# Bayesian Network Methods for Traffic Flow Forecasting with Incomplete Data

Shiliang Sun[1], Changshui Zhang[1], Guoqiang Yu[1], Naijiang Lu[2], and Fei Xiao[2]

[1] State Key Laboratory of Intelligent Technology and Systems,
Department of Automation, Tsinghua University, Beijing, China, 100084
sunsl02@mails.tsinghua.edu.cn
zcs@mail.tsinghua.edu.cn, ygq01@mails.tsinghua.edu.cn
[2] Shanghai Cogent Biometrics Identification Technology Co. Ltd., China
{lunj,xiaof}@cbitech.com

**Abstract.** Traffic flow forecasting is an important issue in the field of Intelligent Transportation Systems. Due to practical limitations, traffic flows recorded can be partially missing or unavailable. In this case few methods can deal with forecasting successfully. In this paper two methods based on the concept of Bayesian networks are originally proposed to cope with this matter. A Bayesian network model and a two-step Bayesian network model are constructed respectively to describe the causal relationship among traffic flows, and then the joint probability distribution between the cause and effect nodes with its dimension reduced by Principal Component Analysis is approximated through a Gaussian Mixture Model. The parameters of the Gaussian Mixture Model are learned through the Competitive EM algorithm. Experiments show that the proposed Bayesian network methods are applicable and effective for traffic flow forecasting with incomplete data.

## 1 Introduction

In recent years, Intelligent Transportation Systems (ITS) have achieved great developments. However, many problems, including traffic management and congestion control, still remain unsolved. Precise analysis of historical trends of traffic flows or transportation forecasting therefore becomes fundamental tasks in order to progress successfully and smoothly in our daily routines. On the other hand, many approaches and methods in the machine learning field were presented recently and got highly developed, which can be instructive to the traffic flow forecasting problem. In this paper, we concentrate on using machine learning methods to deal with the application-orientated problem of short-term traffic flow forecasting in ITS.

Short-term traffic flow forecasting, which is to determine the traffic volume in the next time interval usually in the range of five minutes to half an hour, is an important problem in the research area of ITS. In the past few years, many theories and methods on this theme were proposed including those based on time series models (including ARIMA, seasonal ARIMA), Kalman filter theory,

J.-F. Boulicaut et al. (Eds.): ECML 2004, LNAI 3201, pp. 419–428, 2004.

neural network approaches, non-parametric methods, simulation models, local regression models and layered models [1]∼[6]. Although these theories and methods have alleviated difficulties in traffic flow modelling and forecasting to some extent, they hardly work when the data used for forecasting is partially missing or unavailable, while this case of incomplete data often occurs in practice. Although the historical average (fill up the incomplete data with their historical average) method is often applied to cope with this issue, the forecasting performance is quite limited.

The Bayesian network approach, as studied comprehensively in the community of machine learning, gives us some inspiration on traffic flow forecasting. Considering the nature of short-term traffic flows, we can draw the conclusion that the traffic flow at a given road link is closely related to those of its neighbors, and thus in order to forecast as accurately as possible, we should also take into account the information provided by neighbor links. This is consistent with the ideology of the Bayesian network model. Bayesian network is such a model that can fully take into account the causal relationship between random variables statistically and has the capability to encode incomplete data.

Our main contribution of the paper is that we introduce the concept and approach of Bayesian network in machine learning field to the area of ITS for the first time and effectively solve the traffic flow forecasting problem with incomplete data. For a given road net, we focus on constructing rational Bayesian networks, learn the statistical relations between the cause and effect nodes, and then based on the statistical relationship carry out forecasting. Experiments for real-world short-term vehicular flow forecasting in case of incomplete data are carried out to validate our methods. A comparison with the autoregressive (AR) model and the historical average method shows that our Bayesian network methods are appropriate and effective for this kind of traffic flow forecasting problem.

The remainder of this paper is organized as follows. After introducing Bayesian networks and our two Bayesian network methods for traffic flow forecasting briefly in section 2 and 3 respectively, we describe the approaches and techniques related with our methods in section 4. Section 5 reports the experimental results on real-world traffic flow data, and gives a performance comparison among several methods. Section 6 concludes the paper and discusses some directions of future research work.

## 2 Bayesian Networks

A Bayesian network, also known as a casual model, is simply a directed graphical model for representing conditional independencies between a set of random variables. In a Bayesian network, an arc from node A to B can be informally interpreted as indicating that A "causes" B [7]. Suppose we have several random variables denoted by $x_1, x_2, ..., x_m, y_1, y_2, ..., y_n$ and $z$ respectively. $x_1, x_2, ..., x_m$ are used to forecast $y_1$ and $y_1, y_2, ..., y_n$ are used to forecast $z$ in turn. Then considering the causal relations in variable forecasting, we can construct two

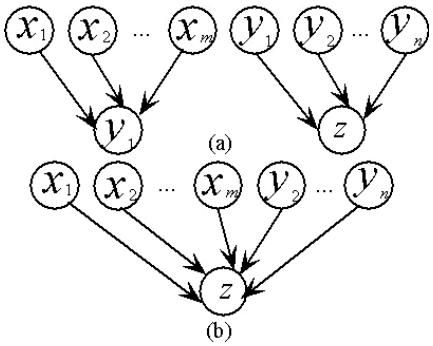


**Fig. 1.** (a) Two Bayesian networks. (b) A newly constructed Bayesian network

Bayesian networks as shown in Fig. 1.a where arrows start from the cause nodes and point to the effect nodes.

## 3 Two Bayesian Network Methods

Fig. 2.a is a patch of one urban traffic map. Each circle node denotes a road link. Each arrow shows the direction of the traffic flow on the corresponding road. We take road link $D_d$ as an example to show our approach. From the view point of Bayesian Network, vehicle flows of $C_e, C_g$ and $C_h$ should have causal relations with vehicle flow of $D_d$. Similarly, vehicle flows of $B_a$ and $B_c$ should have causal relations with vehicle flow of $C_h$. Furthermore, considering the time factor, to predict the vehicle flow of $D_d$ at time $t$ (denoted by $D_d(t)$)we should use values $D_d(t-1), D_d(t-2), ..., D_d(t-d)$ as well. That is, some historical values of $C_e, C_g, C_h$ and $D_d$ could be regarded as the cause nodes of $D_d(t)$ in a Bayesian network. We show this causal relation in Fig. 2.b.

However, if traffic flow data $C_h(t-m)$ is missing, how can we forecast $D_d(t)$? We propose two methods to deal with this issue.

*Method 1.* Construct a new Bayesian network to model the whole causal relationship. We take Fig. 1.a as an example to show the procedure. Suppose data for random variable $y_1$ is missing while data for $x_1, x_2, ..., x_m, y_2, ..., y_n$ is complete (intact), we can construct another Bayesian network to describe the whole causal relation, which is given in Fig. 1.b. The new Bayesian network is constructed and expanded by absorbing the node $y_1$, and thus in the graph, $x_1, x_2, ..., x_m, y_2, ..., y_n$ serve as the cause nodes of $z$ . Similarly, the newly constructed Bayesian network for forecasting $D_d(t)$ with missing data $C_h(t-m)$ can be obtained, as is shown in Fig. 3.

*Method 2.* Two-step Bayesian network method. That is, forecast $C_h(t-m)$ first and then using the result to forecast $D_d(t)$. Both steps use Bayesian network models to describe the causal relations among traffic flows.

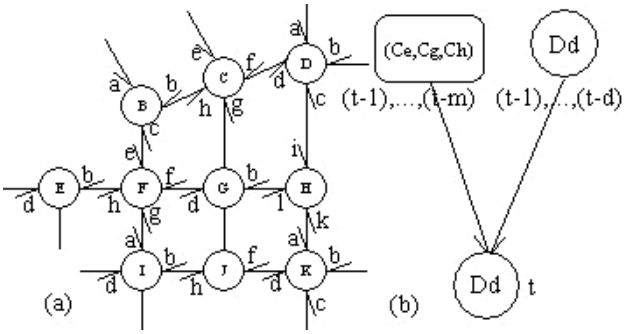


**Fig. 2.** (a) A patch taken from the whole map of Beijing City where UTC/SCOOT systems are placed. (b) The Bayesian network between the object road link and its cause neighbors

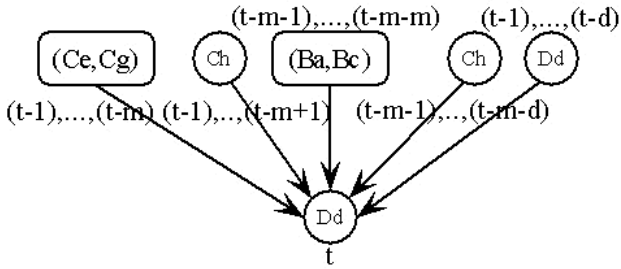


**Fig. 3.** The newly constructed Bayesian network for object road link $D_d$

# 4 Related Methods

## 4.1 PCA for Dimension Reduction

When using Bayesian networks, the joint probability distribution of related variables should be estimated. Usually the dimension of the joint probability distribution using Bayesian network is high and the data is not enough comparatively. So there might be a large error coming from the parameter learning stage. However, if we carry out parameter learning on a lower dimension with the same data, the estimation will be more accurate and efficient.

Principal Component Analysis (PCA) is such an effective tool for linear dimension reduction [8]. When using PCA for dimension reduction, we select some representative principal components from the input nodes, and then estimate the joint probability distribution among these components and the output node. Based on this new and concise relationship, we can carry out traffic flow forecasting more efficiently. In the paper, we using PCA to carry out dimension reduction for both the Bayesian network methods explained in section 3. More detail on how PCA is used will be given in the experiment section.

### 4.2 Representation and Parameter Learning of Joint Probability Distribution

How to estimate the joint probability density function (PDF) among all nodes in a constructed Bayesian network should be considered. The Gaussian Mixture Model (GMM), a combination of several Gaussian functions, is a widely used model applied to approximate an arbitrary PDF with enough accuracy. In this article, the joint probability distribution between cause nodes and effect node in a Bayesian network is approximated through GMM. Let $x$ denote one random variable or multidimensional random vector, and then the GMM form of its probability distribution with M mixture components can be represented as:

$$p(x|\Theta) = \sum_{l=1}^{M} a_l G(x|\theta_l) \tag{1}$$

where the parameters are $\Theta = (a_1, ..., a_M, \theta_1, ..., \theta_M)$ and $M$, s.t. $\sum_{l=1}^{M} a_l = 1$. Each $G(.)$ is a Gaussian PDF parameterized by $\theta_l = (\mu_l, \Sigma_l), l = 1, ..., M$.

Usually we use Maximum Likelihood Estimation (MLE) to carry out parameter learning with given data. Although the EM algorithm is an effective method to carry out MLE, it usually converges to local maxima [9]. To find a global maximum is more significant in most cases. The Competitive EM (CEM) algorithm presented by Zhang et al. overcomes the drawbacks of the basic EM algorithm and can be used for parameter learning of GMM [10]. It includes stages of EM iteration (E-step and M-step), split, merge and annihilation operations. The intial component number and model parameters can be set arbitrarily, and the split and merge operation can be selected efficiently by a competitive mechanism the authors proposed. Using the annihilation operation, CEM algorithm overcomes the problem of converging to the boundaries of parameter space. In other words, CEM algorithm can easily escape all kinds of local extrema and automatically determine the appropriate component number and the model parameters [10]. In this article, the parameters of a GMM which describe the joint PDF of the cause nodes and effect node in a Bayesian network are estimated through CEM algorithm.

### 4.3 Forecasting in Bayesian Networks

Traffic flow forecasting here can be regarded as an inference problem in a Bayesian network. The main goal of inference in a Bayesian network is to estimate the values of hidden nodes, given the values of observed nodes. We use this mechanism to implement forecasting of traffic flows. Suppose $(E, F)$ be a partitioning of the node indices of a Bayesian network into disjoint subsets, and $(x_E, x_F)$ be a partitioning of the corresponding random variables. Under the rule of Minimum Mean Square Error (M.M.S.E.), the optimal estimation of $x_F$ from $x_E$ can be given as [11]:

$$\hat{x}_F = E(x_F|x_E) \ . \tag{2}$$

To deduce the representation of the optimal forecasting $\hat{x}_F$ under the GMM setting, we employ the following lemma.

Lemma [12] Let $G(x;\mu,\Sigma)$ denote a multidimensional normal density function with mean $\mu^T = (\mu_1^T, \mu_2^T)$ and covariance matrix $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. $x^T = (x_1^T, x_2^T)$ is a random vector. Then we have:

$$p(x) = G(x_1;\mu_1,\Sigma_{11})G(x_2;\mu_{x_2|x_1},\Sigma_{x_2|x_1}) \ ,$$

where

$$\begin{aligned} \mu_{x_2|x_1} &= \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}(\mu_1 - x_1) \ , \\ \Sigma_{x_2|x_1} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \ . \end{aligned}$$

If we rewrite $x^T = (x_F^T, x_E^T), \mu_l^T = (\mu_{lF}^T, \mu_{lE}^T)$, $\Sigma_l = \begin{pmatrix} \Sigma_{lFF} & \Sigma_{lFE} \\ \Sigma_{lEF} & \Sigma_{lEE} \end{pmatrix}$, by the above lemma we can obtain:

$$\begin{aligned} p(x_F, x_E) &= \sum_{l=1}^{M} a_l G(x;\mu_l,\Sigma_l) \\ &= \sum_{l=1}^{M} a_l G(x_E;\mu_{lE},\Sigma_{lEE})G(x_F;\mu_{lF|E},\Sigma_{lF|E}) \ . \end{aligned}$$

The conditional probability density function can be represented as follows:

$$p(x_F|x_E) = \sum_{l=1}^{M} \beta_l G(x_F;\mu_{lF|E},\Sigma_{lF|E}) \ ,$$

where

$$\begin{aligned} \beta_l &= \frac{a_l G(x_E;\mu_{lE},\Sigma_{lEE})}{\sum_{j=1}^{M} a_j G(x_E;\mu_{jE},\Sigma_{jEE})} \ , \\ \mu_{lF|E} &= \mu_{lF} - \Sigma_{lFE}\Sigma_{lEE}^{-1}(\mu_{lE} - x_E) \ , \\ \Sigma_{lF|E} &= \Sigma_{lFF} - \Sigma_{lFE}\Sigma_{lEE}^{-1}\Sigma_{lEF} \ . \end{aligned}$$

Thus the optimal forecasting $\hat{x}_F$ under the criterion of M.M.S.E. has the following form:

$$\hat{x}_F = E(x_F|x_E) = \sum_{l=1}^{M} \beta_l \mu_{lF|E} \ , \tag{3}$$

where $\beta_l$ and $\mu_{lF|E}$ have the same meanings as above.

## 5 Experiments

The experimental data for analysis is the vehicle flow rates of discrete time series recorded every 15 minutes on many road links by the UTC/SCOOT system in Traffic Management Bureau of Beijing, whose unit is vehicles per hour (veh/hr). The data is from Mar. 1 to Mar. 25, 2002 and 2400 sample points totally. To evaluate our approach objectively, 2112 points (training set) of them are employed to learn parameters of GMM and the rest (test set) are employed to test the forecasting performance. Fig. 2.a shows the analyzed patch.

The flow chart of our forecasting procedure is described in Table 1.

**Table 1.** The Flow Chart of Our Forecasting Procedure

| |
|---|
| *Step 1:*<br>—Construct the Bayesian network models between input (cause nodes) and output (effect node) for a given road link using two methods explained in section 3.<br>*Step 2:*<br>—Approximate the joint probability distribution of all nodes in the network by PCA and GMM using methods explained in section 4.<br>*Step 3:*<br>—Carry out the optimal estimation of flow rates of the object road in the form of equation (3). |

In the experiment, the forecasting orders from the object road link and from the neighbor links are respectively taken as 4 and 5 empirically (parameters $d = 4, m = 5$). Then for Fig. 3 the joint probability distribution is: $p(C_e(t-j), C_g(t-j), B_a(t-j-5), B_c(t-j-5), C_h(.), D_d(t-j+1), j = 1, ..., 5)$, where $C_h(.) = (C_h(t-l), C_h(t-l-5), l = 1, ..., 4)$.

We can see the dimension of the joint probability distribution is very high (dimension=33), since the dimension for input nodes is high (input dimension=32). We should choose the reduced dimension number for input nodes by means of PCA. Fig. 4 shows the residual variance of PCA with different dimensions for the last joint probability distribution on the training data [13]. We see that using dimension 33 for forecasting is quite redundant and doesn't focus on the few essential dimensions. To effectively make use of data and gain well forecasting performance, we should look for the "elbow" where the curve ceases to decrease significantly with added dimensions. Based on the last curve of residual variance and the rectangle around the elbow, we apply PCA to the input nodes on the training data and select a few principal components (i.e. 5, 6, 7, 8 and 9 ) corresponding to the largest eigenvalues to represent the input data. Then, we respectively reduce the input data to these dimensions, built a GMM, carry out parameter learning through CEM algorithm and implement forecasting on the training set. From these several results, we can select the best one, and its corresponding reduced dimension of input data can also be identified. Experiments

show that the best forecasting accuracy is obtained at input dimension 8 (i.e. the dimension of the joint probability distribution is 9). Thus, we use the parameter configuration corresponding to input dimension 8 to implement traffic flow forecasting on the test set. Fig. 5 gives the final forecasting result of Bayesian network method on the test set for road $D_d$.

For other road links, we also carry out dimension selection by PCA before the parameter estimation of GMM. The final forecasting performances using the Bayesian network method and the two-step Bayesian network method are listed in Table 2 evaluated by root Mean Square Error (RMSE). For a given road link of Table 2, the smaller RMSE corresponds to the better forecasting performance (accuracy).

For vector $X = [x_1, x_2, ..., x_n]$ and its estimation $Y = [y_1, y_2, ..., y_n]$, the performance measure RMSE can be given in the following form:

$$RMSE(X, Y) = [\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}]^{1/2}.$$

Since the parameters $d$ and $m$ is just selected empirically, utilizing autoregressive (AR) model with the same parameter $d$ for comparison is reasonable (AR model is comparable which only uses historical flow rates of the object road link to forecast). The forecasting results through AR model and the historical average method (using the the average value of the historical flow rates at the corresponding time to forecast) are also given in Table 2.

**Table 2.** A Performance comparison of four methods for short-term traffic flow forecasting of four different road links

| Methods | $D_d$ | $J_f$ | $G_d$ | $C_f$ |
|---|---|---|---|---|
| Historical Average | 84.20 | 140.69 | 213.39 | 112.50 |
| AR | 66.14 | 123.65 | 155.20 | 90.76 |
| *Bayesian network* | *57.44* | *110.88* | *138.39* | *86.31* |
| *Two-step Bayesian network* | *53.95* | *113.96* | *140.40* | *87.73* |

From the experimental results we can see the outstanding improvements of forecasting capability brought by using Bayesian networks. For each of the four road links analyzed, the performances of the two Bayesian network methods are quite similar; and they both outperform the other two methods significantly. The Bayesian network method and the two-step Bayesian network method are both appropriate and effective for incomplete short-term traffic flow forecasting.

## 6 Conclusions and Future Work

In this paper, we first successfully introduce the concept and approach of Bayesian networks in the machine learning field to the community of ITS for the application problem of incomplete data forecasting. In the short-term traffic

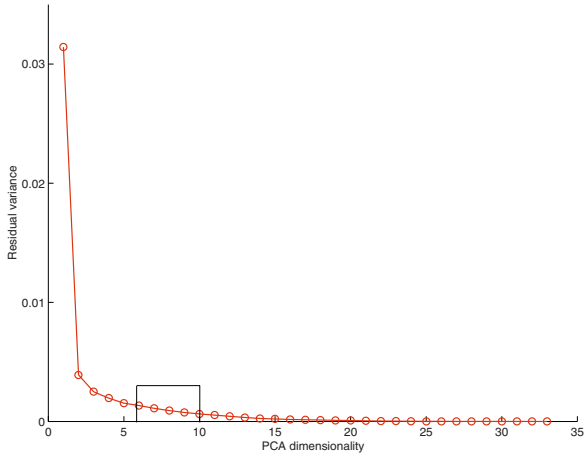


**Fig. 4.** Residual variance of PCA with different dimensions

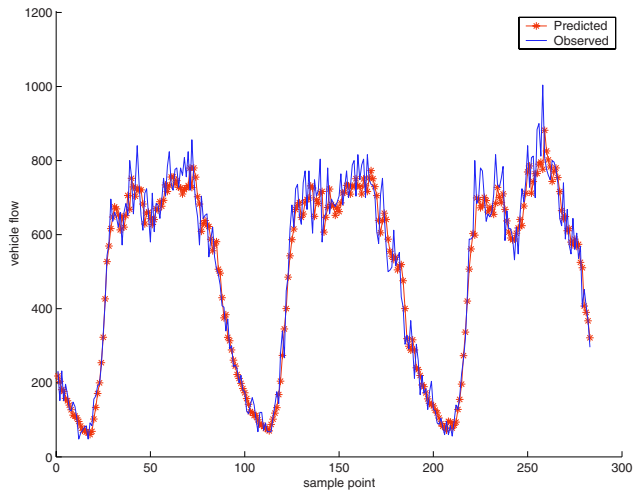


**Fig. 5.** Traffic flow forecasting result (veh/hr) for road $D_d$

flow forecasting theme, as vehicles usually keep travelling from one road link to its neighbor links, the traffic volume of one link at some given time can be regarded as the result of the historical flows of its neighbors and itself. The essence of traffic flow is consistent with the ideology of Bayesian networks. Besides, as Bayesian network encodes dependencies among all variables, it readily handles situations where some data entries are incomplete. Therefore, constructing Bayesian networks for traffic flow forecasting is reasonable. Experiments with real-world data also show that Bayesian network is applicable and effective for short-term traffic flow forecasting.

Though our proposed approach is more complicated than the AR model and some other models used for traffic flow forecasting, with the ever-increasing capability of digital computers coupled with a simultaneous decrease in the cost, the computation cost will not be a burden.

However, there are still some problems to be discussed and improved in the future, two of which are listed below.
1). How to elaborately select the prediction orders $d$ and $m$ for varying road conditions of different road links.
2). How to effectively combine Bayesian network with periodical information of traffic flows, etc.

## Acknowledgements

This work was supported by National High Technology Research and Development Program of China (863 Program) under contract No. 2001AA114190.

## References

1. William, B.M.: Modeling and Forecasting Vehicular Traffic Flow as a Seasonal Stochastic Time Series Process. Doctoral dissertation. University of Virginia, Charlottesville (1999)
2. Okutani, I., Stephanedes, Y.J.: Dynamic Prediction of Traffic Volume through Kalman Filter Theory. Transportation Research, Part B, Vol.18B (1984) 1-11
3. Edwards, T., Transley, D.S.W., Frank, R.J., Davey, N.: Traffic Trends Analysis using Neural Networks. http://homepages.feis.herts.ac.uk/∼nngroup/pubs/papers/ed-wards-iwannt97.pdf (1997)
4. Davis, G.A., Nihan, N.L.: Non-Parametric Regression and Short-Term Freeway Traffic Forecasting. Journal of Transportation Engineering (1991) 178-188
5. Roland, C., Joachim, W., Michael, S.: Traffic Forecast Using Simulations of Large Scale Networks. IEEE Intelligent Transportation Systems Conference Proceedings (2001)
6. Davis, G.A.: Adaptive Forecasting of Freeway Traffic Congestion. Transportation Research Record, No.1287. TRB, National Research Council, Washing, D.C. (1990)
7. Kevin, P.M.: An Introduction to Graphical Models. http://www.ai.mit.edu/∼murphyk/Papers/intro_gm.pdf (2001)
8. Jolliffe, I.T.: Principal Component Analysis. Springer-Verlag, New York (1986)
9. Jeff, A.B.: A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. International Computer Science Institute, Berkeley CA, TR-07-021 (1998)
10. Zhang, B.B., Zhang, C.S., Yi, X.: Competitive EM Algorithm for Finite Mixture Models. Pattern Recognition, Volume: 37, Issue: 1, January (2004) 131-144
11. Andrew, H.J.: Stochastic Processes and Filtering Theory . Academic Press, Inc., New York and London (1970)
12. Rao, C.R.: Linear Statistical Inference and Its Applications (Second Edition). John Wiley&Sons, Inc. (1973)
13. Tenenbaum, J.B., Silva, V.D., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science (2000), 290, 2319