JTE HCMUTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Rule-enriched Decision Tree Classifier for Conditional Sentence Sentiment Analysis

Dinh Tai Pham[1], Nam Hoang Do[1], Huyen Trang Phan[2*]
[1]*Nguyen Tat Thanh University, Vietnam*
[2]*Yeungnam University, South Korea*

[*]Corresponding author. *Email: huyentrangtin@ynu.ac.kr*

ARTICLE INFO

Received:	06/02/2024
Revised:	22/02/2024
Accepted:	23/02/2024
Published:	28/02/2024

KEYWORDS

Sentiment analysis;
Conditional sentence;
ReDTC;
Sentence-level sentiment analysis;
Conditional sentiment analysis.

ABSTRACT

Conditional sentences are often used when people have to choose with some requirements. Conditional sentences account for more than 8% of user opinions. Although accounting for a considerable amount, for sentiment analysis methods, the sentiment expressed in conditional sentences is still analyzed as typical narrative sentences. This causes the approaches to fail to achieve maximum performance. To solve this problem, although some studies have proposed separate approaches to sentiment extraction and analysis for conditional sentences, there are very few, and the performance still needs to improve. This study proposes a new classifier based on a decision tree classifier model enriched with rules (called Rule-enriched Decision Tree Classifier (ReDTC)) to extract and analyze sentiments expressed in conditional sentences. ReDTC has been experimented on a dataset collected from English teaching websites. The performance gain demonstrates that the proposed ReDTC method significantly improved the performance in sentiment extraction and analysis in conditional sentences.

Doi: https://doi.org/10.54644/jte.2024.1530

1. Introduction

Sentiment analysis (SA), a recent area quite active in the Natural Language Processing (NLP) field, is to identify sentiment-related characteristics expressed in user opinions. Sentiment characteristics include sentiment level, sentiment degree, and sentiment target [1]. The sentiment level usually has three main types: document, sentence, and aspect. Sentence-level SA aims to determine user sentiment degree toward a specific target [2]. For example, we have an opinion: *"I love the dessert in the hotel so much."* In this opinion, the user expressed a sentiment degree as positive regarding the target as dessert. Sentences include many types of sentences and conditional sentences are one of them.

Conditional sentences are used to explain that something can happen when the mentioned condition occurs. Most conditional sentences contain "if". A conditional sentence has two clauses: the conditional clause and the consequence clause. The clause often containing "if" is conditional and used to state the condition for the consequence clause to be true. Conditional sentences are usually divided into four main types: zero, first, second, and third. The zero conditional expresses a truth that always happens when a specific condition is met. For example, *"If the sunshine is strong, the temperature is high."* The first conditional is used to express an outcome likely to happen if a specific condition is fulfilled but it's uncertain. For example, *"If my condition is good, I will climb."* The second conditional expresses unlikely scenarios that will never happen because the condition is not met. For example, *"If I had wings, I could fly."* Finally, the third conditional expresses hypothetical situations where the present circumstances might have been different if something else had happened in the past. For example, *"If I had studied harder, I would have passed the exam."* So, with the hypothetical cases presented in these conditional sentences, is the sentiment expressed the different from other common sentences? And can SA methods for sentences that do not contain subjunctive cases be effectively applied to conditional sentence types? In regular sentences, sentiment is determined by the words used and the sentence's context. However, in conditional sentences, the feeling expressed in the user's opinion often depends on the condition mentioned. For instance, if someone says, *"If I were you, I would be thrilled"* analyzing it as a regular sentence would yield a positive result because of the words "thrilled." However, in this case, the user is not expressing positive feelings, and the sentiment is negative in a conditional sentence context. Therefore, using SA methods for regular sentences to determine the sentiment of conditional sentences often obtains inaccurate results because constructing only one approach that works well for all types of sentences is still challenging. In this case, building a SA method suitable to the characteristics of each kind of sentence is necessary. Therefore, some scientists have proposed sentence-level SA methods on conditional sentences, such as [3] and [4]. Narayanan et al. [3] analyzed the sentiment of conditional sentences based on unique characteristics that need special handling from both the linguistic and computational perspectives. In the linguistic research, canonical tense patterns were focused on, which are helpful in classification. In the computational study, Support Vector Machine (SVM) models were built to automatically classify whether sentences are neutral, negative, or positive. Experimental results have proved the performance of the models. Although these studies have proposed separate sentiment extraction and analysis for conditional sentences, there are very few, and the performance is still not high. Therefore, improving the performance of extracting and analyzing sentiment expressed in conditional sentences is still an issue that needs attention in the SA area. In this study, we propose a new classifier based on a Decision Tree Classifier model enriched with rules (called Rule-enriched Decision Tree Classifier (ReDTC) to improve the performance of extracting and analyzing sentiments expressed in conditional sentences. The ReDTC includes two primary components: Conditional sentence detection and conditional sentence-level SA. For conditional sentence detection, first, extracting the linguistic features related to four types of conditional sentences; second, constructing a set of rules to distinguish types of conditional sentences; and finally, building a classifier to learn the rules set for improving the conditional sentence extraction. For conditional sentence-level SA, first, extracting the linguistic and context features related to the sentiment expressed in conditional sentences; second, constructing a set of rules to distinguish types of sentiments (binary and multiple); and finally, building a classifier to learn the rules set for improving the conditional sentence sentiment classification. Our contributions are:

- We introduce a dataset of conditional sentences crawled from English teaching websites. This dataset has been preprocessed and assigned corresponding sentence type labels and sentiment labels.

- We propose novel rule-based extension methods to detect types of conditional sentences and to improve the performance of conditional sentence-level SA.

- We experiment with the ReDTC model on our dataset to evaluate the performance of the proposed model and ablations; data and code are available for a requirement.

We organize the remainder of this paper as follows. Section 2 presents an overview of related research. Section 3 describes the research problem of the proposal. Section 4 offers mathematical models for solving the research problem step by step. The data acquisition, setting up, experimental results, and evaluation of the proposal is provided in Section 5. Conclusions and research directions for the future are presented in Section 6.

2. Related Works

There are many different approaches to constructing and developing sentence-level SA methods. The scope of this article focuses on building a technique for SA of conditional sentences based on sets of rules. Therefore, this section focuses on surveying research on SA of conditional sentences and methods based on rule sets.

2.1. Rule-based SLSA methods

VADER [5] is a lexicon and rule-based SA tool for determining social media sentiment. It is used to analyze the sentiment of text sensitive to the emotion's polarity and intensity (strength). It is especially suitable for microblog-like contexts. Tan et al. [6] proposed a priori polarity vocabulary approach to classify financial news. They applied the sentiment composition rule based on calculating the overall sentiment of a post. This post's overall sentiment is a mathematical formula called the positive/negative ratio of the average sentiment value of all sentences related to financial news. Gao et al. [7] applied a rule-based approach to discover the causes of emotions on a Chinese microblog. Their system activates emotions by following emotional pattern rules to find the corresponding cause. Vo et al. [8] proposed a new model based on the C4.5 decision tree algorithm for sentiment classification at the document level.

This model achieved 60.3% accuracy on the test set. Cepeda and Jaiswal [9] used tweets about the COVID-19 vaccination program in Ireland for SA. First, the VADER tool assigns tweets as negative, positive, and neutral labels. These tweets were then trained using the SVM model. The results showed positive attitudes towards the vaccine at the start of the vaccination drive. However, this sentiment gradually changed into negative in early 2021. Kellert et al. [10] proposed an SA method by combining the adaptation of the recursive and unsupervised composition approach in SA [11] with the universal dependency formalism [12] to generate basic syntactic rules such as modification and negation rules apply to words in the sentiment dictionary. This approach exploits several advantages of unsupervised approaches to SA: (i) interpretability and explainability of SA, (ii) robustness across datasets, languages, and domains, and (iii) usability by non-NLP experts. The results show significant improvement compared to other methods.

We can see although conventional rule-based SA methods do not provide high performance, they all have the advantage that the implementation steps are simple, easy to understand, and easy to reproduce.

2.2. Conditional SLSA methods

Narayanan et al. [3] proposed a method to determine whether the opinions expressed on different topics in conditional sentences are positive, negative, or neutral. This article presents a linguistic analysis of conditional sentences for the first time. It then builds supervised learning models to determine whether the sentiments expressed about different topics in conditional sentences are positive, negative, or neutral by building SVM models. The method was experimented on datasets related to five domains and obtained a promising performance in terms of accuracy at 67.3%. Phan et al. [4] proposed a method for analyzing the sentiment of tweets containing conditional sentences. This method consists of the following steps: First, a set of features related to the syntactic, lexical, semantic, and sentiment of the words is extracted. Second, tweets containing conditional sentences are predicted using the Multilayer Perceptron model. Finally, the sentiment of these tweets is analyzed using the Convolutional Neural Network (CNN) model.

We can see that the previous two methods of detecting and classifying the sentiment of conditional sentences only built sets of rules to detect conditional sentences, and to analyze the sentiment of conditional sentences, they used machine learning models.

3. Research Problem

The main goal of this study is to propose a new classifier based on the CNN model enriched with rules (called Rule-enriched Decision Tree Classifier (ReDTC) to improve the performance of extracting and analyzing sentiments expressed in conditional sentences. The highlight of ReDTC is the ability to take advantage of a simple architecture of rule sets in learning the features and the ability to learn rules automatically on large datasets of the DTC algorithm. The ReDTC includes two strategies as follows:

For conditional sentence detection, a set of n sentences is presented $S = \{s_k | k \in [1, n]\}$, in this work, we treat this work as a multiple classification issue, meaning that each sentence can be given one of the following labels: zero, first, second, or third conditional sentences, or any combination of these labels. Put another way, let's say we have a sentence s_k and we represent the labels of the conditional types of sentences with $Type_{con} = \{0, 1, 2, 3\} \in R^{n \times 1}$. $Type_{con} = 0$, if s_k contains features of the zero conditional sentence. $Type_{con} = 1$ if s_k contains features of the first conditional sentence. $Type_{con} = 2$ if s_k contains features of the second conditional sentence. $Type_{con} = 3$ if s_k contains features of the third conditional sentence. The research topic for this paper's technique is as follows, with the aforementioned notations: Given the labeled vector $Type_{con}$ of conditional sentence types and the features matrix $H_n \in R^{n \times d}$, where d is dimension of vectors. This strategy's goal is to provide an effective rule-based mechanism for predict the vector $\widehat{Type}_{con}$ of unlabeled conditional sentence types.

For conditional sentence-level SA, given a set of m conditional sentences $C = \{c_i | i \in [1, m]\}$, where ci including a set of k features $c_i = \{f_j | j \in [1, k]\}$ where f_j is a j-th feature in c_i. We denote $H_c \in R^{m \times d}$ as the feature matrix, where d is dimension of vectors. This study examines the Conditional SLSA task as a multiple classification issue, wherein each sentence is assigned a single label representing one of the sentiment polarities (positive, negative, or neutral). To put it another way, let's say we have a sentence c_i and we use $SA_{sen} = \{pos, neg, neu\} \in R^{m\times1}$ to represent the labels of the sentence sentiment. Specifically, if c_i expresses a positive sentiment, $SA_{sen} = neg$; if c_i expresses a negative sentiment, $SA_{sen} = neu$; The following is the formulation of the research problem for this study with the aforementioned notations: The labeled sentiment vector SA_{sen} and the contextualized matrix H_c are given. The goal of this strategy is to suggest an effective way to forecast the vector $\widehat{SA}_{sen}$ of the unlabeled sentiments by effectively learning rules drawn from sentiment cues in the sentence.

4. Proposed Method

The ReDTC includes two primary components: Conditional sentence detection and conditional sentence-level SA.

For conditional sentence detection: Input is a set of sentences. The main steps are: (i) first, extracting the linguistic features related to four types of conditional sentences; (ii) second, constructing a set of rules to distinguish types of conditional sentences; and (iii) finally, building a classifier to learn the rules set for improving the conditional sentence extraction. Output is a set of sentences that are assigned conditional sentence labels. For conditional sentence-level SA: Input is a set of sentences that are assigned conditional sentence labels. The main steps are: (i) first, extracting the linguistic and context features related to the sentiment expressed in conditional sentences; (ii) second, constructing a set of rules to distinguish types of sentiments (binary and multiple); and (iii) finally, building a classifier to learn the rules set for improving the conditional sentence sentiment classification. Output is a set of conditional sentences that are assigned sentiment labels. The workflow of ReDTC method is illustrated in Figure 1.

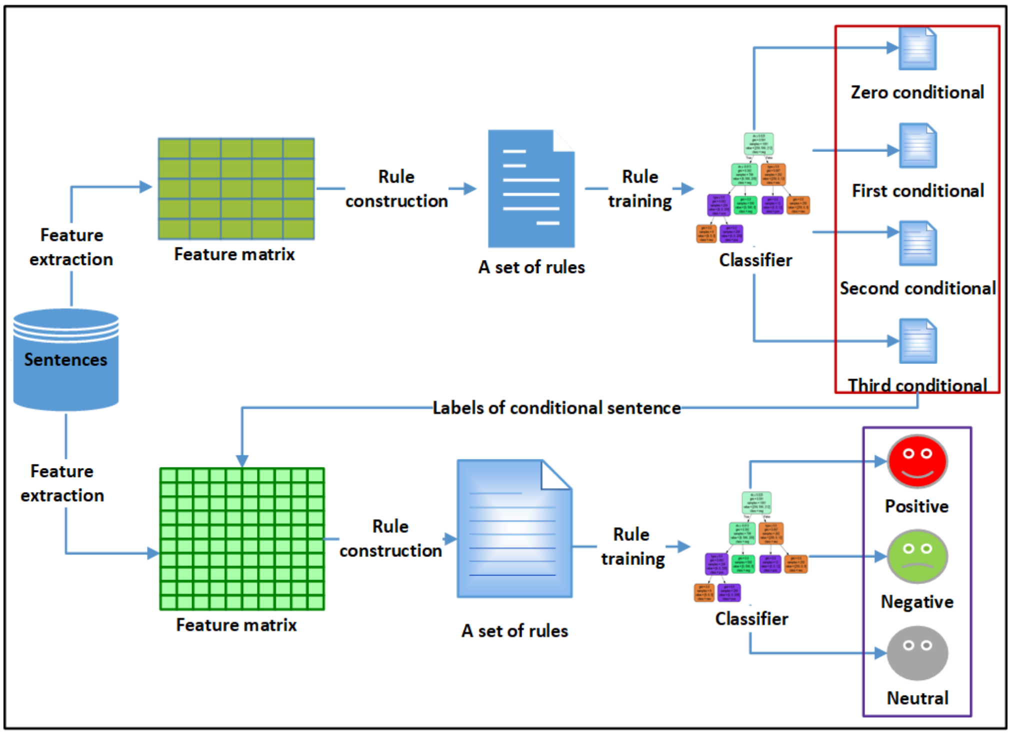


Figure 1. *Workflow of ReDTC method.*

4.1. Conditional sentence detection

This strategy's goal is to distinguish between undefined and conditional sentence types using the following steps:

Feature selection: This step aims to extract features regarding linguistics of conditional sentences related to tense patterns. Difference from the method in [3], in this paper, we only focus on the tense of the consequence clause by extracting features of modals, special words, negation words, and POS tags of verbs as Table 1.

JTE
HCMUTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Table 1. *Tense patterns and corresponding POS tag.*

Type	Tenses of consequence clause	Modal word	Negation	POS of verb
zero	V(-s/es)	None	not	VBP/VBZ
first	Modal +V(bare)	Modal	not	VB
second	PModal+V(bare)	PModal	not	VB
third	PModal+ have + V3/Ved	PModal	not	VBN/VBD

$Modal = \{will, can, shall, may\}$. $PModal = \{would, could, should, might\}$.

Feature extraction: This step aims to find the interactions among features as a premise for formulating rules to distinguish types of conditional sentences. The process to create the feature matrix is illustrated as Figure 2.

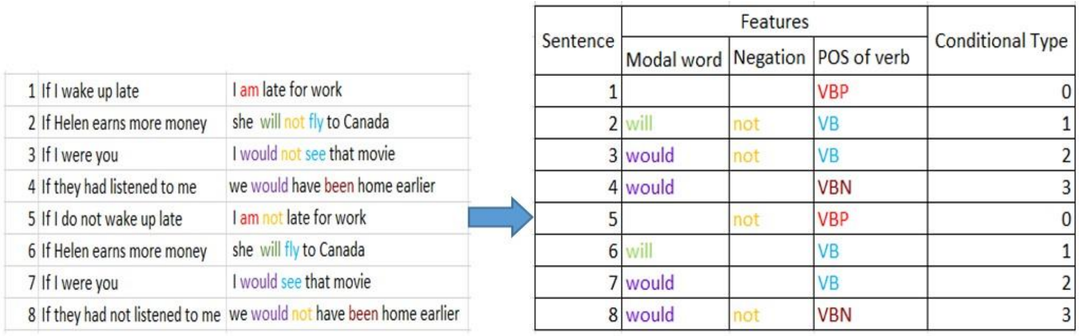


Figure 2. *Example of feature matrix creation.*

Rule construction: This step aims to construct a set of rules to classify the input sentences into type of conditional sentences such as zero, first, second, or third based on identifying the interactions among features shown in the feature matrix. In this study, we have constructed the following conditional sentence extraction rules:

$$If\ \ PoV = \{Y \mid Y \in \{VBP, VBZ\}\} then\ Type_{con} = 0 \quad (1)$$

$$If\ (Modal\ word\ = \{X|X\ \in\ Modal\}) and\ (PoV\ = VB) then\ Type_{con} =\ 1 \quad (2)$$

$$If\ (Modal\ word = \{X|X \in Modal\})\ and\ \big(Negation = \{Z \mid Z \in \{no, not\}\}\big)\ and\ (PoV = VB)\ then\ Type_{con} =\ 1 \quad (3)$$

$$If\ \big((Modal\ word = \{X|X \in PModal\}) and\ (PoV =\ VB)\big) then\ Type_{con} = 2 \quad (4)$$

$$If\ (Modal\ word = \{X|X\ \in\ PModal\})\ and\ \big(Negation =\ \{Z \mid Z\ \in\ \{no, not\}\}\big)\ and\ (PoV = VB)\ then\ Type_{con} =\ 2 \quad (5)$$

$$If\ (Modal\ word = \{X|X \in PModal\}) and\ \big(PoV = \{Y \mid Y \in \{VBN, VBD\}\}\big)\ then\ Type_{con} = 3 \quad (6)$$

$$If\ (Modal\ word\ =\ \{X|X\ \in\ PModal\}) and\ \big(Negation\ =\ \{Z \mid Z\ \in\ \{no, not\}\}\big)\ and\ \big(PoV\ =\ \{Y \mid Y\ \in\ \{VBN, VBD\}\}\big) then\ Type_{con}\ =\ 3 \quad (7)$$

where PoV is an acronym for POS of verb, $Type_{con}$ indicates conditional type.

4.2. Conditional sentence-level sentiment analysis

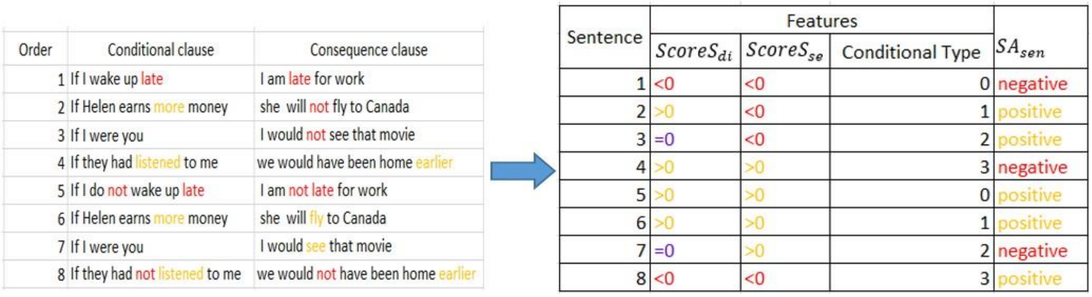
The object of this strategy is to identify the sentiment polarity of conditional sentences with the following steps:

Feature selection: This step aims to extract features regarding sentiment of conditional sentences related to sentiment knowledge and context of conditional types. Difference from the method in [3], in

JTE HCMUTE
ISSN: 2615-9740
JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

this paper, we combine the sentiment expressed on the conditional and consequence clause and the context of the type of conditionals by extracting features of sentiment words and conditional type.

Feature extraction: This step aims to find the interactions among features as a premise for formulating rules to identify the sentiment polarity of conditional sentences. The process to create the feature matrix is illustrated as Figure 3.



Order	Conditional clause	Consequence clause
1	If I wake up late	I am late for work
2	If Helen earns more money	she will not fly to Canada
3	If I were you	I would not see that movie
4	If they had listened to me	we would have been home earlier
5	If I do not wake up late	I am not late for work
6	If Helen earns more money	she will fly to Canada
7	If I were you	I would see that movie
8	If they had not listened to me	we would not have been home earlier

Sentence	Features			SA_{sen}
	$ScoreS_{di}$	$ScoreS_{se}$	Conditional Type	
1	<0	<0	0	negative
2	>0	<0	1	positive
3	=0	<0	2	positive
4	>0	>0	3	negative
5	>0	>0	0	positive
6	>0	>0	1	positive
7	=0	>0	2	negative
8	<0	<0	3	positive

Figure 3. *Example of feature matrix creation.*

Rule construction: This step aims to construct a set of rules to classify the input conditional sentences into type of sentiment polarities such as negative, neutral, and positive based on identify the interactions among features shown in the feature matrix. In this study, we have constructed the following conditional sentence-level sentiment extraction rules:

$$If\ (ScoreS_{di} - ScoreS_{se} < 0) and\ (Type_{con} = 0) then\ SA_{sen} = neg \quad (8)$$

$$If\ (ScoreS_{di} - ScoreS_{se} > 0) and\ (Type_{con} = 0) then\ SA_{sen} = pos \quad (9)$$

$$If\ (ScoreS_{di} - ScoreS_{se} < 0) and\ \left(Type_{con} = \{X \middle| X \in \{1,2,3\}\}\right) then\ SA_{sen} = pos \quad (10)$$

$$If\ (ScoreS_{di} - ScoreS_{se} > 0) and\ \left(Type_{con} = \{X \middle| X \in \{1,2,3\}\}\right) then\ SA_{sen} = neg \quad (11)$$

$$If\ (ScoreS_{di} - ScoreS_{se} = 0) then\ SA_{sen} = neu \quad (12)$$

where $ScoreS_{di}$ and $ScoreS_{se}$ are the sentiment score of the condition clause and consequence clause, respectively; $Type_{con}$ is the type of conditional sentence; SA_{sen} is the sentiment polarity of the conditional sentence; $neg, pos,$ and neu are respectively acronyms for negative, positive, and neutral.

4.3. Rule-based classifier

Rule-based classifiers are Machine Learning algorithms that create a set of IF-THEN rules from training samples to learn them and then apply them to classify new samples into one of the possible classes. They are one of the most classic and basic classifiers, yet practical and widely used in expert systems. In this research, we need to construct two rule-based classifiers to do two classification tasks: one for the conditional sentence prediction and another for the conditional sentence-level SA. After creating two sets of rules to build rule-based classifiers, we use the C4.5 rule-based algorithm [13].

5. Experiment Results

5.1. Dataset

To experiment with our proposed method, we prepare the datasets by ourselves. First, we collect conditional sentences from English teaching websites. Next, we label these sentences into four types: type 0, type 1, type 2, and type 3. Finally, we label these sentences for each conditional type into three sentiment polarities: negative, positive, and neutral. The detailed dataset is listed in Table 2.

Table 2. *Experimental dataset.*

No. Condition	Total	Positive	Negative	Neutral
Type 0	65	16	13	36
Type 1	608	109	119	380
Type 2	462	105	129	228

JTE
HCMUTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Type 3	382	67	83	232

5.2. Experimental Setting

Table 3. *Hyperparameters of the proposed method.*

Key	Value
Training set	n× 70%
Testing set	n× 30%
Way to split dataset	Random
No of features #1	3
No of features #2	3
Decision Creation	Scikit-learn
Optimization	Pre-pruning

n is the number of sentences.
#1 indicates the conditional sentence prediction.
#2 indicates the conditional SLSA.

Evaluation metrics: On the aforementioned dataset, we employed accuracy as the statistic to assess and contrast the effectiveness of our proposal. By contrasting the actual and anticipated test set values, accuracy is determined.

5.3. Results and Discussions

The results of the proposed method are shown in Figure 4.

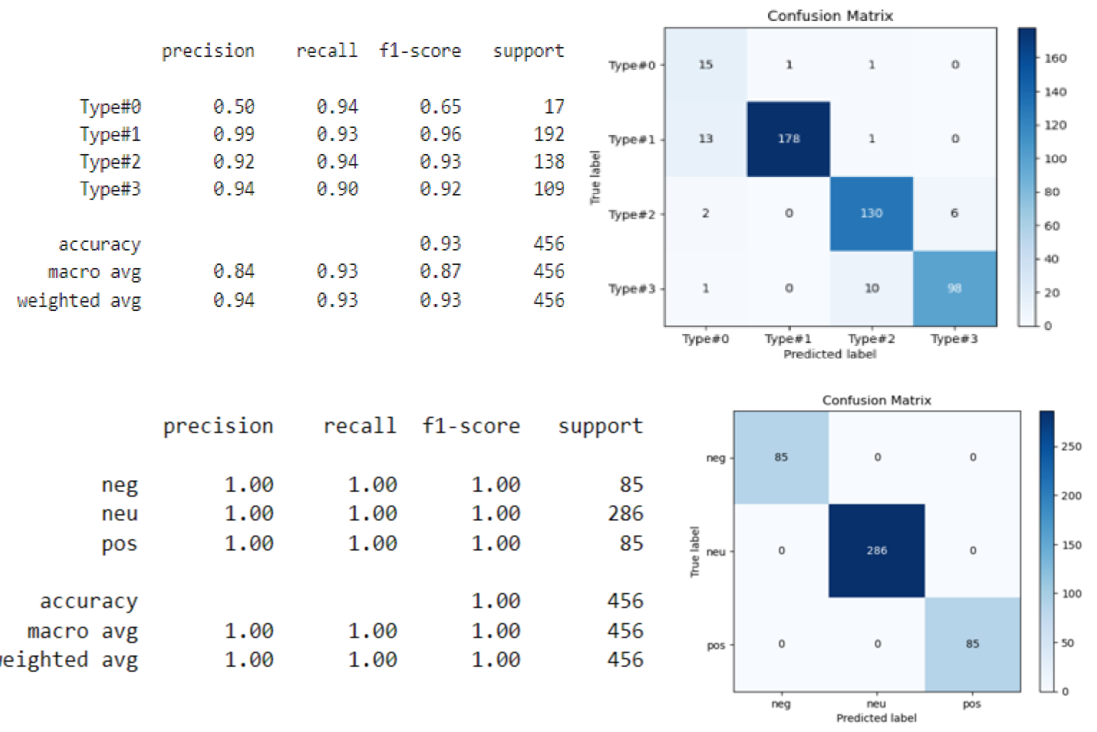


Figure 4. *Results of our proposed method.*

Based on Figure 4, we see clearly that although it was tested on an unbalanced data set in terms of both the number of conditional sentence types and the number of sentences according to sentiment

JTE
HCMUTE
ISSN: 2615-9740

labels, the results achieved by the proposed method are for both Precision, Recall, F1 score, and Accuracy are still very high. In addition, there is no distinction between the three types of sentiment labels for performance. To illustrate the process to obtain the results, we visualize the decision tree created as the following Figure 5.

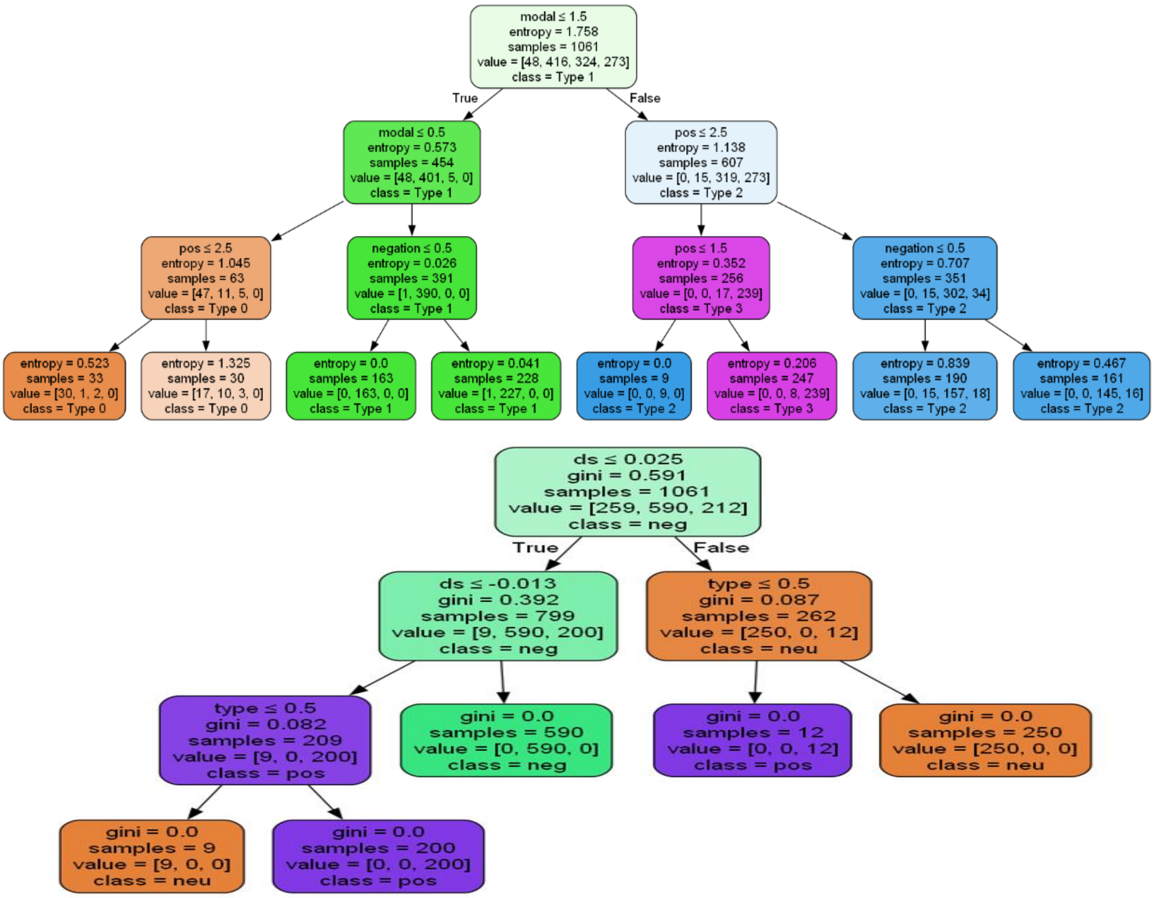


Figure 5. *Results of our proposed method.*

We can see that although we do not use the pruned technique, meaning we only build a decision tree by calculating basic information such as Gain, Gain Ratio, and Gini Index, the model still ensures simplicity and is easy to explain and understand.

To confirm the necessity of all four types of extracted features to build a set of rules used to determine the sentiment of conditional sentence types, we also proceed to develop ablation methods of the proposed method as follows: (i) No disen indicates the ablation method by ignoring the sentiment information of the condition clause. (ii) No sense indicates the ablation method that ignores the sentiment information of the consequence clause. (iii) No type indicates the ablation method by ignoring the information of conditional sentence type. The performance of ablation methods is shown in Table 4 and Figure 6.

Table 4. *Ablation Performance of proposed method.*

Ablation	Negative			Neutral			Positive			Accuracy
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score	
No disen	0.98	0.66	0.79	0.86	1.00	0.92	0.94	0.75	0.84	0.8882
No sesen	0.80	0.47	0.59	0.75	1.00	0.86	0.89	0.28	0.43	0.7654
No type	0.94	0.96	0.95	1.00	1.00	1.00	0.96	0.94	0.95	0.9825

type, disen and sesen mean conditional sentence type, condition and consequence sentiment.

JTE
HCMUTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

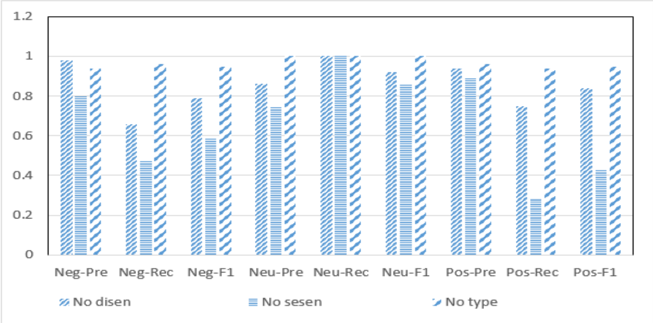


Figure 6. *Ablation performance.*

Looking at Table 4 and Figure 6, we see that the sentiment expressed in the consequence clause plays an essential role in improving the performance of conditional SLSA. In contrast, information about the type of conditional sentence plays a less important role and is more critical. However, all of the above characteristics must be present to ensure that the proposed method can achieve absolute performance.

To prove the performance of the proposed method in Conditional SLSA, we reproduce some deep learning-based SLSA on our dataset, called baselines, and compare the accuracy of the proposed method and these baselines. The baselines include CNN [14], DCNN [15], RNN, RNF [16], and BERT [17]. We do not conduct experiments on other datasets because they are beyond the proposal's scope, which is to focus only on the datasets containing conditional sentences. The comparison performance of the proposed method and baselines is shown in Figure 7:

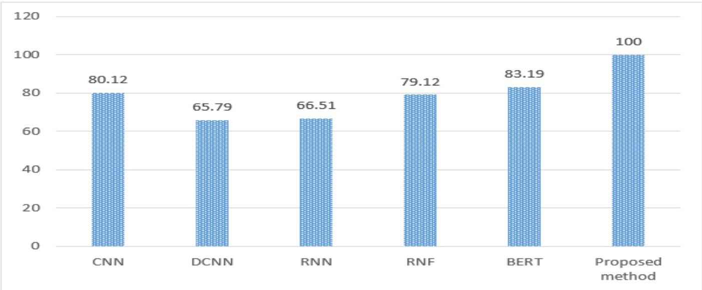


Figure 7. *Comparison performance of the proposed method and baselines.*

Figure 7 shows the proposed method achieves the best accuracy on the conditional sentence dataset. The main reason is that the base methods only consider and process conditional sentences as described sentences.

6. Conclusions

This research proposes a novelty method to detect types of conditional sentences, and analyze the sentence-level sentiment of those conditional sentences. The advantage of the proposed method is that it is not only simple, easy to illustrate, and easy to understand but still has high performance in terms of F1 score, recall, precision, and accuracy. However, the main drawback of this study is that we cannot reproduce the two previously published conditional sentence-level SA methods on our data set to evaluate and compare for confirmation more certainty about the effectiveness of the proposed model. In the future, we intend to use the ReDTC model to develop recommender systems by integrating the user and content feature hashtags [18] or multiple ontologies [19] into the SLSA modular.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

JTE HCMUTE
ISSN: 2615-9740
JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

REFERENCES

[1] H. T. Phan, N. T. Nguyen and D. Hwang, "Aspect-level sentiment analysis: A survey of graph convolutional network methods," *Information Fusion*, vol. 91, pp. 149-172, 2023.

[2] H. T. Phan, D. T. Pham and N. T. Nguyen, "FeDN2: Fuzzy-Enhanced Deep Neural Networks for Improvement of Sentence-Level Sentiment Analysis," *Cybernetics and Systems*, pp. 1--17, 2023.

[3] R. Narayanan, B. Liu and A. Choudhary, "Sentiment analysis of conditional sentences," in Proceedings of the 2009 conference on empirical methods in natural language processing, 2009, pp. 180-189.

[4] H. T. Phan, N. T. Nguyen, V. C. Tran and D. Hwang, "A method for detecting and analyzing the sentiment of tweets containing conditional sentences," *in Asian Conference on Intelligent Information and Database Systems*, Springer, 2019, pp. 177-188.

[5] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *in Proceedings of the international AAAI conference on web and social media*, 2014, pp. 216-225.

[6] L. Im Tan, W. San Phang, K. O. Chin and A. Patricia, "Rule-based sentiment analysis for financial news," *in 2015 IEEE International Conference on Systems, Man, and Cybernetics*, IEEE, 2015, pp. 1601-1606.

[7] K. Gao, H. Xu and J. Wang, "A rule-based approach to emotion cause detection for Chinese micro-blogs," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4517-4528, 2015.

[8] P. V. Ngoc, C. V. T. Ngoc, T. V. T. Ngoc and D. N. Duy, "A C4. 5 algorithm for english emotional classification," *volving Systems*, vol. 10, no. 3, pp. 425--451, 2019.

[9] K. Cepeda and R. Jaiswal, "Sentiment analysis on covid-19 vaccinations in Ireland using support vector machine," *in 2022 33rd Irish Signals and Systems Conference (ISSC)*, IEEE, 2022, pp. 1-6.

[10] O. Kellert, M. U. Zaman, N. H. Matlis and C. G'omez-Rodr'iguez, "Experimenting with ud adaptation of an unsupervised rule-based approach for sentiment analysis of mexican tourist texts," *arXiv preprint arXiv:2309.05312*, 2023.

[11] D. Vilares, C. G'omez-Rodr'iguez and M. A. Alonso, "Universal, unsupervised (rule-based), uncovered sentiment analysis," *Knowledge-Based Systems*, vol. 118, pp. 45-55, 2017.

[12] D. Zeman, J. Nivre, M. Abrams and others, "Universal Dependencies 2.6, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (UFAL), Faculty of Mathematics and Physics, Charles University," 2020.

[13] J. R. Quinlan, C4. 5: programs for machine learning, Elsevier, 2014.

[14] K. Yoon, "Convolutional Neural Networks for Sentence Classification," *in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, EMNLP, Doha, Qatar, ACL, 2014, pp. 1746-1751.

[15] K. Nal, G. Edward and B. Phil, "A Convolutional Neural Network for Modelling Sentences," *in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, MD, USA, ACL, 2014, pp. 655-665.

[16] Y. Yang, "Convolutional Neural Networks with Recurrent Neural Filters," *in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, ACL, 2018, pp. 912-917.

[17] M. Munikar, S. Shakya and A. Shrestha, "Fine-grained sentiment classification using BERT," *in 2019 Artificial Intelligence for Transforming Business and Society (AITB)*, IEEE, 2019, pp. 1-5.

[18] V. C. Tran, D. Hwang and N. T. Nguyen, "Hashtag recommendation approach based on content and user characteristics," *Cybernetics and Systems*, vol. 49, no. 5-6, pp. 368-383, 2018.

[19] T. H. Duong, N. T. Nguyen and G. S. Jo, "A hybrid method for integrating multiple ontologies," *Cybernetics and Systems: An International Journal*, vol. 40, no. 2, pp. 123-145, 2009.

Pham Dinh Tai received a master's degree from Hong Bang International University, Ho Chi Minh City, Vietnam. Currently, he is pursuing a Ph.D. degree and a lecturer in the IT department at Nguyen Tat Thanh University, data science department. He researches natural language processing, recommendation systems, sentiment analysis, machine learning, and deep learning.
Email: pdtai@ntt.edu.vn

Do Hoang Nam received a master's degree in information systems from Graduate University of Sciences and Technology, Vietnam. His current research interests include natural language processing, multimodal sentiment analysis, machine learning, and deep learning.
Email: namdh@ntt.edu.vn

Phan Huyen Trang received an M.S. degree in computer science from the University of Science and Technology, The University of Da Nang, Vietnam, in 2015, and a Ph.D. degree in computer engineering from Yeungnam University, South Korea, in 2020. She is currently an assistant professor in the Department of Computer Engineering, Yeungnam University, South Korea. She has authored ten journal articles and fifteen conference papers as the first author. Her research interests include sentiment analysis, fake news detection, text summarization, and decision support systems.
ORCID: https://orcid.org/0000-0002-7466-9562