pattern recognition, multispectral analysis,
dimensionality reduction, feature selection

Marcin MICHALAK[1], Adam ŚWITOŃSKI[1,2]

LOCAL EMBEDDING AND DIMENSIONALITY REDUCTION IN DETECTION OF SKIN TUMOR TISSUE

This article shows the limitation of the usage of dimensionality reduction methods. For this purpose three algorithms were analyzed on the real medical data. This data are multispectral images of human skin labeled as tumor or non-tumor regions. The classification of new data required the special algorithm of new data mapping that is also described in the paper. Unfortunately, the final conclusion is that this kind of local embedding algorithms should not be recommended for this kind of analysis and prediction.

1. INTRODUCTION

The analysis of big sets of data may become difficult with the increase of the amount of the data. This problem may be considered in two levels: the level of the number of objects or in the level of the number of attributes (features etc.). The first kind of problem is just a computational one: if the current hardware (or software) can not build the model then the faster machine with bigger resources (like the amount of memory or the number of processors) or newer version of software should be used. However, the second kind of this problem is more complicated. Even if we use the fastest computer in the world with the infinite memory it may occur that the chosen algorithm will not build the model.

If we consider a typical parametric model of regression - the polynomial one - we see that for the purpose of building the model of the dependence that is the polynomial of the n^{th} order we have to know at least n+1 points. Building more advanced models where there are more than one independent variables also requires bigger number of training objects. Situation is even more difficult when the nonparametric models are built. It is common criterion that we should have at least 10 objects for each parameter of the model. It may also occur that the number of model parameters is much bigger than the number of object features. It may be easily observed when the artificial neural networks are taken into consideration: the complexity of the model depends on the network configuration (the number of layers and the number of neurons in the hidden layer) and not only on the number of independent variables.

The solution of the problem is commonly known as the dimensionality reduction. There are lot of methods that may be used like principal component analysis (PCA) [3, 4, 7], multidimensional scaling (MDS) [1, 5] statistical rankings [6] or heuristic algorithms [6]. They may be generally divided into two groups: feature selection and feature extraction. Algorithms from the first group find the subset of original features (statistical rankings or heuristic algorithms). The main advantage of this group is that the result may simplify the process of data acquisition. Algorithms from the second group generate the new set of features on the basis of the whole set of original ones. This means that usually the model will still require the whole set of original variables.

[1] Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland,
email: {Marcin.Michalak, Adam.Switonski}@polsl.pl.
[2] Polish-Japanese Institute of Information Technology, ul. Koszykowa 86, 02-008 Warszawa, Poland.

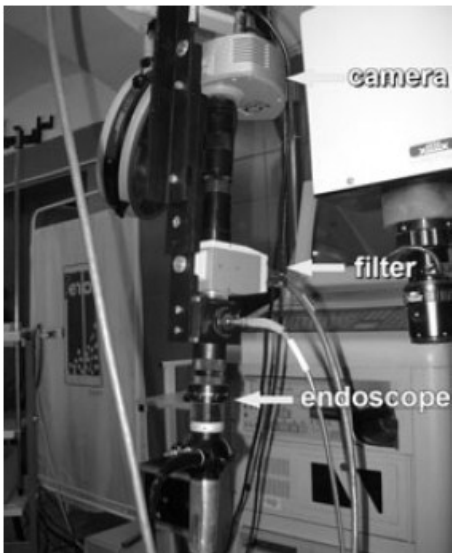

Fig. 1. Acquisition device.

In this paper the problem of feature extraction for the purpose of medical hyperspectral data classification is raised. On the basis of the multispectral images human skin cells are labeled as tumor or non-tumor. The assignment was prepared by medical experts. For the purpose of feature extraction we used three algorithms of nonlinear dimensionality reduction called local embedding.

This paper is organized as follows: it starts with the description of our previous works on hyperspectral images. Then algorithms of dimensional reduction are briefly described. Afterwards the result of experiments are presented and discussed.

2. ANALYSIS BACKGROUND

Medicine is one of the most popular domain of image analysis application. Our research focused on analysis of multispectral images of human skin and the ability of classifying skin regions as tumored or non-tumored. Data were captured with the usage of the device shown on the Fig. 1. It contained camera, liquid crystal filter and the endoscope. Skin fragments were lighted twice: with white and blue light. White light was chosen to observe which components of the visible light spectrum are absorbed by different regions. Blue light was chosen as it was expected that tumor regions should absorb this band of wave more strongly.

Experts marked regions with tumors and then on the basis of the 21 channel multispectral images further analysis was performed. From the various algorithms applied for the data named as "White" and "Blue" it occurred that almost all of them give better results when the "White" data are analyzed. It also occurred that the best algorithm is the artificial neural network (ANN) that gave almost 95% classification accuracy. Details of the experiment may be found in [5]. The visualization of the classification result is shown on the Fig. 2.

As the ANN occurred to be the best algorithm it became interesting to examine whether all 21 channels are essential for the purpose of human skin diagnosis. In the paper [3] several methods of feature selection were applied and the climbing strategy was approved as the best: it needed only 6 components to achieve the comparable ability of correct classification.

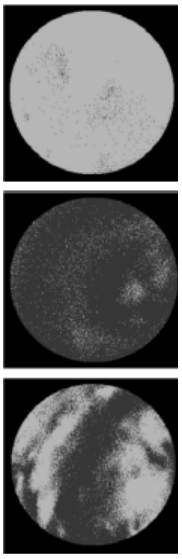

Fig. 2. Sample results visualization (red – tumor, green – healthy).

3. ALGORITHMS OF DIMENSIONALITY REDUCTION

3.1. CLASSICAL MULTIDIMENSIONAL SCALING

Classical Multidimensional Scaling is the method of dimensionality reduction implemented in the Matlab software (`mdscale`). This algorithm performs scaling but in the case when the Euclidean distance is used it also returns the eigenvalues *e* of *YY*' where *Y* is the matrix of transformed coordinates. In general case of *n*-dimensional original data, when $k < n$ elements of *e* are greater than 0 then the *k* first components of *Y* should be considered as the *k*-dimensional approximation of the whole *n*-dimensional space. For more details the Matlab documentation refers to [4].

3.2. NONCLASSICAL MULTIDIMENSIONAL SCALING

Nonclassical Multidimensional Scaling performs non-metric multidimensional scaling on the *N* X *N* dissimilarity matrix (matrix of objects distances). It is possible to define the maximal number of coordinates in the output matrix. The Matlab documentations refers also to [4] but points out two other books as the origin of the algorithm [1, 2].

3.3. NONLINEAR DIMENSIONALITY REDUCTION

In the paper [6] the method of nonlinear dimensionality reduction is described. It is based on analyzing the connection between points considered as the connection in the graph. Generally, the algorithm consists of three steps. The first step is building the graph *G* over all data points. Two points *i*,*j* are connected iff one of the two conditions is fulfilled: they are closer than predefined margin ε (this model is called ε-Isomap) or if the point *i* is one of the *k* nearest neighbors of *j* (the model called *k*-Isomap). The weight of the edge (*i*,*j*) is the distance in the original feature space.

In the second step the shortest path between points is computed. The default algorithm in authors' Matlab Isomap implementation is the Dijkstra's algorithm but it is also possible to use the Floyd's one.

The last step is constructing of *d*-dimensional embedding. If D_G is the matrix of shortest path between points (vertexes) in the graph, the λ_p is the *p*-th eigenvalue (in decreasing order) of the matrix $\tau(D_G)$ and v_p^i is the *i*-th component of the *p*-th eigenvector then the *p*-th component in the *d*-

dimensional coordinate vector y_i equals $\sqrt{\lambda_p} v_p^i$. The operator τ is defined as $\tau(D) = -HSH/2$, where S is the matrix of squared distances ($S_{ij} = D_{ij}^2$) and H is called the "centering matrix" $H_{ij} = \delta_{ij} - 1/N$.

4. EXPERIMENTS AND RESULTS

4.1. DATA DESCRIPTION

The data contain 1250 objects from two classes. Each of the classes has comparable number of objects. This set is the subset of objects that were analyzed in the previous paper [3] due to the complexity of the analysed algorithms. The data are divided into three sets: train (1000 objects), tune (125 objects) and test (125 objects). Classes in every set are balanced in the same way as in the whole data set. Experiments of dimensionality reduction were performed with Matlab. The process of training artificial neural networks was performed with the Statistica Automated Neural Network (SANN) tool.

4.2. THE TRAIN TEST MODEL

Descriptions of dimensionality reduction techniques, presented in the previous point, lead to the conclusion that there are no mathematical formula for data transformation. In other words the reduction of dimensionality of the whole train set would not provide the way of test data transformation. This creates the situation when for every test point the analogical dimensionality reduction must be performed. The algorithm of transformation of the test data into the low dimensionality space is presented below.

Let us denote the train test as *Tr*, test set as *Tt*. The reduced train set is denoted as ρ(*Tr*) and the reduced test set is denoted as ρ (*Tt*).

After reduction of train test *Tr* dimensionality for every test object t_o:

1. Add test object to the iteration set $T_i = T_r \cup \{ t_o \}$.
2. Find the *d*-dimensional coordinates of points from T_i
3. Move the new *d*-dimensional coordinates of t_o to the ρ(T_t).

The pseudocode of generating the low dimensional coordinates for test objects is shown below. The `ALGORITHM` means one of the method of dimensionality reduction.

function REDUCE_TEST(T_t, T_r)
$\rho(T_r) \leftarrow \varnothing$
$n \leftarrow \overline{\overline{T_r}}$
for $i = 1 \rightarrow n$
  $t_o \leftarrow T_r(i)$
  $T_i \leftarrow T_t \cup \{t_o\}$
  $\rho(T_i) \leftarrow$ ALGORITHM(T_i)
  $\rho(T_r) \leftarrow \rho(T_r) \cup \rho(T_i)(i)$
  $i \leftarrow i + 1$
end for
return $\rho(T_r)$
end function

where $\rho(T_i)(i)$ means the coordinates of the *i*-th point from the test set in the dimensionality reduced space.

4.3. RESULTS

For every dimensionality reduction algorithm the maximal number of new coordinates was generated (21). Then for every k first components (k = 1, 2, …, 21) the artificial neural network was built. Figures 3-5 show how the classification accuracy on the train, tune and test set changes due to the increase of the considered coordinates. On the first figure (Fig. 2.) the results for classical multidimensional scaling are presented. Generally for every considered number of components the accuracy of test data classification does not change and remains on the level 55%. It is worth to remind that classes were balanced so this results can be interpreted as the results of ,,random classifier".

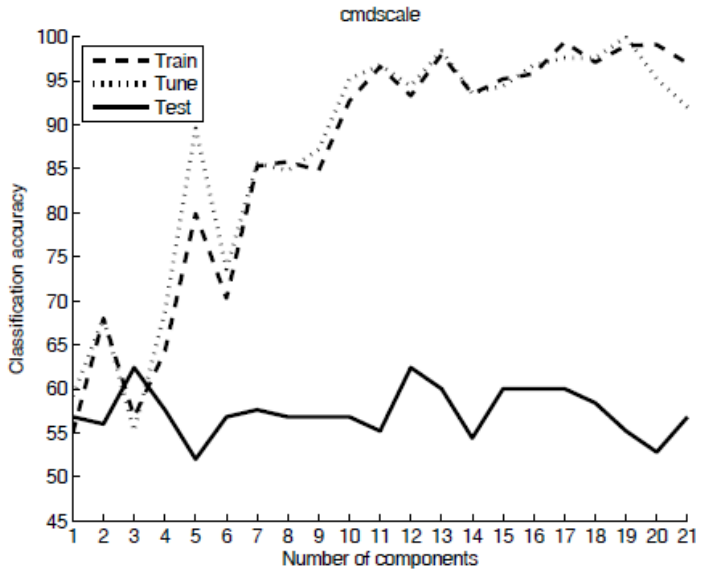

Fig. 3. Graphical presentations of train, tune and test set classification accuracy for cmdscale algorithm.

On the Fig. 4. the results of mdscale algorithm are presented. In this case for several small numbers of components (4, maximally 5) we observe that, indeed, the increase of number of new coordinates improves the classification accuracy. It can even give better results on the test set than on the train or tune set. However, the further increase (six components and more) of the number of components reduces the classifier ability of prediction. It also occurs that the decrease of accuracy does not achieve as low level as in the case of cmdscale algorithm.

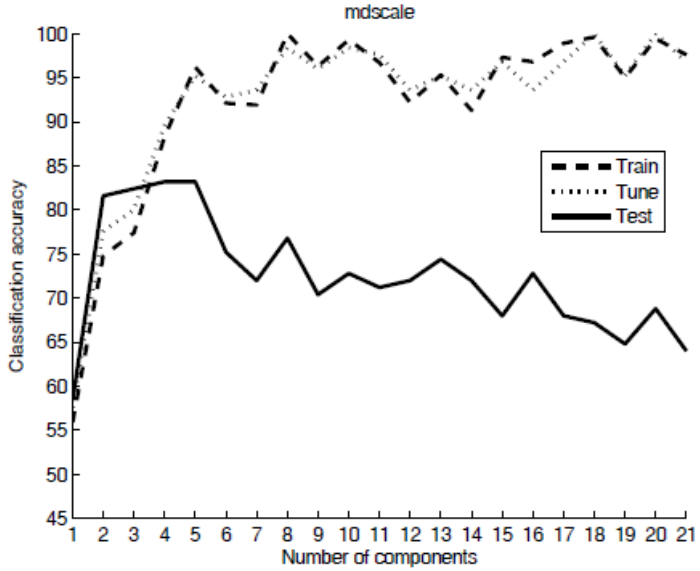

Fig. 4. Graphical presentations of train, tune and test set classification accuracy for mdscale algorithm.

The next figure (Fig. 5.) shows the results of the last analyzed method -- the Isomap. This results are comparable to the results of cmdscale. The accuracy oscillates around the level of 58% what is almost the same as cmdscale 57%.

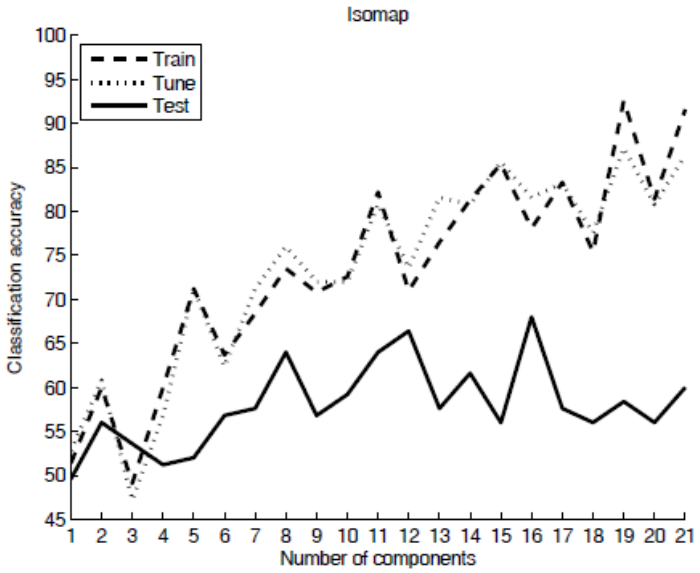

Fig. 5. Graphical presentations of train, tune and test set classification accuracy for Isomap algorithm.

Statistical comparison of each method are shown in the Table 1. In this table the minimal and maximal classification accuracy on the test set are presented, as well as the average accuracy and the standard deviation of the accuracy.

Table 1. Statistical comparison of the classification accuracy (in percent) of three models of dimensionality reduction.

	cmdscale	mdscale	Isomap
min	52.0	58.4	49.6
max	62.4	83.2	68.0
avg	57.37	72.34	58.02
std	2.68	6.39	4.64

More detailed results of algorithms performance are presented in the Table 2. For each algorithm three columns are shown: the error on the train, tune and the test set.

Figures from 4 to 5 are the illustration of the data from the Tab. 2.

5. CONCLUSIONS

In this paper three methods of dimensionality reduction were analyzed due to the aim of tumor tissue detection. On the basis of the previous experiments it was expected that it is possible to limit the number of components without the lose of classification accuracy. After experiments it occured that none of popular methods called local embedding may be applied successfully. One of the most important problems is that for those method there are no explicitly given formula of transformation new data into reduced dimensionality. The only possible way is the algorithm described in the section 4.2: for each new object the new local embedding should be performed and the previously built model should be applied for the transformed components of the test object. General and detailed results show that it is possible to built the accurate model on the known data (accuracy of classification of the tune set reaches the maximal level for cmdscale and mdscale algorithm) but the model would not give satisfactory results for new data.

Table. 2 Results of classification accuracy (in percent) of mapped data.

cmdscale			
#	train	tune	test
1	55.0	59.2	56.8
2	68.0	68.0	56.0
3	56.8	55.2	62.4
4	64.3	68.8	57.6
5	79.9	89.6	52.0
6	70.3	73.6	56.8
7	85.3	85.6	57.6
8	85.8	84.8	56.8
9	84.8	87.2	56.8
10	92.7	95.2	56.8
11	96.6	96.8	55.2
12	93.3	94.4	62.4
13	98.0	98.4	60.0
14	93.5	93.6	54.4
15	95.2	94.4	60.0
16	95.8	96.8	60.0
17	99.4	97.6	60.0
18	97.1	97.6	58.4
19	99.0	100.	55.2
20	99.1	95.2	52.8
21	97.0	92.0	56.8

mdscale			
#	train	tune	test
1	55.0	59.2	56.8
2	68.0	68.0	56.0
3	56.8	55.2	62.4
4	64.3	68.8	57.6
5	79.9	89.6	52.0
6	70.3	73.6	56.8
7	85.3	85.6	57.6
8	85.8	84.8	56.8
9	84.8	87.2	56.8
10	92.7	95.2	56.8
11	96.6	96.8	55.2
12	93.3	94.4	62.4
13	98.0	98.4	60.0
14	93.5	93.6	54.4
15	95.2	94.4	60.0
16	95.8	96.8	60.0
17	99.4	97.6	60.0
18	97.1	97.6	58.4
19	99.0	100.	55.2
20	99.1	95.2	52.8
21	97.0	92.0	56.8

Isomap			
#	train	tune	Test
1	55.0	59.2	56.8
2	68.0	68.0	56.0
3	56.8	55.2	62.4
4	64.3	68.8	57.6
5	79.9	89.6	52.0
6	70.3	73.6	56.8
7	85.3	85.6	57.6
8	85.8	84.8	56.8
9	84.8	87.2	56.8
10	92.7	95.2	56.8
11	96.6	96.8	55.2
12	93.3	94.4	62.4
13	98.0	98.4	60.0
14	93.5	93.6	54.4
15	95.2	94.4	60.0
16	95.8	96.8	60.0
17	99.4	97.6	60.0
18	97.1	97.6	58.4
19	99.0	100.	55.2
20	99.1	95.2	52.8
21	97.0	92.0	56.8

6. ACKNOWLEDGMENTS

This work was supported by the European Union from the European Social Fund (grant agreement number: UDA-POKL.04.01.01-106/09).

BIBLIOGRAPHY

[1] COX T.F., COX M.A., Multidimensional scaling. Chapman & Hall, 1994.
[2] DAVISON M., Multidimensional scaling, Wiley, 1983.
[3] JACKSON J.E., A User's Guide to Principal Components, New York, John Wiley and Sons, 1991.
[4] JOLLIFFE I.T., Principal Component Analysis, Springer-Verlag, 1986.
[5] MARDIA K.V., KENT J.T., BIBBY J.M., Multivariate Analysis, Probability and Mathematical Statistics, Academic Press, 1995.
[6] MICHALAK M., ŚWITOŃSKI A., STAWARZ M., Selection of the most important components from multispectral images for detection of tumor tissue, Journal of Medical Informatics and Technologies, 2011, Vol. 17, pp. 303-307.
[7] PEARSON K., On Lines and Planes of Closest Fit to Systems of Points in Space, Philosophical Magazine, 1901, Vol. 2, No. 6, pp. 559–572.
[8] SEBER G. A., Multivariate Observations, Wiley, 1984.
[9] ŚWITOŃSKI A., MICHALAK M., JOSIŃSKI H, WOJCIECHOWSKI K., Detection of tumor tissue based on the multispectral imaging, Lecture Notes in Computer Science, 2010, Vol. 6375, pp. 325-333.
[10] TENENBAUM J.B., De SILVA V., LANGFORD J.C., A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, 2000, Vol. 290, No. 5500, pp. 2319-2323.