CASESIAN: A Knowledge-based System Using Statistical and Experiential Perspectives for Improving the Knowledge Sharing in the Medical Prescription Process

S.L. Ting, S.K. Kwok*, A.H.C. Tsang, W.B. Lee
Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

*Correspondence author
S.K. Kwok
Department of Industrial and Systems Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
Tel: 852-2766-6578
Fax: 852-2774-9308
Email: mfskkwok@inet.polyu.edu.hk

Summary

Objectives: Knowledge sharing is crucial for better patient care in the healthcare industry, but it is challenging for physicians to exchange their clinical insights and practice experiences, particularly with regard to the issuing of prescriptions for medicine. The aim of our study is to facilitate knowledge sharing and information exchange in this area by means of a knowledge-based system.

Methods: We propose a knowledge-based system, CASESIAN, to automatically model each physician's prescription experience. This is done by collecting as many as possible instances of when the physician has issued a prescription. These occasions will be analyzed from a statistical perspective to form a reciprocal interactive knowledge sharing process for the issuing of medical prescriptions which we will call the prescription process. With the help of the prescription data in medical organizations, the knowledge-based system employs the Bayesian Theorem to correlate the experience of peers in order to evaluate individual prescription knowledge as retrieved through the Case-based Reasoning technique. In addition, a system prototype was implemented in a Hong Kong medical organization to evaluate the feasibility of such an approach.

Results: Our evaluation indicates that there is a significant improvement in knowledge sharing after the adoption of the system. CASESIAN obtains a higher rating in both recall and precision measurement when compared to traditional knowledge-based system. In particular, its information retrieval is much stronger than the baseline in around 40%. Furthermore, regarding the result of the interviews, physicians agree that the system can improve the storing and sharing of medical prescription knowledge.

Conclusion: Compared with conventional knowledge-based systems, CASESIAN provides more peer-based evidence that can enhance the learning and sharing process, transforming it from a single loop to a double loop. The quality of shared knowledge is, in addition, more objective and less biased.

Keywords: Bayesian Theorem, Case-based Reasoning, Knowledge-based System, Knowledge Sharing, Medical Prescription

1. Introduction

Healthcare knowledge sharing is a crucial and promising vehicle for facilitating safer, higher quality patient care [1-3]. According to [4], knowledge sharing can provide efficient and focused assessment, either by directly navigating users to the knowledge artifacts or indirectly providing peer-comparisons to help discover the relevant knowledge artifacts. The sharing of such knowledge resources is particularly valuable for physicians in the area of medical prescription, when they encounter complex and potential drug interactions. It is particular true for the medical organizations which consists of many physicians that specialized in different medical professions. In other words, if a physician practices drug therapy which is not his/her specialty; (s)he cannot provide the required standard of care [5-6].

Numerous methods have been investigated for improving the knowledge sharing process in medical prescription [7]. In essence, the sharing platform is mostly represented in the form of research articles, forum discussions and clinical guidelines. Jabr [8] argues that this kind of knowledge-sharing process is not well constructed and that problems are still mounting. One challenge for physicians is the limited time they have available for acquiring the relevant knowledge because of the demanding nature of their work and the speed and quality of the transfer process. This acknowledges that there is a pressing and burning need to develop a new approach to facilitate time-efficient, effective knowledge sharing and information exchange for medical prescription.

As a backdrop to the above mentioned sharing considerations, knowledge-based systems (KBS) have gained increased attention in recent years both in healthcare knowledge management and in medical prescription. Most KBSs employ artificial intelligence techniques to develop a knowledge-centric healthcare system for gathering prescriptions in a knowledge repository and disseminating the knowledge to all parties for reuse and problem solving [9-11]. Case-based reasoning (CBR) is one of the most prevalent knowledge extraction methods used in developing KBSs because it has a stronger explanation capability than other techniques like neural networks [12]. Related work on using CBR enables physicians to share past experiences stored in the knowledge base to encounter new situations. Generally, physicians have developed their own prescription style and behaviors based on their knowledge and experience. In this situation, the problem solving is presented in a single looping process that generates a solution prescribed by the physician himself/herself previously. As a means of knowledge sharing, this approach is not suitable because physicians do not share what they know with other parties. Even though each physician has the knowledge to make the prescription, it is important for them to learn from others' experiences as well. Thus, an external method is required to enhance the sharing process between physicians, thereby supporting the peer-based comparison determined in statistical perspectives.

In this paper, we intend to construct a KBS for knowledge sharing in the medical prescription process. We propose a state-of-the-art system, CASESIAN (by combining **CASE**-based reasoning and the BAYE**SIAN** theorem), that imitates physicians' prescription decision through reference to electronic medical records (EMR) database and correlate the experience of peers with the same diagnostic information during the prescription process. The system can further be used as a module to assemble other medical diagnostic systems to enhance knowledge sharing, as well as decision making in the prescription process.

2. Outline of CASESIAN

Fig.1 shows the outline of CASESIAN developed in this study. By retrieving the knowledge from numerous medical records, it is possible to derive all prescription decisions. The main purpose of our KBS is to enhance the knowledge sharing process between physicians by taking into consideration their peer-based experiences; therefore, we adopt the CBR technique as the basis of the KBS and employ the Bayesian theorem (BT) for supporting and benchmarking the result from CBR. With the support of CBR, the most relevant cases can be retrieved and reused according to the highest degree of similarity, while the BT allows us to capture and calculate the prescription decisions based on the diagnostic experience of all the physicians within the organization.

2.1 Data Pre-processing

In our KBS, all the patient demographic, treatment and administrative data (including age, sex, treatment date, symptoms, diagnosis, allergies, significant past history, referrals, payment, number and duration of sick leaves, doctor's name, clinic name, and drugs prescribed) are consolidated and stored in the data warehouse of the information system. For facilitating the knowledge sharing in the medical prescription process, a pre-processing method is used to index and extract the specific information from the data warehouse. All the irrelevant information is filtered out, while the remaining information is structured as a data mart for supporting the data analysis from both the statistical and experiential perspectives.

2.2 Construction of Experiential and Statistical Perspectives

As discussed in [9], experience plays a very important role in improving the performance of physicians in the medical prescription process. The selection of each drug dose is heavily reliant on the physician's knowledge of drugs and his/her skill of diagnosis. Even encountering the same disease, the result of prescription is different for different physicians. This is particular true as these medical experts are come from different educational background and have different diagnostic experiences. Therefore, solely considering the experiential perspective is inadequate for providing a better quality of prescription.

In CASESIAN, the statistical perspective is taken into account as an evaluation factor to enhance the prescription result in each diagnostic process. Fig. 2 shows the elements (representing statistical and experiential perspectives) employed in our system. Each physician's prescription decision is represented by the small circle where particular patient past cases are stored inside. Within the medical organization, all these small circles, representing different physicians, are stored in the big rectangle that determines the statistical perspective by pooling the individual experiences. Therefore, this collective wisdom can be shared and transferred through the system as well as provide an evaluation or advisory function to physicians when they make a prescription.

2.3 Experiential Perspectives Modeling by Case-based Reasoning

CBR is a plausible generic model of reasoning based on the view that a significant portion of human problem solving involves recalling prior experiences [13]. CBR has provided an intelligence and cognitive science-based method of adapting the previously experienced and concrete solutions to interpret new situations. According to [14], CBR can be described in a schematic cycle with four central tasks: (1) *retrieve* one or more cases (from the case library) that are similar to the new problems; (2) *reuse* the information or solution in that case; (3) *revise* the proposed case if the new problem does not exactly match the old one; and (4) *retain* the new experience in the case library for future problem solving.

In the case of medical prescription, physicians have a strong tendency to give a similar or even identical dose to that given in past cases. In case of some common diseases (e.g. acute upper respiratory tract infection and gastritis), physicians may reapply previously prescribed solutions in a patient's medical record to a new case if the situation is the same as the previous one. Given this adoption of a prescribed solution based on past knowledge, we can model this situation as a KBS by employing the CBR method. With retrieval of the filtered data from the pre-processing method, the central principle of CBR is to model and retrieve similar prescribed solutions (that is, ones with highest degree of similarity) for the diagnosis determined by the physician. In order to ascertain the case similarity, we used nearest neighbor matching, defined as follows:

$$Similarity = \frac{\sqrt{\sum_{j=1}^{N} (W_j d_{ij})^2}}{\sum_{j=1}^{N} (W_j)^2} \tag{1}$$

where N is the number of inputs, W_j is the weighting of each dimension, d_{ij} is the distance between the ith record's value v_{ij} and the new case's value for the jth field V_j as follows:

$$d_{ij} = \frac{|v_{ij} - V_j|}{|V_j^{max} - V_j^{min}|} \tag{2}$$

where the maximum and minimum values of each field may either be previously specified or determined during index construction

The development of KBSs through CBR has recently been successfully applied in the medical domain for the purpose of diagnostics, classification and treatment planning. Huang et al. [15] propose a model of a chronic disease's prognosis and diagnosis (CDPD) system by integrating data mining and CBR to support chronic disease treatment. Khan and Hoffmann [16] present an approach that allows for the automatic construction of a menu which is strongly tailored to the individual requirements and food preferences of a client. However, in the domain of medical prescription, the complex nature of drug information and interaction makes it challenging to adopt a KBS to assist in the prescription process. Cordier et al. [17] discuss the fact that the retrieved solution may be inappropriate because of insufficient knowledge.

Theoretically, it is possible through CBR to acquire the physician's prescription practices and style, which are formed by their own experiences and knowledge, and hence design a KBS; but it is too complicated, almost impossible in fact, to access the effectiveness of the drug prescribed by physicians because they employ different sets of knowledge in their decision making. In contrast, we may consider the use of external evidence to support personal experience and judgment. One solution is to focus on using the evidences based on large group of peer physicians to evaluate the selection of each medicine for particular diagnoses and symptoms. Therefore, in this paper, we attempt to further extend the KBS by enhancing the knowledge sharing aspect of the medical prescription process by taking statistical perspectives into account.

2.4 Statistical Perspectives Modeling with the Bayesian Theorem

The Bayesian Theorem was originally stated by Thomas Bayes (1702-1761) as a probability theory that could be used to calculate the statistical probability of a proposition based on the original probability plus new relevant factors [18]. This probabilistic approach has been widely applied in the medical domain in areas such as diagnosis classification, drug testing and advice about medicine [19-21]. As far as medical prescription is concerned, Warren et al. [22] developed an anticipative data entry interface (Mediface) to intelligently generate 'hot lists' (by learning through probabilistic models) for general practitioners to reduce the time required for selecting the relevant medicines for the patient. He and his research group demonstrate that applying a probabilistic approach is superior in terms of optimizing medicine selection compared with other statistical methods like multiple linear regression and discriminant analysis models [23]. Mathematically, BT can be defined in the form of a conditional probability which is expressed as follows:

$$P(B|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)} \tag{3}$$

where A and B are two independent events and $P(B_i) > 0$ (i = 1,2,…,n) and $P(A) > 0$

In CASESIAN, we attempt to provide peer-based evidence to enhance knowledge sharing in the medical prescription process. The approach is based on the usual assumptions of the independence Bayesian framework [24]. It automatically learns from the data captured in the diagnostic cycle of selecting a target drug event and hence updates the probabilities in the light of the new evidence. In terms of medical prescription by giving the situation of the problem (such as patient symptoms and diagnosis), Eq. (3) can be rewritten as follows:

$$P(Drug|Evidence) = \frac{P(Drug_i)P(Evidence|Drug_i)}{\sum_{j=1}^{n} P(Drug_j)P(Evidence|Drug_j)} \tag{4}$$

where $Drug$ is the particular drug selected by physician and $Evidence$ is the factors affecting the prescription result (such as patient's symptoms or diagnosis)

According to this process, BT can provide better understanding of the problem in hand by pooling the diagnostic experience of many physicians. Compared with the experiential perspective, BT focuses on the interaction between the physicians and the factors affecting the prescription process. Here, the probabilities in the conditional probability table are learnt automatically from the data stored in the databases. Using computerized electronic medical records to store the information, each visit case (that is, one that includes both the problems and the solution) is segmented into various parts and hence associated with the drug prescribed as a reference. To further illustrate the learning logic in the KBS, consider selection of medicine in treating an upper respiratory tract infection (URTI) as an example. A 15-year-old girl with asthma visits the physician. Patient records indicate that she visited the same physician last month. After diagnosis, the physician discovered that the girl has a fever and a cough as well. At this stage the physician would like to determine whether the past medicine can be prescribed again to treat this case. In this case, the $Evidence$ is defined as:

- Diagnosis = URTI;
- Background = Asthma;
- Symptom = Fever;
- Symptom = Cough;
- Age = Young;
- Sex = F

With specifying the $Evidence$, a set of probable consequent class $\{drug_1, drug_2, \ldots, drug_k\}$ where $drug_i\ (i = 1, \ldots, k)$ can be generated and retrieved for peer-based comparison. This, in turn, provides peer-based evidence and wider professional prescription practice, as opposed to the limit of an individual practice.

2.5 Integrated approach

The application of CBR alone in designing KBS cannot fully achieve the aims for knowledge sharing in the medical prescription process. In the existing solution retrieval design, it is common to see that CBR is a single looping process for learning prescription decision, which is based on the individual physician's knowledge and experience. However, this approach degrades its functionality when the physician meets the patient who has not been visited to the clinic or the physician do not have much knowledge in prescription the diagnosis that he is not familiar. To cope with such situation, it is therefore suggested to consider the peer-based decision by means of statistical perspective to enhance knowledge sharing. The natural proposition, then, is that these two approaches can complement each other to facilitate efficient and effective knowledge sharing and information exchange in medical prescription. As shown in Fig 3, the learning process of existing CBR-based KBS can transform into double loop learning in CASESIAN approach.

The underlying philosophy of the proposed integrated approach, CASESIAN, is to establish a peer-based comparison to benchmark the knowledge repository of past experiences. In particular, once the physician ascertains the patient's clinical information (e.g. age, symptoms and diagnosis) and prescription specifications (e.g. number of days of medication, cost of treatment), the recommended solution can be retrieved and reused through the CBR process, according to the highest degree of similarity between cases. In this way, the experiential prescription data can be modeled and captured. However, to prevent single loop sharing, some CBR parameters are then translated into the evidence of BT to determine whether the medicines retrieved in CBR are in form of statistical perspectives. All these parameters selection in both methods are set by medical experts beforehand. The integration algorithm is described as follow:

Algorithm for retrieving the solution with peer-based evidence

1. Initialize a new medical *case* with patient's clinical information and prescription specifications.
2. Employ CBR process to retrieve the solution in case library based on the highest degree of similarity.
3. Reuse the result generated by CBR and display in the *Case Retrieval* column.
4. Transform the parameters used in CBR into the *evidence(s)* of BT.
5. If there are multiple symptoms or diagnosis, separate them into M x N dimensions (where M is the number of symptoms and N is the number of diagnosis) in the *evidence(s)*.
6. Remove the duplicate medicine(s) generated.
7. Generate the medicine(s) in BT with probability of occurrence according to the *evidence(s)* and display in the *Peer-based Evidence* column.
8. For each medicine generated in BT and if there is same item occurred in the *Case Retrieval* column,
    a. Remove it in the *Peer-based Evidence* column;
    b. List the relevant probability near the medicine.
9. Finalize the prescription by selecting the medicine(s) in *Case Retrieval* column

and *Peer-based Evidence* column.

10. Store the revised prescription decision in case library.

In order to visualize the result, two columns, namely Case Retrieval and Peer-based Evidence, are used to store the suggested medicines in CBR and BT respectively. Then, an “IF-Then” statement is used to match the items in both methods. For example, if Lysozyme 30 mg Tab can be found in both columns, it will be colored and its corresponding possibility of occurrence calculated by BT will be listed near the drug name; whereas the remaining medicines not include in CBR will be listed in Peer-based Evidence column descending. Therefore, such peer-based comparison and sharing can review the logic and rationale behind the solutions to past cases and hence strengthen the experiences and knowledge of physicians when encountering unacquainted or new problems. Since the proposed approach focuses on facilitating knowledge sharing in the medical prescription process, the suggested medicines only serve as advisory information for physicians and the final decision is still relied on their clinical judgment.

3 A Case Study in a Hong Kong Medical Organization

We developed a prototype system to demonstrate the effectiveness of the proposed knowledge sharing system for medical prescription. To validate the feasibility of this solution in an actual operational environment, the system was implemented in a Hong Kong professional multi-disciplinary medical services provider, named Humphrey & Partners Medical Services Limited (HPMS). At HPMS, 10 medical experts work on shift to provide various qualities of medical services to its patients in the four core clinics located in different parts of the city. Since the working hours of physicians are different, they find it difficult to share knowledge with others. As a result, we applied the CASESIAN in the period 1-31 April 2008 to evaluate the effectiveness of the proposed approach. In addition, both quantitative and qualitative measurements were made to compare the performance results with those derived from the existing approach.

3.1 System Implementation

The prototype system was developed from the point of patient registration to the end of medical prescription. Most medical experts were opposed to using information and communication technology in their practices because they did not find the interface to be user-friendly. To cope with this issue, we designed the interface of the prototype on the basis of the paper-based medical patient records, which physicians are more familiar with. Fig. 4 shows the transfer of paper-based medical records to an electronic-based system. Once the physician has identified all the information about the case (e.g. patient's clinical background information, symptoms and diagnosis), the system automatically extracts the relevant results from CBR , allowing the physician to evaluate the solution further by using the statistical result generated by BT (Fig. 5).

3.2 Description of data collected

Every individual medical record in the original data set contains 20 fields relating to three broad categories: patient-related information, diagnostic treatment information and billing information. In this study, only relevant patient-related information and diagnostic treatment information is selected for further analysis.

The selected fields include: patient ID, patient age, patient gender, patient allergies, physician ID, medical record ID, date of service, symptoms, diagnosis (presented as codes) and drugs prescribed. Symptoms are physicians' interpretation to patients' health status in free text format after conducting several focused physical examination to the patient. Figure 6 depicts a typical medical record consisted in this study. Regarding the case of multi-diagnosis (say patient gets two diagnosis in each record), we treat these two diagnoses as an independent variable and hence divide the record into two cases. To obtain a better result in modeling, records in missing values or inconsistent values are deleted. As a result, about 3% of the original records are deleted and 607 medical records are included in the analysis. The attributes and characteristics of each field can be found in Table 1.

Table 2 presents the statistical summary of the pre-processed population data. Patient age is in skewed right distribution, with most patients aged from 25 to 40. Male and female are regularly distributed. On average, physicians usually prescribe four drugs in each diagnosis. Almost 75% of patients have three to five medicines to be prescribed. For the problems distribution, Table 3 presents the 10 most frequent diagnosis experienced by patients.

3.3 Quantitative measurement

To determine the effectiveness of the proposed system, a quantitative measurement was conducted on the basis of two performance measurement criteria - recall rate and precision rate. Details of each criterion are discussed in Table 4. Based on these three criteria, all the medical records (in the dataset) are used for the evaluation. The evaluation is then based on a match between the actual prescription decision by expert groups and the set of medicine(s) generated by the systems. The performance of the CASESIAN is compared with a baseline algorithm which adopts CBR methods.

The results are shown in Fig. 7 to Fig. 10 and in Table 5. The results suggest that CASESIAN gets a higher recall rate (68.09%) when compared with the baseline (27.08%). Since the baseline algorithm generates only the solution from a past medical case, therefore the recall rate is much lower than the proposed method. New cases and not matched cases (i.e. patient does not get the same situation as previous) are the main reason of lower recall rate. Concerning the main purpose of CASESIAN to generate more relevant prescription decision(s) for physicians to encounter different patients' complaints, the higher average precision rate of CASESIAN (37.97%) claims that the medicine(s) generated by the proposed approach is more relevant to the patient's complaint. In other words, the knowledge shared is useful for physicians. In general, CASESIAN successfully associates the medicine(s) with the original prescription, which is promising for its use as the basic domain independent algorithm for serving as an advisory references by collecting the knowledge from peers.

3.4 Qualitative measurement

In evaluating the abovementioned methodology, we adopted a "user-focused" evaluation method [25]. This is because users are the best resources to determine whether the proposed system can satisfy the objective (that is, enhancement of knowledge sharing). Of the various user-focused evaluation methods, conducting interviews with each physician individually was employed. Five general practitioners (GPs) listed in Table 6 were invited to share their comments in the following dimensions:

- What the physicians think about sharing knowledge in medical prescription;
- Whether the use of a KBS helps with the storing and sharing of knowledge; and
- Whether the sharing process can be enhanced by CASESIAN.

The result of the interviews is summarized and presented in Fig. 11. From the results, it is interesting to note that the physicians agree that the system can improve their work in the different dimensions discussed above. In addition, most young physicians (i.e. doctor A and B) report that they welcome CASESIAN since it allows them to acquire more prescription knowledge from their seniors. In particular for the new medicine selection, they commented that more attention has been paid to the peer-based prescription decisions. Although some physicians are refused to use the computerized system as they are not so familiar with general computer skill, they claimed that they can share their prescription decision and experiences to peers interactively. They also commented that they will treat the knowledge retrieved by CASESIAN is a kind of advisory information for them to learn more from a large of peers, especially in the case of encountering unacquainted situations. Although CASESIAN cannot provide the golden standard of prescription and concept of evidence-based medicine (due to the retrieved knowledge does not take any critical examination), one point the physicians all agreed is that the information of CASESIAN is more objective than that in the past knowledge extraction method (e.g. attending seminars).

4 Discussion and conclusion

In this paper, we propose an integrated approach (CASESIAN) that utilizes statistical perspective results in supporting the experiential perspective to enhance knowledge sharing in the medical prescription process. The rationale of integrating BT into CBR is to provide double loop learning that uses peer-based evidence to provide more information about a past solution retrieved in isolation (i.e. single loop learning). Table 7 highlights that, as far as practical aspects of knowledge sharing are concerned, in comparison with the CBR approach alone, CASESIAN presents the advantages of combining the strength and complementing the weakness of conventional CBR-based KBS system.

In the case of medical prescription, physicians rely heavily on their knowledge and experience to select appropriate medicine. As discussed earlier in the paper, it is almost impossible for a physician to utilize only their individual knowledge to consider all the important differences between current and former similar cases. Therefore, it is important to learn from others and consider their peers' experience in making a decision. In particular, some young and inexperienced physicians find the system provides them with better support for their decision making.

One limitation of this study is the small sample size of the physicians using the system. To provide a complete assessment of the system, the results should be examined in combination with external data obtained at other organizations such as hospitals.

In this study, we assume that a prescription is useful when most of the physicians, are seen to make use of it. So we summarize all related prescription information to form peer-based evidence, instead of using individual knowledge. This is a powerful knowledge sharing method that allows for acquiring the knowledge of a large group of physicians and hence model such knowledge through a KBS to support further decision making. These issues constitute interesting and promising directions for future research in how to enhance the quality of knowledge sharing in the decision making context.

Acknowledgments

Acknowledgement is given to Dr. Peter Lo, Dr. Francis Liu, Dr. C.W. Lo and Miss Maggie Poon for their guidance on issues in clinical coding and medical knowledge in general. The authors would also like to express their sincere thanks to the Research Committee of the Hong Kong Polytechnic University for providing the financial support for this research work.

References

1. Abidi S. Knowledge management in healthcare: towards “knowledge-driven” decision-support services. Int J Med Inform. 2001 Aug 1;63(1-2):5-18.
2. Nutkis DS. Webcasting: Knowledge management coming to health care providers. Surgical services management. 1997 November;3(11):18-21.
3. Morrissey J. Principles of knowledge management. Modern Healthcare. 1998;28(7):42.
4. Liebowittz J. Knowledge management handbook. London: CRC Press; 1996.
5. Anis AH, Carruthers SG, Carter AO, Kierulf J. Variability in prescription drug utilization: issues for research. CMAJ. 1996 March 1; 154(5): 635-640.
6. Laurence DR, Bennett PN, Brown MJ. Clinical pharmacology. 8th ed. New York : Churchill Livingstone, 1997.
7. Wickramasinghe N, Gupta JND, Sharma SK. Creating knowledge-based healthcare organizations. Hershey PA: Idea Group Pub., 2005.
8. Jabr NH. Physicians' attitudes towards knowledge transfer and sharing. Competitiveness Review: An International Business Journal incorporating Journal of Global Competitiveness. 2007;17(4):248-60.
9. Van Hyfte D, Van Der Maas A, Tjandra-Maga T, De Vries Robbe P. A formal framework of knowledge to support rational psychoactive drug selection. Artificial Intelligence In Medicine. 2001 June;22(3):261-75.
10. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, Tang PC. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc. 2001;8:527-34.
11. Schmidt R, Montani S, Bellazzi R, Portinale L, Gierl L. Cased-Based Reasoning for medical knowledge-based systems. Int J Med Inform. 2001 December;64(2-3): 355-67.

12. Baesens B, Setiono R, Mues C, Vanthienen J. Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science. 2003 March;49(3):312-29.
13. Kolodner JL. Case-based reasoning. San Mateo, Calif.: Morgan Kaufmann Publishers, 1993.
14. Aamodt A, Plaza E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. Artificial intelligence communication. 1994;7(1):39-59.
15. Huang MJ, Chen MY, Lee SC. Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. Expert Systems with Applications. 2007 April;32(3):856-67.
16. Khan AS, Hoffmann A. Building a case based diet recommendation system without a knowledge engineer. Artificial Intelligence in Medicine. 2003 February; 27(2):155–179.
17. Cordier A, Fuchs B, Lieber J, Mille A. Failure analysis for domain knowledge acquisition in a knowledge-intensive CBR system. In: Weber R, Richter M, editors. Proceedings of the 7th International Conference on Case-Based Reasoning; 2007 Aug 13-16; Belfast, Northern Ireland. Heidelberg: Springer; 2007. p. 463-77.
18. Andrade PJ. Specialized computer support systems for medical diagnosis. Relationship with the Bayes' theorem and with logical diagnostic thinking. Arq Bras Cardiol. 1999 Dec;73(6):537-52.
19. Lian J, Cotrutz C, Xing L. Therapeutic treatment plan optimization with probability density-based dose prescription. Med Phys. 2003 Apr;30(4):655-66.
20. Lee SM, Abbott PA. Bayesian Networks for knowledge discovery in large datasets: basics for nurse researchers. Journal of Biomedical Informatics. 2003 August;36(4):389-99.
21. Park YJ, Kim BC, Chun SH. New knowledge extraction technique using probability for case-based reasoning: application to medical diagnosis. Expert systems. 2006 Feburary;23(1):2-20.
22. Warren JR, Davidovic A, Spenceley S, Bolton P. Mediface: Anticipative data entry interface for general practitioners. In: Calder P, Bruce HT, editors. Proceedings of Australasian Computer Human Interaction Conference. 1998 Nov 30- Dec 4; Adelaide, South Australia., IEEE Computer Society; 1998. P. 192-99.
23. Susan EG, Warren JM. Statistical modelling of general practice medicine for computer assisted data entry in electronic medical record systems. Int J Med Inform. 2000 Jul;57(2-3):77-89.
24. McSherry D. Sequential Diagnosis in the Independence Bayesian Framework.

Soft Computing. 2003 October;8(1):118-25.
25. Davies J, Duke A, Sure Y. OntoShare – An ontology-based knowledge sharing system for virtual communities of practice. Journal of Universal Computer Science. 2004 March;10(3):262-83.

Table captions:

Table 1 Attributes and characteristics of dataset

Field	Field Type	Remark
patient ID	Text	-
patient age	Numeric	Calculated by subtracting date of today with patient date of birth
patient gender	Binary variable	0 – Male; 1 – Female
patient allergies	Binary variable	0 – No; 1 – Yes
pregnant	Binary variable	0 – No; 1 – Yes
physician ID	Text	-
medical record ID	Text	-
number of days of medication	Numeric	-
cost of treatment	Numeric	-
date of service	Date	-
symptoms	Text	Coded by company internal coding schema
diagnosis	Numeric	Coded by ICD-9 codes
drugs prescribed	Numeric	Coded by company internal coding schema

Table 2 Statistical summary of the population data

	Min.	Max.	Mean	St. Dev.
Patient age (full year)	0	80	35.53	16.34
Patient gender	0	1	0.52	-
Number of drug prescribed in each case	0	8	3.55	1.57

Table 3 Distribution of top ten diagnoses in the reference database

Rank	Diagnosis	Number of occurrence
1	Acute upper respiratory tract infection (U.R.T.I.)	288 (47.1%)
2	Others	180 (29.4%)
3	Gastroenteritis	40 (6.5%)
4	Dermatitis	19 (3.1%)
5	Dyspepsia	19 (3.1%)
6	Rhinitis	16 (2.6%)
7	Low Back Pain	12 (2.0%)
8	Urinary tract infection (U.T.I.)	11 (1.8%)
9	Conjunctivitis	10 (1.6%)
10	Skin Allergy	9 (1.5%)

Table 4 Performance measurement criteria used in our study

Criterion	Description	Example
Precision rate	The ratio of the number of correct medicine(s) produced by the system among the total number of medicine(s) generated by the system	Total number of medicine(s): {A,B,C,D,E} Correct medicine(s): {A,B,C,E} Precision rate = 4/5 = 0.8
Recall rate	The ratio of the number of correct medicine(s) produced by the system among the total number of existing relevant medicine(s)	Total number of medicine(s): {A,B,C,D,E} Relevant medicine(s): {A,B} Recall rate = 2/2 = 1

Table 5 Performance measurement criteria used in our study

	Baseline CBR-based KBS	Proposed CASESIAN
Average Precision rate	27.14%	37.97%
Average Recall rate	27.08%	68.09%

Table 6 Characteristics of GPs participated in this study

GP	Year of Experience	Specialty
A	1	Respiratory
B	3	Ear, nose and throat (ENT)
C	10	Pediatrics
D	15	Gynecology
E	30	General surgery

Table 7 Comparison of conventional KBS and CASESIAN

Criteria	CBR-based KBS	CASESIAN
Quality of shared knowledge	More subjective as it is based on individual physician's knowledge and experience	More objective as it is based on large group of physicians
Interactivity	Information is retrieved through physician-patient and physician-diagnosis interaction	Information is retrieved through summarizing the peer evidence
Learning cycle	Mostly single loop but sometimes can be double loop	Double loop
New drug selection	Depend on the physician's knowledge	Take into consideration the peer-based prescription decision to facilitate the own choice

Figure captions:

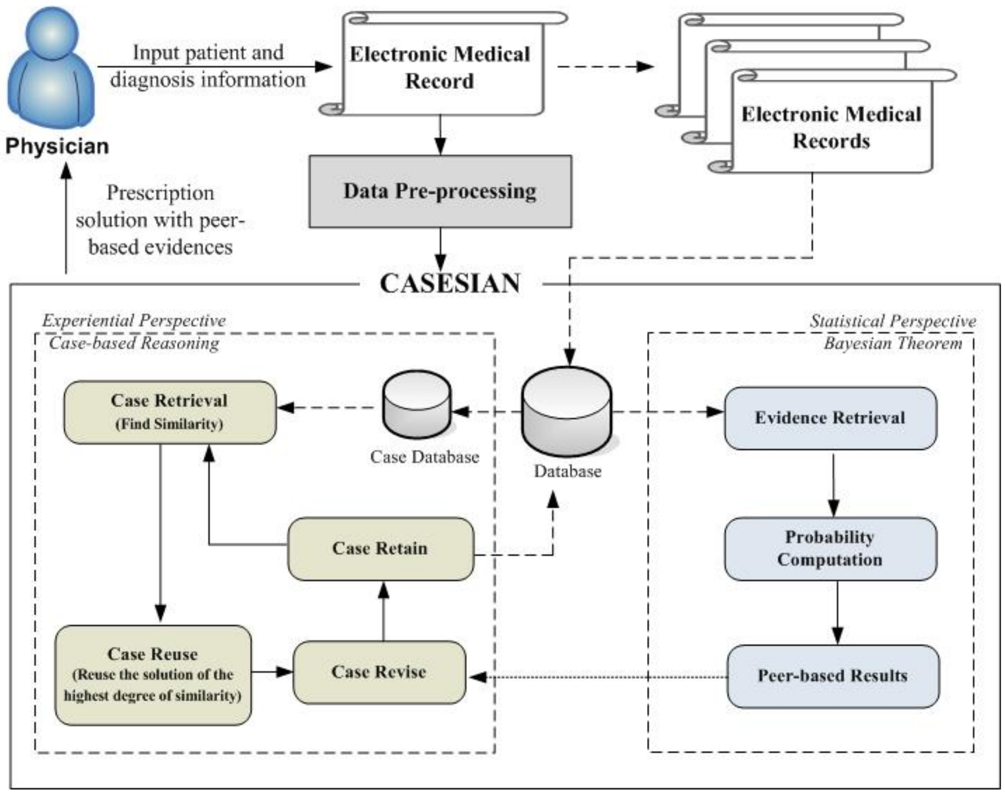

Fig. 1 Architecture of CASESIAN approach

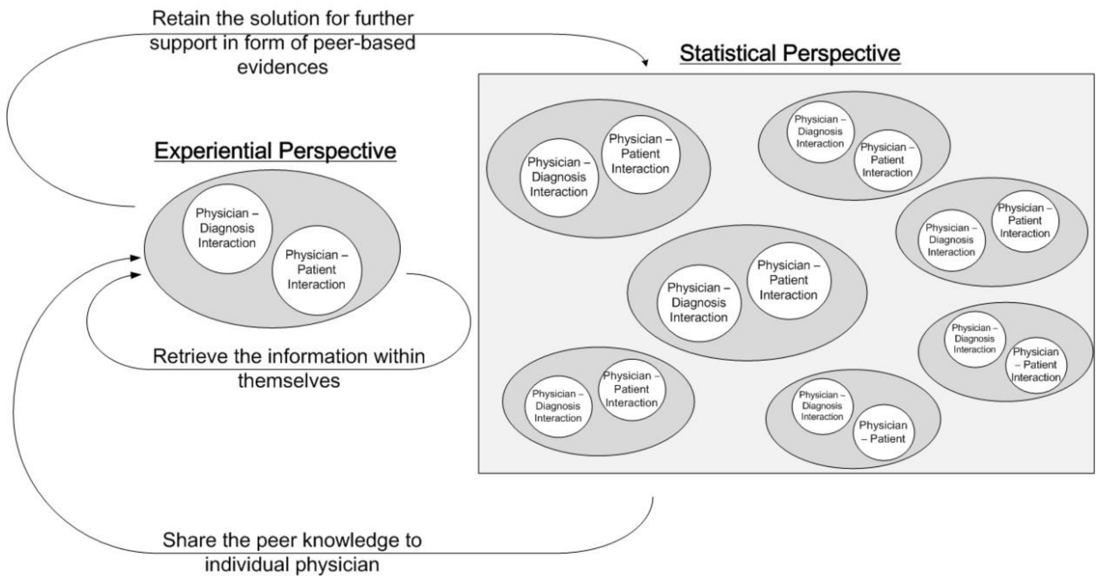

Fig. 2 Ingredients of statistical and experiential perspectives

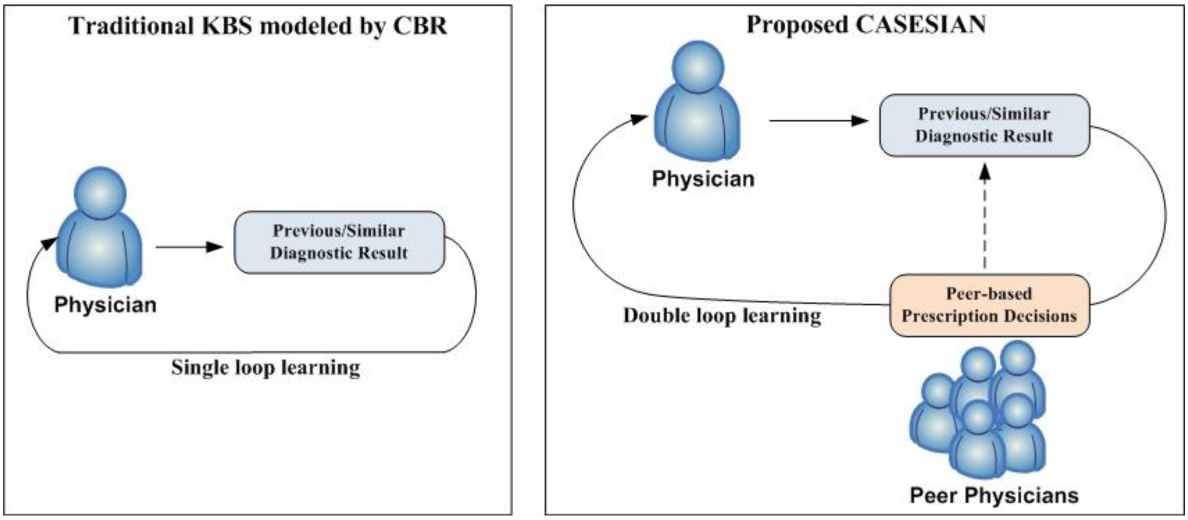

Fig. 3 Learning cycle between traditional KBS and proposed approach

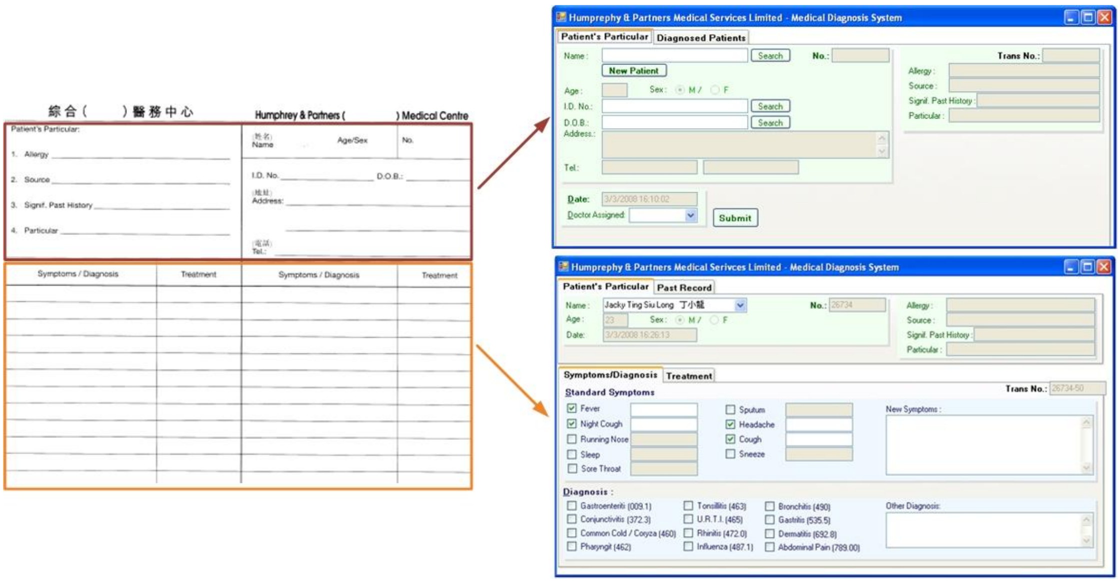

Fig. 4 Paper-based medical record to electronic medical record

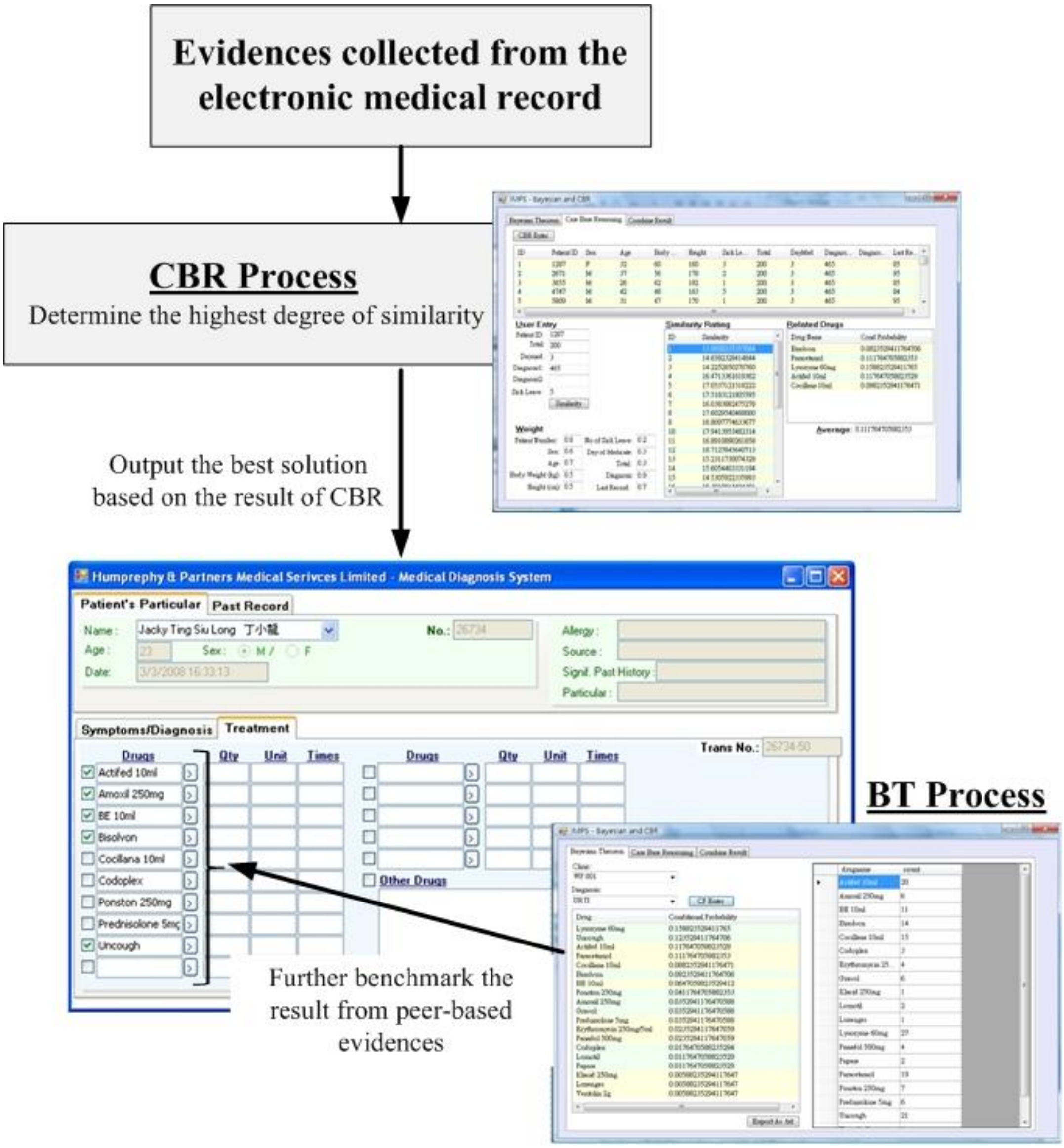

Fig. 5 How to extract the experiential perspective and benchmark the result by statistical perspective

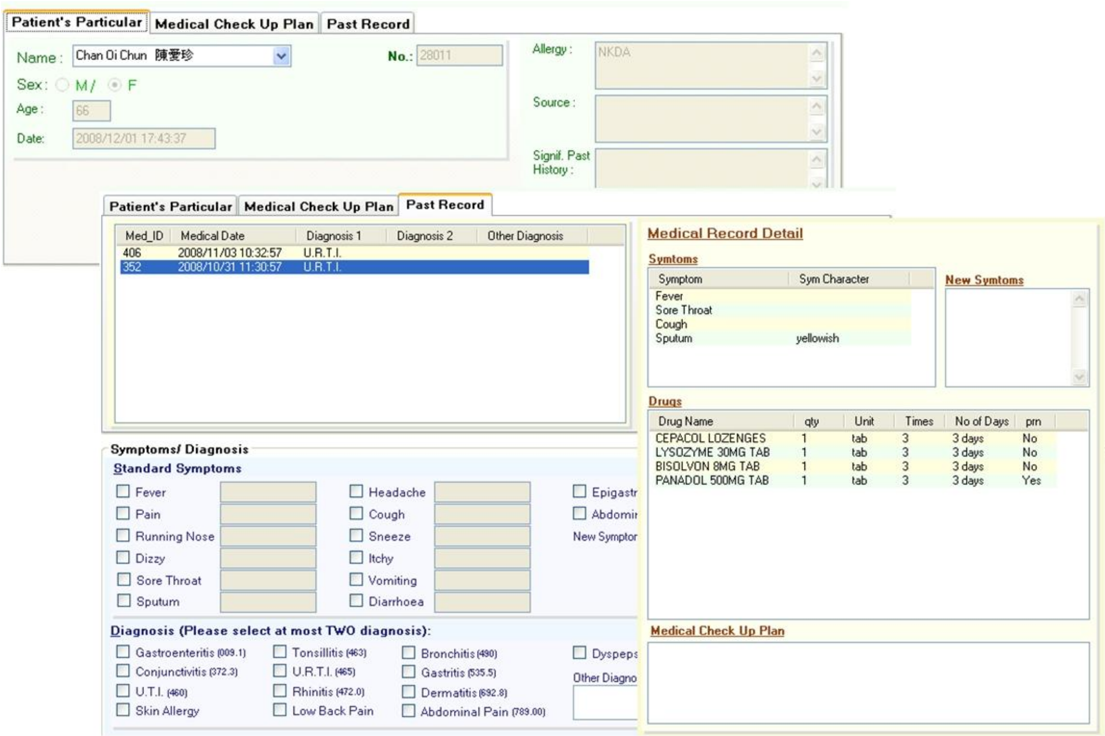

Fig. 6 Patient medical case

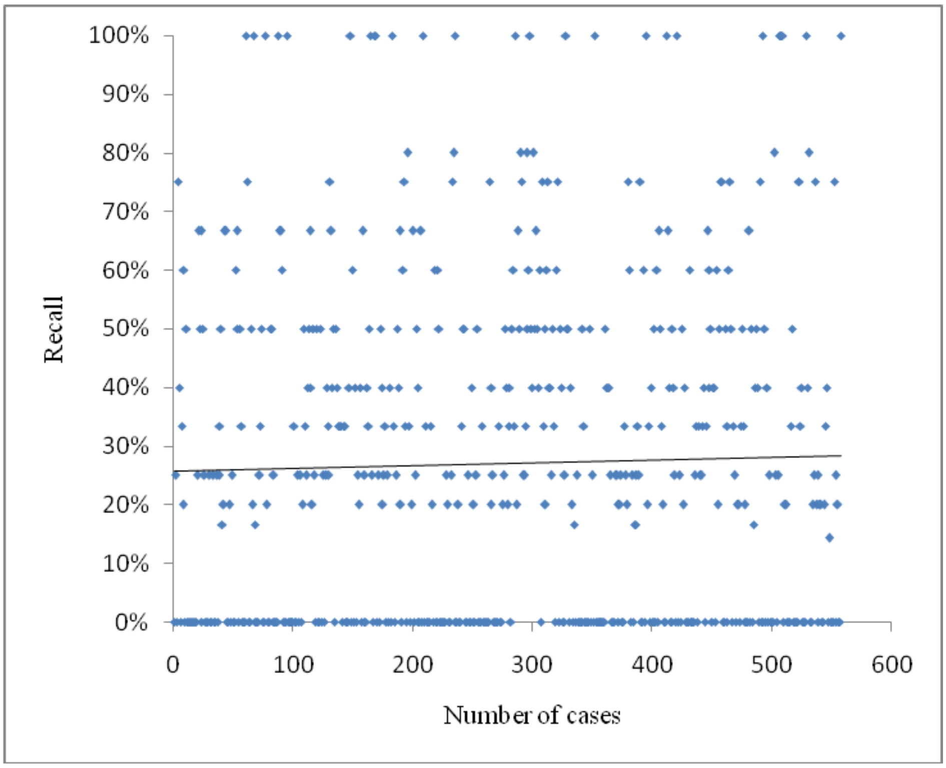

Fig.7 Recall of the baseline (CBR-based KBS)

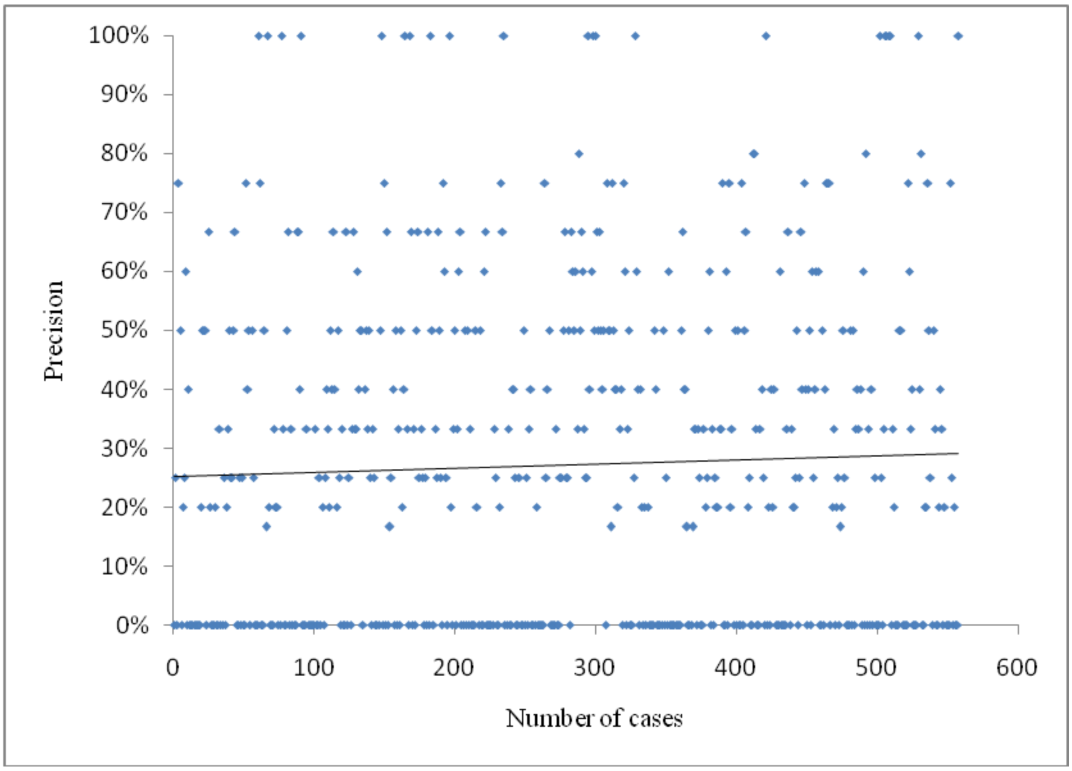

Fig.8 Precision of the baseline (CBR-based KBS)

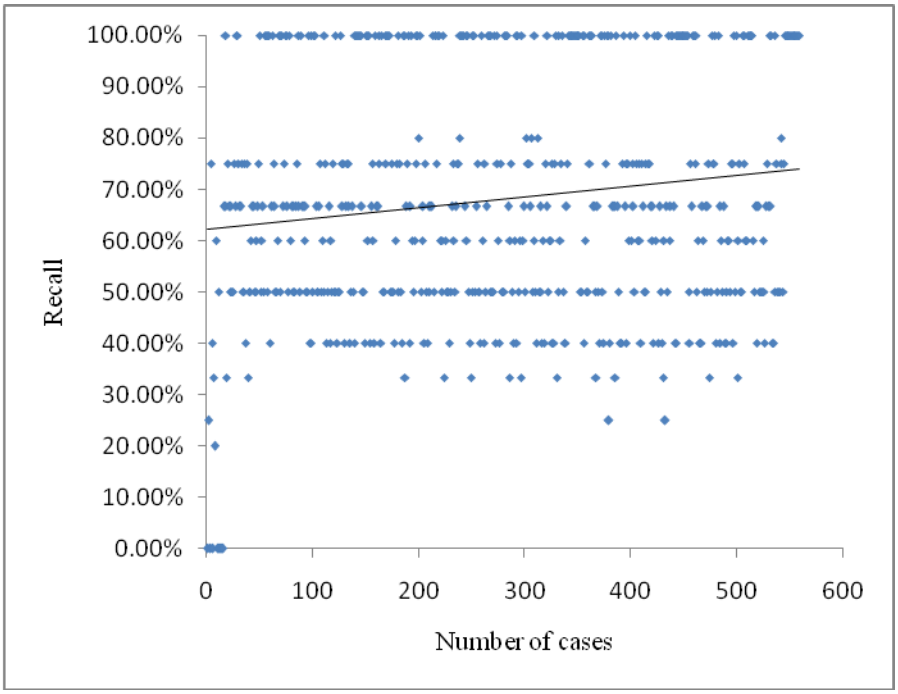

Fig.9 Recall of the proposed method (CASESIAN)

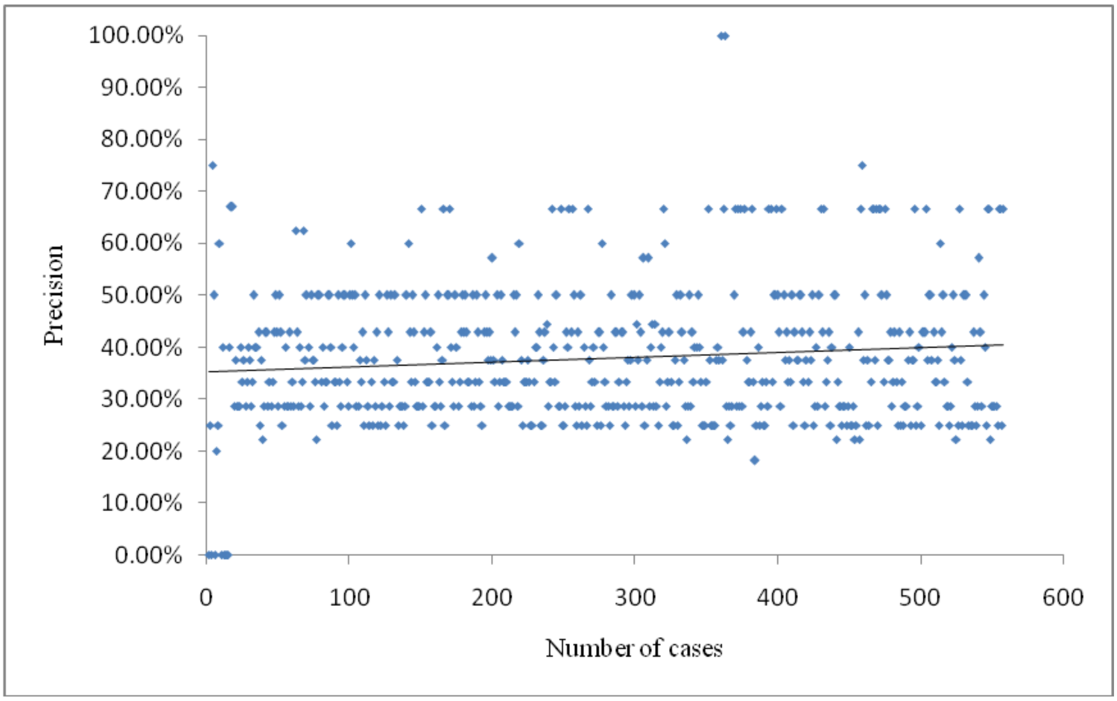

Fig.10 Precision of the proposed method (CASESIAN)

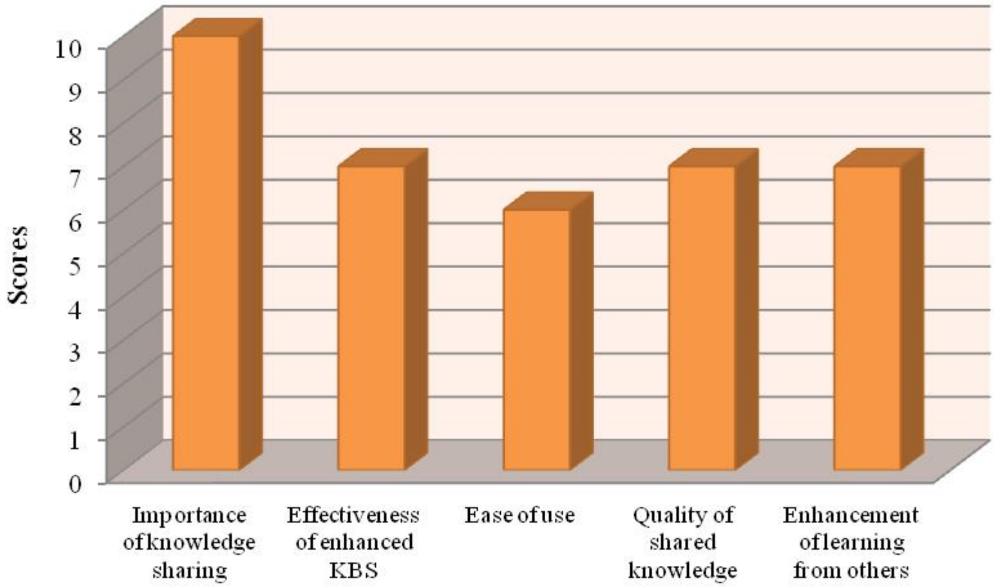

Fig. 11 Results of physicians' feedback