

Hybrid case-based reasoning system by cost-sensitive neural network for classification

Saroj Kr. Biswas¹ · Manomita Chakraborty¹ · Heisnam Rohen Singh¹ ·
Debashree Devi¹ · Biswajit Purkayastha¹ · Akhil Kr. Das²

Published online: 18 August 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Case-based reasoning (CBR) is an artificial intelligent approach to learning and problem-solving, which solves a target problem by relating past similar solved problems. But it faces the challenge of weights assignment to features to measure similarity between cases. There are many methods to overcome this feature weighting problem of CBR. However, neural network's pruning is one of the powerful and useful methods to overcome this feature weighting problem, which extracts feature weights from trained neural network without losing the generality of training set by four popular mechanisms: sensitivity, activity, saliency and relevance. It is habitually assumed that the training sets used for learning are balanced. However, this hypothesis is not always true in real-world applications, and hence, the tendency is to yield classification models that are biased toward the overrepresented class. Therefore, a hybrid CBR system is proposed in this paper to overcome this problem, which adopts a cost-sensitive back-propagation neural network (BPNN) in network pruning to find feature weights. These weights are used in CBR. A single cost parameter is used by the cost-sensitive BPNN to distinguish the importance of class errors. A balanced decision boundary is generated by the cost parameter using prior information. Thus, the class imbalance problem of network pruning is overcome to improve the accuracy of the hybrid CBR. From the empirical results, it is observed that the performance of the proposed hybrid CBR

system is better than the hybrid CBR by standard neural network. The performance of the proposed hybrid system is validated with seven datasets.

Keywords Case-based reasoning · Artificial neural networks · Similarity measure · k-NN similarity measure · Data mining

1 Introduction

CBR is an intelligent technique to learning and problem-solving, which solves problems based on past experiences stored in case base and also captures new knowledge/experiences, making it immediately available for solving next problems. These experiences encode relevant features/attributes, courses of action that were taken and solutions that are ensued. This base of experience forms the memory for the CBR system. CBR is also a variety of reasoning by analogy (Aamodt and Plaza 1994; Leake 1996). Aamodt and Plaza (1994) have described CBR typically as a cyclical process comprising the four REs:

- *Retrieval* retrieves one or more similar cases from case base that can be used to solve a new problem.
- *Reuse* is responsible for proposing solution to the new problem from retrieved cases.
- *Revise* is responsible to evaluate the proposed solution. If the proposed solution is fit for the new problem, then it is possible to learn about the success; otherwise, the solution is repaired/adapted using some problem domain-specific knowledge or any other ways.
- *Retain* consists of a process of integrating the useful information about the new problem's resolution in case base, if it is a case whose solution is not readily available in case base.

Communicated by V. Loia.

✉ Saroj Kr. Biswas
bissarajkum@yahoo.com

¹ Computer Science and Engineering Department, NIT Silchar, Silchar, Assam 788010, India

² Faculty of Technology, Uttar Banga Krishi Viswavidyalaya, Cooch Behar, West Bengal 736165, India

It is observed that the core of CBR methodology is the retrieval of similar cases stored in case base. Therefore, a similarity measure is required to calculate the similarity between the stored cases and a new case, and hence, similarity measure becomes the key element in obtaining a reliable solution for new situations (Nunez et al. 2004; Buta 1994). The task of defining similarity measures for real-world problems is one of the greatest challenges of research in this area as assessing the similarity between cases is a key aspect of CBR. The k-nearest neighbor (k-NN) is one of the most popular similarity measures in CBR, which uses a distance function to find similarity between cases. However, the biggest problem with k-NN is to determine the weight of features because several studies have shown that k-NN's performance is highly sensitive to the definition of its distance function (Watson and Marir 1994; Wettschereck et al. 1997). Many k-NN variants have been proposed to reduce this sensitivity by parameterizing the distance function with feature weights (Wettschereck et al. 1997).

The k-NN considers that each query, q (new case) is represented by b features which are numeric or discrete. The similarity of q with each stored case, $c = \{c_1, c_2, c_3, \dots, c_b, o_c\}$ in a set C is calculated where c_1 to c_b are attribute values or problem description of the case c and o_c is c 's class value. The k-NN then retrieves k most similar (least distance) cases and predicts their weighted majority class or majority class only as the class of q . The distance can be calculated by Eq. (1) given below.

$$\text{Distance}(c, q) = \sqrt{\sum_{i=1}^b w_i \cdot \text{diff}(c_i, q_i)^2} \quad (1)$$

where w_i is the parameterized weight value assigned to feature i and

$$\text{diff}(c_i, q_i) = \begin{cases} |c_i - q_i| & \text{if feature } i \text{ is numeric} \\ 0 & \text{if feature } i \text{ is discrete and } c_i = q_i \\ 1 & \text{Otherwise} \end{cases} \quad (2)$$

The distance given in Eq. (1) is weighted Euclidean distance but can also be weighted absolute or city block distance¹. City block distance examines the absolute differences between co-ordinates of a pair of objects. The detail of Eqs. (1) and (2) could be found in Wettschereck et al. (1997). The concept of equal weights handicaps k-NN as it allows redundant and irrelevant features to have as much impact on distance computations as others. For the cases belonging to the same

class, some features may often have the same value, while others vary their values in most of the cases in that class. Therefore, the features will always have different degrees of impact in similarity measure. Accordingly, different feature weights should be provided to avoid incorrectness in similarity measure. If all of the features are regarded as being equally important, i.e., all the features have the same weight value, then CBR allows redundant or irrelevant features to influence the final solution as well as similarity measure. Therefore, it is very important to solve the feature weighting problem of CBR to work properly where similarity measure is k-NN.

Many methods have been proposed to sort out the feature weighting problem. For example, Daelemans et al. (1994) and Wettschereck and Dietterich (1995) used mutual information to compute coefficients on numeric attributes. Many other feature weighting methods and their analysis could be found in Wettschereck et al. (1997).

Although many feature weighting methods for k-NN have been reported for classification or prediction task, feature weighting methods which can capture generality and domain-specific knowledge together are rare. Artificial neural network (ANN) is a method which captures generality and domain-intensive knowledge to estimate the relative importance of each feature. Therefore, many network pruning tasks are attempted and implemented (Shin and Park 1999; Shin et al. 2000; Im and Park 2007; Park et al. 2004; Ha 2008; Zeng and Martinez 2004; Yang and Jin 2010; Peng and Zhuang 2007; Sarwar et al. 2010; Park et al. 2006) to find feature weights. Sensitivity, activity saliency and relevance mechanisms are used in Shin and Park (1999), Shin et al. (2000), Im and Park (2007) and Park et al. (2004) to find symbolic feature weights from back-propagation neural network (BPNN). They have assumed that the training sets used for learning are balanced in those works. However, this hypothesis is not habitually true in real-world applications. Traditional learning algorithms applied to complex and highly imbalanced training sets do not give satisfactory results when distinguishing between examples of classes. The tendency is to yield classification model that is biased toward the overrepresented (majority) class (Braga and Castro 2013). This paper attempts to overcome the class imbalance problem for sensitivity, activity, saliency and relevance mechanisms, which improves the performance of these feature weighting mechanisms.

Many solutions are proposed to resolve the class imbalance problem both at the data and at the algorithmic levels. The common practice for dealing with imbalanced data at the data level includes different sampling techniques like over-sampling the minority class, or under sampling the majority class, until the classes become balanced. However, while implementing these sampling techniques, some known drawbacks are associated with them. The implementation of under sampling tends to discard the potentially useful data (Gan-

¹ The distance is always greater than or equal to zero. The measurement would be zero for identical points and high for points that show little similarity. It is also known as Manhattan distance or Boxcar distance or absolute value distance

ganwar 2012). Oversampling the minority class leads to increase the number of training examples and hence-over fits the dataset and also increases the learning time as the dataset is rebalanced artificially by using sampling methods.

The solutions at the algorithmic level include various mechanisms. Adjusting the costs of various classes is one of the significant solutions at the algorithmic level to reduce the class imbalance. This is what the core concept of cost-sensitive learning approach (Ganganwar 2012). A cost-sensitive learning approach takes cost, such as misclassification cost, into consideration during model construction and supports to build a classifier that has the minimum cost (Nguyen et al. 2009). The cost-sensitive learning assumes higher misclassification cost with samples in the minority class and hence affords to minimize high cost errors (López et al. 2012). Consequently, this approach reduces the bias toward the majority class in an efficient way while keeping the original dataset size intact. Therefore, the paper proposes a hybrid CBR model, which overcomes the class imbalance problem for sensitivity, activity, saliency and relevance feature weighting mechanisms by cost-sensitive BPNN (Braga and Castro 2013).

The paper is organized as follows: Sect. 2 depicts cost-sensitive BPNN, while Sect. 3 describes weight learning mechanisms from trained neural network and outlines some previous works. Section 4 describes the framework of the proposed hybrid CBR system. Section 5 discusses the results of seven datasets, and finally some conclusions are drawn in Sect. 6.

2 Cost-sensitive BPNN

The dataset, D , of the binary classification task can be written as $D = \{D_1 \cup D_2\}$, where $D_1 = \{\mathbf{x}(p), y(p) | p = 1, \dots, N_1\}$; the vector $\mathbf{x}(p)$ is the p th case of the positive class and $y(p) = +1, \forall \mathbf{x}(p) \in D_1$. The number of positive cases is represented by N_1 . Analogous definitions hold for D_2 , for negative class. In order to achieve solutions that are sensitive to the importance of each class, a cost function for the parameter estimation of BPNN is used. The cost function is defined in Eq. (3) as the weighted sum of functionals J_1 and J_2 which correspond to the sum of squared errors for D_1 and D_2 , respectively

$$J = \lambda J_1 + (1 - \lambda) J_2 \quad (3)$$

where, J_1 and J_2 are described by the following equations:

$$J_1 = \sum_{p=1}^{N_1} e^2(p) \quad \forall \mathbf{x}(p) \in D_1 \quad (4)$$

$$J_2 = \sum_{r=1}^{N_2} e^2(r) \quad \forall \mathbf{x}(r) \in D_2 \quad (5)$$

The vector $\mathbf{x}(r)$ is the r th case of the negative class; $e(p)$ and $e(r)$ denote errors for p th and r th case of D_1 and D_2 , respectively. $N = N_1 + N_2$ and the parameter, $\lambda (0 < \lambda < 1)$ is used to weight the contribution of J_1 and J_2 in the composition of J . The parameter λ modifies the standard formulation of learning by assigning unequal costs to the errors of each class. λ influences the location of the decision boundary obtained from training, compensates the class imbalance and obtains a balanced decision boundary. Thus, if N_2 is greater than N_1 , this parameter can be used to compensate the class imbalance problem (Braga and Castro 2013).

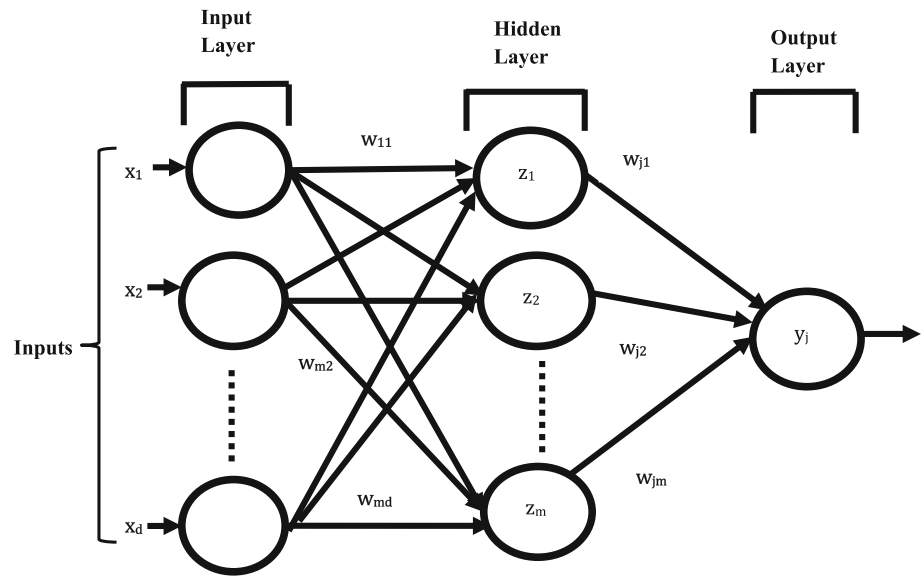
3 Weight learning by neural network

Neural networks are helpful in adjusting the connection weights among the nodes of the network due to their input–output mapping, robustness and adaptive capabilities. Therefore, ANN model can be used to resolve the problem of feature weighting. BPNN model can be used to process the case feature values by two stages: First is learning stage, which also means to build BPNN model and using this model, and samples are taught to get the desired result. Second is the exchanging weight stage, which transfers the nodes value into the desired result to the case feature values. Knowledge exploited by ANNs from the training dataset is stored as connected links of trained neural network nodes. Knowledge extraction from trained neural network is one of the main interests for a long time (Shin and Park 1999). Several methodologies have been attempted for using the connectionist approach (ANN) in CBR system design (Becker and Jazayeri, 1989; Thrift 1989). The work of hybrid CBR with ANN architecture is reinforced to solve complicated problem in Shin and Park (1999), which calculates a set of feature weights from a trained BPNN that plays the core role in finding the most similar cases from case base for prediction or classification.

The four feature weighting mechanisms are proposed: sensitivity, activity, saliency and relevance to obtain a vector of feature weights $\{w_1, w_2, \dots, w_d\}$ from a trained BPNN in Shin and Park (1999), Shin et al. (2000), Im and Park (2007) and Park et al. (2004), where d is the number of input features. Each of the feature weighting mechanisms is briefly described below.

Consider a neural network given in Fig. 1 with a single hidden layer, which consists of d inputs (x_i , where $i = 1, 2, \dots, d$), m number of hidden neurons (z_h , where $h = 1, 2, \dots, m$) and one output, y_j . Here w_{hi} denotes the

Fig. 1 A neural network



weight connecting from x_i to z_h and w_{jh} is the connecting weight between z_h to y_j .

- (i) *Sensitivity* The sensitivity of an input node is calculated by removing the input node from the trained neural network. Sensitivity measure of an input feature is the difference in prediction value between when the feature is removed and when it is left in place. The sensitivity, S_i , of an input feature, i , is defined by Eq. (6):

$$S_i = \frac{\left[\sum_L \frac{|P^o - P^i|}{P^o} \right]}{n} \quad (6)$$

where P^o is the normal prediction value for each training instance after training, P^i is the modified prediction value when the input feature i is removed, L is the set of training data and n is the number of training data.

- (ii) *Activity* The activity of a node is measured by the variance of the activation level in training data. The activation of a hidden node, z_h , is defined by Eq. (7):

$$A_h = [w_{jh}]^2 \cdot \text{var} \left[g \left(\sum_{i=1}^d w_{hi} x_i \right) \right] \quad (7)$$

where $\text{var}()$ is the variance function, w_{jh} is the weight between output node j and hidden node h , w_{hi} is the weight between hidden node h and input node i . The activity of an input node, x_i , is defined by Eq. (8):

$$A_i = \sum_{h=1}^m [(w_{hi})^2 \cdot A_h] \quad (8)$$

- (iii) *Saliency* The saliency of a weight is measured by estimating the second derivative of the error with respect to the weight. The saliency of an input node, i , is defined by Eq. (9):

$$\text{Saliency}_i = \sum_{h=1}^m [(w_{hi})^2 \cdot (w_{jh})^2] \quad (9)$$

- (iv) *Relevance* The variance of weights into a node is a good predictor of the node's relevance, and the relevance of a node is a good predictor of the increase in error expected when the node's largest weight is deleted. The relevance of a hidden node, z_h is defined by Eq. (10):

$$R_h = (w_{jh})^2 \cdot \text{var}(w_{hi}) \quad (10)$$

And the overall relevance of an input node, x_i , is defined by Eq. (11):

$$R_i = \sum_{h=1}^m (w_{hi})^2 \cdot R \quad (11)$$

Shin and Park (1999) have combined MBR and ANN by integrating the calculated feature weights. This hybrid method is also compared with some other machine learning methods. Shin et al. (2000) have used the same concept of weight extraction as Shin and Park (1999) and have built the same hybrid CBR system except decision-making process. The developed hybrid system is validated by high dimensional and dynamic systems like odd parity problem, decision making in sinusoidal task, Wisconsin diagnostic breast cancer

(WDBC) problem, credit card application, classification of sonar signals and auto-mpg regression problem. Im and Park (2007) have also used the same concept of weight extraction and building of hybrid CBR system. They have added the concept of value difference metric (VDM) when feature attributes are non-numeric and the hybrid system is applied to build an expert system for personalization. Ha (2008) has used the same concept of hybrid CBR and also applied in same application as Im and Park (2007), but has reported changes in performance. Sensitivity, activity, saliency and relevance feature weighting mechanisms are used in all those works.

4 The proposed hybrid CBR model

The cost-sensitive BPNN Braga and Castro (2013) is integrated into CBR during the retrieval process that retrieves a set of k -nearest neighbors to a target case. The cost-sensitive BPNN is initially trained, which takes care about class imbalance problem and finds feature weights. The symbolic feature weights are calculated from the trained BPNN using sensitivity, activity, saliency, and relevance mechanisms. When a query (new/target) case is encountered in case retrieval module, k similar cases are retrieved from the case base using k -NN similarity measure where pre-calculated symbolic feature weights are used to measure similarity between cases. Then, case reuse and revision module are used to propose a solution, which is sent to the query (new/target) case that demands a solution. The framework of the proposed hybrid CBR model is shown in Fig. 2. The discussion of the stages of the hybrid CBR model is given below.

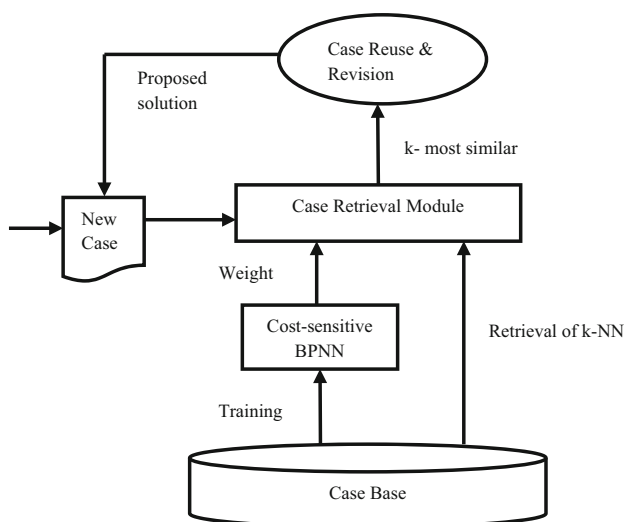


Fig. 2 Framework of hybrid CBR

4.1 STAGE 1: case representation

A case is a piece of knowledge in a particular context that represents an experience. A case stored in case base consists of problem specification and known solution. Case representation is very important in CBR because proper case representation enhances the acceptance of proposed solution. Each case is defined by problem specification having assignment of values $a = (a_1, a_2, a_3, \dots, a_g)$ to a set of features $f = (f_1, \dots, f_g)$, and solution as one of s possible classes, o_1, \dots, o_s to the class label o . In binary case classification tasks, the applicable classes to the class label are $o_1 = 0$ and $o_2 = 1$. The learning of a classifier is inherently determined by the feature values of the cases. In this paper, each case stored in the case base consists of the values of the feature vector representing the problem specification, together with its associated solution or class in the form of '1' or '0' representing either of two classes. Furthermore, a query (new/target) case refers to a problem requiring solution, which is obtained through prediction. Suppose t is the number of attributes that form the problem description of a case. Then, $Q(q_1, q_2, \dots, q_t)$ is the query case and $C(c_1, c_2, \dots, c_t, c_{(t+1)})$ is any stored case, where $c_{(t+1)}$ is solution description of the stored case.

4.2 STAGE 2: case selection/retrieval of k-NN

Retrieval of most similar cases is very critical to the success of a CBR system. The top most similar cases are selected by k -NN, which are presented in the system to measure the performance of the hybrid model. Selection of cases is done by performing the following steps:

- RETRIEVE multi-attribute-based information of past cases by cost-sensitive BPNN in terms of symbolic weights.
- MATCH past cases with the query (new) case by Eq. (1) where weights are calculated from the trained cost-sensitive BPNN by feature weighting mechanisms, i.e., sensitivity, activity, saliency and relevance.
- COMPARE the past cases with each other by dissimilarity score.
- SELECT the past cases having least values of dissimilarity score.

4.3 STAGE 3: case reuse and revision

Top k most retrieved similar cases are taken into consideration to produce the predicted result by observing the frequency of occurrence of similar solutions, i.e., based on the majority of voting of top most similar cases. The top most similar case is initially reused as the proposed solution for the query (new) case, and then, the performance of the model is

measured. Subsequently, the measured performance is compared with the performances of the model when number of similar cases (k) is changed ($k = 2, 3, \dots, 20$). If k is too large, prediction result would taper off because of the inclusion of an increasing number of decreasingly similar cases. Therefore, the value of k is restricted up to 20. The value of k , which produces highest accuracy of the proposed hybrid CBR model, can be adopted as optimal k . Therefore, when a query (new) case arrives, the solution of the query (new) case is determined by majority of voting of the optimal k nearest neighbors.

4.4 STAGE 4: retention

Retention is not adopted based on the requirement of the proposed model but that can be adopted by setting a threshold value in similarity score. If the similarity score is less than the threshold value, then the query (new) case may be considered as totally a new case in the case base. The encountered new case with solution given by domain expert can be retained in the case base for future use.

5 Results and discussion

5.1 Dataset collection

The experimentation of the proposed model is done by MATLAB 7.12.0.635 (R2011a) in Windows environment. The following binary classification tasks which are obtained from the UCI machine learning repository are incorporated into the paper. In addition, Swine flu data are gathered from physicians in several hospitals and from the Internet. With the aid of internet and consulting local medical practitioners, the symptoms of Swine flu are categorized (indexed) and ranged, and accordingly the cases are generated and validated by medical institutions. All the experiments of the paper use three experimental setups to find out the accuracy of the proposed hybrid CBR model: (i) 80% cases of a case base as training set and 20% cases of the case base as test set, which is termed as 80:20 split, (ii) 70% cases of a case base as training set and 30% cases of the case base as test set, which is termed as 70:30 split, and (iii) five-folded validation as five independent runs are executed to validate the comparison results by using Wilcoxon rank-sum nonparametric statistical test. The concise description of the datasets is given below:

5.1.1 Blood transfusion

There are total 748 instances and 5 attributes. One of the 5 attributes is used as solution (decision). The classification task involves in determining whether someone has donated blood or not. 1 stands for donating blood and 0 stands for

not donating blood. The attributes are given below with brief description:

- Recency (R)—months since last donation
- Frequency (F)—total number of donation
- Monetary (M)—total blood donated in c.c.
- Time (T)—months since first donation and
- A binary variable representing whether he/she donated blood in March 2007.

5.1.2 Diabetes mellitus

The database contains 8 attributes and 768 cases. Prediction task involves in determining either presence or absence of diabetes. The attributes are given below:

- Number of times pregnant
- Plasma glucose concentration a 2 h in an oral glucose tolerance test
- Diastolic blood pressure (mm Hg)
- Triceps skin fold thickness (mm)
- 2-h serum insulin (μ U/ml)
- Body mass index (weight in kg/(height in m^2))
- Diabetes pedigree function
- Age (years)

5.1.3 Heart disease

The database contains 13 attributes and 270 cases. Prediction task involves in determining either presence or absence of heart disease. The attributes are given below:

- age
- sex
- chest pain type (4 values)
- resting blood pressure
- serum cholesterol in mg/dl
- fasting blood sugar > 120 mg/dl
- resting electrocardiographic results (values 0, 1, 2)
- maximum heart rate achieved
- exercise-induced angina
- old peak = ST depression induced by exercise relative to rest
- the slope of the peak exercise ST segment
- number of major vessels (0–3) colored by fluoroscopy
- thal: 3 = normal; 6 = fixed defect; 7 = reversible defect

5.1.4 Liver disorder

There are total 345 cases and 7 attributes. One of the attributes is used to determine the classification. The classification problem involves in determining either presence or absence of liver disorder. The attributes are given below:

Table 1 Indexing of attributes based on their ranges

Attribute	Ranges and their index values			
Fast breathing	None	Mild	Moderate	Severe
	0	1	2	3
Sore throat	None	Mild	Moderate	Severe
	0	1	2	3
Temperature	≤98.4	98.5–100.4	100.5–102.4	> 102.4
	0	1	2	3
Runny nose or stuffy nose	None	Mild	Moderate	Severe
	0	1	2	3
Cough	None	Mild	Moderate	Severe
	0	1	2	3
Fatigue	None	Mild	Moderate	Severe
	0	1	2	3
Body aches	None	Mild	Moderate	Severe
	0	1	2	3
Headache	None	Mild	Moderate	Severe
	0	1	2	3
Chills	None	Mild	Moderate	Severe
	0	1	2	3
Nausea	None	Mild	Moderate	Severe
	0	1	2	3

- mcv: mean corpuscular volume
- alkphos: alkaline phosphatase
- sgpt: alanine aminotransferase
- sgot :aspartate aminotransferase
- gammagt: gamma-glutamyl transpeptidase
- drinks: number of half-pint equivalents of alcoholic beverages drunk per day

5.1.5 Swine flu

Swine flu prediction is a problem that involves in determining whether an individual has Swine flu or not. Ten prominent attributes are used as the problem description of a Swine flu case, and class of the case is a decision that is either yes (1) or no (0). This database contains 250 cases. Each attribute of a case is indexed, as given in Table 1 as per their ranges.

5.1.6 Sonar Signals

The database contains 60 attributes and 208 cases. Classification task involves in discriminating between sonar signals bounced off a metal cylinder at various angles and under various conditions as ‘Mine’ and those bounced off a roughly cylindrical rock under similar conditions as ‘Rock.’

5.1.7 Ionosphere

This database contains 34 attributes and 351 cases and describes about the radar returns from the ionosphere. The classification task involves in determining either good or bad where ‘good’ radar returns are those showing evidence of some type of structure in the ionosphere and ‘bad’ returns are those that do not; their signals pass through the ionosphere.

5.2 Results

Tables 2, 3, 4, 5, 6, 7 and 8 show the experimental results of heart, diabetes, blood transfusion, liver disorder, Swine flu, sonar and ionosphere datasets, respectively, for two splits of each dataset: 80:20 and 70:30. The experimental results of Tables 2, 3, 4, 5, 6, 7 and 8 show the performance of the hybrid CBR model with standard BPNN and with cost-sensitive BPNN for sensitivity, activity, saliency and relevance. The performance (accuracy) is shown in percentage in all the tables. The value of k represents the number of similar case(s) taken into consideration to find the performance of the hybrid model. The value of k out of 20, which produces the highest accuracy, is considered as optimal k and is indicated by boldface for each feature weighting mechanism in Tables 2, 3, 4, 5, 6, 7 and 8. The highest accuracy is assumed as system performance. It is observed that the performance of the hybrid CBR model with cost-sensitive BPNN is better than the hybrid CBR model with standard BPNN in all the datasets for both splits. The proposed model achieves 90.74 %, 87.03 %, 85.18 % and, 74.07 % accuracy for sensitivity, activity, saliency and relevance, respectively, in heart disease dataset for 80:20 split where the hybrid CBR with standard BPNN achieves 85.18 %, 83.33 %, 83.33 % and, 68.51 % accuracy for the same. It is observed from Tables 2, 3, 4, 5, 6, 7 and 8 that the similar comparison and improvement hold true in all datasets for both the splits. The performance of the hybrid CBR model is represented by red and black colors for 80:20 and 70:30 splits, respectively, in Tables 2, 3, 4, 5, 6, 7 and 8.

Two statistical measures are also used: mean and mean absolute percentage error (MAPE) to show the improvement of the proposed model for both the splits. It is observed from Table 2 that mean performances for sensitivity, activity, saliency and relevance of the proposed model are 83.98 %, 83.15 %, 82.40 % and 69.26 %, respectively, in heart disease dataset for the 80:20 split, whereas mean performances for sensitivity, activity, saliency and relevance of the hybrid CBR with standard BPNN are 81.39 %, 81.39 %, 81.48 % and 57.87 %, respectively, for the same. Similar improvement is also observed for 70:30 split. For all other datasets, similar improvement by the proposed model is also observed as given in Tables 2, 3, 4, 5, 6, 7 and 8. It is also observed from Table 2 that MAPEs for sensitivity, activity, saliency

Table 2 (color online) Heart

Value s of <i>k</i>	Sensitivity				Activity				Saliency				Relevance			
	Standard BPNN		Cost- sensitive BPNN		Standard BPNN		Cost- sensitive BPNN		Standard BPNN		Cost- sensitive BPNN		Standard BPNN		Cost- sensitive BPNN	
	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30
1	79.62	74.1	90.74	71.6	81.48	74.1	83.33	72.8	81.48	74.1	83.33	72.8	42.59	63.0	66.66	69.1
2	79.62	74.1	90.74	71.6	81.48	74.1	83.33	72.8	81.48	74.1	83.33	72.8	42.59	63.0	66.66	69.1
3	85.18	79.0	87.03	75.3	79.62	76.5	83.33	72.8	81.48	76.5	85.18	74.1	51.85	60.5	62.96	69.1
4	81.48	80.2	87.03	77.8	79.62	77.8	85.18	76.5	79.62	79.0	83.33	76.5	50	63.0	61.11	69.1
5	83.33	80.2	83.33	79.0	79.62	81.5	81.48	77.8	83.33	79.0	81.48	77.8	53.70	64.2	68.51	66.7
6	79.62	77.8	83.33	77.8	77.77	80.2	81.48	77.8	79.62	76.5	81.48	76.5	53.70	67.9	66.66	67.9
7	81.48	77.8	81.48	79.0	81.48	77.8	81.48	77.8	79.62	77.8	81.48	77.8	62.96	67.9	66.66	66.7
8	79.62	76.5	83.33	79.0	81.48	77.8	83.33	77.8	81.48	77.8	83.33	77.8	62.96	65.4	68.51	69.1
9	83.33	77.8	81.48	81.5	81.48	77.8	83.33	79.0	79.62	77.8	83.33	79.0	68.51	64.2	68.51	66.7
10	81.48	77.8	83.33	79.0	79.62	77.8	83.33	77.8	79.62	79.0	81.48	77.8	61.11	63.0	70.37	66.7
11	81.48	80.2	83.33	77.8	81.48	80.2	81.48	79.0	81.48	79.0	77.77	80.2	61.11	65.4	70.37	66.7
12	81.48	77.8	83.33	79.0	83.33	80.2	83.33	80.2	81.48	77.8	79.62	80.2	57.40	61.7	72.22	67.9
13	83.33	79.0	83.33	79.0	83.33	80.2	81.48	80.2	81.48	80.2	81.48	80.2	59.25	65.4	72.22	69.1
14	81.48	77.8	83.33	79.0	81.48	81.5	85.18	80.2	83.33	82.7	83.33	80.2	59.25	63.0	70.37	70.4
15	81.48	77.8	81.48	79.0	83.33	81.5	87.03	79.0	83.33	81.5	83.33	79.0	62.96	67.9	74.07	69.1
16	81.48	76.5	83.33	77.8	81.48	81.5	83.33	79.0	81.48	81.5	81.48	79.0	62.96	64.2	70.37	71.6
17	81.48	77.8	81.48	77.8	83.33	81.5	85.18	80.2	83.33	80.2	83.33	80.2	64.81	65.4	72.22	69.1
18	77.77	76.5	83.33	77.8	81.48	82.7	81.48	81.5	81.48	81.5	81.48	81.5	61.11	63.0	72.22	70.4
19	81.48	77.8	81.48	79.0	81.48	80.2	83.33	84.0	81.48	81.5	83.33	84.0	61.11	64.2	72.22	69.1
20	81.48	77.8	83.33	80.2	83.33	81.5	81.48	82.7	83.33	81.5	85.18	82.7	57.40	64.2	72.22	70.4
Mean	81.3	77.7	83.9	77.9	81.3	79.3	83.1	78.4	81.4	78.9	82.4	78.5	57.8	64.3	69.2	68.7
MAPE	23	29	19	28	23	26	20	28	23	27	21	28	76	56	45	46

Table 3 (color online) Diabetes

Values of <i>k</i>	Sensitivity				Activity				Saliency				Relevance			
	Standard BPNN		Cost- sensitive BPNN		Standard BPNN		Cost- sensitive BPNN		Standard BPNN		Cost- sensitive BPNN		Standard BPNN		Cost- sensitive BPNN	
	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30
1	59.74	65.7	58.44	66.1	65.58	65.2	64.93	65.2	65.58	65.7	64.28	65.2	57.79	61.3	61.03	64.3
2	59.74	65.7	58.44	66.1	65.58	65.2	64.93	65.2	65.58	65.7	64.28	65.2	57.79	61.3	61.03	64.3
3	62.98	71.7	61.68	70.9	67.53	72.2	66.88	72.6	65.58	72.6	67.53	72.2	58.44	59.1	59.09	61.3
4	62.98	71.7	62.98	70.0	67.53	71.7	67.53	71.7	67.53	71.7	66.23	70.9	59.74	61.7	59.09	64.3
5	63.63	70.0	66.23	69.6	70.77	71.3	67.53	71.7	68.18	72.2	66.23	72.6	62.33	65.2	62.33	64.8
6	63.63	69.6	62.98	70.4	69.48	73.0	67.53	73.5	68.18	72.2	67.53	73.0	62.98	63.5	64.28	67.4
7	65.58	66.5	65.58	71.3	70.77	72.2	68.18	71.7	68.83	72.6	69.48	72.2	62.98	64.3	61.68	67.0
8	65.58	67.4	62.98	72.6	69.48	70.9	70.77	72.2	68.18	71.3	69.48	72.2	62.33	66.1	61.03	65.2
9	61.68	66.1	64.28	70.9	70.77	70.9	71.42	71.7	71.42	70.9	70.77	70.9	62.98	64.3	61.68	67.0
10	63.63	66.5	61.68	71.7	69.48	71.3	72.72	70.0	70.77	70.9	70.77	70.0	62.33	66.5	61.03	67.8
11	61.03	67.4	60.38	71.7	70.12	70.4	70.12	70.0	72.07	69.1	70.77	69.1	61.68	64.3	60.38	63.9
12	63.63	65.7	61.68	70.9	69.48	69.6	73.37	69.6	71.42	70.4	74.02	69.1	61.68	66.1	58.44	64.3
13	62.98	65.2	60.38	70.9	69.48	69.6	70.77	70.4	70.77	70.9	70.12	70.4	61.03	64.8	57.14	65.7
14	62.98	66.1	62.33	71.7	66.88	70.0	70.12	69.1	70.12	70.0	68.83	70.0	59.09	64.8	55.19	66.1
15	63.63	67.0	62.98	71.7	68.18	70.0	66.23	70.0	69.48	70.0	65.58	69.6	60.38	65.7	58.44	67.0
16	62.98	66.5	62.33	73.5	65.58	68.7	65.58	69.1	68.83	69.1	65.58	68.3	61.68	67.4	57.79	65.7
17	63.63	67.0	63.63	70.9	67.53	70.9	64.28	71.3	68.18	70.9	65.58	70.4	62.98	64.8	55.84	67.4
18	64.28	67.4	63.63	70.0	66.23	70.4	64.28	71.3	67.53	71.7	64.93	70.4	62.98	63.9	57.14	65.7
19	63.63	67.8	62.98	68.3	66.23	71.7	64.28	71.3	67.53	71.7	64.93	71.3	62.98	64.8	57.79	67.8
20	64.28	68.3	62.98	70.0	65.58	71.3	66.88	71.3	67.53	72.2	66.88	72.2	62.33	65.2	59.09	66.5
Mean	63.1	67.4	62.4	70.4	68.1	70.3	67.9	70.4	68.6	70.5	67.6	70.2	61.3	64.2	59.4	65.6
MAPE	59	48	60	42	47	42	47	42	46	42	48	42	63	56	68	52

and relevance of the proposed model are 19 %, 20 %, 21 % and, 45 %, respectively, in heart disease dataset for the 80:20 split, whereas MAPEs for sensitivity, activity, saliency and relevance of the hybrid CBR with standard BPNN are 23 %,

23 %, 23 % and, 76 %, respectively, for the same. Similar improvement is also observed for 70:30 split. For all other datasets, similar improvement by the proposed model is also observed as given in Tables 2, 3, 4, 5, 6, 7 and 8. Hence, the

Table 4 (color online) Blood transfusion

Values of k	Sensitivity				Activity				Saliency				Relevance			
	Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN	
	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30
1	78	55.80	76.66	52.67	79.3	55.4	80	55.3	79.3	55.8	80	55.4	72.66	53.1	77.33	54.5
2	72.66	55.80	72	52.67	74.6	55.4	75.3	55.3	72.6	55.8	76	55.4	65.33	53.1	71.33	54.5
3	88	54.46	88	54.91	89	53.6	90	52.6	88.6	55.4	90	52.7	78	54.9	80	56.3
4	86	57.14	88	55.80	85.3	55.4	87.3	54.4	86	56.3	88	54.0	74.66	54.0	73.33	56.3
5	89	61.16	89.33	59.82	88.6	62.1	88.6	62.9	89.3	62.5	88.66	62.9	81.33	60.3	82	62.9
6	88.66	62.95	87.33	59.37	87.3	62.5	87.3	62.9	87.3	62.5	86.66	62.5	77.33	62.1	78.66	61.6
7	89	60.71	89.33	61.16	89.3	59.8	89.3	63.8	89.3	63.8	89.33	63.8	84	63.4	86	62.1
8	88	63.84	89.33	62.94	89.3	63.4	88.6	61.6	88	62.1	88.66	61.6	79.33	62.5	80.66	63.8
9	89.33	62.05	89.33	62.05	89.3	61.6	89.3	62.9	89.3	65.6	89.33	62.9	85.33	62.5	88	64.7
10	89.33	64.73	88.66	61.60	89.3	61.6	89.3	64.7	89.3	66.5	89.33	66.1	84.66	62.9	84.66	64.3
11	89.33	61.61	90	62.5	89.3	61.2	90.6	68.3	89.3	65.6	90.66	67.4	87.33	63.4	88	63.4
12	89.33	62.95	90	62.94	89	61.2	90	68.3	89	64.7	90	68.3	85.33	62.5	88	62.9
13	89.33	62.95	90.66	64.73	89	62.5	90	66.9	89.3	66.5	90	67.4	88	63.4	89.33	62.9
14	89.33	63.84	90.66	66.07	89	62.1	90	66.0	89	67.4	90	65.6	86	63.4	89.33	62.9
15	89.33	62.05	90.66	65.62	89.3	63.4	90.6	66.0	89.3	66.5	90.66	65.6	88.66	63.8	90	62.9
16	89.33	63.39	90.66	66.07	89.3	62.5	90.6	64.7	89.3	67.0	90.66	64.7	88.66	63.4	88.66	62.9
17	89.33	62.50	90.66	67.41	89.3	63.4	90.6	65.1	89.3	67.0	90.66	65.2	88.66	63.8	90	62.9
18	89.33	62.95	90.66	66.07	89.3	63.4	90.6	65.1	89.3	67.0	90.66	65.2	88	62.9	89.33	62.5
19	89.33	62.50	90.66	64.73	89.3	63.8	90.6	68.3	89.3	67.0	90.66	67.4	88.66	62.9	89.33	62.5
20	89.33	62.95	90.66	63.83	89.3	63.8	90.6	66.5	89.3	67.0	90.66	66.1	88.66	62.9	89.33	62.5
Mean	87.5	61.3	88.1	61.6	87.6	60.9	88.4	63.0	87.5	63.6	88.5	63.0	83.0	61.0	84.6	61.4
MAPE	15	63	14	63	14	65	13	60	14	58	13	60	21	64	19	63

Table 5 (color online) Liver disorder

Values of k	Sensitivity				Activity				Saliency				Relevance			
	Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN	
	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30
1	47.82	56.3	53.62	53.4	55.07	60.2	57.97	58.3	50.72	59.2	59.42	59.2	53.62	45.6	49.27	47.6
2	47.82	56.3	53.62	53.4	55.07	60.2	57.97	58.3	50.72	59.2	59.42	59.2	53.62	45.6	49.27	47.6
3	56.52	62.1	62.31	62.1	57.97	61.2	59.42	61.2	55.07	59.2	57.97	60.2	44.92	47.6	47.82	55.3
4	55.07	62.1	56.52	61.2	55.07	62.1	62.31	61.2	52.17	61.2	62.31	61.2	49.27	47.6	47.82	53.4
5	57.97	54.4	56.52	54.4	62.31	59.2	57.97	61.2	53.62	59.2	60.86	59.2	50.72	52.4	47.82	55.3
6	55.07	54.4	55.07	54.4	59.42	56.3	57.97	61.2	56.52	55.3	56.52	58.3	49.27	50.5	50.72	50.5
7	55.07	56.3	55.07	57.3	59.42	57.3	59.42	65.0	59.42	58.3	57.97	64.1	52.17	56.3	50.72	54.4
8	50.72	58.3	56.52	54.4	62.31	59.2	57.97	62.1	62.31	59.2	60.86	62.1	53.62	53.4	57.97	49.5
9	53.62	58.3	55.07	55.3	52.17	61.2	60.86	66.0	63.76	61.2	59.42	63.1	50.72	61.2	49.27	58.3
10	53.62	56.3	56.52	59.2	57.97	58.3	62.31	65.0	60.86	55.3	65.21	62.1	53.62	56.3	52.17	54.4
11	53.62	53.4	62.31	58.3	57.97	59.2	59.42	64.1	59.42	57.3	59.42	64.1	52.17	60.2	50.72	56.3
12	50.72	58.3	56.52	59.2	57.97	56.3	60.86	65.0	56.52	55.3	63.76	63.1	53.62	57.3	52.17	55.3
13	52.17	56.3	57.97	60.2	60.86	56.3	57.97	62.1	63.76	55.3	59.42	62.1	50.72	59.2	49.27	62.1
14	47.82	58.3	57.97	58.3	52.17	56.3	63.76	63.1	57.97	58.3	63.76	62.1	55.07	56.3	53.62	56.3
15	53.62	61.2	56.52	60.2	59.42	57.3	62.31	60.2	57.97	59.2	65.21	62.1	53.62	56.3	53.62	55.3
16	50.72	58.3	53.62	56.3	53.62	57.3	63.76	59.2	59.42	58.3	62.31	60.2	53.62	56.3	53.62	55.3
17	53.62	58.3	55.07	58.3	57.97	58.3	63.76	60.2	59.42	59.2	60.86	58.3	52.17	54.4	50.72	54.4
18	50.72	60.2	55.07	54.4	55.07	59.2	63.76	60.2	53.62	61.2	62.31	61.2	52.17	54.4	52.17	51.5
19	52.17	62.1	55.07	58.3	60.86	59.2	60.86	61.2	60.86	58.3	63.76	62.1	55.07	53.4	50.72	52.4
20	49.27	59.2	55.07	59.2	60.86	64.1	63.76	61.2	56.52	62.1	62.31	63.1	52.17	53.4	52.17	51.5
Mean	52.3	58.0	56.3	57.3	57.6	58.9	60.7	61.8	57.5	58.5	61.1	61.3	52.1	53.8	51.0	53.8
MAPE	91	73	78	75	74	70	65	62	75	71	64	63	92	87	96	87

superiority of the proposed model statistically holds true in all the datasets for both the splits.

The performance graphs of the hybrid CBR with cost-sensitive BPNN and with standard BPNN of sensitivity, activ-

ity, saliency and relevance are shown in Figs. 3, 4, 5, 6, 7, 8 and 9 for heart, diabetes, blood transfusion, liver disorder, Swine flu, sonar and ionosphere datasets, respectively, for easy comparison between cost-sensitive BPNN and stan-

Table 6 (color online) Swine flu

Values of k	Sensitivity				Activity				Saliency				Relevance			
	Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN	
	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30
1	88	90.7	88	90.7	88	90.7	90	90.7	90	90.7	90	88.0	50	56.0	52	60.0
2	88	90.7	88	90.7	88	90.7	90	90.7	90	90.7	90	88.0	50	56.0	52	60.0
3	88	88.0	90	90.7	90	92.0	90	92.0	90	92.0	90	92.0	56	58.7	62	65.3
4	90	89.3	90	90.7	90	89.3	92	90.7	90	92.0	92	90.7	56	58.7	60	65.3
5	88	88.0	94	89.3	90	90.7	92	93.3	90	92.0	92	90.7	62	61.3	70	68.0
6	88	88.0	92	90.7	88	90.7	92	92.0	88	92.0	92	90.7	54	61.3	56	70.7
7	88	89.3	92	89.3	88	90.7	92	93.3	90	93.3	92	92.0	64	62.7	64	76.0
8	88	88.0	92	89.3	88	90.7	92	90.7	88	92.0	92	90.7	56	60.0	58	72.0
9	88	89.3	92	89.3	88	90.7	90	93.3	90	93.3	90	93.3	66	66.7	62	72.0
10	88	89.3	92	89.3	88	90.7	90	93.3	88	92.0	90	93.3	64	66.7	60	72.0
11	88	89.3	92	89.3	88	90.7	90	93.3	88	93.3	90	93.3	64	68.0	60	76.0
12	88	89.3	92	89.3	88	90.7	90	93.3	88	93.3	90	93.3	60	69.3	64	74.7
13	88	89.3	92	89.3	88	90.7	90	92.0	88	92.0	90	90.7	64	70.7	66	72.0
14	88	89.3	92	89.3	88	90.7	90	92.0	88	92.0	92	92.0	64	69.3	64	70.7
15	88	89.3	92	89.3	88	90.7	92	92.0	88	92.0	92	92.0	62	73.3	64	73.3
16	88	89.3	92	89.3	88	90.7	90	93.3	88	90.7	94	93.3	60	69.3	60	76.0
17	88	89.3	92	89.3	88	89.3	92	92.0	88	90.7	92	92.0	64	72.0	58	76.0
18	88	89.3	92	89.3	88	90.7	90	90.7	88	92.0	94	90.7	60	72.0	60	73.3
19	88	89.3	92	89.3	88	89.3	90	90.7	88	89.3	92	88.0	62	74.7	58	73.3
20	88	89.3	92	89.3	88	89.3	90	90.7	88	90.7	88	90.7	58	70.7	64	72.0
Mean	88.1	89.1	91.5	89.6	88.3	90.4	90.7	92.0	88.7	91.8	91.2	91.2	59.8	65.8	60.7	70.9
MAPE	14	12	09	12	13	11	10	09	13	09	10	10	68	53	66	42

Table 7 (color online) Sonar

Values of k	Sensitivity				Activity				Saliency				Relevance			
	Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN		Standard BPNN		Cost-sensitive BPNN	
	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30
1	56.09	58.1	58.53	58.1	60.97	56.5	63.41	56.5	60.97	56.5	63.41	56.5	48.78	46.8	63.41	53.2
2	56.09	58.1	58.53	58.1	60.97	56.5	63.41	56.5	60.97	56.5	63.41	56.5	48.78	46.8	63.41	53.2
3	48.78	58.1	53.65	58.1	41.46	48.4	46.34	50.0	43.90	48.4	46.34	48.4	58.53	53.2	68.29	62.9
4	56.09	58.1	58.53	54.8	56.09	53.2	51.21	50.0	48.78	51.6	53.65	51.6	56.09	50.0	68.29	61.3
5	56.09	53.2	53.65	54.8	51.21	50.0	48.78	46.8	48.78	50.0	46.34	50.0	53.65	54.8	65.85	61.3
6	51.21	56.5	53.65	56.5	48.78	56.5	56.09	51.6	48.78	56.5	56.09	56.5	53.65	53.2	63.41	58.1
7	43.90	58.1	51.21	51.6	48.78	51.6	51.21	50.0	48.78	51.6	51.21	51.6	48.78	53.2	60.97	62.9
8	48.78	58.1	53.65	58.1	51.21	51.6	56.09	54.8	48.78	53.2	53.65	53.2	56.09	54.8	63.41	61.3
9	41.46	53.2	46.34	56.5	48.78	53.2	51.21	53.2	51.21	51.6	48.78	51.6	56.09	50.0	58.53	58.1
10	53.65	54.8	56.09	56.5	56.09	56.5	48.78	56.5	46.34	53.2	48.78	53.2	60.97	50.0	60.97	53.2
11	43.90	56.5	46.34	56.5	48.78	53.2	51.21	59.7	51.21	54.8	51.21	54.8	56.09	50.0	53.65	56.5
12	51.21	58.1	51.21	61.3	56.09	53.2	56.09	56.5	53.65	51.6	53.65	53.2	53.65	53.2	63.41	54.8
13	39.02	56.5	36.58	59.7	43.90	48.4	43.90	51.6	46.34	46.8	41.46	46.8	51.21	54.8	63.41	54.8
14	46.34	59.7	46.34	61.3	51.21	51.6	46.34	56.5	48.78	51.6	51.21	51.6	58.53	51.6	65.85	51.6
15	39.02	53.2	41.46	53.2	41.46	48.4	41.46	51.6	36.58	48.4	41.46	48.4	58.53	50.0	65.85	51.6
16	46.34	56.5	48.78	58.1	46.34	53.2	48.78	53.2	43.90	54.8	48.78	54.8	58.53	51.6	60.97	53.2
17	39.02	48.4	39.02	50.0	41.46	45.2	41.46	45.2	41.46	45.2	41.46	45.2	60.97	58.1	58.53	46.8
18	56.09	58.1	53.65	56.5	46.34	45.2	48.78	48.4	48.78	46.8	46.34	46.8	58.53	54.8	58.53	45.2
19	48.78	51.6	51.21	51.6	46.34	43.5	41.46	45.2	46.34	45.2	43.90	45.2	63.41	58.1	58.53	51.6
20	56.09	56.5	53.65	53.2	43.90	46.8	43.90	50.0	43.90	46.8	46.34	46.8	63.41	54.8	60.97	51.6
Mean	48.9	56.0	50.6	56.2	49.5	51.1	50.0	52.1	48.4	51.0	49.8	51.1	56.2	52.4	62.3	55.1
MAPE	108	79	101	78	105	97	103	93	109	97	103	97	79	91	61	83

Table 8 (color online) Ionosphere

Value s of k	Sensitivity				Activity				Saliency				Relevance			
	Standard BPNN		Cost- sensitive BPNN		Standard BPNN		Cost- sensitive BPNN		Standard BPNN		Cost- sensitive BPNN		Standard BPNN		Cost- sensitive BPNN	
	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30	80:20	70:30
1	74.28	81.9	75.71	84.8	74.28	78.1	75.71	78.1	72.85	80.0	75.71	80.0	58.57	54.3	60	56.2
2	74.28	81.9	75.71	84.8	74.28	78.1	75.71	78.1	72.85	80.0	75.71	80.0	58.57	54.3	60	56.2
3	72.85	80.0	77.14	81.0	74.28	77.1	74.28	77.1	70	79.0	74.28	79.0	58.57	58.1	65.71	58.1
4	72.85	79.0	74.28	79.0	74.28	77.1	71.42	77.1	71.42	77.1	71.42	77.1	55.71	58.1	62.85	59.0
5	74.28	78.1	74.28	78.1	71.42	76.2	72.85	76.2	72.85	76.2	72.85	77.1	52.85	60.0	61.42	58.1
6	72.85	78.1	75.71	79.0	70	74.3	71.42	75.2	71.42	76.2	71.42	76.2	55.71	60.0	62.85	61.0
7	72.85	76.2	74.28	78.1	71.42	76.2	72.85	75.2	71.42	73.3	72.85	73.3	54.28	60.0	65.71	60.0
8	71.42	74.3	72.85	78.1	71.42	72.4	72.85	72.4	71.42	73.3	72.85	72.4	55.71	60.0	61.42	61.0
9	71.42	75.2	72.85	74.3	72.85	69.5	74.28	69.5	72.85	68.6	74.28	68.6	55.71	58.1	62.85	60.0
10	71.42	74.3	72.85	76.2	71.42	71.4	74.28	72.4	72.85	71.4	72.85	71.4	54.28	56.2	61.42	60.0
11	72.85	73.3	72.85	75.2	71.42	70.5	75.71	70.5	74.28	70.5	75.71	70.5	57.14	56.2	62.85	61.0
12	72.85	72.4	72.85	76.2	71.42	68.6	72.85	68.6	72.85	67.6	72.85	67.6	55.71	56.2	62.85	60.0
13	72.85	70.5	72.85	74.3	71.42	65.7	74.28	65.7	74.28	64.8	74.28	64.8	55.71	58.1	62.85	60.0
14	72.85	71.4	72.85	76.2	70	68.6	72.85	68.6	74.28	67.6	72.85	67.6	55.71	56.2	61.42	61.9
15	72.85	68.6	74.28	68.6	72.85	62.9	74.28	62.9	72.85	61.9	72.85	61.9	60	59.0	61.42	61.0
16	72.85	71.4	72.85	68.6	70	63.8	72.85	62.9	71.42	62.9	72.85	62.9	58.57	56.2	61.42	60.0
17	72.85	62.9	72.85	66.7	70	61.0	72.85	61.0	71.42	61.0	72.85	61.0	61.42	57.1	61.42	61.0
18	72.85	62.9	72.85	67.6	70	61.0	72.85	61.0	71.42	61.9	72.85	61.9	54.28	56.2	61.42	60.0
19	72.85	60.0	72.85	64.8	65.71	58.1	71.42	58.1	70	59.0	71.42	59.0	55.71	57.1	62.85	60.0
20	70	60.0	71.42	64.8	68.57	58.1	70	58.1	71.42	59.0	72.85	59.0	55.71	57.1	62.85	60.0
Mean	72.7	72.6	73.7	74.8	71.3	69.4	73.2	69.4	72.2	69.5	73.2	69.5	56.5	57.4	62.2	59.7
MAPE	38	39	36	35	40	45	37	45	39	45	37	45	77	74	61	68

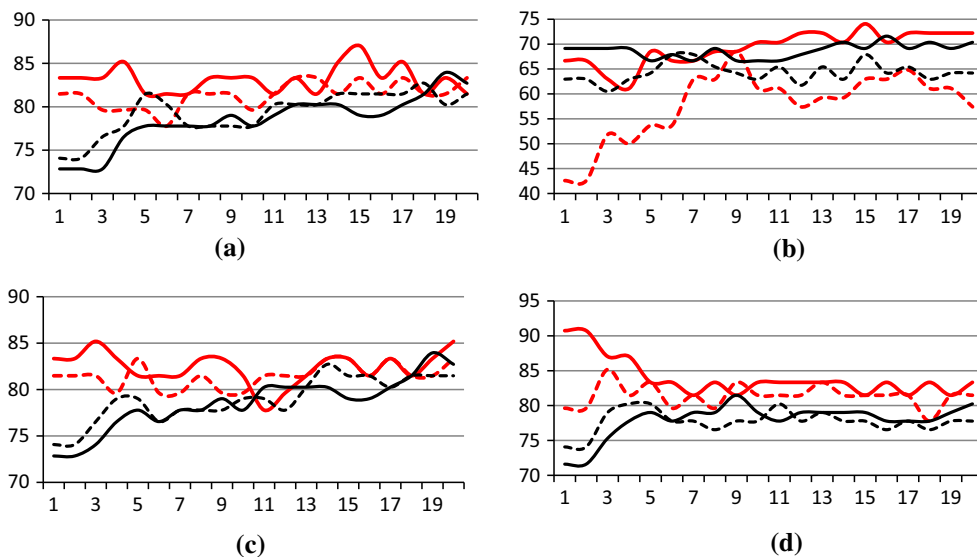


Fig. 3 Solid line is for cost-sensitive and dotted line is for standard BPNN, red color is for 80:20 split and black color is for 70:30, **a** activity, **b** relevance, **c** saliency and **d** sensitivity for heart (color figure online)

standard BPNN in imbalanced datasets for both the splits. The red lines and black lines represent the performance of the hybrid CBR models for 80:20 and 70:30 splits, respectively, in all the performance graphs. Dotted line represents standard BPNN, and solid line represents cost-sensitive BPNN. The X axis stands for the value of k from 1 to 20, and the Y axis stands for accuracy in percentage for each of the k values in Figs. 3, 4, 5, 6, 7, 8 and 9. Graphs for average performances (means) of cost-sensitive BPNN and standard

BPNN are shown in Fig. 10. Solid fill is for cost-sensitive BPNN, and pattern fill is for standard BPNN, and brown color is for 80:20 split and green color is for 70:30 split. Performance graphs and tables are showing the same analysis of the results. Performance graphs are used to improve the clarity and easy understanding of the comparisons.

Table 9 shows the optimal BPNN architectures, and Tables 10 and 11 show the λ values of the corresponding datasets for 80:20 and 70:30 splits, respectively.

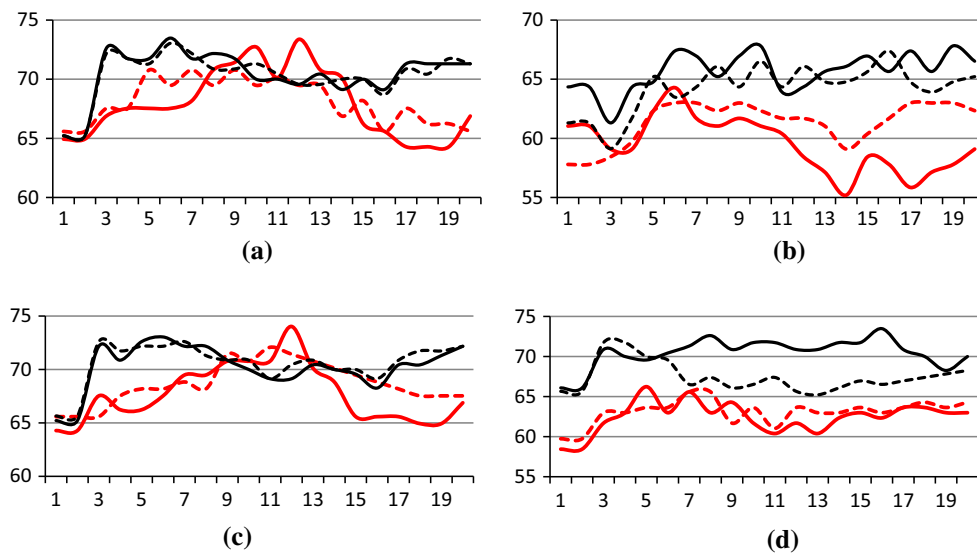


Fig. 4 Solid line is for cost-sensitive and dotted line is for standard BPNN, red color is for 80:20 split and black color is for 70:30, **a** activity, **b** relevance, **c** saliency and **d** sensitivity for diabetes (color figure online)

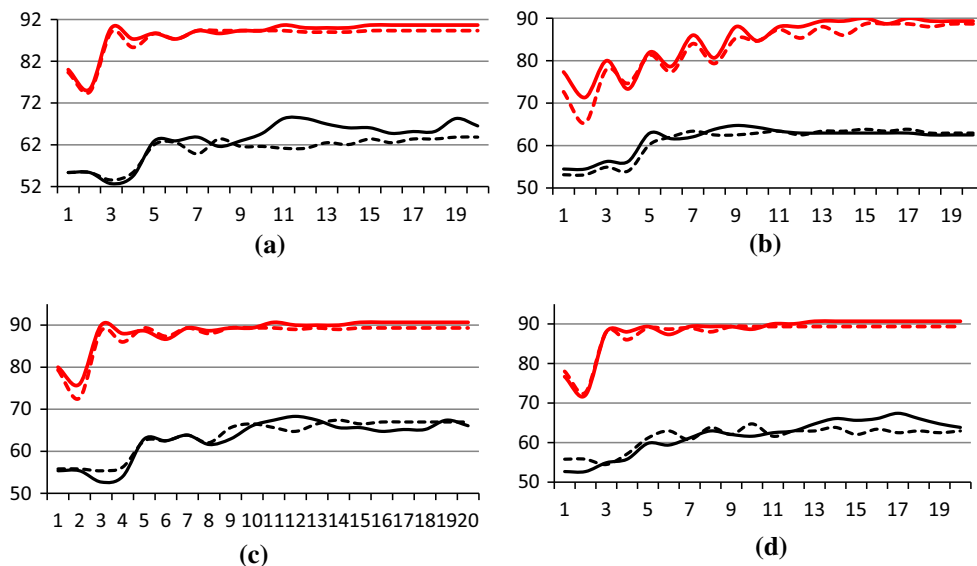


Fig. 5 Solid line is for cost-sensitive and dotted line is for standard BPNN, red color is for 80:20 split and black color is for 70:30, **a** activity, **b** relevance, **c** saliency and **d** sensitivity for blood transfusion (color figure online)

Five independent runs (five-folded validation) are executed to validate the comparison results by using Wilcoxon rank-sum nonparametric statistical test. The Wilcoxon rank sum is a nonparametric test which tests whether two independent samples come from identical continuous distributions with equal medians ($\mu_1 = \mu_2$), against the alternatives, ($\mu_1 > \mu_2$) or ($\mu_1 < \mu_2$) or ($\mu_1 \neq \mu_2$). μ_1 is the median for first sample and μ_2 is the median for second sample. This test is based on the ranks of all $N(=n_1 + n_2)$ observations after the two samples are combined, and the observations are ordered from smallest to largest. Table 12

shows the performance of the hybrid CBR model with cost-sensitive and with standard BPNN for five independent runs in Swine flu dataset by relevance mechanism. Combined performances and their rank are also shown in the table.

H_0 : represents null hypothesis; H_1 : represents alternative hypothesis

For a fixed level α (0.95) test, the test rejects H_0 (null hypothesis), where z -statistic = 1.671145, Upper value (z^*) = -1.64485 and P value (1-NORMSDIST (z)) = 0.047346.

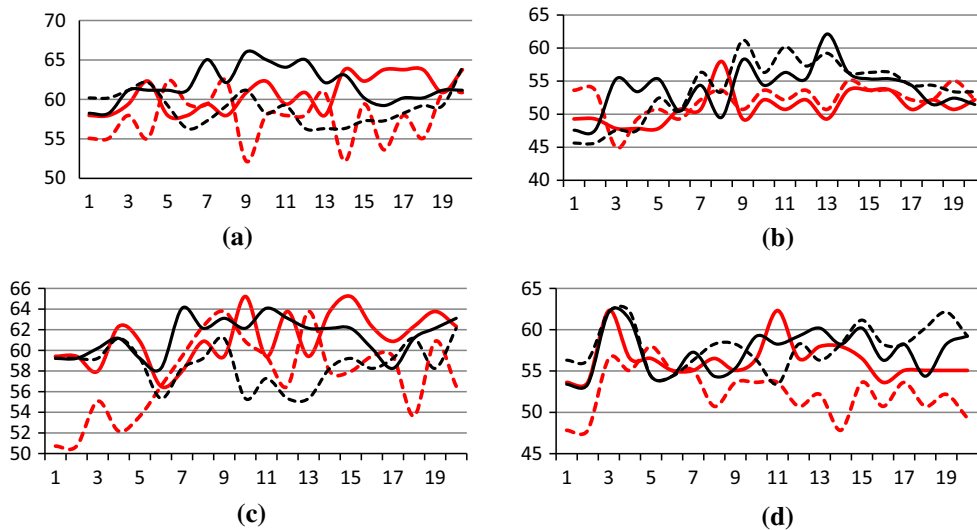


Fig. 6 Solid line is for cost-sensitive and dotted line is for standard BPNN, red color is for 80:20 split and black color is for 70:30, a activity, b relevance, c saliency and d sensitivity for liver disorder (color figure online)

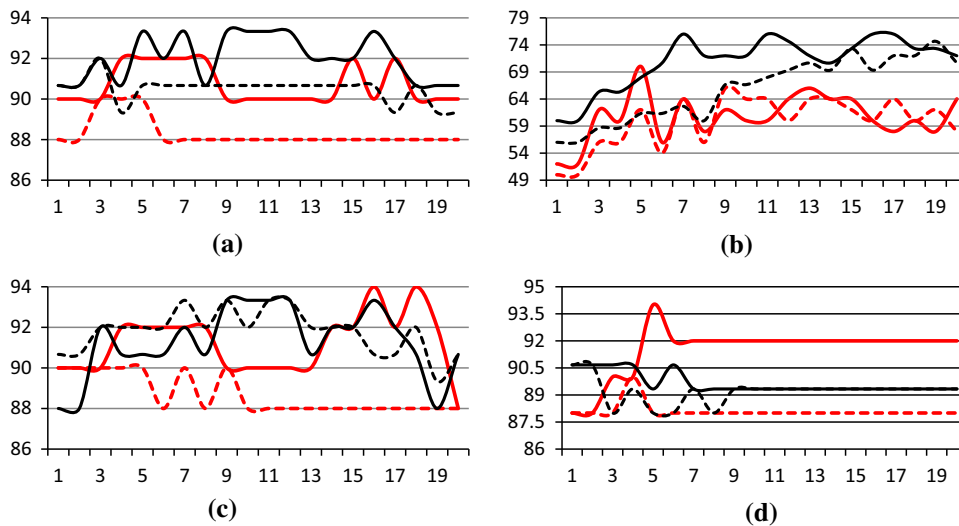


Fig. 7 Solid line is for cost-sensitive and dotted line is for standard BPNN, red color is for 80:20 split and black color is for 70:30, a activity, b relevance, c saliency and d sensitivity for Swine flu (color figure online)

The test result shows that at 95 % confidence level, z -statistic is greater than upper critical value as well as P value is <0.05 .

Hence, H_1 : case retrieval (performance) using cost-sensitive BPNN is better than standard BPNN (i.e., $\mu_1 > \mu_2$).

Table 13 shows the performance of the hybrid CBR model with cost-sensitive and with standard BPNN for five independent runs in ionosphere dataset by activity mechanism.

For ionosphere dataset at a fixed level α (0.95), z -statistic = 1.775592, upper value (z^*) = -1.64485 and P value (1-NORMSDIST (z))=0.0379. Z -statistic is greater than upper critical value as well as P value is <0.05 . Hence, H_0 (null hypothesis) is rejected and H_1 hypothesis (case retrieval (per-

formance) using cost-sensitive BPNN is better than standard BPNN (i.e., $\mu_1 > \mu_2$)) is satisfied.

Similarly, for all the other datasets H_1 hypothesis is satisfied at 95 % confidence level.

5.3 Discussion

It is observed from Tables 2, 3, 4, 5, 6, 7 and 8 and graphs in Figs. 3, 4, 5, 6, 7, 8 and 9 that sensitivity, activity, saliency and relevance from the cost-sensitive BPNN have performed better in retrieving most similar cases than Shin and Park (1999), Shin et al. (2000) and Im and Park (2007) in all the cases. It has also enhanced the performance of the case retrieval mechanism as compared to the weights in Shin and

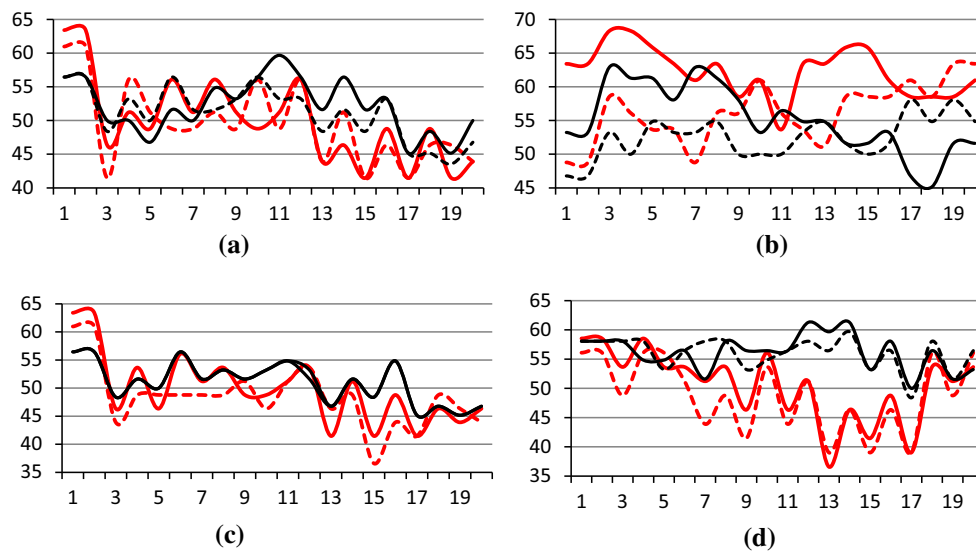


Fig. 8 Solid line is for cost-sensitive and dotted line is for standard BPNN, red color is for 80:20 split and black color is for 70:30, **a** activity, **b** relevance, **c** saliency and **d** sensitivity for sonar (color figure online)

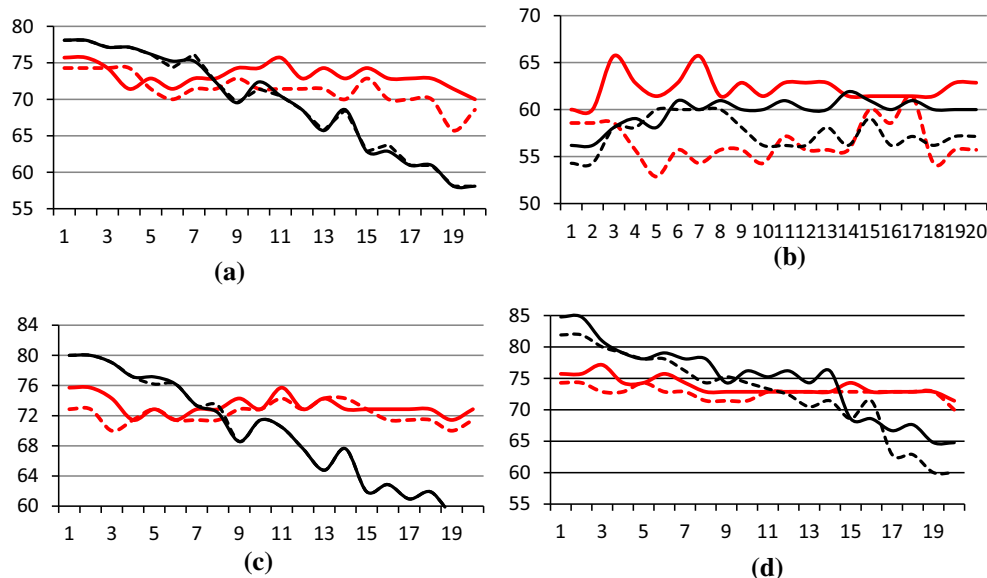


Fig. 9 Solid line is for cost-sensitive and dotted line is for standard BPNN, red color is for 80:20 split and black color is for 70:30, **a** activity, **b** relevance, **c** saliency and **d** sensitivity for ionosphere (color figure online)

Park (1999), Shin et al. (2000) and Im and Park (2007). It is because Shin and Park (1999), Shin et al. (2000) and Im and Park (2007) used standard BPNN and CBR to retrieve similar cases. Standard BPNN assumes that all the datasets (classes) are balanced and therefore has a tendency to get biased toward the majority class. As a result, the accuracy of the retrieved cases by Shin and Park (1999), Shin et al. (2000) and Im and Park (2007) gets affected. To cope with this problem, the proposed hybrid model uses a cost-sensitive BPNN, which uses a cost parameter to identify the importance of each class. Comparison of the proposed model is done only with Shin

and Park (1999), Shin et al. (2000) and Im and Park (2007) as models in Shin and Park (1999), Shin et al. (2000) and Im and Park (2007) have been already compared with other methods of case-based reasoning, and they outperformed all other previously existing case-based reasoning methods. The highest accuracy achieved by the proposed hybrid CBR model is made boldface in Tables 2, 3, 4, 5, 6, 7 and 8 for all the datasets and both the splits. For example, the proposed model produces 90.74, 87.03, 85.18 and 74.07 % accuracy for sensitivity, activity, saliency and relevance, respectively, in heart disease dataset for 80:20 split. A good model desires

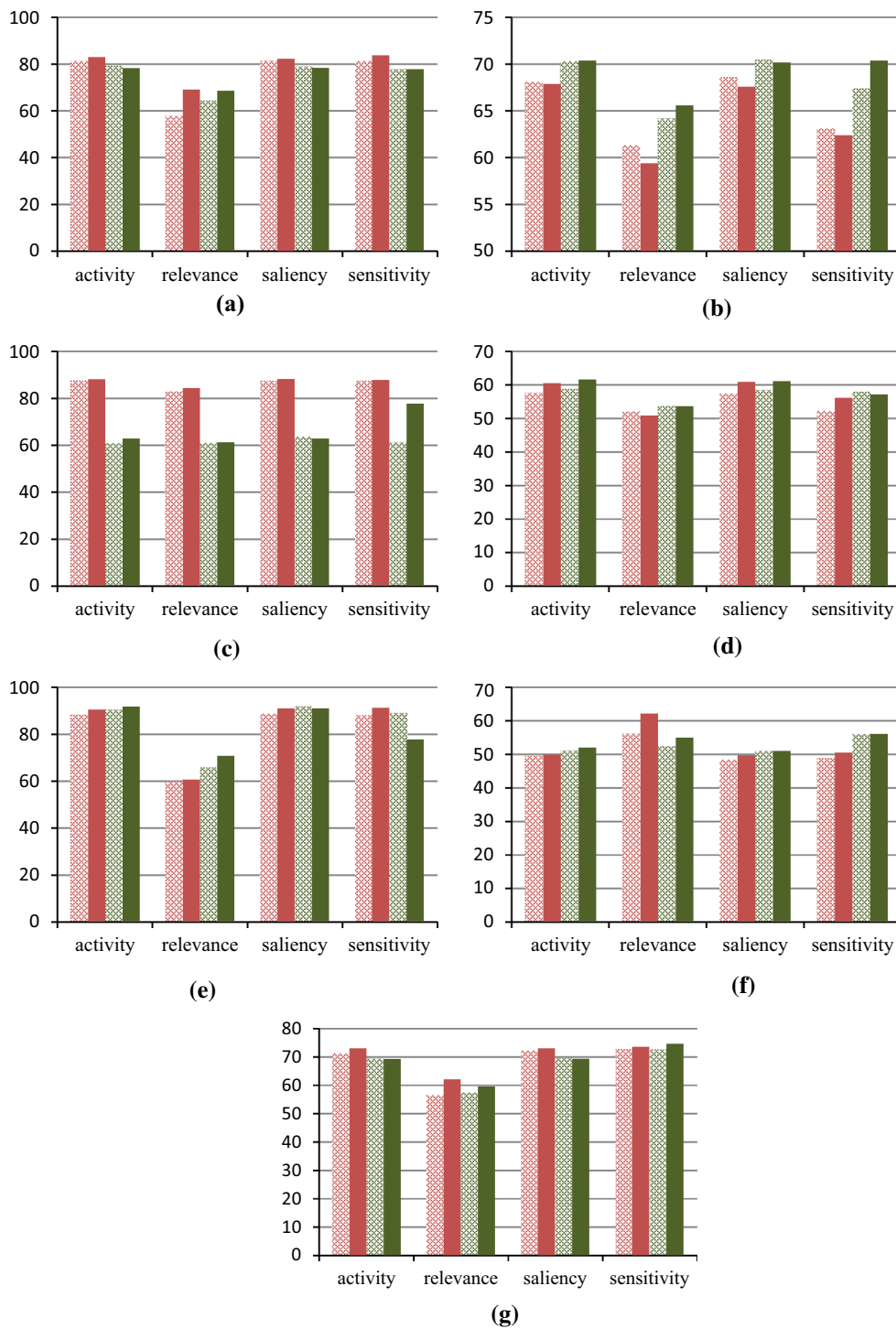


Fig. 10 Comparison of mean values of 20 nearest neighbors for different datasets. *Solid fill* is for cost-sensitive and *pattern fill* is for standard BPNN, and *brown color* is for 80:20 split and *green color* is for 70:30.

a Heart dataset, **b** diabetes dataset, **c** blood transfusion dataset, **d** liver disorder dataset, **e** Swine flu dataset, **f** sonar dataset and **g** ionosphere dataset (color figure online)

higher mean and lower MAPE (mean absolute percentage error). Tables 2, 3, 4, 5, 6, 7 and 8 and Fig. 10 show that the proposed model produces better mean and MAPE than

Shin and Park (1999), Shin et al. (2000) and Im and Park (2007) in most of the cases. For 70-30 split, it is observed from Tables 2, 3, 4, 5, 6, 7 and 8 and Fig. 10 that the aver-

Table 9 Optimal architecture

Dataset	Optimal architecture
Blood transfusion	4-6-1
Diabetes	8-14-1
Heart	13-24-1
Liver disorder	6-9-1
Swine flu	10-18-1
Sonar	60-90-1
Ionosphere	33-55-1

Table 10 λ Value of different datasets for 80:20 split

Dataset	Value of λ
Blood transfusion	0.73
Diabetes	0.57
Heart	0.61
Liver disorder	0.62
Swine flu	0.55
Sonar	0.54
Ionosphere	0.71

Table 11 λ Value of different datasets for 70:30 split

Dataset	Value of λ
Blood transfusion	0.81
Diabetes	0.67
Heart	0.6
Liver disorder	0.59
Swine flu	0.52
Sonar	0.57
Ionosphere	0.7

age performance of the proposed model is better in all the cases except saliency of heart, diabetes, blood transfusion and Swine flu datasets, and activity of heart dataset. Similarly for 80–20 split, the average performance of the proposed model is better in all the cases except sensitivity, activity, saliency

and relevance of diabetes dataset, and relevance of liver disorder dataset. From Tables 2, 3, 4, 5, 6, 7 and 8, it is also observed that for some datasets, the accuracy obtained is less (such as <60 %) and for some datasets accuracy obtained is more (such as more than 90 %). This is because of high or low nonlinearity of the datasets. The datasets producing less accuracy are highly nonlinear. By applying the proposed model with cost-sensitive BPNN, the accuracies for the highly nonlinear datasets are significantly increased. And those datasets producing better accuracy are less nonlinear but after applying the proposed model, significant improvement is also observed. So, it can be said that the proposed model retrieves k most similar cases more accurately. The revision of CBR cycle is implemented by frequency of occurrence of top most similar cases and the value of k , which produces highest accuracy, is only considered to measure the performance of the hybrid model. If more than one k produces highest accuracy, then the first k of them is taken as optimal k .

In order to obtain weights from trained neural network, it is essential to find out the optimal architecture of the network to be trained. A BPNN is used whose architecture is formed with one hidden layer because single hidden layer neural networks are superior to networks with more than one hidden layer with the same level of complexity mainly due to the fact that the later are more prone to fall into poor local minima (de Villiers Barnard 1993). The gradient descent learning algorithm is used, and the learning rate taken is 0.01. The number of input nodes to the network architecture is equal to the number of features of a pattern in a dataset, and there is one output node. The number of hidden nodes in the hidden layer is varied between $(i + 1)$ to $(2i)$ to determine the number of hidden nodes where i is the number of input nodes. Mean square error (MSE) is measured for each of the architectures, and the one which produces minimum MSE is selected for further experimentation and is called optimal architecture of the network. The network architectures that give the highest performance are given in Table 9 with respective datasets. The λ values that give the highest accuracy are given in Tables 10 and 11 with respective datasets. It should be noted that the value of λ is experimental.

Table 12 (color online) Wilcoxon rank-sum table for Swine flu (relevance)

Cost sensitive	Standard back propagation	Combined	Rank
78	76	66	1
84	76	70	2.5
78	74	70	2.5
76	70	74	4
70	66	76	6
		76	6
		76	6
		78	8.5
		78	8.5
		84	10

Table 13 (color online) Wilcoxon rank-sum table for ionosphere (activity)

Cost-sensitive BPNN	Standard BPNN	Combined	Rank
80.66	74.28	72.22	1
75.71	74.28	74.28	3.5
75.71	74.28	74.28	3.5
75.71	75.71	74.28	3.5
74.28	72.22	74.28	3.5
		75.71	7.5
		75.71	7.5
		75.71	7.5
		75.71	7.5
		80.66	10

6 Conclusions

Feature weighting mechanisms: sensitivity, activity, saliency and relevance, from a trained cost-sensitive BPNN are investigated in this paper. It is found that feature weighting mechanisms: sensitivity, activity, saliency and relevance, from a trained cost-sensitive BPNN perform better than feature weighting mechanisms: sensitivity, activity, saliency and relevance, from a trained standard BPNN. This is because the majority of machine learning techniques habitually assume that the training sets used for learning are balanced; however, this hypothesis is not always true in real-world applications. The tendency is to yield a classification model that is biased toward the overrepresented (majority) class.

The use of ANN to provide weights to k-NN module and identification of more exact nearest neighbors to a query case adds more intelligence to the hybrid CBR system. The efficiency of the proposed hybrid CBR model is reflected in obtaining proper retrieval of cases which produces better accuracy.

The proposed model can be used in any binary classification tasks such as diagnosis, prediction and many others. The feature weighting mechanisms with the cost-sensitive BPNN can be used to sort out feature weighting problem of CBR without any hitch. This work can be augmented by introducing the concept of local weighting mechanism. The proposed model is only for binary classification tasks but can also be augmented to multi-classification tasks.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- Aamodt A, Plaza E (1994) Case-based reasoning: foundational issues, methodological variations, and system approaches. *Artif Intell Commun* 7(1):39–52
- Becker L, Jazayeri K (1989) A connectionist approaches to case-based reasoning. In: *Proceedings of case-based reasoning workshop pensacola beach*. Morgan Kaufmann, San Mateo, pp 213–217
- Braga AP, Castro CL (2013) Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data. *IEEE Trans Neural Netw Learn Syst* 24(6):888–899
- Buta P (1994) Mining for financial knowledge with CBR. *AI Expert* 9(2):34–41
- Daelemans W, Gillis S, Durieux G (1994) The acquisition of stress: a data oriented approach. *Comput Linguist* 20(3):421–451
- de Villiers J, Barnard E (1993) Backpropagation neural nets with one and two hidden layers. *IEEE Trans Neural Netw* 4(1):136–141
- Ganganwar V (2012) An overview of classification algorithms for imbalanced datasets. *Int J Emerg Technol Adv Eng* 2(4):42–47
- Ha S (2008) A personalized counseling system using case-based reasoning with neural symbolic feature weighting (CANSY). *Appl Intell* 29(3):279–288
- Im KH, Park SC (2007) Case-based reasoning and neural network based expert system for personalization. *Expert Syst Appl* 32(1):77–85
- Leake DB (1996) *Case-based reasoning: experiences, lessons and future directions*. MIT Press, Cambridge
- López V, Fernández A, Moreno-Torres JG, Herrera F (2012) Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. *J Expert Syst Appl* 39(7):6585–6608
- Nguyen GH, Bouzerdoum A, Phung SL (2009) Learning pattern classification tasks with imbalanced data sets. In: Yin P (ed) *Pattern recognition*, chap 10. In-Tech, Vukovar, Croatia, pp 193–208
- Nunez H, Sanchez-Marre M, Cortes U, Comas J, Martinez M, Rodriguez-Roda I, Poch M (2004) A comparative study on the use of similarity measures in case based reasoning to improve the classification of environmental system situations. *Environ Model Softw* 19(9):809–819
- Park JH, Im KH, Shin CK, Park SC (2004) MBNR: case-based reasoning with local feature weighting by neural network. *Appl Intell* 21(3):265–276
- Park SC, Kim JW, Im KH (2006) Feature-weighted CBR with neural network for symbolic features. *ICIC, LNCS* 4113, pp 1012–1020
- Peng Y, Zhuang L (2007) A case-based reasoning with feature weights derived by BP network. In: *Proceedings of workshop on intelligent information technology application*. IEEE Computer Society, pp 26–29
- Sarwar S, Ul-Qayyum Z, Malik OA (2010) CBR and neural networks based technique for predictive prefetching. *MICAI, Part II, LNAI* 6438, pp 221–232
- Shin CK, Park SC (1999) Memory and neural network based expert system. *Expert Syst Appl* 16(2):145–155
- Shin CK, Yun UT, Kim HK, Park SC (2000) A hybrid approach of neural network and memory-based learning to data mining. *IEEE Trans Neural Netw* 11(3):637–646
- Thrift P (1989) A neural network model for case-based reasoning. In: *Proceedings of case-based reasoning workshop, Pensacola Beach*. Morgan Kaufmann, San Mateo, pp 334–337

- Watson I, Marir F (1994) Case-based reasoning: a review. *Knowl Eng Rev* 9(4):327–354
- Wettschereck D, Dietterich T (1995) An experimental comparison of the nearest-neighbor and nearest-hyper rectangle algorithms. *Mach Learn* 19(1):5–27
- Wettschereck D, Aha DW, Mohri T (1997) A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artif Intell Rev* 11(1–5):273–314
- Yang B, Jin X (2010) Method on determining feature weight in case-based reasoning system. In: *E-Product E-Service and E-Entertainment (ICEEE) international conference on 7–9 Nov 2010*, pp 1–4
- Zeng X, Martinez TR (2004) Feature weighting using neural networks. In: *IEEE international joint conference on neural networks*, pp 1327–1330