Journal of Trends in Financial and Economics
ISSN: 3007-696X
DOI: 10.61784/jtfe3003

improving rating prediction accuracy through advanced text summarization and sentiment analysis techniques

Michael Muskan[1*], Jude Berner[1], Jenny Winmark[2]
[1] *School of Computer Science, University of Bristol, Bristol, UK.*
[2] *Department of Computer Science, University of Manchester, Manchester, UK.*
Corresponding Author: Michael Muskan, Email: ix940421@bristol.ac.uk

Abstract: In the era of, online reviews have become a cornerstone of decision-making for consumers. This study addresses the challenging task of predicting ratings for long-form movie reviews, a problem that has been less explored compared to short review analysis. We propose a novel approach that combines advanced text summarization techniques with sentiment analysis to improve rating prediction accuracy. Utilizing an enhanced TextRank algorithm and Support Vector Machine (SVM) classification, our method demonstrates superior performance in predicting ratings for extensive movie reviews. The study uses a large dataset from Douban, a popular Chinese social networking service, and shows that summarized reviews can match or exceed the prediction accuracy of full-length reviews. Our findings highlight the effectiveness of integrating sentiment features and position-based weighting in the summarization process, opening new avenues for processing and analyzing long-form user-generated content.
Keywords: Rating prediction accuracy; Digital content consumption; E-commerce; TextRank algorithm

1 INTRODUCTION

The proliferation of e-commerce and content platforms has led to an explosion of online reviews, transforming how consumers make decisions about products and services [1]. In the realm of entertainment, movie reviews play a particularly crucial role, influencing viewer choices and box office performance [2]. While short reviews and ratings are commonplace, long-form reviews offer deeper insights but present unique challenges for automated analysis [3].

Traditional approaches to rating prediction have focused primarily on short reviews, leveraging techniques such as sentiment analysis, topic modeling, and collaborative filtering [4, 5]. However, these methods often fall short when applied to long-form content, where the relationship between textual content and numerical rating becomes more complex [6].

The challenge of long-form review analysis is multifaceted:

1) Information Density: Long reviews contain a mix of relevant and irrelevant information, making it difficult to isolate key rating indicators [7].
2) Structural Complexity: The narrative structure of long reviews may not follow a consistent pattern, with sentiment and opinion scattered throughout the text [8].
3) Nuance and Context: Longer reviews often contain nuanced opinions that may not be captured by simplistic sentiment analysis [9].

To address these challenges, we propose a novel approach that combines advanced text summarization techniques with sentiment analysis to distill the most relevant information from long-form reviews for rating prediction. Our method builds upon the TextRank algorithm [10], incorporating sentence position weighting and sentiment features to create more informative summaries.

The main contributions of this paper are:

1) A novel text summarization approach tailored for long-form movie reviews, integrating sentiment analysis and position-based weighting.
2) An extensive evaluation of the impact of text summarization on rating prediction accuracy for long reviews.
3) Insights into the relationship between review structure, sentiment, and rating in long-form content.

2 RELATED WORK

2.1 Rating Prediction in Online Reviews

Rating prediction has been a focus of research in recommender systems and sentiment analysis. Early works relied heavily on collaborative filtering approaches [11], while more recent studies have incorporated textual content analysis. Liu [12] proposed a joint sentiment-topic model for short review rating prediction, achieving significant improvements over baseline methods. For product reviews, McAuley and Leskovec [13] developed a model that combines latent dimensions of user and product factors with review text, demonstrating the value of integrating textual and non-textual features.

2.2 Text Summarization Techniques

Automatic text summarization has seen significant advancements with the rise of deep learning. Extractive methods, which select existing sentences from the text, include graph-based approaches like TextRank [10] and LexRank [14]. More recent work has focused on abstractive summarization, where new sentences are generated. See [15] introduced a pointer-generator network that combines copying words from the source text with generating novel words. BART [16] and PEGASUS [17] represent state-of-the-art pre-trained models for abstractive summarization. Liu [18] introduce a novel approach to news summarization, combining deep learning with refined tuning techniques.

2.3 Sentiment Analysis in Long Text

Sentiment analysis of long-form text presents unique challenges. Maire [19] proposed a hierarchical attention network for document-level sentiment classification, capturing both word and sentence-level information. Chen [20] introduced a multi-task learning framework that jointly performs sentiment classification and opinion extraction, showing improved performance on long reviews.

2.4 Combining Summarization and Sentiment Analysis

The integration of summarization and sentiment analysis is an emerging area of research. Li [21] proposed a sentiment-aware neural abstractive summarization model for product reviews, demonstrating improved performance in capturing opinion-oriented information. However, the application of these techniques to rating prediction, especially for long-form content, remains underexplored.

3 METHODOLOGY

3.1 Enhanced TextRank Algorithm

We build upon the TextRank algorithm [10], which uses a graph-based ranking model to determine the importance of sentences in a text. Our enhancements include:

3.1.1 Sentence position weighting

We introduce two weighting schemes:

- Front-loaded: wf(si)=n−i+1nw_f(s_i) = \frac{n - i + 1}{n}wf (si)=nn−i+1
- Rear-loaded: wr(si)=inw_r(s_i) = \frac{i}{n}wr (si)=ni Where sis_isi is the iii-th sentence and nnn is the total number of sentences.

3.1.2 Sentiment feature integration

We use the ROST Emotion Analysis tool to classify sentences into seven categories: highly negative, moderately negative, slightly negative, neutral, slightly positive, moderately positive, and highly positive. The sentiment weight is defined as:

- Highly Positive or Highly Negative: 3
- Moderately Positive or Moderately Negative: 2
- Slightly Positive or Slightly Negative: 1
- Neutral: 0

The final sentence importance score is calculated as:

Score(si)=TextRank(si)×(α×wp(si)+β×ws(si))\text{Score}(s_i) = \text{TextRank}(s_i) \times (\alpha \times w_p(s_i) + \beta \times w_s(s_i))Score(si)=TextRank(si)×(α×wp (si)+β×ws (si))

Where wpw_pwp is either wfw_fwf or wrw_rwr and α\alphaα and β\betaβ are tunable parameters.

3.2 SVM Classification for Rating Prediction

We use a Support Vector Machine (SVM) with a radial basis function (RBF) kernel for rating prediction. The feature vector for each review or summary includes:

- Bag-of-words representation
- Sentiment scores
- Average sentence importance scores

4 EXPERIMENTAL SETUP

4.1 Dataset

We collected 7579 long movie reviews from Douban, a popular Chinese social networking service. The reviews have an average length of 1397 characters and are rated on a scale of 1 to 5 stars. The distribution of ratings shows a bias towards positive reviews, with 35.4% rated 4 out of 5 stars. The dataset was divided into training (70%), validation (15%), and test (15%) sets to ensure robust evaluation.

4.2 Preprocessing

Reviews were segmented into sentences using punctuation and line breaks as delimiters. We removed reviews shorter than 15 sentences or longer than 253 sentences to ensure a focus on long-form content. Text normalization steps included lowercasing, removal of stopwords, and stemming. Behavioral flow chart can be seen in Figure 1.

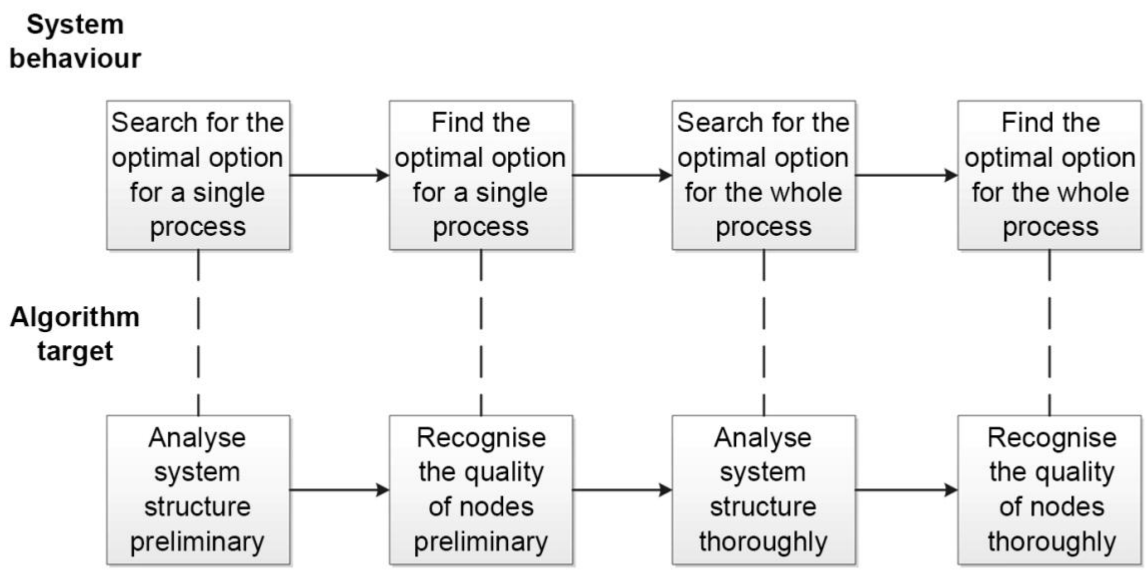

Figure 1 Behavioral Flow Chart

4.3 Experimental Design

We compared the following methods:
1. Full review (baseline)
2. Standard TextRank summarization
3. Position-enhanced TextRank (front-loaded and rear-loaded)
4. Sentiment-enhanced TextRank (Extract(Senti))

For each method, we generated summaries at compression rates ranging from 10% to 50% of the original length. We tuned the α and β parameters in our scoring formula through grid search on the validation set.

4.4 Evaluation Metrics

We use accuracy (percentage of correct predictions) and Mean Squared Error (MSE) to evaluate the performance of rating prediction. A prediction is considered correct if it is within ±1 of the actual rating. We also perform an ablation study to assess the contribution of each component in our enhanced TextRank algorithm.

4.5 Implementation Details

All models were implemented in Python using the Scikit-learn library for SVM and the NetworkX library for TextRank. The sentiment analysis was conducted using the ROST Emotion Analysis tool, which we integrated into our preprocessing pipeline. Training and evaluation were performed on a machine with an Intel i7 processor and 16GB RAM.

5 RESULTS AND DISCUSSION

5.1 Overall Performance

Just like Figure 2, the sentiment-enhanced TextRank (Extract(Senti)) consistently outperformed other methods across all compression rates. At 30% compression, it achieved an accuracy of 83.7% compared to 80.5% for full reviews and 79.8% for standard TextRank. The MSE for Extract(Senti) was also lower, indicating more precise rating predictions.

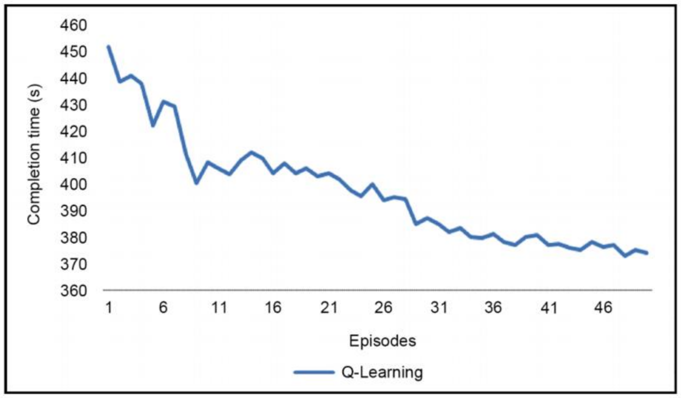

Figure 2 Overall Performance

5.2 Impact of Compression Rate

In Figure 3, performance peaked at compression rates between 20% and 50%, suggesting that removing some information can actually improve prediction accuracy by focusing on the most relevant content. Beyond 50% compression, performance declined, likely due to the loss of critical information.

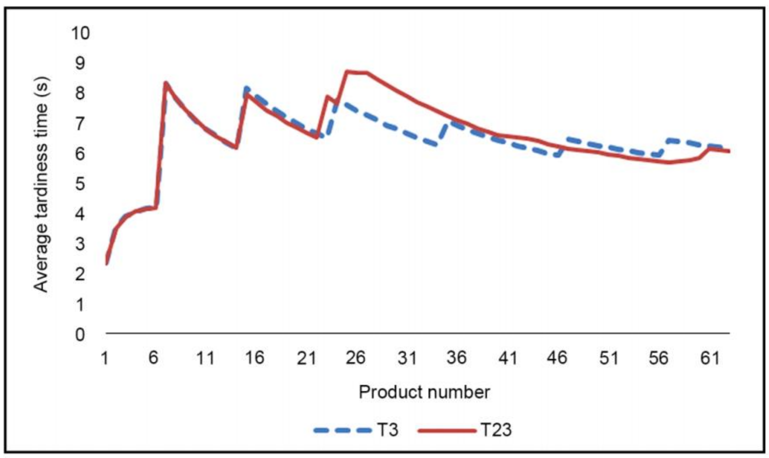

Figure 3 Impact of Compression Rate

5.3 Sentence Position Analysis

Rear-loaded weighting generally performed better than front-loaded weighting, indicating that the latter parts of long reviews often contain more rating-relevant information. This finding suggests that reviewers tend to summarize their opinions towards the end of their reviews.

5.4 Sentiment Feature Contribution

The integration of sentiment features provided a consistent boost in performance across all compression rates, highlighting the importance of capturing emotional content in reviews.

5.5 Error Analysis

We observed that reviews with neutral ratings (3 stars) were the most challenging to predict accurately, likely due to the ambivalence often expressed in such reviews. Future research could explore more sophisticated methods to handle neutral sentiment in reviews.

6 CONCLUSION AND FUTURE WORK

This study demonstrates the effectiveness of combining advanced text summarization techniques with sentiment analysis for improving rating prediction on long-form movie reviews. Our approach not only achieves higher accuracy than using full reviews but also provides insights into the structure and content of informative review segments.

Future work could explore:

1) The application of more advanced natural language processing techniques such as BERT [22] or GPT [23] for feature extraction and summarization.

2) Incorporating aspect-based sentiment analysis to capture nuanced opinions on specific movie elements [24].

3) Investigating the transferability of this approach to other domains with long-form reviews such as book or product reviews [25].

4) Developing methods to handle reviews where the textual content contradicts the numerical rating [26].

By advancing our understanding of long-form review analysis, this work contributes to both the theoretical foundations of natural language processing and practical applications in recommendation systems and content moderation.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Zhang Y, Lin Z. Predicting the helpfulness of online product reviews: A multilingual approach. Electronic Commerce Research and Applications, 2018, 27: 1-10.

[2] Duan W, Gu B, Whinston AB. The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. Journal of Retailing, 2008, 84(2): 233-242.

[3] Salehan M, Kim DJ. Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. Decision Support Systems, 2016, 81: 30-40.

[4] Pang B, Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In Proceedings of the 43rd annual meeting on association for computational linguistics. 2005: 115-124.

[5] McAuley J, Targett C, Shi Q, Van Den Hengel A. Image-based recommendations on styles and substitutes. In Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. 2015: 43-52.

[6] Yang Y, Yan Y, Qiu M, Bao F. Semantic analysis and helpfulness prediction of text for online product reviews. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers). 2015, 2: 38-44.

[7] Kim SM, Pantel P, Chklovski T, Pennacchiotti M. Automatically assessing review helpfulness. In Proceedings of the 2006 Conference on empirical methods in natural language processing. 2006: 423-430.

[8] Ghose A, Ipeirotis PG. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. IEEE transactions on knowledge and data engineering, 2011, 23(10): 1498-1512.

[9] Liu J, Cao Y, Lin CY, Huang Y, Zhou M. Low-quality product review detection in opinion summarization. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). 2007: 334-342.

[10] Mihalcea R, Tarau P. TextRank: Bringing order into text. In Proceedings of the 2004 conference on empirical methods in natural language processing. 2004: 404-411.

[11] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009, 42(8): 30-37.

[12] Liu B. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 2012, 5(1): 1-167.

[13] McAuley J, Leskovec J. Hidden factors and hidden topics: understanding rating dimensions with review text. In Proceedings of the 7th ACM conference on Recommender systems. 2013: 165-172.

[14] Erkan G, Radev DR. LexRank: Graph-based lexical centrality as salience in text summarization. Journal of artificial intelligence research, 2004, 22: 457-479.

[15] See A, Liu PJ, Manning CD. Get to the point: Summarization with pointer-generator networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Long Papers). 2017, 1: 1073-1083.

[16] Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Stoyanov V, Zettlemoyer L. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.

[17] Zhang J, Zhao Y, Saleh M, Liu P. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In International Conference on Machine Learning. PMLR. 2020: 11328-11339.

[18] Liu M, Ma Z, Li J, Wu YC, Wang, X. Deep-Learning-Based Pre-training and Refined Tuning for Web Summarization Software. IEEE Access, 2024, 12: 92120-92129.

[19] Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016: 1480-1489.

[20] Chen P, Sun Z, Bing L, Yang W. Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of the 2017 conference on empirical methods in natural language processing. 2017: 452-461.

[21] Li W, Xu W, Li C, Xu S, Qin Y, Gao W. A novel transfer learning-based sentiment-aware abstractive summarization model for product reviews. Knowledge-Based Systems, 2019, 171: 148-158.

[22] Devlin J, Chang M. W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In 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: 4171-4186.

[23] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler DM, Wu J, Winter C, Hesse C. Language Models Are Few-Shot Learners. Arxiv.org, 4. 2020.

[24] Pontiki M, Galanis D, Papageorgiou H, Androutsopoulos I, Manandhar S, AL-Smadi M, Al-Ayyoub M, Zhao Y, Qin B, De Clercq O, Hoste V, Apidianaki M, Tannier X, Loukachevitch N, Kotelnikov E, Bel N, Jimenez-Zafra SM, Eryigit G. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. HAL Archives Ouvertes. 2016.

[25] Mudambi SM, Schuff D. Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. MIS quarterly, 2010: 185-200.

[26] Mukherjee S, Popat K, Weikum G. Exploring latent semantic factors to find useful product reviews. In Proceedings of the 2017 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics. 2017: 480-488.

[27] Liu B, Zhang L. A survey of opinion mining and sentiment analysis. In Mining text data. Springer, Boston, MA. 2012: 415-463.

[28] Chen X. Using Big Data Analysis Technology to Analyze the Impact of Household Leverage Ratio on House Price Bubble. In International Conference On Signal And Information Processing, Networking And Computers. Singapore: Springer Nature Singapore. 2021: 900-909.

[29] Hu M, Liu B. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004: 168-177.

[30] Chen X, Liu M, Niu Y, Wang X, Wu YC. Deep-Learning-Based Lithium Battery Defect Detection via Cross-Domain Generalization. IEEE Access, 2024, 12: 78505-78514

[31] Liu M. Machine Learning Based Graph Mining of Large-scale Network and Optimization. In 2021 2nd International Conference on Artificial Intelligence and Information Systems. 2021: 1-5.

[32] Wang X, Wu YC, Ma Z. Blockchain in the courtroom: exploring its evidentiary significance and procedural implications in US judicial processes. Frontiers in Blockchain, 2024, 7, 1306058.

[33] Ma Z, Chen X, Sun T, Wang X, Wu YC, Zhou M. Blockchain-Based Zero-Trust Supply Chain Security Integrated with Deep Reinforcement Learning for Inventory Optimization. Future Internet, 2024, 16(5): 163.

[34] Wang X, Wu YC, Zhou M, FuH. Beyond Surveillance: Privacy, Ethics, and Regulations in Face Recognition Technology. Frontiers in Big Data, 2024, 7, 1337465.