

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2015 年 7 月 9 日 (09.07.2015)



(10) 国际公布号
WO 2015/100993 A1

- (51) 国际专利分类号:
G08G 1/00 (2006.01)
- (21) 国际申请号: PCT/CN2014/081350
- (22) 国际申请日: 2014 年 7 月 1 日 (01.07.2014)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201310749807.0 2013 年 12 月 30 日 (30.12.2013) CN
- (71) 申请人: 复旦大学 (FUDAN UNIVERSITY) [CN/CN];
中国上海市杨浦区邯郸路 220 号, Shanghai 200433 (CN)。
- (72) 发明人: 史世雄 (SHI, Shixiong); 中国上海市张衡路 825 号, Shanghai 201203 (CN)。 杨夙 (YANG, Su);
中国上海市张衡路 825 号, Shanghai 201203 (CN)。
- (74) 代理人: 上海专利商标事务所有限公司 (SHANGHAI PATENT & TRADEMARK LAW OFFICE, LLC); 中国上海市桂平路 435 号, Shanghai 200233 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

(54) Title: TIME AND SPACE RELATED DATA MINING-BASED TRAFFIC FLOW PREDICTION METHOD

(54) 发明名称: 一种基于时空关联数据挖掘的交通流预测方法

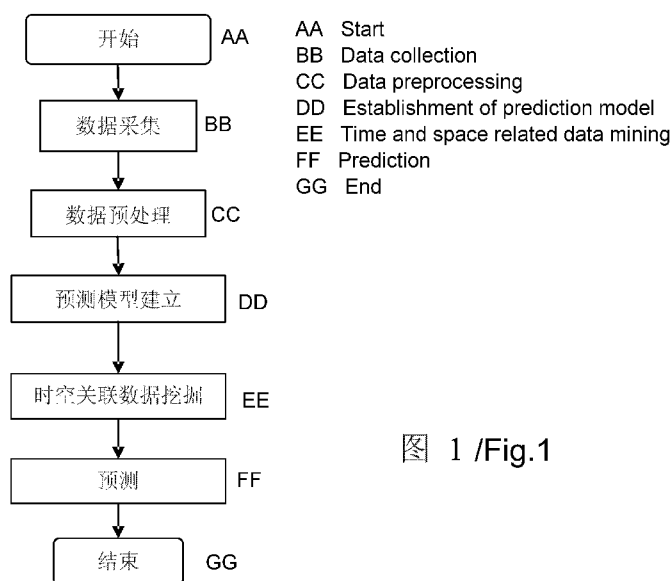


图 1 /Fig.1

(57) Abstract: The present invention relates to the technical field of traffic flow prediction, and particularly to a time and space related data mining-based traffic flow prediction method. The method mainly comprises establishment of a prediction model, time and space related data mining, time and space related data-based traffic flow prediction and the like. The prediction model may adopt a multi-factor linear regression model. The time and space related data mining is automatically selecting data, related to a prediction target, of a time and space related sensor based on the multi-factor linear regression model by a sparse representation optimization method. The time and space related data-based traffic flow prediction is prediction using the data of the time and space related sensor as the input of the prediction model. In the present invention, the time and space related sensor related to a sensor of a prediction target node is automatically determined from the whole traffic network, and the data of the time and space related sensor is used as the input of the prediction model, and the prediction performance of the prediction model is enhanced by the full automatic time and space related data mining.

(57) 摘要:

[见续页]

本发明属于交通流预测技术领域，具体为一种基于时空关联数据挖掘的交通流预测方法。本发明主要包括预测模型建立、时空关联数据挖掘以及基于时空关联数据的交通流预测等。预测模型可采用多因子线性回归模型；时空关联数据挖掘是基于多因子线性回归模型并通过稀疏表达的优化方法自动地选取对于预测目标相关的时空关联传感器的数据；基于时空关联数据的交通流预测是以时空关联传感器的数据作为预测模型的输入进行的预测。本发明从整个交通网络中自动地确定与预测目标节点的传感器相关的时空关联传感器，并以时空关联传感器的数据作为预测模型的输入，全自动时空关联数据挖掘提升了预测模型的预测性能。

一种基于时空关联数据挖掘的交通流预测方法

技术领域

本发明属于交通流预测技术领域，具体涉及一种基于时空关联数据挖掘的交通流预测方法。

背景技术

近年来，交通流预测技术一直在智能交通系统（ITS）中发挥很重要的作用。交通流预测技术能够帮助通行的个人进行智能的出行路线选择，也可以为交通管理者提供决策支持。

早期的一些技术主要针对单序列的交通流预测。可以根据预测模型的有无参数分为有参数预测模型和无参数预测模型。在有参数预测模型中，季节性 ARIMA（Autoregressive Integrate Moving Average）模型是应用最广泛的一种方法（见参考文献[1]）。它在基于单时间序列的交通流预测上可以达到最小的均方误差（MSE）。而在无参数预测模型中，最近邻方法（Nearest Neighbor Method）被认为是可替代季节性 ARIMA 的方案（见参考文献[2]），但是最近邻预测方法的预测效果依赖于历史数据的质量。总体来讲，单序列的交通流预测方法只考虑了序列自身的特征，忽视了不同序列之间的相互作用和关系。

由于交通流的演化是通过交通网中所有节点交通流之间的相互作用而形成的，所以在交通流的预测中，不同节点交通流之间的关系应该被考虑进去。从而，在近些年基于数据时空关联性的多因子交通流预测成为了研究的热点。目前流行的方法大致有三种：1、状态空间模型或卡尔曼滤波器（见参考文献[3]）；2、神经网络（Neural Network）等机器学习方法（见参考文献[4]）；3、时间序列方法，如 Vector Autoregressive Moving Average 模型（见参考文献[5]）（VARMA）。而时空关联数据的选取则是进行多因子交通流预测的必要步骤。在以往的研究中，时空关联传感器的选取大部分是根据人工经验手动选取目标节点一定邻近范围内的传感器。只选取目标节点邻近范围内的传感器的数据作为与预测模型的输入比较主观，没有反映现实中数据之间真正的时空关联性，无法获得最佳的预测性能，基于人工经验的方法缺乏适应性，很难实

施到大型的交通网中。

稀疏表达作为一种数学工具最早应用于信号处理领域，被广泛应用于信号压缩、图像去模糊、特征提取等，本发明提出将其应用于面向交通流预测的时空关联数据挖掘。稀疏表达的主要思想为：一个信号 y 可以通过一个字典 D 包含的 K 个原子信号 $\{d_1, d_2, \dots, d_j, \dots, d_K\}$ 的线性组合表示为 $y = Dx$ ，其中 $y \in R^n$ ， $d_j \in R^n$ ， $D \in R^{n \times K}$ ；或者近似表达为 $y \approx Dx$ ，这里 $\|y - Dx\|_2^2 \leq \varepsilon_0$ ，其中 $x \in R^K$ 为信号 y 的表示系数，稀疏表达要求通过尽可能少的原子信号来表示信号 y ，即要求线性组合系数 x 包含的非零系数尽可能少，因此，求稀疏解的目标函数可以写作

$$\hat{x} = \arg \min_x \|x\|_0 \quad \text{约束条件: } y = Dx$$

或者

$$\hat{x} = \arg \min_x \|x\|_0 \quad \text{约束条件: } \|y - Dx\|_2^2 \leq \varepsilon_0$$

上式中的 $\|x\|_0$ 为 x 的 0 范式，表示向量 x 包含的非零元素的个数。由于已有一些稀疏表达的优化方法方面的数学工具（见参考文献[6]），可以自动从字典中选取有效的原子表示信号，所以本发明将其应用到时空关联性数据挖掘中，以达到从整个交通网络中自动地选取与目标传感器相关的时空关联传感器、并将时空关联传感器的数据作为预测模型的输入用于预测的目的。

参考文献

[1] Williams, B. M., Durvasula, P.K., Brown, D.E., 1998. Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models. Transportation Research Record 1644, 132-144.

[2] Smith, B. L., Williams, B. M., Oswald, R. K., 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. Transportation Research Part C 10, 303 - 321.

[3] Stathopoulos, A., Karlaftis, A., S., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. Transportation Research Part C 11, 121-135.

[4] Vlahogianni, E. I., Karlaftis, M. G., Golias, J. C., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transportation Research Part C 13,

[5] Min, W., Wynter, L., 2011. Real-time road traffic prediction with spatio-temporal correlations. Transportation Research Part C 19, 606 - 616.

[6] Elad, M., 2010. Sparse and redundant representations-From theory to application in signal and image processing. Springer.。

发明内容

本发明的目的在于克服现有交通流预测技术在时空关联数据选取上的不足，提出一种基于时空关联数据挖掘的交通流预测方法，从整个交通网络中自动地确定与预测目标节点相关的时空关联传感器。

本发明提出的基于时空关联数据挖掘的交通流预测方法，具体步骤为：

(a) 通过布局到交通网各个节点的传感器采集交通流量的原始数据；

(b) 通过数据预处理，将采集的原始数据处理为有效的交通流数据；

(c) 建立预测模型：令 v_i^j 表示交通网中的传感器 j 在第 i 个时刻采集的交通流量数据，假设一个交通网中有 m 个传感器，则在第 i 个时刻整个交通网的状态表示为 $V_i = [v_i^1, v_i^2, \dots, v_i^m]$ ，提前 τ 时刻对传感器 j 采集的交通流量数据进行预测的线性回归模型为：

$$\bar{v}_{i+\tau}^j = V_i w^j$$

上式中的权重 $w^j = [w_1^j, w_2^j, \dots, w_k^j, \dots, w_m^j]^T$ 为待优化的模型参数， $\bar{v}_{i+\tau}^j$ 为预测值；

(d) 挖掘时空关联性：通过稀疏表达的优化方法自动得到模型参数 w^j ，权重 $w^j = [w_1^j, w_2^j, \dots, w_k^j, \dots, w_m^j]^T$ 表示整个交通网络的各个传感器的数据对于预测目标传感器 j 的数据而言的时空关联性，当 $w_k^j = 0$ 时，传感器 k 的数据与传感器 j 的数据之间没有关联性，否则 w_k^j 值的大小表示传感器 k 的数据与传感器 j 的数据之间关联程度的强弱， $k=1, 2, \dots, m$ ；

(e) 以时空关联数据为预测模型的输入进行交通流预测。

本发明中，所述的步骤 (d) 有三种优化的目标，分别用 (d1)、(d2) 和 (d3) 表示，具体如下：

(d1) 所述的稀疏表达方法要优化的目标为：

$$\hat{w}^j = \arg \min_{w^j} \|w^j\|_0 \quad \text{约束条件:} \quad \|V w^j - v^j\|_2^2 = \sum_{i=1}^n \|V_i w^j - v_i^j\|^2 \leq \varepsilon_0$$

上式中, $V = (V_1^T, V_2^T, \dots, V_n^T)^T$ 表示从第 1 到第 n 时刻所有传感器采集的交通流数据, 其中第 i 个时刻交通网中 m 个传感器采集的交通流量为 $V_i = [v_i^1, v_i^2, \dots, v_i^m]$, $i=1, 2, \dots, m$, $v^j = [v_{\tau+1}^j, v_{\tau+2}^j, \dots, v_{\tau+n}^j]^T$ 表示传感器 j 记录的从第 $\tau+1$ 时刻到第 $\tau+n$ 时刻的交通流量, $\|w^j\|_0$ 表示 w^j 的 0 范数, 即权重向量 w^j 中非零元素的个数, ε_0 是一个预先设定的门限, 用于控制预测误差 $\|Vw^j - v^j\|_2$, $\|Vw^j - v^j\|_2$ 表示向量 $Vw^j - v^j$ 的 2 范数, 即向量中所有元素的平方和的平方根;

(d2) 所述的稀疏表达方法要优化的目标为:

$$\hat{w}^j = \arg \min_{w^j} \|Vw^j - v^j\|_2^2 \quad \text{约束条件: } \|w^j\|_0 \leq L_0$$

式中, L_0 为预先设定的门限, 用于控制向量 w^j 中非零元素个数; 上式表达的优化目标通过 Orthogonal Matching Pursuit (OMP) 算法进行求解【Tropp, J. A., 2004. Greed is good: algorithmic results for sparse approximation. IEEE Trans. Information Theory 50, 2231-2242.】;

(d3) 所述的稀疏表达方法要优化的目标为:

$$\hat{w}^j = \arg \min_{w^j} \|w^j\|_1 \quad \text{约束条件: } \|Vw^j - v^j\|_2^2 = \sum_{i=1}^n \|V_i w^j - v_i^j\|^2 \leq \varepsilon_0$$

式中, $\|w^j\|_1$ 为 w^j 的 1 范数, 表示向量 w^j 中所有元素绝对值之和, ε_0 是一个预先设定的门限, 用于控制预测误差 $\|Vw^j - v^j\|_2$; 上式表达的优化目标等价地表示为加入拉格朗日乘子 λ 的无约束条件的优化目标:

$$\hat{w}^j = \arg \min_{w^j} \left\{ \lambda \|w^j\|_1 + \frac{1}{2} \|Vw^j - v^j\|_2^2 \right\}$$

上式表达的优化目标通过 Least Angle Regression Stagewise (LARS) 算法进行求解【Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. 2004. Least angle regression. Annals of Statistics 32(2), 407-499.】。

本发明中, 预测模型采用多因子线性回归模型外, 也可采用向量自回归模型、反向传播神经网络模型、径向基函数神经网络模型等。

本发明提出的方法可以从整个交通网络中自动地确定与预测目标节点的传感器相关的时空关联传感器, 并以时空关联传感器的数据作为预测模型的输入, 而不需要根据人工经验手动选取目标节点邻近范围的交通路段流量数据作为预测模型的输入, 相对于凭借人工经验从邻近范围选取的传感器数据作为输入, 本发明提出的全自动时空关联数据挖掘提升了预测模型的预测性能。

附图说明

图 1 为本发明方法的流程图。

具体实施方式

如图 1 所示，本发明涉及的交通流预测可分为五个环节，包括数据采集、数据预处理、预测模型建立、时空关联数据挖掘和以交通流预测。其中涉及的数据采集是通过布局到交通网各个节点的传感器获得的交通流量数据；数据预处理是将采集的原始数据处理为有效的交通流数据；时空关联数据挖掘是通过稀疏表达的优化方法自动地选取对于预测目标相关的时空关联数据的方法；交通流预测是以时空关联数据为预测模型的输入进行的预测。具体实施方式如下：

实施例 1：

（步骤 1）数据采集：通过布局到交通网各个节点的传感器，采集以 30 秒为单位时间的交通数据流，得到整个交通网络各个传感器记录的数据，用矩阵的形式表示为：

$$S = \begin{bmatrix} s_1^1 & s_1^2 & \dots & s_1^M \\ s_2^1 & s_2^2 & \dots & s_2^M \\ \vdots & \vdots & \ddots & \vdots \\ s_N^1 & s_N^2 & \dots & s_N^M \end{bmatrix}$$

其中 M 为整个交通网中传感器的个数， N 为以 30 秒为单位的时间序列的长度， s_i^j 表示传感器 j 在第 i 个时刻对应的 30 秒内记录的交通流量；

（步骤 2）预处理：首先对原始数据进行时间尺度调整，将单位时间为 30 秒的交通流数据转化为单位时间为 10 分钟的数据，即数据转化为如下矩阵形式：

$$\hat{V} = \begin{bmatrix} v_1^1 & v_1^2 & \dots & v_1^M \\ v_2^1 & v_2^2 & \dots & v_2^M \\ \vdots & \vdots & \ddots & \vdots \\ v_n^1 & v_n^2 & \dots & v_n^M \end{bmatrix}$$

其中 $v_i^j = \sum_{k=1}^{20} s_{(i-1) \times 20 + k}^j$ ， $n = \frac{N}{20}$ ；然后对 $\hat{V}$ 的每一列求得标准差，记 $std(j)$ 为第 j

列的标准差，如果 $std(j) < 20$ ，则认为传感器 j 的记录为无关记录，从矩阵 $\hat{V}$ 中删除，从而得到有效交通流数据 $V \in R^{m \times n}$ ，其中 m 为预处理后参与预测的传感器的个数；完成如上所述的预处理后，用于交通流预测的数据转化为如下矩阵形

式：

$$V = \begin{bmatrix} v_1^1 & v_1^2 & \dots & v_1^m \\ v_2^1 & v_2^2 & \dots & v_2^m \\ \vdots & \vdots & \vdots & \vdots \\ v_n^1 & v_n^2 & \dots & v_n^m \end{bmatrix}$$

(步骤3) 建立预测模型：令 v_i^j 表示交通网中的传感器 j 在第 i 个时刻记录的交通流数据，对于交通网中参与预测的 m 个传感器，则在第 i 个时刻整个交通网的状态表示为 $V_i = [v_i^1, v_i^2, \dots, v_i^m]$ ，提前 τ 时刻对传感器 j 的流量进行预测的多因子线性回归模型为：

$$\bar{v}_{i+\tau}^j = V_i w^j$$

上式中的权重 $w^j = [w_1^j, w_2^j, \dots, w_k^j, \dots, w_m^j]^T$ 为待优化的模型参数， $\bar{v}_{i+\tau}^j$ 为预测值， τ 分别取 1、2、3、4、5、6，即分别提前 10 分钟、20 分钟、30 分钟、40 分钟、50 分钟、60 分钟对传感器 j 的流量进行预测；

(步骤4) 时空关联性挖掘：通过稀疏表达的优化方法自动得到模型参数 w^j ，权重 $w^j = [w_1^j, w_2^j, \dots, w_k^j, \dots, w_m^j]^T$ 表示整个交通网络的各个传感器的数据对于预测目标传感器 j 的数据而言的时空关联性，当 $w_k^j = 0$ 时，传感器 k 的数据与传感器 j 的数据之间没有关联性，否则 w_k^j 值的大小表示传感器 k 的数据与传感器 j 的数据之间关联程度的强弱， $k=1, 2, \dots, m$ ；首先对 V 的每一列和 $v^j = [v_{\tau+1}^j, v_{\tau+2}^j, \dots, v_{n+\tau}^j]^T$ 进行如下归一化处理，得到

$$\bar{V} = \begin{bmatrix} \frac{v_1^1}{\sqrt{\sum_{i=1}^n (v_i^1)^2}} & \frac{v_1^2}{\sqrt{\sum_{i=1}^n (v_i^2)^2}} & \dots & \frac{v_1^m}{\sqrt{\sum_{i=1}^n (v_i^m)^2}} \\ \frac{v_2^1}{\sqrt{\sum_{i=1}^n (v_i^1)^2}} & \frac{v_2^2}{\sqrt{\sum_{i=1}^n (v_i^2)^2}} & \dots & \frac{v_2^m}{\sqrt{\sum_{i=1}^n (v_i^m)^2}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{v_n^1}{\sqrt{\sum_{i=1}^n (v_i^1)^2}} & \frac{v_n^2}{\sqrt{\sum_{i=1}^n (v_i^2)^2}} & \dots & \frac{v_n^m}{\sqrt{\sum_{i=1}^n (v_i^m)^2}} \end{bmatrix} \quad \text{和} \quad \bar{v}^j = \begin{bmatrix} \frac{v_{\tau+1}^j}{\sqrt{\sum_{i=1}^n (v_{\tau+i}^j)^2}} \\ \frac{v_{\tau+2}^j}{\sqrt{\sum_{i=1}^n (v_{\tau+i}^j)^2}} \\ \vdots \\ \frac{v_{\tau+n}^j}{\sqrt{\sum_{i=1}^n (v_{\tau+i}^j)^2}} \end{bmatrix};$$

所述的稀疏表达方法要优化的目标为：

$$\hat{w}^j = \arg \min_{w^j} \|\bar{V} w^j - \bar{v}^j\|_2^2 \quad \text{约束条件: } \|w^j\|_0 \leq L_0$$

上式中 L_0 为预先设定的门限，用于控制向量 w^j 中非零元素个数；上式表达的优化目标通过 Orthogonal Matching Pursuit (OMP) 算法进行求解【Tropp, J. A., 2004. Greed is good: algorithmic results for sparse approximation. IEEE Trans. Information

Theory 50, 2231-2242.】;

(步骤5) 将步骤(4)求解得到的权重 $\hat{w}^j$ 代入步骤(3)建立的预测模型, 得到如下预测结果:

$$\bar{v}_{i+\tau}^j = V_i \hat{w}^j$$

上式中, 只有 $\hat{w}^j$ 中不为零的元素才参与预测, 对应于时空关联传感器。

实施例 2:

除步骤 4 外, 其余部分与实施例 1 相同;

(步骤 4) 首先对 V 的每一列和 $v^j = [v_{\tau+1}^j, v_{\tau+2}^j, \dots, v_{\tau+n}^j]^T$ 进行如下归一化处理, 得到

$$\bar{V} = \begin{bmatrix} \frac{v_1^1}{\sqrt{\sum_{i=1}^n (v_i^1)^2}} & \frac{v_1^2}{\sqrt{\sum_{i=1}^n (v_i^2)^2}} & \dots & \frac{v_1^m}{\sqrt{\sum_{i=1}^n (v_i^m)^2}} \\ \frac{v_2^1}{\sqrt{\sum_{i=1}^n (v_i^1)^2}} & \frac{v_2^2}{\sqrt{\sum_{i=1}^n (v_i^2)^2}} & \dots & \frac{v_2^m}{\sqrt{\sum_{i=1}^n (v_i^m)^2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{v_n^1}{\sqrt{\sum_{i=1}^n (v_i^1)^2}} & \frac{v_n^2}{\sqrt{\sum_{i=1}^n (v_i^2)^2}} & \dots & \frac{v_n^m}{\sqrt{\sum_{i=1}^n (v_i^m)^2}} \end{bmatrix} \quad \text{和} \quad \bar{v}^j = \begin{bmatrix} \frac{v_{\tau+1}^j}{\sqrt{\sum_{i=1}^n (v_{\tau+i}^j)^2}} \\ \frac{v_{\tau+2}^j}{\sqrt{\sum_{i=1}^n (v_{\tau+i}^j)^2}} \\ \vdots \\ \frac{v_{\tau+n}^j}{\sqrt{\sum_{i=1}^n (v_{\tau+i}^j)^2}} \end{bmatrix};$$

稀疏表达方法要优化的目标为:

$$\hat{w}^j = \arg \min_{w^j} \|w^j\|_1 \quad \text{约束条件: } \|\bar{V}w^j - \bar{v}^j\|_2^2 = \sum_{i=1}^n \|\bar{V}_i w^j - \bar{v}^j\|^2 \leq \varepsilon_0$$

上式中 $\|w^j\|_1$ 为 w^j 的 1 范数, 表示向量 w^j 中所有元素绝对值之和, ε_0 是一个预先设定的门限, 用于控制预测误差 $\|\bar{V}w^j - \bar{v}^j\|_2$, $\|\bar{V}w^j - \bar{v}^j\|_2$ 表示向量 $\bar{V}w^j - \bar{v}^j$ 的 2 范数, 即向量中所有元素的平方和的平方根; 上式表达的优化目标等价地表示为加入拉格朗日乘子 λ 的无约束条件的优化目标:

$$\hat{w}^j = \arg \min_{w^j} \left\{ \lambda \|w^j\|_1 + \frac{1}{2} \|\bar{V}w^j - \bar{v}^j\|_2^2 \right\}$$

上式表达的优化目标通过 Least Angle Regression Stagewise (LARS) 算法进行求解【Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. 2004. Least angle regression. Annals of Statistics 32(2), 407-499.】; 求解时通过设定 λ 的值来控制 $\hat{w}^j$ 中非零元素的个数, 这里根据经验令 $\lambda = 0.001$, 以获得较优的模型参数;

实施例 3:

除步骤 5 外, 其余部分与实施例 2 相同;

步骤(5): 预测模型为向量自回归 (VAR) 模型【Chandra, S.R., Al-Deek,

H., 2009. Predictions of freeway traffic speeds and volumes using vector autoregressive models. Journal of Intelligent Transportation Systems 13 (2), 53–72.】，具体计算公式为：

$$v_t^j = \sum_{d=1}^p \sum_{i=1}^m w_i^j \psi_i^j B^d v_t^i + u_t^j$$

上式中： B 表示迟滞操作符，即 $B^d v_t = v_{t-d}$ ；如果传感器 i 是传感器 j 的关联传感器，即（步骤4）得到的 $\hat{w}^j$ 向量的第 i 个元素不为零，则令 $\psi_i^j=1$ ，反之令 $\psi_i^j=0$ ； u_t^j 为独立标准高斯噪声； w_i^j 为待优化的参数，可通过最大似然法训练得到；

实施例 4:

除步骤 5 外，其余部分与实施例 2 相同；

步骤（5）：预测模型为 BP 神经网络（参考【Bishop, Ch.M., 1995. Neural Networks for Pattern Recognition. Oxford university press, ISBN 0-19-853864-2.】），采用 3 层网络结构，输入层神经元个数为（步骤 4）得到的 $\hat{w}^j$ 向量中不为零元素的个数，输入数据来自 $\hat{w}^j$ 中不为零元素对应的传感器，隐层神经元设定为 5 个，输出层神经元为 1 个，隐含层传输函数为双极性 S 函数，输出层的传输函数为线性函数；

实施例 5:

除步骤 5 外，其余部分与实施例 2 相同；

步骤（5）：预测模型为 RBF 神经网络（参考【Bishop, Ch.M., 1995. Neural Networks for Pattern Recognition. Oxford university press, ISBN 0-19-853864-2.】），采用的核函数为高斯核函数，输入层神经元个数为（步骤 4）得到的 $\hat{w}^j$ 向量中不为零元素的个数，输入数据来自 $\hat{w}^j$ 中不为零元素对应的传感器，隐层神经元个数和输入层神经元个数相同；

基于实施例 2、3、4、5 所述的方法，利用真实交通数据测试了预测性能，数据下载网址 <http://www.d.umn.edu/tkwon/TMCdata/TMCarchive.html>，数据来自美国明尼苏达州的一个区域交通管理部门，通过几千个传感器记录了明尼苏达州双子城高速公路的交通流量。实验设置如下：训练数据从 2012 年 2 月 4 日到 2012 年 3 月 14 日 总共 40 天，测试数据从 2012 年 3 月 15 日到 2012 年 4 月 3 日 总共 20 天；经过预处理后，利用来自遍布于整个交通网络的 3254 个传感器的数据对 60 个目标传感器的交通流进行预测。

此外，进行了对比实验，除了时空关联传感器的选取方法不同以外，其它实验配置与实施 2、3、4、5 完全相同，对比实验中，根据人工经验选取预测目标节点一定邻近范围内的传感器作为关联传感器，将来自这些传感器的数据作为预测模型的输入，邻近传感器的个数分别取 10、15、20、25、30。

实验表明，本发明提出的时空关联数据挖掘方法与基于人工经验选取邻近传感器作为输入相比，提升了预测模型的预测性能，详细的结果见表 1-6，各表中，第一行的“10、15、20、25、30”分别表示取来自目标节点邻近的 10、15、20、25、30 个传感器的数据作为预测模型的输入，“稀疏”表示使用实施例 2 的方法自动挖掘出的时空关联传感器的数据作为预测模型的输入。表中所展示预测准确率定义如下【Min, W., Wynter, L., 2011. Real-time road traffic prediction with spatio-temporal correlations. Transportation Research Part C 19, 606–616.】：

$$Accuracy = 1 - \frac{1}{n} \sum_{i=1}^n \left| \frac{v_i^j - \bar{v}_i^j}{v_i^j} \right| \times 100\%$$

上式中 n 表示预测过的交通数据的总数， v_i^j 表示传感器 j 采集到的交通流的真实值， $\bar{v}_i^j$ 为预测值。表 1-6 列出的精度是 60 个目标传感器预测精度的平均值。

表 1: 选取 10,15,20,25,30 个邻近传感器作为关联传感器与稀疏表达方法挖掘的关联传感器在提前 10 分钟预测时的准确率

关联传感器	10	15	20	25	30	稀疏
线性回归模型	85.69	85.95	86.12	86.23	86.27	88.59
BP 神经网络	87.23	87.49	87.45	87.63	87.59	88.16
RBF 神经网络	86.71	86.98	87.01	86.96	86.85	88.17
向量自回归模型	87.39	87.39	87.42	87.43	87.46	87.82

表 2: 选取 10,15,20,25,30 个邻近传感器作为关联传感器与稀疏表达方法挖掘的关联传感器在提前 20 分钟预测时的准确率

关联传感器	10	15	20	25	30	稀疏
线性回归模型	83.50	83.61	83.85	83.94	84.02	86.75
BP 神经网络	85.43	85.56	85.51	85.80	85.79	86.25
RBF 神经网络	84.92	85.11	84.93	84.60	84.97	86.39
向量自回归模型	85.38	85.32	85.38	85.34	85.42	85.71

表 3: 选取 10,15,20,25,30 个邻近传感器作为关联传感器与稀疏表达方法挖掘的关联传感器在提前 30 分钟预测时的准确率

关联传感器	10	15	20	25	30	稀疏
线性回归模型	82.05	82.20	82.43	82.50	82.63	85.88
BP 神经网络	84.31	84.55	84.69	84.68	84.95	85.67
RBF 神经网络	83.76	84.01	83.91	83.70	84.04	85.52
向量自回归模型	83.89	83.84	83.88	83.89	83.98	84.69

表 4: 选取 10,15,20,25,30 个邻近传感器作为关联传感器点与稀疏表达方法挖掘的关联传感器在提前 40 分钟预测时的准确率

关联传感器	10	15	20	25	30	稀疏
线性回归模型	80.49	80.62	80.81	80.89	81.01	84.99
BP 神经网络	83.16	83.43	83.64	83.58	83.76	85.08
RBF 神经网络	82.32	82.68	82.65	82.49	82.88	85.15
向量自回归模型	82.07	82.05	82.15	82.17	82.22	84.10

表 5: 选取 10,15,20,25,30 个邻近传感器作为关联传感器与稀疏表达方法挖掘的关联传感器在提前 50 分钟预测时的准确率

关联传感器	10	15	20	25	30	稀疏
线性回归模型	79.18	79.29	79.52	79.55	79.74	84.26
BP 神经网络	82.12	82.50	82.67	82.83	83.15	84.88
RBF 神经网络	81.19	81.69	81.63	81.56	81.88	84.43
向量自回归模型	80.55	80.48	80.59	80.61	80.68	83.44

表 6: 选取 10,15,20,25,30 个邻近传感器作为关联传感器与稀疏表达方法挖掘的关联传感器在提前 60 分钟预测时的准确率

关联传感器	10	15	20	25	30	稀疏
线性回归模型	77.97	78.06	78.32	78.32	78.53	83.92
BP 神经网络	81.58	81.98	82.15	82.22	82.46	84.42
RBF 神经网络	80.31	80.89	81.08	80.98	81.33	84.33
向量自回归模型	79.08	79.09	79.22	79.25	79.34	82.94

权 利 要 求

1. 一种基于时空关联数据挖掘的交通流预测方法，其特征在于具体步骤为：

(a) 通过布局到交通网各个节点的传感器采集交通流量的原始数据；

(b) 通过数据预处理，将采集的原始数据处理为有效的交通流数据；

(c) 建立预测模型：令 v_i^j 表示交通网中的传感器 j 在第 i 个时刻采集的交通流量数据，假设一个交通网中有 m 个传感器，则在第 i 个时刻整个交通网的状态表示为 $V_i = [v_i^1, v_i^2, \dots, v_i^m]$ ，提前 τ 时刻对传感器 j 采集的交通流量数据进行预测的线性回归模型为：

$$\bar{v}_{i+\tau}^j = V_i w^j;$$

上式中的权重 $w^j = [w_1^j, w_2^j, \dots, w_k^j, \dots, w_m^j]^T$ 为待优化的模型参数， $\bar{v}_{i+\tau}^j$ 为预测值；

(d) 挖掘时空关联性：通过稀疏表达的优化方法得到模型参数 w^j ， $w^j = [w_1^j, w_2^j, \dots, w_k^j, \dots, w_m^j]^T$ 表示整个交通网络的各个传感器的数据对于预测目标传感器 j 的数据而言的时空关联性，当 $w_k^j=0$ 时，传感器 k 的数据与传感器 j 的数据之间没有关联性，否则 w_k^j 值的大小表示传感器 k 的数据与传感器 j 的数据之间关联程度的强弱， $k=1, 2, \dots, m$ ；

(e) 以时空关联数据为预测模型的输入，进行交通流预测。

2. 根据权利要求 1 所述的预测方法，其特征在于步骤 (d) 所述通过稀疏表达优化方法要优化的目标为：

$$\hat{w}^j = \arg \min_{w^j} \|w^j\|_0 \quad \text{约束条件:}$$

$$\|V w^j - v^j\|_2^2 = \sum_{i=1}^n \|V_i w^j - v_i^j\|^2 \leq \varepsilon_0;$$

上式中 $V = (V_1^T, V_2^T, \dots, V_n^T)^T$ 表示从第 1 到第 n 时刻所有传感器记录的交通流数据，其中第 i 个时刻交通网中 m 个传感器记录的交通流量为

$V_i = [v_i^1, v_i^2, \dots, v_i^m]$, $i=1, 2, \dots, m$, $v^j = [v_{\tau+1}^j, v_{\tau+2}^j, \dots, v_{\tau+n}^j]^T$ 表示传感器 j 记录的从第 $\tau+1$ 时刻到第 $\tau+n$ 时刻的交通流量, $\|w^j\|_0$ 表示 w^j 的 0 范数, 即权重向量 w^j 中非零元素的个数, ε_0 是一个预先设定的门限, 用于控制预测误差 $\|Vw^j - v^j\|_2$, $\|Vw^j - v^j\|_2$ 表示向量 $Vw^j - v^j$ 的 2 范数, 即向量中所有元素的平方和的平方根。

3. 根据权利要求 1 所述的预测方法, 其特征在于步骤 (d) 所述稀疏表达优化方法要优化的目标为:

$$\hat{w}^j = \arg \min_{w^j} \|Vw^j - v^j\|_2^2 \quad \text{约束条件: } \|w^j\|_0 \leq L_0;$$

上式中 L_0 为预先设定的门限, 用于控制向量 w^j 中非零元素个数; 上式表达的优化目标通过 Orthogonal Matching Pursuit (OMP) 算法进行求解。

4. 根据权利要求 1 所述的预测方法, 其特征在于步骤 (d) 所述的稀疏表达优化方法要优化的目标为:

$$\hat{w}^j = \arg \min_{w^j} \|w^j\|_1 \quad \text{约束条件:}$$

$$\|Vw^j - v^j\|_2^2 = \sum_{i=1}^n \|V_i w^j - v^j\|^2 \leq \varepsilon_0;$$

上式中 $\|w^j\|_1$ 为 w^j 的 1 范数, 表示向量 w^j 中所有元素绝对值之和, ε_0 是一个预先设定的门限, 用于控制预测误差 $\|Vw^j - v^j\|_2$; 上式表达的优化目标等价地表示为加入拉格朗日乘子 λ 的无约束条件的优化目标:

$$\hat{w}^j = \arg \min_{w^j} \left\{ \lambda \|w^j\|_1 + \frac{1}{2} \|Vw^j - v^j\|_2^2 \right\}$$

上式表达的优化目标通过 Least Angle Regression Stagewise (LARS) 算法进行求解。

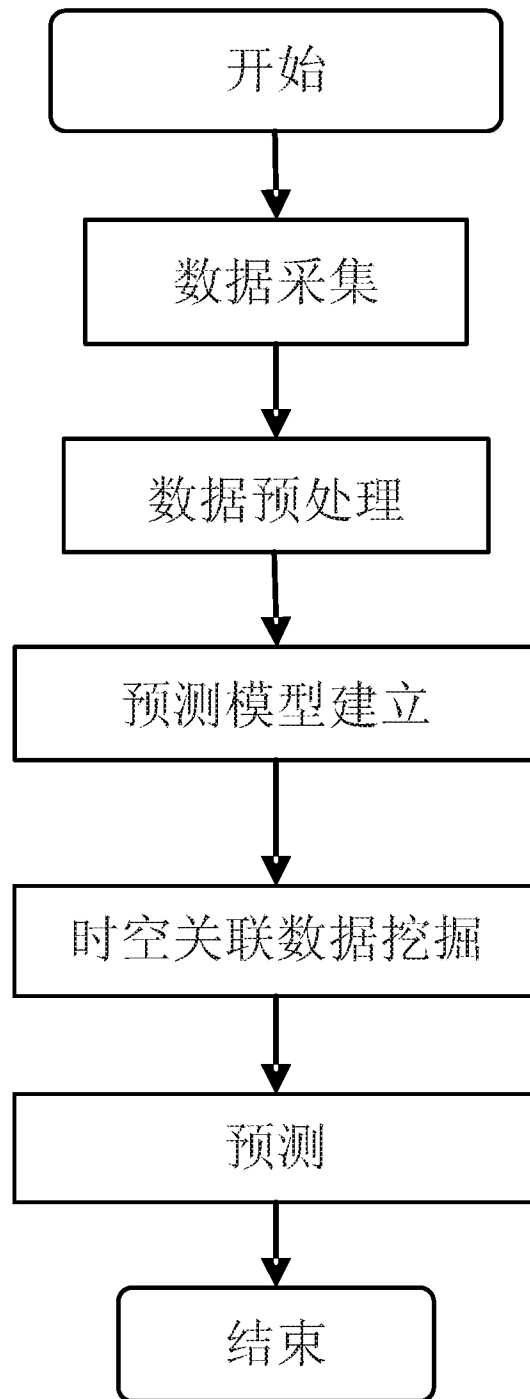


图 1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/081350

A. CLASSIFICATION OF SUBJECT MATTER

G08G 1/00 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G08G; G06F; H04W

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC: TRAFFIC, FLOW, PREDICT+, SPARS+, SPATIAL 3W TEMPORAL, REGRESSION

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 103700255 A (FUDAN UNIVERSITY) 02 April 2014 (02.04.2014) claims 1-4	1-4
PX	SHI, Shixiong et al. "A Novel Traffic Prediction Method Based on Sparse Representation and RBF Neural Network" Microcomputer Applications, no. 5, vol. 30, 31 May 2014 (31.05.2014) sections 1-3	1-4
A	CN 102034350 A (BEIJING STONE INTELLIGENT TRAFFIC SYSTEM INTEGRATION CO., LTD.) 27 April 2011 (27.04.2011) description, paragraphs [0074]-[0184] and figure 10	1-4
A	CN 102169627 A (BEIJING JIAOTONG UNIVERSITY) 31 August 2011 (31.08.2011) the whole document	1-4
A	US 2010306290 A1 (ZHANG, Yin et al.) 02 December 2010 (02.12.2010) the whole document	1-4

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 18 August 2014	Date of mailing of the international search report 05 September 2014
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer YAN, Jie Telephone No. (86-10) 62413393

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/CN2014/081350

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
US 2010306290 A1	02 December 2010	US 8458109 B2	04 June 2013
		WO 2010138536 A1	02 December 2010
CN 102034350 A	27 April 2011	CN 102034350 B	25 July 2012
CN 103700255 A	02 April 2014	None	
CN 102169627 A	31 August 2011	None	

A. 主题的分类		
G08G 1/00 (2006.01) i		
按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类		
B. 检索领域		
检索的最低限度文献(标明分类系统和分类号)		
G08G, G06F, H04W		
包含在检索领域中的除最低限度文献以外的检索文献		
在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))		
CNPAT, CNKI, WPI, EPODOC:交通, 流, 预测, 稀疏, 时空, 回归, TRAFFIC, FLOW, PREDICT+, SPARS+, SPATIAL 3W TEMPORAL, REGRESSION		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 103700255 A (复旦大学) 2014年 4月 02日 (2014 - 04 - 02) 权利要求第1-4项	1-4
PX	史世雄等. "基于稀疏表达和RBF神经网络的交通流预测方法" 微型电脑应用, 第5期卷, 第30卷期, 2014年 5月 31日 (2014 - 05 - 31), 第1-3节	1-4
A	CN 102034350 A (北京四通智能交通系统集成有限公司) 2011年 4月 27日 (2011 - 04 - 27) 说明书第[0074]-[0184]段、附图10	1-4
A	CN 102169627 A (北京交通大学) 2011年 8月 31日 (2011 - 08 - 31) 全文	1-4
A	US 2010306290 A1 (ZHANG YIN ET AL.) 2010年 12月 02日 (2010 - 12 - 02) 全文	1-4
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: "A" 认为不特别相关的表示了现有技术一般状态的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件 "T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 "&" 同族专利的文件		
国际检索实际完成的日期		国际检索报告邮寄日期
2014年 8月 18日		2014年 9月 05日
ISA/CN的名称和邮寄地址		受权官员
中华人民共和国国家知识产权局(ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国		阎洁
传真号 (86-10)62019451		电话号码 (86-10)62413393

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2014/081350

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
US	2010306290	A1	2010年 12月 02日	US 8458109 B2	2013年 6月 04日
				WO 2010138536 A1	2010年 12月 02日
CN	102034350	A	2011年 4月 27日	CN 102034350 B	2012年 7月 25日
CN	103700255	A	2014年 4月 02日	无	
CN	102169627	A	2011年 8月 31日	无	