Chapter 12

Normalized Projection and Graph Embedding via Angular Decomposition

Dengdi Sun
Anhui University, China

Chris Ding
Anhui University, China & University of Texas at Arlington, USA

Jin Tang
Anhui University, China

Bin Luo
Anhui University, China

ABSTRACT

Dimensionality reduction plays a vital role in pattern recognition. However, for normalized vector data, existing methods do not utilize the fact that the data is normalized. In this chapter, the authors propose to employ an Angular Decomposition of the normalized vector data which corresponds to embedding them on a unit surface. On graph data for similarity/kernel matrices with constant diagonal elements, the authors propose the Angular Decomposition of the similarity matrices which corresponds to embedding objects on a unit sphere. In these angular embeddings, the Euclidean distance is equivalent to the cosine similarity. Thus data structures best described in the cosine similarity and data structures best captured by the Euclidean distance can both be effectively detected in our angular embedding. The authors provide the theoretical analysis, derive the computational algorithm, and evaluate the angular embedding on several datasets. Experiments on data clustering demonstrate that the method can provide a more discriminative subspace.

DOI: 10.4018/978-1-4666-1891-6.ch012

INTRODUCTION

Dimensionality reduction is an important problem in pattern recognition, and various methods have been proposed. From the point of view of data embedding, there are two categories of embedding approaches. For vector data embedding, Principal Component Analysis (PCA) for unsupervised data and Linear Discriminate Analysis (LDA) (Duda, Hart & Stork, 2001; Wang Ding, & Huang, 2010) for supervised data are the two most widely used linear algorithms because of their relative simplicity and effectiveness. For graph data embedding, Laplacian Embedding (LE) (Hall, 1971; Belkin & Niyogi, 2003; Luo et al., 2009) is a classical method; in addition, Manifold learning also is one important class of popular approaches such as Isomap (Tenenbaum, de Silva, & Langford, 200), Locally Linear Embedding (LLE) (Roweis & Saul, 2000), Local Tangent Space Alignment (LTSA) (Zhang & Zha, 2004), Locality Preserving Projections (He & Niyogi, 2003), etc.

The most widely used PCA projects data into a subspace using a least square data representation error function. However, in many applications such as information retrieval, image analysis, and genomics, normalized vector data come naturally. PCA does not take advantage of this special nature for the normalized data. Furthermore, in machine learning, many graph data including pairwise similarities are produced by kernel functions (Genton et al., 2001), such as the most widely used RBF kernel, which usually have constant/unit diagonal elements. Most existing embedding methods do not utilize this property.

This motivate us to propose a new embedding method called Angular Decomposition (also called angular embedding) to deal with normalized data or graphs/kernels with constant diagonal elements. The decompositions correspond to embedding data onto a low-dimensional spherical surface. Although Angular Decomposition is best suited to normalized vector data and graph data with constant diagonal elements, it also applies to un-normalized data or graph data with non-constant diagonal. One important feature of angular embedding is that because the embedded data are on the unit sphere, the cosine similarity is equivalent to the Euclidean distance. Thus data structures best described in the cosine similarity and data structures best captured by the Euclidean distance can both be effectively detected in our angular embedding.

Below, we first introduce Angular Decomposition for vector data and for graph data. We then derive computational algorithms for each decomposition respectively. We evaluate these new data decompositions for unsupervised learning. We perform angular embedding on several common datasets. Experiment results demonstrate the effectiveness of these new decompositions as compared to existing approaches.

ANGULAR DECOMPOSITION

We start with a brief discussion of PCA, which is the most widely used dimensionality reduction method. Let the input data matrix $X = (x_1, \cdots, x_n) \in \Re^{p\times n}$ contains the collection of *n* data column vectors in *p* dimension space. In image processing, each column *xi* is a linearized array of pixels' gray levels; in text processing, x_i is a document. PCA finds the optimal low-dimensional (*k*-dim) subspace defined (spanned) by the principal directions $U = (u_1, \cdots, u_k) \in \Re^{p\times k}$. The projected data points in the new subspace are $V = (v_1, \cdots, v_n) \in \Re^{k\times n}$. PCA finds *U* and *V* by minimizing

$$\min_{U,V} J_{PCA} = \left\| X - UV \right\|_F^2 \tag{1}$$

The global optimal solution is rank-k singular value decomposition, $X \approx U\Sigma V$. We absorb Σ into V in Equation 1.

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/normalized-projection-graph-embedding-via/69079

Related Content

Application of Novel Signal Processing Algorithms for the Detection and Minimization of Skywave Interfering Signals in Loran Receivers

Abbas Mohammedand David Last (2011). *Applied Signal and Image Processing: Multidisciplinary Advancements (pp. 28-44).*

www.irma-international.org/chapter/application-novel-signal-processing-algorithms/52110

Application of Software in Soil and Groundwater Recharge Estimation in Ilorin, Nigeria

Afolabi M. Asani, Salihu Lukmanand Isaiah Adesola Oke (2019). *Applications of Image Processing and Soft Computing Systems in Agriculture (pp. 184-207).*

www.irma-international.org/chapter/application-of-software-in-soil-and-groundwater-recharge-estimation-in-ilorin-nigeria/224080

A Metaheuristic Algorithm for OCR Baseline Detection of Arabic Languages

F. Daneshfar, W. Fathyand B. Alaqeband (2018). *Computer Vision: Concepts, Methodologies, Tools, and Applications (pp. 707-734).*

www.irma-international.org/chapter/a-metaheuristic-algorithm-for-ocr-baseline-detection-of-arabic-languages/196975

A Survey on Local Textural Patterns for Facial Feature Extraction

Uma Maheswari V, Golla Vara Prasadand S Viswanadha Raju (2018). *International Journal of Computer Vision and Image Processing (pp. 1-26).*

www.irma-international.org/article/a-survey-on-local-textural-patterns-for-facial-feature-extraction/208175

Ethical Aspects and Considerations of Machine Learning, Artificial Intelligence, and Image Processing in Marketing Research and Branding

Achakkagari Shashidharand Parvez Ahmad (2026). *AI, Machine Learning, and Image Processing in Market Research and Branding (pp. 221-246).*

www.irma-international.org/chapter/ethical-aspects-and-considerations-of-machine-learning-artificial-intelligence-and-image-processing-in-marketing-research-and-branding/403373