Traffic Flow Analysis Based on Singular Vector Decomposition

Waseda University Student Member LIU Xingwei
Waseda University Regular Member Kuniaki SASAKI

1. INTRODUCTION

Singular value decomposition (SVD) was discovered by Beltrami in 1873 and Jordan in 1874 and used as a computational tool in 1960's. SVD has been more than a hundred years. SVD is one of useful data reduction tools where you have high dimensional data. SVD will help us reduce these data into key features which are necessary for data analyzing and describing. It has been applied successfully in many fields, such as computer vision, latent semantic indexing (LST), signal de-nosing, fault diagnosis and analysis of computer network flows. In recent years, the type and dimension of traffic data have been enriched with the diversified data collection ways. What's more, the traffic situation becomes more complicated, more and more traffic data needs to be analyzed. Considering the SVD has been achieved a good results on feature extraction and dimension reduction, we take into account that these properties can be applied in the traffic demand analysis. The following content aims to explain the principle of SVD and explore the possibility about how to understand the variation of traffic demand based on SVD.

2. THE PRINCIPLE OF SVD

SVD is a method of decomposing a matrix into three other matrices. It can be written as equation (1).

$$A = USV^T \tag{1}$$

A is an $n \times m$ matrix; U is an $n \times m$ orthogonal matrix; S is an $m \times m$ diagonal matrix; V is an $m \times m$ orthogonal matrix. There is other form to show the decomposition of SVD. For high dimension, combining vectors with the same subscripts, the equation (1) can be expressed generically as:

$$A = \begin{bmatrix} \\ A_1 & A_2 & \cdots & \cdots & A_m \\ \\ \end{bmatrix} = U\sum V^T = \begin{bmatrix} u_{11} & u_{12} & \cdots & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & \cdots & u_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \\ u_{n1} & u_{n2} & \cdots & \cdots & u_{nm} \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ v_{m1} & v_{m2} & \cdots & v_{mm} \end{bmatrix}^T$$

That is:

$$A = U\sum V^T = \sum_{m=1}^{r} U_m \sigma_m V_m^T \tag{2}$$

Where:

U is left singular vectors;

$\sum$ is the matrix of singular values;

V is right singular vectors;

U is the left singular vectors. In matrix A, A_m is the collection of column vectors of matrix A and its rank is r. $U_m = [u_{1m} \ \cdots \ u_{nm}]^T$ is the collection of column vectors of matrix U. $V_m = [v_{1m} \ \cdots \ u_{mm}]^T$ is the collection of column vectors of matrix V. U_m has the same shape as a column of A_m. If A_m is a n by one vector then the columns of U, U_m , are n by one vectors. Every column in U is the "eigen-features" which are hierarchically arranged. In general, compared the columns in U, previous column in U has stronger ability to extract the features in A.

V is the right singular vectors. In equation (2), it is the V transpose. Every column in V represents the "eigen-time series". When A is the flow data including the time factor, "eigen-features" exist at each "eigen-time series".

V_n^T would be the mixture of columns in U and its total value is A_n which can verified by matrix multiplication as well.

$\sum$ is the matrix of singular values. Each value in sigma is a singular value of A corresponding to singular vector U and

Keywords: Traffic demand, Singular vector decomposition, traffic flow analysis

Contact address: 3-4-1 Okubo, Shinju-ku, Tokyo, 169-8555, Japan, Tel.+81-(0)3-5286-3398

V. For $\sum$, the non-negative diagonal matrix, ordered in decreasing magnitude which means ordered by the importance. It could be wrote as:

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots\cdots \sigma_m \geq 0$$

It is easily to know from equation (2), the first column of U, u_1, corresponding to σ_1; u_2 corresponding to σ_2 and so forth which it represents first columns are more important than the second columns and so on when they describe the data. The matrix V is the identical to the matrix U, v_1, corresponding to σ_1; v_2 corresponding to σ_2 and so forth.

U, V and $\sum$ are given the vary physical meaning according to the meaning of original data matrix A. Among them, U describes more on the meaning of columns in original data matrix and V is for the row of original matrix correspondingly.

3. THE APPLICATION ON TRAFFIC FLOW ANALYSIS

In traffic field, no matter what kind of data format, traffic flow value can be obtained. It can be organized into matrix form. For example, the per hour traffic flow value in one month can be written in the 30×24 which column represents the 30 days, line represents the 24 hours. Each value in the matrix is the traffic flow value of specific moment. We can do SVD based on the temporal matrix of traffic flow.

According to the principle of SVD, the temporal matrix could be decomposed into three matrixes. The 30×24 matrix transfers to 30×24 matrix U, 24×24 matrix $\sum$ and 24×24 matrix V. Combining the principle of SVD mentioned above, we can learn from this application is the left singular vectors U shows the characteristics of traffic flow each day, and right singular vectors V show the characteristic of hourly traffic flow. The matrix of singular value $\sum$ contains the sigma values which can be regarded as the weights. It represents the importance, the bigger the sigma value, the more important it is.

Here, due the limitation is 24 in 30*24 matrix, there are total 24 sigma values in $\sum$. The sigma values are arranged in order from largest to smallest, therefore, the dominant feature is aware by the amount of value. It gives us the points of focus in the next analysis. Besides, the distribution of sigma values can reflect the changing of traffic demand. When there is a significant shift in the share of significant features, this can be considered as a change in the main factor influencing traffic demand. It is a convenient indicator to judge whether the traffic demand is varying or not. In addition, the existing traffic flow value is the result for daily travel and it is tough to know what affects their traffic demands. However, the decomposition results from SVD provide the tool to understand these reasons in multi dimensions, including day to day characters U, hourly to hourly characters V and feature weights $\sum$. We can do statistics on them separately and it allows us to obtain the influence factors as well as the aggregation.

Furthermore, in order to make it clear, we can do classification in response to 24 features and these temporal characters. Due to the general regulation of traffic demand, three traffic demands can be classified, they are: regular traffic demand, trigger-based traffic demand and random one. Traffic demand classification is not set in stone, it should be adjusted according to the research object. Besides, we can explore the traffic demand in different periods, such as understand the variation before, during and after the emergency declaration, knowing the changing before, during and after the golden week holiday.

REFERENCE

Akritas A G, Malaschonok G I. Applications of singular-value decomposition (SVD)[J]. Mathematics and computers in simulation, 2004, 67(1-2): 15-31

Chu Z, Cheng L, Chen H. A review of activity-based travel demand modeling[J]. CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient, 2012: 48-59.

Lakhina, A., Papagiannaki, K., Crovella, M., et al.: 'Structural analysis of network traffic flows', ACM Sigmetrics Perform. Eval. Rev., 2004, 32, (1), pp. 61–72.