Graduate School of Science and Technology Master's Thesis Abstract

Laboratory name (Supervisor)	Mathematical Informatics (Kazushi Ikeda (Professor))		
Student ID	2111413	Submission date	2024 / 7 / 20
Name	CHEN YEN-WEI		
Thesis title	Advances in Dimension Reduction Methods for Interpreting Biological Data: A Comprehensive Survey		

Abstract

Biological systems, often perceived as black boxes with unknown dimensionality, necessitate comprehensive observations to uncover their underlying mechanisms. Dimension reduction techniques play a pivotal role in this endeavor by distilling redundant data into essential components. This study provides a comprehensive survey of dimension reduction methods applied to biological data, including principal component analysis (PCA), independent component analysis (ICA), Fisher discriminant analysis (FDA), canonical correlation analysis (CCA), kernel-based methods, t-Distributed Stochastic Neighbor Embedding (t-SNE), autoencoder, Uniform Manifold Approximation and Projection (UMAP), and CEBRA. We discuss their theoretical foundations and practical applications. Furthermore, numerical experiments validate these methods using synthetic and real-world biological data, highlighting both their effectiveness and limitations. Based on these findings, we showcase the application of various dimension reduction strategies in biological research. The results are presented to serve as a reference for the scientific community, highlighting the role of dimension reduction techniques in enhancing data analysis and interpretation within the biological field.

Keywords: *Biological Systems, Dimension Reduction, Synthetic data, Clustering*