Research Statement

Yi Yu

March 1, 2015
yiyu@nii.ac.jp

1. Background

Mobile technologies have changed the way we communicate with People and world. Many companies have started to provide mobile services via Smartphones. For example, when you want to find places to purchase shoes, monitor children safety, find a place to have lunch, all of them can be done through mobile applications on smartphones. A smartphone is not only a portable computer in our hands. It also serves as a remote dashboard for our life. The widespread of smartphones leads to smart mobile services. Here, the term smart in smart mobile services means understanding a user, and understanding his physical and mental state. Smart services will process user information in the background to make accurate predictions of user intention in real-time and will offer suggestions. By all kinds of smart devices, everyday people create and consume huge amounts of multimedia data. As a result, we are experiencing an era with a rapid increase of data relevant to different aspects of people daily life.

![Diagram of people activities analytics from individual level to society level.](image)

Figure 1: Concept of people activities analytics from individual level to society level.

Fig. 1 shows a concept of people activities analytics from individual level to society level. On the one hand, such data contains personal information of each individual user. On the other hand, more multimedia data and context information are aggregated over time as more users are involved. In this sense, users themselves act as sensors on the movement and form a participatory sensing to gain insights of a society. These data could not only be very beneficial for studying various lifestyle patterns, but also be used to generate more descriptive and explanatory analysis for revealing what relationships hidden in social multimedia data over Internet and user physical activities. We are aware that people activities analytics in the context of social online presences and real physical behaviors in multimedia landscape is still unexplored research area. Our research tries to explore
interesting things of people world and focuses on mining valuable knowledge about user activities from the individual level to the society level across social multimedia landscape to provide mobile smarter services.

2. Distinctive Characteristics of Our Research

Our research focuses on people activities analytics based on the effective collection, analysis and use of multimedia data related to the user-centric physical world and online behaviours and works on data analytics between mobile users and their online behaviours in the multimedia domain, which involves all relevant aspects of people's lives in the physical world, their online presence analytics, and a cohesive set of social multimedia data.

The research distinguishes itself from the existing researches through the following research focus: (i) discovering people-relevant co-occurrences and associations across the Internet multimedia landscape, (ii) extracting mobile user intentions, physical context, and values of local multimedia topics and events, modeling the user-centric physical world and online activities, (iii) gaining insights and understanding of voluminous and diverse user behaviours, (iv) personalizing decisions and services, and (v) building a multimedia ecosystem for Internet-scale multimedia infrastructure.

3. Series of Ongoing Identified Works

The proposed project has an emphasis on how to manage and derive value from multimedia data in the social Internet landscape to facilitate the connections between users' physical world and their online activities. We aim to develop some fundamental algorithms and system to model and predict all aspects of people everyday life. We now have some ongoing novel works as introduced below:

**Personalized E-learning Recommendation**: Large amounts of social media data shared in various online platforms, e.g., Coursera, Videolectures, have fostered an emerging research topic of how to create multimedia-enabled mobile learning and teaching to effectively improve the utilization of online educational resources. The aim of this work is to develop a personalized e-learning recommendation system which can be a plug-in compatible to smartphones to ubiquitously recommend learners appropriate online materials. In particular, we would like to build an education hub to collect all educational resources over the Internet and categorize all obtained educational resources according to, e.g., learners ages or objectives. We classify users initial inputs or query histories and then build a prediction model to identify users personal learning preferences hidden in users initial inputs or query histories.

**Personalized Video Soundtrack Recommendation**: It is very interesting to extract activity data from a user-centric point of view. Exploiting such data could be very beneficial to individual users, especially for preference-aware multimedia recommendations. We categorize user activity logs from different data sources by using semantic concepts. In this way, correlating between preference-aware activities based on categorization of user-generated heterogeneous data complements recommendations of personal multimedia events. This work tries to correlate preference-aware activities from different behavioral signals of individual users, and mainly studied how to recommend soundtracks for user generated outdoor videos. You can imagine the following scenario: a mum brings her son for outdoor activities. She shoots a video when the little boy plays on the beach and swims in the sea. Later they want to add music of their own style to this video to make this video more appealing. As video and audio have quite different low level features, in this work, they are linked via high level semantic concept—moods.
**Personalized Music Discovery:** Personalized music discovery is very important topic in the field of music information retrieval. We plan to develop predictive models for personalized music discovery in mobile: By leveraging users’ listening histories from social music networks, we can predict better music queries and results to improve users’ experiences.

**Automatic Storytelling of Outdoor User-Generated videos:** Location sharing, as an important functionality of mobile services, can be exploited to collect more descriptive and explanatory descriptions from various media data streams over the Internet to explain what is happening in a physical venue. How to gain insights from such location-aware heterogeneous data to provide a smarter service has become very interesting. We intend to develop an algorithm that is able to automatically predict a meaningful description e.g., subtitle, speech, to make a user-generated outdoor video more appealing. Purpose of this work is to predict a meaningful description for a user-generated outdoor video in a real time to make it more meaningful by exploiting knowledge from social networks. There are some related works involving video categorization and annotation based on text information such as video title, description, user tags and viewers comments.

**Participatory Sensing of Venue via Multimedia Events:** People like to use their mobile phones to send to servers venue-related photos, videos, and comments. All these multimedia events can be treated as the sensor data of the venues. Foursquare and Instagram are typical examples of location-based media sharing platforms that also serve as venue-sensing systems. People check in at a venue online using their mobile devices and send information about a venue to media service platforms. For example, the event Metropolitan Museum of Art is launched in Foursquare. Many people post photos and tips there describing this venue. Such multimedia contents, reflecting different aspects of a venue, provide a means of participatory sensing. From these multimedia contents, new users can get a rough image of the venue. Foursquare also tells us how many people have visited this venue. Vast volumes of check-ins, venue photos, and venue comments are aggregated over time. These data are potentially valuable knowledge source about the user physical activities and online sharing behaviors in the era of big data. Such heterogeneous data source of people activities can be used in location-aware recommendation (e.g., provide media advertisement and travel plan) and urban computing (e.g., provide potential sustainability and outlook of urban environment, people life quality, city planning, and social sciences).

### 4. Miscellaneous Topics Working with Collaborators

Some interesting topics that I am involving in:

**Person Re-identification:** We focus on the problem of person of interest re-identification that visually identifies an individual observed by the different camera views in a distributed surveillance environment.

**Music Discovery Based on Psychoacoustic Features:** Using low-level features, such as spectral properties and inter-channel phase/level differences and correlations, to correlate with subjective perception of "spaciousness"/"source width"/"envelopment"/"intimacy". These terms are often used and investigated a lot in the field of concert hall acoustics.

### 5. Other Interesting Topics

I did work on them.
Personalized Geo-Fencing: A geo-fence is a virtual perimeter for a real-world confined geographic area. In the era of big-data, geo-fencing has to pair millions of points and hundreds of polygons or even more in real-time, and its scalability is becoming a very important task. The basic idea behind geo-fencing is very intuitive: when users enter or exit the boundaries of areas based on geo-fencing-enabled location preferences, various notifications are sent out to users. The first personalized geo-fencing application is for social networks. Here, each user defines a geo-fence for an area around himself, for example, a circle defines a school. Then, when the user enters or leaves this geo-fence, a context aware processing is triggered. A notification is sent to his friend for enhancing their social connections. The second geo-fencing application is to keep track of the children. Here we assume that the children have smartphones. When they enter or leave a designated area, parents can use some mobile application to track down their children based on receiving geo-fence-enabled notifications. The third geo-fencing application is to see what is happening at your property in real time, get instant awareness with smart clip capture and never miss a moment with continuous recording. It is easy to check whether your kids or your pets go out of your house, and whether your valuables are stolen out of your house.

Web-Based Map Personalization: People daily life can be visualized in personal map. We hope connections between multimedia data and map can be enhanced through map personalization. A web-based map is usually regarded as a tool that can describe a part of the physical world for the user. There are more and more information and data aggregated over physical world which can be layered on the maps. So, in order to improve the map readability, it becomes very important to simplify and personalize maps for the specific users according to map scale and purpose. Map Customization might allow you to create a suggested route for places and add photos or videos to your personal map, to help categorize the locations of your favorite venues. For example, after you make a personalized Google map, you can publish it for all to see, or just keep it for your own private use. Map simplification can be used in some personal geographic information system and business mapping applications for creating personalized maps with as few vertices as possible. The process of map simplification is to reduce the amount of data that is required to produce maps without losing the general shape of the map. It can simplify and edit map data for display at a particular scale.

Large-Scale Music Data Analytics: Large scale music data processing and analyzing is interdisciplinary science. In my work, approaches for solving this task are related to music signal analysis, audio content indexing, statistical analysis, semantic learning, pattern recognition and mining, music attribute extraction, and music retrieval evaluation, etc. Generally, I have proposed some interesting algorithms for scalable music data processing, analyzing, storing and searching, in my previous works, which are related to two essential research problems: (i) Extracting music characteristic or attribute based on acoustic statistics or music knowledge learning. (ii) Organizing music acoustic or semantic features in the database using an indexable format. These two aspects need joint design so as to improve both efficiency and accuracy of large-scale music audio data management and retrieval.

Pest classification: Pest classification can effectively help save millions of human live loss through malaria interventions, and help grow more food with less pollution, less energy and less environmental damage by crop pest interventions. This project is related to agriculture at the center of social media innovations for communities in National University of Singapore. I am interested in insect sound recognition and participated in the UCR Insect Classification Contest. In the suggested algorithm, I concatenated typical spectral features, STFT, Mel-spectrum and MFCC with proper weights. I used SVM with support vector machines to project the feature sequence into a new feature. Each dimension of the new feature corresponds to one class and includes the likelihood that a pest sound belongs to this class. This algorithm was evaluated by UCR (http://www.cs.ucr.edu/~eamonn/CE/CE_CONTEST_unofficial_results.pdf) and achieved the fourth rank.