Geospatial sentiment analysis using social media hashtags

* **Abstract**

Sentiment analysis, also called opinion mining, is one of the most popular topics in the fields of natural language processing. It refers to a series of procedure dealing with the text data for extracting the sentiment polarity and emotional attitude behind the words. Twitter is a most influent social media service in the world, its service has been described as "the SMS of the Internet"[1]. People has been accustomed to discussing different events, expressing their opinions on Twitter and using location-based services. Emotional analysis on user's tweets based on location information would explore the general feel or impression on different hashtags, it can provide value for various purpose, which is of great significance.

Here we have done the following work on the sentiment analysis using tweets: First, users interact with front-end for typing keywords and adjust the map. Next, Twitter APIs are called to fetch tweets with the geospatial information. Then, tweets will be processed applying the TextBlob library to obtain the sentiment polarity, which is classified into three categories: positive, negative and neutral. Finally, the results will be marked on a map for displaying, making visualized impression for users to easily understand high level emotional attitude.

**Keywords**: Sentiment Analysis, Emotional Analysis, Opinion Mining, Geospatial

* **Introduction**

The number of people who share opinions or attitudes about the products or events on social media, such as Twitter, is increasing dramatically. This text contains an enormous amount of information, which not only can be used by other consumers to make decisions for purchasing a product but also aid merchants understand consumer preferences to subsequently improve products and services. Thus, an increasing number of companies try to apply data mining techniques to help them analyze their customer's comments and reviews on the websites.

Previous research on sentiment analysis has paid more attention to natural language processing algorithms and techniques. And it has achieved great progress, many open source tools are developed based on various programming languages. This report attempts to shift the focus of sentiment analysis from the natural language processing technology itself to other aspects, such as revealing the connection between geographic information and textual data. We try to associate the user's emotional information with the location tags, trying to sort out the pattern between the user's emotional categories and geospatial locations to provide valuable services.

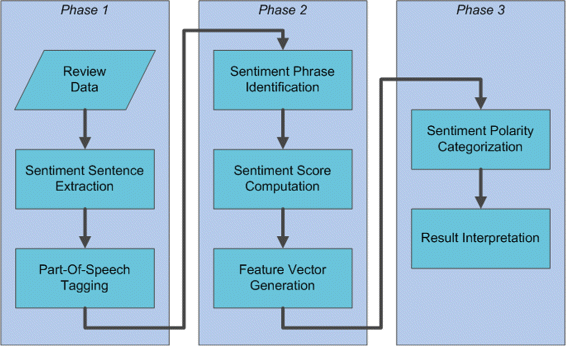
We develop a browser-based application to explore and categorize user’s tweets, displaying generated results on a map. To be more specific, it classifies the public's attitudes or opinions based on their tweets towards a certain keyword displaying the result on google map, so as to help individuals, organizations and governments comprehensively understand the public's opinions.

* **Literature Review**

Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language

Processing (NLP)[]. The goal of sentiment analysis is to extract emotional information from the reviews expressed by the opinion holders through technical methods. Specifically, sentiment analysis mining is to extract the polarity and classification of words, giving sentiment tendency judgment. For example, analyzing the customer’s comments on a product whether is good or bad on Amazon and so on.

In recent years, machine learning algorithms have shown decent performance in many NLP tasks, including sentiment analysis, such as word segmentation and part-of-speech tagging. []



Location-Based Service (LBS) has become more dominant on social media, lots of social media service can obtain user’s location information. The combination of social media and geographic location forms Georeferenced Social Media. Geospatial media covers a variety of social media sites and services with geographical features. After acquiring geospatial media data, model and analyze geographic location and users, mine knowledge, and understand geographic location and users.

In this report, we apply machine learning method to solve the sentiment classification problem, which is implemented by using the Python library named TextBlob. TextBlob is a popular package for handling issues of textual data. It provides a consistent API for diving into common NLP tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, and more[]. The biggest advantages of TextBlob is that it combines two powerful NLP libraries, Natural Language Toolkit and Pattern, and still friendly for programmers.