TELKOMNIKA, Vol. 11, No. 10, October 2013, pp. 5523 ~ 5529
ISSN: 2302-4046

Traffic Prediction Based on Correlation of Road Sections

Xiaodan Huang, Wei Wang*
School of Information & Electrical Engineering
Hebei University of Engineering, Handan 056038, Hebei, China
*Corresponding author, e-mail: wangwei83@hebeu.edu.cn

Abstract

Road section data packet is very necessary for the estimation and prediction in short-time traffic condition. However, previous researches on this problem are lack of quantitative analysis. A section correlation analyzing method with traffic flow microwave data is proposed for this problem. It is based on the metric multidimensional scaling theory. With a dissimilarity matrix, scalar product matrix can be calculated. Subsequently, a reconstructing matrix of section traffic flow could be got with principal components factor analysis, which could display section groups in low dimension. It is verified that the new method is reliable and effective. After that, Auto Regressive Moving Average (A RMA) model is used for forecasting traffic flow and lane occupancy. Finally, a simulated example has shown that the technique is effective and exact. The theoretical analysis indicates that the forecasting model and algorithms have a broad prospect for practical application.

Keywords: *traffic flow, section correlation, metric multidimensional scaling, forecasting*

1. Introduction

Predicting road traffic conditions accurately and in time is the premise of normal operation in intelligent transportation system. Meanwhile it is the foundation of traffic information service, real-time traffic control and induction. It can effectively prevent the urban road network traffic jams happen, and then achieve the purpose of balancing traffic flow. However, during the short-time estimation of road network traffic flow, it is too complex to consider all the section data of road network as a whole. In order to satisfy the real-time traffic flow forecasting requirement, especially in large road networks, the estimation problem of road traffic state needs to be solved well.

In order to resolve this problem, a large-scale road network is divided into several small ones firstly on based of traffic flow section correlation. Then the study is on higher correlation sections in small ones, and predicts the short-term traffic flow. It is necessary, because it can increase short-term traffic flow prediction accuracy as well as ensure real-time in certain degree. In this paper, with the traffic microwave data, we focus on correlation analysis of road sections based on metric multidimensional scaling theory. From the quantitative analysis view, the implementation process is stated in details. Based on the grouped road section, ARMA model is used for forecasting traffic flow and lane occupancy.

Section 2 of this paper describes some related researches about road traffic flow forecasting. Based on the theory of multidimensional scaling and ARMA, Section 3 describes the method of grouping sections according to correlations and forecasting with historical data. Then some experiments are carried on in Section 4. At last, the conclusions are presented in Section 5.

2. Related Works

It experience for nearly 30 years that the traffic flow prediction is from single to multi-section [1]. Figure 1 shows the present situation for prediction researching on short-time traffic flow with multi-section.

Received April 21, 2013; Revised June 19, 2013; Accepted July 8, 2013

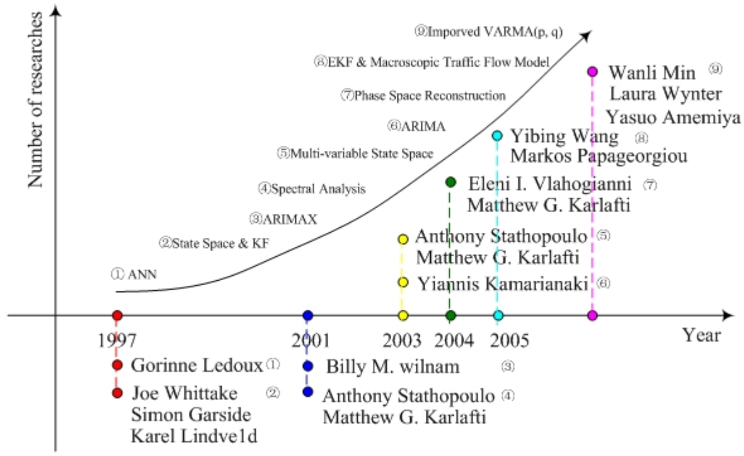


Figure 1. Researches on multi-section short-time traffic flow prediction

The researchers used different methods, from a variety of perspectives on multi-section traffic flow prediction theory. Joe Whittake analyzes a road network in Holland [2], Billy M. Wilnams analyzes the traffic data of the city of Bonn in Germany [3], Anthony Stathopoulos analyzes the traffic flow changes of five different locations in Athens in Greece [4], Wu analyze and predict inter-city traffic mode choice behavior [5], and Wanli Min predicts for a road network of Singapore.

Some extant studies show that there are relationships among each section's distance, position in road network and the network's topological structure. The change rule of these relationships in time and space is the basic gist of multi-section traffic flow forecasting, so it should be researched further more. In addition, some researches manifest that forecasting effect is better for taking into account the mutual influence of the multi-section traffic flow than for using one section data only. Moreover, it can adapt to unexpected situations better.

3. Forcasting with Correlative Information

3.1. Traffic Flow Model

With induction coils, magnetometer, microwave sensors, infrared ray sensors, ultrasonic sensors, and many types of traffic flow measuring equipment, a lot of traffic flow parameters can be got. Among them, there are three basic macroscopic characteristic parameters. They are the traffic flow $Q_l^i(k)$, average speed $V_l^i(k)$, and occupancy $O_l^i(k)$ of i^{th} section in l^{th} traffic lane. According to the second-order macroscopic random traffic flow theory [6-8], relationship among them is:

$$Q_l^i(k) = V_l^i(k) \cdot O_l^i(k) + \xi_l^i(k) \tag{1}$$

where, $\xi_l^i(k)$ is the noise of system model.

By the formula (1), the traffic flow information includes average speed and occupancy, so correlation analysis of road sections can be carried on with traffic flow data.

3.2. Correlation Analysis

When distinguishing correlation of multi-section traffic flow data, correlation coefficient between two sections can be obtained by statistical method, which demonstrates the relevance of two section data [9]. However, there is a problem. They can not directly show the relevance of all sections. During the range selection of road networks, we need to judge the correlation of the whole section traffic flow data, namely to distinguish which correlations between sections are relatively strong or weak. Many road sections are grouped based on the overall correlation. Thus a large-scale road network is divided into several smaller sub-road networks. This article applied the metric multidimensional scaling theory to solve this problem.

According to the theory of metric multidimensional scaling [10], we construct a dissimilarity matrix $\mathbf{\Delta}_{L\times L}$ based on the correlation coefficient of section traffic flow time series. L is the number of road section. The element δ_{ij} in $\mathbf{\Delta}$ is:

$$\delta_{ij} = 1 - \frac{\sum_{k=1}^{N}(\sum_{l=1}^{M_1} Q_l^i(k) - \overline{Q}^i)\cdot(\sum_{l=1}^{M_2} Q_l^j(k) - \overline{Q}^j)}{\left[\sum_{k=1}^{N}(\sum_{l=1}^{M_1} Q_l^i(k) - \overline{Q}^i)^2\right]^{1/2}\cdot\left[\sum_{k=1}^{N}(\sum_{l=1}^{M_1} Q_l^j(k) - \overline{Q}^j)^2\right]^{1/2}} \tag{2}$$

Then calculate scalar product matrix $\mathbf{\Gamma}$. Its element is:

$$\gamma_{ij} = -0.5\times(\delta_{ij}^2 - \delta_{i\bullet}^2 - \delta_{\bullet j}^2 + \delta_{\bullet\bullet}^2) \tag{3}$$

where, $\begin{cases} \delta_{i\bullet}^2 = \frac{1}{L}\sum_j \delta_{ij}^2 \\ \delta_{\bullet j}^2 = \frac{1}{L}\sum_i \delta_{ij}^2 \\ \delta_{\bullet\bullet}^2 = \frac{1}{L^2}\sum_i\sum_j \delta_{ij}^2 \end{cases}$.

Designate the section traffic flow reconstructed matrix in low-dimensional space as $\hat{\mathbf{\Omega}}_{L\times M}$. Correspondingly, its individual dissimilarity matrix is recorded as $\mathbf{D}$. Based on the theory of metric multidimensional scaling, $\mathbf{\Delta}$ is similar to $\mathbf{D}$. Then,

$$\mathbf{\Gamma} = \hat{\mathbf{\Omega}}\hat{\mathbf{\Omega}}' \tag{4}$$

Solving Eq. (4), the section traffic flow reconstructed matrix in low-dimensional space $\hat{\mathbf{\Omega}}$ can be obtained. That is to say the section differences can be represented in low-dimensional space.

It is also easy to obtain the eigenvalues λ_j which is corresponding to the j^{th} dimension of $\hat{\mathbf{\Omega}}$.

$$\lambda_j = \sum_i \hat{\omega}_{ij}^2 \tag{5}$$

where, $\hat{\omega}_{ij}$ are elements of $\hat{\mathbf{\Omega}}$.

3.3. ARMA Prediction Model for Traffic Flow

The traffic flow will be controlled better, if we can predict the traffic information at next time, the value of which is associated with the p numbers of previous values $\{Y_{t-p}, Y_{t-p-1}, ..., Y_{t-1}\}$ and the interference factor e. So the following model can be obtained.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q} \tag{6}$$

In this formula, Y_t satisfies the condition $E[Y(t)] = \mu_t$ and $e \sim N(0,\sigma^2)$, p and q is the order of the model, and $\phi_i\,(i = 1,2,..., p)$, $\theta_j\,(j = 1,2,..., q)$ is the parameters. Obviously, finding the best prediction model matched must sure the values of order (p,q) and parameters at first.

4. Experiment

This article adopts ring road microwave data of a northern city in China, which are recorded in November 2010, to verify the effect of the proposed method above. The traffic flow data includes: the section number, the lane number, the time, the traffic flow, the average speed, the occupancy and the large car traffic flow. Based on the discussion of the previous section, during the section correlation analysis, only the traffic flow data are used.

4.1. Data Preprocessing

Firstly, repair the missing data with the average value of the adjacent period data. So we can get the section traffic flow data from 2001 to 2011 detector per 5 minutes. Secondly, merger traffic flow data of various lane in the same direction into total flow, which is as the basic data for short-time traffic flow forecasting. Lastly, according to the total traffic flow of 11 sections in a day, calculate average flow of the first 15 days in November. Average traffic flow data can be obtained, and the detailed information about the preprocessed data is shown in Table 1.

Table 1. Features of Preprocessed Data

Dataset	Lane Direction	Lane Number	Time	Sequence Length	Usage
A	South to North	Section 3: 2 Lanes. Another section: 3 Lanes	Day 1-15, Nov., 0:00~23:55	288	Analysis
B	North to South	Section 9: 4 Lanes. Another section: 3 Lanes	Day 1-15, Nov., 0:00~23:55	288	Analysis

4.2. Testing

With Dataset A, we can test our method above and demonstrate its effect. The conventional method computes correlation coefficient between every two sections. The result is shown as Figure 2(a). From it, we can not distinguish class, to which sections are grouped together, especially for increasing sections, because they have large amount of information and are not intuitional.

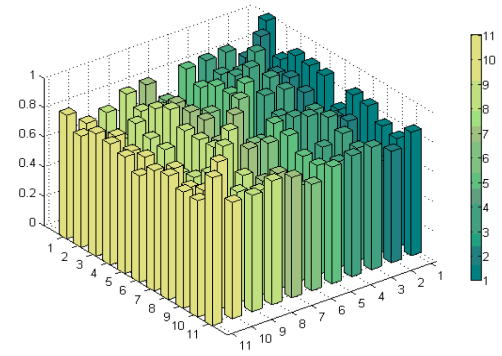
By using metric multidimensional scaling theory, eigenvalues λ_j which is corresponding to the j^{th} dimension of $\hat{\boldsymbol{\Omega}}$ can be obtained. As shown in Figure 2(b), curves illustrating the changing tendency of eigenvalues do not decrease until number of dimensions equals 4. It is seen that all sections could be divided into 4 groups without distortion almost. With reconstruction of the road section in 2 dimension space, shown in Figure 2(c), 4 groups are: Section 1-5, 7, 8, 11, Section 6, Section 9 and Section 10.

Its linear fitting scattergraph is displayed in Figure 2(d). All points are almost in a straight line, so the fitting is better.

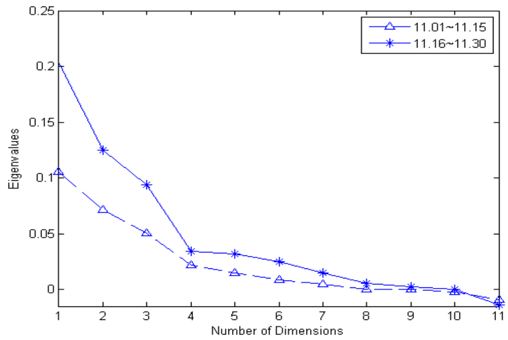
Thus, using the method proposed above, we can group road sections from rich and non-intuitive information which is difficult to distinguish. These works are necessary for later traffic estimation and forecasting.

Correspondingly, we analyze Dataset B. Figure 3(a) is correlation coefficient between every two sections. And Figure 3(b) also demonstrates all sections could be divided into 4 groups without distortion almost. They are not Section 1-5, 7, 8, 11, Section 6, Section 9 and Section 10, but Section 1-4, 7-9, 11, Section 5, Section 6 and Section 10 shown in Figure 3(c). Its linear fitting scattergraph shown in Figure 3(d) displays the fitting is better too.

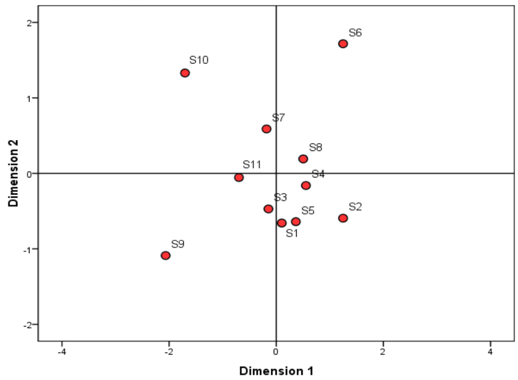
During the analysis, goodness of fittest indicates how well the iterative algorithm fits the model. Kruskal's Stress and squared correlation (RSQ) are usually used. While Stress grows smaller and RSQ gets bigger, effect is better. Their values of the first 15 days in November are shown in Table 2. The RSQ values are smaller than 0.6 both, so results are acceptable.



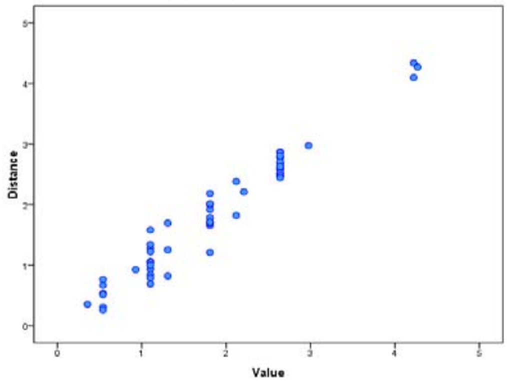
(a) Correlation coefficient



(b) Eigenvalues

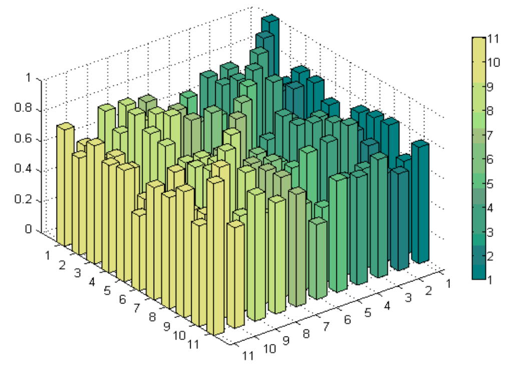


(c) Reconstruction in 2 dimensions

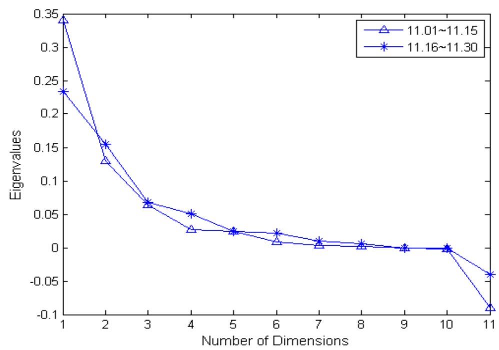


(d) Linear fitting scattergraph

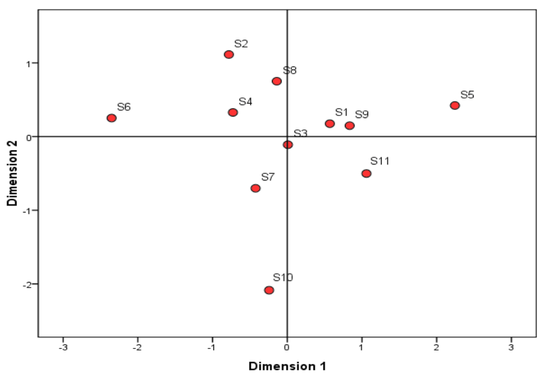
Figure 2. Analysis of dataset A



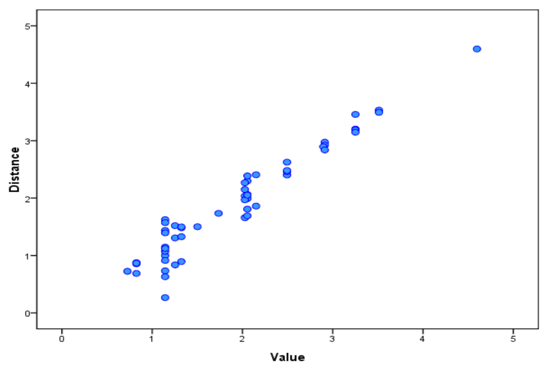
(a) Correlation coefficient



(b) Eigenvalues



(c) Reconstruction in 2 dimensions



(d) Linear fitting scattergraph

Figure 3. Analysis of dataset B

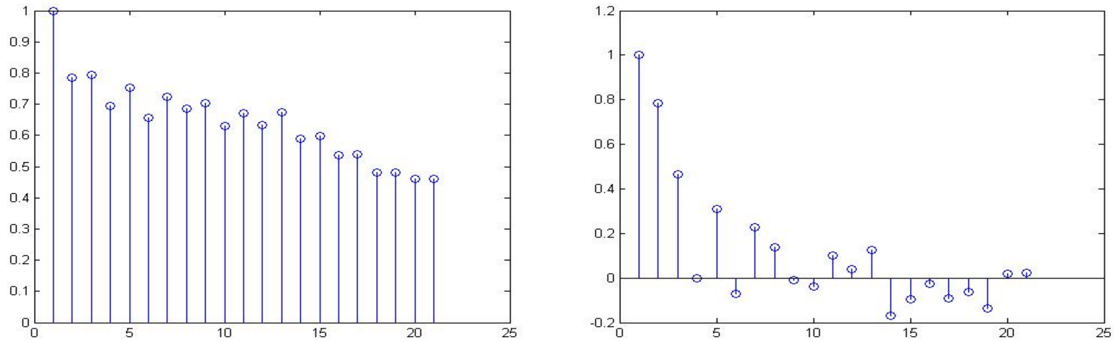
Table 2. Kruskal's Stress and RSQ for the first 15 days in November

Dataset	Stress	RSQ
A	0.09969	0.95444
B	0.11444	0.93146

4.3. Forecasting

In the Dataset A for correlation analysis of road sections, Section 1-5, 7, 8, 11 is a combination. They have strong correlation. The prediction model is set up according to the value of traffic flow and lane occupancy ratio in the section 1.

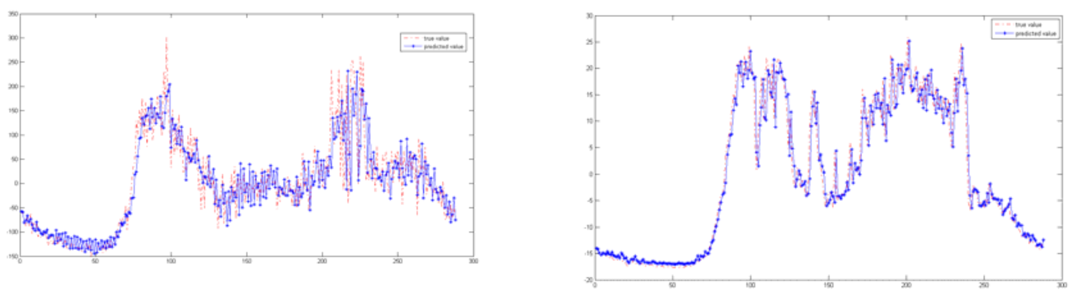
First it needs to analyze the correlation of time series in Section 1. Calculate the self-correlation function and partial self-correlation function of the samples in this series, and the type of model can be estimated by the truncated and trailed characteristics of the two function values. This article selected ARMA model as shown in Figure 4.



(a) Self-correlation (b) Partial self-correlation

Figure 4. Self-correlation and partial self-correlation of Section 1

Then the model can be used to predict the value of traffic flow and lane occupancy ratio in the Section 3 and 8, shown asFigure 5 and 6.

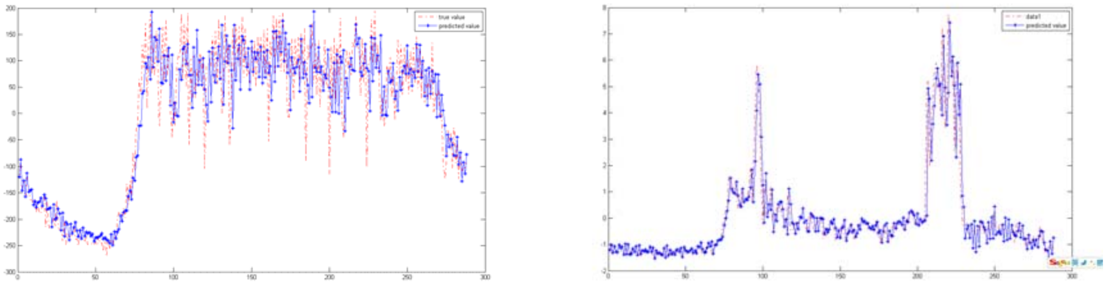


(a) Traffic flow

(b) Lane occupancy ratio

Figure 5. Prediction of traffic flow and lane occupancy ratio in the Section 3

It can be seen that the prediction value in the Section 3 and Section 8 is quite well accorded with the actual values. The results suggest that prediction model set up by Section 1 has a good practicability to the same set of road sections.



(a) Traffic flow (b) Lane occupancy ratio

Figure 6. Prediction of traffic flow and lane occupancy ratio in the Section 8

5. Conclusion

In order to ease the burden on computation of modeling for traffic forecasting, sections correlation between space and time in road networks should be researched. This paper discusses the implementation process of quantitative correlation analysis with the traffic microwave data based on metric multidimensional scaling theory. According to the testing and verifying experiments, the proposed method is effective. Grouping results of 11 sections are acceptable. Then, experiments on traffic forecasting in one group with the same ARMA model are carried out. The result shows that it is very useful in practice.

References

[1] Ma Yilin. Short-term Traffic State Forecasting of Road Network Based on Nonparametric Regression. Beijing: Beijing Jiaotong University; 2008.
[2] Joe Whittaker, Simon Garside, et al. Tracking and Predicting a Network Traffic Process. *International Journal of Forecasting*. 1997; 13(1): 51-61.
[3] Billy M Williams. Multivariate Vehicular Traffic Flow Prediction-Evaluation of ARIMAX Modeling. *Transportation Research Record*. 2001; 1776: 194-200.
[4] Anthony Stathopoulos, Matthew G, et al. Temporal and Spatial Variations of Real-time Traffic Data in Urban Areas. *Transportation Research Record*. 2001; 1768: 135-140.
[5] Yue Hou, Yuemei Mai. Chaotic Prediction for Traffic Flow of Improved BP Neural Network. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(3): 1683-1690.
[6] Ren Shuyun. An Analysis of Traffic State Estimation Based on Particle Filter. Beijing: Beijing Jiaotong University; 2010.
[7] Hegyi A, Mihaylova L, et al. *Parallelized Particle Filtering for Freeway Traffic State Tracking*. Proceedings of the European Control Conference. 2007; 2442-2449.
[8] Zhenyu Na, Bao Peng, Liming Chen. AQM Algorithm with Adaptive Reference Queue Threshold for Communication Networks. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 1062-1072.
[9] Liu Yi, Bao Depei, et al. Research of New Similarity Measure Method on Time Series Data. *Application Research of Computers*. 2007; 24(5): 112-114.
[10] Mark L Davison. Multidimensional Scaling, New York: Wiley, USA. 1983.