International Journal of Information Technology and Knowledge Management
July-December 2011, Volume 4, No. 2, pp. 391-394

MODELING THE DATA STREAM OF A VEHICLES NETWORK AND ITS COMPRESSING BY THE TSVD

S. EL HADAJ[1], S. RAGHAY[2], T. NOEL[3]

Despite of the evolution of Data Stream Management Systems (DSMS), storage and processing of data flow persists as a major problem, especially if it comes from a network of mobile, including a set of vehicles.

One solution proposed for such problems is to summarize the data stream into portions called short summaries. Several selection techniques segmentation of these streams has been proposed based on deterministic or probabilistic algorithms.

Our contribution records the proposal of a new method of data modeling. These data emitted by all vehicles using a section of road. The set of data is represented in the model by a matrix which will be compressed by the Truncated Singular Values Decomposition method (TSVD).

Keywords: DSMS, V2I, Modeling, SVD, TSVD, Data Compression.

1. Introduction

The analysis of information to aid decision making in any sector, has become unavoidable. Thus any information system if it does not generate data, it receives them in large quantities.

Several systems, for the proper functioning of an organism, have applications that use data streams in large volumes and treat them on the fly. The first idea may seem logical is that the use of the structures of classical databases and subsequently (DBMS). However, these structures use queries that have ephemeral effects as if they are submitted. This does not address the need for systems that require answers to queries in real time [5]. The immediate management systems of data stream are needed. We are talking about data stream management systems DSMS [7].

These systems can be satisfied with the treatment of a party reduced and representative of this stream, letting everything else, since it is impossible to store the entire data stream or otherwise deal in real time.

In recent years, several algorithms, classified into probabilistic and deterministic [1, 2, 3], whether these segments have emerged.

A field application of DSMS is the mobile world, including networks of vehicles that produce, process and consume information for security purposes, saving the lives of passengers, reducing CO_2 emissions and time of runs,...[4].

[1]ENCG, Marrakech
[2]FSTG, Marrakech
[3]LSIIT, Université de Strasbourg
Email: [1]elhadajs@yahoo.fr, [2]said.raghay@hotmail. [3]com, noel@unistra.fr

View a number of applications for intelligent transportation systems require vehicle-infrastructure communications, we are interested in this paper by the stock of the data flow generated by all vehicles on a stretch of the road and sent to an infrastructure dedicated to the collection of information and other services.

Our approach to managing data stream from the combination of vehicles is based on modeling of geographic space by a two-dimensional matrix to exploit the principles of matrix calculation and use of data compression techniques. We consider here the TSVD which demonstrated a tolerable fidelity of preservation of the initial information for approximate queries [5].

This paper is organized as follows: Section 2 details the problem. Section 3 describes the TSVD method. Section 4 describes the modeling considered to solve the problem. In Section 5 we present results of numerical simulation and a conclusion in Section 6.

2. Problematic

Today the collection of information about traffic conditions on the road, generated by vehicles or captured by their devices, has become of great importance. Especially to keep a history on the state of the road, increase user safety, maintain the flow of traffic, correct the bad behavior of road users, alert on accident/incident, weather, work in progress (sparse matrices), the caps and limitations on intelligent speed. This is a communication Vehicle-Infrastructure-Vehicle (see Figure 3). Given the large number of vehicles that can travel on a segment, the data flow is very large. However, the treatment and storage tools require large storage capacity and computational very expensive. One solution, used by DSMS, is to summarize the data flow in a representative segment [7-9].

The problem is, therefore, what parts of the data stream that must be kept in order to have a certain image, if not exact, very close to the source stream generated by these vehicles?

We propose in this paper to introduce a method combining modeling of the data stream from a network of vehicles on a portion of the road and the compression of information by the TSVD method based on Vehicle-Infrastructure communication.

The Truncated Singular Value Decomposition (TSVD)

The decomposition of matrices is one of the most important axes of linear algebra. In particular the singular value decomposition SVD is used in many applications tolerate the loss of information without much influence the quality or nature of the objects represented initially by the transformed matrix, such as image processing and meteorology.

We summarize the principle of the method with the following:

$A = (a_{ij})_{m,\,n}$ is a rectangular matrix of rank r. With $r \le \min(m, n)$.

The SVD decomposes A into a product of two orthogonal matrices U and V and a diagonal matrix Σ.

$$A = U\Sigma V^T$$

Since the SVD decomposes the matrix singular value, we first describe the formal definition of singular values, then the SVD theorem.

Definition: The square roots of eigenvalues of matrix $A^T A$ are called singular values of A.

Theorem: Let A be a matrix of type (m, n) of rank $r (r \le \min(m, n))$. Then there exists a matrix U of type (m, n) matrix V(n, n) and Σ a diagonal matrix such that $A = U\Sigma V^T$,

With $$\Sigma_{ij} = \begin{cases} \sigma_i & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

With the σ_i $(1 \le i \le r)$ are the singular values of A and and $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_r > 0$ and $\sigma_{r+1} = \ldots \sigma_n = 0$.

Thus the matrix A can be reconstructed by the formula:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Depending on the desired certainty tolerated, one can always find a matrix B approaching A written as:

$$B = \sum_{i=1}^{k} \sigma_i u_i v_i^T, \text{ where } 1 \le k \le r.$$

So that instead of saving all the singular values we can reconstruct the matrix A with the first k values. This is called the TSVD method.

3. Modeling

To exploit the power of matrix algebra and more specifically that of the Singular Value Decomposition, as mentioned above, we decompose a stretch of road in a set of M sectors. A sector can, at a given time, be borrowed by a set of vehicles we represent by $(V_{i,j\ 1\le i\le n, 1\le j\le k})$(see Fig.1)

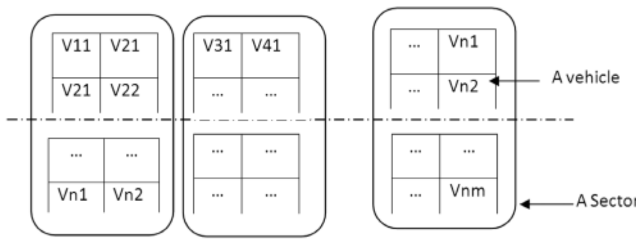


Fig.1: Decomposition of the Section of the Road in Sectors.

A sector is also represented by a matrix of size (n, k) with nxk is the number of vehicles found there at any given time. One sector, in the remainder of our paper, corresponds to the range of a signal exchanged with an infrastructure.

The matrix representing this portion of the road is filled with data carried by the vehicles therein. We can, according to need to use the application, require the memorization of the positions of vehicles on the portion in this data structure. An empty position is symbolized by a zero component. The direction can be symbolized by the negative sign.

In summary, at time T an instance of the road is a matrix Tr.

$$Tr = \begin{pmatrix} P11 & \cdots & P1k \\ \vdots & \ddots & \vdots \\ Pn1 & \cdots & Pnk \end{pmatrix} \quad (1)$$

A data packet, P_{ij}, produced or carried by a vehicle, V_{ij}, is a record of heterogeneous information may be the registration number of the vehicle and also as an identifier of the packet, the physical condition of the road, speed, position, traffic conditions or any other useful information for immediate or later use.

Thus the matrix Tr can be seen as a combination of a set of matrices Tr_{option} which is a projection of matrix Tr in the coefficients relating to a certain option.

Example: $Tr_{Weather}$ = the matrix formed by selection coefficients of the weather in Tr. That we will note

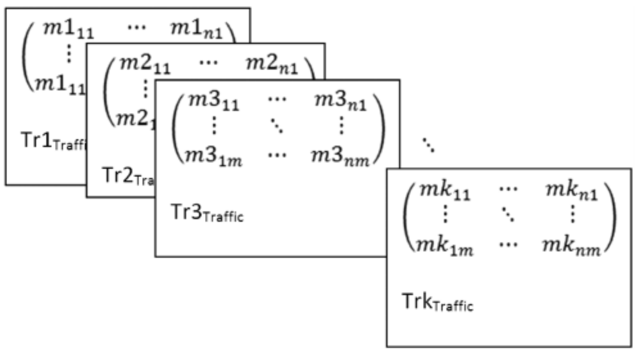
$$Tr_{Weather} = \begin{pmatrix} m_{11} & \cdots & m_{n1} \\ \vdots & \ddots & \vdots \\ m_{1m} & \cdots & m_{nk} \end{pmatrix} \quad (2)$$

Where m_{ij} is the weather coefficient in the packet Pij.

$Tr_{Weather}$ = the matrix formed by selecting coefficients of the traffic in Tr.

In addition to our main objective, which is the application of TSVD to the data matrix created above, we can, initially, operate directly in the original matrix by plotting a graph of variation values of the corresponding matrix weather, traffic, the emission of CO_2, or any other projection matrix constituting the result of original matrix model representing the stretch.

If, for example, we want changes in traffic volume during a time segment, a reasonable number of global arrays at different this time segment, may



4. Numerical Results

(a) Immediate Exploitation of Information

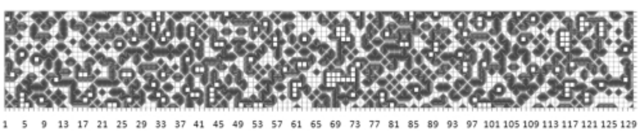


Fig. 2: Dark Areas of the Graph Represent the Parts of the Road Occupied by Vehicles and Clear Areas those Which are not Borrowed.

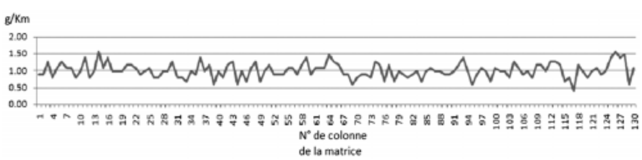


Fig. 3: The Rate of $C0_2$, Corresponding to the State of the Road above.

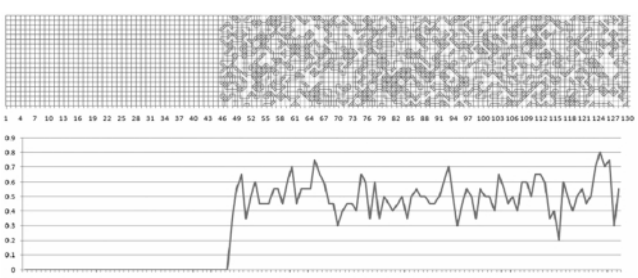


Fig. 4: The Area Not Borrowed, in the Diagram Above, Does Not Correspond to Instantaneous Release of CO_2.

(b) Subsequent Operation

As mentioned above, the major problem encountered in the management of data stream for later use is storage of huge volumes of data collected by road segment for long periods. In our approach this information will be stored in rectangular matrices which we decompose by the singular value decomposition method TSVD to save only approximate matrix of very small size which optimizes the storage of information.

(c) Experiments

In this paper we consider a rectangular (n, m) matrix generated by the formula

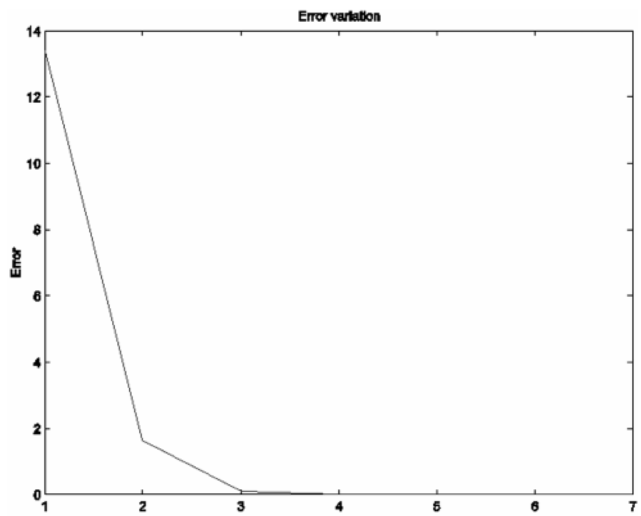
$$A(i, j) = \exp(x(i)*\cos(t(j)))$$

where $x(i) \in]0, \pi[$ and $t(j) \in]0, \pi/2[$ for all i and j.

For n = 60 and m = 1000, we have computed the approximate matrix by using the truncated singular value decomposition (TSVD). In fig 5 we present the variation of the error estimation defined by:

$$err = (\Sigma_{i,j}(a_{ij} - b_{ij})^2)^{(1/2)}$$

This figure show that the error became small than 10^(–7) after just k = 7 iterations. Then the TSVD series can be truncated at k = 7. Thus, instead of stocking all rectangular (60, 1000) matrix, we can stock just the first seven right eigenvectors with the first seven left eigenvector and the first seven singular values. Thus, with this approach we win until 88% of storage space provided for initial flows.



5. Conclusion

In this paper, we propose a data flows modeling using TSVD to approximate the data with the first impotent singular values and the first relative eigenvectors. We adapt the singular value decomposition (SVD) as a dominant compressor. Data flows are decomposed into a column-orthogonal matrix, an orthogonal matrix, and a diagonal matrix that contains singular values. With respect to singular values, some column vectors are dropped.

We implemented our approach for static data and the results show that we can win more than 80% of space storage.

1. References/Bibliographie

[1] Chikly R. (2009), "Résumé De Flux De Données Distributes," Rapport de thèse, à l'ENST de ParisTech.

[2] Csernel B. Résumé Généra liste De Flux De Données, "Rapport de thèse", 2008, ENST Paris.

[3] S. El HADAJ, S. RAGHAY, T. NOEL, "A Vector Algorithm For Managing Data Stream", 3rd IEEE International Conference on Web and Information Technologies (ICWIT'10), Marrakech, Morocco - June 2010.

[4] S. RAGHAY, S. EL HADAJ, T. NOEL, "A Genetic Algorithm for Management Data Stream in Vanet", International Journal of Information Technology and Knowledge Management. IJIT&KM 2010.

[5] Jun-Ki Min and al. "Approximation of MultI-measured Data Using SVD", Information Sciences 179 (2009) 2833–2850

[6] Flip Korn and al. "Efficiently Supporting Ad hoc Queries in Large Datasets of Time Sequences ", International Conference on Management of Data Proceedings of the 1997 ACM SIGMOD, Arizona, United States Pages: 289 - 300

[7] B. Badcock, S. Babu, M. Datar, R. Motwani, and J. Widom. "Models and Issues in Data Stream Systems", "In Symposium on Principles of Database Systems", pages 1–16. ACM SIGACT-SIGMOD, 2002.

[8] V. Goebel and T. Plagemann. "Dsms : Applications, Conceps and Systems", In Proceedings of the 2004 MIPS Conference, 2004.

[9] L. Golab and M.T. Ozsu. "Data stream Management Issues- a Survey", Technical Report cs 2003-08, University of Waterloo, Waterloo, Canada, April 2003.