

ADEMA : A System to Help Physicians in the Asthma Health Care

Icham Sefion^{1,2}, Abdel Ennaji¹, Marc Gailhardou²

¹PSI Laboratory, ²GlaxoSmithKline

¹Place Emile Blondel, 76821 Mt St Aignan, France

²100 route de Versailles, 78163 Marly le roi, France

{icham.sefion, abdel.ennaji}@univ-rouen.fr, marc.gailhardou@gsk.com

Abstract

Asthma is a distressing disease, affecting up to 7% of the French population and causing considerable morbidity and mortality. A medical decision support system such can help physicians to control this chronic disease. Thanks to the health care network (RESALIS[®]) of Alliance Médica (disease management branch from GlaxoSmithKline), asthma consultation data were collected to exploit them. We chose Case-Based Reasoning paradigm to develop our medical decision support system. Intelligent data analysis methods have been used to determine the knowledge models for our system. A Self-Organising Map has been used to analyse medical data to show if homogeneous groups of cases exist. A case is an asthma consultation. Our similarity metric is based on MVDM method. We present our data analysis results and similarity metric from which we designed our Decision Support System for asthmatic patients health care : ADEMA. An evaluation of ADEMA is presented.

Introduction

Asthma is a chronic inflammatory disease of the airways. In spite of the availability of effective drugs, this pathology remains insufficiently controlled. The WHO notes indicate that the prevalence has doubled in the last ten years. There are 2000 deaths per year in France of which more than half are avoidable. One estimates that today 2,500,000 people in France, adults and children, suffer from asthma, this represents 5 to 7 % of the population. This alarming situation of public health shows that it is essential to improve the health care and quality of life of the asthmatic patients notably by providing a decision support system to physicians. This work lies within the scope of the RESALIS[®] project which is an experiment of health care network aiming at improving quality of health care for asthmatic patients. When taking a decision for a patient, physician use both their experience and academic knowledge. Our objective is to improve the resort to the experience.

After studying the principal methods of medical decision support system, we chose the Case Based Reasoning paradigm. Case Based Reasoning is a technique of artificial intelligence that attempts to solve a given problem within a specific domain by adapting established solutions to similar problems. In order to design a case based system, we should determine a case model, case indexing and a similarity metric. To determine this knowledge model, we used intelligent data analysis methods. In the first part, we present the context of this work, the nature of our data and method that we used for intelligent data analysis. A self-organising map has been used to analyse data. After, we present our similarity metric based on the MVDM method.

Next, we show our data analysis results obtained with clustering method, the evaluation of MVDM method and we present finally our Decision Support System : ADEMA, an evaluation of ADEMA and the user interface.

The Context

The RESALIS Project

Our work lies within the scope of the RESALIS[®] project of Alliance Médica. This project is a coordinated health care network which aims at facilitating access to health care for the asthmatic patients. An information system (software, protocols of data exchanges...) ensures the communication between physicians which respects the medical secrecy. A database, containing all the consultations of the patients included in the network (355 patients included to date), is fed daily. This data constitute our raw material for the development of a decision support system for physicians.

The Physician decision

When making a decision for a patient, physicians use both their experience and academic knowledge (Figure 1). Our objective is to improve the experience the physician resorts to with a medical decision support system.

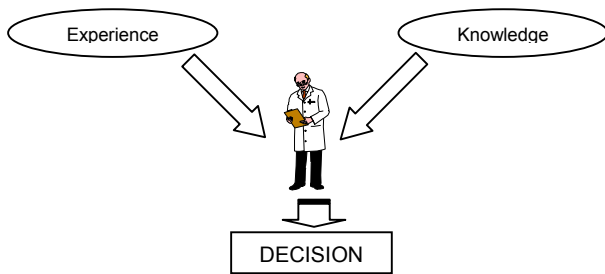


Figure 1 : the physician decision

Data description

Data used to design the CBR system were obtained from the health care network RESALIS[®]. They are stored in a database. This database is today made up of 1200 consultations representing approximately 350 patients.

A physicians group has selected features that they used for asthma medical care. The selected features is a subset of available attributes that directly contribute to health care activities : *asthma attack frequency*, *night symptom frequency*, *beta₂-agonist use frequency*, *exacerbation*, *activity days lost*, *smoking*, *DEP measure*, *patient age*, *asthma age*, *severity of asthma* and *course treatment*. In this item-set, *severity of asthma* and *course treatment* are the outcomes and all other variables are considered to contribute to these outcomes. Concerning his eleven variables, eight are symbolic and three are quantitative.

Before intelligent analysis, the input data must be pre-processed carefully. In the first time, we have transformed Symbolic data into quantitative data. For example, the *asthma attack frequency* variable values are : none, less than once a week, more than one per week and everyday. We transformed those variable values to 1,2,3 and 4 respectively. All symbolic variables are transformed to quantitative variables therefore we have only quantitative variables. However, the scales of the quantitative variables are very different, for example, the *patient age* variable fluctuate from 12 to 80 when the *asthma attack frequency* fluctuate from 1 to 4. All the variables should have an equal influence in the training phase of the SOM, therefore we have normalised all the variables between 0 and 1. In the data, there is a great number of missing values. Consultations with missing values was suppressed from data to not affect the reliability of the data analysis method. We obtained finally a database with approximately 400 consultations with no missing values, with normalised quantitative data. A Self-Organising Map has been used to analyse this data for defining a case and to compile the case library.

Method

After studying the principal methods of medical decision support system, we chose the Case Based Reasoning (CBR) paradigm. A rule-based system breaks a problem down into a set of individual rules that each solves part of the

problem. Rules are combined together to solve a whole problem. To create these rules by hand, it's necessary to know how to solve the problem, and this task can be extremely complex and time consuming. CBR system differ fundamentally in that to use them, we do not need to know how to solve the problem, only to recognise if we have solved a similar problem in the past. The major disadvantage of Neural Network (NN) technology compared with CBR is that an NN system functions as a "black box". The answer given by an NN is a function of the weighted vectors of its neurones. No explanation or justification of any sort can be given by an Neural Network. Remember that CBR retrieves the most similar case and attempts to reuse the solution from case. We chose CBR technique for our medical decision support system for asthmatic patient health care.

The Case-Based Reasoning

Case Based Reasoning is a technique of artificial intelligence that attempts to solve a given problem within a specific domain by adapting established solutions to similar problem. We can describe CBR typically as a cyclical process comprising the four REs :

1. REtrieve the most similar Case(s).
2. REuse the case(s) to attempt to solve the problem.
3. REvise the proposed solution if necessary.
4. REtain the new solution as a part of a new case.

We present the CBR cycle (Figure 2) :

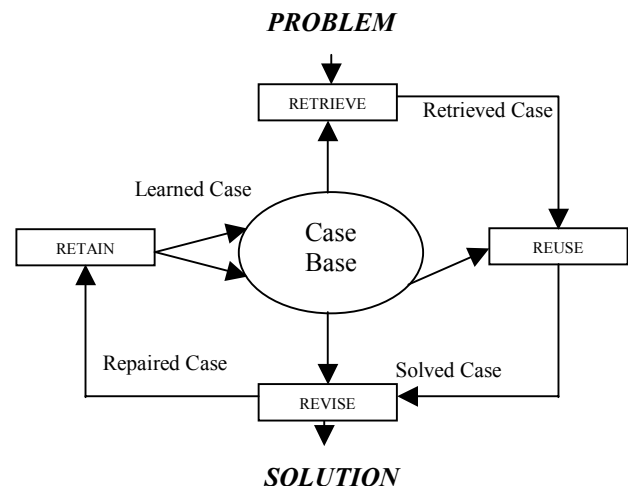


Figure 2 : The CBR cycle

Case representation

A case is a contextualized piece of knowledge representing an experience. It contains the past lesson that is the content of the case and the context in which the lesson can be used. A case can be an account of an event, a story, or some record typically comprising.

- The *problem* that describes the state of the world when the case occurred
- The *solution* that states the derived solution to that problem

To design a Case Based Reasoning system, you should define a set of features for cases in your specific domain. A choice of variables for our cases was made by the physicians group of RESALIS®. We want to analyse the influence of variables choice on the data clustering with the Self-Organising Map. With this method, we want also to estimate the nearness between cases in the case base. To resume, we used this method in order to represent efficiently a case in our CBR system and to compile the case library.

The Self-Organising Map. The Self-Organising Map (SOM) is a powerful neural network for analysis and visualisation of high-dimensional data. It maps non-linear statistical relationships between high-dimensional input data into simple geometric relationships on a usually two dimensional grid. The mapping roughly preserves the most important topological and metric relationships of the original data elements and, thus, inherently clusters the data. Therefore, the SOM can be used at the same time both to reduce the amount of data by clustering, and for projecting the data nonlinearly onto a lower-dimensional display. With this method, we will define our case and the case library.

Case retrieval

Remembering is the process of retrieving a case or a set of cases from case base. In general, two techniques are currently used in CBR tools : nearest neighbour retrieval and inductive retrieval. We chose the first method for its easiness to use. Two problem are similar if they are nearby in data space. You should calculate a similarity metric between cases to find similar case(s). To determine this distance, we chose to use the MVDM (Modified Value Difference Metric) technique. It's a powerful method for measuring distance between values of features in domains with symbolic feature values.

MVDM method. It's a powerful new method for measuring the distance between values of features in domains with symbolic feature values. Using this method, a matrix defining the distance between all values of a feature is derived statistically, based on the examples in the training set. The distance δ between two values for a specific feature is defined in Equation 1 :

$$\delta(V1, V2) = \sum_{i=1}^n \left| \frac{C1_i}{C1} - \frac{C2_i}{C2} \right|^k \quad (1)$$

In the equation, V1 and V2 are two possible values for the feature. The distance between the values is a sum over all n classes. C1_i is the number of times V1 was classified into category i, C1 is the total number of times value 1

occurred, and k is a constant, usually set to 1. The total distance Δ between two instances is defined in Equation 2 :

$$\Delta(X, Y) = \omega_X \omega_Y \sum_{i=1}^N \delta(x_i, y_i)^r \quad (2)$$

where X and Y represent two instances (e.g., two consultations for the asthmatic patient health care), with X being an exemplar in memory and Y a new example. The variables x_i and y_i are values of the ith feature for X and Y, where each example has N features. ω_X and ω_Y are weights assigned to exemplars.

Adaptation process

After a set of similar cases has been retrieved, these cases can be adapted to resemble more closely the case under scrutiny. Adaptation is typically performed based on a set of heuristic rules. These rules can be applied on differences observed between input parameters of the retrieved case and the current case to advise on adaptation of the proposed outcomes according to the current case. Two different adaptation strategies were implemented and tested in the current study :

For the first adaptation strategy, we used the weights of retrieved cases to define the solution of the current case. The weights are selected according to heuristic values specified as follows : 5 for the more similar case, 4 for the second and 1 for the last more similar case. The following formula was used :

$$Solution = \frac{\sum_{i=1}^5 Weight_i * Solution_i}{\sum_{i=1}^5 Weight_i}$$

The second adaptation strategy is based on the weights and on the distance between retrieved cases and the current case. Also, more the distance is little, more the influence will be great according to the following formula :

$$Solution = \frac{\sum_{i=1}^5 Weight_i * Solution_i}{\sum_{i=1}^5 Weight_i * distance_i}$$

Results

Case representation

Globally, one could show that the variables used for the analysis have a rather satisfactory capacity of discrimination taking into consideration our result. One

could see that there were certain regularities within our data. However, we also showed the presence of zones of recovery. This is explained partly by the presence of outliers and with the lack of homogeneity in the manpower of the various groups. It should be recalled that our objective here, was simply to have an idea about the topology of the data in order to direct our choices for the similarity measure on the one hand and to locate the irregularities which can exist in the data on the other hand. Finally, following these various analyses for various configurations of variables, our choice for the representation of the cases remains that proposed by the doctors and the group of experts. It will be necessary however to eliminate the noises in the data in particular the outliers which will be able to have a negative effect on the case retrieval.

Case library

After the choice of the case representation, a case library of 190 patients was compiled. The study group was obtained by a data analysis of our database with the SOM method. Data in the case library included patient information such as age, symptoms such as asthma attack frequency and results such as severity of asthma and course treatment. Therefore, a case in the case library consists of the following features :

Age of the patient and asthma age
Asthma attack frequency, night symptom frequency, Beta₂-agonist use frequency, exacerbation, activity days lost, smoking which represent the *problem*
severity of asthma and course treatment for the *solution*

Case retrieval

We evaluated several methods for the similarity measure and we shown that the MVDM method was the best following our results presenting in table 1 :

	Euclidienne	Manhattan	Chebychev	MVDM
Taux Global	59.0788	61.7628	63.4402	66.7772

Table 1 : recognition rate for several methods

The MVDM technique present a good performance for our type of medical data compared with other techniques. Our similarity measure will be based on this approach for the case-based system.

Adaptation

Two different adaptation algorithms were evaluated that modify the solution assessment of the retrieved cases to fit the new problem. For the three adaptation strategy, we tried to implement adaptation rules based on the solution of the five cases that matched the input case best. These strategies were based on the assumption that information

content may be improved by increasing the number of cases that contribute to the final solution. However, the solution accuracy of this reasoning system was also less than the best match approach. Thus, the solution of the retrieved case with the highest similarity was finally selected as the proposed solution for the prototype implementation of ADEMA. The results obtained by adaptation algorithms following case retrieval are displayed in the table 2 as follows :

Adaptation	Severity	Treatment	Both
No	61%	71.2%	53%
Method 1	54%	63%	49.4%
Method 2	55%	63%	50.1%

Table 2 : Adaptation results

ADEMA

We designed a prototype of our decision support system, ADEMA, with the Matlab software. In the figure 4, we presented the user interface of ADEMA. Thus, the final prototype of ADEMA had to be implemented using the Matlab Language. The case library was organised as a database in a text file and Matlab was used to implement the retrieval, the adaptation processes and the user interface of ADEMA.

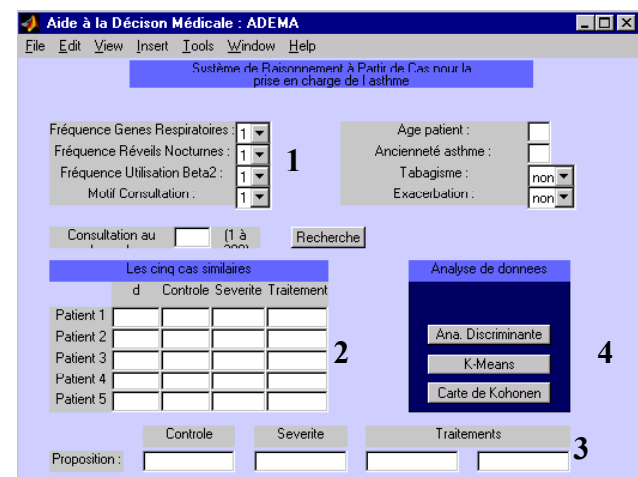


Figure 3 : ADEMA

In step 1, the user should give consultation values concerning the asthmatic patient. After the retrieve step, the system give the five most similar cases in descending order in the step 2. In the step 3, the system propose a global solution after the adaptation step. In 4, you can use a data analysis tool for analysing your data.

Evaluation

To evaluate the accuracy of the CBR application ADEMA, a 'round robin' evaluation method was applied as follows : each of the 190 cases was temporarily removed from the case library and was then presented to ADEMA as a case for evaluation. The output of the CBR system was compared to the solution of tested case. We shown that the best results was obtained with the MVDM method for the similarity metric and with no adaptation. So, we obtained the results as follows :

	Severity	Treatment	Both
accuracy	61%	71.2%	53%

Table 3 : Evaluation results

The solution accuracy of a CBR system for the asthma health care critically depends on the distribution and the number of the study population stored in the case library. Therefore, the accuracy of ADEMA possibly increase provided additional cases are included into the case library.

Conclusion

In this paper, we have presented our work about designing a case-based system for asthmatic patient health care. Case Based Reasoning was chosen for its advantages in medicine domain. We defined our case representation and case library thanks to an intelligent data analysis. This data analysis was made with the self-organizing map. Next, we tested the MVDM technique for the similarity measure and we shown that this approach is the best for our application. A comparison of recognition rate of several methods shown that the MVDM method purpose the best result. Two adaptation algorithms was implemented and we shown that we obtained the best results with no adaptation. These three aspects were integrated in the case-based system. The ADEMA prototype was implemented with Matlab software. Case-based reasoning as a method of artificial intelligence was successfully used to develop a decision support system for asthma health care with good performance. However, further studies are needed to define in more detail the clinical impact of ADEMA.

References

- Degoulet P., Fieschi M. 1998. *Informatique médical*. Paris : Masson.
- Kolodner J. 1993. Case Based Reasoning. San Mateo : Morgan Kaufmann Publishers.
- Aamodt A., Plaza E. 1994. Case Based Reasoning : Foundational Issues, Methodological Variations, and System Approaches. AI Communications 7 : 39-59.

Watson I. 1997. Applying Case-Based Reasoning : Technique for Enterprise Systems, San Francisco : Morgan Kaufmann Publishers.

Fargeas X., Frydman F. 1988. *Les Systèmes Expert en Médecine*, Paris : Hermes.

Ribert A. 1998. Structuration évolutive de données : application à la construction de classifieurs distribués. PhD Thesis, University of Rouen.

Koton, P. 1988. Using experience in learning and problem solving. PhD Thesis, Laboratory of Computer Science MIT.

Bareiss E. R. 1998. PROTON : A Unified Approach to Concept Representation, Classification and Learning. PhD Thesis, University of Texas, 1988.

Kohonen T. 1995. *Self-Organizing Maps*. Berlin: Springer.

Cost S., Salzberg S. 1993. A weighted nearest neighbour algorithm for learning with symbolic features, Machine Learning.