

MobileMiner: A Real World Case Study of Data Mining in Mobile Communication*

Tengjiao Wang[†], Bishan Yang[†], Jun Gao[†], Dongqing Yang[†],
Shiwei Tang[†], Haoyu Wu[†], Kedong Liu[†], Jian Pei[‡]

[†]Key Laboratory of High Confidence Software Technologies (Peking University),
Ministry of Education, China

[†]School of Electronics Engineering and Computer Science, Peking University, Beijing, 100871 China

[‡]School of Computing Science, Simon Fraser University, Canada

{tjwang, bishan_yang, gaojun, dqyang, tsw, why, kdliu}@pku.edu.cn, jpei@cs.sfu.ca

ABSTRACT

Mobile communication data analysis has been often used as a background application to motivate many data mining problems. However, very few data mining researchers have a chance to see a working data mining system on real mobile communication data. In this demo, we showcase our new system MOBILEMINER on a real mobile communication data set, which presents a case study of business solutions using state-of-the-art data mining techniques. MOBILEMINER adaptively profiles users' behavior from their calling and moving record streams. Customer segmentation and social community analysis can be conducted based on user profiles. We show how data mining techniques can help in mobile communication data analysis. Moreover, we also show some interesting observations which still cannot be mined by the current techniques, and thus may motivate new research and development.

Categories and Subject Descriptors

H.2.8 [Database Applications]: Data Mining

General Terms

Algorithms, Performance

Keywords

Mobile communication data, data stream mining, sequence clustering, community discovery

*This work is supported by the National High Technology Research and Development Program of China('863' Program)(No.2007AA01Z191,2009AA01Z150), the Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China(No.708001), and the joint project of China Mobile Communication Corporation and Peking University. Jian Pei's research is supported in part by an NSERC Discovery grant and an NSERC Discovery Accelerator Supplements grant. All opinions, findings, conclusions and recommendations in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

Copyright is held by the author/owner(s).
SIGMOD'09, June 29–July 2, 2009, Providence, Rhode Island, USA.
ACM 978-1-60558-551-2/09/06.

1. APPLICATION BACKGROUND

Mobile communication data analysis has been often used as a background application to motivate many technical problems in data mining research, such as mining frequent patterns and clusters on data streams, social network analysis, collaborative filtering and recommendation. However, very few data mining researchers have a chance to see a working data mining system on real mobile communication data. The lack of this experience prevents those researchers from deeply understanding the business application scenarios in mobile communication as well as the successes and the limitations of the existing techniques.

We are developing MOBILEMINER, a data mining tool for mobile data analysis and business strategy development. Built on the state-of-the-art data mining techniques, MOBILEMINER presents a real case study on how to integrate data mining techniques into a business solution.

In a large mobile communication company like China Mobile Communication Corporation, there are many analytical tasks where data mining can help to address the business interests of the company. Clearly, a system cannot cover all aspects. MOBILEMINER starts with customer relation management, the core component of mobile communication business. In this demo, we focus on two tasks, mobile user segmentation and community discovery from user calling networks.

MOBILEMINER provides a platform for the analytical tasks, where user profiles are extracted continuously from users' moving and calling records. The profiles are extremely important and valuable in business. Based on the profile mining platform, various data mining tasks can be effectively performed using different features of the profiles.

The mobile user segmentation task tries to group customers by their frequent moving patterns. The features used for grouping are obtained by mining users' moving records continuously on the profile mining platform. Knowing the moving patterns for different customer groups, a service provider can dynamically deploy resources to improve the service quality (e.g., adjusting the angles of antennas or re-positioning a mobile station). For example, in Beijing Olympic period, many people are moving from Bird Nest around 9pm. to Olympic Village around 11pm. It is interesting to find the clusters of customers in terms of service areas and time.

The community discovery task aims to discover coherent calling communities. Based on the profile mining platform,

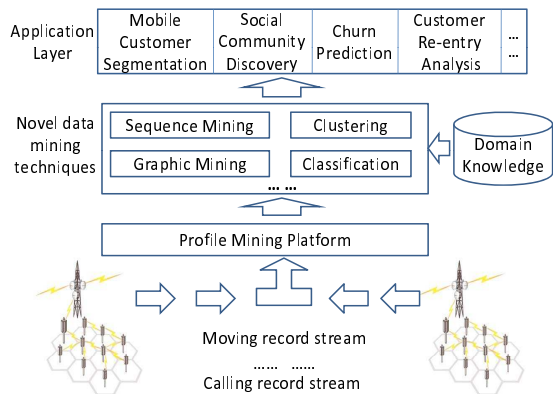


Figure 1: The architecture of MobileMiner.

a social network can be constructed using calls between customers and the calling frequencies. Communities in the social network capture the connectivity and similarity among customers. By considering the properties of the communities, effective market campaign can be designed for targeted customers. For example, the customers with broad social connections should be taken care specially.

We emphasize the following points in our demo. First, we present how we solve the business tasks in mobile communication using novel data mining techniques. Second, we use MOBILEMINER on real data to elaborate what can be done and how the data mining techniques can be integrated in a business-driven model. Third, we show some examples of what still cannot be done satisfactorily using the current data mining techniques, which may motivate future research and development.

2. TECHNOLOGY AND NOVELTY

Figure 1 shows the architecture of MOBILEMINER. Customer records are collected by the mobile communication base stations and fed into MOBILEMINER as data streams, including customer moving trajectories and calling records. A base station serves the cell phones in a specific region, and can detect a mobile customer once she turns on her phone. Once the records are imported into the system, profile mining is performed to generate user profiles for the upper layer data mining tasks.

Specifically, in the profile mining part, customers' moving profiles or their frequent moving patterns are constructed based on their moving records continuously. The core of this task is to mine sequential patterns on data streams, which is challenging since there can be many customers and the sliding window can be large. A customer's moving profile is formed using the set of closed sequential patterns that match the customer's trajectory and the profile is incrementally maintained. We developed a novel algorithm [1] to mine and incrementally maintain on fast data streams closed sequential patterns, which are non-redundant representation of sequential patterns. An effective data structure is designed to keep close sequential patterns in memory and various strategies are proposed to prune search space aggressively. Based on the experiments on both real and synthetic databases, our algorithm outperforms the best existing al-

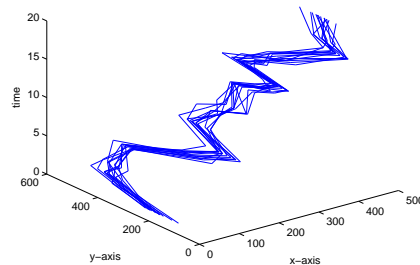


Figure 2: A bicluster in 3D (x, y, time).

gorithms by a large margin. The details of the techniques can be found in [1].

The mobile user segmentation module clusters customers according to their profiles. The goal is to partition customers into groups such that the customers in a group are similar to each other in moving patterns. Importantly, timestamps should be considered. Since each point in a customer trajectory is associated with a timestamp, two trajectories are similar only if they are close to each other in time dimension. The problem is formulated as clustering trajectories in both space and time. The spatio-temporal patterns of clusters are very useful for the company to allocate base stations for specific customer groups. Some related work (e.g., [3]) clusters spatio-temporal patterns in bioinformatics. Here, we adapt the algorithm in [2] to group 2-dimensional trajectories in different time stamps. The main idea is to find biclusters with low mean squared residue through effectively iterative search. The mean squared residue captures the variance of the set of trajectories in a bicluster over time. For example, Figure 2 shows a cluster discovered by the algorithm, where the grouped 2D location trajectories of 13 customers are plotted at 19 consecutive time points.

In mobile communication business, the social relationship among customers often plays a significant role in marketing. For example, losing some customers with broad social connections may cause customer churning. A social network among customers is constructed. Each customer is represented by a node in the network. An edge is drawn to connect two customers if they call each other over a certain number of times in the current sliding window. A social community in the network is a set of nodes such that they are relatively well connected to each other and much less connected to the other nodes in the network. Some previous work (e.g., [4]) discovers communities in a network. In this application, the connection weights on edges in graphs should be considered. We extend the algorithm in [5] to discover communities in the weighted graph in two steps. First, we generate a core set and then expand the core set with affiliated customers. The core set is a set of customers whom are frequently called by other customers. The affiliated customers are the customers surrounding the core with different layers. We use the calling frequencies as the weights in the process of finding core customers and ranking affiliated customers. To control the granularity of the discovering communities, a merging schema is used to merge similar communities to get coarser results.

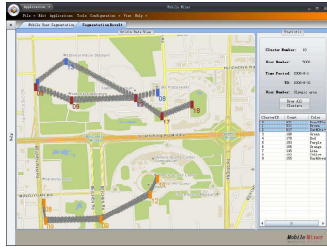


Figure 3: The user interface in the mobile user segmentation module.

3. DEMO SCENARIO

Our demo consists of three parts. First, we showcase how we integrate the state-of-the-art data mining techniques into a business framework and building MOBILEMINER as a business solution. Second, we illustrate how the underlying data mining techniques affect business analysis. Last, we present some interesting observations found from real data which unfortunately still cannot be handled by the existing techniques. Such case studies may motivate novel data mining research and development.

3.1 Techniques Meeting Business Requirements

We will demonstrate some common business analysis tasks in mobile communication companies, including customer segmentation for mobile service bases deployment, and calling community discovery for marketing campaign design.

For example, the user interface of the mobile user segmentation module (Figure 3) is not a simple list of the users grouped in each cluster. MOBILEMINER visualizes the user groups by showing their moving patterns, each group in a different color. Moreover, the moving patterns are shown in temporal order with a local map as the background. With this information, analysts can make informative decisions about how to deploy mobile base stations more effectively.

We will also show how calling community discovery techniques help companies to design marketing campaign. The graph mining results are presented properly in a business driven way. Based on the discovered knowledge, business analysts can identify targeted customers in an effective way.

3.2 Sharpening Business Analysis by Tuning Techniques

Data mining techniques need to be tuned to make business analysis effective. To understand how well the data mining techniques in MOBILEMINER work in practice, we use a real mobile communication data set to show some interesting mining results.

To demonstrate the tuning needs, we will show how the parameters of our sequential pattern mining algorithms may affect the mining results. For example, Figure 4 shows that different size of sliding windows result in different modeling time, and Figure 5 shows the effect of support threshold value on the number of discovering patterns.

Moreover, it is important that the user interface can help business analysts to tune the underlying data mining methods. For example, MOBILEMINER provides an interface for analyzing the social communities found from the social network (Figure 6). A business analyst can interact with the

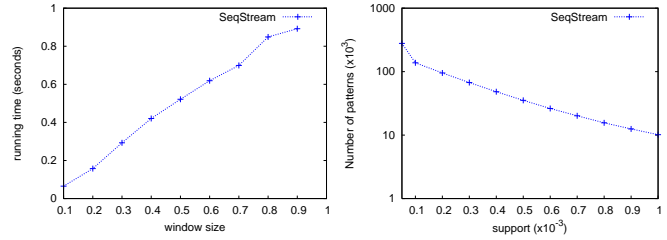


Figure 4: Effect of Time Window Size on Mining Time

Figure 5: Effect of Support Threshold on Number of Patterns

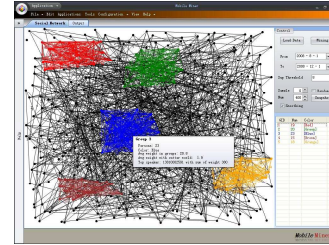


Figure 6: The user interface in the calling community discovery module.

social community visualization to tune the parameters of the social network construction such as the call frequency threshold and the time window.

3.3 Opportunities for Future Research

Mobile communication is a fast growing industry. We demonstrate some patterns found yet from real data by human analysts but cannot be found using the data mining techniques. For example, a new service of low calling charge by the company may negatively effect the sales of another service such as monthly SMS. It is critical to analyze whether such a new service overall improves the business and thus whether it should be introduced. Usually, this decision is based on the experiential analysis on both potential profits and potential customers. This task can be modeled as a hypothesis mining problem, which is highly demanded in business but has not been systematically studied in a practical setting.

4. REFERENCES

- [1] L. Chang, T. Wang, D. Yang, and H. Luan. Seqstream: Mining closed sequential patterns over stream sliding windows. In *Proceeding of ICDM'08*, Pisa, Italy, December 2008.
- [2] Y. Cheng and G. M. Church. Biclustering of expression data. In *Proceedings of ISMB'00. Menlo Park, USA: AAAI*, pages 93–108, 2000.
- [3] D. Jiang, J. Pei, M. Ramanathan, C. Lin, C. Tang, and A. Zhang. Mining gene-sample-time microarray data: a coherent gene cluster discovery approach. *Knowledge and Information Systems: An International Journal*, 13:305–335, November 2007.
- [4] J. Pei, D. Jiang, and A. Zhang. On mining cross-graph quasi-cliques. In *Proceedings of KDD'05*, pages 228–238, 2005.
- [5] W. Zhou, J. Wen, W. Ma, and H. Zhang. A concentric-circle model for community mining in graph structures. In *Microsoft Research, Seattle, Technical. Report MSR-TR-2002-123*, 2002.