

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Extraction of Embedded Image Segment Data Using Data Mining with Reduced Neurofuzzy Systems

Deok Hee Nam

*Engineering and Computing Science
Wilberforce University
1055 N. Bickett Road, OH 45384
USA*

1. Introduction

To realize or implement the large dimensional image data, it may be taking a longer search time to detect the desired target. Recently, for the large amount of data and information in engineering or biomedical applications, various techniques including soft-computing techniques such as neural networks, fuzzy logic, or genetic algorithms, and multivariate analysis techniques like factor analysis, principal component analysis, or clustering analysis, are developed to extract the reduced meaningful information or knowledge from the original raw data.

In this paper, for mining or diminishing the large dimension of the given raw image data, factor analysis, principal component analysis, and clustering analysis are used to make a model using fuzzy logic or neurofuzzy systems, which are applied to predict the characteristics of the images with reduced dimensions. Generally the procedure can produce more precise and reasonable results with reduced dimensions in order to predict the desired images. In addition, all those techniques are useful for searching and saving time for the desired images. Thus, the proposed techniques intend to propose hybrid systems with integrating various multivariate analysis techniques together to establish neurofuzzy or fuzzy logic systems to construct a reasoning system with more accurate and efficient.

2. Literature review

2.1 Multivariate analysis

There are a lot of different kinds of data mining techniques to reduce the large and imprecise raw data into the reduced and precise raw data. Most frequently used techniques are multivariate analyses like factor analysis, principal component analysis, and various

¹ This paper is a revised version with the partial modification from the paper, "Data mining of image segments data with reduced neurofuzzy system" in the Proceedings (LNCS 5620) of 2nd International Conference Digital Human Modeling (ICDHM 2009) in 2009 at San Diego, CA. for the publication of Intech.

clustering analysis. Factor analyses (Gorsuch, 1983) concerns the study of the embedded relationships among the given variables to find or extract new variable sets, which are hidden and fewer in number than the original number of variables from the given data. In general, factor analysis attempts to reduce the complexity and diversity among the interrelationships of the applied data that exist in a set of observed variables by exposing hidden common dimensions or factors. Therefore, those newly extracted factors (or variables) after factor analysis can reform more precise and independent variables with less common dimensions among newly extracted variables, and the more precise information about the embedded structure of the data can be provided by factor analysis.

Principal component analysis (Kendall, 1980) and factor analysis usually produce very similar estimates. However, principal component analysis is often preferred as a method for data reduction, while factor analysis is often preferred when the goal of the analysis is to detect the embedded structure. One of the goals of principal component analysis is to reduce the dimension of the variables, such as transforming a multidimensional space into another dimension (i.e., same or less number of axes or variables), depending upon the given data. Hence, the principal component analysis converts the normalized data to the new data, called **principal component scores**, which represent the original data with a new pattern using the new variables that describe the major pattern of variation among data.

Finally, clustering analysis (Duda et al., 2001) is a method for grouping objects or observations of similar kinds into respective categories. In other words, cluster analysis is an exploratory data analysis tool which aims at sorting or separating different observations into the similar kinds of groups in a way that the degree of association between two observations or objects is maximal if they belong to the same group and minimal otherwise. In addition, cluster analysis can be used to recognize the structures in data without providing an explanation or interpretation.

2.2 Fuzzy logic and neurofuzzy system

Fuzzy logic was originally identified and set forth by Professor Lotfi A. Zadeh.

In general, fuzzy logic (Lin & Lee, 1996) is applied to the system control or the design analysis, since applying fuzzy logic technique is able to reduce the time to develop engineering applications and especially, in the case of highly complicated systems, fuzzy logic may be the only way to solve the problem. As the complexity of a system increases, it becomes more difficult and eventually impossible to make a precise statement about its behavior. Occasionally, it arrives at a point where it cannot be implemented due to its ambiguity or high complexities.

The neurofuzzy system (Yager & Filev, 1994) consists of the combined concepts from neural network and fuzzy logic. To implement the neurofuzzy systems, Adaptive-Network-Based Fuzzy Inference System (ANFIS) (Jang, 1993) is used by implementing the reduced data sets and the actual data set. ANFIS is originally from the integration of TSK fuzzy model (Yager & Filev, 1994), developed by Takagi, Sugeno, and Kang (TSK), using the backpropagation learning algorithm (Duda et al., 2001) with least square estimation from neural networks. TSK fuzzy model proposed to formalize a systematic approach to generating fuzzy rules from and to input-output data set.

3. Data structure

To perform the proposed technique, the selected image segment data provided by Vision Group of University of Massachusetts are used. The instances were drawn randomly from a

database of seven outdoor images. The images were hand segmented to create a classification for every pixel. The selected image segment data are consist of seven different measurement fields such as region centroid column, region centroid row, vedge-mean, hedge-mean, raw red mean, raw blue mean, and raw green mean, and four image classes like brickface, foliage, cement, and grass. The following describes each measurement field and the part of the image segment data.

1. Region centroid column: the column of the center pixel of the region.
2. Region centroid row: the row of the center pixel of the region.
3. Vedge-mean: measure the contrast of horizontally adjacent pixels in the region. There are 6, the mean and standard deviation are given. This attribute is used as a vertical edge detector.
4. Hedge-mean: measures the contrast of vertically adjacent pixels. Used for horizontal line detection.
5. Raw red mean: the average over the region of the R value.
6. Raw blue mean: the average over the region of the B value.
7. Raw green mean: the average over the region of the G value.

	Region centroid column	Region centroid row	Vedge-mean	Hedge-mean	Raw red mean	Raw blue mean	Raw green mean
CEMENT	191	119	1.294	0.77	48.222	35.111	-10.8889
BRICKFACE	140	125	0.063	0.31	7.667	3.556	3.4444
GRASS	204	156	0.279	0.56	25.444	28.333	-19.1111
FOLIAGE	101	121	0.843	0.89	6	3.111	-6.8889
CEMENT	219	80	0.33	0.93	48.222	34.556	-10.1111
BRICKFACE	188	133	0.267	0.08	7.778	3.889	5
GRASS	71	180	0.563	1.87	21.444	27.222	-12
FOLIAGE	21	122	0.404	0.4	1.222	0.444	-1.6667
CEMENT	136	45	1.544	1.53	63.778	47.333	-14.8889
BRICKFACE	105	139	0.107	0.52	7.222	3.556	4.3333
GRASS	60	181	1.2	2.09	17.889	23.889	-7.5556

Table 1. Image segment data (<http://www.cs.toronto.edu/~delve/data/datasets.html>)

4. Proposed algorithm (Nam & Asikele, 2009)

Among implemented algorithms, the preprocessing of principal components analysis (Kendall, 1980) followed by Fuzzy C-means (FCM) clustering analysis (Duda et al., 2001) is shown as a selected algorithm to present in this paper. The following steps summarize the

proposed algorithm to implement the reduced image segment data from the original image segment data.

Step 1. Read the original data set as a matrix format.

Step 2. Normalize the original data from Step 1.

Step 3. Find the correlation matrix of the normalized data from Step 2.

Step 4. Find eigenvalues and eigenvectors of the correlation matrix from Step 3 using characteristic equation.

Step 5. Define a matrix that is the eigenvectors from Step 4 as the coefficients of principal components using the criteria for extracting components.

Step 6. Multiply the standardized matrix from Step 2 and the coefficients of principal components from Step 5.

Step 7. Using the implemented data from Step 6, find the centers of clusters.

Step 8. Initialize the partition matrix, or membership matrix randomly such that $U^{(0)} \in M_{fcn}$.

Step 9. Calculate the cluster centers, v_i , using the equation,
$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\sum_{k=1}^n (u_{ik})^m}.$$

Step 10. Compute the distance, d_{ik} .

Step 11. Update the partition matrix $U^{(new)}$ using the equation
$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}}}$$
 for u_{ik} .

If $d_{ik} > 0$, for $1 \leq i \leq c$, and $1 \leq k \leq n$, then get the new u_{ik} .

Otherwise if $d_{ik} > 0$, and $u_{ik} = [0, 1]$ with $\sum_{i=1}^c u_{ik}^{(new)} = 1$, then $u_{ik}^{(new)} = 0$.

Step 12. Until $\|U^{(new)} - U^{(old)}\| < \varepsilon$ where ε is the termination tolerance $\varepsilon > 0$.

If this condition is not satisfied, then go back to step 9.

5. Analysis and results

Before the proposed data reduction algorithms are applied into the image segment data, the image segment data need to be examined whether the data can be diminished by the redundancy among its original variables with the highly correlated interrelationship. To examine the redundancy, the correlations between the variables of the image segment data set are calculated. As shown in the Table 2, the correlations of the "Brickface" image segment data are presented. The bolded numbers are showing the relatively higher correlation so that there is a possibility to be extracted as a new factor between those measurements.

In addition, there are different criteria to select the reduced dimension for the new reduced variables after extracting new variables from the original data. For this example, two combined criteria are applied. One is the eigenvalues-greater-than-one rule by Cliff (Cliff, 1988) and the second criterion is the accumulated variance that is more than 0.9 from the reduced system.

Using two conditions, the new reduced system with three newly extracted variables is considered. From Table 3, the evaluated analyses of the performance using the proposed algorithms through the neurofuzzy systems (Lin & Lee, 1996) are shown. In the system with three factors, the best result is from the method using the factor analysis with applying the

FCM analysis from the original data. Based upon the results of three newly extracted variables, the proposed algorithm can show the better result from the conventional methods such as factor analysis and principal component analysis.

Comparing to the three newly extracted factors, the four newly extracted variables are also considered among seven different measurement variables. The evaluated analyses of the performance using the proposed algorithms through the neurofuzzy systems are shown with comparing the statistical categories in Table 4. Even though the 4th newly extracted variable did not meet the applied criteria above. But it covers more variance of the original data than the three newly reduced variables. From the results of Table 4, the result from the method, using factor analysis (Gorsuch, 1983) and FCM clustering analysis, shows a relatively better result than other methods including the combinations of principal component analysis and FCM clustering analysis.

	Region centroid column	Region centroid row	Vedge-mean	Hedge-mean	Raw red mean	Raw blue mean	Raw green mean
Region centroid column	1						
Region centroid row	0.333	1					
Vedge-mean	-0.165	-0.266	1				
Hedge-mean	-0.015	-0.194	0.351	1			
Raw red mean	0.008	-0.729	0.33	0.412	1		
Raw blue mean	0.004	-0.691	0.334	0.408	0.993	1	
Raw green mean	-0.017	0.675	-0.248	-0.388	-0.808	-0.747	1

Table 2. Pearson’s correlation values for the “Brickface” image segment data (Nam & Asikele, 2009)

	CORR	TRMS	STD	MAD	EWI
fa	0.3096	0.9049	0.6064	0.8973	3.0991
pca	0.2628	1.4611	1.0579	1.4489	4.7051
fc	0.5386	0.7693	0.86	0.7597	2.8504
pc	-0.2757	1.2729	1.2195	1.257	4.4737

Table 3. Analysis of performance using proposal algorithm and conventional factor analysis and principal component analysis with three newly extracted factors

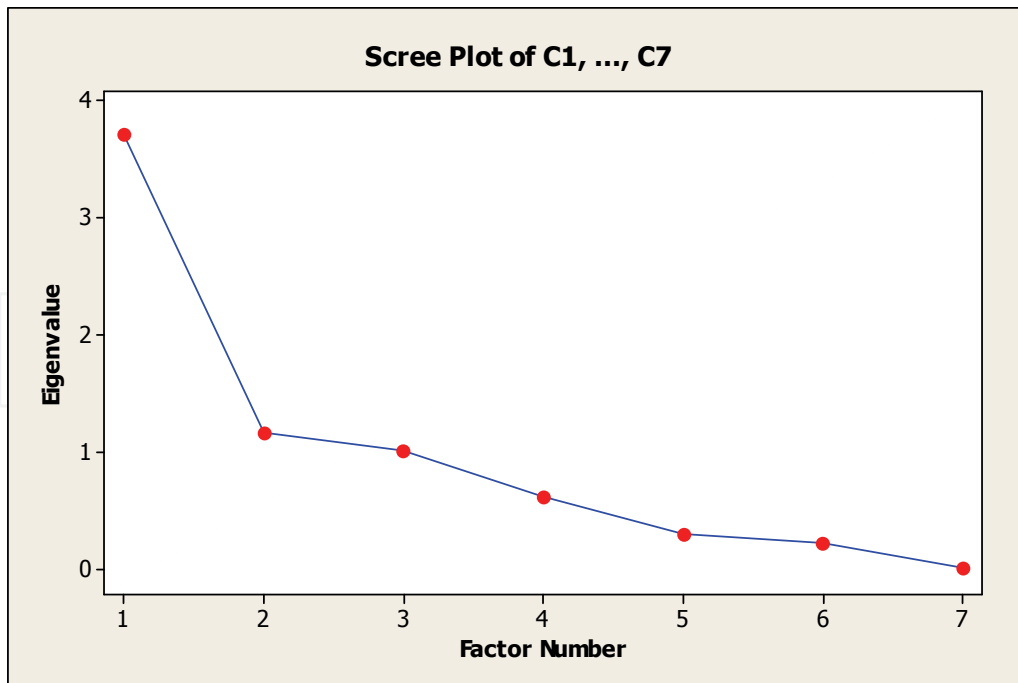


Fig. 1. Scree plot for the newly extracted components for "Brickface" image segment data (Nam & Asikele, 2009)

	CORR	TRMS	STD	MAD	EWI
fa	0.3984	0.8901	0.6664	0.8826	3.0407
pca	0.109	1.5185	1.1062	1.5059	5.0215
fc	0.3652	0.8807	1.3249	0.8697	3.7101
pc	-0.2541	1.2676	1.2878	1.2517	4.5529

Table 4. Analyses of Performance using proposed algorithm and conventional factor analysis and principal component analysis with four newly extracted factors (Nam & Asikele, 2009)

6. Conclusion

The pattern recognition of image segment data is presented and implemented through the neurofuzzy systems using the reduced dimensional data in variables and observations. For the implementation, three and four newly extracted embedded variables from seven original measurements variables are compared. The proposed algorithm performs the relatively better results than using the conventional multivariable techniques alone.

As described in Table 3 and 4, using the combination of factor analysis and FCM clustering analysis, the prediction of the patterns for the image segment data shows the relatively better results than other presented methods. The prediction results using the conventional principal component analyses show relatively worse than using other proposed algorithms.

This result may lead to the conclusion that for a limited number of input-output training data, the proposed algorithm can offer the better performance in comparison with the performance of the other techniques for image segment data.

7. Acknowledgments

This material is based upon the previous work supported by Clarkson Aerospace Corporation.

8. References

- Cliff, N. (1988). The Eigenvalues-Greater-Than-One Rule and the Reliability of Components, *Psychological Bulletin*, Vol. 103, No. 2, pp. 276–279.
- Duda, R.; Hart, P., & Stork, D. (2001). *Pattern Classification*, 2nd ed., John Wiley & Sons, ISBN 0-471-05669-3, New York, NY.
- Gorsuch, R. (1983). *Factor Analysis*, 2nd Ed., Lawrence Erlbaum Associates Inc., ISBN 089859202X, Hillsdale, NJ.
- Jang, J. (1993). ANFIS: Adaptive Network Based Fuzzy Inference System, *IEEE Trans. Systems, Man and Cybernetics*, Vol. 23, No. 3, pp.665–684.
- Kendall, M. (1980). *Multivariate Analysis*, MacMillan Publishing Co. INC., New York, NY.
- Lin, C. & Lee, C. (1996). *Neural Fuzzy Systems: A neuro-fuzzy synergism to intelligent systems*, Prentice Hall, ISBN 0-13-235169-2, Englewood Cliffs, NJ.
- Nam, D. & Singh, H. (2006). Material processing for ADI data using multivariate analysis with neuro fuzzy systems, *Proceedings of the ISCA 19th International Conference on Computer Applications in Industry and Engineering*, Nov.13–15, Las Vegas, Nevada, pp.151–156.
- Nam, D. & Asikele, E.(2009) Data mining of image segments data with reduced neurofuzzy system, *Proceedings (LNCS 5620) of 2nd International Conference Digital Human Modeling (ICDHM 2009)*, held as Part of the 13th International Conference on Human-Computer Interaction, July 19 - 24, 2009, Town and Country Resort & Convention Center, San Diego, CA, pp. 710 - 716.
- Yager, R.& Filev, D. (1994). *Essentials of Fuzzy Modeling and Control*, John Wiley & Sons, ISBN 0-471-01761-2, New York, NY.

Appendix

Abbreviations

CORR: Correlation

TRMS: Total Root Mean Square $TRMS = \frac{\sum_{i=1}^n \sqrt{(x_i - y_i)^2}}{n - 1}$

where x_i is the estimated value and y_i is the original output value.

STD: Standard Deviation

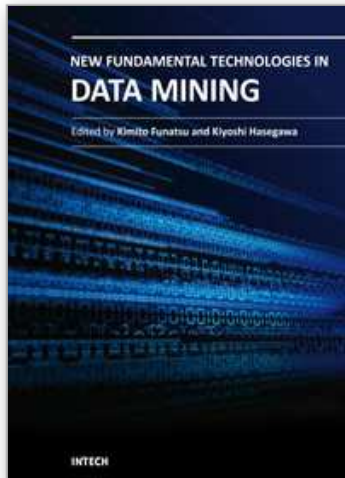
MAD: Mean of the absolute

EWI (Nam & Singh, 2006): Equally Weighted Index, the index value from the summation of the values with multiplying the statistical estimation value by its

equally
weighted potential value for each field
fa: Factor Analysis
pca: Principal Component Analysis
FCM: Fuzzy C-means Clustering Analysis
fc: preprocessing FA and SUBCLUST
pc: preprocessing PCA and SUBCLUST

IntechOpen

IntechOpen



New Fundamental Technologies in Data Mining

Edited by Prof. Kimito Funatsu

ISBN 978-953-307-547-1

Hard cover, 584 pages

Publisher InTech

Published online 21, January, 2011

Published in print edition January, 2011

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by "Data Mining" address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Deok Hee Nam (2011). Extraction of Embedded Image Segment Data Using Data Mining with Reduced Neurofuzzy Systems, New Fundamental Technologies in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-547-1, InTech, Available from: <http://www.intechopen.com/books/new-fundamental-technologies-in-data-mining/extraction-of-embedded-image-segment-data-using-data-mining-with-reduced-neurofuzzy-systems>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen