Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 2 (2010) 5276–5280

Procedia
Social and Behavioral Sciences

WCES-2010

A Web Based Clustering Analysis Toolbox (WBCA) design Using MATLAB

Kenan Savaş[a] *, Kazım Yıldız[a]

[a]*Marmara University Technical Education Faculty, Istanbul, Turkey*

Received November 11, 2009; revised December 1, 2009; accepted January 22, 2010

Abstract

With the development technology, datas which are saved in databases are rising, useful methods and kind of algorithms for making employable data and are build up for processing. For the solution of this problem data mining tools come into existence, to which clustering algorithms belong. For the clustering area this paper supposed to aid education, the tool was designed for clustering algorithms used easy and effective, cut down of the time that is making programme by the students so with the using of data sets that belongs to students, results of the clustering that was supported graphical items shown with the help of user interface and clustering must be so effective and easy. Purpose of this paper, MATLAB software that is used common by the academic area, making web based toolbox using clustering algorithms for the user. This toolbox use kmeans, agnes, fuzzy cmeans algorithms that are exist in MATLAB software and available web interface for the evaluate clustering results with cluster validity criteria. By courtesy of the toolbox, web user that do not need to have MATLAB software and programming knowledge but only a web browser and they can load their own data into the web server then see the result by the chose of algorithm and download the results their own local computers. In this design MATLAB Web Server (MWS) was used that belong to MATLAB software. With the help of this tool web users can send the parameters to server via the internet and the computer that has on the MWS software, used this then working MWS that is developed in this paper , results can have seen with the web server

Keywords: Clustering; Matlab Web Server; MWS; clustering algorithm; cluster validation; user based clustering.

1. Introduction

Internet can be used as virtual learning and also remote learning laboratoary as well. A virtual laboratory gives permanent access permission to a simulate process on computer. In other words, it is not necessary to be only one user at the same time or users don't have to wait other users. (Johansson et al., 1998).

Clustering analysis is a very common method in data mining. With the clustering analysis, a heterogeneous data group can be divided into significant subgroups inside (Berry,2004).

Several clustering algorithms are being used in clustering analysis. K-means (Davidson, 2002), AGNES (Kaufman, 1990) and fuzzy c-means (Azem, 2003) are the most widely used algorithms and they have been discussed in this study.

The objection of this study's are; to become the clustering analysis web based and provide to users to make transaction on data sets at the internet and finally to make active learnings with the help of an easy and useful interface.

* Kenan Savaş Tel.: +90 5327304408
E-mail address: kenan.savas@marmara.edu.tr

1877-0428 © 2010 Published by Elsevier Ltd.
doi:10.1016/j.sbspro.2010.03.860

MATLAB is one of the most popular software especially in technical and engineering fields. Especially in the areas of technical and engineering one of the most popular software used in Matlab. In this study MATLAB, MWS system via Web (Web MWS) is used. This system basically uses the MATLAB Web Server toolbox. This toolbox is an integrated software package to run MATLAB programs using the WWW advantages and to display results in a Web browser (Valera, 2005). In this paper, cluster analysis using the simulation capabilities of MWS is realized as showing graphs, giving numerical values and generating dataset files downloaded by web users.

2. Clustering

Clustering can be explained as the division of a heterogeneous data groups into homogeneous subgroups which is called "cluster" (Berry, 2004). A cluster is a collection consists of datas which are homologous but different from the other cluster's components (Larose, 2005). In this study, K-means, Agnes and Fuzzy C-means from conventional clustering algorithms was used. These clustering algorithms have been told in this chapter.

2.1. Kmeans

K-means is one of the oldest clustering algorithms and was improved by J. B. Mac Queen in 1967 (Mucha, 2009). It is one of the most widely used unattended learning methods. K-means's nominate mechanism lets to the each data to belong only one cluster. It has a tendency to find equal size spherical clusters (Davidson, 2002). K-means algorithm divides n object to k cluster. Firstly, the value of k could be given as an input parameter. It was aimed the similarity of internal cluster to be high but inter-cluster similarity must be low. Cluster similarity is measured with the average value of objects in a cluster. It is also center of gravity of the cluster (Han, 2001).

2.2. Agnes

AGNES (AGglomerative NEsting) algorithm, was presented by Kaufman and Rousseeuw in 1990 (Kaufman,1990). Initially each object is regarded as a separated cluster. Following each step in the algorithm, atomic cluster that show similar attribution is merged. After each merge operation total number of clusters is decreased one. The merge process end when desirable cluster is get or the distance between the two cluster is reached to threshold value (Han,2001).

2.3. Fuzzy C-means

The total membership value for all classes for data that must be "1". Object will be close to a particular cluster center. Membership of this object to belong to that cluster to will be larger than the membership of other clusters'. One of the important features of Fuzzy c-means algorithm, clustering of the membership matrix, has a positive effect on the clustering of the matrix. This matrix facilitates ambigious cases (*Azem,2003*). Also outlier datas effect inconsiderable because of low membership degree. Fuzzy c-means algorithm has flexible structure. Capability of finding overlap clusters is more than other divisive algorithms.

3. The Design of Web Based Clustering Analysis (WBCA) Toolbox

In this study, in the field of clustering, it is considered that designed toolbox could contribute to the education, therefore students could use clustering algorithms in an efficient and easy way. Besides these, the lost time that is lost by students in programming could be prevented. This would be realized uploading the own dataset files through the Internet by the students.Moreover, user interface with the help of clusters would made clustering analysis easier and more effective with graphical results. Designed toolbox is very useful in education. Because WBCA toolbox, education is provided in many ways: simulation, working own data sets, using example data sets, etc. This can be seen in Figure 1.

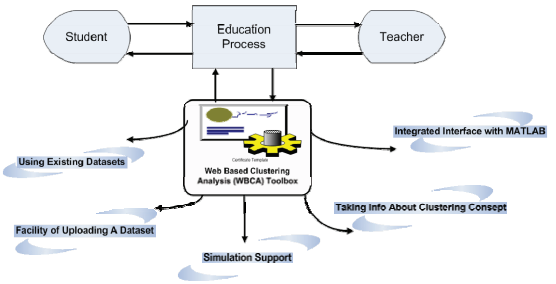


Figure 1. Provided Roles in Education Roles Using **WBCA**

WBCA that has Web-based design, uses web documents and several systems to work together to respond to users via the web. For applications that are included in MATLAB environment WBCA includes mostly web-based design documents and MATLAB source codes. The system is running on a web server environment.. In this system for server software, Apache Web Server v.2.2.8 is used. WBCA toolbox's relationship of the other systems is shows in Figure 2.

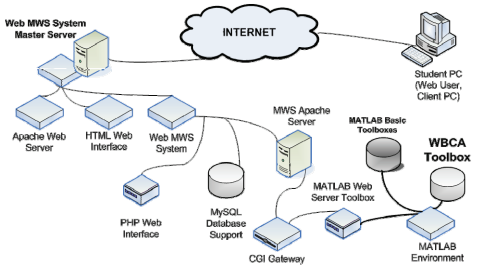


Figure 2. **WBCA** toolbox and the other architectures

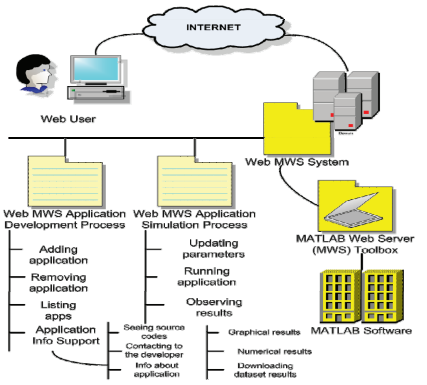


Figure 3. Design Proses with Web MWS System

3.1. Apache Web Server 2.2.8

A server computer needs a server software to publish html structured documents over the web. Web MWS is a web-based system so this system uses a web server computer and Apache Web Server runs on this computer.

3.2. Web MWS System v.5.0.030909.0

This system is designed to allow developing and running applications which use MATLAB environment. It is possible to develop and run a web-based MATLAB applications using Web MWS system that provides an easy way doing this. Which uses PHP-based web documents which are installed on a server computer in which at the same time MATLAB software installed (Retrieved November 16, 2009, from http://www. Controlworld.tk/web_mws/). Thus, a dynamic web page within the support system is used. This system can be accessed at any time via the Internet. Web users can add any application, there are existing applications can list or you can view the source code of applications (Figure 3). MWS System previously installed any of the Web application can run with ease. In addition, the results obtained as a result of MATLAB work may be download to own computers by web users.

3.2.1. The Architecture of Web MWS System

MATLAB users to benefit their own algorithms and can prepare them as a part of the MATLAB libraries become possible. Is easy to make programming with MATLAB (Hercog et al., 2007). Therefore, Web users easily through the Web MWS System develop MATLAB applications. MATLAB software uses an integrated MATLAB Web Server tool box. It can be seen that how Web MWS system works in Figure 4b, It can also be seen that how MWS toolbox can operates MATLAB applications.

3.3. MATLAB Web Server (MWS) Toolbox

MATLAB Web Server is the well-known software package for the MATLAB as a toolbox developed by Mathworks. This tool provides a web front MATLAB interface to establish connections between a client and a server. Thus the client to be viewed from a computer and a server that can start an application on the computer installed MATLAB on simulations will be possible to design (Diez et al., 2002). Outputs of the simulations are HTML documents.

MATLAB Web Server (MWS) and simulations can be done online. These are many technical courses and in particular virtual lab environment is required. Moreover, they are also useful laboratories for mathematics courses. Simulation of the input interface and output of numeric, alphanumeric, or may be graphical. However, there are some disadvantages of using this tool. For example, users can change input values only. Another disadvantage is that high-performance CPU and simulation to shorten the time taken into account that requires much RAM capacity. In addition, MWS development of applications for registration application is required and that application development time increases because of registry of applications. This study, to find solutions to these problems uses a system of Web MWS (Retrieved November 16, 2009, from http://www.controlworld.tk/

web_mws/). When a web user develop and install a Web MWS application to the Web MWS system, the developing steps can be seen in figure 4.

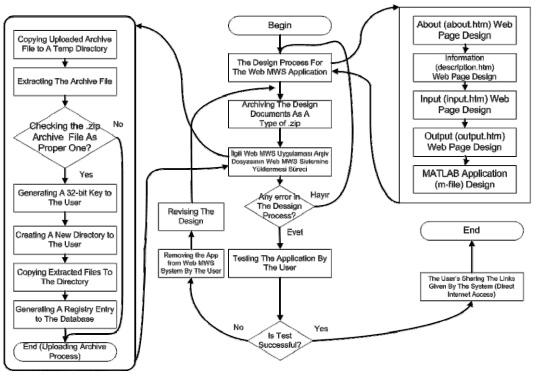


Figure 4. The steps developing applications in the support of Web MWS System

4. WBCA Interface and Simulation Environment

Students or web users via WBCA can reach all kinds of information related to the experiments in clster analysis. Therefore, many users will have the opportunity to be online simultaneously in the system. WBCA user interface in Figure 5a and Figure 5b also shown. Users draw on any page to access the simulation with each data set offered "Cluster Analysis" fields. This fields can be changed from the user as the parameters of the analysis.

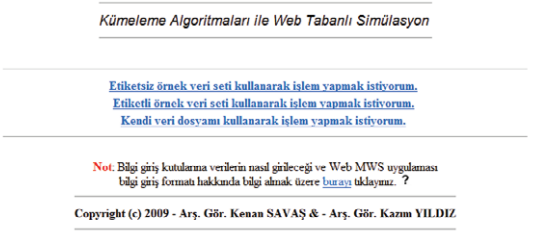
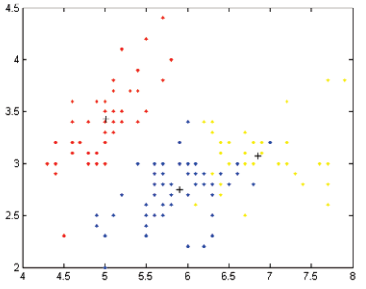


Figure 5a. The first web interface of WBCA toolbox

Figure 5b. The input page of the simulation interface of WBCA toolbox

After students enter values to change the parameters to the cluster analysis, to send the form information they will click on the Send button. Then on the web page in a separate window users can see cluster results graphically and numerically This is illustrated in Figure 6a and Figure 6b.



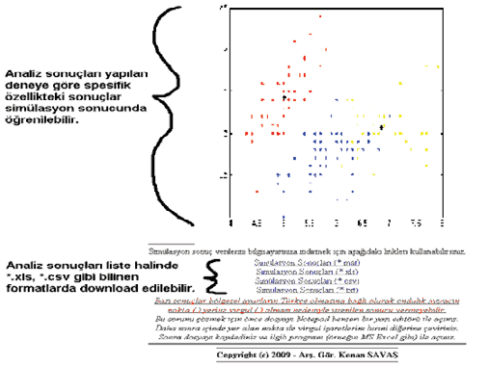


Figure 6a. A result page from the simulation using WBCA toolbox

Şekil 6b. The facility of uploading the result data using the result page

Also, the cluster result values generated by MATLAB users, to their local computer may download and use in any 3rd part software (Figure 6b.).

5. Conclusion

In this study, Web-Based Clustering Analysis (WBCA) toolbox has been designed to be used in MATLAB environment. It is aimed that an easy-to-use tool to help the teaching environment is designed as a supplementary tool. This tool is developed using Web MWS system so that all users can use the WBCA toolbox at any time and from anywhere. In addition to this, users don't need to install MATLAB software on their machines, the only requirement to use this toolbox, is a PC connected to the Internet and a web browser software. Especially among universities and students, as the sharing of MATLAB and the licence cost are considered (Uran & Jezernik, 2008) with Web MWS costs would significantly reduce.

References

Azem, Z: "A Comprehensive Cluster Validity Framework For Clustering Algorithms", MSc Thesis, The University of Guelph, Canada, 2003

Chinchilla, R. & Nurre J. H., Design of an Industrial Controls Laboratory, Proceedings of the 26th Southeastern Symposium on System Theory, IEEE (pp. 618-622), 1994.

Díez, J. L., Vallés M. & Valera A., "A global approach for the remote process simulation and control," in Proc. 15th IFAC World Congress Automatic Control, Barcelona, 2002, pp. 76–81.

Berry, M. J. A. & Linoff, G. S., "Data Mining Techniques: For Marketing, Sales, and Customer Support", Wiley Publishing, Inc,Indiana, 2004.

Davidson, I, Y.: "Understanding K-means Non-hierarchical Clustering", Technical Report, Computer Science Department of State University of New York (SUNY)–Albany, Feb 2002.

Erdal, H., Otomatik Kontrol Dersi Deney Kitabı, İstanbul, 2004, http://posta.marmara.edu.tr/~kenan.savas/categories/ders_notlari/deney_foyleri/oto_kontrol_1_deney_foyu.pdf (16nd Nov, 2009).

Han, J.; Kamber, M.: "Data Mining Concepts and Techniques.", Morgan Kauffmann Publishers Inc.,Aug 2001.

Hercog, D., Gergic B., Uran S. & K. Jezernik, ,A DSP-Based Remote Control Laboratory, Ieee Transactıons On Industrıal Electronıcs,Vol.54,No.6,December 2007.

Johansson, M., Gäfvert M. & Aström K. J., "Interactive tools for education in automatic control,"IEEE Contr. Syst. Mag., vol. 18, no. 3,pp. 33–40, 1998.

Kaufman L.;Rousseeuw, P. J. : "Finding Groups in Data: an Introduction to Cluster Analysis"., *John Wiley and Sons*. 1990

Larose, D., T, "Discovering knowledge in data: an introduction to data mining", John Wiley & Sons, Inc.,New Jersey, 2005.

Mucha, J, M; Sofyan, H: "Nonhierarchical Clustering", http://www.quantlet.com/mdstat/scripts/xag/ html/xaghtmlframe149.htm (29nd Nov 2009)

The Algorithm of c-Means, 2009, http://www.elet.polimi.it/upload/matteucc/Clustering/tutorial_html/cmeans.htm, Nov 15th, 2009.

Uran, S. & Jezernik K., Virtual Laboratory For Creative Control Design Experiments, IEEE Transactions On Education, Vol. 51, No.1, Feb 2008.

Valera, A., Díez J. L., M. Vallés, & P. Albertos, Virtual and Remote Control Laboratory Development, IEEE Control Systems Magazine (pp. 35-39), 2005.

Web MWS System User Manual, 2009, http://www.controlworld.tk/web_mws/ (16th Nov, 2009)