基于Boosting Tree的交通流量预测及可视分析

摘要：交通流预测是智能交通系统中一个热门的研究课题。预测方法主要有统计方法、贝叶斯网络、神经网络模型、混合ARIMA和人工神经网络。广义回归神经网络（GRNN）是用于交通流预测的一个有趣的模型，因为它可以预测性质的动态变化和非线性的数据，这是一般的交通流数据发现。在本研究中，Boosting Tree模型建立过程中的交通流数据，并比较其结果与其他预测方法的结果（ARIMA模型、单指数平滑法、移动平均）。留一交叉验证（LOOCV）是用于检测的交通流数据和平均绝对百分比误差（MAPE）作为测试的评价标准。结果表明，采用广义回归神经网络方法对试验数据进行处理可以提高预测的减少在三其它预测方法的预测值的准确性：有马，单指数平滑法、移动平均。

关键字：

1 引言

交通流数据是分析道路交通状况的主要因素之一。交通流是指在一段道路上通过某一特定点的车辆数。这种流量是用一段时间内的车辆数量来衡量的。短时交通流具有动态性和非线性性特征（1），因此交通流预测在交通管理中起着重要作用。利用神经网络对交通流进行预测（2）。提出了用统计方法预测交通流量的方法（3），将非线性和非平稳方法应用于短时交通流预测。[ 4 ]提出自适应有马预测交通数据。[ 5 ]预测交通数据的贝叶斯模型。[ 6 ] GRNN方法相比，径向基函数神经网络（RBFNN）和反向传播神经网络（BPNN）预测交通流数据。从三种方法，网络提供最高的精度。网络是由[ 7 ] Donald F Specht介绍，属于概率神经网络的分类。网络是一个简单的、稳定的人工神经网络适用于在自然[ 1 ]动态和非线性数据模型。广义回归神经网络的一个特点是在图案层神经元的数目会随着训练数据的增加[ 7 ]，然而太多的神经元会增加过[ 8 ]的机会增多。在本文中，网络是用来预测交通流数据。然后将结果与其他三种预测方法（ARIMA、单指数平滑和滑动平均）进行了比较。

2 数据处理

本文使用的数据是上海市15年的交通数据，其中包括地铁刷卡、公交车刷卡及出租车轨迹等数据。这里选取其中一个月的地铁刷卡数据。原始数据格式如下：

表 地铁原始数据格式

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 用户ID | 日期 | 时间 | 地名 | 乘车方式 | 费用 | 是否优惠 |
| 2601377141 | 2015-04-01 | 08:23:50 | 1号线莘庄 | 地铁 | 0.00 | 非优惠 |
| 2601377141 | 2015-04-01 | 19:18:30 | 1号线莘庄 | 地铁 | 4.00 | 非优惠 |

数据预处理流程：

（1）每20分钟分为一个时间段

（2）将日期转换为星期

（3）计算相同时间段的人流量

2 方法

3 实验结果

4 总结