

文章编号:1007-757X(2020)10-0079-03

基于 BP 神经网络的异常轨迹检测方法

贾瑛

(宝鸡职业技术学院机电信息学院, 陕西 宝鸡 721013)

摘要: 异常轨迹检测算法通常不能依靠轨迹内外部属性而有效地进行检测, 具有较大的应用局限性。针对此类问题, 基于 BP 神经网络提出了一种改进的异常轨迹检测方法。首先对原始轨迹数据去噪, 并上传百度云 LBS 云端储存; 其次基于百度地图轨迹数据可视网站进行了数据归一化处理, 得到了轨迹的属性值; 最后以轨迹内外特征属性代表 BP 神经网络算法的输入层, 以轨迹相似度度量代表输出层, 对隐含层系数调整后获得训练模型。研究针对 Deolfe 项目的两个用户轨迹数据做了仿真实验, 用以检测用户异常轨迹数据。结果证明, 在选取的最优训练方案基础上, 两组数据的异常轨迹检测正确率各达 92.3% 及 100%, 所搭建模型能够作为异常轨迹检测的工具。

关键词: 轨迹数据集; BP 神经网络; 百度 LBS 云服务; 轨迹属性; 异常轨迹监测

中图分类号: TP 3111

文献标志码: A

Abnormal Trajectory Detection Method Based on BP Neural Network

JIA Ying

(College of Mechanical and Electrical Information, Baoji Vocational & Technical College, Baoji 721013, China)

Abstract: Existing abnormal trajectory detection algorithms usually cannot rely on the internal and external attributes of the trajectory for effective detection, which leads to great application limitations. To solve these problems, an improved abnormal trajectory detection method based on BP neural network is proposed. Firstly, the original trajectory data are de-noised and uploaded to Baidu Cloud LBS for storage. Secondly, the data are normalized based on Baidu Map Trajectory Data Visual Website, and the attribute values of the trajectory are obtained. Finally, the input layer of the BP neural network algorithm is represented by the feature attributes inside and outside the trajectory, and the output is represented by the trajectory similarity measure. The training model is obtained by adjusting the coefficients of hidden layers. In this paper, two sets of user trajectory data of Deolfe project are simulated to detect user abnormal trajectory data. The results show that, on the basis of the optimal training scheme, the accuracies of abnormal trajectory detection of the two sets are 92.3% and 100%, respectively. The model can be used as a tool for abnormal trajectory detection.

Key words: trajectory data set; BP neural network; Baidu LBS Cloud Service; trajectory attributes; abnormal trajectory monitoring

0 引言

随着移动通信和无线网络技术、GPS、Wi-Fi、蓝牙室内定位技术的快速发展, 及智能设备的广泛应用, 人们对移动对象位置和轨迹数据获取也非常方便^[1-2]。同时, 随着轨迹数据储存技术的飞速发展, 能将大量移动对象位置及轨迹信息储存到时空数据库^[3-6]。就轨迹来说, 它所指是带时间标签的位置信息而构成的有序位置序列, 且时空信息很丰富, 并对轨迹数据的分析及挖掘起着很重要作用^[7-9]。而异常轨迹检测属于轨迹模式挖掘的主要研究课题, 并普遍用于智能交通与用户行为分析中。

1 异常轨迹监测问题描述

目前, 对异常轨迹检测的研究主要集中在轨迹空间特征上, 极少将轨迹本身属性作为研究基础^[10-12]。对非线性轨迹属性数据的分类, 是检测轨迹数据是否异常的一种可行轨迹分类方法。而 BP 神经网络具备对非线性数据分类功能, 所以文章提出一种基于 BP 神经网络的异常轨迹检测方法。

1.1 传递函数设计

传递函数应用的是非线性函数 $\text{logsig}()$, 而这一对数 S 形函数曲线, 如图 1 所示。

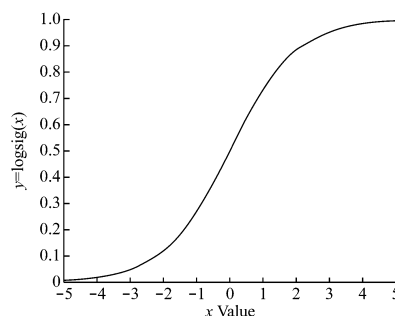


图 1 对数 S 形函数曲线图

其中 $\text{logsig}(x) = \frac{1}{1+e^{-x}}$, 计算出 0 到 1 之间的输出。将轨迹划分成两个类别, 那么这个函数表示为传递函数 0 和 1 的输出, 其中 0 为异常轨迹, 1 为正产轨迹。

1.2 网络结构设计

本文将选用多输入和多隐含的神经元与单输出的 BP 神经网络。单层 BP 神经网络结构图^[13-16], 如图 2 所示。其中, BP 神经网络输入节点由 x_i 表示, 而隐节点由 y_j 表示, 输出节点由 O_i 表示。 w_{ij} 为输入节点和隐节点网络权值, T_j 为隐节点和输出节点间网络权值。

作者简介: 贾瑛(1983-), 女, 硕士, 副教授, 研究方向: 计算机网络和多媒体技术, 物联网。

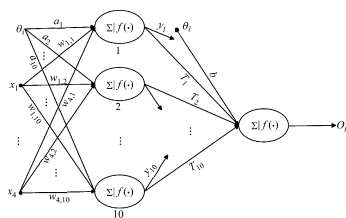


图2 单层BP神经网络结构图

2 基于BP神经网络的异常轨迹监测

本文以BP神经网络作为工具训练,从中获取了轨迹异常检测模型,按照轨迹本身四个属性,实现了用户轨迹是否为异常的目标检测,如图3所示。

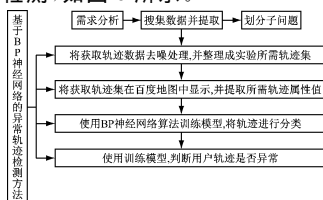


图3 研究方案示意图

2.1 轨迹属性值提取

2.2 轨迹坐标处理

由于轨迹是以经纬度来表示,并且每5秒则会对坐标点采样,所以坐标点间的经纬度值相差极小。为了便于计算,并不丢失轨迹本身的特征,文章运用了将轨迹坐标点直接乘以适当系数的方法,而实验表明这一方法是可行的。具体操作方法为:设置一条轨迹 T (其公式为 $T = \{(t_1, x_1, y_1), (t_2, x_2, y_2), \dots, (t_n, x_n, y_n)\}$,其中时空三元组 (t_i, x_i, y_i) 代表轨迹 T 中的第 i 个位置点,而 t_i 则为这个位置点的采样时刻, (x_i, y_i) 为 t_i 时刻的GPS坐标($1 \leq i \leq n$)),将轨迹集中 x_i, y_i 分别乘以系数 λ ,而这一系数选取取决于实际情况,这样便于计算为原则。在实验中 $\lambda = 1\ 000$,经过处理后的轨迹则为 $T' = \{(x'_1, y'_1), (x'_2, y'_2), \dots, (x'_n, y'_n)\}$ 。因此,对轨迹数据集 TS 处理后的数据集则为 TS' 。

2.3 利用BP神经网络算法训练监测模型

2.3.1 基于BP神经网络的异常轨迹监测算法流程

BP神经网络算法是一种基于梯度下降的学习算法。可以将学习过程划分成两个阶段,即信息正向传递、误差反向传播^[17-19],如图4所示。

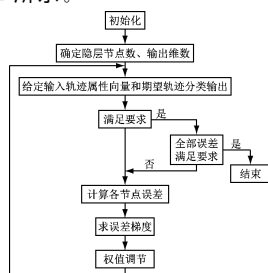


图4 BP神经网络算法流程图

2.3.2 基于BP神经网络的异常轨迹检测算法描述

构建一个三层BP神经网络,其指令为:

$net = \text{newff}(\text{PR}, [S_1, S_2, \dots, S_N], [\text{TF}_1, \text{TF}_2, \dots, \text{TF}_N], \text{BTF}, \text{BLF}, \text{PF})$

其中,PR为输入向量取值范围, S_i 则为第 i ($i = 1, 2, \dots, N$)层神经元个数,共计 N 层; TF_i 为第 i 层传递函数;BTF表示为BP神经网络训练函数,BLF表示BP神经网络权值与阈值学习函数;PF表示性能函数,而试验所用最小均方误差(mse)为性能函数。

执行结果为构建一个 N 层BP神经网络模型。对带分

类标签的数据集,先对设计的模型进行现有轨迹训练,得到稳定模型,然后将未知的轨迹放入模型中检测判断是否为异常轨迹。

3 实验

在文章中的所提出的算法都是在Matlab2016a环境下实现的,硬件和软件环境为: Intel(R) Core(TM) 2 Duo 3.3GHz CPU,内存为4GB,操作系统为Windows10。

3.1 数据集选取

本次实验数据主要源自于GeoLife GPS Trajectories项目中的某两个用户的轨迹数据集。由于轨迹集所使用的是格林威治时间,所以从23:00-00:59、9:00-10:59等这一时段转换为北京时间则为7:00-8:59、17:00-18:59,而这一时段正属于上下班高峰期。因此,本文所选用的轨迹数据集为某两个用户2018年9月23日-12月23日期间每天这一时段的轨迹数据,共计180条。

3.2 试验结果

3.2.1 BP神经网络训练及仿真结果

在本次实验中,对第一个用户的90条轨迹属性数据给予分组,根据分组结果不同,提出了五种训练方案。在方案1中,30条轨迹属性数据用于训练,剩余数据用于仿真;在方案2中,40条轨迹属性数据用于训练,剩余数据用于仿真;在方案3中,50条轨迹属性数据用于训练,剩余数据用于仿真;在方案4中,60条轨迹属性数据用于训练,剩余数据用于仿真;在方案5中,70条轨迹属性数据用于训练,剩余数据用于仿真。此外,由这五种方案所生成的精度、召回率、F度量值柱状图,如图5所示。

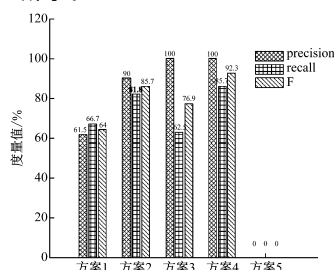


图5 五种方案结果对比

精度是对精准性的度量,也就是标记为正类的元组或异常轨迹元组的实际正类所占百分比,其表达式为:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

公式中的TP代表为正确分类的异常轨迹条数,而FP代表的是异常轨迹错误标记的条数。

所谓的召回率则是指对完全性度量,也就是将正元组或是异常轨迹元组标记为正类的百分比,而其表达式则为:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

公式中的FN代表为将正常轨迹错误标记成异常轨迹的条数。

而F度量所指的是把精度与召回率组合一个度量中,那么其表达式为:

$$F = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

由图5得知,在这五种方案中只有方案4的性能最好,它的F度量值达到92.3%,所以利用此方案对另一用户的轨迹数据集实行仿真实验,也就是使用另一用户的60条轨迹属性数据用作训练,而剩余数据则用作仿真实验。最终获得此次轨迹检测模型的输出类别与实际类别的对比结果。

3.2.2 网络性能分析

采用方案4对另一用户的仿真训练的BP神经网络性

能,如图 6 所示。

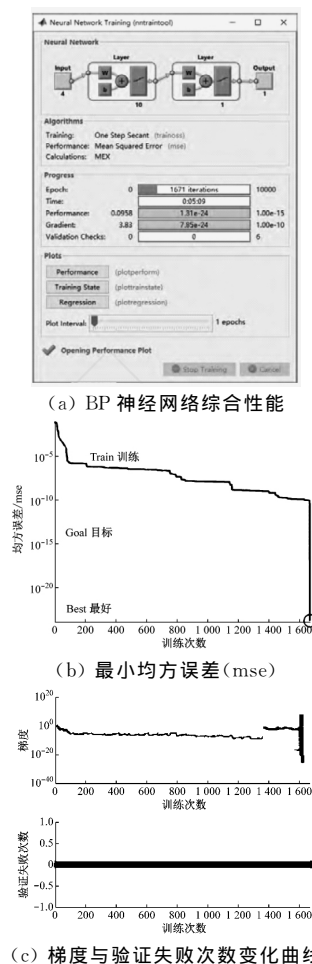


图 6 BP 神经网络性能

通过图 6(a)BP 神经网络综合性能可以得知,此次实验选用的是三层神经网络,第一次为输入层有 4 个输入节点,第二次为隐含层有 10 个隐节点,第三次为输出层有 1 个输出节点。因此得知实验使用的 BP 神经网络训练算法为 trainoss,而使用的性能函数则为均方误差函数 mse,并且这个模型的均方误差 mse 值为 1.31×10^{-24} ,梯度为 7.85×10^{-24} ,最大验证失败 0 次。用时 5 分 9 秒共迭代 1 671 次,说明这个模型非常稳定,而且已经满足精度要求,所以能够作为用户轨迹聚类与异常轨迹检测的模型。由图 6(b)得知,最小均方误差在 10^{-11} 时缓慢下降,然而在训练次数达到最大时,将会突然降到 10^{-24} ,这是因为实验的训练方法为 trainoss,它属于快速训练方法,所以在网络性能到达一定精度时将会快速收敛。由图 6(c)得知,这种算法将会始终保持平稳梯度下降,且验证失败始终为 0 次,且训练次数达到足够大时,将会呈直线下降,究其原因则等同于 mse 突然下降原因。

4 总结

文章提供的基于 BP 神经网络异常轨迹检测方法具有两个阶段。其一,先对轨迹数据收集,然后去除数据噪音并上传到百度云 LBS,云端储存,最后设计轨迹显示系统,实现在百度地图上显示轨迹数据集。其二,利用轨迹属性提取算法对轨迹集进行预处理,从中获取轨迹属性向量,如轨迹长度属性、轨迹角度属性、轨迹时长属性、轨迹平均速度等。其次以轨迹特征属性看作 BP 神经网络算法输入层,以轨迹相似度量作为输出层,然后训练出稳定的异常轨迹检测模型。最

后用异常轨迹检测模型判断用户轨迹是否为异常,便于获取异常轨迹数据。

参考文献

- [1] Zheng Y. Trajectory Data Mining: An Overview[J]. ACM, 2015, 6(3): 29.
- [2] Prelipcean A C, Gidofalvi G, Susilo Y O. Measures of transport mode segmentation of trajectories[J]. International Journal of Geographical Information Science, 2016: 1-22.
- [3] Lü M, Chen L, Xu Z, et al. The discovery of personally semantic places based on trajectory data mining[J]. Neurocomputing, 2016, 173(3): 1142-1153.
- [4] Giannotti F, Nanni M, Pedreschi D, et al. Unveiling the complexity of human mobility by querying and mining massive trajectory data[J]. The VLDB Journal, 2011, 20(5): 695.
- [5] Gupta M, Gao J, Aggarwal C C. Outlier detection for temporal data: A survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 25(1): 1-20.
- [6] Hu W, Li X, Tian G, et al. An incremental DPMM-based method for trajectory clustering, modeling, and retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(5): 1051-1065.
- [7] Cai Y, Wang H, Chen X, et al. Trajectory-based anomalous behaviour detection for intelligent traffic surveillance[J]. IET Intelligent Transport Systems, 2015, 9(8): 810-816.
- [8] Laxhammar R, Falkman G. Online learning and sequential anomaly detection in trajectories[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(6): 1158-1173.
- [9] Shen M, Liu D R, Shann S H. Outlier detection from vehicle trajectories to discover roaming events[J]. Information Sciences, 2015, 294(9): 242-254.
- [10] 王嵘冰,徐红艳,李波. BP 神经网络隐含层节点数确定方法研究[J]. 计算机技术与发展, 2018(4): 1-6.
- [11] 朱燕,李宏伟,樊超. 基于聚类的出租车异常轨迹检测[J]. 计算机工程, 2017, 43(2): 16-20.
- [12] 闫密巧,王占宏,王志宇. 基于 Redis 的海量轨迹数据存储模型研究[J]. 微型电脑应用, 2017, 33(4): 9-11.
- [13] 毛嘉莉,金澈清,章志刚. 轨迹大数据异常检测: 研究进展及系统框架[J]. 软件学报, 2017, 28(1): 17-34.
- [14] 俞庆英,李倩,陈传明,等. 基于 BP 神经网络的异常轨迹检测方法研究[J/OL]. 计算机工程. <https://doi.org/10.19678/j.issn.1000-3428.0051574>
- [15] 时磊,鲁华栋. 基于高斯混合模型的移动目标跟踪[J]. 微型电脑应用, 2017, 33(12): 75-77.
- [16] 崔宝才. 基于 GA 改进 BP 神经网络异常检测方法[J]. 现代电子技术, 2016, 39(3): 90-93.
- [17] 韩旭. 基于车辆轨迹多特征的聚类分析及异常检测方法的研究[D]. 哈尔滨: 哈尔滨工程大学, 2014.
- [18] 鲍苏宁,张磊,杨光. 基于核主成分分析的异常轨迹检测方法[J]. 计算机应用, 2014, 34(7): 2107-2110.
- [19] 龚畅,王华君,李荣,等. 利用两阶段遗传算法的机械手最优轨迹控制仿真[J]. 微型电脑应用, 2016, 32(4): 18-22.

(收稿日期: 2019.06.26)