ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Cognitive Systems Research xxx (2009) xxx–xxx

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Book review

Elden, L (2007). Matrix Methods in Data Mining and Pattern Recognition. Philadelphia (USA): Society for Industrial and Applied Mathematics

Action Editor: Stefan Wermter

Renato Cordeiro de Amorim

School of Computer Science and Information Systems, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK

Matrix Methods in Data Mining is a quite recent book and a short but interesting read for those interested in the application of modern matrix methods in pattern recognition and data mining problems. It is stated in the book that this was written primarily for undergraduate students who have previously taken an introductory scientific computing/numerical analysis course and it is surely a good place for such students to look for final year projects ideas or to start research in the subject. Also for those students and lecturers that are interested, the author provides a website where a collection of exercises and computer assignments are available.

The book is meant not to be primarily a textbook in numerical linear algebra but rather an application-oriented introduction to some techniques in modern linear algebra with the emphasis on data mining and pattern recognition. It should be noted also that the book is of very limited size, around 220 pages, so the reader should not expect a full account of the mathematical and numerical aspects of the algorithms used, but an introduction and its use for real life problems.

A very valid thing to point out is that the book provides MATLAB scripts demonstrating how to implement certain algorithms, but it should not be seen as a book of recipes. Its aim is to provide students with a set of tools that may be tried as they are but most likely will need to be modified to be useful for a particular application. That is specially true if the reader intends to develop a solution that runs as fast as possible in MATLAB as the scripts were clearly designed to focus on cleanliness being then as easy to understand as possible, and by consequence they will not necessarily be fast.

In terms of content, the book has 14 chapters divided in three parts. The first part focuses on linear algebra concepts and matrix decomposition emphasising the existence and properties of matrix decompositions rather than on how they are computed. The second part is on applying those techniques to data mining problems and the last is a very short introduction to eigenvalue and singular value algorithms.

Surely most readers will find the second part of the book the most interesting as it is where the author shows real life problems being solved by the mathematical methods shown in the first part.

Chapter 1: *Linear algebra concepts and Matrix decompositions* presents a short introduction of the usages of vectors and matrices in data mining and pattern recognition which are demonstrated with relevant examples that show the validity and usability of the methods. These examples range from handwritten characters recognition to search engines.

Elden also presents other important concepts, such as flop and floating point errors. To the first, the author gives the appropriate emphasis on this being a crude measure of efficiency and computing time which may be misleading in cases. Other important computing terms such as overflow and underflow are also introduced by the author. Here the reader can also find the book notations and conventions.

In Chapter 2: *Vectors and Matrices* although it is expected that the reader has basic notions of linear algebra, some of the concepts are recapitulated which surely furthers the completeness of the work.

E-mail address: renato@dcs.bbk.ac.uk (R.C. de Amorim)

1389-0417/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2009.03.001

Some of the mathematical formulas are also shown in MATLAB code which helps the reader to see how the different ways of calculating something may have an impact on performance.

Although Chapter 3: *Linear System and Least Squares* presents some results without proofs it points out useful sources where the reader could look for those. The Chapter also makes a number of short introductions, of around 2 pages each in topics such as lower and upper triangular matrix (LU) Decomposition, positive definite matrices, the least squares problem, perturbation theory and condition number.

Chapter 4: *Orthogonality* relates to the separation of the most important information from the least important information, which may be noise, and presents the orthogonal matrix properties.

In Chapter 5: *QR Decomposition* the reader can easily see that one of the main themes of the book is decomposition of matrices to compact form by orthogonal transformations being these triangular or diagonal. The chapter presents an introduction to QR decompositions and explains this to be a factorization of a matrix A in a product of an orthogonal matrix and a triangular matrix. This decomposition is also proved and used as a solution to the least squares problem.

In Chapter 6: *Singular Value Decomposition (SVD)* after explaining that QR Decomposition has the drawback that it treats the rows and column of the matrix differently (it gives a basis only for the column space), the reader can find an introduction to SVD which deals with rows and columns in a symmetric fashion, and therefore it supplies more information about the matrix. It is also argued that SVD is very useful in data mining (and other areas) because of its ability to "order" the information contained in the matrix making the "dominant part" visible.

The chapter also introduces the methods of matrix approximation, Principal Component Analysis and demonstrates that the least squares problem can also be solved using SVD.A short discussion about SVD being the most reliable method for determining the rank of matrices when this is deficient is presented. The author points out the comparative expensiveness of SVD which may be a critical issue for some applications, for instance real time ones. Therefore, methods have been developed that approximate the SVD, the so called *complete orthogonal decomposition.*

Chapter 7: *Reduced-Rank Least Squares models* starts by describing two methods for determining matrices of basis vectors presenting real life examples. First, the author describes Truncated SVD: Principal Component Regression which is a method based on the SVD of the data matrix and second a Krylov subspace method in which the right hand side influences the choice of basis.

The author then explains the desirability of a fast decay of the residual as a function of the number of basis vectors for any right hand side and concludes that it is then necessary to let the right hand side influence the choice of vector. The author describes some algorithms that could do this such as Lanczos-Golub-Kahan (LGK), bidiagonalization and projection to talent structures (PLS), which are described with examples.

In Chapter 8: *Tensor decomposition* the reader will find himself out of the realm of linear algebra and now in multilinear algebra, where for simplicity the text is restricted to tensors with three subscripts. The chapter starts by explaining basic tensor concepts and then goes on to describe a tensor SVD which could have been done in different ways. The author has chosen to present a generalization that is analogous to an *approximate principal component analysis*, which is often referred to as higher order SVD. The chapter shows real life examples which are further developed in the second part of the book.

Chapter 9: *Clustering and Nonnegative matrix factorization* starts by explaining what clustering is, then it describes probably the most well known clustering algorithm: *k*-means providing examples, just like in the other chapters the reader should not expect a full account of the algorithm but a rather concise description. The author then discusses the fairly recent *nonnegative matrix factorizations* gradient descent method, giving the appropriate information about the lack of a good solution for the problem of finding a termination criterion for the iterations in the algorithm. Also a descent method convergence to a global minimum is not guaranteed and in some cases slow. MATLAB scripts are also presented which facilitates the understanding of the topic. This is the last chapter from part I which was intended to provide the reader with a good basis for understanding the applications described in the second part of the book, Data Mining Applications.

In Chapter 10: *Classification of Handwritten Digits* the author starts by explaining the basics of the problem as well as providing a proper problem definition and pointing out useful sources where the reader could go for more information on the subject. It then quickly describes a simple classification algorithm for classification of handwritten digits based on template matching and continues to explain that there are poor results of this algorithm because it attempts to classify digits to an average of the learning data not taking into consideration any information about the variations within each class of digits.

Afterwards the author explains a different method that is based on the modelling of the variation within each digit class using orthogonal basis vectors using SVD. Elden states that the test phase is quite fast and that the algorithm should be suitable for real time computations. The results – using the US postal service database – depend on the number being classified and go from 80 to 93%, which is not as good as it may seem as the best algorithms reach around 97%.

The author goes then to explain a different method, now using the tangent distance which is a distance measure that is invariable under small transformations and this helps to handle digits that deviate considerably in Euclidian distance from the template (ideal) digit. This way small movements in the curves should not influence the distance function. The classification now has a much better performance, reaching 96.9% in the same database but unfortunately it is computationally very expensive since each test digit is compared to all the training digits. The author then concludes that to be competitive it must be combined with some other algorithm that reduces the number of tangent distances comparisons.

Chapter 11: *Text Mining* starts with a definition of text mining as a method for extracting useful information from large and often unstructured collections of texts. The author then demonstrates that in many cases when making searches, straightforward word matching tends not to be good enough as there may be small differences in between the queried string and the target resource. The author then explains the steps of pre-processing as being related to the removal of the words that can be found in virtually any document (stop words) and stemming as the process of reducing each word that is conjugated or has a suffix to its stem, acknowledging that from the point of view of information retrieval in fact no information is lost.

The chapter explains the vector space model and its main idea of creating a term-document matrix where each document is represented by a column vector. Using the cosine distance measure the Chapter presents two examples of query matching with the purpose of finding documents that are relevant to a particular query.

The author then explains latent semantic indexing and states that this is based on the assumption that there is some underlying latent semantic structure in the data, that is corrupted by the wide variety of words used. Elden argues that this semantic structure can be discovered and enhanced by projecting the data (the term document and the queries) onto a lower-dimensional space using SVD, which is then demonstrated with examples.

The next topic in this chapter is a very short introduction to clustering where the author discusses the k-means algorithm as another method for low-rank approximation of the term-document matrix. The chapter then finishes by acknowledging the need of several test collections and a sequence of queries (rather than a single one) when comparing different methods for information retrieval.

In Chapter 12: *Page Ranking for a Web Search Engine* the reader can find an introduction to the problem together with an explanation of the stages of how a search is done exemplified using Google. Following this, Elden explains the impossibility to define a generally valid measure of relevance that would be acceptable for all users of a search engine and explains "Pagerank" with mathematical rigor. In this chapter the reader can also find good explanations of other methods using Random Walk and Markov Chains which are appropriately exemplified together with the power method for pagerank computation. The chapter finalizes with an explanation of HITS (Hypertext Induced Topic Search), which is also based on the link structure of the web.

Chapter 13: *Automatic Key Word and Key Sentence Extraction* starts by arguing that due to the explosion of the amount of textual information available, there is a need to develop automatic procedures for text summarization. The chapter explains what is meant by text summarization and presents a definition of it being the extraction of content from a text document and presentation of the most important content to the user in a condensed form and in a manner sensitive to the user's or application's need. The author then acknowledges this definition as being wide and presents his chapter goal as being much less ambitious: to present a method for automatically extracting key words and key sentences from a text.

The Chapter then shows how to create a term-sentence matrix (term-document matrix). The basis of the presented procedure is the simultaneous but separate ranking of the terms in the sentences: the higher the salience score the more important the term. The procedure is then tested using the chapter 12 of the book as data.

Chapter 14: *Face Recognition Using Tensor SVD* uses the TensorFaces approach, by letting the modes of the tensor represent a different viewing conditions (e.g.: illumination or facial expression). The author argues that it becomes possible to improve the precision of the recognition algorithm compared to the PCA method being a popular technique that often is referred to as "eigenfaces". The chapter also presents a discussion on how the tensor higher order singular value decomposition can also be used for dimensionality reduction to reduce the flop count.

The Chapter 15: *Computing Eigenvalues and Singular Values* aims to give an orientation about what is behind high-level functions (from modern programming environments such as MATLAB). It also briefly describes some methods for computing eigenvalues and singular values of dense matrices and large sparse matrices, pointing out references. In order to have a sound theoretical basis for the decision of when a floating point number is small enough to be considered zero, the author then explains perturbation theory as one may need to know how sensitive the eigenvalues and eigenvectors are in response to small perturbations of the data.

The book finishes with a quick discussion relating to the rather common mistake of underestimating the costs of developing software.

Potential readers should have in mind that Elden has clearly set his target audience. Readers with no scientific computing or numerical analysis background would be unlikely to benefit from the book. Elden's work would be highly recommended to those looking for a compact introduction to the field, that not only describes a number of methods but exemplifies them in real world scenarios.