International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Web Application Using Aspect-Based Sentiment Classification for Improved Tourism Experience

Dr. (Mrs.) N. F. Shaikh and Anindita Ghosh

Department of Computer Engineering, MES College of Engineering, Pune, Maharashtra, India.

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Tourism has become popular across the world in important sectors making it a significant tool in improvement of a country's economy. Nowadays with the boom of the internet services and webpages tourists want to be well prepared and well aware of the place they are planning to visit way ahead of their actual visit. This has become expected and possible because of the availability of large number of opinions and reviews about anything and everything. Similarly tourist reviews are information sources for tourists to know about tourist places beforehand. Users express their views and opinions regarding historical places and services or availability of resources. These opinions are subjective information which represents users' sentiments, feelings or appraisal regarding the same. Unfortunately some reviews are either irrelevant or not from a true source. Thus they often become noisy data. Aspect-Based Sentiment classification has shown promise in suppressing such noise. When the aggregated data about the tourist places is presented in the right way, analyzed and classified by the correct aspect-based algorithm, it could be translated into meaningful information for making vital decisions by tourism enthusiasts. However, not much research has been done on automatic aspect identification, and identification of implicit, infrequent and co-referential aspects, resulting in misclassifications. In this paper we present a framework that employs different aspect-based algorithms to identify aspects and classify sentiments and opinions inherent in tourist reviews with high accuracy. The framework has been implemented as a mobile application that helps tourists find the best historical places of interest, and performance has been evaluated by conducting experiments on real-world datasets.

Keywords: *Decision tree, classification, Clustering, Machine Learning, Aspect-based Sentiment Analysis, Opinion mining, products and services, travel, accommodations, Consumer Reviews, food, hotel rooms, Data mining, tourism etc.*

Introduction

The rapid growth of the tourism industry and its growing success and income rate has made it one of the most important earners to countries globally. With the easy access of internet almost everywhere around the globe and the increasing use of smart phones and PCs, the tourists can instantly share their opinions and views on different social platforms and websites by writing reviews. These reviews provide the travel planners with a broad range of views and opinions regarding a tourist place. This might lead to confusion due to the diversity of sentiments, regarding taking the decision of whether to visit a place or not. The methods for sentiment classification and analysis help categorize and classify sentiments into positive and negative and organize the reviews. However the problem is that each historical place has many aspects that can be reviewed by users so a binary result is not always helpful. So aspect based classification helps in this matter. For example if we take a review "Food was yummy but place was dirty", it has two aspects; "food" and "place". "Yummy" is a positive sentiment so food is classified as positive and "dirty" being a negative word place is classified as negative. The whole process involves two different tasks.1) Aspect Identification and 2) Sentiment

Classification into positive and negative.

This paper studies different works on sentiment classification and proposes a detailed framework of aspect-based sentiment classification by using machine learning algorithms. It uses few algorithms like Naïve Bayes, Support Vector Machines (SVM), Random Forest Tree, Logistic Regression, Maximum Entropy etc. This method uses mainly two modules of tasks. 1) Aspect based sentiment identification of implicit explicit aspects and grouping of co-referential words or synonyms to better evaluate the results using decision tree methods. 2) Next these aspects will be classified into positively or negatively reviewed using machine

learning algorithms. In the first stage word or term tokenization is done by detecting and filtering words or phrases between sentiment words and aspects.

Literature Survey

[1] Juanjuan Zhao, Fan Zhang, Lai Tu, Chengzhong Xu, Dayong Shen. "Estimation of Passenger Route Choice Pattern Using Smart Card Data for Complex Metro Systems".
This paper offers an answer that doesn't require any extra hardware or human mediation other than the AFC frameworks. It is building up a probabilistic model that can appraise how traveler streams are delivered to different courses and prepares from experimental examination. It utilizes a huge scale informational collection from the Shenzhen Metro framework to confirm the methodology used. The deliberate outcomes give us helpful information when constructing the model of decision for the traveler's way.

[2] Yu Lu, Archan Misra, Wen Sun, and Huayu Wu, "Smartphone Sensing Meets Transport Data: A Collaborative Framework for Transportation Service Analytics".
Famous unknown city-scale informational collections, (for example, taxi appointments and GPS directions) give bits of knowledge into the total movement of transportation foundation, however don't uncover singular transportation communications, (for example, holding up times in taxi lines); while versatile detecting data that catch individual explicit driving related exercises, yet experiences accuracy and overhead force.

[3] Hongzhi Yin, Weiqing Wang, Hao Wang, Ling Chen and Xiaofang Zhou, "Spatial-Aware Hierarchical Collaborative Deep Learning for POI Recommendation,"
They remain on late advances in profound figuring out how to handle these issues and propose a Spatial-Aware Hierarchical Collaborative Deep Learning (SH-CDL) framework. Together, the model directs profound portrayal taking in for POIs from heterogeneous attributes and progressively added substance portrayal learning for individual inclinations that are spatially mindful. To request to battle data scarcity to spatialmindful client particular demonstrating, both the open's aggregate inclinations in an objective district and the client's close to home inclinations in neighboring areas are misused as social regularization and spatial smoothing. We fuse a late element combination methodology into our SH-CDL framework to manage the multi-modular heterogeneous highlights of the POIs. The far reaching test examination shows that, especially in out - of-town and cold-start suggestion situations, our proposed model beats condition of - the-workmanship suggestion models.

[4] Meng Qu, Hengshu Zhu, Junming Liu. "A Cost Effective Recommender System for Taxi Drivers." The worldview of versatile administrations has changed with GPS innovation and better approaches for urban arranging. The plenitude of GPS follows has in this manner permitted better approaches for directing taxi business. Ongoing endeavors to create phone proposal frameworks for cabbies utilizing hints of Taxi GPS have really been made. To expand the likelihood of finding a customer with the most brief driving separation, these frameworks may propose a succession of get focuses. Nonetheless, cabbies ' pay is unequivocally associated with the powerful driving hours in reality. At the end of the day, understanding the real driving courses to lessen driving time before arriving at a client is progressively significant for cab drivers. To this end, we recommend making a savvy arrangement of rules for cabbies right now. The improvement reason for existing is to boost their benefits on the off chance that they follow the proposed courses to discover travelers. In particular, to decide the potential benefits of driving streets, we first model a net benefit target strategy. Rather, by mining the chronicled taxi GPS follows, we make a chart portrayal of street arranges and give a Brute-Force system to create ideal proposal driving way.

So as to adequately scan for ideal applicant courses, we build up another recursion methodology dependent on the exceptional type of the net benefit work. Specifically, rather than proposing a grouping of get focuses and requesting that the driver decide how to arrive, our suggestion framework is fit for conveying a full driving way, so drivers can find a client with the best potential advantage by following the proposals. At long last, we perform point by point contemplates on a true informational collection acquired from the San Francisco Bay zone and the exploratory outcomes plainly show the adequacy of the proposed suggestion framework.

[5 Carl Yang, Lanxiao Bai, Chao Zhang, .an Yuan, "Bridging Collaborative Filtering and Semi-Supervised Learning: A Neural Approach for POI Recommendation,"
Recommender structure is one of the most well-known points in information mining which keeps on drawing huge enthusiasm from both scholarly community and industry. Between them, the guidance for POI (focal point) is very viable however testing: it significantly benefits the two customers and organizations, all things considered, yet it is troublesome because of data shortage and distinctive setting. While various calculations try to take care of the specific information and issue settings of the issue w.r.t., they additionally bomb when the circumstances change. Right now, propose the production of a general and principled framework for SSL (semi-directed learning), the easing of information shortage by smoothing between neighboring clients and POIs, and the treatment of

various settings by regularizing client inclination dependent on setting graphs.

Proposed Methodology

Data collection – Data sets were collected from online reviews about tourist places and/or hotels from related websites. Different algorithms were applied to evaluate the effect of various factors like dataset size, feature size, number of features and types of features on the performance of the proposed methodology framework. Five machine learning algorithms have been used. These are, 1. Naïve Bayes, 2. SVM, 3. Decision Tree, 4. J48, 5. Random Forest. Equal number of positive and negative reviews are collected for model training **Data Preprocessing** – Extraction of sentences takes place followed by tokenization of words and identification of aspects based sentiment words using Natural Language Processing techniques

Architecture

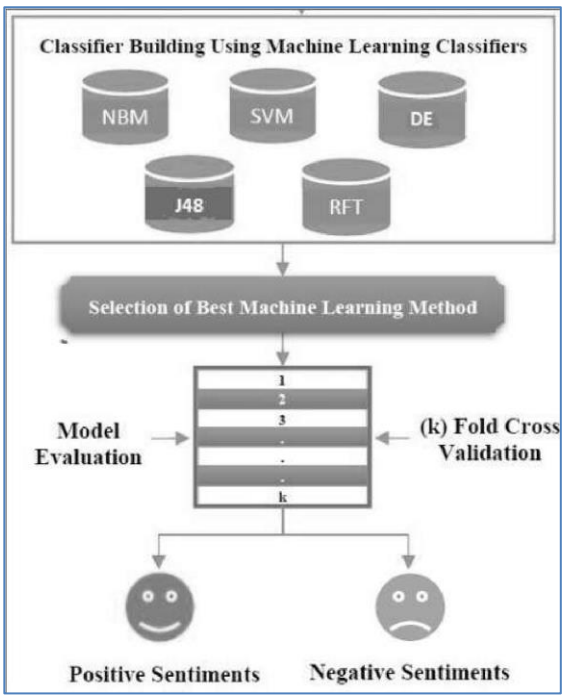

Fig.1 System Architecture

A. Algorithms

1. Naïve Bayes
2. SVM
3. Decision Tree
4. J48
5. Random Forest

Evaluation

In this section the results by all the classifiers are compared by calculating their Precision, Recall and FMeasure values.

Result and Discussions

It has been found that Naïve Bayes algorithms family gives the optimal result in our case considering all factors like accuracy and complexity etc.

A. Figures and Tables

Table.1 Algorithm Accuracy Comparison

SrNo	Algorithm	Accuracy
1	Naive Bayes	65
2	Decision Tree	70
3	J48	60
4	SVM	60
5	Random Forest	55

Conclusions

An interactive and informative user interface is developed to help access and visualize all the analytics results. On a broader canvas, the proposed framework demonstrates he feasibility of recognizing and analyzing different groups of public commuters, such as tourists, business travellers, local citizens, or even foreign workers. We believe that many other insights of practical interest (e.g., the different travel demands and behaviors between tourists and business travellers) can be investigated using the proposed framework and the public transport data. Moreover, this work reveals many unique advantages of transport data over other information sources , typically including a good coverage of population, timeliness of information, and the usefulness of the transportation infrastructures (e.g., subway gantries or bus stops can be potentially used to distribute the analytics results). Tourist will get result automataically by using the system without anyone assistant. Based on classification he/she will get find the place easily with anyone assistant tourist saving money , time.

References

[1] Juanjuan Zhao, Fan Zhang, Lai Tu, Chengzhong Xu, Dayong Shen: Estimation of Passenger Route Choice Pattern Using Smart Card Data for Complex Metro Systems.
[2] Yu Lu, Archan Misra, Wen Sun, and Huayu Wu: Smartphone Sensing Meets Transport Data: A Collaborative Framework for Transportation Service Analytics.
[3] Hongzhi Yin, Weiqing Wang, Hao Wang, Ling Chen and Xiaofang Zhou: Spatial-Aware Hierarchical Collaborative Deep Learning for POI Recommendation
[4] Meng Qu, Hengshu Zhu, Junming Liu : A Cost-Effective Recommender System for Taxi Drivers.
[5] Carl Yang, Lanxiao Bai, Chao Zhang, .an Yuan: Bridging Collaborative Filtering and Semi-Supervised Learning: A
Neural Approach for POI Recommendation [6] K. Cheverst, N. Davies, and et al. Developing a contextaware electronic tourist guide: some issues and experiences. In the SIGCHI Conference on Human Factors in Computing Systems, pages 17{24, 2000.
[7] M. Deshpande and G. Karypis. Item-based top-n recommendation algorithms. ACM Transactions on Information Systems (TOIS), 22(1):143{177, 2004.
[8] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani. An energy-e_cient mobile recommender system. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 899{908. ACM, 2010.
[9] Y. Ge, H. Xiong, Z.-H. Zhou, H. Ozdemir, J. Yu, and K. Lee. Top-eye: Top-k evolving trajectory outlier detection. In CIKM'10.
[10] D. Grosu and A. T. Chronopoulos. Algorithmic mechanism design for load balancing in distributed