

学校编码: 10384
学 号: 23020111153041

分类号 _____ 密级 _____
UDC _____

厦 门 大 学

硕 士 学 位 论 文

基于 Gn 信令的移动网络用户行为研究

Research of Mobile Customers' Behavior Based on

Gn Interface Message

陈 子 弘

指导教师姓名: 黄晓阳 助理教授

郑灵翔 高级工程师

洪学敏 副教授

专 业 名 称: 计 算 机 技 术

论文提交日期: 2014 年 4 月

论文答辩时间: 2014 年 5 月

学位授予日期: 2014 年 6 月

答辩委员会主席: 

评 阅 人: _____

2014 年 5 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（）课题（组）的研究成果，获得（）课题（组）经费或实验室的资助，在（）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）： 陈子弘

2014年5月16日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：陈子弘

2014年5月16日

摘要

近年来，移动终端的普遍程度逐渐超过 PC。移动互联网的浪潮掀起了运营商大数据时代的序幕。一方面由于 OTT 的兴起使得通信网络逐渐被管道化；另一方面，电信行业内同质化竞争严重，以往简单的粗放式经营已不再适用。对管道资源的争夺是保障用户发展的关键，运营商期待通过对通信网络中闲置的海量用户数据进行建模分析，此举有助于运营商建立智能化的通信管道，以差异化经营在行业竞争中谋求发展。相比移动通信网络中服务器上的数据，Gn 信令含有较为丰富的信息，能准确记录用户使用网络场景，使得用户建模结果更加精确。

本文首先介绍了国内外的网络用户建模研究现状，并对现有的用户建模工作按照基于移动终端、内容服务提供商和通信网络运营商三个类别进行综述。之后，本文采用 Gn 信令对用户流量使用场景进行建模，以挖掘出不同场景下的用户流量使用模式。

本文主要工作分为以下两部分：第一部分为海量数据挖掘平台的设计；第二部分为基于 Gn 接口信令的用户流量信息聚类分析。

首先，本文针对海量用户数据环境，设计了基于 Hadoop 2.x 生态环境的系统平台，平台中集成了 Hive 和 Mahout 用于挖掘分析。在数据存储模块上，本文分析对比原有 HDFS 以及常见 DFS 在架构上的缺陷，采用 Ceph 系统来替代原有的 HDFS 以提供高性能的数据存储服务。实验结果表明 Ceph 系统具有线性的可扩展性和良好的高性能。

其次，本文对采集的 Gn 接口信令进行预处理，从流量层次和移动性两个角度来将一个用户流量状态描述成一个六维的概率分布向量。由于流量数据关系和用户类别特点尚不明确，因而本文采用基于 Canopy 预处理的 K-Means 算法从不分时段和分六个时段两方面对移动用户群体的流量信息进行聚类分析。

最后，从统计分析结果中，本文发现第三时段和第四、五时段，以及第二时段和第六时段的用户流量使用情况具有很高的相关性，说明用户在第三时段到第五时段，以及第二时段和第六时段具有相似的流量使用模式。

从聚类结果分析中，本文结合相关业务经验，发掘出不同用户在不同时段下

对于手机流量的使用模式，并定位出流量营销的优质和潜在客户群体。

关键词：Gn 接口信令；客户细分；流量使用模式

厦门大学博硕士论文摘要库

Abstract

The dawn of mobile internet starts the information explosion age. The emergence of massive data brings huge challenge to storage and data mining. Nowadays, mobile internet is popularized with the rapid development of kinds of mobile terminals and the number of mobile users is increasing with a much faster rate than that of traditional internet.

In the meantime, the mobile industry is amidst a transformation. As market segmentation has become a new driving force to promote the growth of mobile users, management philosophy of user-centered asks for data support. The massive user data is rich in user related information that is very important for market segmentation and business promoting.

Compared with other interfaces of GPRS network, Gn interface message provides much richer information that records mobile users' business scenarios more accurately. Thus, picking up network data through Gn interface message contributes to mine mobile user behavior characters.

This paper will analyze about mobile users' behavior through Gn interface message, and mine the traffic flow floating model in different time periods and scenarios. A data mining platform based on Hadoop 2.x environment that integrated with hive and mahout is designed for experiments. Ceph storage system is used as a drop-in replacement in this platform to provide high ability and linear scalability of storage and data mining.

In user modelling, each user vector is preprocessed and integrated with other heterogeneous data sources, then mapping to flow floating probability of six states. A clustering method -- K-Means which is preceeded by Canopy is used to mine flow floating models and analyze users' behavior in different time periods and scenarios.

Finally, we sum up some features about mobile flow floating model with the

clustering results and experience in business promoting, then target the customer group for flow promoting.

Key Words: Gn Interface Message; Market Segmentation; Traffic Flow Floating Model;

厦门大学博硕士学位论文摘要库

目录

摘要	I
Abstract	III
目录	V
Contents	IX
第 1 章 绪论	1
1.1 背景和意义	1
1.2 国内外研究现状	2
1.2.1 用户建模的研究现状	2
1.2.2 现阶段用户建模存在的问题	3
1.2.3 基于 Gn 信令的研究和意义	4
1.3 本文主要工作及组织结构	6
1.3.1 本文主要工作	6
1.3.2 本文组织结构	6
第 2 章 网络用户建模综述	9
2.1 GPRS 网络介绍	9
2.1.1 GPRS 网络架构	9
2.1.2 Gn 接口信令	11
2.2 通信行业用户建模现状	11
2.3 用户数据特征	13
2.4 用户建模的数据来源	15
2.4.1 OTT 服务器的数据	16
2.4.2 终端设备的数据	17
2.4.3 通信运营商的数据	17
2.5 用户建模方法和应用场景	18

2.5.1 用户数据建模方法	18
2.5.2 用户建模应用场景	19
2.6 运营商建模的优势和机遇	19
第 3 章 数据挖掘平台设计	21
3.1 数据存储现状概述	21
3.2 分布式文件系统	22
3.3 MR 并行计算模型	23
3.4 平台设计	25
3.4.1 数据存储模块	26
3.4.2 数据处理框架	31
3.5 本章小结	32
第 4 章 基于 Gn 信令的移动用户流量分析	33
4.1 本章概述	33
4.2 移动用户向量描述	34
4.3 数据预处理	35
4.4 聚类方法选择	37
4.5 个体差异性度量	45
4.6 基于流量和位置的流量信息聚类	46
4.6.1 用户向量数据集	46
4.6.2 聚类结果及分析	49
4.7 分时段用户流量信息聚类	52
4.7.1 第一时段聚类分析	53
4.7.2 第二时段聚类分析	55
4.7.3 第三时段聚类分析	57
4.7.4 第四时段聚类分析	59
4.7.5 第五时段聚类分析	61
4.7.6 第六时段聚类分析	64
4.7.7 分时段流量使用模式分析	66
4.8 本章小结	67

第 5 章 结论和未来展望	69
5.1 全文总结	69
5.2 研究展望	70
参考文献	71
致谢	75
攻读硕士学位期间发表的论文及所做工作.....	77

厦门大学博硕士论文摘要库

Contents

Abstract in Chinese	I
Abstract in English	III
Contents in Chinese	V
Contents in English	IX
Chapter 1 Introduction	1
1.1 Background	1
1.2 Review of research	2
1.2.1 Research of User Modelling	2
1.2.2 Main Currently Problem of User Modelling	3
1.2.3 User modelling based on Gn interface.....	4
1.3 Main Contents and structure	6
1.3.1 Main Contents.....	6
1.3.2 Dissertation Structure	6
Chapter 2 Survey of User Modelling in Network	9
2.1 Introduction to GPRS Core Network	9
2.1.1 GPRS Network Architecture	9
2.1.2 Gn Interface Message	11
2.2 User modelling in Telecom Industry	11
2.3 Characteristics of User Data	13
2.4 Data Source of User Modelling	15
2.4.1 Data from OTT Servers.....	16
2.4.2 Data from End-user Devices.....	17
2.4.3 Data from Mobile Telecom Carriers	17
2.5 Modelling Methods and Scenarios	18
2.5.1 Modelling Methods.....	18
2.5.2 Modelling Scenarios	19

2.6 Advantages of User Modelling to Mobile Telecom Carriers	19
Chapter 3 Design of Data Mining System	21
3.1 Survey of Data Storage	21
3.2 Introduction to Distributed File System.....	22
3.3 Parallel Programming Model	23
3.4 Framework of Data Mining System	25
3.4.1 Storage Module	26
3.4.2 Processing Module	31
3.5 Summary	32
Chapter 4 Analysis of Flow Based on Gn Interface Message	33
4.1 Introduction to Data Mining	33
4.2 Description of Mobile User Data.....	34
4.3 Data Preprocessing	35
4.4 Clustering Methods	37
4.5 User Measurements.....	45
4.6 User Clustering Based on information of Flow and Location.....	46
4.6.1 Nature of User Data.....	46
4.6.2 Analysis of Clustering Results	49
4.7 User Clustering on Six time periods	52
4.7.1 Clustering on First period.....	53
4.7.2 Clustering on Second period	55
4.7.3 Clustering on Third period	57
4.7.4 Clustering on Fourth period	60
4.7.5 Clustering on Fifth period.....	63
4.7.6 Clustering on Sixth period.....	65
4.7.7 Analysis of Flow mode	67
4.8 Summary	68
Chapter 5 Conclusion and Prospect	69
5.1 Conclusion	69
5.2 Prospect	70

References	71
Acknowledgement	75
Published Paper and Research During Pursuing Master Degree.....	77

厦门大学博硕士学位论文摘要库

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库