Master Thesis
A Crowdsourcing Based Crime Mapping System

Hairihan Tong

Thesis submitted in partial fulfillment of
the requirements for the degree of

Master in Information and Computer Science

<table>
<thead>
<tr>
<th>Student ID</th>
<th>5112BG08-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submission Date</td>
<td>July 21, 2014</td>
</tr>
<tr>
<td>Supervisor</td>
<td>Professor Tatsuo Nakajima</td>
</tr>
</tbody>
</table>

Department of Computer Science
Faculty of Science and Engineering
Waseda University
A Crowdsourcing Based Crime Mapping System

Tong Hairihan

學籍番号 5112BG08-2
提出 2014年7月21日
指導教員 中島 達夫
Abstract

In this paper, we deployed a real time crime mapping system service BianYi, in the field in China. We make a combination of crowdsourcing crime incidents from client applications and extract crime information contained contents from the social network services Sina Weibo. The goal of this paper is to establish a near real-time service via crowdsourcing approach and discover a way to collect crime information in a timely fashion by which will make a contribution to crime of geographical research and local police offices. The system mainly provided three features to the users: collect and distribute crime reports, visualize crime and send email alerts. We carried out an online survey to figure out user behavior so that evaluates each feature of the system. Finally, we discussed the main factors which will be effected on the user motivation and present the usability and the future directions based on this system.
概要

本論文で我々はリアルタイム犯罪マッピングシステムサービス BianYi を開発し、中国で本システムをリリースした。我々はクラウドソーシングを用いたクライアントアプリケーションからの犯罪事件の供給と、ソーシャルネットワークサービス Sina Weibo の犯罪情報を含めた内容を抽出するという二つの方法の組合せを用いた。この研究の目的は、クラウドソーシング・アプローチを通してリアルタイム・サービスに近いサービスを作る事と、犯罪地理学研究と地方警察署の犯罪に貢献できる、リアルタイムに犯罪情報を収集する新たな方法を探す事である。本システムは次の三つの機能からなる：犯罪報告の集めと配信、犯罪データの視覚化、電子メール警報の送信。我々は評価としてシステムの各々の特徴を評価する事とユーザのふるまいを理解するためのオンライン調査を行い、ユーザのモチベーションに影響する主因を議論し、本システムの有用性と今後の課題を提示した。
1 Introduction
  1.1 Background ......................................................... 2
  1.2 Motivation ......................................................... 2
  1.3 Structure of this Thesis .......................................... 3

2 Related Research
  2.1 Crowdsourcing ..................................................... 4
  2.2 Crowdsourcing in Social Network Services ..................... 4
  2.3 Mobile Crowdsourcing ............................................. 5
  2.4 Crowdsourcing Crime Incidents ................................ 5

3 System Design and Implementation
  3.1 Background: Microblogging Services in China ................. 6
  3.2 System Architecture .............................................. 6
  3.3 System Requirement ............................................... 7
  3.4 Main component of Client Applications ......................... 8
    3.4.1 Crime Report Constructor .................................. 8
    3.4.2 Data Visualizer ............................................. 9
    3.4.3 Email Alert Sender .......................................... 10
  3.5 Weibo Data Extractor ........................................... 11

4 Result and Evaluation
  4.1 Survey Result .................................................... 15
  4.2 Data Result ...................................................... 17

5 Discussion
  5.1 Security and Trust .............................................. 18
  5.2 User Interface Design ........................................... 18
  5.3 Limitation ....................................................... 19
    5.3.1 Open Government Data ..................................... 19
    5.3.2 Application Programming Interface ........................ 19

6 Future Work
  6.1 Online “Lost and Found” Service ............................... 20
  6.2 Limited-time Discount Information ............................. 20
List of Figures

3.1 High level system architecture ........................................ 7
3.2 Screenshot of website page: submit a crime report .................... 8
3.3 Screenshot of credibility score of a crime description ................ 9
3.4 Screenshot of crime map in BianYi website ............................. 9
3.5 Screenshot of crime list on the BianYi iPhone application ............ 10
3.6 Screenshot of submit a crime report on the iPhone application .... 10
3.7 The main procedure of the Weibo data extractor module ............ 10
3.8 Screenshot of email alert setting in website page ..................... 11
3.9 Screenshot of crime map on application ............................... 12
3.10 Screenshot of email alert on a iPhone email client .................. 12
3.11 Screenshot of the Sina Weibo topic search result .................... 13
4.1 Screenshot of part of survey form ....................................... 14
4.2 Results of questions: users’ intention to post a microblog on the social network service towards each type of crime ........................................ 15
4.3 Results of survey: users’ decision to post crime incident on BianYi or the social network services ........................................ 15
4.4 4 crime types classification ............................................... 16
List of Tables

3.1 Classification of crime ............................................. 12
3.2 Main comparison of micro-blog and crime report ..................... 13
Chapter 1

Introduction

Nowadays, social media services, especially micro-blogs, have become an indispensable part of people’s daily life, where they can share events around them, such as what they have seen and what events have occurred around them, by posting text messages, geo-information, photos or even videos on social media. As a result, there has been a growing interest in using the Internet crowdsourcing to solve crimes [1, 18]. For instance, in Beijing, since the way of using Sina Weibo¹ to call the police has been activated from beginning of May 2013, citizen can call to police by just put the context in a micro-blog or sent a message to the Weibo account “pinganbeijing”², which run by Beijing Municipal Public Security Bureas³. It is considered reduce the intermediate links and improve the efficiency of the scheduling command by the police. In addition, the Large Emergency Event Digital Information Repository (LEEDIR)⁴ platform allows any law enforcement and relief agencies in the nation to requests for eyewitness photos and videos from the public. LEEDIR is now available to all law enforcement and relief agencies around the US. Moreover, after the Boston Marathon bombings, police used online crowdsourcing, asked the public to submit photos, videos or anything suspicious that they may have seen or heard, to search hypothetical suspects and crime scenarios⁵.

Crime map is an instrument that visualizes crime information based on the geographical location of the crimes. Crime mapping can play an important role in the policing and crime reduction process, from the first stage of data collection through to the monitoring and evaluation of any targeted response [14]. With the progress of Geographic Information System (GIS) and Internet, the geo-information of a crime incident can be easily recorded at a digital map, which makes crime mapping more and more widely used for crime prevention and crime reduction program by police officer all around the world. Since the popularization of Global Positioning System (GPS) in smart devices, people can use a smart phone to easily share a geo-information contained multimedia message which describing a crime incident, like “I found a guy robbed a backpack from an old people”, “there is one guy seems stealing someone else’s bicycle”, “find a cutpurse pick one’s pocket now”, on the social media services. Because, the location of the crime occurred and crime committed could be the same place or is often close to where the crime was perpetrated [14, 6, 5, 4]. Therefore,

¹http://www.weibo.com
²http://www.weibo.com/pinganbeijing
³http://www.bjgaj.gov.cn
⁴http://www.leedir.us
⁵http://www.cbsnews.com/feature/boston-bombings
a commit location of a crime can be as near to the location of the crime. Thus, it creates an opportunity for us to develop a crime map, which can collect the information of the crime, such as a micro-blog describe a crime incident at that place, a news report on the internet, from the public by the social network services or directly on the internet.

In this paper, we deployed a real time crime mapping system service BianYi, in the field in China. As well as other crowdsourcing crime map, BianYi allow ordinary citizens to mark the crime incidents, according to its location and temporal information. However, BianYi provides a number of features that differ from a traditional crime map. Firstly, not only obtaining data from users uploaded, also getting crime information by analyzing messages from social network services. Secondly, provides a feature that can allow anyone to submit a comment on crime incidents and allow users to commit the credibility option of a crime report. Thirdly, alarm people about the crime information in a timely fashion, when a new incident marked in the area they cared about on the crime map.

1.1 Background

In earliest time, the police use a crime map to recognize the inherent geographical component of crimes by sticking pins into maps displayed on walls, where each pin on the map represents a crime incident. As a result, the police officer used crime map for crime analysis, thus helping prevent, reduce and solve crime problems. Since the Mapping and Analysis for Public Safety (MAPS) program held much of the innovation in crime mapping in the United States, the awareness of crime mapping has been so raised that many other countries, including the United Kingdom, Australia, South Africa and across South America is running crime mapping development.

Recently, there has been a growing interest in developing a crime map which obtaining crime data via using the methods of crowdsourcing or the government open data. There are a number of sites such as CrimeReports\textsuperscript{6}, CrimeMapping.com\textsuperscript{7} located in America; and SpotCrime\textsuperscript{8}, which are located in United Kingdom; and WikiCrimes\textsuperscript{9} are mapping the crime data, upload by users, to a digital map, and then analyzing crime incident patters, so that and ordinary citizens can use these maps to obtain an understanding of neighborhood crime or even to receive alert notices when a crime occurs in the place in which he or she is interested.

1.2 Motivation

However, almost all the crime data used in the crime map coming from the data obtained by the police officer, and it is considered take too long time and efforts from the take place of incident to confirm and fill it into the crime map. Furthermore, as the dark figure of crime problem \cite{23, 24, 25}, there are a lot number of many unrecorded and unreported crime, so the police crime data is problematic. Is there any way that can be obtained these data in a near real-time to get a crowd crime information database to the local police office and make it allow all the citizens to contribute to it?

\textsuperscript{6}http://www.crimereports.com
\textsuperscript{7}http://www.crimemapping.com
\textsuperscript{8}http://www.spotcrime.com
\textsuperscript{9}http://www.wikicrimes.org
In this paper, we introduced the crowdsourcing approach to obtain crime data to solve those problems, in which all the citizens are inspired to report crime incidents positively. Moreover, the system monitoring information flow of Sina Weibo to give a more choice of data sources. To the best of our knowledge, there is no similar system that has been deployed in the field in China before. Thus, BianYi is the first crowdsourcing based crime mapping system in China.

According to the crime statistics [19, 20, 21, 22], there are still have a huge number of crime incidents happened in the world every year. Worse, the trend is still keeps growing in some countries. In this paper, we would like to introduce a crowdsourcing based crime mapping system. We hope the system will raise police office working efficiency of control crime and increase the prevention awareness of people, by inspiring the citizens to report the crime incidents positively and accurately to the system. Furthermore, we hope the way we obtained data will provide sources for geo-criminal research in geography of crime, especially in the country that cannot obtain open government data, get an understating of local crime and even prediction of crime in the future. The goal of this research is to find out a way to establish a crime database by collecting timely crime data from the social networks and by utilizing a crowdsourcing approach, and make use of it to establish a series of near real-time services.

1.3 Structure of this Thesis

The rest of this paper is organized as follows. In Chapter 2, some related researches are described. In Chapter 3, the system design and implementation is explained in details. In Chapter 4, we analyze the survey results and usages data findings and the discussion and evaluation outcomes are discussed in Chapter 5. We put our expectation of the system and future work in Chapter 6 and the conclusion is given in Chapter 7 finally.
Chapter 2

Related Research

In this chapter, we would like to present some earlier efforts that were related to crowdsourcing or social network services and introduce some of the relevant system that utilizes the crowdsourcing crime incidents, which similar to the BianYi system.

2.1 Crowdsourcing

There are many previous works that utilize the contributions of the crowd on the World Wild Web (WWW) and achieved good results.

The most typical representative of it is Wikipedia\(^1\), which collect the knowledge of a huge number of people to benefit everybody on the Internet. Another successful example is OpenStreetMap\(^2\) (OSM), a free wiki world map, is a form of collective intelligence with location tracks from volunteers all over the world. There have been several efforts that using crowdsourcing to solve the language translation problem. In [2], Liu Y et al. proposed a crowdsourcing based mobile image translation system that towards people who unfamiliar with Chinese character's writing and reading. The system will distribute the words contained photos from the requester to the crowd translator and return back the answer to the requester. Besides, Lang-8\(^3\) is a language exchange social networking service held by Lang-8 Inc., offer a platform that gives a chance to all the language learners that complete a translation task helped by the native speakers.

2.2 Crowdsourcing in Social Network Services

There have been many research before BianYi, take a lot advantage of the geo-information contained contents and real-time attribute of the social networking service. In [3], Cox J et al. present research that using a method that by crowdsourcing citizen report on the real-world weather from the Twitter stream to improve the automatic weather observations. In [7], Liu Y et al. present MoboQ, a temporal and geo-sensitive question answering system, can solve the problems which are difficult to captured by sensors, such as whether a restaurant is crowded, whether a bank has a long waiting line. MoboQ user’s location-based questions will be sent to Sina Weibo users whose latest

\(^1\)http://www.wikipedia.org
\(^2\)http://www.openstreetmap.org
\(^3\)http://www.lang-8.com
check-in location is approached to the questioned aimed location, and return the answer back to the questioner.

2.3 Mobile Crowdsourcing

In [8], Stevens M et al. conducted a research that makes use of mobile-sourcing approach to measuring and mapping urban noise pollution, to build a collective noise map. The paper deployed a mobile phone application on top of the NoiseTube\(^4\) system, which collects the noise information with a geo-localized information from the users. Similar to weather crowdsourcing projects mentioned above, Mojitianqi\(^5\), a localized weather sharing smart phone application in China, which collect the temporal local weather photos (e.g., such as a sky photo, the streets in the rain) from the client users, to provide an intuitive feeling of weather for the people who concerned about it at that place.

2.4 Crowdsourcing Crime Incidents

A study conducted by Furtado V et al. is very similar to our work[9]. The paper has described the WikiCrimes\(^6\), a Web 2.0 application that creates a collaborative environment based on the use and direct manipulation of maps, in order to register and research criminal events. However, there are two major differences between their work and BianYi system. The first difference is that we are additional providing a near real-time email alert service. The next difference is the source of crime data, Wikicrimes take advantages of government open data, whereas it is hardly obtained by BianYi, since it is difficult to find government open data in China. In [10], Shah S et al. proposed CROWDSAFE, a smartphone application that crowdsourcing crime incidents and find out a safe routing between two points in the map for users by analyzing and calculating the obtained crime data. The service is different from BianYi because it is only extracted from its mobile user, while BianYi provides both data from crowdsourcing of mobile devices and extracting of real-time social network services.

\(^4\)http://noisetube.net
\(^5\)http://www.mojichina.com/mjsoft/
\(^6\)http://www.wikicrimes.org
Chapter 3

System Design and Implementation

In this section, we present the design and implementation of the BianYi system: the high level of the system architecture, the system requirement, the design and key features of client applications and Weibo data extractor module.

3.1 Background: Microblogging Services in China

Sina Weibo is a Chinese Twitter-like micro-blog platform, which launched by Sina Corporation\(^1\) in August 2009. The main features of Sina Weibo are very similar to Twitter\(^2\), also providing a number of Application Programming Interfaces (APIs) to third parties. As of end-March 2013, Sina Weibo reached 556 million registered users, 54 million daily-active users and own 3.3 billion times API called per day on an average, which makes it the most popular microblogging service in China.

Furthermore, similar to Twitter’s verified account for known people and organizations, Sina Weibo also has an identification policy. However, Sina Weibo has a more complex identification policy and verification process than Twitter. Several types of verification accounts (i.e., people or organization need to upload their national identification number or an authorization letter with company’s seal or legal person’s signature to declare their account) are supplied and mainly can be divided into personal accounts and corporate accounts, which includes media platform, company and government agencies and so on. According to China Daily Asia\(^3\), China has the largest number of government micro blog accounts in the world, among which about 100 thousand government accounts had been verified on the Sina Weibo, by the end of the 2014. BianYi can take the advantages of both these Sina Weibo accounts public timeline to get the crime incident information.

3.2 System Architecture

BianYi is a crowdsourcing crime map system that composed of a server running in the cloud and client application that include a website and a iPhone application to present the interactive crime contents to the users. The high level of the BianYi system architecture is shown in Figure 3.1.

---

\(^1\)http://www.sina.com  
\(^2\)http://www.twitter.com  
\(^3\)http://www.chinadailyasia.com
As shown in Figure 3.1, the main components of client application are the crime data constructor, the data visualizer, the email alert sender and the Weibo data extractor. The crime data constructor plays a role in creating a crime report; The data visualizer is used for display the crime information on a Google map; The email alert sender takes a responsibility for sending an email to the BianYi user who have set an email alert in the BianYi system; The Weibo data extractor is a component that runs in the cloud to ensure obtaining data from Sina Weibo.

In addition, it is not difficult to find that BianYi obtain data from two ways: the first one is crowdsourcing from people who are used BianYi client applications, that is, BianYi users can upload a crime report, which should be included geo-information of the crime location, a description of the crime and the crime happened date, by using the iPhone application or the web application. On the other hand, BianYi extract crime information from Sina Weibo. More specifically, BianYi extract crime containied micro-blogs from Sina Weibo’s public timeline, especially from verified police office accounts and under several specified topics, to the ensure obtain the data in timely. The details of each component of system APIs and algorithms shall be described in the section 3.4 and 3.5.

### 3.3 System Requirement

We made a full use of open source crowd map system Ushahidi Platform[^4]. A traditional web services solution stack LAMP, consisting of Linux (3.2.0–29–generic x86_64), the Apache(2.2.22) HTTP Server, the MySQL(5.5.37) database engines, and the PHP(5.3.10) programming language, was chosen for BianYi system development. On the server side is running in the Amazon Elastic

[^4]: [http://www.ushahidi.com](http://www.ushahidi.com)
Compute Cloud (Amazon EC2)\(^5\). In addition, the objective-c programming language was used in the iPhone application development.

3.4 Main component of Client Applications

The client applications are composed of three modules, crime report constructor, data visualizer and email alert sender. The main features of each module are: submit and receive a crime application, show crime reports and send email alerts.

3.4.1 Crime Report Constructor

![Screenshot of website page: submit a crime report](image)

Figure 3.2: Screenshot of website page: submit a crime report

This module can easily extend to the following four features: submit reports to the crime map, view and comment on a crime report, change a credibility score of a crime report. A crime report is required five values: a report title, a description of the crime, a date of the crime occurred, a location of the crime and a name of the crime categories. Besides, user information (i.e., name, name,

\(^5\)http://aws.amazon.com/cn/ec2
contact) and media resources of crime (i.e., photos, videos, link) are optional. The screenshot of submitting a crime report on the web application side and the iPhone application side are given in Figure 3.2 and Figure 3.6.

A credibility score option to represent a model of trust and reputation of the users, is a significant difference between BianYi and other crime map systems. Users can add the credibility score if they trust the report or reduce it if they not trust it. The screenshot of credibility score of a description of a crime on the web application side looks like Figure 3.3.

In general, it is not as easy as for the user to fill the geo-information of the crime location by using a device not equipped GPS. In order to solve this problem, BianYi allows these users to directly fill in the location name, then uses Google Geo-encoding to translate the location name into geo-information.

![Figure 3.3: Screenshot of credibility score of a crime description](image)

![Figure 3.4: Screenshot of crime map in BianYi website](image)

### 3.4.2 Data Visualizer

All the crime data will be displayed as a color spot on the digital map, according to its location and classification. When there are multi-incidents took place in a region, the number of criminal incidents will be showing on the spot on the web sites. Certainly, the users can use the zoom in or zoom out operation on the map, to see the overall situation or the detail description of a crime in a place of their interest. Furthermore, the users also can utilize report incidents filter to view the specified types of crime. The screenshot of a crime map of BianYi web application and the iPhone
application is shown in Figure 3.4 and Figure 3.9. Certainly, all the crime data also give a list view to the users, the screenshot of crime list in BianYi iPhone application is shown in Figure 3.5.

3.4.3 Email Alert Sender

BianYi allow the users to set up an email alert sender, to get the latest crime reports occurred in the selected zone. As shown in Figure 3.8, specified area and email address used for receiving email alerts is required. Users have to select a location by moving the spot on the map or input a
location name first, and then choose a radius as the center of the circle to set an area. In addition, users can select a specified category of crime is optional for the users to filter no need messages. The setting will be activated after accomplish the email authentication. The screenshot of email alert on a iPhone email client is given in Figure 3.10.

3.5 Weibo Data Extractor

This module extracts crime information contained micro-blogs from Sina Weibo. As shown in Figure 3.7, the main procedure of this module is following three sub-procedure:

Sources manager

Manage the sources of crime information contained micro-blogs, including police office verified accounts and the specified topic titles. The screenshot of the website page of Sina Weibo topic search is given in Figure 3.11. Because Sina Corporation has closed Weibo search API to all the developers since the beginning of 2013 so that in this paper, we extract the data by using the topic search API and from police officer public timeline instead of Weibo search APIs. In fact, even if Weibo search API is available, the results of its will required Natural Language Processing (NLP) to filter the error messages, which considered to have a low accuracy. Because of lack of the related APIs and documents, we could not obtain all these verified accounts and topics in automated. Instead, we do the collection all of them are manual and insert it into the database via human coding.
Key Conversations Extraction

Extract the micro-blog from the sources obtained in the previous step by using the keyword database. The keyword database is stored the relationship between search used keywords and its crime types. As shown in Table 3.1, BianYi classified the crime into following four types: theft, robbery, violence and others and each of it continue to sub classified to several crime types. In order to identify the crime type of micro-blog contents, we decide to use the name of crime type as keywords, thus the result of a keyword will be identified as that crime types. For example, the crime type of all the results extracted from the keyword “pickpocketing” will be classified as “pickpocketing”. If there are several keywords matched to one micro-blog, BianYi will choose the more serious crime as its crime type. The priority order of all crime is violence, robbery, theft and others.

Table 3.1: Classification of crime

<table>
<thead>
<tr>
<th>Crime Types</th>
<th>Sub-Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theft</td>
<td>Pickpocketing, Burglary, Vehicle theft, Bicycle theft, Other theft</td>
</tr>
<tr>
<td>Robbery</td>
<td>Home invasion, Handbag Forcible, Seizure, Other Robbery</td>
</tr>
<tr>
<td>Violence</td>
<td>Homicide, Indecent assault, Rape, Intentional injury, Other violence</td>
</tr>
<tr>
<td>Others</td>
<td>Affray, Fraud, Other</td>
</tr>
</tbody>
</table>

After using the keywords database and topic search API, user timeline API, the results will
Figure 3.11: Screenshot of the Sina Weibo topic search result

output to the next step.

<table>
<thead>
<tr>
<th>Sina micro-blog</th>
<th>BianYi crime report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro-blog contents</td>
<td>Crime descriptions</td>
</tr>
<tr>
<td>Micro-blog geo-information</td>
<td>Crime geo-informationery</td>
</tr>
<tr>
<td>Micro-blog published date</td>
<td>Crime happened date</td>
</tr>
<tr>
<td>Matched Keywords Category</td>
<td>Crime type</td>
</tr>
<tr>
<td>Category + Location City</td>
<td>Crime title</td>
</tr>
</tbody>
</table>

Table 3.2: Main comparison of micro-blog and crime report

Data Cleaning

Processing the data obtained at the key conversations extraction step. Mainly included repeated micro-blogs cleaning, useless information cleaning (e.g., Emoji or URL in a micro-blog) and format to the BianYi crime Report Format. The main comparative relationship between the micro-blog and crime report is shown in Table 3.2.
Chapter 4

Result and Evaluation

In this chapter, we would like to describe the online survey conducted with this crime mapping system in the field in China to understand the behavior and decision of users to the each type of crime and each feature of the BianYi system. The screenshot of parts of the survey form is given in Figure 4.1.

Figure 4.1: Screenshot of part of survey form
4.1 Survey Result

We asked a series of questions towards the users to understand if they would post a microblog when they met each type of crime incident. The question is like “Suppose you met a theft, in your daily life (e.g., on the road, subway, shops), will you share this matter on the social network services”. As the result shown in figure 4.2, we find that the percentages of people said they will not post a microblog when they met a theft, robbery and violence are 16.7%, 50% and 66.7%. Whereas, the percentages that people answered they may post a microblog when they met type of crime: a theft, robbery and violence are 50%, 33.3% and 0%. It refers that, Chinese users have a tendency that the more serious crime, the less reported on social networks. Conversely, the lesser crime incident is more likely to be reported on social networks. We believe that the probability of citizen met a serious crime in their daily life is lower than a lesser crime.

Figure 4.3: Results of survey: users’ decision to post crime incident on BianYi or the social network services
In order to investigate if the citizens want to customize a crime alert service like BianYi and if their decisions will change when the BianYi system is run by the police officers, we asked following two questions: “Would you like to customize a crime alert service on the service like BianYi”; “If BianYi is run by the public security organizations, what will be your decision”. As shown in Figure 4.3, the result of the first question showed the half the users have said that “Sure, I will do” and the half of the user said “I will do in general”. However, all of them answered they would like to customize this alert service when BianYi is run by the public security organization. It is shows that the users surely need alert services to remind them of the neighbor crime incident in timely. In addition, we consider that the users may care about the security and accuracy of the BianYi system, since we did not cooperate with any of local police officer.

The result of question we asked “The frequency of your check in on the crime map like BianYi”, are 83.3% participants chose “Occasion” and 16.6% of them chose “Never”, which indicate that the user not likely to check or view the crime map. Even we think it is the intuition of crime, the crime map may be more useful to the professional than the ordinary citizens.

For the purpose of investigating the user behavior of preferring to submit it to the like BianYi services or post it to the social networks, we request the users to answer the following two questions: “Suppose you met a criminal behavior in your daily life (e.g., theft, robbery, murder, fights), will you take this matter to BianYi or on social networks” and “If BianYi was run by the public security organizations, what will be your decision”. The result show that 50% of the participants said they want to share it both the BianYi and social networks services no matter the system is run by a police officer or not. However, 33.3% users showed they will more likely to report it on the BianYi system than social networks, if it is run by the police organizations. Constantly, none of them would like to post it on the BianYi services if it is not run by a police officer. The number of people who want to post crime on crime map system will update if it is run by a government organization. However, the social networks remain a hub of crime incidents reports.

Figure 4.4: 4 crime types classification
4.2 Data Result

We have randomly chose to identify 1000 reports from all the data and classify it based on the crime type by human coding. As shown in Figure 4.4, we have find that the most collected crime data’s crime type is “theft”, which own almost 80% of all the reports. In addition, “robbery” and “violence” type of crime data represent 8.2% and 5.9% of all the crime data. It can be matched the result referred to the previous Figure 4.2 that the lesser crime incidents are more likely to be reported on social networks. Moreover, we have randomly chose 200 records, which obtained from the Weibo data extractor module to determine how many percent of the crime reports will be meaningful. We got a result of only 9.5% percent of result is meaningful, and we find that most of the micro-blog are posted by police verified accounts are police tips and propaganda purpose micro-blogs.
Chapter 5

Discussion

In this chapter, the main factors of which will effects the user motivation are discussed. Furthermore, the limitations of this thesis are considered and suggestions for further version are presented.

5.1 Security and Trust

The trust is one of the most important factors to motivate users contribute data to the crowdsourcing service [16, 17]. In this paper, we take a lot advantages from the verified police accounts on the Sina Weibo to give people a trust of the data sources. However, accuracy and authenticity of crowd-sourced data are still a problem. Although we add the credibility attributes to each crime description for users to edit, the verification of the authenticity to one incident is unable to do for anyone except the witness. A limitation (e.g., identity authentication, verified accounts) to the users who can edit credibility option should be a better solution, but the level of openness will be reduced. Furthermore, on the other hand, although, we required user name and contact for optional when they submit crime reports, some of the users still worried about their personal information and the security of the BianYi system.

5.2 User Interface Design

User interface design is playing a major role in user determination [15]. We find from the surveys that some users considered that the design of the BianYi system is still need to be updated, because they think it is troubling for them to fill many forms to upload a crime report. Since the birth of wearable devices, there are many applications tried to combine the smartphone and devices to make it more easily used for human computer interaction [13]. We believe that a wearable interface will overcome deficiencies of current crime reports form it the future. On the other hand, because of the battery consumption and security reason, we cannot be real-time monitoring the GPS information of the users smart devices, so that the users have to customize the geographic area of email alerts each time they required a new area alerts. In addition, we consider to add a push technology instead of an email alerts on the smart devices will be more easy to accept by the users.
5.3 Limitation

Actually, there have several limitations on the BianYi system, mainly summarized as the following two aspects:

5.3.1 Open Government Data

A spatial analysis of the result of the data and a comparison and contrast between the official data will prove and evaluate the obtained data quality very well [11, 12]. However, as our best of knowledge, there are no open government data in the field in China, so we are unable to prove the worth of crowdsourced data very well. Nonetheless, we believe the social networks services and the crowdsourcing approach can provide a hub of crime information, according to the user behavior shown in the survey described at previous chapter, and the police officer verified account on Sina Weibo.

5.3.2 Application Programming Interface

The data extract module is depends much on the corporation of social network services. In our cases, the real time attributes required streaming APIs, which not provided to ordinary developer by the Sina Corporation. Moreover, the topic search API is limited to only provided 200 micro-blog results under one topic.
Chapter 6

Future Work

In this chapter, we would like to describe the future direction of our research and the system can be deployed on top of the BianYi system.

6.1 Online “Lost and Found” Service

When people lost their thing in their daily life, they have a trend to express it on the social network services, especially in a small circle (e.g., campus, lab, gym, mess hall). They do not know exactly where they lose their thing, but what they can ensure is that it must at the somewhere of their circle environment. The finder maybe puts the thing to a local Lost and Found. However, because of unknown the owner of the thing, the local Lost and Found is waiting for the owner to pick up it in most of the time. A mapping system with crowdsourcing from social networks could be a solution for this problem. The finder can take a picture or write a description of the thing they find and submit it to the map according to the location of the thing. So that the owner can be search the location and category of the lost thing on the system.

6.2 Limited-time Discount Information

The sellers opened a limited-time discount might be want to dispose the products in entirely (e.g., vegetable and fruit) or attract buyers to the mall or some other reasons. On the contrary, the buyers might want to buy cheaper products. However, the limited-time discount information seldom available on the Internet in general, because it is relying heavily on the trading conditions of current day. Sometimes when the business is good and the products are sold entirely, or many buyers in the mall that day, sellers did not need to open a limited-time discount. Certainly, the condition of current day’s business is hard to predict in general. So the social media and the crowdsourcing will be a good solution for both the buyers and sellers.
Chapter 7

Conclusion

In this paper, we deployed BianYi, a crime map system, which investigated the theory of crowd sourcing crime incidents. BianYi, gathered both crime reports from the crowd and crime information contained contents (e.g., I lost my wallet today, must be stolen at the train station), from Chinese microblogging service Sina Weibo. We take advantages of the verified police officer accounts and several topics about crime incidents on the Sina Weibo, where have many real world micro-blog describing crime incidents posted by the Weibo the users. BianYi users can share a crime report on crime map, review and comment on a crime incidents on a digital map which crowdsourced from users and social network services, in addition, are able to customize an email alerts a timely fashion with the specified area and crime type.

In order to verify the advantages of building crowd sourcing crime applications, we discussed the people’s decision towards the neighbor crime incidents and proposed a list of future ideas for the crime sourcing applications, by analyzing the obtained data and survey results.

We sincerely hope that the crowdsourcing crime applications will be widely used in the future, and will help police work and bring convenience and benefits to the people all over the world.
Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor Professor Tastuo Nakajima for the continuous support of this thesis and 2 years study and life in Japan, for his patience, motivation, enthusiasm, and immense knowledge. Besides, I would also like to convey thanks to the Ting Hsin International Group, Tingyi (Cayman Islands) Holding Corp. and Waseda University for providing the scholarship and tuition waiver for me to which support my two years economic costs.

Special thanks also to all members of Distributed and Ubiquitous Computing Lab, Waseda University, especially Yefeng Liu, Ning Li, Han Chen, Chye Connsynn, Kanako Nakajima, Ryusuke Nomoto, Heiki, Rui Tao. Without their encouragement and editing assistance, I would not have finished this thesis.

Last but not the least, I would like to thank my family: my parents Baode Tong and Xingmin Shi, for giving birth to me at the first place and supporting me spiritually throughout my life.
References


