



# 国际漫游通信产品用户识别模型及实证分析

申志伟, 崔瑞媛, 吕恒

(中国联合网络通信有限公司研究院, 北京 100032)

**摘要:**针对国际漫游通信产品使用率低的问题,基于移动用户在通信行为和属性上的全量数据,提出统计分析和建立识别模型相结合的方法,运用 SPSS 和 Clementine 工具,以高命中率和广覆盖率为目标,提炼出国际漫游通信产品用户的潜在特征,对这些潜在特征逐一进行了实证对比分析,得出了非常高的精准识别率,这对运营商更有针对性地开展国际漫游业务具有重要参考意义。

**关键词:**国际漫游;用户行为特征;识别模型;实证分析

中图分类号:TN929.5

文献标识码:A

doi: 10.11959/j.issn.1000-0801.2016022

## User identification model and empirical analysis of international roaming communication products

SHEN Zhiwei, CUI Ruiyuan, LV Heng

China Unicom Research Institute, Beijing 100032, China

**Abstract:** For the use rate of international roaming communications products are too low, based on full volume data of mobile user in communication behavior and attribute, a method of combining statistical analysis and building recognition model was proposed. Using SPSS and Clementine tools, with high hit rate and wide coverage as the target, the user potential features of the international roaming communications products were extracted. Subsequently, the empirical analysis of the potential features also was given, and the very high accuracy rate was obtained. This has important reference value for operators to carry out international roaming service better.

**Key words:** international roaming, user behavior characteristics, identification model, empirical analysis

### 1 引言

随着经济的不断发展,中国在出境旅游、商务活动、出国留学等出国行为上的人数及消费费用呈现出逐年增加的趋势:在出境游方面,根据金棕榈咨询发布的数据<sup>[1]</sup>,2014年中国出境人数达到1.17亿人次,同比增长18%,出境游花费约1400亿美元;根据国家旅游局的数据<sup>[2]</sup>,仅仅2015年春节期间中国出境游人数就达到了518.2万人次,并且出境人数首次超过境内旅游,2015年

上半年,出境旅游达到6190万人次<sup>[3]</sup>;而根据世界旅游组织的报告<sup>[4]</sup>,自2012年起,中国已经连续3年成为世界最大的出境游市场。在商务出境方面,中国是澳大利亚第一大商务交流来源国<sup>[5]</sup>。在出国留学方面,据国家教育部统计,2014年度我国出国留学人员总数为45.98万人<sup>[6]</sup>。对于电信运营商而言,这些数据意味着潜在的国际漫游用户的规模和频次不断上升,有着广阔的国际漫游市场,从2015年开始,三大运营商虽然都在对国际漫游资费进行大幅下调,开通国际漫游权限的用户比例也在上升,但

是使用国际漫游产品的用户占比仍然极少,通过有效提取国际漫游用户通信行为特征,从而有针对性地推荐国际漫游产品是运营商提升国际收入的重要手段。本文选取运营商本地网3个月的全量移动数据来建立国际漫游通信产品用户识别模型,再用后续一个月的数据来验证模型提取到的特征准确性。

## 2 国际漫游通信产品用户总体分析

从实际数据上来看,使用国际漫游产品的用户几乎都是稳定用户,和稳定用户相比,新用户在较短的时间内使用国际漫游产品的规模基本可以忽略不计,因此本文所选取的用户是剔除三无、公测等无效用户并且在网时长超过3个月的用户。以这些本地网1200多万全量移动用户数据为基础,把移动用户按照2G、3G、4G来划分,同时选取语音、流量、短信、属性等基本数据为维度进行分析。根据3个月的实际数据,虽然开通国际漫游权限和使用国际漫游产品的占比都较低,但是2G、3G、4G用户的占比均呈现递增态势,并且国际漫游收入中以国际漫游语音产品为主,同时有一定的国际漫游流量需求,但是没有国际漫游短信的需求,这意味着移动用户国际漫游产品提升空间较大,一方面需要加快2G向3G、4G用户迁移,3G向4G用户的迁移,另一方面要向移动用户重点推荐国际漫游语音产品,同时也要不断挖掘用户国际漫游流量需求并加大其营销力度。

在开通国际漫游权限同时使用国际漫游产品的用户中,从消费能力上看,3G、4G用户中高端用户是其主要组成部分,占比在90%以上,并且多数都属于稳定用户,有着较长的在网时长,一般超过3年;从使用终端上看,使用3G、4G终端的用户占比也都在90%以上,其中使用苹果、三星高端品牌的终端占比高达80%;从用户的基本属性上看,18~50岁用户占比在85%以上,国际漫游用户呈现年轻化态势,同时男性群体较多。

基于以上数据分析,总体上国际漫游用户呈现出中高端、语音流量偏好、年轻化、稳定等特点,具体而言,2G、3G、4G用户的特征如下。

### (1)2G用户

使用三星、苹果、小米、华为和诺基亚的3G、4G终端;向出账收入在300元之内的用户重点推荐国际语音产品。

### (2)3G用户

3G用户有一定的国际漫游需求,有进一步激发的潜力,

尤其是国际流量需求;重点对有国际长途收入的用户推荐国际漫游产品;使用苹果、三星、华为和小米的3G、4G终端;向3G用户同时推荐国际语音和国际流量产品。

### (3)4G用户

4G用户有一定的国际漫游需求,但占比呈现下降趋势,需要增强国际漫游产品营销力度;重点对有国际长途收入的用户推荐国际漫游产品;使用苹果、三星、华为和小米的3G、4G终端,尤其是4G终端,需要加大4G终端营销力度;向4G用户同时推荐国际语音和国际流量产品。

## 3 国际漫游通信产品用户识别模型

### 3.1 模型说明

根据有国际漫游收入用户通信行为特征,总结归纳用户有国际漫游收入动因,形成判定模型及判别规则;基于此规则,为在移动用户群中推荐国际漫游产品提供模型支撑依据,在原始语音、流量、短信、终端等数据指标基础上,增加均值、标准差、变异系数3个基本衍生指标形成模型所需的数据输入,模型以高覆盖率和命中的高命中率作为目标。

覆盖率 =  $\frac{\text{预测命中的有国际漫游收入用户数}}{\text{实际有国际漫游收入用户数}}$ , 考察模型

全面性。

命中率 =  $\frac{\text{预测命中的有国际漫游收入用户数}}{\text{预测有国际漫游收入用户数}}$ , 考察模型

精确性。

### 3.2 建模过程

对初始数据指标利用SPSS工具<sup>[7]</sup>进行方差分析、相关性分析,剔除相关性强以及对目标变量不显著的数据指标后作为模型的输入文件,在Clementine工具中<sup>[8]</sup>运用决策树C5.0进行建模。图1给出了国际漫游通信产品用户识别模型的建模框架。

### 3.3 特征规则提取

#### (1)3G用户

把用户分为有国际漫游收入和没有国际漫游收入两种情况分别进行建模。在变量重要性方面,图2和图3给出了这两种情况下的指标重要性排序。在有国际漫游收入用户所有指标变量中,标准差国际语音主叫时长、均值国际语音被叫时长、标准差出账总收入和国际漫游总流量最为重要,在没有国际漫游收入用户所有指标中,标准差出账总收入、均值国际长途通话收入和均值出账总收入最重要。

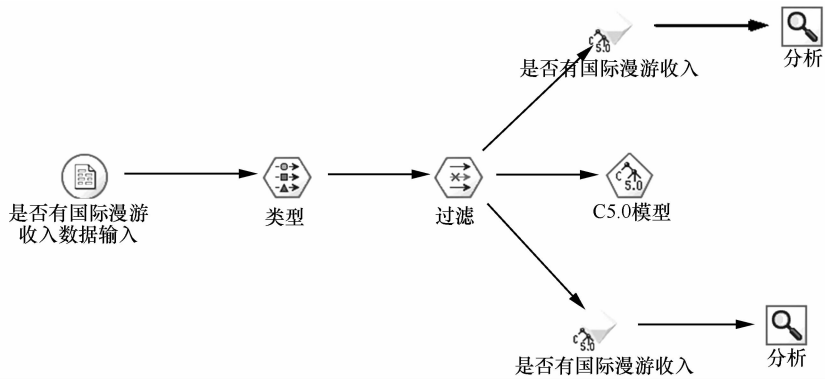


图1 国际漫游通信产品用户识别模型框架

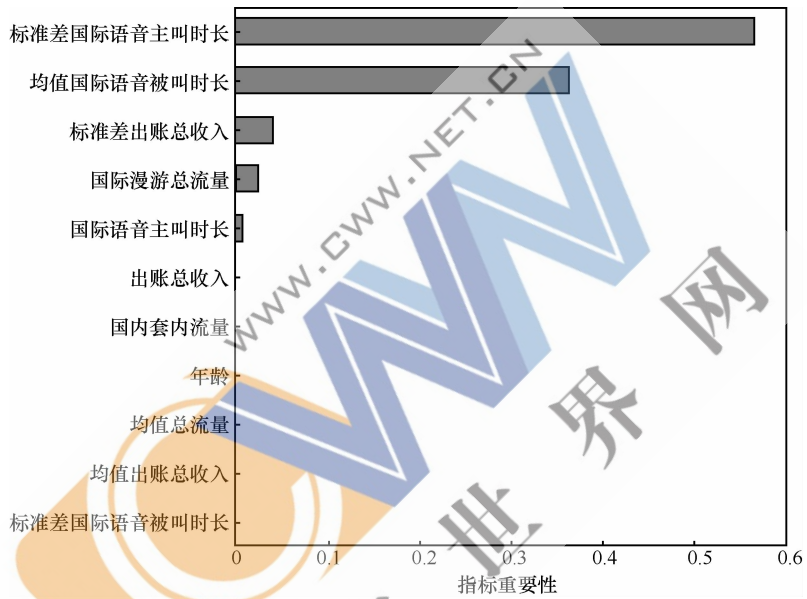


图2 3G有国际漫游收入用户指标重要性

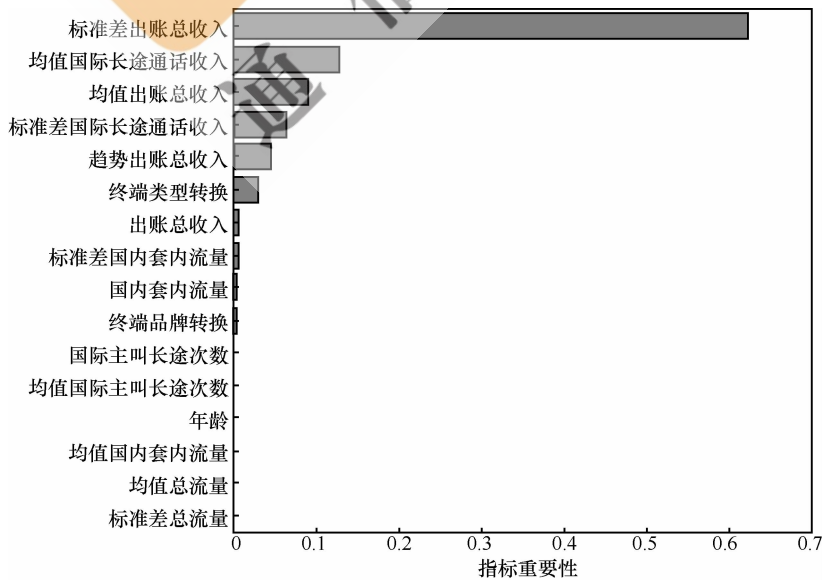


图3 3G没有国际漫游收入用户指标重要性

在有国际漫游收入用户方面,共提取 3 条规则,命中率为 93.7%,覆盖率为 99.1%,具体如下:

- 有任意国际漫游通信行为的用户是国际漫游首要目标用户群体;
- 有国际语音通信行为的用户,出账总收入有较大波动,向其推荐国际漫游产品;
- 有国际语音通信行为的用户,出账总收入较为平稳,有一定的国际语音被叫行为且总流量在 300 MB 以上,向其推荐国际漫游产品。

在没有国际漫游收入用户方面,同样提取 3 条规则,命中率 72.5%,覆盖率为 69.4%,有国际长途语音行为,满足下列条件之一,重点推荐国际漫游产品:

- 出账总收入有较大波动;
- 出账总收入较为平稳,同时有较为波动的国际长途语音行为,且使用 3G、4G 终端;
- 出账总收入和国际长途语音行为较为平稳且年龄在 19~50 岁使用 3G 终端。

### (2)4G 用户

同样把用户分为有国际漫游收入和没有国际漫游收入两种情况分别进行建模。在变量重要性方面,图 4 和图 5 给出了这两种情况下的指标重要性排序。在有国际漫游收入用户所有指标变量中,均值国际语音被叫次数、均值国际语音主叫次数最为重要;在没有国际漫游收入用户所有指标中,标准差出账总收入、标准差国际主叫长途次数和终端类型最重要。

在有国际漫游收入用户方面,共提取 3 条规则,命中率为 99.2%,覆盖率为 99.3%,具体如下:

- 有任意国际漫游通信行为的用户是国际漫游首要

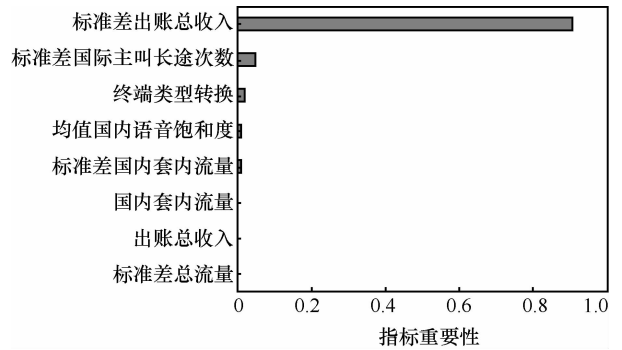


图 5 4G 没有国际漫游收入用户指标重要性

目标用户群体;

- 向有被叫国际语音通信行为的用户推荐国际漫游产品;
- 向无被叫无国际语音通信行为,且有主叫国际语音通信行为的用户推荐国际漫游产品。

在没有国际漫游收入用户方面,同样提取 3 条规则,命中率为 73%,覆盖率为 75.7%,具体如下:

- 向近 3 个月有非常大的出账总收入波动的用户,重点推荐国际漫游产品;
- 向有国际长途通信行为,有较大的出账总收入波动且有国内流量使用行为的用户,重点推荐国际漫游产品;
- 向无国际长途通信行为,出账总收入有较大波动且在 60 元以上,使用 4G 终端有国内流量使用行为的用户,重点推荐国际漫游产品。

### (3)特征总结

综合 3G、4G 特征,可提取以下主要共有特征。

- 有过国际漫游收入历史记录的用户是使用国际漫游产品的首要目标用户。

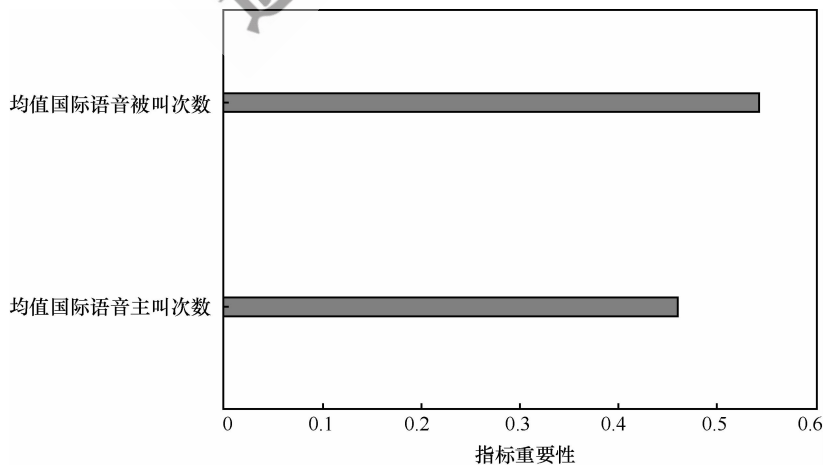


图 4 4G 有国际漫游收入用户指标重要性



- 在没有国际漫游收入历史记录的用户中,满足以下条件建议重点推荐国际漫游产品:有国际长途语音行为;出账总收入较高,在 56 元以上,同时近期有较大波动;流量在 300 MB 以上;使用 3G、4G 终端的男性。

## 4 国际漫游用户实证对比

选取建立国际漫游通信产品用户模型后,后续一个月 1 200 多万本地网全量移动用户数据,对在总体分析和模型中提取的特征进行实证分析,以覆盖率和提升率两个指标来进行验证,这两个验证指标和建立模型时的含义不同,以下为其具体定义。

**覆盖率:**用于验证总体分析中提取的用户特征的准确性,在所需类型(有国际漫游收入用户)中符合特征的用户在其中的占比,占比越大,说明在该类型群体中该特征越理想,覆盖越广。

**提升率:**用于验证国际漫游通信产品用户模型提取的用户特征的准确性,在全量移动用户中提取符合特征的用户,考察这些用户中实际是该类型的用户(全量符合特征用户中实际有国际漫游收入用户)在其中的占比,将这个占比和没有根据这个特征的实际占比进行对比,这两个占比之差为提升率。

### 4.1 总体分析实证

表 1、表 2、表 3 分别为 2G、3G、4G 用户总体分析中提取特征的覆盖率。

表 1 2G 用户总体分析覆盖率

提取特征	覆盖率
使用三星、苹果、小米、华为和诺基亚的 3G、4G 终端	40% (缺失值占比近 60%)
向出账收入在 300 元之内的用户重点推荐国际语音产品	96%

表 2 3G 用户总体分析覆盖率

提取特征	覆盖率
有国际长途收入用户	94%
使用苹果、三星、华为和小米的 3G、4G 终端	80%

表 3 4G 用户总体分析覆盖率

提取特征	覆盖率
有国际长途收入用户	97%
使用苹果、三星、华为和小米的 3G、4G 终端	88%

### 4.2 识别模型实证

表 4 给出了国际漫游通信产品用户识别模型中提取特征的提升率,在其中提取的有国际漫游收入的用户必然会使用国际漫游产品,无需再做验证。

表 4 国际漫游通信产品用户识别模型提升率

提取特征	3G 用户提升率	4G 用户提升率
有国际长途语音行为	3 500%	2 700%
出账总收入较高,在 56 元以上,同时近期有较大波动	40%	70%
流量在 300 MB 以上	130%	130%
使用 3G、4G 终端的男性	40%	30%

### 4.3 实证分析小结

从实际的验证效果上看,本文建立的国际漫游通信产品用户识别模型所提取的特征规则,无论从覆盖广度上还是从精准识别上都有很大的契合度和提升率,向根据这些规则选择的移动用户推荐国际漫游产品会更有针对性,同时节省运营商营销成本,这里需要说明的是模型中提取的规则是根据特定本地网和近期 3 个月的数据进行的,因此属于动态模型,会随着不同本地网和不同时期的数据进行一些调整,但其中所建立的指标体系和识别模型方法可以固化成模块,放在经分系统上进行数据的自动输入和特征提取,虽然可能在具体的数据指标识别规则的阈值上会有所微调,但从实际数据上来看,在一定的时间内会比较稳定,因此从定量上给出方法,从定性上给出方向,本文所提取的使用国际漫游通信产品的用户典型特征具有一定的实际指导意义,也比较符合当前运营商的总体发展策略,比如 2G 用户向 3G、4G 的快速迁移,有国际漫游收入记录、有国际长途语音通信行为、使用中高端的 3G 或 4G 终端、出账收入较高的中青年群体等都属于出国人群的典型特征。

## 5 结束语

当前国际漫游用户在运营商移动用户中的占比仍然很低,随着国际漫游资费的不断下调,出国人群中必然会有越来越多的用户使用国际漫游通信产品,根据出国人群的通信行为特征持续优化和丰富国际漫游通信产品,将会逐步消除用户顾虑,培养出国人群使用国际漫游通信产品的习惯,增强运营商国际竞争力,提升国际漫游通信产品资费的议价能力,从而可以进一步下调国际漫游产品资费,国际漫游用户也就会越多,最终形成正向循环,促进运营商国际业务的拓展以及提升国际漫游用户的满意度。

参考文献:

[1] 2014年中国公民出境游达1.17亿人次[EB/OL]. [2015-01-13]. <http://www.traveldaily.cn/article/88133>.  
The outbound travel of Chinese citizens amounted to 1.17 billion in 2014 [EB/OL]. [2015-01-13]. <http://www.traveldaily.cn/article/88133>.

[2] 国家旅游局. 2015年春节中国出境旅游报告—信息图[EB/OL]. [2015-03-16]. <http://www.199it.com/archives/333033.html>.  
National Tourism Administration of the People's Republic of China. 2015 spring festival China outbound tourism report-information map[EB/OL]. [2015-03-16]. <http://www.199it.com/archives/333033.html>.

[3] 中国旅游研究院: 2015年上半年出入境旅游总人数1.27亿人次, 同比增长9.8%[EB/OL]. [2015-07-22]. <http://www.199it.com/archives/368089.html>.  
China Tourism Research Institute. The first half of 2015 the total number of entry and exit travel 127 million passengers, an increase of 9.8% [EB/OL]. [2015-07-22]. <http://www.199it.com/archives/368089.html>.

[4] 世界旅游组织. 中国连三年成为世界最大出境游市场[EB/OL]. [2015-02-04]. [http://news.xinhuanet.com/overseas/2015-02/04/c\\_127455722.htm](http://news.xinhuanet.com/overseas/2015-02/04/c_127455722.htm).  
World Tourism Organization. China has become the world's largest outbound tourism market in three years [EB/OL]. [2015-02-04]. [http://news.xinhuanet.com/overseas/2015-02/04/c\\_127455722.htm](http://news.xinhuanet.com/overseas/2015-02/04/c_127455722.htm).

[5] 2015中国国际移民报告 [R/OL]. [2015-03-20]. <http://edu.sina.com.cn/a/2015-03-20/1146258457.shtml>.  
2015 China international migration report [R/OL]. [2015-03-20]. <http://edu.sina.com.cn/a/2015-03-20/1146258457.shtml>.

[6] 2014年中国出国留学人员超45万 [EB/OL]. [2015-03-05]. <http://www.chinanews.com/edu/2015/03-05/7103900.shtml>.  
Chinese students abroad over 45 million in 2014 [EB/OL]. [2015-03-05]. <http://www.chinanews.com/edu/2015/03-05/7103900.shtml>.

[7] 杜强, 贾丽艳, 严先锋. SPSS统计分析从入门到精通 (第2版)[M]. 北京: 人民邮电出版社, 2014.  
DU Q, JIA L Y, YAN X F. SPSS Statistical Analysis from Entry to the Master (Second Edition)[M]. Beijing: Posts & Telecom Press, 2014.

[8] 薛薇, 陈欢歌. 基于Clementine的数据挖掘[M]. 北京: 中国人民大学出版社, 2012.  
XUE W, CHEN H G. Data Mining Based on Clementine [M]. Beijing: China Renmin University Press, 2012.

[作者简介]



申志伟(1980-),男,博士,中国联合网络通信有限公司研究院高级工程师, 主要研究方向为大数据应用、移动互联网产品研发、用户行为研究、物联网及云计算等。



崔瑞媛(1985-),女,博士,中国联合网络通信有限公司研究院工程师, 主要研究方向为物联网、移动互联网、通信业务与商业模式等。



吕恒(1983-),女,中国联合网络通信有限公司研究院工程师, 主要研究方向为大数据、用户行为研究、市场策略、互联网金融等。