Kernel-based Sammon Mapping for Dimensionality Reduction & Data Visualization

Georgios C. Anagnostopoulos
Electrical & Computer Engineering
Florida Institute of Technology
http://my.fit.edu/~georgio
georgio@fit.edu
Abstract, Overview
Abstract

- In this talk a novel generalization of Sammon’s Mapping (SM) will be presented. The SM is a popular, metric multi-dimensional scaling technique used in data analysis and visualization. The new approach, namely the Kernel-based Sammon Mapping (KSM), yields the classic SM and other much related techniques as special cases.

- Apart from being able to approximate distance-preserving projections, it can also learn to metrically represent arbitrarily-defined dissimilarities or similarities between samples. Moreover, it can handle equally well numeric, categorical or mixed-type data. It is able to accomplish all this by modeling its projections as linear combinations of appropriate interpolation kernels.

- Experimental results will be reported, which showcase KSM’s capabilities in visually representing several meaningful relationships between samples of selected datasets. These capabilities enable KSM to be used for exploratory data analysis, manifold learning, and as a preprocessing tool to successfully tackle supervised learning tasks.
Overview

- Sammon’s Mapping (SM)
  - Description, Characteristics, Extensions
- Kernel-based Sammon’s Mapping (KSM)
  - Motivation, Description, Learning, Characteristics
- Experimental Results
  - MSTAR, Teapot, Mushroom, Congressional Voting Recs
- Discussion
  - Summary of findings, Future Work
- Q&A
Sammon’s Mapping
Sammon’s Mapping

- Sammon’s Mapping (SM)
  - A metric MDS method
  - Visual depiction of data, where distances between projected points reflect magnitudes of dissimilarity.
  - SM acts as a nonlinear isometry between the original high-dimensional space to the low-dimensional embedding space.

Sammon’s Mapping

- Inputs
  - Data \( \{x_n\} \) the original (high-dimensional) space
  - Specify a dissimilarity measure \( \delta \)

- SM parameters
  - Point configuration \( \{y_n\} \) in embedding space

- Cost Function
  \[
  E = \frac{1}{2} \sum_{1 \leq i < j \leq N} u_{ij} (d_{ij} - \delta_{ij})^2
  \]
Sammon’s Mapping

Abstract
Overview
Sammon’s Mapping (SM)
Kernel-based Sammon’s Mapping
Experimental Results
Discussion
Credits
Q&A

Original Space

Projection Space

Sammon’s Mapping

Manifold

$\delta_{ij}$

$d_{ij}$
Sammon’s Mapping

- SM is used for Dimensionality Reduction
  - Data Visualization

- Different dissimilarities measures in the original space will yield different results

- Selecting appropriate $u_{ij}$ weights can be very important for manifold unfolding
Sammon’s Mapping

- Main Drawbacks
  - No interpolation/extrapolation capability
  - Number of parameters are $O(N)$; increased training time

- Extensions to SM


Kernel-Based Sammon’s Mapping
Kernel-based SM

- Motivation
  - Previously mentioned MLP-based methods have all the MLP-related drawbacks:
    - Training can be computationally intensive
    - Non-local knowledge representation
    - Choice of model structure
  - Webb’s RBF approach is using strict interpolation
    - Poor generalization
Kernel-based SM

Abstract

Overview

Sammon’s Mapping (SM)

Kernel-based Sammon’s Mapping

Experimental Results

Discussion

Credits

Q&A

KSM Model

\[ y = W^T k(x; \theta) \]
 Kernel-based SM

- **KSM Model**

  \[ y = W^T k(x; \theta) \]

- **Characteristics**
  - Kernel form: \( k(x, c_h; \psi) \)
  - Choice of kernel (similarity measure) implies a dissimilarity measure
  - Number of kernels \( H \) is less or equal to number of training patterns \( N \).
Kernel-based SM

- Hit-or-miss kernel
  \[ k(x, c_h) = \begin{cases} 
    1 & x = c_h \\
    0 & \text{otherwise} 
  \end{cases} \]

- Gaussian kernel
  \[ k(x, c_h; \psi) = e^{-\psi \| x - c_h \|^2_2} \]

- Hyperbolic tangent kernel
  \[ k(x, c_h; \psi) = \tanh(x^T c_h + \psi) \]

- Exponential kernel
  \[ k(x, c_h; \psi) = \psi^{\delta(x, c_h)} \]
Kernel-based SM

- **Training**
  - Weight training
  - Based on SMACOF algorithm
  - Guttman Transform (super-linearly convergent)

\[
W_t = A^{-1} B(W_{t-1}) W_{t-1}
\]

\[
A \doteq \sum_{1 \leq i < j \leq N} u_{i,j} \Delta k_{i,j} \Delta k_{i,j}^T
\]

\[
B(W) \doteq \sum_{1 \leq i < j \leq N, d_{i,j} \neq 0} u_{i,j} \frac{\delta_{i,j}}{d_{i,j}(W)} \Delta k_{i,j} \Delta k_{i,j}^T
\]
Kernel-based SM

- The weight update process (Guttman Transform) is an Iterative Majorization (IM) algorithm
- The IM concept:
  - Find $g(\cdot)$ such that
    \[ f(w) \leq g(w, v), \quad g(w, w) = f(w) \]
  - Sandwich Inequalities
    \[ f(w^{k+1}) \leq g(w^{k+1}, w^k) \leq g(w^k, w^k) = f(w^k) \]
  - IM iteration
    \[ w^{(k+1)} = \arg \min_{w} g(w, w^{(k)}) \]
Kernel-based SM

- **Training**
  - Prototype vector training
    - Gradient
      \[
      \frac{\partial E}{\partial \mathbf{c}_h} = \sum_{1 \leq i < j \leq N} u_{i,j} \left(1 - \frac{\delta_{i,j}}{d_{i,j}}\right) \frac{\partial \Delta \mathbf{k}_{i,j}^T}{\partial \mathbf{c}_h} \mathbf{W} \mathbf{W}^T \Delta \mathbf{k}_{i,j}
      \]
  - Practical updating methods
    - Trust region based Quasi-Newton (small to medium sized datasets)
    - Line Search based Conjugate Gradient methods (medium to very large sized datasets)
Kernel-based SM

- **Advantages**
  - Interpolation/extrapolation capability
  - Generalizes original, RBF-based, and MLP-based SM approaches
  - Local knowledge representation
  - Model complexity control by adjusting H
  - Able to handle both similarities and dissimilarities

- **Drawbacks**
  - Non-statistical basis
  - Limited to cases when preservation of dissimilarities is meaningful to the problem at hand.
  - May be sensitive to “outliers”; the user needs to identify the unimportant dissimilarities via appropriate choice of the U matrix.
  - Slow training of prototype vectors when dealing with large datasets
Kernel-based SM

- **Utility**
  - Explorative Data Analysis
    - In lieu of prior knowledge
    - Interactive exploration
  - Data Visualization
    - Non-numeric data
      - E.g. criminal network investigation, plagiarism detection
    - Graphs
      - E.g. social networks, complex networks(?)
  - Manifold Learning
    - Known degrees of freedom
Experimental Results
Experimental Results

Teapot Toy Problem

- Manifold learning with the Teapot dataset
  - 560 × 420 pixels (~235K dim)
  - 100 poses (every 3.6 degrees)

Experimental Results

Teapot Toy Problem

- $H=50$ (50%)
- Geodesic distances
- Exponential kernel
Experimental Results
Tank Manifold Learning

- Manifold learning with the MSTAR dataset
  - 158 × 158 pixels (~25K dim), 1 square foot per pixel
  - 72 poses (every 5 degrees)

Experimental Results
Tank Manifold Learning

Abstract
Overview
Sammon’s Mapping (SM)
Kernel-based Sammon’s Mapping
Experimental Results
Discussion
Credits
Q&A
Experimental Results
Tank Manifold Learning

- H=18 (50%)
- Gaussian kernel
- Mod 6 weights
Experimental Results

Texture Orientation Manifold Learning

- **Shape from Texture**
  - By Mehdy Bohlool & Dr. Eraldo Ribeiro (FIT CS Dept.)

- **Goals**
  - Visualize manifold(s) of textures under different orientations (tilt and rotation)
  - Predict the orientation from position on manifold

- **Far-reaching implications**
  - Extract features from an image composed of textures, estimate surface normals and reconstruct shapes of objects in the scene
Experimental Results
Texture Orientation Manifold Learning

- Test textures
- Tilt & rotation deformations

Abstract
Overview
Sammon's Mapping (SM)
Kernel-based Sammon's Mapping
Experimental Results
Discussion
Credits
Q&A
Tilt manifold for a single texture

- $H=0.4N$, Euclidian distance (patch from center); training samples in red, test patterns in green
Experimental Results
Texture Orientation Manifold Learning

- Rotation-tilt manifold for a single texture
  - $H=0.2N$, Euclidian distance (patch from center); training samples in red, test patterns in green

![Graph showing rotation-tilt manifold for single texture](image.png)
Experimental Results

Mushroom Dataset Visualization

- Visualization of the *Mushroom* dataset
  - 22 categorical attributes
  - 2 classes: poisonous/edible

Experimental Results
Mushroom Dataset Visualization

- H=40 (20 from each class)
- Hamming kernel
Experimental Results

Voting Records Visualization

- Visualization of the *US Congressional Voting Records*
  - 16 categorical attributes
  - 2 classes: republicans/democrats

Experimental Results

Voting Records Visualization

- N=20, H=10
- Tanimoto distance
- Exponential kernel
Discussion
The KSM family of models allows for the visual representation of potentially high-dimensional data as projections into a 2- or 3-dimensional space.

The projection map strives to preserve as much fidelity in representing inter-sample dissimilarities or even similarities in the original feature space as Euclidean distances in the visualization space.

KSM subsumes the classic SM and other related models as special cases and allows for interpolation/extrapolation.
Discussion

- It also extends SM’s original idea and is able to handle data, whose attributes are not necessarily of purely numeric nature.

- Can handle both dissimilarities and similarities.

- Experimental results showcase KSM’s capabilities, in which KSM emerges as, potentially, a very useful tool for exploratory data analysis, data visualization and manifold learning.
Future Work

- Extension to Curvilinear Component Analysis
- Using KSM for Dimensionality Reduction (other than visualization)
- Model Regularization approaches
- Acceleration schemes for weight updates
- Iterative maps for kernel parameter estimation

For localized kernels, a tantalizing prospect:

\[
\mathbf{c}_h^{(k+1)} = \frac{\sum_{n=1}^{N} \zeta_n \left( \mathbf{c}_h^{(k)} \right)^T \mathbf{x}_n}{\sum_{n=1}^{N} \zeta_n \left( \mathbf{c}_h^{(k)} \right)}
\]
Credits, Q&A
Credits

- Preliminary work by Mingbo Ma (MS advisee), Ryan Gonet (Univ. of South Florida), RuiZhi Yu (Princeton Univ.) during the 2009 AMLTHEA REU program

- Partial support from the following US National Science Foundation grants:
  - No. 0647018, No. 0647120, No. 0717680 and No. 0717674.

- Cong Li (Univ. of Central Florida) for aiding with the Mushroom dataset experiments.

- Mehdy Bohlool and Dr. Eraldo Ribeiro (FIT) for the work on learning texture manifolds.
I will be happy to entertain a few quick questions from the audience.
And a “plug”...
SECES

- **SECES:** Series of Electrical & Computer Engineering Seminars

- Started last fall
- 12 talks per academic year
- Invited talks & distinguished lectures
- Partnership with IEEE Melbourne Section
- Everybody is welcome to join; we only ask you to RSVP first
May be of particular interest to you:

Evolutionary Multi-objective Optimization of ART Neural Network Architectures

by Prof. Michael Georgiopoulos, Professor of EE & CS, Department of EECS, University of Central Florida

Mon, Nov 1st 2010, 2:00 PM - 3:30 PM
Room EC276, Olin Engineering Building
SECES

- my.fit.edu/seces
- Follow us on Twitter
- @mySECES
- We are iPhone & Android phone friendly!
End of “plug”...
Thank you for attending my presentation...

These slides are available at my.fit.edu/~georgio