Expert Systems with Applications 39 (2012) 12358–12370

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An intelligent decision support system for forecasting and optimization of complex personnel attributes in a large bank

Ali Azadeh [a,*], Morteza Saberi [b], Zahra Jiryaei [a]

[a] Department of Industrial Engineering, College of Engineering, University of Tehran, P.O. Box 11365-4563, Iran
[b] Institute for Digital Ecosystems & Business Intelligence, Curtin University of Technology, Perth, Australia

ARTICLE INFO

Keywords:
Decision support system
Personnel efficiency
Data mining
Data envelopment analysis
ANN
K-Means algorithm
Optimization

ABSTRACT

Developing decision support system (DSS) can overcome the issues with personnel attributes and specifications. Personnel specifications have greatest impact on total efficiency. They can enhance total efficiency of critical personnel attributes. This study presents an intelligent integrated decision support system (DSS) for forecasting and optimization of complex personnel efficiency. DSS assesses the impact of personnel efficiency by data envelopment analysis (DEA), artificial neural network (ANN), rough set theory (RST), and K-Means clustering algorithm. DEA has two roles in this study. It provides data to ANN and finally it selects the best reduct through ANN results. Reduct is described as a minimum subset of features, completely discriminating all objects in a data set. The reduct selection is achieved by RST. ANN has two roles in the integrated algorithm. ANN results are basis for selecting the best reduct and it is used for forecasting total efficiency. Finally, K-Means algorithm is used to develop the DSS. A procedure is proposed to develop the DSS with stated tools and completed rule base. The DSS could help managers to forecast and optimize efficiencies by selected attributes and grouping inferred efficiency. Also, it is an ideal tool for careful forecasting and planning. The proposed DSS is applied to an actual banking system and its superiorities and advantages are discussed.

1. Introduction

Generally, data mining is the process of analyzing and summarizing data from different viewpoints into valuable information. This area presents new theories and methods for processing large volumes of data and has obtained noteworthy consideration among researchers. Several immeasurable influences and complex relationships among attributes impact efficiency in organizations. Rough set theory (RST) proposed by Pawlak, is one of the techniques for the identification and recognition of common patterns in data (Pawlak, 1982, 1991). This technique has found applications in knowledge discovery from data bases, data mining, fault diagnosis, machine learning, knowledge acquisition, expert systems and decision support systems (Błaszczyński, Greco, & Słowiński, 2007; Fan, Liu, & Tzeng, 2007; Inuiguchi & Miyajima, 2007). It is also used to study uncertainty (Beynon & Peel, 2001; Lili & Zhi, 2001; Ton Su & Hsu, 2006; Ziarko, 1993), prediction (Becerra-Fernandez, Zanakis, & Walczak, 2002; Kusiak & Tseng, 2000; Sanchis, Segovia, Gil, Heras, & Vilar, 2007), service organizations (Chou, Cheng, & Chang, 2007; Hassanien, 2007; Kowalczyk & Slisser, 1997; Sikder & Gangopadhyay, 2007; Tsumoto, 1997), financial firms (Ravi Kumar & Ravi, 2007; Ruhe, 1996; Shyng, Wang, Tzeng, & Wu, 2007), and scheduling problems (Liu, Chen, Wu, & Li, 2006; Triantaphyllou, Liao, & Iyengar, 2002).

Efficiency is a key concept for financial institutions. As personnel specifications have greatest impact on efficiency, they can help us designing work environments for maximizing efficiency. Providing information on multiple input and output factors are a complicated and time consuming procedure. Developing expert system in this situation is hard. So, available attributes must be reduced. Rough set theory is a candidate for this. At the present, the study on rough set theory is focusing on feature selection techniques with much success. Stefanowski and Slowinski have studied rough sets as a tool for feature selection by studying attribute dependencies (Stefanowski & Slowinski, 1997). Kusiak and Tseng have proposed two independent algorithms for accurate feature selection in medical, industrial and engineering case studies (Kusiak, Kern, Kernstine, & Tseng, 2000; Kusiak & Tseng, 2000). Others like Xia and Wu discusses feature extraction technique of rough set theory for supplier selection to select best suppliers according to different tangible and intangible attributes (Xia & Wu, 2007). Moreover, there are some other application of rough set theory to feature selection in customer relationship management (Tseng & Huang, 2007), product quality evaluation (Zhai, Khoo, & Fok, 2002) and healthcare (Xiangyang, Jie, Jensen, & Xiaojun, 2006). However, existing heuristic rough set approaches to feature selection are insufficient at finding optimal reductions. On the other hand, it is

* Corresponding author.
E-mail address: aazadeh@ut.ac.ir (A. Azadeh).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.04.056

not feasible to search for optimal in even in average sized datasets. Therefore, the combination of this method by other robust data mining tools may help practitioners to go further into feature selection to obtain more accurate results.

In this paper, a new DSS approach for feature selection and forecasting and optimization of personnel efficiency amongst various branches within a large bank is introduced. It is accomplished by integration of Data Mining tools (RST, ANN, MLP, GA, and CVTT), DEA and K-Means algorithm is proposed. By using ANN and DEA, the integrated approach decides on which feature subsets (reducts) produced by rough set theory are more important to decision making procedure. ANN and DEA have too many applications in engineering case studies (Al-Omari & Al-Jarrah, 2004; Azadeh, Amalnick, Ghaderi, & Asadzadeh, 2007; Azadeh, Ghaderi, Anvari, & Saberi, 2006; Azadeh, Ghaderi, Anvari, & Saberi, 2007; Azadeh, Ghaderi, Anvari, Saberi, & Izadbakhsh, 2006; Azadeh, Ghaderi, & Izadbakhsh, 2007; Azadeh, Ghaderi, & Sohrabkhani, 2007; Azadeh, Ghaderi, Tarverdian, & Saberi, 2006; Azadeh, Ghaderi, Tarverdian, & Saberi, 2007; Fonseca & Navaresse, 2002).

After selecting the best reduct, construction of DSS is initiated. At first, K-Means algorithm is used for grouping efficiency values. Then, the preferred ANN is executed with arbitrary values of inputs. With proposed rule base, the position of this output is determined in available groups and a new group is constructed.

This study is an extension of a previous study by Azadeh, Saberi, Reza, and Leili (2011). Furthermore the previous study presented an integrated data envelopment analysis-artificial neural network-rough set algorithm for assessment of personnel efficiency. However, this study presents a DSS approach based on previous approaches and K-Means algorithm for optimization of personnel attributes. Moreover, the intelligent DSS approach utilizes data envelopment analysis (DEA), and data mining tools including rough set theory (RST), artificial neural network (ANN), cross validation test technique (CVTT), and K-Means algorithm for forecasting and optimization of personnel efficiency. The paper is organized as follows: DEA, ANN, RST and CVTT are discussed in the first section. The methodology or the intelligent DSS is discussed in the second section. Third section explains the details of experimentations and results of DSS. Section four deals with the execution of K-Means algorithm which is the core of DSS. Finally, last section presents the conclusions of this study.

1.1. Data envelopment analysis

DEA is a non-parametric method that uses linear programming to calculate the efficiency in a given set of decision-making units (DMUs). The DMUs that make up a frontier envelop, the less efficient firms and the relative efficiency of the firms is calculated in terms of scores on a scale of 0–1, with the frontier firms receiving a score of 1. DEA models can be input or output oriented and can be specified as constant returns to scale (CRS) or variable returns to scale (VRS).

1.2. Basic models of DEA

The original fractional CCR model (1) evaluates the relative efficiencies of n DMUs ($j = 1,\dots,n$), each with m inputs and s outputs denoted by $x_{1j}, x_{2j},\dots,x_{mj}$ and $y_{1j}, y_{2j},\dots,y_{sj}$, respectively (Charnes, Cooper, & Rhodes, 1978). This is done so by maximizing the ratio of weighted sum of output to the weighted sum of inputs:

$$\begin{aligned}
&\text{Max} \quad \theta = \frac{\sum_{r=1}^{S} u_r y_{ro}}{\sum_{r=1}^{m} v_i x_{io}},\\
&\text{s.t.} \quad \frac{\sum_{r=1}^{S} u_r y_{rj}}{\sum_{r=1}^{m} v_i x_{ij}} \le 1, \quad j = 1,\dots,n, \quad r = 1,\dots,s,\\
&u_r,\ v_i \ge 0, \quad i = 1,\dots,m, \quad r = 1,\dots,s.
\end{aligned} \tag{1}$$

In model (1), the efficiency of DMU_o is θ_o and u_r and v_i are the factor weights. However, for computational convenience the fractional programming model (1) is re-expressed in linear program (LP) form as follows:

$$\begin{aligned}
\text{Max} \quad & \theta = \sum_{r=1}^{s} u_r y_{r0},\\
\text{s.t.} \quad & \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1,\dots,n,\\
& \sum_{i=1}^{m} v_i x_{io} = 1,\\
& u_r,\ v_i \ge \varepsilon, \quad i = 1,\dots,m, \ r = 1,\dots,s,
\end{aligned} \tag{2}$$

where ε is a non-Archimedean infinitesimal introduced to ensure that all the factor weights will have positive values in the solution. The model (3) evaluates the relative efficiencies of n DMUs ($j = 1,\dots,n$), respectively, by Minimizing inputs when outputs are constant. The dual of linear program (LP) model for input oriented CCR is as follows:

$$\begin{aligned}
\text{Min} \quad & \theta,\\
\text{s.t.} \quad & \theta x_{io} \ge \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1,\dots,m,\\
& y_{ro} \le \sum_{j=1}^{n} \lambda_j y_{rj} \quad r = 1,\dots,s,\\
& \lambda_j \ge 0.
\end{aligned} \tag{3}$$

The output oriented CCR model is as follows:

$$\begin{aligned}
\text{Max} \quad & \theta,\\
\text{s.t.} \quad & x_{io} \ge \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1,\dots,m,\\
& \theta y_{ro} \le \sum_{j=1}^{n} \lambda_j y_{rj} \quad r = 1,\dots,s,\\
& \lambda_j \ge 0,
\end{aligned} \tag{4}$$

If $\sum \lambda_j = 1 (j = 1,\dots,n)$ is added to model (3), the BCC model is obtained which is input oriented and its return to scale is variable. The calculations provide a maximal performance measure using piecewise linear optimization on each DMU with respect to the closest observation on the frontier. The linear programming system for the BCC input-oriented model is given in expression (5), and the output-oriented model in expression (6) (refer to Charnes et al. (1994) for more detail (Charnes et al., 1978).

$$\begin{aligned}
\text{Min} \quad & \theta,\\
\text{s.t.} \quad & \theta x_{io} \ge \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1,\dots,m,\\
& y_{ro} \le \sum_{j=1}^{n} \lambda_j y_{rj} \quad r = 1,\dots,s,\\
& \sum_{j=1}^{n} \lambda_j = 1,\\
& \lambda_j \ge 0, \quad j = 1,\dots,n,
\end{aligned} \tag{5}$$

$$\begin{aligned}
\text{Max} \quad & \theta,\\
\text{s.t.} \quad & x_{io} \ge \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1,\dots,m,\\
& \theta y_{ro} \le \sum_{j=1}^{n} \lambda_j y_{rj}, \quad r = 1,\dots,s,\\
& \sum_{j=1}^{n} \lambda_{j=1},\\
& \lambda_j \ge 0, \quad j = 1,\dots,n.
\end{aligned} \tag{6}$$

1.3. Artificial neural networks

An artificial neural network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. Although ANNs arose to model the brain, they have been applied when there is no theoretical evidence about the functional form. In this way, ANNs are data-based, not model-based. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. ANNs are normally arranged in three layers of neurons, the so-called multilayer structures are input, hidden and output layers. Input layers which are neurons (also called nodes or processing units) introduce the model inputs. Hidden layers combine the inputs with weights that are adapted during the learning process. Output layer provides the estimations of the network. Multi layer perceptron (MLP) which is the most important ANNs is introduced in the Appendix I.

1.3.1. Neural network modelling

Usually train data set contains 70–90% of all data and remaining data are used for test data set (Azadeh, Ghaderi, Anvari, et al., 2007). One of the problems that occur during neural network training is called overfitting. The error on the training set is driven to a very small value, but when new data is presented to the network the error is large. The network has memorized the training examples, but it has not learned to generalize to new situations. Early stopping method is used for this problem. In this method the available data is divided into three subsets. The first subset is the training set, which is used for computing the gradient and updating the network weights and biases. The second subset is the validation set. The error on the validation set is monitored during the training process. The validation error will normally decrease during the initial phase of training, as does the training set error. However, when the network begins to over fit the data, the error on the validation set will typically begin to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are returned. The test set error is not used during the training, but it is used to compare different models. It is also useful to plot the test set error during the training process. If the error in the test set reaches a minimum at a significantly different iteration number than the validation set error, this may indicate a poor division of the data set.

2. Method: the DSS

Several immeasurable influences and complex relationships among attributes impact human efficiency in organizations. Efficiency relevant to human attributes is a goal that is rarely questioned in contemporary organizations. As personnel specifications have greatest impact on efficiency, they can help us designing work environments for maximizing efficiency. The DSS described in this section is capable of forecasting and optimization of personnel efficiency. The DSS is applicable for all problems associated with decision making in organizations composed of decision making units (DMUs) and will be a valuable asset for executives and senior managers. The stages involved in the proposed algorithm are illustrated in Fig. 1. The components of the DSS are presented in the following sections.

2.1. Efficiency calculation by DEA

This stage is involved with efficiency calculation of DMUs. Efficiency is a key concept for large organizations and hence DEA is utilized in the integrated approach. DEA is a multi-factor analysis tool that measures the relative efficiencies of a set of DMUs and its benefits are clearly understood. It effectively considers multiple inputs and output factors in computing the efficiency scores. As efficiency scores vary on different selection of inputs and outputs, we should utilize an accurate DEA specification for each particular case (Serrano Cinca, Mar Molinero, & Chaparro García, 2002).

2.2. Decision system

A data set or information system is a table, where each row indicates an object and every column shows an attribute that can be measured for each object. The input features are called conditional and the output is called decision attribute respectively. In this study, the information system is comprised of decision making unites as its objects and efficiency as decision attribute. Conditional attributes vary according to each specific case of decision making, to measure their impact on DMU's efficiencies.

2.3. Data pre-processing

Pre-processing is a fundamental stage in whole modeling process. The accuracy of the final constructed model primarily depends on how well this stage is developed. Incomplete,

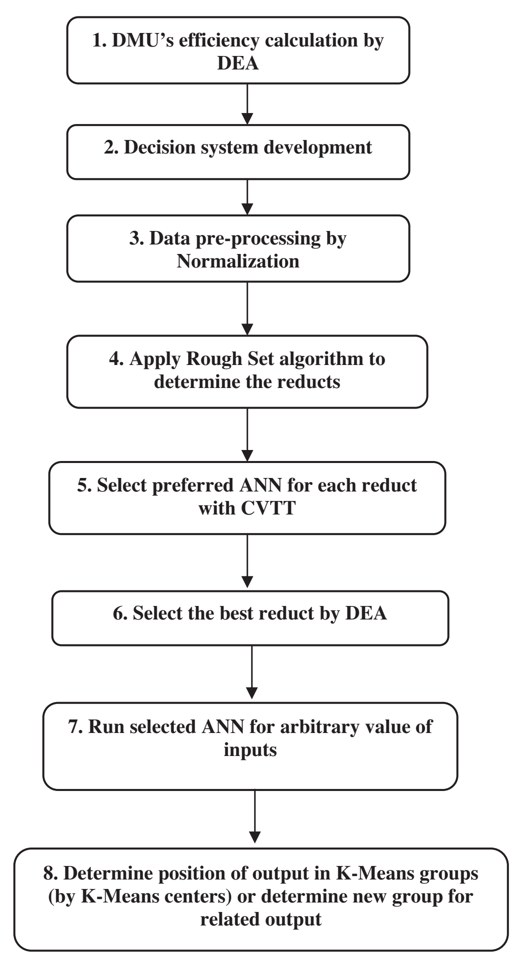

Fig. 1. The intelligent DSS for optimization of personnel efficiency.

inconsistent, noisy and large data sets are problems, highlighting the need for pre-processing. We have utilized clustering methods, rough set theory, ANN and cross validation test technique as data mining tools in our pre-processing stage. Further explanation of this stage is discussed in the following sections.

2.4. Rough set theory: feature selection

The central concept of feature selection technique is called reduct that can be described as a minimum subset of features, completely discriminating all objects in a data set. Decision tables with input and output features are the best case for reduct application. In this situation we compute reducts according to the value of the output or decision features. Considering the data in Table 1 with four input features (F1–F4) and one output feature with three possible answers, we can obtain a reduct containing the single feature of F2 that can cluster decision class into two subsets of objects according to F2 value in data set. One of the goals of feature selection technique is to find reducts with the smallest number of features. Total number of reducts in an information system with m attributes may be equal to $\binom{m}{[m/2]}$. Computing all reducts cannot be done simply and is Np-hard problem (Banker, Charnes, & Cooper, 1984). But, fortunately, good heuristic algorithms are developed for this reason that genetic algorithm is the most famous one (Cheng & Titterington, 1994).

2.5. Selecting preferred ANN

2.5.1. Test by ANN and CVTT

Much of the emphasis here is selecting a good subset of conditional features according to different data sets constructed by feature selection and extraction algorithms. As all reducts have classification quality of 100% in training data set, we may differentiate their performance in the form of accuracy of prediction on unseen data. This approach can help us to identify best subset of parameters; effecting DMUs efficiencies the most. To estimate critical parameters, ANN and CVTT, including training a network several times are used to provide reliable information for ranking the reducts.

To estimate the quality of constructed neural network we have employed cross validation test technique (CVTT). CVTT is often used as a method for evaluation of classification models. It comprises of several training and testing runs. The data set is first split into several parts. Then, one part is utilized for testing and the rest are saved for training purpose. These steps are repeated until all parts used as testing set. The final product of CVTT is the mean accuracy of total runs.

2.5.2. Selecting the preferred ANN

As discussed by Cybenko and Patuwo et al., a single hidden layer is sufficient in constructing neural nets (Cybenko, 1989; Patuwo, Hu, & Hung, 1993). Therefore a single hidden layer neural network is selected in this study. To find the appropriate numbers of hidden nodes in ANN analysis of each reduct, following steps are performed to construct networks with one to q nodes, where q is an optional parameter and will be changed until the desired goal error met by the algorithm[1]:

- Training step, using scaled conjugate gradient training algorithm (Moller, 1993).
- Evaluate the model using the test data and obtaining MAPE[2] error.

Where,

– Variable $MAPE_{ijk}$ is defined as:

$$Mean\ absolute\ percentage\ error\ of\ ANN,\ with\ regard\ to\ kth\ part\ of\ data,\ used\ as\ test\ data\ and\ j\ node\ in\ hidden\ layer,\ both\ related\ to\ ith\ reduct.$$

– Variable ERR_{ij} is defined as:

$$ERR_{ij} = Average\{MAPE_{ijk} : k = 1, 2 \ldots\}\ or\ average\ of\ MAPE\ in\ all\ constructed\ ANN\ for\ ith\ reduct\ with\ regard\ to\ j\ node\ in\ hidden\ layer.$$

– $AERR_i$ is defined as:

$$AERR_i = Average\ (ERR_{ij}).$$

– $VarERR_i$ is defined as:

$$VarERR_i = Variance\ (ERR_{ij}).$$

– $MaxERR_i$ is defined as:

$$MaxERR_i = Max\ (ERR_{ij}).$$

– $MinERR_i$ is defined as:

$$MinERR_i = Min\ (ERR_{ij}).$$

Finally these variables will use to rank the performance of each reduct according to their constructed ANN.

2.6. Selection of the preferred reduct

We have utilized DEA method instead of more time-consuming machine learning techniques to rank reducts in this section. As we do not know which attributes are more important for efficiency analysis, it makes sense to use this method. Since mean scores of error (*AERRi*) alone may not be appropriate in ranking reducts thus we need to utilize other variables defined previously for constructed ANN, involving variance (*VarERRi*), minimum (*MinERRi*) and maximum (*MaxERRi*) of error rate for each reduct. DEA method will effectively take into account these variables to select a good subset of features after calculating reduct's efficiency score. We treated scores of *AERRi*, *VarERRi*, *MaxERRi* and *MinERRi* for each reduct as inputs of efficiency with the specific output of 1 to calculate efficiency of each reduct.

2.7. K-Means algorithm

K-Means Algorithm is used to identify homogenous groups for the branches. K-Means clustering can best be described as a partitioning method. That is, partitions the observations in data into K mutually exclusive clusters. K-Means uses an iterative algorithm

Table 1
Five-object data set.

Object No.	F1	F2	F3	F4	D
1	0	1	0	2	0
2	1	1	0	2	2
3	0	0	0	1	0
4	0	1	1	0	1
5	0	0	1	3	0

[1] In this study the value of the desired minimum error has been defined between 6% and 8% (94–92% confidences) and the value of q has been defined 20. The error is estimated by mean absolute percentage error (MAPE).

[2] Mean absolute percentage error $MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\text{Actualvalue}_i - \text{Setpointvalue}_i}{\text{Setpointvalue}_i}\right|$ (N : the *number of rows*).

that minimizes the sum of distances from each object to its cluster centroid, over all clusters. This algorithm moves objects between clusters until the sum cannot be decreased further. The result is a set of clusters that are as compact and well-separated as possible. Without some knowledge of how many clusters are really in the data, it is a good idea to experiment with a range of values for k. Silhouette plot or value could use to determine best number of cluster. The silhouette value for each point is a measure of how similar that point is to points in its own cluster compared to points in other clusters, and ranges from −1 to +1. It is defined as:

$$S(i) = (\min(b(i,:),2) - a(i))./\max(a(i),\min(b(i,:),2)) \tag{7}$$

where $a(i)$ is the average distance from the ith point to the other points in its cluster, and $b(i,k)$ is the average distance from the ith point to points in another cluster k. Finally we calculate average of silhouette value for all of point and compare each cluster with this parameter. This parameter is called *mean silhouette*. Cluster with highest *mean silhouette* value is selected. Branch's efficiency is used as K-Means algorithm input. Thus, the center of each created group is one dimensional.

2.8. Run selected ANN for arbitrary value of inputs

This step is related to manager and each manager could run the selected ANN in each reduct for arbitrary inputs value. In fact, manager is eager to know the efficiency of bank when encountering with various conditions in his mind. Related output is called as *out value*. Also, managers prefer linguistic outputs than numerical outputs. In next section, an algorithm is proposed to satisfy this preference.

2.9. Optimum position of outputs

With regard to the number of cluster that yields after running K-Means, created groups are called as $Group_1$, $Group_2$... and $Group_k$. K is the number of created groups. Entitling is done in which state that $center_i < center_{i-1}$. Reader should note that $center_i$ is the position of $Group_i$ on the x axis. Following rule is defined to determine output degree of membership in related groups:

1. *If Lower < out value < Upper then*
   $degmemoutvalue_i = 1 - [(out\ value - center_i)/(center_{i-1} - center_i)]$
   And
   $degmemoutvalue_{i-1} = 1 - [(center_{i-1} - out\ value)/(center - center_{i-1})]$
2. *If Upper < out value then*
   $degmemoutvalue_1 = 1 - [(outvalue - center_1)/c]$
3. *If out value < Lower then*
   $degmemoutvale_k < 1 - [(center_k - out\ value)/c]$

That
$degmemoutvalue_i$*: degree membership of out value in* $Group_i$
Lower: min $\{center_i: 1 < i < k\}$
Upper: max $\{center_i: 1 < i < k\}$

Also, following rule is defined to determine position of output in K-Means group or determine new group for related output:

1. If $Lower < out\ value < Upper$ and $degmemoutval_i > 0.7$ or $degmemoutval_i < 0.3$ then yield output belongs to Group I and Group I-1 with mentioned degree membership. Else, yielded output belongs to new group that is between Group I and Group I-1.
2. If $Upper < outval$ and $degmemoutval_1 > 0$ then yield output belongs to Group1 with mentioned degree membership. Else, yielded output belongs to new group that is bigger than Group 1.
3. If $out\ value < Lower$ and $degmemoutval_1 > 0$ then yield output belongs to Group k with mentioned degree membership. Else, yielded output belongs to new group that is smaller than Group 1.

3. Experiments

The proposed approach of DSS is applied to a large private bank. Efficiency relevant to human attributes is a goal that is rarely questioned in contemporary organizations. As personnel attributes have greatest impact on efficiency, they can help us designing work environments for maximizing efficiency. The case study focuses on 102 branches to analyze the effect of personnel attributes on bank branches efficiency. Evaluation of the decision making framework with explanation of each stage are described by following sections.

Athanassopoulos (1997) discuses two models of intermediation and production for financial firms (Athanassopoulos, 1997). Intermediation institution collects deposits as input and placing loans as output in order to make profit. Examples of intermediation firms are presented by Berger and Humphrey (1991). Production model utilizes physical resources as input and collected deposits and loans as outputs which its examples are discussed by Soteriou and Zenios (1999). In this paper we have calculated efficiency on major features of production model according to the nature of financial firms in this bank. Table 2 shows DEA inputs and outputs and Table 3 shows efficiency scores calculated for 102 branches. Output-oriented BCC Model is used for efficiency calculation. It is based on maximization of the following objective function:

$$\begin{aligned} &\text{Max} \quad \theta, \\ &\text{s.t.} \quad x_{io} \geq \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1,\ldots,2, \\ &\qquad \theta y_{ro} \leq \sum_{j=1}^{n} \lambda_j y_{rj} \quad r = 1,\ldots,3, \\ &\qquad \sum_{j=1}^{n} \lambda_j = 1, \\ &\qquad \lambda_j \geq 0, \quad j = 1,\ldots,102 \end{aligned} \tag{8}$$

3.1. Decision system

We have identified four groups of personnel in each branch. First group are associated as tellers who conduct most of a bank's routine transactions. Among the responsibilities of tellers are accepting deposits, loan payments, and processing withdrawals. They also may sell savings bonds; accept payment for customers' utility bills, charge cards and process necessary paperwork for certificates of deposit. Some tellers specialize in handling foreign currencies or commercial or business accounts. The second group consists of supervisors who cashing checks and performing controlling task on performing transactions. Branch mangers and their assistants are in third and fourth groups. Each branch has one branch manager and may have several personnel assigned to other

Table 2
DEA inputs and outputs.

DEA inputs	DEA outputs
Number of employees Fixed assets	Deposits Operating income Loans

Table 3
Bank branch's efficiency scores.

Branch ID	BCC Eff.	Branch ID	BCC Eff.	Branch ID	BCC Eff.	Branch ID	BCC Eff.	Branch ID	BCC Eff.
1	1.00	22	0.96	43	0.93	64	0.87	85	0.94
2	1.00	23	1.00	44	0.9	65	0.87	86	0.89
3	1.00	24	0.9	45	0.81	66	0.91	87	0.88
4	1.00	25	0.92	46	1.00	67	0.92	88	1.00
5	1.00	26	1.00	47	0.70	68	1.00	89	0.91
6	1.00	27	1.00	48	1.00	69	0.93	90	0.84
7	1.00	28	0.93	49	0.71	70	0.91	91	0.91
8	1.00	29	0.89	50	1.00	71	0.89	92	0.82
9	1.00	30	1.00	51	1.00	72	0.87	93	0.82
10	1.00	31	0.95	52	0.95	73	0.93	94	0.80
11	0.98	32	1.00	53	0.72	74	0.93	95	0.94
12	1.00	33	0.92	54	1.00	75	0.93	96	0.74
13	0.94	34	0.96	55	0.89	76	1.00	97	0.81
14	1.00	35	0.95	56	0.94	77	0.89	98	0.72
15	1.00	36	0.86	57	0.91	78	0.93	99	0.82
16	1.00	37	0.98	58	1.00	79	0.95	100	0.69
17	0.88	38	0.99	59	0.96	80	0.82	101	0.88
18	0.94	39	0.89	60	1.00	81	0.99	102	0.71
19	1.00	40	0.97	61	0.94	82	0.92		
20	0.84	41	0.91	62	0.94	83	0.91		
21	0.92	42	1.00	63	1.00	84	0.85		

Eff. = Efficiency.

groups. We have recognized 28 conditional attributes of personnel attributes in this study with the decision attribute of efficiency. Personnel attributes are categorized in Table 4. Moreover Fig. 2 show the relationship between personnel groups and defined personnel specification.

3.2. Data pre-processing

Naturally, we perform the data pre-processing tasks by applying data normalization. Data in decision system are normalized as follows:

- $\textbf{\textit{Quantity}}: \xrightarrow{(Divided\ by)} \underline{\text{Number of personnel in each branch}}$
- $\textbf{\textit{Education}}: \xrightarrow{(Divided\ by)} \underline{\text{Number of personnel in each branch}}$
- $\textbf{\textit{Work experience}}: \xrightarrow{(Divided\ by)} \underline{\text{Maximum work experience}}$ $\underline{\text{existing in each branch}}$

As Age data are comparable by themselves, they did not need to be normalized.

3.3. Feature selection

In this step the reducts of personnel decision system are calculated. The reducts are generated by genetic algorithm (Soteriou & Zenios, 1999). Twelve reducts were extracted which are shown in Table 5. It can be simply seen that the number's of males has maximum frequency and maximum age of assistant to branch manager has minimum frequency in the reduct set.

3.4. Selection of the preferred ANN

Preferred ANN is selected with the aid of error variable. In order to calculate error value, CVTT is employed with four folds. The data set is firs split into 4 divisions and then one of the parts is taken as validation and test sample and the remainder becomes the training set. In this study the value of the desired minimum error has been defined between 6% and 8% (92–94% confidence) and the value of q has been defined 40. The error is estimated by mean absolute percentage error (MAPE). Table 6(a–l) shows error variables of *MAP-Eijk* calculated for each reduct versus number of nodes in hidden layer. Also, architect of each preferred ANN is shown in Fig. 3(a-l).

3.5. Best reduct

We have utilized DEA method instead to rank reducts in this section. Since mean scores of error (*AERRi*) alone may not be appropriate in ranking reducts thus we need to utilize other variables, involving variance (*VarERRi*), minimum (*MinERRi*) and maximum (*MaxERRi*) of error rate for each reduct. Error variables are shown in Table 6. DEA method will effectively take into account the values of these variables to select a good subset of features after calculating reduct's efficiency score. We treated scores of *AERRi*, *VarERRi*, *MaxERRi* and *MinERRi* for each reduct as inputs of efficiency with the specific output of 1 to calculate efficiency of each reduct.

Table 4
Personnel attributes of the 102 branches.

Quantity	Education	Age	Work experience
Number of male	Number of below diploma	Age of branch manager	Work experience of branch manager
Number of singles	Number of diploma	Average, maximum and minimum age of assistant to branch manager	Average, maximum and minimum work experience of assistant to branch manager
Number of tellers	Number of upper diploma	Average, maximum and minimum age of supervisor	Average, maximum and minimum work experience of supervisor
	Number of bachelor of science	Average, maximum and minimum age of Teller	Average, maximum and minimum work experience of Teller
	Number of master of science		

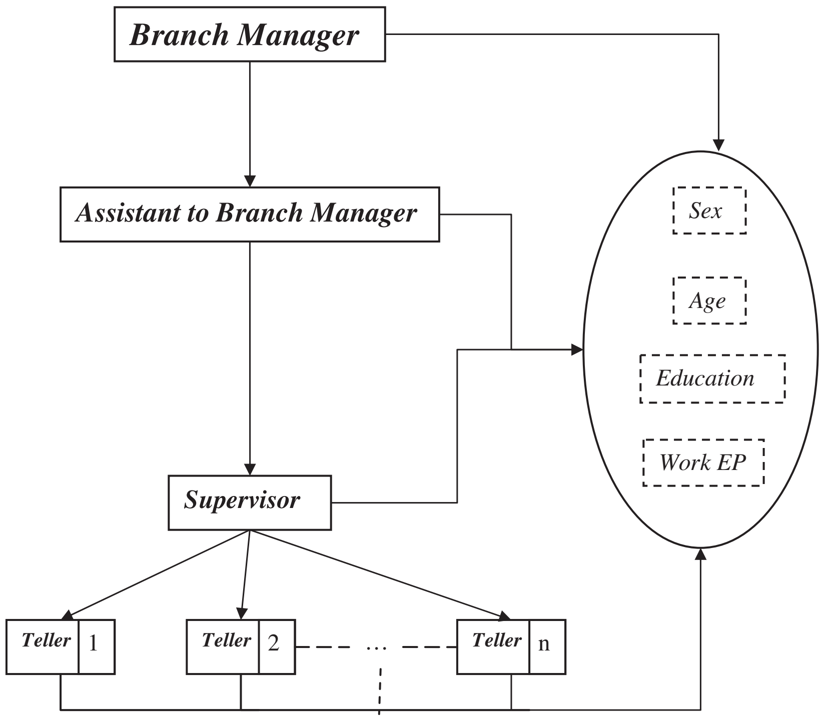

Fig. 2. Relationship between personnel groups and defined personnel specification.

Table 5
The set of reducts in the bank.

Reduct ID	Reducts	Reduct size
1	{Average work experience of supervisor, average age of teller, number of male}	3
2	{Number of male, average age of supervisor, average age of teller}	3
3	{Number of male, number of teller, average work experience of assistant to branch manager}	3
4	{Minimum age of teller, maximum work experience of teller, minimum work experience of assistant to branch manager}	3
5	{Number of male, work experience of branch manager, minimum work experience of supervisor}	3
6	{Number of male, maximum age of supervisor, maximum work experience of supervisor}	3
7	{Number of single, number of bachelor of science, average work experience of teller}	3
8	{Maximum age of supervisor, average work experience of assistant to branch manager maximum work experience of supervisor}	3
9	{Number of single, maximum age of teller, work experience of branch manager}	3
10	{Work experience of branch manager, minimum age of supervisor, max age of assistant to branch manager}	3
11	{Average work experience of teller, number of bachelor of science, number of single, number of male}	4
12	{Number of teller, number of bachelor of science, number of single, number of male}	4

Table 6
Mean absolute percentage error as DEA inputs.

	AERR	*VarERR*	*MaxERR*	*MinERR*
Reduct 1	0.2529	0.0339	0.6974	0.1042
Reduct 2	0.1933	0.0086	0.4147	0.1121
Reduct 3	0.2014	0.0070	0.3993	0.1045
Reduct 4	0.1841	0.0105	0.4645	0.0938
Reduct 5	0.2084	0.0109	0.4872	0.1016
Reduct 6	0.2634	0.0241	0.6350	0.0995
Reduct 7	0.2319	0.0206	0.6834	0.1213
Reduct 8	0.1923	0.0084	0.4693	0.1132
Reduct 9	0.2009	0.0039	0.3099	0.1291
Reduct 10	0.1820	0.0077	0.4095	0.1142
Reduct 11	0.2865	0.0122	0.5380	0.1648
Reduct 12	0.2013	0.0061	0.3493	0.1179

Table 7
Mean absolute percentage error.

	Full rank efficiency	Rank
Reduct 1	1.058994	1
Reduct 2	1.037797	2
Reduct 3	1.014341	3
Reduct 4	1.01156	4
Reduct 5	1.000493	5
Reduct 6	0.990344	6
Reduct 7	0.987629	7
Reduct 8	0.97908	8
Reduct 9	0.977991	9
Reduct 10	0.959753	10
Reduct 11	0.898111	11
Reduct 12	0.843254	12

Calculated full rank efficiency scores along with reduct's ranks are shown in Table 7, so 9th reduct identified as the best one, containing attributes of number of singles, maximum age of tellers and Work experience of branch manager, have the greatest impact

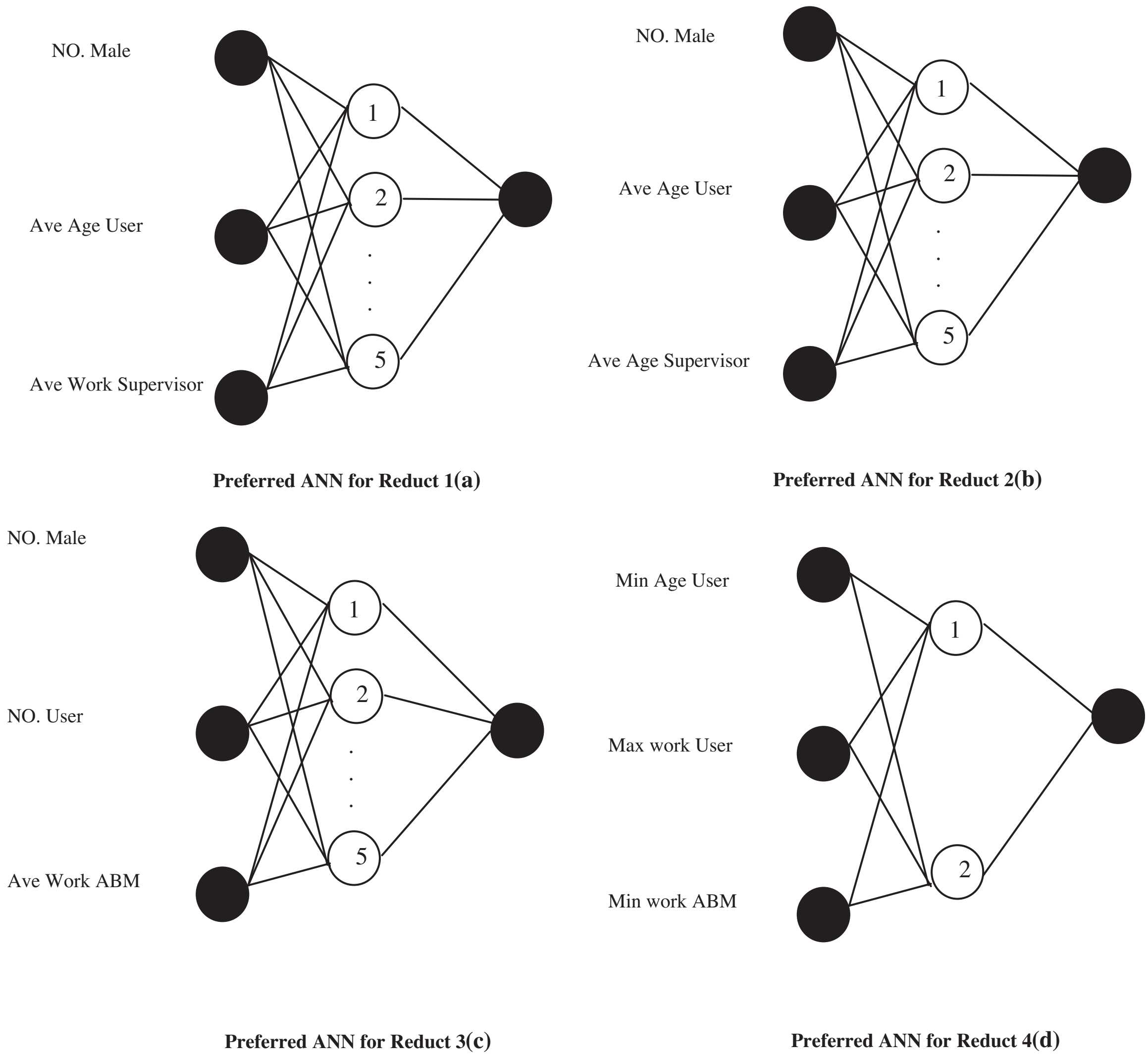

Fig. 3. Architect of Reduct's preferred ANN.

on efficiency of bank branches. By constructing neural network on 9th reduct with two neurons in hidden layer (Fig. 3(i)), we will obtain effective DSSs, which can be utilized by senior manager of the bank for sensitive analysis or efficiency prediction of inefficient or new bank branches according to their personnel specifications values.

3.6. Development of DSS

DSS is now ready to be executed. The DSS is run and 9th reduct was selected as preferred reduct. Also, ANN is ready to be used in DSS. However, K-Means algorithm must also be executed.

4. Execution of K-Means algorithm

K-Means is executed to identify homogenous groups for the branches. Efficiency value is used as K-Means algorithm input. At first, it is needed to determine correct number of clusters. The *mean silhouette* parameter is suitable criterion for this. The value of *mean silhouette* is reported in Table 8 for $k = 2$ to $k = 6$. As Table 10 shows, five clusters is the preferred number of cluster as it provide the best *mean silhouette* value. Table 9 shows the center of five clusters.

4.1. Execution of the selected ANN

As we saw in the previous section, 9th reduct was selected as preferred reduct. Number of single, maximum age of teller and work experience of branch manager are attributes of 9th reduct. This reduction in features number decrease the time of decision making and consequently reduces the cost of efficient expert analysis. Seven arbitrary values for present inputs are used as ANN input. The selected ANN is used as the engine of DSS. Table 10 shows the outputs of ANN. Moreover, Table 10 shows *outvalue*.

4.2. Position of out value

Table 11 shows that three and four rows are new situation. Manager can analysis the new situations. The value of input for four and three rows leads to new groups. Of course, analyses of occurred situation require some tools. Future studies can proceed to identify this problem.

5. Conclusion

This paper provided an intelligent DSS to help banks formulate an effective decision-making procedure for improvement and optimization of complex personnel efficiency. The DSS approach is proposed to determine important attributes. The proposed DSS approach is composed of an integrated mechanism and eight-stage analysis. Truly, the stated DSS approach is proposed to demonstrate critical attributes impacting efficiencies of various bank branched. In fact this study is an extension of a previous study by Azadeh, Javanmardi, and Saberi (2010), Azadeh et al. (2011). Furthermore the previous study presented an integrated data envelopment analysis-artificial neural network-rough set algorithm for assessment of personnel efficiency. However, this study presents a DSS approach based on previous approaches and K-Means algorithm for optimization of personnel attributes. Moreover, the intelligent DSS approach utilizes data envelopment analysis (DEA), and data mining tools including rough set theory (RST), artificial neural network (ANN), cross validation test technique (CVTT), and K-Means algorithm for forecasting and optimization of personnel efficiency. It is introduced to asses the impact of personnel attributes on efficiency. DEA is used for DMUs efficiency evaluation. RST is used for data preprocessing. ANN and CVTT are used for precision testing and forecasting and finally DEA is utilized again for identification of attributes importance. DSS is initiated after selecting the best reduct. At first, K-Means algorithm is used for grouping efficiency values. Then, the preferred ANN is executed with arbitrary values of inputs. With the proposed rule base, the position of this output is determined in available groups and a new group is constructed.

The purpose of DSS is to alert management to the important attributes which should be considered if an effective decision is to be made to enhance personnel efficiency. This is because there is a great desire to identify the critical attributes for sensitivity analysis of inefficient bank branches regarding efficiency attributes. The outcome helps managers to construct helpful system to forecast and optimize branches' efficiencies by selected attributes and yielded grouping results. In comparison with ordinary methods the new system decreases the time of testing and consequently reduces the cost of efficiency evaluation. Fig. 4 shows the processes of the intelligent DSS.

The DSS was developed for an actual banking system and its superiorities and advantages are discussed. Moreover, the DSS was successfully developed for 102 branches of a large private bank. The intelligent approach evaluates and optimizes personnel attributes impact on bank branches efficiency.

Appendix I. Artificial neural networks

An ANN is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. ANNs have been applied when there is no theoretical evidence about the functional forms. Therefore, ANNs

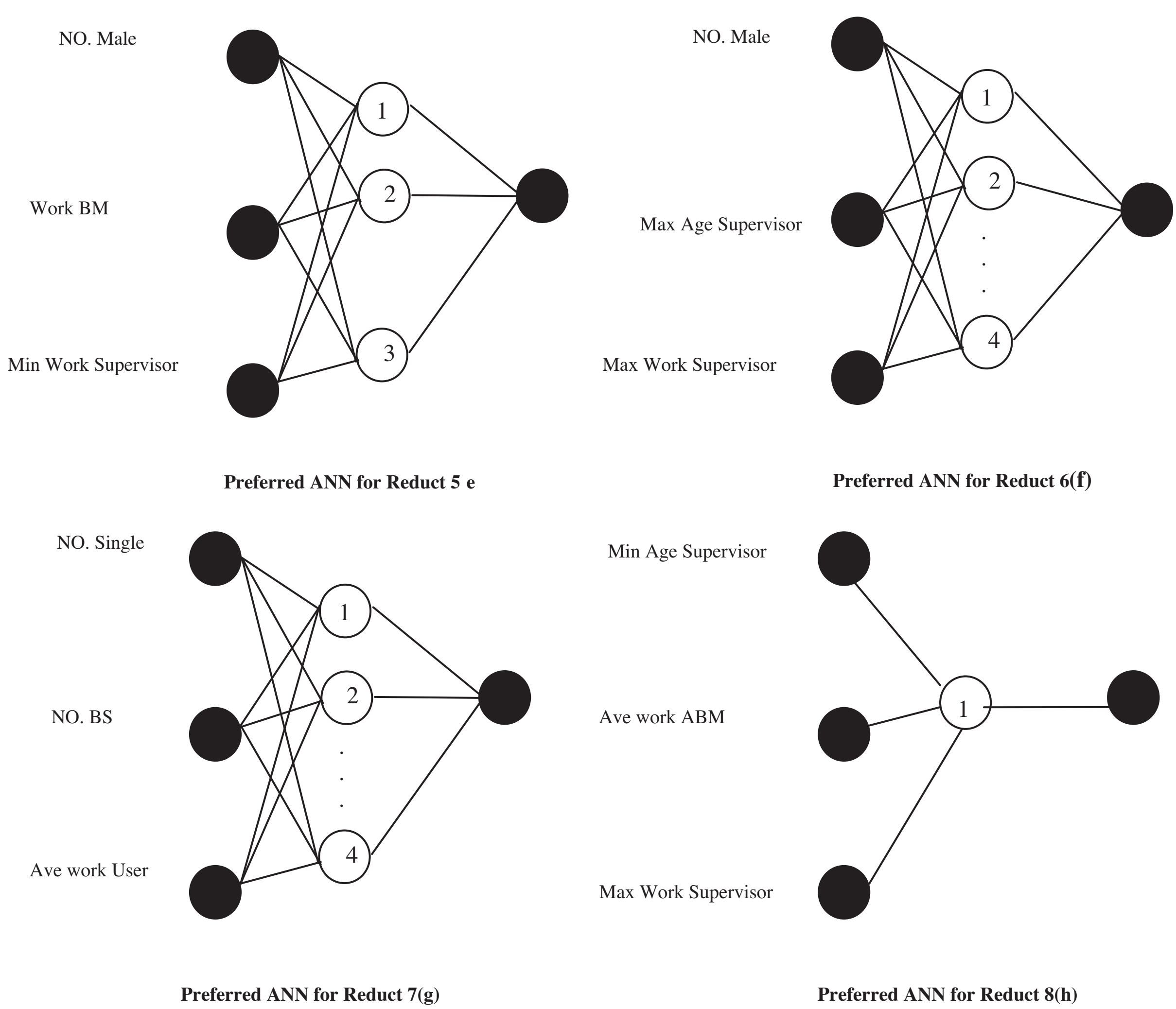

Fig. 3 (*continued*)

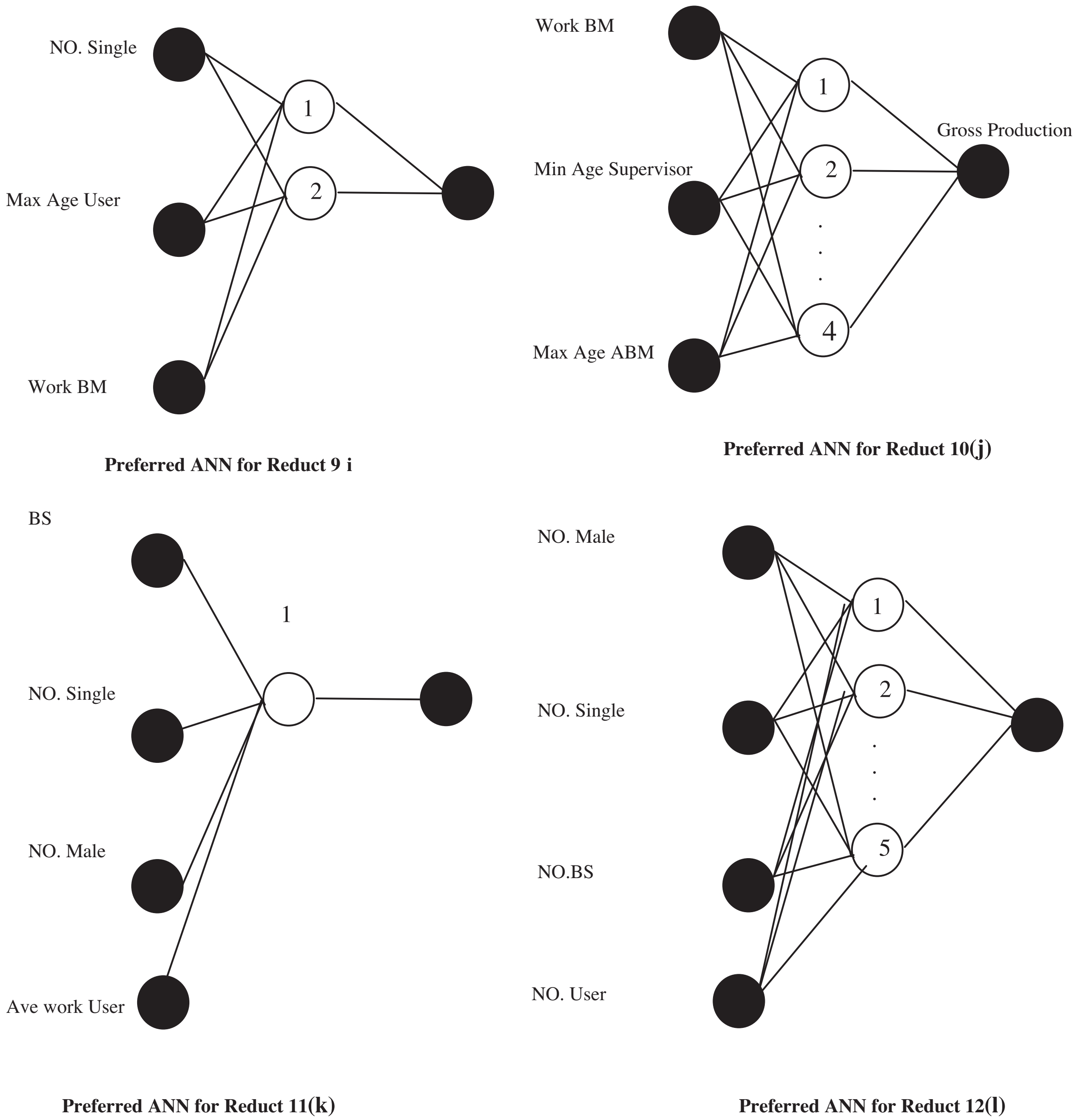

Fig. 3 (*continued*)

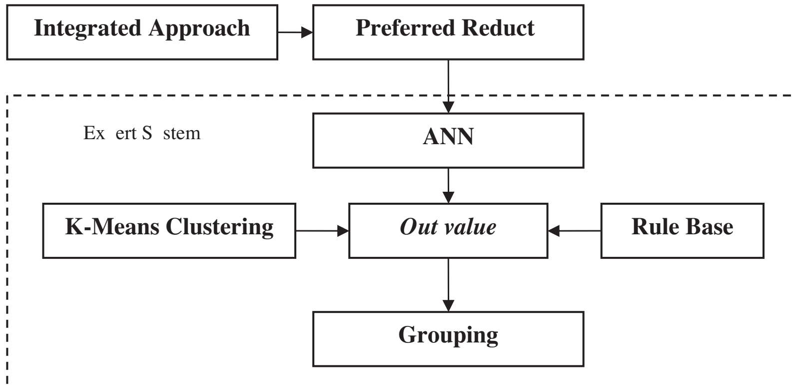

Fig. 4. Developing stages of DSS.

Table 8
The *mean silhouette* value for different *k* value.

	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$
Mean silhouette value	0.77	0.74	0.8	0.82	0.79

Table 9
Center of five clusters.

$Center_1$	$Center_2$	$Center_3$	$Center_4$	$Center_5$
0.9983	0.9427	0.8985	0.8264	0.7129

Table 10
The *outvalue* (reduct 9).

	Single	Max age user	Work boss	*Outvalue*
1	0.2	0.4	0.7	0.912139
2	0.5	0.7	0.4	0.933977
3	0.4	0.5	0.6	0.92593
4	0.7	0.6	0.7	0.965406
5	0.2	0.3	0.2	0.845798
6	0.2	0.8	0.1	0.894381
7	0.5	0.2	0.4	0.876782

Table 11
The *out value* (reduct 9).

	Outvalue	$Degmemoutvalue_i$	$Degmemoutvalue_{i-1}$	Final group
1	0.912139	0.691419	0.308581	Group 3
2	0.933977	0.197353	0.802647	Group 2
3	0.92593	0.379406	0.620594	New group
4	0.965406	0.591622	0.408378	New group
5	0.845798	0.730963	0.269037	Group 4
6	0.894381	0.057125	0.942875	Group 3
7	0.876782	0.301215	0.698785	Group 3

are data-based, rather than model-based. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems.

Among the different networks, the feed forward neural networks or multi layer perceptron (MLP) are the most commonly used in engineering science. MLP networks are normally arranged in three layers of neurons, the input layer and output layer represent the input and output variables of the model and between them lie one or more hidden layers which hold the networks ability to learn non-linear relationships.

In these networks, the output is function of the linear combination of hidden units' activations; each one is a non-linear function of the weighted sum of inputs:

$$y = f(x,\theta) + \varepsilon \tag{1}$$

where x is the vector of explanatory variables, ε is the random error component. $f(x,\theta) = \hat{y}$ is the unknown function for estimation and prediction from the available data. Consider a MLP with three layers and one output. The network consists of the following form:

$$\hat{y} = F\left(\vartheta_0 + \sum_{j=1}^{m} H\left(\lambda_j + \sum_{i=1}^{n} x_i\theta_{ij}\right)\vartheta_j\right) \tag{2}$$

where
$\hat{y}$: network output,
F: output unit activation function,
H: hidden unit activation function,
n: number of input units,
m: number of hidden units,
x_j: input vector for unit j (x_{ji} = ith input to the jth unit),
θ_{ij}: weight from input layer i to hidden unit j,
υ_o: output bias,
λ_j: hidden units biases ($j = 1,\ldots,m$),
υ_j: weights from hidden unit j to output ($j = 1,\ldots,m$).

From Eq. (2), it can be observed that MLPs are mathematical models often equivalent to conventional models in econometrics (linear regression, auto regressive moving average (ARMA) models for time series analysis), but with specific estimation methods (Cheng & Titterington, 1994). In Fig. A1 a MLP with three layers and one output is shown.

The activation function for output layer is generally linear. The non-linear feature is introduced at the hidden transfer function. From the previous universal approximation studies, these transfer functions must have mild regularity conditions: continuous, bounded, differentiable and monotonic increasing. The most popular transfer function is sigmoid or logistic, nearly linear in the central part. Architecture selection is one major issue with implications on the empirical results and consists of:

1. Input and output variables number.
2. Hidden layers' number.
3. Hidden and output activation function.
4. Learning algorithm

All of the above issues are open questions today and there are several answers to each one. The hidden units' number is determined by a trial–error process considering $m = 1,2,3,4,\ldots$. Finally, it is common to eliminate 'irrelevant' inputs or hidden units (White, 1989). Too few neurons in hidden layers (hidden units) can lead to under fitting. However, too many neurons can cause over fitting. The actual number of neurons required in the hidden layer must be found by trial and error. Moreover, the inputs are used by the network must be effective on the value of output(s), in fact the input and output variables should be identified carefully, because enable the network to learn relationships quicker and use fewer hidden units.

Another critical issue in ANNs is the neural learning or model estimation based upon searching the weights that minimize some cost function such as square error:

$$\mathrm{Min}\left[E(y - f(x,\theta))^2\right]\theta \in \Theta. \tag{3}$$

The most popular learning algorithm is the back proportion (BP). BP learning is a kind of supervised learning introduced by Werbos (1974) and later developed by Rumelhart and McClelland (1986). Desirable output for input set is made by this algorithm. Error in each neuron is the difference between ANN output and real output. The interconnections weight and threshold value in each neuron is adjusted to minimize the error. Let:
E: denote error function,
o_j: output of unit j,
θ_{ij}: weight from input layer i to hidden unit j,
z_j: θ_j, x_j the weighted sum of inputs for unit j,

In this algorithm for reducing error, the weights vector (θ) is adjusted. For this $\frac{\partial E}{\partial \theta}$ is calculated. By applying chain rule, we have:

$$\frac{\partial E}{\partial \theta_{ji}} = \frac{\partial E}{\partial z_j}\cdot\frac{\partial z_j}{\partial \theta_{ji}}$$

It is proofed that:

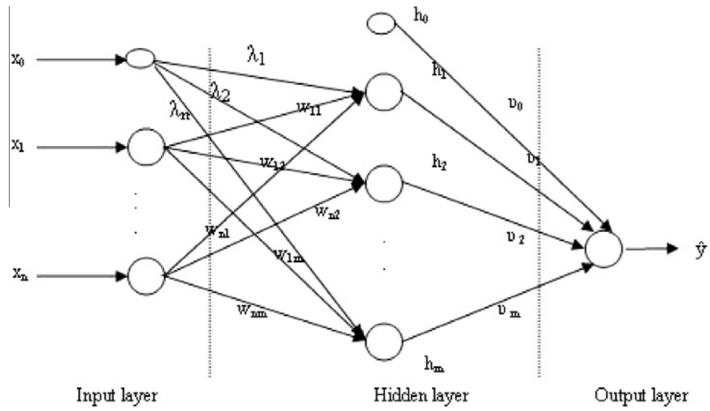

Fig. A1. Single-output, three layer feed-forward neural network MLP $(n, m, 1)$.

$$-\frac{\partial E}{\partial \theta_{ji}} = \delta_i \cdot o_j \quad \text{when} \quad \delta_i = \frac{\partial E}{\partial z_j}.$$

Thus:

$$\Delta\theta_{ji} = \eta\delta_i o_j,$$
$$\theta_{ji}(k+1) = \theta_{ji}(k) + \eta\delta_i o_j.$$

So this algorithm process can be express as:

$$\theta(k+1) = \theta(k) - \eta\frac{\partial E}{\partial \theta}(k) \tag{4}$$

BP is an iterative process (k indicates iteration). Parameters are revised from the error function (E) gradient by the learning rate η, constant or variable. The error propagates backwards to correct the weights until some stoppage criterion – epoch, error goals – is reached.

After neural training (training set), new observations (validation and/or test sets) are presented to the network to verify the so-called generalization capability (Schiffman, 1992). ANNs have advantages, but logically they also have several drawbacks. Therefore, ANNs can learn from experience and can generalize, estimate, predict, with few assumptions about data and relationships between variables. These attributes have made the ANN approach fairly efficient for problem solving. Hence, ANNs have an important role when these relationships are unknown (non-parametric method) or non-linear (non-linear method), provided there are enough observations with flexible form and universal approximation property. Algorithm convergence and trial and error process are also some relevant drawbacks.

Acknowledgement

This study was supported by a grant from University of Tehran (Grant No. 8106013/1/04). The authors are grateful for the support provided by the College of Engineering, University of Tehran, Iran. The authors are grateful for the valuable comments and suggestion from the respected reviewers. Their valuable comments and suggestions have enhanced the strength and significance of our paper.

References

Al-Omari, F., & Al-Jarrah, O. (2004). Handwritten Indian numerals recognition system using probabilistic neural networks. *Advanced Engineering Informatics, 18*(1), 9–16.

Athanassopoulos, A. D. (1997). Service quality and operating efficiency synergies for management control in the provision of financial services: Evidence from Greek bank branches. *European Journal of Operational Research, 98*(2), 300–313.

Azadeh, A., Amalnick, M. S., Ghaderi, S. F., & Asadzadeh, S. M. (2007). An integrated DEA PCA numerical taxonomy approach for energy efficiency assessment and consumption optimization in energy intensive manufacturing sectors. *Energy Policy, 35*(7), 3792–3806.

Azadeh, A., Ghaderi, S. F., Anvari, M., & Saberi, M. (2006). Measuring performance electric power generations using artificial neural networks and fuzzy clustering. In G. A. Capolino, L. G. Franquelo, (Eds.), *Proceedings of the 32nd annual conference of the IEEE industrial electronics society*. IECON, Paris, France.

Azadeh, A., Ghaderi, S. F., Anvari, M., & Saberi, M. (2007). Performance assessment of electric power generations using an adaptive neural network algorithm. *Energy Policy, 35*, 3155–3166.

Azadeh, A., Ghaderi, S. F., Anvari, M., Saberi, M., & Izadbakhsh, H. (2006). An integrated artificial neural network and fuzzy clustering algorithm for performance assessment of decision making units. *Applied Mathematics and Computation*.

Azadeh, A., Ghaderi, S. F., & Izadbakhsh, H. (2007). Integration of DEA and AHP with computer simulation for railway system improvement and optimization. *Applied Mathematics and Computation*.

Azadeh, A., Ghaderi, S. F., & Sohrabkhani, S. (2007). Forecasting electrical consumption by integration of neural network, time series and ANOVA. *Applied Mathematics and Computation, 186*, 1753–1761.

Azadeh, A., Ghaderi, S. F., Tarverdian, S., & Saberi, M. (2007). Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. *Applied Mathematics and Computation, 186*, 1731–1741.

Azadeh, A., Ghaderi, S. F., Tarverdian, S., & Saberi, M. (2006). Integration of artificial neural networks and GA to predict electrical energy consumption. In G. A. Capolino & L. G. Franquelo (Eds.), *Proceedings of the 32nd annual conference of the IEEE industrial electronics society – IECON'06*. Paris, France: Conservatoire National des Arts & Metiers.

Azadeh, A., Javanmardi, L., & Saberi, M. (2010). The impact of decision making units features on efficiency by integration of DEA, ANN, fuzzy C-means and ANOVA. *International Journal of Operation Research, 7*(3).

Azadeh, A., Saberi, M., Tavakkoli-Moghaddam, R., & Javanmardi, L. (2011). An integrated data envelopment analysis-artificial neural network-rough set algorithm for assessment of personnel efficiency. *Expert Systems with Applications, 38*(3), 1364–1373.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in DEA. *Management Science, 32*, 1078–1092.

Becerra-Fernandez, I., Zanakis, S. H., & Walczak, S. (2002). Knowledge discovery techniques for predicting country investment risk. *Computers & Industrial Engineering, 43*(4), 787–800.

Berger, A. N., & Humphrey, D. B. (1991). The dominance of inefficiencies over scale and product mix economies in banking. *Journal of Monetary Economics, 28*, 117–148.

Beynon, M. J., & Peel, M. J. (2001). Variable precision rough set theory and data discretisation: An application to corporate failure prediction. *Omega, 29*(6), 561–576.

Błaszczyński, J., Greco, S., & Słowiński, R. (2007). Multi-criteria classification – A new scheme for application of dominance-based decision rules. *European Journal of Operational Research, 181*(3), 1030–1044.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of DMUs. *European Journal of Operational Research, 2*(6), 429–444.

Charnes, A., Cooper, W. W., Lewin, A. Y., & Seiford, L. M. (Eds.). (1994). *Data envelopment analysis: Theory, methodology, and applications*. Boston: Kluwer.

Cheng, B., & Titterington, D. M. (1994). Neural networks: A review from a statistical perspective. *Statistical Science, 9*(1), 2–54.

Chou, H. C., Cheng, C. H., & Chang, J. R. (2007). Extracting drug utilization knowledge using self-organizing map and rough set theory. *Expert Systems with Applications, 33*(2), 499–508.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems, 2*, 303–314.

Fan, T. F., Liu, D. R., & Tzeng, G. H. (2007). Rough set-based logics for multi-criteria decision analysis. *European Journal of Operational Research, 182*(1), 340–355.

Fonseca, D. J., & Navaresse, D. (2002). Artificial neural networks for job shop simulation. *Advanced Engineering Informatics, 16*(4), 241–246.

Hassanien, A. (2007). Fuzzy rough sets hybrid scheme for breast cancer detection. *Image and Vision Computing, 25*(2), 172–183.

Inuiguchi, M., & Miyajima, T. (2007). Rough set based rule induction from two decision tables. *European Journal of Operational Research, 181*(3), 1540–1553.
Kowalczyk, W., & Slisser, F. (1997). Modeling customer retention with rough data models. In *Proceedings of the first European symposium, PKDD'97* (pp. 4–13), Trondheim, Norway.
Kusiak, A., & Tseng, B. T. L. (2000). Data mining in engineering design: A case study. In *Proceedings of the IEEE international conference on robotics and automation* (Vol. 1, pp. 206–211).
Kusiak, A., Kern, J. A., Kernstine, K. H., & Tseng, B. T. L. (2000). Autonomous decision-making: A data mining approach. *IEEE Transactions on Information Technology in Biomedicine, 4*(4), 274–284.
Lili, W., & Zhi, P. (2001). Fuzziness in rough sets based on general relations. *Journal of Ningxia University (Natural Science), 22*(4), 352–355.
Liu, M., Chen, D., Wu, C., & Li, H. (2006). Reduction method based on a new fuzzy rough set in fuzzy information system and its applications to scheduling problems. *Computers & Mathematics with Applications, 51*(9–10), 1571–1584.
Moller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks, 6*, 525–533.
Patuwo, E., Hu, M. Y., & Hung, M. S. (1993). Two-group classification using neural networks. *Decision Sciences, 24*(4), 825–845.
Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences, 11*, 314–356.
Pawlak, Z. (1991). *Rough sets: Theoretical aspects of reasoning about data*. Boston, MA: Kluwer.
Ravi Kumar, P., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research, 180*(1), 1–28.
Ruhe, G. (1996). Qualitative analysis of software engineering data using rough sets. In *Proceedings of the fourth international workshop on rough sets, fuzzy sets, and machine discovery* (pp. 292–299), Tokyo, Japan.
Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Foundations, Cambridge, MA: MIT Press.
Sanchis, A., Segovia, M. J., Gil, J. A., Heras, A., & Vilar, J. L. (2007). Rough sets and the role of the monetary policy in financial stability (macroeconomic problem) and the prediction of insolvency in insurance sector (microeconomic problem). *European Journal of Operational Research, 181*(3), 1554–1573.
Schiffmann, W., Joost, M., & Werner, R. (1992). Optimization of the backpropagation algorithm for training multilayer perceptrons. Technical Report 16/92. Institute of Physics. University of Koblenz.
Serrano Cinca, C., Mar Molinero, C., & Chaparro García, F. (2002). Behind DEA efficiency in financial institutions. *Accounting and Finance, 40*, 30–45.
Shyng, J. Y., Wang, F. K., Tzeng, G. H., & Wu, K. S. (2007). Rough set theory in analyzing the attributes of combination values for the insurance market. *Expert Systems with Applications, 32*(1), 56–64.
Sikder, I. U., & Gangopadhyay, A. (2007). Managing uncertainty in location services using rough set and evidence theory. *Expert Systems with Applications, 32*(2), 386–396.
Soteriou, A., & Zenios, S. A. (1999). Operations, quality and profitability in the provision of banking services. *Management Science, 45*(9), 1221–1238.
Stefanowski, J., & Slowinski, K. (1997). Rough sets as a tool for studying attribute dependencies in the urinary stones treatment data set. In *Rough sets and data mining: Analysis and imprecise data* (Vol. 10, pp. 177–196).
Ton Su, C. T., & Hsu, J. H. (2006). Precision parameter in the variable precision rough sets model: An application. *Omega, 34*(2), 149–157.
Triantaphyllou, E., Liao, T. W., & Iyengar, S. S. (2002). A focused issue on data mining and knowledge discovery in industrial engineering. *Computers & Industrial Engineering, 43*(4), 657–659.
Tseng, T. L. B., & Huang, C. H. (2007). Rough set-based approach to feature selection in customer relationship management. *Omega, 35*(4), 365–383.
Tsumoto, S. (1997). Extraction of experts' decision process from clinical databases using rough set model. In *Proceedings of the first European symposium, PKDD'97* (pp. 58–67), Trondheim, Norway.
Wang, X. Y., Yang, J., Jensen, R., & Liu, X.-J. (2006). Rough set feature selection and rule induction for prediction of malignancy degree in brain glioma. *Computer Methods and Programs in Biomedicine, 83*(2), 147–156.
Werbos, P. I. (1974). Beyond regression: New tools for prediction and analysis in the behavior sciences. Ph.D. Thesis, Harvard University, Cambridge, MA.
Xia, W., & Wu, Z. (2007). Supplier selection with multiple criteria in volume discount environments. *Omega, 35*(5), 494–504.
Zhai, L. Y., Khoo, L. P., & Fok, S. C. (2002). Feature extraction using rough set theory and genetic algorithms – An application for the simplification of product quality evaluation. *Computers & Industrial Engineering, 43*(4), 661–676.
Ziarko, W. (1993). Variable precision rough set model. *Journal of Computer and System Sciences, 46*, 39–59.