Performance of Case-Based Reasoning Retrieval Using Classification Based on Associations versus Jcolibri and FreeCBR: A Further Validation Study

Performance of Case-Based Reasoning Retrieval Using Classification Based on Associations versus Jcolibri and FreeCBR: A Further Validation Study

Ahmed S. Aljuboori[1], Frans Coenen[2], Mohammed Nsaif[3], David J. Parsons[4]

[1] The School of Computing, Science and Engineering, University of Salford, Salford, M5 4WT, UK
College of Education for Pure Science Ibn-Al-Haitham, University of Baghdad, Iraq.
[2] The Department of Computer Science, The University of Liverpool, Liverpool L69 3BX, UK
[3] College of Education for Pure Science Ibn-Al-Haitham, University of Baghdad, Iraq.
[4] A visiting professor with the School of Computing, Science and Engineering, University of Salford, Salford, M5 4WT, UK

[1] a.s.aljuboori@edu.salford.ac.uk.
[2] coenen@liverpool.ac.uk
[3] mohammed.k.n@ihcoedu.uobaghdad.edu.iq
4 d.j.parsons@salford.ac.uk

Abstract. Case-Based Reasoning (CBR) plays a major role in expert system research. However, a critical problem can be met when a CBR system retrieves incorrect cases. Class Association Rules (CARs) have been utilized to offer a potential solution in a previous work. The aim of this paper was to perform further validation of Case-Based Reasoning using a Classification based on Association Rules (CBRAR) to enhance the performance of Similarity Based Retrieval (SBR). The CBRAR strategy uses a classed frequent pattern tree algorithm (FP-CAR) in order to disambiguate wrongly retrieved cases in CBR. The research reported in this paper makes contributions to both fields of CBR and Association Rules Mining (ARM) in that full target cases can be extracted from the FP-CAR algorithm without invoking P-trees and union operations. The dataset used in this paper provided more efficient results when the SBR retrieves unrelated answers. The accuracy of the proposed CBRAR system outperforms the results obtained by existing CBR tools such as Jcolibri and FreeCBR.

***Index Terms*— case-based reasoning, class association rules, partial trees, similarity-based reasoning.**

1 INTRODUCTION

This fundamental idea underpinning CBR is a technique which uses the experience of previous cases to help solving new problems [1]. [2] stated that A case is a group of experience that are collected, defined and stored in a case base. Principally, the cases are defined by a problem description and its own solution. Among the four main stages in Case Based Reasoning, retrieval is a key step, with success being heavily dependent on its performance. Its aim is to retrieve the most similar cases that can be positively utilised to assist solve a target problem.

Basically, retrieval is performed through a specific strategy referred to as 'Similarity of Based Retrieval' (SBR) [2]. In SBR technique, measures are subroutines can be utilised to identify cases importance by their similarity to the case problem. The solution is basically "associated" to the closest case to enable users to determine the rank of cases [3].

Class Association Rule (CAR) mining is one of the most significant and common data mining approaches. It can be efficiently utilised in any decision making as a classifier [4]. It produces classification rules based on ARs as an integration of both classification and association. The integrated framework of CARs suggested by [5] is achieved by discovering a special subset of Association Rules (ARs) where the right-hand side of the implication equation is confined to the classification class label. The concept of CARs was used in previous work to show that patterns of classed rules can be combined with a similar pattern in the context of CBR [6], [7]. ARs have also been used by [8] for deriving association information from a certain case base using different ARM approaches.

The originality of the work presented in this paper is the focus on the contribution when 'Data Mining' (DM) is integrated into CBR in a merged system. The retrieved cases will be more effective to one correct class rather than being just similar to multi classes. In addition, this paper is a further validation based on a former study which highlighted the use of ARs as one of the DM methods that can be used in order to enhance the outcome of the retrieval process. A key point of this paper, is that the P-trees and set union concept which have been developed in [7] are not invoked to let different rules to be joint to produce a more correct cases. Therefore, the proposed CBRAR can produce remove the ambiguity of the retrieved answers of existing CBR systems.

One of the shortcomings of our former approaches [6], [7] were the reliability of the outcomes. Therefore, a further objective of this paper is to provide more improved and validated outcomes using CBRAR. TABLE 1 shows that in our earlier research the CBRAR only resolved 3 out of 4 cases achieving 75% accuracy [6], and 4 out of 5 cases achieving 80% accuracy [7]. In this paper, without invoking the P-trees and union of two rules which gave rise to the FP-CAR algorithm, CBRAR resolved 12 cases of the dataset.

2 A Brief Review of Related Work

CBR can be listed into four phases: retrieve – is to search and find the most similar cases, reuse – is to find what old cases can be reused, revise – is to apply the retrieved solution in a real world field and make sure it is correct, and retain – is to save the new solution and maintain it in CBR library as an experience [1]. Retrieval is a significant phase in the CBR cycle because if the system retrieves a wrong case, this could lead to a wrong decision. The main objective of this phase is to explore comparable cases that could be positively used to resolve a target problem. The process of retrieving a case starts with a (partial) matching of a new case, and finishes when the most similar cases are retrieved.

Some CBR methods retrieve a former case mainly based on the similarities among problem descriptors [9], [10]. Some approaches focus on deeper feature retrieval, [11], [12] and [13], while more recent methods attempt to utilise other knowledge to enhance SBR [14], [15]. The work presented in this paper further validates a new technique developed to improve the retrieval strategy. It also explores various methods by integrating other knowledge types into the CBR process. The cases build on similarities and the relative significance of features as a large part of the domain knowledge is required to explain the nature of why two cases are matched and how reliable the match is. In addition, the method of matching a case is described as hard or unachievable to obtain, because of the poor representation of the knowledge. By contrast, combined methods are capable of using the meaning of the problem. Therefore the description and its meaning make the similarity of matching cases obtainable [1].

CAR is one type of AR algorithms, which integrates association rule mining (finding all rules existing in the dataset that satisfy some constraints) and classification rule mining, (discovering a small set of rules in the database that forms an accurate classifier by focusing on mining a special subset of the existing ARs) [5]. It can be applied not only to linearly separable cases, but also to linearly inseparable cases, or where other linear classification approaches are not applicable [16]. One of CAR mining's advantages over conventional methods, for example support vector machine, is its interpretability. This is because classifiers are generated as a set of simple rules without much sacrifice of accuracy [17]. In addition, when applied to a medical dataset, for instance gene data, the CAR algorithm, which predicts a class label based on specific sets of differentially genes that are actually noticed in training samples, are expected to generate more biologically reasonable classifiers. This is because it is generally not single genes but groups of these genes that jointly define phenotypes i.e. drug responses [18]. In a CBR context, a CAR is represented as an AR in which a consequent holds the item built as a pair of a solution attribute and its value. This might be called a solution item. A CAR therefore has the form $X \Rightarrow y$, where $X \subseteq I$ is an itemset and $y \in I$ is a solution item. Considering this, it should be noted that the form of a CAR X⇒ y allows the representation of correlation between an itemset X (i.e., a set of problem attributes) and a solution item y (i.e., the resultant solution) in an easy method. CAR mining is also considered to be an extension of the Apriori algorithm. In other words, the goal of CAR is to find all rules of the items of the form < cond_set, y > where cond_set is a set of items, and y ∈ Y where Y is the set of class labels. In addition, Association Rules can be employed in order to find significant relationships from case bases [8].

Accordingly, derived subsets of a set of CARs are the CARs whose resultant is limited to one label only. In the CBR scope, a CAR is a rule whose importance consists of the element is designed as a pair included a solution attribute and its real value [15]. In a specific case base storage, AK is programmed to describe the cause behind the specific problem's features and then associate with a solution. The proposed CBRAR strategy which presented in the earlier research [6] has been enhanced by containing the union set theory as well as the FP-CAR algorithm using a new dataset.

3 Material and Methods

This section gives a brief review of the characteristics of the newly validated dataset and its validated performance compared with the dataset used in [6], [7].

3.1 Balloon Dataset Characteristics

The Balloon dataset is utilised to test the FP-CAR algorithm. It has been used to determine the influence of prior knowledge e.g. experimental results, and experimental psychology, including learning and memory. Adult and stretch is one of the four dataset provided by Michael Pazzani within the UCI repository website. It is used for determining the status of inflating a balloon to acknowledge whether it is true or false according to its characteristics.

This dataset is used in the third series of experiments to support the theory behind our proposition. The attributes' values and characteristics that were used in the hash table and FP-CAR tree are as follows {A= act {Stretch, Dip}}, {B= age {Adult, Child}}, {C = colour {yellow, purple}}, {D=size {Small, Large}}. The values of attributes appear in the FP-CAR tree as a string. In addition, the class characteristic is {Class =c{True=c1, False=c2}}.

3.2 Performance of CBRAR using Balloon versus Acute Inflammation and Space Shuttle Datasets

The main objective of this paper is to present further validation results when integrating CARs into CBR to disambiguate the wrongly retrieved cases where a full case can be extracted from the CBRAR. A block diagram illustrating the proposed Strategy is given in Figure 1. It starts by combining existing data mining algorithms i.e. CAR and FPGrowth to obtain the FP-CAR algorithm. The diagram also shows that the FPCAR is prioritized to find a full pattern that can be compared with a target case, if a full match is recognized, a correct solution can be extracted from the frequent built tree and there is no need to use other knowledge i.e. P-trees and union to obtain an optimum tree. This Strategy was developed and explained in details by the author in [6], [7]. The FP-CAR algorithm consists of two stages. Firstly, it produces a FP-tree from a set of CARs [19]. Secondly, the P-tree concept [20] and table of implications are used to optimize the resultant FP-tree. The former two phases will then be processed to obtain an optimum tree. This tree can be likened to a new case referred as Q which is a super pattern to improve the performance of the Similarity-Based Reasoning.

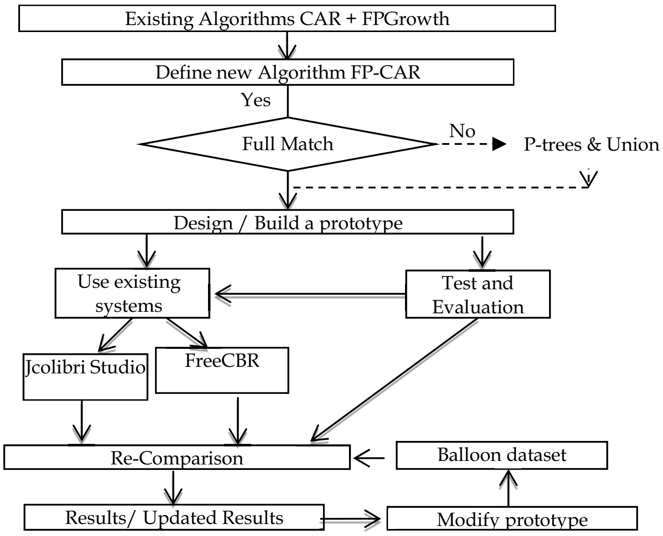

Figure 1. CBRAR Flow Diagram

In [6], a voting process was achieved by counting the value of longest length of edges that connect nodes of the updated FP-tree when a partial match is found. In addition, the P-trees technique was used so as to compensate any missing nodes in the tree if required to create an similar pattern to CBR queries. Therefore, 3 out 4 cases were resolved by the CBRAR when voting and P-trees are used. In [7], CBRAR was enhanced by the author when union is utilized at the top of P-trees. The shortcomings of [6] was the results accuracy. Thus, the purpose of Enhanced CBRAR was to suggest a better result utilising a new method based on union theory. Whilst, previous study only resolved two out of five cases, employing the joint of 2 rules has generated in the FP-CAR algorithms resolving four out of five cases. Therefore, the error is decreased and the accuracy rate is doubled with respect to the space shuttle dataset.

In this research, we remove one case from the CBR base storage till the system returns two different classes with the same percentage of similarity as shown in

TABLE 2. In the new conducted experiments an attempt to validate the FP-tree has been done and the rules allocated to one label as a root class have been obligated. Furthermore, this paper is implemented using Balloon dataset, where the CBRAR resolved more cases with less run time memory being used where no P-tree and union are required to resolve a target case. The mentioned datasets i.e. acute inflammations and space shuttle were previously tested to assess the former work of the researchers in [6], [7].

Ultimately, the outcomes gained utilizing the new validated dataset are likened with the outcomes of the Similarity Based Reasoning phase in order to choose the best answer. The chosen case is then likened with the outcome of the retrieved cases in order to ignore unassociated solutions.

4 EXPERIMENTAL RESULTS

In order to prove that the CBRAR has achieved better results, the accuracy of CBRAR is examined by conducting more experiments on a new dataset benchmarked from the UCI website, namely the Balloon dataset. Typically, one case is drawn from the CBR base to be counted as a new case for each experiment conducted by Jcolibri and FreeCBR. The Balloon dataset was utilised to calculate the accuracy rate of both CBRAR and CBR considering the same case base source as the input for both systems. Basically, when using Jcolibri the SBR returns the 5 most similar cases for a new target case. However, the preordained cases (1,2), 3, 4, 6, 8, 9, (11,12), 13, 14, (16,17), 18 and 19 all retrieved unrelated cases that misled the decision maker. This is because the whole returned cases had an identical ratio of similarity with two classes i.e. (*c1, c2*). Practically, the FreeCBR tool produces more cases than those returned by Jcolibri tool. In

TABLE 2, the outcomes are presented in the first column denotes the new arrived case *Q* and then the retrieved other cases using the Case-Based Reasoning tools i.e. NewCase(1,2). Then cases (3, 4, 6, 7 and 11 for Jcolibri) and one more case 12 for FreeCBR. The other column denotes to attributes which start with *A* then followed by 3 extra attributes *B,* *C* and *D*. The class section associated with labels (c1, c2). Those "Accuracy" columns recap the assessment of Jcolibri, FreeCBR as welll as CBRAR. The terms TP and FP allude on true positive and false negative distinctly.

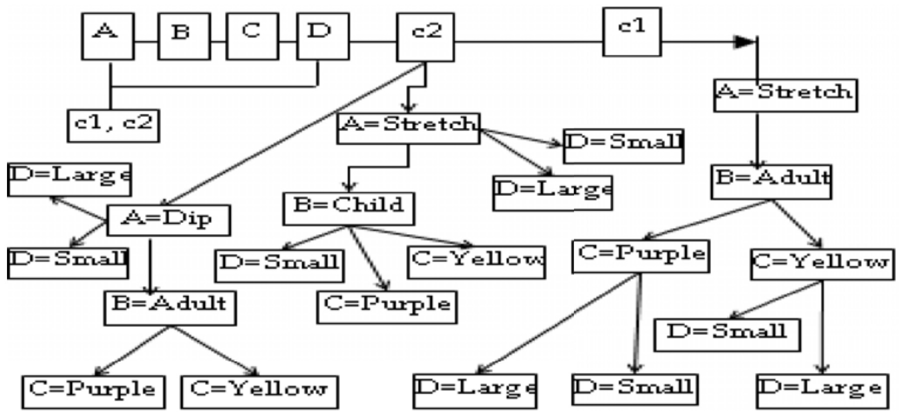

Figure. 2. FP-CAR Tree

TABLE 2 Indicates that to each new examined case utilizing CBR, five cases for those same similitude measure for 0. 866 would recouped by Jcolibri once those NewCases(1,2) is utilized. Jcolibri retrieved 2 TP also 3 FP cases with those same comparability proportion

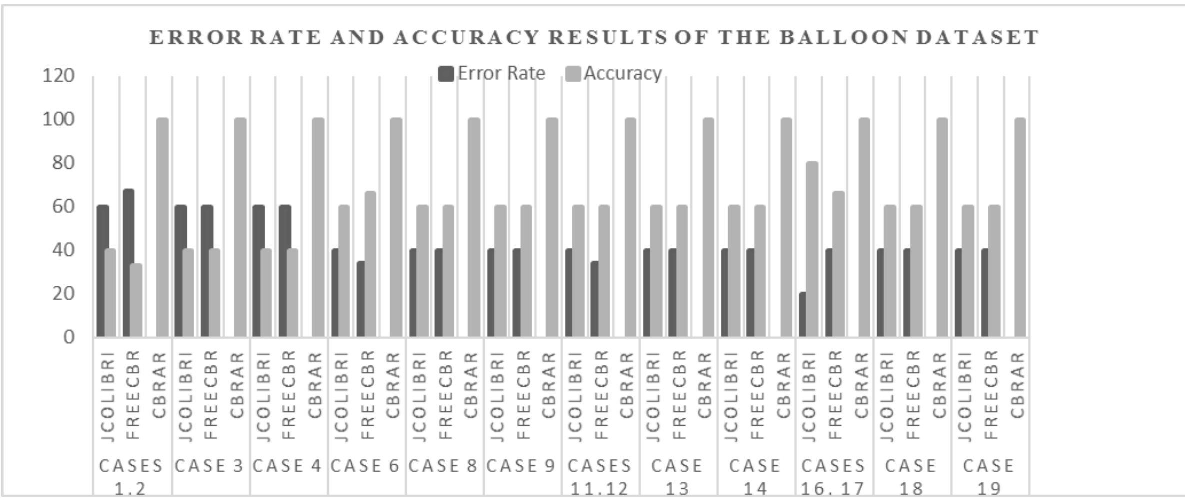

Figure. 3. Error Rate and Accuracy of Balloon Dataset

equal to 40%, whereas FreeCBR retrieved 6 cases, 2 TP Furthermore 4 FP for same similitude proportion of 50.0%, giving an accuracy rate of 33%. whilst CBRAR returned 1 TP case of evidence from the suggested model, giving a exactness about 100%.

In the second experiment, the NewCase3 has been applied to the CBR. Both Jcolibri and FreeCBR returned two TN and three TN with an accuracy of 40%. CBRAR returned 1 TN from the new system, achieving an accuracy of 100%. Third experiment, used NewCase4, Jcolibri and FreeCBR retrieved 2 TN and 3 FN cases for same similitude rate, equivalent to 40% accuracy and CBRAR returned 1 TN from the FP-CAR tree. inside the fourth test, Newcase6, Jcolibri retrieved 3 TP moreover 2 FP cases for the ones identical similitude ration, giving 60% accuracy, even as FreeCBR retrieved 6 instances, 4 TP what is extra 2 FP, undertaking an exactness charge from claiming 66%. CBRAR retrieved 1 TP case, giving an accuracy of 100%. In the 5th experiment, NewCase8 is applied, Jcolibri retrieved three TN and a couple of FN instances with the equal similarity ratio, giving 60% accuracy, and FreeCBR retrieved three

TN and a pair of FN, giving 60% accuracy. CBRAR retrieved 1 TN case from the new model which is the correct case hence achieving 100% accuracy and outperforming the performance of the CBR tools used. In the sixth experiments, NewCase9, Jcolibri retrieved 3 TN and 2 FN cases with the same similarity ratio, providing 60% accuracy, and FreeCBR retrieved 3 TN and 2 FN, providing 60% accuracy. CBRAR retrieved 1 TN case from the new strategy which is the correct case therefore giving 100% accuracy and outperforming the performance of the CBR tools used.

In the seventh experiment, using NewCases (11 ,12), Jcolibri returned three TP and two FP responses with similar percentages, presenting 60% accuracy, whilst FreeCBR returned four TP and two FP, registering 66% accuracy. CBRAR returned one TP case from the new version which is the correct case therefore giving 100% accuracy. In experiment 8, using NewCase13, Jcolibri and FreeCBR retrieved 3 TN and 2 FN cases with the same similarity percentage, and this achieved 60% accuracy. CBRAR retrieved 1 TN case from the new model which is the correct case hence achieving 100% accuracy. In experiment 9, which used NewCase14, Jcolibri and FreeCBR retrieved 3 TN and 2 FN cases with the same similarity, giving 60% accuracy. CBRAR retrieved 1 TN case from new model, and this equals to 100% and showing a better performance when compared with the CBR tools used. In the tenth experiment, using NewCases (16, 17), Jcolibri returned 4 TP and 1 FP instances registering same percentage of similarity, giving 80% accuracy, FreeCBR returned four TP and two FP, achieving 66% precision. CBRAR returned 1 TP instance which is the correct case hence achieving 100% accuracy. In the experiment 11, NewCase18 was used, Jcolibri retrieved 3 TN and 2 FN cases with the same similarity, and this achieved 60% accuracy, and FreeCBR retrieved 3 TN and 2 FN, giving 60% accuracy, whereas 1 TN case was retrieved by the CBRAR which is the correct case hence achieving 100% accuracy. In the twelfth experiment, NewCase19 applied to the CBR, Jcolibri and FreeCBR retrieved 3 TN and 2 FN cases with the same similarity ratio, giving 60% accuracy. From the new model CBRAR retrieved 1 TN case which is the correct case hence achieving 100% accuracy and outperforming the performance of the CBR tools used.

Table 2 also shows that, the NewCases((1,2), 6, (11,12) and (16,17)) have matched a full pattern within the FP-CAR algorithm without invoking the P-tree procedure to compensate the missing nodes as used in [6], [7]. The novel strategy (CBRAR) is a significant step in the machine learning field, where a target case can be drawn directly from FP-CAR for a further validation.

The consequences display that 34 out of the 60 Jcolibri retrieved instances are categorized as TP and TN giving fifty six% accuracy. via evaluation, 34 of the 64 instances retrieved through FreeCBR are categorized as TP and TN giving 53% accuracy. Nevertheless, both Jcolibri and FreeCBR provide confusing set of consequences. The proposed CBRAR system exhibits favorable element through both Jcolibri and FreeCBR through determining 12 out of 12 giving one hundred percent exactness as well as no ambiguity. Cases (1,2), 3, 4, 6, 8, 9, (11,12), 13, 14, (16,17), 18 and 19 in

TABLE 2 can be reworked in Figure. 2 to prove that CBRAR identifies a correct case using a frequent classed tree for a further validation study.

The most remarkable outcomes in this study is the significantly reduced error rates attained in [6],[7]. In Figure. 3 , the chart illustrates a comparison between Jcolibri and FreeCBR as traditional tools versus the suggested CBRAR explaining the error and accuracy rate. It can be seen that CBRAR offered many advantages and registered zero error percentage, which is considered the least amongst other rates. Thus, giving the highest accuracy and notably resolving 12 out of 12 cases accurately on the Balloon dataset as shown in Figure. 3.

TABLE 1. Validation of Balloon Dataset Vs Acute Inflammation and Space Shuttle Datasets

Dataset	Solved Cases			CBRAR	CBRAR
	CBRAR	Jcolibri	FreeCBR	Accuracy	Error
Acute inflammation	3/4	14/20	29/35	75%	25%
Space Shuttle	4/5	10/18	13/21	80%	20%
Balloon	11/11	34/60	34/64	100%	0%

TABLE 2. Results of Wrong Retrieved Cases

Cases	Attributes					Accuracy		
	A	B	C	D	Class	Jcolibri	FreeCBR	CBRAR

NewCase1,2	**STRETCH**	**ADULT**	**YELLOW**	**SMALL**	**c1**	**0.866**	**50.0**	**100**
Case3	STRETCH	CHILD	YELLOW	SMALL	c2	TP	TP	TP
Case4	DIP	ADULT	YELLOW	SMALL	c2	TP	TP	
Case6	STRETCH	ADULT	YELLOW	LARGE	c1	FP	FP	
Case7	STRETCH	ADULT	YELLOW	LARGE	c1	FP	FP	
Case11	STRETCH	ADULT	PURPLE	SMALL	c1	FP	FP	
Case12	STRETCH	ADULT	PURPLE	SMALL	c1		FP	
NewCase3	**STRETCH**	**CHILD**	**YELLOW**	**SMALL**	**c2**	**0.866**	**50.0**	**100**
Case1	STRETCH	ADULT	YELLOW	SMALL	c1	TN	TN	
Case2	STRETCH	ADULT	YELLOW	SMALL	c1	TN	TN	TN
Case5	DIP	CHILD	YELLOW	SMALL	c2	FN	FN	
Case8	STRETCH	CHILD	YELLOW	LARGE	c2	FN	FN	
Case13	STRETCH	CHILD	PURPLE	SMALL	c2	FN	FN	
NewCase4	**DIP**	**ADULT**	**YELLOW**	**SMALL**	**c2**	**0.866**	**50.0**	**100**
Case1	STRETCH	ADULT	YELLOW	SMALL	c1	TN	TN	
Case2	STRETCH	ADULT	YELLOW	SMALL	c1	TN	TN	TN
Case5	DIP	CHILD	YELLOW	SMALL	c2	FN	FN	
Case9	DIP	ADULT	YELLOW	LARGE	c2	FN	FN	
Case14	DIP	ADULT	PURPLE	SMALL	c2	FN	FN	
NewCase6	**STRETCH**	**ADULT**	**YELLOW**	**LARGE**	**c1**	**0.866**	**50.0**	**100**
Case1	STRETCH	ADULT	YELLOW	SMALL	c1	TP	TP	
Case2	STRETCH	ADULT	YELLOW	SMALL	c1	TP	TP	
Case8	STRETCH	CHILD	YELLOW	LARGE	c2	FP	FP	TP
Case9	DIP	ADULT	YELLOW	LARGE	c2	FP	FP	
Case16	STRETCH	ADULT	PURPLE	LARGE	c1	TP	TP	
Case17	STRETCH	ADULT	PURPLE	LARGE	c1		TP	
NewCase8	**STRETCH**	**CHILD**	**YELLOW**	**LARGE**	**c2**	**0.866**	**50.0**	**100**
Case3	STRETCH	CHILD	YELLOW	SMALL	c2	TN	TN	
Case6	STRETCH	ADULT	YELLOW	LARGE	c1	FN	FN	TN
Case7	STRETCH	ADULT	YELLOW	LARGE	c1	FN	FN	
Case10	DIP	CHILD	YELLOW	LARGE	c2	TN	TN	
Case18	STRETCH	CHILD	PURPLE	LARGE	c2	TN	TN	
NewCase9	**DIP**	**ADULT**	**YELLOW**	**LARGE**	**c2**	**0.866**	**50.0**	**100**
Case4	DIP	ADULT	YELLOW	SMALL	c2	TN	TN	
Case6	STRETCH	ADULT	YELLOW	LARGE	c1	FN	FN	TN
Case7	STRETCH	ADULT	YELLOW	LARGE	c1	FN	FN	
Case10	DIP	CHILD	YELLOW	LARGE	c2	TN	TN	
Case19	DIP	ADULT	PURPLE	LARGE	c2	TN	TN	
NewCase11,12	**STRETCH**	**ADULT**	**PURPLE**	**SMALL**	**c1**	**0.866**	**50.0**	**100**
Case1	STRETCH	ADULT	YELLOW	SMALL	c1	TP	TP	
Case2	STRETCH	ADULT	YELLOW	SMALL	c1	TP	TP	
Case13	STRETCH	CHILD	PURPLE	SMALL	c2	FP	FP	TP
Case14	DIP	ADULT	PURPLE	SMALL	c2	FP	FP	
Case16	STRETCH	ADULT	PURPLE	LARGE	c1	TP	TP	
Case17	STRETCH	ADULT	PURPLE	LARGE	c1		TP	
NewCase13	**STRETCH**	**CHILD**	**PURPLE**	**SMALL**	**c2**	**0.866**	**50.0**	100
Case3	STRETCH	CHILD	YELLOW	SMALL	c2	TN	TN	TN
Case11	STRETCH	ADULT	PURPLE	SMALL	c1	FN	FN	
Case12	STRETCH	ADULT	PURPLE	SMALL	c1	FN	FN	
Case15	DIP	CHILD	PURPLE	SMALL	c2	TN	TN	
Case18	STRETCH	CHILD	PURPLE	LARGE	c2	TN	TN	
NewCase14	**DIP**	**ADULT**	**PURPLE**	**SMALL**	**c2**	**0.866**	**50.0**	**100**
Case4	DIP	ADULT	YELLOW	SMALL	c2	TN	TN	TN
Case11	STRETCH	ADULT	PURPLE	SMALL	c1	FN	FN	
Case12	STRETCH	ADULT	PURPLE	SMALL	c1	FN	FN	

Case15	DIP	CHILD	PURPLE	SMALL	c2	TN	TN	
Case19	DIP	ADULT	PURPLE	LARGE	c2	TN	TN	
NewCase16,17	**STRETCH**	**ADULT**	**PURPLE**	**LARGE**	**c1**	**0.866**	**50.0**	**100**
Case6	STRETCH	ADULT	YELLOW	LARGE	c1	TP	TP	TP
Case7	STRETCH	ADULT	YELLOW	LARGE	c1	TP	TP	
Case11	STRETCH	ADULT	PURPLE	SMALL	c1	TP	TP	
Case12	STRETCH	ADULT	PURPLE	SMALL	c1	TP	TP	
Case18	STRETCH	CHILD	PURPLE	LARGE	c2	FP	FP	
Case19	DIP	ADULT	PURPLE	LARGE	c2		FP	
NewCase18	**STRETCH**	**CHILD**	**PURPLE**	**LARGE**	**c2**	**0.866**	**50.0**	**100**
Case8	STRETCH	CHILD	YELLOW	LARGE	c2	TN	TN	TN
Case13	STRETCH	CHILD	PURPLE	SMALL	c2	TN	TN	
Case16	STRETCH	ADULT	PURPLE	LARGE	c1	FN	FN	
Case17	STRETCH	ADULT	PURPLE	LARGE	c1	FN	FN	
Case20	DIP	CHILD	PURPLE	LARGE	c2	TN	TN	
NewCase19	**DIP**	**ADULT**	**PURPLE**	**LARGE**	**c2**	**0.866**	**50.0**	**100**
Case9	DIP	ADULT	YELLOW	LARGE	c2	TN	TN	TN
Case14	DIP	ADULT	PURPLE	SMALL	c2	TN	TN	
Case16	STRETCH	ADULT	PURPLE	LARGE	c1	FN	FN	
Case17	STRETCH	ADULT	PURPLE	LARGE	c1	FN	FN	
Case20	DIP	CHILD	PURPLE	LARGE	c2	TN	TN	
Average						**56%**	**53%**	**100%**

5 CONCLUSION

To conclude, the P-trees and union of two rules was used in CBRAR [6], [7]. The assessment on the on the acute inflammation dataset recorded three out for four situations with seventy five percent of accuracy. On the space go back and forth dataset, it became possible to promote the previous system. The FP-CAR method resolved four out of five cases instead of two out of five cases via doubling the accuracy from forty to eighty percent, wherein it constructs far fewer common classed subsets than would be built from an actual FP-tree.The CBRAR approach has offered a better performance of Similarity Based Reasoning but needed more time and memory for resolving cases on the previous study. It makes use of a length voting technique likened to the TFPC set of rules wherein nodes preserve a value of object node at the same time as building the tree. moreover, the subsets of the common tree that meet the support and confidence, and longest length can be used to suit the pattern while indexed in a hash table. In this paper, a further validation achieved on the Balloon dataset showed that in NewCases ((1,2), 6, (11,12) and (16,17)) a superset was drawn directly from the FP-CAR in order for it to be matched with other CBR target cases. Furthermore, New Cases (3, 4, 8, 9, 13, 14, 18 and 19) were resolved using a P-tree has shown the ability of CBRAR to resolve more cases compared to our previous work. Jcolibri and FreeCBR as a CBR tools were contrasted to the CBRAR and an improved grade of accuracy was gained with the least error percentage up to this point. All unclear answers with similar percentage have been avoided by CBRAR as an advantage of using it over Jcolibri and FreeCBR. As a result of the outcomes offered on this research paper, extra datasets will be examined in order to generalise the proposed strategy.

References

[1] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI Commun.*, vol. 7, no. 1, pp. 39–59, 1994.

[2] R. Lopez De Mantaras *et al.*, "Retrieval, reuse, revision and retention in case-based reasoning," *Knowl. Eng. Rev.*, vol. 20, no. 3, pp. 215–240, 2005.

[3] P. Perner, "Introduction to Case-Based Reasoning for Signals and Images. Case-Based Reasoning on Signals

and Images," in *Case-Based Reasoning on Images and Signals*, P. Perner, Ed. Springer Verlag, 2008, pp. 1–24.

[4] D. Nguyen, L. T. T. Nguyen, B. Vo, and T.-P. Hong, "A novel method for constrained class association rule mining," *Inf. Sci. (Ny).*, vol. 320, pp. 107–125, 2015.

[5] B. Ma, W. Liu, and Y. Hsu, "Integrating classification and association rule mining," in *Proceedings of the 4th Knowledge Discovery and Data Mining*, 1998.

[6] A. S. Aljuboori, F. Meziane, and D. J. Parsons, "A new strategy for case-based reasoning retrieval using classification based on association," in *12th MLDM International Conference*, 2016, p. pp 326-340.

[7] A. Aljuboori, "Enhancing case-based reasoning retrieval using classification based on associations," *2016 6th International Conference on Information Communication and Management (ICICM)*. pp. 52–56, 2016.

[8] Y.-B. Kang, S. Krishnaswamy, and A. Zaslavsky, "A Retrieval Strategy for Case-Based Reasoning Using Similarity and Association Knowledge," *IEEE Trans. Cybern.*, vol. 44, no. 4, pp. 473–487, 2014.

[9] P. Thagard, K. J. Holyoak, G. Nelson, and D. Gochfeld, "Analog retrieval by constraint satisfaction," *Artif. Intell.*, vol. 46, no. 3, pp. 259–310, 1990.

[10] Y. Orbach, M. E. Lamb, K. J. Sternberg, J. M. G. Williams, and S. Dawud-Noursi, "The effect of being a victim or witness of family violence on the retrieval of autobiographical memories," *Child Abuse Negl.*, vol. 25, no. 11, pp. 1427–1437, 2001.

[11] J. L. Arcos and E. Plaza, "A reflective architecture for integrated memory-based learning and reasoning," in *Topics in Case-Based Reasoning*, Springer, 1993, pp. 289–300.

[12] B. Li and H. Johan, "3D model retrieval using hybrid features and class information," *Multimed. Tools Appl.*, vol. 62, no. 3, pp. 821–846, 2013.

[13] J. Kendall-Morwick and D. Leake, "A Study of Two-Phase Retrieval for Process-Oriented Case-Based Reasoning," in *Successful Case-based Reasoning Applications-2*, Springer, 2014, pp. 7–27.

[14] M.-J. Huang, M.-Y. Chen, and S.-C. Lee, "Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis," *Expert Syst. Appl.*, vol. 32, no. 3, pp. 856–867, 2007.

[15] V. / Aparna and M. Ingle, "Enriching Retrieval Process for Case Based Reasoning by using Vertical Association Knowledge with Correlation," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 2, no. 12, pp. 4114–4117, 2014.

[16] D. L. Sampson, T. J. Parker, Z. Upton, and C. P. Hurst, "A comparison of methods for classifying clinical samples based on proteomics data: a case study for statistical and machine learning approaches," *PLoS One*, vol. 6, no. 9, p. e24973, 2011.

[17] F. P. Pach, A. Gyenesei, and J. Abonyi, "Compact fuzzy association rule-based classifier," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2406–2416, 2008.

[18] A. R. Bateman, N. El-Hachem, A. H. Beck, H. J. W. L. Aerts, and B. Haibe-Kains, "Importance of collection in gene set enrichment analysis of drug response in cancer cell lines," *Sci. Rep.*, vol. 4, 2014.

[19] "TFPC Classification Association Rule Mining (CARM) Software." [Online]. Available: https://cgi.csc.liv.ac.uk/~frans/KDD/Software/Apriori-TFPC/Version2/aprioriTFPC.html. [Accessed: 15-Dec-2015].

[20] F. Coenen, P. Leng, and S. Ahmed, "Data structure for association rule mining: T-trees and P-trees," *IEEE Trans. Knowl. Data Eng.*, no. 6, pp. 774–778, 2004.