



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Expert Systems with Applications xxx (2005) xxx–xxx

Expert Systems
with Applications

www.elsevier.com/locate/eswa

Comparison of dimension reduction methods using patient satisfaction data

Mevlut Ture ^{a,*}, Imran Kurt ^a, Zekeriya Akturk ^b

^a *Trakya University, Medical Faculty, Department of Biostatistics, 22030 Edirne, Turkey*

^b *Trakya University, Medical Faculty, Department of Family Medicine, 22030 Edirne, Turkey*

8 Abstract

9 In this study, we compared classical principal components analysis (PCA), generalized principal components analysis (GPCA), linear
10 principal components analysis using neural networks (PCA-NN), and non-linear principal components analysis using neural networks
11 (NLPCA-NN). Data were extracted from the patient satisfaction query with regard to the satisfaction of patients from hospital staff,
12 which was applied in 2005 at the outpatient clinics of Trakya University Medical Faculty. We found that percentages of explained var-
13 iance of principal components from PCA-NN and NLPCA-NN were highest for doctor, nurse, radiology technician, laboratory tech-
14 nician, and other staff using a patient satisfaction data set. Results show that methods using NN which have higher percentages of
15 explained variances than classical methods could be used for dimension reduction.

16 © 2005 Elsevier Ltd. All rights reserved.

17 *Keywords:* Principal components analysis; Artificial neural networks; Generalized principal components analysis; Dimension reduction; Patient satisfac-
18 tion
19

20 1. Introduction

21 High number of variables and interactions between vari-
22 ables makes it harder to interpret and summarize the
23 results as well as to apply multivariate statistics. Principle
24 Components Analysis (PCA) is a method developed to
25 remove the dependence between the variables and to
26 reduce p variables to m variables ($p > m$) with longitudinal
27 components (Ozdamar, 2004; Tatlıdil, 1996). PCA only
28 examines continuous relationships between ordinal or con-
29 tinuous variables. Generalized Principal Components
30 Analysis (GPCA), which is developed as an alternative
31 for PCA, is the optimal scaling method for data sets with
32 continuous, ordinal, and nominal variables.

33 Another method currently used for dimension reduction
34 is the neural networks (NN). Inspired by biological neural
35 networks, the NN imitates functions of the human brain

such as knowledge transfer and storage. This method can
determine the basic variables by evaluating linear and
non-linear relationships without any restriction to the type
of the variables.

Some researchers have investigated the performance of
these methods with different data sets. Dong and MacAvoy
(1996) have compared PCA and non-linear principal com-
ponents analysis using neural networks (NLPCA-NN) in
image compression. Monahan (2000) has compared the
dimension reduction performances of PCA and NLPCA-
NN using a data set related with climate. Albanis and
Batchelor (1999) have compared PCA, PCA-NN, and
NLPCA-NN in dimension reduction of financial rates in
evaluating a data set with long term credit continuity.
Hsieh (2001) has compared PCA, rotated PCA, and
NLPCA-NN using surface temperature data of the Pacific
Sea.

This study aimed to compare the PCA, GPCA, PCA-
NN, and NLPCA-NN methods in the dimension reduction
of questions examining satisfaction of the 294 patients
applying at Trakya University Medical Faculty in 2005.

* Corresponding author. Tel.: +90 284 2357641x1631; fax: +90 284 2357652.

E-mail address: ture@trakya.edu.tr (M. Ture).

57 **2. Material and methods**58 *2.1. Data*

59 Our study included 294 consecutive patients (46.3%
60 males, 53.7% females) admitted to the outpatient clinics
61 of Trakya University Medical Faculty on 12 January
62 2005. A patient satisfaction questionnaire consisting of 31
63 items with a 5-point Likert scale (1—very bad, 2—bad,
64 3—average, 4—good, 5—excellent) evaluating the satisfac-
65 tion of patients from the hospital staff was applied to the
66 patients.

67 The reliability coefficient of the questionnaire (Cron-
68 bach α) was 0.96. The 31 items were subdivided as satisfac-
69 tion from the doctor, nurse, radiology technician,
70 laboratory technician, and other staff. The reliability coef-
71 ficients for doctor, nurse, radiology technician, laboratory
72 technician, and other staff were 0.96, 0.93, 0.91, 0.93, and
73 0.84 respectively (Table 1).

2.2. *Principal components analysis*

74

Essentially, PCA maximizes the correlation between the
original variables to form new variables that are mutually
orthogonal, or uncorrelated (A project funded by the Tsu-
nami Initiative, 1999; Ozdamar, 2004). PCA was used to
describe the variance in data sets of n observations on p
variables (Jolliffe, 1986; Manly & Bryan, 1986). PCA is
a statistical technique that linearly transforms the original
set of variables into a substantially smaller set of uncorre-
lated variables that represents the maximum amount of
information in the original set of variables. A small set of
uncorrelated variables is much easier to understand and
use in further analyses than a larger set of correlated
variables.

Principal components are able to describe different
dimensions of a given $n \times p$ data set of n observations on
 p variables. The first principal component (Y_1) can be
defined as a linear combination of the elements of the data
matrix, \mathbf{X} :

75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92

Table 1
Reliability coefficients of the different groups

Question	Cronbach α
<i>Items related with satisfaction from the doctor</i>	
1. Giving you enough time during the consultation	0.96
2. Facilitating you to explain you problems	
3. Involving you in clinical decisions related with your care	
4. Listening to you	
5. Having fast relief for your complaints	
6. Examining you	
7. Explaining the aims of tests ant treatments	
8. Responding to your information requests	
9. Helping you to deal with your anxieties	
10. Helping to become aware of the importance of his suggestions	
11. Remembering what he/she said and did during the previous encounters	
12. His/her respect to you	
13. The interest/listening of the doctor	
14. Cheerfulness of the doctor	
15. Giving you information	
<i>Items related with the satisfaction from the nurse</i>	
1. Respect to you	0.93
2. Interest/listening to you	
3. Cheerfulness of the nurse	
4. Giving you information	
<i>Items related with satisfaction from the radiology technician</i>	
1. Respect of the radiology technician	0.91
2. Interest/listening to you	
3. Cheerfulness of the radiology technician	
4. Giving you information	
<i>Items related with satisfaction from the laboratory technician</i>	
1. Respect of the laboratory technician	0.93
2. Interest/listening of the laboratory technician	
3. Cheerfulness of the laboratory technician	
4. Giving you information	
<i>Items related with satisfaction from other staff</i>	
1. Interest and listening of the registration desk staff to you	0.84
2. Cheerfulness of the registration staff	
3. Respect of the registration staff to you	
4. Attitudes of the security staff	

$$94 \quad Y_1 = e_{11}X_1 + e_{12}X_2 + \dots + e_{1p}X_p$$

95 where coefficients that chosen so as to maximize the vari-
96 ance represented by the first principal component are sim-
97 ply the eigenvectors of the symmetric covariance matrix.

98 The eigenvalues of the covariance matrix represent the
99 variation of each principal component, where

$$101 \quad \text{Var}(Y_i) = \lambda_i$$

102 Ideally, a principal component analysis will yield several
103 components that describe the majority of the total varia-
104 tion of the data set. Geometrically, the first PC is the line
105 of closest fit to the n observations. It minimizes the sum
106 of the squared distances of the n observations from the line
107 where the distance is defined in a direction perpendicular to
108 the line. The second PC a line of closest fit to the residuals
109 from the first PC, the third PC is a line of closest fit to the
110 residuals from the second PC, and so on (Ozdamar, 2004;
111 Sharma, 1996; Tatlidil, 1996).

112 2.3. Generalized principal components analysis

113 The GPCA procedure quantifies categorical variables
114 using optimal scaling, resulting in optimal principal com-
115 ponents for the transformed variables. The variables can
116 be given mixed optimal scaling levels and no distributional
117 assumptions about the variables are made (Gifi, 1990;
118 Michailidis & De Leeuw, 1998, 2000; SPSS Inc., 1999).

119 In GPCA, dimensions correspond to components, and
120 object scores correspond to component scores (Gifi, 1990;
121 SPSS Inc., 1999).

122 The GPCA objective is to find object scores X and a set
123 of \underline{Y}_j (for $j = 1, 2, \dots, m$) so that the function

$$\sigma(\mathbf{X}; \underline{\mathbf{Y}}) = n_w^{-1} \sum_j c^{-1} \text{tr}((\mathbf{X} - \mathbf{G}_j \underline{\mathbf{Y}}_j)' \mathbf{M}_j \mathbf{W} (\mathbf{X} - \mathbf{G}_j \underline{\mathbf{Y}}_j))$$

125 with c is p if $j \in J$,

126 is minimal, under the normalization restriction
127 $\mathbf{X}' \mathbf{M}_* \mathbf{W} \mathbf{X} = n_w m_w \mathbf{I}$ (\mathbf{I} is the $p \times p$ identity matrix). The
128 inclusion of \mathbf{M}_j in $\sigma(\mathbf{X}; \underline{\mathbf{Y}})$ ensures that there is no influence
129 of passive missing values. \mathbf{M}_* contains the number of active
130 data values for each object. The object scores are also cen-
131 tered. $\underline{\mathbf{Y}}$ is that collection of category quantifications for
132 variables with multiple nominal scaling level, and vector
133 coordinates for non-multiple scaling level. \mathbf{G}_j is indicator
134 matrix for variable j , of order $n_{\text{total}} \times k_j$. n_w is the weighted
135 number of analysis cases, m_w is the weighted number of
136 analysis variables, and \mathbf{W} is diagonal $n_{\text{total}} \times n_{\text{total}}$ matrix,
137 with w_i on the diagonal (Gifi, 1990; Michailidis & De
138 Leeuw, 1998, 2000).

139 2.4. Linear principal components analysis using neural 140 networks

141 PCA-NN is mainly used for classification and feature
142 extraction. The goal of PCA is to find a set of orthogonal
143 components that minimize the error in the reconstructed

144 data. An equivalent formulation of PCA is to find an
145 orthogonal set of vectors that maximize the variance of
146 the projected data (Diamantras & Kung, 1996).

147 Sanger proved that one-layered linear neural network is
148 equivalent to the linear standard PCA. And the neural net-
149 works which implement this learning algorithm is called
150 PCA-NN. We are assuming that the network has m out-
151 puts, each given by

$$y_j(n) = \sum_{i=1}^p w_{ij}(n) x_i(n), \quad j = 1, 2, \dots, m \quad 153$$

154 and p inputs ($m < p$). To apply Sanger's rule the weights
155 ($w_{ij}(n)$) are updated according to

$$\Delta w_{ji}(n) = \eta [y_j(n) x_i(n) - y_j(n) \sum_{k=1}^j w_{ki}(n) y_k(n)],$$

$$i = 1, 2, \dots, p \quad 157$$

158 where η is the step size. In this rule, the input to each neu-
159 ron is modified by subtracting the product of the outputs
160 from the preceding neurons and the respective weights.
161 This implements the deflation method after the system con-
162 verges. That is, after convergence of the first neuron
163 weights will the second neuron weights converge com-
164 pletely to the eigenvector that corresponds to the second
165 largest eigenvalue (Albanis & Batchelor, 1999; Hassoun,
166 1995; Principe, Euliano, & Lefebvre, 2000).

167 2.5. Non-linear principal components analysis using neural 168 networks

169 The NLPCA-NN is a general purpose feature extraction
170 algorithm producing features that retain the maximum
171 possible amount of information from the original data
172 set. If non-linear correlations between variables exist and
173 sufficient data to support the formulation between more
174 complex mapping functions are available, then NLPCA
175 will describe the data with greater accuracy than PCA
176 (Albanis & Batchelor, 1999).

177 Kramer presented a NLPCA-NN method based on
178 autoassociative neural networks that are trained by back-
179 propagation (Krammer, 1991). NLPCA-NN uses five-layer
180 feed-forward network with a bottleneck layer of nodes to
181 reduce the dimension of the input variables and each layer
182 fully connected to the next (Krammer, 1991; Oja, 1992).
183 The second and fourth layers of the network have sigmoidal
184 activation functions, so layers 1–3 and layers 3–5 model
185 non-linear functions. The activation functions of the third
186 and fifth layers are linear. The input (first) and output
187 (fifth) layers have p units (the number of variables in the
188 data set). The third layer has fewer nodes ($m < p$) than
189 the first or fifth. The values of the output nodes in layer
190 5 are trained to approximate the inputs. After the network
191 has been trained, bottleneck node activation values in layer
192 3 give a lower dimensional representation of the inputs
193 (Fotheringham & Baddeley, 1997; Monahan, 2000).

194 The goal of the network is to minimize the error term
 195 (e). The network is then trained to try and reproduce the
 196 input pattern at the output layer by using an error term
 197 which is simply the squared difference between the network
 198 prediction (X_i) and input pattern (X'_i);

$$200 \quad e = \sum_{i=1}^p (X_i - X'_i)^2 \quad (i = 1, 2, \dots, p)$$

201 NLPCA-NN reduces the dimension of the inputs by fitting
 202 a curve through the data. The first three layers of the net-
 203 work project the original data onto the curve and the acti-
 204 vation values of the bottleneck layer, called scores, give the
 205 location of the projection. The last three layers define the
 206 curve (Albanis & Batchelor, 1999; Daszykowski, Walczak,
 207 & Massart, 2003; Hsieh, 2001; Michailidis & De Leeuw,
 208 2000).

209 2.6. Package programs

210 In this study, data were analyzed using SPSS 10.5 (PCA,
 211 GPCA, Hierarchical Cluster Analysis) and NeuroSolutions
 212 5.0 (PCA-NN, NLPCA-NN).

213 3. Results

214 The whole data set was used for PCA and GPCA.
 215 Before building NN, the data set was randomly split into
 216 two parts: 80% ($n = 235$) of the data for a training set
 217 and 20% ($n = 59$) for a cross validation set.

218 In Table 2, we present the percentages of variance
 219 explained by dimension reduction methods for a principal
 220 component which represented items of each group. The
 221 minimum–maximum levels for percentage of variance
 222 explained with PCA, GPCA*, GPCA+, PCA-NN and
 223 NLPCA-NN were 62.2–82.1%, 60.2–82.1%, 63.4–84.0%,
 224 84.9–91.7%, and 84.9–96.1% respectively (Table 2).

225 As it can be seen from Table 2 and Fig. 1, PCA-NN and
 226 NLPCA-NN had the highest percentages of variance
 227 explained for doctor, nurse, radiology technician, labora-
 228 tory technician, and other staff. Ordinal GPCA performed
 229 better than numeric GPCA and PCA.

230 Percentages of variance explained were used as input
 231 variables in the Hierarchical Cluster Analysis (HCA)
 232 (Ture, Kurt, Kurum, & Ozdamar, 2005). HCA was done

to identify homogenous groups of dimensionality reduction 233
 techniques based on percentages of variance explained. The 234
 dendrogram from centroid clustering method that was 235
 obtained is shown in Fig. 2. In the dendrogram, the data 236
 points appear to cluster in two groups. The first cluster 237
 includes PCA, GPCA*, and GPCA+. The second cluster 238
 includes PCA-NN and NLPCA-NN. We found that both 239
 PCA-NN and NLPCA-NN explain a higher amount of 240
 variation in the original set of variables than other 241
 techniques. 242

4. Discussion 243

Frequently the investigated issues are under the effect of 244
 multiple variables and thus it is mandatory to evaluate the 245
 effecting variables together in order to ascertain reliability 246
 and validity. However, having multiple variables and rela- 247
 tionship between variables makes data analysis more diffi- 248
 cult. Dimension reduction methods are used frequently to 249
 reduce the data into m dimensions instead of working in 250
 a p dimension environment ($p > m$). Since PCA is a method 251
 developed for determining linear relationships, it does not 252
 consider non-linear relationships. Therefore it remains 253
 insufficient in determining representative variables in the 254
 data set in case of non-linear relationships. 255

Albanis and Batchelor (1999) have demonstrated that 256
 the PCA-NN and NLPCA-NN are superior to PCA in 257
 dimension reduction using the long term credit continuity 258
 data set. Dong and MacAvoy (1996) have shown that 259
 NLPCA-NN is better than PCA in image compression. 260
 Monahan (2000) has used a data set related with climate 261
 and compared the dimension reduction performances of 262
 PCA and NLPCA-NN where he has shown that 263
 NLPCA-NN is a more superior method in doing this. 264
 Hsieh (2001) compared PCA, rotated PCA, and NLPCA- 265
 NN using surface temperature data set of the Pacific Sea 266
 where he demonstrated a better performance of NLPCA- 267
 NN. In our study using the patient satisfaction data set, 268
 the highest variance explanation rates for variables deter- 269
 mined for doctor, nurse, radiology technician, laboratory 270
 technician, and other staff was with the PCA-NN and 271
 NLPCA-NN methods. Ordinal scaled GPCA method 272
 showed a slightly higher performance than the proportion- 273
 ally scaled GPCA and PCA methods. 274

Table 2
 Percentage of variance explained after dimensionality reduction

Principal component	Variable	Percentage of variance explained (%)				
		PCA	GPCA*	GPCA+	PCA-NN	NLPCA-NN
Doctor	15	62.2	60.2	63.4	84.9	84.9
Nurse	4	82.0	80.5	82.7	88.4	88.7
Radiology technician	4	79.4	79.2	84.0	91.7	96.1
Laboratory technician	4	82.1	82.1	83.8	90.9	90.9
Other staff	4	67.6	66.5	69.8	85.9	86.1

* Variables were determined as continuous.

+ Variables were determined as ordinal.

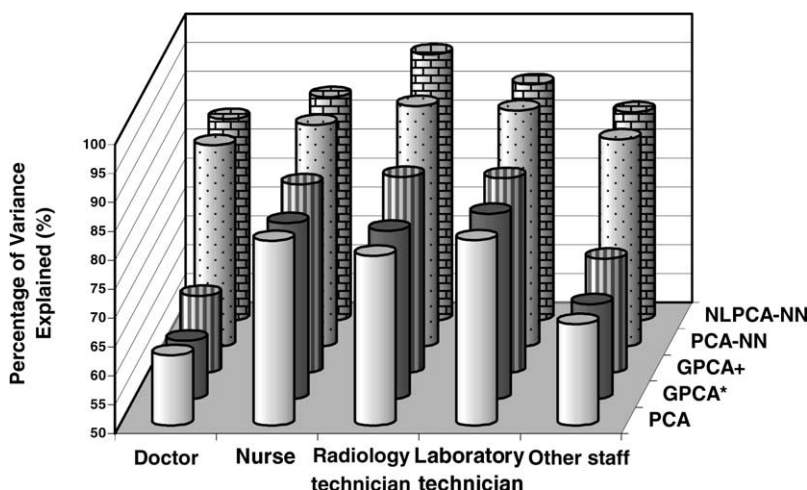


Fig. 1. Percentages of variance explained according to principal component of methods.

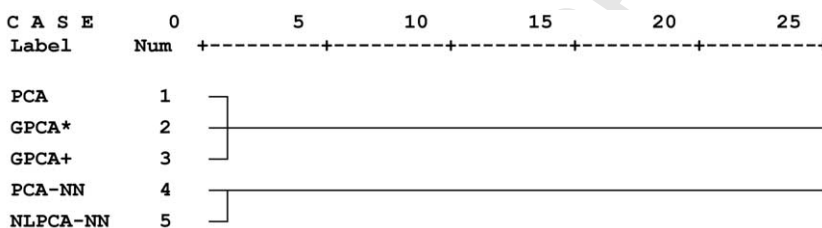


Fig. 2. Dendrogram showing relationship among dimension reduction methods.

275 It can be concluded from this study that instead of
 276 working with variables of lower explanatory rates, methods
 277 containing NN should be implemented in dimension reduc-
 278 tion due to its advantage of containing alternative methods
 279 in contrast to many classical methods. It should be kept in
 280 mind that NN has the ability to determine the best vari-
 281 ables due to its advantage of considering the non-linear
 282 relationships along with the linear relationships.

283 **References**

284 Albanis, G. T., & Batchelor, R. A. (1999). Assessing the long-term credit
 285 standing using dimensionality reduction techniques based on neural
 286 networks—an alternative to overfitting. In *The proceedings of the SCI*
 287 *99/ISAS 99 conference, Orlando, US*.
 288 A project funded by the Tsunami Initiative (1999). Investigation of the risk
 289 from climate variability and change over Northern Europe. Available
 290 from <http://www.nerc-bas.ac.uk/public/tsunami/>.
 291 Daszykowski, M., Walczak, B., & Massart, D. L. (2003). A journey into
 292 low-dimensional spaces with autoassociative neural networks. *Talanta*,
 293 *59*, 1095–1105.
 294 Diamantras, K. I., & Kung, S. Y. (1996). *Principal component neural*
 295 *networks*. New York: John Wiley and Sons.
 296 Dong, D., & MacAvoy, T. J. (1996). Batch tracking via nonlinear
 297 principal component analysis. *AIChE Journal*, *42*(8), 2199–2208.
 298 Fotheringham, D., & Baddeley, R. (1997). Nonlinear principal compo-
 299 nents analysis of neuronal spike train data. *Biological Cybernetics*,
 300 *77*(4), 283–288.
 301 Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester: Wiley.

Hassoun, M. H. (1995). *Fundamentals of artificial neural networks*. 302
 Cambridge, MA: MIT Press. 303
 Hsieh, W. W. (2001). Nonlinear principal component analysis by neural 304
 networks. *Tellus*, *53A*, 599–615. 305
 Jolliffe, I. T. (1986). *Principal component analysis*. New York: Springer- 306
 Verlag. 307
 Krammer, A. M. (1991). Non-linear principal component analysis using 308
 autoassociative neural networks. *AIChE Journal*, *37*(21), 223–243. 309
 Manly, B. F., & Bryan, F. J. (1986). *Multivariate statistical methods: A* 310
primer. London: Chapman and Hall. 311
 Michailidis, G., & De Leeuw, J. (1998). The Gifi system of descriptive 312
 multivariate analysis. *Statistical Science*, *13*(4), 307–336. 313
 Michailidis, G., & De Leeuw, J. (2000). Multilevel homogeneity analysis 314
 with differential weighting. *Computational Statistics & Data Analysis*, 315
32(3–4), 411–442. 316
 Monahan, A. H. (2000). Nonlinear principal component analysis by 317
 neural networks: Theory and applications to the Lorenz system. 318
Journal of Climate, *13*, 821–835. 319
 Oja, E. (1992). Principal components, minor components, and linear 320
 neural networks. *Neural Networks*, *5*, 927–935. 321
 Ozdamar, K. (2004). *Paket programlar ile istatistiksel veri analizi-2*. 322
 Eskişehir: Kaan Kitabevi. 323
 Principe, J. C., Euliano, N. R., & Lefebvre, W. C. (2000). *Neural and* 324
adaptive systems: Fundamentals through systems. New York: John 325
 Wiley & Sons. 326
 SPSS Inc. (1999). *SPSS Categories 10.0*. Chicago: SPSS Inc. 327
 Tatlıdil, H. (1996). *Uygulamalı çok değişkenli istatistiksel analiz*. Ankara: 328
 Akademi Matbaası. 329
 Ture, M., Kurt, I., Kurum, A. T., & Ozdamar, K. (2005). Comparing 330
 classification techniques for predicting essential hypertension. *Expert* 331
Systems with Applications, *29*(3), 583–588. 332
 333