Trajectory Understanding and Spatial-temporal Profile Learning Based on Mobile Crowdsourced Data

Chih-Yuan Chen, Hao-Chun Hsu, You-Ming Xu, Pei-Yu Luh, You-Ru Lin, Chih-Chi Kao
Department of Geography, Chinese Culture University

MOST 科技部 Ministry of Science and Technology 勝義科技股份有限公司 HyXen Technology Co., Ltd.
Project Name: Crowd AI: Cooperative Positioning, Geographical Knowledge and Edge Profile Learning, and Cognitive Interaction. 108-2634-F-155 -001-

Abstract

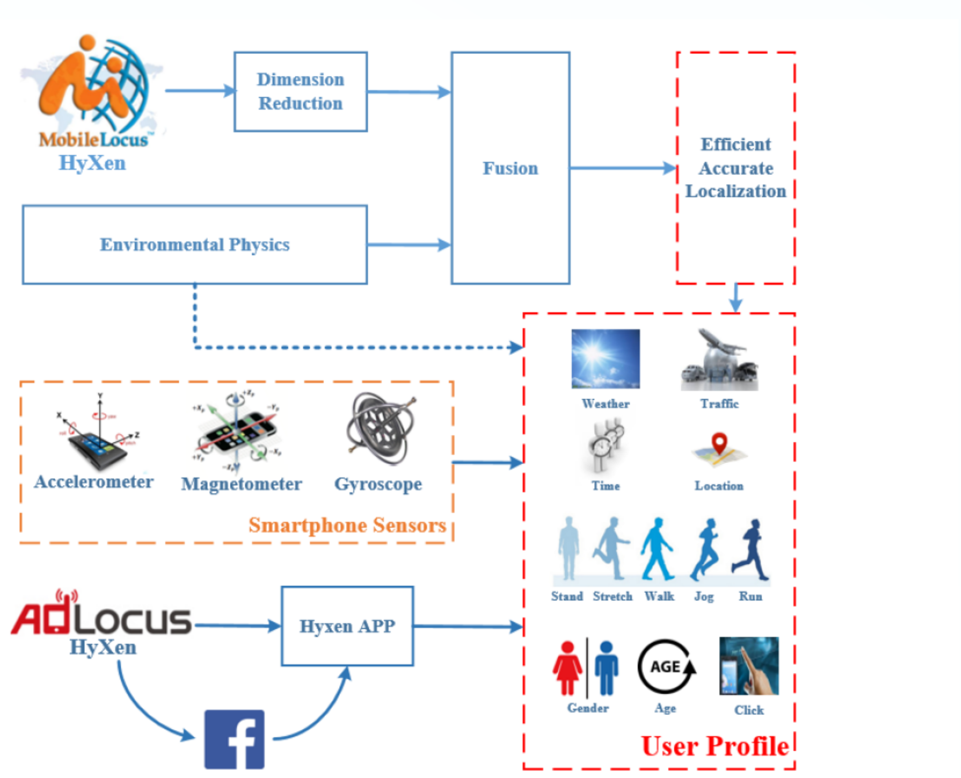
Crowdsoured data from mobile devices such as smartphones and internet of things gadgets can provide meaningful information about users' spatio-temporal context. By analyzing their locations and trajectories, we are able to extract meaningful information and hidden patterns from data thereafter to provide users better context-aware services. In this project, we devided spatial-temporal profiling of mobile crowdsoured data into two categories, fast dynamics and slow dynamics. The fast-dynamic profile includes the user trajectories, transportation patterns, and the other environmental factors such as nearest POIs while the slow-dynamic profile contains the user age, sex, interests, and the other attributes based on predicted and pre-collected data. The "Divide and Group" strategy is the main idea for our research to find the knowledge for the fast-dynamic profile and slow-dynamic profile. We used this strategy on real smartphones' sensory data and concluded several spatio-temporal contexts of designated user groups.

Introduction

The explosion in Big Data has lied a singular opportunity to explore and depict human dynamics in urban environments. Fed by mobile users' billions of inputs of their locations as well as trajectories, there are many potential applications of geospatial Artificial Intelligence (GeoAI) research can help both governments and also private company. While novel AI or machine learning techniques are flourishing in recent years, how to discover useful geographic information hidden in multidimensional geographic and temporal data are still evolving fields. In this project, we cooperate with an industry partner to develop a environment for mining spatial knowledge such as spatial relations, patterns, and characteristics in order to build up a GeoAI-ready environment for processing and analyzing data.

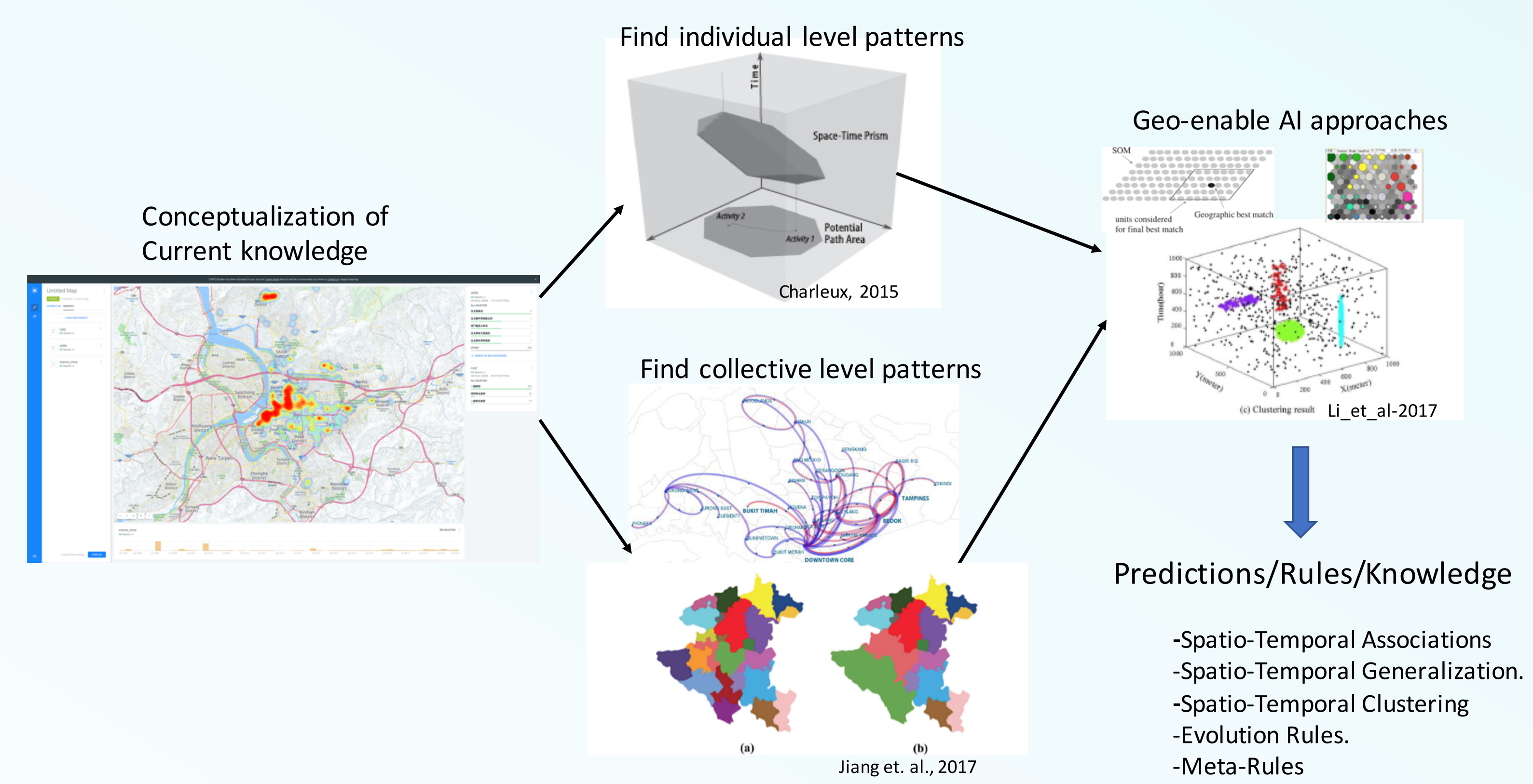
Materials

In this project, we have one major industry partner: HyXen Technology (勝義科技). HyXen Technology is a leading information technology company in location-based services (LBSs) in Taiwan. HyXen Technology adopts self-owned positioning technology in various location-based applications, and has accumulated 3 billion location-based data entries from their users. The crowd AI technology developed in this project will be used to enhance HyXen's personalized recommender system for automatic delivery of advertisement or coupons, and to enable stores or shopping centers to perform group recommendation and long-term behavioral influence. We focus on studying how to incorporate these modules to provide customized services for different personal and group scenarios. For our industry partners HyXen Technology, this project shall utilize the vast amount of user location information as well as additional contextual information to produce personalized advertisement (or recommendation) that are closely coupled with users' physical environment and behavior.

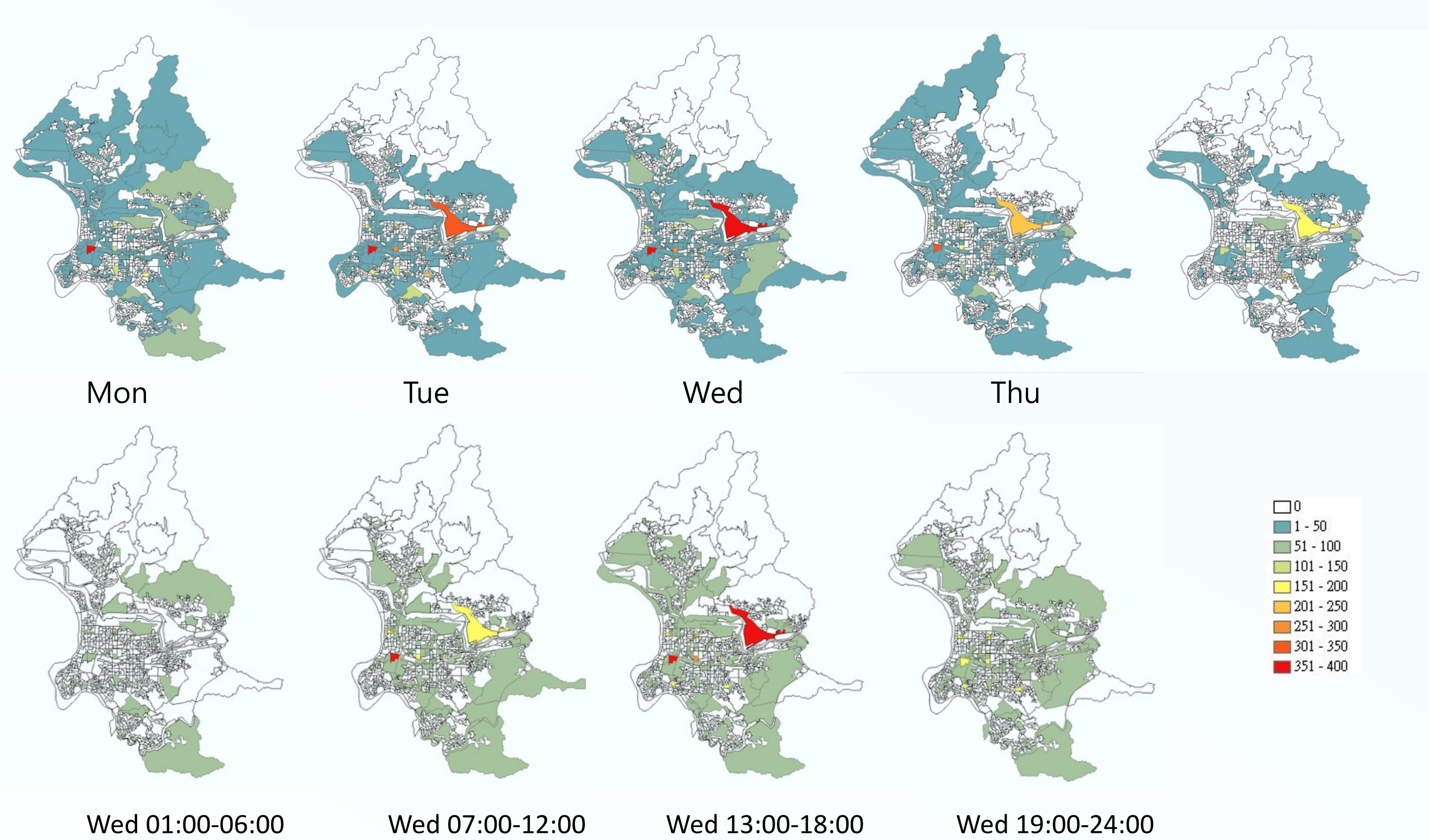


Utilize the vast amount of user location information from our industry partner

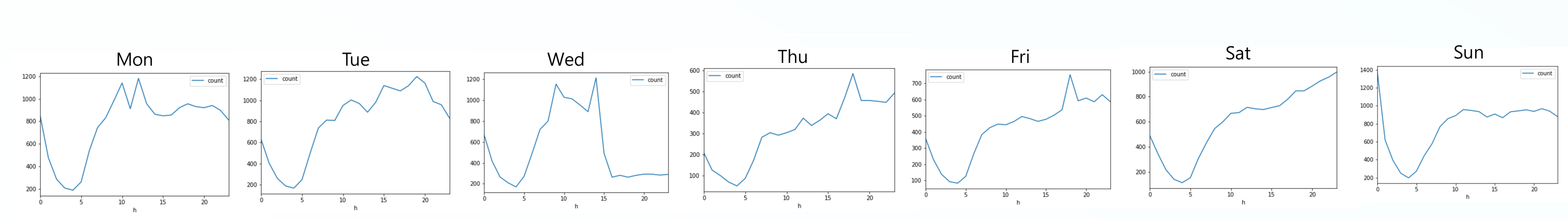
Methodology



A Geo-Visualization environment to support geographic knowledge discovery



An example of collective level geographic knowledge discovery

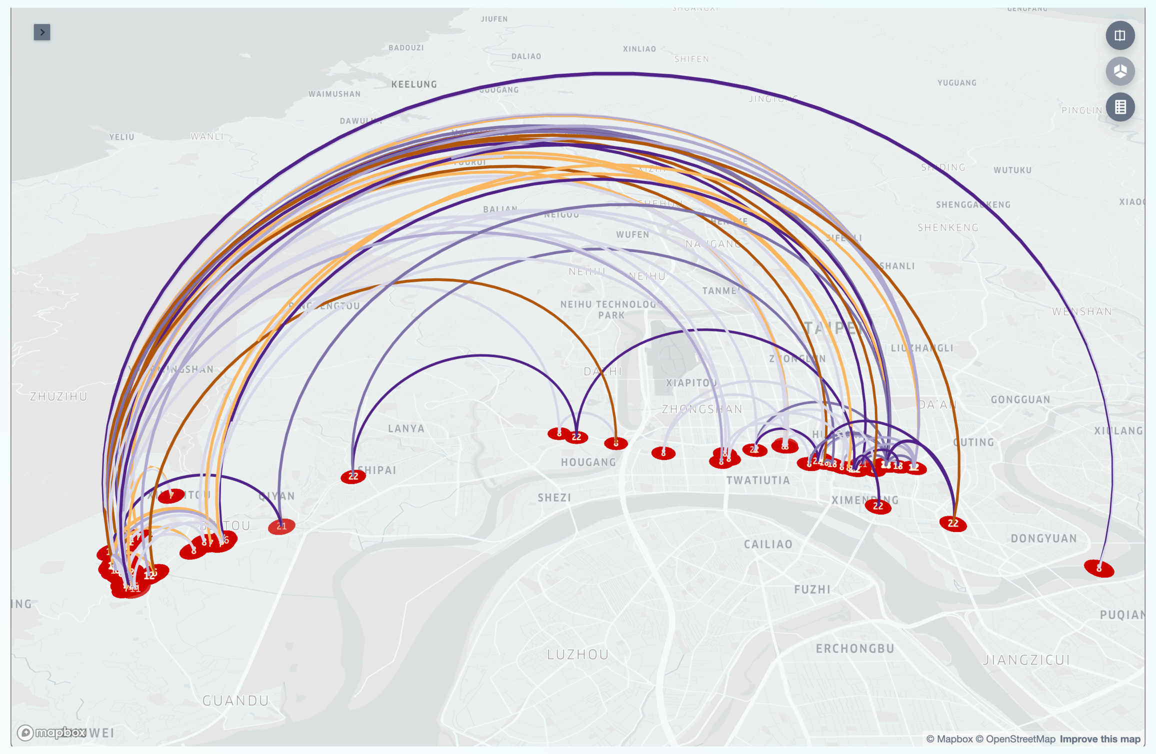


An example of collective level time profiling

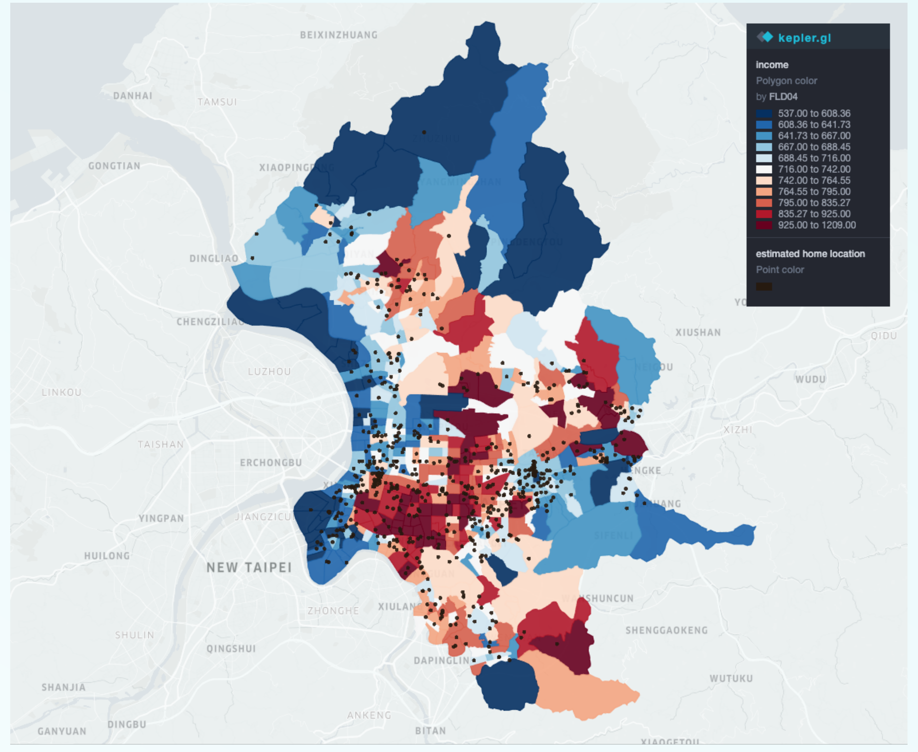
The location-based data from HyXen Technology contains volunteered location and time information. The data size is huge and the spatial and temporal intervals of individual user could be sparse or dense, which makes the user profiling very difficult to conduct. In order to perform fast dynamics and slow dynamics profiling, there are two main tasks to be accomplished. First, to develop geographical knowledge extraction algorithms and information visualization methods to discover and represent the physical meaning of the mobility. Second, use those geographical knowledge acquired from crowdsoured data to label location and trajectory data for crowd/edge profiling. Among the different strategies to perform GKD, the "Divide and Group" is used here. The strategy is often used when data is big and highly complex and can be characterized as a combination of several homogeneous parts. The homogeneous part mentioned here may also refer to spatial patterns in geo-spatial space. Geographical visualization, or geo-visualization, is one of the most important tools in GKD which help researchers to explore spatial knowledge via interaction with map and graph. It has been widely used for exploring geo-spatial data and proven useful in finding spatial relations or patterns. In this project we will utilize the power of geo-visualization techniques to discover particular data features and construct location-based knowledge for fast-dynamic profile and slow-dynamic profile. We started from using the current knowledge to divide the data into collective and individual levels. Each level can be used for exploring useful spatial-temporal attributes to label location and trajectory data.

For collective level, we use the 2nd Dissemination Area as our spatial unit for calculating and visualizing the density of data. Each 2nd DA has an approximate population of 3000 persons which differs from administrative division (e.g. township or county), which is ideal units for summarizing and visualizing the point-base data. By using 2nd DA, we can get a baseline understanding of human density of our study area. To divide the data by different day of week also reveal the dynamics of vitality of the users.

For individual level, we analyze user trajectories, transportation patterns, and the other environmental factors. While our industry partner provides the user age, gender, interests, and the other attributes based on their pre-collected data, those attributes are often incomplete because only a very small amount of users are will to provide their information. To cope with this limitation, we analyze the trajectory of users to portrait the fast dynamics and also use government's open data such as income map to label users' slow dynamic profile. If user's data is not sufficient or in good quality, the collective level attributes can act as a supplement to give the best estimation.



Examples of individual level geographic knowledge discovery of user



User estimated home locations overlapped with income data

Next steps

1. Collect more POIs and open data from multiple sources (Google, Foursquare, Facebook, IG, government)
2. Label users' locations and trajectories and discover spatio-temporal associations, clusters, or rules
3. Group related labels to form target user types (e.g. car owners, has children)
4. Collect ground truth data for selected target user types to build up Geo-enable AI analytical approaches.