

Mining Association Language Patterns for Negative Life Event Classification

Liang-Chih Yu¹, Chien-Lung Chan¹, Chung-Hsien Wu² and Chao-Cheng Lin³

¹Department of Information Management, Yuan Ze University, Taiwan, R.O.C.

²Department of CSIE, National Cheng Kung University, Taiwan, R.O.C.

³Department of Psychiatry, National Taiwan University Hospital, Taiwan, R.O.C.

{lcyu, clchan}@saturn.yzu.edu.tw, chwu@csie.ncku.edu.tw, linchri@gmail.com

Abstract

Negative life events, such as death of a family member, argument with a spouse and loss of a job, play an important role in triggering depressive episodes. Therefore, it is worth to develop psychiatric services that can automatically identify such events. In this paper, we propose the use of association language patterns, i.e., meaningful combinations of words (e.g., <loss, job>), as features to classify sentences with negative life events into predefined categories (e.g., Family, Love, Work). The language patterns are discovered using a data mining algorithm, called association pattern mining, by incrementally associating frequently co-occurred words in the sentences annotated with negative life events. The discovered patterns are then combined with single words to train classifiers. Experimental results show that association language patterns are significant features, thus yielding better performance than the baseline system using single words alone.

1 Introduction

With the increased incidence of depressive disorders, many psychiatric websites have developed community-based services such as message boards, web forums and blogs for public access. Through these services, individuals can describe their stressful or negative life events such as death of a family member, argument with a spouse and loss of a job, along with depressive symptoms, such as depressive mood, suicidal tendencies and anxiety. Such psychiatric texts (e.g., forum posts) contain large amounts of natural language expressions related to negative life events, making them useful resources for build-

ing more effective psychiatric services. For instance, a psychiatric retrieval service can retrieve relevant forum or blog posts according to the negative life events experienced by users so that they can be aware that they are not alone because many people have suffered from the same or similar problems. The users can then create a community discussion to share their experiences with each other. Additionally, a dialog system can generate supportive responses like “Don’t worry”, “That’s really sad” and “Cheer up” if it can understand the negative life events embedded in the example sentences shown in Table 1. Therefore, this study proposes a framework for negative life event classification. We formulate this problem as a sentence classification task; that is, classify sentences according to the type of negative life events within them. The class labels used herein are presented in Table 1, which are derived from Brostedt and Pedersen (2003).

Traditional approaches to sentence classification (Khoo et al., 2006; Naughton et al., 2008) or text categorization (Sebastiani 2002) usually adopt bag-of-words as baseline features to train classifiers. Since the bag-of-words approach treats each word independently without considering the relationships of words in sentences, some researchers have investigated the use of n -grams to capture sequential relations between words to boost classification performance (Chitturi and Hansen, 2008; Li and Zong, 2008). The use of n -grams is effective in capturing local dependencies of words, but tends to suffer from data sparseness problem in capturing long-distance dependencies since higher-order n -grams require large training data to obtain reliable estimation. For our task, the expressions of negative life events can be characterized by *association language patterns*, i.e., meaningful combinations of words, such as <worry, children, health>, <break up, boyfriend>, <argue, friend>, <loss, job>, and

| Label | Description | Example Sentence |
|--------|---|--|
| Family | Serious illness of a family member; Son or daughter leaving home | I am very worried about my children's health . |
| Love | Spouse/mate engaged in infidelity; Broke up with a boyfriend or girlfriend | I broke up with my dear but cruel boyfriend recently. |
| School | Examination failed or grade dropped; Unable to enter/stay in school | I hate to go to school because my teacher always blames me. |
| Work | Laid off or fired from a job; Demotion and salary reduction | I lost my job in this economic recession a few months ago. |
| Social | Substantial conflicts with a friend; Difficulties in social activities | I argued with my best friend and was upset. |

Table 1. Classification of negative life events.

<school, teacher, blame> in the example sentences in Table 1. Such language patterns are not necessarily composed of continuous words. Instead, they are usually composed of the words with long-distance dependencies, which cannot be easily captured by n -grams.

Therefore, the aim of this study is two-fold: (1) to automatically discover association language patterns from the sentences annotated with negative life events; and (2) to classify sentences with negative life events using the discovered patterns. To discover association language patterns, we incorporate the measure mutual information (MI) into a data mining algorithm, called *association pattern mining*, to incrementally derive frequently co-occurred words in sentences (Section 2). The discovered patterns are then combined with single words as features to train classifiers for negative life event classification (Section 3). Experimental results are presented in Section 4. Conclusions are finally drawn in Section 5.

2 Association Language Pattern Mining

The problem of language pattern acquisition can be converted into the problem of association pattern mining, where each sales transaction in a database can be considered as a sentence in the corpora, and each item in a transaction denotes a word in a sentence. An association language pattern is defined herein as a combination of multiple associated words, denoted by $\langle w_1, \dots, w_k \rangle$. Thus, the task of association pattern mining is to mine the language patterns of frequently associated words from the training sentences. For this purpose, we adopt the Apriori algorithm (Agrawal and Srikant, 1994) and modified it slightly to fit our application. Its basic concept is to identify frequent word sets recursively, and

then generate association language patterns from the frequent word sets. For simplicity, only the combinations of nouns and verbs are considered, and the length is restricted to at most 4 words, i.e., 2-word, 3-word and 4-word combinations. The detailed procedure is described as follows.

2.1 Find frequent word sets

A word set is frequent if it possesses a minimum support. The support of a word set is defined as the number of training sentences containing the word set. For instance, the support of a two-word set $\{w_i, w_j\}$ denotes the number of training sentences containing the word pair (w_i, w_j) . The frequent k -word sets are discovered from $(k-1)$ -word sets. First, the support of each word, i.e., word frequency, in the training corpus is counted. The set of frequent one-word sets, denoted as L_1 , is then generated by choosing the words with a minimum support level. To calculate L_k , the following two-step process is performed iteratively until no more frequent k -word sets are found.

- **Join step:** A set of candidate k -word sets, denoted as C_k , is first generated by merging frequent word sets of L_{k-1} , in which only the word sets whose first $(k-2)$ words are identical can be merged.
- **Prune step:** The support of each candidate word set in C_k is then counted to determine which candidate word sets are frequent. Finally, the candidate word sets with a support count greater than or equal to the minimum support are considered to form L_k . The candidate word sets with a subset that is not frequent are eliminated. Figure 1 shows an example of generating L_k .

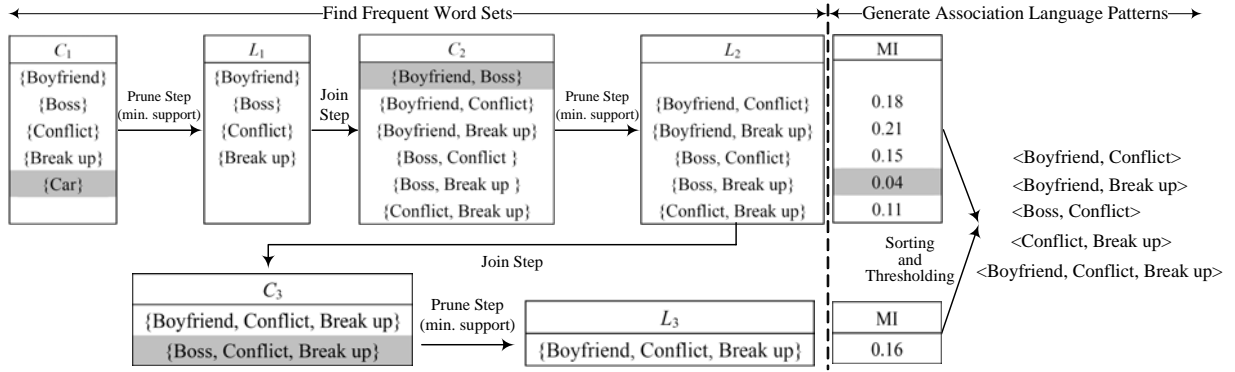


Figure 1. Example of generating association language patterns.

2.2 Generate association patterns from frequent word sets

Association language patterns can be generated via a confidence measure once the frequent word sets have been identified. The confidence of an association language pattern of k words is defined as the mutual information of the k words, as shown below.

$$\begin{aligned} Conf(\langle w_1, \dots, w_k \rangle) &= MI(w_1, \dots, w_k) \\ &= P(w_1, \dots, w_k) \log \frac{P(w_1, \dots, w_k)}{\prod_{i=1}^k P(w_i)} \quad (1) \end{aligned}$$

where $P(w_1, \dots, w_k)$ denotes the probability of the k words co-occurring in a sentence in the training set, and $P(w_i)$ denotes the probability of a single word occurring in the training set. Accordingly, each frequent word set in L_k is assigned a mutual information score. In order to generate a set of association language patterns, all frequent word sets are sorted in the descending order of the mutual information scores. The minimum confidence (a threshold at percentage) is then applied to select top N percent frequent word sets as the resulting language patterns. This threshold is determined empirically by maximizing classification performance (Section 4). Figure 1 (right-hand side) shows an example of generating the association language patterns from L_k .

3 Sentence Classification

The classifiers used in this study include Support Vector Machine (SVM), C4.5, and Naïve Bayes (NB) classifier, which is provided by Weka Package (Witten and Frank, 2005). The feature set includes:

Bag-of-Words (BOW): Each single word in sentences.

Association language patterns (ALP): The top N percent association language patterns acquired in the previous section.

Ontology expansion (Onto): The top N percent association language patterns are expanded by mapping the constituent words into their synonyms. For example, the pattern <boss, conflict> can be expanded as <chief, conflict> since the words *boss* and *chief* are synonyms. Here we use the HowNet (<http://www.keenage.com>), a Chinese lexical ontology, for pattern expansion.

4 Experimental Results

Data set: A total of 2,856 sentences were collected from the Internet-based Self-assessment Program for Depression (ISP-D) database of the PsychPark (<http://www.psychpark.org>), a virtual psychiatric clinic, maintained by a group of volunteer professionals of Taiwan Association of Mental Health Informatics (Bai et al., 2001). Each sentence was then annotated by trained annotators with one of the five types of negative life events. Table 2 shows the break-down of the distribution of sentence types.

The data set was randomly split into a training set, a development set, and a test set with an 8:1:1 ratio. The training set was used for language pattern generation. The development set was used to optimize the threshold (Section 2.2) for the classifiers (SVM, C4.5 and NB). Each classifier was implemented using three different levels of features, namely BOW, BOW+ALP,

| Sentence Type | % in Corpus |
|---------------|-------------|
| Family | 28.8 |
| Love | 22.8 |
| School | 13.3 |
| Work | 14.3 |
| Social | 20.8 |

Table 2. Distribution of sentence types.

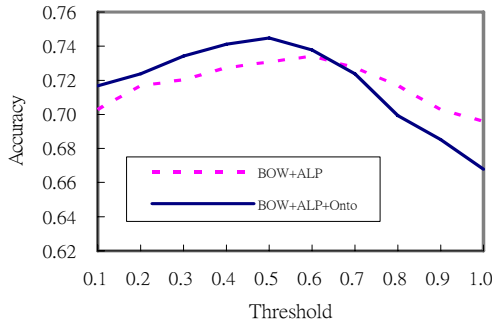


Figure 2. Threshold selection.

and BOW+ALP+Onto, to examine the effectiveness of association language patterns. The classification performance is measured by *accuracy*, i.e., the number of correctly classified sentences divided by the total number of test sentences.

4.1 Evaluation on threshold selection

Since not all discovered association language patterns contribute to the classification task, the threshold described in Section 2.2 is used to select top N percent patterns for classification. This experiment is to determine an optimal threshold for each involved classifier by maximizing its classification accuracy on the development set. Figure 2 shows the classification accuracy of NB against different threshold values.

When using association language patterns as features (BOW+ALP), the accuracy increased with increasing the threshold value up to 0.6, indicating that the top 60% discovered patterns contained more useful patterns for classification. By contrast, the accuracy decreased when the threshold value was above 0.6, indicating that the remaining 40% contained more noisy patterns that may increase the ambiguity in classification. When using the ontology expansion approach (BOW+ALP+Onto), both the number and diversity of discovered patterns are increased. Therefore, the accuracy was improved and the optimal accuracy was achieved at 0.5. However, the accuracy dropped significantly when the threshold value was above 0.5. This finding indicates that expansion on noisy patterns may produce more noisy patterns and thus decrease performance.

4.2 Results of classification performance

The results of each classifier were obtained from the test set using its own threshold optimized in the previous section. Table 3 shows the comparative results of different classifiers with different levels of features. The incorporation of association language patterns improved the accuracy of NB, C4.5, and SVM by 3.9%, 1.9%, and 2.2%,

| | NB | C4.5 | SVM |
|--------------|-------|-------|-------|
| BOW | 0.717 | 0.741 | 0.787 |
| BOW+ALP | 0.745 | 0.755 | 0.804 |
| BOW+ALP+Onto | 0.759 | 0.766 | 0.815 |

Table 3. Accuracy of classifiers on testing data.

respectively, and achieved an average improvement of 2.7%. Additionally, the use of ontology expansion can further improve the performance by 1.6% in average. This finding indicates that association language patterns are significant features for negative life event classification.

5 Conclusion

This work has presented a framework that uses a data mining algorithm and ontology expansion method to acquire association language patterns for negative life event classification. The association language patterns can capture word relationships in sentences, thus yielding higher performance than the baseline system using single words alone. Future work will focus on devising a semi-supervised or unsupervised method for language pattern acquisition from web resources so as to reduce reliance on annotated corpora.

References

- R. Agrawal and R. Srikant. 1994. Fast Algorithms for Mining Association Rules. In *Proc. Int'l Conf. Very Large Data Bases (VLDB)*, pages 487-499.
- Y. M. Bai, C. C. Lin, J. Y. Chen, and W. C. Liu. 2001. Virtual Psychiatric Clinics. *American Journal of Psychiatry*, vol. 158, no. 7, pp. 1160-1161.
- E. M. Brostedt and N. L. Pedersen. 2003. Stressful Life Events and Affective Illness. *Acta Psychiatrica Scandinavica*, vol. 107, pp. 208-215.
- R. Chitturi and J. H.L. Hansen. 2008. Dialect Classification for online podcasts fusing Acoustic and Language based Structural and Semantic Information. In *Proc. of ACL-08*, pages 21-24.
- A. Khoo, Y. Marom and D. Albrecht. 2006. Experiments with Sentence Classification. In *Proc. of Australasian Language Technology Workshop*, pages 18-25.
- S. Li and C. Zong. 2008. Multi-domain Sentiment Classification. In *Proc. of ACL-08*, pages 257-260.
- M. Naughton, N. Stokes, and J. Carthy. 2008. Investigating Statistical Techniques for Sentence-Level Event Classification. In *Proc. of COLING-08*, pages 617-624.
- F. Sebastiani. 2002. Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, vol. 34, no. 1, pp. 1-47.
- I. H. Witten and E. Frank. 2005. Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition, Morgan Kaufmann, San Francisco.