Research Article
Case-Based Reasoning Method in Cost Estimation of Drilling Wells

Hossein Shams Mianaei and Seyed Hossein Iranmanesh
Department of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran

Abstract: Aim of study is the cost estimation of drilling wells using Case-Based Reasoning (CBR) method which is created based on the viewpoint of using presented solutions for previous solved problems in order to solve new similar problems. In companies or organizations which cost estimation, scheduling, design, planning and project activities sequences and etc., are considered as an iterative activity, project managers naturally use previous experiences in order to estimate the instances mentioned above in new similar case. Thus, offering a comprehensive approach which can support this natural behavior, will significantly contribute to the appropriate decision making. CBR method could be an appropriate tool for decision making in this area. The major findings of this study is that, in the proposed CBR model despite limited data, the error of method according to the performance indicators was very low. Therefore, obtained estimation accuracy of the proposed CBR model is high and the model is useful. On the other hand, given that the available estimation methods spend much time to estimate cost, we could save time using the CBR method and hence the speed of drilling which is very important is increased.

Keywords: Case Based Reasoning (CBR), cost estimation

INTRODUCTION

Engineering design and management of cost, not only play an essential role in project feasibility studies, but also are fundamental for final success of a project. Costs appropriate estimation is key factor in projects. Scientific methods should be employed during project planning and design to increase costs estimation accuracy. Improved cost estimation techniques, which are available for project managers, will facilitate the more effective control of time and costs in projects. The cost estimation model of Case-Based Reasoning (CBR), which in preliminary stages estimates the costs with minimum project information, is useful in preliminary design stage of the drilling wells project which is our subject.

Case-based reasoning method is formed based on using response of problems solutions to solve new similar problems. CBR is known as a method that is inspired from how humans behave in dealing with new problems. As follows which uses earned experiences in solving past problems as a guide for solving new problems.

Solving problem by using CBR method is done in one cycle and it includes four processes as follows:

- Retrieve the similar case with new problem
- Reuse the solution of retrieved similar problem to provide the proposed solution for new problem
- Revise the proposed solution if there is conflict in conditions of the new problem and retrieved problem
- Retain new case (new problem and its solution) for use in future

Figure 1 shows the CBR-Cycle.

Each case is composed of two parts: The first part of each case addresses into expression specifications of problem in related to it. The second part includes the presented problem solution. Specification of a case is determined with its constituent features and the assigned values into these features, shows the status of that case.
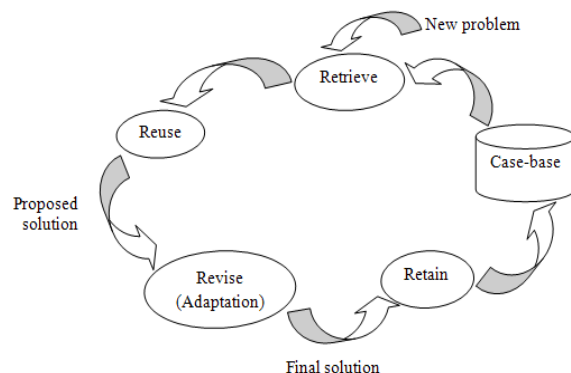

Fig. 1: The CBR-cycle

Corresponding Author: Hossein Shams Mianaei, Department of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran, Tel.: +98 2188021067, Fax.: +98 2188013102

When a new problem is presented, its conditions is compared with status of previous solved problems and using mechanisms of similarity, the most similar previous cases are retrieved. Then, the retrieved cases are applied to provide a solution for the new problem and subsequently proposed solution is provided. If required, the proposed solution will be revised according to the position of the new problem and finally, new case (i.e., the considered problem and its solution) is retained in case base for future usage. In this study, we seek to estimate the cost of drilling wells using CBR method.

LITERATURE REVIEW

In this section, we briefly review the prior fundametal studies on cost estimation and CBR in cost estimation. At the end of this section, we address into the justification of research in this field.

Cost estimation: Some approaches for cost estimation have been developed since 1970s based on statistics (Wilson, 1982; Singh, 1990) and analysis of linear regression (Kouskoulas and Koehn, 1974; Bowen and Edwards, 1985; Khosrowshahi and Kaka, 1995; Skitmore and Thomas, 2003).

An alternative branch of artificial intelligence, Neural Networks (NN), has appeared as an appropriate alternative for cost estimation of construction during 1990s (Garza and Rouhana, 1995; Creese and Li, 1995; Adeli and Wu, 1998; Yeh, 1998; Bode, 2000). However, the NN approach is a black box technique and process of its knowledge acquisition is very time-consuming. These are the almost highlighted disadvantages in previous researches (Li, 1995; Yeh, 1998; Hegazy et al., 1994; Boussabaine, 1996; Hegazy and Ayed, 1998).

Jurgensen and Wallace (2000) discuss that how costs of project (if they are based on models) can be estimated better by taking into account how projects are actually implemented, in particular, how managerial flexibility adds value to a project. The purpose is not to advocate the better ways of implementing a project, but purpose is to obtain better estimations for costs of project.

Contractors traditionally estimate the cost probability based on subjective judgment, such as 5-10% from estimated cost by considering past similar project. However, such method does not have a stable and accurate basis and is difficult to justify. Therefore, objective methods have been presented for estimating occurrence probability of the cost of project. However, the most methods still have relied on formal modeling

techniques, which are not easy to be applied in construction industry. In this field, Idrus et al. (2011) proposed a method for estimating occurrence probability of cost using a flexible and logical approach that can matches the subjective judgment of contractors based on analysis of risk and concept of fuzzy expert system. The proposed method includes the development of occurrence probability model of cost for construction and infrastructures of project in Malaysia.

CBR in cost estimation: In late 1980s, a new approach, called expert systems, was introduced to estimate cost. However, the use of expert systems did not reach to its maximum potential. Therefore, CBR systems were proposed as an alternative for expert system in cost estimation. For example, Perera and Watson (1998) proposed a prototype system, NIRMANI, based on CBR in order to support design and cost estimation.

During design stage, 80% of the cost of a product is planned. Therefore, it is essential that the designer gives some efficient cost estimation methods. In this field, Duverlie and Castelain (1999) showed the application of parametric methods and CBR (induction) method for cost estimation in design phase of mechanical production and more particularly, the production of pistons for railway diesel engines.

Kim et al. (2004) examined the performance of three cost estimation models. The examinations are based on Multivariate Regression Analysis (MRA), Neural Networks (NNs) and Case-Based Reasoning (CBR) from the data of 530 historical costs. Estimation model of CBR performed better than estimation model of NN with regard to long-term use, available information from result and tradeoffs of time vs. accuracy.

Chou (2008) outlined the modeling of CBR estimation which it compares and retrieves the most similar case across the case library. Four CBR approaches were presented and assessed based on their Mean Absolute Prediction Error (MAPE) rates. The result of model demonstrates the ability of costs estimation of pavement maintenance project with satisfactory accuracy at the preliminary stages.

For any successful construction project, it is very important to accurately estimate the construction cost during preliminary stage of project. So Raphael et al. (2007) investigated three CBR systems which have been developed during 7 years with collaboration of two industrial partners. In their research, CBR was used for calculating cost of construction projects.

An et al. (2007) proposed a CBR model that the experience is taken into account by Analytical

Hierarchy Process (AHP) in all processes of construction cost estimation.

In the contrary to the available CBR-based construction cost prediction models, Ji *et al.* (2010) developed a CBR revision model that reflects the "revise" stage of CBR cycle (retrieve, reuse, revise and retain) based on nine housing projects executed by "A" Housing Corporation.

Kim and Kim (2010) proposed a preliminary cost estimation model using CBR and Genetic Algorithm (GA). In similarity measurement and retrieving similar cases from a case-base for minimum prediction error, it is a key process in determining factors with maximum weight among features of cases in case-base.

As far as we know, up to 2010, there was no model which could be applied directly in forecasting manufacturing costs. Research of Chang *et al.* (2010) made the first attempt for development of a hybrid system by integrated CBR and Artificial Neural Networks (ANN) as a forecasting model of Product Unit Cost (PUC) in Mobile Phone Company. The proposed model in their research was compared with other five models. Their findings indicated that MAPE value of the proposed model was smallest.

Since success of a CBR system mainly depends on effective maintenance of its case-base, Kim and Han (2001) proposed a Genetic Algorithms (GAs) approach for maintenance of CBR systems. Their approach automatically determines the representation of cases and indexes of relevant features to understand the rapidly changing environment around the system. In their study, the proposed model is applied to analyze the stock market. Experimental results show that the proposed model outperforms the conventional CBR and Case-Base Maintenance (CBM) systems.

Justification: Most of the methods used for project cost estimation have static mode. That is, they (for each special problem) must possess the data about it and we can't use the planning for a period in other periods. But, The CBR method is an intelligent learner method and it has dynamic mode that uses data of previous problems in new problem and plans intelligently. Indeed, we can use the CBR within used methods for project cost estimation which they have the static mode. Estimation models generally are used to help decision making in cost estimation. In decision making process, the CBR model is quite successful in explaining procedures for obtaining cost of a new project because it can retrieve the costs and descriptions of a similar project from case-base. Understanding how CBR works, is very simple and this is one of the reasons that why it has been easily adopted. The CBR also can complete the other calculation techniques, not just data-bases. We

should revise the method which we design the computers for solving problems, because if we again give the same problem into system, it will not be different with last time of view time of solving. But CBR solve this problem. Before starting to solve a problem, we should see if we've solved the problem already. If not just, we'll have to start from scratch.

To plan the new case, almost all methods applied for costs estimation of project use the data of quite similar features of previous cases with new case. But in CBR method, it doesn't need to define the quite similar features. For example, suppose that a factory produces different parts which may be different at manufacturing features in some cases. In such cases, the CBR is the good method to estimate manufacturing costs, because, the similar cases are retrieved through similarity measure of new case with previous cases. If a new part (not necessarily the similar of previous parts) is going to be made, the most similar case can be easily achieved and planning could be done accordingly.

THE PROPOSED CBR MODEL

In this study, a CBR model is developed for cost estimation of drilling wells based on features of drilling well (case). Figure 2 shows the process of proposed CBR model.

Initially, features and performance indices of project and their scale are identified in three steps. In the next stage, the CBR-based model is developed in ten steps for cost estimation of drilling wells. A case-base is established by considering CBR-based model. Methods of calculating feature distinction and similarity, feature weight, case distinction and similarity and adaptation can be different according to viewpoint of researcher's who is developing the CBR model.

Step 1: Data preparation:
Step 1.1: Data collection: Data collection and preparation is the most important and time consuming stage in projects of data mining. Since the data are the input of projects, the more accurate the input, the more accurate the work output. Poor quality data leads to poor quality data mining and accordingly poor quality decisions.

There are final reports at the end of project for each project which is defined. In this study which we are going to use past data, paying attention to the issue discussed above is important and the art of a data analyst is that he or she be able to extract the best and appropriate characteristics or features according to the available data. Selecting correct and accurate data for each feature is the most important stage in data

```
┌─────────────────┐
│   Collecting    │
└─────────────────┘
         ↓
┌─────────────────┐
│ Defining features│
│    of case      │
└─────────────────┘
         ↓
      ⌠ Case- ⌡
         ↓
┌─────────────────┐
│ Calculating features│
│ distinction and similarity│
└─────────────────┘
         ↓
┌─────────────────┐
│  Calculating    │
│   features      │
└─────────────────┘
         ↓
┌─────────────────┐
│ Calculating cases│
│ distinction and similarity│
└─────────────────┘
         ↓
┌─────────────────┐
│ Cases retrieval │
└─────────────────┘
         ↓
┌─────────────────┐
│ Selecting the most similar cases│
│  (Proposed solutions)│
└─────────────────┘
         ↓
┌─────────────────┐
│   Adaptation    │
└─────────────────┘
         ↓
┌─────────────────┐
│ Final solution  │
└─────────────────┘
         ↓
┌─────────────────┐
│     Output      │
└─────────────────┘
```
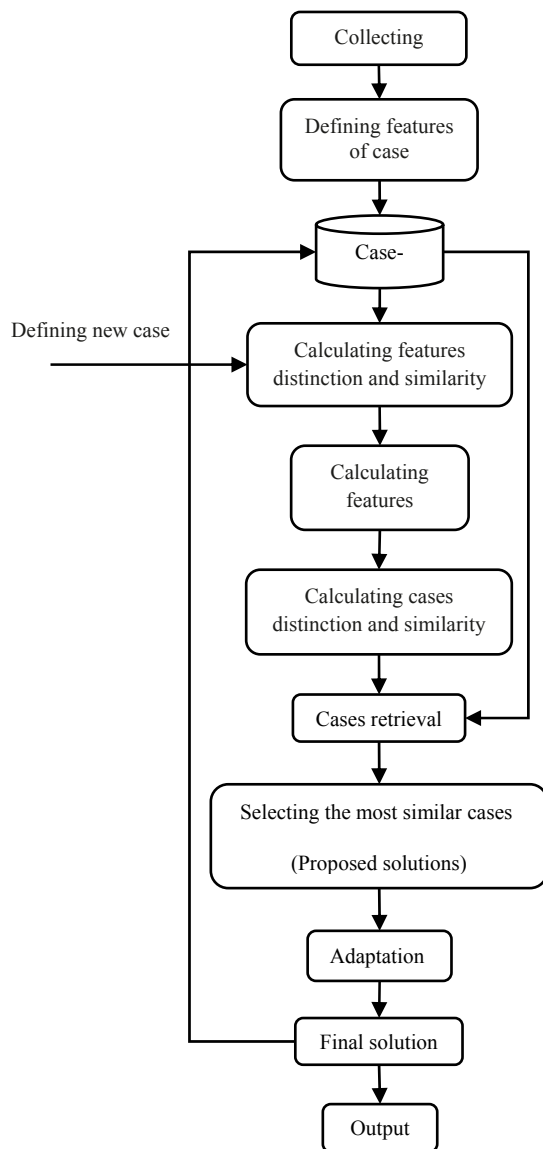
Defining new case

Fig. 2: The process of CBR proposed model

collection. On the other hand, obtaining accurate and actual results by analyst depends on this stage.

In CBR, the term of feature is used for exhibition of stored cases and term of case is used for exhibition of row's records.

Step 1.2: Filtering data: To analyze the collected data in step 1.1, it was necessary to ensure their compatibility and consistency through filtering data. Therefore, we survey the used data types.

A data set consists of data objects. Data set is usually a file in which objects (cases) are the file rows and each column corresponds to one feature.

With regard to distinct (= and ≠), sequence (<, > , ≤ and ≥), taking the sum (+ and −) and multiply (× and ÷) properties, four types of features can be defined:

Nominal: Names are different merely and only provides the information to distinguish case (=, ≠). Such as sex, student number.

Rank: It provides the enough information to sort the cases (<, >, ≤, ≥). Such as education {bad, average, good}.

Interval: Difference between values is meaningful. That is, there is the unit of measurement (+, -). Such as centigrade, date calendar.

Ratio: Differences and ratios are meaningful (×, ÷). Such as length, value of money, Kelvin.

Nominal and rank features are known as class or qualitative feature all together. These features have some limited modes. Even if these features are expressed with a number for example an integer number, they should be treated as a symbol. Two other types (i.e., interval and ratio) are known as a numerical or quantitative feature. Quantitative features are expressed with numbers and they hold the most of number's properties. These features can adopt the integer or continuous value.

The difference of interval and ratio scales is how the exposure of zero point in scale. Zero point in interval scale is defined contractually and optionally. Inversely, a ratio scale has an absolute zero point.

Discrete features have finite and or countable infinite set of values. These features can be the class type, such as the number of students and or they can be the numerical type, such as enumeration. Discrete features usually are displayed with the integer variables. Binary features are the special mode of discrete features that they only have two values, such as true/false, yes/no. Binary features are often expressed into the form of Boolean variables or 0 and 1 variable.

Symmetric binary is a feature that two obtained modes by it have the equal values from the view of similarity, such as sex feature which male and female have the same worthiness. In asymmetric binary feature, different modes 0 and 1 don't have the same worthiness values and each of them have their special importance. Such as, positive and negative answer of experiment for a patient, so that positive answer be more important (If experiment answer of two people is

positive for diagnosing a rare disease, those two people will be very similar to each other, but the negative answer is reason for the similarity of two people).

Step 1.3: Data standardization: Scale of features measurement in data may be different, as an example a feature like Operation Time contains a greater range than Percent H_2S. With regard to the difference amount aggregated in features of distance function, high-scale features eliminate the effect of low-scale features. To solve this problem, values should be standardized (normalized) before comparison. Standardization with regard to the type of feature is explained in step 2.2.

Step 2: The CBR-based model: This step shows the CBR-based model for inferring feature distinction and similarity, feature weight, case distinction and similarity, selecting proposed solutions, adaptation, model validation, testing case base and retaining CBR-based model.

Step 2.1: Organizing data: In this step, we convert the data to a form that is needed in CBR and consists of two parts: case base and new case (test case). The case base includes No. and name of case, input features and output features.

Step 2.2: Calculating distinction and similarity with regard to the type of feature: There are large numbers of distinction and similarity measurement. All of them are very different and are applied to particular demands. Most applications development based on CBR, requires the development of a particular measurement. There are much published works on this topic (Duverlie and Castelain, 1996). In CBR, the purpose of numerical distinctions and similarities is to quantify the differences and resemblances which exist between two or more cases or structures.

In this section, the distinction and similarity of cases are explained according to the type of feature or used variables. In CBR, two types of special data structure have the great importance which they are in form of matrix. These data structures are: data matrix and distinction matrix (distance matrix)

Data matrix (case-feature matrix): This type of data structure displays m cases with n features. This matrix is shown in Fig. 3.

Figure 3 Shows a matrix which consists of m different data (data-base records) and each of them have n dimension.
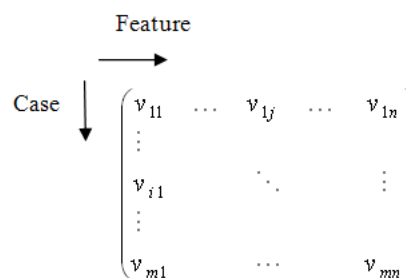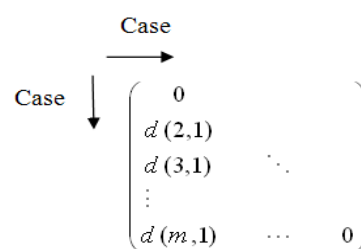

Fig. 3: Data matrix

Fig. 4: Distinction matrix

Distinctions matrix (case-case matrix): This matrix specifies the distance or distinction between two cases and is often m×m. The d(i, z) displays the distinction between cases i and z. This matrix is shown in Fig. 4.

Now, we examine the distinction for each type of scale of feature.

Interval and ratio features: For each feature j of case i^{th}, we normalize the feature the following:

$$v'_{ij} = \frac{v_{ij} - M_j}{S_j}$$

(1)

$$M_j = \frac{1}{m}(v_{1j} + v_{2j} + ... + v_{mj})$$

(2)

$$S_j = \frac{1}{m}\left(\left|v_{1j} - M_j\right| + \left|v_{2j} - M_j\right| + \cdots + \left|v_{mj} - M_j\right|\right)$$

(3)

In which v_{ij} is the feature value j in case i^{th}.

Distances are good criteria for distinction between cases with this type of features. One of famous distances is Minkowski distance:

$$d_{iz}^{Interval \& Ratio} = \left(|v'_{i1} - v'_{z1}|^q + |v'_{i2} - v'_{z2}|^q + \cdots + |v'_{ir} - v'_{zr}|^q\right)^{\frac{1}{q}}$$

(4)

In which r(r≤n) is the number of interval and ratio features in two cases i and z and $d_{is}^{Interval\&Ratio}$ is distinction between cases i and z in interval and ratio mode. For q = 1, Manhattan distance is obtained and for q = 2, Euclidean distance is obtained. For $q > 2$, this relation doesn't have the special meaning, but it can give the better answers in some conditions. For example, when we want to give more weight to far distances, we can use larger numbers and or even decimal.

And if we define W_j for each feature:

$$d_{iz}^{Interval\&Ratio-Weighted} = (W_1|v'_{i1}-v'_{z1}|^q + W_2|v'_{i2}-v'_{z2}|^q + ... + W_r|v'_{ir}-v'_{zr}|^q)^{\frac{1}{q}}$$

(5)

In features with interval scale, in addition to specifying sequencing features, also is determined the distance between them.

Nominal features: Nominal features are divided to two modes:

- **The single mode (combo box):** We can choose only one mode and value from the total number of modes of the considered features.
- **The multi mode (list box):** We are allowed to choose several modes and value from the total number of modes of the considered features.

In the Single Mode, if s is the number of modes of the nominal feature, then the position of these modes can be demonstrated with 1, 2,...., s numbers. For measuring the distinction between cases with regard to nominal features in this mode, we use the following equation:

$$d_{iz}^{Nominal(C)} = \frac{t-s}{t}$$

(6)

In which s is the number of features (in the Single Mode) that the cases I and z have the same modes from that feature and t (t≤n) is the total number of nominal features in the Single Mode.

And if we define W_j for each feature:

$$d_{iz}^{Nominal(C)-Weighted} = \frac{t-s}{t}(W_1+W_2+...+W_t)$$

(7)

And in the Multi Mode, if s_j is the number of modes of the nominal feature j^{th}, then the position of

these modes can be demonstrated with 1, 2,..., s_j numbers. To measure the distinction between cases with regard to nominal features in this mode, we use the following equation:

$$d_{iz}^{Nominal(L)} = \frac{\sum_{j=1}^{p} Max(t_j^i, t_j^z) - \sum_{j=1}^{p} s_j}{\sum_{j=1}^{p} Max(t_j^i, t_j^z)}$$

(8)

In which p(p≤n) is the number of nominal features in the Multi Mode, s_j is the number of the same modes of feature j^{th} between cases i and z, t_j^i and t_j^z, are respectively the number of obtained modes (not the number of total modes (t_j)) of nominal feature j^{th} in cases i and z are.

And if we define W_j for each feature:

$$d_{iz}^{Nominal(L)-Weighted} = \frac{\sum_{j=1}^{p} Max(t_j^i, t_j^z) - \sum_{j=1}^{p} s_j}{\sum_{j=1}^{p} Max(t_j^i, t_j^z)}(W_1+W_2+...+W_p)$$

(9)

Rank features: In these features, the sequential value of each position is specified but the distance between these positions is meaningless. For example, in feature of acquiring medal, gold medal has better position than silver medal, but it is not specified that how much this excellence is. Suppose that the number of different modes of rank features j is 1, 2,..., O_j. Calculating the distinction of cases based on these features includes three steps:

- **Step 1:** Replace v_{ij} with number of its sorted position in j. i.e., $r_{ij} \in \{1, 2,..., O_j\}$ which r_{ij} is a rank that is assigned to v_{ij}.
- **Step 2:** Since rank features have different ranges, so we mapped them to [0, 1] through the following equation:

$$Z_{ij} = \frac{r_{ij}-1}{O_j-1}$$

(10)

That O_j is the maximum of possible modes of rank feature j.
- **Step 3:** Now each of the methods of measuring distance can be used:

$$d_{iz}^{Rank} = (|Z_{i1}-Z_{z1}|^q + |Z_{i2}-Z_{z2}|^q + ... + |Z_{il}-Z_{zl}|^q)^{\frac{1}{q}}$$

(11)

Fig. 5: Distinction matrix for binary features

In which l(l≤n) is the number of features of the type of rank in two cases i and z.

And if we define W_j for each feature:

$$d_{iz}^{\text{Rank-Weighted}} = (W_1|Z_{i1} - Z_{z1}|^q + W_2|Z_{i2} - Z_{z2}|^q + ... + W_l|Z_{il} - Z_{zl}|^q)^{\frac{1}{q}}$$

(12)

Symmetric and asymmetric binary features: As mentioned in previous sections, the features which have only two values 0 or 1, are called binary. These features are two types, symmetric and asymmetric. Unlike symmetric features in asymmetric features, only existence (or non-zero value) is important. To measure the distinction between cases i and z, we form the distinction matrix of Fig. 5 for both symmetric and asymmetric.

In which g = a + b + c + d is the number of features of the type of symmetric or asymmetric and *a* is the number of features that their values in both case i and z is equal to 1 and similarly, b, c and d are defined according to Fig. 5. To calculate the distinction between case i and z, if all of binary features are symmetric, we use the following equation:

$$d_{iz}^{\text{Symmetric Binary}} = \frac{b + c}{a + b + c + d}$$

(13)

and if we define W_j for each feature:

$$d_{iz}^{\text{Symmetric Binary-Weighted}} = \frac{b + c}{a + b + c + d}(W_1 + W_2 + ... + W_g)$$

(14)

To calculate the distinction between case i and z for asymmetric binary features, we use the following equation:

$$d_{iz}^{\text{Asymmetric Binary}} = \frac{b + c}{a + b + c}$$

(15)

(Note that d has ommitted because the negative features or with zero value for both cases i and z have the little importance aaccording to the contract) and if we define W_j for each feature:

$$d_{iz}^{\text{Asymmetric Binary-Weighted}} = \frac{b + c}{a + b + c}(W_1 + W_2 + ... + W_{g-d})$$

(16)

Distinction and similarity for each index of feature: Table 1 shows methods of calculating distinction and similarity of each of the feature j. In Table 1 v_{ij} and v_{zi} are the feature value j, respectively, in cases i and z. Also d_{iz}^j shows the features distinction j between two cases i and z and S_{iz}^j shows the feature similarity j between two cases i and z.

Step 2.3: Calculating feature weight: Determining the importance of features is necessary for calculating similarity index. Several methods can be used for determining weight of features that using direct opinions of expert and or analytical methods such as Analytical Hierarchy Process method which it is based on paired comparisons, are samples of it. Another way is using algorithmic methods. One of these algorithmic methods is the genetic algorithm in which set of weights as chromosomes and fitness is calculated through error of classification.

In this study, we use the following method (Montazemi and Gupta, 1997) to calculate the weight of feature:

If we consider the formula S_{ik}^j as the similarity of feature between new case k and case i in case-base. The standard deviation σ_j of the feature similarity along with a feature is a measurement for its variation. The greater the value of σ_j, the more ability S^j has in distinguishing the previous candidate cases:

$$\sigma_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(S_{ik}^j - \bar{S}^j)^2} \quad \forall j.$$

(17)

$$\bar{S}^j = \frac{1}{m}\sum_{i=1}^{m}S_{ik}^j$$

(18)

and m is the number of available previous cases in case-base. Therefore, the importance of feature j^{th} is determined as following:

$$W_j = \frac{\sigma_j}{\sum_I \sigma_j} \quad I \in A^k \cap A^i, \forall j.$$

(19)

In which A^k and A^i are respectively, a set of the features of new case and previous cases (case-base).

Since we use d_{ik}^j in our formulas (steps 2.2) that d_{ik}^j is the distinction (dissimilarity) of feature between new case k and case i in case-base. With placement $S_{ik}^j = 1 - d_{ik}^j$ in the formula 18:

Table 1: Calculating distinction and similarity for each index of feature

Type of feature	Distinction (dissimilarity)	Similarity
Nominal in single mode (and symmetric binary)	$d_{iz}^{j} = \begin{cases} 0 & if \quad v_{ij} = v_{zi} \\ 1 & if \quad v_{ij} \neq v_{zi} \end{cases}$	$S_{iz}^{j} = \begin{cases} 1 & if \quad v_{ij} = v_{zj} \\ 0 & if \quad v_{ij} \neq v_{zj} \end{cases}$
Nominal in multi mode	$d_{iz}^{j} = \dfrac{Max(t_j^i, t_j^z) - S_j}{Max(t_j^i, t_j^z)}$	$S_{iz}^{j} = 1 - d_{iz}^{j}$
Asymmetric binary	When feature j_{th} in both cases i and z is zero (mode d of Fig. 5), we can't define the feature distinction or the feature similarity. Because the value zero for feature of two cases is not the reason for their feature similarity. In other modes, we perform like symmetric binary mode. (of course, we can obtain the distinctions or similarities of all asymmetric binary features with Eq. (15) or (16))	
Rank	$d_{iz}^{j} = \dfrac{|r_{ij} - r_{zj}|}{(O_j - 1)} = |Z_{ij} - Z_{zj}|$	$S_{iz}^{j} = 1 - d_{iz}^{j}$
Interval or ratio	$d_{iz}^{j} = \dfrac{|v_{ij} - v_{zj}|}{max_h v_{hi} - min_h v_{hi}}$ h is the index of all cases of feature j^{th}	$S_{iz}^{j} = \dfrac{1}{1 + d_{iz}^{j}}$ $S_{iz}^{j} = 1 - \dfrac{d_{iz}^{j} - min d_{iz}^{j}}{max d_{iz}^{j} - min d_{iz}^{j}}$

$$\bar{S}^{\,j} = 1 - \frac{1}{m} \sum_{i=1}^{m} d_{ik}^{\,j}$$

(20)

With placement $S_{ik}^{j} = 1 - d_{ik}^{j}$ and formula 20 in formula 17:

$$\sigma_j = \frac{1}{m} \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\sum_{i=1}^{m} d_{ik}^{\,j} - m d_{ik}^{\,j})^2} \quad \forall j.$$

(21)

and

$$\bar{d}^{\,j} = \frac{1}{m} \sum_{i=1}^{m} d_{ik}^{\,j} \Rightarrow \sigma_j = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\bar{d}^{\,j} - d_{ik}^{\,j})^2} \quad \forall j.$$

(22)

Step 2.4: Calculating cases distinction and similarity: If cases have the various features, the distinction between cases i and z is defined as following equation:

$$d(i, z) = \frac{d_{iz}^{Interval\&Ratio} + d_{iz}^{Nominal(C)} + d_{iz}^{Nominal(L)} + d_{iz}^{Rank} + d_{iz}^{Symmetric\ Binary} + d_{iz}^{Asymmetric\ Binary}}{NOF}$$

(23)

In which $1 \leq NOF \leq 6$ is the number of the type of feature in our model.
And if we define W_j for each feature:

$$d(i, z) = \frac{d_{iz}^{Interval\&Ratio-Weighted} + d_{iz}^{Nominal(C)-Weighted} + d_{iz}^{Nominal(L)-Weighted} + d_{iz}^{Rank-Weighted} + d_{iz}^{Symmetric\ Binary-Weighted} + d_{iz}^{Asymmetric\ Binary-Weighted}}{\sum_{j=1}^{n} W_j}$$

(24)

Now if we want to obtain distinction between two cases i and z according to Table 1, we perform as follows. In here, the distinction of features summation is their average individual distinctions. Suppose the set of data consists of n features:

$$d(i, z) = \frac{\sum_{j=1}^{n-g} d_{iz}^{\,j} + d_{iz}^{Asymmetric\ Binary}}{n - g + 1}$$

(25)

In which:

$$d_{iz}^{j} = \begin{cases} 0 & if\ v_{ij}\ or\ v_{zj}\ is\ no\ value \\ 1 & if\ v_{ij}\ and\ v_{zj}\ is\ no \\ d_{iz}^{j} & value\ otherwise \end{cases}$$

And g is the number of features of the type of asymmetric.

If importance of features is different, the weight W_j is given to each feature:

$$d(i,z) = \frac{\sum\limits_{j=1}^{n-g} W_j\ d_{iz}^{j} + d_{iz}^{Asymmetric\ Binary-Weighted}}{\sum\limits_{j=1}^{n} W_j} \quad (26)$$

And to obtaining similarity between cases i and z through Table 1, use the following formula:

$$S(i,z) = \frac{\sum\limits_{j=1}^{n-g} S_{iz}^{j} + S_{iz}^{Asymmetric\ Binary}}{n-g+1}, \quad S_{iz}^{Asymmetric\ Binary} = \frac{a}{a+b+c} \quad (27)$$

In which:

$$S_{iz}^{j} = \begin{cases} 0 & if\ v_{ij}\ or\ v_{zj}\ is\ no\ value \\ 1 & if\ v_{ij}\ and\ v_{zj}\ is\ no \\ S_{iz}^{j} & value\ otherwise \end{cases}$$

And *a*, b and c is explained in section of symmetric and asymmetric binary features.
If importance of features is different, then:

$$S(i,z) = \frac{\sum\limits_{j=1}^{n-g} W_j\ S_{iz}^{j} + S_{iz}^{Asymmetric\ Binary-Weighted}}{\sum\limits_{j=1}^{n} W_j} \quad (28)$$

In which:

$$S_{iz}^{Asymmetric\ Binary-Weighted} = \frac{a}{a+b+c}\left(W_1 + W_2 + ... + W_g - d\right)$$

Note that all features initially are normalized.

Step 2.5: Selecting proposed solutions: In this section, the cases are ranked from the lowest to the highest according to their percentage of distinction which are created in step 2.4. The case with the lowest distinction among Case-Base is most similar to new case. To use adaptation, we need several similar cases. So the cases which their percentage of distinction is not more than a percentage which is important for us, are considered as proposed solutions (retrieved proposed cases).

Step 2.6: Adaptation: Adaptation applies formulas or rules that outstanding differences between retrieved case and new case take into account when suggesting a final solution. While adaptation is useful in many situations, it is by no means essential. Complex adaptation is knowledge intensive and since CBR often is applied to problems which is not good understanding about it, complex knowledge-intensive adaptation would not be possible. In general, we can use two types of manual (interactively by the user) and automatic adaptation for our CBR adaptation.

In manual adaptation, user adapts the retrieved solutions. The reason is that if we want to use the most similar case as the final solution, nature of two cases may be quite different in a series of common features and, thus we will not be allowed to use the feature value of the most similar case in planning new case. So here, the user needs some analysis with regard to the experience and knowledge which he or she has about cases and their features which makes him or her able to reach the features value which require adaptation to acceptable value until the percentage of distinction of two cases comes down as possible (with respect to allowance limits).

In here, we use the automatic adaptation method in most stages because the CBR system which we write its program in C#, can adapt the retrieved solutions. And at the end, we use the manual adaptation method to avoid features value which may change the nature of new case. Adaptation method is as following:

We use the proposed solutions in step 2.5. We put the most similar case (with the lowest percentage of distinction) as criterion. In our proposed model, Percentage of distinction reduces as following:

Manager believes that the percentage of distinction of the most similar case with new case should not exceed β%. Thus, we should try to reduce the percentage of distinction to β% (if it is more than β%) by use of adaptation method. If the percentage of distinction of each feature of the most similar case is more than β%, is substituted the value of that feature as following:

We also examine the percentage of distinction of feature in cases of the proposed solutions and we substitute the feature value of most similar case with the feature value of a case which it has the lowest percentage of distinction of feature. At the end, we calculate the percentage of distinction of modified case.

If is not more than β%, problem solves. But if is more than β%, is substituted its value as following:

We examine other features (that are not examined at the above) of the most similar case, respectively, the highest weight. Like the procedure mentioned at the above, substitution is done. By examining feature with the highest weight, we calculate the percentage of distinction of modified case. If it is not more than β%, problem solves. But if it is more than β%, we examine feature with the second weight and we iterate this work for lower weights until the percentage of distinction of modified case is not more than β%. If we examine all feature of the most similar case and the percentage of distinction of modified case is more than β%, we perform as following:

We examine the percentage of distinction of each feature of modified case, respectively, the highest weight. If the percentage of distinction for examined feature is more than β%, is substituted its value as following:

We also examine the percentage of distinction of examined feature in other cases of case-base (Apart from the proposed solutions). Substitution is done like the procedure mentioned at the above. By examining feature with the highest weight, we calculate the percentage of distinction of new modified case. If is not more than β%, problem solves. But if is more than β%, we examine feature with the second weight and we iterate this work for lower weights until the percentage of distinction of new modified case is not more than β%. In here, the algorithm finishes.

At the end, we use manual adaptation method as follows: we compare each of feature of new modified case with its corresponding feature in new case and if feature value of the new modified case naturally conflicts with its corresponding feature value in new case, we'll modify its value with regard to experience and knowledge which we have about the whole system and consequently from the whole case. Consequently, the final solution is obtained.

Step 2.7: Obtaining output feature of final solution: In this section, we obtain the estimated output or cost of final solution as following:

It easily is specified that value of each features of final solution obtained in Step 2.6 is associated with which case. For example, if our output is the cost, then we extract the cost of each of features from relevant case (which most of them are from the most similar case to new case). Of course, it is noteworthy that we should calculate these costs in the present with regard to time and date of implementing and cost present value of it feature. At the end, we sum all costs.

Consequently, the estimated cost of new case is obtained.

Step 2.8: Model validation: Step 2.8 calculates the performance indicators of CBR model in cost estimation and specifies the validity of model. This section compares the new case with final solution (final case), which is obtained after the adaptation stage and calculate the Standard Error Rate (SER) and the Estimation Accuracy (EA). Equation (29) is used for calculating the SER:

$$SER = \frac{\left| C_{Actual-New\,Case} - C_{CBR-Final\,Solution} \right|}{C_{Actual-New\,Case}} \times 100\%$$
(29)

In which $C_{Actual-New\,Case}$ is the actual cost of new case and $C_{CBR-Final\,Solution}$ is the cost of final solution using CBR that is obtained in step 2.7. Equation (30) is used for calculating the EA:

$$EA = 100 - SER$$
(30)

Now we obtain the Mean Absolute Estimation Error (MAEE) using Eq. (31):

$$MAEE = \frac{\sum_N \frac{\left| C_{Actual-New\,Case} - C_{CBR-Final\,Solution} \right|}{C_{Actual-New\,Case}}}{N} \times 100\%$$
(31)

In which N is the number of new cases (we can consider the representative cases in step 2.9 as new cases and or can consider all of cases as new cases).

Other performance indicator is the Normalized Mean Square Error (NMSE) that is obtained as following:

$$NMSE = \frac{\sum_N \left(C_{Actual-New\,Case} - C_{CBR-Final\,Solution} \right)^2}{\sum_N \left(C_{Actual-New\,Case} - \bar{C} \right)^2}$$
(32)

In which $\bar{C}$ is the average $C_{Actual-New\,Case}$ for new cases.

Step 2.9: Testing case base: This step should be repeated at time intervals during our development of CBR system and should not be left to think that our system is completed. This step involves the following stages:

- **Checking retrieval accuracy:** A case should exactly match itself. For a sample of cases in our case-base, we make a copy of the resource case and use it as a new case. If that case does not retrieve itself or does not match exactly, there is the error in our program in step of retrieval and or adaptation. If we are using the nearest neighbor, distinction measurement should be 0%.

- **Checking retrieval consistency:** If we exactly do the same search again, we should retrieve the same source cases with the same accuracy. If this does not happen, there is again the error in our program in step of retrieval and or adaptation. Thus, we several times repeat a set of searches and check for consistency.

- **Checking case coverage:** It is better to have the same distribution of cases across the space of problem. It is really hard to achieve it, but we can perform as following:

o If we have quantitative features, we check their standard deviation. It should be as low as possible. If the standard deviation is high, we have outlying cases. We should either try to remove them or obtain more cases for improving the coverage of that feature. Average feature value should also be near the middle of the feature range. If the average is toward one end of the range, we have too many cases at one end and not enough at the other.

o If we have qualitative features, check frequency of occurrence of each qualitative value in the case-base. If the distribution is very uneven, then we have outlying cases. We should either try to remove them or obtain more cases for improving the coverage of that feature.

o If we have a feature that will be used as the output of inductive retrieval, we check its distribution very exactly. We want cases to cover that feature value at regular intervals.

- **System overall verification tests:** After checking the above stages, it is important to verify the overall performance of our system:

o We should obtain a number of representative cases (perhaps 5 to 10% of the size of our overall case-base). These cases should be representative of the full range of cases in our case-base.

o Should use each of our representative cases as new cases.

o Should evaluate the system performance. And we may need to help experts to do it. Questions which we should consider as follows:

   o Did the system retrieve a useful case or a set of useful cases?

o Did retrieval time was acceptable?

o If was relevant, did adaptation was successful?

o We should record the results of these tests to compare for next time which we repeat them. Since we will likely retain more cases by case-base, we will exactly not obtain the same answers but we hope to see an improvement.

If we follow this simple step, we make a better CBR system, more reliable and particularly, more useful.

Step 2.10: Retaining case base: CBR system can grow over time. In fact, the growth over time is one of original ability of CBR to learn and improve its performance by acquiring new cases. This step is very similar to verification of case-base. They will not guarantee success, but they certainly will help. This step should be repeated at time regular intervals. The exact period depends on the application which our CBR system provides and the number of new cases which it requests. By no means, size of our CBR system should increase more than 20% without implementing stages of obtaining statistics of case use and verification tests repeat. Therefore, developing of a good case-base needs care.

IMPLEMENTING THE PROPOSED CBR FOR COST ESTIMATION OF DRILLING WELL

This section implements the proposed CBR model mentioned in previous sections for cost estimation in drilling wells.

Data preparation:
Data collection: At the present study, we show the application of CBR method for cost estimation of the drilling of oil and gas wells in design phase in Petro Pars Company and National Iranian Drilling Company (NIDC). Features of drilling project of oil and gas wells and their performance indicators are inferred through extensive literature review and more importantly, based on final reports of different parts and based on available data in reports and important and variable effect on which those features have in different wells cost (the parts which almost are used in all drilling projects both onshore and offshore and or have fixed cost, are not considered as feature) and also interview with experts in the field drilling. The extracted features (67 features) of the drilling of oil and gas wells are shown in Table 2.

Fig. 6: Onshore drilling

Fig. 7: Offshore drilling

Figure 6 and 7 show, respectively, drilling in land (onshore) and sea (offshore) in two above companies.

Filtering data: Type of scale of all features and types each qualitative feature related to drilling wells are shown in Table 2.

Data standardization: This step is inserted in step 2.2.

CBR-based model:
Organizing data: In this step, data consists of two parts: case-base and new case (test case). The case-base includes No. or name of case, input features and output feature (well cost). A case-base is made with 11 cases (wells) and is shown in Table 3 along with new case and completely, in Table 1, 2 and 3 in Appendix.

Calculating distinction with regard to type of feature: Results of this step are shown in Table 4 in Appendix using mentioned equations in step 2.2 of previous sections. Table 4 in Appendix shows d_{ik}^{j} or the distinction of new well k with wells of case-base ($1 \leq i \leq 11$) with regard to type of feature j up to two decimal places.

Calculating feature weight: Results of this step are shown in Table 4 in Appendix.

Calculating cases distinction: Results of this step are shown in Table 4 using mentioned equations in step 2.4

Table 2: Features and their scale

Name of feature	Type of scale	Types of qualitative features
Drilling position	Nominal (C[1])	Onshore; offshore
Well type	Nominal (C)	Exploration; appraisal; development
Exploratory material	Nominal (C)	Oil; gas
Rig drow works (horse power)	Rank	2000; 3000; 4000
Drilling length (m)	Ratio	
Operation time (h)	Ratio	
Non productive time (h)	Ratio	
Average GPM[2] interval of 32" section	Ratio	
Average GPM interval of 24" section	Ratio	
Average GPM interval of 16" section	Ratio	
Average GPM interval of 12 1/4" section	Ratio	
Average GPM interval of 8 1/2" section	Ratio	
Average GPM interval of 5 7/8" section	Ratio	
Average pressure interval of 32" section	Ratio	
Average pressure interval of 24" section	Ratio	
Average pressure interval of 16" section	Ratio	
Average pressure interval of 12 1/4" section	Ratio	
Average pressure interval of 8 1/2" section	Ratio	
Average pressure interval of 5 7/8" section	Ratio	
Rig rotary system of 32" section (average RPM[3] interval)	Ratio	
Rig rotary system of 24" section (average RPM interval)	Ratio	
Rig rotary system of 16" section (average RPM interval)	Ratio	
Rig rotary system of 12 1/4" section (average RPM interval)	Ratio	
Rig rotary system of 8 1/2" section (average RPM interval)	Ratio	
Rig rotary system of 5 7/8" section (average RPM interval)	Ratio	

Table 2: (Continue)

Name of feature	Type of scale	Types of qualitative features
Rig mud pump type	Rank	Duplex; triplex
Logging	Nominal (L[6])	Wireline; LWD[4]; TLC; MWD[5]
Cementing	Nominal (C)	One stage; two stage
Cementing additive (cement class)	Nominal (L)	A; B; C; D; E; F; G
Consumed cement (MT[7] per bulk)	Ratio	
Stimulation services	Nominal (L)	Matrix acidizing; acid wash
Stimulation material	Nominal (L)	Acid HCL 15%; acid HCL 28%; VDA 15%
DST[8]	Asymmetric binary	Yes; no
Well testing	Asymmetric binary	Yes; no
Coring	Asymmetric binary	Yes; no
Perforation	Nominal (L)	Wireline; TCP[9]; drill pipe; CTTCP[10]; CIRP; SPF; openhole with 7" liner
Well profile	Nominal (L)	Vertical; horizontal; deviated; multi lateral
Casing/tubing running hole section	Rank	26" CP[11]; 18 5/8" CSG[12]; 13 3/8" CSG; 10 3/4" CSG; 9 5/8" CSG; 7" liner; 4 1/2" slotted/blank liner CSG; 4 1/2" slotted liner TBG[13]; 7" tie back liner TBG; 2 7/8"-3 1/2" TBG; macaroni pipe 1 3/4"
Fishing services	Asymmetric binary	Yes; no
Slick line	Asymmetric binary	Yes; no
Downhole completion services	Nominal (L)	Because abounding the modes, is not shown here.
Downhole completion equipment	Nominal (C)	Tapered; monobor
Bits	Nominal (L)	PDC[14]; roller cone (milled tooth); roller cone (insert); rock; button; diamond
MUD	Nominal (L)	Water base (sea water); water base (bentonite); water base (polymer); oil base;gas base (compressed air); gas base (foam)
Wellhead services	Asymmetric binary	Yes; no
MLS[15]-services	Asymmetric binary	Yes; no
Special tools	Nominal (L)	MWD; power drive; whipstock; RTTS
Percent CO_2	Ratio	
Percent H_2S	Ratio	
Transport	Nominal (L)	Supply boats; tugs; barges; helicopter; fixed wing airplanes
Producing reservoir	Nominal (L)	K-1; K-2; K-3; K-4; K-5
Water depth	Ratio	
Inclination of 32" section (α)	Interval	
Inclination of 24" section (α)	Interval	
Inclination of 16" section (α)	Interval	
Inclination of 12 1/4" section (α)	Interval	
Inclination of 8 1/2" section (α)	Interval	
Inclination of 5 7/8" section (α)	Interval	
Azimuth of 32" section (α)	Interval	
Azimuth of 24" section (α)	Interval	
Azimuth of 16" section (α)	Interval	
Azimuth of 12 1/4" section (α)	Interval	
Azimuth of 8 1/2" section (α)	Interval	
Azimuth of 5 7/8" section (α)	Interval	
Max reservoir pressure (psi)	Ratio	
Pressure gradient of gas in tubing (psi/ft)	Ratio	
Pressure gradient of oil in tubing (psi/ft)	Ratio	

[1]: Combo box; [2]: Gallons per minute; [3]: Revolutions per minute; [4]: Logging while drilling; [5]: Measurement while drilling; [6]: List box; [7]: Metric tone; [8]: Drill stem test; [9]: Tubing conveyed perforator; [10]: Coiled tubing TCP; [11]: Conductor pipe; [12]: Casing; [13]: Tubing; [14]: Polycrystalline diamond compact; [15]: Mud line suspension

of previous sections. Table 4 shows d (i, k) or the distinction of new well k with wells of case-base ($1 \leq i \leq 11$) up to four decimal places.

Selecting proposed solutions: In previous section, case or well No. 11 is the most similar case to new case. To use adaptation, we need several similar cases. So cases are considered as proposed solutions that their distinction percentage is not more than 35% (a percentage which is important for us) that hereby, cases No. 7, 10 and 11 are retrieved.

Adaptation: Manager believes that the percentage of distinction for final solution should not be more than entage of distinction for final solution should not be more than 15%. The CBR system which we write its program in C#, adapted the retrieved solutions. The final solutiondid not need the manual adaptation because there was no feature value that could change the nature of new case. The percentage of distinction reduced to 10.68% Value of each of features of final solution is shown in Table 5.

Table 3: Case-base

Name of feature	Well 1	…	Well 11	New well (test)
Drilling position	Onshore	…	Offshore	Offshore
Well type	Development	…	Development	Development
Exploratory material	Gaz	…	Gaz	Gaz
Cementing	Two stage	…	Two stage	One stage
Downhole completion equipment	Monobor	…	Monobor	Monobor
Logging	LWD, MWD	…	LWD, MWD	LWD, MWD
Cementing additive (cement class)	D, G	…	E, F, G	G
Stimulation services	Acid_wash	…	Matrix_acidizing	Acid_wash, matrix_acidizing
Stimulation material	Acid_HCL_28%	…	Acid_HCL_28%	Acid_HCL_15%, acid_HCL_28%, VDA_15%
Perforation	CIRP, drill_pipe	…	CTTCP, SPF	Openhole_with_7"_liner
Well profile	Vertical	…	Deviated	Vertical
Downhole completion services	Flowing (high_rate_liner), multiple_zone (concentric_string), perforated_casing_for_vertical/deviated_wells (standard)	…	Flowing (high_rate_liner), multiple_zone (concentric_string), perforated_casing_for_vertical/deviated_wells (standard)	Flowing (high_rate_liner), multiple_zone (multiple_strings), perforated_casing_for_vertical/deviated_wells (standard)
Bits	Rock, roller_cone (milled tooth)	…	Roller_cone (insert), roller_cone (milled tooth)	pdc, rock, roller_cone (insert), roller_cone (milled tooth)
MUD	Oil_base, water_base (bentonite)	…	Water_base (sea water)	Gas_base (compressed_air), oil_base, water_base (bentonite)
Special tools	MWD, RTTS	…	MWD, power_drive	MWD, power_drive
Transport	Helicopter, tugs	…	Barges, helicopter, supply_boats, truck	Barges, fixed_wing_airplanes, helicopter, supply_boats, tugs
Rig drow works (horse power)	3000	…	3000	4000
Rig mud pump type	Duplex	…	Triplex	Triplex
Casing/tubing running hole section	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 7" liner, 9 5/8" CSG	…	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 7" Liner, 9 5/8" CSG	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 4 1/2" slotted liner TBG, 4 1/2" slotted/blank liner CSG, 7" liner, 9 5/8" CSG
Producing reservoir	K-1, K-2, K-3, K-4	…	K-1, K-2, K-3, K-4	K-1, K-2, K-3, K-4
DST	No	…	No	No
Well testing	No	…	No	Yes
Coring	No	…	No	No
Fishing services	Yes	…	Yes	Yes
Slick line	Yes	…	Yes	Yes
Wellhead services	Yes	…	Yes	Yes
MLS-services	No	…	No	No
Drilling length (m)	4100.000	…	4650.000	4840.000
Operation time (h)	1900.000	…	2100.000	2600.000
Non productive time (h)	120.0000	…	102.0000	94.00000
Average GPM interval of 32" section	850.0000	…	925.0000	1050.000
Average GPM interval of 24" section	800.0000	…	900.0000	800.0000
Average GPM interval of 16" section	1000.000	…	950.0000	925.0000
Average GPM interval of 12 1/4" section	750.0000	…	975.0000	950.0000
Average GPM interval of 8 1/2" section	400.0000	…	385.0000	410.0000

Table. 3: (Continue)

Name of feature	Well 1	…	Well 11	New well (test)
Average GPM interval of 5 7/8" section		…	850.000	
Average pressure interval of 32" section	1150.000	…	1210.00	1250.000
Average pressure interval of 24" section	1200.000	…	1280.00	1290.000
Average pressure interval of 16" section	2000.000	…	2100.00	2225.000
Average pressure interval of 12 1/4" section	2300.000	…	2245.00	2525.000
Average pressure interval of 8 1/2" section	2150.000	…	2300.00	2300.000
Average pressure interval of 5 7/8" section		…	1950.00	
Rig rotary system of 32" section (average RPM interval)	45.0000	…	61.0000	50.0000
Rig rotary system of 24" section (average RPM interval)	45.0000	…	41.5000	50.0000
Rig rotary system of 16" section (average RPM interval)	55.0000	…	57.0000	50.0000
Rig rotary system of 12 1/4" section (average RPM interval)	50.0000	…	55.0000	49.5000
Rig rotary system of 8 1/2" section (average RPM interval)	37.5000	…	35.0000	35.0000
Rig rotary system of 5 7/8" Section (average RPM interval)		…	37.5000	
Consumed cement (MT per bulk)	400.000	…	387.000	376.000
Percent CO_2	2.00000	…	2.00000	2.00000
Percent H_2S	0.50000	…	0.50000	0.50000
Water depth	67.8000	…	67.8000	60.0000
Max reservoir pressure (psi)	4800.000	…	5050.00	5200.000
Pressure gradient of gas in tubing (psi/ft)	0.11300	…	0.11500	0.12100
Pressure gradient of oil in tubing (psi/ft)				
Inclination of 32" section (α)	0.20000	…	0.23000	0.31000
Inclination of 24" section (α)	3.70000	…	3.50000	4.50000
Inclination of 16" section (α)	56.0000	…	50.0000	64.0000
Inclination of 12 1/4" section (α)	68.0000	…	63.0000	75.0000
Inclination of 8 1/2" section (α)	77.0000	…	85.0000	83.0000
Inclination of 5 7/8" section (α)		…	88.0000	
Azimuth of 32" section (α)	246.400	…	208.190	232.500
Azimuth of 24" section (α)	9.40000	…	315.050	293.700
Azimuth of 16" section (α)	203.500	…	314.390	345.000
Azimuth of 12 1/4" section (α)	56.4000	…	313.900	311.900
Azimuth of 8 1/2" section (α)	164.600	…	312.000	287.300
Azimuth of 5 7/8" section (α)		…	307.500	
Cost (million dollars)	11.0000	**…**	44.0000	45.2000

Table 4: Distinction of cases

	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8	Well 9	Well 10	Well 11
Distinction	0.4223	0.4789	0.4268	0.4406	0.4183	0.4610	0.3232	0.4127	0.3878	0.3211	0.2660

Obtaining output feature of final solution: In this step, we obtain the estimated drilling cost of new well or cost of final solution as follows:

We extracted the cost of each feature of final solution from relevant case (which most of them are from case No. 11). These costs are available in the files of well cost of company in department of planning and project management. These costs are calculated at present with regard to time and date of implementing and cost present value and along with the relevant case number are shown in Table 5. At the end, we sum all costs. Consequently, the estimated cost of new well is obtained 43.1 million dollars.

Model validation: SER and EA are calculated using mentioned equations in step 2.8 of previous sections. For new (test) case:

$$SER = \frac{|45.2 - 43.1|}{45.2} \times 100\% = 4.65\% \quad , \quad EA = 100 - 4.65 = 95.35\%$$

Table 5: The final solution and the cost of each features of final solution along with the relevant case number

Name of feature	Final solution	The case number of each feature	The extracted cost with regard to base year
Drilling position	Offshore	11	0.221478000
Well type	Development	11	0.261869879
Exploratory material	Gaz	11	0.380901642
Cementing	Two stage	11	0.535733000
Downhole completion equipment	Monobor	11	1.337824000
Logging	LWD, MWD	11	0.740458000
Cementing additive (cement class)	D, G	10	0.133794000
Stimulation services	Acid_wash, matrix_acidizing	7	0.129300000
Stimulation material	Acid_HCL_15%, acid_HCL_28%, VDA_15%	10	0.183606000
Perforation	CTTCP, SPF	11	0.781834000
Well profile	Deviated, vertical	7	0.904641400
Downhole completion services	Flowing (high_rate_liner), multiple_zone (concentric_string), perforated_casing_for_vertical/deviated_wells (standard)	11	0.629260000
Bits	Roller_cone (insert), roller_cone (milled tooth)	11	0.497374000
MUD	Gas_base (compressed_air), water_base (bentonite)	10	1.295586000
Special tools	MWD, power_drive	7	0.009913000
Transport	Barges, helicopter, supply_boats, truck	11	1.956740000
Rig drow works (horse power)	4000	10	0.552124075
Rig mud pump type	Triplex	11	0.023806353
Casing/tubing running hole section	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 4 1/2" slotted/blank liner CSG, 7" liner, 9 5/8" CSG	10	8.439842000
Producing reservoir	K-1, K-2, K-3, K-4	11	1.333155747
DST	No	11	-
Well testing	Yes	10	0.635725000
Coring	No	11	-
Fishing services	Yes	11	0.139213000
Slick line	Yes	11	0.208173000
Wellhead services	Yes	11	0.087493000
MLS-services	No	11	-
Drilling length (m)	4650.00	11	0.238063526
Operation time (h)	2300.00	7	14.54305711
Non productive time (h)	102.000	11	0.767815769
Average GPM interval of 32" section	1110.00	7	0.012973100
Average GPM interval of 24" section	750.000	7	0.095035500
Average GPM interval of 16" section	950.000	11	0.114646000
Average GPM interval of 12 1/4" section	975.000	11	0.368375700
Average GPM interval of 8 1/2" section	390.000	10	0.473518150
Average GPM interval of 5 7/8" section	850.000	11	0.168952000
Average pressure interval of 32" section	1210.00	11	0.007413200
Average pressure interval of 24" section	1280.00	11	0.054306000
Average pressure interval of 16" section	2205.00	10	0.065512000
Average pressure interval of 12 1/4" section	2450.00	7	0.210500400
Average pressure interval of 8 1/2" section	2300.00	11	0.270581800
Average pressure interval of 5 7/8" section	1950.00	11	0.096544000
Rig rotary system of 32" section (average RPM interval)	55.0000	10	0.009266500
Rig rotary system of 24" section (average RPM interval)	44.0000	7	0.067882500
Rig rotary system of 16" section (average RPM interval)	55.0000	7	0.081890000
Rig rotary system of 12 1/4" section (average RPM interval)	53.5000	10	0.263125500
Rig rotary system of 8 1/2" section (average RPM interval)	35.0000	11	0.338227250
Rig rotary system of 5 7/8" section (average RPM interval)	37.5000	11	0.120680000

Table 5: (Continue)

Name of feature	Final solution	The case number of each feature	The extracted cost with regard to base year
Consumed cement (MT per bulk)	387.00	11	0.757945000
Percent CO_2	2.0000	11	0.142838116
Percent H_2S	0.5000	11	0.230154000
Water depth	67.800	11	0.857028695
Max reservoir pressure (psi)	5050	11	0.571352463
Pressure gradient of gas in tubing (psi/ft)	0.1200	7	0.047612705
Pressure gradient of oil in tubing (psi/ft)			-
Inclination of 32" section (α)	0.2800	10	0.004447920
Inclination of 24" section (α)	42.000	10	0.032583600
Inclination of 16" section (α)	59.000	10	0.039307200
Inclination of 12 1/4" section (α)	68.000	10	0.126300240
Inclination of 8 1/2" section (α)	85.000	11	0.162349080
Inclination of 5 7/8" section (α)	88.000	11	0.057926400
Azimuth of 32" section (α)	255.70	10	0.002965280
Azimuth of 24" section (α)	315.05	11	0.021722400
Azimuth of 16" section (α)	314.39	11	0.026204800
Azimuth of 12 1/4" section (α)	313.90	11	0.084200000
Azimuth of 8 1/2" section (α)	312.00	11	0.108233000
Azimuth of 5 7/8" section (α)	307.50	11	0.038618000
Total cost (million dollars)			43.1

Table 6: The estimated cost of wells use of CBR method

Well number	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8	Well 9	Well 10	Well 11	Well 12
The estimated cost (million dollars)	9.6	11.53	44.13	46.93	11.15	47.27	43.16	15.05	37.78	38.86	41.67	43.1

In this step, are chosen all of cases in case-base as new or test cases. The estimated cost of all cases (wells) is shown in Table 6.

Now, performance indicators MAEE and NMSE are calculated using Eq. (31) and (32):

$$MAEE = \frac{\frac{|11-9.6|}{11} + \frac{|12.9-11.53|}{12.9} + ... + \frac{|45.2-43.1|}{45.2}}{12} \times 100\% = 7.63\%$$

$$NMSE = \frac{(11-9.6)^2 + (12.9-11.53)^2 + ... + (45.2-43.1)^2}{(11-33.483)^2 + (12.9-33.483)^2 + ... + (45.2-33.483)^2} = 2.09\%$$

Testing case base: This step involves the following stages:

• **Checking retrieval accuracy:** We chosen the sampels of cases (case of the previous step) in our case-base and each of them used separately in the case-base as well as a new case. Each of them exactly adapted itself and percentage of distinction was 0%.

• **Checking retrieval consistency:** We exactly repeated search several times for new case. In each time, we retrieved the same proposed solutions with the same accuracy.

• **Checking case coverage:** To have the same distribution of cases across the space of problem, we performed as following:

o For quantitative features in our case-base, we checked their standard deviation. The value of each

of them is as low as possible and desirable. The average value of each feature is also near the middle of the feature range and desirable.

o For qualitative features in our case-base, we checked the frequency of occurrence of each qualitative value in the case-base. Distribution is not very uneven.

o Since we did not use inductive retrieval in this study, we do not need to exactly check distribution of the output feature i.e., cost of drilling wells.

• **System overall verification tests:** After checking the above stages, we address to check verification of our system overall performance:

o We obtain a number of representative cases (5% of the size of our overall case-base). These cases are representative of the full range of cases in our case-base (cases of the previous step).

o We used each of our representative cases as new cases.

o We evaluate the system performance. System retrieved a set of useful cases for each of representative cases. Retrieval time was acceptabl and Adaptation was successful.

o We recorded the results of these tests for comparison of next time (which our case-base has more cases) that we repeat these representative cases as new cases.

Finally, we made a better CBR system, more reliable and particularly, more useful.

Retaining case base: This step doesn't have significant difference with verification of case-base. This step is repeated at time intervals when new data of drilling wells is obtained. By no means, size of our CBR system should increase more than 20% without implementing stages of obtaining statistics of case use and verification tests repeat.

RESULTS ANALYSIS

The estimated cost value of the CBR model along with the actual costs is shown graphically in Fig. 8 (horizontal axis is the number of well and vertical axis is the cost (million dollars)).

As you have seen, we used the performance indicators MAEE and NMSE for validation of CBR model. Indicator MAEE is obtained 7.63%. Also indicator NMSE is obtained 2.09%. So CBR model is powerful and accurate method in field of costs estimation.

At the end, it is noteworthy that as regards the number of features in model is very much greater than the number of wells, the accuracy of method is lower than the its actual accuracy. The more wells we use the data, the less the error of model. In here, main
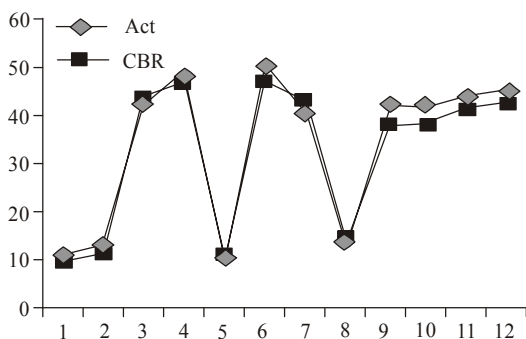

Fig. 8: Comparison of the estimated cost with the actual cost of wells

advantage of in the proposed CBR is that in spite of limited data (low number of cases), the error of method is very low.

CONCLUSION

In this research, we estimated the cost of drilling wells using CBR method. In the proposed CBR system, we didn't need to know how the problem is solved. To build a CBR system, we extract only previous cases and their solutions and we were not required to extract the rules from experts.

Most estimation methods are good in areas which data displays purely numerical. On the contrary, the proposed CBR system is less good for pure numerical data and it is much better for complex data with qualitative structure. In this study, which independence of the features may not be recognized, the proposed CBR system is most appropriate.

Unlike other methods in the proposed CBR system, we did not need output data for all defined cases in case base, which output data in this study is the cost of drilling wells, because our purpose is primarily to obtain the cases similarity which for this purpose, we did not need the wells cost. In this stage, we needed only the cost of the most similar case. Then in stage of adaptation, the cases are specified which we need their costs.

The proposed CBR system, is made where no model exist and can learn and grow by aquiring new cases in order to reflect its experience. After the proposed CBR system was developed, when the new case is added to the case base, they are updated. In other methods, it is very difficult to update.

In the proposed CBR method, if a feature doesn't have the value, it does not affect the model. While in other methods, if a feature doesn't have the value, the model is not solved. And if we put zero instead of that

APPENDIX

Table 1: Case-base

Name of feature	Well 1	Well 2	Well 3	Well 4
Drilling position	Onshore	Onshore	Offshore	Offshore
Well type	Development	Development	Development	Appraisal
Exploratory material	Gaz	Oil	Gaz	Gaz
Cementing	Two stage	Two stage	Two stage	Two stage
Downhole completion equipment	Monobor	Monobor	Monobor	Tapered
Logging	LWD, MWD	TLC, wireline	LWD	LWD, MWD
Cementing additive (cement class)	D, G	E, G	G	E, F, G
Stimulation services	Acid_wash	Acid_wash, matrix_acidizi	Matrix_acidizing	Matrix_acidizing
Stimulation material	Acid_HCL_28%	Acid_HCL_28%	Acid_HCL_28%	Acid_HCL_28%
Perforation	CIRP, drill_pipe	CTTCP, TCP	CIRP, SPF	CIRP, TCP
Well profile	Vertical	Deviated, vertical	Deviated, horizontal	Deviated

Table 1: (Continue)

Name of feature	Well 1	Well 2	Well 3	Well 4
Downhole completion services	Flowing (high_rate_liner), multiple_zone (concentric_string), perforated_casing_for_vertical/deviated_wells (standard)	Flowing (high_rate_liner), multiple_zone (concentric_string), perforated_casing_for_vertical/deviated_wells (standard)	Flowing (high_pressure), multiple_zone (multiple_strings), perforated_casing_for_vertical/deviated_wells (standard)	Flowing (high_rate_liner), multiple_zone (concentric_string), perforated_casing_for_vertical/deviated_wells (standard)
Bits	Rock, roller_cone (milled tooth)	Roller_cone (insert), roller_cone (milled tooth)	Button, PDC, roller_cone (milled tooth)	Button, roller_cone (insert), roller_cone (milled tooth)
MUD	Oil_base, water_base (bentonite)	Gas_base (compressed_air), oil_base	Gas_base (compressed_air), water_base (sea water)	Gas_base (compressed_air), water_base (bentonite), water_base (polymer)
Special tools	MWD, RTTS	MWD, power_drive	Power_drive, RTTS, whipstock	MWD, power_drive, whipstock
Transport	Helicopter, tugs	Helicopter, supply_boats	Barges, helicopter, truck	Helicopter, tugs
Rig drow works (horse power)	3000	2000	3000	4000
Rig mud pump type	Duplex	Triplex	Triplex	Triplex
Casing/tubing running hole section	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 7" liner, 9 5/8" CSG	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 7" liner, 9 5/8" CSG	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 4 1/2" slotted/blank liner CSG, 7" liner, 9 5/8" CSG	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 4 1/2" slotted/blank liner CSG, 7" liner, 9 5/8" CSG
Producing reservoir	K-1, K-2, K-3, K-4	K-1, K-2, K-3, K-4	K-1, K-2, K-3	K-1, K-2, K-3, K-4
DST	No	No	Yes	No
Well testing	No	No	No	No
Coring	No	No	No	Yes
Fishing services	Yes	Yes	No	Yes
Slick line	Yes	Yes	Yes	Yes
Wellhead services	Yes	Yes	Yes	No
MLS-services	No	No	No	Yes
Drilling length (m)	4100.000	3400.000	5100.000	4615.000
Operation time (h)	1900.000	2050.000	2400.000	2700.000
Non productive time (h)	120.0000	114.0000	149.0000	86.40000
Average GPM interval of 32" section	850.0000	850.0000	975.0000	850.0000
Average GPM interval of 24" section	800.0000	800.0000	900.0000	825.0000
Average GPM interval of 16" section	1000.000	975.0000	850.0000	825.0000
Average GPM interval of 12 1/4" section	750.0000	750.0000	825.0000	850.0000
Average GPM interval of 8 1/2" section	400.0000	375.0000	365.0000	335.0000
Average GPM interval of 5 7/8" section				800.0000
Average pressure interval of 32" section	1150.000	1150.000	1150.000	1165.000
Average pressure interval of 24" section	1200.000	1250.000	1350.000	1150.000
Average pressure interval of 16" section	2000.000	2000.000	2050.000	2075.000
Average pressure interval of 12 1/4" section	2300.000	2225.000	2375.000	2300.000
Average pressure interval of 8 1/2" section	2150.000	2550.000	2400.000	2525.000
Average pressure interval of 5 7/8" section				2100.000

Table 1: (Continue)

Name of feature	Well 1	Well 2	Well 3	Well 3
Rig rotary system of 32" section (average RPM interval)	45.0000	50000	53.0000	60.0000
Rig rotary system of 24" section (average RPM interval)	45.0000	40000	42.5000	35.0000
Rig rotary system of 16" section (average RPM interval)	55.0000	50.0000	60.0000	55.0000
Rig rotary system of 12 1/4" section (average RPM interval)	50.0000	55.0000	52.0000	47.5000
Rig rotary system of 8 1/2" section (average RPM interval)	37.5000	34.0000	40.0000	42.5000
Rig rotary system of 5 7/8" section (average RPM interval)				32.5000
Consumed cement (MT per bulk)	400.000	450.000	380.000	390.000
Percent CO_2	2.00000	1.5000	2.0000	2.0000
Percent H_2S	0.50000	0.7000	0.5000	0.5000
Water depth	67.8000	0.0000	67.800	60.000
Max reservoir pressure (psi)	4800.00	4500.00	4700.00	4300.00
Pressure gradient of gas in tubing (psi/ft)	0.11300		0.1080	0.1050
Pressure gradient of oil in tubing (psi/ft)		0.11000		
Inclination of 32" section (α)	0.2000	0.30000	0.2400	0.4000
Inclination of 24" section (α)	3.7000	4.50000	4.0000	5.0000
Inclination of 16" section (α)	56.000	60.0000	57.000	64.000
Inclination of 12 1/4" section (α)	68.000	70.0000	70.000	75.000
Inclination of 8 1/2" section (α)	77.000	80.0000	84.000	86.000
Inclination of 5 7/8" section (α)				89.000
Azimuth of 32" section (α)	246.40	237.000	162.69	186.40
Azimuth of 24" section (α)	9.4000	120.400	3.2600	64.600
Azimuth of 16" section (α)	203.50	110.800	12.410	83.900
Azimuth of 12 1/4" section (α)	56.400	115.000	13.260	76.340
Azimuth of 8 1/2" section (α)	164.60	194.000	18.950	54.350
Azimuth of 5 7/8" section (α)				112.60
Cost (million dollars)	11	12.9	42.300	48.900

Table 2: Case-base

Name of feature	Well 5	Well 6	Well 7	Well 8
Drilling position	Onshore	Offshore	Offshore	Onshore
Well type	Development	Appraisal	Development	Development
Exploratory material	Oil	Oil	Gaz	Oil
Cementing	One stage	Two stage	Two stage	One stage
Downhole completion equipment	Monobor	Tapered	Monobor	Tapered
Logging	LWD, wireline	MWD, wireline	LWD, MWD	LWD, MWD
Cementing additive (cement class)	D, G	C, E, F, G	C, E, G	C, F, G
Stimulation services	Matrix_acidizing	Matrix_acidizing	Acid_wash, matrix_acidizing	Acid_wash
Stimulation material	Acid_HCL_28%	Acid_HCL_15%, VDA_15%	Acid_HCL_28%	Acid_HCL_28%
Perforation	CIRP, CTTCP, SPF	SPF, TCP	TCP, wireline	CIRP, SPF
Well profile	Deviated, vertical	Deviated, horizontal	Deviated, vertical	Deviated, horizontal
Downhole completion services	Flowing (high_rate_liner), multiple_zone (multiple_strings), perforated_liner_for_vertical/deviated_wells	Flowing (high_pressure), multiple_zone (concentric_string), perforated_casing_for_vertical/deviated_wells (standard)	Flowing (high_rate_liner), multiple_zone (concentric_string), perforated_liner_for_vertical/deviated_wells	Flowing (high_pressure), multiple_zone (multiple_strings), perforated_casing_for_vertical/deviated_wells (standard)
Bits	Roller_cone (insert), roller_cone (milled tooth)	Button, diamond, rock, roller_cone (insert), roller_cone (milled tooth)	Button, diamond, rock	Diamond, rock, roller_cone (milled tooth)

Table 2: (Continue)

Name of feature	Well 5	Well 6	Well 7	Well 8
MUD	Gas_base (foam), oil_base, water_base (sea water)	Gas_base (compressed_air), water_base (polymer)	Gas_base (compressed_air)	Water_base (bentonite), water_base (polymer)
Special tools	Power_drive, RTTS, whipstock	MWD, power_drive, RTTS, whipstock	MWD, power_drive	MWD, power_drive, whipstock
Transport	Barges, helicopter, supply_boats, truck	Fixed_wing_airplanes, helicopter, supply_boats, truck	Helicopter, tugs	Helicopter, truck, tugs
Rig drow works (horse power)	3000	2000	3000	3000
Rig mud pump type	Duplex	Duplex	Triplex	Triplex
Casing/tubing running hole section	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 7" liner, 9 5/8" CSG	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 4 1/2" slotted liner TBG, 4 1/2" slotted/blank liner CSG, 7" liner, 7" tie back liner TBG, 9 5/8" CSG	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 7" liner, 9 5/8" CSG	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 4 1/2" slotted/blank liner CSG, 7" liner, 9 5/8" CSG
Producing reservoir	K-1, K-2	K-1, K-2, K-3, K-4	K-1, K-2	K-1, K-2, K-3, K-4
DST	No	Yes	No	No
Well testing	Yes	No	No	Yes
Coring	No	Yes	No	No
Fishing services	Yes	Yes	No	Yes
Slick line	No	Yes	Yes	Yes
Wellhead services	Yes	Yes	Yes	No
MLS-services	No	Yes	No	No
Drilling length (m)	3700	4585.00	4510.00	3300.00
Operation time (h)	2500.00	2400.00	2300.00	1800.00
non productive time (h)	110.000	135.600	98.4000	76.0000
Average GPM interval of 32" section	1050.00	1025.00	1110.00	835.000
Average GPM interval of 24" section	925.000	800.000	750.000	900.000
Average GPM interval of 16" section	900.000	875.000	850.000	950.000
Average GPM interval of 12 1/4" section	1000.00	850.000	975.000	800.000
Average GPM interval of 8 1/2" section	425.000	450.000	385.000	350.000
Average GPM interval of 5 7/8" section			775.000	
Average pressure interval of 32" section	1300.000	1350.00	1150.00	1250.00
Average pressure interval of 24" section	1275.000	1250.00	1280.00	1240.00
Average pressure interval of 16" section	2200.000	2300.00	2085.00	2315.00
Average pressure interval of 12 1/4" section	2475.000	2500.00	2450.00	2550.00
Average pressure interval of 8 1/2" section	2400.000	2375.00	2225.00	2415.00
Average pressure interval of 5 7/8" section			1825.00	
Rig rotary system of 32" section (average RPM interval)	60.0000	62.500	57.500	48.000
Rig rotary system of 24" section (average RPM interval)	42.5000	49.500	44.000	42.000
Rig rotary system of 16" section (average RPM interval)	61.0000	53.000	55.000	55.000
Rig rotary system of 12 1/4" section (average RPM interval)	60.0000	54.000	61.000	65.000

Table 2: (Continue)

Name of feature	Well 5	Well 6	Well 7	Well 8
Rig rotary system of 8 1/2" section (average RPM interval)	36.5000	44.500	39	37.50
Rig rotary system of 5 7/8" section (average RPM interval)			40	
Consumed cement (MT per bulk)	367.000	500.00	450.00	437.00
Percent CO_2	1.5000	2.0000	2.0000	1.5000
Percent H_2S	0.7000	0.5000	0.5000	0.7000
Water depth	0.0000	60.000	67.800	0.0000
Max reservoir pressure (psi)	5000.00	5400.0	5100.0	4850.0
Pressure gradient of gas in tubing (psi/ft)			0.1200	
Pressure gradient of oil in tubing (psi/ft)	0.1190	0.1400		0.1000
Inclination of 32" section (α)	0.3400	0.2700	0.2300	0.3300
Inclination of 24" section (α)	4.6000	3.8000	3.7000	4.8000
Inclination of 16" section (α)	60.000	50.000	50.000	68.000
Inclination of 12 1/4" section (α)	73.000	60.000	63.000	82.000
Inclination of 8 1/2" section (α)	82.000	73.000	74.000	89.000
Inclination of 5 7/8" section (α)			80.000	
Azimuth of 32" section (α)	218.840	175.70	199.60	245.80
Azimuth of 24" section (α)	242.400	240.70	119.23	319.00
Azimuth of 16" section (α)	255.420	253.40	100.88	332.70
Azimuth of 12 1/4" section (α)	262.750	260.50	114.77	300.60
Azimuth of 8 1/2" section (α)	204.350	225.70	92.600	336.80
Azimuth of 5 7/8" section (α)			136.80	
Cost (million dollars)	10.2000	50.000	40.700	13.300

Table 3: Case-base

Name of feature	Well 9	Well 10	Well 11	New well (test)
Drilling position	Offshore	Offshore	Offshore	Offshore
Well type	Development	Development	Development	Development
Exploratory material	Gaz	Gaz	Gaz	Gaz
Cementing	Two stage	Two stage	Two stage	One stage
Downhole completion equipment	Monobor	Monobor	Monobor	Monobor
Logging	LWD, MWD	LWD	LWD, MWD	LWD, MWD
Cementing additive (cement class)	G	D, G	E, F, G	G
Stimulation services	Acid_wash, matrix_acidizing	Acid_wash, matrix_acidizing	Matrix_acidizing	Acid_wash, matrix_acidizing
Stimulation material	Acid_HCL_28%	Acid_HCL_15%, acid_HCL_28%, VDA_15%	Acid_HCL_28%	Acid_HCL_15%, acid_HCL_28%, VDA_15%
Perforation	CIRP, CTTCP	CTTCP, SPF	CTTCP, SPF	Openhole_with_7"_liner
Well profile	Horizontal	Deviated	Deviated	Vertical

Table 3: (Continue)

Name of feature	Well 9	Well 10	Well 11	New well (test)
Downhole completion services	Flowing (high_rate_liner), multiple_zone (concentric_string), perforated_casing_f or_vertical/deviated_wells (standard)	Flowing (high_pressure), multiple_zone (concentric_string), perforated_liner_for_ver tical/deviated_wells	Flowing (high_rate_liner), multiple_zone (concentric_string), perforated_casing_for_v ertical/deviated_wells (standard)	Flowing (high_rate_liner), multiple_zone (multiple_strings), perforated_casing_for_ve rtical/deviated_wells (standard)
Bits	Rock, roller_cone (insert)	Button, roller_cone (milled tooth)	Roller_cone (insert), roller_cone (milled tooth)	PDC, rock, roller_cone (insert), roller_cone (milled tooth)
MUD	Water_base(sea water)	Gas_base(compressed_a ir), water_base(bentonite)	Water_base(sea water)	Gas_base(compressed_ai r), oil_base, water_base(bentonite)
Special tools	MWD	Power_drive, whipstock	MWD, power_drive	MWD, power_drive
Transport	Helicopter, supply_boats, truck	Barges, supply_boats	Barges, helicopter, supply_boats, truck	Barges, fixed_wing_airplanes, helicopter, supply_boats, tugs
Rig drow works (horse power)	4000	4000	3000	4000
Rig mud pump type	Duplex	Triplex	Triplex	Triplex
Casing/tubing running hole section	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 7" liner, 9 5/8" CSG	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 4 1/2" slotted/blank liner CSG, 7" liner, 9 5/8" CSG	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 7" liner, 9 5/8" CSG	10 3/4" CSG, 13 3/8" CSG, 18 5/8" CSG, 26" CP, 4 1/2" slotted liner TBG, 4 1/2" slotted/blank liner CSG, 7" liner, 9 5/8" CSG
Producing reservoir	K-1, K-2, K-3	K-1, K-2, K-3, K-4	K-1, K-2, K-3, K-4	K-1, K-2, K-3, K-4
DST	No	No	No	No
Well testing	No	Yes	No	Yes
Coring	No	No	No	No
Fishing services	Yes	Yes	Yes	Yes
Slick line	No	Yes	Yes	Yes
Wellhead services	Yes	Yes	Yes	Yes
MLS-services	No	No	No	No
Drilling length (m)	4756.800	4515.0	4650.000	4840.00
Operation time (h)	2037.000	1950.0	2100.000	2600.00
Non productive time (h)	61.50000	85.000	102.0000	94.0000
Average GPM interval of 32" section	1000.000	965.00	925.0000	1050.00
Average GPM interval of 24" section	825.000	700.00	900.0000	800.000
Average GPM interval of 16" section	900.000	1000.0	950.0000	925.000
Average GPM interval of 12 1/4" section	775.000	925.00	975.0000	950.000
Average GPM interval of 8 1/2" section	365.000	390.00	385.0000	410.000
Average GPM interval of 5 7/8" section			850.0000	
Average pressure interval of 32" section	1200.000	1350.0	1210.000	1250.000
Average pressure interval of 24" section	1300.000	1150.0	1280.000	1290.000
Average pressure interval of 16" section	2125.000	2205.0	2100.000	2225.000
Average pressure interval of 12 1/4" section	2400.000	2260.0	2245.000	2525.000
Average pressure interval of 8 1/2" section	2450.000	2475.0	2300.000	2300.000
Average pressure interval of 5 7/8" section			1950.000	
Rig rotary system of 32" section (average RPM interval)	55.0000	55.000	61.0000	50.0000

Table 3: (Continue)

Name of feature	Well 9	Well 10	Well 11	Well 12
Rig rotary system of 24" section (average RPM interval)	42.500	43.50	41.500	50.000
Rig rotary system of 16" section (average RPM interval)	60.000	57.50	57.000	50.000
Rig rotary system of 12 1/4" section (average RPM interval)	55.000	53.50	55.000	49.500
Rig rotary system of 8 1/2" section (average RPM interval)	40.000	46.00	35.000	35.000
Rig rotary system of 5 7/8" section (average RPM interval)			37.500	
Consumed cement (MT per bulk)	366.000	395.00	387.00	376.000
Percent CO_2	2.0000	2.000	2.0000	2.0000
Percent H_2S	0.5000	0.500	0.5000	0.5000
Water depth	60.000	67.80	67.800	60.000
Max reservoir pressure (psi)	5200.00	49600	5050.00	5200.00
Pressure gradient of gas in tubing (psi/ft)	0.1170	0.110	0.1150	0.1210
Pressure gradient of oil in tubing (psi/ft)				
Inclination of 32" section (α)	0.2900	0.280	0.2300	0.3100
Inclination of 24" section (α)	3.6000	4.200	3.5000	4.5000
Inclination of 16" section (α)	49.000	59.00	50.000	64.000
Inclination of 12 1/4" section (α)	59.000	68.00	63.000	75.000
Inclination of 8 1/2" section (α)	70.000	80.00	85.000	83.000
Inclination of 5 7/8" section (α)			88.000	
Azimuth of 32" section (α)	273.700	255.70	208.190	232.500
Azimuth of 24" section (α)	60.5600	56.700	315.050	293.700
Azimuth of 16" section (α)	183.400	316.50	314.390	345.000
Azimuth of 12 1/4" section (α)	244.600	304.50	313.900	311.900
Azimuth of 8 1/2" section (α)	11.4600	287.50	312.000	287.300
Azimuth of 5 7/8" section (α)			307.500	
Cost (million dollars)	41.4000	41.900	44	45.200

Table 4: Distinction and weight of features

Name of feature	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8	Well 9	Well 10	Well 11	Feature weight
Drilling position	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.032
Well type	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.016
Exploratory material	0.00	1.00	0.00	0.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.032
Cementing	1.00	1.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	1.00	0.016
Downhole completion equipment	0.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.024
Logging	0.00	1.00	0.50	0.00	0.50	0.50	0.00	0.00	0.00	0.50	0.00	0.024
Cementing additive (cement class)	0.50	0.50	0.00	0.67	0.50	0.75	0.67	0.67	0.00	0.50	0.67	0.015
Stimulation services	0.50	0.00	0.50	0.50	0.50	0.50	0.00	0.50	0.00	0.00	0.50	0.016
Stimulation material	0.67	0.67	0.67	0.67	0.67	0.33	0.67	0.67	0.67	0.00	0.67	0.008
Perforation	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.000
Well profile	0.00	0.50	1.00	1.00	0.50	1.00	0.50	1.00	1.00	1.00	1.00	0.020
Downhole completion services	0.33	0.33	0.33	0.33	0.33	0.67	0.67	0.33	0.33	1.00	0.33	0.011
Bits	0.50	0.50	0.50	0.50	0.50	0.40	0.75	0.50	0.50	0.75	0.50	0.003
MUD	0.33	0.33	0.67	0.33	0.67	0.67	0.67	0.67	1.00	0.33	1.00	0.035
Special tools	0.50	0.00	0.67	0.33	0.67	0.50	0.00	0.33	0.50	0.50	0.00	0.032
Transport	0.60	0.60	0.60	0.60	0.40	0.40	0.60	0.60	0.60	0.60	0.40	0.013
Rig drow works (horse power)	0.50	1.00	0.50	0.00	0.50	1.00	0.50	0.50	0.00	0.00	0.50	0.004
Rig mud pump type	1.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.032
Casing/tubing running hole section	0.20	0.20	0.10	0.10	0.20	0.10	0.20	0.10	0.20	0.10	0.20	0.004
Producing reservoir	0.00	0.00	0.25	0.00	0.50	0.00	0.50	0.00	0.25	0.00	0.00	0.012
Asymmetric binary[1]	0.25	0.25	0.60	0.67	0.25	0.57	0.50	0.25	0.50	0.00	0.25	0.011
Drilling length (m)	0.41	0.80	0.14	0.13	0.63	0.14	0.18	0.86	0.05	0.18	0.11	0.020
Operation time (h)	0.78	0.61	0.22	0.11	0.11	0.22	0.33	0.89	0.63	0.72	0.56	0.007
Non productive time (h)	0.30	0.23	0.63	0.09	0.18	0.48	0.05	0.21	0.37	0.10	0.09	0.014

Table 4: (Continue)

Name of feature	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8	Well 9	Well1 0	Well 11	Feature weight
Average GPM interval of 32" section	0.73	0.73	0.27	0.73	0.00	0.09	0.22	0.78	0.18	0.31	0.45	0.004
Average GPM interval of 24" section	0.00	0.00	0.44	0.11	0.56	0.00	0.22	0.44	0.11	0.44	0.44	0.017
Average GPM interval of 16" section	0.43	0.29	0.43	0.57	0.14	0.29	0.43	0.14	0.14	0.43	0.14	0.015
Average GPM interval of 12 1/4" section	0.80	0.80	0.50	0.40	0.20	0.40	0.10	0.60	0.70	0.10	0.10	0.029
Average GPM interval of 8 1/2" section	0.09	0.30	0.39	0.65	0.13	0.35	0.22	0.52	0.39	0.17	0.22	0.008
Average GPM interval of 5 7/8" section	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.000
Average pressure interval of 32" section	0.50	0.50	0.50	0.43	0.25	0.50	0.50	0.00	0.25	0.50	0.20	0.015
Average pressure interval of 24" section	0.45	0.20	0.30	0.70	0.08	0.20	0.05	0.25	0.05	0.70	0.05	0.020
Average pressure interval of 16" section	0.71	0.71	0.56	0.48	0.08	0.24	0.44	0.29	0.32	0.06	0.40	0.001
Average pressure interval of 12 1/4" section	0.69	0.92	0.46	0.69	0.15	0.08	0.23	0.08	0.38	0.82	0.86	0.033
Average pressure interval of 8 1/2" section	0.38	0.63	0.25	0.56	0.25	0.19	0.19	0.29	0.38	0.44	0.00	0.028
Average pressure interval of 5 7/8" section	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.000
Rig rotary system of 32" section (average RPM interval)	0.29	0.00	0.17	0.57	0.57	0.71	0.43	0.11	0.29	0.29	0.63	0.023
Rig rotary system of 24" section (average RPM interval)	0.33	0.67	0.50	1.00	0.50	0.03	0.40	0.53	0.50	0.43	0.57	0.006
Rig rotary system of 16" section (average RPM interval)	0.45	0.00	0.91	0.45	1.00	0.27	0.45	0.45	0.91	0.68	0.64	0.006
Rig rotary system of 12 1/4" section (average RPM interval)	0.03	0.31	0.14	0.11	0.60	0.26	0.66	0.89	0.31	0.23	0.31	0.003
Rig rotary system of 8 1/2" section (average RPM interval)	0.21	0.08	0.42	0.63	0.13	0.79	0.33	0.21	0.42	0.92	0.00	0.033
Rig rotary system of 5 7/8" section (average RPM interval)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.000
Consumed cement (MT per Bulk)	0.18	0.55	0.03	0.10	0.07	0.93	0.55	0.46	0.07	0.14	0.08	0.018
Percent CO_2	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.024
Percent H_2S	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.024
Water depth	0.12	0.88	0.12	0.00	0.88	0.00	0.12	0.88	0.00	0.12	0.12	0.016
Max reservoir pressure (psi)	0.36	0.64	0.45	0.82	0.18	0.18	0.09	0.32	0.00	0.22	0.14	0.015
Pressure gradient of gas in tubing (psi/ft)	0.50	1.00	0.81	1.00	1.00	1.00	0.06	1.00	0.25	0.69	0.37	0.028
Pressure gradient of oil in tubing (psi/ft)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.000
Inclination of 32" section (α)	0.55	0.05	0.35	0.45	0.15	0.20	0.40	0.10	0.10	0.15	0.40	0.012
Inclination of 24" section (α)	0.53	0.00	0.33	0.33	0.07	0.47	0.53	0.20	0.60	0.20	0.67	0.027
Inclination of 16" section (α)	0.42	0.21	0.37	0.00	0.21	0.74	0.74	0.21	0.79	0.26	0.74	0.027
Inclination of 12 1/4" section (α)	0.30	0.22	0.22	0.00	0.09	0.65	0.52	0.30	0.70	0.30	0.52	0.015
Inclination of 8 1/2" section (α)	0.32	0.16	0.05	0.16	0.05	0.53	0.47	0.32	0.68	0.16	0.11	0.015
Inclination of 5 7/8" section (α)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.000
Azimuth of 32" section (α)	0.13	0.04	0.63	0.42	0.12	0.51	0.30	0.12	0.37	0.21	0.22	0.005
Azimuth of 24" section (α)	0.90	0.55	0.92	0.73	0.16	0.17	0.55	0.08	0.74	0.75	0.07	0.039
Azimuth of 16" section (α)	0.43	0.70	1.00	0.79	0.27	0.28	0.73	0.04	0.49	0.09	0.09	0.031
Azimuth of 12 1/4" section (α)	0.85	0.65	0.99	0.78	0.16	0.17	0.66	0.04	0.22	0.02	0.01	0.036
Azimuth of 8 1/2" section (α)	0.38	0.29	0.82	0.72	0.25	0.19	0.60	0.15	0.85	0.00	0.08	0.028
Azimuth of 5 7/8" section (α)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.000

[1]: Because we can't define the distinction of feature for each asymmetric binary feature, we consider all the features together and a distinction of feature is defined for all the asymmetric binary features

value (which generally put) may affect some features which need to calculate the maximum and minimum.

Finally, we used the performance indicators MAEE and NMSE for validation of the model. Indicators MAEE and NMSE are obtained as 7.63 and 2.09%, respectively. Moreover, in the proposed CBR model in spite of limited data, the error of method was very low. Therefore the proposed CBR model accuracy is high and the model is useful.

In future researches, the K-Modes Clustering algorithms could be used in similarity measurement and retrieval of CBR. The three parameters time, cost and resource could be estimated simultaneously because now in project management, these three parameters should be simultaneously considered.

ACKNOWLEDGMENT

REFERENCES

Adeli, H. and M. Wu, 1998. Regularization neural network for construction cost estimation. J. Constr. Eng. Manage., 124(1): 18-24.

An, S.H., G.H. Kim and K.I. Kang, 2007. A case-based reasoning cost estimating model using experience by analytic hierarchy process. Build. Environ., 42: 2573-2579.

Boussabaine, A.H., 1996. The use of articcial neural networks in construction management: A review. Constr. Manage. Econ., 14(5): 427-36.

Bowen, P.A. and P.J. Edwards, 1985. Cost modeling and price forecasting: Practice and theory in perspective. Constr. Manage. Econ., 3: 199-215.

Bode, J., 2000. Neural networks for cost estimation: simulations and pilot application. International J. Prod. Res., 38(6): 1231-54.

Chang, P.C., J.J. Lin and W.Y. Dzan, 2010. Forecasting of manufacturing cost in mobile phone products by case-based reasoning and artificial neural network models. J. Intell. Manuf., 10.1007/s10845-010-0390-7.

Chou, J.S., 2008. Applying AHP-based CBR to estimate pavement maintenance cost. Tsinghua Sci. Technol., 13(S1): 114-120.

Creese, R.C. and L. Li, 1995. Cost estimation of timber bridge using neural networks. Cost Eng., 37(5): 17-22.

Duverlie, P. and J.M. Castelain, 1996. A Measure of the Filming Similarite'applique'e. 5e`me Seminaire Franc¸ais du Raisonnement a` partir de cas.

Duverlie, P. and J.M. Castelain, 1999. Cost estimation during design step: Parametric method versus case based reasoning method. Int. J. Adv. Manuf. Technol., 15: 895-906.

Garza, J. and K. Rouhana, 1995. Neural network versus parameter-based application. Cost Eng., 37(2): 14-8.

Hegazy, T., P. Fazio and O. Moselhi, 1994. Developing practical neural network applications using back-propagation. Microcomputer Civil Eng., 9: 145-59.

Hegazy, T. and A. Ayed, 1998. Neural network model for parametric cost estimation of highway projects. J. Constr. Eng. Manage., 124(3): 210-8.

Idrus, A., M.F. Nuruddin and M.A. Rohman., 2011. Development of project cost contingency estimation model using risk analysis and fuzzy expert system. Exp. Syst. Appl., 38: 1501-1508.

Ji C.Y., T. Hong and C.T. Hyun, 2010. CBR Revision Model for Improving Cost Prediction Accuracy in Multifamily Housing Projects. 10.1061/ (ASCE) ME.1943-5479.0000018.

Jurgense, T. and S.W. Wallace, 2000. Improving project cost estimation by taking into account managerial flexibility. Eur. J. Oper. Res., 127: 239-251.

Khosrowshahi, F. and A.P. Kaka, 1995. Estimation of Project Total Cost and Duration for Housing Projects in the U.K. Build. Environ., 31(4): 373-383.

Kim, K.J. and I. Han, 2001. Maintaining case based reasoning systems using a genetic algorithm approach. Exp. Syst. Appl., 21: 139-145.

Kim ,G.H., S.H. An and K.I .Kang, 2004. Comparison of construction cost estimating models based on regression analysis, neural networks and case-based reasoning. Build. Environ., 39: 1235-1242.

Kim, K.J. and K. Kim, 2010. Preliminary cost estimation model using case-based reasoning and genetic algorithms. J. Comput. Civil Eng., 24(6): 499-505.

Kouskoulas, V. and E. Koehn, 1974. Predesign cost estimation function for building. J. Construct. Div., 589-604.

Li, H., 1995. Neural networks for construction cost estimation. Build. Res. Inf., 23(5): 279-84.

Montazemi, A.R. and K.M. Gupta, 1997. A framework for retrieval in case-based reasoning systems. Ann. Oper. Res., 72: 51-73.

Perera, S. and I. Watson, 1998. Collaborative case-based estimating and design. Adv. Eng., 29(10): 801-808.

Raphael, B., B. Domer, S. Saitta and I.F.C. Smith, 2007. Incremental development of CBR strategies for computing project cost probabilities. Adv. Eng. Inform., 21: 311-321.

Singh, S., 1990. Cost model for reinforced concrete beam and slab structures in building. J. Constr. Eng. Manage., 116(1): 54-67.

Skitmore, R.M. and N.S. Thomas, 2003. Forecast models for actual construction time and cost. Build. Env., 38(8): 1075-1083.

Wilson, A.J., 1982. Experiments in Probabilistic Cost Modelling. In: Brandon, P.S. (Ed.), Building Cost Techniques. New Direction, E& FN Spon, London, pp: 169-180.

Yeh, I.C., 1998. Quantity estimating of building with logarithm-neuron networks. J. Constr. Eng. Manage., 124(5): 374-380.