SPRINGER BRIEFS IN COMPUTER SCIENCE

Harry Strange
Reyer Zwiggelaar

Open Problems in Spectral Dimensionality Reduction

Springer

SpringerBriefs in Computer Science

For further volumes:
http://www.springer.com/series/10028

Harry Strange · Reyer Zwiggelaar

Open Problems in Spectral Dimensionality Reduction

Springer

Harry Strange
Department of Computer Science
Aberystwyth University
Aberystwyth
UK

Reyer Zwiggelaar
Department of Computer Science
Aberystwyth University
Aberystwyth
UK

ISSN 2191-5768 ISSN 2191-5776 (electronic)
ISBN 978-3-319-03942-8 ISBN 978-3-319-03943-5 (eBook)
DOI 10.1007/978-3-319-03943-5
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013956626

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

> The ability to simplify means to eliminate the unnecessary
> so that the necessary may speak.
>
> —Hans Hofmann, 1880–1966

The last few years have seen a great increase in the amount of data available to scientists, engineers, and researchers from many disciplines. Datasets with millions of objects and hundreds, if not thousands, of measurements are now commonplace in areas such as image analysis, computational finance, bio-informatics, and astrophysics. This large volume of data does, however, come at a price, more often than not many computational techniques used to analyze these datasets cannot cope with such large data. Therefore, strategies need to be employed as a pre-processing step to reduce the number of objects, or measurements, whilst retaining important information inherent to the data. One of the key problems with such datasets is how to reduce the number of measurements, often referred to as dimensions, in such a way that the reduced set of measurements captures the main properties of the original data. Spectral dimensionality reduction is one such family of methods that has proven to be an indispensable tool in the data processing pipeline. In recent years, the area has gained much attention; thanks to the development of nonlinear spectral dimensionality reduction methods, often referred to as manifold learning algorithms.

Spectral dimensionality reduction methods can be broadly split into two categories; those that seek to maintain linear properties in the data, and those that seek to maintain nonlinear, manifold, properties. Both linear and nonlinear methods achieve the reduction in dimensionality through the careful construction of a feature matrix, the spectral decomposition of which gives rise to the reduced dimensionality dataset. Ever since the first nonlinear spectral dimensionality reduction methods were proposed over a decade ago, numerous algorithms and improvements have been proposed for the purpose of performing spectral dimensionality reduction. Although these algorithms may improve and extend existing techniques, there is still no gold standard technique. The reasons for this are many; however, one of the core problems with the area is that there are still many obstacles that need to be overcome before spectral dimensionality reduction that can be applied to a specific problem area. These obstacles, referred to herein as *open problems*, have implications for those without a background in the area who wish to employ spectral dimensionality reduction to their problem domain.

Those wish to use spectral dimensionality reduction without prior knowledge of the field will immediately be confronted with questions that need answering; what parameter values to use? how many dimensions should the data be embedded into? how are new data points incorporated? what about large-scale data? For many, a search of the literature to find answers to these questions is impractical, as such, there is a need for a concise discussion into the problems themselves, how they affect spectral dimensionality reduction and how these problems can be overcome.

This book provides a survey and reference aimed at advanced undergraduate and postgraduate students as well as researchers, scientists, and engineers in a wide range of disciplines. Dimensionality reduction has proven useful in a wide range of problem domains, and so this book will be applicable to anyone with a solid grounding in statistics and computer science seeking to apply spectral dimensionality to their work.

Acknowledgments

We both would like to thank the love and support of Jo, Diana, and our families. As well as to this, we would like to acknowledge both NISCHR and the Aberystwyth APRS funding schemes for partially supporting this work.

Aberystwyth, October 2013

Harry Strange
Reyer Zwiggelaar

Contents

Chapter 1
Introduction

Abstract A brief introduction to dimensionality reduction and manifold learning is provided and supported by a visual example. The goals of the book and its place in the literature is given, while the chapter is concluded by an outline of the remainder of the book.

Keywords Manifold learning · Spectral dimensionality reduction · Medical image analysis

Many problems in the machine learning, computer vision, and pattern recognition domains are inherently high-dimensional, that is, the number of measurements taken per observation is considered 'high'. There are both practical and theoretical reasons for wanting to reduce the number of measurements to a more manageable amount. One such motivating factor is that visualising information that contains more than three dimensions is a near impossible task. Although methods have been presented to aid in multi-variate visualisation, humans are still limited to visualising data in three dimensions or less. As such, reducing the number of measurements, or dimensionality, of a dataset is an important process when seeking to visually analyse and understand the data. From a theoretical perspective, high-dimensional spaces have a number of properties that can pose real difficulties. The well known "curse of dimensionality" [1] and "empty space problem" [2] are two examples of such difficulties; as the number of dimensions increases, the properties often associated with 2 or 3-dimensional Euclidean spaces disintegrate and are replaced by strange and complex phenomena [3].

Such practical and theoretical motivations drive the need for automatic methods that can reduce the dimensionality of a dataset in an intelligent way. Spectral dimensionality reduction is one such family of methods. Spectral dimensionality reduction seeks to transform the high-dimensional data into a lower dimensional space that retains certain properties of the subspace or sub manifold upon which the data lies. This transformation is achieved via the spectral decomposition of a square symmetric

H. Strange and R. Zwiggelaar, *Open Problems in Spectral Dimensionality Reduction*,
SpringerBriefs in Computer Science, DOI: 10.1007/978-3-319-03943-5_1,