The role of an IoT Platform in the Design of Real-time Recommender Systems

Sangwhan Cha, Marta Padilla Ruiz, Monica Wachowicz, Loc Hoang Tran, Hung Cao, and Ikechukwu Maduako
People in Motion Lab, Department of Geodesy and Geomatics Engineering
University of New Brunswick
Fredericton, Canada
{scha, mpadilla, monicaw, ltran, hcao3, imaduako} @unb.ca

Abstract—The growth of Internet of Things (IoT) brings the promise of a wide range of new recommender systems due to the expected 57 billion smart connected devices by 2025. In this paper, we propose a new IoT platform for supporting a real-time recommender system. To illustrate the effectiveness of our proposed IoT platform, we present a prototype implementation and a tourism application to demonstrate the entire process from user event data collection to notification/recommendations provision. We conducted several experiments including notification and system performance tests to illustrate the use and performance of our real-time recommender system.

Keywords—IoT platform; recommender system; tourism; cloud computing.

I. INTRODUCTION

The Internet of Things (IoT) has been largely recognized as a global network interconnecting humans with RFIDs, sensors, actuators, smartphones, computers, buildings, home/work appliances, cars and any other device with the goal of unlocking a new combination of applications and services in the near future[1,2]. From a conceptual point of view, IoT is about devices acting as providers and consumers of data related to a specific user context. From a networking point of view, an IoT is a system architecture that supports point-to-point communications, preferably in real-time.

Due to the large device heterogeneity, IoT is also about exchanging and analyzing massive amounts of data and generating information users need at the right time and on the move [3]. In general, IoT applications require to be adaptable to a highly diversity of contexts, by responding in an intelligent way to the presence of users, as well as to their location, time of the day, and tasks at hand. This requires the support of IoT platforms for generating context intelligence that will discover new user-centric interests and provide personalized information associated with them. While devices would be able to support limited and lightweight services, one key aspect of IoT is the need for a cloud architecture which can leverage big data streaming from millions of devices and at the same time, generate mobility patterns, and provide value-added services to end users according to their specific context.

In [4], Jiang et al. propose a Hadoop system as suitable data storage for IoT devices since they generate data very rapidly. They argue that any cloud architecture would be able to process massive structured and unstructured data efficiently. Tyagi et al. also propose a cloud computing infrastructure for IoT data [5]. They assert that a Hadoop & NoSql are the dominant massive data technologies in use today.

Many general purpose IoT platforms with specific technologies such as Representational State Transfer (REST) [6], Software Defined Radio (SDR) [7], and plug and play module connection [8] have been previously developed. Additionally, there are many specialized IoT platforms such as health-IoT platform [9, 10] and IoT-lab experiment platform [11]. However, no research has been found in the literature that deploys IoT platforms for real-time recommender systems, and in particular for tourism applications. Most of the solutions for these systems have been focused on planning an itinerary prior to traveling and supporting a search based on keywords and preferences of places to visit [12-15].

This paper aims to describe the design of an IoT platform for a real-time recommender system that enables streaming data collection from smartphones in order to recommend new items on the fly using geofencing as a user context. Geofencing is a virtual circle defined by a centre point and a radius. The geofencing can be generated by using two approaches: (a) without IoT devices: using the Google Location API for defining a-priori coordinates of a point of interest and its radius; or (b) with IoT devices: using the physical position of IoT devices and their signal range as the radius. In this research we have applied both approaches. The first approach is used to identify regions of interest around Points of Interest (POI) located outdoors. When users are carrying a smartphone with a GPS sensor, the system detects when their current geographical coordinates are within the geofence boundaries. The second approach is applied for regions of interest around POIs, but in this case, they are located indoors. The Bluetooth technology available in the users’ smartphones can be used to detect the signal of IoT devices such as beacons, allowing the system to know whether or not their current locations are within the geofence boundaries.

One of the main challenges when developing an IoT platform for a real-time recommender system is choosing the cloud architecture that would allow scalability, resiliency to failure, and fast processing of massive data sets according to a user context such as a user current location in relation to other users, where users are heading to, staying at and leaving from a geofence boundary as well as the means of transportation, weather conditions, and time of the day. Towards this challenge, we have chosen the cloud-based Apache Hadoop cluster for our proposed IoT platform.
The remaining of this paper is structured as follows. Section 2 presents related work. Section 3 describes the proposed system architecture. The experiments for evaluating the proposed real-time recommender system are described in Section 4. Section 5 provides the concluding remarks.

II. RELATED WORK

IoT has been recognized as a revolution in ICT during the past few years and the number of IoT developments such as mobile applications, wireless networking protocols and IoT platforms has grown in an unprecedented rate [16]. As part of an IoT platform, many devices such as smartphones, RFID tags, sensors, and actuators have been transformed to “connected things” and various applications such as smart homes, environmental monitoring, health care and smart cities have emerged in the market [9,17]. This requires IoT platforms to be responsive in nature, anticipate users’ needs according to different contexts they are in by means of intelligent components, devices, and applications.

It is strategic to design a specialized IoT platform that supports proper data collection, data interpretation/analysis, and data visualization. Furthermore, it is important that this IoT platform is able to provide personalized recommendations in real-time. At this point, there is a lack of research on designing specialized IoT platforms for providing real-time recommendations when users are moving around visiting a tourist destination. In its most common design, the recommender systems are reduced to the problem of estimating ratings for the items that have not been seen by a user [18].

One of the main disadvantages of previous approaches is that of only being capable of recommending items that score highly against a user’s profile. As a result, the users are limited to being recommended items that are similar to those already rated. Our research proposes a user context-based recommendations such as a user current location in relation to other users, where users are heading to, staying at and leaving from a geofence boundary as well as the means of transportation, weather conditions, and time of the day. For a tourism application, our approach will improve how tourists explore tourist attractions.

III. THE SYSTEM ARCHITECTURE

As mentioned in [19], there are two perspectives in the vision of IoT. They are “Internet” centric or “Thing” centric. In this research, we have developed a system architecture based on the “Thing” centric perspective, mainly because geo-fencing has the potential of modelling a user context in space and time since it can combine awareness of a user’s current location and proximity to a particular POI as well as of a group of users and POIs. We are making use of geofences to represent virtual circles around Points of Interest (POI) such as sites, monuments, paintings, and museums, both indoors and outdoors. Indoors POI are represented using geofences with IoT devices (e.g. beacons), which work using Bluetooth technology. Outdoors POI are represented using geofences without IoT devices, which work using Google Location API and smartphones with GPS sensors.

The overall architecture it represented in Figure 1. It consists of a mobile application for Android smartphones, a Notification Server and a Recommendation System. This architecture supports four different user events: (E1) enter geofence with IoT devices, (E2) exit geofence with IoT devices, (E3) enter geofence without IoT devices, (E4) exit geofence without IoT devices. Each of these user events is related to a task that the system run when events occur: (T1) Get Recommendations and notifications from the notification server (http request), (T2) Post user event E1 and E2 data to the cloud, (T3) Post user event E3 and E4 data to the cloud.

There are two additional task, (T4) which is running in the background in order to collect GPS sensing data from the smartphone, posting the data to the cloud; and (T5), which creates an interface in the mobile application to show the notifications received. This interface can be a simple Android notification or a Card created by Google Cards library.

A. Mobile Application

There are ten components in the architecture of our mobile application (Figure 2): one related to location data, three related to geofences with IoT devices (beacons), two related to geofences without IoT devices and two to handle the recommendations.

There are two additional task, (T4) which is running in the background in order to collect GPS sensing data from the smartphone, posting the data to the cloud; and (T5), which creates an interface in the mobile application to show the notifications received. This interface can be a simple Android notification or a Card created by Google Cards library.

A. Mobile Application

There are ten components in the architecture of our mobile application (Figure 2): one related to location data, three related to geofences with IoT devices (beacons), two related to geofences without IoT devices and two to handle the recommendations.

The Location Data Handler (LDH) has two functionalities, which are collecting GPS sensing data from the smartphone each ten seconds and conducting T4, transmitting these data to the FluentD of our cloud system.

The Beacon Signal Monitor (BSM) is monitoring the geofences generated by beacons (without IoT devices). Each geofence is monitored creating a region, which is identified by three numbers:

- **UUID**: a string, e.g. “B9407F30-F5F8-466E-AFF9-25556B57FE6D”.
- **Major number**: an unsigned short integer, i.e. an integer ranging from 1 to 65535.
• Minor number: an unsigned short integer, like the major number.

These numbers allow to monitor beacons with the same UUID, and major or adding the minor number too. Therefore, it is possible to define regions containing more than one beacon in order to increase the signal range, and as a result, the radius of geofencing. However, this approach was not considered in the proposed architecture and we have only used the UUID+major+minor to create one region for each beacon, monitoring them individually.

The Beacon Signal Distiguisher (BSD) is responsible for distinguishing the user events enter a geofence (E1) and exit a geofence (E2). It detects all the geofences the user is entering to or exiting from, retrieving them as an array of beacons. The Beacon Data Transmitter (BDT) transmits user event data to the FluentD of our cloud system, conducting T2.

The Geofence Handler (GH) is responsible for monitoring the geofences generated by the Google API and detecting the user’s events enter a geofence (E3) and exit from a geofence (E4). The Geofence Data Transmitter (GDT) transmits these data to the FluentD of our cloud system, conducting T3.

The Recommendation Receiver (RR) receives recommendations associated with user’s specific events generated by the recommender system and pre-defined notifications, conducting T1. The Recommendation Creator (RC) conducts T5, creating a user interface based only on text recommendations if it is an Android notification, or adding also a multimedia associated, if it is a Google Card.

B. Notification Server

As illustrated in Figure 3, there are five components in the architecture of the notification server. The Notification Receiver (NR) receives initial recommendations from the RT via HTTP Post method. There is a pre-defined text notification associated to each recommendation about a POI. The Notification Modifier (NM) produces proper recommendations by combining the predefined text based notifications and recommendations. The Notification Store (NS) stores modified recommendations into the database, in this case, the MySQL database. The Text Notification Provider (TNP) and Image Notification Provider (INP) provides text and image notifications to the mobile applications.

C. Recommender System

Once the data have been stored in the HBase for a certain period time, we can explore mobility patterns to provide personalized recommendations to users in real-time. As illustrated in Figure 4, there are four components of our proposed real-time recommender system.

The Data Processor (DP) extracts data from the HBase using a time window, which is represented as a sequence of visiting POIs geographical coordinates that a user is visiting and the duration time that a user is spending at each POI. The Mobility Pattern Analyzer (MPA) runs a spectral clustering algorithm to partition users into groups based on their mobility patterns. Consider we are given a set of POIs. Then the mobility of users can be represented by an un-directed graph. The set of POIs is the set of vertices of the graph. Each vertex (i.e. POI) of the graph is connected to all of its adjacent vertices by edges. The Recommendation Generator (RG) generates recommendations based on collaborative filtering for each cluster of users. In this research, we propose the frequency of a user visiting each POI and the duration of time spent in each POI as a rating.

IV. EXPERIMENT DESIGN AND RESULTS

In the recent tourism literature there is an increasing emphasis on the quality of tourist’s experiences when exploring a tourism destination. From a marketing perspective, it is well known that a key component is to offer a unique and differentiated service that can lead to memorable experiences that add value for visitors. Our proposed real-time recommendation system is aimed to support new services in the near future that will create new experience concepts through the use of user contexts that intensify engagement of a tourist with his/her surroundings and emotional connections with POIs. We have selected for our experiment the city of Saint John in New Brunswick, Canada, mainly because tourism is critical to Saint John’s culture, heritage, arts, recreation, and entertainment industries, and it also contributes significantly to city’s service industries including transportation and travelling services, accommodations, and food and beverage services. It is a 20 billion industry, with the Port Authority expecting over 60 cruises with a total of 140,000 passengers and 50,700 crew members in 2016.

We have selected the existing self-guided Art in Public Places walking tour. The tour consists of 9 POIs where features such as sculptures, figures, paintings, buildings are located outdoors and indoors. The geofences were designed to create virtual circles around these features. The POIs are shown on the map and can be described as one of the following:

• St. Andrew’s Bicentennial Square (Barbour’s General Store) People Waiting, by John Hooper These life-sized...
figures were originally installed at the Saint John main branch of Canada Post on Rothesay Avenue.

- Market Square Boardwalk: The Moosehead Challenger, by Forest Hart. This life-sized bronze moose was commissioned by Moosehead Breweries Ltd. and given as a gift to the City.
- Market Square Entrance: Timepiece, 1984, by John Hooper. This intricate carving acts as both a symbol of time and a functioning clock.
- Harbour Passage: Along Harbour Passage, with 3 pieces of public art.
- Free Public Library, 2nd Level Market Square. It features a number of sculptures and paintings.
- City Hall (Level L): City Hall boast two prominent pieces of art, both inside and outside the building.
- Saint John Arts Centre, Peel Plaza: A number of pieces are installed inside and outside this former library, which also houses the City of Saint John Gallery.
- Bell Aliant Building. This work was commissioned by Parks Canada in 1985, for the 200th anniversary of the City.
- Saint John City Market. The concept for this piece came from the market itself. The ceiling timbers supporting the market roof are similar to the hull of bygone wooden vessels. The artist designed a figurehead, such as those placed on these historic ships.

We are planning to mount twenty beacons and ten geofences through the tour, covering all the places and sculptures. The experiment was chosen to evaluate the functional requirements of the mobile application, the real-time recommender system and the notification server, as well as the system performance.

A. Functional Requirements

The first functional evaluation is to test if the mobile application could create a Google card based on received JSON formatted data from the notification server, when user’s events E1, E2, E3 or E4 occurs. This process includes the following tasks: T1 (get the notification from notification server) and T5 (to display the notification in the smartphone). The second functional evaluation is to test if the cloud receives the data of user’s events (E1, E2, E3 or E4) and GPS data. This process includes the following tasks: T2 and T3 (post geofence data) and T4 (post GPS data).

B. System Performance

The system performance is evaluated by conducting a non-functional test comparing the average time taken for receiving a notification and creating a Google card after users generate events on geofences. The test was performed using debug mode and logcat/monitoring from Android Studio 2.1.1, and the mobile device to load the application. Table 1 describes the specifications for the cloud app, mobile device used, the IDE, the gateway and the Images.

### Table I. Specifications of the Experiment Environment

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Cloud App</th>
<th>Mobile device</th>
<th>Gateway</th>
<th>Images</th>
<th>Beacon SDK</th>
</tr>
</thead>
</table>

In this environment we performed four experiments described in Table 2. The experiments test the performance of the different tasks developed for architecture elements. The last experiment (4) is conducted once per each image.

### Table II. Descriptions of the Experiments

<table>
<thead>
<tr>
<th>ID</th>
<th>Architecture elements tested</th>
<th>Tasks tested</th>
<th>Description</th>
<th>Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cloud, application performance</td>
<td>T2</td>
<td>Time taken to post geofence (without IoT devices) data to cloud after generating enter/exit event (E1, E2)</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3</td>
<td>Time taken to post geofence (without IoT devices) data to cloud after generating enter/exit event (E3, E4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T4</td>
<td>Time taken to post user location data to cloud after the thread is run each 10 seconds</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Network connection</td>
<td>T2</td>
<td>Time taken to perform the http connection to send geofence (without IoT devices) data to cloud</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T3</td>
<td>Time taken to perform the http connection to send geofence (without IoT devices) data to cloud</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T4</td>
<td>Time taken to perform the http connection to send user location data to cloud</td>
<td></td>
</tr>
</tbody>
</table>
The results in Table 5 show that in average, it takes more time for geofences without IoT devices to display Google Cards than with IoT devices. Here again, network connection is a variable to consider. Images and JSON files have to be downloaded and after that, it is needed to build the Google Card. The implementation to build the Google Card is the same in both cases. We are expecting a maximum of approximately 3.3 seconds after a user event to display the Google Card.

The overall results reveal that it takes more time for geofences without IoT devices than without IoT devices to display google cards and to post data. This could indicate that a geofence service is slower than a beacon service, due to the differences when using GPS rather than Bluetooth. However, for improving the results, we should also monitor the network connection (WiFi) during the experiment, to be able to subtract this variable to analyze the performance differences between Bluetooth and GPS.

V. CONCLUDING REMARKS

We have developed a prototype of our proposed real-time recommender system with pre-defined notifications/recommendations and have demonstrated its effectiveness. However, there are some considerations for deploying the mobile application on the target mobile devices. First, the mobile application needs to be running in the background of a smartphone in order to detect user events and transmit data regularly to the cloud, it can rapidly consume the battery. Second, it is of paramount importance to have a high bandwidth wireless network infrastructure at a tourist destination in order to be able to transmit streaming data in real-time.

In order to cope with the battery consumption on mobile devices and network bandwidth, we need to reduce background monitoring to detect Bluetooth signal from beacons and transmitting data to our cloud system. There could be many approaches to solve this battery issues.
However, cost-benefit analysis should be conducted before we choose the right approach, since there are always pros and cons of each approach.

Also, both GPS and Bluetooth sensing are services running in the background in order to detect user events, reducing the battery life of the phone. However, geofencing using GPS has more impact on the battery life of mobile devices than beacons as it determines the location using satellites services. Beacons do not need to know the exact location of a mobile user; it just uses proximity detection to estimate if a mobile user is within the range of its signal. Moreover, Bluetooth 4.0 is a low energy technology which is optimized for long battery life. In order to copy with the battery limitations, the use of beacons should be preferred against GPS. Also, the experiments shows that the wireless network is of paramount importance in the performance of the system. In order to be able to transmit notifications in real time is necessary to have a high bandwidth wireless network infrastructure at a tourist destination. Future research work will focus on conducting more experiments, monitoring the use of network and the use of battery, to be able to get more conclusions.

Another issue is that provided recommendations need to be meaningful to the end users. A way to measure this satisfaction could be to analyze if the users are actually following the recommendations we give to them. This would be easy since we have the GPS coordinates of their path. Recommendations could be updated in real time, changing the ones that the users do not follow. We need to consider the followings for our future research. Furthermore, we need to consider the followings for our future research.

a. Personalized recommendations
b. Context-aware (i.e. weather, a number of visitors, etc) recommendations

Currently, we have been evaluating our proposed IoT platform. Future research work will focus on improving the clustering and recommendation algorithms for finding clusters of users moving around a tourist destination in real-time. The recommendations will then be sent to a group of users moving with similar mobility patterns. With respect to the collaborative filtering for our real-time recommender system, we aim to introduce other rating factors rather than the visiting time in each POI and the frequency of visits of a POI. Finally, we will include more information about a user context such as transportation, weather conditions, and time of the day.

ACKNOWLEDGMENTS

This research is being funded by the NSERC/Cisco Industrial Research Chair in Real Time Mobility Analytics.

REFERENCES


