

RACER: Rule-Associated Case-based Reasoning for Supporting General Practitioners in Prescription Making

S.L. Ting, W.M. Wang, S.K. Kwok*, Albert H.C. Tsang, W.B. Lee

Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong

Abstract

Prescription is an important element in the medical practice. An appropriate drug therapy is complex in which the decision of prescribing is influenced by many factors. Any discrepancy in the prescription making process can lead to serious consequences. In particular, the General Practitioners (GPs), who need to diagnose and treat a wide range of health conditions and diseases, must be knowledgeable enough in deciding what type of medicines should be given to the patients. With the widespread computerization of medical records, GPs now can make use of accumulated historic clinical data in retrieving similar decisions in therapeutic treatment for treating the new situation. However, the applications of decision support tools are rarely found in the prescription domain due to the complex nature of the domain and limitations of the existing tools. It was argued that existing tools can only solve a small amount of the cases on the real world dataset

This paper proposes a new revised Case-based Reasoning (CBR) mechanism, named Rule-Associated Case-based Reasoning (RACER), which integrates CBR and association rules mining for supporting GPs prescription. It aims at leveraging the two most common techniques in the field and dealing with the complex multiple values solution. 800 real cases from a medical organization are collected and used for evaluating the performance of RACER. The proposed method was also compared with CBR and association rules mining for testing. The results demonstrate that the combination leads to increased in both recall and precision in various settings of parameters. The performance of RACER remains stable by using different sets of parameters, which shows that the most important element of the mechanism is self-determined.

Keywords: Association rules mining, Case-based reasoning, Decision support, Hybrid intelligent system, Prescription

* Corresponding Author

1. Introduction

Prescription is an essential element of medical practice. It functions as a pharmacological therapy to relieve pain of people, resolve individuals with unusual medical conditions by controlling their symptoms, and refreshing their bodies into healthy condition. The aim is to administer the most appropriate medicines for a particular patient or a population of patients in order to achieve the desired therapeutic results with minimal adverse drug effects and to improve their conditions by given the available clinical information (Galland, 1997). Quaglini et al. (1992) described that the therapy consists of repetitive cycles that the physicians are capable to: (i) obtain data regarding the state of patient; (ii) interpret these data to make diagnostics hypotheses and therapy for remedy; (iii) evaluate and refine the therapy; (iv) predict the progress; and (v) remove contra-indicated therapy. In each action within the process, good observation and experiences of the physicians are demanded. In contrast, patient data misinterpretation and insufficient knowledge in medicines are the main reasons for the cause of medication errors.

Furthermore, the decision to prescribe is influenced by many other factors, such as interactions between physician and patient, cost issues, uncertainty of the diagnosis, and complex nature of medicine information (Bradley, 1991; Bradley, 1992; Chen and Landefield, 1994; Gill et al., 1995; Denig et al., 1998; Coscelli, 2000; Lundin, 2000; Wazana, 2000). Bradley (1991), and Greenhalgh and Gill (1997) discussed that the act of issuing medicines is the culmination of a complex chain of decisions along with biomedical, historical, psychosocial, and commercial influences. Substantially, an appropriate prescription is difficult to be made and medication errors often occur when there is any improper use of medication.

According to the study conducted by the Institute of Medicine (2006), around 1.5 million people are injured and 7,000 died each year in the United States because of medication errors. On average, every hospital is probably subjected to at least one medication error every day. Carter (2004) discussed that the most likely prescription mistakes made by physicians are: (i) interactions between the prescribed medicines and the medicines the patient is already taking, or the foods that the patient commonly eats; (ii) lack of the considerations of medicine allergy; (iii) failure to recognize the side effect; and (iv) incorrect dose. In particular, General Practitioners (GPs) need to diagnose and treat a wide range of health conditions and diseases. Most of the patients go to consult GPs instead of specialists during their first visit. In other words, GPs must be knowledgeable in interpreting patients' conditions as well as deciding which kind of treatments should be conducted (i.e. either prescribing medicines or referring the patient to other health professionals). Such complex demand of services increases the challenge of GPs to provide effective treatment, especially in cases where they are not familiar with.

Attempts to respond to these issues, Oren et al. (2003) investigated that technology-based intervention plays an important role in avoiding medication errors and improving patient safety. Decision Support System (DSS) has been proposed as one of the most effective ways of medication errors reduction, since it integrates both knowledge-based and expert-based concepts to support GPs in selecting and deciding appropriate medicines to cure the patient (Garg et al., 2005). DSS is a computerized system which provides an interactive and user-friendly interface. It makes use of historic patient data and elements of relevant medical knowledge (such as the information provided in biomedical literatures) to reach the required conclusion. The historic patient data is always stored in the DSS in the form of Electronic Medical Record (EMR). Each EMR consists of patient information,

symptoms and diagnoses encountered, and resulted treatment, etc. When a General Practitioner (GP) faces a new medical problem, s/he looks up the DSS. The DSS then deduces recommendations based on the given problem and the previous successful cases stored in the DSS. As a result, the GP is provided with extra information and evidences for supporting his/her decision so as to improve the service quality.

For recently decade, DSS has been widely applied to various kinds of problem solving in medical domains (Banning, 2008). One of the possible reasons to explain this phenomenon is that the decision making process of GPs is similar to the inference process employed in the DSS. Deutsch et al. (2001) pointed out that physicians wish to rely directly on the past experience that stored in the historic patient data, select similar cases that had reliable outcomes and reuse the solution accordingly, which works similar to the inference process of DSS. Therefore, the quality of a DSS is highly depended on its inference mechanism. Among the numerous inference techniques, Case-Based Reasoning (CBR) and association rules mining are the two common techniques used in the medical industry. CBR utilizes the specific knowledge of previously experienced and concrete problem situations (cases), while association rules mining relies on general knowledge of a problem domain and making associations along generalized relationships between problem descriptors and conclusions (Zhuang, et al., 2009). They are two distinct techniques that consist of their own strengths and limitations. And they are also seldom integrated together, particularly in the prescription domain.

Unlike other medical domains (such as cancer diagnosing), the conclusion of decision support of prescription is more complex that consists of a number of medicines. Comparing to diagnosis, which always considers only two classes (e.g. either positive or negative) or multiple classes (e.g. one disease out of different diseases), each medicine out of hundreds of medicines can be a part of the solution in prescription making. However, the multiple values solution received lack of concern in the domain. Therefore, this paper proposes a new revised CBR mechanism, named Rule-Associated CasE-based Reasoning (RACER), to integrate CBR with association rules mining for supporting GP prescription. It aims at leveraging the strengths and compensating the limitations of CBR and association rules mining, in order to improve the accuracy of selecting the appropriate medicines for prescription making. Furthermore, RACER is the first model attempted to handle the multiple values solution by considering both specify knowledge and general knowledge. The model has been tested through a series of real data experiments and encouraging results are yielded.

2. Related work

In order to describe the process of developing the RACER, it is helpful to first discuss the principle and related work of CBR, association rules mining, and the integrated approach.

2.1 Case-based reasoning (CBR)

CBR is used to solve a new problem by remembering a previous similar situation and by reusing information and knowledge of that situation (Aamodt and Plaza, 1994). Similar to human problem-solving process, CBR requires a knowledge-based learning mechanism to learn from old cases and reuse the most specific case or set of cases to explain the new situations (Hammond, 1989). Compared with other problem-solving techniques (such as Bayesian networks and neural networks), CBR does not have the tendency to over-generalize (Mitchell, 1997), and thus CBR can achieve excellent accuracy provided that it generates the solutions from the memorized cases (Bichindaritz and Marling, 2006).

CBR is argued to be very effective in the medical domain. Bichindaritz and Marling (2006) stated that CBR is an essential tool in decision support in the health sciences because reasoning from historical examples is natural for healthcare professionals and case histories have long been used in the training of health care professionals. Huang et al. (2007) explained that the favor of CBR adoption in medicines are due to its cognitive adequateness, explicit experience, duality of objective and subjective knowledge, automatic acquisition, and system integration. Dussart et al. (2008) also argued that CBR is an effective reasoning strategy for optimizing clinical practice in which it learns through experiences and matches the natural reasoning model of human.

Over the decades, CBR has widely applied in medical domain ranged from supporting diagnosis, prescription to treatment planning. For example, Marling and Whitehouse (2001) developed AUGUSTE to support treatment planning in Alzheimer's disease by using CBR to determine if a neuroleptic medicine should be prescribed and hence select the approved medicines for a patient via a rule-based mechanism. Hartge et al. (2006) proposed a similarity measurement algorithm for a CBR system to help minimizing inappropriate selection of medicines that will cause adverse drug-drug interaction. Hartge et al. (2006) modeled the patient treatments as groups of vectors representing discrete time intervals to explore similarity in treatment of different patients. With the enhancement in the flexibility and speed of wireless computing, O'Sullivan et al. (2007) proposed that caregivers can input patients' symptoms to a mobile device for quickly retrieving similar profiles in supporting effective diagnoses and prognoses by comparing symptoms, treatments, diagnosis, test results and other patient information. In-depth review of applying CBR in medical domain can be found in Schmidt et al. (2001), and Yusof and Buckingham (2009).

Despite numerous researches showing CBR is effective in problem-solving in medical domain, several researchers argued that the chance of reusing a case from CBR is not high in some areas (Atzmueller et al., 2003), such as insurance claims prediction (Daengdej et al., 1999) and multiple medical disorder cases (Shi and Barnden, 2005). According to the study of Atzmueller et al. (2003), CBR can only solve about 3% of the cases on their real world dataset. It limits the power to explain and address the new problems. This is also true in the domain of prescription support. Since the solution of a prescription case typically involves multiple medicines (usually 5 to 7 medicines), not all the medicines are effective in addressing the problem in the new case. Thus, further modification of CBR is required to improve the accuracy of selecting the appropriate set of medicines in prescription support.

2.2 Association rules mining

Association rules mining is another common techniques used in DSS (Lee et al., 2001; Cho et al., 2002; Chien and Chen, 2008; García et al., 2008). It aims to extracts interesting correlations, frequent patterns, associations or casual structures among sets of items in databases (Kotsiantis and Kanellopoulos, 2006). A famous example of applying association rules is the market basket analysis. Agrawal and Srikant (1996) introduced the Apriori algorithm for discovering regularities between products in large scale transaction data recorded by point-of-sale systems. The rules can be expressed as “ $\{X, Y\} \Rightarrow \{Z\}$ [support: 60% and confidence: 80%]” meaning that X , Y and Z occur in 60% of all transactions (i.e. support) and 80% of the transactions containing X and Y contains Z (i.e. confidence). In general, a rule regards as interesting if it satisfies both the minimal support and confidence thresholds that pre-defined by experienced users or domain experts.

Similar to CBR, association rules mining is also widely applied in medical domain. The

main reasons is due to its ability in uncovering new information and relationships embedded in the large databases, and generating new patterns and relationships throughout learning mechanism (Abidi, 2001; Hung et al., 2006). Kuo et al. (2006) employed clustering techniques to cluster the medical database into several groups, and hence apply the association rules mining algorithm to discover the hidden relation in the groups easier. Jiang and Gruenwald (2004) proposed to use association rules to mine the association relationships among different genes under the same experimental conditions. Shan et al. (2008) presented an application of association rule mining to detect fraud and inappropriate practice in the health service management domain.

All in all, association rules mining discovers important rules which provide useful references for GPs in making decisions. However, it is always difficult to handle the redundant frequent association when the database is large. Particularly in the prescription domain, the complex nature and large variety of medicines make the association rules mining difficult to identify useful and meaningful rules, and hence limits its ability in deriving solutions accurately. Furthermore, the threshold values of minimum support and confidence are difficult to be determined. It requires rigorous training of the parameters before real life applications.

2.3 Integrated approach of CBR and association rules mining

CBR has been widely integrated with other mechanisms. One of the possible options is to integrate with rules derived by the domain experts or generated by mining from the databases as such integration can achieve new synergies and significantly improve the problem-solving capabilities in CBR (Kumar et al., 2009). Several researches have been done by integrating the two approaches. Montani et al. (2003) integrated CBR, rule-based reasoning (RBR) and model-based reasoning to provide physicians with a reliable decision support tool in the context of type 1 diabetes mellitus management. Apart from approaches integration, Rossille et al. (2005) proposed a multi-modal reasoning decision-support system based on the RBR-first CBR-last approach to automatically compare the patient's case to the corresponding guideline, then to other cases, and retrieve similar cases for breast cancer. In recent years, Park et al. (2009) integrated CBR with rule induction (RI) techniques for case filtering. They applied their method to three medical diagnosis datasets and their findings demonstrated that the hybrid approach significantly outperforms the results in either CBR or RI.

Important contributions have been made by integrating CBR and rules in numerous applications. However, lack of researches and empirical investigations have been done for the prescription related topics. Therefore, the RACER method proposed in this paper is focused on improving the solution extracted in CBR (especially the missing medicines) and providing relevant and objective evidences in the prescription support. This paper also proposes a novel measure to rank the multiple values solution by combining the results from CBR and association rules mining, so as to assist the physicians in identifying which medicines are more appropriate for the patients.

3. Rule-Associated Case-based Reasoning (RACER)

The proposed RACER methodology is mainly composed of three parts: cases retrieval, association rules mining, and suggestions combination. RACER starts from the point where the GP interprets the diagnosis of patient. As shown in Figure 1, a new case (the diagnosis) is firstly codified based on a predefined EMR. The medical data recorded in EMR consists of all the examination data and patient particular information which is voluminous and

heterogeneous. It is important to preprocess the data by selecting attributes or features which are useful for prescription making. The codified new case is then processed by comparing with the previous cases retained in the knowledge base. Association rules mining and case retrieval are then applied. Association rules mining is used to extract the most interesting association rules based on support and confidence measure. Weightings are then assigned to the associated medicines. Simultaneously, the most similar cases are extracted based on a similarity measure for cases retrieval. Weightings are then assigned to the retrieved medicines. Then, the weightings of the associated medicines and the retrieved medicines are combined based on a simple rule of combination which is adapted from the Dempster's rule of combination (Dempster, 1968). Based on the combination, a consolidated medicine list is provided as suggestion for the new case. The suggestion is then reviewed and revised by the GP. When the case and results are verified, they are then retained to the knowledge base for future reuse.

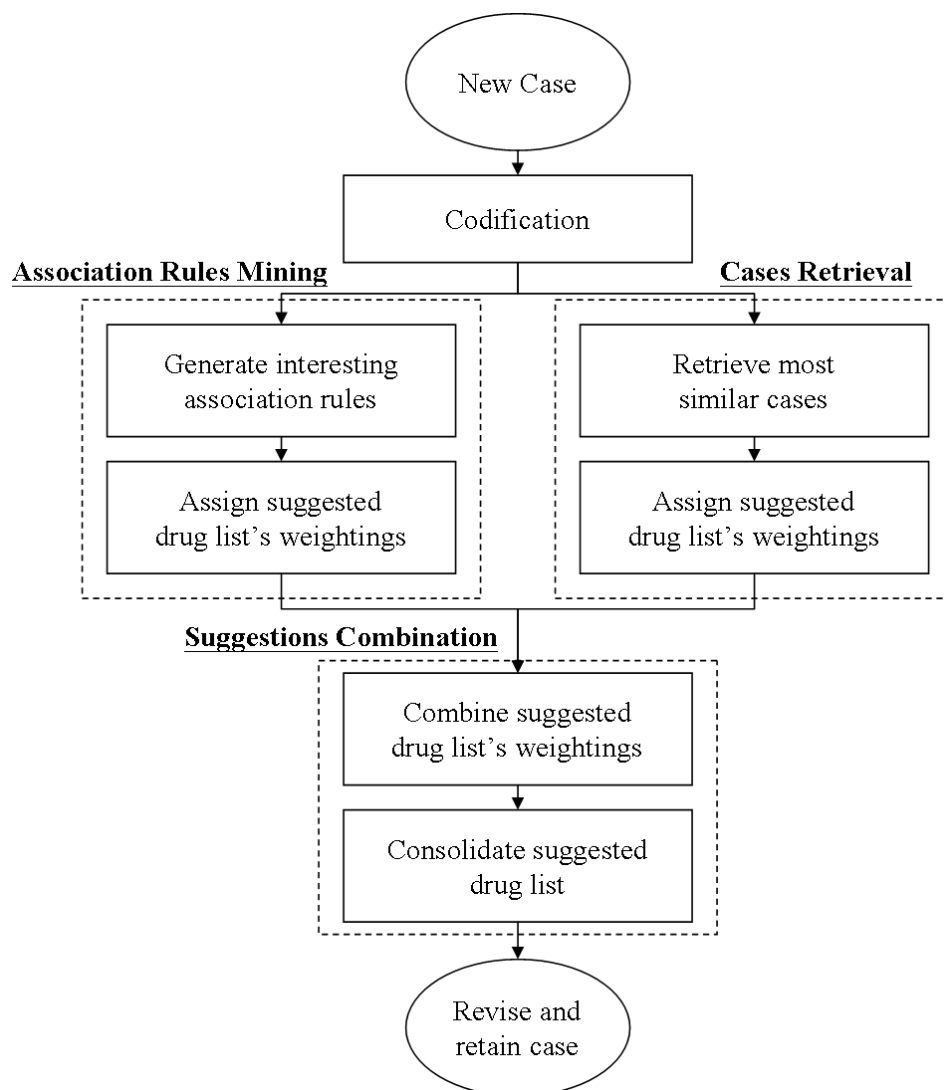


Figure 1: Architecture of rule-associated case-based reasoning (RACER)

3.1 Cases retrieval in RACER

In general, CBR consists of case retrieval, adaption, reuse, and retain. However, Schmidt et al. (2001) discussed that adoption of complete CBR cycle are rather exceptional in the medical field. Zhuang et al. (2009) also mentioned that it is almost impossible to generate adaptation rules to consider all possible important differences between current and former

similar cases in medical application. Since the adaptation knowledge is difficult to be acquired, the present study is focused only on the retrieval of similar cases.

A case consists of features for describing the problem and solution. In the present study, the information of diagnosis and patient particular information (such as age and gender) are representing the problem features of a case, while the medicines to be prescribed are described as the solution of the case. Mathematically, each case is represented in the following notation:

Let Ω_E be the set of all cases, Ω_A be the set of all problem features, and Ω_D be the set of all medicines (i.e. the solution attribute), where each case $c \in \Omega_E$, each attribute $a \in \Omega_A$, and each medicine $d \in \Omega_D$. Thus,

$$c = (A_c, D_c) \quad (1)$$

where $A_c \subseteq \Omega_A$ is the set of problem features observed in the case c . The set $D_c \subseteq \Omega_D$ is the set of medicines to be prescribed for this case.

Medical records are codified and stored as cases in a knowledge base for case retrieval. The present study employs a similarity measure approach, nearest-neighbor retrieval (NNR), for determining the degree of similarity between the new case and old case. This method is used due to its simplicity and good performance in case indexing (Sun and Finnie, 2004). During comparison, the features of a new case are matched to their corresponding features of all cases stored in the knowledge base.

The algorithm of cases retrieval is shown in Figure 2. A threshold γ is set for determining the maximum number of similar cases being retrieved. The contributions of the retrieved cases are weighted by their corresponding similarity, so that the similar cases contribute more to the average than the less similar ones. It is accomplished by the following steps:

- (i) n most similar cases are retrieved based on equations (2) and (3)
- (ii) A list of unique medicines is extracted from the retrieved cases
- (iii) Weightings of the unique medicines are determined based on the occurrence of the corresponding medicines prescribed in each retrieved case, which is shown in equations (4) and (5)

The similarity for each case is calculated by equations (2) and (3):

$$\text{similarity}(c^I, c^R) = \frac{\sum_{i=1}^n w_i \times \text{sim}(a_i^I, a_i^R)}{\sum_{i=1}^n w_i} \quad (2)$$

$$sim(a_i^I, a_i^R) = \begin{cases} 0 & \text{if } a_i^I \text{ and } a_i^R \text{ are categorical and } a_i^I \neq a_i^R \\ 1 & \text{if } a_i^I \text{ and } a_i^R \text{ are categorical and } a_i^I = a_i^R \\ \frac{a_i^R}{a_i^I} & \text{if } a_i^I \text{ and } a_i^R \text{ are numerical and } a_i^I > a_i^R \\ \frac{a_i^I}{a_i^R} & \text{if } a_i^I \text{ and } a_i^R \text{ are numerical and } a_i^I < a_i^R \end{cases} \quad (3)$$

where c^I and c^R represent the new case and the old case respectively, a_i^I and a_i^R represent the i -th feature value of the new case and the old case respectively, the similarity function $sim(a_i^I, a_i^R)$ computes the similarity between a_i^I and a_i^R , and w_i represents the feature weighting for each i -th feature.

The weightings of the unique medicines are calculated by equations (4) and (5):

$$W_j^{cbr} = \frac{\sum_{i=1}^n similarity(c_i^I, c^R) \times s_{i,j}}{\sum_{i=1}^n similarity(c_i^I, c^R)} \quad (4)$$

$$s_{i,j} = \begin{cases} 1 & \text{if } d_j \text{ is prescribed in } c_i \\ 0 & \text{if otherwise} \end{cases} \quad (5)$$

where W_j^{cbr} is the weighting of medicine j , n is the number of retrieved cases, $similarity(c_i^I, c^R)$ is the similarity between the i -th retrieved case and the new case, and $s_{i,j}$ is the occurrence of medicine d_j being prescribed in i -th case c_i . $s_{i,j}$ is determined by the following equation:

As a result, a unique medicines list with weightings is generated as the suggestions of CBR.

Input: Examination data of disease determined by the GP

Output: A set of medicines in a ranking list

Preprocessing

Set the threshold γ as the maximum number of cases retrieved

Set the weightings w_i for each i th feature

Case retrieval algorithm

Do while (a new case is ready)

Trigger Similarity Analysis

 Compute similarity for each cases in the knowledge base

End Trigger

Sort the cases by their similarities in descending order

Extract the first γ most similar cases

Extract the unique medicines list from the retrieved cases

<p>Trigger Weighting Assignment Compute the weights for each medicine in the unique medicines list</p> <p>End Trigger Sort the unique medicines list by their weights in descending order</p> <p>End Do Report the results</p>
--

Figure 2: Algorithm of cases retrieval in RACER

3.2 Association rules mining in RACER

A standard association rule consisting of an antecedent (i.e. X) and consequent (i.e. Y) is implicated as follow:

$$X \Rightarrow Y \text{ where } X, Y \subset I \text{ is a itemset} \quad (6)$$

In the present study, the association rules mining aims to discover interesting association rules between the medicines and the problem features of the new case by analyzing the previous cases stored in the knowledge base. Thus, X is the set of problem features of the new case, and Y is the set suggested medicines. The interestingness of a rule is measured by its Support (i.e. the probability that the antecedent and consequent occur among cases in the knowledge base) and its Confidence (i.e. the conditional probability that the consequent occurs given the occurrence of the antecedent). A rule is considered as interesting when it satisfies both the minimum thresholds of support and confidence. Support and confidence are determined by equations (7) and (8), respectively.

$$\text{Support}(X \Rightarrow Y) = \frac{\text{Number of cases containing both } X \text{ and } Y}{\text{Total number of cases}} \quad (7)$$

$$\text{Confidence}(X \Rightarrow Y) = \frac{\text{Number of cases containing both } X \text{ and } Y}{\text{Number of cases containing } X} \quad (8)$$

The algorithm of association rules mining is shown in Figure 3. Apriori algorithm (Agrawal and Srikant, 1996) is applied to identify the associations. It is the best-known algorithm to mine association rules. It uses a breadth-first search strategy to counting the support of rules and uses a candidate generation function which exploits the downward closure property of support. It is applied in the present study for speeding up the mining process. Similar to the consolidated of similar cases in CBR, the mined rules are consolidated to extract a list of unique medicines. The weightings of the medicines in the list are determined by the maximum confidence of the rules associated with the corresponding medicines, which is shown in equation (9):

$$W_j^{arm} = \text{Max}\{\text{Confidence}(a_1 \Rightarrow d_j), \text{Confidence}(a_2 \Rightarrow d_j), \dots, \text{Confidence}(a_m \Rightarrow d_j)\} \quad (9)$$

where W_j^{arm} is the weighting of medicine j , a_i is the i -th problem feature, d_j is the j -th medicine, and m is the number of problem features.

<p>Input: Examination data of disease determined by the GP</p> <p>Output: A set of medicines in a ranking list</p>
--

Preprocessing

Set the minimal support α and minimal confidence β

Association rules mining algorithm

Do while (a new case is ready)

Trigger Apriori algorithm

Measure the support of the features of the new case

Remove the features that do not satisfy α

Measure the support of the medicines

Remove the medicines that do not satisfy α

Trigger Rule Extraction

Associate filtered medicines with the filtered features

Measure the support and confidence of the association rules

Remove the rules that do not satisfy α and β

End Trigger

End Trigger

Extract the unique medicines list from the associated rules

Trigger Weighting Assignment

Compute the weights for each medicine in the unique medicines list

End Trigger

Sort the unique medicines list by their weights in descending order

End Do

Report the results

Figure 3: Algorithm of association rules mining in RACER

3.3 Suggestions combination

The algorithm of combining the suggestions of CBR and that of association rules mining is shown in Figure 4. Before combining the suggestions, the weightings of the suggestions are needed to be normalized by equation (10):

$$N_j = \frac{W_j}{\sum_{i=1}^n W_i} \quad (10)$$

where N_j and W_j are the normalized weighting and suggested weighting of medicine j of CBR (or association rules mining) respectively, and n is the number of medicines in the suggested medicines list of CBR (or association rules mining).

A simple rule of combination is proposed to integrate the normalized weightings of CBR and association rules mining into one single solution. The combination method is adapted from the Dempster's rule of combination (Dempster, 1968), which compensates the missing medicines in the solutions of CBR or that of association rules mining, and updates the weightings of the medicines when new evidences are available. The combination weights of the medicines are calculated from the aggregation of normalized weightings of CBR and association rules mining as shown in equation (11):

$$N_i^{com} = \frac{w_i^{cbr} N_i^{cbr} + w_i^{arm} N_i^{arm}}{w_i^{cbr} + w_i^{arm}} \quad (11)$$

where N_i^{com} , N_i^{cbr} , and N_i^{arm} are the combined weighting of medicine i , normalized weighting of CBR of medicine j , and normalized weighting of association rules mining of medicine i , respectively, and w_i^{cbr} and w_i^{arm} are weighting of CBR and association rules mining for combination of medicine i . The final solution is then sorted by the combined weightings of the medicines in descending order.

Input: A set of medicines in a ranking list from CBR and a set of medicines in a ranking list from association rules mining

Output: A set of medicines in a ranking list

Preprocessing

Set the threshold γ as the maximum number of medicines of the output medicines list

Suggestions combination algorithm

Do while (the input is ready)

 Normalize the weighting of medicines list of CBR

 Normalize the weighting of medicines list of association rules mining

 Combine the weighting of medicines lists of CBR and association rules mining

 Sort the unique medicines list by their weights in descending order

 Extract the first γ medicines

End Do

Report the results

Figure 4: Algorithm of suggestion combination in RACER

4. Experiments and results

Figure 5 depicts the experiment setup for measuring the performance of RACER. Real case data is collected from a medical organization, named Humphrey & Partners Medical Services Limited (HPMS), in Hong Kong. HPMS is one of the largest multi-disciplinary medical services providers (in both general practice and specialist services) founded by a team of dedicated medical practitioners. Each patient's medical history, including personal information, information on medical allergy, past visit's diagnosis and therapeutic result, is recorded in a secure Electronic Medical Record (EMR) system. Since the study scope of this paper focuses on GP prescription, therefore only GP-related patient records are retrieved and used in this experiment. In total, 800 cases which ranged from October 2008 to January 2009 are used in this experiment.

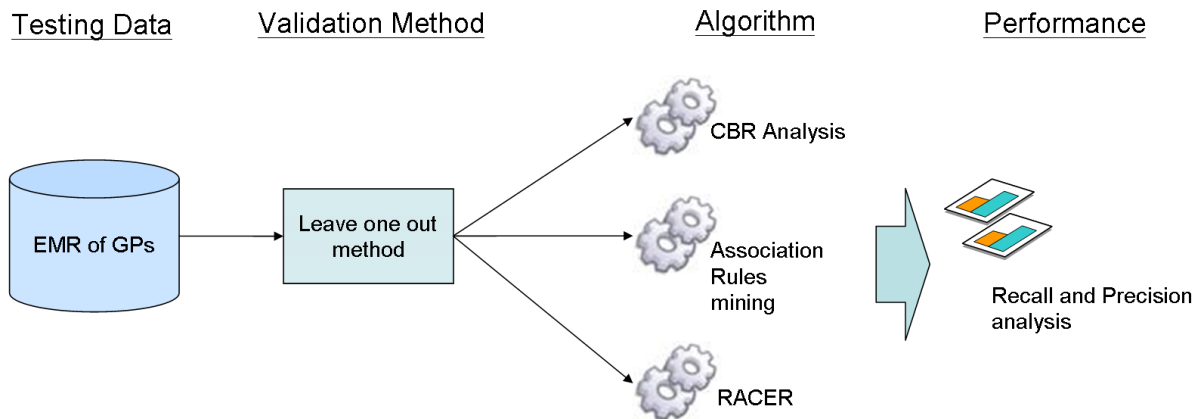


Figure 8: The experiment setup for measuring the performance of RACER

Numerous data are stored in the EMR in which not all the data is useful in supporting the prescription making. After discussing with the medical practitioners of HPMS, four categories of data are used in this experiment. The four categories are:

- Demographic category: Patient’s sex and age
- Allergic category: Patient’s allergy on medication
- Diagnostic category: Symptoms and diagnosis in each case
- Therapeutic category: Medicines prescribed in each case

Table 1 elucidates the features used in each category and illustrates whether the feature is a problem feature or a solution feature.

Table 1: Features of URTI diagnosis that are used in the experiment

Category	Feature	Possible values	Problem feature or solution feature
Demographic	● Sex	Boolean value (M, F)	Problem
	● Age	Single value (Baby, children, youth, adult, elderly)	Problem
Allergic	● Allergy on medication	Single or Multi-value (e.g. NKDA)	Problem
Diagnostic	● Symptoms	Single or Multi-value (e.g. itchy, nasal discharge, nasal congestion, and so on)	Problem
	● Diagnosis	Single or Multi-value (e.g. URTI)	Problem
Therapeutic	● Medicines prescribed	Single or Multi-value (e.g. Dexophen 30mg, Bisolvon Co, Actifed Co, and so on)	Solution

As shown in Figure 8, leave-one-out method is used as the validation method for determining how accurately a learning algorithm will be able to predict data that it was not trained on. Leave-one-out cross validation is useful because it does not waste data. When using the leave-one-out method, the learning algorithm is trained multiple times, using all but one of the training cases. The form of the leave-one-out method is shown as Figure 9:

For $i = 1$ to N (where N is the number of training cases)
 Temporarily remove the i -th case from the training set

Train the learning algorithm on the remaining $N - 1$ points Test the removed case and note the accuracy End For Calculate the overall accuracy over all N cases

Figure 9: Algorithm of suggestion combination in RACER

A series of experiments have been carried out for measuring the performance of RACER. The experiment setting is shown in Table 2. To verify the scalability of RACER, the experiments are carried out with different number of training cases (i.e. 100 to 800 cases with a 100 cases increment). Three different sets of minimum support and confidence are used in the association rules mining (i.e. 0, 0; 0.1, 0.4; and 0.2, 0.6). Three different sets of threshold values for determining the maximum number of retrieved cases are used (i.e. no. of training cases/10, no. of training cases/5, and no. of training cases/2). Three different sets of threshold values for determining the maximum number of suggested medicines are used (i.e. 5, 6 and 7). Equal feature weightings are used in the CBR and RACER analysis. Only the first most similar case is retrieved in the CBR analysis. Recall and precision analysis are applied for the performance measurement by comparing the suggested solutions of the three analysis method against the actual solution. The recall and precision are defined as equations (12) and (13), respectively.

$$recall = \frac{|d_s \cap d_p|}{|d_p|} \quad (12)$$

$$precision = \frac{|d_s \cap d_p|}{|d_s|} \quad (13)$$

where d_s and d_p are the medicine lists of the suggested solution and the actual solution, respectively; $|d_s|$ is the number of medicines in d_s ; $|d_s \cap d_p|$ is the number of medicines jointly appearing in d_s and d_p .

Table 2: Experiment setting for measuring the performance of RACER

Test ID	No. of training cases	Minimum support, Minimum confidence			Maximum no. of retrieved cases			Maximum no. of suggested medicines		
		0, 0	0.1, 0.4	0.2, 0.6	10	20	50	5	6	7
1	100	0, 0	0.1, 0.4	0.2, 0.6	10	20	50	5	6	7
2	200	0, 0	0.1, 0.4	0.2, 0.6	20	40	100	5	6	7
3	300	0, 0	0.1, 0.4	0.2, 0.6	30	60	150	5	6	7
4	400	0, 0	0.1, 0.4	0.2, 0.6	40	80	200	5	6	7
5	500	0, 0	0.1, 0.4	0.2, 0.6	50	100	250	5	6	7
6	600	0, 0	0.1, 0.4	0.2, 0.6	60	120	300	5	6	7
7	700	0, 0	0.1, 0.4	0.2, 0.6	70	140	350	5	6	7
8	800	0, 0	0.1, 0.4	0.2, 0.6	80	160	400	5	6	7

The results are summarized in Figures 10 to 14. Figure 10 shows the precision and recall of the algorithms with the minimum support = 0, minimum confidence = 0, maximum no. of

retrieved cases = no. of training cases/10, and maximum no. of suggested medicines = 5. The figure reveals that RACER outperforms the other two approaches in both recall and precision in this setting. Figures 11 and 12 show the precision and recall of the algorithms with a higher minimum support and confidence. The recall and precision of CBR and RACER remain steady, whereas association rules mining has a higher precision rate but a very low recall. Figures 13 and 14 show the effects on precision and recall of RACER by different sets of maximum number of retrieved cases and different sets of maximum number of suggested medicines, respectively. The results show that the precision and recall remain very stable with the maximum number of retrieved cases and it has a higher recall but lower precision when using a higher number of suggested medicines. For association rules mining, two parameters (i.e. the minimum support and confidence) are needed to be adjusted to control the recall and precision. RACER is only required to adjust one single parameter that is the maximum number of suggested medicines. In addition, the meaning of support and confidence is technical and difficult to be understood, while the meaning of maximum number of suggested medicines is much more simple and obvious.

All in all, the results exhibits that the performance of RACER remains very stable by using different sets of parameters. The results are almost the same (i.e. only a few percentage differences) when different settings of parameters are used. It is not necessary to know what the appropriate settings for the RACER are in advance, which makes RACER robust.

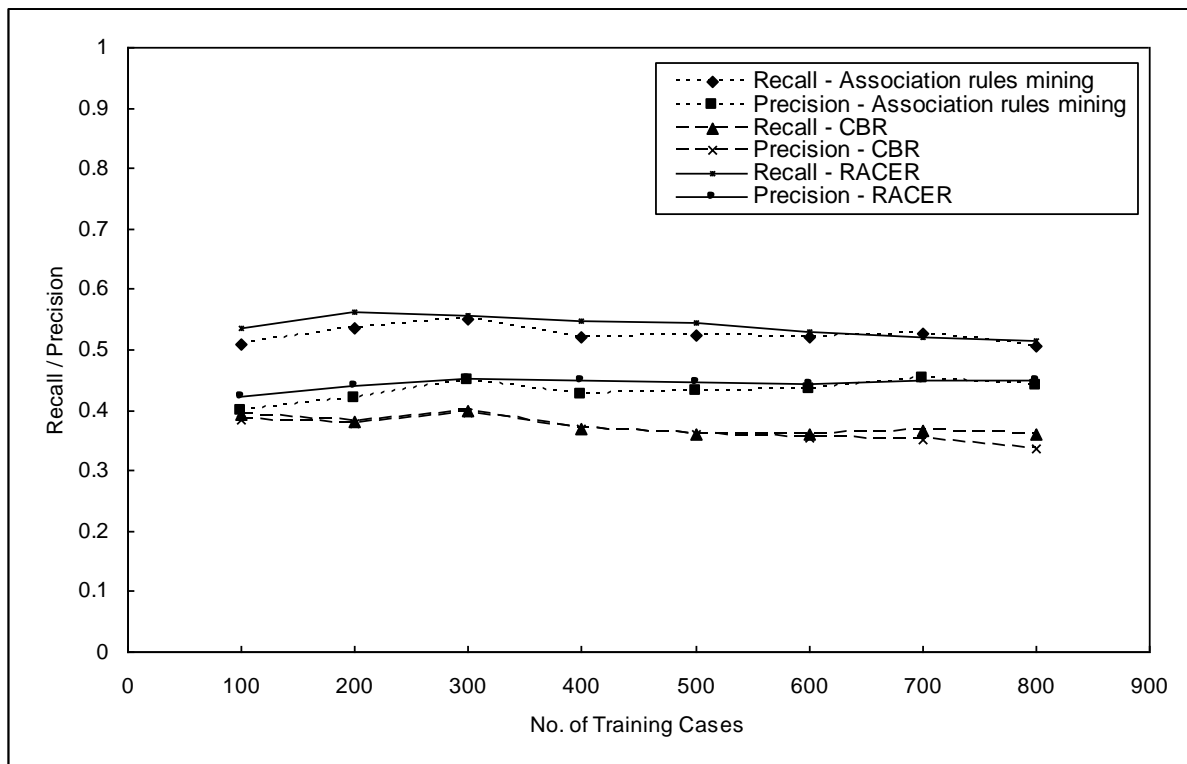


Figure 10: The precision and recall of the algorithms (Minimum support = 0, Minimum confidence = 0, Maximum no. of retrieved cases = no. of training cases/10, Maximum no. of suggested medicines = 5)

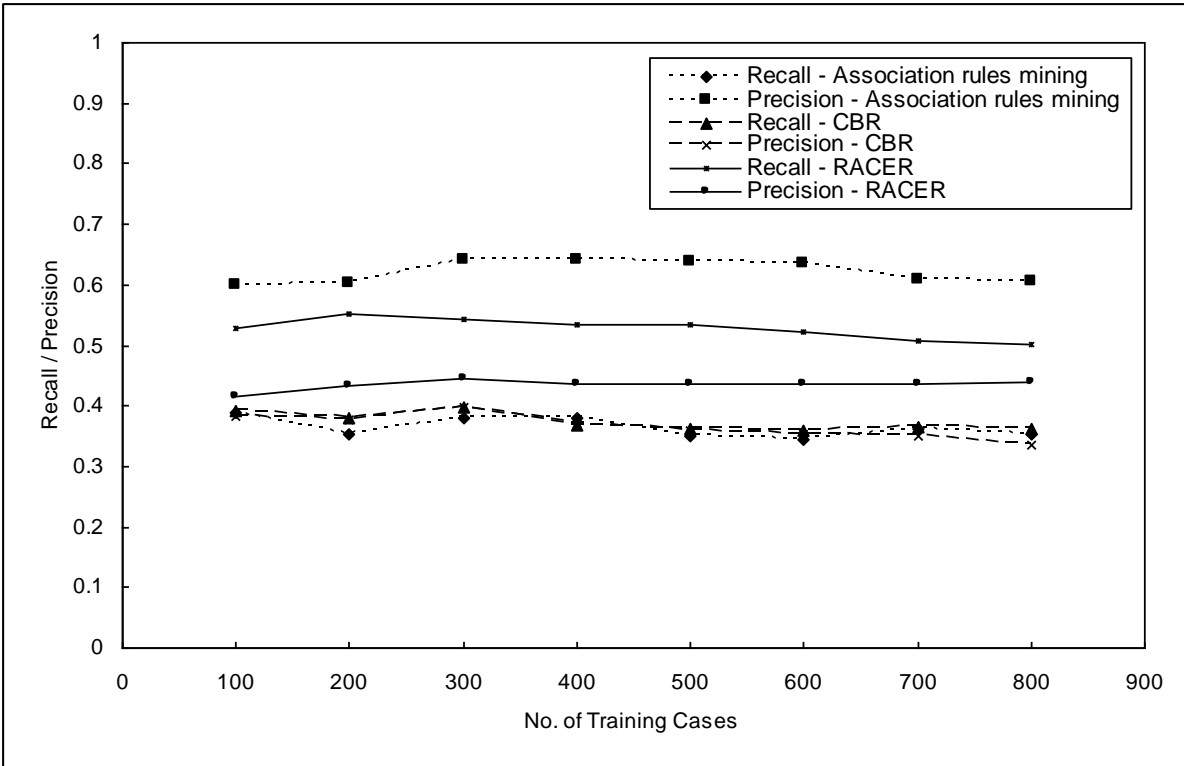


Figure 11: The precision and recall of the algorithms (Minimum support = 0.1, Minimum confidence = 0.4, Maximum no. of retrieved cases = no. of training cases/10, Maximum no. of suggested medicines = 5)

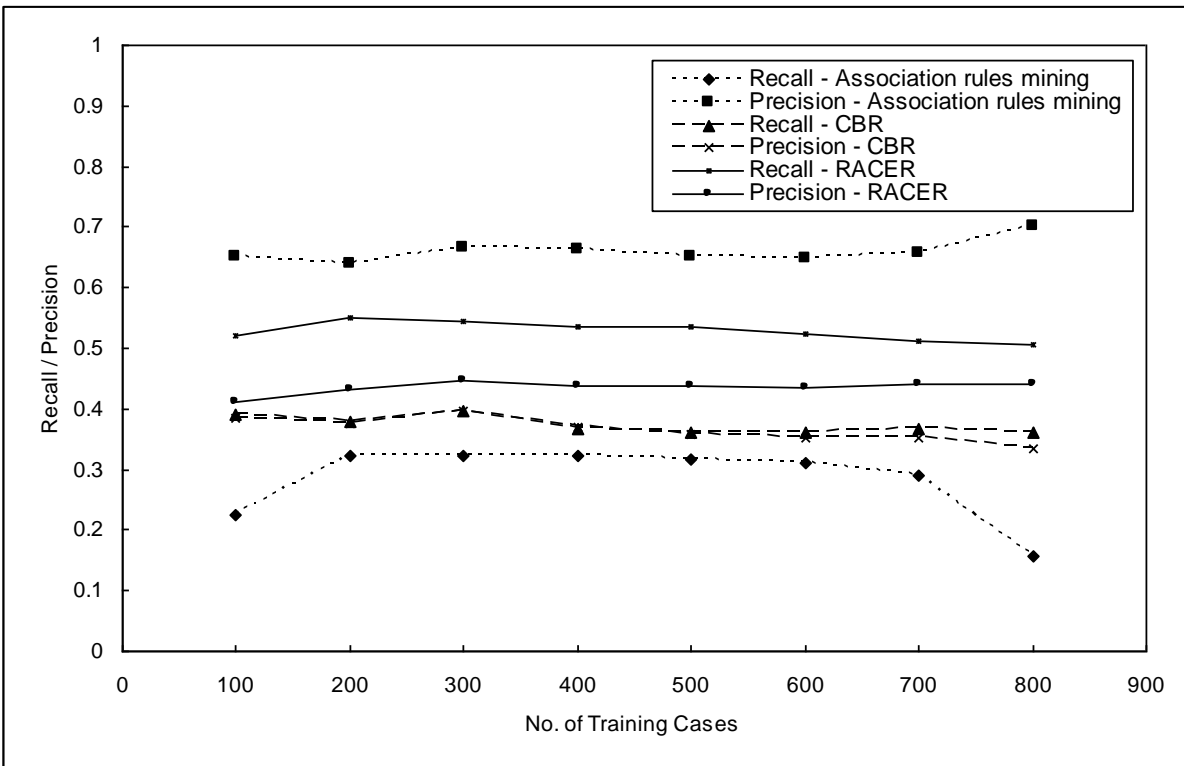


Figure 12: The precision and recall of the algorithms (Minimum support = 0.2, Minimum confidence = 0.6, Maximum no. of retrieved cases = no. of training cases/10, Maximum no. of suggested medicines = 5)

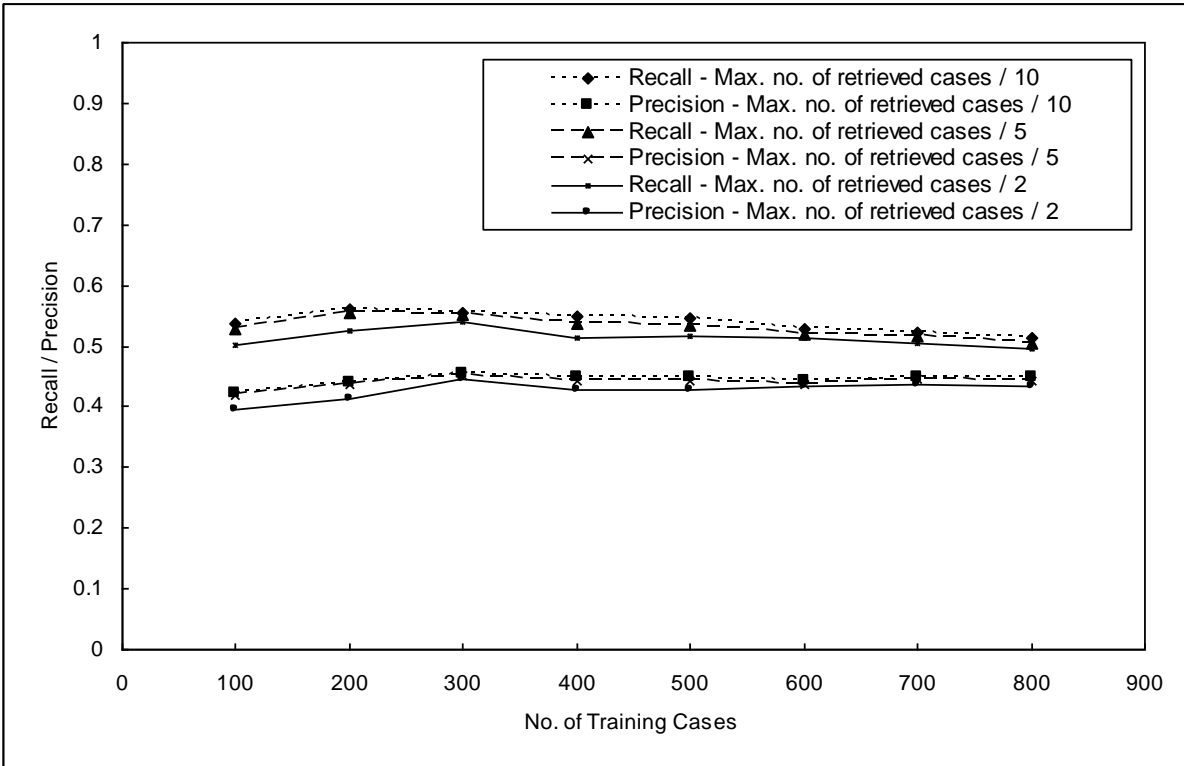


Figure 13: The precision and recall of RACER with different sets of maximum of retrieved cases (Minimum support = 0, Minimum confidence = 0, Maximum no. of suggested medicines = 5)

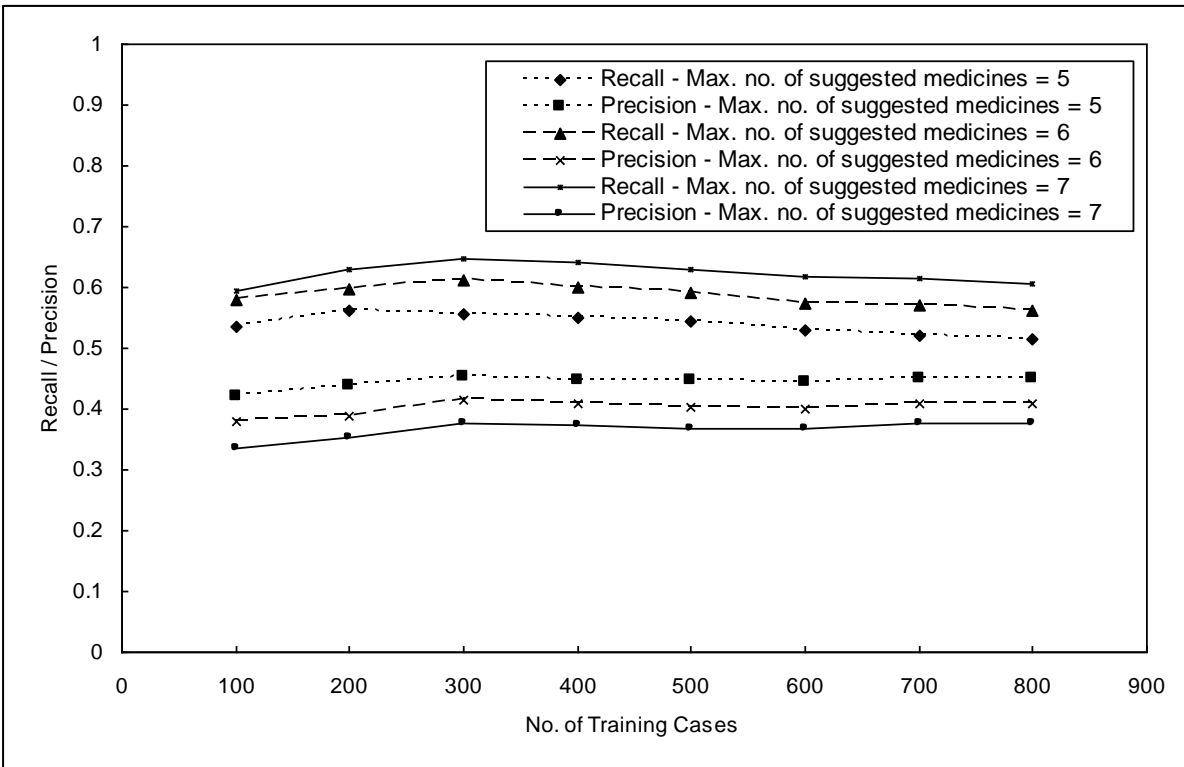


Figure 14: The precision and recall of RACER with different sets of maximum of suggested medicines (Minimum support = 0, Minimum confidence = 0, Maximum no. of retrieved cases = no. of training cases/10)

5. Conclusions

This paper presents a hybrid approach, RACER, which integrates CBR and association rules mining for supporting the prescription making of GPs. By taking the specific experiential knowledge (i.e. from cases) and general knowledge in the medical records (i.e. from the associative relationship between clinical findings and medicines being prescribed) into considerations, the proposed approach is able to leverage and compensate both kinds of knowledge, so as to provide a better decision support. This paper also introduces a new ranking measurement for assigning a likelihood ratio for each medicine extracted from the cases. A series of experiments has been carried out for measuring the performance of RACER against CBR and association rules mining by using real prescription data. The results showed that RACER is outperforming than the other two approaches in various settings. The performance of RACER remains stable by using different sets of parameters, which shows that it is not necessary to know what the appropriate settings for the RACER are in advance. The next research stage will be to select the appropriate features and parameters in order to optimize the accuracy of the algorithm and also to design a user-friendly interface for the GPs to apply RACER in their daily operations.

Acknowledgement

Acknowledgement is given to Dr. C.W. Lo and Miss Maggie Poon for their guidance on issues in clinical coding and general practice medicine in general. Also, the authors would like to express their sincere thanks to the Research Committee of the Hong Kong Polytechnic University for financial support of the research work.

References

- Aamodt, A., & Plaza, E. (1994). Case-based reasoning: foundational issues, methodological variations, and system. *AI Communications*, 7(1), 39-59.
- Abidi, S. (2001). Knowledge management in healthcare: Towards 'knowledge-driven' decision-support systems. *International Journal of Medical Informatics*, 63(1-2), 5-18.
- Agrawal, R., & Strikant, R. (1994). Fast algorithm for mining association rules. In *Proceeding of the 20th International Conference on Very Large Data Bases*, Morgan Kaufmann, Santiago (pp. 487-499).
- Atzmueller, M., Baumeister, J., & Puppe, F. (2003). Evaluation of two strategies for case-based diagnosis handling multiple faults. In *Proceedings of the 2nd German Workshop on Experience Management (GWEM 2003)*, Luzern, Switzerland.
- Banning, M. (2008). A review of clinical decision making: models and current research. *Journal of Clinical Nursing*, 17, 187-195.
- Bichindaritz, I., & Marling, C. (2006). Case-based reasoning in the health sciences: what's next? *Artificial Intelligence In Medicine*, 36(2), 127-135.
- Bradley, C. P. (1991). Decision making and prescribing patterns - a literature review. *Family Practice*, 8, 276-287.
- Bradley, C. P. (1992). Factors which influence the decision whether or not to prescribe: the dilemma facing general practitioners. *British Journal of General Practice*, 42, 454-458.

Carte, J. H. (2004). How EMR software can help prevent medical mistakes, Available at <http://www.acpinternist.org/archives/2004/09/emr.htm>.

Chen, M. M., & Landefield, C. S. (1994). Physicians' behavior and their interactions with drug companies. *Journal of the American Medical Association*, 271, 648-649.

Chien, C. F., & Chen, L. F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems with Applications*, 34(1), 280-290.

Cho, Y. H., Kim, J. K., & Kim, S. H. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert Systems with Applications*, 23(3), 329-342.

Coscelli, A. (2000). The importance of doctors' and patients' preferences in the prescription decision. *Journal of Industrial Economics*, 48, 349-369.

Daengdej, J., Lukose, D., & Murison, R. (1999). Using statistical models and case-based reasoning in claims prediction: experience from a real-world problem. *Knowledge-based Systems*, 12(5-6), 239-245.

Dempster, A. P. (1968). Generalization of Bayesian inference. *Journal of the Royal Statistical Society*, 30, 205-247.

Denig, P., Haaijer-Ruskamp, F. M., & Zijsling, D. H. (1988). How physicians choose drugs. *Social Science & Medicine*, 27(12), 1381-1386.

Deutsch, T., Cramp, D., & Carson, E. (2001). *Decisions, computers, and medicines: the informatics of pharmacotherapy*. New York: Elsevier.

Dussart, C., Pommier, P., Siranyan, V., Grelaud, G., & Dussart, S. (2008). Optimizing clinical practice with case-based reasoning approach. *Journal of Evaluation in Clinical Practice*, 14(5), 718-720.

Galland, L. (1997). *The four pillars of healing*. New York: Random House.

García, M. N. M., Román, I. R., Peñalvo, F. J. G., & Bonilla, M. T. (2008). An association rule mining method for estimating the impact of project management policies on software quality, development time and effort. *Expert Systems with Applications*, 34(1), 522-529.

Garg, A. X., Adhikari, N. K. J., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., Sam, J., & Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *Journal of the American Medical Association*, 293, 1223-1238.

Gill, P., Scrivener, G., Lloyd, D., & Dowell, T. (1995). The effect of patient ethnicity on prescribing rates. *Health Trends*, 4, 111-114.

Greenhalgh, T., & Gill, P. (1997). Pressure to prescribe. *British Medical Journal*, 315, 1482-1483.

- Hammond, K. J. (1989). *Case-based planning*. Academic Press Inc.
- Hartge, F., Wetter, T., & Haefeli, W. E. (2006). A similarity measure for case based reasoning modeling with temporal abstraction based on cross-correlation. *Computer Methods and Programs in Biomedicine*, 81(1), 41-48.
- Huang, M. J., Chen, M. Y., & Lee, S. C. (2007). Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. *Expert Systems with Applications*, 32(3), 856-867.
- Hung, S. Y., Yen, D., & Wang, H. Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3), 515-524.
- Institute of Medicine. (2006). *Preventing Medication Errors: Quality Chasm Series*.
- Jiang, X. R., & Gruenwald, L. (2004). Microarray gene expression data association rules mining based on BSC-tree and FIS-tree. *Data & Knowledge Engineering*, 53(1), 3-29.
- Kotsiantis, S., & Kanellopoulos, D. (2006). Association rules mining: A recent overview. *GESTS International Transactions on Computer Science and Engineering*, 32(1), 71-82.
- Kumar, K. A., Singh, Y., & Sanyal, S. (2009). Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU. *Expert Systems with Applications*, 36(1), 65-71.
- Kuo, R. J., Lin, S. Y., & Shih, C. W. (2006). Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan. *Expert Systems with Applications*, 33(3), 794-808.
- Lee, C. H., Kim, Y. H., & Rhee, P. K. (2001). Web personalization expert with combining collaborative filtering and association rule mining technique *Expert Systems with Applications*, 21(3), 131-137.
- Lundin, D. (2000). Moral hazard in physician prescription behavior. *Journal of Health Economics*, 19(5), 639-662.
- Marling, C. R., & Whitehouse, P. J. (2001). Case-based reasoning in the care of Alzheimer's disease patients. In D. W. Aha, & I. Watson (Eds), *Proceedings of the 4th international conference on case-based reasoning, ICCBR-01*, Springer, Berlin (pp. 702-715).
- Mitchell, T. M. (1997). *Machine learning*. Massachusetts: McGraw-Hill.
- Montani, S., Magni, P., Bellazzi, R., Larizza, C., Roudsari, A. V., & Carson, E. R. (2003). Integrating model-based decision support in a multi-modal reasoning system for managing type 1 diabetic patients. *Artificial Intelligence in Medicine*, 29(1-2), 131-151.
- Oren, E., Shaffer, F., & Guglielmo, B. (2003). Impact of emerging technologies on medication errors and adverse drug events. *American Journal of Health-System Pharmacy*, 60, 1447-1458.

- O'Sullivan, D., McLoughlin, E., Bertolotto, M., & Wilson, D. C. (2007). Mobile case-based decision support for intelligent patient knowledge management. *Health Informatics Journal*, 13(3), 179-193.
- Park, Y. J., Choi, E., & Park, S. H. (2009). Two-step filtering datamining method integrating case-based reasoning and rule induction. *Expert Systems with Applications*, 36(1), 861-871.
- Quaglioni, S., Bellazzi, R., Berzuini, C., Stefanelli, M., & Barosi, G. (1992). Hybrid knowledge-based systems for therapy planning. *Artificial Intelligence in Medicine*, 4(3), 207-226.
- Rossille, D., Laurent, J. F., & Burgun, A. (2005). Modeling a decision-support system for oncology using rule-based and case-based reasoning methodologies. *International Journal of Medical Informatics*, 74(2-4), 299–306.
- Schmidt, R., Montani, S., Bellazzi, R., Portinale, L., & Gierl, L., Case-based reasoning for medical knowledge-based systems. *International Journal of Medical Informatics*, 64(2-3), 355–367.
- Shan, Y., Jeacocke, D., Murray, D. W., & Sutinen, A. (2008). Mining medical specialist billing patterns for health service management. In J. F. Roddick, J. Li, P. Christen, & P. Kennedy (Eds.), *Proceedings of the Conferences in Research and Practice in Information Technology*.
- Shi, W., & Branden, J. A. (2005). How to combine CBR and RBR for diagnosing multiple medical disorder cases. In *Proceedings of the 6th international conference on case-based reasoning*, Chicago, IL, USA (pp. 477–491).
- Sun, Z., & Finnie, G. R. (2004). *Intelligent techniques in E-Commerce: a case based reasoning perspective*. New York: Springer-Verlag.
- Wazana, A. (2000). Physicians and the pharmaceutical industry: is a gift ever just a gift? *Journal of the American Medical Association*, 283(3), 373-380.
- Yusof, M. M., & Buckingham, C. D. (2009). Medical case-based reasoning: A review of retrieving, matching and adaptation processes in recent systems. In M. H. Hamza (Eds.), *Proceedings of the Artificial Intelligence and Applications (AIA 2009)*, Innsbruck, Austria.
- Zhuang, Z. Y., Churilov, L., Burstein, F., & Sikaris, K. (2009). Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. *European Journal of Operational Research*, 195(3), 662-675.