



一种博弈论辅助的机器学习算法检测用户流失行为

欧阳晔¹, 杨爱东¹, 孟凡语²

(1. 亚信科技(中国)有限公司通信人工智能创新实验室, 北京 100193;

2. 加州大学伯克利分校电子工程与计算机科学系, 旧金山 伯克利 94720)

摘要: 中国在 2019 年 11 月底正式实施已经试行了 9 年的携号转网政策。该政策会加强通信市场的流动性和竞争性, 使运营商用户流失的问题更加突出。提出、验证并产品化了一种博弈论辅助的机器学习方案, 以帮助运营商主动应对携号转网市场的竞争。所提方案为运营商提供了一种机器学习模型, 检测用户的携转倾向, 并给予差异化待遇。实验结果证明, 所提方案能够指引运营商制定有针对性的携号转网策略, 准确识别出有携入或者携出倾向的“异常”用户。此外, 所提方案已被成功地产品化, 极大地提高了运营商的营销效率, 增加了用户的满意度, 为中国某主要运营商减少了大约 50% 的用户流失。

关键词: 用户流失; 携号转网; 博弈论; 机器学习

中图分类号: TP181

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2020164

A game theory-assisted machine learning methodology for subscriber churn behaviors detection

OUYANG Ye¹, YANG Aidong¹, MENG Fanyu²

1. Telco Artificial Intelligence Labs, AsiaInfo Technologies (China) Co., Ltd., Beijing 100193, China

2. Electrical Engineering & Computer Science Department, University of California, Berkeley 94720, US

Abstract: At the end of November 2019, China officially implemented the number portability policy (MNP) that has been in trial for 9 years. The policy will strengthen the liquidity and competitiveness of the telecommunication market, making the problem of subscriber churn more prominent. A game theory-assisted machine learning methodology was proposed, verified and commercialized timely, which could help mobile network operator (MNO) actively respond to competition in the MNP market. The proposed methodology provides MNO with a machine learning model to detect subscriber portability and give differentiated treatment. Experimental results show that the proposed methodology can guide MNOs to make a targeted MNP strategy, and precisely identify “abnormal” subscribers who tend to churn-out and potential new subscribers who may churn-in. In addition, the proposed methodology has been successfully commercialized, greatly improving the marketing efficiency of operators, increasing user satisfaction, and reducing the loss of users by about 50% for a tier-1 MNO in China.

Key words: subscriber churn, mobile number portability, game theory, machine learning



1 引言

手机号码被认为是当今信息社会最重要的个人资产之一。传统上,即使服务不令人满意,移动用户也不愿意更换运营商,因为不愿意放弃已有电话号码^[1]。为了降低转换成本,增加电信市场的竞争,越来越多的国家/地区,包括日本^[2]、美国^[3]和欧洲^[4]等,已发布 MNP (mobile number portability, 移动号码可携带性)或 WLNP (wireless local number portability, 无线本地号码可携带性)政策,即携号转网。WLNP 是由美国 FCC (Federal Communications Commission, 联邦通信委员会)定义的^[5],允许无线用户从一个 MNO (mobile network operator, 移动网络运营商)切换到同一区域的另一个 MNO,并保留其现有电话号码。

中国在 5 个试点省份进行 9 年的 MNP 试点后,于 2019 年 11 月底正式实施 MNP。2020 年受“新冠肺炎”疫情影响,经济增长备受压力,MNP 对市场注入了另一个不确定因素。中国 3 家一级运营商中有两家在 2019 年财务报告出现净利润同比下滑(中国移动下降 9.5%,中国电信下降 3.3%,中国联通增长 11.1%)^[6-8],MNP 市场变化的另一个明显迹象是,3 家 MNO 均宣布从 2019 年 8 月开始结束“无限量计划”^[9]。因此,运营商在 MNP 政策下将采取更具战略性的行动,而不仅仅是像过去那样无休止地打价格战。

在这种情况下,运营商只采取固定的营销策略并不能适应日新月异的环境。即使运营商采用了一些灵活的策略,MNP 个体消费者的行为仍然很难识别和预测。MNP 用户流失受到多种因素的影响,本文利用博弈论辅助机器学习方法从以下两个方面帮助运营商:

- 为运营商制定有针对性的 MNP 营销策略,以应对同一市场中对手的竞争;
- 为运营商制定一套精准的 MNP 策略,即识别和检测用户流失行为、挽留出现 MNP

流失早期迹象的用户,并锁定可能从竞争对手处携入的潜在用户。

本文工作的主要贡献包括以下 5 个方面。

- 针对我国出台的 MNP,本文首次提出并运用博弈论的方法来解决运营商在开放、自由竞争市场中的最优营销策略问题,以实现自身利益的最大化。
- 为了帮助运营商在波动的市场环境中动态地实现最优决策,本文综合分析了 3 家运营商 27 种可能的市场营销博弈组合以及 3 种可能市场决策下的预期利润。
- 将运营商的市场策略简化为对用户携转行为的预测,即转化为一个三分类问题。利用机器学习模型,结合我国三大运营商的大数据,构造出任意用户的转移矩阵,从而得到用户的转移倾向。
- 通过比较常用模型和神经网络,发现 SNN (stacking neural network, 叠加神经网络)模型具有优越的性能。在极不平衡的数据集上,该模型可提高运营商的营销效率,减少近 50%的用户流失。
- 通过对 SNN 模型分析,发现具有携转倾向用户具有呼叫时间较长、呼叫频率较高的特点。他们可能对数据和语音通话服务质量有更高要求。因此,提高数据和语音服务质量可有效地保留这些用户。

2 相关研究

用户流失预测是一个备受关注的研究热点。由于吸引新用户的成本比留住现有用户的成本高出约 5~6 倍^[10],因此,提前预测用户流失对运营商至关重要,特别是在携号转网政策下。在这一部分中,本文对相关的研究进展进行了简要的概述。

2.1 影响用户流失的因素

在高度饱和的竞争性电信市场中,用户将根据他们认为重要的任意因素从一个运营商迁移到

另一个运营商。在西班牙进行的一项纵向调查发现,与固定服务捆绑的移动服务消费者离开现有运营商的意愿较低,捆绑越复杂,流失率越低^[11]。自从美国引入 MNP 后,相关研究表明,通话质量可能不是一个重要的因素^[3],这一发现与德国电信市场的研究结果一致^[12]。然而,在中国北方地区,服务质量对用户的流失行为有着显著的影响^[13]。用户的社交网络信息也是影响用户流失的另一个关键因素。顾客并不是孤立的,他们的行为会受到周围人的影响^[14]。基于这一观点,研究者探讨了社交网络在用户保持中的作用,发现在紧邻的用户流失之后,用户流失率增加了 80%^[15]。

2.2 用户流失预测模型

在一个饱和的电信市场中,吸引新用户比留住现有用户更困难。据估计,移动用户的平均流失率为每月 2.2%^[10]。为了降低用户流失率,运营商需要找出潜在的流失用户。这点可以通过建立用户流失预测模型来实现,它利用历史数据对当前用户是否是潜在的流失用户进行分类。在电信领域,有许多算法被用来构建这样的用户流失预测模型,包括决策树、基于规则算法、朴素贝叶斯、支持向量机和神经网络^[10]。此外,为了获得更准确的预测结果,参考文献[16]中,利用树方法和遗传规划相结合的混合学习模型来推导分类规则,并运用博弈论的方法分析用户流失的社区效应。

通过文献调研,不难发现,现有的大多数研究仅集中在用户流失行为预测上,没有考虑到 MNP 政策下的市场博弈行为。MNP 的引入不仅使运营商之间的市场博弈更加复杂,用户的携转倾向也变得难以预测。本文首先运用博弈论对运营商的 MNP 策略进行了理论分析和推导,然后提出了一个博弈论辅助的堆叠神经网络模型,将用户分类为流失者和非流失者,以便运营商制定更加高效的营销策略。

3 基于博弈论的携号转网决策算法

当多家公司向同一市场提供竞争性产品时,

它们的营销行为会相互影响。博弈论为分析理性决策者之间的战略互动提供了一种自然的数学分析工具,并被广泛应用于经济学、政治学、逻辑学和计算机科学等领域^[17]。中国移动、中国联通和中国电信是主导电信市场的 3 家公司,为了避免泄露用户隐私,本文在下面使用 A、B 和 C 来指代中国的三大运营商。

3.1 博弈论算法

博弈论已成功应用于销售价格战^[18]、交易决策^[19]和欺诈惩罚^[20]等方面。本节介绍了一种基于博弈论的 MNO 攻防营销行为的策略评估方法,并将其应用于 MNP 下运营商的营销决策。

本文所提出的基于博弈论的攻防营销策略方法的工作流程如图 1 所示,其中包括核心业务数据的特征提取、业务价值的综合评分、与竞争对手的市场博弈以及最佳决策策略的输出。为了最大程度地提高 MNO 的效益,针对中国 MNP 设计了一种基于博弈论的决策算法,见算法 1。该算法动态输出当前的最佳营销决策,包括最大携入($\max_{in}$)、最小携出($\min_{out}$)或兼并最大携入和最小携出($\max_{in/out}$),具体何种策略取决于当前的营销目标 and 市场条件。

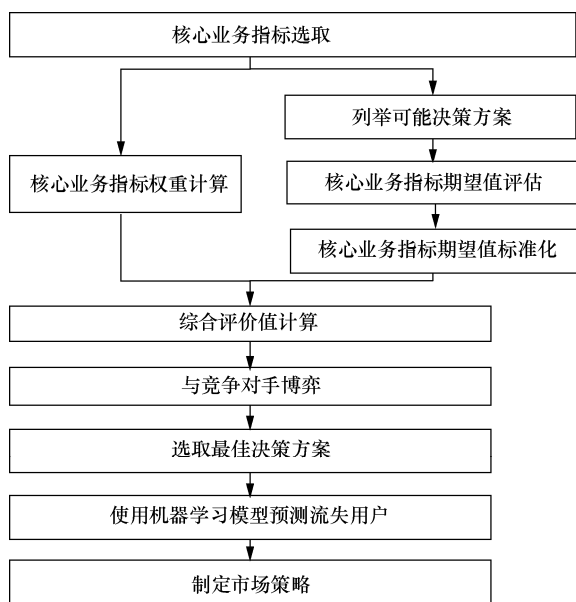


图 1 基于博弈论的攻防营销决策策略评价方法的工作流程



算法1给出了运营商在MNP业务中的博弈算法过程。首先,选择运营商公开和/或私有业务数据的重要特征 $X_i, i=1,2,\dots,n$ 作为输入,并计算 X_i 和 X_j 之间的相对重要性矩阵 β :

$$\beta_{i,j} = \begin{cases} 1, & X_i \text{ 比 } X_j \text{ 重要} \\ 0, & X_i \text{ 不如 } X_j \text{ 重要} \end{cases} \quad (1)$$

特征重要性的得分 α_i 和权重 ω_i 可以通过式(2)和式(3)计算:

$$\alpha_i = \sum_{j=1}^n \beta_{i,j} \quad (2)$$

$$\omega_i = \frac{\alpha_i}{\sum_{i=1}^n \alpha_i} \quad (3)$$

算法1 中国运营商携号转网博弈算法

选择核心指标 $X_i, i=1,2,\dots,n$

通过式(1)~式(3)分别计算参数 $\beta_{i,j}$ 、 α_i 和 ω_i

for $v \in \{\max_{in}, \min_{out}, \max_{in/out}\}$ do

对 i^{th} 特征和 v^{th} 营销策略, $i=1, 2, \dots, n$, 设置操作目标期望值 $\tilde{x}_i^v$

通过表1计算得分 $\tilde{s}_i^v, i=1,2,\dots,n$ 计算综合评价值 $Z_v = \sum_{i=1}^n \omega_i \tilde{s}_i^v$

end

for 所有 $v \in \{\max_{in}, \min_{out}, \max_{in/out}\}$ do

for 所有 $o \in \{A, B, C\}$ do

通过式(5)计算得分 $S_{v(i)}^{o(m)}$

end

end

对所有 $v \in \{\max_{in}, \min_{out}, \max_{in/out}\}$ 和所有 $o \in \{A, B, C\}$ 计算数学期望 $E\{S_{v(i)}^{o(m)}\}$

对所有 $o \in \{A, B, C\}$ 输出最优的 i of $\max[E\{S_{v(i)}^{o(m)}\}]$

假设MNP业务中特征 i 和营销策略 v 的营销目标为 $\tilde{x}_i^v$, 其中, v 表示 $\max_{in}$ 、 $\min_{out}$ 或者 $\max_{in/out}$, 营销目标的基准分见表1, $\tilde{s}_i^v$ 为 $\tilde{x}_i^v$ 的基准得分, 营销策略 v 的综合价值为:

$$Z_v = \sum_{i=1}^n \omega_i \tilde{s}_i^v \quad (4)$$

表1 评分基准

特征	得分 ≥ 5	得分 ≥ 4	得分 < 4	得分 ≥ 3	得分 < 3	得分 ≥ 2	得分 < 2	得分 ≥ 1	得分 < 1	得分 < 0
$\tilde{x}_1$	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}
$\tilde{x}_2$	S_{20}	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}
$\tilde{x}_3$	S_{30}	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
$\tilde{x}_n$	S_{n0}	S_{n1}	S_{n2}	S_{n3}	S_{n4}	S_{n5}	S_{n6}	S_{n7}	S_{n8}	S_{n9}

由于中国三大电信运营商A、B、C在实际博弈中都有3种可能的营销策略,即 $\max_{in}$ 、 $\min_{out}$ 和 $\max_{in/out}$, 并且任何两家都会相互博弈,因此相互的博弈策略有27种可能的排列组合。为了得到每个博弈者的最优解,所有可能的博弈论结果都通过式(5)计算。

$$S_{v(i)}^{o(m)} = 2Z_{v(i)}^{o(m)} - Z_{v(j)}^{o(n)} - Z_{v(k)}^{o(l)} \quad (5)$$

其中, $S_{v(i)}^{o(m)}$ 是运营商 $o(m)$ 选择策略 $v(i)$ 后和其他两个运营商博弈后的最终得分。 Z_v^o 是由式(4)

计算的运营商 o 选择营销策略 v 的综合得分,其中 $o \in \{A, B, C\}, v \in \{\max_{in}, \min_{out}, \max_{in/out}\}, i, j, k, m, n, l$ 的取值分别为1、2或3,并且 $m \neq n \neq l$ 。

通过对所有 v 和 o 计算期望值 $E\{S_{v(i)}^{o(m)}\}$ 可以得到对运营商 $o(m)$ 的最优策略。

$$v\{i\} \leftarrow \max[E\{S_{v(i)}^{o(m)}\}] \quad (6)$$

这里考虑了所有 $v \in \{\max_{in}, \min_{out}, \max_{in/out}\}$ 和所有 $o \in \{A, B, C\}$ 。

3.2 用户流失预测算法

在中国通信市场上,用户只有中国移动、中国联通和中国电信3个运营商可供选择。在MNP政策实施后,用户可以自由地从一个运营商转到另一个运营商。为了最大化利润,运营商必须提前估算出用户流失率。

假设用户 i 决定从运营商 m 转移到运营商 n ,那么转移决策 p_{mn}^i 可表示成:

$$p_{m,n}^i = \sum_k \delta_{m,k} p_{k,n}^i \quad (7)$$

其中, $\delta_{m,k}$ 是狄拉克 Delta 公式。根据式 (7) 可以构造出转移矩阵。对于任意一个用户来说, 他的选择都是确定并且唯一的。因此, 用户的转移决策 $P_{m,n}$ 可以写成:

$$P_{m,n} = \begin{cases} 1, & \text{用户从 } m \text{ 转向 } n \\ 0, & \text{其他} \end{cases} \quad (8)$$

从式 (8) 可以看出, 转移矩阵中只有一个元素是 1, 其他元素都是 0。假设整个通信市场上有 N 个用户, 在 $t+1$ 时间 (单位: 月), 用户携入运营商 m 总的数量为:

$$\text{in}(m, t+1) = \sum_{i=1}^N \sum_{n \neq m} p_{nm}^i(t) \quad (9)$$

而从运营商 m 携出的总的用户数为:

$$\text{out}(m, t+1) = \sum_{i=1}^N \sum_{n \neq m} p_{mn}^i(t) \quad (10)$$

为了吸引用户携入, 减少用户携出, 需要提前知道任意用户 i 的转移决策矩阵 M^i 。根据先前的研究^[1-4, 8, 21-22], 影响用户转移决策的因素有很多, 包括受教育程度、套餐绑定状态、携转成本、通话质量、月租费和通话时长等。因此, 用户的转移决策会依赖于用户的属性, 也就是:

$$p_{mn}^i(t) = p_{mn}^i(t, D^i) \quad (11)$$

其中, D^i 是指用户 i 的特征。

现在的挑战是如何从 (t, D^i) 得到式 (11) 中的 p_{mn}^i 。最近, 一些研究提出用 Markov 链^[21]、Logistic、社会网络分析^[16]以及其他的一些统计方法^[11, 23]来预测 p_{mn}^i 。考虑到用户有 3 种选择, 留在原运营商和转向另两个运营商, 转移决策问题实际上是一个三分类问题。在本文中, 为了获得优异的预测性能和可解释性, 采用了堆叠神经网络模型 (SNN) 来预测用户的转移决策。此外, 一些其他的模型, 比如随机森林 (RF)、决策树 (DT)、长短期记忆网络 (LSTM)^[24]、梯度提升树 (GBDT)、XGBoost^[22]、ExtraTree 和全连接神经网络 (NN) 也被用来作为对比。由于用户携出数

量相对较小, 重采样和下采样的技术被用来解决数据不均衡的问题。在构建好用户流失预测模型后, 每种策略期望收益可以表达为:

$$\max \{ \text{in}(m, t+1) \} = \max \left\{ \sum_{i=1}^N \sum_{n \neq m} (p_{nm}^i)^* (t, D^i) \right\} \quad (12)$$

$$\min \{ \text{out}(m, t+1) \} = \min \left\{ \sum_{i=1}^N \sum_{n \neq m} (p_{mn}^i)^* (t, D^i) \right\} \quad (13)$$

$$\max \left\{ \frac{\text{in}(m, t+1)}{\text{out}(m, t+1)} \right\} = \max \left\{ \frac{\sum_{i=1}^N \sum_{n \neq m} (p_{nm}^i)^* (t, D^i)}{\sum_{i=1}^N \sum_{n \neq m} (p_{mn}^i)^* (t, D^i)} \right\} \quad (14)$$

其中, p^* 是机器学习预测的用户转移决策。

使用这种方法, 无论博弈论得到的策略是什么, 首先都需要通过机器学习模型得到转移决策矩阵。对于策略 $\max_{\text{in}}$ 、 $\min_{\text{out}}$ 和 $\max_{\text{in/out}}$, 预测用户转移倾向的机器学习模型是相同的。在识别出将要流失的用户后, 运营商可针对性开展营销活动。与此同时, 机器学习模型的结果也可以作为博弈论决策的依据。

4 仿真和讨论

为了深入研究基于博弈论的机器学习方法的性能, 并验证第 3 节理论分析的有效性, 本文以实际市场环境下运营商 A 动态营销决策的博弈论为例进行了蒙特卡罗仿真分析, 并评估基于营销决策的用户流失预测模型的算法性能。

4.1 博弈论分析

以 MNP 为例, 展现某运营商 A 如何通过博弈论算法获得最优营销策略。博弈论算法 1 列出了 A 运营数据的月度核心特征 X_i (可根据业务和营销目标选择), 见表 2。

将表 2 特征代入式 (1) ~ 式 (3) 可得特征重要性矩阵 β , 见表 3, 其中 $X_1 \sim X_5$ 是运营商 A 在本月 MNP 营销中的重要特征, X_6 是可选扩展特征。

为了在与其他竞争对手博弈的同时实现营销目标, 营销运营商需要为表 4 的 3 种营销运营策



表 2 电信核心月度业务数据特征提取

运营商	特征							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
A	-5.88%	-70.4	58.50%	-29.7	68.20%	4 701.9	3.093	162.5
B	-0.79%	-14.8	13.90%	1.3	5.60%	389.5	0.467	38.8
C	-3.91%	-15.3	27.60%	1.3	26.20%	1 807.2	1.391	76.6

注： X_1 ：月收入变化率； X_2 ：每月新增用户数变化（万户）； X_3 ：每月新增用户变化率； X_4 ：包月用户变动（万户）； X_5 ：留存用户月变化率； X_6 ：月留存用户数量（万户）； X_7 ：月收入（10 亿元）； X_8 ：新增用户数量（万户）。

表 3 特征重要性矩阵

特征	特征						得分 a_i	权重 w_i
	X_1	X_2	X_3	X_4	X_5	X_e		
X_1	0	1	1	0	0	$1-\beta_{e_1}$	2	0.2
X_4	1	1	0	0	0	$1-\beta_{e_4}$	2	0.2
X_5	1	1	1	1	0	$1-\beta_{e_5}$	4	0.4
X_e	β_{e_1}	β_{e_2}	β_{e_3}	β_{e_4}	β_{e_5}	0	$\sum_j \beta_{e_j}$	$\frac{\alpha_e}{\sum_{i=1}^e \alpha_i}$

表 4 运营商 A 不同营销策略所选特征的营销目标

营销目标 v	特征					
	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$\tilde{x}_5$	$\tilde{x}_A$
$\max_{in}$	2%	11	57%	-5	66%	NA
$\min_{out}$	-4%	20	61%	-8	64%	NA
$\max_{in/out}$	5%	-14	54%	-1	69%	NA

略 $v(i)$ 设置营销目标值 $\tilde{x}_i^v$ ，其中 $\tilde{x}_A$ 是可选特征。将表 4 中营销目标 $\tilde{x}_i^v$ 代入表 1，可得不同策略下的综合评价，见表 5。

4.2 营销策略的纳什均衡

电信运营商在制定市场营销策略时，其他两个运营商的选择也会对其战略决策产生很大影响。纳什均衡可以解决这个策略的影响问题，纳什均衡是一种基于博弈论的方法，可以精确模拟所有理性决策者之间的战略互动。通过与其他运

营商市场博弈模拟，可得到任一市场策略下期望营销得分，进而得到运营商当前最有利的市场策略。由于每个运营商有 3 种营销策略，即 $\max_{in}$ 、 $\max_{out}$ 和 $\max_{in/out}$ ，三大运营商在不同营销策略下的博弈收益见表 6，其排列组合总共有 27 种博弈场景。根据对竞争者选用策略的了解程度（已知、部分已知、未知），可得各运营商在不同策略下的预期收益。运营商 A 在不知道任何竞争对手的策略 v 的情况下的预期收益见表 7。

表 5 运营商 A 不同营销策略所得的综合评价

营销策略 v	X_1	X_2	X_3	X_4	X_5	X_A	综合得分 Z_v
	0.20	0.10	0.10	0.20	0.40	0	
$\max_{in}$	3	4	4	2	3	NA	3.00
$\min_{out}$	1	5	5	1	2	NA	2.20
$\max_{in/out}$	5	1	1	2	4	NA	3.20

表 6 三大运营商在不同营销策略下的博弈收益

博弈 轮数	项目	运营商		
		A	B	C
第 1 轮	营销策略	$\max_{in/out}$	$\max_{in/out}$	$\max_{in/out}$
	综合评分	0.3	-0.3	0
第 2 轮	营销策略	$\max_{in/out}$	$\max_{in/out}$	$\max_{in}$
	综合评分	1.5	0.9	-2.4
第 3 轮	营销策略	$\max_{in/out}$	$\max_{in/out}$	$\min_{out}$
	综合评分	-0.1	-0.7	0.8
第 4 轮	营销策略	$\max_{in/out}$	$\max_{in}$	$\max_{in/out}$
	综合评分	-0.3	0.9	-0.6
第 5 轮	营销策略	$\max_{in/out}$	$\max_{in}$	$\max_{in}$
	综合评分	0.9	2.1	-3
第 6 轮	营销策略	$\max_{in/out}$	$\max_{in}$	$\min_{out}$
	综合评分	-0.7	0.5	0.2
第 7 轮	营销策略	$\max_{in/out}$	$\min_{out}$	$\max_{in/out}$
	综合评分	0	0.3	-0.3
第 8 轮	营销策略	$\max_{in/out}$	$\min_{out}$	$\max_{in}$
	综合评分	1.2	1.5	-2.7
第 9 轮	营销策略	$\max_{in/out}$	$\min_{out}$	$\min_{out}$
	综合评分	-0.4	-0.1	0.5
第 10 轮	营销策略	$\max_{in}$	$\max_{in/out}$	$\max_{in/out}$
	综合评分	-1.3	0.5	0.8
第 11 轮	营销策略	$\max_{in}$	$\max_{in/out}$	$\max_{in}$
	综合评分	-0.1	1.7	-1.6
第 12 轮	营销策略	$\max_{in}$	$\max_{in/out}$	$\min_{out}$
	综合评分	-1.7	0.1	1.6
第 13 轮	营销策略	$\max_{in}$	$\max_{in}$	$\max_{in/out}$
	综合评分	-1.9	1.7	0.2
第 14 轮	营销策略	$\max_{in}$	$\max_{in}$	$\max_{in}$
	综合评分	-0.7	2.9	-2.2
第 15 轮	营销策略	$\max_{in}$	$\max_{in}$	$\min_{out}$
	综合评分	-2.3	1.3	1
第 16 轮	营销策略	$\max_{in}$	$\min_{out}$	$\max_{in/out}$
	综合评分	-1.6	1.1	0.5
第 17 轮	营销策略	$\max_{in}$	$\min_{out}$	$\max_{in}$
	综合评分	-0.4	2.3	-1.9
第 18 轮	营销策略	$\max_{in}$	$\min_{out}$	$\min_{out}$
	综合评分	-2	0.7	1.3
第 19 轮	营销策略	$\min_{out}$	$\max_{in/out}$	$\max_{in/out}$
	综合评分	0.7	-0.5	-0.2
第 20 轮	营销策略	$\min_{out}$	$\max_{in/out}$	$\max_{in}$
	综合评分	1.9	0.7	-2.6
第 21 轮	营销策略	$\min_{out}$	$\max_{in/out}$	$\min_{out}$
	综合评分	0.3	-0.9	0.6
第 22 轮	营销策略	$\min_{out}$	$\max_{in}$	$\max_{in/out}$
	综合评分	0.1	0.7	-0.8
第 23 轮	营销策略	$\min_{out}$	$\max_{in}$	$\max_{in}$
	综合评分	1.3	1.9	-3.2
第 24 轮	营销策略	$\min_{out}$	$\max_{in}$	$\min_{out}$
	综合评分	-0.3	0.3	0
第 25 轮	营销策略	$\min_{out}$	$\min_{out}$	$\max_{in/out}$
	综合评分	0.4	0.1	-0.5
第 26 轮	营销策略	$\min_{out}$	$\min_{out}$	$\max_{in}$
	综合评分	1.6	1.3	-2.9
第 27 轮	营销策略	$\min_{out}$	$\min_{out}$	$\min_{out}$
	综合评分	0	-0.3	0.3

根据表 7 可知, 运营商可获得当前市场中最优营销策略, 以实现自身利益最大化。如运营商 A 在当前市场环境下, 应选择 $\min_{out}$, 将以 100% 正收益概率获得 13.7 亿元收入。需注意的是, 当竞争对手的营销策略已知或部分已知时, 表 4 中的统计数据会发生变化, 最优营销策略选择也会发生改变。

4.3 用户流失预测

根据博弈论分析, 对于任一用户来说, 只需区分他是留下还是离开, 无须知晓他转向哪方。因此, 预测用户流失的三分类问题退化成二分类问题。

用来构建模型的数据收集来自不同月份, 这些数据分成两组 I 和 II。在组 I 中, 用户特征取自于 t_1-2 至 t_1 月, 标签取自于 t_1+1 月, 而在组 II 中, 用户特征取自于 t_2-2 至 t_2 月, 标签取自于 t_2+1 月。每组数据包含 31 000 000 个用户, 数据有 216 个特征。组 I 数据用来训练和验证模型, 组 II 数据用来测试。由于每组数据仅含 20 000 个正样本, 把它们选出来并随机插入 60 000 个负样本形成总量为 80 000 个训练和验证集 (比例为 4:1)。

SNN 模型的结构如图 2 所示, SNN 模型结构分成两层, 第一层包含 5 个子模型, 即 RF、ExtraTree、GBDT、XGBoost 和 NN, 第二层是神经网络模型。在训练阶段, 第一层 5 个模型都使用了五重交叉验证, 它们的输出是第二层输入。在模型训练前, 训练集中正样本进行重采样, 使其和负样本数量一致, 验证集正负样本保持为 1:3。

模型性能用 AUC 评估, 不同模型在测试集上的 AUC 值如图 3 所示, 不难发现, 本文所提出的 SNN 模型表现出了优越的性能。作为对比, DT、RNN、RF、ExtraTree、GBDT、XGBoost 和 NN 模型结果也呈现在图 3 中。SNN 模型 AUC 值是 0.88, 比其他模型大, DT 模型 AUC 值最小, 效果最差。此外, 循环神经网络模型 (RNN) 也有少量研究者提出用来尝试预测用户的流失



表 7 不同营销策略下各运营商的收益值

营销策略 v	A			B			C		
	$*Y_1$	$*Y_2$	$*Y_3$	$*Y_1$	$*Y_2$	$*Y_3$	$*Y_1$	$*Y_2$	$*Y_3$
$\max_{in}$	0.17	44%	56%	-0.1	33%	56%	0.27	44%	44%
$\min_{out}$	1.37	0	100%	-2.5	100%	0	-1.33	100%	0
$\max_{in/out}$	0.77	22%	78%	0.7	0	89%	0.67	11%	78%

注: $*Y_1$: 期望收益 (10 亿元); $*Y_2$: 负收入率; $*Y_3$: 正收入率。

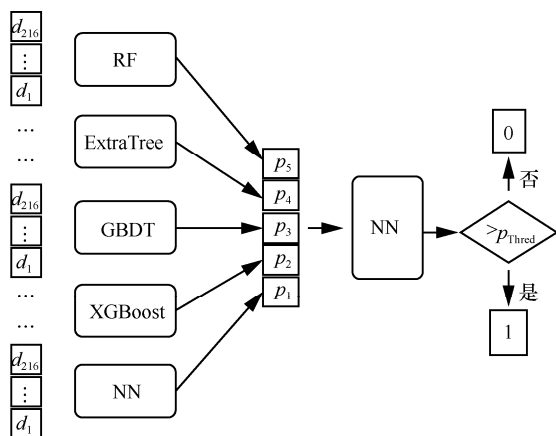


图 2 SNN 模型的结构示意图

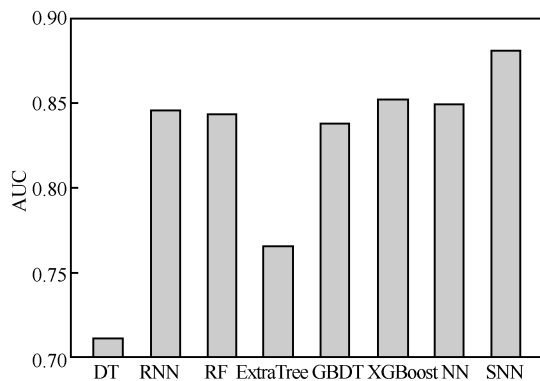


图 3 不同模型在测试集上的 AUC 值

行为，为了综合评估，本文采用了 RNN 中的 LSTM 来构造二分类模型，并获得最佳 AUC 值为 0.85，其性能与 XGBoost 与 NN 模型相当，但由于 RNN 结果不稳定，因此本文未在实际项目中选用 RNN 模型。

考虑到运营商的营销成本，实际项目中经常采用变化阈值来获得不同数量的目标客户群。图 4 评估了变化阈值对所提出的 SNN 模型性能影响。

在这次实验中，测试集包含 21 582 个正样本和 31 000 000 个负样本，并统计了不同阈值下的模型分类精确率、召回率和 $F1$ -score 值。

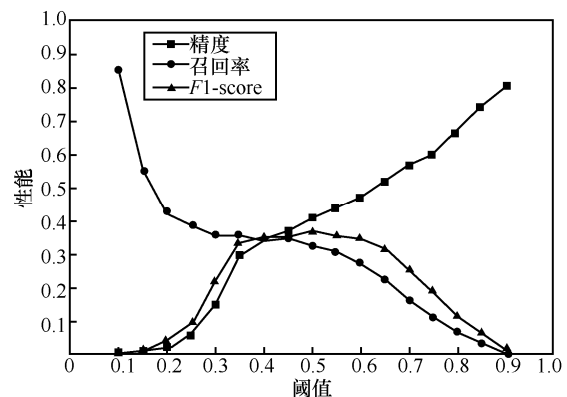


图 4 测试集上，SNN 模型在不同阈值下的分类结果

从图 4 可以看出，当阈值减小时，精度降低而召回率增大。在阈值为 0.55 时，SNN 的 $F1$ -score 最优为 0.37，而在验证集上 $F1$ -score 为 0.68。测试集的结果看起来比验证集差很多，是由于验证集正负样本比例为 1:3，而真实的测试集样本比例为 1:1 550。相比于传统的营销策略，基于 SNN 模型的策略会高效得多。此外，在阈值为 0.35~0.65 时，SNN 模型的性能几乎不变，说明模型的稳健性比较好。在实际营销过程中，根据运营商 A 投入的资源 and 策略，模型的阈值设为了 0.15，大约有 55% 的携号转网用户会被识别出来。如果想要找出更多携出用户，需要进一步降低阈值，但这会增加营销成本。

为了进一步分析倾向携出用户的特征，本文给出了对模型最重要的 8 个特征和它们对应的相对重要性 (RI)，如图 5 所示。

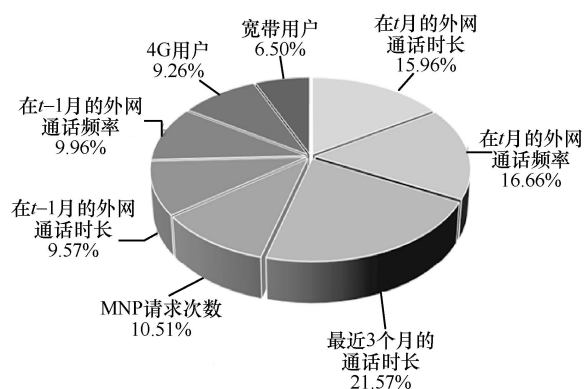


图5 对SNN模型影响最大的8个特征RI值

$$RI^i = \frac{F_1^i - F_0}{F_0} \quad (15)$$

其中, F_0 是 $F1$ -score, F_1^i 是特征 i 设为 0 时的 $F1$ -score。在图 5 中, 最重要的 8 个特征的 RI 被归一化为 1。 F_0 和 F_1^i 都是在阈值为 0.55 时计算得到的。

5 产品: MNP 分析平台 MAP

总体解决方案是在 MNP 分析平台 MAP 上开发的, 该平台已在中国多个省份一级运营商中推出。MAP 系统结构如图 6 所示, 由三大部分的 6 个组件, 即大数据平台、决策平台、算法引擎、策

略匹配、营销执行和效果评估构成。

- 第一部分为大数据平台部分, 它包含大数据处理的基础组件, 负责从外网和内网源数据中提取、转换、加载、聚合 MNP 相关联的数据。典型的数据包括营销数据、用户行为、概况和属性数据、网络和应用程序性能数据以及其他竞争对手策略数据。
- 第二部分为模型引擎部分, 它包含决策平台、算法引擎和策略匹配 3 个组件。决策平台负责 MAP 策略制定, 博弈论算法动态输出营销策略和所有可能策略选项。算法引擎负责 MNP 的流失预测, 一旦确定了最佳营销策略, MNP 流失预测组件中的机器学习模型将开始运行, 实时输出可能的流失客户群。策略匹配组件负责营销策略的匹配, 该组件内设计了精准策略匹配模型, 策略推荐算法将流失用户与策略库最佳策略进行匹配。
- 最后一部分为执行和评估部分, 它包含了营销执行和效果评估两个组件。营销执行组件负责在用户中心根据营销策略与最佳营销渠道、时机对目标用户开展营销活

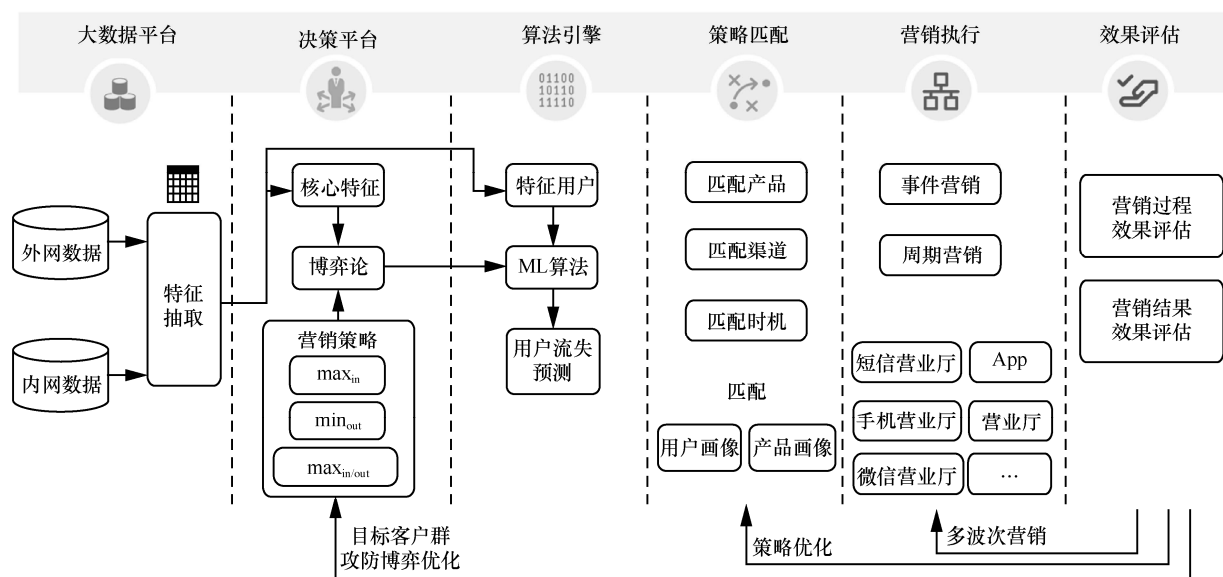


图6 MAP的系统架构（MNP分析平台）



动,实现降低用户流失意愿。效果评估组件负责后验评估和反馈功能,它对整个工作流程和结果进行监控和反馈,将用户最终反馈回传给前面各组件以进行算法调优。

这三大部分的6个组件对MAP平台同等重要,本文只关注算法引擎部分。MAP平台其他部分即将推出。在使用MAP平台后,A运营商在某省份携号转网用户从每月平均约20 000人减少到约10 000人,减少了近50%(6个月平均值)。

6 结束语

携号转网政策加剧了电信市场的流动性和竞争性,本文提出了一种基于博弈论辅助的机器学习算法来分析用户的流失行为,从而辅助运营商来适应新的市场环境。首先,提出了采用了博弈论算法来帮助运营商进行最佳市场决策,并针对3家运营商,根据纳什均衡原理,推导了全部市场博弈组合和预期收益。接着,利用机器学习模型SNN检测MNP用户流失行为,通过模型预测结果针对性地开展营销活动来减少用户流失。最后,将所提出的算法全部产品化,并在中国的多个省份一级运营商中推出该产品,不仅可提高运营商的营销效率,还能提高营销质量。此外,从实际的产品运营数据来看,策略匹配推荐算法对减少客户流失起到很重要的作用,未来的研究方向将集中于最佳策略匹配推荐算法,以实现智能化的精准营销。

参考文献:

- [1] MAICAS J P, POLO Y, SESE J. Reducing the level of switching costs in mobile communications: the case of mobile number portability[J]. *Telecommunications Policy*, 2009, 33(9): 544-554.
- [2] OTSUKA T, MITOMO H. User benefits and operator costs of mobile number portability in Japan and impact on market competitiveness[J]. *Telecommunications Policy*, 2013, 37(4-5): 345-356.
- [3] SHIN D H. A study of mobile number portability effects in the united states[J]. *Telemat Inform*, 2007, 24(1): 1-14.
- [4] BUHLER S, DEWENTER R, HAUCAP J. Mobile number portability in Europe[J]. *Telecommunications Policy*, 2006, 30(7): 385-399.
- [5] FCC. Wireless local number portability (WLNP)[EB]. 2016.
- [6] China Mobile. Announcements[EB]. 2020.
- [7] China Telecom. Investor calendar[EB]. 2020.
- [8] China Unicom. Presentations & Webcasts[EB]. 2020.
- [9] SINA. Operators will cancel the unlimited packages. Telecom will no longer reach the speed limit from next month[EB]. 2019.
- [10] BERSON A, SMITH S, THEARLING K. Building data mining applications for CRM[Z]. 1999.
- [11] GARCIA-MARTINOSO B, SUÁREZ D. Switching mobile operators: evidence about consumers behavior from a longitudinal survey[J]. *Telecommunications Policy*, 2019, 43(5): 426-433.
- [12] GERPOTT T J, RAMS W, SCHINDLER A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market[J]. *Telecommunications Policy*, 2001, 25(4): 249-269.
- [13] LIANG D, MA Z, QI L. Service quality and customer switching behavior in China's mobile phone service sector[J]. *Journal of Business Research*, 2013, 66(8): 1161-1167.
- [14] BACKIEL A, BAESENS B, CLAESKENS G. Predicting time-to-churn of prepaid mobile telephone customers using social network analysis[J]. *Technological Forecasting and Social Change*, 2016, 67(9): 1135-1145.
- [15] NITZAN I, LIBAI B. Social effects on customer retention[J]. *Journal of Marketing*, 2011, 75(6): 24-38.
- [16] YESHWANTH V, RAJ V V, SARAVANAN M. Evolutionary churn prediction in mobile networks using hybrid learning[C]//*Proceedings of Twenty-Fourth International FLAIRS Conference*. Palo Alto: AAAI Press, 2011.
- [17] ZHANG R, ZHU Q. A game-theoretic defense against data poisoning attacks in distributed support vector machines[C]//*Proceedings of 2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. Piscataway: IEEE Press, 2017: 4582-4587.
- [18] SLADE M E. Vancouver's gasoline-price wars: an empirical exercise in uncovering supergame strategies[J]. *Rev. Economic Studies*, 1992, 59(2): 257-276.
- [19] FERRERO R, SHAHIDEPOUR S, RAMESH V. Transaction analysis in deregulated power systems using game theory[J]. *IEEE Transactions on Power Systems*, 1997, 12(3): 1340-1347.
- [20] MAGALONI B. The game of electoral fraud and the ousting of authoritarian rule[J]. *American Journal Political Science*, 2010, 54(3): 751-765.
- [21] ADEBIYI S O, OYATOYE E O, MOJEKWU J N. Predicting customer churn and retention rates in Nigeria's mobile telecommunication industry using Markov chain modeling[J]. *Acta Universitatis Sapientiae, Economics and Business*, 2015, 3(1): 67-80.
- [22] CHEN T, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//*Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2016: 785-794.

- [23] DORFFNER G. Neural networks for time series processing[C]// Proceedings of Neural Network World. [S.l.: s.n.], 1996.
- [24] CZAJKOWSKI M, SOBOLEWSKI M. How much do switching costs and local network effects contribute to consumer lock-in in mobile telephony?[J]. Telecommunications Policy, 2016, 40(9): 855-869.

[作者简介]



欧阳晔(1981—)，男，博士，亚信科技(中国)有限公司首席技术官，国家特聘专家，北京市政府特聘专家，主要研究方向为移动通信、人工智能、数据科学、科技研发创新与管理。



杨爱东(1984—)，男，博士，亚信科技(中国)有限公司通信人工智能实验室数据科学家，主要研究方向为5G通信人工智能技术、超宽带无线通信技术、高级信号处理技术、机器学习与深度学习。



孟凡语(1998—)，男，美国加州大学伯克利分校电子工程与计算机科学系在读，主要研究方向为机器学习与深度学习。