# Effective Polling System for On-Site Social Media Users with Bias Elimination Using Sentiment Analysis

[1]M.Janani, [2]M.A.Mukunthan,

, [1]PG student [2]Associate Professor Jaya Engineering College, Chennai, India
janani@gmail.com mamukunthan@gmail.com

Abstract — We are implementing opinion polling system, we are gathering people's thought and opinion of any using sentiment analysis. We gather opinion of each individual person based on their location. Data Analysis is achieved by Big Data for analyzing the polling / opinion from the common public for any common issue. Participation of recasting is avoided. Final decision is derived by FEGOR Enhancement.

Ke words – Sentiment Analysis, On-site User-Social Event Di stance, FEGOR Enhancement, Gaussian Process Regression, Parameter Learning.

## 1 INTRODUCTION

The wide spread use of social network services, especially location based services, has transformed social networks into an important information source of real-world events. In the age of social network services, where Twitter[1] changed its prompt from "What are you doing" to "What's happening" in 2009 and Mark Zuckerberg made it clear that he wanted Facebook[2] to serve as the "World's Newspaper", the way we consume information and the types of data we generate have changed significantly. Social network users are more likely to post newsworthy materials such as on-site photos, videos, or other types of data about social events, instead of only text-based status from themselves. Such revolutionary changes have led us into a new world which can make great use of social and community intelligence. Especially with the development of location based services, geo-tagged posts from social networks users offer rich information about our society, and have inspired a line of works aiming at real-world event detection, intelligent location based systems, and other applications, which can be of great importance to capture city dynamics for social goods.

## 2 RELATED WORK

Our work is relevant to two lines of research as follows: Mobility Models for Ad Hoc Network Research and Human Mobility Pattern in Network.

### 2.1 Mobility Models for Ad Hoc Network Research [2]

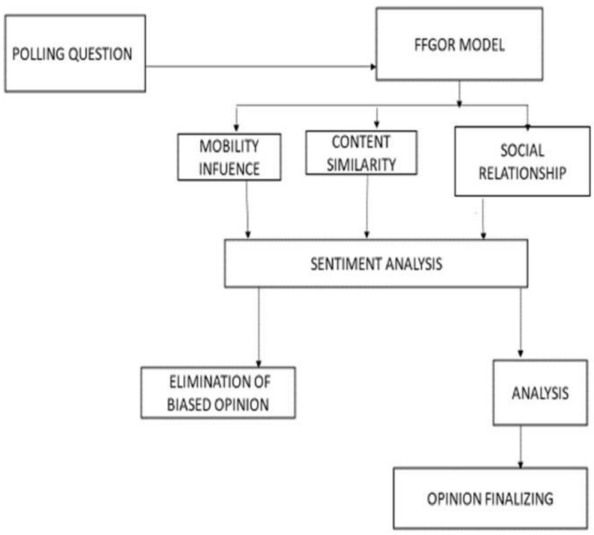
In the performance evaluation of a protocol for an ad hoc network, the protocol should be tested under realistic conditions including, but not limited to, a sensible transmission range, limited buffer space for the storage of messages, representative data traffic models, and realistic movements of the mobile users. Survey of mobility models that are used in the simulations of ad hoc networks. Several mobility models that represent mobile nodes whose movements are independent of each other and several mobility models that represent mobile nodes whose movements are dependent on each other. The goal is to present a number of mobility models in order to offer researchers more informed choices when they are deciding upon a mobility model to use in their performance evaluations. Lastly, presenting simulation results that illustrate the importance of choosing a mobility model in the simulation of an ad hoc network protocol. Specifically, illustrated how the performance results of an ad hoc network protocol drastically change as a result of changing the mobility model simulated.

### 2.2 Human Mobility Pattern in Network [3]

Previous studies demonstrated empirically that human mobility exhibits Lévy flight behaviour. However, knowledge of the mechanisms governing this Lévy flight behaviour remains limited. Analysing over 72 000 people's moving trajectories, obtained from 50 taxicabs during a six-month period in a large street network, and illustrate that the human mobility pattern, or the Lévy flight behaviour, is mainly attributed to the underlying street network. In other words, the goal-directed nature of human movement has little effect on the overall traffic distribution. Simulating the mobility of a large number of random walkers, and find the simulated random walkers can reproduce the same human mobility pattern, and the simulated mobility rate of the random walkers correlates pretty well with the observed human mobility rate.

## 3 SYSTEM OVERVIEW

This proposed framework, which consists of Mobility Influence, Content Similarity, Social Relationship and Sentiment Analysis and opinion poll.




### Mobility Influence

This module describes about the mobility of user. through social media we will gather all the information about the each individual user. Because user information plays a vital role in our system. Here analyzing the mobility of the user and what is their actual location.

### Content Similarity

In this module to check the content similarity then analyzing previous history of user's posts.

Because this implementation is gather the opinion of public people. for that discard the poll which is given by the person who is related to the issue posted by the admin. in the other word, if admin post politics related issues this system will discard the poll posted by politics related people.

**Social Relationship**

In this module to check the user whether they are member in any other group like politics, sports or any other groups. if they are available in any other this system will not consider the poll posted by them.

**Sentiment Analysis and Opinion Poll**

This is the main module it contains all the module which is described above. In this module we analyze the opinion of each individual person based on the issue and as well as location also. If the issue is based on any particular area then take the opinion of that particular place living people. because they only know the actual issue of that place so discard other people opinion.

## 4 INFERENCE ALGORITHM

We propose a Fused fEeature Gaussian prOcess Regression (FEGOR) model to estimate the User-Social Event Distance. And an information entropy based genetic algorithm is proposed for parameter learning.

### 4.1 Gaussian Process Regression Model

Previous studies have shown that human mobility can be modelled by Brownian motion and the jump length of Brownian motions exhibits a Gaussian distribution. Based on this observation, we model the distance between user's trajectory and social event location as a Gaussian distribution, whose mean and variance can be influenced by miscellaneous factors. In this work, we aim to determine some influential factors by extracting user's mobility, content and social relationship features. Hence, we apply Gaussian process regression (GPR) as the inference model for our work since every user's mobility pattern obey Gaussian distribution, and their combination has a multivariate normal distribution. Based on GPR model, we are able to infer an unknown target output $y_*$ conditioned on the known value of $y$, $x$; and its corresponding input $x_*$, where $y_*$ is the distance that need to be estimated, $x_*$ is the integration of three features mentioned above, $y$ and the corresponding $x$ are the existing knowledge which is leaned from the Active Users.

### 4.2 Parameter Learning

We apply Genetic Algorithm (GA) in our work to learn the most suitable parameters. Instead of randomly initializing the first generation of population for searching, we utilize the information entropy of features to build "seed" for possible solutions.

## 5 EXPERIMENTAL RESULTS

### 5.1 Dataset Description

Using the TwitterAPI, we collected raw data of user tweets that were posted at the New York City. To check whether there are significant bias between active and inactive users, we also compare the behavioral patterns in terms of the tweet's creation time distribution and some other statistical metrics as shown. The distribution of tweet's creation time shows a similar pattern between active and inactive users.

### 5.2 Detailed Results

5.2.1 Locating the Social Event

We adopt the algorithm for event localization. The method originally considers “loca l words” to match the locations with content, which c an be adapted for our system by replacing the “local words” with “topic words”. Topic words refer to the hot/key words that are generated by users near a social event.

5.2.2 Feature Evaluation

**Determining participants and historical records**

Given the location of each event, the corresponding participants are defined to be persons who stay within a certain radius from the event location. Without loss of generality, the radius for a social event is usually lesser than 500 meters. However, different event may differ in the area range it covers. Therefore, we need to select an appropriate range of radius to cover most of the situations to model the events’ area.

5.2.3 Inferring the Parameters for Genetic Algorithm

After we obtain the three features calculated from event patterns and event-related users’ patterns, we then infer the model parameters based on the training dataset. As mentioned above, the parameters $\alpha$, $\beta$ and $\gamma$ in equation 17 determine the different influence weights for constructing the input of the FEGOR model.

5.2.4 Feature Weight Analysis and Distance Estimation

**Analyzing Feature Weight**

In order to better understand the reasons/factors that may motivate ordinary people to participate in social events, we illustrate the learned feature weight, values for parameter $\alpha$, $\beta$, and $\gamma$, of the three events. Different events have different parameter values indicating that people attracted to the events are influenced by the three proposed factors differently. As for event 1, we can see that $\beta$ is the largest among the three factors which indicates that people attend that event mainly because of the content preference. And for event 2, apart from content similarity, $\alpha$ also has a relatively high value indicating the importance of user mobility pattern to motivate users’ attendance of the target event.

**Distance Estimation and Model Comparison**

Using the FEGOR model and the parameter values learned for a target social event, we can then predict/estimate the user-social event distance for every event-related individual. Figure 16 compares the distance estimation performance of our FEGOR model and several other regression models, including Simple Linear Regression, RBF Network, Random Forest and traditional Gaussian Regression.

## 6 CONCLUSION

The process of the system is to propose an application to get the opinion from people about any issues. Using bigdata analysing the comment posted by people and discarding repeated reviews from same id. Hence implementing opinion polling system, so gathering people’s thought and opinion of any using sentiment analysis. We gather opinion of each individual person based on their location. Data Analysis is achieved by Big Data for analyzing the polling / opinion from the common public for any common issue. Final decision is derived by ignoring biases polling from people similarity with respect to Content, Social Relationship & Mobility Influence.

**REFERENCES**

1. Zhan, Nag li, and Zhi-hun zhou,fellow (2014)"B-Planner: Planning Bidirectional Night Bus Routes Using Large-Scale Taxi GPS Traces",ieee transactions on intelligent transpotation systems.

2. Tracy Camp, Jeff Boleng, Vanessa Davies," (2002) A Survey of Mobility Models for Ad Hoc Network Research", http://toilers.mines.edu,toilers[pdf].

3. Bin Jiang, Junjun Yin and Sijian Zhao," Characterizing (2019) the Human Mobility Pattern in a Large Street Network ", [pdf] arxiv.org.

4. Daqing Zhang and Bin Guo," (2011) The emergency of social and community intelligence", IEEE Compute.

5. Zhiwen Yu, Hui Wang, Bin Guo, Tao Gu, and Tao Mei," (2015) Supporting Serendipitous Social Interaction Using Human Mobility Prediction", IEEE Transaction and human machine.

6. http://www.journalism.org/2016/05/26/news-useacross-social-media-platforms-2016/

7. Hecht B, Hong L, Suh B, (2011:237-246) et al. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles.

8. Jurgens D (2013). That's What Friends Are For: Inferring Location in Online Social Media Platforms Based on Social Relationships. ICWSM.

9. Yang D, Zhang D, Qu B, et al. (2016) PrivCheck: privacy preserving check-in data publishing for personalized location based services. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM.

10.DuR,YuZ,MeiT,etal(2014). Predicting activity attendance in event-based social networks: Content, context and social influence. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing.