Sentiment Classification using Enhanced Contextual Valence Shifters

Vo Ngoc Phu

Ho Chi Minh City University of Technology

Ho Chi Minh City, Vietnam

vongocphu03hca@gmail.com

Phan Thi Tuoi

Ho Chi Minh City University of Technology

Ho Chi Minh City, Vietnam

tuoi@cse.hcmut.edu.vn

***Abstract*— We have explored different methods of improving the accuracy of sentiment classification. The sentiment orientation of a document can be positive (+), negative (-), or neutral (0). We combine five dictionaries from [2, 3, 4, 5, 6] into the new one with 21137 entries. The new dictionary has many verbs, adverbs, phrases and idioms, that are not in five ones before. The paper shows that our proposed method based on the combination of Term-Counting method and Enhanced Contextual Valence Shifters method has improved the accuracy of sentiment classification. The combined method has accuracy 68.984% on the testing dataset, and 69.224% on the training dataset. All of these methods are implemented to classify the reviews based on our new dictionary and the Internet Movie data set.**

***Keywords*—sentiment classification, sentiment orientation, valence shifters, contextual valence shifters, term counting.**

I. INTRODUCTION

To understand the sentiment of documents or speeches, it is very important for us to categorize documents by various ways such as subject, genre, or the sentiment expression in the documents.

At present, numerous researchers have used many methods such as support vector machines method (SVM), Naïve Bayes method, Contextual Valence Shifters, Maximum entropy classifiers to classify the sentiments as positive, as negative or as neutral.

We have one approach to classify the sentiments, which is represented and compared each other in this paper. We propose the method based on the combination of Term-Counting and enhanced Contextual Valence Shifters methods (shortly, TC-CVS method), which classifies the 25000 documents of the testing dataset and the 25000 documents of the training dataset.

The experiment of the paper will show the high accuracy of our combined method.

The rest of the paper is organized as follows: Section 2 discusses the related works about CVS in [7,8]. In Section 3, we represent the data set for supporting the classification with 25000 movie reviews of the training dataset and 25000 movie reviews of the testing dataset. Section 4 represents the methodology of the proposed TC-CVS method and an advantage of our new dictionary. The experiments show the result accuracy of TC-CVS method with our dictionary in comparation with some other methods in section 5 .Section 6 is the conclusion of our proposed method.

II. RELATED WORK

According to [7], to define the emotions of an article, a sentence, a paragraph is done based on the valence of the word or term only, which never has an absolute accuracy, but always gets incorrect results. However, the identification of emotions in an article, a sentence, a paragraph is be able to perform based on the chemotherapy of words or terms with their context. This method identifying the emotion is called Contextual Valence Shifters (CVS), that can give results with a high accuracy.

Normally, the terms are classified as positive, or as negative, or as neutral emotions. The simple calculation of emotions is a total number of negative and positive subjects appeared in a text; then the classification result is based on the high valence of negative or positive. [7] found out the simple count is not efficient for classifying the emotions. Because, many words or terms have positive or negative valences, which valences depend on the context of those words or those terms.[7]counted the emotions by CVS for a sentence or an article level, but the authors of [7] did not give the accuracy of the results.

[8] used two methods to determine the sentiments expressed by reviews of the movies. [8] examined the effect of CVS on the classification of all reviews. Moreover, [8] tested the sentiments by three factors: negation, intensifier and diminisher.

Negation is used to invert the sentiment polarity of a particular term. Intensifier, diminisher are used to increase and decrease the emotion level such as positive or negative of the term.

The first method is to count the terms and to classify all reviews based on the number of positive and negative terms in the article. [8] applied the result of [2] to identify the positive and negative terms, as well as the negation, diminishing and intensifier terms with the data set of the film reviews [1] [8] showed that the expansion of term-counting method with CVS is able to improve the accuracy of a classification.

The second method is the combination of Maximum Likehood (ML) and SVM algorithms. The authors of [8] started to train terms with the unigram feature. Then, they trained the terms with bigram feature, which involves a valence shifter (a contextual valence shifter) and other feature. The higher accuracy of the classification of words in the positive or negative terms is due those features.[8] showed that the combination of two methods improved the result of the emotion classification

III. DATA SET

We use a public available large dataset of classified movie reviews from the Internet Movie Database (IMDb) [1]. The dataset has 25000 highly polar movie reviews for training and 25000 for testing. The training dataset contains 12500 positive movie reviews and 12500 negative movie reviews. The testing dataset contains 12500 positive movie reviews and 12500 negative movie reviews.

We use the enhanced Contextual Valence Shifters method to classify the 25000 reviews for training and 25000 reviews for testing. Therefore, the result of the enhanced Contextual Valence Shifters gets the better accuracy. The accuracy of the training set is approximately equal the accuracy of the testing set.

IV. METHODOLOGY

A. Evaluation Methodology

Like any other text classification problem, sentiment classification has main four measures of classification effectiveness; namely, accuracy, precision, recall and f-score.

Accuracy (A)

The accuracy of a classifier is defined as the percent of correctly classified objects. It is calculated as:

$$A = \frac{TP+TN}{TP+TN+FP+FN}$$

with:

TP denotes the number of positively-labeled test documents that were correctly classified as positive.

TN denotes the number of negative-labeled test documents that were correctly classified as negative.

FP denotes the number of positive-labeled test documents that were incorrectly classified as positive.

FN denotes the number of negative-labeled test documents that were incorrectly classified as negative.

Precision (P)

The precision for a class is defined as the probability that if a random document should be classified with this class, then this is the correct decision.

Precision for the positive class for instance is calculated as:

$$P = \frac{TP}{TP+FP}$$

Precision for the negative class for instance is calculated as:

$$P = \frac{TN}{TN+FN}$$

Recall (R)

The recall for a class is defined as the probability that if a random document should be classified with this class, then this is the taken decision.

Recall for the positive class for instance is calculated as follows:

$$R = \frac{TP}{TP+FN}$$

Recall for the negative class for instance is calculated as follows:

$$R = \frac{TN}{TN+FP}$$

F-Score (F)

In statistical analysis of binary classification, F-score or F-measure is a measure of a test's accuracy. It considers both the precision (P) and the recall (R) of the test to compute the score. The F-score can be interpreted as a weighted average of the precision and recall, where an F- score reaches its best value at 1 and worst score at 0.

The traditional F-measure or balanced F-score is the harmonic mean of precision and recall:

$$F = 2x\frac{P*R}{P+R}$$

B. Contextual Valence Shifters

The idea of the Term-Counting method is simple to count number of negative and positive terms in the sentence. If the positive terms are more than negative ones, then the sentence will be considered positive emotion and vice versa. If numbers of positive and negative terms are equal, then the sentence is neutral emotion.

We do not implement Contextual Valence Shifters method with subtopics, genre constraints, Cultural constraints the same as [7].

Our proposed method, that is the combination of Term-Counting and Contextual Valence Shifters methods (shortly, TC-CVS) is better than Term-Counting method and Contextual Valence Shifters method. Consider the following example.

Example: The sentence “Alex is very brilliant to solve most of problems but she solves her loving problem very badly and terribly”. To assume, we have the dictionary which contains term ‘very’ is an intensifier with 50%, term ‘brilliant’ is a positive polarity with valence = +2, ‘problem’ is a neutral polarity, ‘solve’ is a positive polarity with valence =

+1, ‘loving’ is a neutral polarity, ‘badly’ term is a negative polarity with valence = -2 and ‘terribly’ is a negative polarity with valence = -2.

According to Term-Counting method, that sentence has three positive polarities ($brilliant^+$ and $solve^+$occurred 2 times, so the valence = +4) and two negative polarities ($badly^-$ and terribly, so valence= -4); therefore, that sentence is classified as a positive polarity (valence = 0).

According to Contextual Valence Shifters method:

very brilliant (but) 2+1 = 3 => 0 (adjust)

solve (but) +1 => 0 (adjust)

solve +1

very badly and terribly -2 + -2*50% + -2 = -5

Total score: - 4; so, the sentence has a valence -4.

Our proposed method, that is a combination of Term-Counting and Contextual valence Shifters methods:

very brilliant (but) 2+1 = 3 => 0 (adjust)

solve (but) +1 => 0 (adjust)

solve +1

very badly and terribly (-2 + -2) + (-2 + -2)*50% = -6

Total score: - 5;so, the sentence has a valence -5.

In our proposed method, Term-Counting method must be implemented first, and then Contextual Valence Shifters method is performed. If Contextual Valence Shifters method is performed first, after that Term-Counting method is implemented, then the result will be wrong. See example above, if we implement Contextual Valence Shifters method first, term ‘but’ does not affect the rest of the sentence.“Alex is very brilliant to solve most of problems” will have valence = +3, “very badly and terribly” will have valence = -2 + -2*50% + -2 = -5 and “she solves her loving problem very badly and terribly” will have valence = -5 + 1 (valence of “solve” is +1) = -4; therefore, valence of the sentence “Alex is very brilliant to solve most of problems but she solves her loving problem very badly and terribly.” will be : +3 -4= -1, so the result will be wrong.

The improved accuracy of our proposed method relies on our dictionary, which is built from five dictionaries [2, 3, 4, 5, 6]. Besides, this dictionary is added many terms such as verb, adjective, adverb, noun, phrases, idioms, which are not absolutely in these five dictionaries.

Our new dictionary takes General Inquirer dictionary [2] as a core. Each word and each phrase in the five dictionaries have not a valence, they just have their polarity such as strong positive, positive, strong negative, negative.

We assign the polarity of words, phrases, and idioms in our new dictionary by a valence of them as input parameters. Because if word, phrases and idioms have not a valence, we cannot apply the Contextual Valence Shifters method for them to classify emotions.

The dictionary for sentiment classification has a certain accuracy, which depends on the sentiment polarity of each term in this dictionary. Hence, the accuracy of the dictionary always affects the result accuracy of the emotion classifying methods.

Emotion classification result of document depends on sentence level and emotion classification result of sentence depends on words’ sentiment in this sentence. Word’s sentiment polarities depend on the dictionary, in which they are.

Each element of the new dictionary has a valence. This value is determined accurately or not will affect the classification results of Contextual Valence Shifters method as well as our proposed TC-CVS method.

Algorithm 1: TC-CVS method classifying at sentence level.

Input: a sentence

Output: valence of the sentence

Method:

Begin

Perform separating sentence into individual words;

Apply Term-Counting method to identify polarity and valence of words.

Apply Contextual Valence Shifters method to identify valence of the sentence.

End

Example, to determine a emotion polarity of the sentence “Peter is very good at Math but today he does not do his test well” by two methods CVS and TC-CVS, as follows.

According to CVS method

- with C dictionary, phrase “Peter is very good at Math” has valence +4 (‘very’ is intensifier with 100%, ‘good’ has valence +2, ‘well’ has valence +4. So, “very good” has valence: +2+2*100%= +4)

and phrase “he does not do his test well”has valence-4.

So, the sentence has neutral polarity.

According to our proposed TC-CVS method

- with D dictionary, phrase “Peter is very good at Math” has valence +3 (‘very’ is intensifier with 50%, ‘good’ has valence +2, ‘well’ has valence +4. So, “very good” has valence: +2+2* 50%= +3)

and phrase “he does not do his test well” has valence

-4.

So, the sentence has negative polarity.

The new combined dictionary takes full advantage of five dictionaries. Because, there are a lot of words, phrases and idioms in one dictionary but they are not in the other. Each of five dictionaries is not adequate to words, phrases, and idioms. For example, term ‘queue’ is not in [2] but that term is in [3], [4], [5], [6]. Terms ‘bellow’, ‘shape’, ‘sin’ are not in [2,3,4,5,6], but term ‘rejoice’ is in all of five dictionaries [2,3,4,5,6]. This also occures in the polarity problem. When to classify a polarity of document, if a term is not in the dictionary, then it is calculated as a neutral polarity. Therefore, the lack of terms or polarities in dictionary affects result of sentiment classification.

Algorithm 2. Our Proposed TC-CVS method for document level.

Input: a movie reviews as a document, a paragraph, a sentence, a pharse, a word

Output: valence of the movie review

Method

Begin

1 Perform to separate reviewinto sentences,if this is a document or paragraph;

2 For each sentence in step 1, calculate valence;

2.1 if sentence contains presuppositional items then perform the same as in [7];

2.2 if the sentence contains connectors then go to 3;

2.3 if the sentence contains comma (,) or semicolon (;) then go to 4;

2.4 if the sentence contains ‘either .. or..’ then go to 5;

2.5 if the sentence contains ‘neither .. or .. ‘ then go to 6;

2.6 if the sentence contains comparative or superlative then go to 7;

2.7 if the sentence contains intensifier or diminisher then go to 8;

2.8 if the sentence contains negative forms (not, never, nothing, no one, no more, none, none of) then go to 9.

2.9 perform the following:

Calculate valence of pharse, sentence and idiom; then, remove them out of sentence;

Perform split the sentence into individual words based on ‘ ‘; For each individual words, calculate valence of each word; Sum all valence; Return valence of the sentene.

3 Processing connectors the same as [7], [8].

4 Processing comma (,) or semicolon (;) the same as [7], [8].

5 Processing ‘either..or..’ the same as [7], [8].

6 Processing ‘neither ..or..’ the same as [7], [8].

7 Processing comparative, superlative the same as [7], [8].

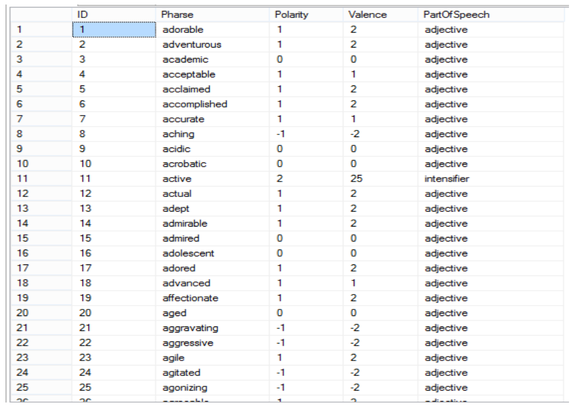
8 Processing intensifier, diminisher the same as [7], [8].

9 Processing negative forms the same as [7], [8]

10 Sum all valence of each sentence.

11 Return valence of the document.

End



	ID	Pharse	Polarity	Valence	PartOfSpeech
1	1	adorable	1	2	adjective
2	2	adventurous	1	2	adjective
3	3	academic	0	0	adjective
4	4	acceptable	1	1	adjective
5	5	acclaimed	1	2	adjective
6	6	accomplished	1	2	adjective
7	7	accurate	1	1	adjective
8	8	aching	-1	-2	adjective
9	9	acidic	0	0	adjective
10	10	acrobatic	0	0	adjective
11	11	active	2	25	intensifier
12	12	actual	1	2	adjective
13	13	adept	1	2	adjective
14	14	admirable	1	2	adjective
15	15	admired	0	0	adjective
16	16	adolescent	0	0	adjective
17	17	adored	1	2	adjective
18	18	advanced	1	1	adjective
19	19	affectionate	1	2	adjective
20	20	aged	0	0	adjective
21	21	aggravating	-1	-2	adjective
22	22	aggressive	-1	-2	adjective
23	23	agile	1	2	adjective
24	24	agitated	-1	-2	adjective
25	25	agonizing	-1	-2	adjective

Figure 1.Our new dictionary

V. EXPERIMENTS

To perform our proposed TC-CVS method, that is a combination of Term Counting and Context Valence Shifters methods, we use the Microsoft SQL SERVER 2008 R2 to store the our new combined dictionary and the results of the classification.

The Microsoft Visual Studio 2010 (C# programming language) is used to code programs to build the new combined dictionary from five dictionaries and to classify the 25000 reviews of the training set and the 25000 reviews of the testing set by TC-CVS method.

Our programs ran on a laptop with an Intel Core i3 at 1.8 GHz processor Memory 6 GB. Our Operating System is Microsoft Windows 8.

Results of 25000 reviews for training are presented on the Table 1.

TABLE 1. RESULTS OF THE 25000 REVIEWS FOR TRAINING DATASET

	Dataset	CC	IC	CT
Negative	12500	7638	4862	9.36128
Positive	12500	9668	2832	13.748
Summary	25000	17306	7694	11.55464

with:

CC : Correct Classification

IC : Incorrect Classification

CT : Classification Time (second).

TABLE 2. THE ACCURACY OF TC-CVS METHOD FOR 25000 REVIEWS FOR TRANING

PM	Class	A	P	R
TC-CVS-D	Negative	69.224%	0.61104	0.72951
	Positive		0.7744	0.6654

with:

PM: Proposed Method.

A : Accuracy

P : Precision

R : Recall

TC-CVS-D: TC-CVS method with using new combined dictionary

Results of the 25000 reviews for the testing dataset are presented on the Table 3.

]

TABLE 3. RESULTS OF THE 25000 REVIEWS FOR THE TESTING DATASET

	Dataset	CC	IC	CT
Negative	12500	7638	4862	9.36128
Positive	12500	9668	2832	13.748
Summary	25000	17306	7694	11.55464

TABLE 4. THE ACCURACY OF THE 25000 REVIEWS FOR THE TESTING DATASET

PM	Class	A	P	R
TC-CVS-D	Negative	68.984%	0.55552	0.75957
	Positive		0.82416	0.64964

The Table 5. Shows the accuracy of our proposed TC-CVS method and the accuracy of the methods in [7] and [8].

TABLE 5. COMPARE OUR COMBINING TERM-COUNTING AND CONTEXTUAL VALENCE SHIFTERS METHODS to [7], [8]

System	Class	A	P	R	F-score
No results on any dataset to test [7]					
(1)	Positive	61.1%	.599	.798	.684
	Negative		.700	.425	.529
(2)	Positive	61.2%	.600	.794	.684
	Negative		.699	.430	.533
(3)	Positive	66.5%	.667	.693	.680
	Negative		.693	.637	.664
(4)	Positive	62.8%	.604	.785	.683
	Negative		.694	.476	.565
(5)	Positive	63.0%	.606	.784	.684
	Negative		.694	.476	.565
(6)	**Positive**	**67.8%**	**.673**	**.701**	**.687**
	Negative		**.691**	**.655**	**.673**
(7)	Positive	58.9%	.882	.209	.338
	Negative		.552	.969	.703
(8)	Positive	63.4%	.606	.832	.701
	Negative		.728	.437	.546
(9)	**Positive**	**68.984%**	**.5552**	**.7596**	**.639**
	Negative	**(testing dataset)**	**.8242**	**.6496**	**.727**
(10)	**Positive**	**69.224% (training dataset)**	**.611**	**.7295**	**.665**
	Negative		**.7734**	**.6654**	**.715**

(1) Basic:GI - The system used Counting Terms method method with General Inquirer dictionary [2] in [8].

(2) Basic: GI & CTRW - The system used Counting Terms method with General Inquirer dictionary [2] and CTRW dictionary in [8].

(3) Basic: GI & CTRW & Adj - The system used Counting Terms method with General Inquirer dictionary [2], CTRW dictionary and ADJ (list of positive/negative adjectives) in [8].

(4) Enhanced: GI - The system used combining Counting Terms and Contextual Valence Shifters method with General Inquirer dictionary [2] in [8].

(5) Enhanced: GI & CTRW - The system used combining Counting Terms and Contextual Valence Shifters method with General Inquirer dictionary [2] and CTRW dictionary in [8].

(6) **Enhanced:GI & CTRW & Adj** - The system used combining Counting Terms and Contextual Valence Shifters method with General Inquirer dictionary [2], CTRW dictionary and ADJ (list of positive/negative adjectives) in [8].

(7) Enhanced: GI & SO-PMI 1 - The system used combining Counting Terms and Contextual Valence Shifters method with General Inquirer [2] and SO-PMI 1 (a longer list of positive/negative terms) in [8].

(8) Enhanced: GI & SO-PMI 2 - The system used combining Counting Terms and Contextual Valence Shifters method with General Inquirer [2] and SO-PMI 2 (a longer list of positive/negative terms, different from SO-PMI 1) in [8].

(9) **TC-CVS method with combined new dictionary** - Our proposed method, which is

a combination of Term Counting (TC) and Context Valence Shifters (CVS) methods.

(10) TC-CVS method with combined new dictionary

We assume that, a document or a paragraph has m sentences, each sentence is ended by dot (.) or exclamation mark (!) or question mark (?); each sentence in the document or a paragraph, has n word phrases; each phrase has k words or subphrases.

The complexity of our proposed TC-CVS method using new combined dictionary is: O (m * n * k).

VI. CONCLUSION

This paper proposes TC-CVS method, which is a combination of Term Counting and Context Valence Shifters methods. TC-CVS method uses new combined dictionaryto recognize and classify sentiment in the document. TC-CVS method has accuracy 68.984% of the testing dataset, and 69.224% of the training dataset.

With the same dictionary, TC-CVS method has the classification result better than Term-Counting method. For example, in Table 5, with the same General Inquirer dictionary, TC-CVS method has an accuracy 62.8%, it is better than Term-Counting method with accuracy 61.1%. With the same General Inquirer dictionary and CTRW dictionary, Our proposed TC-CVS method has an accurary 63%, it is better than Term-Counting method with accuracy 61.2%.

With the same method for sentiment classification, the different dictionaries give different results. For example, in Table 5, with the same TC-CVS method, using General Inquirer, CTRW and Adj has an accuracy 67.8%, using GI and SO-PMI 2 has an accurcy 63.4%; but using our combined dictionary has an accuracy 68.984% (testing dataset).

Our proposed TC-CVS method using our combined dictionary has the best accuracy in Table 5.

TC-CVS method depends on the dictionary and identifying valence of terms in the dictionary ,but it does not depend on the training dataset domain . So, it can classify different domains such as films, hotels, electronics, books, apparel, music, dvd..

The accuracy of the dictionary always affects the result accuracy of the emotion classifying methods. The new combined dictionary takes full advantage of five dictionaries.

Our proposed TC-CVS method with our combined dictionary has improved the sentiment result for document.

REFERENCES

[1] Large Movie Review Dataset, http://ai.stanford.edu/~amaas/data/sentiment/

[2] General Inquirer, http://www.wjh.harvard.edu/˜inquirer/homecat.htm

[3] Sentiment Lexicon (Hu and Liu,2004), http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

[4] MPQA subjectivity lexicon (Wilson,Wiebe and Hoffmann,2005) http://www.cs.pitt.edu/mpqa/subj_lexicon.html

[5] SentiWordNet (Esuli and Sebastiani,2006), http://sentiwordnet.isti.cnr.it/

[6] Emotion lexicon (Mohammad and Turne,2010), http://www.purl.org/net/emolex

[7] Livia Polanyi and Annie Zaenen, *Contextual valence shifters*, In Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text, pp 106–111, 2004

[8] A. Kennedyand D. Inkpen, *Sentiment classification of movie reviews using contextual valence shifters*. Computational Intelligence, 22(2): pp 110–125, 2006.

[9] Vivek Narayanan, Ishan Arora and Arjun Bhatia, *Fast and accurate sentiment classification using an enhanced Naive Bayes model,* Journal of Intelligent Data Engineering and Automated Learning IDEAL 2013 Lecture Notes in Computer Science, Volume 8206, pp 194-201, May 2013.

[10] Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze, *Introduction to information retrieval.* Vol. 1. Cambridge: Cambridge University Press, 2008.

[11] Yelena Mejova , *Sentiment Analysis: A Overiewe*, Computer Science Department, University of Iowa, 2009.

[12] Pang and Lee, *Opinon Mining and sentiment Analysis*, Foundations and Trends in Information Retrieval, Vol 2, No 1-2, pp 1-135, 2008

[13] Bing Liu, *Sentitment Analysis and Subjectivity,* Handbook of Natural Language Processing, Second Edition, 2010

[14] A. Abbasi, H. Chen, and A. Salem. *Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums*. ACM Trans. Inf. Syst., 26:12:1– 12:34, June 2008.

[15] Daniel Jurafsky and James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition* (2ed.), Prentice Hall, 2008.

[16] Vo Ngoc Phu, *The combined dictionary,* vongocphu03hca@gmail.com