首先，从仿真结果中我们可以看出，当在对交通流量进行预测时，对于预测精度的影响主要是来源于对于空间相关的检测线圈节点的交通数据，而时间上并不会造成很大的影响；其次，我们对于道路交通中的20个检测线圈进行了总计8000次的分析中发现，对于5分钟的交通流量预测问题来说，不需要很长的时间间隔便能够得到较好的预测精度，反而当时间间隔增加后，预测精度不但不会增加，反而会降低预测准确度。

季节性*SARIMA(1,1,1)(1,1,1)288* [Billy M. Williams]，通过In this paper, the Akaike Information Criterion (AIC) (Chen et al., 2012) is used to determine the appropriate order of the comparative model，对于20个不同路段进行了仿真实验，*SARIMA*模型的参数拟合在数据不断获取后更新。

时空数据分布对预测结果的不平衡性，如对于5分钟的流量预测而言，预测误差主要来自于空间数据，而时间序列数据的贡献比较小，对于10-20分钟的交通流量预测来说，时间上的滞后和空间的采样点个数对于预测结果都有明显的影响，为此我们对于时间和空间的检测点的数目进行选取，以达到充分挖掘交通流量的时间和空间相关关系。

为什么要从Rinitial开始

首先，我们应该说明，第一个提交版本中确实没有对这个部分叙述地清晰，这可能会导致读者的误解，对于图6中的结论是对应于5分钟的流量预测问题，对于不同的预测问题（如预测长度发生变化）应该单独进行数据选择和模型训练。从第N章的图F可以看出，实际对于5分钟的交通流量预测而言，延长时滞对提升预测效果并没有带来有效的提升，更进一步地，如果增加时间间隔能够明显地（这取决于不敏感系数α的选取）降低预测误差的话，STFSA是可以将预测时间滞后延长到24h甚至更长。我们知道在时间预测问题中加入一定的季节性分析可能会给预测结果带来一定的提升，但是对于高速路交通问题来说，我们的研究希望最后筛选出相对较小的输入数据以降低模型的复杂度，从而能让模型在小计算能力的设备上也能处理。

STFSA给出的步骤：

给出STFSA的完整工作过程流程图 🡪 对于时空数据MAPE和MAE分布图的分析引出时空两个维度信息不平衡 🡪 给出STFSA目标函数并对其中的不敏感系数加以解释 🡪 给出最终的寻优算法。



目前在文章中发现了时空数据选取的不平衡性，某些时间或者空间数据的加入并不能明显地提升预测的准确度，我们的预测算法中并不会将这些数据作为模型的输入数据。因此在STFSA中我们引入了一个不敏感系数α来权衡输入数据对结果的影响，降低输入数据的维度大小，减少模型的复杂度。在本文中受限制与篇幅只对预测数据的选取作定量分析，在我们提出的时空特征选取方法并不能找到输入数据的最优解，也不能描述与最优解的近似程度，是一种启发式的算法，能够在有限的计算时间内给出一个可以接受的解。而在以后的工作中将会对在考虑车辆的空间可到达性和车流时间相关长度的定量分析。

显著性，然而，上述分析并没有回答我们在MAPE的基础上有多少统计信心认为一种方法比另一种方法“更好”的问题。因此,绝对的错误百分比从每个k=20预测进行了分析,以及绝对的错误百分比naïve方法和季节性ARIMA预测,相关使用弗里德曼样品测试和Wilcoxon显著性等级检验。一个α=0.05显著性水平是用于假设检验。这些相关样本分析的结果如表2和表3所示。Friedman和Wilcoxon的signed-rank测试操作的是相关度量的秩(在本例中是绝对百分比错误)，而不是实际度量。换句话说，测试将每个预测点的最低绝对百分比错误转换为第1级，第2级最低转换为第2级，依此类推。表2和表3中所列的八种预测方法均按预测平均秩的升序排列。这种顺序不一定与表中给出的MAPE一致。

The temporal evolutions and spatial dependencies of network traffic are considered and applied simultaneously in traffic prediction problems by exploiting the proposed image-based method and deep learning architecture of CNNs. naïve



In the feature subset selection problem, a learning algorithm is faced with the problem of selecting a relevant subset of features upon which to focus its attention, while ignoring the rest.

The feature subset selection algorithm conducts a search for a good subset using the induction algorithm itself as part of the function evaluating feature subsets. The induction algorithm is run on the dataset, usually partitioned into internal training and holdout sets, with different sets of features removed from the data. The feature subset with the highest evaluation is chosen as the final set on which to run the induction algorithm. The resulting classifier is then evaluated on an independent test set that was not used during the search.

In Section 3, we investigate the search engine used to search for feature subsets and show that greedy search (hill-climbing) is inferior to best-first search. we modify the connectivity of the search space to improve the running time.

在本节中，我们将研究如何找到一个好的特征子集及其与相关特征集的关系。我们展示了现有的相关性定义存在的问题，并展示了如何将相关特性划分为弱的和强的两个系列，以帮助我们更好地理解这个问题。我们研究了用于特征子集选择的两种通用方法:筛选器方法和包装器方法，然后详细研究每种方法。

This paper is organized as follows. In Section 2, we review the feature subset selection

problem, investigate the notion of relevance, define the task of finding optimal features,

and describe the filter and wrapper approaches. In Section 3, we investigate the search

engine used to search for feature subsets and show that greedy search (hill-climbing) is

inferior to best-first search. In Section 4, we modify the connectivity of the search space

to improve the running time. Section 5 contains a comparison of the best methods found.

In Section 6, we discuss one potential problem in the approach, over-fitting, and suggest

a theoretical model that generalizes the feature subset selection problem in Section 7.

Related work is given in Section 8, future work is discussed in Section 9, and we

conclude with a summary in Section 10.

引用文献包括：

1. 统计显著性分析
2. KNN
3. SARIMA

In practical learning scenarios, however, we are faced with two problems: (1) the learning algorithms are not given access to the underlying distribution, (2) most practical algorithms attempt to find a hypothesis by approximating NP-hard optimization problems. The first problem is closely related to the bias-variance tradeoff [ 36,611: one must trade

off estimation of more parameters (bias reduction) with accurately estimating these parameters (variance reduction). This problem is independent of the computational power available to the learner. The second problem, that of finding a “best” (or approximately best) hypothesis, is usually intractable and thus poses an added computational burden. For example, decision tree induction algorithms usually attempt to find a small tree that fits the data well, yet finding the optimal binary decision tree is NP-hard [ 42,451. For neural networks, the problem is even harder; the problem of loading a three-node neural network with a training set is NP-hard if the nodes compute linear threshold functions [ 12,481.

数据集有限，预测算法无法知道底层的数据的真实分布。

预测算法的优化过程不能充分利用数据，most practical algorithms attempt to find a hypothesis by approximating NP-hard optimization problems