

02. ZHOU Qiqi.

Summary

This project addresses the image inpainting problem using Principal Component Analysis (PCA) to construct a dictionary for the compact/sparse representation of the image. Components learned from PCA were used to fill in missing or damaged parts of an image, specifically in the handwritten digits dataset. The project performs PCA on the training images and determines that $k=40$ by selecting the top components that account for at least 99% of the variance importance. The project successfully restores semantic information for moderate damage scenarios, but may not be stable when dealing with highly damaged images.

Pros

1. It is effective to restore missing or damaged parts of an image by PCA to construct a dictionary for image inpainting, which also provides a solution to a common and important problem in computer vision and image processing.
2. Using components learned from PCA as the dictionary is a more efficient and effective solution compared to traditional methods of finding a suitable dictionary for the compact/sparse representation of the image.

Cons

This method may not be stable when dealing with highly damaged images, which limits the generalizability of the method to real-world scenarios.

Scoring

Evaluation on Clarity and Quality of Writing: 5

Evaluation on Technical Quality: 5

Overall Rating: 5

Confidence on my Assessment: 4

05. XU Yingxue; Jiaxin ZHUANG; Fengtao ZHOU.

Summary

The project aims to analyze the complex phenomenon of crime through innovative multivariate analysis techniques, which reduce the complex dataset to two or three significant factors, providing valuable insights for the development of effective solutions to reduce crime rates. The study uses a dataset of 59 major cities across the United States, spanning a period from 1969 to 1992, with 36 variables capturing various socio-economic, demographic, and environmental factors that could potentially influence crime patterns. The data underwent three preparatory computations, and Principal Component Analysis (PCA) was applied to reduce the variable dimensions from 24 to 2/3. PCA had good performance in terms of data reduction and visualization, making it a useful tool for analyzing the crime dataset. The study also used other visualization methods, including Sparse PCA, Multidimensional Scaling (MDS), Isometric Mapping (ISOMAP), and Locally Linear Embedding (LLE), to gain a more nuanced understanding of the data.

Pros

The project uses innovative techniques to analyze a complex phenomenon, providing valuable insights for policymakers and law enforcement agencies. Comprehensive information is considered, such as socio-economic, demographic, and environmental factors, which can provide a more holistic view of crime patterns and help identify underlying causes and risk factors associated with crime occurrence.

Cons

The underlying parameters have different effects on different crime types.

Scoring

Evaluation on Clarity and Quality of Writing: 5

Evaluation on Technical Quality: 5

Overall Rating: 5

Confidence on my Assessment: 4

09. HUANG, Zhanmiao; Wencan XIA; Yuanhui LUO

Summary

This project explores the relationship between genetic variation and geographical variation using single-nucleotide polymorphism (SNP) data. The project applies different dimension reduction methods, including Principal Component Analysis (PCA), Multidimensional Scaling (MDS), Random Forest (RF), and Extra Trees (ET), to extract crucial information from the high-dimensional SNP data. The project investigates the SNPs with top importance and evaluates the number of important SNPs needed for good prediction on regions with statistical learning methods. The project also includes a case study focusing on the genetic variability among populations in China and its neighboring areas.

Pros

1. The use of dimension reduction methods, such as PCA and MDS, can effectively distinguish people from different regions based on essential principal components of SNPs, providing insights into the relationship between genetic variation and geographical variation.
2. Several statistical learning methods, such as RF and ET, are helpful in identifying key features for classification and predict regions where people come from based on genetic information.
3. The case study on the populations of China and its neighboring areas reveals that the similarity of SNPs principal components can reflect the relationship between their geographical locations.

Cons

The project's dataset is limited to 1064 people around the world, which may not be representative of the global population and may limit the generalizability of the findings.

Scoring

Evaluation on Clarity and Quality of Writing: 5

Evaluation on Technical Quality: 5

Overall Rating: 5

Confidence on my Assessment: 4

13. LI Haobo, CHEN Zixin, TENG fei, SHENG Rui.

Summary

This project proposes a visual analytics approach to help researchers explore the evolution of word usage across time in accepted academic papers. The dataset used is a co-occurrence matrix of 11463 words and 5811 NIPS conference papers from 1987 to 2015, which presents challenges such as sparsity and large search space. The project visualizes the progression of word distribution in academic writing and conducts experiments to predict the probability of word occurrences in conference papers using the previous word distribution. The project uses a sunburst chart to organize the top 100 possible words and applies clustering algorithms to predict word probabilities.

Pros

1. The proposed visual analytics approach provides a useful tool for researchers to stay tuned with changes in language usage and adapt their writing style accordingly, increasing their chances of success in the publishing process.
2. The clustering algorithms to predict word probabilities can help researchers predict the possibly frequent words or phrases in future papers, providing valuable insights for academic writing.
3. The word embeddings and visualization techniques, such as the sunburst chart and time series visualization, can provide a clear and intuitive understanding of the evolution of word usage across time in academic writing.

Cons

1. The clustering algorithms for predicting word probabilities may not be accurate in all cases and may overlook important factors such as context and semantics.
2. The project mainly focuses on predicting word probabilities, which may overlook other important aspects of academic writing such as the structure and coherence of the paper.

Scoring

Evaluation on Clarity and Quality of Writing: 4

Evaluation on Technical Quality: 3.5

Overall Rating: 3.75

Confidence on my Assessment: 3

16. Chris HC Nguyen and James M Shihua

Summary

This project demonstrates the use of RPCA with an augmented Lagrange multiplier to separate a video into a foreground and background component, allowing for the detection and removal of moving objects from the video background. The algorithm is implemented using Intel optimized Python and Numpy library with Intel MKL for optimal efficiency. The video is represented by a matrix of N-pixels by M-frames, and the algorithm converges efficiently within tens of iterations.

Pros

1. RPCA with an augmented Lagrange multiplier is significantly more efficient than traditional methods like ADMM.
2. The algorithm converges efficiently within tens of iterations, making it suitable for real-time applications.

Cons

This algorithm relies on a regularization parameter, λ , which may require tuning for optimal performance in different applications.

Scoring

Evaluation on Clarity and Quality of Writing: 4

Evaluation on Technical Quality: 4

Overall Rating: 4

Confidence on my Assessment: 3