Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 683–685

Evolutionary Nonlinear Data Transformation for Visualization and Classification Tasks

Kamil Ząbkiewicz
Vilnius University, Institute of Mathematics and Informatics
Akademijos str. 4, LT-08663 Vilnius, Lithuania
Email: Kamil.Zabkiewicz@mii.vu.lt

Polish Academy of Sciences, Systems Research Institute
ul. Newelska 6, 01-447 Warsaw, Poland
Email: K.Zabkiewicz@ibspan.waw.pl

University of Bialystok, Faculty of Economics and Informatics
in Vilnius
Kalvariju str. 135, LT-08221 Vilnius, Lithuania
Email: K.Zabkiewicz@uwb.edu.pl

***Abstract*—In this paper we propose new approach in data set dimensionality reduction. We use classical principal component analysis transformation. Instead of rejecting features we generate new one by using nonlinear feature transformation. The values of transformation weights are changed evolutionary by using genetic algorithms. Results show better classification rates in smaller feature space. Visualization results also look better.**

I. Introduction

NOWADAYS data visualization is one of the important fields in knowledge discovery. Human beings can perceive data up to three dimensions. Although dimensionality of datasets is often enormous. To solve this issue, dimensionality reduction methods were proposed. There are two main directions: feature selection and feature extraction. The subject of this paper will be related only with second approach. In this work we propose new approach by extending functionality of PCA transformation.

The paper is organized as follows. In section II the most popular data visualization methods will be described. In section III proposed nonlinear data transformation method will be presented. In section IV the results of experiments with UCI repository data will be shown. Finally in the section V main conclusions and directions for future research will be presented.

II. Data visualization methods

There are many different methods to visualize multidimensional data. One of the oldest and popular also in nowadays is principal component analysis. First proposed by Karl Pearson in 1901. Good review of this method is given by Jollife [2]. Another data visualization technique is multidimensional scaling, which firstly was used in psychometry. It is based on minimization of squared differences between distances of points in original and reduced feature space. Overview of this method is given in [4].There is also neural network approach called self-oragnizing maps (SOM) introduced by Kohonen in [3]. Good overview of recent visualization techniques is also given in [1].

III. Proposed nonlinear transformation

Main idea of this work is extension of PCA transformation with additional information provided with new feature. Our goal is to encode rejected features after principal component transformation into one number and append it to reduced data set. Encoding scheme is based on principles of fundamental theorem of arithmetic, i.e. that every natural number can be written as a product of prime numbers rised to appropriate power. This product is unique

$$a = \prod_{i=1} p_i^{k_i},$$

where p_i is prime number and k_i is power. In this case powers are natural numbers (including zero). It can be extended to rational powers due to the same cardinality. In this case numbers are also unique. In this work we will simplify initial requirement by using rational fractions between 0 and 1 instead of prime numbers. These numbers will be called weights. The initial formula is now written in such way

$$a = \prod_{i=1} w_i^{x_i},$$

Due to the fact that we do not know values of weigths we have to solve optimization problem. We will use genetic algorithm to optimize weights. Our fitness function will be result of classification made with nearest neighbour classifier. The algorithm is presented on page 684.

Let us comment several moments of this algorithm. First we perform initial data normalization into interval [0,1]. It is done because of possible different scales of features. We have chosen the nearest neighbour classifier because of small number of

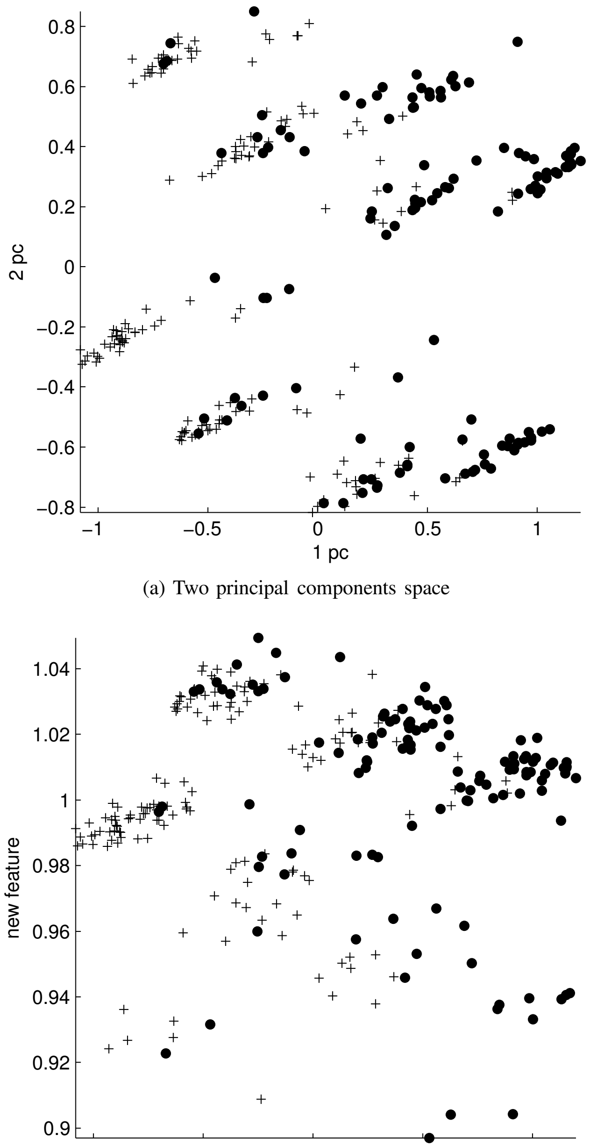
Algorithm 1 Classification and visualization using nonlinear evolutionary transformation

Require: $data$ - initial dataset, n_c - number of components for visualization, n_{iter} - number of iterations, n_{folds} - number of folds for crossvalidation, $n_{trfolds}$ - number of folds for internal crossvalidation
Ensure: $data_r$ - reduced transfomed dataset, w - vector of weights
  normalize dataset features into interval [0,1] using min-max normalization
  perform PCA transformation
  split transformed dataset into n_{folds} folds
  for $i = 1$ to n_{folds} **do**
    generate population of initial weights
    $iter = 1$ {current iteration}
    while $stopping\_criterium \neq$ true or $iter \leq n_{iter}$ **do**
      split training set into $n_{trfolds}$ folds
      for $j = 1$ to $n_{trfolds}$ **do**
        classify $fold_j$
        measure average classification error err_{iter}
      end for
      if $iter > 1$ **then**
        if $err_{iter} > err_{iter-1}$ **then**
          $stopping\_criterium \rightarrow$ true
        end if
      end if
      $iter \rightarrow iter + 1$
    end while
    perform evaluation on test fold tst_fold_i
  end for
  compute average algorithm performance $perf$
  $w \leftarrow w_{best}$
  $data_r \leftarrow$ transform($data, w$)
  return $perf, data_r$
  visualize($data_r$)

parameters (only number of neighbours and distance measure, in this case it is Euclidean distance). First tests of algorithm showed that proposed method has tendency to overfit the data. As a result new task appeared - to formulate early stop criterium. Brumen et al work [5] gave an idea how to solve this problem. In this work we use more simple criterium, i.e. "if error starts to grow break evolution process". To measure change of the error rate we use internal 5-fold crossvalidation. Only by using early stopping criterium we reduce number of iterations to perform. Internal crossvalidation provides also very important property, i.e. that classification error does not increase very rapidly. To prevent rapid increasing to infinity (decreasing to zero) of the generated feature value we let weights to vary only in interval $[1 - \varepsilon, 1 + \varepsilon]$, where $\varepsilon = \exp(1/number\_of\_rej\_comps) - 1$

IV. Experimental results

The experiments were performed on datasets from UCI Repository. These are: Glass, Teaching Assistant Evaluation, Heart and Chess. More details about each set is given in table I.



(a) Two principal components space

(b) First principal component and new feature space

Fig. 1. Heart dataset visualizations.

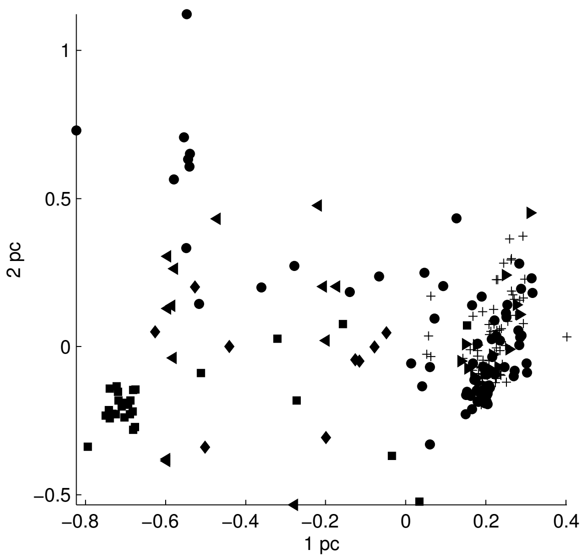
Datasets were chosen because the percent of total variance is not mostly covered by the several first principal components. The experiments were performed by reducing principal component space into 2 and 3 dimensions. Later the same experiments were made with the same datasets but with new additional feature appended. The number of features in both cases is equal. Results are shown in table II. We can notice that additional feature generated by nonlinear transformation in many cases gives better results.Visual analysis of datasets was also performed. Due to limited space we will show only several comparisons in two dimensional space. Let us look at plots of Heart dataset shown on figure 1. Classical PCA

TABLE I
DETAILS OF UCI DATASETS

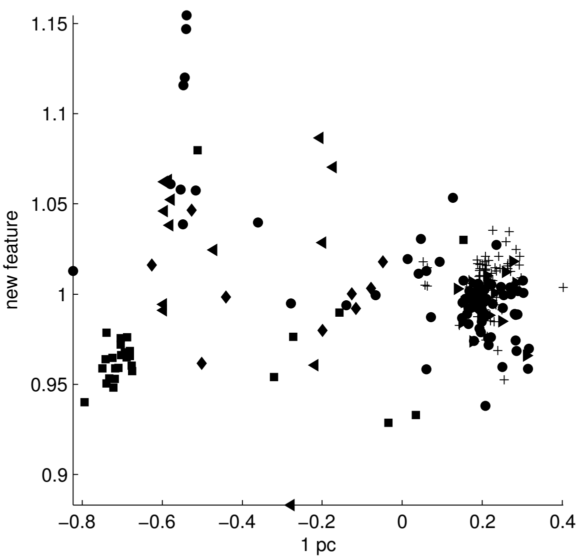
Dataset	Num. of feat.	Num. of classes	Num. of vectors
Glass	9	6	214
Teaching Assistant Evaluation	5	3	151
Heart	13	2	270
Chess	36	2	3196

TABLE II
UCI DATASETS CLASSIFICATION RESULTS

Dataset	1-NN				3-NN				5-NN			
	2 feat.		3 feat.		2 feat.		3 feat.		2 feat.		3 feat.	
	with	without	with	without	with	without	with	without	with	without	with	without
Glass	**0.5882**	0.5557	**0.5795**	0.5364	**0.5573**	0.5506	0.5672	**0.5688**	0.6053	**0.6136**	**0.6115**	0.6017
Teaching Assistant Evaluation	**0.6394**	0.6179	0.5763	**0.5911**	**0.5114**	0.4518	0.5143	**0.5241**	**0.5274**	0.4946	**0.4725**	0.4625
Heart	**0.7385**	0.7333	0.7330	**0.7481**	**0.7874**	0.7519	**0.7670**	0.7444	**0.8078**	0.7963	**0.8070**	0.7852
Chess	**0.6928**	0.5942	0.7020	**0.7030**	**0.6815**	0.6142	**0.7331**	0.6609	**0.6918**	0.6230	**0.7379**	0.6559



(a) Two principal components space



(b) First principal component and new feature space

Fig. 2. Glass dataset visualizations.

shows that classes are scattered, our approach on the other hand makes points of one class to be concentrated in certain part of space. Now let us analyze vizual data of Glass dataset. Both cases are shown on figure 2. As we can see objects of some classes are groupped more closer to each other. As a result using e.g. k-nearest neighbour classifier we can obtain better classification rates.

V. CONCLUSIONS

Proposed approach extends popular and quite simple data transformation. The computational cost of new feature is not that large as e.g. computing pairwise distance matrix (in case of multidimensional scaling) As it has been presented earlier proposed transformation made positive influence on analyzed data sets. First it increased classification rates with small number of components. In next step it provided more clear visualization results. Unfortunately this approach has one minus - it can be used only with numerical features. Of course we can encode categorical features into numbers, but this method is rather artificial and classification results can depend on way how features were encoded. The problem of big data sets is also actual. In this work small datasets were rather used. In future the proposed technique will be also tested on larger ones, e.g. KDD Cup 1999 data set. For solving such type of problem we will make some optimizations. One of them could be use of architectures that are based on parallel computing paradigm, such as clusters or GPUs.

REFERENCES

[1] G. Dzemyda, O. Kurasova and J. Zilinskas, *Multidimensional Data Visualization,* Springer, 2013.
[2] I. Jolliffe, *Principal Component Analysis, Second Edition,* Springer-Verlag New York, 2002.
[3] T. Kohonen, *Self-Organizing Maps, Third Edition,* Springer-Verlag, 2001.
[4] I. Borg, P. Groenen, *Modern Multidimensional Scaling: theory and applications (2nd ed.),* Springer-Verlag, New York, 2005.
[5] B. Brumen et al., "Learning process termination criteria," Informatica, vol. 23, No. 4, pp. 521-536, Vilnius University, 2012.