Sentiment Analysis for Restaurant Tweets

Team Name: TasteLens AI

1.Introduction:

The field of sentiment analysis has become essential for businesses to interpret unstructured text data and understand customer emotions, opinions, or attitudes. With the increasing influence of social media, especially in the restaurant industry, sentiment analysis offers a powerful way to gain real-time insights into customer satisfaction. By analyzing tweets about a restaurant, we can capture public perception and respond promptly to customer needs. In this project, we aim to leverage NLP and machine learning to build a model specifically tailored for classifying restaurant-related tweets as positive, negative, or neutral, thereby providing the restaurant with valuable insights to improve its services.

2.Problem Statement:

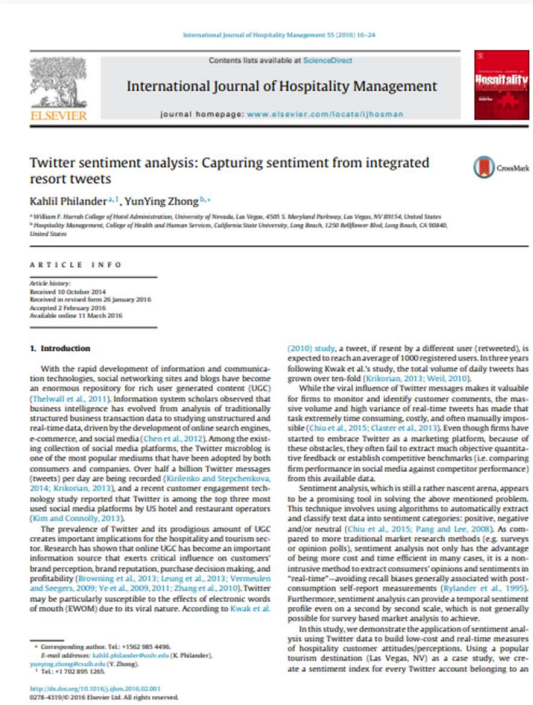
Monitoring customer satisfaction is a continuous challenge in the restaurant industry, especially when dealing with large volumes of social media feedback. Many existing sentiment analysis tools lack industry-specific language processing and often struggle with nuances such as sarcasm, mixed sentiments, or restaurant-related jargon. This results in misinterpretation or loss of valuable feedback insights. To solve this, our project will develop a machine learning model specifically designed to capture and classify sentiment in restaurant-related tweets, allowing the restaurant to understand customer trends and concerns more accurately.

3.Goals:

- **Customized Machine Learning Model:** Develop a machine learning model to classify tweets related to the restaurant as positive, negative, accounting for unique restaurant-industry language and sentiment nuances.
- **Identification of Trends and Key Themes**: Recognize common themes and satisfaction trends in customer feedback, helping the restaurant track changing opinions and common issues.
- **Actionable Insights for Service Improvement:** Use sentiment data to generate targeted recommendations that can guide the restaurant in enhancing customer satisfaction and addressing common feedback concerns.

4.Related Work:

- Twitter sentiment analysis: Capturing sentiment from integrated resort tweets:



International Journal of Hospitality Management 55 (2016) 16–24

Contents lists available at ScienceDirect

International Journal of Hospitality Management

journal homepage: www.elsevier.com/locate/ijhosman

Twitter sentiment analysis: Capturing sentiment from integrated resort tweets

Kahlil Philander[a,1], YunYing Zhong[b,*]

[a] William F. Harrah College of Hotel Administration, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, Las Vegas, NV 89154, United States
[b] Hospitality Management, College of Health and Human Services, California State University, Long Beach, 1250 Bellflower Blvd, Long Beach, CA 90840, United States

ARTICLE INFO

Article history:
Received 10 October 2014
Received in revised form 26 January 2016
Accepted 2 February 2016
Available online 11 March 2016

1. Introduction

With the rapid development of information and communication technologies, social networking sites and blogs have become an enormous repository for rich user generated content (UGC) (Thelwall et al., 2011). Information system scholars observed that business intelligence has evolved from analysis of traditionally structured business transaction data to studying unstructured and real-time data, driven by the development of online search engines, e-commerce, and social media (Chen et al., 2012). Among the existing collection of social media platforms, the Twitter microblog is one of the most popular mediums that have been adopted by both consumers and companies. Over half a billion Twitter messages (tweets) per day are being recorded (Kirilenko and Stepchenkova, 2014; Krikorian, 2013), and a recent customer engagement technology study reported that Twitter is among the top three most used social media platforms by US hotel and restaurant operators (Kim and Connolly, 2013).

The prevalence of Twitter and its prodigious amount of UGC creates important implications for the hospitality and tourism sector. Research has shown that online UGC has become an important information source that exerts critical influence on customers' brand perception, brand reputation, purchase decision making, and profitability (Browning et al., 2013; Leung et al., 2013; Vermeulen and Seegers, 2009; Ye et al., 2009, 2011; Zhang et al., 2010). Twitter may be particularly susceptible to the effects of electronic words of mouth (EWOM) due to its viral nature. According to Kwak et al. (2010) study, a tweet, if resent by a different user (retweeted), is expected to reach an average of 1000 registered users. In three years following Kwak et al.'s study, the total volume of daily tweets has grown over ten-fold (Krikorian, 2013; Weil, 2010).

While the viral influence of Twitter messages makes it valuable for firms to monitor and identify customer comments, the massive volume and high variance of real-time tweets has made that task extremely time consuming, costly, and often manually impossible (Chiu et al., 2015; Claster et al., 2013). Even though firms have started to embrace Twitter as a marketing platform, because of these obstacles, they often fail to extract much objective quantitative feedback or establish competitive benchmarks (i.e. comparing firm performance in social media against competitor performance) from this available data.

Sentiment analysis, which is still a rather nascent arena, appears to be a promising tool in solving the above mentioned problem. This technique involves using algorithms to automatically extract and classify text data into sentiment categories: positive, negative and/or neutral (Chiu et al., 2015; Pang and Lee, 2008). As compared to more traditional market research methods (e.g. surveys or opinion polls), sentiment analysis not only has the advantage of being more cost and time efficient in many cases, it is a non-intrusive method to extract consumers' opinions and sentiments in "real-time"—avoiding recall biases generally associated with post-consumption self-report measurements (Rylander et al., 1995). Furthermore, sentiment analysis can provide a temporal sentiment profile even on a second by second scale, which is not generally possible for survey based market analysis to achieve.

In this study, we demonstrate the application of sentiment analysis using Twitter data to build low-cost and real-time measures of hospitality customer attitudes/perceptions. Using a popular tourism destination (Las Vegas, NV) as a case study, we create a sentiment index for every Twitter account belonging to an

* Corresponding author. Tel.: +1562 985 4496.
E-mail addresses: kahlil.philander@unlv.edu (K. Philander), yunying.zhong@csulb.edu (Y. Zhong).
[1] Tel.: +1 702 895 1265.

http://dx.doi.org/10.1016/j.ijhm.2016.02.001
0278-4319/

- Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques:

Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques

Akshay Krishna, V. Akhilesh, Animikh Aich and Chetana Hegde

Abstract Evolution of the Internet in the past decade resulted in generation of voluminous data in all sectors. Due to these advents, the people have new ways of expressing their opinions about anything in the form of tweets, blog posts, online discussion forums, status updates, etc. Sentiment analysis deals with the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude toward a particular topic is positive, negative, or neutral. Knowing the opinion of customers is very important for any business. Hence, in this paper, we analyze the reviews given by the customers of the restaurant with the help of machine learning classification algorithms. This paper mainly focuses on the implementation of various classification algorithms and their performance analysis. The simulation results showed that SVM classifier resulted in the highest accuracy of 94.56% for the given dataset.

Keywords Decision Tree · K-Nearest Neighbor · Naïve Bayes · Random Forest · ROC · Supervised learning · Support Vector Machine

1 Introduction

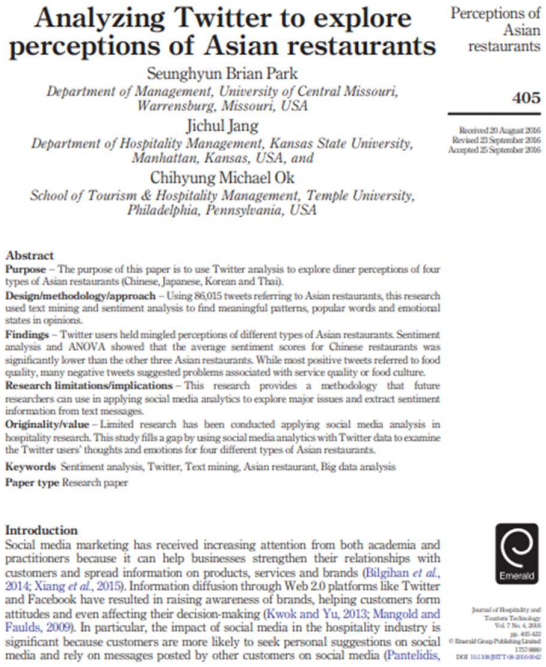
The exponential increase in the use of the Internet has led massive online activities (like chatting, video calling, conferencing, surveillance, ticket booking, e-commerce, online transactions, social media communications, and blogging). This enforces the need to extract, transform, load, and analyze a huge amount of unstructured, structured, and heterogeneous data, at a fast pace [1]. The process of analyzing such a

A. Krishna (✉) · V. Akhilesh · A. Aich
RNS Institute of Technology, Bangalore 560098, India
e-mail: akshay.krishna154@gmail.com

C. Hegde
Manipal Global Education Services, Bangalore 560100, India
e-mail: chetanahegde@ieee.org

V. Sridhar et al. (eds.), *Emerging Research in Electronics, Computer Science and Technology*, Lecture Notes in Electrical Engineering 545,
https://doi.org/10.1007/978-981-13-5802-9_60

- Analyzing Twitter to explore perceptions of Asian restaurants:



Analyzing Twitter to explore perceptions of Asian restaurants

Seunghyun Brian Park
Department of Management, University of Central Missouri, Warrensburg, Missouri, USA
Jichul Jang
Department of Hospitality Management, Kansas State University, Manhattan, Kansas, USA, and
Chihyung Michael Ok
School of Tourism & Hospitality Management, Temple University, Philadelphia, Pennsylvania, USA

Received 20 August 2016
Revised 22 September 2016
Accepted 25 September 2016

Abstract

Purpose – The purpose of this paper is to use Twitter analysis to explore diner perceptions of four types of Asian restaurants (Chinese, Japanese, Korean and Thai).

Design/methodology/approach – Using 86,015 tweets referring to Asian restaurants, this research used text mining and sentiment analysis to find meaningful patterns, popular words and emotional states in opinions.

Findings – Twitter users held mingled perceptions of different types of Asian restaurants. Sentiment analysis and ANOVA showed that the average sentiment scores for Chinese restaurants was significantly lower than the other three Asian restaurants. While most positive tweets referred to food quality, many negative tweets suggested problems associated with service quality or food culture.

Research limitations/implications – This research provides a methodology that future researchers can use in applying social media analytics to explore major issues and extract sentiment information from text messages.

Originality/value – Limited research has been conducted applying social media analysis in hospitality research. This study fills a gap by using social media analytics with Twitter data to examine the Twitter users' thoughts and emotions for four different types of Asian restaurants.

Keywords Sentiment analysis, Twitter, Text mining, Asian restaurant, Big data analysis

Paper type Research paper

Introduction

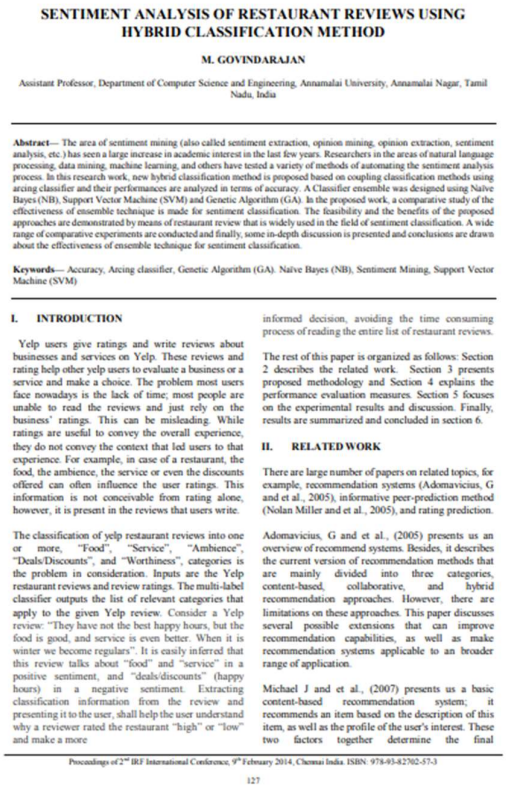
Social media marketing has received increasing attention from both academia and practitioners because it can help businesses strengthen their relationships with customers and spread information on products, services and brands (Bilgihan *et al.*, 2014; Xiang *et al.*, 2015). Information diffusion through Web 2.0 platforms like Twitter and Facebook have resulted in raising awareness of brands, helping customers form attitudes and even affecting their decision-making (Kwok and Yu, 2013; Mangold and Faulds, 2009). In particular, the impact of social media in the hospitality industry is significant because customers are more likely to seek personal suggestions on social media and rely on messages posted by other customers on social media (Pantelidis,

Emerald

Journal of Hospitality and Tourism Technology
Vol. 7 No. 4, 2016
pp. 405-422

1757-9880
DOI 10.1108/JHTT-08-2016-0042

- Sentiment Analysis of Restaurant Reviews Using Hybrid Classification Method:



SENTIMENT ANALYSIS OF RESTAURANT REVIEWS USING HYBRID CLASSIFICATION METHOD

M. GOVINDARAJAN

Assistant Professor, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, Tamil Nadu, India

Abstract— The area of sentiment mining (also called sentiment extraction, opinion mining, opinion extraction, sentiment analysis, etc.) has seen a large increase in academic interest in the last few years. Researchers in the areas of natural language processing, data mining, machine learning, and others have tested a variety of methods of automating the sentiment analysis process. In this research work, new hybrid classification method is proposed based on coupling classification methods using arcing classifier and their performances are analyzed in terms of accuracy. A Classifier ensemble was designed using Naïve Bayes (NB), Support Vector Machine (SVM) and Genetic Algorithm (GA). In the proposed work, a comparative study of the effectiveness of ensemble technique is made for sentiment classification. The feasibility and the benefits of the proposed approaches are demonstrated by means of restaurant review that is widely used in the field of sentiment classification. A wide range of comparative experiments are conducted and finally, some in-depth discussion is presented and conclusions are drawn about the effectiveness of ensemble technique for sentiment classification.

Keywords— Accuracy, Arcing classifier, Genetic Algorithm (GA), Naïve Bayes (NB), Sentiment Mining, Support Vector Machine (SVM)

I. INTRODUCTION

Yelp users give ratings and write reviews about businesses and services on Yelp. These reviews and rating help other yelp users to evaluate a business or a service and make a choice. The problem most users face nowadays is the lack of time; most people are unable to read the reviews and just rely on the business' ratings. This can be misleading. While ratings are useful to convey the overall experience, they do not convey the context that led users to that experience. For example, in case of a restaurant, the food, the ambience, the service or even the discounts offered can often influence the user ratings. This information is not conceivable from rating alone, however, it is present in the reviews that users write.

The classification of yelp restaurant reviews into one or more, "Food", "Service", "Ambience", "Deals/Discounts", and "Worthiness", categories is the problem in consideration. Inputs are the Yelp restaurant reviews and review ratings. The multi-label classifier outputs the list of relevant categories that apply to the given Yelp review. Consider a Yelp review: "They have not the best happy hours, but the food is good, and service is even better. When it is winter we become regulars". It is easily inferred that this review talks about "food" and "service" in a positive sentiment, and "deals/discounts" (happy hours) in a negative sentiment. Extracting classification information from the review and presenting it to the user, shall help the user understand why a reviewer rated the restaurant "high" or "low" and make a more informed decision, avoiding the time consuming process of reading the entire list of restaurant reviews.

The rest of this paper is organized as follows: Section 2 describes the related work. Section 3 presents proposed methodology and Section 4 explains the performance evaluation measures. Section 5 focuses on the experimental results and discussion. Finally, results are summarized and concluded in section 6.

II. RELATED WORK

There are large number of papers on related topics, for example, recommendation systems (Adomavicius, G and et al., 2005), informative peer-prediction method (Nolan Miller and et al., 2005), and rating prediction.

Adomavicius, G and et al., (2005) presents us an overview of recommend systems. Besides, it describes the current version of recommendation methods that are mainly divided into three categories, content-based, collaborative, and hybrid recommendation approaches. However, there are limitations on these approaches. This paper discusses several possible extensions that can improve recommendation capabilities, as well as make recommendation systems applicable to an broader range of application.

Michael J and et al., (2007) presents us a basic content-based recommendation system; it recommends an item based on the description of this item, as well as the profile of the user's interest. These two factors together determine the final