A Machine Learning Framework to Map User Classification for Online Map Community

Shaohua wang[1], Yang Zhong[4,*], Ershun Zhong[1,2], Erqi Wang[1,2],Yinna Zhang[2], Wenwen Cai[2]

[1] Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China
[2]SuperMap Software Co. Ltd., Beijing 100015, China
[3]Department of Geography, University of California, Santa Barbara, CA,93117, USA
[4] Claremont Graduate University, Claremont, CA 91711, USA

*Email: andyz1985@hotmail.com

Abstract

The online map communities have been attracting much more attention among students, teachers, researchers, and other volunteers. In there, people could use its functions to create, manage, and customize maps easily. The maps and related information for the map communities have important implications for map and spatial cognition, especially for top map contributors. It is very important to understand the map users' classification so that the different contributors of the map can be identified. In this study, a machine learning framework is designed to user classification for online map community. Gradient Boosted Decision Trees is used as the machine learning algorithm to process the data and generate the outcome. Four major categories of map users are discovered in given data, which are top map contributors(experts), sophisticated users, normal users, and entry level users. They are accounted for 4.46%, 24.58%, 42.39%, and 29.57% of the total users respectively. The Results shows that the machine learning framework is able to provide feedbacks on key aspects of the map user classification for online map community.

Keywords: Geographic Information System (GIS), Machine Learning Framework, Online Map Community, dituhui.

1. Introduction

Over the last few years, online map communities have attracted much more attention of students, teachers, researchers and other volunteers. It has the ability to make maps and customize personal map and data more convenient. The Maps and related information for the map communities have important implications for map and spatial cognition, especially for top map contributors. Map users' classification is so important that can be used for identify different contributors.

Machine learning techniques have been applied in the research and application of GIS (Lu et al. 2012, Chen et al. 2014, Ibrahim et al. 2014, Allen et al. 2016), which also can be used for user classification (Pennacchiotti and Popescu 2011, Pal, 2012, Campbell et al. 2013). In this study, we explore a machine learning framework to user classification for Online map community.

2. Methods

There are four types of information which can be used for map users' classification: profile, maps, messaging behavior and linguistic content. Map user profile information including name, location, age, etc. Maps information is characterized by color, symbol and other elements. Messaging behavior is assessed by the average number of maps per day, number of message, number of replies. Linguistic content information shows the role of classifying users for the context, because a wide variety of linguistic content features can be explored.

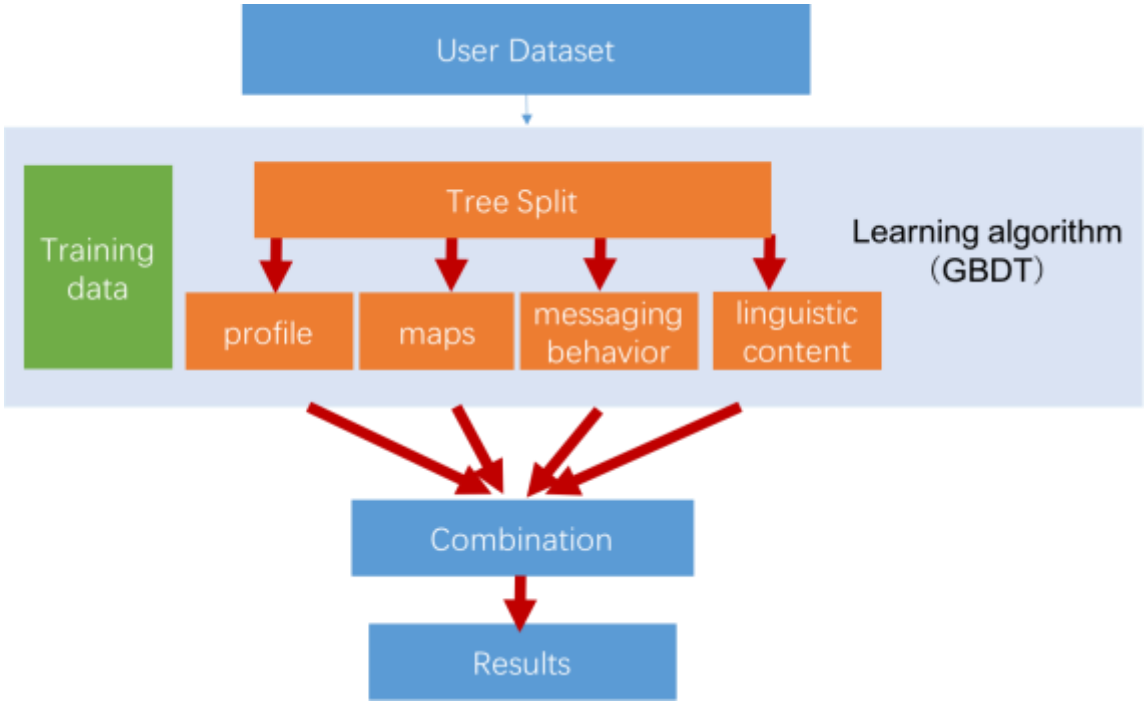

Figure 1. Flowchart of machine learning framework to map user classification.

In this study, we choose Gradient Boosted Decision Trees (Friedman, 2001) as learning algorithm. The flowchart in Figure 1. shows how these steps interact in the machine learning framework.

3. Experimental evaluation

The data of this study came from dituhui(www.dituhui.com). Dituhui is an online map community founded by SuperMap Software Co. Ltd [Wang. 2016]. After analysed over 400 thousand users and more than 800 thousand maps, we find the top map contributors(experts), sophisticated users, normal users and entry level users are the four major categories of map users, holding 4.46 per cent, 24.58 per cent, 42.39 and 29.57 per cent, respectively.

4. Conclusion and Future Work

In this research, the machine learning framework was able to provide feedback of key aspects of the map user classification for online map community, making the methodology suitable for personal user classification. Further work in this area will implement this methodology in customized spatial cognition analytical software with a simple to use graphical interface. In addition, we will use parallel framework to improve performance, the computation efficiency in the framework will be evaluated in future work.

Acknowledgements

This research has been partially supported by the project of Beijing Excellent Talents (201500002685XG242), National Postdoctoral International Exchange Program (Grant No. 20150081).

References

Lu, Y., Zhu, S., & Zhang, L. (2012, May). A Machine Learning Approach to Trip Purpose Imputation in GPS-Based Travel Surveys. In 4th Conference on Innovations in Travel Modeling, Tampa, Fla.

Chen, A. Y., & Lu, T. Y. (2014). A GIS-based Demand Forecast using Machine Learning for Emergency Medical Services. In Computing in Civil and Building Engineering (2014) (pp. 1634-1641).

Ibrahim, A. M., & Bennett, B. (2014, April). Point-based model for predicting mineral deposit using GIS and machine learning. In Proceedings of the 2014 First International Conference on Systems Informatics, Modelling and Simulation, SIMS (Vol. 14, pp. 83-88).

Allen, C., Tsou, M. H., Aslam, A., Nagel, A., & Gawron, J. M. (2016). Applying GIS and Machine Learning Methods to Twitter Data for Multiscale Surveillance of Influenza. PloS one, 11(7), e0157734.

Pennacchiotti, M., & Popescu, A. M. (2011). A Machine Learning Approach to Twitter User Classification. ICWSM, 11(1), 281-288.

Pal, A. (2012). User classification in online communities (Doctoral dissertation, University of Minnesota).

Campbell, W. M., Baseman, E., & Greenfield, K. (2013). Content+ context networks for user classification in twitter. In Frontiers of Network Analysis: Methods, Models, and Applications Workshop at Neural Information Processing Systems.

Friedman, J. H. 2001. Greedy function approximation: A gradient boosting machine. Annals of Statistics 29(5):1189–1232.

Wang S.H. (2016) Innovations of SuperMap GIS platform software: Introduction to the Dituhui of SuperMap GIS. Journal of Geo-information Science, 18(4): 575-576.