



基于网络数据实现对物联网终端网络行为监控的方法

凌颖, 邱芸

(中国电信股份有限公司上海研究院, 上海 200122)

摘 要: 物联网是通过部署具有一定感知、计算、执行和通信等能力的各种设备, 获得物理世界的信息或对物理世界的物体进行控制, 通过网络实现信息的传输、协同和处理, 从而实现人与物通信、物与物通信的网络, 每一个物联网传感节点都是一个信息源, 数据流源源不断。电信网络运营商可以通过对网络数据的分析, 及时发现物联网终端的异常情况, 从而保证物联网终端的正常使用。阐述了一种通过分析电信网络中的网络数据、制定物联网终端的网络行为标签、对物联网终端网络行为进行实时监控的方法。并以穿戴行业应用为例, 描述了对物联网终端网络行为进行监控的分析过程。

关键词: 物联网; 网络数据; 网络行为监控

中图分类号: TP277

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017333

A method of IoT terminal's network behavior monitoring based on network data

LING Ying, QIU Yun

Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

Abstract: The internet of things (IoT) is the network of physical devices, vehicles, home appliances and other items embedded with electronics, software, sensors, actuators and network connectivity which enable these objects to connect and exchange data. Every sensor node is a source of information. In the offered method, telecom network operators could analyze the network data and discover the abnormal situation of the IoT terminals, so as to ensure the normal use of the IoT terminals. A network behavior tag for the IoT terminals by analyzing the network data in the telecom network, and a real-time monitoring method for the IoT terminal network behavior were described. And take the wearable industry application for example, the analysis process of the internet of things network behavior was described.

Key words: internet of things, network data, network behavior monitoring

1 引言

物联网是一种全球化的信息服务基础设施, 它基于现有和未来演进的通信和信息技术, 提供

互联物理和虚拟物体的先进服务^[1]。物联网是指通过部署具有一定感知、计算、执行和通信等能力的各种设备, 获得物理世界的信息或对物理世界的物体进行控制, 通过网络实现信息的传输、

协同和处理,从而实现人与物通信、物与物通信的网络^[2]。

物联网应用的数据产生于物联网产业链的各个环节,包括感知层、网络层、平台层和应用层上都会产生大量的数据^[3]。物联网应用产生的数据具有海量性、多态性等特性。每一个物联网传感节点都是一个信息源,数据流源源不断^[4];物联网涉及的应用范围广泛,不同领域、不同行业的物联网应用数据通常具有不同的类型和格式,因此物联网中数据多样性更为突出^[5]。

物联网应用提供商拥有物联网感知层和应用层上的数据,而电信网络运营商则拥有物联网的网络层和平台层上的数据,这些数据不是应用层面的数据,但能够通过对这些数据的分析结果,掌握与物联网终端所属行业应用相关的网络行为特征。目前,很多物联网大数据分析是基于感知层和应用层的数据进行的^[6],而很少有运营商对物联网终端在网络中产生的数据进行分析。

电信网络运营商可以通过对网络数据的分析,及时发现物联网终端的异常情况,从而保证物联网终端的正常使用,而且还可以将这些与行业应用相关的物联网终端网络行为数据作为咨询报告形式提供给第三方行业研究机构。

本文主要描述了从网络数据中可以分析出哪些物联网终端网络行为特征以及如何为不同行业应用的物联网终端设计网络行为标签,并基于这些标签进行物联网终端异常行为检测的方法。最后还以可穿戴行业应用为例,描述了对物联网终端网络行为进行监控的分析过程。

2 物联网终端网络行为特征

物联网感知设备基于计算机和通信技术,利用蜂窝移动网络、有线网络、无线网络等完成信息的传输、协同和处理,从而实现物与物通信、物与人通信的网络^[7]。因此,物联网终端相关的网络信息数据可以从电信网络数据中采集

清洗获得,物联网相关的网络数据主要包括:物联网终端语音、短信业务的通信数据、物联网终端上网数据、物联网终端的基本属性数据和位置数据等。

通过这些数据可以获得不同行业应用的物联网终端网络行为特征,如下所示。

(1) 物联网终端的数据通信行为特征

随着 3G、4G 网络的发展,越来越多的物联网终端使用数据上网的通信方式进行信息传输,不同行业应用的物联网终端数据通信行为特征各不相同,主要体现在以下几个方面:终端数据上网访问的时段、频次、在线时长、活跃时长、流量大小以及数据访问方向等。

- 数据上网访问的时段每周可以分为工作日或休息日,每天可以分为 8:00—18:00 的工作时间或 18:00—8:00 的非工作时间。如:税控机行业的物联网终端通常是在工作日的工作时间进行信息传输;视频监控行业的物联网终端可以是在非工作时段内进行信息传输。
- 数据上网的频次可以分为高频次的数据传输、低频次的数据传输或偶发的数据传输。如:金融 POS 终端每日会高频次地进行信息传输;智能井盖在被移动或破坏时才会有信息传输,属于偶发的信息传输行为。
- 数据上网的在线时长是指物联网终端建立数据连接的时间长度,而活跃时长是指物联网终端在连接状态下进行信息传输的时间长度。如:金融 POS 终端虽然长时间处于连接状态,在线时长通常很长,但是金融 POS 终端的活跃时长却很短,通常是在交易发生时,用户在 POS 机上划卡时才会产生信息传输。
- 不同行业应用的物联网终端的数据上网流量大小差别较大。如:智能家居行业应用终端传输的数据量较大;而金融 POS 终端传输的数据量通常都较小。



- 物联网终端的数据访问分为定向和非定向两种。很多物联网终端是向特定地址的应用平台传输数据的，如：智能抄表的终端就是向特定的抄表应用服务器上传抄表数据，这类数据访问就是定向的数据访问；而也有一些物联网终端数据访问的目的地不是特定的地址，而是任意的网络地址，如：车载 Wi-Fi 终端的数据访问地址就不是特定地址，这类数据访问就是非定向的数据访问。

(2) 物联网终端的语音和短信通信行为特征

目前网络中还有一些使用 2G 的语音和短信的通信方式进行信息传输的物联网终端。物联网终端的语音和短信的通信行为特征与数据通信行为特征相类似，包括：语音通话时段、频次、时长以及被叫号码是否为定向号码等；短信通信时段、频次以及接收短信号码是否为定向号码等。

(3) 物联网终端的移动性行为特征

如果物联网终端利用移动网络实现信息传输，那么其在网络中就具备移动性特征。物联网终端的移动性行为特征包括固定位置类型、游牧类型和移动类型等。

有些物联网终端的位置是固定的，而有些物联网终端的位置是不固定的。

对于不连续移动的物联网终端可以称为是游牧类型的终端，如：物流行业的物联网终端，其位置变化过程是不连续的，可能上午在北京进行了信息传输，而下午就出现在上海。对于游牧类型的物联网终端而言，可以根据不同游牧地的个数进行分类，分为游牧位置多的终端和游牧位置少的终端。

对于连续移动的物联网终端可以称为是移动类型的终端，如：车联网行业的物联网终端，其位置变化过程是连续的。对于移动类型的物联网终端而言，可以根据移动速度快慢进行分类，分为快速移动的终端和慢速移动的终端。

表 1 列举了 8 个物联网行业应用的网络行为特征，这些特征都是基于网络数据分析获得的。

综上所述，根据物联网终端在移动网络中所具备的网络特征，可以对物联网终端的网络行为进行描述，并结合物联网终端所属行业、应用分类等基本属性信息，归纳出不同行业物联网终端的网络行为特征。而这些不同行业物联网终端的网络行为特征可以通过网络行为标签进行描述。

表 1 8 个物联网行业应用的网络行为特征

行业类别	网络行为特征
可穿戴	可穿戴终端的上网时间和位置不固定，根据个人行为的不同而不同；可穿戴终端的上下行流量并不大，不超过 10 Mbit/s 的数据流量，数据波动比较大；如果终端不活跃，只有下行不到 1 kbit/s 的数据流量
智能抄表	智能抄表终端长时间在线，终端位置固定，上行流量（120~200 kbit/s）大于下行流量（70~90 kbit/s），终端持续保持活跃，数据传输比较稳定
车载 Wi-Fi	车载 Wi-Fi 终端位置可以移动，终端持续在线，数据传输时间通常从上午 6:00—8:00 左右开始，到晚上 22:00—23:00 结束。上行流量小于下行流量
智能家居	智能家居终端位置固定，通常是白天 9:00 开始，晚上 23:00 或第二天 0:00—2:00 结束，或者是一整天持续数据传输。数据传输量保持稳定，大小一致。数据上下行流量都较大
车联网	车联网终端可以移动，终端持续在线，上网时间和位置并不固定，根据个人行为的不同而不同，车联网终端的上下行流量并不大，但是数据波动比较大
金融 POS	金融 POS 终端可以移动，终端持续在线，数据传输时间较短，上行数据流量大于下行数据流量，数据量小于 100 kbit/s，数据量变化较大
智能监控	智能监控终端位置固定，终端持续在线，数据持续传输，数据流量并不大，上行数据流量（小于 300 kbit/s）大于下行数据流量（小于 250 kbit/s），但是发现有凌晨 0:00 时，下行数据流量（2 Mbit/s 左右）大于上行数据流量（140 kbit/s）。会在凌晨 0:00—2:00 左右出现下行数据流量大于上行数据流量的情况
税控机	税控机终端位置固定，工作时间非常有规律，在工作日（9:00—16:00 左右）终端处于在线状态，上行数据流量（小于 10 Mbit/s）小于下行数据流量（最大可以到 51 Mbit/s 左右）

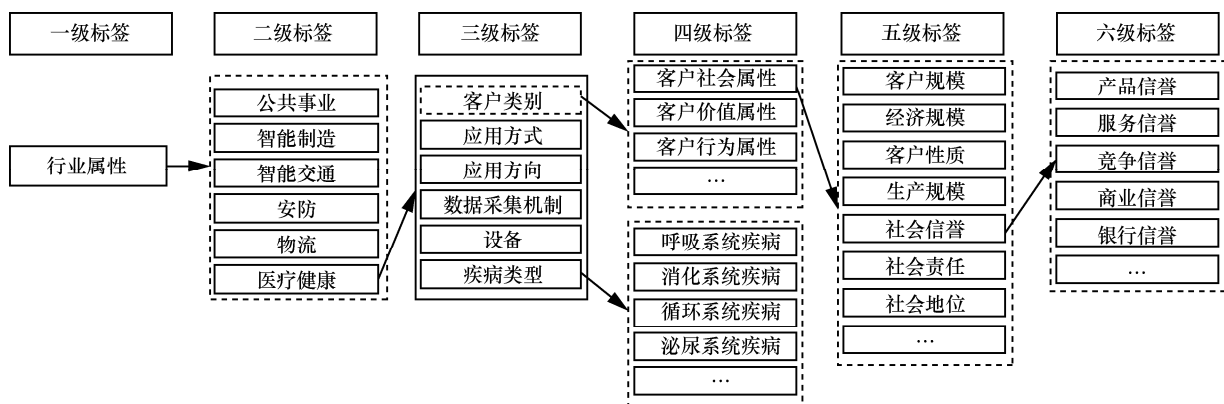


图1 物联网行业属性标签

3 网络行为标签设计

电信网络运营商设计的物联网行业用户的标签体系,可以通过物联网行业属性标签、物联网终端网络行为静态标签、物联网终端网络行为动态标签三大标签模块共同构成。静态标签是直接从原始数据中提取出来的标签,而动态标签是通过原始数据计算汇总后得到的统计数据中提取出来的标签。

物联网行业属性标签如图1所示,这类以行业应用特征为主的标签模块,来源于对物联网行业的深入研究,不依赖于电信所拥有的用户数据。

物联网终端网络行为静态标签如图2所示。

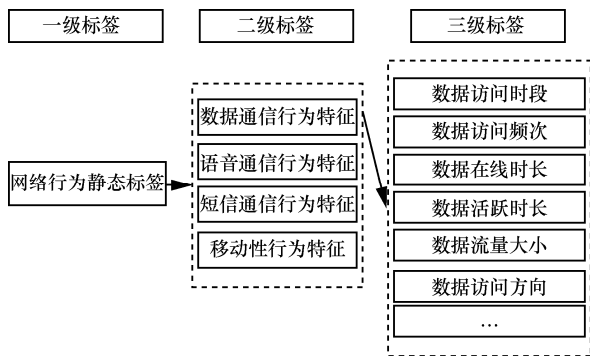


图2 物联网终端网络行为静态标签

物联网终端网络行为动态标签如图3所示。

综上所述,通过物联网行业属性标签和网络行为标签结合的定性分析结果,可以提取出物联网终端的关键特征,形成物联网行业群体的网络行为画像。

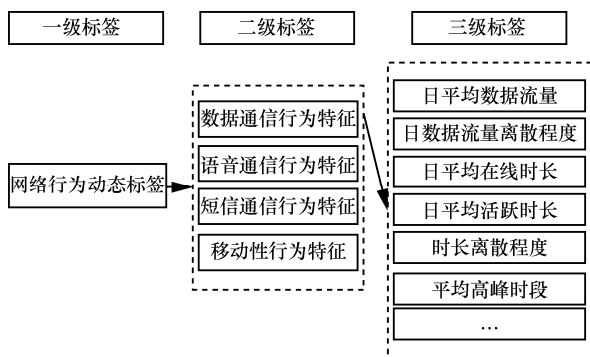


图3 物联网终端网络行为动态标签

4 网络行为监控模型设计

基于物联网行业应用的网络行为标签体系,可以实现各种应用,对物联网网络行为进行监控就是其中一种重要的应用。

物联网网络行为监控应用需要对物联网终端异常行为进行分析,旨在寻找具有异常特征的物联网终端。首先是基于物联网终端的网络数据,对各行业物联网行为进行归纳,提取特征,建立分析维度;然后对各分析维度进行数据清洗和抽取,建立分析维度字段;接下来选择合适的算法进行异常行为分析,找到离群点,建立分析模型;最后进行模型的应用和评估。

不同行业的物联网终端有不同的网络行为特征,因此,物联网终端异常行为的分析需要分行业进行,需要依据现有样本及物联网终端的网络行为标签,通过对不同行业物联网终端离群点的



检测,分行业标注异常样本点,建立模型训练集,从而构建机器学习分类器,预测物联网终端是否异常。

依据上述物联网终端的网络行为标签,找到离群点,作为具有异常行为的终端,为下一步构建分类器做准备。寻找离群点的方法大致可以分为两类:一类是基于统计分布模型的方法,例如,假定总体服从高斯分布,落在 3 倍标准差之外的样本,认为是离群点,这种方法比较适合进行单变量分析;另一类是基于距离的方法,主要有 k NN 算法、聚类算法以及局部异常因子(local outlier factor, LOF)算法。

k NN 算法是有监督的学习,不适合现在的场景;局部异常因子算法的时间复杂度很高,不适合大数据集;而聚类算法效率较高,适用于大数据集,所以这里采用的是基于聚类算法的离群点检测。基于聚类算法的离群点检测的主要步骤如下。

步骤 1 对样本数据进行聚类。

步骤 2 计算每个样本的离群因子,将离群因子较大的对象判定为离群点。

假设样本集 D 被聚类算法划分为 k 个簇 $C=\{C_1, C_2, \dots, C_k\}$, 样本 x 的离群因子(outlier factor) $OF(x)$ 定义为 x 与所有簇间距离的加权平均值:

$$OF(x) = \sum_{j=1}^k \frac{|C_j|}{|D|} \cdot d(x, C_j) \quad (1)$$

进一步计算所有样本离群因子均值(AVG_OF)以及离群因子标准差(STD_OF),从而得到离群因子阈值:

$$THRESHOLD_OF = AVG_OF + \beta \cdot STD_OF \quad (2)$$

大于该阈值的样本认为是异常样本。

上述是基于聚类算法找到现有样本集中具有异常行为的终端,然后根据终端行为特征标签,对正负样本集进行训练,构建有监督学习的分类器,并固化终端异常行为检测模型。

5 可穿戴终端的网络行为分析

可穿戴终端的上网时间和位置并不固定,根据个人行为的不同而不同;可穿戴终端的日均流量均值达到 71.9 Mbit/s,中位数是 11.4 Mbit/s,数据波动比较大;如果终端不活跃,只有下行流量不到 1 kbit/s 的数据流量。

5.1 数据流量

从数据流量角度分析,对可穿戴物联网终端使用的日均流量进行统计,日均流量是用户使用的总流量与发生流量的天数之比。表 2 反映了可穿戴终端 4G 话单下日均流量统计指标值的情况。

表 2 可穿戴日均流量统计指标值(4G)

统计指标	日均流量指标值/(Mbit·s ⁻¹)
1/10 分位数	0.001
上四分位数	1.053
中位数	11.419
下四分位数	44.830
9/10 分位数	169.813
均值	71.923
标准差	169.926

在 4G 情形下,可穿戴终端日均流量数据偏差较大,有 10%的终端日均流量在 1 Mbit/s 以下;同时,有 10%的终端日均流量在 169 Mbit/s 以上。低流量终端可能是不活跃的终端,而对于流量大于 169 Mbit/s 的终端可能是存在异常的高流量终端。

在 3G 情形下,通过极大似然估计得到的可穿戴终端日均流量均值和标准差结果见表 3。

表 3 可穿戴日均流量统计指标值(3G)

统计指标	日均流量指标值/(Mbit·s ⁻¹)
均值	2.690
标准差	41.828

基于高斯分布模型,在 3G 情形下,日均流量大于 87 Mbit/s 的概率是 0.022 8,因此,根据统计分布模型异常检测思想,当可穿戴终端在 3G 下的业

务数据流量大于 87 Mbit/s 时,可能是高流量异常终端。

5.2 在线时长

从可穿戴终端在线时长维度分析,可穿戴终端在线时长因人而异,其在线时长包含终端和基站交互的静默时长。对可穿戴物联网终端日均在线时长进行统计,统计指标值结果见表 4。

表 4 可穿戴日均在线时长统计指标值 (4G)

统计指标	日均在线时长指标值/h
1/10 分位数	1.573
上四分位数	7.760
中位数	17.002
下四分位数	24.000
9/10 分位数	24.407
均值	16.267
标准差	10.850

在 4G 情形下,可穿戴终端日均在线时长中位数为 17 h,均值为 16 h,偏度较小,在线时长分布较为集中。对于日均在线时长大于 24 h 的终端,是数据质量异常。

在 3G 情形下,通过极大似然估计得到的可穿戴终端日均在线时长均值和标准差见表 5。

表 5 可穿戴日均在线时长统计指标值 (3G)

统计指标	日均在线时长指标值/h
均值	4.187
标准差	9.445

基于高斯分布模型,在 3G 情形下,日均在线时长大于 22 h 的概率是 0.022 8,因此,根据统计分布模型异常检测思想,当可穿戴终端在 3G 下的在线时长大于 22 h 时,可能是存在异常终端。

5.3 活跃时长

对可穿戴终端的活跃时长进行统计分析,活跃时长是终端产生业务数据流量的时长,反映终端的活跃程度。各指标统计结果见表 6。

在 4G 情形下,终端活跃时长均值为 5 h,中位数为 4 h,在统计的样本中,有 10%的终端活跃时长不足 1 h,属于低活跃终端;另有 10%的终端活跃时长超过 14 h,属于异常高活跃终端。

表 6 可穿戴日均活跃时长统计指标值 (4G)

统计指标	日均活跃时长指标值/h
1/10 分位数	0.278
上四分位数	1.338
中位数	4.082
下四分位数	8.411
9/10 分位数	14.223
均值	5.368
标准差	5.674

在 3G 情形下,通过极大似然估计得到的可穿戴终端日均活跃时长均值和标准差见表 7。

表 7 可穿戴日均活跃时长统计指标值 (3G)

统计指标	日均活跃时长指标值/h
均值	3.613
标准差	8.187

基于高斯分布模型,在 3G 情形下,日均活跃时长大于 19 h 的概率是 0.022 8,因此,根据统计分布模型异常检测思想,当可穿戴终端在 3G 下的活跃时长大于 19 h,可能是异常高活跃终端。

5.4 上网频次

对可穿戴终端上网频次进行统计,通过极大似然估计得到的可穿戴终端日均上网频次的均值和标准差,4G 和 3G 情形下的日均上网频次分布模型参数见表 8。

表 8 可穿戴日均上网频次分布模型参数

统计指标	日均上网频次 (4G) /次	日均上网频次 (3G) /次
均值	12.388	50.549
标准差	90.786	55.621

基于高斯分布模型,在 4G 情形下,日均上网频次大于 192 次的概率是 0.022 8,因此,根据统计分布模型异常检测思想,当可穿戴终端在 4G 下的上网频次大于 192 次时,可能是异常高频次终端。同理,在 3G 情形下,上网频次大于 160 次时,可能是异常高频终端。

6 结束语

本文提出的通过分析电信网络数据,制定物



联网终端的网络行为标签，并对物联网终端网络行为进行实时监控的方法。可以广泛应用在物联网业务运营分析系统中，可以为不同行业的物联网应用打上网络行为标签，从而实现对物联网业务运营情况的实时分析和监控，有利于物联网应用在趋势分析、产品规划和营销决策等层面的工作。

参考文献：

- [1] ITU. Overview of the Internet of things: ITU-T Y.2060[S]. 2012.
- [2] 工业和信息化部. 物联网总体框架与技术要求：YD/T 2437-2012, CCSA[S]. 2012.
MIIT. General framework and technical requirements of IoT: YD/T 2437-2012, CCSA[S]. 2012.
- [3] 罗伯特·斯特科维卡, 艾特·利希特, 维诺·曼萨, 等. 大数据与物联网：企业信息化建设新时代[M]. 刘春容, 译. 北京：机械工业出版社, 2016.
STACKOWIAK R, LICHT A, MANTHA V, et al. Big data and Internet of things: enterprise information architecture for a new age[M]. Translated by LIU C R. Beijing: China Machine Press, 2016.
- [4] 梁循, 杨小平, 赵吉超. 大数据物联网复杂信息系统[M]. 北京：清华大学出版社, 2017.
LIANG X, YANG X P, ZHAO J C. Big data IoT complex information system[M]. Beijing: Tsinghua University Press, 2017.
- [5] 王桂玲, 王强, 赵卓峰, 等. 物联网大数据处理技术与实践[M]. 北京：电子工业出版社, 2017.
WANG G L, WANG Q, ZHAO Z F, et al. The technology and

practice of large data processing in the Internet of things[M]. Beijing: Publishing House of Electronics Industry, 2017.

- [6] 尼克·贝瑟斯, 斯普莱恩·杜布里. 大数据与物联网：面向智慧环境路线图[M]. 郭建胜, 周竞赛, 毛声, 等译. 北京：国防工业出版社, 2017.
BESSIS N, DOBRE C. Big data and Internet of things: a roadmap for smart environment[M]. Translated by GUO J S, ZHOU J S, MAO S, et al. Beijing: National Defense Industry Press, 2017.
- [7] 魏颖琪, 林玮平, 李颖. 物联网智能终端技术研究[J]. 电信科学, 2015, 31(8): 146-152.
WEI Y Q, LIN W P, LI Y, et al. Study on key technologies of intelligent IoT device[J]. Telecommunications Science, 2015, 31(8): 146-152.

[作者简介]



凌颖（1973—），女，中国电信股份有限公司上海研究院高级工程师，主要研究方向为大数据、物联网网络数据分析、数据管理等。



邱芸（1980—），女，中国电信上海研究院网络数据系统架构师、中级工程师，主要从事基于网络数据的应用研究开发、物联网网络数据分析方面的工作。