

深度学习在运营商固网存量业务保障方面的研究及应用

Research and Application of Deep Learning in Operators' Fixed Broadband Service Guarantee

赵越,刘芳琦,郑冠鸣,兰婷(中国联通江苏分公司,江苏南京210000)

Zhao Yue,Liu Fangqi,Zheng Guanming,Lan Ting(China Unicom Jiangsu Branch,Nanjing 210000,China)

摘要:

使用人工智能中的深度学习神经网络及机器学习决策树技术,对用户固网存量业务进行预测,并且对每个影响固网业务质量的变量进行重要度排名,以期提高运营商维系客户的效率和精准度,并且优化用户在使用固网业务时的体验。验证表明,人工智能技术在生产环境中可以帮助客服人员进行有效的客户维系。最后,因为人工智能技术在通信领域生产环境的应用属于初期阶段,还存在许多的问题需要改进,包含数据的“干净”、完整程度及建模时候的数据清洗整理难度等。

关键词:

深度学习;固网业务;机器学习;智能化运维
doi: 10.12045/j.issn.1007-3043.2018.12.012
中图分类号:TP181
文献标识码:A
文章编号:1007-3043(2018)12-0057-04

Abstract:

It explores the application of deep learning and machine learning technology to predict the user's fixed broadband off-grid behavior, and ranks the importance of each variable that affects the fixed broadband quality, in order to improve the efficiency and accuracy of operators to maintain customers, and to improve user experience. Verification shows that artificial intelligence technology can help customer service to maintain the customer. Finally, because the application of artificial intelligence technology in the communication field is in its infancy, there are still many problems that need to be improved, including the "cleanliness" of the data, the completeness and the difficulty of data cleaning during modeling.

Keywords:

Deep learning; Fixed-broadband service; Machine learning; Intelligente operation and maintenance

引用格式:赵越,刘芳琦,郑冠鸣,等.深度学习在运营商固网存量业务保障方面的研究及应用[J].邮电设计技术,2018(12):57-60.

1 概述

随着5G及人工智能时代的到来,通信行业的竞争也越发激烈,各大运营商都希望能在每个业务中为客户提供优质的服务,提升用户满意度,保留更多的用户,其中就包含固网业务。人工智能技术同5G技术一样刚刚登上历史舞台,都处于“婴儿期”,日后也都会茁壮成长,本文讨论人工智能算法模型今后在通信运营商实际生产环境中对固网业务用户保留维系的应用及存在的问题。

当前,各个领域都开始向更智能化、通信更快的方向前进,比如美国Google、特斯拉等企业开始研发的无人驾驶汽车,中国阿里巴巴、京东等电商使用的精准营销系统,这些都是通过人工智能深度学习及机器

学习算法加上快速的通信网络融合而成的新技术。对于通信行业,5G与人工智能技术同样和这个领域摩擦碰撞出了火花,例如智能化的故障排查检测、智慧城市等一系列项目。在众多的5G、人工智能技术与通信行业融合的项目中,智能化运营维护就是一个非常有应用前景,并且是实实在在产生经济效益的方向,通过人工智能算法达成的智能化运维将会大大提升通信运营商的运行效率,减少之前非智能化时代的不必要开支。

从通信运营商角度出发,如果可以预知用户行为,例如固网业务行为、手机离网、换机等,并且有针对性地得知用户产生一系列行为的主要原因是什么,然后,对用户进行有针对性的维系挽留,有目标性的故障问题解决,那么,对于通信运营商来说,这一个运行系统将会为其保留更多的用户,带来更多的经济效益。从用户角度出发,如果通信运营商可以提早有针对性地帮助其解决使用通信产品时的问题,让其产品体验感

收稿日期:2018-11-19

更加舒适的话,对用户来说也是一个福音,也愿意继续使用原本的产品服务。

上述的这些智能运维系统设想,都可以通过人工智能算法来实现,预知用户行为,并采取有目标的高效运维。本文以固网存量保障业务为例,阐述人工智能算法与通信运营相互合作的应用过程及后续智能化发展还需解决的问题。

固网业务行为预测的智能化运维实施主要分为2个部分,第1部分为用户离网行为的预测,该部分将使用深度学习领域的人工神经网络算法,最终给出用户离网的概率,例如一个用户的离网概率为78.5%。在建立神经网络的过程中会遇到的数据清洗、变量筛选都会在下文中详细说明。第2部分为影响用户固网业务离网行为变量的重要度排名,比如,对某一用户离网行为影响较大的前3个变量为MOS告警数、用户上行带宽、费用。这些变量的重要度排名将会给营销维系人员一个参考,让他们有针对性地对用户进行挽留。该部分使用机器学习中决策树分支的随机森林算法。

图1示出的是固网用户离网预测算法。

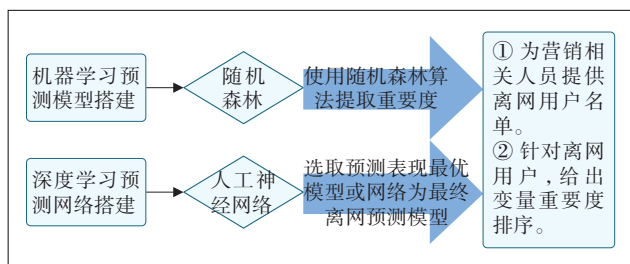


图1 固网用户离网预测算法

在建立模型的过程中将会选取O侧指标、B侧指标、故障信息、投诉信息、附属业务、用户家庭圈等数据。模型算法建立的过程中依旧存在许多问题,例如数据库的复杂程度、原始数据库的填充错误以及设计人工智能算法的工作人员对通信类数据不了解等问题,这些问题都会影响最后智能化运维的效果,具体细节将在下文中讨论。

图2示出的是固网用户离网预测模型。

根据前期阶段的实验结果、人工智能算法可以帮助提高通信运营商进行智能化运维,提高运维效率。

2 深度学习算法模型研究

2.1 深度学习应用设计概述

在通信行业进行人工智能化运维的改造主要包含

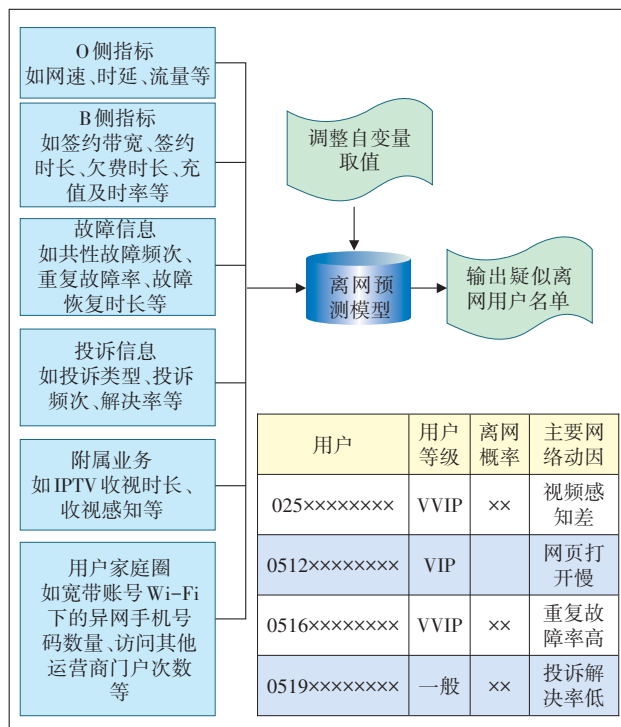


图2 固网用户离网预测模型

几个方面。

第1,设计人工智能算法,仔细依照业务实际情况来设计,要深入了解通信业务的过程及最终目标。

第2,进行有效的数据选取,清洗整理建模所需要的数据,根据原始数据生成新的字段变量。

第3,将生成验证完成的模型进行自动化运行,这一过程包含建模数据库的定时自动生成以及模型的定时自动调取数据、更新输出。

以固网业务离网行为预测为例,该项目的目标是帮助运营商有针对性地挽留维系固网业务用户,主要由2个部分组成。

第1,由模型算法给出用户的离网概率。

第2,给出对用户离网行为影响较大的几个变量排名,方便一线运维人员有目标性地与用户进行交流。

2.2 神经网络算法

在固网业务离网行为预测项目的第1个部分,使用了深度学习分支的人工神经网络,使用这个算法的原因是该算法较机器学习的回归模型更具有包容性,深度学习算法不需要像机器学习算法一样,需要前期对数据进行复杂的清洗来达到减少共线性,减少误差率的目的。再者,神经网络也比传统的回归模型具备更高的准确度。

神经网络是模仿人类大脑神经元而形成的,

一个神经网络由多层神经元构成,每一层神经元都有输入与输出,每一层神经元接受上一层神经元的输入,经过加权处理再传入下一层神经元当中,以此类推。人工神经网络主要分为输入层、输出层及隐藏层3个部分。

图3示出的是神经网络示意图。

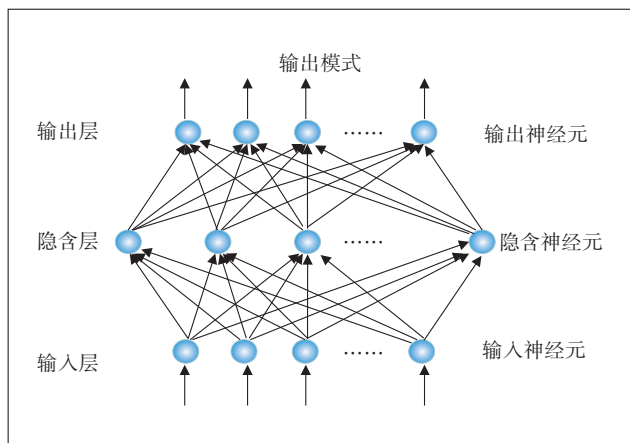


图3 神经网络示意图

第1部分输入层的作用是向算法传输数据变量的值,通常来说,输入层的输入节点数量与数据变量数量相等。这一层的节点都是被动地接受外来数据,不做任何改动,复制获取到的数据并向下一层中的多个隐藏节点传输数据。

第2部分隐藏层的任务就是对输入的数据进行处理改变。隐藏层可以有一层或者多层,隐藏层的节点数目不定,但是越多的节点会使得模型的非线性更加突出。

第3部分的输出层就是输出经过神经元处理的数据,针对设计变量输出相应的值,在固网业务项目中输出0~1的数值代表概率。

2.3 随机森林算法

固网业务离网预测的第2个部分是变量重要度的排名,该部分要对所有变量进行排名筛选,选择出几个会影响离网行为的重要特征变量。选择随机森林算法来完成,随机森林对特征值重要度评估的大体思路不复杂,在每一棵形成的树上观察每一个特征变量的贡献度,然后将这些贡献度平均一下,即获得了特征变量的重要度排名。每一棵树观察的贡献度即是基尼指数:

$$Gini(p) = \sum_{k=1}^k p_k(1 - p_k)$$

随机森林算法的一大优势在于其解决了每棵树之

间存在强共线性问题,在决策树方法中,假设有一个对于宽带离网行为非常重要的变量,然后还有一些其他的重要变量存在,那么在决策树算法中,几乎每一次每棵树都会选用影响力最大的那个变量作为最顶层树的分类标准,其次是其他重要变量,那么,按这个思路所形成的所有决策树都会十分相似,这样会导致结果存在高误差。但是随机森林算法不同,该算法很好地克服了这个问题,随机森林算法强制每一次的分叉都选用一部分的变量,所以,这样的做法不仅将重要的变量列入考虑范畴,其他的变量也享受了公平的“待遇”,一视同仁地选择分叉处的变量。所以,利用随机森林算法得出的重要度排名是有最小误差保证的。

3 深度学习应用与不足

3.1 深度学习算法应用

利用神经网络输出的概率大大提升了一线运维人员的工作效率,固网业务离网预测项目的建模数据库包含了200万名固网业务用户,如果选取离网概率高于70%的用户,那么一线运维人员可以将运维的范围从200万名用户缩小至30万名用户,运维工作量从人数上来说降低了87%,大大提升了工作效率以及运维的精准度。利用随机森林算法给出的特征变量重要度排名中,排名前3位的是上网总上行流量、tcp建立时长、上网总下行流量。

图4示出的是特征变量重要度。

3.2 深度学习算法的不足

虽然,利用人工智能的方法可以帮助实现运维的智能化,但是在实现的过程中依旧存在许多问题。其中最主要的问题是数据库的完整度与清洗难度。大部分通信运营商都会选择让第三方公司帮助其采集数据,但是在采集数据的过程中因为缺少监管,同时对数据完整度的重视程度不高,导致许多第三方公司在采集导入数据的时候出现问题,比如少采集、错误采集数据、导入填充数据时存在数据字段错行的现象、错误的值填入了错误的字段名称下方。在数据处理及模型设计方面,存在不相符的问题,换言之,懂得人工智能算法的人才不清楚通信业务,明白通信业务的人才不明白人工智能算法,导致在项目进行过程中有重复无效的动作发生,严重拖慢了项目进程,甚至导致项目设计错误。综上所述,在实施真正的智能化运维之前,一定要保证数据源的准确度,不然再优秀的模型算法设计及团队协作都不可能完成任务,数据一定要有质量

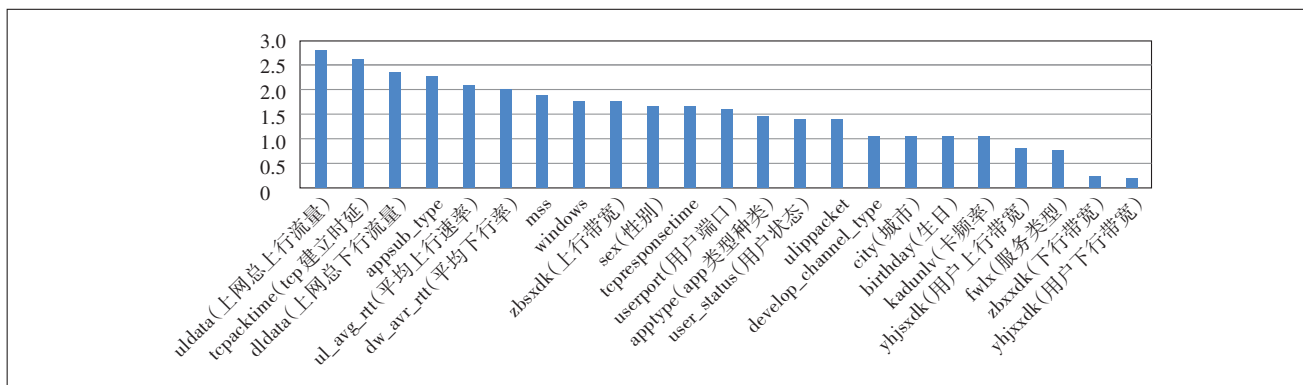


图4 特征变量重要度

保证。所以,通信运营商可以对采集数据的第三方公司进行更加严格的监管,协商签署可以保证数据源头质量的合作协议。对于复合型人才缺少的问题,当前解决的方法是增强团队协作能力,在管理智能化项目团队的时候要有明确的分工安排,让通信人才和人工智能人才紧密地融合团结在一起。

除此之外,还存在前期模型验证的问题,每一个算法模型在前期建立完毕之后,一定需要人员去尝试验证,这样才可以方便之后的模型迭代与优化,这个工作一定要在设计之初完成,计划好每一次需要验证的版块。例如,固网业务行为预测,当第一个人工神经网络搭建完毕之后,一定需要将其输出的结果投放至市场上去验证,如果准确率不高的话,需要工作人员及时地对数据库及算法本身进行检查,尝试重新整理,清洗数据,再建立更新模型。

人工智能算法在智能运维方面的价值一定要在保证数据库尽量完整准确,工作人员深度了解业务情况的前提下,才可以发挥出来。

4 人工智能未来展望

从用户体验角度出发,利用人工智能实现的智能化运维一定会创造出新的通信产品体验,从经济角度出发,智能化运维会为通信运营商带来新的经济增长点。如何在这个新老时代交接之际把握住机会,学会将人工智能技术运用到通信运维当中就是非常必要的一步。在电信运营商的发展战略中,学会挽留维系客户是非常重要的,例如固网业务等的用户维系挽留,如果可以预知用户行为,并进行维系挽留,那么对于运营商来说是非常有利的。

所以,深度学习及机器学习算法在通信运营领域是有用武之地的,是可以帮助运营商提高工作效率,在

不同的业务中实现智能化的,但是,在彻底实现智能化之前,要做的就是保证数据源的完整性与准确性,保证复合型人才和培养速度。那么,在不久的将来,由智能化运维武装的通信企业必定会为自己带来可观的收益,同时,为通信行业的用户带来人工智能及5G时代的优质服务。

参考文献:

- [1] 向军. 宽带业务发展趋势及策略研究[D]. 北京:北京邮电大学, 2009:25-29.
- [2] 程伟卿. 电信运营商宽带业务发展策略研究[J]. 中国新通信, 2013(6):52-52.
- [3] 张朝阳. IT、通信与互联网的融合正在加速[J]. 通信世界, 2004(1):39-39.
- [4] 章文波. 突破宽带 享受宽带[J]. 世界电信, 2005, 18(10):56-56.
- [5] 徐红辉. 电信企业的差异化竞争战略的制定[J]. 商场现代化, 2006(22):210-211.
- [6] 王建莹,刘志斌. 下一代电信业务网络[J]. 电信技术, 2003(8): 14-16.
- [7] 曾涛涌. 广电运营商进入宽带业务市场的策略研究[D]. 成都:西南财经大学, 2010.
- [8] 何勇生,新疆电信宽带业务发展策略研究[J], 新疆大学, 2012: 14-16.
- [9] 李华. BP神经网络在软件项目风险评估中的应用[J]. 计算机仿真, 33-37.
- [10] 徐晓霞,李金林. 基于决策树法的我国商业银行信用风险评估模型研究[J]. 北京理工大学学报. 2006, 8(3): 71-74.

作者简介:

赵越,毕业于南京邮电大学,硕士,主要从事固网存量业务保障与大数据分析等工作;刘芳琦,毕业于南京邮电大学,学士,主要从事固网业务支撑与大数据分析等工作;郑冠鸣,毕业于密歇根大学安娜堡分校,学士,主要从事固网用户离网模型建立工作;兰婷,毕业于西南民族大学,学士,主要从事固网大数据采集平台维护与支撑等工作。