

Research Service Issues of Chiang Mai Chinese Restaurants to Chinese Tourists Based on Text Analysis and Sentiment Analysis

1. Problem

The problem addressed in the study "Research Service Issues of Chiang Mai Chinese Restaurants to Chinese Tourists Based on Text Analysis and Sentiment Analysis" is identifying specific service issues that Chinese tourists face when dining at Chinese restaurants in Chiang Mai, Thailand. The motivation behind this research stems from the significant number of Chinese tourists visiting Chiang Mai and the importance of satisfying their dining experiences to foster cultural exchange and understanding between Thai and Chinese communities. The study aims to provide insights into how Chinese restaurants can improve their service quality to meet the needs and expectations of Chinese tourists, thereby enhancing their satisfaction and potentially increasing revenue for the restaurants.

2. Motivation

The motivation for this research is further driven by the recognition that the growth of Chinese tourism in Chiang Mai has contributed to the success of the city's Chinese restaurants. However, there have been reports of service problems that have left Chinese tourists dissatisfied, such as language barriers, cultural misunderstandings, and concerns about overcharging or poor food quality. Addressing these issues is crucial for the restaurants to maintain a positive reputation and attract more Chinese tourists, who often seek familiar cuisine while traveling.

The study aims to identify the main service problems for Chinese tourists in Chiang Mai Chinese restaurants and propose solutions based on the identified issues. This is achieved through a combination of text analysis and sentiment analysis of reviews from Dianping.com, a popular Chinese-language restaurant review and social networking site, to understand Chinese tourists' preferences and behaviors and uncover the factors that most affect their satisfaction with the dining experience.

3. Activity

- (1) Use text analysis to filter related Chinese restaurants in Chiang Mai with review conditions
- (2) Use Python crawler to obtain all reviews, restaurants, user names, content, and other information of 20 restaurants.
- (3) Preprocess the obtained text
- (4) Use 'weiciyun' for sentiment analysis
- (5) Use Google API for Sentiment Analysis
- (6) Create a sentiment analysis model based on NLP for analysis
- (7) Compare the differences between weiciyun, Google API, and new models
- (8) Statistics on specific service issues of 20 Chinese restaurants in Chiang Mai
- (9) Statistics on the specific service issues of the top 10 Chinese restaurants in Chiang Mai

- (10) Statistics on the specific service issues of the bottom 10 Chinese restaurants in Chiang Mai
- (11) Analyze the causes and solutions to problems

4. Outputs

- (1) A list of Chinese restaurants in Chiang Mai that meet specific review criteria.
- (2) A dataset containing reviews, restaurant details, usernames, and content for 20 selected restaurants.
- (3) Cleaned and normalized text data ready for analysis.
- (4) Sentiment classifications for each review using Weiciyun.
- (5) Sentiment scores and classifications for each review using Google API.
- (6) A custom-built sentiment analysis model tailored for this specific dataset.
- (7) A comparative analysis of sentiment analysis results from Weiciyun, Google API, and the custom NLP model.
- (8) A detailed report on specific service issues identified across the 20 restaurants
- (9) A focused analysis of the top 10 Chinese restaurants, highlighting prevalent service issues.
- (10) A focused analysis of the bottom 10 Chinese restaurants, highlighting prevalent service issues.
- (11) A comprehensive analysis report that outlines the root causes of the identified service issues and proposes specific, actionable solutions tailored to each problem.

5. Outcomes

- (1) Enhanced precision in targeting relevant restaurants, ensuring the study focuses on the most pertinent establishments to improve the relevance and applicability of the analysis.
- (2) Creation of a comprehensive database providing a rich foundation for in-depth analysis, facilitating the development of customized solutions.
- (3) Improved data quality ensures the accuracy of analyses, laying a solid foundation for subsequent sentiment analysis and model training.
- (4) Initial sentiment orientation analysis provides baseline data for comparative analysis.
- (5) Advanced API tools validate and supplement Weiciyun analysis results, adding depth and breadth to the analysis.
- (6) Custom model development enhances analysis precision and adaptability, providing more accurate sentiment analysis for the specific dataset.
- (7) Identify the most suitable sentiment analysis tool for the project's needs, guiding future research and practice.
- (8) Clear identification of service shortfalls provides restaurants with directions for improvement, directly promoting service quality enhancement.
- (9) In-depth analysis of service quality in top-performing restaurants sets higher service standards for the industry benchmark.
- (10) Help these establishments understand their critical service flaws and enable them to prioritize improvements, ultimately striving to elevate their service standards to meet customer expectations more effectively.
- (11) Implementing effective solutions based on this analysis is expected to address and mitigate

the service directly challenges the restaurants face. This strategic approach aims to enhance overall customer satisfaction, improve service quality, and potentially elevate the restaurants' ratings and reputation among Chinese tourists in Chiang Mai.

6. Data source

Chiang Mai Chinese restaurant reviews on Dianping.com.

7. Methodology

(1) Crawler

A crawler is an automated program used to scrape information from the Internet. It includes components such as a crawler engine, scheduler, downloader, parser, and storage to access the target website and extract the required data by simulating the behavior of human users. The basic workflow of a crawler covers selecting targets, analyzing web page structure, selecting a crawler framework, writing crawler code, simulating user behavior, handling exceptions, crawling data, and storage and analysis. When using crawlers, you need to abide by laws, regulations and website usage agreements, respect privacy and intellectual property rights, handle anti-crawler mechanisms, regularly update the code to adapt to website changes, and ensure the quality of crawled data.

(2) Weiciyun

Weiciyun is a highly practical and simple online word cloud and online word cloud generator. Compared with other products, this product has more powerful functions. It not only supports online word segmentation, but also supports word frequency statistics and word frequency analysis, and can produce high-quality Works and drawing boards provide common copyright-free fonts such as simplified Chinese, traditional Chinese, and English in various formats. Whether it is design, operation, or data analysis, we can create eye-catching word cloud designs for users.

(3) Google Cloud Natural Language API

Google Cloud Natural Language API provides a sentiment analysis service that allows you to perform sentiment analysis on text through the API interface. Its main functions are:

- **Sentiment Analysis:** The Google Cloud Natural Language API is able to analyze sentiment in text and return the overall sentiment polarity of the text. Sentiment polarity typically includes positive, negative, or neutral.
- **Entity analysis:** In addition to sentiment analysis, the API can also identify entities in text (such as names of people, places, organizations, etc.) and provide categories and related information of the entities.
- **Syntax analysis:** Provides grammatical structure analysis of the text, including syntax tree and dependency analysis, which helps to understand the relationship between words in the text.

(4) Self-built sentiment analysis model

- Data collection
 - 1) Get a labeled sentiment analysis dataset. This can be a dataset containing text and corresponding sentiment labels, such as positive, negative, or neutral.
 - 2) Ensure the diversity and representativeness of the data set to improve model performance.
- Data preprocessing:
 - 1) Clean text data, including removing special characters, punctuation, and stop words.
 - 2) Tokenize the text and split the text into words or sub-words.
 - 3) Encode labels, such as positive, negative, neutral, into numerical values.
- Partition the data set:
 - 1) Divide the data set into training set, validation set and test set. The training set is used to train the model, the validation set is used to adjust the hyperparameters of the model, and the test set is used to evaluate the performance of the model.
- Select model architecture:
 - 2) Choose a model architecture suitable for sentiment analysis tasks, such as Recurrent Neural Network (RNN), Long Short-Term Memory Network (LSTM) or Transformer (Transformer), etc.
 - 3) You can try to use pre-trained models, such as BERT, GPT, etc., for fine-tuning.
- Modeling:
 - 1) Train a sentiment analysis model on the training set using the selected model architecture.
 - 2) Set appropriate parameters such as batch size, etc.
- Model evaluation:
 - 1) Evaluate model performance on the validation set and adjust model structure and hyperparameters to improve performance.
 - 2) Use evaluation metrics such as accuracy, precision, recall and F1 score to evaluate the effectiveness of the model.
- Optimization model
 - 1) Optimize the model, including adjusting hyperparameters, using regularization techniques, adding data enhancement and other methods.
- Model testing:
 - 1) Use the test set to evaluate the performance of the final model to ensure that the model performs well on unseen data.
- keep improve:
 - 1) Monitor the performance of the model and make updates and improvements based

on actual conditions.

8. Evaluation metrix

1. Accuracy: Measures the proportion of correctly predicted observations to the total observations, offering a general sense of model performance across all classes.
2. Confusion Matrix: Provides a detailed breakdown of predictions versus actual labels, helping to identify models' performance with respect to each class, including true positives, false positives, true negatives, and false negatives.
3. Precision: The ratio of correctly predicted positive observations to the total predicted positives, important for understanding the model's quality of positive predictions.
4. Recall (Sensitivity): The ratio of correctly predicted positive observations to all observations in the actual class, crucial for gauging the model's ability to find all relevant cases.
5. F1 Score: A weighted average of Precision and Recall, useful for comparing model performance when there are class imbalances.

9. Final product

- A detailed report with insights on the service issues and their impact on customer satisfaction.
- A set of recommendations tailored to address the identified issues, including strategies for environmental and hygiene improvements, food quality enhancement, and language service provision.
- A dashboard or visualization tool that presents the analysis results in an easy-to-understand format for stakeholders.

Team name: Freshman

Team Icon:



Team member:

First Name	Last Name	ID
YuanKun	Gu	662131025
HaiTao	Liu	662131023
ShengZhe	Huang	662131024

Research Link: <https://ieeexplore.ieee.org/document/10329710>