Reducing the dimensionality of data using different techniques

Ahmed Fadhil Mohammed[1]
Al-Muthanna Education Directorate, Al- Muthanna Governorate, Iraq
Dhulfiqar Abbas Hoiji[2]
Al-Muthanna Education Directorate, Al- Muthanna Governorate, Iraq

Abstract :

High-dimensional data processing is really the key issue in many systems, including content-based extraction, voice signals, fMRI analyses, electroencephalogram object detection, multimedia extraction, market-based technologies, etc. In order to enhance the system's effectiveness, the data dimensions must be minimized to a low dimensional space. Inthis paper, we analyzed linearization, nonlinear and network embedding dimensionality reduction methods. A few of these methods are ideal and used for linear data that have linear relationship between data points, and many other dimensionality reduction methods are used for nonlinear data that have nonlinear relationship among data points.From an analysis of this paper, we found that Structural Deep Network Embedding (SDNE), LINE (Large-Scale Network Embedding) and Nod2Vec are the best techniques for dimensionality reduction in network data.Furthermore, every approach has its own characteristics and drawbacks. This study presents different methods utilized to minimize the high dimensional data into low dimensional space.

Introduction

Dimensionality reduction can be defined as the process of mining the important information from the data. Real-world data typically seems to have a high dimensionality. Examples of these datasets are speech signals, digital images, or fMRI scan. However, their dimensionality required to be reduced as to be handled [1].The main objective of Reduction of dimensionality is to reduce the high-dimensional data into some kind of desirable low-dimensional representation [2].The decreased representation is assumed to correspond to the fundamental dimensionality of the data that is really the minimal number of features required to represent the original data features [3]. There are a number of techniques for reducing the dimensionality of data [1] that can be classified as per several viewpoints like unsupervised , supervised , nonlinear, linear, global and local .Traditional models include standard techniques like Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA) these techniques are called as linear approaches. By a linear transformation, the low-dimensional projections produced by all these approaches are connected to the input data features vector. Therefore, if the input feature vector locates on low-dimensional manifold, the linear approaches cannot handle the complexity because its structure of the data appears extremely nonlinear[4].Withcomparing the classical linear and nonlinear dimensionality reduction methods, the nonlinear methods are capable of dealingwith complicated nonlinear data. So that nonlinear dimensionality reduction approaches can give an advantage particularly for real-world data that have a highly nonlinear structure [5, 6]. Networks have become commonplace for sharing data and so many real-world applications really have to analyze the data among these applications. In instance, Twitter's recommendation system tends to extract the popular tweets for clients from social networks.Internet advertisement also requires aggregating the clients in social network into groups. It is also very necessary to leverage the data in the network. How and when to practice useful network configurations is one of the underlying issues [7]. In order to recreate the network in the trained embedding space, the accurate method is to incorporate networks in a low-dimensional space, for example to practice vector representations within each edge. The data mining processes such as classification [8], information retrieval [9] and clustering [10] in the network can be performed in low dimensional space.

2. Dimensionality reduction.

This sub section deal with the definition of dimensionality reduction and its function for transforming the high dimensionality data into low dimensionality space. Suppose we have dataset consists of X = $\{x_1 , x_2 , x_3 \ldots . x_N\}, x_i \in R^D$.

Where D denotes the total number of features and N represents the number of data samples. Dimensionality reduction can be well-defined as discovery a mapping function Z: x→ y which it utilize toconverts$x \in R^D$ to the preferred lowdimensional exemplification$y \in R^D$. The inspiration of dimensionality reduction is recovering the fundamental dimensionality of the original data.

3. Dimensionality reduction methods.

World data has linear and nonlinear relationship so that the dimensionality reduction techniques can be classified into two types (1) linear dimensionality reduction methods and (2) non-linear dimensionality reduction methods.

3.1 Linear dimensionality reduction methods.

This sub section presents the most used linear dimensionality methods that are principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA)

3.1.1 PCA

Principal Component Analysis (PCA) is one of the most common unsupervised methods that used to handle the problem of data's dimensionality. It plays a major role in data mining and machine learning. The PCA projections are conducted by optimizing the relationship between the data points. The projections offers a low dimensional space that really can reflect all data points without missing any information [11]. The central concept of the PCA is transforming the high dimensional space into low input space where the maximum Covariance is depicted. The equation for PCA is presented below.

$$\widehat{W} = argmax\ Tr\ (W^T\ Cov(X)W) \quad (1)$$

$$W^T W = 1 \quad (2)$$

Where $\widehat{W}$ represents the linear transformation matrix $\widehat{W} \in R^{d \times D}$. In addition to linearity assumptions, PCA suggests that main components with greater correlated variances reflect an interesting structure, whereas those with smaller variances indicate noise [12].PCA also believes that the principal components are in orthogonal form, which creates PCA reversible through eigen decomposition methods.

3.1.2 LDA

Linear Discriminant Analysis (LDA) is one of supervised linear dimensionality reduction methods. The selection of features throughout the conventional LDA is performed by maximizing the gap between classes and reducing distances throughout classes. For better discrimination, the high dimensional complexity is decreased to a low dimensional subspace [13].LDA propose that the representation of every class is generalized linear and that the linear transform matrix $\widehat{W} \in R^{d \times D}$ is computed by optimizing the proportion in between the interclasses scatters that are S_B and S_W [12]. The formula of LDA is given below.

$$\widehat{W} = argmaxTr\{(W^T S_W W)^{-1}\ (W^{-T} S_B W)\} \quad (3)$$

Both issues can be overcome by eigen decomposition. A low - dimensional d can also be derived by identifying a massive difference in the values spectrum [13]. The low-dimensional depiction can be represented by $Y = W^T X$.Figure 1 displays the comparison between LDA PCA.Two-dimensional data mainly categorized into two classes displayed in red and red. Blue is transformed into a single dimension. A PCA selects the position of the highest variance, as seen by the teal curve that cause to a high overlap whereas a LDA takes into consideration the marked classes and corresponds to a representation of the green curve providing a very much better class distinction.

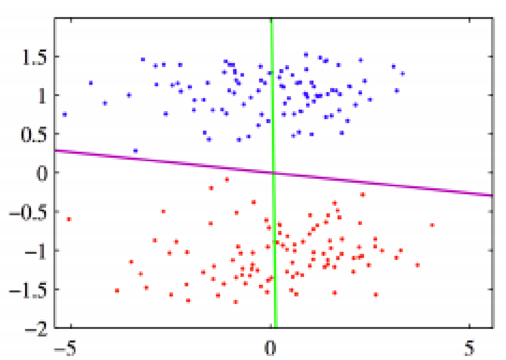

Figure 1: LDA vs PCA [14]

3.2. Non-Linear dimensionality reduction methods.

In actual world data, many of the datasets are in the representation of nonlinear relationship. Controlling such representations of data besides more evaluation is challenging. Several methods can tackle this form of non-linear data that are the following.

3.2.1 Independent Component Analysis.

Independent Component Analysis is one of important unsupervised nonlinear technique for mining individual components from high dimensional data. A main objective of ICA would be to retrieve independent features from observations, which are based on other certain data.ICA discovers the significant relation between the data, as well as decorrelates the same data via minimizing the dissimilarity information [15]. The formula for ICA is presented below.

$$(x_1 , \ldots\ldots , x_m)^T = f(s_1 \ldots. , s_k)^T \qquad (4)$$

Where, f represents the function of actual m-dimensional vectors, x_m is input features.ICA requires reducing the linearity between the data elements and catches the covariance thru the novel independent data vectors. Consequently that determined independency is attained [16].

3.2.2. Multi-Dimensional Scaling.

MDS isa group of non-linear methods to turn high-dimensional data points to low-dimensional data. Throughout the stress function [17], the fault among the pair wise difference between both the low dimensional data as well as a high dimensional data is represented. The formula for stress function is given below.

$$R(Y) = \sum_i^j (\|x_i - x_j\| - \| y_i - y_j \|) \quad (5)$$

Where, $\|x_i - x_j\|$ refers to the Euclidean space among the high dimensional data items and $\| y_i - y_j \|$ represents low dimensional data items.

3.2.3 Isomap

Isomap is one of nonlinear dimensionality reduction method. The traditional scaling methods has been proved their efficiency in more implementationsbut these methods fails again from assumption that they primarily seeks to preserve pairwise Euclidean distances and therefore do not take into consideration the representation of adjacent data points.Therefore, the Isomapmethod [18] has provided a solution for this problem by trying to reserve pairwise geodesic distancesamongstthe data points.The performance of Isomap can be summarized in three points: a. Creates neighborhood map on the manifold. b. Calculates the shortest distanceamong pairwise points. c. Creates low-dimensional embedding thru using multi-dimensional scaling [19]. The formula for Isomap is presented below.

$$W = T(D_G) = -\frac{HSH}{2}(6)$$

Where H = $-\frac{1}{N}ee^T$ and e indicates N-dimensional vector. S referred to squared distance matrix.

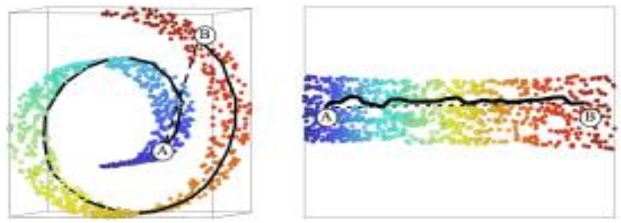

Figure 2: Isomap for dimensionality reduction in Swiss Roll Dataset [20]

3.2.4 Network embedding dimensionality reduction technique

Jian Tang et al. (2015) have proposed Large-scale InformationNetwork Embedding (LINE) model .this model is based on first and second orders proximitiesfor transforming high dimensionality real world data networks into low dimensionality spaces. They reported objective function that can be used to maintain both global and local network architecture. Furthermore,A vertex-sampling mewas appliedfor improving the objective function. This method solves the drawbacks of the traditional Stochastic Gradient Decent (SGD) and increasesthe efficiency and proficiency of the interpretation. For conducting an experiment, the authors have used different real network datasets such as citation networks(AuthorCitation and PaperCitation),Network language (Wikipedia pages classification), and social network (Flickr,Youtube).They have compared another two dimensionality techniques like Deep Walk , Graph factorization with their LINE model on same datasets. From experimental results, it is observed that LINE model outperforms all others techniques. The classification results by the LINE model wereas depicted in following table 1.

Name of dataset	Macro F1-score (%)	Micro F1-score (%)	Technique	Typeof classification
AuthorCitation dataset PaperCitation dataset	65.14 53.02	66.05 62.80	LINE(1st+2nd) LINE(1st+2nd)	Multi-labels Multi-labels
Youtube dataset	39.40	46.43	LINE(1st+2nd)	Multi-labels
Flickr network dataset	63.68	64.74	LINE(1st+2nd)	Multi-labels
Wikipedia dataset	**83.66**	**83.74**	LINE(1st+2nd	Multi-labels

With comparing the performance of LINE on the used dataset,It noted that LINE model provides better results in case of the Wikipedia dataset.

B. Perozzi et al.(2014) have reported Deep Walk dimensionality reduction technique. DeepWalk have introduced solution to social network embedding that is only appropriate to binary vertex networks. As to achieve contextual data, reduced random walks beginning first from edge can be utilized for each node in the network. It is based on second-order proximity method. A.

Ahmed et al. (2013) have proposed Graph Factorization (GF) dimensionality reduction approach fordata networks that can be formed by affinity matrix and can define each edge by matrix factorization with something like a low-dimensional matrix. Using stochastic gradient descent, network factorization is improved and is capable of managing large size networks. It also only implements to networks that are undirected.

Palash Goyal et al.(2017) have presented comparative analysis study for the performance of differentgraph and network embedding dimensionality reduction techniques. These techniques are Structural Deep Network Embedding (SDNE), Large-scale Information Network Embedding (LINE), Locally Linear Embedding (LLE), Node2Vec, Laplacian Eigenmaps (LE) and High-Order Proximity preserved Embedding (HOPE). They implemented all of these techniques on four datasets such as Synthetic, Social Network, Collaboration Network and Biology Network. From experimental results, the best performance was provided by Node2Vec and SDNE techniques onSynthetic(SBM) dataset with obtained 100 % in the term macro F1-score.

Wang et al.2016 have used Structural Deep Network Embedding (SDNE) technique toachieve network embedding and reduce the dimensionality in network data. The SDNE approachmaintain the network framework and has an ability to filter the network information to an extremely non-linear subspace that is reliable for sparse networks. The authors have evaluated five real different networks datasets using SDNE technique. These datasets were social network of online users dataset (Blogcatalog, Flickr and Youtube), Arxiv GR-QC (papers collaboration network) and 20-Newsgroup (20000 newsgroup documents). From the results analysis of this study , it is noted that the SDNE technique presented better classification performance on Arxiv GR-QC dataset with obtained 83 % in the term of MAP (Mean Average precision). With comparing the performance of DeepWalk, LINE, Laplacian Eigenmaps and Common Neighbor techniques with the SDNE , it was observed that the SDNE outperforms other techniques

4. Conclusions

High-dimensional data processing is the key problem in many systems. In this study, we analyzed different methods for minimizing the features of the original data. From this paper , we found that the Principal Component Analysis and linear Discriminant Analysis can be very effective methods to control linear dimensional data representation, where the Multi Dimensional Scaling, Independent Component Analysis and Isomap methods can perform efficiently on nonlinear relationship data.From an analysis of previous studies that related to network embedding and the dimensionality reduction, we observed that theSDNE, LINE, Nod2Vec techniques provide the best performance in dimensionality reduction in network data.

References.

[1]. Van Der Maaten, L., Postma, E., & Van den, J. (2009). Dimensionality reduction: a comparative. J Mach Learn Res, 10(66-71), 13.

[2]. S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin. (2007) “Graph embedding and extensions: a general framework for dimensionality reduction,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 1, pp. 40–51.

[3]. K. Fukunaga, Introduction to statistical patternrecognition. Academic press, 1990.

[4]. T. F. Cox and M. A. Cox, Multidimensional scaling. CRC Press. (2000).

[5] L. K. Saul, K. Q. Weinberger, J. H. Ham, F. Sha, and D. D. Lee. (2006). "Spectral methods for dimensionalityreduction," Semi supervised learning, pp. 293–308.

[6]. C.J.C. Burges. DataMining and Knowledge Discovery Handbook. (2005).A Complete Guide for Practitioners and Researchers, chapter Geometric Methods for Feature Selection and Dimensional
Reduction: A Guided Tour. Kluwer Academic Publishers.

[7]. S. Chang, W. Han, J. Tang, G.-J. Qi, C. C. Aggarwal, and T. S.Huang. (2015). Heterogeneous network embedding via deep architectures. InProceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 119–128. ACM.

[8]. A. Krizhevsky, I. Sutskever, and G. E. Hinton (2012). Image netclassification with deep convolutional neural networks. In Advancesin neural information processing systems, pages 1097–1105.

[9]. Y. Weiss, A. Torralba, and R. Fergus (2009). Spectral hashing. In Advancesin neural information processing systems, pages 1753–1760.

[10]. A. Y. Ng, M. I. Jordan, Y. Weiss, et al (2002). On spectral clustering: Analysis and an algorithm. Advances in neural informationprocessing systems, 2:849–856.

[11]. Arunasakthi, K., &KamatchiPriya, L. (2014). A review on linear and non-linear dimensionality reduction techniques. Machine Learning and Applications: An International Journal (MLAIJ), 1(1), 65-76.

[12]. Yan, S., Xu, D., Zhang, B., Zhang, H. J., Yang, Q., & Lin, S. (2006). Graph embedding and extensions: A general framework for dimensionality reduction. IEEE transactions on pattern analysis and machine intelligence, 29(1), 40-51.

[13]. R. Fisher, "The statistical utilization of multiple measurements.(1998).Ann.Eugenics, vol. 8, no. 4, pp. 376 386.

[14]. C. M. Bishop et al., Pattern recognition and machine learning.(2006). vol. 1. Springer. New York.

[15]. A. Hyvärinen, J. Karhunen, and E. Oja. (2001). Independent Component Analysis. New York: Wiley.

[16]. A. Hyvarinen and E. Oja.(2000). "Independent component analysis: Algorithms and applications," Neural Netw., vol. 13, nos. 4–5, pp. 411–430.

[17]. T. Cox and M. Cox. (1994). Multidimensional scaling. Chapman & Hall, London, UK.

[18]. J.B. Tenenbaum, V. de Silva, and J.C. Langford.(2000). A global geometric framework for nonlinear Dimensionality reduction. Science, 290(5500):2319–2323.

[19]. T. F. Cox and M. A. Cox.(2000) Multidimensional scaling. CRC Press.

[20]. L. K. Saul, K. Q. Weinberger, J. H. Ham, F. Sha, and D. D. Lee. (2006). "Spectral methods for dimensionalityreduction," Semisupervised learning, pp. 293–308.

[21]. A. Ahmed, N. Shervashidze, S. Narayanamurthy,V. Josifovski, and A. J. Smola.(2013). Distributed large-scale

natural graph factorization. In Proceedings of the 22ndinternational conference on World Wide Web, pages

37–48. International World Wide Web Conferences Steering Committee.

[22]. B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk.(2014):Online learning of social representations. In

Proceedings of the 20th ACM SIGKDD internationalconference on Knowledge discovery and data mining,

pages 701–710. ACM.

[23]. Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey. Knowledge-Based Systems, 151, pp.78-94.

[24]. Wang, D., Cui, P., & Zhu, W. (2016, August). Structural deep network embedding. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1225-1234).