

Identifying Emotion Labels from Psychiatric Social Texts Using Independent Component Analysis

Liang-Chih Yu^{1,2} and Chun-Yuan Ho¹

¹Department of Information Management

²Innovation Center for Big Data and Digital Convergence

Yuan Ze University, Chung-Li, Taiwan

lcyu@saturn.yzu.edu.tw, s986304@mail.yzu.edu.tw

Abstract

Accessing the web has been an efficient and effective means to acquire self-help knowledge when suffering from depressive problems. Many mental health websites have developed community-based services such as web forums and blogs for Internet users to share their depressive problems with other users and health professionals. Other users or health professionals can then make recommendations in response to these problems. Such communications produce a large number of documents called psychiatric social texts containing rich emotion labels representing different depressive problems. Automatically identify such emotion labels can make online psychiatric services more effective. This study proposes a framework combining *latent semantic analysis (LSA)* and *independent component analysis (ICA)* to extract concept-level features for emotion label identification. LSA is used to discover latent concepts that do not frequently occur in psychiatric social texts, and ICA is used to extract independent components by minimizing the term dependence among the concepts. By combining LSA and ICA, more useful latent concepts can be discovered for different emotion labels, and the dependence between them can also be minimized. The discriminant power of classifiers can thus be improved by training them on the independent components with minimized term overlap. Experimental results show that the use of concept-level features yielded better performance than the use of word-level features. Additionally, combining LSA and ICA improved the performance of using each LSA and ICA alone.

1 Introduction

Sentiment analysis has been successfully applied for many applications (Picard, 1997; Pang and Lee, 2008; Calvo and D'Mello, 2010; Liu, 2012; Johansson and Moschitti, 2013; Balahur et al., 2014). Analysis of online psychiatric or mental health texts (Wu et al., 2005; Yu et al., 2009) is also an emerging field that could benefit from sentiment analysis techniques because more and more people search for help from the web when they suffered from depressive problems, which boost the development of online community-based services for Internet users to share their depressive problems with other users and health professionals. Through these services, individuals can describe their depressive symptoms via web forums and blogs. Other users or health professionals can then make recommendations in response to these problems. Figure 1 shows an example psychiatric social text collected from PsychPark (<http://www.psychpark.org>), a virtual psychiatric clinic, maintained by a group of volunteer professionals belonging to the Taiwan Association of Mental Health Informatics (Bai et al., 2001; Lin et al., 2003).

This example shows a subject's depressive problems and the responses recommended by the experts. Some meaningful tags called emotion labels herein are also annotated by the experts to indicate which categories the text belongs to. These emotion labels are useful information and can make online psychiatric services more effective. For instance, psychiatric retrieval systems are able to retrieve relevant documents according to the depressive problems (emotion labels) described in user queries so that the

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

<i>User Problem:</i>	
I broke up with my dear but cruel boyfriend recently. Since then, I have often felt like crying out of nowhere, and I feel pain every day. Also, it takes me a long time to fall asleep at night. So, I think that continuing to live like this is meaningless	<Depression> <Insomnia> <Suicide>
<i>Recommendation:</i>	
Feeling this way is normal when going through these kinds of struggles, but over time your emotions should level out. Suicide doesn't solve anything; think about how it would affect your family. It's only when we learn to face our despair that we can learn the value of life, and also how to help other people. There are a few things you can try to help you get to sleep at night, like doing some light exercise in the evening, drinking warm milk, and listening to relaxing music; all of these can be conducive to sleep. If you still have trouble dealing with the pain, and you feel as if your mood is getting worse, it wouldn't hurt to get seek help from a healthcare professional, who can help you work through your emotions.	
<i>Emotion Label:</i> <Depression>, <Insomnia>, <Suicide>	

Figure 1. Example of a psychiatric social text.

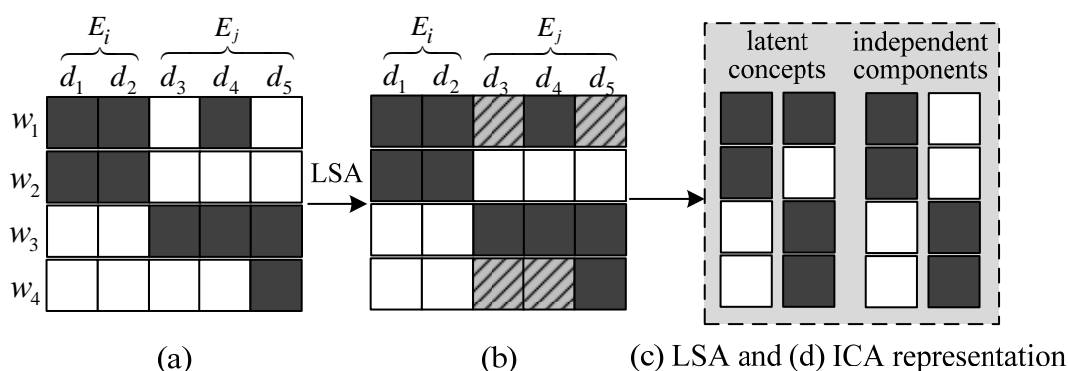


Figure 2. Comparison of LSA and ICA for feature representation.

users can learn self-help knowledge from the responses. Therefore, this study aims to identify emotion labels from psychiatric social texts. We cast this problem into a multi-label text classification task because a psychiatric social text may contain multiple emotion labels. Additionally, we propose the use of concept-level features to build classifiers instead of using surface-level features such as words, n-grams and dependency structure commonly used in the previous studies (Naughton et al., 2008; Chituri and Hansen, 2008; Li and Zong, 2008; Kessler and Schütze, 2012; Post and Bergsma, 2013; Yu et al., 2011).

In extraction of concept-level features, latent semantic analysis (LSA) (Landauer et al., 1998) has been demonstrated its effectiveness in exploring the latent structure from a collection of documents. It uses singular value decomposition (SVD) (Golub and Van Loan, 1996) to discover latent features that do not frequently occur in the documents through the indirect associations between words and documents. Figure 2 shows an example. The original matrix, as shown in Figure 2(a), is built using five documents with two different emotion labels E_i and E_j . Suppose that the words w_1, w_2 are the useful features for E_i , and w_3, w_4 are useful for E_j , but w_4 is a latent feature because it does not frequently occur in the documents of E_j . After applying SVD, the latent features can be identified by replacing the zero entries in the original matrix with non-zero real values through the indirect associations between words and documents. For instance, w_4 originally does not occur in d_3 and d_4 , but it does co-occur with w_3 in the matrix (e.g., in d_5), which means that w_4 might also occur in the documents where w_3 occurs

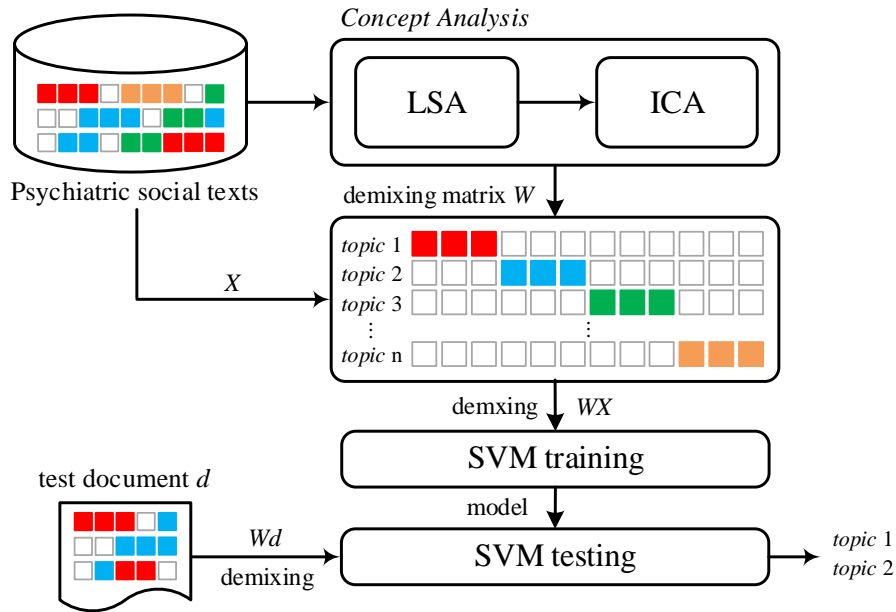


Figure 3. Framework of emotion label identification.

(e.g., d_3 and d_4). Therefore, the zero entries (w_4, d_3) and (w_4, d_4) are replaced with a non-zero value through the indirect associations between w_3 and w_4 in d_5 , as shown in Figure 2(b). This helps identify a useful latent feature w_4 for E_j . However, identifying latent features through the indirect associations cannot avoid feature overlap when different emotion labels share common words. For instance, in Figure 2(a), w_1 , which is useful for E_i , still occurs in the document of E_j (e.g., d_4). Through the indirect associations between w_1 and w_3 in d_4 , the frequency of w_1 increases in the document of E_j because it may also occur in the documents where w_3 occurs (e.g., d_3 and d_5), as shown in Figure 2(b). Therefore, when all word features are to be accommodated in a low-dimensional space reduced by SVD, term overlap may occur between the latent concepts. As indicated in Figure 2(c), the two sample latent concepts which contribute to two different emotion labels share a common feature w_1 . Classifiers trained on such latent vectors with term overlap may decrease classification performance.

To reduce the term overlap among concepts, we used the *independent component analysis* (ICA) (Lee, 1998; Hyvärinen et al., 2001; Naik and Kumar, 2011) because it can extract independent components from a mixture of signals and has been used in various text applications (Kolenda and Hansen, 2000; Rapp, 2004; Honkela et al., 2010; Yu and Chien, 2013). For our task, the psychiatric social texts are a mixture of emotion labels, which can be separated by ICA to obtain a set of independent components (concepts) with minimized term dependency for different emotion labels. Instead of using ICA alone, we propose a framework combining LSA and ICA for emotion label identification. The LSA is used to discover latent features that do not frequently occur in psychiatric texts, and ICA is used to further minimize the dependence of the latent features such that overlapped features can be removed, as presented in Figure 2(d). Based on this combination, the proposed framework can discover more useful latent features for different emotion labels, and the dependence between them can also be minimized. The discriminant power of classifiers can thus be improved by training them on the independent components with minimized term overlap. In experiments, we evaluate the proposed method to determine whether the use of concept-level features could improve the classification performance, and determine whether the combination method could improve the performance of using each LSA and ICA alone.

The rest of this paper is organized as follows. Section 2 describes the overall framework including LSA and ICA for emotion label identification. Section 3 summarizes comparative results. Conclusions are finally drawn in Section 4.

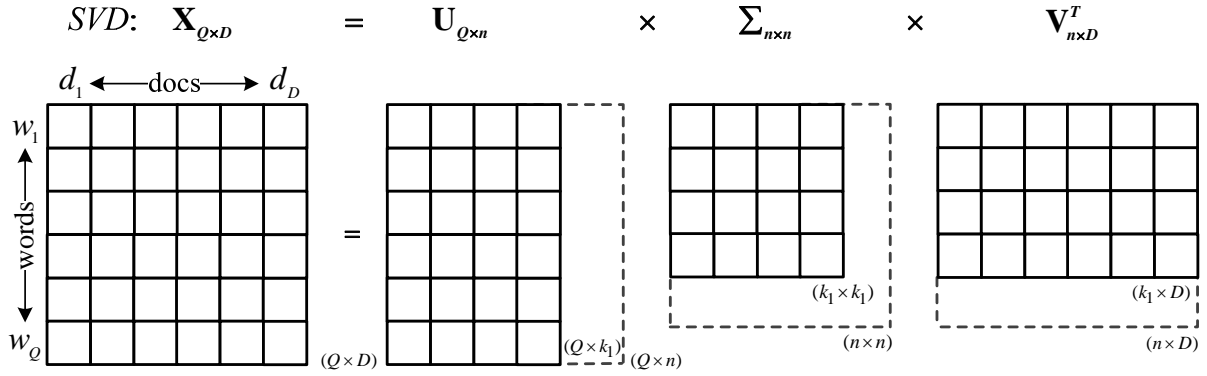


Figure 4. Illustrative example of singular value decomposition for latent semantic analysis.

2 Framework of Emotion Label Identification

Figure 3 shows the overall framework for emotion label identification. A corpus of psychiatric social texts with annotation of emotion labels are first collected from the web. This corpus which is a mixture of different emotion labels is then sequentially analyzed by LSA and ICA to generate a demixing matrix composed of a set of concepts with minimized term dependency for different emotion labels. The demixing matrix is used to separate the psychiatric social texts with mixed emotion labels into independent components for building a support vector machine (SVM) classifier. The classifier can then be benefit from the independent components to identity multiple emotion labels contained in each test example.

2.1 Latent Semantic Analysis (LSA)

LSA is a technique for analyzing the relationships between words and documents. For our task, LSA is used to identify useful latent concepts for emotion labels through indirect associations between words and documents. The first step in LSA is to build a word-by-document matrix from a corpus of psychiatric texts with different emotion labels, as shown in the sample matrix \mathbf{X} in Figure 4.

The columns in $\mathbf{X}_{Q \times D}$ represent D psychiatric texts in the corpus, and the rows represent Q distinct words occurring in the corpus. Singular value decomposition (SVD) is then used to decompose the matrix $\mathbf{X}_{Q \times D}$ into three matrices as follows:

$$\mathbf{X}_{Q \times D} = \mathbf{U}_{Q \times n} \mathbf{\Sigma}_{n \times n} \mathbf{V}_{n \times D}^T, \quad (1)$$

where \mathbf{U} and \mathbf{V} respectively consist of a set of latent vectors of words and documents, $\mathbf{\Sigma}$ is a diagonal matrix of singular values, and $n = \min(Q, D)$ denotes the dimensionality of the latent semantic space. Each element in \mathbf{U} represents the weight of a word, and the higher-weighted words are the useful features for the emotion labels. By selecting the largest k_1 ($\leq n$) singular values together with the first k_1 columns of \mathbf{U} and \mathbf{V} , the word and documents can be represented in a low-dimensional latent semantic space. The matrix $\mathbf{V}_{n \times D}^T$ can then represented with the reduced dimensions, as shown in Eq. (2).

$$\mathbf{V}_{k_1 \times D}^T = \mathbf{\Sigma}_{k_1 \times k_1}^{-1} \mathbf{U}_{k_1 \times Q}^T \mathbf{X}_{Q \times D}, \quad (2)$$

In SVM training and testing, each input psychiatric text first transformed into the latent semantic representation as follows:

$$\hat{\mathbf{t}}_{k_1 \times 1} = \mathbf{\Sigma}_{k_1 \times k_1}^{-1} \mathbf{U}_{k_1 \times Q}^T \mathbf{t}_{Q \times 1}, \quad (3)$$

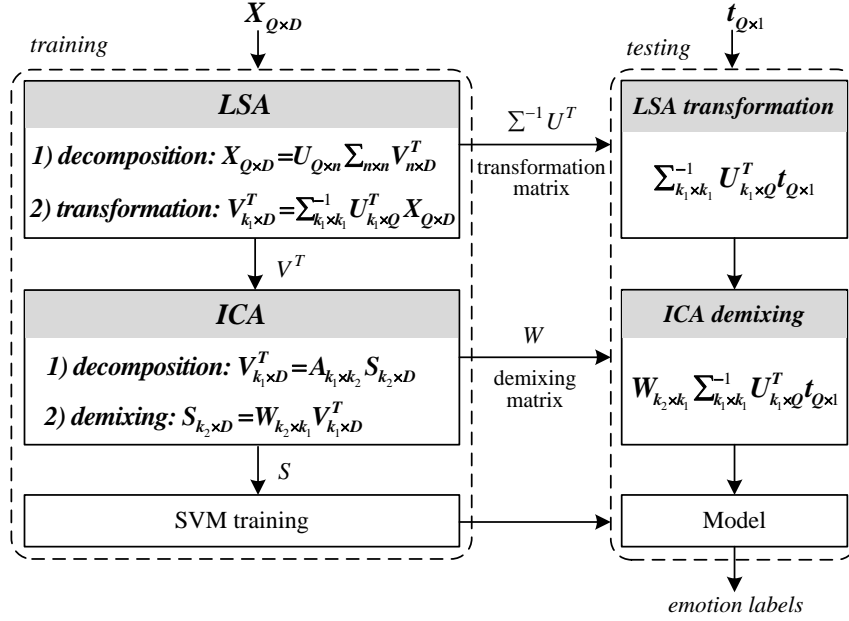


Figure 5. ICA-based method for emotion label identification.

where $\mathbf{t}_{Q \times 1}$ denotes the vector representation of an input instance, and $\hat{\mathbf{t}}_{k_1 \times 1}$ denotes the transformed vector in the latent semantic space. An SVM classifier is then trained with the transformed training vectors.

2.2 Independent Component Analysis (ICA)

ICA is a technique for extracting independent components from a mixture of signals and has been successfully applied to solve the blind source separation problem (Saruwatari et al., 2006; Chien and Hsieh, 2012). The ICA model can be formally described as

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (4)$$

where \mathbf{X} denotes the observed mixture signals, \mathbf{A} denotes a mixing matrix, and \mathbf{S} denotes the independent components. The goal of ICA is to estimate both \mathbf{A} and \mathbf{S} . Once the mixing matrix \mathbf{A} is estimated, the demixing matrix can be obtained by $\mathbf{W} = \mathbf{A}^{-1}$, and Eq. (4) can be re-written as

$$\mathbf{S} = \mathbf{W}\mathbf{X} \quad (5)$$

That is, the observed mixture signals can be separated into independent components using the demixing matrix. For our problem, psychiatric texts can be considered as a mixture of signals because each of them may contain multiple emotion labels. Therefore, ICA used herein is to estimate the demixing matrix so that it can separate the psychiatric texts with mixed emotion labels to derive the independent components for each emotion label. Figure 5 shows the diagram of the proposed method.

2.1.1 LSA decomposition and transformation

In the training phase, the original matrix $\mathbf{X}_{Q \times D}$ is first processed by SVD using Eq. (1) and (2) Useful latent features that do not frequently occur in the original matrix can thus be discovered in this step.

2.1.2 ICA decomposition and demixing

The matrix $\mathbf{V}_{k_1 \times D}^T$ decomposed by SVD is then passed to ICA to estimate the demixing matrix. ICA accomplishes this by decomposing $\mathbf{V}_{k_1 \times D}^T$ using Eq. (6). Figure 6 shows an example of the decomposition.

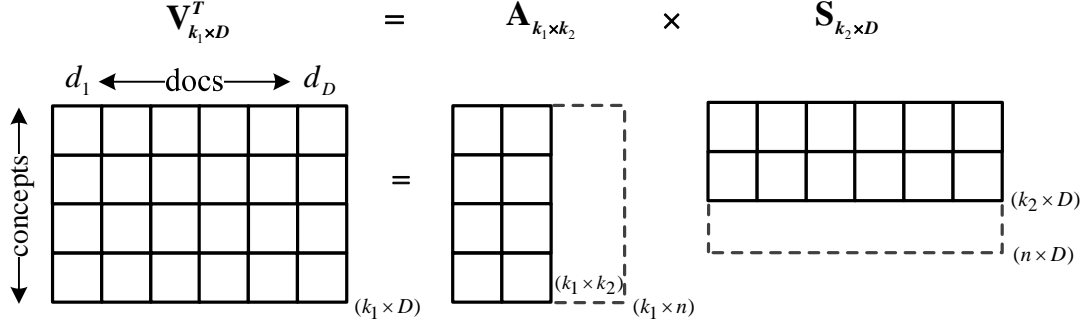


Figure 6. Illustrative example of ICA decomposition.

$$\mathbf{V}_{k_1 \times D}^T = \mathbf{A}_{k_1 \times k_2} \mathbf{S}_{k_2 \times D}. \quad (6)$$

Based on this decomposition, the demixing matrix can be obtained by $\mathbf{W}_{k_2 \times k_1} = \mathbf{A}_{k_1 \times k_2}^{-1}$, where k_2 ($\leq k_1$) is the number of independent components. The demixing matrix is then used to separate $\mathbf{V}_{k_1 \times D}^T$ to derive the independent components as follows:

$$\mathbf{S}_{k_2 \times D} = \mathbf{W}_{k_2 \times k_1} \mathbf{V}_{k_1 \times D}^T, \quad (7)$$

An SVM classifier is then trained with the independent components $\mathbf{S}_{k_2 \times D}$, as shown in Figure 5. In testing, each test instance $\mathbf{t}_{Q \times 1}$ is transformed using both LSA and ICA, and then predicted with the trained SVM model.

3 Experimental Results

3.1 Experiment Setup

3.1.1 Data

The data set used for experiments included 1,711 Chinese psychiatric social texts collected from the PsychPark. Each psychiatric social text was manually annotated with an emotion label by a group of volunteer mental health professionals. Table 1 shows the proportions of the emotion labels in the corpus. In calculating the proportion of each emotion label, a psychiatric social text was counted for multiple emotion labels depending on the number of emotion labels contained in it. In evaluation, 20% of psychiatric social texts in the corpus were randomly selected as a test set, and the remaining 80% were used for training.

No.	Emotion Label	Proportion
1	Depression	35.26%
2	Drug	13.38%
3	Insomnia	5.79%
4	Mood	30.04%
5	OCD (Obsessive compulsive disorder)	4.51%
6	Schizophrenia	5.36%
7	Social Anxiety	5.65%

Table 1. Distribution of emotion labels in experimental data

3.1.2 Classifiers

The classifiers involved in this experiment included PureSVM, LSA, ICA, and LSA+ICA. The PureSVM was trained on word-level features, and the others were trained on concept-level features derived using LSA, ICA, and combination of them, respectively. The implementation details for each classifier are as follows:

- **PureSVM:** An SVM classifier trained with bag-of-words features.
- **LSA:** An SVM classifier trained with the latent vectors obtained from the word-by-document matrix built from the training corpus.
- **ICA:** An SVM classifier trained with the independent components obtained by demixing the word-by-document matrix built from the training corpus.
- **LSA+ICA:** An SVM classifier trained with the independent components obtained by demixing the word-by-document matrix produced by LSA.

To identify multiple emotion labels contained in test examples, each emotion label presented in Table 1 was trained a binary classifier in the training phase. That is, for each method presented above, we built seven binary classifiers so that they can output multiple positive results to indicate that a test example contained multiple emotion labels.

3.1.3 Evaluation Metrics

The metrics used for performance evaluation included *recall*, *precision*, and *F-measure*, respectively. Recall was defined as the number of emotion labels correctly identified by the method divided by the total number of emotion labels in the test set. Precision was defined as the number of emotion labels correctly identified by the method divided by the number of emotion labels identified by the method. The F-measure (F1) was defined as $2 \times recall \times precision / (recall + precision)$.

3.2 Evaluation of LSA and ICA

This experiment compared the performance of LSA and ICA using different settings for the parameters k_1 and k_2 , which respectively represent the dimensionality of the latent semantic space and the

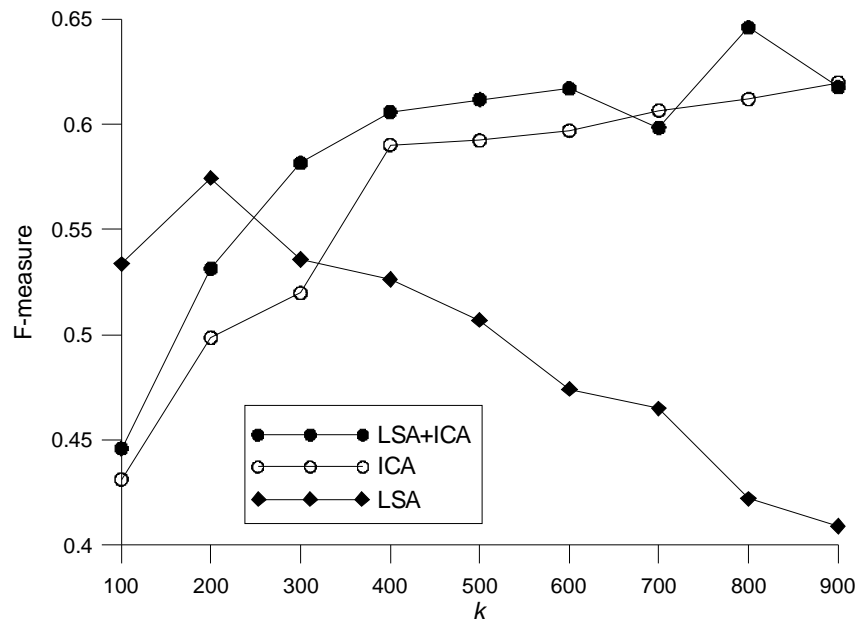


Figure 7. Performance of the LSA, ICA and LSA+ICA, as a function of k .

number of independent components. Figure 7 shows the F-measure of LSA, ICA, and combination of them with the setting $k = k_1 = k_2$. The F-measure is the average F-measure over the seven emotion labels. The results show that the optimal settings of LSA was $k=200$. The performance of LSA dropped dramatically as $k>200$, indicating that most useful latent features were discovered within the first 200 concepts and the remaining concepts may contain noisy features, thus reducing performance. The respective optimal settings for ICA and LSA+ICA were $k=900$ and $k=800$. In addition, both ICA and LSA+ICA outperformed LSA for most settings of k . The best settings of the parameters were used in the following experiments.

3.3 Comparative Results

This section reports the classification performance of PureSVM, LSA, ICA, and LSA+ICA. Table 2 shows the comparative results. Compared to the use of word-level features (i.e., PureSVM), LSA, ICA, and LSA+ICA achieved a higher F-measure. Additionally, LSA yielded a much greater recall than did PureSVM, whereas ICA yielded much greater precision. These findings indicate that the concept-level features are useful for emotion label identification. Among the three concept-based methods, LSA can discover latent concepts for emotion labels, whereas ICA can extract independent components that can minimize the term dependence within them. The results show that ICA yielded higher recall and F-measure but lower precision than did LSA. By combining LSA and ICA, the performance was improved on all measures because LSA+ICA can not only discover latent concepts but also minimize term overlap among the concepts.

Another observation is that the emotion label Depression yielded the highest F-measure while both OCD and Schizophrenia yielded the lowest. One possible reason for these results is the distribution of emotion labels in the test set (e.g., Depression and Mood are the major classes). However, the skewed distribution was just a minor factor. For example, the test set included four small classes (Insomnia, OCD, Schizophrenia and Social Anxiety) with similar proportions (5.79%, 4.51%, 5.36% and 5.65%), but their F-measures were quite different (70%, 57%, 57% and 64%). Terms overlap emotion labels could have a significant impact on classification performance. For example, Insomnia had a much higher classification performance than the other three minor classes because the words used in this class were quite distinct from those used for other classes. Conversely, the words used for OCD and Social Anxiety overlapped significantly, thus yielding lower performance. Table 3 shows some representative words (with higher weights) in the independent components for the emotion labels.

Class	PureSVM			LSA			ICA			LSA+ICA		
	R	P	F	R	P	F	R	P	F	R	P	F
Depression	58	59	59	68	74	71	72	75	73	73	78	75
Drug	60	38	47	57	71	63	51	69	59	55	72	62
Insomnia	53	66	59	49	76	60	65	76	70	66	75	70
Mood	63	48	54	61	56	58	65	59	62	67	61	64
OCD	58	39	47	53	53	53	53	53	53	56	59	57
Schizophrenia	63	23	34	56	64	60	56	47	51	58	57	57
Social Anxiety	34	40	37	24	78	37	52	71	60	56	74	64
Avg.	56	45	48	53	67	57	59	64	61	62	68	64

Table 2. Performance for different classifiers. The columns R, P, and F represent recall, precision, and f-measure, respectively. (in %)

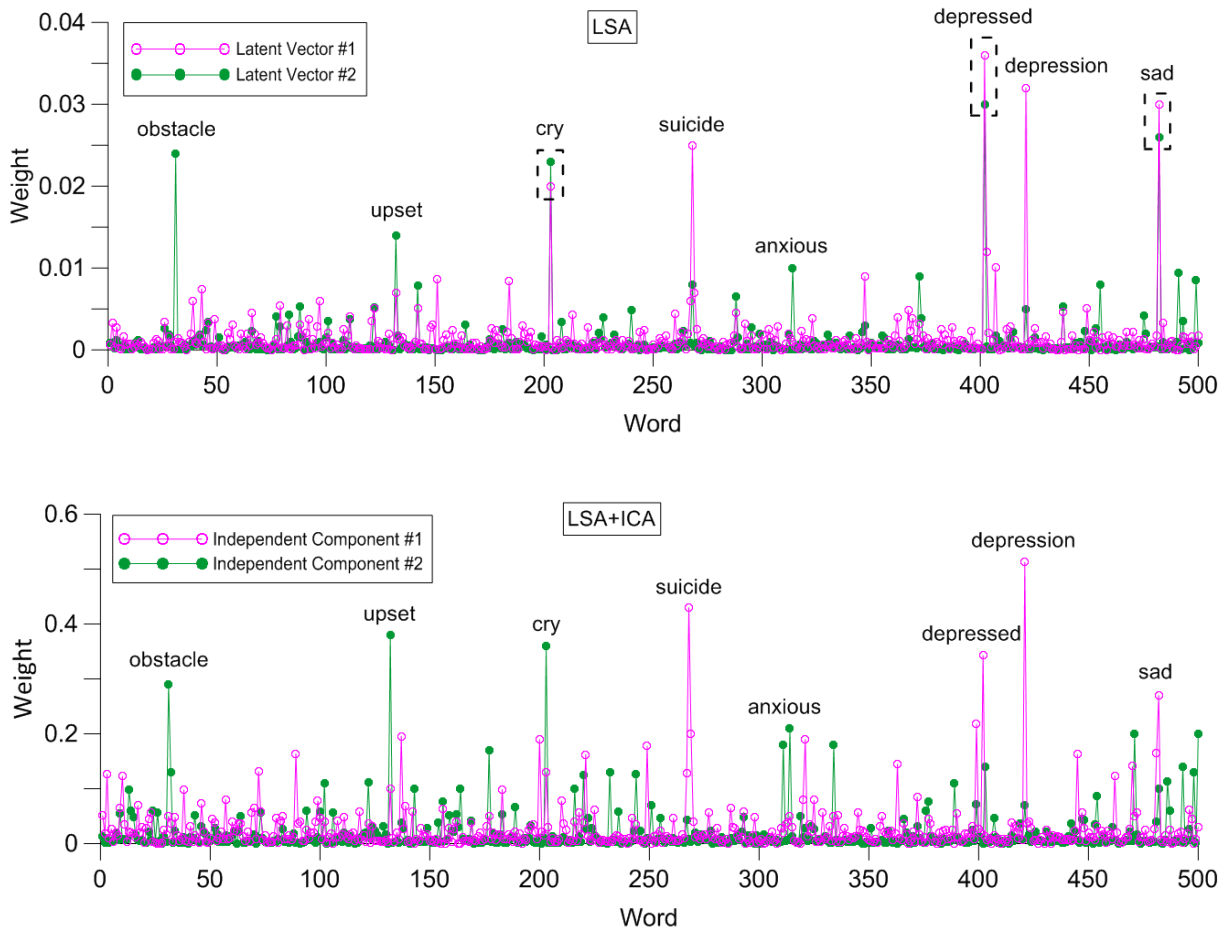


Figure 8. Examples of latent vectors, selected from $\mathbf{U}_{Q \times k}$, and independent components, selected from $\mathbf{W}_{Q \times k}^T$, for the emotion labels <Depression> and <Mood>.

3.4 Term Overlap Analysis

In order to investigate the term overlap in LSA and LSA+ICA, we analyze their respective corresponding matrices $\mathbf{U}_{Q \times k}$ and $\mathbf{W}_{Q \times k}^T$ where $\mathbf{W}_{Q \times k}^T$ is the transpose of the demixing matrix obtained with the input of $\mathbf{X}_{Q \times D}$ reconstructed using LSA. Each column of $\mathbf{U}_{Q \times k}$ and $\mathbf{W}_{Q \times k}^T$ represents a latent vector/independent component of Q words, and each element in the vector is a word weight representing its relevance to the corresponding latent vector/independent component. Figure 8 shows two sample latent vectors for LSA and two independent components for LSA+ICA, where the weights shown in this figure are the absolute values.

The upper part of Fig. 8 shows parts of the words and their weights in the two latent vectors, where latent vector #1 can be characterized by *depressed*, *depression*, and *sad* which are the useful features for identifying the emotion label <Depression>, and latent vector #2 can be characterized by *depressed*, *sad*, and *cry* which are useful for identifying <Mood>. Although the two latent vectors contained useful features for the respective emotion labels, these features still had some overlap between the latent vectors, as marked by the dashed rectangles. The overlapped features, especially those with higher weights, may reduce the classifier’s ability to distinguish between the emotion labels. The lower part of Fig. 8 also shows two independent components for the emotion labels <Depression> and <Mood>. As indicated, the term overlap between the two independent components was relatively low. Table 3 shows some representative words (with higher weights) in the independent components for the emotion labels.

Emotion Label	Representative Words
Depression	depression, depressed, sad, down, suicide
Drug	Medication, drug, dose, sedative, withdrawal,
Insomnia	sleep, insomnia, dream, nightmare, awake
Mood	cry, upset, anxious, energy, obstacle
OCD	OCD, compulsion, weight, overeating, behavior
Schizophrenia	paranoia, fantasy, memory, split, genetic
Social Anxiety	crowd, tense, friend, stiffness, ridicule, shivering

Table 3. Representative words for the emotion labels.

4 Conclusions

This work has presented a framework combining LSA and ICA for emotion label identification. Both LSA and ICA are used to analyze concept-level features, where LSA is used to discover latent concepts that do not frequently occur in psychiatric texts, and ICA is used to further minimize the term dependence among the concepts. The experimental results show that the use of concept-level features yielded better performance than the use of word-level features. Additionally, ICA can reduce the degree of term overlap of LSA so that combining LSA and ICA can discover more useful latent concepts with minimized term dependence for different emotion labels, thus improving classification performance. Future work will focus on investigating the use of the machine-labeled emotion labels as meta-information to improve online psychiatric services such as information retrieval for self-help knowledge recommendation.

Acknowledgments

This work was supported by the Ministry of Science and Technology, Taiwan, ROC, under Grant No. NSC102-2221-E-155-029-MY3.

Reference

- Y. M. Bai, C. C. Lin, J. Y. Chen, and W.C. Liu. 2001. Virtual psychiatric clinics. *American Journal of Psychiatry*, 158(7): 1160-1161.
- A. Balahur, R. Mihalcea, and A. Montoyo. 2014. Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications. *Computer Speech & Language* 28(1): 1-6.
- R. A. Calvo and S. D'Mello. 2010. Affect Detection: An interdisciplinary review of models, methods, and their applications. *IEEE Trans. Affective Computing*, 1(1): 18-37.
- J. T. Chien and H. L.Hsieh. 2012. Convex divergence ICA for blind source separation. *IEEE Trans. Audio, Speech, and Language Processing*, 20(1): 302-313.
- R. Chitturi and J. H. L. Hansen. 2008. Dialect classification for online podcasts fusing acoustic and language based structural and semantic information. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL-08)*, pages 21-24.
- G. H. Golub and C. F. Van Loan. 1996. *Matrix Computations*, Third Edition, Johns Hopkins University Press, Baltimore, MD.
- T. Honkela, A. HyvÄärinen, and J. J. VÄärynen. 2010. WordICA — emergence of linguistic representations for words by independent component analysis. *Natural Language Engineering*, 16(3): 277-308.

- A. Hyvärinen, J. Karhunen, and E. Oja. 2001. *Independent Component Analysis*. Wiley, New York.
- R. Johansson and A. Moschitti. 2013. Relational features in fine-grained opinion analysis. *Computational Linguistics*, 39(3): 473-509.
- W. Kessler and H. Schütze. 2012. Classification of inconsistent sentiment words using syntactic constructions. *Proceedings of the 24th International Conference on Computational Linguistics (COLING-12)*, pages 569-578.
- T. Kolenda and L. K. Hansen. 2000. Independent components in text. *Advances in Neural Information Processing Systems* 13: 235-256.
- T. K. Landauer, P. W. Foltz, and D. Laham. 1998. An introduction to latent semantic analysis. *Discourse Processes*, 25(2&3): 259-284.
- T. W. Lee. 1998. *Independent Component Analysis—Theory and Applications*. Kluwer, Norwell, MA.
- S. Li and C. Zong. 2008. Multi-domain Sentiment Classification. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL-08)*, pages 257-260.
- C. C. Lin, Y. M. Bai, and J. Y. Chen. 2003. Reliability of information provided by patients of a virtual psychiatric clinic, *Psychiatric Services*, 54(8): 1167-1168.
- B. Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool, Chicago, IL.
- G. R. Naik and D. K. Kumar. 2011. An overview of independent component analysis and its applications. *Informatica* 35(1): 63-81.
- M. Naughton, N. Stokes, and J. Carthy. 2008. Investigating statistical techniques for sentence-level event classification. *Proceedings of the 22nd International Conference on Computational Linguistics (COLING-08)*, pages 617-624.
- B. Pang and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2: 1-135.
- R. W. Picard. 1997. *Affective Computing*, MIT Press, Cambridge, MA.
- M. Post and S. Bergsma. 2013. Explicit and implicit syntactic features for text classification. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL-13)*, pages 866–872.
- R. Rapp. 2004. Mining text for word senses using independent component analysis. *Proceedings of the 4th SIAM International Conference on Data Mining (SDM-04)*, pages 422-426.
- H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano. 2006. Blind source separation based on a fast-convergence algorithm combining ICA and beamforming. *IEEE Trans. Audio, Speech, and Language Processing*, 14(2): 666-678.
- C. H. Wu, L. C. Yu, and F. L. Jang. 2005. Using semantic dependencies to mine depressive symptoms from consultation records. *IEEE Intelligent System*, 20(6): 50-58.
- L. C. Yu, C. L. Chan, C. C. Lin, and I. C. Lin. 2011. Mining association language patterns using a distributional semantic model for negative life event classification. *Journal of Biomedical Informatics*, 44(4): 509-518.
- L. C. Yu and W N. Chien. 2013. Independent component analysis for near-synonym choice. *Decision Support Systems*, 55(1): 146-155.
- L. C. Yu, C. H. Wu, and F. L. Jang. 2009. Psychiatric document retrieval using a discourse-aware model. *Artificial Intelligence*, 173(7-8): 817-829.