Emotion Mining and Analysis of Reviews Based on Emotion Models: A Case Study of Douban Book Reviews (Postprint)

Authors: Nie Hui, Liu Mengyuan

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose/Significance] This study aims to explore methods for extracting and visualizing user emotions from unstructured user-generated content, conduct an in-depth analysis of user-generated content from a perceptual perspective, and discuss its application prospects.

[Method/Process] Taking book reviews from the Douban website as the analysis object, this study employs Chinese emotion lexicons and LDA latent topic modeling methods to achieve fine-grained extraction of emotional elements, and utilizes visualization techniques to analyze the emotional elements reflected in the review content.

[Results/Conclusion] The study finds that both topic analysis methods and lexicon-based methods can effectively extract user emotional elements from review content, but differences exist. Emotional topic modeling can provide more nuanced user emotions and perceptual information. Through fine-tuning of application scenarios, the methods involved in this study can be applied to various forms of review perception utility mining tasks, such as experience-based product recommendations.

Full Text

Preamble

ChinaXiv Cooperative Journal *Academic Exploration*
Emotion Mining and Analysis of Comments Based on Emotional Model — A Case Study on Douban Book Reviews
Nie Hui, Liu Mengyuan
School of Information Management, Sun Yat-sen University, Guangzhou 510006

Abstract

[**Purpose/Significance**] This study aims to explore methods for extracting and visualizing user emotions from unstructured user-generated content, enabling in-depth analysis of such content from a perceptual perspective and discussing its application prospects. [**Method/Process**] Taking book reviews from Douban as the analytical object, this research employs a Chinese-domain emotion lexicon and LDA latent topic modeling to refine fine-grained emotional elements, and utilizes visualization techniques to analyze the emotional components reflected in review content. [**Result/Conclusion**] The study finds that both topic modeling and lexicon-based methods can effectively extract user emotional elements from reviews, though with differences: emotional topic modeling can provide more nuanced user emotions and perceptual information. With scenario-specific adjustments, the methods involved in this research can be applied to various perceived utility mining tasks for reviews, such as experience-based product recommendations.

Keywords: user-generated content, emotion perception, review mining, information visualization
Classification Number: TP391
Citation Format: Nie Hui, Liu Mengyuan. Emotion mining and analysis of comments based on emotional model: A case study on Douban book reviews [J/OL]. *Knowledge Management Forum*, 2018, 3(6): 313-324 [citation date]. http://www.kmf.ac.cn/p/150/.

1 Introduction

As user-generated content has grown exponentially, sentiment analysis has become increasingly prevalent in both theoretical research and practical applications. Broadly defined, sentiment analysis involves computational research on identifying, extracting, and classifying opinions, emotions, sentiments, and attitudes expressed in text [1], encompassing tasks such as subjective/objective analysis, attitude analysis, emotion analysis, and opinion mining. It has been widely applied in public opinion monitoring [2], market forecasting [3], customer satisfaction surveys, and other domains.

Overall, sentiment analysis research has evolved from document-level to sentence-level analysis, from word frequency to topic modeling, and from binary sentiment to multi-dimensional emotion analysis. However, most sentiment analysis studies still focus primarily on determining positive/negative polarity, with relatively few addressing emotional intensity quantification or emotional states (such as joy, anger, anxiety, sadness, etc.), particularly in the Chinese domain. The English domain has seen more active exploration in this regard. Representative studies have examined the impact of fine-grained emotional states (or "emotions") on the perceived usefulness of reviews: L. Martin and P. Pu [4] demonstrated that emotional features positively contribute to review quality prediction; A. Felbermayr and A. Nanopoulos

[5] found that fine-grained emotional features significantly outperform other structural text features in influencing perceived review usefulness; and D. Yin et al. [6] discovered that "anxiety" significantly enhances perceived review usefulness while "anger" does not. These recent studies all indicate that more refined emotional states embedded in user-generated content have significant and direct impacts on receivers' perceptions and behaviors, and that the role of fine-grained emotions in the generation, dissemination, and effective utilization of user reviews cannot be ignored.

This research focuses on this topic, using Chinese book reviews as the analytical object to explore methods for extracting fine-grained emotions from book reviews. By introducing an LDA model to construct emotional topics in book reviews and visualizing the emotional states within user-generated content, this study also uses literary work reviews as a case study to explore the application prospects of fine-grained sentiment analysis.

2.1 Sentiment Analysis and Emotion Lexicons

Early sentiment analysis focused on document-level granularity, considering only positive and negative states. Subsequent research delved into the sentence level, emphasizing not only sentiment polarity but also sentiment intensity, sentiment targets, and contextual situations. Shi et al. [7] used a self-built fuzzy sentiment ontology to identify product features in mobile phone and wedding photography reviews, distinguish sentiment polarity, and refine sentiment intensity calculations by introducing semantic elements such as modifiers. Nie Hui [8] used latent topic models to aggregate user opinions on digital product reviews while employing syntactic analysis for sentiment quantification, achieving concise summarization of online word-of-mouth. Jin Yan [9] explored the impact of emotional factors on the quality of user-generated content in microblogs, quantifying positive, neutral, and negative sentiments and further subdividing positive and negative emotions into high, medium, and low levels before introducing them into a research model and using the ROST content mining system for sentiment analysis.

As sentiment analysis research has deepened, researchers have drawn on psychological studies to refine positive/negative sentiments into emotional features. Psychology provides clear definitions of "emotion." R. Bagozzi et al. [10] define emotion as "a mental state resulting from cognitive appraisal of events," which is "concrete, has clear origins, and can lead to specific behavioral tendencies." Based on this understanding of emotion, scholars in psychology have proposed "emotion models," among which the most representative is Plutchik's emotion model [5], which depicts eight basic emotion categories as the famous "Wheel of Emotions." Emotion models have extensive applications in psychology.

Psychological theory has facilitated the development of emotion lexicons. In the English domain, the NRC (National Research Council Canada) lexicon is widely recognized. With emotion lexicons, fine-grained emotion analysis can be conducted. R. Ullah et al. [11] used the SentiWordNet emotion lexicon to conduct an in-depth analysis of user emotion distributions in reviews of 17 product categories on Amazon. L. Martin and P. Pu [4] explored the predictive effect of emotional variables on multiple review quality dimensions based on the GALC emotion lexicon, finding that multi-dimensional emotion analysis outperforms binary sentiment polarity analysis. In the Chinese domain, the Chinese Sentiment Lexical Ontology from the Information Retrieval Laboratory of Dalian University of Technology (http://ir.dlut.edu.cn/), referencing Ekman's emotion model [12], categorizes emotions into seven major categories (joy, fear, surprise, sadness, disgust, anger, and goodness) with 21 subcategories, and has been widely applied in sentiment analysis. Cao Yu et al. [13] started from emoticons, identified emotional words in sentences, and combined them with the *Tongyici Cilin* (Chinese synonym dictionary) to expand existing multi-emotion ontologies.

2.2 Review Analysis

Product reviews are the primary object of review analysis. Product reviews can be categorized into search-type reviews (represented by electronic products) and experience-type reviews (represented by movies and books) [14]. The former focuses on product features and tends to use objective indicators for description, while the latter emphasizes user participation experiences, tends toward subjective perceptual descriptions, and contains rich personal feelings and experiential descriptions with abundant emotional content. S. M. Mudambi and D. Schuff [14] studied the impact of review length and user star ratings on perceived review usefulness, finding that conclusions differ between experience-type and search-type product reviews. A. Ghose and P. G. Ipeirotis [15] examined how readability elements such as grammatical errors, word choice, and sentence length affect movie review quality. J. H. Lee et al. [16] focused on the role of emotions, finding that removing negative reviews has little effect on boosting movie box office sales. M. Ko et al. [17] screened key features with positive emotions based on details in movie reviews for refined movie recommendations. In these studies, sentiment indicators are often represented by user ratings or obtained through statistical analysis of positive/negative vocabulary.

Domestic research on experience-type product reviews has primarily focused on analyzing factors affecting review quality, information extraction, and opinion mining. Yin Guopeng et al. [18] introduced reviewer characteristics to explore the impact of various movie review elements, including user sentiment based on review star ratings, on perceived review usefulness from a conformity perspective. Xue Bozhao [19] used the text mining tool LingPipe to refine sentiment features in Amazon book review data into indicators such as positive sentiment difference, subjectivity expression, and differences between review ratings and product ratings. Zhang Li et al. [20] manually identified high-frequency words in Dangdang book reviews and conducted user attention analysis on review content from dimensions including emotional intensity, subjective/objective expression, evaluation, and book content, finding that book content received the most attention. Zhu Zhenyuan [21] analyzed reviews of different book types on Amazon, extracting nine categories of elements including content, sentiment orientation, readers, and reviewers based on classification, syntactic analysis, and the SentiWordNet emotion lexicon, and generated book review summaries after quantification and integration. These studies demonstrate in-depth exploration of experience-type product reviews such as book and movie reviews. However, such research tends to consider broad review features with many non-content factors and does not deeply engage with content itself, let alone conduct specialized fine-grained subjective emotion analysis based on content.

3 Research Core and Framework

As an analytical task for subjective text, this research focuses on content-based emotional element analysis, aiming to explore methods for extracting "emotions" from review content and conducting emotional topic analysis of book reviews from a user perception perspective. This study expands Chinese-domain sentiment analysis research and explores application prospects of fine-grained sentiment analysis through case studies.

This research selects long-form book reviews from Douban (https://www.douban.com/) as analytical data, addressing three core questions: (1) How to effectively extract emotional features from text content; (2) How to condense emotional features and construct an emotion model based on review content; and (3) How to visualize user emotions and topics embedded in review content.

The entire research process consists of six modules: data collection, preprocessing, emotion and emotional feature extraction, emotional topic modeling, emotional topic description, and visualization, as shown in [Figure 1: see original paper]. The collection module selects third-party review websites as data sources and filters comments based on vote counts and usefulness during the collection phase to reduce noise and ensure later modeling effectiveness. Next, the raw corpus undergoes preprocessing including sentence/word segmentation, stopword filtering, and part-of-speech tagging. Emotion and emotional feature identification is achieved through part-of-speech-based filtering and dictionary matching. Identified emotional and sentiment features are then used for emotion classification calculations and emotional topic construction, with various feature results aggregated and output in visualized form, thereby condensing natural language review information into structured product characteristics.

4.1 Emotion Theory and Emotion Lexicon

Emotion represents human subjective feelings and thoughts [22]. Due to the complexity of human emotions, emotion categorization has been a key focus of research. The famous Plutchik Wheel of Emotions [5] divides emotions into eight basic types: anger, disgust, fear, sadness, anticipation, joy, surprise, and trust, which are pairwise opposite (e.g., joy vs. sadness) and vary in intensity.

Complex emotions derive from basic emotions—for example, combining "joy" and "trust" produces "love." Based on emotion theory, emotion lexicons have emerged. In the Chinese domain, the "Chinese Sentiment Lexical Ontology" from the Information Retrieval Laboratory of Dalian University of Technology has a similar structure to the NRC lexicon, with complete annotations, and is currently a widely used Chinese sentiment/emotion lexicon. This lexical ontology references Ekman's emotion model [12], which classifies emotions into six categories (joy, fear, surprise, sadness, disgust, and anger), and adds a "goodness" category. The "goodness" category is further subdivided into positive emotional states including respect, praise, belief, and affection. The entire lexicon covers 27,466 entries, divided into 7 major categories and 21 subcategories, with each entry annotated with detailed attributes including part of speech, emotion type, emotional intensity, and polarity. The author uses this Chinese lexicon to identify emotional words from review content through matching, and classifies and counts them according to their emotional categories in the lexicon to obtain emotion state distributions based on review content.

[Figure 2: see original paper] Plutchik's Wheel of Emotions [5]

4.2 LDA Latent Topic Model

LDA (Latent Dirichlet Allocation) is a multi-level generative probabilistic model comprising three layers: word, topic, and document. In the LDA model, D is the document set (|D|=N), W is the word set (|W|=M), and represents latent topics (| |=K). Document d D is generated by random mixing of i, represented as a sequential pattern on the word set d{w1,w2,…,wi,…}, where wi W; while topic i(i=1,2,…,K) is a multinomial distribution on W. The graphical model of LDA is shown in [Figure 3: see original paper].

[Figure 3: see original paper] Graphical model representation of LDA [8]

α and β are important parameters of the model. α relates to the prior Dirichlet distribution of topics, reflecting the relative strength of latent topics in document set D; β characterizes the probability distribution of topics themselves. At the topic level, i (i=1,2,…,K) is the "topic-word" distribution, parameterized by a K×M matrix β. At the document level, d(d=1,2,…,N) is the "document-topic" distribution, representing document d's distribution in the topic space, determined by α. At the word level, zi (i=1,2,…,M) represents the latent topic component assigned to each word item in document d, following a multinomial distribution with parameter d. The core of constructing an LDA model is inferring latent variables, i.e., determining and . Given the word distribution W of the document set, the model reversely derives zi(i=1,2,…,M), and then infers t (t=1,2,…,K) and d (d=1,2,…,N). Model construction involves multiple unknowns and generally uses approximation methods for solution, such as using Gibbs sampling to estimate the topic distribution of w [8].

This study extracts emotional words using the "Chinese Sentiment Lexical Ontology" to construct review descriptions based on emotional word sequences, then uses the LDA latent topic model to conduct emotional topic analysis on the review set, mining user emotions in review content.

4.3 Emotion Visualization

Visualization, as a method that can intuitively display data characteristics, provides convenience for people to integrate massive data and extract patterns and rules. There are many information visualization models, with the classic one being the information visualization reference model proposed by S. K. Card [23] (see [Figure 4: see original paper]). The information visualization reference model involves four parts: data collection, data processing and transformation, visual mapping, and human visual perception. Visual mapping is the key to the transformation process. Visualization techniques are developed according to data object characteristics. Basic statistical visualization methods include bar charts, pie charts, radar charts, etc., while text visualization methods focusing on how to visually represent semantic features in text and their evolution over time include static word clouds and dynamic topic rivers.

[Figure 4: see original paper] S. K. Card's information visualization reference model [23]

The visualization techniques involved in this study include statistical bar charts and pie charts, as well as text-based word clouds and topic clustering visualization. For topic clustering visualization, the author applied LDAvis [24]. LDAvis is an interactive LDA latent topic modeling visualization tool developed by C. Sievert and K. E. Shirley based on R language, which can dynamically display the intuitive effects of LDA latent topic clustering and uses interactive methods to adjust parameters to observe topic convergence effects and the distribution of key terms on topics.

5 Experiments and Analysis

This study takes book reviews as its object to conduct identification, extraction, aggregation, and visual analysis of user emotions. The research involves three main experiments: emotion extraction based on emotion lexicons, user emotion mining and analysis based on emotional topic modeling, and related visualization. The analysis language is R, and the data visualization part mainly uses word clouds from Tagxedo (http://www.tagxedo.com/) and LDAvis diagrams.

5.1 Data Acquisition and User Emotion Identification Based on Emotion Lexicon

The corpus is sourced from Douban Books (https://book.douban.com/). As a third-party review platform, Douban has generally higher credibility than e-commerce platforms. Based on book reputation rankings from major book purchasing websites, this study selected 12 bestselling literary works including

The Devotion of Suspect X as analysis objects. Between October and November 2016, a web crawler was used to capture reviews from the "Book Reviews" section (distinct from the "Short Reviews" section) of relevant book pages, collecting a total of 23,683 reviews. Among these, 12,316 were long-form reviews with more than 300 characters, accounting for approximately 52% of the captured reviews. To ensure review quality, only reviews with usefulness votes greater than 5 were retained for actual analysis, resulting in 1,555 review data points, accounting for about 13% of the long-form reviews. The average valid data for the 12 books was approximately 126 reviews each.

Raw book reviews underwent preprocessing including word segmentation, sentence segmentation, and part-of-speech tagging. Nouns, adjectives, and verbs were retained as candidates, and dictionary matching was used to identify emotional words from candidate terms, assigning them emotion labels and intensity values based on lexicon annotations. The relevant experiments involved 1,829 emotional ontology terms, accounting for approximately 7% of the total emotional ontology vocabulary.

5.2 Emotional Topic Modeling

Based on identified emotional words, this study employed the latent topic model LDA to cluster topics in emotional content descriptions, thereby better condensing emotional features embedded in review content.

5.2.1 Establishing Emotional Topic Model

First, parameter optimization experiments were conducted for model selection. According to research by Hu Jiming and Chen Guo [25], LDA hyperparameters α and β achieve good model effects when set to {0.1, 0.5, 0.9} and {0.01, 0.1} respectively. Therefore, this study combined these two sets of values to compare the effects of six parameter settings to identify optimal parameters. Model evaluation used inter-topic similarity, as shown in formula (1):

Formula (1)

where ci and cj correspond to feature vectors of clustered topics ti and tj respectively, and win is the weight of word n in topic i. The model's similarity is the mean of clustered topic similarities; the smaller this metric, the more significant the differences between topic clusters and the better the clustering effect. Taking *The Little Prince* as an example (see [Figure 5: see original paper]), when α=0.9 and β=0.01, the topic model achieves the best clustering effect. In the figure, the X-axis corresponds to the number of topics, and the Y-axis corresponds to inter-topic similarity in the model. Further, using LDAvis and dynamic parameter adjustment, the optimal number of topics K was determined. As shown in [Figure 6: see original paper], when K=5, the topics in the LDA model are relatively far apart from each other, achieving the best topic cohesion effect.

[Figure 5: see original paper] Parameter settings and model similarity index distribution
[Figure 6: see original paper] Topic clustering visualization LDAvis (K=5) (screenshot)

5.2.2 Book Review Emotional Topics Based on LDA Model

User emotions in book reviews mainly involve liking, friendship, loneliness, love, and memory, representing overall positive emotional states. In contrast, reviews of *Journey Under the Midnight Sun*, which features desolate love and calm reasoning, express emotions of despair, understanding and affirmation, loss and misfortune, and dedication, showing overall negative emotional states, as shown in Table 2. The topic clustering results reveal people's perceptions and emotions about works. Using *The Little Prince* and *Journey Under the Midnight Sun* as examples, Table 1 and Table 2 list the emotional topic analysis results for these two books. Although word distribution in long reviews tends to be relatively dispersed and the top few items have low proportions, they can still reflect the emotional color of reviews to some extent. As shown in Table 1, *The Little Prince*'s emotional topics align well with the book's content from a reading experience perspective.

The Little Prince emotional topics and example sentences
Journey Under the Midnight Sun emotional topics and example sentences

5.3 Visualization of User Emotion Extraction Content Based on Emotion Lexicon and Emotional Topic

To more intuitively reflect research results, further experiments visually describe user emotion extraction content through visualization methods. For lexicon-based review content sentiment analysis, taking *The Little Prince*—which showcases human loneliness and solitude while praising truth, goodness, and beauty—as an example, the review content emotion distribution is shown in [Figure 7: see original paper]. The figure shows that the book is dominated by positive and warm emotions (joy, goodness), while also containing some sadness (sorrow) and negative emotions (disgust). The emotions in user reviews are generally consistent with the emotional state presented in the book content. However, for individual reviews, the author randomly extracted two cases and used pie charts to plot the emotion category distributions (see [Figure 8: see original paper]). Clearly, the reviewer on the left felt sadness, while the reviewer on the right felt more beauty and kindness. Through visualization, differences in reader perception states can be clearly observed.

Further word cloud analysis of book review sentiment distribution is shown in [Figure 9: see original paper]. The left side of the figure shows a word cloud generated based on all content words, while the right side shows a word cloud generated based on emotional words. From the word clouds, we can observe that *The Little Prince* brings people feelings of "happiness," "loneliness," "beauty," "joy," etc., consistent with the work's theme. Comparing the two word clouds reveals that the emotional word cloud can more intuitively present readers' feelings in book reviews. Evidently, for experience-type product reviews, analysis based on emotional words can better characterize themes.

For emotional topic-based review content sentiment analysis, word clouds are used to visualize clustered topics. As seen in [Figure 10: see original paper], the word cloud for *The Little Prince* emotional topic 3 intuitively presents the core term "loneliness" for the corresponding topic, while also including important topic words such as "friend," "lonely," and "pain." This shows that emotional topic analysis can better reveal reviewers' perceptions of books—not only overall perceptions but even perceptions of plot or characters.

[Figure 7: see original paper] *The Little Prince* emotion attribute statistics per review (separate)
[Figure 8: see original paper] *The Little Prince* review emotion attribute distribution (case analysis)
[Figure 9: see original paper] Comparison of all-content word cloud and emotional word cloud for *The Little Prince*
[Figure 10: see original paper] Word cloud for *The Little Prince* emotional topic 3

6.1 Book Style Comparison Based on User Perception

Different books have different emotional styles and evoke different perceived emotions in readers, with user-perceived emotions being consistent with the emotional state of the book content. By analyzing the emotional distribution and topics of reviews, we can refine book styles and achieve perception-based book retrieval and recommendation. [Figure 11: see original paper] shows a comparison of emotional polarity distributions across 12 books, visually presenting differences in emotional styles among different books. For instance, books like *Journey Under the Midnight Sun*, *One Hundred Years of Solitude*, and *To Live* tend toward negative emotions, while *The Storied Life of A.J. Fikry* and *The Miracles of the Namiya General Store* tend toward positive emotions—conclusions that basically align with the emotional states reflected in the books' thematic content. [Figure 12: see original paper] shows a comparison of fine-grained emotion distributions for the 12 books, depicting readers' feelings in greater detail. *The Little Prince*, themed around "love," blends joy and sorrow; *One Hundred Years of Solitude* and *To Live* tell stories of people against the backdrop of domestic and international social transformation, depicting the rise and fall of people and families and their suffering history, making "sadness" the main tone of both works; while *Ordinary World* reflects truth, goodness, and beauty in heavy life, gaining more positive recognition and feelings from readers.

6.3 Review Analysis Based on Emotional Features

Emotion is unique to subjective text, and fine-grained sentiment analysis has important application value in subjective text mining, such as analyzing differences in reader-perceived emotions and concerns under different star ratings. Taking *Journey Under the Midnight Sun* as an example, as shown in [Figure 14: see original paper], in 5-star reviews, reviewers tend to provide more detailed descriptions of their perceptions, while in 1-star reviews, reviewers tend to vent their dissatisfaction. From this perspective, for experience-type products, businesses can obtain more diverse user experiences from positive reviews, making positive reviews more valuable for utilization. We therefore believe that emotional characteristics are essential traits of subjective review information, and once these traits are grasped, such information can be effectively controlled and managed to better realize its information value.

By mining emotions in reviews of controversial books, we can capture the direction and reasons for the flow of mainstream reader opinions, thereby positively impacting book evaluation and marketing. Taking *The Three-Body Problem* as an example, [Figure 13: see original paper] shows that reader opinions on this book are polarized, but opposing views emerge at different time periods and gradually converge. The reduction or increase in differences may be closely related to events such as book awards and film adaptations.

[Figure 11: see original paper] Comparison of positive/negative sentiment tendencies across 12 books
[Figure 12: see original paper] User review emotion distribution (percentage chart)
[Figure 13: see original paper] Positive/negative sentiment tracking for *The Three-Body Problem*
[Figure 14: see original paper] Relationship between average user emotion values and review star ratings for *Journey Under the Midnight Sun*

7 Conclusion and Future Research Plans

This study takes Douban book reviews as its analytical object, using a Chinese-domain emotion lexicon and LDA latent topic modeling to refine fine-grained emotional elements, and employs visualization techniques to analyze emotional components reflected in review content. The study found that: (1) Both topic modeling and lexicon-based methods can effectively extract user emotional elements from review content, each with its own strengths—the former can provide more nuanced emotional descriptions and user perceptions, while the latter can intuitively present emotional distributions embedded in review content. (2) The emotional states extracted from all user reviews are consistent with the emotional themes of the books, enabling the clarification of books' emotional style positioning based on emotional states in reviews. Meanwhile, reader-perceived emotions vary, requiring further understanding of readers' emotional appeals during reading. Combining both approaches, businesses can implement emotion-based advertising recommendations and emotional marketing for experience-type products. (3) Various visualization methods based on emotional features—including word clouds, bar charts, and pie charts—are better ways to present user perceptions. (4) Describing books based on user emotion perception is an effective method for exploring book analysis, retrieval, and recommendation based on emotions.

In future work, the author will further improve experiments by attempting to identify the opposition relationships between the existing seven emotions in the emotional ontology and emotional topics. Building upon word matching, the author will attempt to incorporate grammatical features such as degree adverbs and negation words to more accurately and realistically reveal emotional characteristics reflected in reviews, and will introduce more visualization methods. Additionally, the author will explore practical applications of fine-grained emotion mining, such as multiple comparisons of book styles based on user-perceived emotions and tracking mainstream opinions on controversial books, and will provide suggestions for deeper applications. On the other hand, the author will also introduce fine-grained sentiment analysis into search-type product review analysis and explore issues related to emotional marketing from a business operations perspective.

References

[1] RAVI K. A survey on opinion mining and sentiment analysis: tasks, approaches and applications[J]. *Knowledge-based systems*, 2015, 89(C):14-46.
[2] LAN Y X, XIA Y X, LIU B Y, et al. Research on the evolution mechanism and trend prediction of netizens' emotions for public opinion big data[J]. *Journal of Intelligence*, 2017, 36(11):134-140.
[3] ZHANG X D, YUAN D L. Research on the impact of investor sentiment based on Weibo on the stock market[J]. *Journal of Intelligence*, 2017, 36(8):81-87.
[4] MARTIN L, PU P. Prediction of helpful reviews using emotions extraction[C]// Twenty-Eighth AAAI conference on artificial intelligence. Palo Alto: AAAI Press, 2014:1551-1557.
[5] FELBERMAYR A, NANOPOULOS A. The role of emotions for the perceived usefulness in online customer reviews[J]. *Journal of interactive marketing*, 2016, 36(11):60-76.
[6] YIN D, BOND S D, ZHANG H. Anxious or angry? effects of discrete emotions on the perceived helpfulness of online reviews[J]. *MIS quarterly*, 2014, 38(2): 539-560.
[7] SHI W, WANG H W, HE S Y. Chinese online review sentiment analysis based on semantics[J]. *Journal of the China Society for Scientific and Technical Information*, 2013, 32(8):860-867.
[8] NIE H. Aggregation and quantification of product review opinions under latent topic model[J]. *Journal of the China Society for Scientific and Technical Information*, 2017, 36(6):565-573.

ChinaRxiv 中X

[9] JIN Y. UGC quality evaluation model based on emotion analysis[J]. *Library and Information Service*, 2017, 61(20):131-139.
[10] BAGOZZI R, GOPINATH M, NYER P U. The role of emotions in marketing[J]. *Journal of the academy of marketing science*, 1999, 24(3):149-167.
[11] ULLAH R, AMBLEE N, KIM W, et al. From valence to emotions: exploring the distribution of emotions in online product reviews[J]. *Decision support systems*, 2016, 81(5):41-53.
[12] EKMAN P. An argument for basic emotions[J]. *Cognition & emotion*, 1992, 6(3/4):169-200.
[13] CAO Y, WANG M Y, HE H X. Research on multi-emotion classification of Weibo text based on emotion dictionary expansion[J]. *Journal of Intelligence*, 2016, 35(10):185-189.
[14] MUDAMBI S M, SCHUFF D. What makes a helpful online review? a study of customer reviews on amazon.com[J]. *Social science electronic publishing*, 2010, 34(1):185-200.
[15] GHOSE A, IPEIROTIS P G. Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics[J]. *Social science electronic publishing*, 2010, 23(10):1498-1512.
[16] LEE J H, SUN H J, PARK J H. The role of entropy of review text sentiments on online WOM and movie box office sales[M]. New York: Elsevier Science Publishers B. V., 2017.
[17] KO M, KIM H W, YI M Y, et al. MovieCommenter: Aspect-based collaborative filtering by utilizing user comments[C]// International conference on collaborative computing: networking, applications and worksharing. Piscataway: IEEE, 2012:362-371.
[18] YIN G P, LIU W W, ZHU S. Research on the influencing model of online review helpfulness in online communities—Based on the perspective of information adoption and social network[J]. *Library and Information Service*, 2012,56(16): 140-147.
[19] XUE B Z. Research on influencing factors of consumer online review helpfulness in the book domain[D]. Harbin: Harbin Institute of Technology, 2011.
[20] ZHANG L, ZHANG L, ZHANG Y, et al. Text analysis of online book reviews based on Chinese word segmentation and word frequency statistics[J]. *Information System Engineering*, 2011(7):71-73.
[21] ZHU Z Y. Research on content mining and integration of online book reviews based on information classification[J]. *Library and Information Service*, 2016, 60(1):114-124.
[22] LIU B. Sentiment analysis and opinion mining[J]. *Synthesis lectures on human language technologies*, 2012, 5(1): 1-167.
[23] CARD S K, MACKINLAY J D, SHNEIDERMAN B. Readings in information visualization - using vision to think[C]// Series in interactive technologies. San Francisco: Morgan Kaufmann Publishers Inc., 1999.
[24] SIEVERT C, SHIRLEY K E. LDAvis: A method for visualizing and interpreting topics[EB/OL].[2018-01-23]. https://www.researchgate.net/publication/265784473_{LDAvis}A_{{{m
[25] HU J M, CHEN G. Content theme mining and evolution based on dynamic

LDA topic model[J]. *Library and Information Service*, 2014, 58(2):138-142.

Author Contributions:
Nie Hui: Overall design and conceptual guidance, data collection and organization, paper revision;
Liu Mengyuan: Experimental analysis, initial paper drafting.

Emotion Mining and Analysis of Comments Based on Emotional Model — A Case Study on Douban Book Reviews

Nie Hui, Liu Mengyuan
School of Information Management, Sun Yat-sen University, Guangzhou 510006

Abstract: [Purpose/significance] This study aims to explore the methods on extracting and visualizing users' emotions from unstructured user-generated content, analyze user-generated content from a perceptual level, and discuss the related application prospects. [Method/process] The research took book reviews of Douban as analysis object. Emotional dictionary in Chinese domain and LDA latent topic model were used to refine the fine-grained emotional elements. And further, visualization techniques helped to analyze the emotional elements reflected in the review content. [Result/conclusion] The study found that both latent topic model and emotion dictionary can effectively extract the user emotion elements in the content of the review, even though some difference still exists, such as the emotional topic model can provide more exquisite results. By fine-tuning the application scenario, the methods used in this study can be applied to various forms of perceived utility mining tasks about reviews, like experience-based products recommendation.

Keywords: user-generated content, emotion perception, review mining, information visualization

Note: Figure translations are in progress. See original paper for figures.