International Journal of Information and Education Technology, Vol. 1, No. 1, April 2011
ISSN: 2010-3689

WKMSDI: An Optimal Algorithm by Evaluation of Effective Parameters on Data Mining Algorithm

Arash Ghorbannia Delavar, Esmaeil Zeynali Khesraghi, Majid Feizollahi, and Nasim Anisi

***Abstract*—In this paper, we will present an optimal algorithm by evaluation of effective parameters on data mining algorithm. Using effective parameters on interactions of clustering method, this algorithm, WKMSDI, increases speed of data recovery. It occurs when our qualification function can optimize or improve clustering scope with given parameters.**
Attached parameters to basic function of k-means help us to increase system efficiency. For this purpose, recommended algorithm, compared to the preceding algorithms, increases system reliability for high processing. Effective parameters increase algorithm quality more than the compared algorithm. This needs integration of specific parameters to enhance system performance.

For enhancing quality in distributed integrated systems, we require the procedures which have direct access to some parameters. Usually, effective parameters increase system efficiency in very high interactions.

***Index Terms*—WKMSDI ; ERPSD ; ERPASD; Data Mining ; Knowledge Base ; Scheduling ; Dependability.**

I. Introduction

In the twentieth century, environmental conditions for entering customers are among powerful managerial tools which are used in information technology. For processing cumulative percentage of information, we need to do high level data investigations at data bases.

Since implementation of ERP systems has increased over the recent years, scientific researchers in the fields of software and information technology have been preoccupied with rapid information search and reduction of survey time.

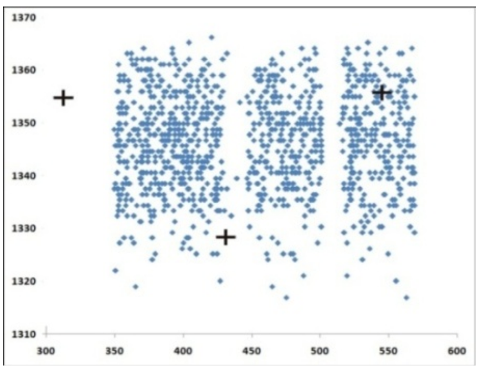
The software and hardware aspects of this issue have been always considered so that, using proper methods and assessments, the system has a mechanism by which we can reduce data base survey scheduling and improve relationships with the clients in order to enhance their level of satisfaction.

In distributed integrated systems which are based on ERPSD, two phases of software and hardware have been used and they have a high processing capacity within the distributed environment.[2] In this method, by integrating combined methods a suitable strategy has been presented which enhances the level of clients' satisfaction. Therefore, in order to increase the level of clients' satisfaction, we have replaced integrated central systems distributed integrated systems so that their dependability and security will be improved.

By studying previously Proposed frameworks and models, we succeeded in presenting suitable strategies for creating the new algorithm. Considering the new Proposed algorithm, we devised new technical methods in which a new methodology has been used because although all the above-mentioned independent algorithms have some advantages but after combing effective factors, their performance decreases. Using the Proposed WKMSDI algorithm, we succeeded in improving the performance of the base K-Means algorithm.

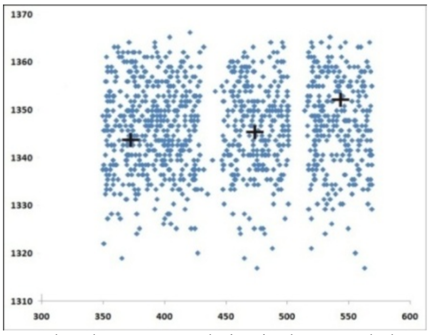
II. Distributed Systems With Data Mining Mechanism

Successful implementation of ERP systems with regard to their capabilities has always been studied and in this matter the most effective factors are selection of suitable process and algorithm.[11] Knowledge discovery process, extracting suitable information, and using optimal algorithms are among the most important challenges in this regard. Taking into consideration their unique requirements, various companies have presented different strategies for data extraction and knowledge discovery.[12]

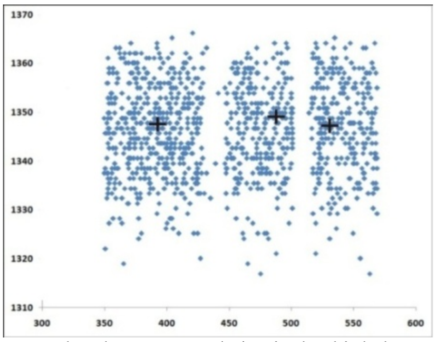


Data abundance accumulation in the first phase

Manuscript received March 20, 2011
Payam Noor University, faculty of engineering, Tehran, Iran , email:a_ghorbannia@pnu.ac.ir
Islamic Azad University,Qazvin branch (qiau), qazvin , Iran , email: zeinali@qiau.ac.ir
Payam Noor University, faculty of engineering, Tehran,email:IranMfeiz78@gmail.com
Payam Noor University, faculty of engineering, Iran, Tehran, email: n.anisi@oiic-ir.com



Data abundance accumulation in the second phase



Data abundance accumulation in the third phase

Figure 1. Function of *K*-Means algorithm in the information bank of members of Khane Kargar center

ERPSD is a new method for integrating all current processes and optimizing the employment of system resources with regard to competitive level and clients' satisfaction using ERP. In this method, security and performance levels have increased compared with the previous model and, thus, time and cost have decreased. Acquiring the new methodology requires identification of ERPSD phases and data mining algorithms.[2] Following the review of the previous literature, it was found out that this methodology is better implemented in distributed rather than central systems. Therefore, in order to increase the efficiency, we use dependable distributed integrated systems so that we can remove problems observed in central ERPs.[8,9]

In a previous study, the ERPASD algorithm has increased security and performance levels as comparison with previous methods and also considering this algorithm, in the distributed information bank, a technical method has been presented which reduces the total cost of the distributed integrated system, calculates repetitive data using Apriori ERPSD, and optimizes the ERP system. [1]

In order to control the data base and accessibility, there are various algorithms for data mining such as: K-Means, Page Rank, EM, and SVM Apriori. [3] But in this research, following an accurate and comprehensive study of various algorithms, we used K-Means algorithm which produces practical efficient results.

III. Base K-Means Algorithm

Base K-Means algorithm is one of the preferable data mining algorithms in which clustering method is used. In a simple type of this method, first, random points equal to required clusters are selected. Then the data are related to one of these clusters according to their proximity to these clusters and new clusters are formed. [3]

TABLE I. Users Respond Time In K-Means Algorithm(s)

		number of records	Users Respond Time In K-Means Algorithm(s)
1	S1	5000	0.022
2	S2	7000	0.033
3	S3	10000	0.037
4	S4	20000	0.039
5	S5	30000	0.041
6	S6	50000	0.043
7	S7	70000	0.048
8	S8	100000	0.049
9	S9	200000	0.254
10	S10	300000	0.277
11	S11	400000	0.436
12	S12	500000	0.491
13	S13	700000	0.691
14	S14	1000000	0.745
15	S15	2000000	1.155

Repeating the same procedure, in each repetition by averaging the data new centers can be estimated for them and again the data can be related to new clusters. This procedure continues until no change occurs in the data.

The following function is the target function:

$$j= \sum_{j=1}^{k} \sum_{i=1}^{n} \| x_i^{(j)} - c_j \|^2 \quad (1)$$

Number of fields: F_c

Number of clusters: K

$$K = F_c + 1 \quad (2)$$

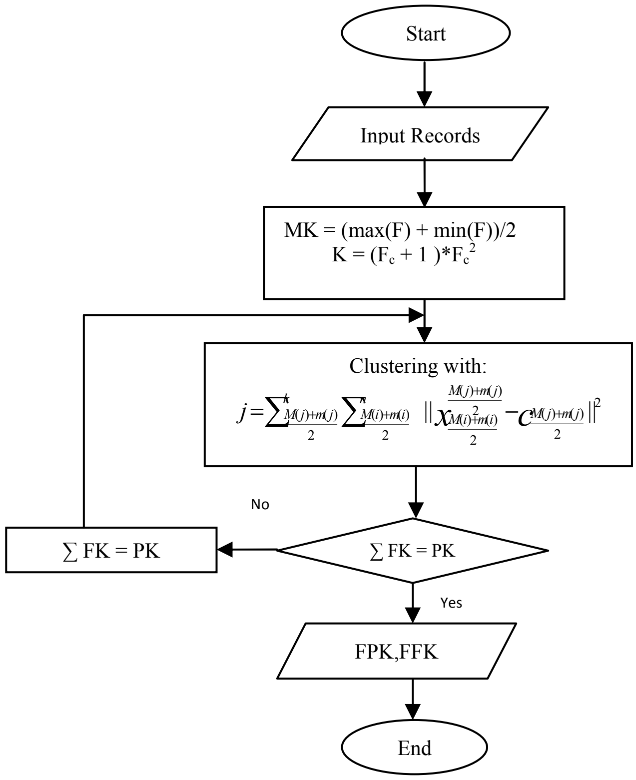
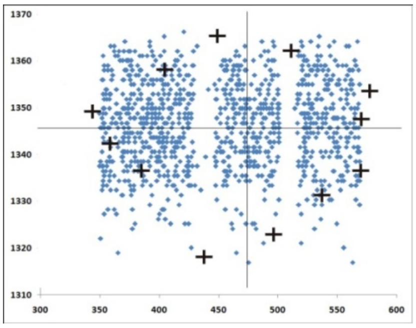


Figure 2. WKMSDI algorithm flowchart

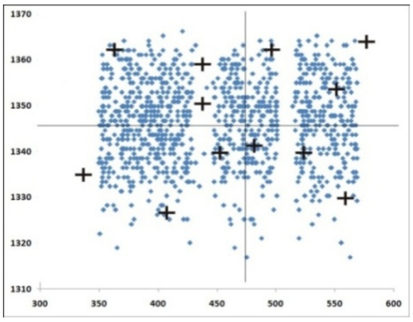
$\| xi(j) - cj \| 2$ is a criterion for the distance between xi(j) data points and cj cluster centre and j is the distance between n data points from their respective cluster centers. The problem with implementing this algorithm is that by determining the points and clustering in voluminous data, still data abundance accumulation percentage in each cluster will be high and survey time in information bank will be higher compared with similar algorithms and we must present an algorithm which may reduce its search time.

IV. WKMSDI Proposed Algorithm

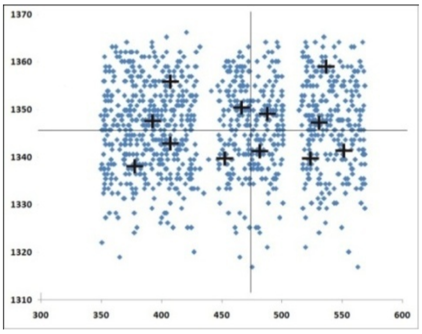
Using WKMSDI algorithm, we create a new technique by which data accumulative percentage at central points will be reduced. In this case, time survey in the information bank and accumulation of information is presented by breaking up the areas. We will illustrate these phases by defining all the parameters affecting the WKMSDI algorithm.



Data abundance accumulation in the first pha



Data abundance accumulation in the second phase



Data abundance accumulation in the third phase

Figure 3. Function of WKMSDI algorithm in the information bank of members of Khane Kargar center

WKMSDI algorithm has been compared with K-Means algorithm and the Proposed algorithm is more efficient than the Base K-Means algorithm. We have evaluated the amount of specified fields in WKMSDI algorithm and then by combining its Min and Max, we create the vertical bisection condition which is estimated using their average mechanism. Then, by this average, in each area, information search survey is separately evaluated using the above technique and new data and clustering in relation to K are established. This procedure continues until we reach a stable condition in various simulations and when it approximates the Max and Min points of the function, it will reach a more stable condition.

The following function is the target function:

$$j=\sum_{\frac{M(j)+m(j)}{2}}^{k}\sum_{\frac{M(i)+m(i)}{2}}^{n}\left\| x_{\frac{M(i)+m(i)}{2}}^{\frac{M(j)+m(j)}{2}} - c_{\frac{M(j)+m(j)}{2}} \right\|^2 \quad (3)$$

M=Maximum ,m=Mininum
Number of fields: F_c
Number of clusters: K

$$K = (F_c + 1)*F_c^2 \quad (4)$$

$\left\| x_{\frac{M(i)+m(i)}{2}}^{\frac{M(j)+m(j)}{2}} - c_{\frac{M(j)+m(j)}{2}} \right\|^2$ is a criterion for the distance between $x_{\frac{M(i)+m(i)}{2}}^{\frac{M(j)+m(j)}{2}}$ data points and $c_{\frac{M(j)+m(j)}{2}}$ cluster centre and j is the distance between n data points from their respective cluster centers.

In this method, when the information amount is high data abundance is broken in each area and survey time increases.

V. Conclusion

Taking into consideration the limitations which Base K-Means algorithm has caused in the data abundance accumulation percentage, we succeeded in reducing this problem by presenting a new technique with WKMSDI algorithm and, using this algorithm, we created a target function in order to improve the survey method in the information bank. Several simulations which have been studied using simulation software and information bank as well as comparisons made indicate that the Proposed algorithm has proved its efficiency in comparison with Base K-Means algorithm and point finding survey in the Proposed algorithm for data accumulation has been performed with higher accuracy and a lower amount of survey.

TABLE II. Users Respond Time In WKMSDI Algorithm(s)

		number of records	Users Respond Time In WKMSDI Algorithm(s)
1	S1	5000	0.019
2	S2	7000	0.029
3	S3	10000	0.032
4	S4	20000	0.034
5	S5	30000	0.036
6	S6	50000	0.037
7	S7	70000	0.041
8	S8	100000	0.044
9	S9	200000	0.175
10	S10	300000	0.254
11	S11	400000	0.416
12	S12	500000	0.482
13	S13	700000	0.624
14	S14	1000000	0.714
15	S15	2000000	0.984

This algorithm has a higher efficiency compared with its previous similar algorithm and is more reliable than Base K-Means algorithm. Finally, we have evaluated abundance accumulation in all three phases using Base K-Means and WKMSDI algorithm and the results obtained from WKMSDI algorithm are much better than those of the Base algorithm.

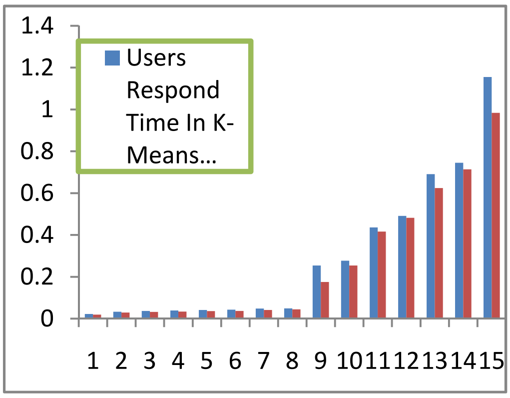


Figure 4. Result Respond Times of Implementation Algorithm

Figure 5. Data abundance accumulation percentage in *K*-Means final clustering

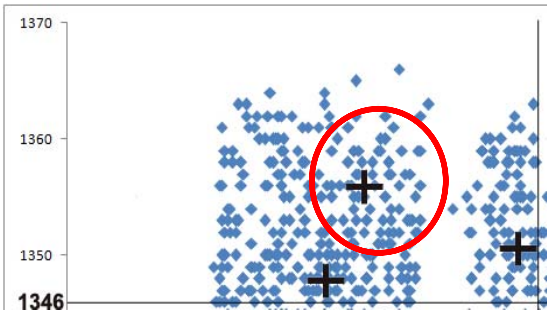


Figure 6. Data abundance accumulation percentage in WKMSDI final clustering

REFERENCES

[1] A. Ghorbannia Delavar, M. Zekriyapanah Gashti" ERPASD: A Novel Algorithm for Integrated Distributed Reliable Systems Using Data Mining Mechanisms" September 2010,Chongqing china, PP.483-487.

[2] A.Ghorbannia Delavar , V.Aghazarian , S.Sadighi "ERPSD: A New Model for Developing Distributed,Secure, and Dependable Organizational Softwares" IEEE CSIT 2009 Yerevan, Armenia, September 28 2009 ,PP.404-408.

[3] X.Wu , V.Kumar , J. R.Quinlan ,J.Ghosh , Q.Yang , H.Motoda, G.J. McLachlan , A.Ng · Bing Liu · Philip S. Yu , Z.H.Zhou , M.Steinbach , D.J.Hand "Top 10 algorithms in data mining" Dan Steinberg Received: 9 July 2007 / Revised: 28 September 2007 / Accepted: 8 October 2007 Published online: 4 December 2007© Springer-Verlag London Limited 2007

[4] A.Ghorbannia Delavar ,L.Saiedmoshir "Analysis of a New Model for Developing Concurrent Organizational Behaviors with Distributed, Secure and Dependable ERPSD Technology" IEEE CSIT 2007 September 24-28 2007,PP.321-324.

[5] A. Ghorbannia Delavar and M. Nejadkheirallah and M. Motalleb, A New Scheduling Algorithm for Dynamic Task and Fault Tolerant in Heterogeneous Grid Systems Using Genetic Algorithm, Chengdu, China, 2010 , PP.408-412.

[6] A.Abdullah , S.Al-Mudimigh , B.Zahid Ullah , C.Farrukh Saleem "A Framework of an Automated Data Mining Systems Using ERP Model" International Journal of Computer and Electrical Engineering (IJCEE 2009) , Vol. 1, No. 5 December 2009 , PP.651-655.

[7] Ch.Liu , Jiancheng An " Fast Mining and Updating Frequent Itemsets " International Colloquium on Computing, Communication, Control, and Management , IEEE CCCM 2008.

[8] T.Lan, Catherine, Huang, D.Erdogmus , " A Comparison of Temporal Windowing Schemes for Single-trial ERP Detection " Proceedings of the 4th International IEEE EMBS Conference on Neural Engineering Antalya, Turkey, April 29 - May 2, 2009

[9] K.B. Hendricks a,1, V.R.Singhal b,*, J.K.Stratman b,2a R.Ivey "The impact of enterprise systems on corporate performance:A study of ERP, SCM, and CRM system implementations" School of Business, The University of Western Ontario, London, Ont., Canada N6A-3K7b College of Management, Georgia Institute of Technology, 800 West Peachtree St., NW, Atlanta, GA 30332-0520, United StatesAvailable online 23 March 2006

[10] N.A. Morton, Q.Hu "Implications of the fit between organizational structure and ERP:A structural contingency theory perspective" Department of Information Technology and Operations Management, Barry Kaye College of Business,Florida Atlantic University, Boca Raton, FL 33431-2007, USA

[11] Neil A. Morton, Qing Hu]Implications of the fit between organizational structure and ERP:A structural ontingency theory perspective Department of Information Technology and Operations Management, Barry Kaye College of Business,Florida Atlantic University, Boca Raton, FL 33431-0991, USA International Journal of Information Management 28 (2008) 391–402

[12] Michel Pouly, Jean-Charles Binggeli, Rémy Glardon Ecole Polytechnique Fédérale de Lausanne (EPFL),ERP data sharing framework using the Generic Product Model (GPM) Souleiman Naciri, Naoufel Cheikhrouhou *, Laboratory for Production Management and Processes, Station 9, 1015 Lausanne, Switzerland 2010

Arash Ghorbannia Delavar Received the MSc and Ph.D. degrees in computer engineering from Sciences and Research University, Tehran, IRAN, in 2002 and 2007. He obtained the top student award in Ph.D. course. He is currently an assistant professor in the Department of Computer Science, Payam Noor University, Tehran, IRAN. He is also the Director of Virtual University and Multimedia Training Department of Payam Noor University in IRAN. Dr. Arash Ghorbannia Delavar is currently editor of many computer science journals in IRAN. His research interests are in the areas of computer networks, microprocessors, data mining, Information Technology, and E-Learning.

Majid Feizollahi Payame Noor University,Jame elmi karbordi university,(Khane karghar centre),Tehran, Iran ,Mfeiz78@gmail.com .

Nasim Anisi Payame Noor University,OIIC Company ,Tehran, Iran ,n.anisi@oiic-ir.com.