# Emotional dictionary

We create this Vietnamese emotional dictionary which contains 5 parts: Noun dictionary, verb dictionary, adjective dictionary, adverb dictionary and intensiﬁcation dictionary. Our dictionary is essentially based on the English SO-CAL dictionary. Moreover, we added by ourselves some words to make our product consistent with the Vietnamese grammar and the concise spelling that people are using on social network.

The number of words in each dictionary of noun, verb, adjective and adverb is 1546 words, 1108 words, 2357 words , 749 words respectively and each word is paired with an integer which describes the corresponding emotional value (EV) from the most negative (-5) to the most positive (+5). Notice that no word has SO emotional value at zero value (0).

Table 1‑1 Some examples from dictionary of noun, verb, adjective and adverb

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Noun dictionary | | Verb Dictionary | | Adjective Dictionary | | Adverb dictionary | |
| Noun | EV | Verb | EV | Adjective | EV | Adverb | EV |
| hoàn hảo  (perfection) | 5 | tôn kính  (respect) | 4 | tuyệt vời  (perfect) | 5 | thú vị  (interestingly) | 5 |
| lộng lẫy  (luxury) | 4 | hoan hỉ  (delight) | 4 | cao cấp  (high-grade) | 4 | huy hoàng  (splendidly) | 4 |
| chiến thắng  (victory) | 3 | thành công  (succeed) | 3 | bổ ích  (helpful) | 3 | giỏi  (well) | 3 |
| phước lành  (blessing) | 2 | sáng tạo  (create) | 2 | chặt chẽ  (close) | 2 | tươi  (freshly) | 2 |
| độc lập  (liberty) | 1 | tăng  (increase) | 1 | hợp lý  (agreed) | 1 | sạch  (clean) | 1 |
| tội phạm  (crime) | -1 | vùi dập  (ruin) | -1 | cũ  (old) | -1 | kỳ quặc  (weirdly) | -1 |
| điểm yếu  (weakness) | -2 | xấu hổ  (shame) | -2 | đần độn  (silly) | -2 | thô  (crudely) | -2 |

The intensiﬁcation dictionary has 185 words and each word also has a accompanied decimal to demonstrate the increase or decrease of its EV.

Table 1‑2 Some examples from the intensiﬁcation dictionary

|  |  |
| --- | --- |
| Intensiﬁcation | Value |
| ít  (Slenderly) | -1.5 |
| chút ít  (Slightly) | -0.9 |
| hơi  (a little) | -0.5 |
| khá  (rather) | -0.2 |
| chắc  (surely) | 0.2 |
| Siêu  (super) | 0.4 |
| hoàn toàn  (completely) | 0.5 |

Example : If EV of “nhếch nhác” (messy) is (-3), EV of “khá nhếch nhác” (rather messy) is (-3)\*(1 – 0,1) = (-2.7). If EV of “xuất sắc” (excellent) is 5, EV of “xuất sắc nhất” (the most excellent) is 5\*(1 + 1) = 10.

# The training data

As we know, our emotion is extremely complicated. Hence to build a manageable data, we conducted collecting comments and opinions of the social network user and label those sentences to analyze them. Each comment has subjective nature of every person. The first task is to classify which comment is emotional or non-emotional (also known as subjectivity classification) and the second task is to classify which comment is negative or positive (also known as sentiment classification).

Our data source was chosen from 3 topics: Education, Movies and Sports. Each database contains from 250 to 350 comments of those topics. Consequently, we created a bigger synthetic database from 3 topics which includes 885 comments.

Firstly, we divided manually the synthetic database into 2 parts: subjective and objective sentences. Result of subjective manual classification is shown in the following table:

Table 2‑1 Result of subjective manual classification

|  |  |  |  |
| --- | --- | --- | --- |
| Number | Topic | Training data | |
| Subjective sentences | Objective sentences |
| 1 | Education | 173 | 99 |
| 2 | Movies | 194 | 95 |
| 3 | Sports | 248 | 76 |
| 4 | Summary | 615 | 270 |

After that, the subjective sentences were classified manually into 2 parts: negative and positive sentences. Result of sentiment manual classification is shown in the following table:

Table 2‑2 Result of sentiment manual classification

|  |  |  |  |
| --- | --- | --- | --- |
| Number | Topic | Training data | |
| Positive sentences | Negative sentences |
| 1 | Education | 133 | 40 |
| 2 | Movies | 115 | 79 |
| 3 | Sports | 201 | 47 |
| 4 | Summary | 449 | 166 |

# The Experimental Model

Compared with other researches related to our topic, our research has been attributes that were selected, were tested many times and emotional value in the Vietnamese emotional dictionary from -5 to +5 will help assess detail the emotional value of sentence. It helps improve accuracy in sentiment analysis for Vietnamese comments.

## Subjective classification method:

This method uses 6 attributes to classify which comment is emotional or non-emotional:

* 1st: the amount of word in the comments. It partly displays what the users want to express through the comments. If a large number of words are appeared, the user is really interested in this topic.
* 2nd, 3rd, 4th and 5th: The summary of EV of: noun, adjective, verb, adverb in the comments. The EV of a comment depends on the type of word and the EV of each word or phrase which was compared with the Vietnamese emotional dictionary. We realize the emotional value in sentence is mainly dependent on the following word types: adverb, adjective, noun and verb.
* 6th: The summary of EV of a sentence is basically total of 4 attributes that is 2nd, 3rd, 4th, 5th.

Moreover, the EV of a sentence does not only depend on the emotional words, but also depend on the type of the sentences. The emotional value of a sentence will be 0 point if this sentence is a condition sentence or a question.

## Sentiment Classification Method:

After the subjective classification method process, we continued to apply the ECM to these comments .We developed and upgraded our group of attributes which was based-on the English emotional analysis method and the consistence with traditional Vietnamese language.

* Firstly, the emotional value of a sentence depends on the emotional value of each emotional word or phrase. The most basic attributes inherited from subjective analysis. Includes: The emotional value of word kinds in sentence: adjective, adverb, noun and verb. The summary of EV of a sentence from all kinds above.
* Secondly, the emotional value of a sentence which depends on the EV of the intensiﬁcation will be calculated by this formula :

*Emotional value = Value of intensiﬁcation \* Value of emotional word*

The total of these values will be the new value of the emotion after review intensification. In the absence of intensification in sentence, this value is the total value of all kinds of emotional words in a sentence.

* Thirdly , The emotional value of a sentence which depends on the negative words: “không”, “không có”, “không phải”, “không được”, “chẳng”, “chẳng có” và “chẳng phải” will be calculated by this formula:

*Emotional value = (-1) \* Value of emotional word*

* Fourthly , The emotional value of a sentence which depends on the imperfect words : “nên” , “phải” , “có thể” ....will be calculated by the formula:

*Emotional value = (0.5) \* total value of all kinds of emotional words in a sentence*

* Fifthly, the emotional value of a positive sentence: In fact, traditional vietnamese culture , people avoid using negative words to express their opinions so that the positive words are commonly used. Hence the emotional value of a positive word will be calculated by the formula:

*Emotional value = (1 + 0.5) \* Value of positive word*

* Lastly, the emotional value of a sentence which has a contrasting-linked word likes: “nhưng”, “nhưng mà”, “mà” và “cơ mà” will be calculted by total of the EV of words that subtract the EV of the words before the contrasting – linked word.

*Emotional value = total of the EV of words – total of the EV of the words before the contrasting – linked word*

# Result and Discussion

## The result of subjective classification method

From the training data classficated manually, SVM classification method and the training data is presented above we examined the accuracy of subjective classification method. Results are presented in the following table:

Table 4‑1 Results of subjective classification method

|  |  |  |
| --- | --- | --- |
| Number | Topic | Result (precision: %) |
| 1 | Education | 92.6% |
| 2 | Movies | 89.7% |
| 3 | Sports | 89.5% |
| 4 | Summary | 89.8% |

## The result of sentiment classification method

We continue to assess the accuracy of the sentiment classification method. Results are presented in the following table:

Table 4‑2 Results of sentiment classification method

|  |  |  |
| --- | --- | --- |
| Number | Topic | Result (precision: %) |
| 1 | Education | 90.8% |
| 2 | Movies | 79.2% |
| 3 | Sports | 95.0% |
| 4 | Summary | 89.5% |