THESIS INFORMATION

Title**: Sentiment Analysis with Contextual Valence Shifters for Vietnamese**

Major: Computer Science

Major Code: 62.48.01.01

PhD student: Tran Khai Thien

Advisors: Prof. Phan Thi Tuoi

University: Ho Chi Minh City University of Technology, VNU-HCMC

1. ABSTRACT

Various sentiment analysis works, tools, and applications have been developed to exploit opinions in user-generated content on social media. However, the performance of these systems is not great because sentiment analysis itself is a complex natural language processing problem. These works are still ineffective in dealing with some linguistic phenomena, such as context valence shifting and mixed opinion text.

The dissertation explores cases of contextual valence shifting in Vietnamese text to accomplish two objectives: 1) to build a sentiment vocabulary database in Vietnamese and 2) to combine the semantic-oriented approach with machine-learning techniques, and deep-learning methodology to handle the sentiment classification challenge.

Experiments show that paying attention to contextual valence shifting and using a combination of various methods are key for the system to yield more accurate results.

The thesis has 12 published articles, among which 06 articles in the list of international journals (3 articles in SCIE), 01 article in domestic journals, and 05 articles in the proceedings of international scientific conferences.

2. MAIN CONTRIBUTIONS

Building a sentiment dictionary for Vietnamese words and phrases.

Through analyzing linguistic features and emotional shifting in Vietnamese comments, the thesis translates into Vietnamese emotional English words based on SentiWordnet emotion dictionary using Logistic Regression and applying fuzzy computation proposed by Zadeh to propose an effective model for determining the emotional metric of Vietnamese words and phrases. The thesis adjusts fuzzy functions for calculating phrase sentiment based on the syntactic structure of Vietnamese phrases to be in harmony with Vietnamese language features.

Proposing an effective ensemble learning model with component learning sets learned on a dataset composed of many different characteristics of Vietnamese.

The different features of the data set are identified by methods oriented towards semantics, machine learning, and deep learning. The choice of the word embedding model Word2Vec and the deep learning method for base-learners of the ensemble learning model improve the effectiveness of the sentiment classification model. The proposed model of the thesis can be well applied to the English language as well.

3. QUESTION ISSUES TO CONTINUE THE RESEARCH

The thesis's research results have deciphered some problems in solving the problem of emotion analysis and yet it needs carrying out further studies to ameliorate the work's quality:

- Conducting more profound research on emotional shifting, then putting emotional analysis problems into application, in spite of this being a big challenge and involving many a field of linguistics.
- Considering upgrading some preprocessor tools such as a dependent grammar and parsing. These tools can greatly affect the accuracy of the system.
- Paying attention to the processing of nouns and noun phrases as well as expanding the emotional dictionary in following studies also constitutes essential work as the data source for Vietnamese emotion analysis is currently very limited.

– The thesis focus on processing for Vietnamese language, but the ideas and practical methods of the model proposed by the thesis can still be applied to other languages, such as English.

ADVISORS	PHD STUDENT
PROF. PHAN THI TUOI	TRAN KHAI THIEN