

UMA COMPARAÇÃO DE MÉTODOS DE REDUÇÃO DE DIMENSIONALIDADE UTILIZANDO ÍNDICES DE PRESERVAÇÃO DA TOPOLOGIA

Cláudio J. F. de Medeiros¹ e José Alfredo F. Costa^{1,2}

¹ Departamento de Engenharia Elétrica – Centro de Tecnologia
Universidade Federal do Rio Grande do Norte – 59.072-970 – Natal – RN
c.j.franca@ig.com.br; alfredo@ufrnet.br

² School of Electrical and Electronic Engineering, The University of Manchester,
Sackville Street, Manchester, UK, M60 1QD

Abstract— Efficient methods for data dimensionality reduction represented in high-dimensional spaces are important not only for visualization purposes but also, for instance, for eliminating redundant data features in pattern recognition systems. This work presents comparisons between the unsupervised methods, such as principal components analysis, Sammon mapping, auto-associative neural networks, self-organizing maps, isomap and locally linear embedding. Two indexes to measure the topological preservation are proposed. Results are presented for the methods when applied for reducing high-dimensional data into two-dimensional projections.

Keywords— Dimensionality reduction, Projections, Visualization, Data Mining, Adaptive Systems.

Resumo — Métodos eficientes de redução de dimensionalidade de dados representados em elevada dimensão são importantes, não apenas para viabilizar a visualização de dados em dimensões adequadas para a percepção humana, como também em sistemas automáticos de reconhecimento de padrões, como por exemplo, na eliminação de características redundantes. Este trabalho apresenta comparações entre os métodos não-supervisionados de redução de dimensionalidade, como análise de componentes principais, projeção de Sammon, redes auto-associativas, mapas auto-organizáveis, Isomap e LLE. Esses métodos foram aplicados à redução de dados de alta dimensionalidade para obtenção de projeções em duas dimensões. Resultados são apresentados para os métodos citados e avaliados utilizando dois índices de manutenção da ordem topológica propostos.

Palavras-chave – Redução de dimensionalidade, Projeções, Visualização, Mineração de dados, Sistemas Adaptativos.

1. Introdução

A redução da dimensionalidade é uma operação fundamental para viabilizar a visualização de dados multidimensionais em processos de descoberta do conhecimento e mineração de dados (*data mining*). Baseia-se na aplicação de transformações sobre os dados projetando-os em espaços de menor dimensão com máxima manutenção da topologia, i.e., relações de vizinhança entre os dados. Métodos de redução de dimensionalidade (RD) objetivam reduzir o espaço de características preservando ao máximo as relações topológicas dos dados [1].

Métodos de RD são utilizados para diversas finalidades, principalmente como um estágio de pré-processamento para outros algoritmos. Uma das aplicações peculiares de RD é a visualização de dados multidimensionais. Tipicamente, visualização de dados é usada em análise exploratória para descobrir indícios sobre a estrutura de conjuntos de dados.

Devido à grande diversidade de métodos atualmente disponíveis, é útil procurar estabelecer formas de comparar a sua funcionalidade em condições de trabalho semelhantes, buscando conhecer suas virtudes e deficiências em determinados tipos de aplicações. O presente trabalho procura uma comparação empírica entre métodos representativos, aplicando-os a bases de dados naturais (compostas de dados do mundo real ou de dados que os simulem). Atendendo ao critério de serem já consagrados na literatura e gozarem de amplo reconhecimento por parte da comunidade científica, foram selecionados seis métodos representativos de redução de dimensionalidade. Dentre eles, dois métodos clássicos e

bastante antigos, Análise de Componentes Principais (PCA) [2] e Projeção de Sammon [13]; dois métodos conexionistas clássicos, Rede Neural Auto-associativa [4] e Mapas Auto-organizáveis de Kohonen (SOM) [5]; e dois métodos relativamente recentes, baseados em preservação de manifold, Isomap [6] e LLE (*Locally Linear Embedding*) [7]. Cada um desses métodos é descrito sucintamente na seção 2.

O interesse do trabalho se concentra, particularmente, nas projeções em duas dimensões que podem ser visualizadas diretamente em gráficos bidimensionais. A idéia é comparar os métodos quanto à capacidade de preservação da topologia, utilizando dois índices que buscam avaliar a preservação das vizinhanças dos pontos do espaço original. Os dois índices propostos são descritos na seção 3. A seção 4 descreve os resultados dos testes utilizando os métodos selecionados para a redução de dimensionalidade de uma base de alta dimensionalidade e a seção 5 apresenta as conclusões do trabalho.

2. Métodos de redução de dimensionalidade

A **projeção de Sammon** é uma variedade de Escalonamento Multidimensional (MDS), um dos métodos tradicionais de redução de dimensionalidade. MDS pode ser definido como a busca por um conjunto de pontos num espaço de baixa dimensão no qual cada ponto representa um objeto de um espaço n -dimensional, de tal forma que as relações de distâncias entre os pontos no espaço projetado representem, o mais próximo possível, as dissimilaridades existentes entre os respectivos objetos no espaço original [3]. Para isso, o algoritmo de Sammon utiliza o método de Newton de segunda ordem para otimizar a função de custo

$$S = \frac{1}{\sum_{i < j} d_{ij}} \sum_{i < j} \frac{(d_{ij} - D_{ij})^2}{d_{ij}},$$

onde d_{ij} representa a distância entre os pontos i e j no espaço de dados original e D_{ij} , a distância entre os pontos i e j mapeados no espaço projetado.

A **Análise de Componentes Principais (PCA)** [2] é um método clássico de projeção linear, ainda largamente utilizado em redução de dimensionalidade. O método aplica uma transformação linear sobre um conjunto n -dimensional de dados de entrada e encontra um novo sistema de coordenadas de forma que a projeção de maior variância possível do conjunto de dados de entrada coincida com o primeiro eixo desse novo sistema (denominado primeiro componente principal), a de segunda maior variância, com o segundo eixo e assim sucessivamente, para os n novos eixos. PCA pode ser usado para obter uma redução de uma dimensão original n para uma dimensão reduzida m ($m < n$) selecionando-se os m primeiros componentes principais de um determinado conjunto de dados e ignorando os demais.

As **Redes Auto-associativas** (ou redes-gargalo) [4] são redes neurais artificiais, tipo MLP (*Multi-Layer Perceptron*) [12], em que são aplicadas restrições específicas de arquitetura. A utilização dessa rede na redução de dimensionalidade de dados n -dimensionais é obtida colocando-se n neurônios na camada de entrada para receberem as n variáveis de um vetor de entrada, n neurônios na camada de saída e m ($m < n$) neurônios na camada central. Normalmente existem pelo menos mais duas camadas intermediárias, com funções de ativação não lineares, entre a camada central e as camadas externas. O número m representa a dimensão reduzida desejada e o sinal de erro será a diferença entre a saída e a entrada da rede. Tal arquitetura, após o treinamento da rede, fornece um mecanismo de redução de dimensionalidade, sendo a projeção de cada ponto de entrada representada pelas saídas dos m neurônios da camada central.

Os **Mapas Auto-organizáveis de Kohonen (SOM)** definem um mapeamento de um espaço p -dimensional contínuo para um conjunto finito de vetores referência, ou neurônios, dispostos na forma de um arranjo espacial regular, normalmente bidimensional. O objetivo principal do treinamento é reduzir a dimensionalidade dos sinais ao mesmo tempo em que se busca preservar ao máximo a topologia do espaço de entrada [5]. A proximidade de pontos num mapa treinado tende a representar a similaridade dos objetos correspondentes, no espaço original.

Uma característica importante do SOM, derivada da quantização vetorial inerente ao algoritmo, é que a densidade dos neurônios em um mapa treinado é uma aproximação da densidade dos dados [16].

Assim, é possível obter informações dos agrupamentos analisando as relações geométricas dos neurônios após o treinamento. Para efeito de visualização, Ultsch [15] descreveu a *U-matrix* que contém o resultado do cálculo das distâncias entre neurônios (pesos) adjacentes no mapa treinado. A projeção do conteúdo da *U-matrix* no espaço de saída permite gerar uma imagem representando as regiões que concentram neurônios similares, separados por regiões de fronteira com baixa densidade de neurônios.

O algoritmo *Isomap*, proposto por Tenenbaum *et al* [6], pode ser visto como uma extensão do método MDS em que a dissimilaridade entre os objetos é representada pela distância geodésica, em princípio mais apropriada do que a distância euclidiana. Dado um espaço X de pontos de entrada, é determinada a matriz de distâncias $d_X(i, j)$ entre todos os pares de pontos $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$. Baseando-se nas distâncias $d_X(i, j)$ define-se a vizinhança de cada ponto \mathbf{x}_i , encontrando, por exemplo, os seus k vizinhos mais próximos. Em seguida, é construído um grafo G de todos os pontos do espaço X , por meio da interligação de cada ponto aos seus vizinhos, através de ramos de peso $d_X(i, j)$. O algoritmo estima a distância geodésica entre cada par de pontos, calculando o menor caminho $d_G(i, j)$ entre esses pontos, no grafo G . Finalmente, é aplicado o conceito clássico do algoritmo MDS para obter uma projeção cartesiana do espaço de entrada numa dimensão reduzida. Para tanto, minimiza-se a função clássica do MDS,

$$S = \frac{\sum_{i,j} (d_{ij} - D_{ij})^2}{\sum_{i,j} D_{ij}^2}$$

pelo método do gradiente, com a diferença que a distância d_{ij} entre dois pontos no espaço original, será representada pela sua distância geodésica $d_G(i, j)$.

O método *Locally Linear Embedding (LLE)*, proposto por Roweis e Saul [7], busca uma projeção global dos dados, captando as características “locais” de um *manifold*. LLE modela o *manifold* tratando-o como uma união de vários pequenos pedaços (retalhos) assumindo que os dados se encontrem sobre o *manifold* ou muito próximos dele e que esse *manifold* apresente aproximada linearidade local em todos os pontos [17]. Considerando que cada ponto $\mathbf{x}_i \in \mathbb{R}^D$ possua um número de vizinhos mais próximos indexados pelo conjunto $N(i)$ e que $\mathbf{y}_i \in \mathbb{R}^d$ seja a representação de \mathbf{x}_i na baixa dimensão, a idéia é expressar cada \mathbf{x}_i como uma combinação linear dos seus vizinhos e depois construir os \mathbf{y}_i de tal forma que eles sejam expressos pela mesma combinação linear dos seus correspondentes vizinhos, também indexados por $N(i)$.

3. Descrição dos índices utilizados

Para este trabalho, foram definidos dois índices inter-relacionados, denominados *índices de coincidência ordenada e não ordenada de vizinhanças*. Além de uma definição simples e intuitiva, esses índices refletem, de alguma maneira, o grau de manutenção da topologia no conjunto de dados de saída.

O processo de obtenção dos índices parte das duas matrizes de distâncias inter-pontos computadas para cada um dos conjuntos de dados de entrada e de saída, respectivamente. A partir de cada uma dessas matrizes obtém-se uma matriz das vizinhanças de todos os pontos. A comparação das respectivas vizinhanças dos dados de entrada e dos dados projetados permite calcular os índices, de acordo com a sua formulação.

O índice de **coincidência não ordenada de vizinhanças**, $c(t)$ é definido como:

$$c(t) = \frac{1}{nt} \sum_{i=1}^n u(i, t),$$

onde, t é um parâmetro livre que define o tamanho da vizinhança considerada em torno de cada um dos pontos i.e. o número de vizinhos mais próximos de cada ponto, seja no espaço de entrada, seja no de saída; n é o número de pontos do conjunto de dados; i é o índice que indica a posição de cada ponto dentro desse conjunto de dados; $u(i, t)$ é uma função interna que quantifica o índice para cada ponto \mathbf{x}_i no

espaço de entrada (alta dimensão) comparado ao respectivo ponto \mathbf{x}'_i no espaço de saída (baixa dimensão). A função $u(i,t)$ depende do tamanho da vizinhança e é definida para cada ponto \mathbf{x}_i como $u(i,t) = \# \{ V_t(\mathbf{x}_i) \cap V_t(\mathbf{x}'_i) \}$, sendo, $\# \{A\}$ a operação que computa a cardinalidade do conjunto A ; $V_t(\mathbf{x}_i)$, a vizinhança de tamanho t do ponto \mathbf{x}_i , o conjunto composto pelos índices dos pontos que sejam os t vizinhos mais próximos de \mathbf{x}_i no espaço de entrada; e $V_t(\mathbf{x}'_i)$ o equivalente para o ponto \mathbf{x}'_i no espaço de saída.

A interpretação do índice é simples. Para um tamanho de vizinhança t , previamente definido, o índice reflete a proporção de vizinhos dos pontos do espaço de entrada que, depois de projetados, ainda se mantêm vizinhos dos pontos correspondentes do espaço de saída, respectivamente.

Por outro lado, o **índice de coincidência ordenada de vizinhanças**, $o(t)$, leva em consideração a ordem relativa dos pontos dentro da vizinhança. Ele é definido pela mesma expressão do índice anterior, porém com a modificação da função $u(i,t)$ para refletir a exigência de manutenção do ordenamento dentro da vizinhança dos pontos de saída. Assim, teremos:

$$o(t) = \frac{1}{nt} \sum_{i=1}^n u(i, t),$$

sendo mantidas as mesmas definições anteriores para os parâmetros n , t , i e para os conjuntos $V_t(\mathbf{x}_i)$, $V_t(\mathbf{x}'_i)$. Definamos uma vizinhança restrita $V_{\perp}(\mathbf{x}_i)$ como um subconjunto dos índices contidos em $V_t(\mathbf{x}_i)$ que também fazem parte de $V_t(\mathbf{x}'_i)$ ou, de outra forma, $V_{\perp}(\mathbf{x}_i) = V_t(\mathbf{x}_i) - [V_t(\mathbf{x}_i) - V_t(\mathbf{x}'_i)]$ e $V_{\perp}(\mathbf{x}'_i) = V_t(\mathbf{x}'_i) - [V_t(\mathbf{x}'_i) - V_t(\mathbf{x}_i)]$ será o equivalente para uma vizinhança $V_t(\mathbf{x}'_i)$. Fica evidente que $V_{\perp}(\mathbf{x}_i)$ e $V_{\perp}(\mathbf{x}'_i)$ possuem os mesmos elementos (índices) porém não necessariamente dispostos na mesma ordem. A função $u(i,t)$ será então o número de correspondências perfeitas (mesmos índices nas mesmas posições) entre os dois conjuntos $V_{\perp}(\mathbf{x}_i)$ e $V_{\perp}(\mathbf{x}'_i)$. Percebe-se que este segundo índice retrata não só a presença dos mesmos vizinhos nos dois conjuntos de pontos, mas exige uma correspondência na ordem das distâncias desses vizinhos. Nesse caso, um valor igual a um refletiria uma perfeita manutenção do ordenamento topológico original.

Dificuldades particulares aparecem quando se deseja aplicar tais tipos de índices aos mapeamentos originados pelo algoritmo SOM. Sabe-se que o SOM, entre outras capacidades, é também um quantizador vetorial, o que significa que vários pontos de entrada poderão estar projetados num único ponto de saída (protótipo). Esse fato produzirá incertezas de interpretação dos fatores vizinhança e ordenamento. Portanto, é necessário estabelecer algumas convenções para os referidos fatores. Definiu-se, então, para um mapa SOM, que (1) os pontos representados no mesmo neurônio sejam tratados como vizinhos no espaço de saída e (2) pontos em tal condição, por não poderem ser distinguidos no espaço projetado, sejam considerados como corretamente ordenados. Essa é uma abordagem otimista e, portanto, o uso desses critérios tende a favorecer o SOM nos resultados comparativos com os outros métodos. Por outro lado, deve-se considerar que as convenções adotadas são coerentes com a percepção intuitiva da visualização dos dados projetados (por exemplo, numa análise visual não serão distinguíveis eventuais erros de ordenamento entre pontos representados no mesmo neurônio). Particularmente, por esses critérios, haverá uma dependência entre os valores do índice e as dimensões escolhidas para o mapa. Para tornar a escolha do tamanho do mapa menos arbitrária, optou-se por utilizar, os tamanhos *default* computados pelo pacote de software *SOM toolbox* [20]. Nesse caso, é utilizada uma fórmula heurística que, em função do conjunto dos dados de entrada, define o número de neurônios e a proporção entre os tamanhos dos lados do mapa [20].

4. Descrição dos testes e resultados

Para os testes comparativos dos métodos de redução de dimensionalidade enfocados neste trabalho foi utilizada a base de dados *Animals* disponível para pesquisas na área de aprendizagem de máquina e reconhecimento de padrões no repositório da *Universidade da Califórnia em Irvine, UCI KDD Archive* [11]. A base de dados *Animals* representa quatro classes de mamíferos por meio de 72 atributos referentes em maior parte a medidas morfológicas dos animais. Os atributos já são normalizados para a faixa de

magnitudes entre zero e um. A quantidade original de amostras era de 500.000 pontos. Por razões computacionais, uma amostra aleatória de 2.000 pontos foi extraída da base original, para utilização neste trabalho. A base *Animals* caracteriza-se pelo grande volume de amostras e por um número significativo de atributos.

Para os testes foram utilizados os pacotes de software *SOM toolbox* [20] e *DRtoolbox* [8], no ambiente *Matlab*. Para o algoritmo de Sammon foi usada a métrica euclidiana. A rede SOM utilizou geometria hexagonal e função de aprendizagem linear. Para o treinamento na fase de aproximação, foi estabelecido o limite de 1.000 épocas, um raio de vizinhança inicial de 5 unidades e valor inicial da taxa de aprendizagem igual a 0,05. Para a fase de refinamento foi utilizado um limite de 10.000 épocas, um raio de vizinhança inicial de 5 unidades e um valor inicial da taxa de aprendizagem igual a 0,01. Como os resultados do SOM dependem dos tamanhos escolhidos para o mapa, com o fim de tornar a escolha menos arbitrária, o tamanho do mapa utilizado (22 x 10) foi obtido pelo procedimento padrão mencionado na subseção 3. Para os métodos LLE e Isomap, os valores do parâmetro livre k , tamanho da vizinhança, foram definidos empiricamente. Os valores escolhidos foram $k = 100$, para o LLE, e $k = 200$, para o Isomap. A arquitetura da Rede Neural Autoassociativa utilizada foi fixada em 5 camadas. As camadas de entrada/saída contêm 72 neurônios, as camadas intermediárias contêm 7 neurônios e a camada central (gargalo) contém 2 neurônios.

Os testes compreenderam a aplicação dos índices descritos na seção 3 às projeções bidimensionais da base *Animals*, fornecidas pelos diversos métodos. Os valores dos índices foram computados para uma faixa de tamanho da vizinhança (t) variando de 1 a 50. As figuras 1 e 2 mostram os resultados dos testes por meio do gráfico que apresenta os valores dos índices em função da vizinhança t .

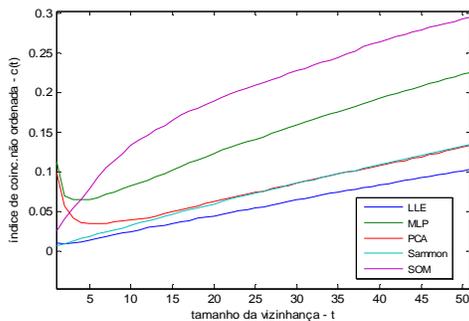


Figura1 - índice de coincidência não ordenada x tamanho da vizinhança

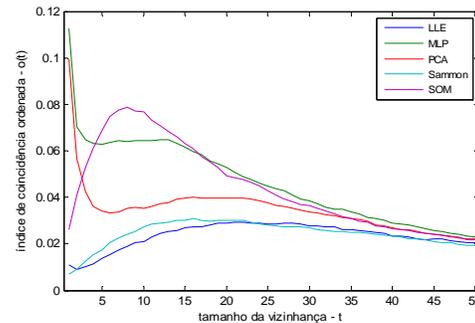


Figura 2 - índice de coincidência ordenada x tamanho da vizinhança

Deve-se observar que o método Isomap está ausente do gráfico, pois para essa base o método não teve bom desempenho, chegando a perder duas classes inteiras (ver figura 6). Diante dessa deficiência, torna-se sem sentido tentar mensurar a preservação topológica. A observação dos índices mostrados nas figuras 1 e 2 demonstra que os métodos RNAA, SOM e PCA tiveram desempenho superior. Os métodos RN-AA e PCA superaram o SOM para pequenos valores de vizinhança (entre 1 e 5 aproximadamente). No caso não ordenado o SOM os ultrapassou, a partir desses valores. No caso ordenado, o RNAA foi superior ao SOM, exceto na faixa de vizinhança entre 5 e 17, aproximadamente. O método LLE teve o pior desempenho, próximo do Sammon.

É interessante ilustrar esses resultados com as visualizações obtidas a partir dos dados projetados em duas dimensões, conforme mostram as figuras 3 a 8. Para efeito de verificação, foi acrescentada uma informação de cor para identificar as classes (conhecimento prévio). Na figura 6 pode-se constatar o problema de perda de dados, quando o Isomap é aplicado a conjuntos de dados com clusters bem definidos e descontinuidades marcantes da densidade de pontos. O algoritmo LLE (figura 7) mostrou uma visualização muito pobre, produzindo um congestionamento de todo o conjunto de dados em apenas três

pontos. Segundo van der Maaten [10] esse método tende a apresentar esse tipo de deficiência em situações de reduções drásticas de dimensionalidade, particularmente quando a dimensão final é muito baixa (como no presente caso).

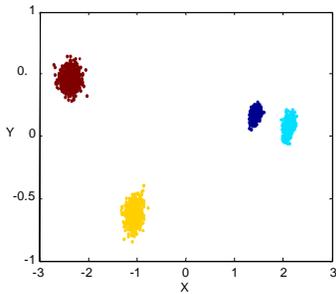


Fig. 3 - Proj. PCA - base *Animals*

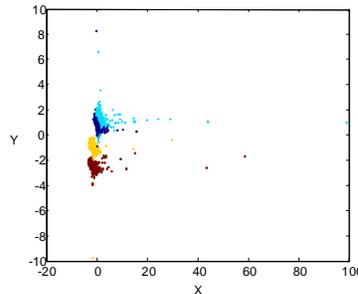


Fig. 4 - Proj. Sammon - base *Animals*

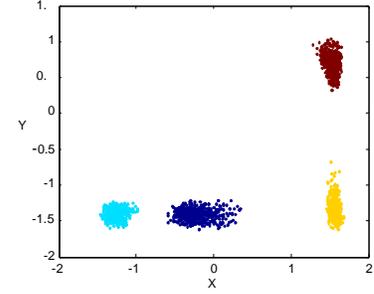


Fig. 5 - Proj. RN-AA da base *Animals*

A projeção de Sammon (figura 4) apresentou certa confusão entre as classes o que tornaria difícil a interpretação do gráfico monocromático.

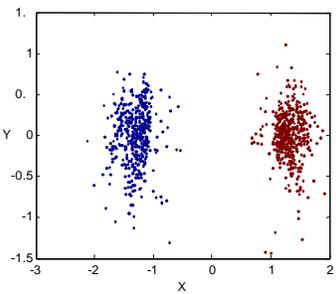


Fig. 6 - Proj. Isomap - base *Animals*

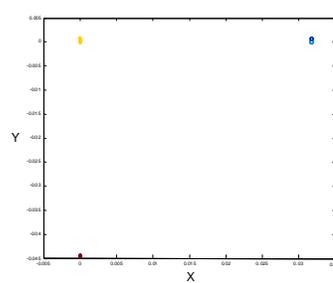


Fig. 7 - Proj. LLE - base *Animals*

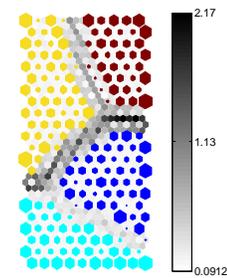


Fig. 8 - Proj. SOM - base *Animals*

As projeções do SOM, PCA e Rede Autoassociativa mostraram-se muito efetivas, distinguindo de forma inequívoca os quatro *clusters* presentes nos dados.

6. Conclusão

Neste trabalho, seis diferentes métodos de redução de dimensionalidade, incluindo alguns métodos clássicos e outros, mais recentes, foram testados com uma base de dados de dimensão elevada, reduzindo-a a duas dimensões. A avaliação do desempenho baseou-se em medidas de manutenção de topologia, quantificadas por meio de dois índices relacionados: índices de coincidência ordenada e não ordenada de vizinhanças.

Os métodos *Isomap* e LLE mostraram fragilidade nos tipos de testes realizados. Os métodos que apresentaram melhor desempenho foram o método PCA e os métodos adaptativos (conexionistas), ou seja, o Mapa Auto-organizado de Kohonen (SOM) e as Redes Neurais Autoassociativas.

Os resultados observados não determinam uma superioridade ou inferioridade absoluta de quaisquer métodos. Os testes aqui enfocados são bem definidos: projeção bidimensional para a visualização da estrutura de agrupamentos, em bases naturais com estrutura em *clusters* bem definidos (descontinuidades ou variações marcantes de densidade de pontos), contendo amostras em alta dimensão, de grandezas do mundo real. Também se deve ter em consideração que alguns critérios foram estabelecidos para permitir uma comparação razoável de elementos não perfeitamente homogêneos (mapeamentos contínuos *versus* discretos).

Futuros trabalhos envolverão pesquisas de outros índices de preservação topológica que permitam expandir o número de critérios para aferir a qualidade das projeções.

Agradecimento: Os autores agradecem ao CNPq, processos 480043/2008-6 e 201382/2008-3.

Referências:

- [1] Yang J., Ward M.O., Rundensteiner, E.A., Huang S., “Visual Hierarchical Dimension Reduction for Exploration of High Dimensional Datasets”, *Joint EUROGRAPHICS - IEEE TCVG Symposium on Visualization*, 2003.
- [2] Jolliffe, J. *Principal Component Analysis*. Springer Verlag, 1986.
- [3] Cox, T. F. and Cox, M.A.A., *Multidimensional Scaling* /2nd ed, Chapman and Hall/CRC, 2001.
- [4] Kramer, M.A. “Nonlinear principal component analysis using Auto-associative Neural Networks”, *AICHE Journal*, Vol. 37, No. 2, 1991.
- [5] Kohonen, T., *Self Organizing Maps*. 3^a. Ed. Springer-Verlag, Berlin, 2003.
- [6] Tenenbaum, J. B., Silva, V., Langford, J. C., “A Global Geometric Framework for Nonlinear Dimensionality Reduction”, *Science*, Vol 290, Pg. 2319-2323, 2000.
- [7] Roweis, S.T. and Saul, L.K., “Nonlinear Dimensionality Reduction by Locally Linear Embedding”, *Science*, Vol 290, Pg. 2323–2326, 2000.
- [8] van der Maaten, L.J.P., *An Introduction to Dimensionality Reduction Using Matlab*, Report MICC 07-07, Universiteit Maastricht, 2007.
- [9] van der Maaten, L.J.P., Postma, E.O. and van den Herik, H.J., “Dimensionality Reduction: A Comparative Review”, *Neurocomputing*, 2008 (subm.).
- [10] Yin, H. Nonlinear Dimensionality Reduction and Data Visualization: A Review, *International Journal of Automation and Computing*, N. 4 vol. 3, 294-303, 2007.
- [11] *UCI Machine Learning Repository*, Irvine, CA, University of California. [www.ics.uci.edu/~mllearn/MLRepository.html].
- [12] Haykin S., *Neural Networks, 2nd. Ed.*, Prentice Hall, 1999.
- [13] Sammon, J. W. “A Nonlinear Mapping for Data Structure Analysis”, *IEEE Transactions on Computer*, vol. C-18, no.5, pp. 401/409, 1969.
- [14] Kaski, S., Nikkilä, J. and Kohonen, T., “Methods for exploratory cluster analysis.”, *Proc. of SSGRR*, L'Aquila, 2000.
- [15] Ultsch, A., “Self-Organizing Neural Networks for Visualization and Classification.”, *Information and Classification*. Springer, Berlin, 307-313, 1993.
- [16] Gaetan, K. and Golival, J.C. “Feature extraction using auto-associative neural networks”. *Smart Mater. Struct.* **13**, 211–219 . 2004.
- [17] Saul, L. K., Roweis, S. T. “Think Globally, Fit Locally: Unsupervised Learning of Nonlinear Manifolds”. *Journal of Machine Learning Research* 4: 119-155, 2003.
- [18] Balachander T, Kothari R, Cualing H. “An empirical comparison of dimensionality reduction techniques for pattern classification”. *Artificial Neural Networks ICANN 97*. 1327: 589-594, 1997.
- [19] de Backer, S., Naud, A., Scheunders, P., “Nonlinear dimensionality reduction techniques for unsupervised feature extraction”, *P.R. Letters*. 19(8):711–720, 1998.
- [20] Vesanto, J *et al* , *Somtoolbox for Matlab*, Report A 57, Helsinki University of Technology, 2000.