Matrix Decomposition Algorithms for Feature Extraction

Derrick Pisani

Department of Computer Science and AI,
University of Malta

Abstract. Clinical decision support software is a delicate system which, can potentially be the physician's closest friend. The aim of such systems is to be able to cleverly recommend a list of treatment options which closely matches the patient. We envisage a system which learns from experts without ever needing to ask them for feedback, and thus one which learns from past patient encounters. The system needs to be adaptive as well as dynamic, since all patients are different even if they may exhibit very similar symptoms. This paper proposes using matrices to capture such data, and algorithms using Singular Value Decomposition to predict treatments.

1 Introduction

The aim of the project is that of achieving the level of machine intelligence required for delivering clinical decision support[6]. The traditional meaning of learning is the gaining of knowledge or information about an area, and thus the acquiring of an understanding of the topic[7]. An understanding is the product of the learning process, and it is very often an interpretation of the knowledge gained. As a result, in complex areas such as those found in the medical field, learning does not lead to all experts, the physicians, to have a common understanding. An analysis of the actions of several physicians, however, reveals trends or patterns of prescriptions to tests and medications given to patients. Although this is not equivalent to an understanding of the area, which is satisfactory to substitute the expert, such knowledge can be used as training data for clinical decision support systems. This paper introduces techniques which use singular value decomposition to extract this training data dynamically.

2 Basic Definitions

Observations of a working system or natural phenomenon may be described by a set of simultaneous equations. These, in turn, may be represented in matrix notation.

In the case of a system with m simultaneous equations in n variables, we may write

$$\mathbf{A}x = b \tag{1}$$

where x is an unknown solution column vector and $\mathbf{A}$ is an $(m \times n)$ co-efficient matrix. Thus:

$$x = \mathbf{A^{-1}}b$$

For a matrix $\mathbf{A}$, where $b = 0$, the system is said to be *homogenous*. The solution here is apparently $x = 0$, but when the $det\mathbf{A} = 0$, the inverse $\mathbf{A}^{-1}$ does not exist, and thus the solution for x is non-trivial. This is also always the case when $m > n$. Matrix $\mathbf{A}$ is then said to be *singular*.

The *rank* of a matrix is the maximum number of linearly independent rows or columns. When the $det\mathbf{A} = 0$, the rows of $\mathbf{A}$ are linearly dependent and thus $rank(\mathbf{A}) < n$. This is termed *rank-deficiency*, and leads to the system potentially having infinitely many solutions. Solving a system where the coefficient matrix is rank-deficient is equivalent to solving m simultaneous equations in n unknowns where $m < n$, and thus traditionally considered intractable.

Let us finally consider the same system of simultaneous equations described by equation (1) when:

$$b = \lambda x$$

λ is a scalar and its values are called *eigenvalues*, whilst the corresponding solutions are called *eigenvectors*. Using the above substitution for b in $\mathbf{A}x = b$:

$$\begin{aligned} \mathbf{A}x &= \lambda x \\ (\mathbf{A} - \lambda I)x &= 0 \end{aligned}$$

Therefore, the *characteristic equation* for systems with non-trivial solutions is:

$$det(\mathbf{A} - \lambda I) = 0 \tag{2}$$

Its solution gives the eigenvalues of $\mathbf{A}$ and, in turn, the solutions of x.

3 Singular Value Decomposition

Most systems of simultaneous equations may be solved using algebraic manipulations involving elementary row operations. In those cases where Gaussian Elimination or Triangular (LU) Decomposition do not succeed to give satisfactory results, we can use *Singular Value Decomposition (SVD)* to diagnose and possibly solve the problem[8].

SVD techniques deal with singular matrices, or ones which are very close to being singular. They are an extension of eigen decomposition to suit non-square matrices. Any matrix may be decomposed into a set of characteristic eigenvector pairs called the component factors, and their associated eigenvalues called the singular values [3].

The SVD equation for an $(m \times n)$ singular matrix $\mathbf{A}$ is:

$$\mathbf{A} = \mathbf{U \Sigma V^T} \tag{3}$$

where $\mathbf{U}$ is an $(m \times m)$ orthogonal matrix, $\mathbf{V}$ is an $(n \times n)$ orthogonal matrix and $\mathbf{\Sigma}$ is an $(m \times n)$ diagonal matrix containing the singular values of $\mathbf{A}$ arranged in descreasing order of magnitude.

A vector in an orthogonal matrix, which can be expressed as a linear combination of the other vectors. The vectors in this space are thus, also mutually independent, and thus a solution for x may be now calculated.

The equation for Singular Value Decomposition needs to be further explained by showing the significance of the eigenvectors and the singular values.

Each of the orthogonal matrices is achieved by multiplying the original $(m \times n)$ data matrix $\mathbf{A}$ by its transpose, once as $\mathbf{AA^T}$ to get $\mathbf{U}$, and once as $\mathbf{A^TA}$ to get $\mathbf{V^T}$. The data is then available in square matrices, on which eigen decomposition may be applied. This involves the repeated multiplication of a matrix by itself, forcing its strongest feature to predominate. In SVD, this is done on both the orthogonal matrices. The pairs of eigenvectors are the rows in **U** and the columns in **V**. These are sorted in order of their strength; the respective singular value acts as the required numerical indicator as further explained below. [3]

Let us first explain the relationship between the singular values and the co-efficient matrix's eigenvalues:

$$\mathbf{A^T A} = (\mathbf{U\Sigma V^T})^T(\mathbf{U\Sigma V^T})$$
$$\mathbf{A^T A} = \mathbf{V\Sigma^T U^T U\Sigma V^T}$$
$$\mathbf{A^T A} = \mathbf{V\Sigma^T\Sigma V^T}$$
$$\mathbf{A^T A} = \mathbf{V\Sigma^2 V^T}$$

$\mathbf{\Sigma^2}$ contains the eigenvalues of $\mathbf{A^T A}$. Similarly, it may be shown that they are also the eigenvalues of $\mathbf{AA^T}$. They may be arranged in decreasing order of magnitude:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ... \geq \lambda_n \geq 0$$

Singular values of a matrix **A** are defined as the square root of the corresponding eigenvalues of the matrix $\mathbf{A^T A}$:

$$\sigma_{\mathbf{j}} = \sqrt{\lambda_j}$$

As with the eigenvalues, they are sorted according to their magnitude, and the corresponding eigenvectors in the orthogonal matrices **U** and **V** follow the same order. The SVD equation (3) takes the form:

$$\mathbf{A} = \begin{pmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1m} \\ u_{21} & u_{22} & u_{23} & \dots & u_{2m} \\ u_{31} & u_{32} & u_{33} & \dots & u_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & u_{m3} & \dots & u_{mm} \end{pmatrix} \begin{pmatrix} \sigma_{11} & 0 & 0 & \dots & 0 \\ 0 & \sigma_{22} & 0 & \dots & 0 \\ 0 & 0 & \sigma_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{nn} \end{pmatrix} \begin{pmatrix} v_{11} & v_{21} & v_{31} & \dots & u_{n1} \\ v_{12} & v_{22} & v_{32} & \dots & u_{n2} \\ v_{13} & v_{23} & v_{33} & \dots & u_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{1n} & v_{2n} & v_{3n} & \dots & u_{nn} \end{pmatrix}$$

An alternative way to what was initially described, exists for defining the size of the matrices: $\mathbf{U}_{m\times n}$, $\mathbf{V}_{n\times n}$ and thus $\mathbf{\Sigma}_{n\times n}$ [1]. This is due to the fact that when $m > n$, then n rows are sufficient to solve for x, and the main diagonal of Σ still crosses the top left $(n \times n)$ subsection of the matrix.

Some singular values are equal or very close to zero. A singular value σ_j describes the importance of u_j and v_j, and when its value approaches zero it indicates that these associated vectors are less significant.

A proof that SVD can be used to decompose singular matrices may be found in Hourigan and McIndoo's work [1].

[1] Hourigan J S, McIndoo L V. The Singular Value Decomposition. http://online.redwoods.cc.ca.us/instruct/darnold/laproj/Fall98/JodLynn/report2.pdf

4 Applications in Clinical Decision Support

4.1 The Dataset

The dataset being utilised for this project originates from collections of anonymous patient-encounters from various acute care hospitals during the year 2001. The hospitals selected are all medium to large ones and geographically sparse over a very large country. As shown in Table 1, they have accommodated thousands of patients.

Table 1. Hospitals in Dataset

Hospital	**Size in no of beds**	**Patient Encounters**
A	200-399	80,000
B	200-399	55,000
C	400+	175,000
D	400+	90,000

For the scope of this study only inpatients who have spent at least one night in hospital are being considered. The sex distribution is about 4 males to 6 females, and there is a very even and broad distribution of ages.

Each patient is represented by a unique identifier that masks the person's identity, but real demographic and medical information is available. This includes sex, age, diagnosis in ICD-9CM format, procedure information in ICD-9CM and CPT-4 format, and hospital billing information. Since hospital-specific codes are used to describe the latter, when patient-encounters from more than one hospital are used, the data needs to be standardised in advance.

4.2 Use of SVD Techniques

Singular Value Decomposition is an excellent unsupervised learning tool to process source data in order to find clusters, reduce dimensionality, and find latent variables. This means that field or experiment data can be converted into a form which is free of noise or redundancy and that it can be better organised to reveal hidden information.

Although a lot of work on information retrieval (IR) has been done using SVD, most of it is in the areas of genetic patterns and semantic analysis of text and hypertext documents. Clinical decision support software affects human lives, and thus, since SVD techniques are still being tested with less delicate applications, no current attempts at using it to help the physician's decision making process are known to the author. The many current efforts to harness SVD techniques in the field of IR have been instrumental for this project, and good examples are provided by Wall's work [9–11] and others such as [2, 4].

The various prescriptions (orders) received by a hypothetical number of patients exhibiting a certain condition can be visualised in Fig. 1. The orders are however not mutually exclusive, and no information is given regarding which orders were prescribed together to which patients. SVD is not a straightforward statistical technique, and manages to mine such information.

In the following sections we will briefly describe the benefits of using Singular Value Decomposition to process the source data originating from patient encounters.

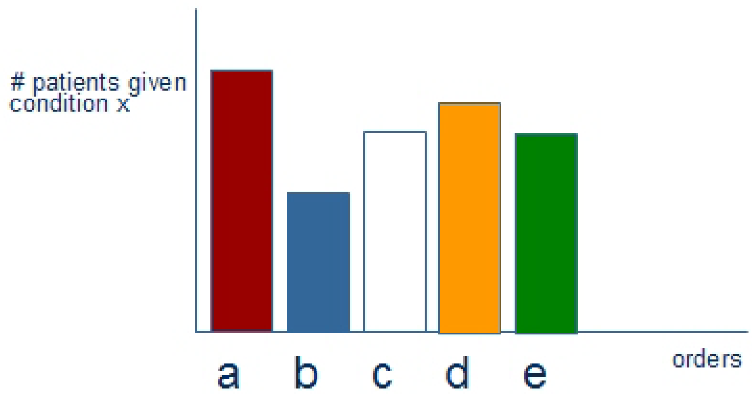

Fig. 1. Different Orders for a Particular Condition

Redundancy Reduction: The diagram in Fig. 2 shows the process of converting field data into matrix format. The chosen rows are those of patients who fit the demography and symptom profile chosen by the physician.

Since we are using field data, and not one specifically collected for this project, the matrix containing the patient-encounter data contains a degree of noise and redundancy. The sorted singular values in a Singular Value Decomposition help represent the original matrix in less rows, by pushing the less significant rows to the bottom of the orthogonal matrices. The matrix may thus be smoothed by eliminating those vectors whose respective singular value is equal to, or approaches zero. As explained in section 3, these are the vectors which least impact on the original matrix.

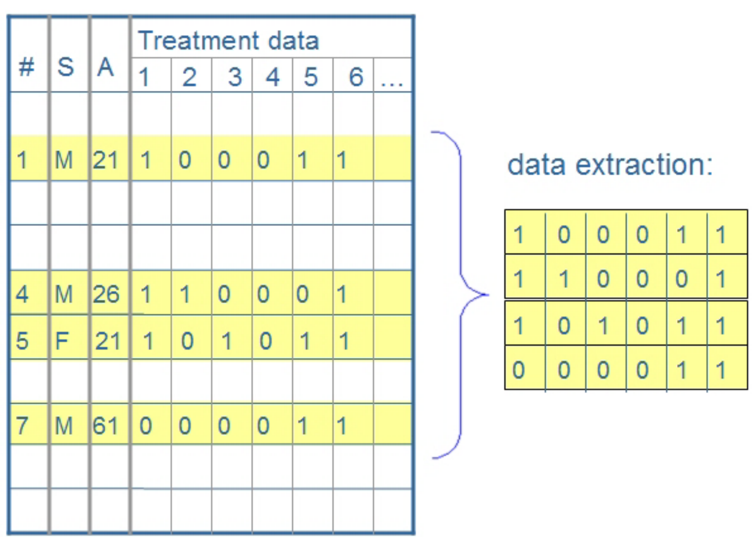

Fig. 2. Source Data converted to Matrix-format

Exposing Components: As already explained in Section 3, Singular Value Decomposition results in sorted eigenvectors (components). These are very useful in analysing patient encounters and revealing the most predominant components. These are representative of the procedures which most significantly describe the typical patient's treatment for the chosen patient profile.

We believe that the clusters formed when the source data is projected in the dimensions of the first two components is indicative of their constituents. An example is seen in Fig 3.

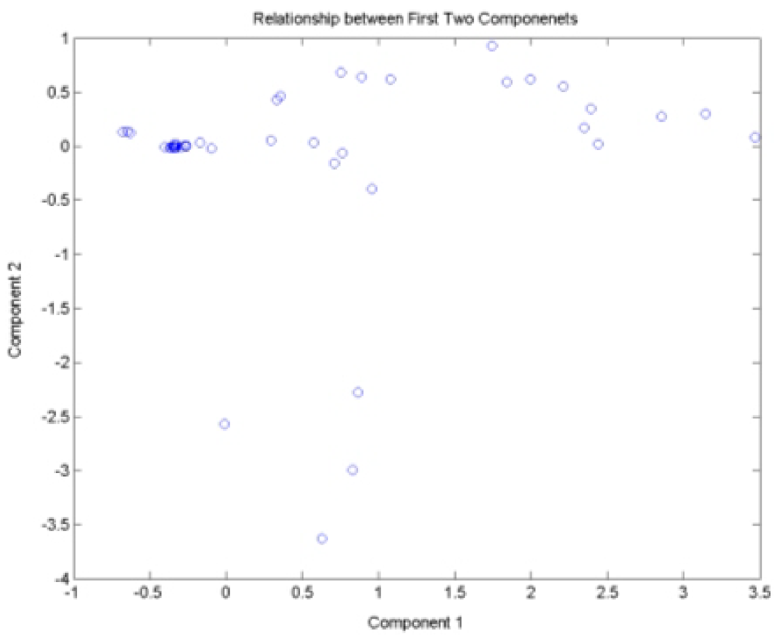

Fig. 3. Clustering when Source Data is Projected in two Dimensions

Understanding Latent Variables: Latent variables are hidden sources in the original data; they explain the common traits in patient encounters, but cannot be measured directly. For the observed data represented by matrix $\mathbf{A}$, in n variables, the latent variables may be expressed in l dimensions where $l << n$. If we assume that the initial n variables capture the whole domain, the product of SVD will be a set of variables which includes the latent ones. What may initially be evident as a prescription to a set of tests and medications for particular symptoms may thus have been due to a combination of latent variables including demography, diagnosis, or the *'school of thought'* of the physician or expert. [5, 4, 2]

Following SVD, this hidden information will not become transparent, but it will be captured by the system as shown in Fig. 4. The 'Controls' act for the Latent variables and for every patient where the same conditions apply, the relevant ones are toggled on. In this sense, every 'Control' is a description of the effects of a set of Latent Variables, and certain medical procedures go ordered when a control is toggled on.

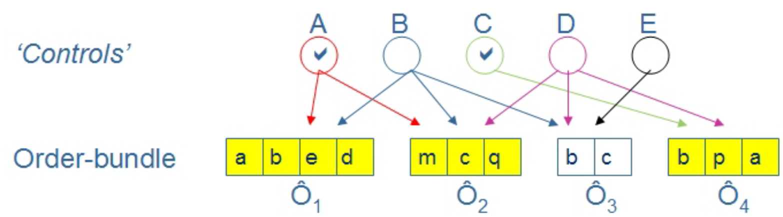

Fig. 4. Hidden sources may be represented as 'Controls'

5 Conclusion

SVD techniques will be used to predict a patient's treatment based on all the information known about the patient, without having to assign static weights to a particular characteristic. This means that the importance of age or a particular symptom, for instance, may be very significant in one encounter and disregarded in another as the system's decision support is based on a non-trivial dynamic distance measure between the current patient profile and past cases.

References

1. Glover D M, Jenkins W J, Doney S C. Modeling, Data Analysis and Numerical Techniques for Geochemistry (Notes). Massachusets, USA: Woods Hole Oceonographic Institution; 1998,2000,2002
2. Gong Y, Liu X. Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis. In: SIGIR'01; 09 December 2001. San Jose, USA: NEC USA, C & C Research Laboratories; 2001
3. Gorrell G. Word Co-occurrence and Singular Value Decomposition. Sweden: Institutionen for datavetenskap, Linkopings universitet. http://w3.msi.vxu.se/ nivre/teaching/gslt/mlpapers/gorrell.ps
4. Jiangsheng Y. Singular Value Decomposition With Applications to IR and Text Clustering. Beijing, China: School of Electronics Engineering and Computer Science, Peking University; 2003. p12-14
5. Landauer T K, Foltz P W, Laham D. An Introduction to Latent Sematic indexing. Discourse Processes, 25, p259-284; 1998
6. Pisani D. Achieving Intelligent Clinical Decision Support through Orderset Management. In: Pace G. and Cordina J., editors. CSAW 2003; 10 July 2003; Malta: Department of Computer science and AI, University of Malta; 2003. p129-132.
7. Porter N editor. Webster's Revised Unabridged Dictionary. 1998 edition (MICRA, Inc.). Springfield, Massachusetts, USA: C. & G. Merriam Co.; 1913.
8. Press W H, Flannery B P, Teukolsky S A, Vetterling W T. Numerical Recipes in C : The Art of Scientific Computing. 2nd ed. Cambridge: Cambridge University Press; 1992
9. Wall M E, Rechsteiner A, Rocha L M. Singular value decomposition and principal component analyis. In: Berrar D P, Dubitzky W, Granzow M., editors. A Practical approach to Microarray Data Analysis; Kluwer: Norwell Massachusetts, USA; 2003. p91-109
10. Wall M E, Dyck P A, Brettin T S. SVDMAN - Singular value decomposition analysis of microarray data. Los Alamos, USA: Bioscience Division, Los Alamos National Laboratory; 2001
11. Wall M E. Validation of Singular Value Decomposition analysis of Global expression Data. Los Alamos, USA: Bioscience Division, Los Alamos National Laboratory