Mobile WEKA as Data Mining Tool on Android

Pengfei Liu[1], Yanhua Chen[1], Wulei Tang[2], and Qiang Yue[3]

[1] College of Science, South China Agricultural University, Guangzhou, 510642, China
[2] SoftPark of Guangdong, Guangzhou, 510663, China
[3] Guangdong Nortel Telecommunications Co. Ltd, 510665, China
pfliu@scau.edu.cn, {chen_yanhua123,atota}@126.com,
stevenyue@gdnt.com.cn

Abstract. Mobile data mining is an exciting research area that aims at finding interesting patterns from datasets on mobile platform. Limited to the computing power and operating system of traditional mobile devices, mobile data mining lacks attention before. Nowadays mobile devices have a stronger and stronger computation power also the advanced operating system supporting the demand of data mining anywhere and anytime. This paper presents and implements a Java based framework to extend data mining tool Weka to mobile platform. It provides a friendly graphic user interface and simplifies the classification, clustering and associate rule mining functions on android platforms. As an example of usage, we test the model on some datasets and illustrate the feasibility of the proposed approach. A Java implementation of the model demonstrated in this article is available from mobileWeka project website. http://mobileweka.googlecode.com/files/MobileWeka.zip.

Keywords: Mobile, Data mining, Machine learning.

1 Introduction

Mobile computing increasingly present in recent years with stronger and stronger computation power however being ignored, but with right computing models it could be used to solve practical problems such as data mining. At present we focus on the Android platform running on Linux kernel due to its advantages such as open source and free, that make it the best choice for mobile devices.

Data mining plays an important role in many research areas and is mostly done on computers in the past. Supposing that one day every mobile device can easily connect to other devices including data generating devices and can exchange data freely via Bluetooth or WIFI, and then we can process data just by mobile devices. This idea of data mining anywhere and anytime is very amazing but not crazy.

Most of the popular data mining models are designed to work on computers and they often do not pay attention to mobile computing environments. In support of mobile data mining, we proposed an extensible model, which does not need to send local data to other workstations or servers and receive results from them thereby not causing extra communication cost.

The goal of this work is to provide a user-friendly model to simplify data mining on mobile devices, based on an extensible design for enabling adding new data

A. Xie & X. Huang (Eds.): Advances in Electrical Engineering and Automation, AISC 139, pp. 75–80.
springerlink.com © Springer-Verlag Berlin Heidelberg 2012

mining algorithms. These algorithms can be of any kind, including classical ones such as Bayes, Neural Network and K-Means, as the cases we will present.

2 State of the Art

Mobile computing enables the processing with a more convenient way than traditional desktop processor architectures. At present there are many interesting applications in mobile devices. Many of them are task-specific and for research purpose, such as measuring generalization of visuomotor perturbations in wrist movements using mobile phones [1]; Shih G *et al.* proposed the possibility of imaging informatics on Android or IPhone platforms [2].

Data mining on mobile platform is developing in recent years. Several data mining tools on mobile platform have been proposed. For example, The AnalyticDroid from Togaware is Rattle for Android essentially [3], that is an experimental application for controlling analytics from a mobile device using an R Server.

One client programs on mobile devices can invoke remote execution of data mining tasks and show analysis results locally [4]. A distributed and mobile algorithm for global association rule mining was proved to work well for large problem sizes [5].

A model performs minor data analysis and summary early before the source data arrives to the data mining machine. It aims to reducing the amount of further processing required in order to perform data mining [6]. In addition, there is another Java-based project, being the main reference Weka-for-Android (https://github.com/rjmarsan/Weka-for-Android), designed to be a port of weka 3 to the Android platform; the project is to get GUI components removed from Weka project to make it work on Android.

On traditional desktop computers, Weka is very popular and is a landmark in data mining and machine learning [7]. It is widely admitted within academia users as a useful tool for data mining research, especially for bioinformatics. However, it lacks the mobile platform version and is not applied to the mobile data mining domain, which is worth devoting time.

We think it is reasonable now to localize mobile data mining with a user friendly graphic user interface while mobile devices have a stronger and stronger computation power; also there are a lot of minor or medium size datasets, and the mobile processors can tackle them without taking too much time.

In this way, the mobile data mining would be a very attractive technique to increase the flexibility of data mining, especially for bioinformatics researchers.

3 Framework

This section dives in the functionality and design of the proposed framework, called mobileWeka. It is designed to easily launch data mining on mobile device, and be capable of seamlessly employ all of the algorithms embedded in Weka.

MobileWeka should support standard data mining tasks, e.g. clustering, classification and association rule mining. The main issue to solve in developing course is to

abstract basic control flow of data mining, and providing users with graphic interface that having some key features: simple and extensible.

The framework composes of four layers. The first one is Android OS; the second one is Android Java SDK that depends on the first one; the third one is Weka which is in charges of bottom data mining algorithms. The last one is the presentation layer, which provides a friendly user interface and control logic. Fig. 1 presents an illustration of the proposed system architecture.

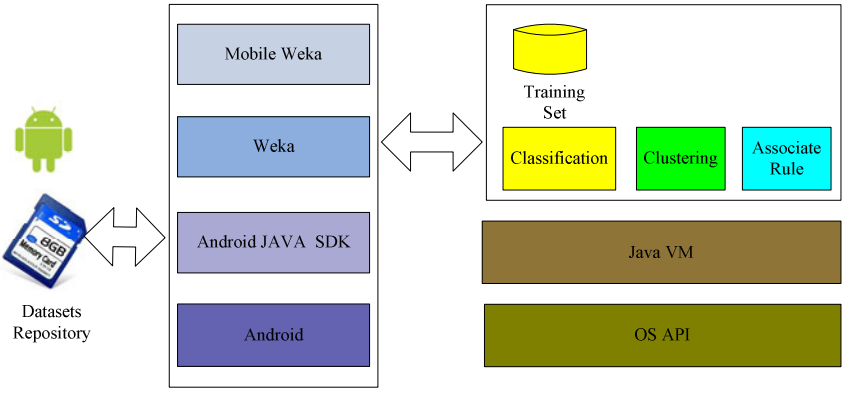


Fig. 1. Architecture of the proposed mobile data mining model

MobileWeka is written in Java since it is supported by Android SDK also platform independent. The developing environment is Eclipse IDE for Java Developers, with Java SDK 1.6 and android SDK 1.5. What we mainly to do list as following.

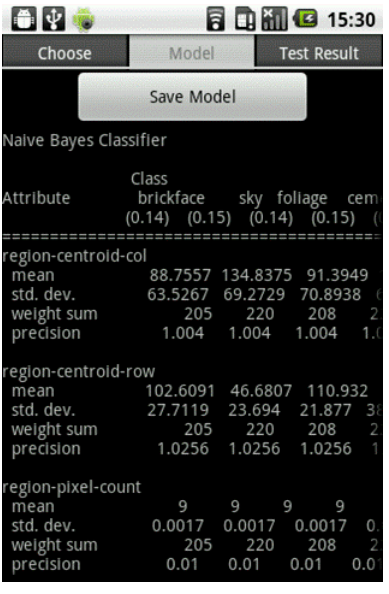
(i). Write XML configuration files according to the design of GUI.
(ii). Fulfill the file chooser.
(iii). Import Weka package to mobileWeka.
(iv). Enroll Weka's embedded algorithms to mobileWeka.
(v). Design suitable processing logic of data mining.

4 Result

As we have expressed, mobileWeka is the mobile version of Weka and the main differences between mobileWeka and Weka are the running efficiency and user interface. So this section mainly presents efficiency test results between both of them.

To fully test mobileWeka, some datasets contained in WEKA are introduced. The experimental mobile device is the HTC Dream mobile phone (also known as TMobile G1), which was the first phone to the market using the Android OS. The processor of G1 is ARM-based MSM7201A with 528 Mhz speed and a 192MB RAM memory; while computer's processor is Intel® Core™2 Duo T6670 with 2.20 GHz.

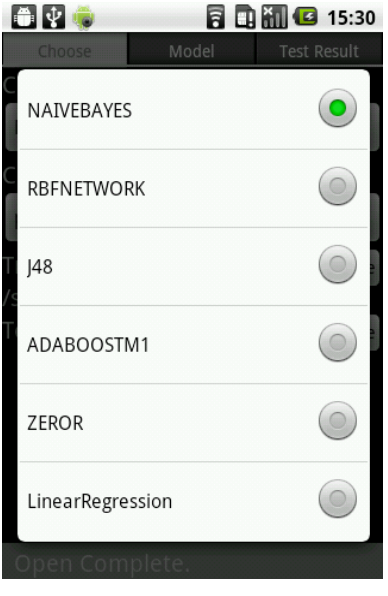
During test procedure, we take some screenshots of mobileWeka. Figure 2 shows the classifier model option setting, algorithms selection, model parameters setting and result evaluation. More screenshots such as clustering and association rule mining can refer to supplementary data.



(a)

(b)

(c)



(d)

Fig. 2. (a): Classifier models evaluation result. 2 (b): Classifier model option setting. 2(c): Classifier model parameters setting. 2(d): Classifier algorithm selection.

The test datasets of bank-data, vote and segment-test have the sizes of 37k bytes, 40k bytes and 108k bytes respectively. During the classification tests, we choose the 66% percentage split method as default standard. Every experiment runs on G1 (Mobile-Weka) and computer (Weka) both. The clustering results are shown in table 1 while the results of classification (sum of training time and test time, in seconds) of are shown in table 2 and the results of association rule mining are shown in table 3. More test results can refer to the supplementary data.

It can be seen that computing time of MobileWeka varied from about 0.01 second to hundreds of seconds, depending on the mining algorithms and the sizes of the data. Some test results have the computing time of 0, because the processing time on computer is too short to record.

The results indicates that mobileWeka uses more computing time than desktop computer but still falls within normal range in most cases. As the sizes of datasets increase, mobileWeka's processing time grows dramatically.

Without doubts the computing models on mobile platforms have lower performance than in computers. Nevertheless, they have great room of improvement with the rapid development of mobile technology.

Table 1. Result of clustering

Algorithm	Dataset	MobileWeka	Weka
SimpleKMeans	bank-data	8.8	0
EM	bank-data	unfinished	49
FarthestFirst	bank-data	1.23	0
DBScan	bank-data	410	1

Table 2. Result of classification

Algorithm	DataSet	MobileWeka	Weka
NAIVEBAYES	vote	1.36	0.01
RBFNETWORK	vote	16.41	0.07
J48	vote	1.89	0.02
ADABOOSTM1	vote	6.02	0.02
ZEROR	vote	0.28	0
NAIVEBAYES	segment-test	34.76	0.03
RBFNETWORK	segment-test	1300.45	6.29
J48	segment-test	60.2	0.11
ADABOOSTM1	segment-test	11.49	0.02
ZEROR	segment-test	1.56	0.01

Table 3. Result of association rule mining

Algorithm	Dataset	MobileWeka	Weka
Apriori	vote	22.76	0
FilteredAssociator	vote	15.8	0
FPGrowth	vote	13.6	0

5 Conclusion

We have presented a mobile data mining model named mobileWeka and demonstrated the feasibility of data mining on mobile devices. MobileWeka is easy to use also works well when datasets are small size or medium size; On the other hand, it does sacrifice execution time while working on large datasets due to low computing capability.

Mobile data mining is not being used as often as would be expectable, taking into account that mobile devices' capabilities. However, this situation will change with the rapid development of mobile hardware in short years, and then mobile data mining will enjoy widespread acceptance in both academia and business.

As future developments of the proposed model we are studying how to make the model more useful and high efficiency, also completing other functions such as data editing and graph presentation.

Furthermore, MobileWeka can be transplanted to any other mobile platform that supporting Java environment. We plan to release this work as an open source project in future.

Acknowledgements. We would like to express our sincere thanks to all the people who have devoted their time to Weka project. Without their effort, it would be impossible for us to finish our study and this paper.

Funding: Supported by Scientific Research Foundation of South China Agricultural University (4900-k11045) and Natural Science Foundation of Guangdong Province, China (S2011040004387).

References

1. Fernandes, H.L., Albert, M.V., Kording, K.P.: Measuring Generalization of Visuomotor Perturbations in Wrist Movements Using Mobile Phones. Plos ONE 6(5), e20290 (2011)
2. Shih, G., Lakhani, P., Nagy, P.: Is Android or iPhone the Platform for Innovation in Imaging Informatics. Journal of Digital Imaging 23(1), 2–7 (2010)
3. Williams, G.: Rattle: A Data Mining GUI for R. The R Journal 1, 45–55 (2009)
4. Talia, D., Trunfio, P.: Mobile Data Mining on Small Devices through Web Services. In: Mobile Intelligence, John Wiley & Sons, Inc., NJ (2010)
5. Wang, F., Helian, N., Guo, Y., Jin, H.: A distributed and mobile data mining system. Parallel and Distributed Computing, Applications and Technologies 27-29, 916–918 (2003)
6. Goh, J.Y., Taniar, D.: An Efficient Mobile Data Mining Model. In: Cao, J., Yang, L.T., Guo, M., Lau, F. (eds.) ISPA 2004. LNCS, vol. 3358, pp. 54–58. Springer, Heidelberg (2004)
7. Frank, E., Hall, M., Trigg, L., Holmes, G., Witten, I.H.: Data mining in bioinformatics using Weka. Bioinformatics 20(15), 2479–2481 (2004)