The 2016 Conference on Computational Linguistics and Speech Processing
ROCLING 2016, pp. 204-213

Computing Sentiment Scores of Verb Phrases for Vietnamese

Thien Khai Tran[1], Tuoi Thi Phan[2]
Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology
Ho Chi Minh City, Vietnam
[1] thientk@cse.hcmut.edu.vn, [2] tuoi@cse.hcmut.edu.vn

Abstract. Sentiment analysis is an emerging research field. One of the major tasks of sentiment analysis is building sentiment lexicons and calculating their scores, which is an essential job that provides "material" for all sentiment analysis problems. In this paper, we propose a fuzzy language computation by taking linguistic context into account to provide an effective method for computing the sentiment polarity of verb phrases. The positive results, which come from an experimental period, will provide us with a basis from which to build an effective sentiment analysis system by making use of the contextual valence shifter.

Keywords: sentiment lexicons; language computation; linguistic variable; fuzzy logic; fuzzy function; approximate reasoning

1. Introduction

Sentiment analysis (or opinion mining) is a new research field, but it is an important area that attracts the attention of not only researchers but also businesses and organizations. Building sentiment lexicons is an essential task that provides "material" for all sentiment analysis levels: document-based, sentence-based, concept-based, and aspect-based. One of the biggest English sentiment lexicons is SentiWordNet [15]. It contains opinion terms extracted from WordNet [3] with a semi-supervised learning method and is available for research purposes. SenticNet [2] is a lexical resource used in concept-level sentiment analysis. It provides sentiment scores for 14,000 common sense concepts. To tackle the problem of mining verb expressions to identify opinions from customer reviews, there also have been a large number of works discovered the semantics of verbs and verb phrases. For example, Sokolova and Lapalme [13] incorporate semantic verb categories including verb past and continuous forms into features sets. Neviarouskaya et al. [9] built a rule-based approach to incorporate verb classes from VerbNet [12] to detect the sentiment orientation of sentences.

For Vietnamese, Vu et al. [18] built VietSentiWordNet, which contains 1,000 words; it also includes syntactic rules for extracting sentiments from review sentences. Hong et al. [4] built an opinion dictionary for product domains based on a combination of a statistical method, a machine translation technique, and WordNet. Their work outperformed VietSentiWordNet. Recently, in 2016, Son et al. [14] built a Vietnamese opinion dictionary that contains five sub-dictionaries: verb, adjective, adverb, noun, and proposed features. The sub-dictionaries are based on the English emotional analysis approach and adapted to traditional Vietnamese language. The support vector

machine classification technique was then used to identify the emotional content of the user's message. However, the authors calculated the sum of the emotional values of the linguistic variables based on feelings.

In this paper, based on Vietnamese linguistic characteristics and the fuzzy computation proposed by Zadeh [6,8,19], we present an effective method for computing the sentiment polarity of verb phrases. From this, we built a fine-grained linguistic sentiment analysis for Vietnamese. Zadeh developed the concept of fuzzy linguistic variables that modify the meaning and intensity of their operands, and we developed a modified fuzzy function suitable for use with the Vietnamese language. In our experiments, we showed that our system provides good results.

In this paper, we describe our research contributions, as follows:

- The mining of Vietnamese linguistic characteristics to propose sentiment computing rules for verb phrases.
- Proposing the modified fuzzy functions suitable for Vietnamese linguistic variables.
- Taking steps toward building an effective sentiment analysis system with fine-grained scores.

The outline of the rest of this paper is as follows: in section 2, we present the linguistic characteristics of Vietnamese; in section 3, the proposed model is described; in section 4, we report our experiments; and finally, we conclude the paper and discuss possibilities for future work.

2. Linguistic Characteristics of Vietnamese

Vietnamese is an isolating language with lexical tones and monosyllabic word structure. These characteristics are evident in all aspects: phonetic, vocabulary, and grammar. For vocabularies, Le [7] and Nguyen [11] proposed three common standards used to classify them: 1) essential meaning of the word type, 2) the function of the word in the sentence, and 3) the ability to combine with other words. Both Vietnamese and English words can be divided into content words and function words. Content words carry lexical meaning; while, function words relate lexical words to each other. For both languages, content words may be further divided into nouns, adjectives, and verbs. Nouns are words that represent entities; adjectives represent qualities or characteristics; and verbs represent actions or states. In English, most adverbs are content words, but Vietnamese adverbs are function words. Generally, these words modify any part of speech other than a noun. Adverbs can modify verbs, adjectives, clauses, sentences, and other adverbs. In this paper, we only focus on verbs and adverbs.

2.1 Vietnamese Verbs

Verbs denote action, state, or occurrence, and form the main part of the predicate of a sentence. In Vietnamese, there are some types of verbs [1,10] as follows:

- Intransitive verb (denotes Vin): Intransitive verbs are not used with an object; they relate only to the subject. For example: ngủ sleep, ngồi sit, khóc cry, cười smile etc.

- Transitive verb (denotes Vex1): Transitive verbs are action verb that have an object to receive that action. For example: làm do, trồng plant, xây build, phát triển develop, đàn áp suppress, mua bán purchase etc.
- Verb of giving and receiving (denotes Vex2): For example: cho give, gửi send, tặng offer, biếu donate etc.; nhận get, vay lend etc.
- Verb of command (denotes Vex3): this type of verb presents activities that promote or prevent one from doing something else. For example: khuyên advice, bắt buộc obligatory, đề nghị suggest, đình chỉ suspend etc.
- Verb of moving, direction (denotes Vdr). For example: vào in, ra out, lên up, xuống down, đến come, lại back etc.
- Modal verb (denotes Vt): is a type of verb that is used to indicate modality, that is: likelihood, ability, permission, and obligation. For example: cần need, muốn want, ước wish etc. There are some kind of modal verbs:
    - A need (denotes Vt1): nên should, phải have to…
    - An ability (denotes Vt2): có lẽ may, có thể can, không thể cannot...
    - A volition (denotes Vt3): dự định intend, dám dare...
    - A wishing (denotes Vt4): hy vọng hope, ước wish, mơ dream...
    - A recipient, stand (denotes Vt5): đạt obtain, nhận get...
    - A judge (denotes Vt6): cho, thấy...
- Verb of mentality, awareness (denotes Vin1): hối tiếc regret etc.
- Verb of emotion (denotes Vin2): hạnh phúc happy, buồn sad, giận angry etc.
- Verb of physiology (denotes Vin3): mong want etc.
- Verb of nature, morality, personality (denotes Vin4): nhịn condescend, tha thứ forgive etc.

2.2 Vietnamese Adverbs

Adverbs are words that modify or describe verbs, adjectives, clauses, sentences, and other adverbs. Generally, these words modify any part of speech other than a noun.

The following observations relate to Vietnamese adverbs when comparing them with English adverbs.

Morphology. English adverbs are content words but Vietnamese adverbs are function words. To the best of our knowledge, there are approximately 600 Vietnamese adverbs while English has more than 6,000 adverbs.

Syntactic. In English, the adverb is the head of the phrase, can appear alone, or can be modified by other words. An adverb phrase is a subordinate clause in a sentence. In Vietnamese, adverbs do not have primary grammatical functions in a clause (subject, predicate).

Function. English adverbs modify a verb, adjective, or another adverb. The adverb typically expresses the manner, time, place, cause, or circumstance in which something has happened. In

Vietnamese, adverbs do not have real meaning for describing the name, action, status, nature, and quantity of things. Adverbs only contain grammatical meaning based on the part of speech they modify.

Position. There are three normal positions for adverbs in an English sentence: before the subject, between the subject and the verb, and at the end of the clause. Vietnamese adverbs can precede or follow the words they modify.

Classification. English adverbs have the following types: time adverbs, degree adverbs, manner adverbs, frequency adverbs, and place adverbs.

For Vietnamese, a selection of the types of adverbs and their ability to combine with verbs are presented in Table 1.

Table 1. Vietnamese adverbs and their ability to combine with verbs.

Types	Adverbs	Kinds of verbs	Verb phrases
PV1 the same, similar	đều both cũng too cùng jointly	PV1 + (Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3)	cùng chuẩn bị (prepared jointly)
PV2 continuation	vẫn still	PV2 + (Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3)	vẫn cười (still smile)
PV31 time relation (present+ future)	sẽ will đang - ing	PV3 + (Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3)	Anh sẽ thi rớt. (He will fail the exam.)
PV32 time relation (pass)	vừa just đã -ed từng	PV3 + (Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3)	Anh ấy từng thi rớt. (He has failed the exam.)
PV41 frequency (increase)	thường usually hay often năng always	PV4 + (Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3)	hay ăn trễ (often eat lately)
PV42 frequency (decrease)	ít rarely hiếm rarely	PV4 + (Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3)	ít đi trễ (rarely go late)
PV5 degree	rất very, hơi a bit quá too, lắm much cực extremely	PV5 + (Vin1, Vin2, Vin3)	rất yêu (very love)
PV6 confirmation	có to	PV6 + (Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3)	có tồn tại (to exist)
PV7 command	đừng don't chớ shouldn't	PV7 + (Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3)	chớ hiểu lầm (shouldn't misconceive)
PH negation	không don't chưa yet	PH + (Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3)	không đi (don't go)

PV9 immediateness	ngay right liền right tức khắc right tức thì right	(Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3) + PV9	quay ngay (spin right)
PV10 mediateness	dần dần gradually dần gradually từ từ slowly	(Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3) + PV10	dần dần cải thiện (gradually improved)
PV11 direction	vào into	Vdr + PV11	cứ nói vào (talk into)
PV12 direction	ra out lùi back	Vdr + PV12	lại bàn ra (talk out)
PV13 activity direction	đi away, về back tới to, qua over lại back	Vdr + PV13	mang lại (bring back)
PV14 quickview	qua through	(Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3) + PV14	Đọc qua (read through)
PV15 ascription	cho for	(Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3) + PV15	Người ta cười cho. (people can laugh)
PV16 joint	với along	(Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3) + PV16	Đi với (go along)
PV17 do for himself or he does it with himself	lấy out	(Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3) + PV17	Cầm lấy. (Take it out)
PV18 describe a positive	được obtain	(Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3) + PV18	Tôi mua được cái áo đẹp. (I acquire a nice shirt.)
PV19 describe a negative	phải ought	(Vdr, Vin1, Vin2, Vin3, Vin4, Vex1, Vex2, Vex3) + PV19	Cô ấy tin phải người xấu. (She has trusted a bad guy.)

3. Proposed Model

In this model, we try to compute the sentiment scores for word phrases that include verbs and adverbs based on Vietnamese linguistic characteristics. By combining with some adverbs, the verb phrases will have a smoother sentiment scaling.

3.1 System architecture

Our system architect is presented in Figure 1. We used the English sentiment dictionary, SentiWordNet, and the translate tools Vdict* and Google Translate** to build the core verb lexicons with sentiment scores for Vietnamese. The fuzzy rules then computed the sentiment scores for the whole phrase, which included the verbs and associated adverbs.

* http://vdict.com ** https://translate.google.com/

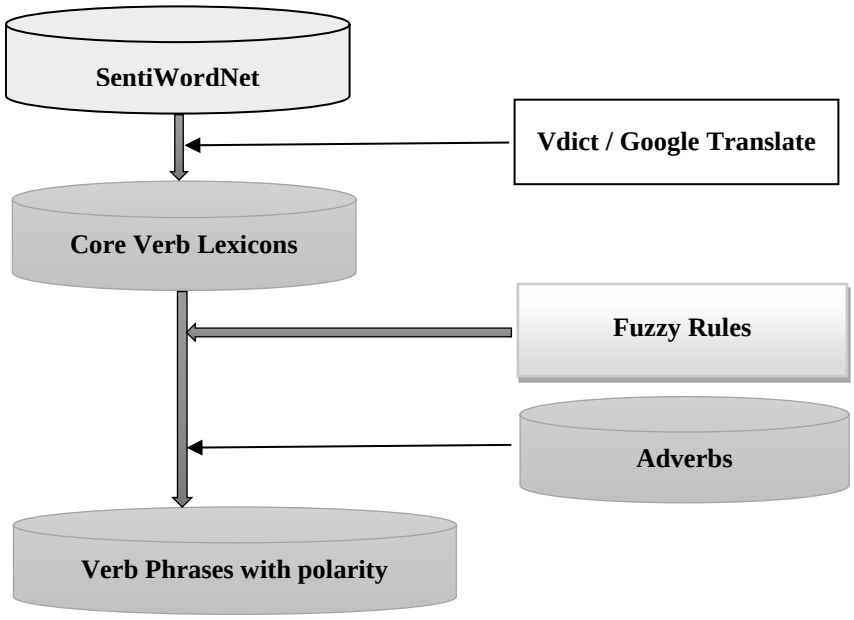

Fig.1. System architecture.

Building core verb lexicons.

We constructed a handcrafted opinion dictionary containing approximately 1,000 verbs. The number of words was high enough to cater to the problem we sought to solve. These words:

- appeared in the review corpus obtained from [16,17].
- are matched with corresponding English words in SentiWordNet; we used Vdict and Google Translate to check this. To meet the scope of this project, we assigned opinion word scores that were the same as the scores of words in SentiWordNet.

In Table 2, we describe some of the opinion words that appear in this core dictionary.

Table 2. Fragment of Core Opinion Dictionary.

Term	Positive Score	Negative Score	POS	Tag
yêu $_{\text{love}}$	0.375	0	Verb	Vin2
ghét $_{\text{hate}}$	0	0.75	Verb	Vin2
tin tưởng $_{\text{trust}}$	0.625	0	Verb	Vin2
kính nể $_{\text{respect}}$	0.5	0	Verb	Vin2

3.2 Fuzzy Rules

Overall sentiment scores for the verb phrases were calculated thanks to fuzzy rules that were associated with the combination between the verb (denotes x) and the adverb (denotes y). We used fuzzy functions to incorporate the effect of the adverbs in the verb phrases. We considered the sentiment score of a verb to be its initial fuzzy score $\mu(x)$. Based on Vietnamese linguistic characteristics, we realized five sentiment shifting scalings for adverbs that go along with verbs;

these were intensifier, booster, diminisher, minimizer, and modifier. General principles for classifying adverbs are as follows:

1. Adverbs of degree: There are five levels: intensifier, booster, diminisher, minimizer, and modifier. Some Vietnamese adverbs of degree are presented by Table 3.
2. Other adverbs: There are three levels that are booster, diminisher, and modifier:
   - Booster: PV1, PV2, PV31, PV41, PV6, PV9, PV10, PV13, PV16, PV17, PV18.
   - Diminisher: PV32, PV42, PV10, PV12, PV14, PV15.
   - Modifier: PH, PV19.

   Some of these adverbs are presented by Table 4.

Table 3. Some Vietnamese adverbs of degree with their scalings.

intensifier	booster	diminisher	minimizer	modifier
cực kỳ extremely	rất very	khá rather	cũng seemingly	không no
cực strongly	quá too	tương đối relatively	hơi a bit	chẳng no
siêu super	lắm much	tạm rather	rồi already	chả no

Table 4. Some other adverbs with their scalings.

booster	diminisher	modifier
đều both	phải to	chả no
vẫn still	hiếm rarely	không not
hay often	từng already	chưa yet

In our system, Vietnamese adverbs are organized in a database. In Table 5, we describe some of the adverbs that appear in our adverb database. In the table, "Tag" is the scaling category to which an adverb can belong.

Table 5. Some Vietnamese adverbs with their tags.

Adverbs	Types	Tag
cực kỳ extremely	PV5	intensifier
không no	PH	modifier
phải to	PV19	modifier
hay often	PV41	booster
hiếm rarely	PV42	diminisher

Similar to Zadeh's proposition [6,8,19], if the verb phrase had an adverb, its modified fuzzy score was computed by (1):

$$f(\mu(x)) = 1 - (1 - \mu(x))^{\delta} (1)$$

We chose δ = 4, 2, 1/2, or 1/4 if the adverb was a(n) intensifier, booster, diminisher, or minimizer. which gives us a modified fuzzy score, as indicated in (2).

$$f(\mu(x), y) = \begin{cases} 1 - \sqrt[4]{1-\mu(x)} & y.tag \in minimizer \\ 1 - \sqrt[2]{1-\mu(x)} & y.tag \in diminisher \\ 1 - (1-\mu(x))^2 & y.tag \in booster \\ 1 - (1-\mu(x))^4 & y.tag \in intensifier \\ -\mu(x) & y.tag \in modifier \cap y.type \in pv19 \\ 0 & y.tag \in modifier \cap y.type \in ph \end{cases} \quad (2)$$

with

- f(μ(x),y) is the sentiment score of a verb phrase, in which x: verb, y: adverb.
- μ(x) is the sentiment score of a verb.

Table 6 presents an example of verb phrases and their sentiment scores.

Table 6. Sentiment score of verb phrases.

f(μ(x),y)					**μ(x)**
intensifier	**booster**	**diminisher**	**minimizer**	**modifier**	**verb**
cực kỳ (yêu) extremely (love)	rất (yêu) very love	khá (yêu) rather love	cũng (yêu) seemingly love	không (yêu) doesn't love	yêu love
0.85	0.61	0.21	0.11	0	0.375

According to the formula (2), if the adverb was a modifier (y.tag = modifier), we had two cases. For example:

f(*tin* ***phải*** trust (a bad guy)) = - f(*tin* *trust*) = - 0.625, but

f(***đừng*** *hiểu lầm* shouldn't misconceive) = 0

4. Experiments

Cohen's kappa coefficient. Two judges participated in categorizing the adverbs as intensifier, booster, diminisher, minimizer, or modifier. To compute the "between judges' agreement," we used the Cohen's kappa coefficient [5], as follows:

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (3)$$

where

Pr(a) is the relative observed agreement among the judges and Pr(e) is the hypothetical probability of a chance agreement. The Cohen's kappa coefficient of our corpus $k = 0.80$.

104 Vietnamese verb phrases from Agoda.com were randomly collected to evaluate the system performance. The system was capable of handling 100 phrases. The highest sentiment score was +0.98 *(cực kỳ tin tưởng extremely trust)*, and the lowest one was -0.99 *(vô cùng ghét extremely hate)*. Obviously, the adoption of fuzzy logic for computing sentiment scores of verb phrases helps the sentiment valences have a smoother sentiment scaling, not only 1, -1, and 0. In Table 7, we describe the eleven levels of sentiment polarities that obtained from the testing.

Table 7. The eleven levels of sentiment polarities.

Level	5[‡]	4	3	2	1	0	-1	-2	-3	-4	-5
Number of phrases	2	12	6	11	13	13	14	13	5	10	1

[‡] *5: extremely positive; 4: very positive; 3: positive; 2: rather positive; 1: a little positive; 0: neutral; -1: a little negative; -2: rather negative; -3: negative; -4 very negative; -5: extremely negative.*

Application. By identifying the fine-grained scores of phrase in sentence, the system can deal with many multi-class sentiment classification problems. For example, to classify the sentences, we simply counted the mean scores of sentiment phrases in each sentence. If the final score was more than +0.1 the sentence was considered to show a positive emotion. If the score was less than -0.1 the sentence was considered to show a negative emotion. Otherwise, the sentence was considered to show a neutral emotion.

For example: Rất tin tưởng vào dịch vụ khách sạn, cực yêu phong cảnh nơi đây *(Very trust in the hotel services, extremely love the scenery).* Total score: *(f(rất tin tưởng very trust) + f(cực yêu extremely love)) / 2 = (0.86 + 0.85) / 2 = 0.855.* Therefore this sentence is considered to show a extremely positive emotion.

5. Conclusions

This paper has presented a mechanism for computing the sentiment scores of verb phrases by mining the Vietnamese linguistic characteristics and using fuzzy functions. We have shown this approach to be effective. By identifying the opinion phrase polarity automatically, the method can be useful to deal with many sentiment analysis problems. Still, there are a number of challenges to indentify, classify, and calculate the sentiment scores of verbs and verb phrases because of linguistic challenges and the rule based approaches often suffer from domain-specificity problem. Future work will expand our research with more data and adopt this approach for developing Vietnamese sentiment lexicons with adjective phrases and noun phrases. We will also consider using machine learning methods to help the system become more robust.

6. Acknowledgment

This paper was supported by the research project C2016-20-32 funded by Vietnam National University Ho Chi Minh City (VNU-HCM).

7. References

[1] Diep Van Ban and Hoang Van Thung, (1998).: Ngữ pháp tiếng Việt, "Vietnamese Grammar", Vietnam Education Publishing House.

[2] Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal, (2014). SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. AAAI, pp. 1515–1521.

[3] Princeton University "About WordNet." WordNet. Princeton University. 2010. <http://wordnet.princeton.edu>.

[4] Hong Nam Nguyen, Thanh Van Le, Hai Son Le, Tran Vu Pham, (2014). Domain Specific Sentiment Dictionary for Opinion Mining of Vietnamese Text. The 8th Multi-Disciplinary International Workshop on Artificial Intelligence (MIWAI 2014): 136-148.

[5] J. Carletta (1996). Assessing agreement on classification tasks: the Kappa statistic, Computational Linguistics, 22, 249-254.

[6] L. A. Zadeh, (1975). "The concept of a linguistic variable and its application to approximate reasoning-II," Information Sciences,vol.8, no.4, part3, pp.301–357.

[7] Le Van Ly, (1972) .:Sơ thảo ngữ pháp Việt Nam, "Vietnamese Essentials: Grammar". Vietnam Education Publishing House.

[8] Mita K. Dalal and Mukesh A. Zaveri, (2014). "Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges," Applied Computational Intelligence and Soft Computing. Volume 2014, Article ID 735942.

[9] Neviarouskaya A., Prendinger H., Ishizuka M, (2009) "Semantically distinct verb classes involved in sentiment analysis", IADIS International Conference Applied Computing 2009.

[10] Nguyen Kim Than, (1997) .: Nghiên cứu ngữ pháp tiếng Việt, "Vietnamese Grammar". Vietnam Education Publishing House.

[11] Nguyen Tai Can, (1975) .: Ngữ pháp tiếng Việt (Tiếng - Từ ghép - Đoản ngữ), "Vietnamese Grammar". Hanoi Publisher.

[12] Schuler, K. K., Korhonen, A., and Brown, S. W. (2009). VerbNet overview, extensions, mappings and applications. In HLT-NAACL, 13–14.

[13] Sokolova, M., and Lapalme, G. 2008. Verbs Speak Loud: Verb Categories in Learning Polarity and Strength of Opinions. In Canadian Conference on AI, 320–331.

[14] Son Trinh , Luu Nguyen, Minh Vo, Phuc Do, (2016) "Lexicon-Based Sentiment Analysis of Facebook Comments in Vietnamese Language," Recent Developments in Intelligent Information and Database Systems Volume 642 of the series Studies in Computational Intelligence. pp 263-276.

[15] Stefano Baccianella, A.E., Sebastiani, F. (2010).: Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: LREC'10 (May 2010).

[16] Thien Khai Tran and Tuoi Thi Phan, (2015) "An upgrading SentiVoice – a system for querying hotel service reviews via phone," Proceedings of the 19th International Conference on Asian Language Processing (IALP 2015), Suzhou, China. pp. 115-118.

[17] Thien Khai Tran and Tuoi Thi Phan, (2015) "Constructing Sentiment Ontology for Vietnamese Reviews," Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS2015), Brussels, Belgium. December 11 - 13, 2015. pp.281-285. ISBN: 978-1-4503-3491.

[18] Tien-Thanh Vu, Huyen-Trang Pham, Cong-To Luu, Quang-Thuy Ha (2011).: A feature-based opinion mining model on product reviews in Vietnamese. In: Semantic Methods for Knowledge Discovery and Communication, Polish-Taiwanese Workshop, Springer Berlin Heidelberg. pp.22–23.

[19] V. N. Huynh, T. B. Ho, and Y. Nakamori, (2002) "A parametric representation of linguistic hedges in Zadeh's fuzzy logic," International Journal of Approximate Reasoning, Vol. 30, No. 3, pp.203–223.