Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identifying and Characterizing Eating Disorder Discourse on Chinese Social Media: A Machine Learning Approach

Yuchen Zhang
University of Toronto
Nanyu Luo
University of Toronto
Xiaoya Zhang
University of Florida
Feng Ji
f.ji@utoronto.ca

University of Toronto
Jinbo He
Xi'an Jiaotong-Liverpool University

Research Article

Keywords: machine learning, deep learning, eating disorder topics, topic modelling, social media

Posted Date: November 12th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7852043/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Eating disorders (EDs) are severe psychiatric conditions with high mortality and substantial medical complications. In China, underdiagnosis and low treatment engagement hinder timely intervention. Social media platforms provide a naturalistic lens into ED-related experiences, yet research on Chinese-language data remains scarce. Advances in machine learning (ML) and deep learning (DL) offer new opportunities to identify and characterize such ED discourse, informing the development of scalable detection methods and culturally tailored prevention and intervention strategies in the Chinese context.

Methods

We collected ED-related posts from Weibo via keyword-based API searches and manually annotated them into three groups: irrelevant, promotional/educational content, and layperson posts. Five ML/DL methods, including Convolutional Neural Networks (CNNs), Random Forests, XGBoost, Support Vector Machines (SVMs), and Logistic Regression, were trained to identify ED-related posts in a two-stage framework: (1) filtering out irrelevant posts and (2) distinguishing promotional/educational posts from layperson posts. Classifier performance was evaluated on additional posts from the same users. Latent Dirichlet Allocation (LDA) was applied to the layperson subset to extract underlying ED-related themes.

Results

CNN consistently outperformed other models, achieving high F1-scores in both classification stages (0.87 and 0.98, respectively). Topic modelling revealed five themes: restrictive symptomatology and physical distress, binge eating and body-health concerns, relapse and coping narratives, emotional venting, and chronic ED patterns with identity impact.

Conclusions

This study demonstrates that CNN-based classification combined with topic modeling provides a scalable framework for detecting ED-related discourse on Chinese social media. Beyond methodological advances in non-English NLP, the findings highlight culturally specific symptom expressions and psychosocial concerns, offering novel insights for public health surveillance. These insights can inform the development of early detection tools and culturally sensitive interventions to address the unmet needs of individuals with EDs in China.

1. Introduction

Eating disorders (EDs), such as anorexia nervosa (AN), bulimia nervosa (BN), and binge eating disorder (BED), are complex and severe psychiatric conditions marked by high mortality [1–3], and extensive adverse physical and psychological consequences, such as malnutrition [4]; metabolic disturbances [5]; comorbidity with major depression and generalized anxiety disorders [6]. EDs typically manifest during adolescence and early adulthood and often persist throughout life, bringing significant distress to both patients and their families [7].

Although incidence rates of EDs in China have been rising, their prevalence remains underestimated, and treatment engagement is low [8–10]. One plausible explanation is that individuals with EDs tend to exhibit self-denial and concealing behaviors for self-protection, which creates substantial challenges in assessing and treating these conditions [11]. Given China's large population base and substantial unmet need, identifying risk factors and patterns of disordered eating is critical for prevention [12]. Yet, research on ED-related discourse in the Chinese context remains limited. Thus, there is an urgent need for scalable methods to capture risk patterns beyond clinical samples

Social media provides a unique window into hidden or unreported mental health experiences [13], including EDs [14–15]. Platforms (e.g., Meta and Twitter/X in Western countries, and Weibo in China) have provided open and (semi-)anonymous spaces for people to share their daily thoughts and feelings, seek information, and discuss topics through various media (e.g., text, pictures, videos; [15–16]. These digital traces have been shown to reveal novel behavioral patterns and can complement epidemiological and clinical approaches [13]. In China, Weibo functions much like Twitter/X, which allows users to rapidly exchange information and form interactive social networks. Moreover, Chinese natural language processing (NLP) presents unique challenges, including word segmentation and idiomatic expressions [17–18]. Such linguistic and cultural features necessitate context-specific approaches rather than extrapolating from English-language research. Thus, rigorous study of Chinese platforms like Weibo is essential to uncover more culturally specific signals of EDs at the population scale.

Moreover, because people in ED communities often share their thoughts, behaviors, and daily lives online [19–20], the large volumes of user-generated content on social media provide a strong basis for large-scale analysis in digital health [21]. Although Chinese datasets have linguistic and cultural features that differ from those of English, powerful toolkits, such as Machine Learning (ML), can still extract key signals useful for detecting risk factors and behavioral patterns [16, 22–23]. In particular, advanced deep-learning models can accurately identify users showing ED-related symptoms from social media posts [10, 20]. These methods help narrow broad user pools to specific at-risk groups, enabling closer examination of distinct patterns.

ED-related topics are crucial to understanding the complexities of individuals' mental health conditions. A retrospective literature review examined key trends in ED topics, including treatment, symptoms, and risk factors, from 1990 to 2021, emphasizing the importance of ongoing investigation in this field [24]. Moreover, researchers are also beginning to utilize social media data and big data analysis to enhance these insights [10, 19]. For example, Wang and colleagues focused on detecting and characterizing ED

communities on Twitter, which provided insights into community structures and interactions among ED individuals [20].

Machine learning can also further accelerate topic discovery at scale by minimizing the need for manual labeling. For instance, algorithms such as Latent Dirichlet Allocation (LDA) automatically extract latent themes related to social support and ED-specific content across large datasets, eliminating the need for intensive manual categorization that was previously required [25]. Similarly, Zhou and colleagues used the Correlation Explanation (CorEx) models to identify and cluster topics related to ED symptoms (e.g., body image, the consequences of EDs; [10]). Together, these methods demonstrate the potential of social media data to enhance prevention and treatment efforts in the public health sector. However, most existing work focuses on English-language datasets and evaluates a limited set of models. This leaves a gap in the understanding of culture-specific risk factors associated with EDs.

Our study extends Zhou et al.'s method on Twitter data by adapting their ML classifiers and topic modeling methods to Weibo, thereby addressing the specific cultural and linguistic differences of Chinese-language social media content. Although Zhou et al. demonstrated the feasibility of English datasets, it remains unclear whether they apply equally well to other unique content [10]. Compared with Twitter, Weibo differs in most topic distributions and dynamics of social attention, as the entertainment context dominates the topics, and hashtag use also follows Weibo-specific conventions [26]. In this environment, body-image norms and shared dieting or eating habits can spread rapidly, often via celebrities, influencers, and trending tags, which might shape public perceptions and potentially amplify the risk of EDs. Additionally, Chinese natural language processing tasks require addressing word segmentation and Part-of-Speech tagging issues, which introduce challenges that are less prominent in English [27]. Moreover, ED-related topics in East Asian contexts are distinct from those in Western settings. For example, studies have found that Chinese young women reported higher levels of baseline stigmatization compared with Australian students (Yan et al., 2018). Additionally, research has demonstrated that East Asians are more likely to communicate negative experiences through somatic expression than Westerners [28–29]. Therefore, these specifications in cultural-linguistic features and platform-specific norms underscore the significance of examining ED-related content patterns in Chinese social media.

Thus, we utilize data from Weibo to develop machine-learning techniques for exploring ED-related content in a culturally specific Chinese context. Specifically, our research addresses two key questions. The first step is to systematically assess the effectiveness of five distinct statistical and computational methods (e.g., Convolutional Neural Networks [CNNs], Random Forests [RFs], Extreme Gradient Boost [XGBoost], Support Vector Machines [SVMs], and Logistic Regression [LR]) in identifying ED-related content. The second research question aims at exploring and evaluating ED-related topics extracted from Weibo posts via Latent Dirichlet Allocation (LDA) so that risk factors or behavioural patterns associated with ED symptoms can be further revealed.

By combining supervised classification with unsupervised topic modeling, we aim to demonstrate a scalable framework for ED surveillance in a non-English context. This approach not only advances methodological innovation but also provides culturally grounded insights into disordered eating, with potential applications for early detection and intervention in China.

2. Method

2.1 Study Design

Our study methodology follows a two-stage pipeline shown in Fig. 1. In the first stage, we collected an initial dataset of 6,589 Weibo posts containing keywords related to EDs via the Weibo Application Programming Interface (API). These posts are manually annotated into three categories (i.e., ED-laypeople, ED-promotion and education, and ED-irrelevant), which yields 3,469 posts in the ED-laypeople subset.

Using these 3,469 posts as our training dataset, we developed and compared several traditional ML-classifiers alongside a deep learning (DL) model. The power of ML is utilized to identify ED-related posts and analyze ED-related topics within them, resulting in a twofold development. All models had two main tasks: first, to filter out ED-irrelevant posts and then to sort the remaining posts into either ED-promotion and education or ED-laypeople. Second, to evaluate the effectiveness of the classifiers, we identified the user IDs associated with the 3,469 laypeople posts and randomly sampled 20,000 additional posts from these same users as our testing datasets. Afterwards, the best-performing classifier retrieved 8,665 new laypeople subsets.

In the second stage, we applied LDA topic modelling to both the manually labelled ED-laypeople subset (N = 3,469) and the classifier-identified subset (N = 8,665). The outcome of the topic modelling was scrutinized to assess how well it assisted in generating ED-relevant topics for each post. This was achieved by comparing each topic's theme word to the content of posts under that topic, aiming to validate the relevance and accuracy of the automated topic modelling process.

2.2 Weibo Posts Collection

This study was conducted in January 2023. We used the Weibo streaming API to search and collect posts relevant to our research. Utilizing a carefully selected list of keywords related to EDs (e.g., "□□□ Anorexia", "□□□□□□ Anorexia Nervosa", "□□□□□ Bulimia Nervosa", "□□□ Bulimia", "□□□ Binge-eating Disorder", "□□□□ Binge eating", "□□□□ Eating disorder", "□□□ Compulsive overeating", "□□□□ Eating Disorder", "□□□ Excessive Hunger") in Chinese, we were able to retrieve a comprehensive pool of posts for manual screening on Weibo.

2.3 Identification of Target Posts

2.3.1 Manually Annotated ED-laypeople Posts and ED-irrelevant Posts

Our framework for identifying target posts was based on the “gold-standard” dataset distribution reported by Zhou et al., which annotated posts into three categories: *ED-irrelevant, ED-promotional and education*, and *ED-laypeopl*e. Posts that were unrelated to EDs were categorized as ED-irrelevant; posts aimed at promoting products or educating the public about EDs were classified as *ED-promotional and education* posts; and *ED-laypeople* posts referred to ED-relevant posts made by individual users. In this study, we focused specifically on *ED-laypeople* posts [10].

Three reviewers and two supervisors formed the annotation team for the manual review tasks in the study. All team members were pre-trained on the annotation principles to ensure consistency. Initially, the three reviewers independently annotated the posts. Then, they cross-checked their results to ensure the reliability of the results. Next, any inconsistencies were discussed with the supervisors to reach a final consensus on all posts.

2.3.2 Machine Learning Classifier Development

To explore latent patterns within the manually labeled corpus, we employed ML algorithms to design automatic text classifiers. Since these algorithms only process numerical data, we pre-processed the texts. This involved removing elements that were potentially devoid of significant information, such as Chinese stop words (e.g., “□” [dàn], translating to “but”); user mentions (e.g., @username); hyperlinks (e.g., https://abccc.com); hashtags (e.g., #gym); and emojis. The refined text was then segmented into individual words using the ‘*jieba*’ Chinese text segmentation toolkit. Subsequently, a comprehensive lexicon was established, assigning a unique numerical identifier to each word. This conversion transformed each post into a numeric vector representation.

Our study centered on the construction of a classifier utilizing four conventional machine learning models, including Logistic Regression (LR), Support Vector Machines (SVMs), Random Forests (RFs), and Gradient Boosting Trees (XGBoost), alongside Convolutional Neural Networks (CNNs), a DL model. To enhance performance and avoid overfitting, we meticulously tuned the hyperparameters of the traditional models by amalgamating grid search with cross-validation.

The CNNs’ structure consisted of three layers: an embedding layer, a convolutional block, and an output layer. The embedding layer, pre-populated with terms from both corpora mentioned above, facilitates the initial transformation of text into vector form. The convolutional block sequentially applied 1D convolutions with 64 filters of kernel size 3, followed by the ReLU activation function, max pooling with a kernel size of 2 for dimensionality reduction, and a dropout with a rate of 0.2 to deter overfitting. The last linear output layer was responsible for projecting the flattened features from the CNN onto the spectrum of class categories, yielding logits corresponding to each class.

Our classification framework was designed in two stages. The initial phase entailed differentiating ED-related posts (including ED-layperson discussions and ED-promotion and education posts) from those irrelevant to ED, resulting in a binary classification. The ensuing stage engaged classifiers to distinguish between posts of ED-laypeople and those categorized under ED-promotion and education.

2.4 Identifying Topics from Posts of Potential ED Users

Topic modeling techniques were used to extract topics and explore the mental status and daily concerns of potential ED users. Specifically, we utilized Latent Dirichlet Allocation (LDA; [30]), a probabilistic model designed to identify latent thematic structures within sizable textual datasets. This model assumes that documents are composed of a blend of latent topics, where each topic is defined by its unique word distribution. In practice, LDA interprets each document as a mixture of topics, assigning each word to a topic based on the distribution of words across topics and vice versa.

We applied the LDA model to two distinct sets of data: (1) the initial dataset's posts that were manually labeled as ED-related discussions by laypeople and (2) specific posts from potential ED users, which were categorized as ED-laypeople discussions by our automated classifier.

Determining the appropriate number of topics in LDA is a pivotal step. We experimented with different topic quantities, ranging from 4 to 10. While statistical measures such as perplexity and relevance scores offer guidance on the number of topics, they do not always yield thematically rich topics. Therefore, we conducted a series of testimonies and meticulously examined the resultant topics, allowing us to identify the most informative and representative topic number for subsequent in-depth analysis.

2.5 Topic evaluation

Our team was instructed by domain experts to manually identify each topic from the top 30 most salient keywords generated by topic modeling and assign a corresponding key theme to each topic. Next, we then manually reviewed the top 10 most relevant posts associated with these keywords and evaluated whether the post was coherent with the generated topic theme. The relevance scores were defined as the percentages of posts that were coherent with each topic theme.

3. Results

3.1 Post Collection and Manual Identification of ED-related Content

Using the keyword search feature via the Weibo API, we collected 6,589 posts. Of these, 3,489 were classified as discussions by ED laypeople, 1,338 as ED promotional and educational materials, and 1,782 as unrelated to EDs. The results of Cohen’s Kappa range from .58 to .79, indicating a moderate to substantial agreement among reviewers. The post collections of users whose posts were labeled as ED

laypeople are the source of the second dataset, which consists of 20,000 posts randomly selected from these collections.

3.2 Machine Learning Classifier Development

Our analysis encompassed five machine learning algorithms: CNNs, RFs, XGBoost, SVMs, and LR, mentioned above. The manually labeled dataset was randomly divided into a training set, encompassing 69.98% of the posts (4,611 out of 6,589), and a validation set, containing the remaining 30.02% (1,978 out of 6,589). Stratified sampling ensured equivalent distribution across both sets relative to the original dataset. We designed a two-tiered classification system to separate ED-irrelevant posts and to differentiate between ED-promotional and ED-laypeople discussions. Preliminary results indicate the classifiers' performances as shown in the subsequent Table 1.

Table 1
Performance comparison of classifiers under different thresholds

Classifier	Precision	Recall	F1-score	Precision	Recall	F1-score
ED-irrelevant versus other 2 labels (Threshold 0.5)						
	ED-irrelevant (535)			ED-promotional and education + ED-laypeople (1443)		
CNN	0.68	0.69	0.68	0.89	0.88	0.88
RF	0.83	0.55	0.66	0.85	0.96	0.9
XGBoost	0.81	0.51	0.63	0.84	0.96	0.9
SVM	0.68	0.36	0.47	0.8	0.94	0.86
LR	0.57	0.46	0.51	0.82	0.87	0.84
ED-irrelevant versus other 2 labels (Threshold 0.9)						
	ED-irrelevant (535)			ED-promotional and education + ED-laypeople (1443)		
CNN	0.64	0.75	0.69	0.9	0.85	0.87
RF	0.35	0.96	0.51	0.96	0.34	0.5
XGBoost	0.29	0.98	0.45	0.96	0.12	0.21
SVM	0.37	0.83	0.51	0.89	0.48	0.62
LR	0.31	0.92	0.46	0.89	0.25	0.39
ED-promotional and education versus ED-laypeople						
	ED-promotional and education (402)			ED-laypeople (1041)		
CNN	0.99	0.91	0.95	0.97	1	0.98
RF	0.95	0.92	0.93	0.97	0.98	0.97
XGBoost	0.91	0.87	0.89	0.95	0.97	0.96
SVM	0.87	0.75	0.8	0.91	0.96	0.93
LR	0.88	0.67	0.76	0.88	0.96	0.92

In the initial stage, we examined two methodologies to distinguish ED-irrelevant posts from the other two groups. For any post input, all the models were regarded as statistical models that output the probability of the input belonging to ED-irrelevant posts and the probability of the opposite. If the traditional 0.5 threshold is set, events with higher probability are more likely to occur, so the post should belong to the higher probability group. By setting the conventional threshold at 0.5, our models exhibited a

commendable average F1-score of 0.87 but a modest average recall of 0.51 for the ED-irrelevant category, indicating that although the classifier successfully found ED-relevant posts, the posts predicted as no ED-irrelevant posts would be mixed with a considerable number of ED-irrelevant posts. Therefore, we set the threshold at a high value of 0.9, meaning the classifier considers the post ED-irrelevant only if the output probability exceeds 0.9. The adjustment of the threshold diminished the F1-scores for the traditional models; however, the CNN sustained its high F1-score (0.87) and enhanced the recall for ED-irrelevant posts to 0.75.

During the second phase, the CNNs surpassed their traditional counterparts with an F1-score of 0.98. Consequently, we selected CNNs for automated text classification, which identified 8,665 posts as ED laypeople discussions within the second dataset of 20,000 posts of potential ED users.

3.3 Topic Modeling and Evaluation

Applying the LDA model to the aforementioned corpora, 3,489 manually tagged ED laypeople posts and 8,665 CNN-classified ED laypeople posts, enabled topic extraction. Following extensive experimentation and analysis, we determined that five topics are optimal for clustering ED-related symptoms into corresponding ED topics. The emblematic themes associated with each topic are presented subsequently.

The five topic themes that we extracted based on our manual labeling training data were (1) restrictive ED symptomatology & physical distress, (2) binge eating behaviors & body-health concerns, (3) ED episodes, relapse, & coping narratives, (4) emotional venting during ED episodes, and (5) chronic ED patterns & identity impact. The relevance scores for these topics were 90%, 90%, 90%, 80%, & 100%, respectively. According to the intertopic distance map (Fig. 2), topic theme 1 and topic theme 2 were positioned far apart, with large and non-overlapping bubbles, indicating two highly prevalent but lexically distinct subtopics. In contrast, topics 3, 4, and 5 formed a tight cluster with partial overlaps. This reflected shared topics related to ED episodes but with differentiated focuses. Specifically, topic theme 3 focused on diary-style and relapse-related language; topic theme 4 centered on affect-laden expression; and topic theme 5 emphasized the habitual-related impacts of chronic ED patterns (Table 2). Additional figures also showed the top-30 most relevant terms for each topic (Fig. 3, 4, 5, 6, 7).

Table 2: Five topic themes and their 10 most relevant contents

Topic Theme 1: restrictive ED symptomatology & physical distress

Post	Content keywords	Keywords translation	Relevance
1	□□□□, □□□□, □□□□□	nausea due to hunger/ plain nausea, anorexia tendency*	1
2	□□□□, □□□, □□□□	dizzy, short of breath, no appetite	1
3	□□,□□□□□□□□, □□□□□	diet, no hunger after a day without food, anorexia tendency	1
4	□/□□□□□, □□□□□ □□	insomnia due to hunger/bloating; binge eating, want to vomit	1
5	□□□□□, □□□, □□, □□□	sequela of covid-19, anorexia*, nausea, lost 4kg	1
6	□□□□□□□, □□□□□□□	feel I have anorexia, eating like a task	1
7	□□,□□□, □□□□□□□□□	nausea, no appetite, sequela of covid-19, anorexia	1
8	□□□,□□□□	huge appetite, binge eating	0
9	□□□□□□, □, □□□□□□	full after several bites, vomit, suspect of anorexia	1
10	□□□□, □□□□□□□	No hunger, suspect of anorexia	1
Relevance Scores			90%

* Anorexia tendency and anorexia means low appetite in China, incidating restrictive ED symptomology. People may use sentences like "I feels like I have gotten the anorexia that I have no appetite these days.

Topic Theme 2: binge eating behaviors & body-health concerns

Post	Content keywords	Keywords translation	Relevance
1	□□□□□□□□, □□	binge eating then diet, stay up	1
2	□□□□□□□, □□□	binge eating over 2 weeks, high weight	1
3	□□□□□□	binge eating daily	1
4	□□□□	binge eating	1
5	□□□□	binge eating	1
6	□□□□□□□□□□	severe depression, anorexia	0
7	□□□□□	night-eating syndrome	1
8	□□□□□□□□□	binge eating triggered hives	1
9	□□□□□□□□□□□□□□□	poor immune system, somatization reaction, ED	1

10	□□□,□□□□	binge eating disorder, eating obsessively	1
Relevance Scores			90%

Topic Theme 3: ED episodes, relapse, & coping narratives			
Post	Content keywords	Keywords translation	Relevance
1	***□□□□□	f*** bulimia nervosa	0
2	□□, □□□	meal log, binge eating disorders	1
3	□□□□□□□□,□□□□	bulimia recovery diary, ED	1
4	□□□□□,□□□□□□	relapse of binge eating, mood affecting	1
5	□□□, □□	binge eating disorders, tests	1
6	□□□□□□	stabilize cravings	1
7	□□□	bulimia nervosa	1
8	□□□□□, □□□□□□	no binge today, no regret	1
9	□□□□□□□□...□□, □□□□, □□□□	severe binge eating recently, stuffing stomach until full, give up thinking	1
10	□□□□□□□□,□□□□, □□□□, □□□	binge eating recovery dairy, pale, stop binge, love yourself	1
Relevance Scores			90%

Topic Theme 4: emotional venting during ED episodes			
Post	Content keywords	Keywords translation	Relevance
1	□□□	anorexia nervosa	0
2	□□□□□□□, □□□□	like anorexia nervosa, nothing to live for	1
3	□□□	anorexia nervosa	0
4	□□□□□□□□□□□,□□□	I feel sick from eating too much, eating syndrome	1
5	□□□□□□	Counterexample of binge eating disorders	0
6	□□□, □□□, □□	Anorexia nervosa, eating syndrome, heart breaking	1
7	□□, □□□	I can’t even, eating syndrome	1
8	□□□, ~!	eating syndrome with excited tones	1
9	□□□, !!	eating syndrome with excited tones	1

10	□□□, □□□	eating syndrome, so unfair	1
Relevance Scores			70%

Topic Theme 5: chronic ED patterns & identity impact			
Post	Content keywords	Keywords translation	Relevance
1	□□□□, □□□, □□□□, □□□□□□	hobbies, overspending, binge eating, staying in bed all day	1
2	□□□□, □□□□, □□□□	Staying up late drinking, binge eating, planning to drop dead early	1
3	□□□□, □□	binge eating, bowel movement	1
4	□□□	eating syndrome	1
5	□□□□□, □□□□	strain, binge eating, feel unbearably stuffed	1
6	□□□□, □□□, □□□□, □□□□□□	hobbies, overspending, binge eating, staying in bed all day	1
7	□□□□, □□□, □□□□, □□□□□□	hobbies, overspending, binge eating, staying in bed all day	1
8	□□□□, □□□	overwhelming stress, eating syndrome	1
9	□□□□	eating syndrome, oversleeping	1
10	□□□, □□□□	feeling down, binge eating	1
Relevance Scores			100%

In addition to the five topic themes described above, our analysis also revealed several cross-cutting patterns in the post contents. Many posts linked ED symptoms with comorbid conditions (e.g., sleeping difficulties, anxiety, endocrine dyscrasia) and with bodily sensations (e.g., lowered immune function, urticaria, somatization). Furthermore, given the specific years represented in our dataset, a notable subset of posts referenced the sequelae of COVID-19.

4. Discussion

This study developed a two-tier text classification and topic modeling pipeline to identify and characterize ED-related social media posts from laypeople. In Stage 1, the CNN maintained a strong F1-score (0.87). In Stage 2, the CNN outperformed traditional models and was used to classify 20,000 additional posts, subsequently identifying 8,665 posts from ED laypeople. LDA on the combined corpora yielded five interpretable topic themes related to EDs.

Our CNN accurately classified social-media posts, showing that deep models learn predictive local phrase patterns that simpler baselines miss. This finding is consistent with reports that CNNs and other deep neural models perform strongly in text-based mental health detection [10, 31]. In parallel, our topic modeling yielded coherent, clinically interpretable themes from unselected posts, adding information that complements the supervised classifier. Prior studies also show that LDA and related models reveal latent concerns, track shifts in discourse, and describe support dynamics in online mental health communities [25, 32].

The distance between bubbles of Topic themes 1 and 2 suggests that restriction-focused ED symptomology (e.g., “no appetite”, “anorexia tendency”) is linguistically apart from binge-eating-focused ED symptoms (e.g., “binge eating”, weight complaints). The clustering of Topic themes 3, 4, and 5 indicates that laypeople often narrate their ED episodes alongside emotion dysregulation and personal consequences (e.g., staying overnight, overwhelming stress), such that episodic descriptions of EDs and coping strategies (e.g., recovery diaries, “staying in bed all day”) appear together. These findings align with the literature, which suggests that emotional dysregulation (e.g., emotional nonacceptance, emotional distress, poor impulse control, deficits in emotion awareness) is a central feature across various EDs [33–34]. Meanwhile, the presence of diary-style self-monitoring and logs aligns with the CBT-E practice, which encourages the use of a real-time self-monitoring mechanism to modify eating behaviors. In this setting, users can anonymously use social media to track their eating behaviors and support their self-regulation, while often seeking community as well, without disclosing their formal diagnosis or treatment. These five topic themes suggest a need for interventions that not only address cognitive reframing but also provide psychoeducation related to body-focused management (i.e., understanding hunger cues and techniques to soothe nausea), connect individuals to chronic recovery resources (i.e., routine-based self-care and narrative-based techniques), and ensure sustained long-term support.

Our findings are based on ED-related content in user-generated posts and are therefore not suitable for clinical settings. However, as social media offers a medium to reach broader public audiences, including individuals who may conceal or underreport maladaptive eating behaviors, these data could help characterize behavioral patterns in those lay populations and inform targeted psychoeducation courses, interventions, and prevention programs. Additionally, using a stringent 0.9 threshold for the ED-irrelevant class trades precision for recall to balance the results, which may potentially place borderline and/or ambiguous posts in the ED-relevant stream (i.e., ED-promotional and educational content, ED-laypeople). Moreover, although overall relevance scores are high, the topic coherence measure remains partly subjective, which prevents interpretability and leaves room for refinement [35]. We could incorporate human-in-the-loop to adjust for those borderline cases to improve reliability in our future work. Further gains may be achieved by using more advanced models. Recurrent models such as LSTM can improve classification accuracy [31], and topic models like BERTopic [36] can better capture latent patterns in text. At the same time, social-media posts, especially on platforms such as Weibo, often contain noise, including slang, idioms, and ambiguous internet expressions. More careful preprocessing and cleaning,

along with the use of higher-quality text sources, when possible, are therefore crucial for obtaining reliable and meaningful results.

5. Conclusion

Our study provides actionable insights that CNN can reliably classify large volumes of social media posts and combine with LDA to cluster coherent themes that point to intervention targets. Practically, this approach can help public-health teams prioritize designs targeted to psychoeducational content for lay communities who may not disclose clinical diagnoses. For example, this offers a novel direction for implementing interventions focused on physical sensation, such as hunger cues, in EDs, and initiating long-term recovery management (i.e., routine-based recovery supports, narrative-based self-monitoring tools). For future work, we recommend incorporating human-in-the-loop review for borderline cases, utilizing more advanced topic and text models, refining platform-specific preprocessing, and validating findings against clinical samples to enhance reliability and inform the design of intervention trials more effectively.

Abbreviations

AN
Anorexia Nervosa
BED
Binge Eating Disorder
BN
Bulimia Nervosa
CNN
Convolutional Neural Network
DL
Deep Learning
ED
Eating Disorder
LDA
Latent Dirichlet Allocation
LR
Logistic Regression
ML
Machine Learning
NLP
Natural Language Processing
RF
Random Forest
SVM

Support Vector Machine
XGBoost
Extreme Gradient Boosting
Table 2
Five topic themes and their 10 most relevant contents

Declarations

Ethical approval and consent to participate

Ethical approval

The project was approved by the Institutional Review Board (Applied Psychology) of The Chinese University of Hong Kong, Shenzhen (Approval No. EF20241018001).

Informed consent

Not Applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data are available from the corresponding author upon request.

Competing Interest

The authors(s) declare that they have no competing interests.

Funding

Jinbo He was supported by the National Natural Science Foundation of China (Grant Number 72204208). Feng Ji was supported by the Connaught Fund (Grant Number 520245) and Social Sciences and Humanities Research Council (SSHRC) of Canada (Grant Number 215119, CRC-2024-00169).

Authors' Contributions

Yuchen Zhang: Investigation, Project administration, Writing – original draft, Writing – review & editing; Nanyu Luo: Formal analysis, Writing – original draft, Writing – review & editing; Xiaoya Zhang: Writing – review & editing, Methodology; Feng Ji: Conceptualization, Supervision, Funding acquisition, Project administration, and Writing – review & editing; Jinbo He: Conceptualization, Supervision, Funding acquisition, Project administration, and Writing – review & editing.

Declaration of generative AI in scientific writing

During the preparation of this work, the authors utilized GPT-5 to enhance the language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Acknowledgements

Not Applicable.

References

1. Klump KL, et al. Academy for eating disorders position paper: eating disorders are serious mental illnesses. Int J Eat Disord. 2009;42(2). http://doi.org/10.1002/eat.20589.
2. Krug I et al. *A meta-analysis of mortality rates in eating disorders: an update of the literature from 2010 to 2024.* Clinical psychology review, 2025: p. 102547. https://doi.org/10.1016/j.cpr.2025.102547
3. Smink FR, Van Hoeken D, Hoek HW. Epidemiology of eating disorders: incidence, prevalence and mortality rates. Curr psychiatry Rep. 2012;14(4):406–14. https://doi.org/10.1007/s11920-012-0282-y.
4. Carlson JL, Lemly DC. Medical Considerations and Consequences of Eating Disorders. Focus. 2024;22(3):301–6. https://doi.org/10.1176/appi.focus.20230042.
5. Mitchell JE. Medical comorbidity and medical complications associated with binge-eating disorder. Int J Eat Disord. 2016;49(3):319–23. https://doi.org/10.1002/eat.22452.
6. Keski-Rahkonen A, Mustelin L. Epidemiology of eating disorders in Europe: prevalence, incidence, comorbidity, course, consequences, and risk factors. Curr Opin Psychiatry. 2016;29(6):340–5. http://doi.org/10.1097/YCO.0000000000000278.
7. Chavez M, Insel TR. Eating disorders: National Institute of Mental Health's perspective. Am Psychol. 2007;62(3):159. https://doi.org/10.1037/0003-066X.62.3.159.
8. Li Z, et al. Burden of eating disorders in China, 1990–2019: an updated systematic analysis of the global burden of disease study 2019. Front Psychiatry. 2021;12:632418. https://doi.org/10.3389/fpsyt.2021.632418.
9. Qian J et al. *An update on the prevalence of eating disorders in the general population: a systematic review and meta-analysis.* Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 2022. 27(2): pp. 415–28. https://doi.org/10.1007/s40519-021-01162-z
10. Zhou S, et al. Exploring eating disorder topics on Twitter: machine learning approach. JMIR Med Inf. 2020;8(10):e18273. http://doi.org/10.2196/18273.
11. Howard LM, et al. *The role of denial in eating disorder development, assessment, and treatment*, in *Eating Disorders*. Springer; 2023. pp. 367–83.

12. Wu J, et al. Secular trends in the incidence of eating disorders in China from 1990 to 2017: a joinpoint and age–period–cohort analysis. Psychol Med. 2022;52(5):946–56. https://doi.org/10.1017/S0033291720002706.
13. De Choudhury M et al. *Predicting depression via social media*. in *Proceedings of the international AAAI conference on web and social media*. 2013. https://doi.org/10.1609/icwsm.v7i1.14432
14. Chancellor S et al. *Quantifying and predicting mental illness severity in online pro-eating disorder communities*. in *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 2016. https://doi.org/10.1145/2818048.2819973
15. Derenne J, Beresin E. Body image, media, and eating disorders—a 10-year update. Acad Psychiatry. 2018;42(1):129–34. https://doi.org/10.1007/s40596-017-0832-z.
16. Wang X, et al. Depression risk prediction for chinese microblogs via deep-learning methods: Content analysis. JMIR Med Inf. 2020;8(7):e17958. http://doi.org/10.2196/17958.
17. Sproat R, Emerson T. *The first international Chinese word segmentation bakeoff*. in *Proceedings of the second SIGHAN workshop on Chinese language processing*. 2003.
18. Zheng C, Huang M, Sun A. *ChID: A large-scale Chinese IDiom dataset for cloze test.* arXiv preprint arXiv:1906.01265, 2019.
19. Park E, Kim W-H, Kim S-B. *What topics do members of the eating disorder online community discuss and empathize with? an application of big data analytics*. in *Healthcare*. 2022. MDPI. https://doi.org/10.3390/healthcare10050928
20. Wang T et al. *Detecting and characterizing eating-disorder communities on social media*. in *Proceedings of the Tenth ACM International conference on web search and data mining*. 2017. https://doi.org/10.1145/3018661.3018706
21. Mathews SC, et al. Digital health: a path to validation. NPJ Digit Med. 2019;2(1):38. https://doi.org/10.1038/s41746-019-0111-3.
22. Guan L, et al. Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model. JMIR mental health. 2015;2(2):e4227.
23. Huang X et al. *Detecting suicidal ideation in Chinese microblogs with psychological lexicons*. in. 2014 *IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops*. 2014. IEEE. http://doi.org/10.1109/UIC-ATC-ScalCom.2014.48
24. Park E, Kim W-H. A retrospective literature review of eating disorder research (1990–2021): application of bibliometrics and topical trends. Int J Environ Res Public Health. 2022;19(13):7710. https://doi.org/10.3390/ijerph19137710.
25. Moessner M, et al. Analyzing big data in social media: Text and network analyses of an eating disorder forum. Int J Eat Disord. 2018;51(7):656–67. https://doi.org/10.1002/eat.22878.
26. Zhang Q, Gonçalves B. *Topical differences between chinese language Twitter and sina weibo*. in *Proceedings of the 25th International Conference Companion on World Wide Web*. 2016.

https://doi.org/10.1145/2872518.2890562
27. Xu Q, Wang Z. A Data-Driven Model for Automated Chinese Word Segmentation and POS Tagging. Comput Intell Neurosci. 2022;2022(1):7622392. https://doi.org/10.1155/2022/7622392.
28. Choi E, Chentsova-Dutton Y, Parrott WG. The effectiveness of somatization in communicating distress in Korean and American cultural contexts. Front Psychol. 2016;7:383. https://doi.org/10.3389/fpsyg.2016.00383.
29. Ryder AG, et al. The cultural shaping of depression: somatic symptoms in China, psychological symptoms in North America? J Abnorm Psychol. 2008;117(2):300. http://doi.org/10.1037/0021-843X.117.2.300.
30. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res. 2003;3(Jan):993–1022.
31. Kour H, Gupta MK. An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM. Multimedia Tools Appl. 2022;81(17):23649–85. https://doi.org/10.1007/s11042-022-12648-y.
32. Low DM, et al. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. J Med Internet Res. 2020;22(10):e22635. http://doi.org/10.2196/22635.
33. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. J Psychopathol Behav Assess. 2004;26(1):41–54. https://doi.org/10.1023/B:JOBA.0000007455.08539.94.
34. Lavender JM, et al. Dimensions of emotion dysregulation in bulimia nervosa. Eur Eat Disorders Rev. 2014;22(3):212–6. http://doi.org/10.1016/j.cpr.2015.05.010.
35. Röder M, Both A, Hinneburg A. *Exploring the space of topic coherence measures*. in *Proceedings of the eighth ACM international conference on Web search and data mining*. 2015. https://doi.org/10.1145/2684822.2685324
36. Grootendorst M. *BERTopic: Neural topic modeling with a class-based TF-IDF procedure.* arXiv preprint arXiv:2203.05794, 2022.

Figures

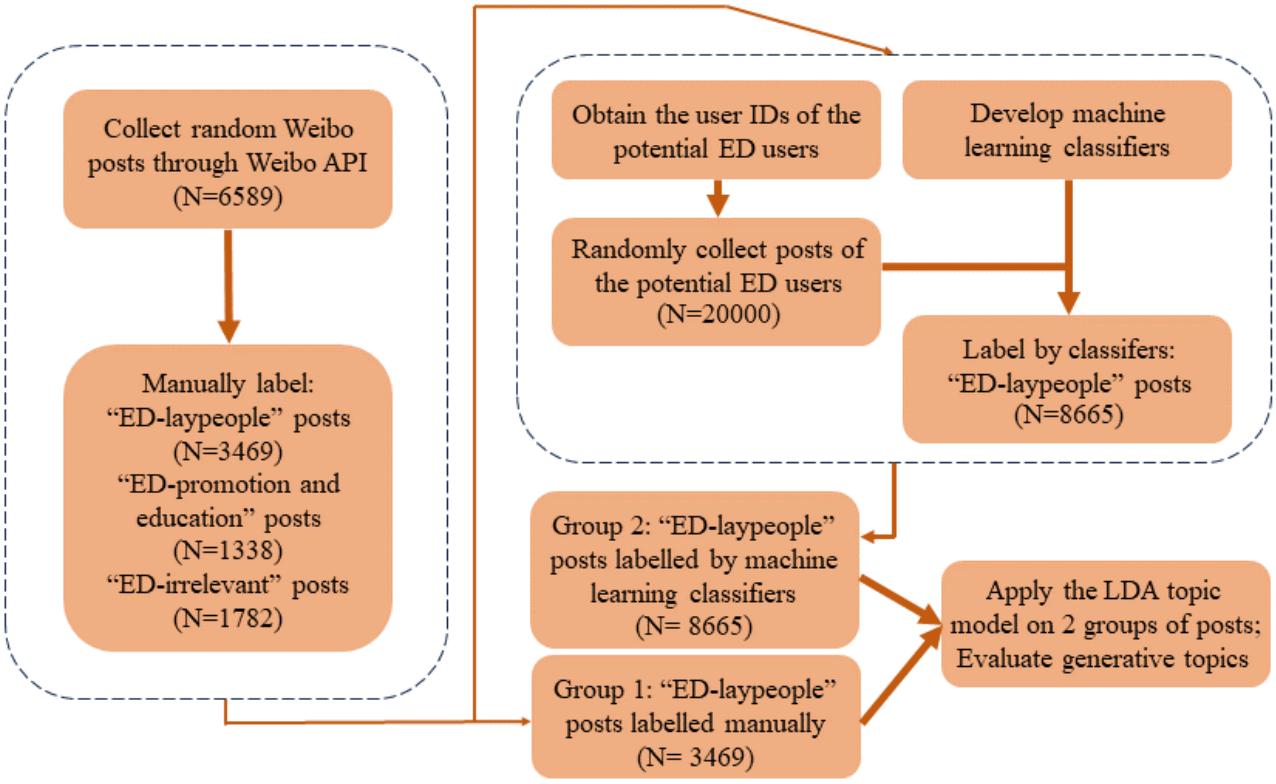


Figure 1

A two-stage pipeline of our study methodology

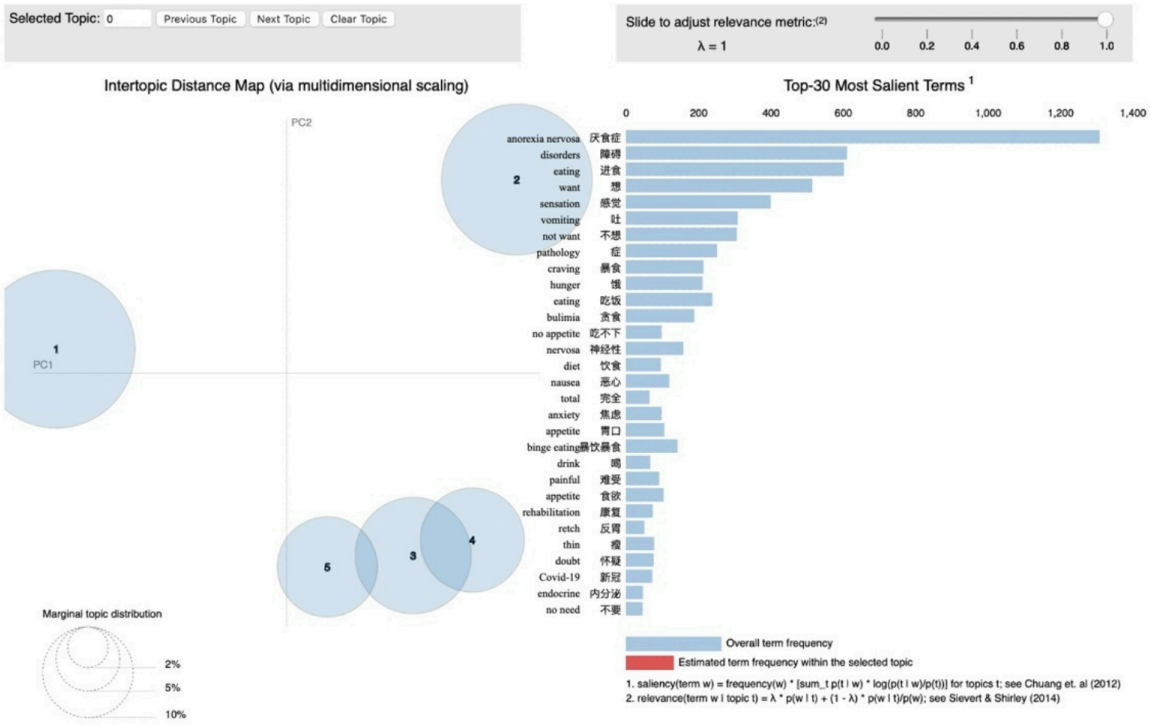


Figure 2

Intertopic Distance Map for Labeling Dataset Preview

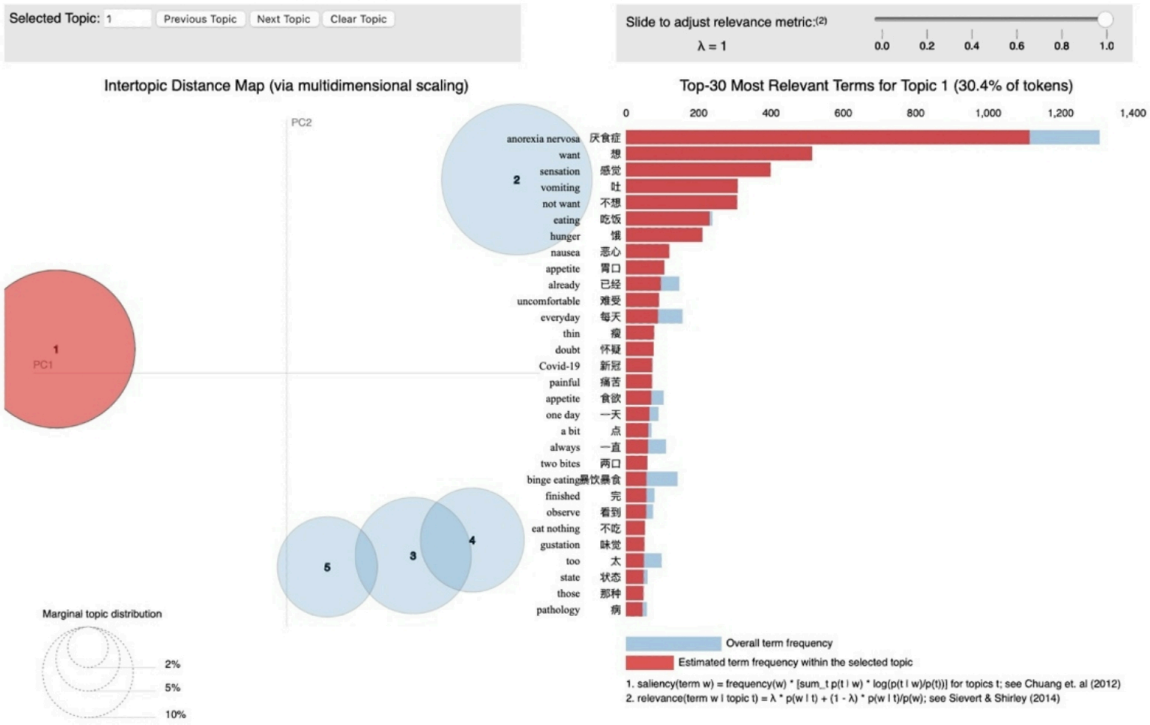


Figure 3

Intertopic Distance Map for Labeling Dataset (Top-30 most relevant terms for Topic 1)

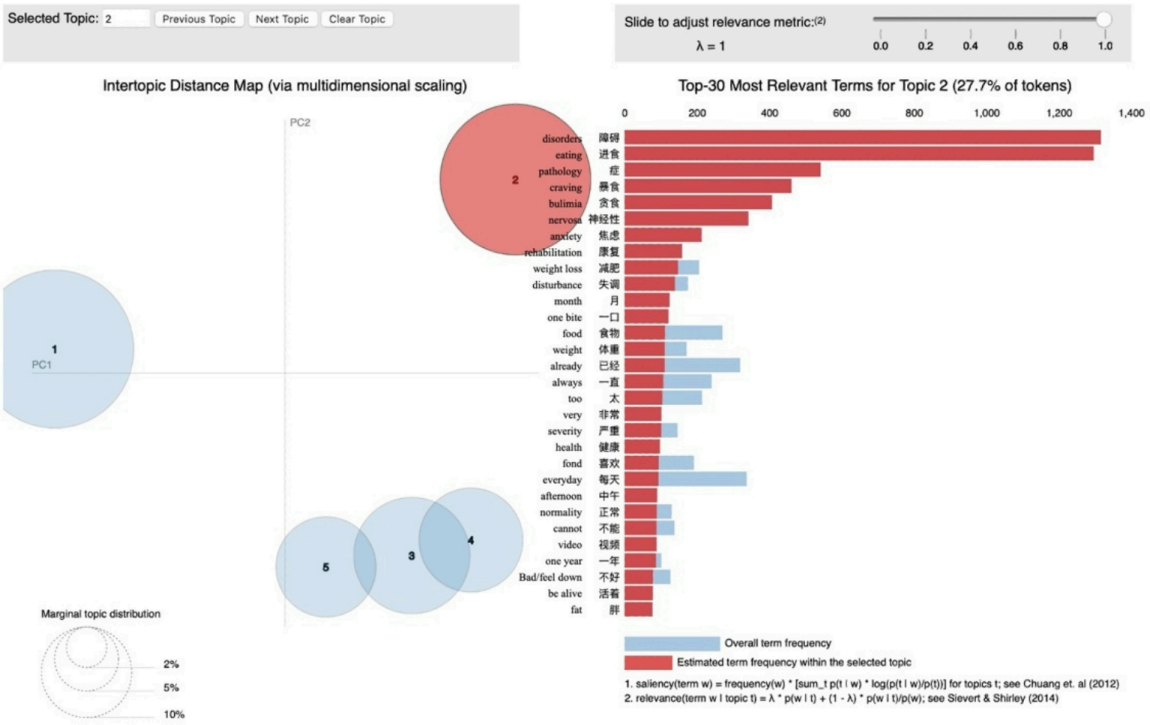


Figure 4

Intertopic Distance Map for Labeling Dataset (Top-30 most relevant terms for Topic 2)

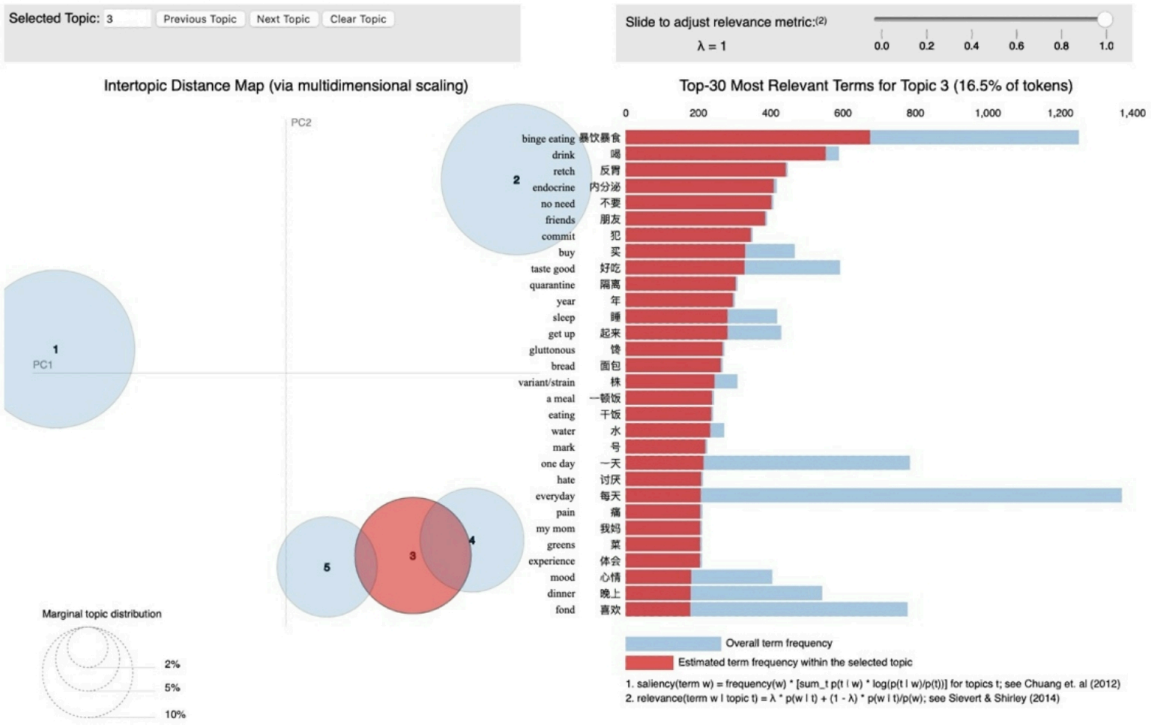


Figure 5

Intertopic Distance Map for Labeling Dataset (Top-30 most relevant terms for Topic 3)

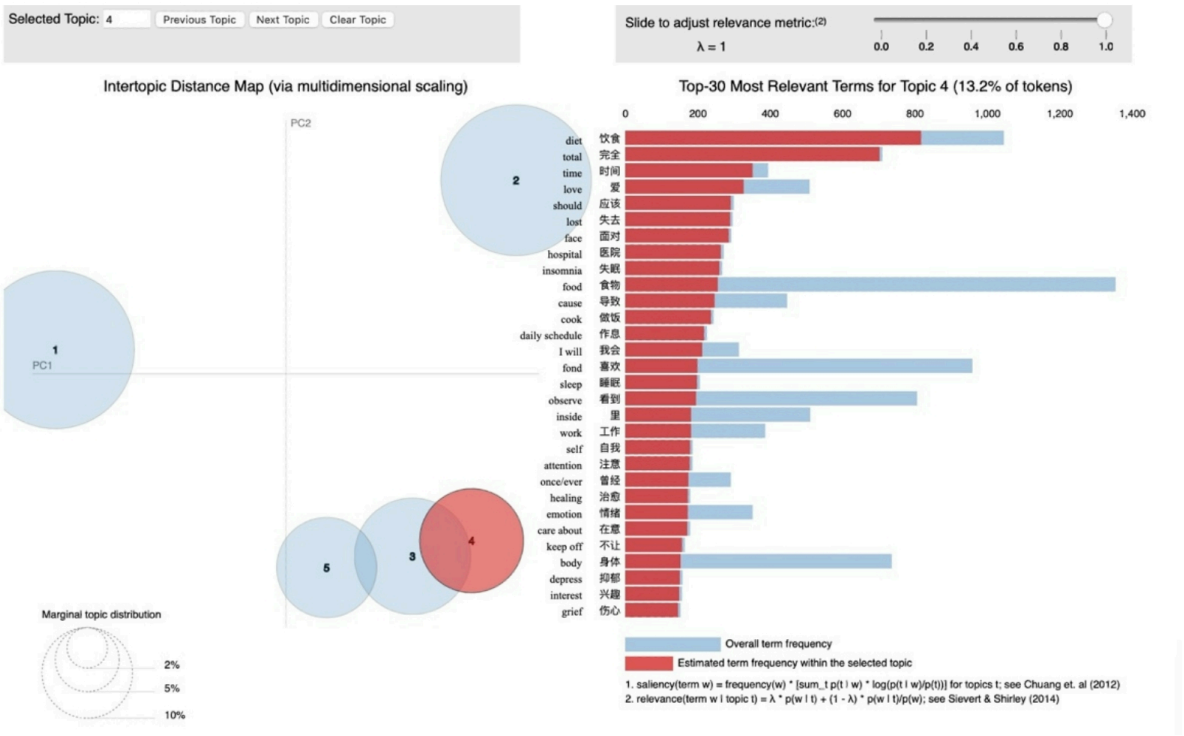


Figure 6

Intertopic Distance Map for Labeling Dataset (Top-30 most relevant terms for Topic 4)

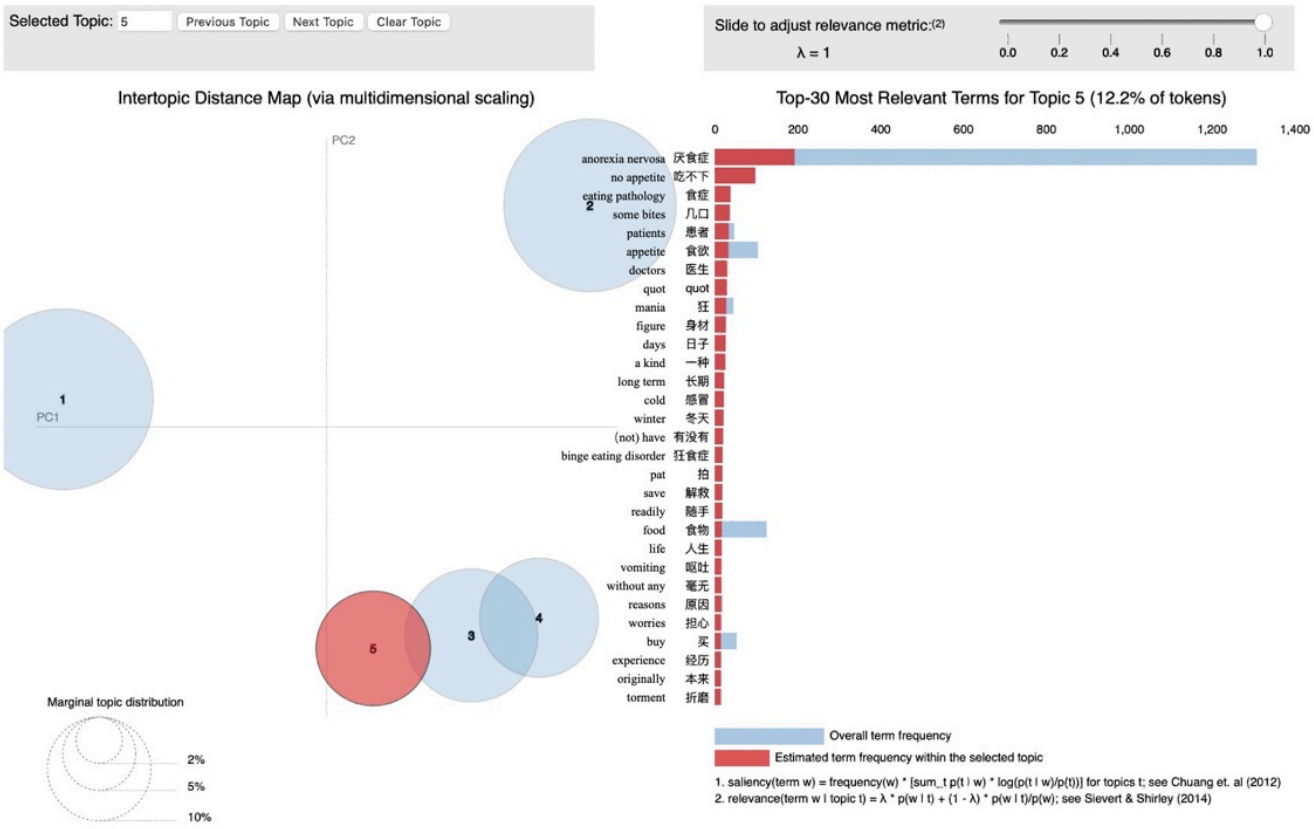


Figure 7

Intertopic Distance Map for Labeling Dataset (Top-30 most relevant terms for Topic 5)