GLOBAL RESEARCH IMMERSION PROGRAM FOR YOUNG SCIENTISTS

Qingyun Go: Analyzing and Predicting Mobile User Emotions Using Location-Based Social Networks

Yuchen Zhang[1]

[1]Fudan University

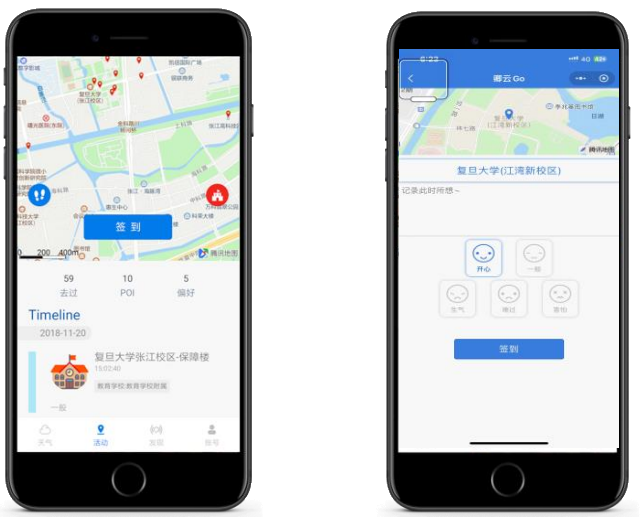
Research objectives

User emotions are key to creating context-aware applications. In this paper, we investigate whether we can predict user emotions based on their smartphone activities. Using Wechat mini-app as our platform, we use Qingyungo, a self-developed mini-app to collect data on user emotions, locations and activities. This app has gethered extensive user data. We analyze this data for statistical insights and correlations between emotions and other features. We also build classifiers to predict user emotions, achieving high accuracy. This paper details the data collection, analysis, and classifier performance.

Background

Mobile devices, especially smartphones, now provide pervasive context-aware services that consider factors like location, user activity, and environmental conditions for customized experiences. Understanding user emotions can significantly enhance these services, such as suggesting music based on mood. Traditionally, detecting emotions required intrusive methods like physiological or body sensing. However, with the widespread use of smartphones, a new question arises: can we detect emotions through smartphone activities?

With 350 million people suffering from depression globally, recognizing and managing negative emotions is crucial to prevent health risks and improve quality of life. Mobile phones, used by over 2 billion people who check them frequently, offer a unique opportunity for real-time emotion tracking and mental health intervention. Studies show their potential for monitoring and managing emotional states, providing a promising tool for mental health support.



Two figures showing how Qingyun Go works

Methods

We gathered data from two main sources: location-based behaviors via the LBSLab platform and demographic information through online questionnaires.

User Activity Data Collection

We utilized QingyunGo, a WeChat mini-program, to collect location-based data from 467 Fudan University students over 11 days. WeChat's
extensive user base and easy sharing features enhanced user engagement. The platform supported check-ins at selected POIs, weather
checks, and exploring nearby locations. User actions, including logins and rank list checks, were recorded with timestamps. Location data was gathered using WeChat's API, and weather data was obtained via the QWeather API.

Demographic Data Collection

Demographic information was collected from the 467 university students through online questionnaires with their consent. This data includes basic characteristics such as gender and grade.

Model Selection

To identify the best classifier for predicting user moods, we experimented with several supervised learning algorithms:

Logistic Regression: Estimates the probability of a categorical dependent variable using the logistic function.

Naive Bayes: A probabilistic classifier assuming independence between predictors, performing well with many features.

Support Vector Machine (SVM): Finds the optimal hyperplane for class separation, using the RBF kernel for non-linear relationships.

Random Forest: An ensemble method using multiple decision trees to enhance predictive accuracy and manage high-dimensional data effectively.

Results

Emotion	Precision	Recall	F1-Score	Support
angry	0.88	0.79	0.83	177
fearful	0.86	0.86	0.86	496
happy	0.91	0.92	0.91	3163
ordinary	0.9	0.91	0.9	3242
sad	0.86	0.82	0.84	743
accuracy			0.9	7821
macro avg	0.88	0.86	0.87	7821
w. avg	0.9	0.9	0.9	7821

After testing out these four methods, we find random forest has the highest accuracy compared to other methods. Overall, the random forest model demonstrates high accuracy and balanced performance across different emotions, with particularly strong results in predicting 'happy' and 'ordinary' emotions. The slight variations in precision and recall for 'angry' and 'sad' categories suggest areas for potential improvement, but the model performs robustly in general.

Discussion

The random forest model demonstrated robust predictive performance with a high overall accuracy, indicating its effectiveness in classifying user moods. The results highlight the potential of supervised learning algorithms in mood prediction based on location and demographic data.

Conclusion

The study successfully employed supervised learning algorithms to predict user moods based on location and demographic data, with SVM showing the best performance. Future work will focus on enhancing prediction accuracy and expanding the dataset to include a more diverse user base.

Key References

Zhang, Yuwei, et al. "A human mobility dataset collected via LBSLab." Data in Brief 46 (2023): 108898.

Acknowledgments

This work is guided by Prof. Yang Chen and Shaobin Wang. Special thanks to them.