基于神经网络的交通流预测方法研究

王东 鲍立

湖南大学信息科学与工程学院

**概要（Abstract）**

交通拥堵评价不仅是道路交通服务水平的重要依据，也是交通管理与控制的前提。准确地评价交通拥堵状态，对道路拥堵程度评价、交通疏导以及最佳路径规划均具有非常重要的意义。但是，由于实际道路中交通数据获取的难度较大，信息共享程度较低，大多数研究工作还是停留在理论以及实验阶段。因此对实际道路的交通数据分析还有较大的提升空间。本文基于实际道路的大量交通数据（卡口数据）对交通拥堵状态的评价与预测方法进行了进一步的研究，主要取得了以下两个方面的研究成果：

1. 对该场景实际交通数据获得拥堵状态的评价方法。首先使用了数据挖掘过程中比较高效的聚类方法，针对一般的K-means等聚类方法对初始簇中心严重依赖、聚类结果较不稳定的问题，提出用ROCK算法来进行聚类，最终依据聚类结果得出交通拥堵状态的划分数值。拥堵状态是交通状态分析的重点，通过聚类方法从大量交通数据中处理得到该场景不同状态对应的划分数值，显得更加灵活和可信。
2. 对短时交通拥堵状态进行预测的过程当中，针对常见BP神经网络模型直接预测交通拥堵状态时存在有些情况预测结果不够准确的问题，提出了在神经网络的预测阶段使用BP神经网络结合统计学理论进行联合预测得到最终预测模型的短时交通拥堵状态预测方法。直接通过神经网络模型的预测正确率为95.1%，而在预测阶段结合统计学方法联合校验得到最终预测模型后进行预测的正确率为97.5%。该方法对交通拥堵状态的预测的正确率经试验结果对比证明有了一定的提升。

**关键词：城市道路；交通拥堵状态评价；短时交通拥堵预测；神经网络；统计方法**

**引言（Introduction）**

随着我国社会经济的快速发展，城市化进程的不断加深以及人民生活水平的不断提高，城市交通拥堵问题也越加明显。交通拥堵制约了城市发展，严重延长了居民的出行时间，提高了出行成本，降低了生活效率。同时，低速行驶会增加燃油消耗，浪费大量能源，大量尾气排放也是城市环境质量恶化的重要原因。因此，有必要对交通拥堵状况进行评价与预测，用于信号控制与交通诱导，从而缓解交通拥堵状况。

城市交通建设是一项系统化的工程。既要满足当前的交通供需平衡，还要考虑城市未来的整体发展规划。当前城市交通拥堵的重要原因之一，是缺乏科学整体的交通发展规划。有些城市过于热衷大型交通工程，如地铁、轻轨，缺忽视了普通公共交通建设，引起结构性的“负效应”。城市交通需要整体全面的规划。在建设大型交通工程的同时，也不能忽视了普通公共交通的建设。城市交通拥堵问题的传统解决方法有两种：一是，控制需求，最直接的方法是控制车辆数量增长速度；二是，增加供给，也就是加强基础道路建设。但这两种方法都存在局限性。我国汽车产业正处于发展的关键阶段，限制车辆增长会使其收到沉重的打击；另外，对于城市周边并不十分拥堵，“一刀切”的限制车辆增长也是不合适的。而在资源、土地、环境矛盾越来越突出的今天，道路的修建也受到各方面的限制。这就迫切需要新的方法来满足日益增长的道路交通需求。随着“智慧城市”、“智慧交通”概念的提出，将新一代信息技术用于解决城市道路交通拥堵问题，实现城市交通的现代化，成为专家学者研究的重点。其中，智能交通系统方面的研究成为工作的重点。而在智能交通系统中，对道路交通流情况预测的准确程度起着一定的关键性作用，为此，我们有必要对交通流的预测方法进行不断的研究和改进。

通过神经网络建模对交通流拥堵情况进行预测是一个常见且比较高效的方法。基于神经网络建模进行预测的优点:(1)神经网络有很强的非线性拟合能力，可映射任意复杂的非线性关系;(2)学习规则简单，便于计算机实现。(3)具有很强的鲁棒性、记忆能力、非线性映射能力以及强大的自学习能力，因此有很大的应用市场。基于神经网络建模进行预测的缺点:（1）最严重的问题是没能力来解释自己的推理过程和推理依据。（2）不能向用户提出必要的询问，而且当数据不充分的时候，神经网络就无法进行工作。（3）把一切问题的特征都变为数字，把一切推理都变为数值计算，其结果势必是丢失信息。（4）理论和学习算法还有待于进一步完善和提高。

本文在使用基于聚类的方法用来得到判断某个时间段的拥堵程度的依据值后，提出了一种在经典神经网络的预测阶段结合统计学理论得到最终预测模型来进行预测的思路对道路交通流短时拥堵情况进行预测。即在经典BP神经网络方法训练出预测模型的基础上，预测过程中结合统计方法联合校验进行预测。在第一阶段，训练得到经典的BP神经网络模型；在第二个阶段，在训练得到的神经网络模型的基础上，预测阶段结合统计学中的平均值以及标准差联合判断校验得到最终的预测结果。

文章接下来的部分章节安排如下：参考文献部分介绍了文献查阅情况；方法提出部分介绍了本文提出的方法；提出算法部分介绍了本文提出的详细算法；实验结果部分给出了实际实验的结果情况；最后，本文以总结模块结尾。

**研究背景（Literature survey）**

2004年,McDonald等人以高速公路.上浮动车为研究对象，提出了一种基于浮动车数据的交通拥堵判别算法。该算法原理是交通量增加造成的交通拥堵，导致车辆通过该路段的行程时间会急剧增加，当行程时间增加幅度大于该路段正常通过时间时，判定发生交通拥堵状况;2007年，Georgiou等人通过在特定的高速公路上安装的检测器得到的交通数据，并设计了一种随机宏观交通流模型，借助扩展卡尔曼滤波算法来对实际道路拥堵状况进行估计。由于作者对该方法使用到的交通模型参数初始值、相关噪声的偏差值都进行了认真的调查，所以该方法取得了十分理想的效果;2013年，林信明等人对城市环路的交通拥挤状态进行了分析研究,提出了一种交通状态自适应判别算法。该算法结合了模糊C均值聚类方法以及灰色聚类方法；通过模糊C均值聚类方法得到各个交通状态聚类中心以及边界阈值，然后利用灰色聚类的方法进行交通状态判别。该方法避免了模糊C均值聚类方法可能陷入局部最优解的问题，同时对灰色聚类方法的参数选择进行了一定的约束。实践显示该方法对不同道路的交通状况，能够有针对性的进行交通拥堵状态评价；2014年，倪升华在对短时交通拥堵状态进行预测时，针对一阶马尔科夫模型预测交通拥堵状态，存在预测拥堵状态滞后的问题，提出了一种高阶马尔科夫模型的短时交通拥堵状态预测方法。该方法对交通拥堵状态的预测准确度有了一定提升达到92.7%，并且有效的消除了预测滞后问题；2014年，管硕采用K-均值聚类算法的RBF神经网，对市区内的交通流进行预测，相对于普通的BP神经网络，获得了更高的预测精度和准确度；2017年，张亚茹将模糊c均值聚类算法与Canopy算法结合，用Canopy算法初始化聚类中心与聚类数量，同时引入了Xie-Beni指标来自适应确定模糊加权指数，取得了更好的聚类效果。

查阅文献可以看到，大部分研究都是针对某个具体类型模型或者算法本身的改善，以达到提高判断准确度的目的。而在本文中，利用统计学中的部分理论，结合经典的神经网络建立得到最终预测模型，提出了一种在神经网络训练得到预测模型的基础上结合统计学方法得到最终模型进行联合校验的预测方案。直接通过神经网络模型的预测正确率为95.1%，进一步结合统计学理论联合预测的正确率为97.5%。在对实验数据的预测结果表明，提出的方法较一般的神经网络模型的预测方法在正确率上有了一定的提高。

**交通数据集（Traffic database）**

本文实验的数据集为在OpenITS网站上获得的符合实验需求的数据集：获得了宣城市（位于安徽省东南部）的卡口12月3日到12月9日，为期一周的卡口车辆通过信息数据集。本文通过对获得交通数据集基础之上进行处理得到实验需要的数据格式。

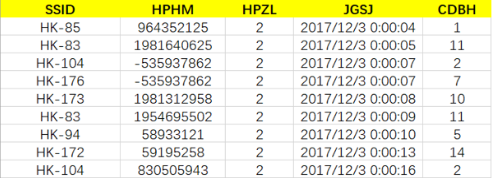


图1 卡口数据集截图

|  |  |  |
| --- | --- | --- |
| 字段名 | 数据类型 | 字段说明 |
| SSID | VARCHAR(20) | 卡口编号 |
| HPHM | VARCHAR(20) | 车辆编号 |
| HPZL | NUMBER(8) | 号牌种类 |
| JGSJ | DATE | 车辆经过卡口时刻 |
| CDBH | NUMBER(8) | 车辆经过车道编号 |

表1 卡口数据集字段说明

**方法提出（Proposed methodology）**

第一部分：基于ROCK算法对交通数据进行聚类

ROCK 算法中用到的四个关键概念：

邻居（Neighbors）：如果两个样本点的相似度达到了阈值（θ），这两个样本点就是邻居。阈值（θ）由用户指定，相似度也是通过用户指定的相似度函数计算。

链接（Links）：两个对象的共同邻居数量。

目标函数（Criterion Function）：最大化下面目标函数以获得最优的聚类结果（最终簇之间的链接总数最小，而簇内的链接总数最大）。Ci：第i个簇，k:簇的个数，ni:Ci的大小（样本点的数量）。一般可使用f (θ) = (1-θ)/(1+θ). f(θ)一般具有以下性质：Ci中的每个样本点在Ci中有nif(θ)个邻居。

相似性的度量（Goodness Measure）：使用该公式计算所有对象的两两相似度，将相似性最高的两个对象合并。通过该相似性度量不断的凝聚对象至k个簇，最终计算上面目标函数值必然是最大的。

第二部分：结合神经网络模型和统计学方法对交通流的拥堵状况进行预测的方法

本文提出了一种神经网络建模结合统计学方法得到最终模型的思路来对上述获取的交通数据集进行实验和预测。在对数据集进行处理的过程中，通过对获得的数据集中车辆路过卡口的时间进行整合，最后处理得到每单位时间（5分钟）卡口通过的车辆的数量。然后进一步对着处理的到的数据进行聚类处理，根据聚类结果得到判别拥挤（busy）、一般（normal）或是空闲（free）的数量上的标准，并以此判断出各单位时间的交通拥堵状态。最后我们将获得的所有数据以及其对应的交通拥堵状态划分为训练集和测试集两部分，进行接下来的实验。



图2 本文中神经网络简单模拟图示

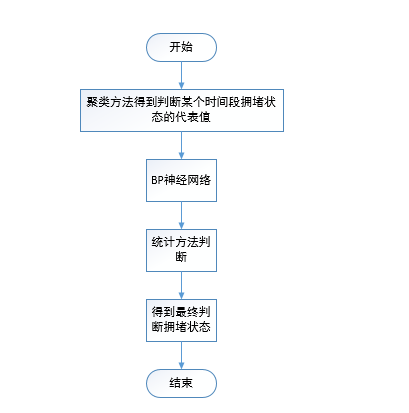


图3 本文实验框架流程图

**第一阶段**

在第一阶段中，通过训练得到经典的BP神经网络模型，如图一所示，输入数据为4个，是4个连续单位时间内通过的车辆数量，预测结果为3个在0到1之间的数值，分别对应拥堵、一般、空闲的代表值，通过这3个数据来判断接下来一个连续单位时间内的交通拥堵状态，通过训练可得到经典的神经网络模型。

**第二阶段**

通过获得的数据集的大部分数据作为训练集，第一阶段训练得出了经典的神经网络模型。如图1所示，模型对输入会得到3个对应的输出。本文在训练出第一阶段的神经网络模型后，再次将训练集数据输入，然后得到所有训练集数据通过模型预测得到的3个代表数值和数据对应交通拥堵状态的对应值。之后再将所有同一拥堵状态下的这3个数值及其对应着的状态进行归类整合在一起。再对所有属于每个分类下的数据进行处理，获得每个状态分类下对应的最大数值的平均值和标准差。最后在对测试集的预测中，再对预测得到的3个数值进行状态归类的环节，结合这两个数据，判断这3个数值相对每一种状态的平均值的偏离程度，并以此作为辅助判断预测结果的手段。

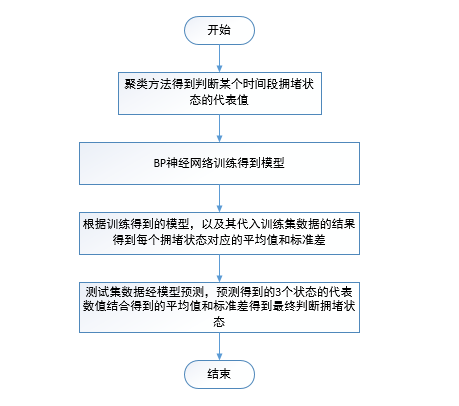


图4 算法流程图

**算法提出（Proposed algorithm）**

**第一部分：ROCK算法对交通数据进行聚类**

输入：

需要聚类的个数-k，和相似度阈值-θ

输出：

　　簇和异常值（不一定存在）

算法：

　　开始每个点都是单独的聚类，根据计算点与点间的相似度，生成相似度矩阵；

　　根据相似度矩阵和相似度阈值-θ，计算邻居矩阵-A。如果两点相似度>=θ,取值1（邻居），否则取值0；

　　计算链接矩阵-L=A x A；

　　计算相似性的度量（Goodness Measure），将相似性最高的两个对象合并。回到第2步进行迭代直到形成k个聚类或聚类的数量不在发生变换。

**第二部分：结合神经网络模型和统计学方法对交通流的拥堵状况进行预测的算法**

具体过程如下：

输入：连续4个单位时间（5分钟）内的车流辆数据

输出：判断得到的交通拥堵状态，比如拥挤、一般和空闲。

第一步：将所有数据集数据集分为训练集和测试集两部分；

第二步：通过聚类得到的交通拥堵状态判断标准，对训练集中所有数据进行状态的归类。

第三步：通过第二步中的归类结果，建立一般的BP神经网络模型，该模型包含4个输入和3个输出，3个输出的数值为3种状态的代表数值，并以此判断交通拥堵状态，开始进行训练；

第四步：通过训练得到的神经网络模型反过来对训练集数据进行处理得到各自状态对应的3个数值；并通过3个数值种取决定性作用的那个数值来确定每个状态对应的平均值和标准差。

第五步：最终通过下面的公式对3个数值所属于的交通拥堵状态进行判别

，



，y为通过模型后得到的每种状态的代表数值，比如判断是否属于拥堵状态，在公式中通过比较预测得到的数值时候在这个区间可以进行判断，



第六步：重复前面的步骤依次对一般状态以及空闲状态进行判别，通过第五步的辅助判断标准，便更能确定各状态的判别结果。

**实验结果（Experimental results）**

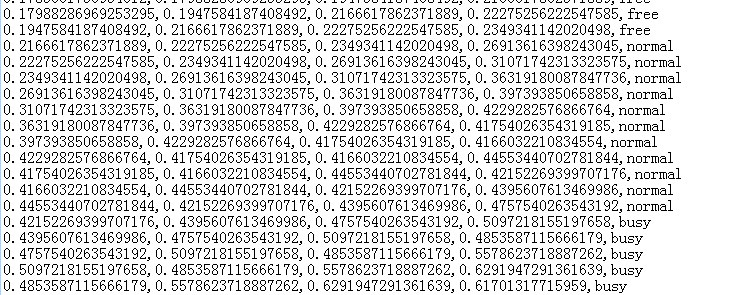


图5 预测结果存到txt文件中图

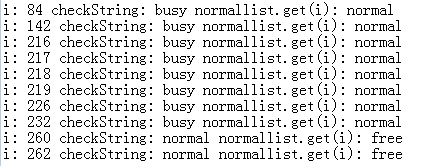


图6 预测结果不正确部分数据截图

实验结果表明本文提出的算法是有效的。在测试集中，根据一天24小时的车流量，判别每4个单位连续时间（5分钟）后的交通拥堵状态，通过第四步得到的神经网络模型直接进行预测的判断正确率为95.1%，而通过第5步的结合统计学方法的辅助判断后，拥堵状态的判断正确率达到了97.5%。很显然结果表明，本文提出的算法不仅在逻辑上可行，在实际中也是可行的。

**本文总结（Conclusions and discussion）**

本文主要提出了一种通过神经网络训练得到预测模型，联合统计方法校验得到预测结果的思路对交通流的拥堵状况进行预测，即在神经网络建模的基础上结合统计学的理论进行辅助判断。

本文在第一阶段使用了ROCK算法对交通数据进行聚类，第二阶段在对数据聚类的基础之上，提出了结合神经网络模型和统计学方法对交通流的拥堵状况进行预测的算法。实验结果表明，本文提出的思路在正确率上有了一定的提高，即本文提出的思路，不仅在理论上，在实际中也是可行的。

本文提出的思路仅针对车流量数据处理，可以考虑加入更多的特征考量，增强模型的全面性，以及辅助判断分析时的可靠性。

**参考文献（References）**

[1]X. Wang, L. T. Yang, X. Xie, J. Jin, and M. J. Deen, “A cloud-edge

computing framework for cyber-physical-social services,” IEEE Commun.

Mag., vol. 55, no. 11, pp. 80–85, Nov. 2017.

[2]M. R. Dileep & Ajit Danti (2018) Human Age and Gender Prediction Based on

Neural Networks and Three Sigma Control Limits, Applied Artificial Intelligence, 32:3, 281-292, DOI:10.1080/08839514.2018.1451217

[3]Jiabin Qu, Xinyu Gu, Lin Zhang Improved UGRNN for Short-Term Traffic Flow

Prediction with Multi-Feature Sequence Inputs

[4]LEI REN (Member, IEEE), YAQIANG SUN 1, HAO WANG4, (Member, IEEE),AND LIN ZHANG(Senior Member, IEEE) Prediction of Bearing Remaining Useful Life With Deep Convolution Neural Network

[5]Huaxiu Yao, Xianfeng Tang, Hua Wei(2018) Modeling Spatial-Temporal Dynamics for Traffic Prediction

[6]SHIN-JYE LEE, TONGLIN CHEN, LUN YU, AND CHIN-HUI LAI(2018) Image Classification Based on the Boost Convolutional Neural Network

[7]HILAL TAYARA, KIM GIL SOO, AND KIL TO CHONG(2018) Vehicle Detection and Counting in High-Resolution Aerial Images Using Convolutional Regression Neural Network

J. J. WANG1, S. G. HU, X. T. ZHAN1, Q. LUO, Q. YU1, ZHEN LIU,T. P. CHEN, Y. YIN, SUMIO HOSAKA, AND Y. LIU Predicting House Price With a Memristor-Based Artificial Neural Network

[8]Ze-Guo Zhang, Jian-Chuan Yin & Cheng Liu (2018) A Modular Real-time Tidal Prediction Model based on Grey-GMDH Neural Network, Applied Artificial Intelligence, 32:2,165-185, DOI: 10.1080/08839514.2018.1451220

[9]Y. Lecun, Y. Bengio, and G. Hinton.“Deep learning,” Nature, vol. 521,

no. 7553, pp. 436–444, 2015.

[10]J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural

Netw., vol. 61, pp. 85–117, Jan. 2015.

[11]Ma, X., and Z. B. Liu. 2015. Research on the novel recursive discrete multivariate grey prediction model and its applications. Applied Mathematical Modelling 40 (7–8):4876–90. doi:10.1016/j.apm.2015.12.021.

[12]Yao, H., Jiang, X., Sun, T., and Wang, S. (2017). 3d human action recognition based on the spatial-temporal moving skeleton descriptor. In IEEE International Conference on Multimedia and Expo (ICME), pages 937{942. DOI:10.1109/ICME.2017.8019498.

[13]Martín Abadi et al. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. (2015).

[14] Xingyi Cheng, Ruiqing Zhang, Jie Zhou, and Wei Xu. 2017. DeepTransport:

Learning Spatial-Temporal Dependency for Trafc Condition Forecasting. arXiv

preprint arXiv:1709.09585 (2017).

[15]Zhiyong Cui, Ruimin Ke, and Yinhai Wang. 2016. Deep Stacked Bidirectional

and Unidirectional LSTM Recurrent Neural Network for Network-wide Trafc

Speed Prediction. In ACM SIGKDD Workshop on Urban Computing.

[16]Dingxiong Deng, Cyrus Shahabi, Ugur Demiryurek, Linhong Zhu, Rose Yu, and

Yan Liu. 2016. Latent space model for road networks to predict time-varying

trafc. KDD (2016).

[17]Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2017. Spatio-temporal Graph Convolutional Neural Network: A Deep Learning Framework for Trafc Forecasting.

arXiv preprint arXiv:1709.04875 (2017).

[18]Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep Spatio-Temporal Residual

Networks for Citywide Crowd Flows Prediction. AAAI (2017).

[19]张亚茹. 基于车联网大数据分析的实时路况信息系统关键技术研究 [D]. 南京：南京邮电大学，2017

[20]王浩坤. 高速公路多源异构交通数据融合与预测方法研究吉林大学 [D]. 长春：吉林大学， 2015

[21]孟强. 基于rest架构的交通信息服务关键技术研究 [D]. 长安：长安大学，2016

[22]曹云舒. 多源异构交通数据质量控制关键技术研究 [D]. 长春：吉林大学,2015

[23]葛志鹏. 基于多源数据的高速公路短时交通状态预测方法研究 [D]. 南京：东南大学，2016

[24]倪升华. 基于数据的交通拥堵评价与评价方法 [D]. 杭州：浙江工业大学，2017

[25]顾超然. 基于模糊C均值的城市道路交通状态判别研究 [D] 北京：北京交通大学，2012

[26]李琦. 基于多源数据的交通状态监测与预测方法研究[D]. 吉林大学，2013

[27]覃艳，王洪，周全华. 数据挖掘中聚类算法的研究[J]. 网络安全技术与应用，2014,1:65-66.

[28]管硕. 基于K-均值聚类算法RBF神经网络交通流预测[D]. 青岛大学，2014.

[29]王春安. Hadoop环境下基于神经网络的交通流预测方法研究[D]. 青岛大学，2014.