IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 2, 2021, accepted August 1, 2021, date of publication August 9, 2021, date of current version August 25, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3103512

A Template Approach for Summarizing Restaurant Reviews

YENLIANG CHEN[1], CHIALING CHANG[2], AND JERYEU GAN[1]

[1]Department of Information Management, School of Management, National Central University, Taoyuan 320317, Taiwan
[2]Department of Information and Library Science, Tamkang University, New Taipei City 251301, Taiwan

Corresponding author: Chialing Chang (lindam@mail.tku.edu.tw)

ABSTRACT In the era of rapid development of social networks, user reviews of restaurant review websites have grown rapidly. In order to allow users to quickly grasp the key points of review information on review sites, this paper provides an abstractive multi-text summary method that can automatically generate template-based review summaries based on predefined topics and sentiments. In particular, for each predefined topic and each type of sentiment (positive or negative), this study uses the TextRank algorithm to find the most representative sentences to form a summary. This method allows users to quickly grasp the positive and negative opinions of each important aspect of the restaurant. The previous research on generating abstracts from reviews either did not generate abstracts based on topics, or they were based on topics generated by random models. However, the latter method cannot guarantee that the topics generated by the random model are really the topics that the user needs. For a restaurant review, some topics are indispensable. In order to ensure that abstracts can be generated for these essential topics, our method predefines the topics that must be generated, and then generates abstracts for these topics. In the evaluation, this study compared the template method with the Refresh and Gensim systems based on criteria such as informativeness, clarity, usefulness and likes. The results show that the method proposed in this paper is superior to the other two summary methods.

INDEX TERMS Restaurant reviews, sentiment analysis, summarization, template, TextRank.

I. INTRODUCTION

Today, people usually consume digital content by reading online reviews, photos, and videos on social media sites to support their shopping decisions and decision-making models. However, in the face of a large amount of digital content, people usually spend more energy and time digesting them. Therefore, text summarization technology is a strategy to deal with information overload in a big data environment. Review abstracts are currently widely used in many fields.

According to the summary results, the text summary approaches can be divided into two methods, namely extractive summarization and abstractive summarization [1]. The extractive summary method selects important sentences and paragraphs from the text, and then summarizes them into a shorter form as a summary. In contrast, the abstractive summary method first understands the main concepts in the document, and then changes the original words while retaining the original meaning to form the summary [1], [2]. When the text is to be retained to form a summary, the extraction method is mainly used. On the other hand, the abstractive summary method is mainly based on the understanding of the main concepts/topics in the document. For each major concept/topic in the document, the corresponding content must be extracted to form a summary.

To produce a good summary of documents, no matter which approach is used, a very important point is to find the main concepts/topics contained in the documents. After finding the topics, both approaches extract important sentences or content which retain the original meaning from each topic. This will help to generate a summary with content that can cover all major concepts contained in the document. For example, some research [3], [4] has used the PCA (principle component analysis) model to find the hidden topics in the document, and some [5], [6] has used the LDA (latent Dirichlet allocation) model, while others have used the LSA [7], [8] (Latent semantics Analysis) model. After that, they then find important content from each topic to form a summary of the document.

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wang.

However, the traditional topic finding methods have a very serious problem. That is, these topics are found by a mathematical method that is either based on variance-reduction (PCA), matrix approximation (LSA), or the generative stochastic model (LDA). For all these topic generation methods, there is no guarantee that the generated topics are truly topics that exist in the real world. It is very likely that the topics generated are not what the business is concerned with or the user needs. For example, for a restaurant review, some topics are indispensable. In order to ensure that abstracts can be generated for these indispensable topics, the summary must be extracted from these topics. Unfortunately, none of the topic generation models are able to produce topics according to these predefined topics. As a result, the generated summary may not correctly cover the important content of the original document.

In order to achieve this goal, an abstractive multi-text summary method will be used, and a system that can automatically generate restaurant review summary templates based on predefined topics and emotions is proposed. Unlike normal templates, we summarize comments based on predefined topics and emotional attitudes. For each predefined topic and each type of sentiment (positive or negative), we find the most representative sentences to form a summary. In addition, we point out the degree of positive and negative emotions in the comments on each topic. This method not only allows users to quote their favorite aspects (topics) more clearly, but it also divides reviews into common aspects (topics) of restaurants.

The difference between restaurants and other places is that they have some main basic aspects. Therefore, in this paper, we have pre-defined four aspects of the restaurant. However, in addition to these four basic aspects, each restaurant has an ''others'' aspect to include other possible topics that may be mentioned in the reviews. Then, for each aspect, we get the most representative positive and negative sentences and positive and negative sentiment analysis results, and finally get a template with complete information. This paper makes the following three contributions:

(1) It provides a technique that can automatically generate restaurant summaries, and will generate a template with complete information, including the restaurant's basic information, aspects, its most positive and negative sentences, and the proportion of positive and negative emotions.

(2) It uses a pre-defined topic method which allows users to see the topics according to the topics adopted on most restaurant review sites, so that users can get useful information faster.

(3) It provides the most important positive and negative sentences for each aspect so that users can see the advantages and disadvantages of each aspect at a glance.

II. RELATED WORK

A. TEXT SUMMARIZATION APPROACHES

According to the summary content, text summarization methods can be divided into extraction methods and abstractive methods [9]–[11]. The extraction method mainly selects important sentences from the content, and then summarizes them [12]. It mainly uses the TFIDF method and graph theory method [1]. The TFIDF method defines the sentence frequency as the number of sentences that contain the word in the document, and scores the sentence vector through a similarity query, and then selects the sentence with the highest score as part of the summary [13], [14]. The graph theory method uses graph theory as the subject recognition method [15]. The sentences in the preprocessed document are represented as nodes of the undirected graph.

Abstractive methods can be divided into two categories: structure-based methods and semantic-based methods [1]. Structure-based methods mainly use templates or extract some rules and other structures (such as trees, ontologies) to make the abstract look structured [16]. On the other hand, semantic-based methods use semantics to represent the summary [17]. This method uses linguistic data to identify noun phrases and verb phrases. In addition, according to the source of the text, the text summary method can be divided into a single text summary and a multi-text summary [9], [10], [18]. A single text summary literally generates a summary from a single text [19]. In contrast, multi-text summaries use multiple texts to generate summaries [20].

The method we proposed belongs to the multi-text structured method in the abstractive approach, because we use templates to make the summary look structured, and our goal is to aggregate multiple reviews of a restaurant into one abstract. In addition, when doing summarization, we must understand the topic (theme) and the emotional tendency of each sentence. In other words, our method must first understand the main concepts in the document, and then select the most representative sentence to describe the semantics of each specific concept.

B. TEXT SUMMARIZATION IN RESTAURANT REVIEWS

In the past, the research on automatic summarization of restaurant reviews was divided into two types, one to generate summaries based on topics through random models, and the other to generate summaries not based on topics. However, the topic is the main focus of restaurant reviews, that is, certain topics are indispensable. Therefore, generating summaries based on topics is critical to the success of automatic summarization of restaurant reviews. The main content of our research is a summary of restaurant reviews based on topics. Similar studies have also been conducted in previous studies.

Titov and McDonald [21] proposed a new method of extracting topics from online reviews. Their method automatically mines product reviews from the web and generates a summary of user reviews based on opinions. These reviews are concentrated in the areas of restaurants, hotels, and electronics. In order to extract topics, they proposed an unsupervised topic modeling based on the extension of standard topic modeling methods (such as LDA and PLSA) to induce multi-granular topics. The results show that their multi-granular topic model outperforms the standard topic model in extracting ratable aspects from online reviews. Xu *et al.* [22] considered the problem of extracting abstracts from online reviews based on aspects, and introduced the representativeness and diversity requirements for generating good abstracts. In order to meet these two requirements, they proposed a summary generation method which considers the inherent relationship between sentences in the entity review set and the dependency relationship between the extracted sentences. They use MG-LDA to extract and model aspects from a collection of restaurant reviews. Das *et al.* [6] introduced a two-step LDA-based aspect extraction technology for topic modeling of restaurant, hotel, and product review summaries. In the first step, they use the traditional latent Dirichlet allocation (LDA) to select the appropriate clusters for identifying seed words. These seed words are then set as training words to guide the second step of LDA. Since the traditional LDA extraction topic does not always reflect the desired result, the second step uses guided LDA technology to find the final topic. This method helps to identify potential, implicit aspects and explicit aspects. Their method is applicable to product reviews and restaurant reviews. Perera *et al.* [23] studied a new method based on aspect-based opinion summarization, which is a hierarchical aspect aggregation in the restaurant and laptop domains. They employed an amalgamation of domain-specific and domain-independent word embeddings along with agglomerative clustering to output a multi-granular structure of aspects. In order to construct effective word embeddings, they adopted principal component analysis (PCA). They used PCA for dimensionality reduction to further improve the accuracy of the distance between vectors. The number of PCA components is selected empirically, and these components are able to capture more than 80% of the variance in the aspect vectors. Shelke *et al.* [24] proposed a system that uses an expectation maximization algorithm to explore statistical methods for sentiment analysis of product and restaurant reviews. They used LSA (Latent Semantic Analysis) to find the underlying meaning or concepts of these documents. Their experiments were conducted on customer reviews such as for antivirus software, computers, restaurants, vehicles, digital cameras, etc.

Although the above studies generated summaries with content that could cover all major concepts by random models, their methods do not guarantee that the topics generated by the random model can match the topics that the company or customers care about. Although many summarization methods provide good efficiency in online reviews, due to the uniqueness of restaurant reviews, some topics that companies or customers pay attention to are indispensable. In order to ensure that these essential topics can generate abstracts, our method predefines the topics that must be generated and then generates abstracts for these topics.

Moreover, none of the above methods use templates to form summaries. If we use a template to generate a summary, it can make the information more structured and accurate. Therefore, our system in this paper is to create a template-based, topic-based, and sentiment-based summary. The system not only generates a complete template, but also lists the most representative sentences according to each topic, and has a positive and negative emotion ratio display for each topic. Therefore, compared with previous restaurant summaries, the template proposed in this paper enables users to quickly understand the advantages and disadvantages of the restaurant.

C. PREDEFINED TOPICS

In order to find predefined topics for restaurants, we looked at some online restaurant forums, such as TripAdvisor, OpenTable, Zomato, and Zagat. Then, we found that some topics appeared in most of these forums. Finally, we chose food, service, atmosphere, and value as our predefined topics. In addition, we added a fifth topic, "Others," to accommodate various other issues that may be mentioned in the reviews.

TABLE 1. The topics of restaurant forums.

Website	TripAdvisor	OpenTable	Zomato	Zagat
Topic	Food Service Value Atmosphere	Food Service Ambience Value	Food Service Look & Feel	Food Décor Service

III. METHODOLOGY

Throughout the method, our input is a restaurant review, and the output is a template. The entire methodology is an unsupervised approach. In other words, we do not need to label the topic of sentences in the training set in advance. So, the entire process does not depend on a training set of certain restaurant reviews. Our methodology can therefore produce a summary for any restaurant.

In the execution process, we will first find a candidate keyword set for each predefined topic, and then based on this we build a classification model to classify each word in the review to a topic. Next, we propose an algorithm to mark the topic tag on each review sentence. After that, we calculate the sentiment score of each sentence, and then calculate the sentiment score of each topic based on the sentence score. Finally, we find the most representative sentences in each topic and output all the information on the template.

A. FIND K KEYWORDS FOR A GIVEN TOPIC

This step will find k (k = 50) keywords from each of the 4 predefined topics (Food, Service, Ambience, Value). This can be done by the following steps: (1) For each topic, use the Gensim suite to find the 100 words most similar to the topic as candidate sets. In this step, each word and topic name is converted into word2vec format so that the similarity can be calculated by vector cosine similarity. (2) Query the number of search results for "restaurant" and candidate set words in Google search. Finally, select the top k (k = 50) keywords based on the number of search results and expert judgment.

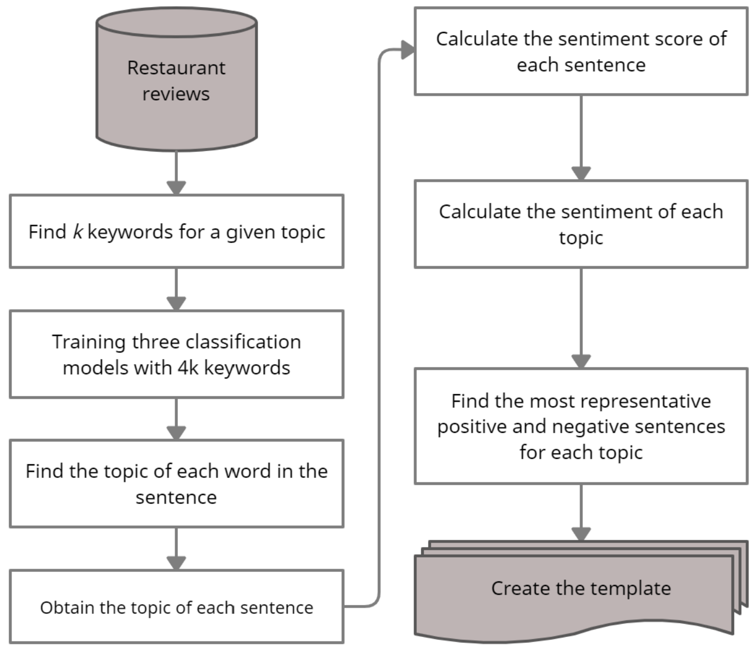


FIGURE 1. Flow chart of the proposed framework.

Due to the universality of the term "service," the term may appear in many fields, such as connection services on the Internet. Therefore, if we only want to find "service" topics related to restaurants, we must narrow the scope. On the website (thesaurus.com) that specifically looks for word synonyms, we first chose synonyms for "services" that are more relevant to restaurant services (such as assistance, benefits, duties, supply, utility, appropriateness, and courtesy). Repeating the above steps for these synonyms, we finally identified 50 keywords.

B. TRAINING THREE CLASSIFICATION MODELS WITH 4 × K KEYWORDS

After finding the keywords for each topic, we use these 200 keywords to train classifiers that can classify word topics. After feature extraction of the candidate keyword set, each keyword is converted into a 300-dimensional vector in word2vec format. Then, we use these 200 sample data with topic tags to train three word classification models based on machine learning.

When training the classification algorithm, we use 10-fold cross-validation (180 words as the training set and 20 words as the test set) to verify the trained model. Then we use three classifiers, KNN, SVM, and MLP, to train the classification model.

KNN (k nearest neighbor) is a machine learning method that can classify according to the distance between different feature values. KNN is mainly used to judge the category of unknown things and assign it to the category closest to its characteristics. We found in experiments that when k = 22, the best accuracy rate is 0.835.

food	service	ambience	value
foodstuff	customer	vibe	price
meal	provide	atmosphere	worth
nutrition	provider	hominess	cost
nutritious	delivery	homey	overvalued
meat	convenience	decor	valuated
beverage	training	rustic	premium
seafood	help	quaintness	salability
gourmet	support	spacious	holdings
edible	counseling	funkiness	undervalued
snack	obligation	conviviality	asset
perishable	responsibility	lounge	desirability
grocery	scarcity	splendor	undepreciated
eat	shortage	amenity	tangibility
biscuit	propriety	elegance	intangible
vegetable	suitability	homeliness	amount
milk	compliments	aura	attractiveness
dairy	kind	airy	depreciated
cook	thanks	liveliness	quantity
drink	organization	interiors	product
maize	assistance	flavor	profit

FIGURE 2. Some of the 50 keywords for the 4 topics.

SVM is a supervised learning method. It will find a hyperplane to separate different label sets. It has good applicability even in high-dimensional space. The advantage is that only a part of the samples (support vectors) can build a classification model. Compared with methods such as KNN, resource overheads such as memory are smaller, and since kernel functions can be specified, multiple situations can be handled, and different kernel functions can achieve different results.

We use grid search to find the best parameters: C = 1, gamma = 0.1, kernel = rbf, and get the best accuracy: 0.835. Grid search is a method for performing hyperparameter

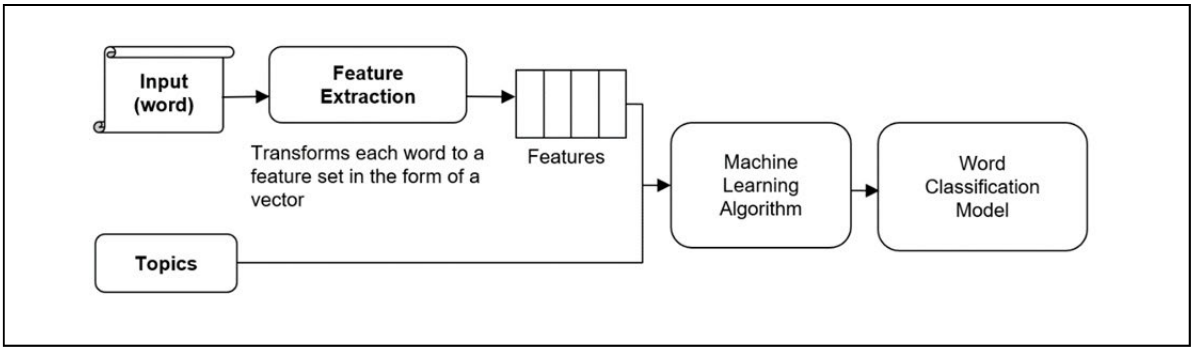


FIGURE 3. Training classification model architecture.

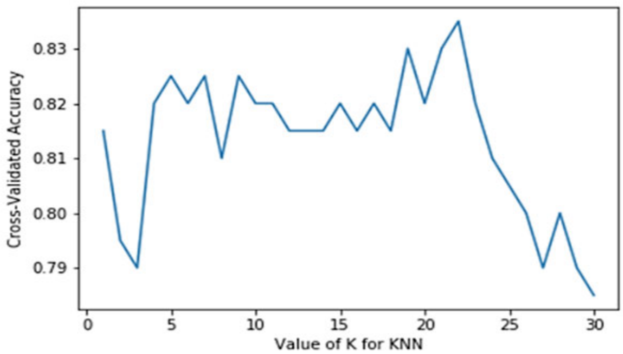


FIGURE 4. Relationship between k value and accuracy of KNN.

optimization in the sklearn kit, which is a Python module for machine learning built on top of SciPy.

```
grid.best_estimator_

SVC(C=1, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma=0.1, kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
```

FIGURE 5. The best parameters of SVM.

MLP is a type of feed-forward artificial neural network (ANN). It consists of at least three layers of nodes: the input layer, the hidden layer and the output layer. The advantage of MLP is that it can build a nonlinear model. On the contrary, the disadvantage is the need to adjust neurons, the number of layers and the number of iterations of each layer. We found in experiments that when the hidden layer = 42, the best accuracy rate is 0.86.

Although the above three algorithms can classify words with similar accuracy (as shown in Table 2), they can only classify words into four different topics. However, we have another topic, "others," that needs to be classified. The "others" topic is not easy to handle because it represents all other possible topics that do not belong to the four predefined topics. Therefore, we do not have word samples to learn the classifier to classify the "others" topics. In the next section,

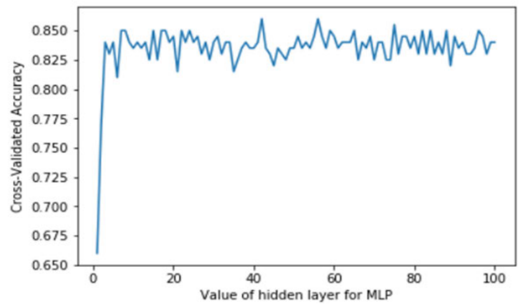


FIGURE 6. The relationship between MLP's hidden layer value and accuracy.

we will present the design of three other algorithms that divide words into five topics including "others" based on the above three methods.

TABLE 2. Accuracy of the three classification methods.

	KNN	SVM	MLP
Accuracy	0.835	0.835	0.860

C. FIND THE TOPIC OF EACH WORD IN THE SENTENCE

Real restaurant reviews not only contain the four topics mentioned above, but also include the "others" category. For example, "I highly recommend this restaurant" or "It provides a good choice for family and friends hanging out here." Therefore, when we look for the topic to which the sentence belongs, we must also consider the "others" topic. The difficulty in determining the "others" topic lies in the fact that all other possible topics that may be mentioned in reviews which do not belong to the predefined four topics belong to the category "others." This category may therefore contain a great diversity of words.

In this section, three possible algorithms will be introduced, namely Algorithm WT-1 (Word Topic-1), Algorithm WT-2, and Algorithm WT-3, to determine the topic of each word in the sentence. They will be included in the algorithm ST (Sentence Topics) in the next section to determine

```
Algorithm WT-1
Input:
K, S, M : The prediction of topic using KNN, SVM, MLP classifier model
Output:        T: The topic of the word
if K≠S and S≠M and K≠M then
               T ←Others
else if K=S then
               T←K
else if K=M then
               T←M
else if S=M then
               T←S
end if
```

FIGURE 7. Algorithm WT-1.

the topic of each sentence. Therefore, in the next section, we conduct a pre-test to determine which method can find 5 topics with higher accuracy (4 topics + others).

The WT-1 algorithm uses three classifiers to make predictions, and uses topics that appear multiple times as the results. If the three classifiers predict different topics, then we set the topic of the word to "others," and finally we can get the topic of each word.

Algorithm WT-2 finds the average word vector of the above 50 keywords for each 4 predefined topics. If the cosine similarity of the word vector and the average word vector is less than a given threshold, the word is classified as "others." After that, the most accurate MLP in the classification model is used to classify the remaining words, and finally the topic to which each word belongs is obtained.

Algorithm WT-3 uses a hybrid method to combine the methods of algorithms WT-1 and WT-2. First, we find the average word vector of 50 keywords for 4 predefined topics. If the cosine similarity of the word and the average word vector is less than a given threshold, the word is classified as "others." Then, we use three classifiers to predict the remaining words, and use the topic that appears most often as the result. If the three classifiers predict different topics, the word is classified as "others."

D. LABEL SENTENCES WITH TOPIC TAGS

After finding the topics to which all words belong, we need to mark all sentences with topic tags, so that when calculating the sentiment score, we can add up according to each topic. First, based on the topic of all words, we check which topic tag appears most in the sentence, and then the sentence will belong to that topic. However, since there may be more than one topic that appears most often in a sentence, a sentence is not necessarily limited to one topic, but may contain multiple topics.

Our sentence labeling algorithm belongs to an unsupervised model. In other words, we do not need to rely on many labeled sentences to train the classification system. The advantage of using unsupervised methods is that we do not have to train the model. In addition, unsupervised models are not limited to specific data sets. They can be easily extended to any data set or any type of restaurant. The steps of our sentence labeling algorithm, Algorithm ST (Sentence Topic) are given as follows.

Here, we conduct a pre-test to determine the accuracy of the sentence labeling algorithm. The test data set is selected by experts and includes 250 sentences and 5 topics for pre-testing (each topic contains 50 sentences). There are three algorithms to test, including ST + WT-1, ST + WT-2

```
Algorithm WT-2
Input: F, S, A, V : The cosine similarity of the word vector and the food, service, ambience, and values' average word
       vectors
Th: Threshold of the similarity
M: The prediction of topic using MLP classifier model
Output:        T: The topic of the word
Method:
if F<Th and S<Th and A<Th and V<Th then
        T←Others
else
        T←M
end if
```

FIGURE 8. Algorithm WT-2.

```
Algorithm WT-3
Input: F, S, A, V : The cosine similarity of the word vector and  the food, service, ambience, and values' average word
        vectors
Th: Threshold of the similarity
K, S, M: The prediction of topic using KNN, SVM, MLP classifier model
Output:  T: The topic of the word
Method:
if  F<Th and S<Th and A<Th and V<Th then
          T ←Others
else
      if K≠S and S≠M and K≠M then
          T ←Others
      else if K=S then
          T ←K
      else if K=M then
          T ←M
      else if S=M then
          T ←S
end if
```

FIGURE 9. Algorithm WT-3.

```
Algorithm ST
Input: S: sentence
         List(S): the topics appear most often in sentence S
Output:   T(S): A list of the topics attached to sentence S
Method:
For all sentences S
            T(S)←List(S)
```

FIGURE 10. Algorithm ST.

and ST + WT-3. The metric is how many sentences can be found correctly from all sentences of the class to which it belongs. For example, 40 means that forty sentences marked with X can be found correctly from all fifty sentences marked with X. The following shows the results of the pre-test.

TABLE 3. The result of Algorithm ST + WT-1.

Topic	Food	Service	Ambience	Value	Others
Found	44	46	32	46	0
Accuracy	0.67				

Based on the above test results, we can see that the number of "others" found will increase as the threshold increases. Of all these possible combinations, we prefer the combination with higher accuracy, but at the same time we need to find the largest minimum number of found sentences in all five topics. Finally, we chose the algorithm ST + WT-2 (only using MLP, and set the threshold to 0.25) as our classification algorithm model, which makes the number of sentences correctly found in each topic greater than 35 (70%), and the overall accuracy is also the highest.

TABLE 4. The result of Algorithm ST + WT-2.

Threshold Topic	0.2	0.225	0.25	0.275	0.3
Food	44	46	45	43	38
Service	42	36	35	33	33
Ambience	37	36	37	35	30
Value	48	47	45	42	39
Others	10	25	39	43	48
Accuracy	0.72	0.76	0.80	0.78	0.75

TABLE 5. The result of Algorithm ST + WT-3.

Threshold Topic	0.2	0.225	0.25	0.275	0.3
Food	44	46	45	43	38
Service	45	40	38	36	36
Ambience	34	33	34	32	27
Value	46	45	43	41	38
Others	10	20	34	41	48
Accuracy	0.72	0.74	0.78	0.77	0.75

E. CALCULATE THE SENTIMENT SCORE OF EACH SENTENCE

After calculating the topic to which each sentence belongs, next we need to calculate the sentiment score of the sentence in each topic. In the text summary, the sentiment score is an important indicator that can indicate the degree of positive and negative emotions in various aspects of the summary. In our template, there is also a ratio of positive emotion scores and negative emotion scores, which can help readers quickly grasp the emotional tendencies of various aspects of the summary.

Existing sentiment analysis methods can be divided into three categories: knowledge-based techniques, statistical methods, and hybrid methods [25]. Knowledge-based technology uses emotional words that appear in the text (such as "happy," "sad," "fear," "boring," etc.) to determine classification. Statistical methods use machine learning (such as latent semantic analysis, support vector machine, word bags, etc.) to calculate sentiment. The hybrid approach combines machine learning and knowledge representation (such as ontology and semantic networks) in order to be able to find more subtle emotion expressions between words, and to analyze the emotions.

In addition to the above methods, there are kits that third parties produce to calculate the sentiment score, such as OpinionFinder [26]. OpinionFinder is a system that processes documents and automatically recognizes subjective sentences and all aspects of subjectivity within sentences. It can identify the source of opinions, direct subjective expressions, emotional expressions, etc. The sentiment words in each sentence are divided into positive (+1), negative (−1), and neutral (0).

There is also a novel method VADER (Valence Aware Dictionary for sEntiment Reasoning) [27] for text emotion recognition based on thesaurus and grammar rules. VADER uses a combination of qualitative and quantitative methods, and then verifies the gold standard sentiment dictionary based on experience, which is particularly suitable for microblog-like environments.

The reason we chose VADER is that it works very well on social media type texts (such as reviews). It also does not require any training data, but is constructed from a generalizable, valence-based, human-curated gold standard sentiment dictionary. Moreover, it is fast enough to process streaming data online.

VADER's sentiment score (also called the compound score) is a metric that calculates the sum of all the lexicon ratings which have been normalized between −1 (most extreme negative) and +1 (most extreme positive). If the compound score is greater than 0.05, then the sentence tends to have a positive sentiment. Otherwise, if the compound score is less than −0.05, then the sentence tends to have a negative sentiment. If the score is between 0.05 and −0.05, then the sentence tends to have a neutral sentiment

F. CALCULATE THE SENTIMENT OF EACH TOPIC

On the template of our system, the proportion of sentiment scores of each topic will be listed, so in this step we will sum up the sentiment scores of each topic (positive and negative) and calculate the proportion.

Through VADER, we can easily divide a sentence into positive or negative ($>$ 0.05 is positive, $<$-0.05 is negative), and then calculate the total number of all positive sentences and negative sentences in a topic, and finally get the ratio between them.

G. FIND THE MOST REPRESENTATIVE POSITIVE AND NEGATIVE SENTENCES FOR EACH TOPIC

In this step, we will find the most representative positive and negative sentences in each topic. Therefore, we must apply an algorithm to rank important sentences in the text.

In the text summary, it is very important to find the most important sentence as the summary. Therefore, in order to find the most important sentences, we will rank these sentences. There are many ways to rank sentences. The most typical method is to use graph-based algorithms [28], of which HITS and PageRank were the earliest. HITS (Hyperlink Induced Topic Search) [29] is an iterative algorithm designed to rank web pages according to their authority. PageRank [30] is also a popular ranking method. It is designed as an algorithm for analyzing web page links. Unlike other algorithms, PageRank concentrates the influence of incoming links, and generates a set of scores and sorts by score.

The methods mentioned above are currently used by browsers to find page rankings. In addition, there are methods such as SemanticRank, iSpreadRank, and TextRank to find out the sentence ranking. SemanticRank [31] is a graph-based sorting algorithm that extracts keywords and sentences from text and uses implicit links to construct semantic graphs. Implicit links are based on the semantic correlation between text nodes, so different sorting algorithms are used to sort the nodes. iSpreadRank [32] models a set of documents related to the topic into a sentence similarity network, and expresses the importance of nodes (sentences) in the network, not only depending on the number of connected nodes, but also depending on the importance of the node. Therefore, the importance of sentences is recursively weighted to calculate and rank. TextRank [33] is a graph-based ranking model that uses a method similar to PageRank, replacing sentences with web pages, storing sentence similarity scores in a similarity matrix, using sentences as vertices, and similarity scores as edges. The generated graph will be highly connected, and is used to calculate the weight of each edge, and to get the final ranking.

This paper uses TextRank to find the most important sentences in each topic. TextRank has a good processing effect on unsupervised keyword and sentence extraction. One of the advantages is that it does not need deep linguistic knowledge or a corpus (an annotated corpus specific to the domain or language), so TextRank can be transferred to other fields, genres or languages.

There are some steps in the TextRank algorithm. First, we will connect all the text contained in the paper. Then, we split the text into sentences. After that, we will find the vector representation (word embedding) of each sentence. In the next step, the similarity between sentence vectors is calculated and stored in the matrix. Then, the similarity matrix is converted into a graph with the sentence as the vertex and the similarity score as the edge, which is used to calculate the sentence ranking. Finally, we can find the sentence with the highest ranking.

When ranking sentences, we found that although a sentence ranks higher, it may not be very positive or negative, and may be emotionally neutral. By checking restaurant reviews, we found that there are more positive sentences in the reviews than negative sentences. Therefore, when we look for sentence rankings, positive sentences must meet the overall score $>= 0.5$, and negative sentences must meet the overall score $<= -0.25$, to ensure that the sentences found are not only the most representative but are also emotionally positive or negative sentences.

H. CREATE THE TEMPLATE

Finally, we put all the information in the template. The template will first have basic information such as the restaurant name, address, and phone number. In addition, it also has a positive and negative total emotional ratio. After that, the template will contain the emotional proportion in each topic, and list the most important positive and negative sentences separately to complete the template.

IV. EVALUATION

A. DATA SETS

For the data set, we crawled restaurant reviews from the Tripadvisor website and turned them into templates. Tripadvisor is a well-known website with a large number of users. It has reviews and ratings for hotels, restaurants, etc. Many users comment on their experiences and provide restaurant information. Therefore, we used website reviews to obtain restaurant summaries to create templates.

For the collection of data sets, we considered five different types of restaurant. These restaurants are: 5 Napkin Burger, Applebee's, Bea of Bloomsbury, Serafina 77th Upper West, and The Ledbury. 5 Napkin Burger is an American burger restaurant, Applebee's is a bar-style restaurant, Bea of Bloomsbury is a coffee shop, Serafina 77th Upper West is an Italian restaurant, and The Ledbury is a creative gourmet restaurant in the UK. Finally, we invited more than one hundred people to participate in our experiment. We asked them to evaluate our method and the other two summarization algorithms one by one for the five selected restaurants by questionnaire.

B. EXPERIMENT DESIGN

Most traditional summaries use rouge indicators to assess the quality of the summary results. Rouge (Recall-Oriented Understudy for Gisting Evaluation) is an indicator for evaluating automatic summaries. It compares the summaries generated by automatic summaries with the so-called gold standard based on public datasets evaluated by experts, checks whether unigrams and bigrams meet the gold standard, and obtains automatic summary scores. The reason why we do not use the rouge indicator is because our dataset was found in Tripadvisor's review, not the public dataset, so there is no comparable gold standard. In addition, the template we want to present is to give users a good subjective feeling, so that they can clearly understand the quality and condition of the entire restaurant at a glance. Compared with the word "co-occurrence" in Rouge, the user experience of the entire template is more important, so we evaluated the template based on other indicators.

This study investigated peoples' subjective perceptions of the three summary methods. Its indicators include whether they felt that the content of the restaurant summary was informative, whether they felt that they could clearly understand the content of the restaurant summary at a glance, whether they felt that the content of the restaurant summary could help them understand the actual situation, and whether they felt that they liked this content of the restaurant summary. Zhang *et al.* [34] indicated that informativeness is an important indicator that affects online reviews and can be used as important supplementary information to help consumers reduce uncertainty and improve their purchasing decisions. Moreover, Mani [35] mentioned that informativeness is an indicator that can be used to measure the summary. The purpose is to assess whether the content of the summary retains sufficient information. If there is more informativeness in a summary, it enables the user to obtain more information, and also allows the user to obtain the most information in the shortest amount of time. Clarity means that when viewing papers or essays, users need to be able to see them clearly and concisely. In other words, users can see all the information they provide in a very short time or even at a glance. Hardy *et al.* [36] believed that clarity is an important indicator for evaluating summaries. They indicated that poor clarity will directly affect the comprehension of the article. Helpfulness is an indicator that can often be seen in website comments today. Helpfulness is often seen on a website to specify whether the review is helpful to the reader, so it can be regarded as whether the summary will give the user some substantive help or be useful to the user after reading [37]. Likes is another common metric to evaluate reviews. Likes is an indicator that measures whether users like the summary template. When the average "likes" score of the template is high, it means that the template is generally favored by users; furthermore, its content may have more opportunities to be adopted by users as a reference basis for consumption decisions [38].

In this study, we compared our template with other summary methods and administered a questionnaire to ask participants for evaluation according to informativeness, clarity, helpfulness, and likes, with 1 to 5 points to choose a better summary method. After removing people who were not interested in our experiment and were invalid samples, we were left with a total of 103 participants. To determine whether our method is good, we chose Refresh [39] and Gensim [40] as the methods to compare with our template method. Refresh and Gensim are both techniques for summarizing an entire paper, and they work well. Among them, Refresh divides the concept of abstract extraction into sentence ordering tasks, proposes a novel algorithm, and optimizes rouge evaluation indicators through reinforcement learning, and finally obtains abstract sentences. Gensim has been used in thousands of research papers. This is a very

TABLE 6. The results for 5 Napkin Burger.

Avg./SD	Our method	Refresh	Gensim
Informativeness	4.3350/0.6926	3.0388/1.0135	3.9437/0.9056
Clarity	4.2524/0.8047	3.3534/0.9564	3.2990/0.9596
Helpfulness	4.2175/0.7865	3.1602/1.0729	3.6738/0.9588
Likes	4.0621/0.8767	3.0087/1.0829	3.3816/0.9729
Avg.	4.2168/0.7901	3.1403/1.0314	3.5745/0.9492

TABLE 7. The results for Applebee's.

Avg./SD	Our method	Refresh	Gensim
Informativeness	4.2244/0.7047	2.8687/0.9447	3.7486/0.7821
Clarity	4.1526/0.7714	3.2303/0.9741	3.3012/0.9320
Helpfulness	4.0896/0.7862	3.0663/1.0392	3.5799/0.8471
Likes	3.9878/0.8258	2.7887/0.9853	3.2586/0.8794
Avg.	4.1136/0.7720	2.9885/0.9858	3.4721/0.8602

TABLE 8. The results for Bea's of Bloomsbury.

Avg./SD	Our method	Refresh	Gensim
Informativeness	4.1341/0.7582	2.7380/0.9304	3.8399/0.7347
Clarity	4.1196/0.7570	3.0487/0.9218	3.4099/0.8332
Helpfulness	4.0595/0.8070	2.7527/1.0284	3.6294/0.7584
Likes	3.9402/0.8214	2.6043/0.9888	3.3713/0.8608
Avg.	4.0633/0.7859	2.7859/0.9673	3.5626/0.7968

effective unsupervised method for extracting abstracts from papers.

C. EXPERIMENT RESULTS

We applied three algorithms (the template model, Refresh and Gensim) to summarize 200 reviews in each of the five restaurants. When using Gensim for summarization, we chose the top ten sentences as summaries, because our template uses ten sentences to describe five topics. However, the "refresh" system always summarizes the text with three sentences, so the summary in Refresh always has three sentences. The summary results of the reviews for 5 Napkin Burger are listed in Appendices A, B, and C. After summarization, we randomly selected samples from the Internet and invited them to participate in the experiment. After removing people who were not interested in our experiment, we were left with a total of 103 participants. In this experiment, we asked them to rate these four indicators (informative, clear and helpful, and likes), ranging from 1 to 5. The final results are shown in Tables 6 to 10.

Looking at the results of the above five restaurants, we found that the method of this paper outperforms on all restaurants in terms of information, clarity, helpfulness, and likes (as shown in Tables 6-10). Table 11 shows the average results of the three summary methods for the four indicators by averaging all five restaurants. Moreover, we wanted to examine if there were statistical differences among our method, Refresh, and Gensim for the four indicators by using ANOVA. The results revealed that all three summary methods have significant differences for the four indicators (informativeness: $F = 369.494$, $p < 0.000$; clarity: $F = 193.325$, $p < 0.000$; helpfulness: $F = 243.775$, $p < 0.000$; likes: $F = 236.742$, $p < 0.000$).

In order to obtain a more comprehensive result, we used descriptive statistical analysis, ANOVA tests, and post-hoc

TABLE 9. The results for Serafina.

Avg./SD	Our method	Refresh	Gensim
Informativeness	4.1914/0.7485	2.7195/1.0309	3.7719/0.7844
Clarity	4.0992/0.7342	2.9614/0.9755	3.4245/0.9051
Helpfulness	4.1420/0.7221	2.7149/1.0613	3.5702/0.8047
Likes	4.0082/0.7856	2.7169/1.0860	3.3790/0.8562
Avg.	4.1102/0.7476	2.7782/1.0384	3.5364/0.8376

TABLE 10. The results for The Ledbury.

Avg./SD	Our method	Refresh	Gensim
Informativeness	4.1013/0.8507	2.7263/0.9965	3.8148/0.8377
Clarity	4.0508/0.8593	2.9089/1.0111	3.5225/0.8008
Helpfulness	4.0304/0.8282	2.7840/1.0320	3.6925/0.8441
Likes	3.9702/0.8441	2.6652/1.0576	3.5031/0.8862
Avg.	4.0382/0.8456	2.7711/1.0243	3.6332/0.8422

TABLE 11. The average results for the five restaurants.

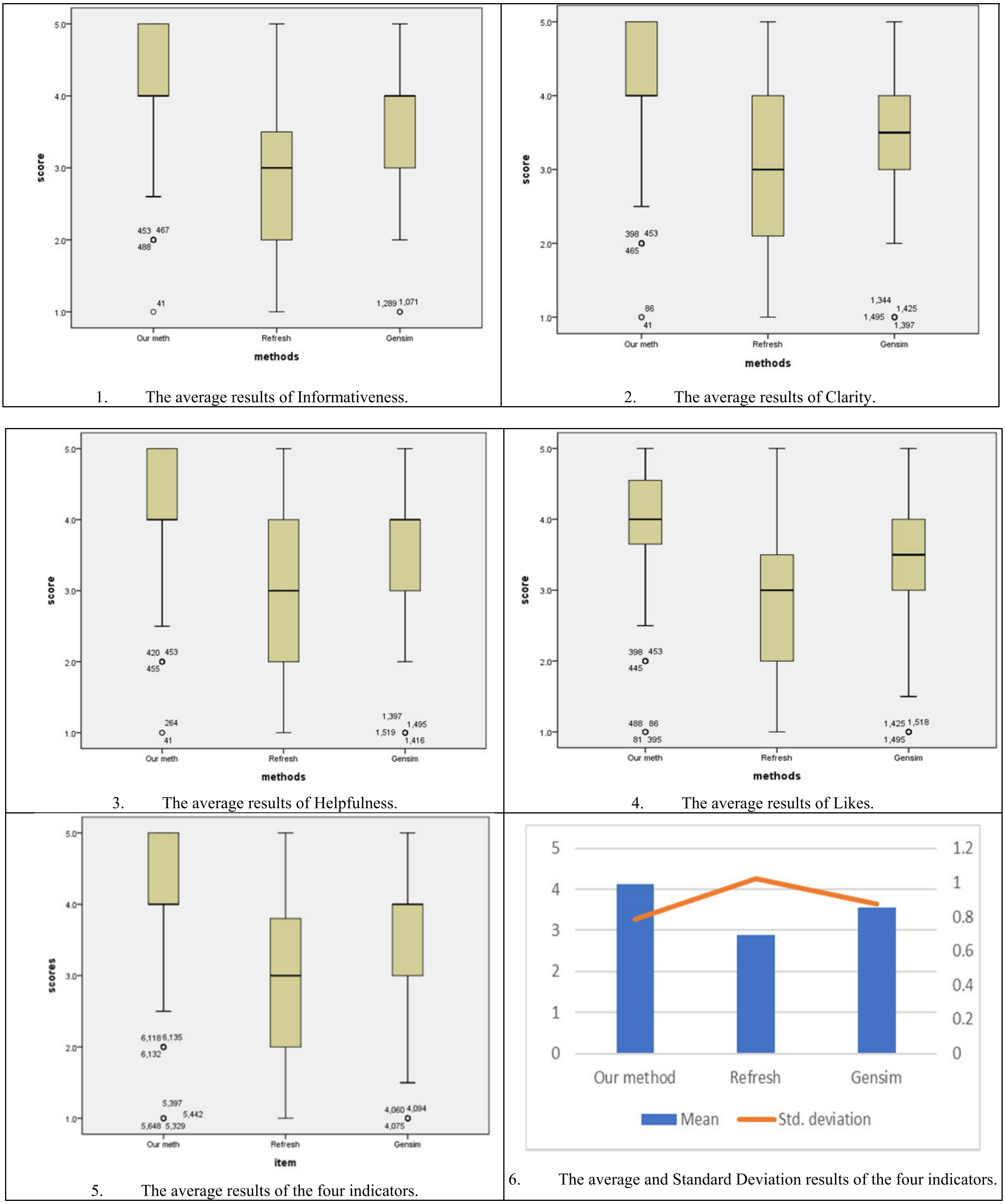
Avg./SD	Our method	Refresh	Gensim
Informativeness	4.2128/0.7420	2.8208/0.9852	3.8303/0.8009
Clarity	4.1505/0.7758	3.1027/0.9744	3.3988/0.8834
Helpfulness	4.1233/0.7765	2.8977/1.0548	3.6384/0.8359
Likes	4.0091/0.8218	2.7588/1.0429	3.3860/0.8874

TABLE 12. Descriptive statistics of our method, Refresh, and Gensim.

Group(I)	Mean	SD	Minimum	Maximum
Refresh	2.895	1.0224	1.00	5.00
Gensim	3.563	0.8717	1.00	5.00
Our method	4.124	0.7825	1.00	5.00

TABLE 13. Post-hoc Scheffé test results for our method, Refresh, and Gensim.

Group(I)	Group(J)	Mean Difference (I-J)	Std. Error	Sig.
Refresh	Refresh	-0.6684	0.0280	0.000
	Our method	-1.2289	0.0280	0.000
Gensim	Refresh	0.6684	0.0280	0.000
	Our method	-0.5605	0.0280	0.000
Our method	Refresh	1.2289	0.0280	0.000
	Gensim	0.56605	0.0280	0.000



1. The average results of Informativeness.
2. The average results of Clarity.
3. The average results of Helpfulness.
4. The average results of Likes.
5. The average results of the four indicators.
6. The average and Standard Deviation results of the four indicators.

FIGURE 11. Results of the four indicators.

tests to study the differences among Refresh, Gensim, and our proposed method. Table 12 shows that our method has a better mean and standard deviation than other methods. The ANOVA results of the three methods have a significant difference ($F = 967.672$, $p < 0.000$). The Scheffé post hoc results are shown in Table 13. The results showed that all

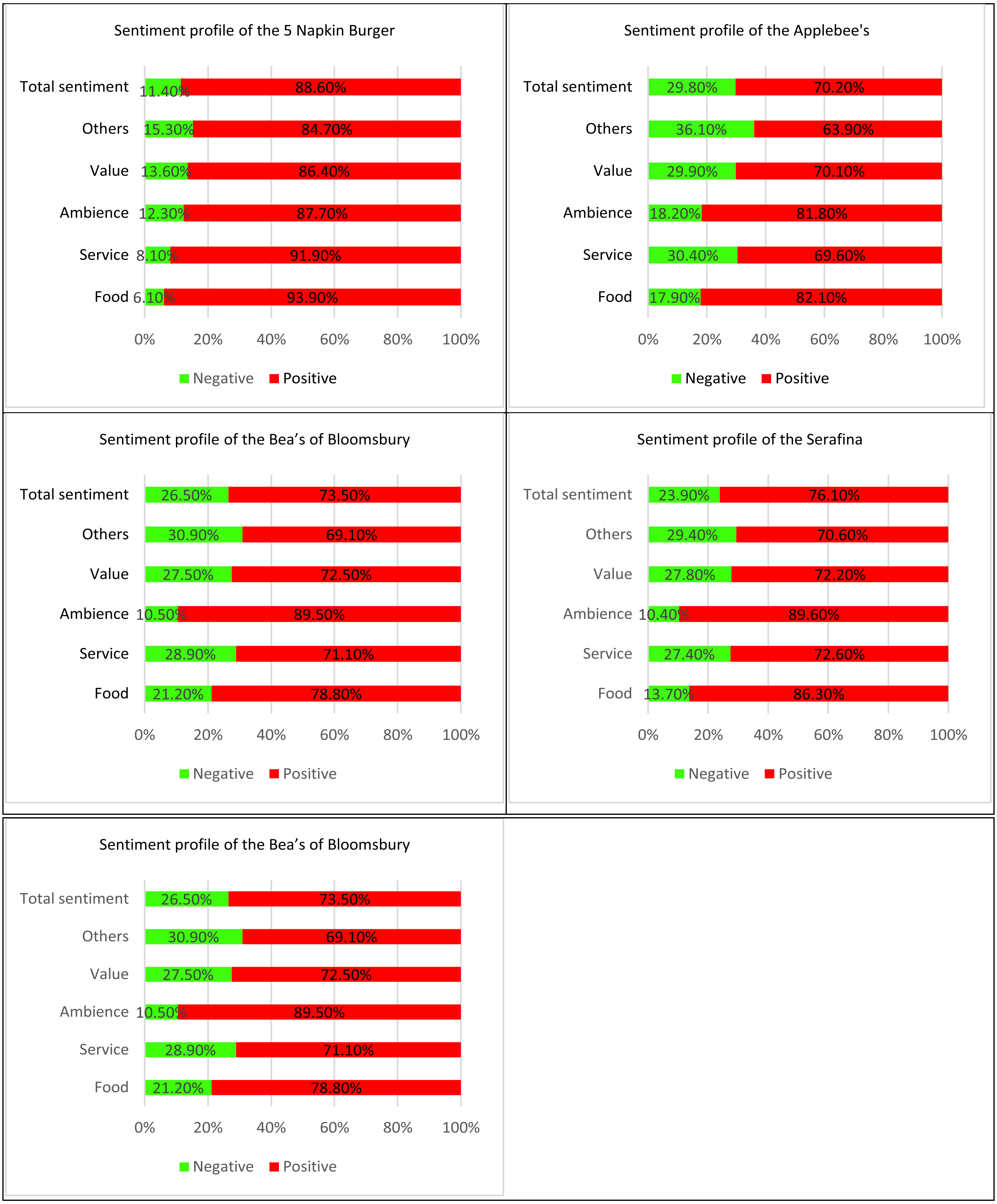


FIGURE 12. Sentiment profile of the five restaurants.

groups had significant differences. That is, the three summary methods are statistically different from each other.

Moreover, to help readers quickly grasp the emotional tendencies of various aspects of the summary, we calculated a

TABLE 14. The summary for 5 Napkin Burger by our method.

Restaurant name: 5 Napkin Burger
Location: 630 9th Ave, New York City, NY 10036
Tel: +1 212-757-2277
Total sentiment: #Top 26% in overall restaurants.

Food	#Top 10% in Food.	*Pos.:* Lots of choices, but we got the Beyond Meat burger with Vegan cheese which was really good can't even tell it's Vegan, along with the fresh cut fries and sweet potato fries. Neg.: We had no idea there was so much meat!
Service	#Top 16% in Service.	*Pos.:* The service was friendly and quick and we would definitely recommend if you're looking for a good burger. Neg.: We were not offered deserts which I thought was odd, they seemed to want the table.
Ambience	Top 28% in Ambience.	*Pos.:* Fabulous fun American atmosphere, very friendly staff and absolutely delicious burgers! Neg.: It's quite noisy inside, which isn't a problem, but just be aware
Value	Top 32% in Value.	*Pos.:* Great ambient, really great service and a huge burger... it is worth to try it, not cheap but worth every taste! Neg.: Overall a very poor experience.
Others	Top 37% in Others.	*Pos.:* Simply put, one of the best burgers I've ever had! Neg.: Being there with two kids was a not good experience since they let us sit next to the door where people where coming in and out

TABLE 15. The summary for 5 Napkin Burger by Refresh.

- But it is a clean, tasty place that doesn't lose quality and keeps making better options.
- Both were good portion sizes and super super tasty.
- We had a lovely meal here on New Years Eve 2019.

TABLE 16. The summary for 5 Napkin Burger by Gensim.

- We arrived slightly early and took a seat at the bar, drinks and service were both great!
- We were then called to our table, perfectly on time :)The waiter was very helpful and friendly and the food was really good!
- Had so much fun last time we went and I recommend the Jerk chicken Burger with loaded fries.
- Service was good, except the waiter memorized our order but when the food was brought out it was all mixed up at the table.
- We fancied a decent burger and checked out the reviews on tripadvisor so booked, was shown to our table by a really friendly girl and had a great meal in a busy restaurant with buzzing atmosphere.
- And walked around for a while then found this place, it was busy but we were told to wait at the bar where we ordered a beer, waited no more than 5 minutes and we were shown our table.
- The waiters was spot on, would you like some water etc and recommended the best burgers.
- The food was excellent, I had a 5N veggie burger, highly recommended but maybe a bit expensive for what it is.
- If you are looking for a great burger and fries, this is the place to go when on the West Side.
- Amazing burger restaurant on 9th avenue, great service with very tasty food, it is best to book a place in advance so you don't wait.

ratio of positive emotion scores and negative emotion scores as shown in Figure 12.

The findings of Table 6-13 and Figure 11 show that the method in this paper is significantly superior to the other two summary methods in terms of informativeness, clarity, helpfulness, and likes. Compared with the other two methods, our method can provide users with more information. This additional information in our template includes two-dimensional information classification based on topic and emotion. Therefore, users can obtain the necessary information about each topic and understand the customers' emotional tendencies. Experimental results prove that the template we proposed brings a good subjective feeling to users. It provides sufficient and clear information to help users understand the quality and condition of the entire restaurant at a glance. In addition, this template was well received by users.

V. CONCLUSION

This paper makes three contributions. First, it proposes a technique that can automatically generate a restaurant template with complete information, including the restaurant's basic information, pre-defined topics, and the most positive and negative representative sentences, as well as their positive and negative emotional ratio. Second, this paper uses four pre-defined topics to classify the information. These four topics are based on the choices that most users see on most websites, so that users can grasp the information faster. Finally, the method of this paper is superior to the other two summary methods in terms of informativeness, clarity, helpfulness, and likes. Therefore, it can be concluded that the method in this paper can provide sufficient and clear information about the target restaurant, and is helpful for users to understand the actual situation of the restaurant. Moreover, this template is well received by users.

In terms of future prospects, we hope that the target area is not limited to restaurants, but can be expanded to more areas and aspects. Although the method in this paper is designed for restaurants based on pre-defined topics, it should be able to achieve the same effect if it is regarded as other pre-defined areas.

APPENDIX

A. THE SUMMARY FOR 5 NAPKIN BURGER BY OUR METHOD

See Table 14.

B. THE SUMMARY FOR 5 NAPKIN BURGER BY REFRESH

See Table 15.

C. THE SUMMARY FOR 5 NAPKIN BURGER BY GENSIM

See Table 16.

REFERENCES

[1] D. K. Gaikwad and C. N. Mahender, "A review paper on text summarization," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 5, no. 3, pp. 154–160, 2016.
[2] V. Gupta and G. S. Lehal, "A survey of text summarization extractive techniques," *J. Emerg. Technol. Web Intell.*, vol. 2, no. 3, pp. 258–268, 2010, doi: 10.4304/jetwi.2.3.258-268.
[3] S. Bhatia, "A comparative study of opinion summarization techniques," *IEEE Trans. Comput. Social Syst.*, vol. 8, no. 1, pp. 110–117, Feb. 2021.
[4] O. Alcón and E. Lloret, "SemPCA-summarizer: Exploiting semantic principal component analysis for automatic summary generation," *Comput. Inform.*, vol. 37, no. 5, pp. 1126–1148, 2018, doi: 10.4149/cai_2018_5_1126.
[5] W.-J. Ye and A. J. T. Lee, "Mining sentiment tendencies and summaries from consumer reviews," *Inf. Syst. e-Bus. Manage.*, vol. 19, no. 1, pp. 107–135, Mar. 2021.
[6] S. J. Das, R. Murakami, and B. Chakraborty, "Development of a two-step LDA based aspect extraction technique for review summarization," *Int. J. Appl. Sci. Eng.*, vol. 18, no. 1, pp. 1–18, 2021.
[7] J. Khairnar and M. Kinikar, "Latent semantic analysis method used for mobile rating and review summarization," *Int. J. Comput. Sci. Telecommun.*, vol. 4, no. 6, pp. 61–67, 2013.
[8] Y. Gong and X. Liu, "Generic text summarization using relevance measure and latent semantic analysis," in *Proc. 24th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 2001, pp. 19–25.
[9] Y. J. Kumar, O. S. Goh, H. Basiron, N. H. Choon, and P. C. Suppiah, "A review on automatic text summarization approaches," *J. Comput. Sci.*, vol. 12, no. 4, pp. 178–190, Apr. 2016, doi: 10.3844/jcssp.2016.178.190.
[10] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: A survey," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 1–66, Jan. 2017, doi: 10.1007/s10462-016-9475-9.
[11] S. Ghodratnama, A. Beheshti, M. Zakershahrak, and F. Sobhanmanesh, "Extractive document summarization based on dynamic feature space mapping," *IEEE Access*, vol. 8, pp. 139084–139095, 2020.
[12] V. Priya and K. Umamaheswari, "Aspect-based summarisation using distributed clustering and single-objective optimisation," *J. Inf. Sci.*, vol. 46, no. 2, pp. 176–190, Apr. 2020.
[13] M. Fachrurrozi, N. Yusliani, and R. U. Yoanita, "Frequent term based text summarization for Bahasa Indonesia," in *Proc. Int. Conf. Innov. Eng. Technol. (ICIET)*, Bangkok, Thailand, Dec. 2013, pp. 25–26.
[14] A. P. Siva Kumar, P. Premchand, and A. Govardhan, "Query-based summarizer based on similarity of sentences and word frequency," *Int. J. Data Mining Knowl. Manage. Process*, vol. 1, no. 3, pp. 1–12, May 2011, doi: 10.5121/ijdkp.2011.1301.
[15] M. Litvak and M. Last, "Graph-based keyword extraction for single-document summarization," in *Proc. Workshop Multi-Source Multilingual Inf. Extraction Summarization (MMIES)*, 2008, pp. 17–24.
[16] Y. Kikuchi, T. Hirao, H. Takamura, M. Okumura, and M. Nagata, "Single document summarization based on nested tree structure," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics (Short Papers)*, vol. 2, 2014, pp. 315–320, doi: 10.3115/v1/P14-2052.
[17] L. Plaza, A. Díaz, and P. Gervás, "A semantic graph-based approach to biomedical summarisation," *Artif. Intell. Med.*, vol. 53, no. 1, pp. 1–14, Sep. 2011, doi: 10.1016/j.artmed.2011.06.005.
[18] R. Alqaisi, W. Ghanem, and A. Qaroush, "Extractive multi-document Arabic text summarization using evolutionary multi-objective optimization with K-medoid clustering," *IEEE Access*, vol. 8, pp. 228206–228224, 2020.
[19] C. S. Yadav, A. Sharan, R. Kumar, and P. Biswas, "A new approach for single text document summarization," in *Proc. 2nd Int. Conf. Comput. Commun. Technol.*Hyderabad, India: Springer, 2016, pp. 401–411, doi: 10.1007/978-81-322-2523-2_39.
[20] S. Ou, C. S.-G. Khoo, and D. H. Goh, "Design and development of a concept-based multi-document summarization system for research abstracts," *J. Inf. Sci.*, vol. 34, no. 3, pp. 308–326, Jun. 2008.
[21] I. Titov and R. McDonald, "Modeling online reviews with multi-grain topic models," in *Proc. 17th Int. Conf. World Wide Web (WWW)*, 2008, pp. 111–120.
[22] X. Xu, T. Meng, and X. Cheng, "Aspect-based extractive summarization of online reviews," in *Proc. ACM Symp. Appl. Comput. (SAC)*, 2011, pp. 968–975, doi: 10.1145/1982185.1982396.
[23] R. Perera, T. Malepathirana, Y. Abeysinghe, Y. Albar, and U. Thayasivam, "Amalgamation of general and domain specific word embeddings for improved hierarchical aspect aggregation," in *Proc. IEEE 13th Int. Conf. Semantic Comput. (ICSC)*, Jan. 2019, pp. 55–62.
[24] N. M. Shelke, S. Deshpande, and V. Thakre, "Exploiting expectation maximization algorithm for sentiment analysis of product reviews," in *Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Mar. 2017, pp. 390–396.
[25] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 15–21, Mar. 2013, doi: 10.1109/MIS.2013.30.
[26] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan, "OpinionFinder: A system for subjectivity analysis," in *Proc. HLT/EMNLP Interact. Demonstrations*, 2005, pp. 34–35, doi: 10.3115/1225733.1225751.
[27] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. 8th Int. AAAI Conf. Weblogs Social Media*, 2014, pp. 216–225.
[28] R. Mihalcea, "Graph-based ranking algorithms for sentence extraction, applied to text summarization," in *Proc. ACL Interact. Poster Demonstration Sessions*, 2004, pp. 170–173, doi: 10.3115/1219044.1219064.

[29] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999, doi: 10.1145/324133.324140.
[30] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, nos. 1–7, pp. 107–117, Apr. 1998.
[31] G. Tsatsaronis, I. Varlamis, and K. Nørvåg, "SemanticRank: Ranking keywords and sentences using semantic graphs," in *Proc. 23rd Int. Conf. Comput. Linguistics*, 2010, pp. 1074–1082.
[32] J.-Y. Yeh, H.-R. Ke, and W.-P. Yang, "ISpreadRank: Ranking sentences for extraction-based summarization using feature weight propagation in the sentence similarity network," *Expert Syst. Appl.*, vol. 35, no. 3, pp. 1451–1462, Oct. 2008, doi: 10.1016/j.eswa.2007.08.037.
[33] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2004, pp. 404–411.
[34] K. Z. K. Zhang, S. J. Zhao, C. M. K. Cheung, and M. K. O. Lee, "Examining the influence of online reviews on consumers' decision-making: A heuristic–systematic model," *Decis. Support Syst.*, vol. 67, pp. 78–89, Nov. 2014.
[35] I. Mani, "Summarization evaluation: An overview," in *Proc. NAACL Workshop Autom. Summarization*, 2001.
[36] Hardy, S. Narayan, and A. Vlachos, "HighRES: Highlight-based reference-less evaluation of summarization," 2019, *arXiv:1906.01361*. [Online]. Available: http://arxiv.org/abs/1906.01361
[37] N. Korfiatis, E. García-Bariocanal, and S. Sánchez-Alonso, "Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. Review content," *Electron. Commerce Res. Appl.*, vol. 11, no. 3, pp. 205–217, May 2012, doi: 10.1016/j.elerap.2011.10.003.
[38] S. Bhattacharyya and I. Bose, "S-commerce: Influence of Facebook likes on purchases and recommendations on a linked e-commerce site," *Decis. Support Syst.*, vol. 138, Nov. 2020, Art. no. 113383.
[39] S. Narayan, S. B. Cohen, and M. Lapata, "Ranking sentences for extractive summarization with reinforcement learning," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., (Long Papers)*, vol. 1, 2018, pp. 1747–1759, doi: 10.18653/v1/N18-1158.
[40] R. Rehurek and P. Sojka, "Software framework for topic modelling with large corpora," in *Proc. LREC Workshop New Challenges NLP Frameworks*. Princeton, NJ, USA: Citeseer, 2010, pp. 46–50, doi: 10.13140/2.1.2393.1847.

YENLIANG CHEN received the Ph.D. degree in computer science from the National Tsing Hua University, Hsinchu, Taiwan. He is currently a Professor of information management with the National Central University, Taiwan. He has published articles in *Decision Support Systems*, *Information & Management*, *Operations Research*, *Decision Sciences*, IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, *Electronic Commerce Research and Applications*, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A AND B, *Transportation Research—Part B*, *European Journal of Operational Research*, *Naval Research Logistics*, *Journal of American Society for Information Science and Technology*, *Information Processing & Management*, and many others. His research interests include data mining, information retrieval, knowledge management, and decision-making models. He is the Former Editor-in-Chief of *Journal of Information Management* and *Journal of E-Business*.

CHIALING CHANG received the Ph.D. degree from the National Central University, Taiwan. She is currently a Professor of information and library science with Tamkang University, Taiwan. She has published in *Decision Support Systems*, *Expert Systems with Applications*, *Journal of Electronic Commerce Research*, *Sustainability*, and *Advanced Science Letter*. Her current research interests include EC technologies, information retrieval, and data mining.

JERYEU GAN received the master's degree from the Department of Information Management, National Central University, Taiwan. His current research interests include data mining, information retrieval, and EC technologies.