Weakly Supervised Analysis of Vietnamese Review Aspects

Vũ Minh Hiếu
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam
student628@ictu.edu.vn

***Abstract*—This undergraduate thesis investigates weakly supervised analysis of Vietnamese reviews. Vietnamese review data need preprocessing and labelled examples before aspect and sentiment models can be evaluated. The student collects and preprocesses two Vietnamese datasets, represents comments with count vectors, and evaluates a Naive Bayes classifier for positive and negative labels in an exploratory study. The manuscript reports approximately 0.81 accuracy and F1 values near 0.80 to 0.81 for its sentiment classification example. These figures are the student's reported test results, not a verified aspect-extraction benchmark. Although the cover states aspect detection, the detailed conclusion mainly describes positive and negative sentiment classification. The source itself says the results are insufficient for practical use and notes data quality and label imbalance. The contribution is a documented student study or prototype whose stated scope can guide further testing.**

***Index Terms*—Vietnamese reviews, weak supervision, aspect detection, sentiment classification**

I. Introduction

Vietnamese review data need preprocessing and labelled examples before aspect and sentiment models can be evaluated. This thesis investigates weakly supervised analysis of Vietnamese reviews. It begins from a practical engineering need and explains the technologies and workflow chosen for the case. The objective is to document a student implementation or comparative investigation, together with the evidence available in the manuscript. Broader operational or economic effects require separate measurements and are not assumed from the existence of a prototype [1]. Earlier ICTU student work on related designs is included here for context [2]. Related published studies provide context for the subject and methods [3], [4], [5], [6], [7], [8], [9], [10]. These contextual works were selected for this English summary; they are not asserted to appear in the original thesis bibliography.

II. Methods

The student collects and preprocesses two Vietnamese datasets, represents comments with count vectors, and evaluates a Naive Bayes classifier for positive and negative labels in an exploratory study. The source is summarized through its stated design, implementation material, examples, and conclusion. Features shown in the manuscript are distinguished from proposals for later development. This review does not add a new experiment, fill missing measurements with estimates, or treat a planned capability as an observed result. The source thesis remains the provenance for the reported work [1].

Supervisor: Đoàn Thị Bích Ngọc, Thai Nguyen University of Information and Communication Technology (dtbngoc@ictu.edu.vn).

III. Results

The manuscript reports approximately 0.81 accuracy and F1 values near 0.80 to 0.81 for its sentiment classification example. These figures are the student's reported test results, not a verified aspect-extraction benchmark. The report's screens, configurations, examples, or measurements constitute the evidence for this summary. A capability described only as a design intention is not counted as a validated operational outcome. No additional performance, adoption, financial, or safety figure has been introduced here. The result is interpreted within the scope of the student's documented setup [1].

IV. Discussion

Although the cover states aspect detection, the detailed conclusion mainly describes positive and negative sentiment classification. The source itself says the results are insufficient for practical use and notes data quality and label imbalance. Further work could define aspect labels explicitly, clean the corpus, publish the evaluation split, and compare an aspect-level baseline with the sentiment classifier. These limitations affect how the evidence should be read. The current manuscript can support understanding of the design and reported demonstration, while independent tests would be needed for stronger claims. Any later deployment should document its environment and separate observed outcomes from intended benefits [1].

V. Conclusion

The thesis documents weakly supervised analysis of Vietnamese reviews. The manuscript reports approximately 0.81 accuracy and F1 values near 0.80 to 0.81 for its sentiment classification example. These figures are the student's reported test results, not a verified aspect-extraction benchmark. Although the cover states aspect detection, the detailed conclusion mainly describes positive and negative sentiment classification. The source itself says the results are insufficient for practical use and notes data quality and label imbalance. Further work could define aspect labels explicitly, clean the corpus, publish the evaluation split, and compare an aspect-level baseline with the sentiment classifier. Accordingly, the

present contribution is a source-grounded account of the student work and its tested or illustrated scope. Claims about broader effectiveness should be assessed in later work using explicit evaluation conditions [1].

References

[1] Vũ Minh Hiếu, ”Weakly Supervised Analysis of Vietnamese Review Aspects,” internal undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, 2024.

[2] ICTU Student 197, ”Application of Genetic Algorithm in Text Classification of Vietnamese Language,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2010.

[3] Bishrul Haq et al., ”A Semi-Supervised Approach for Aspect Category Detection and Aspect Term Extraction from Opinionated Text,” Computers, Materials & Continua, 2023. doi:10.32604/cmc.2023.040638.

[4] Anna Koufakou, ”Deep learning for opinion mining and topic classification of course reviews,” Education and Information Technologies, 2023. doi:10.1007/s10639-023-11736-2.

[5] Tu Nguyen Thi Ngoc et al., ”Mining aspects of customer’s review on the social network,” Journal of Big Data, 2019. doi:10.1186/s40537-019-0184-5.

[6] Huang Huang et al., ”Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions,” IEEE Access, 2023. doi:10.1109/access.2023.3307308.

[7] Khairullah Khan et al., ”Mining opinion components from unstructured reviews: A review,” Journal of King Saud University - Computer and Information Sciences, 2014. doi:10.1016/j.jksuci.2014.03.009.

[8] Thien Khai Tran et al., ”Capturing Contextual Factors in Sentiment Classification: An Ensemble Approach,” IEEE Access, 2020. doi:10.1109/access.2020.3004180.

[9] Vo Ngoc Phu et al., ”A CURE Algorithm for Vietnamese Sentiment Classification in a Parallel Environment,” Journal of Computer Science, 2019. doi:10.3844/jcssp.2019.1355.1377.

[10] Mayur Wankhade et al., ”A survey on sentiment analysis methods, applications, and challenges,” Artificial Intelligence Review, 2022. doi:10.1007/s10462-022-10144-1.