Emotion Classification by Incremental Association Language Features

Jheng-Long Wu, Pei-Chann Chang, Shih-Ling Chang, Liang-Chih Yu, Jui-Feng Yeh and Chin-Sheng Yang

***Abstract*—**The Major Depressive Disorder has been a burden of medical expense in Taiwan as well as the situation around the world. Major Depressive Disorder can be defined into different categories by previous human activities. According to machine learning, we can classify emotion in correct textual language in advance. It can help medical diagnosis to recognize the variance in Major Depressive Disorder automatically. Association language incremental is the characteristic and relationship that can discovery words in sentence. There is an overlapping-category problem for classification. In this paper, we would like to improve the performance in classification in principle of no overlapping-category problems. We present an approach that to discovery words in sentence and it can find in high frequency in the same time and can't overlap in each category, called Association Language Features by its Category (ALFC). Experimental results show that ALFC distinguish well in Major Depressive Disorder and have better performance. We also compare the approach with baseline and mutual information that use single words alone or correlation measure.

***Keywords*—**Association language features, Emotion Classification, Overlap-Category Feature, Nature Language Processing.

I. INTRODUCTION

The Global Burden of Disease 2000 (GBD2000) project shows that unipolar depressive disorders place an enormous burden on society and are ranked as the fourth leading cause of burden among all diseases. If still trends for demographic and epidemiological continue, the burden of depression disorder becoming the second in 2020.[1] Therefore, we need to attention the emotion because the social culture was changed become social relationship complex between peoples. In actuality environment, may people did not mange emotion by itself that cause it is not confidence for the future life. In fact, the emotion classification can help medical personnel effectively.

In the past the emotion recognition focused on the writer's state, but this paper mid the problem from reader's observation. For example, medical personnel diagnose is interpreted by the doctor either by patient. Many psychiatric websites have developed community-base services such as message boards, web forums and blogs for public access, form these services we can give larger corpora to analysis. And the class of sentence is identified by reader. Such psychiatric texts contain large amounts of natural language expressions related to Major Depressive Disorder, making them useful resources for building more effective psychiatric services.

Many people present some sentence on web, these data has potential information about the moment person emotion state, in this paper to recognize emotion for training very interesting. For instance, a user speak some sentences about feeling expressions on blog, we want to recognize emotion of sentence into category, moreover, response some suggestion to user.

Nature Language Processing (NLP) is a field of computer science and linguistics concerned with the interactions between computers and human languages. Text information just as nature language is suitable for express the emotion better than questionnaire quantity measure, because it contain copious information now. Nonetheless, text was very complex as nature language, through computer to assist identifiable meaning that the process such classification or summarization to create. How to use text data to analysis that needed to transform data structure to feature set that through suitable representation or extraction process. The feature set may consist of all possible language from transformation process such as Part-Of-Speech (POS) tagger or *n*-grams sequence, in this paper use the POS tagger to segmentation word that in this paper each tagger of word called "term". On the other hand, feature selection and weighting methods such as term frequency (TF) or inverse document frequency (IDF) or both combine TF×IDF are used for classification analysis [6].

Traditional approaches in sentence classification [1], [12] or text categorization [4] usually use bag-of-word to make baseline features for train text classification model, the bag-of-word treats each words independently, but and it did not consider the relationships of words in sentences. On the other hand, n-grams approach was useful to capture sequential relations between words to improve classification performance [14], [15]. The association may across one word or more in sentences, but the n-grams is capturing local dependencies of words requires large

Jheng-Long Wu is with the Department of Information Management, Yuan Ze University, Taiwan, R.O.C. (corresponding author, phone: 886-3-4638800 Ext.2768; fax: 886-3-4352077; e-mail: jlwu.yzu@gmail.com).

Pei-Chann Chang is with the Department of Information Management, Yuan Ze University, Taiwan, R.O.C. (e-mail: iepchang@saturn.yzu.edu.tw).

Shih-Ling Chang is with the Department of Information Management, Yuan Ze University, Taiwan, R.O.C. (e-mail: s976232@mail.yzu.edu.tw).

Liang-Chih Yu is with the Department of Information Management, Yuan Ze University, Taiwan, R.O.C. (e-mail: lcyu@saturn.yzu.edu.tw).

Jui-Feng Yeh is with the Department of Computer Science and Information Engineering, National Chiayi University, Taiwan, R.O.C (e-mail: ralph@mail.ncyu.edu.tw).

Chin-Sheng Yang is with the Department of Information Management, Yuan Ze University, Taiwan, R.O.C. (e-mail: csyang@saturn.yzu.edu.tw).

[1] The World Health Report 2001, WORLD HEALTH ORGANIZATION □ WHO□

TABLE I CLASSIFICATION OF MAJOR DEPRESSIVE DISORDER

Category	Description	Example sentence
Depressed	depressed mood most of the day, nearly every day, as indicated by either subjective report	It makes me very sad
Interested	markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day	It makes me disappear and can not to do anything
Weight	significant weight loss when not dieting or weight gain, or decrease or increase in appetite nearly every day	this thing makes me cry and lose 10 kg weight in a week
Insomnia	insomnia or hypersomnia nearly every day	I can't sleep and stay up all night
Energy	fatigue or loss of energy nearly every day	I find I'm so tired and weak easily
Guilt	feelings of worthlessness or excessive or inappropriate guilt nearly every day	When I heard that thing, I feel so guilty.
Concentrate	diminished ability to think or concentrate, or indecisiveness, nearly every day	I usually cannot concentrate
Death	recurrent thoughts of death ,recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide	Living is only delaying time.

training data to obtain reliable estimation, so it is not suitable capturing long-distance dependencies in data sparseness problem, just like words very little in a sentence.

For our aim, we want to incremental feature of emotion characterized by association language mining, i.e., meaningful combinations of terms, such as {spirit} and {tired} are association language features, it presents an association between terms of emotion. Table I shows the Major Depressive Disorder as category, there are 8 kinds of emotion of MDD (it is a kind of emotion as behavior retardation doesn't use, because it is difficult observation from text data) [2] according this category to develop supervised method of association mining. The question we want to answer is whether there is something to be gained by association mining of unsupervised revise to supervised learning when we have incremental feature. We want to discover the information of discrimination, according to the language type that it be not necessarily composed of continuous words, so it usually composed of the words with long-distance dependencies, yet cannot easily captured by n-grams.

The main research questions addressed in this paper are:

1) How to discovery association language features from the emotion sentences with Major Depressive Disorder automatically?
2) What's important condition for association language features to classify sentences with Major Depressive Disorder?

The remainder of this paper is organized as follows. The concept of association in nature language process helps for classification analysis in Section 2. How to mine frequency item set of words of association language features by its category is in Section 3. We apply multiple classifiers algorithms for classification with combine multiple features set in Section 4. Section 5 is presents their results and discussion. Conclusions and opportunities for future work in Section 6.

II. ASSOCIATION LANGUAGE FEATURES

The meaning in statistics is to measure the relationships between two quantities [8]. Association language is a similar type, it finds any relationship between words such as relationship between peoples, so we use the concept of association for incremental feature. For example, the association of two terms that {myself} and {feeling} is high frequency of co-occurrence in training data when people express many words. In the language domain, to understand specific words in the same emotion is an important information in text meaning. Sometimes many researches use correlation confidence measures such as Pointwise Mutual Information (PMI) or Mutual Information (MI).The front is a tool to measure degree about the association between two terms [5]. The posterior can gain good effect of classification [9]-[11].

However the association language has a problem that association mining of data mining measures high frequency of co-occurrence between terms or correlation confidence, these features may exist overlapping problem (such covariance in statistics). The accuracy drops down when overlapping categories occurred. This idea from the cluster concept single categories of association language features to the correlation of high, instead overlap clusters dependent. For example, an item set almost appeared when the itemset in two or more categories such as "A"and "B" and "D", if we find some itemsets of association language, they have high co-occurrence and correlation but don't know which association language feature is particular of category in nature language, so it doesn't help for classification analysis, it's an important concept in this paper. Thus, this paper wants to build a framework that mining association language features approach and filter overlapping feature to decrease the classification interference.

III. METHODOLOGE

Current problem is how to acquaint association language; it can be converted into the problem of association rule mining in data mining. It is a well-known mining technique that to discover interesting pattern from transactional databases. Association language from each sentence in the corpora is discovered similarly, and item in a transaction denotes term in a sentence, only the combinations of nouns and verbs are considered, and the length is fully automatic that it satisfies min co-occurrence. It uses Apriori algorithm [3], [13] and modify it slightly to fit our problem. It also discovers the association language as frequency association terms in sentence of training data to improve the classification accuracy. The association language feature is defined as a combination of multiple associated terms, denoted by $< t_1,...,t_k >$.

There are two stages to create association language features (see Fig. 1). The first part is finding frequent termsets to discover frequent termsets in each training data of categories reduplicate. The second part is to generate association language features by its category according to each frequent termsets of every category from training data and it has be removed the overlap-category frequent termsest. The detailed procedure is described as follows.

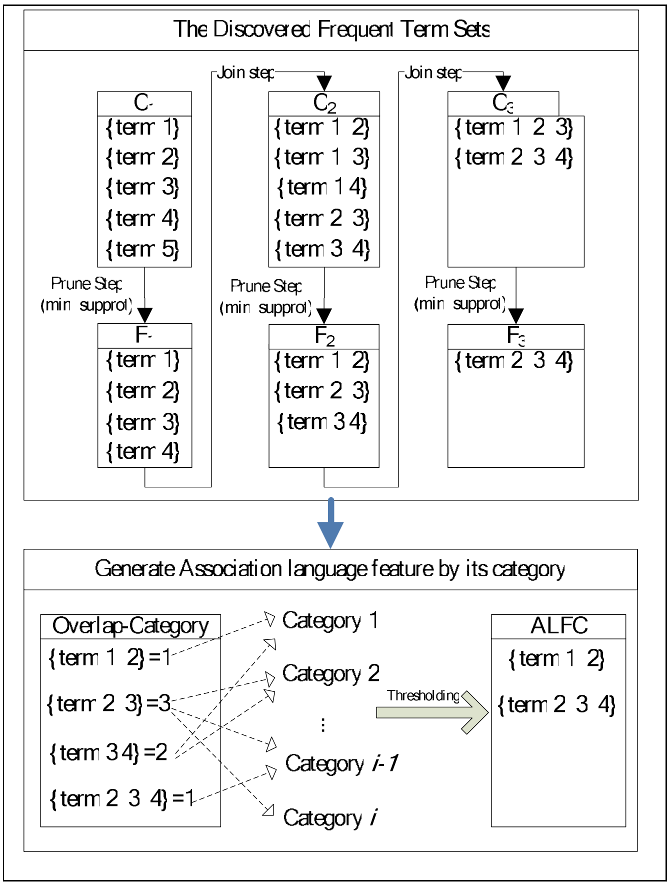

Fig. 1 Discover frequent term sets and generate association feature by its category

A. Discover frequent term sets

In this part, we want to find high frequent of co-occurrence between terms in same sentence of corpus. First, counts all the 1-termsets from the set of sentences and finds supports which exceed the minimum support (this is called the frequent 1-termsets). Second, combines the frequent 1-termsets to form candidate 2-termsets, and cunts all the candidate 2-termsets from the set of sentences and determines the frequent 2-termsets. Finally, repeats frequent k-termsets until the candidate k-termsets or the frequent is empty. The code of the algorithm is listed in Fig. 2. It repeats each category until discovering all termsets when reaching satisfying condition.

F_1 = {frequent 1 - termstes};
for ($k = 2$; $F_{k\text{-}1} \neq \varnothing$; $k++$) do begin
  C_k = appriori − gen($F_k - 1$); //generate candiates
  forall sentence $s \in D$ do begin
    $C_s = subset(C_k, s)$; //candidates contained in s
    forall candidate $c \in C_s$ do
      c.count ++;
    End;
    $F_k = \{c \in C_k \mid c.count \geq minsupport\}$;
End;
Return $\cup_k F_k$;

Fig. 2 an algorithm for discovery frequent term sets

B. Generate Association Language Features by its Category (ALFC) from frequent term sets

It is to measure the overlap of feature bases on predefined categories, association language features can be generated via an overlap measure each frequent term set occurs in raw data in each category in training data. The overlapping value of association language features in k-terms is defined as below.

$$y_j = \sum_{i}^{n} \sum_{j}^{m} x_{ij} \tag{1}$$

where

y_j denotes the count number of t words co-occurring as termset j in sentence of category in training set,

$$x_{ij} = \begin{cases} 1 & \text{if termset } j \text{ in catefoery } i \text{ occurs} \\ 0 & \text{ohterwise,} \end{cases}$$

The overlap-category was calculated in formula (1). It counts the quantity that occurs in each category. According to each frequent term set in L_k is assigned a number of overlap-category score, set a threshold if each y_j *of termsets* is 1 (that setting 1 meaning that each termset only occurs in a category) than can be Association Language Features by its Category (ALFC). Each termsets satisfies the threshold is only in particular ones category. So, feature set from association language features by its category, they are the important discrimination degree.

IV. Emotion as Classification

In this paper, we test combination of the proposed three feature types. Classification accuracy is used as the performance metric, classification technique applies three classification algorithms included Naïve bayes (NB), C4.5 and Support Vector Machine (SVM) classifier. The evaluation platform is the machine learning software toolkit WEKA [7]. These feature sets includes:

- **Bag-of-Words (BoW):** Each term in sentences corresponds to a feature after Part-Of-Speech tagger process.
- **Association Language Features by MI (ALFMI):** The top N percent association language terms acquired with original frequency features by Mutual Information

measure.

- **Association Language Features by its Category (ALFC)**: The association language terms acquired through verification of category in the previous Section 3.

V. Experiments results

In our experiments, in order to discover important information of discrimination, we needed to split data into train and test data. Train data is used for training association language features and test data is used for testing accuracy of classification. Different feature sets used to determine the experiment design. Moreover, compares difference incremental feature which affects the accuracy of classification.

A. Data Set

The total number of sentence data is 1,331 that collected from the PsychPark (http://www.psychpark.org), a virtual psychiatric clinic, maintained by a group of vol-unteer professionals of Taiwan Association of Mental Health Informatics [16]. The text type is the traditional Chinese words. In our experiment, the data sets in each category are partitioned randomly into training data, testing data with the proportion of 90%, 10% respectively. All of data were preparing for the category definition by medical personnel (reader's observation), the training data was used for language feature generation. The Table 2 shows the distribution of sentence category about the Major Depressive Disorder corpus. Furthermore, according to Diagnostic and statistical Manual of Mental Disorder (DSM-IV) referred to emotion that there are four categories most discovery of symptom as the Depressed, Interested, Insomnia and Death [2], so the sentence distribution was unevenness in the corpus.

TABLE II Distribution of sentence category

Sentence category	Distribution in Corpus %
Depressed	24.5
Interested	13
Weight	2
Insomnia	24.5
Energy	3
Guilt	5
Concentrate	3
Death	25

B. Baseline

We test three different feature set to construct our feature vector. The first feature set of baseline uses the Bag-of-Words (BoW) to construct feature vector, data segmentation by POS tagger only include nouns and verb for construct. The second feature set was BoW plus ALFMI that they did not remove any frequent termset with overlap, simple to measure correlation of termset by MI method According to [11] it selects top N percent features for classification. And through experiments to determine an optimal threshold for each classifier by maximizing its classification accuracy on the test set. Finally, feature set using BoW plus ALFC of this paper method to construct feature vector. Each classifier implements effect and compare accuracy ratio among three feature sets.

C. Results and Discussions

Table III lists the results of the classification of individual feature sets from test data. We implement the baseline just BoW feature sets for the accuracy is lower than other feature sets. The better classifier is SVM form each feature set. The BoW+ALFMI is used Mutual Information to measure correlation, a threshold at percentage is applied to select top *N* percent frequent term sets, the BoW plus ALFMI feature set improved accuracy of NB , C4.5 and SVM by 1.98%, 0.66% and 0.66% best than BoW feature set, and achieved an average improvement of 1.1%. The BoW plus ALFC (association language features by its category) was the better feature set that they improved accuracy of NB, C4.5 and SVM by 2.64%, 1.97 and 2.63 best than feature set of BoW and BoW+ALFMI, and an average improvement of 2.4% and 1.3%.

In our experiments, it discusses the effect of incremental feature for emotion classification analysis in Major Depressive Disorder. The simple association language features (only used frequent termsets) make accuracy of classification about 86.18% in SVM. Then we tried to test famous measure method as Mutual Information it makes good accuracy of classification best than simple association language features. To avoid a low discrimination that select high correlation of association language features. Our approach that Association Language Features by its Category has been best classification results, many researchers have mistakes in the correlation confidence because mined association language features correlation has overlapping state among category.

TABLE III Classification performance: accuracy of emotion classification on testing data

Feature Set	NB	C4.5	SVM
BOW	76.97	78.29	86.84
BOW+ALFMI	78.95	78.95	87.50
BOW+ALFC	79.61	80.26	89.47

VI. Conclusion

This paper has been presented an approach for emotion classification using incremental association language features. The association mining as unsupervised learning into supervised learning that each incremental feature that only exist corresponding individual of category. Therefore every classifier technique of machine leaning to acquire optimized accuracy for emotion of Major Depressive Disorder, the association language features by its category should capture important information of discrimination that increases precision very effective. Using ALFC approach that discovers the overlap outcomes, thus according experiments to confirm they are noisy feature due to the discrimination. In the future, we plan to "integrate" emotion classifier into a medical system. We would also like to improve emotion identify performance.

References

[1] A. Khoo, Y. Marom and D. Albrecht, "Experiments with Sentence Classification," In *Proc. of Australasian Language Technology Workshop*, 2006, pp. 18-25.

[2] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author, 1994.

[3] C. H. Wu, Z. J. Chuang, and Y. C. Lin, “Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models,” *ACM Trans. Asian Language Information Processing*, vol. 5, no. 2, pp. 165-182, 2006.

[4] F. Sebastiani, “Machine Learning in Automated Text Categorization,” *ACM Computing Surveys*, vol. 34, no. 1, pp. 1-47, 2002.

[5] G. Mishne, “Experiments with Mood Classification in Blog Posts” In *proc. 1st Workshop on Stylistic Analysis of Text for Information Access*, at SIGIR 2005, pp. 53–60.

[6] H. He, J., Y. Xiong, B. Chen, W. Sun and L. Zhao, “Language Feature Mining for Music Emotion Classification via Supervised Learning from Lyrics,” *ISICA 2008*, 2008, pp. 426–435.

[7] I. H. Witten and E Frank, *Data Mining: Practical Machine Learning Tools and Techniques* (2nd ed.), Elsevier, San Francisco, 2005

[8] J. Han and M. Kamber, *Data Mining: Concepts and Techniques* (2nd ed.). Morgan Kaufmann,San Francisco, 2007.

[9] L. C. Yu, C. H. Wu, and F. L. Jang, “HAL-based Cascaded Model for Variable-Length Semantic Pattern Induction from Psychiatry Web Resources,” *in Proc. of the COLING/ACL 2006*, Sydney, Australia, 2006, pp. 945–952.

[10] L. C. Yu, C. H. Wu, J. F. Yeh and Fong-Lin Jang, “HAL-based Evolutionary Inference for Pattern Induction from Psychiatry Web Resources,” *IEEE Trans. Evolutionary Computation*, vol. 12, no. 2, 2008, pp. 160-170.

[11] L.-C. Yu, C.-L. Chan, C.-H. Wu and C.-C. Lin, “Mining Association Language Patterns for Negative Life Event Classification,” In *Proc. ACL-IJCNLP 2009 Conference Short Papers*, Singapore, 2009, pp. 201-204.

[12] M. Naughton, N. Stokes, and J. Carthy, “Investigating Statistical Techniques for Sentence-Level Event Classification,” In *Proc. of COLING-08*, 2008, pp. 617-624.

[13] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules,” In *Proc. Int’l Conf. Very Large Data Bases (VLDB)*, 1994. pp. 487-499.

[14] R. Chitturi and J. H. L. Hansen, “Dialect Classification for online podcasts fusing Acoustic and Language based Structural and Semantic Information,” In *Proc. of ACL-08*, 2008, pp. 21-24.

[15] S. Li and C. Zong, “Multi-domain Sentiment Classification, ” In *Proc. of ACL-08*, 2008, pp. 257-260.

[16] Y. M. Bai, C. C. Lin, J. Y. Chen, and W. C. Liu, “Virtual Psychiatric Clinics,” *American Journal of Psychiatry*, vol. 158, no. 7, pp. 1160-1161, 2001.