International Association of Scientific Innovation and Research (IASIR)
(An Association Unifying the Sciences, Engineering, and Applied Research)

ISSN (Print): 2279-0063
ISSN (Online): 2279-0071

International Journal of Software and Web Sciences (IJSWS)
(Open Access, Double Blind Peer-reviewed, Refereed and Indexed Journal)
www.iasir.net

Rules Generation for Impacts and Challenges of Enterprise Resource Planning

Prafulla Bafna, Nida Saiyed, Tejendra Khatri, Shubham Jain, Ayush Soni
Symbiosis Institute of Computer Studies and Research (SICSR),
Symbiosis International University (SIU), Atur Centre, Gokhale Cross Road,
Model Colony, Pune – 411 016, Maharashtra State, INDIA

Abstract: *The proposed idea talks about Enterprise Resource Planning (ERP) systems offering various business benefits such as real-time data availability, improvement in visibility, and increment in task automation and for generating minimal overhead cost. The mode of deployment is considered as one of the biggest advantage of on-demand enterprise resource planning systems. Feature Selection (FS) is the fundamental process for identifying the features that play an important role in decision-making process. Data mining techniques are widely used for data driven decision support. Feature selection is an important part of preprocessing that is carried out prior to data mining. FS algorithm or technique helps to reduce dimensions by removing unwanted or noisy data and the data, which is not having any influence on the outcome. In this paper, we present a novel approach, which facilitates groups of similar factors based on their degree of excellence in ERP output. It also selects the top N factors, which are affecting ERP practices using a feature selection technique. Cluster quality is used as decision support for the optimum feature set. This approach will inform the quantified measure about overall ERP practice in Maharashtra state.*

Keywords: *ERP, hierarchical clustering, , dimension reduction, optimal features,*

I. Introduction

Resources (vital part) in ERP helps in prevention of Inventory reduction and keep its record for wider scope.ERP lays focus on entire business process and strategies as a whole and generates results enhancing productivity. Integrating core business processes together in a single application, it helps the company to accelerate the efficiency of business process across the entire organization. Plant manufacturing can make speedy production . Increase on time delivery, productivity, ability to forecast demand to supplies, order capacity, and in improving customer service .As the ERP system clusters all business management functions, it reduces level of inconsistency . Thus, by using ERP system, managers can gather correct information and make a correct decision.ERP facilitates improved services to customer and suppliers.[24][25]Data mining is a form of knowledge discovery essential for solving problems in specific domain. It is used in financial banking for credit risk covering, market segmenting, health care etc. Clustering is one of the data mining techniques, used to discover hidden patterns when relation between the data is unknown. Clustering is to be used on complete data. There are different methods of clustering viz. partitioning, hierarchical, density based, model based and grid based.[14] The decision making process heavily depends on the quality of clusters and the algorithm should be chosen appropriately. The suitability of the algorithm depends on the domain and usually several algorithms need to be tried before choosing the best which is expensive. For most applications not only data is of dynamic nature that is continuously growing in size, but also is multidimensional in nature leading to scalability problems. Data size cannot be reduced but irrelevant dimensions can be definitely reduced. So before applying any algorithm, it is essential to use some preprocessing technique to reduce unnecessary dimensions/features and select the most relevant features. Feature selection is one of the approaches for dimension reduction. Dimension reduction limits storage requirements. It speeds up the running time of the learning algorithms and improves its accuracy. Domain knowledge can be used to select the features [11], but there is also possibility of missing relevant features in the absence of domain expertise. Clustering based technique could be one of the ways to achieve feature selection. We propose a novel approach in which rules generate by apriori algorithm is used as an indicator to select the set of features. We have carried out experiment on different research papers discussing ERP stories. All these research papers are based on ERP implementation strategies in different enterprises. The research papers are classified based on their topics and questionnaire is prepared which consists of questions based on the critical success factors. There are at least 10-15 factors which are observed as crucial. The prosed technique selects main factors based on quantitative measure. These factors are mutually exclusive and rest factors are dependent on the selected factors. So if selected critical factors are removed then ERP performance can be improved. . The factors really not significant in the improving ERP performance can ignore. It's based on clustering and to asses clustering quality entropy is used.

II. Background

Table of LR

Author's year	Parameters	Challenges
Varma, V. A., Reklaitis, G. V., Blau, G. E., & Pekny, J. F. (2007). Enterprise-wide modeling & optimization—An overview of emerging research challenges and opportunities. *Computers & Chemical Engineering*, *31*(5), 692-711.	1) Strategic and Tactical decision making 2) Integration of dynamic resource allocation and scheduling of decisions	[1] Non comprehensive model can lead non-uniform methodologies in various department of organization. [2] Project productivity can be decelerated by varying financial and inaccurate physical resource allocation to processing ERP activities.
Ehie, I. C., & Madsen, M. (2005). Identifying critical issues in enterprise resource planning (ERP) implementation. *Computers in industry*, *56*(6), 545-557.	1) Top management support 2) Cost/budget 3) Feasibility/evaluation of ERP project 4) Consulting services	[1] Delay in decision marking [2] Improper distribution of tangible and intangible cost [3] Unsuccessful integration and modularization [4] Selection of inefficient consultant .Casual approach while changing consultant.
Hwang, C. L., & Yoon, K. (2012). *Multiple attribute decision making: methods and applications a state-of-the-art survey* (Vol. 186). Springer Science & Business Media.	1) Intention to Use / j Use and User Satisfaction 2) Information Quality 3) System Quality 4) Service Quality	[1] Exploited system functionality [2] Integrated Quality information is not generated Absence/inadequate a) Flexibility b) Interoperability c) Usability [3] Reliability is hampered. Unavailability of services.
Seethamraju, R. (2015). Adoption of software as a service (SaaS) enterprise resource planning (ERP) systems in small and medium sized enterprises (SMEs). *Information Systems Frontiers*, *17*(3), 475-492.	1) Vendor-related factors 2) Technology-related factors: 3) Organization-related factors:	[1] Vendor does not carry good reputation. Value is not created with co-relations. [2] Configurability of software is hampered. [3] IT-readiness and Change management ability is not there.

Technology advancement has led to exponential growth in the data with respect to dimensionality and sample size. Manual processing for these datasets is really impractical. Data mining and machine learning tools were proposed to automate pattern recognition and knowledge discovery process. Knowledge discovery from this data really requires storage and processing, which is a great challenge in the field of pattern recognition, statistics, and data mining. However, using data mining techniques on the collected data is mostly useless due to the high level of noise, missing, irrelevant attributes, variations etc. associated with collected samples[9]. The reason is either imperfection in the technologies that collected the data or the nature of the source of this data itself e.g. datasets crawled from the internet, are noisy by nature because it generally contains grammatical mistakes, misspelling, and improper punctuation. So extracting useful knowledge from such huge and noisy datasets is need of decision makers and challenge in front of researchers.[16]

Dimensionality reduction is one popular technique to remove noisy (i.e. irrelevant) and redundant attributes. This preprocessing technique can be categorized mainly into feature extraction and feature selection. Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Singular Value Decomposition (SVD), are few of widely used feature extraction techniques. On the other hand, the feature selection approach aims to select a small subset of features that minimize redundancy and maximize relevance to the target (i.e. class label). Popular feature selection techniques include Relief, Fisher Score etc[4]. Information Gain, Chi Squares are test to evaluate selected features [1]. There are two main models of FS i.e Wrapper and Filter .Wrapper model uses classifier and a strategy. Wrapper is more superior to filter model in terms of classification accuracy but it is expensive and biased to the chosen classifier, so filter model is used for large datasets, which uses certain criteria to filter out the features.

Aprioriis an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis [14]. The wide variety is also due to wide range of applications of clustering approach. It is used to identify the Data distribution, Outlier detection, Data reduction, Summarization, Preprocessing for regression, PCA, Classification, finding characteristics/patterns for each group. Thus clustering is used for data reduction which is a preprocessing step for other algorithms. [10]

There are several issues that need to be addressed before using the clustering approach for a given Dataset. The interpretability and usability of results heavily depends on the cluster quality and several measures need to be taken to improve cluster quality. The noisy data, high dimensionality, input order, data characteristics such as

types of attributes are some of the factors that affect quality of clusters and choosing the right algorithm and resolving some of these issues plays an important role in the success of clustering technique for decision making.[6] Clustering methods can be developed to learn more accurate cluster labels of the input samples, which guide feature selection simultaneously. Meanwhile, the cluster labels are also predicted by exploiting the hidden structure shared by different features, which can uncover feature correlations to make the results more reliable [19]. Choosing the initial centroids is always issue for all types of clustering. A careful and comprehensive study of data is required for the same. Also, if the initial clusters are not properly chosen, then after a few iterations it is found that clusters may even be left empty.[20]

If Feature selection technique is applied before applying any algorithm on sample data, then it reduces the cost and improves performance of learning of an algorithm by reducing the number of attributes. K-means Clustering is an unsupervised learning data mining technique that groups objects based upon distance or similarity[1][5]. Feature selection is used with k-means clustering algorithm and applied on dataset having 14 attributes in [8]. The result obtained when K-means clustering (Unsupervised grouping algorithm) is applied with feature selection is 50 % more accurate than just applying K-means clustering algorithm without using feature selection. Different algorithms are studied with respect to different parameters in literature and it is observed that improving stability of an algorithm by reducing noise remains a challenge. Applying iterative algorithm to identify maximum relevant feature set is used in[2]. Feature subset selection method iteratively selects subset of features from the database. Results based on 140,000 records show that the performance of the method exceeds by 70% than those of the other methods. The feature selection algorithm works in the iterations to find out feature subset. Error gets reduced at each iteration. Error calculation at each iteration indicates that, it is being minimized to get accuracy in the result.[3][7][12]. Feature selection method is mainly used to improve quality of clusters. Initialization parameters for identifying different clusters are an unresolved issue. Generally many search strategies are being used for feature selection. Existing feature selection algorithms involve search strategies which reduces degree of optimality of final feature subset.

III. Data Collection and Pre-processing:

Collected dataset is of different research papers. If the entire data set is given as an input it not only needs more time to run the experiment but also produces inaccurate results. Decision support system can be fast, if representative and important data from input dataset is being chosen. Data consists of 5 X 34 dimensions and selected dimensions are 5. Base set attributes selected from domain knowledge.(csv file)

IV. Selection of clustering algorithm

We carried out extensive literature survey of different clustering algorithms.. Clustering algorithms depends on nature of data .So Apriori algorithms used to find out important critical success factors as well as interdependent factors.

VII. Table 1 shows the rules generated by Weka using apriori algorithm

=== Run information ===
Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: ERP paper questions
Instances: 5
Attributes: 34
=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.85 (4 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 3
Best rules found:
1. Type of ERP=FALSE 4 ==> Identification of objective=FALSE 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
2. Identification of objective=FALSE 4 ==> Type of ERP=FALSE 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
3. Identification of degree of modification in management plan=FALSE 4 ==> Identification of objective=FALSE 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
4. Identification of objective=FALSE 4 ==> Identification of degree of modification in management plan=FALSE 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
5. Integration of ERP with other Softwares=FALSE 4 ==> Identification of objective=FALSE 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
6. Identification of objective=FALSE 4 ==> Integration of ERP with other Softwares=FALSE 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)

7. Identification of Upgradation ability of ERP software=FALSE 4 ==> Identification of objective=FALSE 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
8. Identification of objective=FALSE 4 ==> Identification of Upgradation ability of ERP software=FALSE 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
9. Identification of Deadline=FALSE 4 ==> Identification of objective=FALSE 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
10. Identification of objective=FALSE 4 ==> Identification of Deadline=FALSE 4 <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)

Table 2: Interdependent factors

Total Factors	**Important factors**
Identification of objective	Type Of ERP
Identification of expertise of team member	Identification of Objective
Identify degree of customization of ERP modules	Identification of degree of modification in management plan
Identifying correct vendor	Identification of Deadline
Identify delay in decision making by management	Integration of ERP with other Softwares
Identify user interference at pre implementation stage	
Choice of efficient consultant	
Improper identification of problems	
Identification of technical failures	
Team members communication	
Identification of incorrect methods/approaches	
Choosing correct vendor	
Type of ERP	
Impact of external environment	
Identification of extent of management support	
Identification of degree of modification in management plan	
ERP Training to the employees	
Identification of resources	
System Design of ERP Software	
Identify service level agreements	
Integration of ERP with other Softwares	
ERP Packet Selection	
Cost of implementation	
Interruption in working processes	
Company Size	
Identification of Upgradation ability of ERP software	
Identification of Deadline	
Identifying Strategies for work processes	
Reduction in Inventory	
Centralization of audit process	
Maintenance Reduction	
Visibility of information	
Customer Responsiveness	

V. Conclusion

Any application where clustering is to be used on input dataset can adopt this generic approach. In literature multiple critical factors are focused in order to get successful ERP implementation. Most of them are dependent on each other. This dependency is removed by quantifying the factors. It will help to get effective and fast decision making from ERP software. Quantification of critical success factors will help to improve ERP process and cluster validity parameters proves the efficiency of the technique.

References

[1] S Alelyani, J Tang, H Liu,Feature Selection for Clustering: A Review. public.asu.edu 2013.
[2] Ritu Ganda ,Vijay Chahar ,A Comparative Study on Feature Selection Using Data Mining Tools, International Journal of Advanced Research in Computer Science and Software Engineering, 2013 .

[3] Ren Diao ,Qiang Shen ,Feature Selection With Harmony Search ,IEEE Systems, Man, and Cybernetics, 2012.
[4] Salem Alelyani, Thesis On Feature Selection Stability: A Data Perspective,2012.
[5] SunitaBeniwal, Jitender Arora Classification and Feature Selection Techniques in Data Mining, International Journal of Engineering Research & Technology (IJERT),Vol. 1 Issue 6, August – 2012.
[6] Gauthier Doquire and Michel Verleysen,s,Mutual information for feature selection with missing datacomputational Intelligence and Machine Learning,2011.
[7] Farahat, A.K. ,Data Mining (ICDM), An Efficient Greedy Method for Unsupervised Feature Selection,IEEE,2005.
[8] Ren Diao ,Qiang Shen, Two New Approaches to Feature Selection with Harmony Search, IEEE World Congress on Computational Intelligence ,2010.
[9] M. Ramaswami and R. Bhaskaran,A Study on Feature Selection Techniques in Educational Data Mining,journal of computing,2009.
[10] Liu, H. Torkkola, K. ,Evolving feature selection ,Intelligent Systems, IEEE,2005.
[11] Huan Liu and Lei Yu, Toward Integrating Feature Selection Algorithms for Classification and Clustering,
[12] Hiroshi Mamitsuka ,Principles of Data Mining and Knowledge Discovery ,Lecture Notes in Computer Science springer 2002.
[13] Lesh, N. MERL, Zaki, M.J.,Scalable feature mining for sequential data, Intelligent Systems and their Applications, IEEE (Volume:15 , Issue: 2) ,2000.
[14] Xiaofei He, Ming Ji, Chiyuan Zhang, and HujunBao,A Variance Minimization Criterion to Feature Selection Using Laplacian Regularization, IEEE Transaction on Pattern Analysis and machine Intelligence, VOL. 33, NO. 10, 2011.
[15] L.V. Bijuraj,Clustering and its Applications ,Proceedings of National Conference on New Horizons in IT - NCNHIT ,2013.
[16] Lan H. Witten, Eibe Frank, Data miming:practical machine tools and techniques.
[17] HuanLiu,Evolving Feature Selection,Intelligent Systems, IEEE (Volume:20 , Issue: 6) ,2005.
[18] Jiawei Han, Micheline Kamber ,Data Mining: Concepts and Techniques .
[19] Zechao Li et al , Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection ,IEEE Transactions on Knowledge and Data Engineering (TKDE),2014.
[20] Parul Agarwal ,M. AfsharAlam,RanjitBiswas,Issues, Challenges and Tools of Clustering Algorithms , IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011.
[21] Prafulla Bafna1 , Hema Gaikwad , A hybrid approach to measure design improvement factor of website., American International Journal of Research in Science, Technology, Engineering & Mathematics" (ISSN (Print): 2328-3491, ISSN (Online): 2328-3580, ISSN (CD-ROM): 2328-362) December-2014 to February-2015.
[22] A novel clustering approach to select optimal usability principles for educational websites ,,Prafulla Bafna, International Journal of Software and Web Sciences , ISSN (Print): 2279-0063, ISSN (Online): 2279-0071, Issue 11, December-2014 to February-2015 Sciences(IJSWS).
[23] Prafulla Bafna,, Pravin Metkewar, ShailajaShirwaikar, ,Novel Clustering approach for Feature Selection , American International Journal of Research in Science, Technology, Engineering & Mathematics,pp62-67,2014
[24] Lewandowski, J., Salako, A. O., & Garcia-Perez, A. (2013, September). SaaS enterprise resource planning systems: challenges of their adoption in SMEs. In e-Business Engineering (ICEBE), 2013 IEEE 10th International Conference on (pp. 56-61). IEEE.
[25] Singla, A. R. (2008). Impact of ERP systems on small and mid sized public sector enterprises. Journal of Theoretical and Applied Information Technology, 4(2), 119-131.