Proceedings of CONF-SEML 2026 Symposium: Importance of Machine Learning Methods and Analysis in Engineering
DOI: 10.54254/2755-2721/2026.AS32482

Discovering Latent Emotional Stages in Heartbreak-Related Social Media Texts via Unsupervised Representation Learning

Rui Huang

Department of Computer Science and Engineering, Jishou University, Zhangjiajie, China
3608423526@qq.com

Abstract. This study presents an unsupervised computational framework for discovering latent emotional stages from large-scale heartbreak-related social media texts. The proposed approach leverages representation learning to identify stage structures directly from real-world textual data. A total of 12,908 comments were collected, and 10,675 valid samples were retained after preprocessing. Text representations were extracted using BERT embeddings, followed by UMAP for dimensionality reduction and K-means clustering for stage partitioning. The optimal number of clusters was determined using the silhouette coefficient. Experimental results demonstrate clear structural separability (silhouette = 0.411) and significant inter-stage differences in emotional intensity ($F = 364.24$, $p < 0.001$, $\eta^2 = 0.064$), as well as systematic variations in temporal expression patterns. These findings indicate that unsupervised representation learning can effectively uncover psychologically interpretable stage structures from large-scale social media texts.

Keywords: Unsupervised learning, Text clustering, Representation learning, Social media mining, Sentiment analysis

1. Introduction

Romantic relationship dissolution is often accompanied by intense emotional impacts. Multi-stage emotional response theories and relationship disengagement models suggest that individuals typically experience stages such as shock, immersion, and recovery [1,2]. However, existing studies predominantly rely on questionnaire-based surveys or small-scale interviews, which yield small samples and rely heavily on subjective reports, thus lacking empirical validation in real-world contexts at scale.

With the widespread adoption of social media platforms, an increasing number of individuals publicly express their emotional experiences following breakups in online spaces, providing novel data sources for studying emotional evolution. Prior research has demonstrated that social media texts can reflect individuals' psychological states and serve as predictors of mental health indicators such as depression risk [3,4]. Furthermore, systematic reviews of social media–based mental health detection have summarized the application pathways of machine learning and deep learning techniques in psychological state recognition [5]. Nevertheless, most existing studies focus on emotion classification or risk prediction tasks, with limited attention devoted to modeling structured stages of emotional processes.

DOI: 10.54254/2755-2721/2026.AS32482

In recent years, pre-trained language models have achieved substantial progress in sentiment analysis tasks. BERT and its Chinese pre-trained variants effectively capture contextual semantic information [6,7], while methods such as Sentence-BERT further enhance sentence-level representation capability [8]. Deep learning–based emotion recognition and intensity prediction models have been widely applied in diverse scenarios [9,10]. Meanwhile, unsupervised clustering and semantic representation learning techniques have been extensively employed for text structure discovery [11,12], and contrastive learning approaches have been introduced to short-text clustering to enhance semantic discrimination [13]. Comprehensive surveys of text clustering algorithms provide theoretical foundations for these methods [14]. However, research applying unsupervised text structure discovery to psychological stage modeling remains limited, particularly when integrating temporal features for stage interpretation [15].

Therefore, this study aims to explore the feasibility of automatically discovering latent emotional stage structures from large volumes of heartbreak-related social media texts using unsupervised representation learning with no manual annotation. To this end, we develop a unified framework that integrates semantic representation learning, dimensionality reduction and clustering, sentiment intensity modeling, and temporal expression analysis. Statistical tests are conducted to validate the significance of inter-stage differences, and robustness analysis is performed to examine structural stability.

2. Methodology

2.1. Data collection and preprocessing

A total of 12,908 user comments related to romantic breakups were collected from comment sections on short-video platforms. To ensure data quality and reduce noise, a systematic preprocessing pipeline was applied.

Duplicate entries were first removed to avoid bias in clustering results. Texts dominated by emojis or containing only symbolic characters were filtered out. Semantically irrelevant content, including isolated "@" mentions, single-character replies, and automated interaction patterns, was also excluded. In addition, temporal expressions were normalized to facilitate subsequent analysis of temporal semantics.

After preprocessing, 10,675 valid samples were retained. The average text length was 28.6 characters, indicating typical short-text characteristics with limited contextual information and high semantic sparsity. These properties impose challenges on representation learning and clustering stability, motivating the use of context-aware semantic encoding models.

2.2. Unsupervised emotional phase discovery framework

To enable automatic discovery of latent emotional stage structures, we design an unsupervised computational framework that integrates representation learning and clustering analysis. The framework consists of five components: (1) semantic representation learning, (2) dimensionality reduction, (3) clustering-based stage partitioning, (4) sentiment intensity modeling, and (5) temporal expression analysis.

All experiments were conducted in a GPU-enabled Kaggle environment and implemented in Python with PyTorch to ensure reproducibility.

DOI: 10.54254/2755-2721/2026.AS32482

2.2.1. Semantic representation learning

To obtain contextualized semantic representations, the pre-trained BERT-base Chinese model [6,7] was adopted to encode textual inputs. BERT is built upon the Transformer architecture and leverages bidirectional self-attention mechanisms to capture deep contextual dependencies among tokens.

Given the short-text nature of the dataset, mean pooling was applied to the token-level hidden representations to derive fixed-length sentence embeddings. Specifically, for an input text

$$T=\{w_1,w_2, \ldots,wn\} \tag{1}$$

BERT produces contextualized hidden vectors

$$h_i \in \mathbb{R}^{768},\ i=1,2,\ldots,n \tag{2}$$

The sentence-level representation is computed as the average of all token embeddings:

$$v=\frac{1}{n}\sum_{i=1}^{n} h_i \tag{3}$$

The resulting 768-dimensional vector serves as the semantic representation of each text sample, preserving contextual information while maintaining computational efficiency for subsequent clustering.

2.2.2. Dimension reduction mapping

High-dimensional embedding spaces suffer from the curse of dimensionality, which can adversely affect clustering performance. To address this issue, Uniform Manifold Approximation and Projection (UMAP)[10] was employed to project the 768-dimensional sentence embeddings into a lower-dimensional space.

UMAP is a manifold learning–based dimensionality reduction technique that preserves local neighborhood structures while maintaining global topological consistency. This property improves cluster separability and the structural stability of embeddings in the reduced space.

In our implementation, the target dimensionality was set to 50. The number of neighbors was fixed at 15, and the minimum distance parameter was set to 0.1 to balance local structure preservation and global distribution alignment.

2.2.3. Clustering strategy and phase division

Clustering was performed in the reduced-dimensional embedding space using the K-means algorithm [12]. K-means partitions samples by minimizing the within-cluster sum of squared distances. The objective function is defined as:

$$J=\sum_{i=1}^{k}\sum_{x\in C_i}\|x-\mu_i\|^2 \tag{4}$$

where C_i denotes the $i-th$ cluster and μ_i represents the centroid of cluster C_i .

DOI: 10.54254/2755-2721/2026.AS32482

To determine the optimal number of clusters, the silhouette coefficient [11] was adopted for model selection. For a given sample iii, the silhouette score is defined as:

$$s(i)=\frac{b(i)-a(i)}{\max(a(i),b(i))} \tag{5}$$

Where $a(i)$ denotes the average distance between sample i and all other points within the same cluster, and $b(i)$ represents the minimum average distance between sample i and points in the nearest neighboring cluster.

The number of K-means initializations was set to 20 to reduce sensitivity to random initialization. A fixed random seed was used to ensure reproducibility. By comparing silhouette scores across different values of k, the optimal cluster number was determined as $k=3$.

2.2.4. Sentiment intensity modelling

To quantify emotional differences across clusters, a pre-trained sentiment classification model was used to estimate the probability distribution over sentiment categories for each text sample. Let P_{pos} and P_{neg} denote the predicted probabilities of positive and negative sentiment, respectively.

A continuous sentiment intensity score was defined as the difference between these two probabilities:

$$S=P_{positive}-P_{negative} \tag{6}$$

This formulation captures the relative polarity of each sample while avoiding the information loss associated with discrete sentiment labels. The resulting continuous variable was subsequently used for one-way ANOVA and Kruskal–Wallis testing.

2.2.5. Temporal expression extraction and statistical analysis

To examine the relationship between stage structure and temporal semantics, a rule-based matching approach was employed to identify temporal expressions in each text sample. Based on temporal scope, expressions were categorized into three types: short-term (e.g., "just", "today"), medium-term (e.g., "several months"), and long-term (e.g., "several years", "all along").

A stage-by-temporal-expression-type contingency table was constructed, and a chi-square test was performed to assess differences in distribution across stages. The chi-square statistic is defined as:

$$\chi^2=\sum_i \sum_j \frac{(O_{ij}-E_{ij})^2}{E_{ij}} \tag{7}$$

where O_{ij} and E_{ij} denote the observed and expected frequencies for cell (i,j), respectively.

DOI: 10.54254/2755-2721/2026.AS32482

3. Experimental results

3.1. Cluster structure analysis

To examine the structural characteristics of the clustering results, we first report the sample size and sentiment intensity statistics for each cluster, as summarized in Table 1.

Table 1. Cluster statistical summary

Cluster	Size	Mean Sentiment	Std. Dev	Median
0	3926	0.4661	0.2752	0.4736
1	1009	0.6410	0.1712	0.6469
2	5740	0.6033	0.2778	0.6449

As shown in Table 1, the clusters exhibit clear differences in both sample size and sentiment intensity. Cluster 0 has the lowest mean sentiment score, corresponding to a relatively moderate emotional stage. Cluster 1 demonstrates the highest mean sentiment with a smaller standard deviation, indicating more concentrated emotional expression. Cluster 2 contains the largest number of samples, with sentiment intensity distributed in the medium-to-high range and greater variability.

The clustering model achieved its highest silhouette score of 0.411 at $k=3$, suggesting favorable separability and internal consistency of the three-cluster structure.

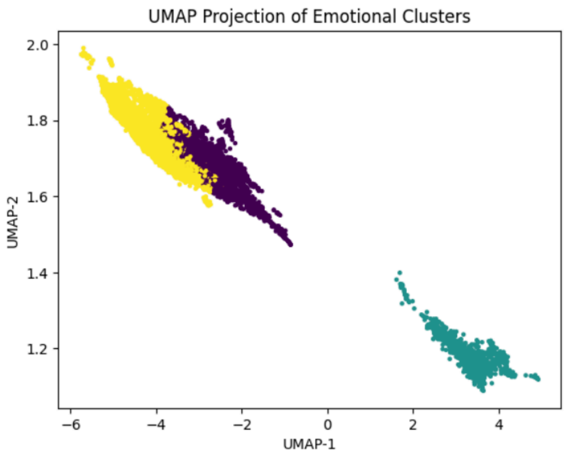

Figure 1. UMAP visualisation of clustering results

As illustrated in Fig. 1, the UMAP projection reveals three clearly separated clusters in the two-dimensional embedding space. The clusters exhibit distinct geometric boundaries, indicating strong structural separability after dimensionality reduction.

The spatial distribution further supports the selection of $k=3$ as the optimal clustering configuration. The three clusters correspond to distinct emotional patterns identified through semantic inspection.

Specifically, the high-intensity cluster is characterized by immediate expressions (e.g., "just broken up", "can't take it anymore"), reflecting acute emotional reactions. The sustained immersion cluster contains a higher proportion of long-term temporal expressions, suggesting prolonged emotional engagement. In contrast, the low-intensity cluster predominantly includes relatively moderated or rational expressions.

DOI: 10.54254/2755-2721/2026.AS32482

3.2. Differences in emotional intensity

To assess whether sentiment intensity differed significantly across clusters, both one-way ANOVA and the non-parametric Kruskal–Wallis test were conducted.

The ANOVA results indicate a statistically significant effect of cluster membership on sentiment intensity:

$$F(2,10672)=364.24,\ p<0.001,\ \eta 2=0.063. \tag{8}$$

The Kruskal–Wallis test further confirmed the presence of significant differences across clusters ($p<0.001$).

Table 2 summarizes the statistical test results.

Table 2. Statistical test results

Test	Statistic	p-value	Effect Size
ANOVA	364.24	< 0.001	$\eta^2 = 0.063$
Kruskal–Wallis	—	< 0.001	—

The effect size ($\eta 2=0.063$) suggests a moderate explanatory contribution of cluster structure to sentiment intensity.

Figure 2. Mean sentiment intensity across clusters

As shown in Fig. 2, clear differences in mean sentiment intensity are observed among the three clusters. Cluster 1 exhibits the highest average sentiment score, Cluster 0 the lowest, while Cluster 2 lies in the intermediate range. This gradient pattern is consistent with the stage interpretation derived from clustering analysis.

DOI: 10.54254/2755-2721/2026.AS32482

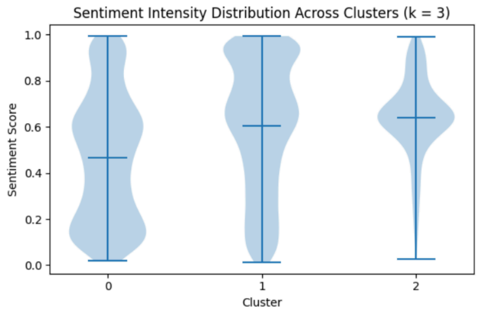

Figure 3. Sentiment intensity distribution across clusters (k = 3)

Fig. 3 further illustrates the distributional characteristics of sentiment intensity within each cluster. Beyond differences in mean values, the clusters demonstrate distinct distributional patterns in terms of dispersion and central tendency. Cluster 1 shows relatively concentrated high-intensity values, whereas Cluster 0 presents a broader and lower-intensity distribution. These distribution-level differences provide additional support for the structural validity of the identified stage configuration.

3.3. Distribution of temporal expressions

To investigate the relationship between stage structure and temporal semantics, a contingency table of stage × temporal expression type was constructed and evaluated using a chi-square test.

The results indicate significant differences in temporal expression distribution across stages ($p<0.001$) .

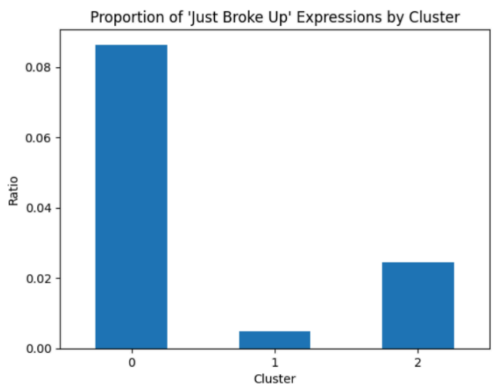

Figure 4. (a) Proportion of short-term expression ("just broke up")

As shown in Fig. 4(a), the high-intensity stage exhibits a substantially higher proportion of short-term temporal expressions compared to other clusters, reflecting immediate emotional reactions. In contrast, Fig. 4(b) demonstrates that long-term temporal expressions are more prevalent in the sustained immersion stage, suggesting prolonged emotional engagement.

DOI: 10.54254/2755-2721/2026.AS32482

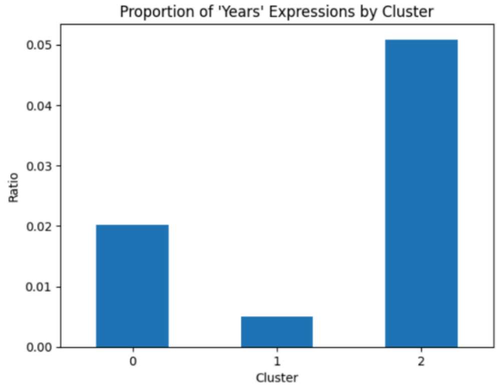

Figure 4. (b) Proportion of long-term expression ("years")

As shown in Fig. 4(b), long-term temporal expressions are substantially more prevalent in the sustained immersion stage, indicating prolonged emotional engagement.

These findings indicate that the phase structure exhibits distinctiveness not only in emotional intensity but also in systematic differences across temporal semantic dimensions.

4. Robustness analysis

To evaluate the stability of the identified stage structure with respect to clustering scale, we compared clustering performance across $K=2$ to $K=5$.

The silhouette scores under different cluster numbers are shown in Fig. 5. The highest silhouette score (0.6824) was obtained at $K=2$, indicating strong geometric separability under a dichotomous partition. However, this configuration only distinguishes between high and low emotional intensity, failing to capture finer-grained structural differences.

When $K=3$, the silhouette score decreases to 0.4088, yet a three-stage structure emerges with clear gradient differentiation. These clusters correspond to low-intensity, sustained immersion, and high-impact stages, demonstrating consistent separation in both sentiment intensity and semantic interpretation.

For $K\geq4$, additional subclusters are formed, but the differences in mean sentiment intensity between newly introduced groups become marginal, suggesting potential over-segmentation.

Overall, although $K=2$ achieves optimal geometric separability, $K=3$ provides superior structural interpretability and theoretical alignment. This balance indicates that the three-stage configuration exhibits stronger interpretative robustness and scale stability.

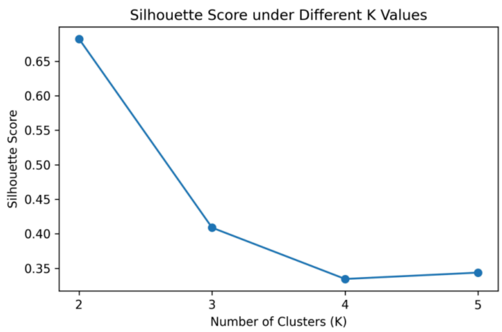

Figure 5. Silhouette score under different values of KKK

DOI: 10.54254/2755-2721/2026.AS32482

As shown in Fig. 5, the silhouette score varies across different values of KKK.

The results indicate that $K=2$ achieves the highest geometric separability, whereas $K=3$ provides better structural interpretability.

5. Discussion

The experimental results demonstrate that unsupervised representation learning can uncover structurally differentiated emotional stages from large-scale social media texts. The identified clusters exhibit distinct sentiment intensity profiles and temporal semantic patterns. One stage is characterized by short-term expressions and immediate emotional reactions, whereas another stage shows sustained reflection and long-term temporal references. These patterns show partial correspondence with classical stage theories [1,2], yet a clearly defined recovery stage was not observed. This absence may be attributed to sampling bias toward users who actively express emotional distress online.

The proposed framework provides a scalable computational approach for modeling emotional stage structures in real-world text data, complementing prior research in social media–based psychological analysis [11,12]. Nevertheless, several limitations should be acknowledged. The dataset was collected from a single platform, and user-level longitudinal data were unavailable.

To enhance interpretability, keyword-based semantic profiling using TF–IDF was conducted for each cluster. Distinct lexical patterns were observed across stages. The high-intensity cluster contains expressions such as "just broken up" and "heartbroken," reflecting acute emotional responses. The sustained immersion cluster frequently includes phrases such as "can't let go" and "what should I do," indicating prolonged emotional engagement. The low-intensity cluster exhibits comparatively moderated or reflective expressions. Together with the temporal expression analysis, these findings suggest a weak sequential progression from short-term to long-term temporal focus, supporting the plausibility of the identified stage configuration.

6. Conclusions

This paper proposes an unsupervised representation learning–based framework for the automatic discovery of emotional stages from large-scale heartbreak-related social media texts. Unlike traditional stage models derived from questionnaires or small-sample interviews, the proposed approach leverages authentic online textual data and enables stage structure identification without manual annotation.

Experimental results demonstrate that the optimal clustering configuration was achieved at $k=3$ (silhouette score = 0.411). Significant differences in sentiment intensity were observed across stages $(F(2,10672)=364.24,\ p<0.00)$, and temporal expression distributions also varied systematically among clusters $(\chi^2 \text{test}, p<0.001)$. These findings indicate that the identified stage structure is both statistically robust and psychologically interpretable.

Robustness analysis further confirms that while $k=2$ achieves stronger geometric separability,the three-stage configuration provides superior structural interpretability and theoretical alignment.

Although the study is limited by single-platform data and the absence of user-level longitudinal tracking, the proposed framework demonstrates strong extensibility.

References

[1] E. Kübler-Ross, On Death and Dying. New York, NY, USA: Macmillan, 1969.

[2] Duck, Steve. "A topography of relationship disengagement and dissolution." Personal relationships 4 (1982): 1-30.

DOI: 10.54254/2755-2721/2026.AS32482

[3] De Choudhury, Munmun, et al. "Predicting depression via social media." Proceedings of the international AAAI conference on web and social media. Vol. 7. No. 1. 2013.

[4] Chancellor, Stevie, Jessica L. Feuston, and Jayhyun Chang. "Contextual gaps in machine learning for mental illness prediction: The case of diagnostic disclosures." Proceedings of the ACM on Human-Computer Interaction 7.CSCW2 (2023): 1-27.

[5] Devlin, Jacob, et al. "BERT: Pre-training of deep bidirectional transformers for language understanding." Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies, volume 1 (long and short papers). 2019.

[6] Cui, Yiming, et al. "Revisiting pre-trained models for Chinese natural language processing." Findings of the Association for Computational Linguistics: EMNLP 2020. 2020.

[7] Reimers, Nils, and Iryna Gurevych. "Sentence-bert: Sentence embeddings using siamese bert-networks." Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). 2019.

[8] Calvo, Rafael A., and Sidney D'Mello. "Affect detection: An interdisciplinary review of models, methods, and their applications." IEEE Transactions on Affective Computing 1.1 (2010): 18-37.

[9] Abdullah, Tariq, and Ahmed Ahmet. "Deep learning in sentiment analysis: Recent architectures." ACM Computing Surveys 55.8 (2022): 1-37.

[10] McInnes, Leland, John Healy, and James Melville. "UMAP: Uniform manifold approximation and projection for dimension reduction." arXiv preprint arXiv: 1802.03426 (2018).

[11] Rousseeuw, Peter J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." Journal of computational and applied mathematics 20 (1987): 53-65.

[12] MacQueen, J. "Some methods for classification and analysis of multivariate observations." Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability/University of California Press. 1967.

[13] Tahir, Waleed Bin, et al. "Depression detection in social media: A comprehensive review of machine learning and deep learning techniques." IEEE Access 13 (2025): 12789-12818.

[14] Aggarwal, Charu C., and ChengXiang Zhai. "A survey of text clustering algorithms." Mining text data. Boston, MA: Springer US, 2012. pp. 77–128.

[15] Zhu, Zhengzhong, et al. "Matrix contrastive learning for short text clustering." International Conference on Neural Information Processing. Singapore: Springer Nature Singapore, 2023.