



研究与开发

集成 BP 神经网络预测模型的研究与应用

赵会敏, 雒江涛, 杨军超, 徐正, 雷晓, 罗林

(重庆邮电大学电子信息与网络工程研究院, 重庆 400065)

摘要: BP 神经网络对逼近实数值提供了一种稳健有效的学习方法, 适合对路口交通流量进行预测。针对 BP 神经网络存在易陷入局部最小值和收敛速度慢的问题, 提出了一种集成 BP 预测模型。该模型集成多个具有不同初始权值和训练集的 BP 模型, 并以加权平均值的方法作为结合方法。其中的每个 BP 模型是以一种改进的 MapReduce 方法实现的。将该模型运用到交通路口车辆分流流量大小的预测实例中, 并依次与单机实现的单个 BP 模型和 MapReduce 实现的单个 BP 模型进行比较。结果表明, 集成 BP 模型在路口车辆分流流量大小的预测中有较高的准确率和实时性。

关键词: BP 神经网络; 集成预测; MapReduce; 加权平均值; 交通路口分流流量大小的预测

中图分类号: TP183

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016060

Research and application of prediction model based on ensemble BP neural network

ZHAO Huimin, LUO Jiangtao, YANG Junchao, XU Zheng, LEI Xiao, LUO Lin

Electronic Information and Networking Research Institute, Chongqing University of Posts and
Telecommunications, Chongqing 400065, China

Abstract: BP neural network provides a robust and effective learning method for approximating real values. It is fit for intersection traffic flow prediction. In order to resolve its slow convergence speed and easily falling into local minimum problem, an ensemble prediction model was proposed. This model integrated multiple BP neural networks which had different initial weights and training set and used weighted average as combination method. It had used an improved MapReduce method to implement every BP neural network of the ensemble prediction model. This ensemble prediction model took traffic shunt flow prediction at intersection as an example. At last, it was compared with simple single implementation BP model and MapReduce implementation BP model respectively. Finally, results prove that ensemble prediction model has a higher accuracy rate and real-time performance in traffic shunt flow prediction at intersection.

Key words: BP neural network, ensemble prediction, MapReduce, weighted average, traffic shunt flow prediction at intersection

收稿日期: 2015-10-07; 修回日期: 2016-01-08

基金项目: 重庆市新一代信息网络与终端协同创新中心经费支持项目; 基于云计算的网络流量分类标识和深度分析关键技术研究项目 (No.cstc2013yykfA40006)

Foundation Items: Collaborative Innovation Center for Information Communication Technology Foundation of Chongqing, Key Technology Research on Classification, Identification of Network Traffic and Depth Analysis Based on Cloud Computing (No.cstc2013yykfA40006)

1 引言

随着国内城市的不断高速发展,城市交通拥堵、交通事故频发、能源浪费等已成为各个城市发展所面临的问题,很多城市将大力发展智能交通系统(intelligent transport system, ITS)。在交叉路口对车辆分流流量大小(路口左转、右转、直行车流量的大小)的预测是 ITS 中非常关键、基础的一部分,对它的研究将使交通信号灯智能控制成为可能,对交通服务行业的发展也有至关重要的影响。目前已经有很多种预测模型运用到车流量预测中,如非参数回归预测方法、自回归移动滑动平均(autoregressive moving average, ARMA)模型,参考文献[1]对这些模型做了介绍和对比,除此还有 BP(back propagation)神经网络^[2]、RBF 神经网络^[3]、Elman 神经网络^[4]、支持向量机(support vector machine, SVM)等算法。其中 BP 神经网络模型是应用最广泛的预测模型。

BP 神经网络^[6]由输入层、隐含层和输出层构成,隐含层的层数可以是一层或多层,每一层的神经元之间没有相互连接。任意的函数可以用一个有 3 层结构的网络以任意精度逼近。BP 神经网络是一种误差反向传播的模型,将网络的误差由输出层反向传播到输入层,在此过程中使用随机梯度下降的法则不断优化网络参数,调整网络的权值,使得网络误差下降,直到网络误差值开始有上升趋势时停止,得到最优网络。BP 神经网络可以实现大规模的并行处理,可以用 Hadoop 的分布式计算框架 MapReduce 编程来处理海量的数据^[7],因此可以用来处理大量的交通数据。但 BP 神经网络采用梯度下降的方法搜索网络权值空间,这决定了其容易陷入局部极小值和收敛速度慢的问题,为此提出了一种集成 BP 神经网络预测的模型。

集成学习是机器学习的一个发展成果,它将多个单个的模型按照某种结合方法结合在一起,来获得比单个模型更好的预测效果^[8]。目前,集成学习已经被广泛应用于面部识别^[9]、图像分析^[10]、疾病诊断^[11]等领域,并已取得良好的效果。本文的集成预测模型集成了多个以 MapReduce 方式实现的 BP 神经网络模型,为了解决易陷入局部极小值的问题,对每个网络用不同的随机权值初始化,这是因为网络初始权值不同,则算法训练权值的方向不同,算法最终会收敛于不同的局部极小值或全局极小值。集成模型的输出是每个 BP 模型输出的加权平均值,这样就降低了某个

模型陷入局部极小值,从而对预测结果的严重影响。此外,每个 BP 模型在计算更新权值量时引入了冲量项,使其越过误差曲面的局部极小值或平坦区域,在梯度不变的区域逐渐增大搜索步长的效果,从而可以加快收敛速度。

2 集成 BP 神经网络预测模型

集成 BP 神经网络预测模型的建立主要分为数据集构建、单个 BP 模型的建立和预测结果的汇聚 3 个步骤,如图 1 所示。

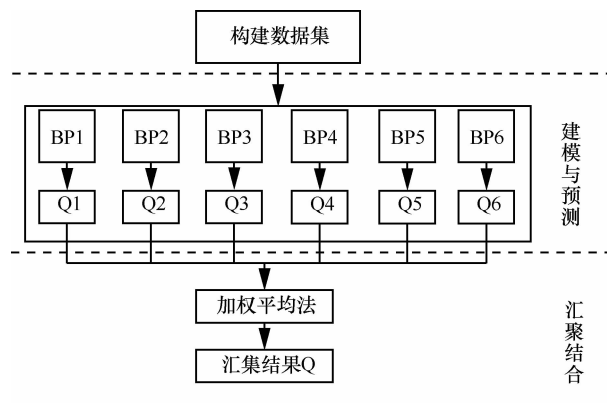


图 1 集成 BP 神经网络预测模型

2.1 数据集的构建

构建数据集首先需要确定特征字段,本文以交叉路口车辆分流流量大小的预测为例来验证模型的预测性能。参考文献[12]指出路口各个方向的车流量不仅和它前 3 个时间段的车流量有关,还和其他方向的流量有关。参考文献[13]指出当前车流量与之前一个月以内的车流量关联最为紧密,并且必须选取交通流量特征相似的时间段的数据。因此,选取路口前 3 个时间段的总流量、3 个方向的流量以及下一个时间段路口的总流量,共 13 个特征字段作为输入特征,以下一个时间段的 3 个方向的流量,共 3 个特征字段作为输出特征。然后,选取一个特定的时间段(如上下班高峰期)的一个月的数据来建立训练集,选取接下来一天该时间段的数据来建立验证集和预测集。

由于 BP 神经网络的隐含层一般采用 Sigmoid 转换函数,为提高训练速度和灵敏性以及有效避开函数的饱和区,一般要求输入数据的值在 0~1 之间。因此要对输入数据进行归一化处理。本文采用 min-max 标准化线性归一化处理方法,如式(1)所示:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$



2.2 单个 BP 神经网络模型的建立

集成预测模型的关键在于构建出具有差异性的个体模型,这也是获得较好集成效果的保证^[14]。

BP 神经网络模型的建立,首先要确定 BP 网络的网络结构,然后确定构建差异性个体模型的方法,最后用一种改进的 MapReduce 编码方式实现模型的建立。

2.2.1 BP 网络结构的确定

一个 BP 神经网络是由网络层数、每层节点数、激活函数、学习效率、冲量系数和网络的初始连接权值等构成的。

(1) 网络层数和激活函数的确定

Hornik 等人早已证明若输入层和输出层采用线性转换函数,隐含层采用 Sigmoid 转换函数,则包含一个隐含层的 MPL 网络能够以任意精度逼近任何有理函数^[15]。因此采用 3 层 BP 网络,即一个输入层、一个隐含层、一个输出层。网络的激活函数采用 Sigmoid 函数。

(2) 网络各层节点数的确定

根据输入层和输出层的特征字段的个数,网络的输入层节点数是 13 个,输出层的节点数是 3 个。隐含层的节点数的选择对整个网络的性能至关重要。若隐含层的节点数过少,网络可能不能被训练或网络的性能很差;若节点数过多,虽然网络的系统误差可能很低,但一方面使网络的训练时间很长,另一方面很容易导致训练陷入局部极小值而得不到最优极小值,这也是训练出现过拟合的内在原因,从而影响了网络的泛化能力。所谓泛化能力是指训练的网络不仅对于训练的数据能很好地拟合,并且对于不在训练集内的新的预测数据也能很好地拟合,即有较高的预测准确率。因此,在满足精度的要求下,隐含层的节点数越少越好。隐含层节点数不仅与输入、输出层的节点数有关,更与需要解决的问题的复杂程度、转换函数的形式以及样本数据的特性等因素有关。目前广泛用于隐含层节点数确定的方法有两种:一是 Kolmogorov 定理中指出,隐含层节点数一般为输入层节点数的两倍加 1;另一种如式(2)所示, u 是隐含层节点数, m 、 n 分别是输入层和输出层的节点数, a 为 3~7 之间的某个值。依照这两种方法确定隐含层的节点数范围大致在 7~11 或 7~19 中。

$$u = \sqrt{m+n} + a \quad (2)$$

(3) 学习率、冲量系数和初始连接权值

学习率 η 会影响系统学习过程的稳定性。大的学习率会导致网络权值的修正量过大,甚至导致权值超出误

差的极小值,呈不规则跳跃而不收敛。虽然过小的学习率会导致学习训练时间过长,但是能保证收敛到某个极小值。所以本文的 η 在 0.1~0.8 中。增加冲量项可以减少网络训练时间而又不导致震荡,可以加快收敛速度,通常冲量系数在 0~1 中,而且一般比学习率大。BP 算法决定了误差函数一般存在多个局部极小值,网络初始权值不同,则算法训练权值的方向不同,这直接决定了算法收敛于哪个局部极小值或全局极小值,权重的初始化强烈影响着最终的解。

基于以上理论分析,结合车辆分流量大小预测的数据集,单机 Java 编程 BP 神经网络,确定最终 BP 神经网络的结构:隐含层节点数为 7,学习率为 0.25,冲量系数为 0.3,初始权值可以分布在 $[-0.3, 0.3]$ 、 $[-0.4, 0.4]$ 、 $[-0.5, 0.5]$ 、 $[-0.6, 0.6]$ 和 $[-0.7, 0.7]$ 几个区间中。

2.2.2 构建多个差异性个体 BP 神经网络模型

在引言部分已经说明使用不同的初始权值来构建 BP 神经网络,所以使用已经验证的 $[-0.3, 0.3]$ 、 $[-0.4, 0.4]$ 、 $[-0.5, 0.5]$ 、 $[-0.6, 0.6]$ 和 $[-0.7, 0.7]$ 这几个分布区间来初始不同的 BP 神经网络。使用不同的随机种子来生成两组在分布区间 $[-0.5, 0.5]$ 中的初始权值。共构建 6 个有差异性的单个 BP 神经网络的模型。

Bagging 是一种能够提高不稳定学习算法的预测精度的集成技术,对于决策树和人工神经网络这样的学习算法特别有效^[16]。借鉴 Bagging 集成技术,从构建的训练集 S 中抽取 6 个子训练集 $\{S_1, S_2, S_3, S_4, S_5, S_6\}$,分别用这些训练集训练 6 个 BP 神经网络模型,即使用不同的数据集训练网络来增加单个 BP 神经网络的差异度,从而有效地提高集成 BP 预测模型的泛化能力。

2.2.3 以 MapReduce 的方式实现 BP 神经网络模型

参考文献[17]对 BP 神经网络的 MapReduce 的实现思路是在 map 阶段计算并输出每个权值的变化量,然后在 reduce 阶段统计每个权值的总变化量,再对权值进行调整。MapReduce 过程中产生大量的中间结构,这就导致了大量的 I/O 操作,大大延长了作业的完成时间。本文采用了一种改进的 MapReduce 实现方式,能显著减少算法的中间结果的数量,加快算法收敛速度。针对每个样本 map 阶段输出的所有权值,定义了一个继承于 Hadoop 序列化标准类的 TwoDArrayWritable 的 WeightWritable 类,用于传递权值,并且对于每个输入样本,经过数次的本地迭代后,才输出多次迭代后的权值。

对 BP 神经网络的 MapReduce 化实现主要通过两个 map 函数和一个 reduce 函数来实现,第二个 map 函数用验证集来测试预测模型,计算测试误差,用来判断是否可以停止迭代。其实现系统如图 2 所示。

在第一个 map 函数中,网络从文件系统中读取权值,初始化网络,读入样本。对每个样本迭代 20 次之后, map 输出一个<LongWritable,WeightWritable>键值对,其中 WeightWritable 作为 map 的值类型,记录了经过训练后的网络的所有权值。每一个回合的学习结果都要送到一个 reduce 中进行综合,对输出类型为 LongWritable 的键设一个统一的值,因为网络连接权值包括输入层到隐含层的连接权值和隐含层到输出层的连接权值,因此有两类权值矩阵,为将这两个权值矩阵分开,将它们的键分别设为 1、2。

在 reduce 函数中,接收 map 的输出<LongWritable,WeightWritable>作为输入,reduce 函数的输出仍为<LongWritable,WeightWritable>。收集 map 函数输出的 WeightWritable,根据不同的键值(1 或 2),累加不同的值和

值的数目,计算出两个权值矩阵的算术平均值 WeightWritable 作为 reduce 函数的输出值,即获得经过训练后网络的新权值。

在另一个 map 函数 map1 中,将 reduce 函数计算得到的权值作为初始权值,初始化一个神经网络。并用验证数据集测试网络,计算验证集的预测误差,如果此次迭代的预测误差小于上次迭代的预测误差,则用 reduce 函数得到的权值矩阵更新文件系统中的权值文件。此外验证集的预测误差将会用来验证模型的预测性能是否达到要求,从而判断是否可以停止迭代。

另外还要设置一个驱动函数来保证程序的正常运行。驱动函数的输入包括两个路径,一是样本的输入路径,另一个是 Hadoop 作业的输出路径。该函数先判断是否存在权值文件,如果没有则创建一个 WeightWritable 实例,并赋予已选定区间的随机值作为网络的初始权值,写入 HDFS 中供后续使用。然后根据参数中的样本路径、输出路径创建 Hadoop 作业。在创建作业之前,根据 map1 函数计算的预测误差来判断是否继续进行下一次作业的循环迭代,评

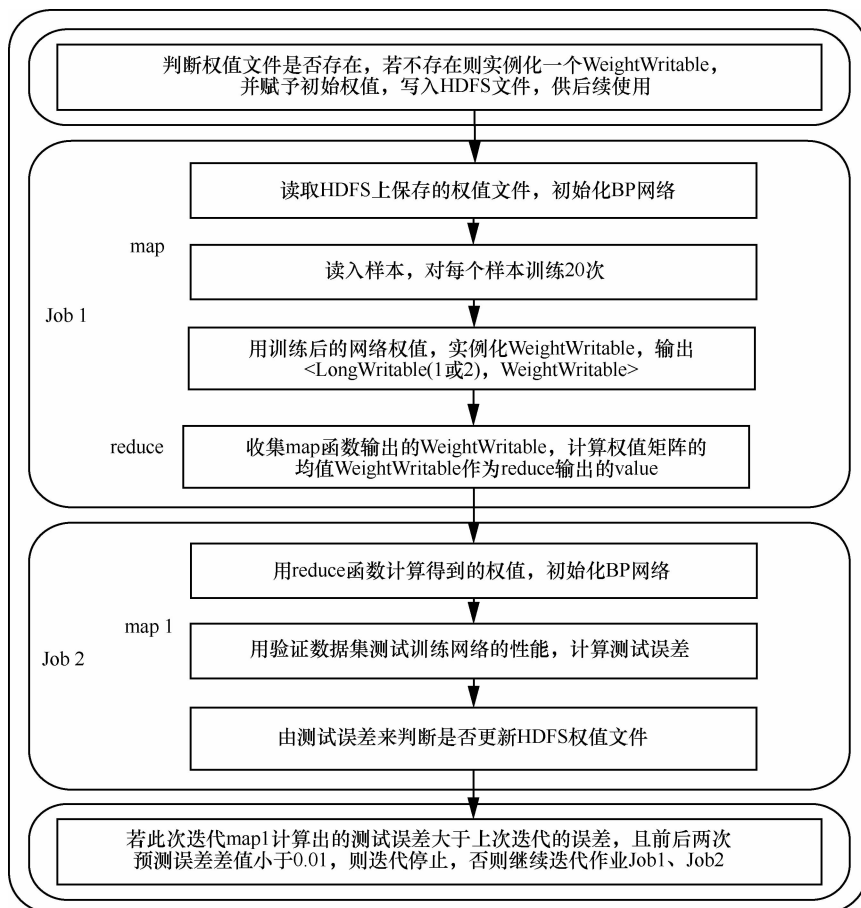


图2 BP神经网络以MapReduce方式实现的系统



价标准是前后两次迭代的预测误差的差值小于 0.01,且此次预测误差大于上次误差,则结束迭代。此评价标准与参考文献[18]中提出的以训练样例的误差达到某个阈值以下或达到一定训练次数作为训练的终止条件相比,解决了反向传播算法容易过度拟合训练样例,降低对其他新的预测实例的泛化精度的问题。

2.3 预测结果的汇集结合

运用已经训练好的 6 个 BP 神经网络模型(模型 BP1、BP2、BP3、BP4、BP5、BP6)对预测集中的每个样本实例进行预测,并保存每个模型的预测结果。然后根据加权平均值的汇集结合方法,用每个 BP 模型验证集误差值最低的一次的验证集的平均准确率除以所有模型平均准确率之和,计算出集成的加权系数。将 6 个模型的预测结果汇集,得到最终的集成预测结果。集成预测结果的汇集结合方法如式(3)所示,其中 Q 代表集成预测的流量大小, Q_n (n 为 1~6)代表单个 BP 模型的预测结果, T_n (n 为 1~6)代表每个模型的加权系数。 T_n 的计算方法如式(4)所示,其中 R_n (n 为 1~6)代表每个 BP 模型验证集误差值最低的一次的验证集的平均准确率。

$$Q = T_1 Q_1 + T_2 Q_2 + T_3 Q_3 + T_4 Q_4 + T_5 Q_5 + T_6 Q_6 \quad (3)$$

$$T_n = \frac{R_n}{R_1 + R_2 + R_3 + R_4 + R_5 + R_6} \quad (4)$$

3 模型测试与结果分析

3.1 测试平台

本文的 BP 神经网络模型是以 MapReduce 方式实现的,需要在 Hadoop 平台上进行运行测试。在虚拟机 CentOS6.5 系统中分别部署了 Hadoop-2.6.0 伪分布式平台和主要开发平台 Eclipse-Jee-Luna-SR2-Linux。其中,为使 Eclipse 运行于 Hadoop 平台上,使用插件 Hadoop-Eclipse-Plugin-2.6.0。

3.2 数据集的来源及构建

数据来源于开源网站数据堂^[9],该数据是南京市一天的 RFID 交通数据,包含的特征字段为身份信息(电子牌编号)、车型信息(大、中、小)、时间信息(通过某一采集点时间)、位置信息(采集点编号及方位信息)和车道号(车辆处在哪个车道上)。利用源数据的这些特征字段构建集成预测模型的特征集。

分析这一天的南京市 RFID 数据的规律,编程仿造出 30 天的 RFID 数据,并将其构建为数据集,随机分为 6 份,作为 6 个 BP 神经网络的训练集;将一天的南京市的 RFID 数据构建为数据集分为两份,分别作为模型的验证集和预测集。

3.3 预测结果与分析

预测的准确率和所需时间的长短是评估路口车辆分流流量大小预测模型的重要指标,流量预测具有实时性,如果预测的时间过长,将会使流量预测失去意义。准确率的计算式如下:

$$q = 1 - \frac{|Q - G|}{G} \quad (5)$$

其中, q 代表准确率, Q 代表预测流量大小, G 代表实际目标流量大小。用式(5)计算所有样例的准确率,取平均作为模型的准确率。

本文将单机实现(即未用 MapReduce 方式)的 BP 神经网络预测模型、MapReduce 方式实现的 BP 神经网络预测模型以及集成 BP 神经网络预测模型三者的预测结果进行比较,见表 1。

由表 1 可知,单机 BP 神经网络预测模型的准确率略低于 MapReduce 方式实现的 BP 神经网络预测模型的准确率,在左转方向上低出 1.4%~13.2%,在右转方向上低出 0.2%~8.7%,在直行方向上低出 5.1%~8.3%。而它们的准确率都低于集成 BP 预测模型的准确率,在左转方向上单机 BP 低出 18.7%,MapReduce 方式实现的 BP 低出 5.5%~

表 1 各个模型的路口车辆分流流量大小预测性能

模型	左转准确率	右转准确率	直行准确率	训练时间/s	预测时间/s
单机 BP	67.2%	69.7%	78.3%	64.696	17.626
BP ₁	72.3%	70.4%	85.3%	40.577	10.105
BP ₂	80.4%	75.7%	86.6%	35.042	9.907
BP ₃	73.3%	71.2%	85.9%	25.981	9.873
BP ₄	68.8%	69.9%	83.4%	28.036	10.038
BP ₅	71.9%	78.4%	85.1%	27.992	8.991
BP ₆	70.7%	75.7%	86.3%	25.992	9.658
集成 BP	85.9%	90.2%	88.9%		13.112

17.1%;在右转方向上单机BP低出20.5%,MapReduce方式实现的BP低出11.8%~20.3%;在直行方向上单机BP低出10.6%,MapReduce方式实现的BP低出2.3%~5.5%。

出现以上结果的原因是单机BP神经网络模型在训练时,将样本连续地输入网络,每个样本每次经过网络,网络的权值都会被更新,这种训练方式,有可能对后输入的样本训练效果较好,而对最开始输入的样本因训练时间久而“遗忘”;而MapReduce实现的BP神经网络预测模型在训练时,对每个实例进行多次本地迭代后更新权值,然后再求每个实例更新的权值的平均值作为神经网络的权值,这兼顾了每个实例,因此后者的预测准确率高于前者的预测准确率;而集成预测模型,通过训练多个BP神经网络预测模型并将其结果集成输出,提高了模型的泛化能力,因此本文提出的集成BP预测模型的预测准确率最高。各个模型准确率的比较结果如图3所示。因为3个方向在各个模型中的预测准确率情况大致相似,因此,只展示出了各个模型左转弯准确率的情况。

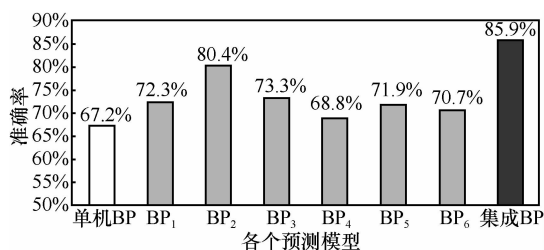


图3 各个模型左转弯准确率比较情况

由表1可知,MapReduce方式实现的BP神经网络预测模型的训练时间和预测时间与单机BP神经网络预测模型的训练时间和预测时间相比,都大大缩短了。MapReduce方式实现的6个BP模型的平均训练时间为30.063 s,与单机BP的训练时间64.696 s相比,训练时间减少了一半;

MapReduce方式实现的6个BP模型的平均预测时间为9.762 s,与单机BP的预测时间17.626 s相比,预测时间将近减少了一半。而集成预测模型的预测时间为13.112 s,要稍微比MapReduce方式实现的6个BP神经网络预测模型的平均预测时间9.762 s长,是因为在预测开始时,需要从HDFS上读取已经训练好的6个MapReduce方式实现的BP神经网络预测模型的权值,来初始化6个BP神经网络,需要消耗一定的时间,但这个时间开销不是很大,在对路口车辆分流流量大小预测的实时性要求范围内。各个模型预测时间的比较结果更直观地展示如图4所示。

4 结束语

集成学习可以提高单个神经网络的泛化能力,即对新样本有较高的预测准确率,从而提高模型的稳定性。本文提出了集成多个BP神经网络的预测模型,并以MapReduce的方式实现,将其分别与单机实现(即未用MapReduce方式)的BP神经网络预测模型和MapReduce方式实现的BP神经网络预测模型的性能进行比较。测试表明,集成预测模型的预测准确率较高、实时性合理,适合处理大量的交通流量预测数据。为路口车辆分流流量大小的预测提供了一种可靠的参考方法,在实际系统中,可以利用历史RFID数据训练建立集成BP神经网络预测模型,确定各个BP模型的连接权值和汇聚结合计算式的加权系数,来预测路口实时车辆分流流量大小,并判断预测的准确率是否达到要求,如果不符合要求便更新建模的数据和各个BP模型的连接权值、汇聚结合计算式的加权系数。但本文对路口车辆分流流量大小预测的准确率在0.9左右,仍有待提高。在本文的工作基础上继续深入研究,力争建立性能更好的预测模型。

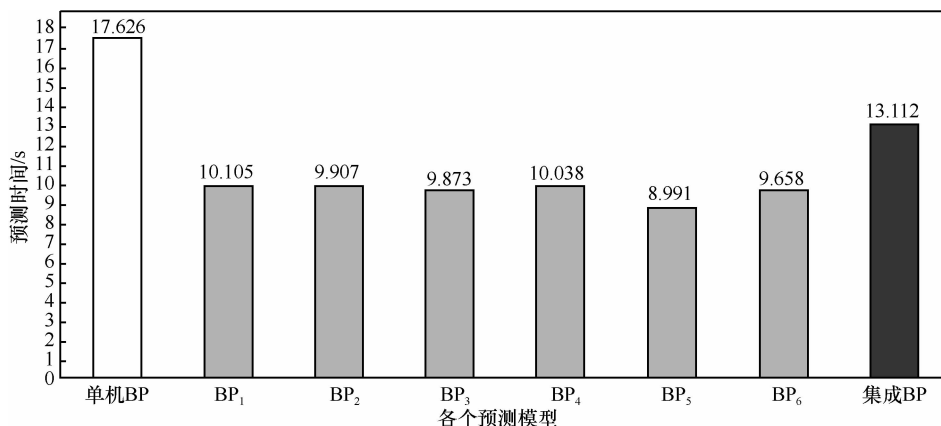


图4 各个模型预测时间比较情况



参考文献:

- [1] SMITH B L, DEMESKY M J. Traffic flow forecasting: comparison of modeling approaches [J]. Journal of Transportation Engineering, 1997, 123(4): 262-266.
- [2] 朱学明. 基于神经网络的短时交通流预测方法的研究与应用 [D/OL]. 兰州: 兰州理工大学, 2013[2013-10-30]. http://xueshu.baidu.com/s?wd=paperuri%3A%28cccf11993959c64f0b0fa9321a680311%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http%3A%2F%2Fd.wanfangdata.com.cn%2FThesis%2FY2335877&ie=utf-8.
ZHU X M. Research and application on the short-term traffic flow prediction based on neural network [D/OL]. Lanzhou: Lanzhou University of Technology, 2013 [2013-10-30]. http://xueshu.baidu.com/s?wd=paperuri%3A%28cccf11993959c64f0b0fa9321a680311%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http%3A%2F%2Fd.wanfangdata.com.cn%2FThesis%2FY2335877&ie=utf-8.
- [3] ZHAO J Y, JIA L, WANG X Q. The forecasting model of urban traffic flow based on parallel RBF neural network [C]//The International Conference on Intelligent Computing and Applications, April 4-6, 2015, Wuhan, China. [S.l.:s.n.], 2005: 515-520.
- [4] 董春娇, 邵春福, 熊志华, 等. 基于 Elman 神经网络的道路网短时交通流预测方法[J]. 交通运输系统工程与信息, 2010, 10(1): 145-151.
DONG C J, SHAO C F, XIONG Z H, et al. Short-term traffic flow forecasting of road network based on Elman neural network [J]. Transportation Systems Engineering and Information Technology, 2010, 10(1): 145-151.
- [5] 徐启华, 杨瑞. 支持向量机在交通流量实时预测中的应用[J]. 公路交通科技, 2005, 22(12): 131-134.
XU Q H, YANG R. Traffic flow prediction using support vector machine based method[J]. Highway and Transportation Research and Development, 2005, 22(12): 131-134.
- [6] MITCHELL T M. 机器学习[M]. 北京: 机械工业出版社, 2003: 60-90.
MITCHELL T M. Machine Learning [M]. Beijing: China Machine Press, 2003: 60-90.
- [7] 朱晨杰, 杨永丽. 基于 MapReduce 的 BP 神经网络算法研究 [J]. 微型电脑应用, 2012, 28(10): 9-12.
ZHU C J, YANG Y L. Research of MapReduce based on BP neural network algorithm[J]. Microcomputer Applications, 2012, 28(10): 9-12.
- [8] BREIMAN L. Bagging predictors [J]. Machine Learning, 1996, 24(2): 123-140.
- [9] CONNOLLY J F, GTANGER E, SABOURIN R. Dynamic multi-objective evolution of classifier ensembles for video face recognition [J]. Applied Soft Computing, 2013, 13 (6): 3149-3166.
- [10] SAUX B L, SANFOURCHE M. Robust vehicle categorization from aerial images by 3d-template matching and multiple classifier system[C]//The 7th International Symposium on Image & Signal Processing & Analysis, September 4-6, 2011, Dubrovnik, Croatia. New Jersey: IEEE Press, 2011: 466-470.
- [11] KRAWCZYK B, SCHAEFER G. Effective multiple classifier systems for breast thermogram analysis [C]//The 21st International Conference on Pattern Recognition (ICPR), November 11-15, 2012, Tsukuba, Japan. New Jersey: IEEE Press, 2012: 3345-3348.
- [12] 尚宁, 覃明贵, 王亚琴, 等. 基于 BP 神经网络的路口短时交通流量预测方法 [J]. 计算机应用与软件, 2006, 23(2): 32-33.
SHANG N, QIN M G, WANG Y Q, et al. A BP neural network method for short-term traffic flow forecasting on crossroads[J]. Computer Applications and Software, 2006, 23(2): 32-33.
- [13] 崔德光, 吴淑宁, 徐冰. 空中交通流量预测的人工神经网络和回归组合方法[J]. 清华大学学报(自然科学版), 2005, 45(1): 96-99.
CUI D G, WU S N, XU B. Air traffic flow forecasts based on artificial neural networks combined with regression methods[J]. Journal of Tsinghua University(Science and Technology Edition), 2005, 45(1): 96-99.
- [14] DIETTERICH T G. Ensemble Methods in Machine Learning[M]. Berlin: Springer Berlin Heidelberg, 2000: 1-15.
- [15] CHU C T, SANG K K, LIN Y A, et al. Map-reduce for machine learning on multicore [C]//The Twentieth Annual Conference on Neural Information Processing Systems, December 4-7, 2006, Vancouver, Canada. [S.l.:s.n.], 2006, 281-288.
- [16] QUINLAN J R. Bagging, boosting, and C4.5 [EB/OL]. (1996-08-04) [2015-3-11]. http://xueshu.baidu.com/s?wd=paperuri%3A%28c35bca131cc209561cc4f533a6cfd1a0%29&filter=sc_long_sign&sc_ks_para=q%3DBagging%2C%20boosting%2C%20and%20C4.5&tn=SE_baiduxueshu_c1gjeupa&ie=utf-8.
- [17] 鞠训光, 邵晓根, 鲍蓉, 等. Hadoop 下并行 BP 神经网络路马湖水水质分类 [J]. 南京师大学报 (自然科学版), 2014, 37(1): 52-56.
JU X G, SHAO X G, BAO R, et al. Based on parallel BP neural network of classification on water quality of Luoma lake under hadoop [J]. Journal of Nanjing Normal University(Natural Science Edition), 2014, 37(1): 52-56.
- [18] 朱启敏. 基于云计算平台的神经网络计算方法及其应用研究 [D/OL]. 广州: 华南理工大学, 2014 [2014-12-28]. <http://xueshu>.

baidu.com/s wd=paperuri% 3A% 28376cd6d6a7a74296 f30ac 11139993edc% 29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http% 3A% 2F% 2Fedmd.cnki.com.cn% 2FArticle%2FCDMD-10561-1014064144.htm&ie=utf-8.

ZHU Q M. Research on computation methods of neural networks and applications based on a cloud computing platform (master dissertation). Guangzhou: South China University of Technology, 2014 [2014-12-28].http://xueshu.baidu.com/s wd=paperuri% 3A% 28376cd6d6a7a74296f30ac 11139993edc% 29& filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http% 3A% 2F% 2Fedmd.cnki.com.cn% 2FArticle% 2FCDMD-10561-1014064144.htm&ie=utf-8.

[19] 南京交通 RFID 及车牌识别数据集 (样例版)[EB/OL]. (2015-04-20)[2015-05-21].<http://www.datatang.com/data/47187>.

[作者简介]



赵会敏(1988-),女,重庆邮电大学电子信息与网络工程研究所硕士生,主要研究方向为数据挖掘、交通大数据分析。



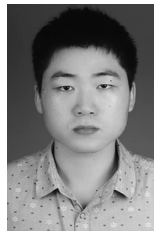
雒江涛(1971-),男,博士,重庆邮电大学电子信息与网络工程研究所副院长、教授、博士生导师,IEEE 高级会员,主要研究方向为移动互联网数据挖掘和新一代网络技术。



杨军超(1988-),男,重庆邮电大学电子信息与网络工程研究所博士生,主要研究方向为数据挖掘。



徐正(1989-),男,重庆邮电大学电子信息与网络工程研究所硕士生,主要研究方向为数据挖掘。



雷晓(1989-),男,重庆邮电大学电子信息与网络工程研究所硕士生,主要研究方向为数据挖掘。



罗林(1990-),男,重庆邮电大学电子信息与网络工程研究所硕士生,主要研究方向为数据挖掘。