

Data Visualization using Linear and Non-linear Dimensionality Reduction Methods

Junsuk Kim*, Joosang Youn**

Abstract

As the large amount of data can be efficiently stored, the methods extracting meaningful features from big data has become important. Especially, the techniques of converting high- to low-dimensional data are crucial for the 'Data visualization'. In this study, principal component analysis (PCA; linear dimensionality reduction technique) and Isomap (non-linear dimensionality reduction technique) are introduced and applied to neural big data obtained by the functional magnetic resonance imaging (fMRI). First, we investigate how much the physical properties of stimuli are maintained after the dimensionality reduction processes. We moreover compared the amount of residual variance to quantitatively compare the amount of information that was not explained. As result, the dimensionality reduction using Isomap contains more information than the principal component analysis. Our results demonstrate that it is necessary to consider not only linear but also nonlinear characteristics in the big data analysis.

▶ Keyword: Dimensionality reduction, Principal component analysis, Isomap, Residual variance

1. Introduction

정보통신 기술의 발전과 함께 대규모 데이터의 효율적인 저장이 가능해짐에 따라, 다량의 데이터 (빅데이터; Big data) 에서 의미 있는 가치를 추출하고 자료를 시각화하여 정보를 효과적으로 전달하는 기술에 대한 관심이 증대되고 있다 [1,2]. 빅데이터는 다양한 분야에서 빠르게 생산되고 있다. 소셜 네트워크 서비스를 통하여 얻어지는 이용자들의 커뮤니케이션 정보, 스마트 기기의 사용에 따른 이용자의 위치 및 어플리케이션 사용 정보, 도시전체의 대중교통 이용 정보 등 생산되는 데이터는 양적으로 매우 가파르게 증가하고 있다 [3]. 뿐만 아니라, 최근 각광받고 있는 증강현실, 가상현실, 인공지능, 사물 정보통신 기술의 발달은 디지털 정보량의 증가 속도를 더욱 가속화 시킬 것으로 예상된다.

이처럼 디지털화 되어 저장된 거대한 데이터의 분석은 기존의 방법으로는 효과적으로 이루어지기 어렵다. 생체신호 빅데이터를 예로 들어보면, 기능적 자기공명영상 (functional Magnetic Resonance Imaging; fMRI) 방식을 통하여 일반인의 뇌 신호를 데이터화 하는 경우, 10분의 데이터 수집만으로 뇌신경 활성화 정보, 공간 정보, 시간 정보 등이 기록된 대략

10^6 개의 복셀 (voxel) 데이터가 저장된다. 이와 같은 방대한 양의 데이터를 효율적으로 처리하기 위하여 기존의 일변량 분석방법 (Uni-variate analysis) 보다는 특정 패턴이나 의미 있는 관계정보를 추출하는 다변량 분석 (Multi-variate analysis) 이 많이 사용되는 추세이다 [4]. 또한 숨겨져 있는 유용한 상관관계를 찾아내는 데이터 마이닝 (Data mining) 및 학습된 분류기를 통하여 새롭게 입력되는 정보를 예측하는 기계 학습 (Machine learning) 과 같은 통계적 기법들이 빅데이터 분석에 널리 사용되고 있다 [5].

빅데이터 분석에는 의미 있는 정보만을 선택하고 추출하는 기술과 함께, 데이터를 시각화 (Data visualization) 하여 정보를 효과적으로 전달하는 기술도 필수적이다 [6]. 인간은 일반적으로 3차원 이상의 정보를 시각화 하는데 익숙하지 않으므로 고차원의 데이터를 낮은 차원의 데이터로 근사시키는 차원축소 (Dimensionality reduction) 기술이 데이터 시각화라는 측면에서 매우 중요하다. 대표적인 차원축소 방법으로 주성분분석 (Principal component analysis; PCA) [7], 다차원 척도법

• First Author: Junsuk Kim, Corresponding Author: Joosang Youn

*Junsuk Kim (junsuk.kim@skku.edu), Center for Neuroscience Imaging Research, Institute for Basic Science.

**Joosang Youn (jsyou@deu.ac.kr), Dept. of Industrial ICT Engineering, Dong-Eui University

• Received: 2018. 10. 12, Revised: 2018. 11. 15, Accepted: 2018. 11. 21.

(Multidimensional scaling; MDS) [8] 등이 있다. 현재 다양한 분야에서 많은 연구들이 주성분분석과 차원축소법을 사용하여 의미 있는 결과를 얻어냈지만, 두 방법 모두 선형적인 차원축소 방법으로서 데이터의 잠재적인 비선형 특성 (Non-linear feature) 까지 효과적으로 추출하지는 못하였다는 한계가 있다. 최근 데이터의 기하학적 특성을 비선형 공간으로 확장시켜 근사시키는 비선형 차원 축소 기술이 개발되었다 (예, Isomap [9], Manifold learning [10]). 본 논문에서는 선형과 비선형 차원 축소 기법 중 대표적인 방법인 주성분분석과 Isomap 방법을 각각 소개하고 뇌 신호를 저장하고 있는 빅데이터에 적용하여 그 결과를 비교하였다. 실제로 진행되고 있는 뇌 신호 분석 연구들에서 가장 보편적으로 사용되는 방법은 생체 신호의 비선형적 특성을 고려하지 않은 단순화된 혈류역학 함수 (Hemodynamic function) 를 사용하여 연관성을 측정하는 방법이다. 하지만 생체신호들은 본질적으로 시공간적 상관성이 매우 높고 이로 인한 비선형적 특성이 뇌 신호에 반드시 존재하게 된다. 이런 비선형 특성까지 내포하는 저차원 신호를 얻어내기 위해서는 다양한 비선형 차원 축소 기술을 뇌 신호에 적용해 볼 필요가 있다. 특히, 같은 차원에서 담고 있는 정보량인 잔차분산 (Residual variance) 을 통하여 정량적인 비교가 가능하게 하였다.

II. Dimensionality reduction methods

많은 양의 정보를 가지고 있는 데이터일수록 변수 (Variables) 와 관측치 (Observations) 의 개수가 방대해지므로, 데이터는 필연적으로 차원이 매우 높아지게 된다. 데이터의 차원이 커지면 분석의 정확도를 일정 수준으로 유지하기 위한 통계적, 계산적 부하가 지수적으로 증가하는 문제가 발생한다. 이처럼 기존의 차원에 별도의 차원이 더해지면 더해질수록 데이터양과 이를 처리하기 위한 부하가 기하급수적으로 늘어나는 현상을 ‘차원의 저주 (Curse of dimensionality)’ 라고 일컫는다 [11]. ‘차원의 저주’ 를 피하기 위한 방법으로 여러 가지 통계 기법들이 개발되었는데, 가장 일반적인 기법은 핵심이 되는 파라미터를 선별하여 이를 기준으로 데이터의 차원을 낮춰서 처리하는 방법이다 [12]. 대표적인 방법이 이 논문에서 소개할 주성분분석과 Isomap 이다.

2.1. Principal component analysis (PCA)

주성분분석은 데이터의 분산을 최대한으로 보존하면서 서로 직교하는 방향을 찾고 이 벡터에 고차원 공간의 샘플들을 사영시켜 저차원 공간으로 변환하는 기법이다.

변수가 p 개, 관측치가 m 개 있는 행렬로 이루어진 데이터 X 에 대한 주성분은 다음과 같은 선형결합 (Linear combination) 으로 나타난다.

$$\vec{y}_1 = a_{11}\vec{x}_1 + \dots + a_{1p}\vec{x}_p = \vec{a}_1^T X \quad (1)$$

...

$$\vec{y}_m = a_{m1}\vec{x}_1 + \dots + a_{mp}\vec{x}_p = \vec{a}_m^T X$$

이 수식을 행렬 꼴로 변환시키면 $Y = A^T X$ 라고 나타낼 수 있고, 이는 X 를 \vec{a}_1 라는 새로운 벡터에 사영시켜 \vec{y}_i 를 얻어내는 과정이라고 할 수 있다. 주성분분석의 목적은 데이터 X 의 분산을 최대한 보존하는데 있기 때문에 변환된 데이터 Y 의 분산 역시 최대가 되어야한다. X 의 공분산행렬을 Σ 라고 한다면,

$$\begin{aligned} \max_a [Var(Y)] &= \max_a [Var(\vec{a}^T X)] \quad (2) \\ &= \max_a [\vec{a}^T Var(X) \vec{a}] \\ &= \max_a [\vec{a}^T \Sigma \vec{a}] \end{aligned}$$

위의 조건을 만족하는 \vec{a} 는 Σ 의 고유벡터 (Eigenvector) 가 된다. 여기서 Σ 의 고유벡터를 주성분 (Principal component) 이라고 한다.

2.2. Isomap

Isomap 은 주성분분석의 기하학적 해석을 고차원 데이터에 내재되어있는 비선형구조에 까지 확대시킨 방법이다. 이 논문에서 사용한 Isomap 기법은 국소 영역은 선형적 특성을 가지며 이웃 포인트들 (Neighbor points) 간의 선형생성 (Linear spanning) 으로 비선형 특성이 설명가능하다고 가정한다. 구체적으로 이웃 포인트들의 개수 또는 일정 크기의 반지름을 정하고, 매니폴드 (Manifold) 를 따라서 이웃 포인트들 사이의 거리를 더한 총 거리를 선형적 기법을 이용하여 특징 짓는다 [13].

포인트 i 와 j 가 이웃이라고 할 때 두 포인트 간의 유클리드 거리는 $d_x(i, j)$, 측지 거리 (Geodesic distance) 는 $d_M(i, j)$ 로 정의하고, 각 포인트들 간의 가중관계 (Weighted relation) 을 이용하여 고차원상의 가장 가까운 거리 $d_G(i, j)$ 을 계산한다. 이렇게 얻어진 $D_G = d_G(i, j)$ 행렬에 차원축소법을 적용하여 매니폴드의 기하학적 특성을 최대한 보존하는 d -차원의 유클리드 공간 Y 를 생성해낸다. 공간 Y 의 각 포인트들에 대한 좌표벡터 (Coordinate vector) 는 비용함수를 최소화하는 방식으로 얻어진다.

$$E = \| \tau(D_G) - \tau(D_Y) \|_{L^2} \quad (3)$$

여기서 D_Y 는 y 벡터들의 유클리드 거리 행렬을 의미하고 함수 τ 는 유클리드 거리를 내적으로 변환시켜준다 [9].

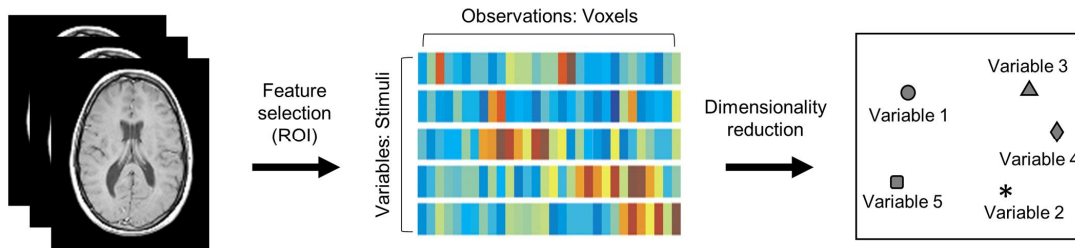


Fig. 1. Overview of neural data visualization procedure

III. Experiment and evaluation

본 연구에서는 기능적 자기공명 영상 기법 (fMRI) 을 이용하여 피험자들이 다양한 강도의 거칠기 자극을 만질 때 동반되는 대뇌에서 일어나는 변화를 데이터로서 수집하였다. 기능적 자기공명 영상 기법은 뉴런의 활동과 상관관계가 있다고 알려진 혈류 활성화 영역, 혈류량의 변화, 혈류 내 산소량의 변화에 대한 정보를 저장한다. 약 1 시간에 걸쳐서 얻어진 방대한 양의 대뇌 신경 활성화 정보는 두 단계를 통하여 분석이 이루어진다 (그림 1). 먼저 촉각 인지와 관련이 있다고 알려진 일차체성감각영역 (Primary somatosensory cortex; S1) 의 관심영역 (Region of interest; ROI) 마스크를 만들고, 여기에 속하는 영역의 신호를 추출해낸다. 이와 같은 특징선택 (Feature selection) 과정을 거친 데이터는 차원축소 기법을 통하여 데이터의 시각화를 하게 된다.

3.1. fMRI experiment

3.1.1. Neural data acquisition

본 실험에는 15 명 (남자 6 명, 여자 9 명) 이 참여하였으며, 모두 폐쇄공포증이 없고 급속 삽입 수술 경험이 없는 평균나이 26.7 세의 정상 성인이었다. 대상자의 선정은 신경과적 병력, 정신과적 병력, 인지 기능에 영향을 미칠 수 있는 질환, 촉각 인지에 문제가 있는 대상은 배제하였으며 모든 피험자가 설문지 및 실험 동의서 작성에 동의하였다.

기능적 자기공명 영상은 독일 막스플랑크 연구소 (Max Planck Institute for Biological Cybernetics, Tuebingen, Germany) 에 위치한 3 테슬라의 Siemens Prisma 장비를 통하여 얻어졌다. 사용된 영상 변수는 TR (Repetition time) = 1,520 ms, TE (Time to echo) = 30 ms, 숙임각 (Flip angle) = 68o, Field of view = 192 mm, Voxel size = 3 × 3 × 3 mm³, 절편수 46 개로 설정하였다. 절편가속 멀티밴드 경사에코 (Slice-accelerated multiband gradient-echo-based echo planar) 펄스파형을 사용하여 뇌영상을 획득하였으며 뇌 신호 전처리에는 SPM8 (Statistical parametric mapping; Wellcome Department of Imaging Neuroscience, UCL, London, UK)을 통하여 이루어졌다. 영상 획득간의 머리 움직임을 보정하기 위하여 재배열 (Realignment) 과정을 진행하였고 고해상도의 해부학적 정보를 이용하기 위하여 T₁-weighted 영상과 상관정립 (Coregistration) 을 시행하였다.

이후 개인 간의 뇌의 모양 및 크기 편차를 보정하기 위하여 정규화 (Normalization) 하였고 마지막으로 2 mm 등방성 가우시안 커널을 사용하여 평활화 (Smoothing) 을 진행하였다.

3.1.2. Experimental design

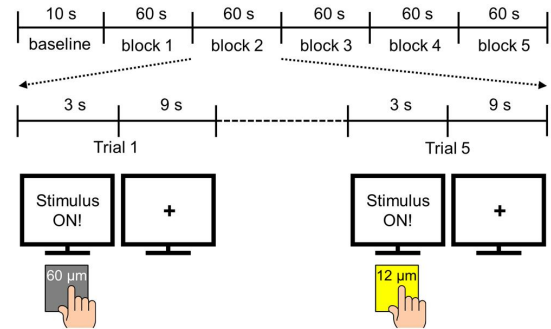


Fig. 2. fMRI experimental design

다양한 강도의 거칠기 자극을 제시하기 위하여, 이전 연구를 통하여 인체에 무해함과 자극 전달의 효율성이 입증된 5 가지 종류의 사포를 준비하였다. 각 사포의 입자 크기는 0.3, 12, 40, 60, 100 μm 로서 큰 입자 크기의 사포일수록 표면은 더 거칠어진다.

뇌 신호가 평형기 (Equilibrium state) 를 이루기 전의 신호를 보정하기 위하여 뇌 활성화 영상을 얻기 전에 10 초간 베이스라인 영상을 추가로 얻었으며 이 영상은 분석에서 제외하였다. fMRI 영상 획득 중에 피험자는 누워있는 상태로 오른손 두 번째 손가락 끝마디를 이용하여 주어진 거칠기 자극을 만지는 과제를 수행하였다. 정확한 자극 제시시점을 알려주기 위하여 프로젝터를 사용하여 자극 제시 시점과 관련된 영상을 피험자에게 보여주었다. 실험은 총 5 블록으로 이루어졌으며 각 블록은 5 가지 강도의 거칠기 자극을 한번씩 제시하는 5 번의 시행으로 구성되었다. 이 실험은 총 3 회 반복되었다. 각각의 시행에서 피험자는 3 초간 주어진 거칠기 자극을 만지고 이어진 9 초 동안은 다음 자극을 기다리는 방식의 과제를 수행하였다 (그림 2).

3.1.3. Neural data analysis

뇌 영상 데이터 분석은 특징선택과 차원축소, 두 단계를 통하여 이루어졌다.

먼저 특징선택은 관심 있는 뇌 영역의 데이터만을 추출해내는 과정이다. 인간의 뇌에서 촉각 정보를 인코딩하고 있다고 알려진 영역은 일차체성감각영역이다 (그림 3). Martuzzi 등의 고

해상도 (7 Tesla) 기능적 자기공명 영상 연구에서 손가락에 촉각 자극을 제시했을 때 일차체성감각영역에서 유의미한 활성화가 일어남을 확인하였다 [14]. 본 연구에서는 일차체성감각영역을 포함하는 관심영역 마스크를 만들어서 특징선택을 진행하였다. 이렇게 선택된 뇌 신호는 1071 차원으로, 기존의 매우 높은 차원보다는 낮아졌지만 아직도 높은 차원의 데이터이다.

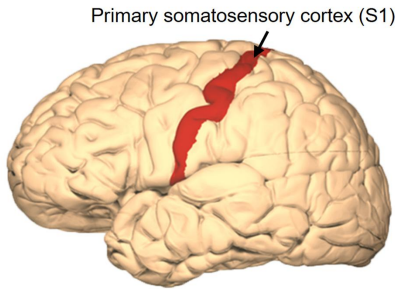


Fig. 3. Anatomical region of primary somatosensory cortex

이후 데이터 시각화를 위한 차원 축소 기법이 적용되었다. 선형, 비선형 기법으로 주성분분석과 Isomap 기법이 각각 적용되었다. 차원 축소 결과는 1 차원과 2 차원 공간 매핑을 이용하여 보고하였다. 이와 같은 데이터 시각화를 통하여 일차체성감각영역에 인코딩되어있는 거칠기 강도의 정보가 실제 물리적 강도와 얼마나 비슷한지를 비교할 수 있게 된다. 예를 들어, 실제 거칠기 자극의 물리적 강도는 입자의 크기를 기준으로 결정되므로 1 차원 공간에서 매핑이 가능하다 (그림 4). 즉, 뇌 신호에 인코딩되어있던 거칠기 자극 강도 정보를 1 차원 공간으로 매핑 시킨다면, 주성분분석과 Isomap 차원 축소 기법이 얼마나 그 물리적 특성을 효과적으로 보존하고 있는가에 대한 정성적 비교를 할 수 있는 것이다. 3 차원 이상의 공간 매핑 결과는 데이터 시각화

가 쉽지 않으므로, 잔차분산의 양을 비교하는 정량적 비교 방식을 사용하였다. 차원 축소 기법을 통하여 설명하지 못하는 분산의 양을 비교하므로 차원 축소 이후에 얼마나 많은 정보를 포함하고 있는지를 직접적으로 비교할 수 있게 하였다.

3.2 Results

그림 4 는 피험자에게 제시한 거칠기 자극의 물리적 특성에 따른 1 차원 매핑과 함께, 다양한 강도의 거칠기 자극에 의하여 유발된 뇌 신호를 차원 축소 기법을 통하여 1 차원 또는 2 차원으로 근사시킨 데이터 시각화 결과를 보여준다. 자극물의 물리적 공간 매핑 결과 (그림 4 의 왼쪽 위 그래프) 를 뇌신경 공간 매핑 결과 (그림 4 의 왼쪽 아래, 오른쪽 그래프) 와 비교해보면 주목할 점이 몇 가지 있다. 뇌 신호를 1 차원으로 근사시켰을 경우, 주성분분석과 Isomap 방법 모두, 위치는 차이가 있지만 거칠기 강도의 순서에 따라 매핑 (0.3-12-40-60-100 의 오름차순) 된 것을 확인할 수 있다. 2 차원으로 근사시킨 결과를 살펴보면, 선형 차원 축소 방법인 주성분분석의 경우 첫 번째 성분 (1st principal component) 에 따라서는 거칠기 강도의 순서에 따라 매핑 되었지만 두 번째 성분을 따라서는 12-100-40-0.3-60 순서로 거칠기 자극의 물리적 특성을 보존하지 못하였다. 하지만 비선형 차원 축소 방법인 Isomap 의 두 번째 성분의 경우 0.3-12-40-100-60 의 순서로 매핑이 되어 주성분분석의 두 번째 성분 보다 제시한 자극의 물리적 특성을 좀 더 효율적으로 보존하고 있음을 확인할 수 있었다.

3 차원 이상의 공간 매핑 결과는 효율적인 시각화가 힘들기 때문에 두 차원 축소 기법의 정량적 비교를 위하여 차원의 변화 (1 ~ 10 차원) 에 따른 잔차분산의 양을 보고하였다 (그림 5). 1 차원에서의 주성분분석과 Isomap 의 잔차분산의 양을 살펴보면 각각 56.1 %, 50.6 % 로 큰 차이를 보이지 않는다. 하

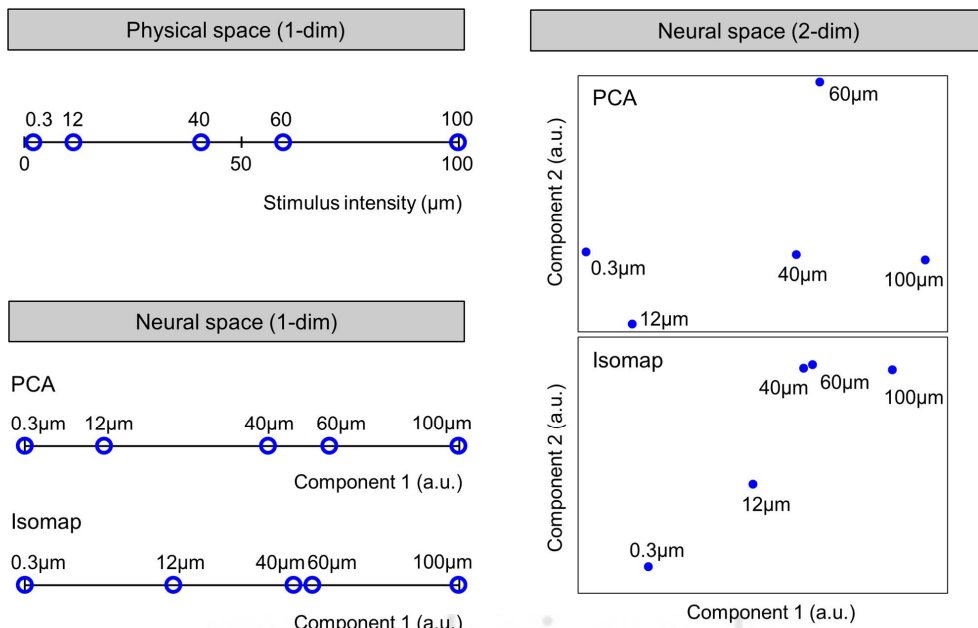


Fig. 4. Visualization results using two different dimensionality reduction methods

지만 점점 차원이 증가함에 따라 잔차분산의 양이 줄어드는 정도가 Isomap 기법의 경우가 확연히 빠름을 확인할 수 있다. 4 차원에서의 주성분분석과 Isomap의 잔차분산의 양을 살펴보면 각각 23.8%, 6.5%로 큰 차이가 나기 시작한다. Isomap 기법의 경우 4 차원 이후로는 잔차분산의 양이 큰 폭으로 줄어들지 않는 이른바 'Point of Elbow'가 존재함을 볼 수 있다. 차원이 점점 증가함에 따라서 주성분분석 기법의 잔차분산의 양도 점차 줄어들어 9 차원 이상에서는 Isomap의 잔차분산의 양과 거의 비슷하게 맞춰졌다.

3.3. Discussion and limitations

본 연구에서는 빅데이터의 시각화 과정에서 발생하는 '차원의 저주' 문제 해결 및 정보를 보다 분명하고 효과적으로 전달하기 위한 차원 축소 기법을 비교하였다. 특히, 선형 차원 축소 기법으로서 주성분분석, 비선형 차원 축소 기법으로 Isomap 기법을 기능적 자기공명 영상을 통하여 얻어진 뇌 신호 빅데이터에 적용하였다. 그 결과, 1 번째 성분에 대해서는 주성분분석, Isomap 기법 모두 제시된 자극의 물리적 특성을 잘 보존하는 차원 축소가 이루어졌다. 하지만 2 번째 성분까지 포함한 매핑의 결과는 Isomap이 주성분분석보다 효과적인 차원 축소 결과를 보여주었다 (그림 4). 또한 차원의 수에 따른 잔차분산의 양을 살펴보면 Isomap이 주성분분석보다 같은 차원에서 보다 많은 정보를 포함하고 있음을 알 수 있다 (그림 5).

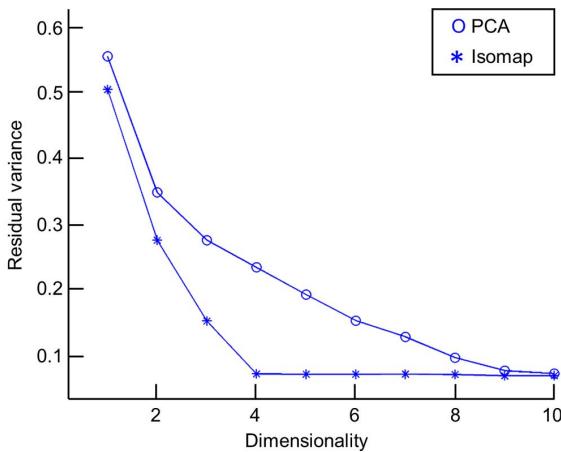


Fig. 5. Performance comparison using residual variances

비선형 차원 축소 기법은 빅데이터에 내재되어 있는 매니폴드의 특징을 보다 효과적으로 모델링하기 위하여 개발되고 사용되어 왔다. 예를 들어, Hannachi 등은 아시아 지역의 문순 현상의 특이점을 진단하기 위하여 거대한 양의 대기 정보 데이터를 수집하고 빅데이터의 비선형적 특징을 고려한 차원 축소를 위하여 사용하였다 [15]. 또한 Zhang 등은 사람의 얼굴 이미지에 포함되어 있는 비선형적 특징을 효율적으로 추출하여 분류에 사용하기 위하여 Isomap 기법을 비롯하여 Locally linear embeddings (LLE) 등의 다양한 비선형 기법을 적용하였다

[16]. 뿐만 아니라, Yang 등은 Isomap과 Laplacian Eigenmap을 함께 활용하여 생물학 데이터에서 특정 기능을 가진 유전자 조각 (Biobrick)을 효과적으로 시각화하는 사례를 소개하였다 [17]. 하지만 Isomap 기법을 뇌 신호 빅데이터의 차원 축소에 적용한 사례는 많지 않다. Hu 등은 기능적 자기공명 영상을 통해 얻은 뇌 신호 데이터에 주성분분석과 Isomap 기법을 적용하였다. 이 연구에서는 두 기법을 비교하는 방법으로 잔차분산의 양을 보고하였는데 본 연구결과와 마찬가지로 Isomap 기법이 주성분분석보다 같은 차원에서 포함하고 있는 정보의 양이 많았다 [13]. 이전의 연구들과 본 연구 결과를 종합해보면, 빅데이터 분석에 있어서 선형적 특징뿐만 아니라 비선형적 특징을 고려한 차원 축소가 보다 효과적인 데이터의 시각화를 가능케 함을 알 수 있었다.

본 연구는 아래와 같은 한계가 있기에 추후에 이를 고려한 분석이 이루어져야 할 것이다. 그림 4에서 뇌 신호 데이터의 시각화 결과를 자극물의 물리적 특성과 연관지어 정성적 평가를 진행하였다. 하지만 뇌 신호 데이터는 자극물의 물리적 특성 자체보다는 피험자들이 거칠기의 강도를 어떻게 인지하였는가, 즉 개인의 인지적 특성과 연관이 있을 확률이 더 높다. 향후에 이어질 연구에서 자극물의 물리적 특성과 함께 피험자들의 인지적 특성을 함께 고려하여 뇌 신호 데이터와 비교를 한다면 차원 축소 기법들에 대한 보다 체계적이고 명확한 비교가 가능하리라고 생각된다.

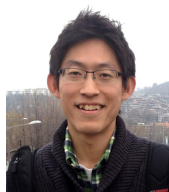
IV. Conclusions

지금 현재에도 다양한 종류의 방대한 데이터가 디지털화 되어 데이터베이스에 저장되고 있다. 때 시간 빠른 속도로 축적되는 데이터의 효율적이고 의미 있는 활용을 위해서는 빅데이터 안에 숨겨져 있는 데이터의 특성을 정확히 파악하고 가치를 추출해내는 과정이 필수적이다. 본 연구에서는 빅데이터의 차원 축소 단계에서 데이터의 선형적 특성뿐만 아니라 내재되어 있는 비선형 특성 (예, 매니폴드)까지도 추출해내는 것이 데이터 시각화라는 측면에 얼마나 중요한지를 보여주었다. 특히, 각 차원에서 잔차분산의 양을 비교해봄으로써 비선형 차원 축소 기법이 기존의 고차원 데이터 모델을 보다 정확하게 표현함을 확인할 수 있었다. 본 연구 결과는 뇌 신호 빅데이터를 대상으로 하였지만 비선형 차원 축소 기법은 고차원의 모든 종류의 빅데이터에도 적용이 가능하다. 이 연구 결과를 기반으로 비선형 차원 축소 기법의 연구가 많아지고 연구 결과가 축적된다면 향후 빅데이터의 보다 효율적인 시각화가 가능할 것이다. 추후 이 연구결과를 바탕으로 기존의 선형적인 차원 축소 기법들의 한계점을 극복할 수 있는, 즉, 데이터의 잠재적인 비선형적 특성을 효율적으로 추출할 수 있는 새로운 차원 축소 기법을 개발하여 기고할 예정이다.

REFERENCES

- [1] Kim, Y, Cho, KH, "Big Data and Statistics", Journal of the Korean Data and Information Science Society, Vol. 24, No. 5, pp.959-974, 2013.
- [2] Park, H-J, Gwon, Y-H, An, Y-M, "Big Data and Its Refining Technology", Korean Society of Computer Information Review, Vol. 21, No. 1, pp.1-8, 2013.
- [3] Kim, ST, "3 Elements for the Successful Big Data Usage: Resources, Technology, Man Power", IT & Future Strategy, Vol. 3, 2012.
- [4] Pereira, F. and Botvinick, M., "Machine learning classifiers and fMRI: a tutorial overview." *Neuroimage*, Vol.45, No.1, pp.S199-S209, 2009.
- [5] Lee, J-H, Je, M-G, Jo, M-J, Son, H-S, "Trends of Big Data Use in Medical Field", *Information and Communications Magazine*, Vol. 32, No. 1, pp.63-75, 2014.
- [6] Choe, G-S, Ham Y-G, Kim S-H, "Visualization of Big Data", *Korean Society of Computer Information Review*, Vol. 21, No. 1, pp.33-43, 2013.
- [7] Pearson, K., "On Lines and Planes of Closest Fit to Systems of Points in Space." *Philosophical Magazine*, Vol.2, No.11, pp.559-572, 1901.
- [8] Bronstein AM, Bronstein MM, Kimmel R, "Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching." *Proc. Natl. Acad. Sci. U.S.A.*, Vol.103, No.5, pp1168-1172, 2006.
- [9] Tenenbaum, J.B., De Silva, V. and Langford, J.C., "A global geometric framework for nonlinear dimensionality reduction." *Science*, Vol.290, No.5500, pp.2319-2323, 2000.
- [10] Gorban, A. N., Kégl, B., Wunsch, D. C., and Zinovyev, A. Y., *Principal manifolds for data visualization and dimension reduction*, Springer, Berlin-Heidelberg, 2008.
- [11] Bellman. R. *Adaptive control processes: A guided tour*, Princeton University Press, Princeton, 1961.
- [12] Hastie, T., Tibshirani, R. and Friedman, J., *Overview of supervised learning*. In *The elements of statistical learning*, Springer, New York, 2009.
- [13] Hu, J., Tian, J. and Yang, L., "Functional feature embedded space mapping of fMRI data." In *Engineering in Medicine and Biology Society (EMBS)*, pp.1014-1017, 2006.
- [14] Martuzzi, R., Zwaag, W., Farthouat, J., Gruetter, R. and Blanke, O., "Human finger somatotopy in areas 3b, 1, and 2: a 7T fMRI study using a natural stimulus." *Human brain mapping*, Vol.35, No.1, pp.213-226, 2014.
- [15] Hannachi, A. and Turner, A.G., "Isomap nonlinear dimensionality reduction and bimodality of Asian monsoon convection." *Geophysical Research Letters*, Vol.40, No.8, pp.1653-1658, 2013.
- [16] Zhang, C., Wang, J., Zhao, N. and Zhang, D., "Reconstruction and analysis of multi-pose face images based on nonlinear dimensionality reduction." *Pattern Recognition*, Vol.37, No.2, pp.325-336, 2004.
- [17] Yang, J., Wang, H., Ding, H., An, N. and Alterovitz, G., "Nonlinear dimensionality reduction methods for synthetic biology biobricks' visualization." *BMC bioinformatics*, Vol.18, No.47, pp.1-10, 2017.

Author



Junsuk Kim received the B.S. degree in Computer and Radio Communications Engineering from Korea University, Korea, in 2010 and received the M.S. and Ph.D. degrees in Brain and Cognitive Engineering from Korea University,

Korea, in 2012 and 2016, respectively. Dr. Kim is a research professor at the Center for Neuroscience Imaging Research, Institute for Basic Science (IBS). He is interested in machine learning and neuroscience.



Joosang Youn received the B.S. degree in Electrical Engineering from Korea University, Korea, in 2001 and received the M.S. and Ph.D. degrees in Electrical Computer Engineering from Korea University, Korea, in 2003 and 2008,

respectively. Dr. Youn is a professor at the Department of Industrial ICT Engineering, Dong-Eui University. He is interested in IoT, AI, and Edge Computing.