm in order to process large natural language corpora (consist for thousands to millions of natural sentence in a specific language) in order to extract useful information from human natural spoken language. The fields range from speech recognition, natural language understanding, natural language generator or dialog system (where computer are capable of performing natural conversation between with the human). In the field of natural language processing, there are four tasks often considered as the fundamental problems which are: Word tokenization; Part-of-speech tagging; Name entity recognition and Text classification [2-4].

3.1 Word tokenization

The electronic text is a linear sequence of characters, words, number and other different symbols. To accomplish satisfied the result from text processing, the electronic text has been segmented into smaller linguistic units such as words, number, punctuation called token. The act of segmenting electronic words is call word tokenization and word segmentation [5]. Token is a language unit which has to be linguistically significant and methodologically useful [6].

The tokenization process requires many linguistic feature and data, and can definitely varies between languages. The diagram below illustrates the English language feature date for English tokenization (**Fig.1**).

3.2 Part-of-speech tagging

In the field of linguistics, part-of-speech tagging, also known as grammatical tagging is the process of marking up a word in a text corresponding to a specific part of speech (noun, verb, adjective …) [7]. A program used for the same task is referred to as Part-of-speech tagger or POS tagger. In natural language processing, the POS tagging process always comes after the tokenization process. In order to correctly tag the input sentences, the tagger use several of particular language information: dictionary, lexicons, rules… Dictionary has a category of a specific word, which means that a word can also be a different part of speech depend on the sentence where it is used in. For instance, the word run and also is a noun and verb. In order to tackle this ambiguity, the tagger uses a statistical model to make a prediction of which tag or label most likely applies in this specific context [8]. This model consists of binary data which generated from the task of supervised learning, where training data consist of many part-of-speech labeled examples.

3.3 Named entity recognition

In the field of information extraction, name entity is a real-world object such as locations (New York City), persons (Bill Gates), organizations (Microsoft) or products, email address, a currency that appear in a sentence. Name entity recognition is the act of recognizing such words in a given sentence. Named entity recognition is essential steps in natural language processing to understand and extract information from text [9]. To achieve the robust performance of this process, many different algorithms are applied such as conditional random field, averaged perceptron algorithm to train on the labeled input data (supervised learning).

4. VIETNAMESE NATURAL LANGUAGE PROCESSING

The Vietnamese language, which also is the official language of the Socialist Republic of Vietnam and is spoken as the mother tongue language by more than 100 million users around the world, by both domestic as well as overseas Vietnamese. It is among the top 20 most spoken language of the world. Back in the ancient times, as a part East Asian (with China, Japan, Korea and Ryukyu), Vietnamese people relied on Hán Tự (漢字 – Chinese script) or Hán Nôm (漢喃 – a Vietnamese invention that was heavily based on the Chinese script). But since the beginning of the 20 centuries, Vietnamese gradually gave the conventional script and switched to Chữ Quốc Ngữ - Latin script introduced by a French missionary Alexandre de Rhodes. Vietnamese speaking language, however, is classified in the Viet Muong group of the Mon-Khmer branch that belongs to the Austro-Asiatic language family. Because of different history and geography factors, Vietnamese vocabulary heavily influenced by Chinese with estimated as much as 60% of the vocabulary derived from Chinese used in formal texts. Furthermore, by being in contact with the French language as well as English, the vocabulary of Vietnamese was enriched not only in vocabulary but also in syntax.

***Table 1**. Structure of Vietnamese Syllable*

TONE MARK		
Initial Consonant	Vowel	Final Consonant

***Table 2.** Distinct tones in the Vietnamese language*

Name	**Contour**	**Diacritic**
Ngang/ bằng	Mid-level	unmarked
Sắc	High rising	´ - acute accent
Huyền	Low failing	` - grave accent
Hỏi	Mid failing	? - hook above
Ngã	Glottalized rising	~ - tide
Nặng	Glottalized falling	. – dot below

***Table 3**. Analysis of some Vietnamese syllable*

Syllable	**Initial Consonant**	**Vowel**	**Final Consonant**
Tuần (week)	T	uâ	n
Trứng (egg)	Tr	ư	Ng
Anh (brother)	None	a	nh
Ngã (fall)	Ng	a	none

Word segmentation is one of the fundamental preprocessing steps for developing any natural language processing in any specific language. This problem is even more essential and challenging in some Asian languages. For instance, Chinese and Japanese languages don't have white spaces to divide each separate word. For the Vietnamese case, a similar problem is noticed since a word can contain more than one white-space separated syllables. Hence, white space is not considered as a word separator in the Vietnamese language. Each Vietnamese syllable is elementary units have only one pronunciation. They are not undivided elements but a structured as described by the following table 1. Vietnamese is a tonal language, each word meaning is depended on the tone in which it is pronounced. There are six distinct tones in the Vietnamese language like Table 2. Furthermore, in the Vietnamese language, the main vowel of each syllable is required but not for other parts of the word. Initial consonant, final consonant as well as main vowel can either be a single or double consonant. This is illustrated by examples of some common Vietnamese syllable in this table. Vietnamese words, on the other hands, contain one or more syllable. Statically, 80% of Vietnamese words consist of two syllables (disyllabic). Between each syllable as well as each word is a white space separator, this makes Vietnamese tokenization just as

challenging as Chinese and Japanese tokenization. Because of the necessity of word segmentation, many Vietnamese computer science individuals and teams have release different tool for this same purpose of tokenizing Vietnamese words, each one has its own approach and most of them perform with excellent results (the F1 score of all approaches hits more the 90%) when trained on the same tokenized input data set (the VLSP Corpus).

As can be seen from the table above, tools for Vietnamese tokenization are mostly written in python, which can pose serious compatibility in issues with python environments. The Pyvi tool is the only tool written in python 2 and the accuracy rate is significantly high. This is the reason why it is chosen to be applied in the NNVLP toolkit as a word segmenter [3]. The package can be downloaded directly as a python package.

After having installed the package from the pip installer, we can execute the package in order to observe the result of the Pyvi in term of word tokenizing with these simple snippet codes in Python. Table 6 shows another example of sentences taken from a Vietnamese website.

Table 4. *Comparison between several Vietnamese word tokenizers*

Tools	Author	Published Year	F1 (%)	Programming Language
UETsegmenter	Nguyen Tuan Phong	2016	98.98	Java
Pyvi	Tran Viet Trung	2016	97.8	Python2
Vitk	Le Hong Phuong	2016	97	Java
vnTokenizer	Le Hong Phuong	2009	96	Java
JVnSegmenter	Nguyen Cam Tu Phan Xuan Hieu	2007	92.4	Java

Table 5. *Word tokenizing of a sentence*

#coding:utf-8 **from** pyvi.pyvi **import** ViTokenizer, ViPosTagger **print** (ViTokenizer.tokenize(**u"Việt Nam là cực nam của văn hóa Đông Á."**))	
Input sentences	Việt Nam là cực nam của văn hóa Đông Á.
Output sentences	Việt_Nam là cực nam của văn_hóa Đông Á .

5. EXPERIMENTS AND DISCUSSION

Vietnamese is a quite fixed order language, with general word order SVO subject-verb-object). As for most languages which have relatively restrictive word orders, Vietnamese relies on the order of constituents to convey important grammatical information. Although Vietnamese text is written in a variant of the Latin alphabet (a script that exists in its current form since the 17th century, and has become the official writing system since the beginning of the 20th), there are three principal characteristics of Vietnamese which distinguish it from occidental languages. First, Vietnamese is an inflectionless language in which its word forms never change. Since there is no inflection in Vietnamese, all the grammatical information is conveyed through word order and tool words. The inflectionless characteristic makes a special linguistic phenomenon common in Vietnamese: type mutation, where a given word form is used in a capacity that is not its typical one (a verb used as a noun, a noun as an adjective. . .) without any morphological change. This leads to the fact that Vietnamese word forms are usually highly ambiguous in their part-of-speech. The same word may be a noun in a context while it may be a verb or a preposition in other contexts. For example, the word yêu may be a noun (the devil) or a verb (to love) depending on the context.

As the robot main responsibility is to serve as the receptionist of an educational

institution - STEM education center – which provide technical courses for elementary and middle school children, JCBot is required to process capability of informing parents of the center current and future courses, thus recommending them which classes is the best to suit for their need.

JCBot also must be capable of giving direction on classrooms and offices within the center for its customers and guests. Students of the center might need to be informed of their timetable as well as their final grades. Also, as for employees and teachers of the center, JCBot should be able to remind them their working schedule.

Furthermore, JCBot should offer user entertainment utilities such as music listening, games based on his/her appearance, emotions and preferences. As the customers of the center might speak different languages, the ability to offer information in many different languages is necessary as well.

Last but not least, JCBot most important objective is being perceived as friendly, human-like robot in order to attract users' attention. Therefore, the capability of understanding Vietnamese and English natural language is essential. Furthermore, JCBot is designed with two arms that can handle a wide range of motions and a screen capable of displaying various facial expressions. JCRobot software needs to be able to coordinate the action of their actuators so that the response to user command as naturally as possible. JCBot also possesses a touch screen on her chest for displaying information and interacting with her users. The user interface to this touch screen also needs and designes adaptively and responsively for users convenience.

Table 6. *Vietnamese word tokenizing of a paragraph*

Input sentences	Trong không khí vui tươi, rộn ràng của các hoạt động chào mừng ngày Phụ nữ Việt Nam 20/10, hội thi như một điểm nhấn là nơi để quý cô thể hiện sự đảm đang, chu đáo và sáng tạo của mình qua việc chăm lo bữa ăn cho gia đình, là người "Giỏi việc nước, đảm việc nhà", niềm tự hào của phụ nữ Việt Nam.
Output sentences	Trong không_khí vui_tươi , rộn_ràng của các hoạt_động chào_mừng ngày Phụ_nữ Việt_Nam 20/10 , hội thi như một điểm nhấn là nơi để quý cô thể_hiện sự đảm_đang , chu_đáo và sáng_tạo của mình qua việc chăm_lo bữa ăn cho gia_đình , là người " Giỏi việc nước , đảm việc nhà " , niềm tự_hào của phụ_nữ Việt_Nam .

Table 7. *Part of speech tag of the NNVLP toolkit*

Key	Meaning	Key	Meaning
N	Noun	**Np**	Proper Noun
Nc	Classifier Noun	**Nu**	Unit Noun
Ny	Abbreviate Noun	**V**	Verb
Vy	Abbreviate Verb	A	Adjective
M	Numeral	**E**	Preposition
C	Conjunction	**R**	Adverb
P	Pronoun	**L**	Determiner
T	Auxiliary Word	**I**	Emotivity Word
B	Loan-word	**S**	Component Stem
CH	Punctuation	**Y**	Abbreviate
X	Undetermined		

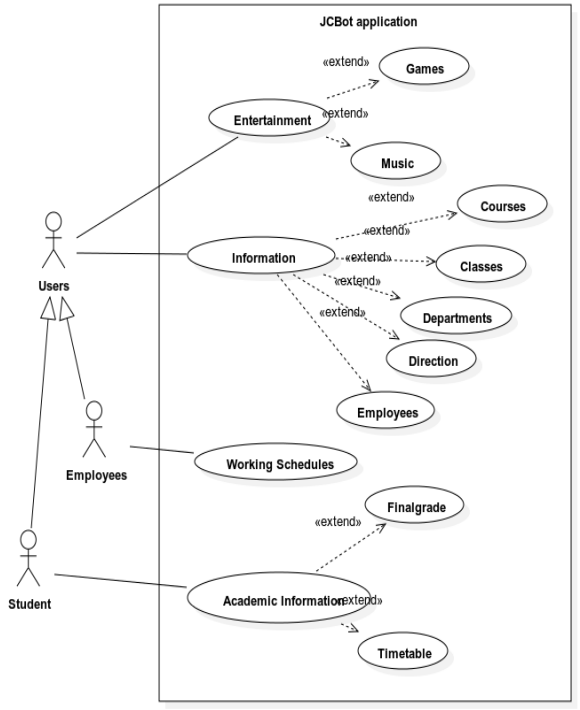

Fig. 3. *ULM uses case diagram of service robot application*

6. CONCLUSIONS

This paper proposed a new application for giving precisely predictable according to its learning data. This processing resembles the human learning process. Investigated were four key technologies: Word tokenization; Part-of-speech tagging; Name entity recognition and Text Classification. We obtained promising results and grasped issues to be tackled. Moreover, training robots will be a good research platform or test bed for human-robot interaction or multimodal spoken dialogue systems because it will not require difficult spoken Vietnamese understanding and users will continuously interact with them spontaneously, which enables researchers to collect huge data and to test their technologies over a longer span.

The results suggest that it is helpful to allow human to communicate with the robot and to observe correct demonstrations by the robot. Therefore, it would become important to voice recognition process with advantages of both the aspects, to achieve effective learning and to maintain the interests of the audience at the same time.

ACKNOWLEDGMENT

We would like to thank very much Ho Chi Minh City Department of Science and Technology, to have accepted the proposal of this research and to have prepared and managed the robot during the experimentation. We would also thank very much Ho Chi Minh City University of Technology and Education, Viet Nam supported financially the project, in which the study took place, particularly where was actively involved in the project.

REFERENCES

[1] Nguyễn, T.M.H., Romary, L., Rossignol, M. and Vũ, X.L., 2006. A lexicon for Vietnamese language processing. *Language Resources and Evaluation*, *40*(3-4), pp.291-309.

[2] Pham, Thai-Hoang, and Phuong Le-Hong. "End-to-end Recurrent Neural Network Models for Vietnamese Named Entity Recognition: Word-level vs. Character-level." In *International Conference of the Pacific Association for Computational Linguistics*, pp. 219-232. Springer, Singapore, 2017.

[3] Pham, Thai-Hoang, Xuan-Khoai Pham, Tuan-Anh Nguyen, and Phuong Le-Hong. "Nnvlp: A neural network-based vietnamese language processing toolkit." *arXiv preprint arXiv:1708.07241* (2017).

[4] Joulin, Armand, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hérve Jégou, and Tomas Mikolov. "Fasttext. zip: Compressing text classification models." *arXiv preprint arXiv:1612. 03651* (2016).

[5] Huyen, Nguyen Thi Minh, Azim Roussanaly, and Hô Tuong Vinh. "A hybrid approach to word segmentation of Vietnamese texts." In *Language and Automata Theory and Applications*, pp. 240-249. Springer, Berlin, Heidelberg, 2008.

[6] Le, Ngoc Minh, Bich Ngoc Do, Vi Duong Nguyen, and Thi Dam Nguyen. "VNLP: an open source framework for Vietnamese natural language processing." In *Proceedings of the Fourth Symposium on Information and Communication Technology*, pp. 88-93. ACM, 2013.

[7] Le-Hong, Phuong, Azim Roussanaly, Thi Minh Huyen Nguyen, and Mathias Rossignol. "An empirical study of maximum entropy approach for part-of-speech tagging of Vietnamese texts." In *Traitement Automatique des Langues Naturelles-TALN 2010*, p. 12. 2010.

[8] Hall, Gary. *Pro WPF and Silverlight MVVM: effective application development with Model-View-ViewModel*. Apress, 2011.

[9] Sokolova, Marina, and Guy Lapalme. "A systematic analysis of performance measures for classification tasks." *Information Processing & Management* 45, no. 4 (2009): 427-437.

Corresponding author:
Nguyen Truong Thinh
Ho Chi Minh City University of Technology and Education
Email: thinhnt@hcmute.edu.vn