

# Resolving Mobile Communique App Using Classification of Data Mining Technique

M.Mohamed Suhail,  
Lecturer, Department of Computer Science,  
Jamal Mohamed College, Trichy,

K.SyedKousarNiasi,  
Assistant Professor, Department of Computer Science,  
Jamal Mohamed College, Trichy,

**Abstract-** A mobile app is a software application developed specifically for use on small, wireless computing devices, such as smartphones and tablets, rather than desktop or laptop computers. On the Internet and mobile app, chatting is talking to other people who are using the Internet at the same time you are. Usually, this "talking" is the exchange of typed-in messages requiring one site as the repository for the messages (or "chat site") and a group of users who take part from anywhere on the Internet. In this work, we apply classification techniques of data mining using a dataset of some mobile applications (eg. whatsapp, Line...etc.) to analyze the messengers. The aim of this work is analyze and summerize relevant information from social media with the help of the Data Mining Techniques.

**Keywords:** Messengers, Mobile Communication, Analytics, Classification.

## 1. INTRODUCTION

The term "mobile app" used to describe Internet applications that run on smartphones and other mobile devices. Mobile applications usually help users by connecting them to Internet services more commonly accessed on desktop or notebook computers, or help them by making it easier to use the Internet on their portable devices. A mobile app may be a mobile Web site bookmarking utility, a mobile-based instant messaging client, Gmail for mobile, and many other applications. Instant messaging is a type of communications service that enables you to create a kind of private chat room with another individual in order to communicate in real time over the Internet, analogous to a telephone conversation but using text-based, not voice-based, communication. Typically, the instant messaging system alerts you whenever somebody on your private list is online. You can then initiate a chat session with that particular individual. Nowadays we are frequently using some mobile apps for instant messaging (Whatsapp, Line, Viber, etc...). These apps are making it easier for people to find and communicate with individuals who are in their networks using the Web as the interface. The purpose of this work is to compare and analyze the social networks for the people to find the best social network by using the Data mining Technique.

Data Mining refers to extracting or "mining" knowledge from large amount of data. Many other terms carry a similar or slightly different meaning to Data mining, such as knowledge mining from data, knowledge extraction, data/pattern analysis, Data archaeology, and Data dredging. Many people treat data mining as a synonym for another popularly used term, Knowledge Discovery from Data, or KDD. Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration. Data mining can be viewed as a result of the natural evolution of information technology [1].

Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Data mining requires data preparation which can uncover information or patterns which may compromise confidentiality and privacy obligations. A common way for this to occur is through data aggregation. Data aggregation involves combining data together (possibly from various sources) in a way that facilitates analysis.

Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

The term "Real Time" is used to describe how well a data mining algorithm can accommodate an ever increasing data load instantaneously. However, such real time problems are usually closely coupled with the fact that conventional data mining algorithms operate in a batch mode, where having all of the relevant data at once is a requirement. Data mining algorithms are used in a vast range of fields and situations yet, in almost all cases of their

application, there are inevitable limitations that arise that significantly constrain successful and desirable outcomes. Frequently, these problems are associated with large increases in the rate of generation of data, the quantity of data and the number of attributes (variables) to be processed. Increasingly, these data situations are beyond the capabilities of conventional data mining methods.

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system.

## 2. MOTIVATION

There are lots of mobile messaging apps available in Appstore. Mobile messaging apps are growing ever more popular as they add social networking features and compete to meet the growing demand for free mobile calling and SMS texting services. Established mobile apps like WhatsApp Messenger and older Internet calling services such as Skype increasingly are competing with upstart mobile apps, including Viber, Line and WeChat, etc. Mobile communication apps allow users to share ideas, pictures, posts, activities, events, and interests with people in their network. Nowadays, the Mobile apps used by all people (eg. WhatsApp, WeChat, MessageMe, Line, etc...). In this research, using the data mining technique to analyzing which network is highly preferred and used by the people.

## 3. RELATEDWORKS

### 1. Influence Propagation in Social Networks: A Data Mining Perspective

Francesco Bonchi [3] proposed to apply a data mining perspective and we discuss what (and how) can be learned from the available traces of past propagations. While doing this we provide a brief overview of some recent progresses in this area and discuss some open problems. In this area is the identification of influential users, by targeting whom certain desirable marketing outcomes can be achieved.

### 2. Distilling Massive Amounts of Data into Simple Visualizations: Twitter Case Studies

Miguel Rios and Jimmy Lin [4] proposed to different patterns of activity between the four cities. For example, waking/sleeping times are relatively constant throughout the year in Tokyo, but the other cities exhibit seasonal variations. The Japanese users' activities are concentrated in the evening, whereas in the other cities there is more usage during the day. In Istanbul, nights get shorter during August; Sao Paulo shows a time interval during the afternoon when tweet volumes goes down, and also longer

nights during the entire year compared to the other three cities.

3. Detecting Sentiment Change in Twitter Streaming Data  
Albert Bifet, Geo Holmes, Bernhard Pfahringer, Ricard Gavald [5] proposed to real-time system to read tweets in real time, to detect changes, and to find the terms whose frequency changed. Twitter is a micro-blogging service built to discover what is happening at any moment in time, anywhere in the world. Twitter messages are short, and generated constantly, and well suited for knowledge discovery using data stream mining. MOA-TweetReader is a software extension to the MOA framework. Massive Online Analysis (MOA) is a software environment for implementing algorithms and running experiments for online learning from evolving data streams.

### 4. Data Mining in Social Networks

David Jensen and Jennifer Neville [6] proposed to make good use of research in other areas, such as social network analysis and statistics. Cross-disciplinary efforts and joint research efforts should be encouraged to promote rapid development and dissemination of useful algorithms and data representations. In particular, this work should focus on the unique statistical challenges raised by relational data.

### 5. Data Warehousing and Analytics Infrastructure at Facebook

Ashish Thusoo, Dhruva Borthakur, Raghotham Murthy, Zheng Shao, Namit Jain, Hao Liu, Suresh Anthony and Joydeep Sen Sarma [7] proposed to analysis of data and creation of business intelligence dashboards by analysts across the company, a number of Facebook's site features are also based on analyzing large data sets. These include Scribe, Hadoop and Hive which together form the cornerstones of the log collection, storage and analytics infrastructure at Facebook. They will present how these systems have come together and enabled us to implement a data warehouse that stores more than 15PB of data (2.5PB after compression) and loads more than 60TB of new data (10TB after compression) every day. They discuss the motivations behind our design choices, the capabilities of this solution, the challenges that we face in day today operations and future capabilities and improvements that we are working on.

### 6. Mining Social Media: A Brief Introduction

Pritam Gundecha, Huan Liu [8] proposed to use of social media has generated unprecedented amounts of social data. Social media provides easily an accessible platform for users to share information. Mining social media has its potential to extract actionable patterns that can be beneficial for business, users, and consumers. Social media data are vast, noisy, unstructured, and dynamic in nature, and thus novel challenges arise. This tutorial reviews the basics of data mining and social media, introduces representative research problems of mining social media, illustrates the application of data mining to social media using examples, and describes some projects of mining social media for

humanitarian assistance and disaster relief for real-world applications.

#### 7. Scaling Big Data Mining Infrastructure: The Twitter Experience

Jimmy Lin and Dmitriy Ryaboy [9] proposed to discuss the evolution of our infrastructure and the development of capabilities for data mining on "big data". One important lesson is that successful big data mining in practice is about much more than what most academics would consider data mining: "life in the trenches" is occupied by much preparatory work that precedes the application of data mining algorithms and followed by substantial effort to turn preliminary models into robust solutions. It has two goals: For practitioners, we hope to share our experiences to attend bumps in the road for those who come after us. For academic researchers, we hope to provide a broader context for data mining in production environments, pointing out opportunities for future work.

#### 4. EXISTING MODEL

In the existing model is only survey about Mobile applications. For example, the previous research work, how many users used mobile applications (e.g. WhatsApp, WeChat, etc...) in a particular country or city. It was done by using some data mining technique.

#### 5. PROPOSED MODEL

Mobile communication apps refers to the means of interactions among people in which they create, share, and/or exchange information and ideas in virtual communities and networks. Mobile messengers depends on mobile and web-based technologies to create highly interactive platforms through which individuals and communities share, co-create, discuss, and modify user-generated content. It introduces substantial and pervasive changes to communication between organizations, communities, and individuals. A mobile communication application has positive effects such as allowing the democratization of the internet while also allowing individuals to advertise themselves and form friendships.

In the proposed system is to find out the best Mobile communication App, which one is highly preferred and used by the people, when analyze and compare the apps with the help of the Data Mining Technique.

#### 6. METHODS

##### 6.1. Data Mining Technique

Data mining is the process of automatically collecting large volumes of data with the objective of finding hidden patterns and analyzing the relationships between numerous types of data to develop predictive models.

Data Mining is an analytic process designed to explore data. Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses.

Data mining is an interdisciplinary field, the confluence of a set of disciplines, including database systems, statistics, machine learning, visualization, and

information science. Moreover, depending on the data mining approach used, techniques from other disciplines may be applied, such as neural networks, fuzzy and/or rough set theory, knowledge representation, inductive logic programming, or high performance computing.

In this work, we use the classification techniques and text analytics for social network analytics. The most rapid data growth is not in numerical data, but in text - Twitter feeds, the contents of Facebook pages, emails, etc. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. Such analysis can help provide us with a better understanding of the data at large.

Classification technique is defined as the learning from the data of a function to map, or classify, data points into classes. Such a function can be used straightforwardly in deciding which action to take if the classes are naturally associated to different actions.

Data mining systems can be categorized according to the kinds of knowledge they mine, i.e., based on data mining functionalities, such as characterization, discrimination, association, classification, clustering, trend and evolution analysis, deviation analysis, similarity analysis, etc. A comprehensive data mining system usually provides multiple and/or integrated data mining functionalities.

A data mining system can be classified according to the kinds of databases mined. Database systems themselves can be classified according to different criteria (such as data models, or the types of data or applications involved), each of which may require its own data mining technique. Data mining systems can therefore be classified accordingly.

Data mining creates classification models by examining already classified data (cases) and inductively finding a predictive pattern. A number of data mining algorithms have been introduced to the community that perform summarization of the data, classification of data with respect to a target attribute, deviation detection, and other forms of data characterization and interpretation.

Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. Such analysis can help provide us with a better understanding of the data at large. In this work using the technique in data mining is decision tree classifier.

The C4.5 technique is one of the decision tree families that can produce both decision tree and rule-sets; and construct a tree. Besides that, C4.5 models is easy to understand as the rules that are derived from the technique have a very straightforward interpretation. J48 classifier is among the most popular and powerful decision tree classifiers. C5.0 and J48 are the improved versions of C4.5 algorithms. WEKA toolkit package has its own version known as J48. J48 is an optimized implementation of C4.5. The J48 classifier produced the analysis of the training dataset and the classification rules.

### 6.2. Data Mining Tool Selection

Data mining tool selection is normally initiated after the definition of problem to be solved and the related data mining goals. However, more appropriate tools and techniques can also be selected at the model selection and building phase. Selection of appropriate data mining tools and techniques depends on the main task of the data mining process. In this paper we have used WEKA software for extracting rules and built decision tree. The selected software should be able to provide the required data mining functions and methodologies. The data mining software selected for this research is WEKA (to find interesting patterns in the selected dataset).

Weka is developed at the University of Waikato in New Zealand. "Weka" stands for the Waikato Environment of Knowledge Analysis. The suitable data format for Weka data mining software are MS Excel and arff formats respectively. Scalability-Maximum number of columns and rows the software can efficiently handle. However, in the selected data set, the number of columns and the number of records were reduced.

The system is written in Java, an object-oriented programming language that is widely available for all major computer platforms, and Weka has been tested under Linux, Windows, and Macintosh operating systems. Java allows us to provide a uniform interface to many different learning algorithms, along with methods for pre and post processing and for evaluating the result of learning schemes on any given dataset. Weka expects the data to be fed into to be in ARFF format.

### 6.3. Classification of Mobile Communique App

#### WhatsApp:

WhatsApp Messenger is an instant messaging app for smartphones that operates under a subscription business model. The proprietary, cross-platform app enables users of select feature phones to use the Internet to communicate. In addition to text messaging, WhatsApp can be used to send images, video and audio media messages. WhatsApp has also started rolling out the much awaited voice calling feature. Locations can also be shared through the use of integrated mapping features.

#### Line:

Line (styled "LINE") is a proprietary application for instant communications on electronic devices such as smartphones, tablet computers and personal computers. Line users exchange texts, images, video and audio, and conduct free VIP conversations and video conferences. Line was originally developed as a mobile application for Android and iOS smartphones.

#### WeChat:

WeChat is a mobile text and voice messaging communication service developed by Tencent QQ in China, first released in January 2011. It is the largest standalone messaging app by monthly active users. The app is available on Android, iPhone, BlackBerry, Windows Phone and Symbian phones, and there are also Web-based and OS X clients but these require the user to have the app installed on a supported mobile phone for authentication. As of

August 2014, WeChat has 438 million active users; with 70 million outside of China.

#### MessageMe:

MessageMe is a newer entrant into the crowded field of mobile messaging. The app launched early in 2013 and is focused on multimedia and context, allowing people to mix media by drawing on photos and send instant messages that blend media types. It has a slick interface designed to simplify the process of creating and sharing multimedia messages in real time. MessageMe is available for iPhone/iPad and Android devices.

#### Viber:

Viber is a popular free app that piggybacks on WiFi or the data plan a cell phone user has to make voice phone calls and send SMS messages that are not counted as part of their carrier's paid calling plan or text messaging plan. The company's tag line is "Connect Freely". So its main appeal is the ability to use WiFi or a paid mobile data plan on a smart phone to circumvent the limits of a paid voice-calling plan and/or paid SMStext messaging plan.

#### Telegram

Telegram is an instant messaging system focusing on privacy and multi-platform availability. Telegram clients exists for both mobile (Android, iOS, Windows Phone) and desktop systems (Windows, Mac OS X, Linux). Telegram users can exchange messages encrypted end-to-end, self-destructing messages, along with photos, videos and files of any type up to 1.5GB in size.

### 6.4. Architectural Diagram

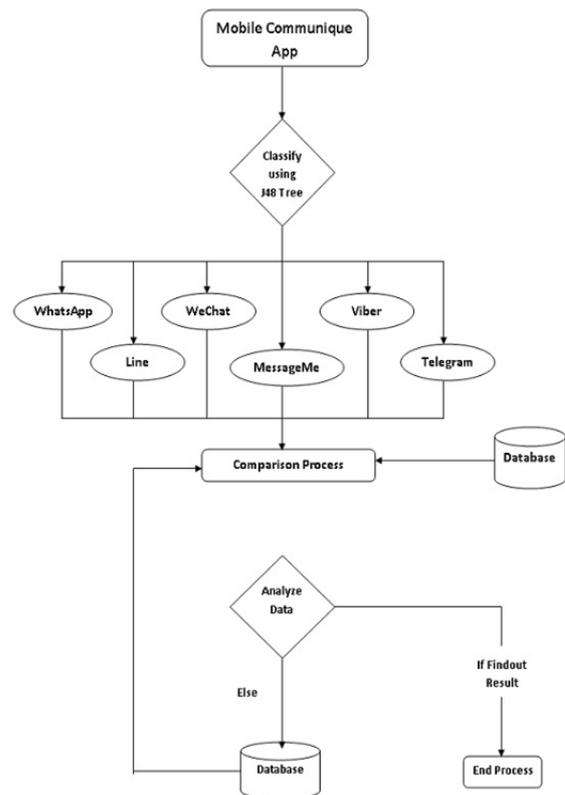


Fig 1.1. Architectural Diagram for analytics of Mobile Communique App

**7. RESULT AND DISCUSSION**

Messaging apps help people connect with one another, share photos, video chat, and, increasingly - play games, buy digital goods, and even shop offline. It has become a fiercely competitive space, into which Internet giants have poured billions of dollars. The most popular are WhatsApp, Line WeChat, etc.

Table 1.1: Mobile Messaging Apps Growth in Last Six Months based on percentage

Apps	Growth (%)
WhatsApp	75
Line	44
WeChat	22
MessageMe	10
Viber	35
Telegram	9



Fig 1.2 Mobile Messaging Apps Growth in Last Six Months

The above figure has shown which mobile messaging app is highly growth in the last six months. From this survey WhatsApp is highly preferred and used by the lot of people. Next highly preferred app by the user is Line.

**8. CONCLUSION**

Based on these results will be able to find out which Mobile communicate app is the best and mostly preferred and used by the user. We conclude WhatsApp is mostly preferred by the people. It is very user friendly, secure. The Next and Second place of the Messaging app is Line. These works done by using the data mining classification technique for analyze and compare the data. Future models can be used in the design of decision support system or cluster analytics technique.

**9. REFERENCES**

- [1] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Second Edition, University of Illinois at Urbana-Champaign, 2006.
- [2] Arun K. Pujari, "Data Mining Techniques", Computer Science, Universities Press (India) Private Limited, 2001.
- [3] Francesco Bonchi, "Influence Propagation in Social Networks: A Data Mining Perspective", IEEE Intelligent Informatics Bulletin, Vol.12 No.1, December 2011.
- [4] Miguel Rios and Jimmy Lin, "Distilling Massive Amounts of Data into Simple Visualizations: Twitter Case Studies", AAAI Technical Report WS-12-03 Social Media Visualization, 2010.
- [5] Albert Bifet, Geo Holmes, Bernhard Pfahringer, Ricard Gavald, "Detecting Sentiment Change in Twitter Streaming Data", JMLR: Workshop and Conference Proceedings 17, 5-11, 2011.
- [6] David Jensen and Jennifer Neville, "Data Mining in Social Networks", Knowledge Discovery Laboratory, Computer Science Department, University of Massachusetts, Amherst, MA 01003, 2003.
- [7] Ashish Thusoo, Dhruba Borthakur, Raghobam Murthy, Zheng Shao, Namit Jain, Hao Liu, Suresh Anthony and Joydeep Sen Sarma, "Data Warehousing and Analytics Infrastructure at Facebook", Indianapolis, Indiana, USA, 2010.
- [8] Pritam Gundecha, Huan Liu, "Mining Social Media: A Brief Introduction", Arizona State University, Tempe, Arizona 85287, INFORMS ISBN 978-0-9843378-3-5, 2012.
- [9] Jimmy Lin and Dmitriy Ryaboy, "Scaling Big Data Mining Infrastructure: The Twitter Experience", SIGKDD Explorations, Volume 14, Issue 2, 2012.