INTELLIGENT DATA ANALYSIS USING MULTIPLE CRITERIA DECISION MAKING

Crina Grosan[1,2], Ajith Abraham[1]
[1]*Center of Excellence for Quantifiable Quality of Service*
Norwegian University of Science and Technology
Trondheim, Norway
[2]*Department of Computer Science*
Babes-Bolyai University Cluj-Napoca, Romania
{crina, ajith}@q2s.ntnu.no

ABSTRACT.

Multiple Criteria Decision Making or multi-attribute (MCDM) problems refer to performing preference decisions (e.g. ranking, ordering, prioritization, selection) over several available alternatives or objects that are characterized by multiple, usually conflicting, attributes or criteria. MCDM techniques have been and are still being encountered in a wide range of human activities, from the professional choices to managerial decisions. The typical examples arise from project management, facility location, capital budgeting, and, last but not least, ranking and evaluating in medical field (medical treatments, risk factors for a disease, diseases, medical units, etc.). The goal of this paper is to present some methods in this area which could be useful for medical applications as well as to describe an evolutionary MCDM model for ranking the risk factors for Bronchial asthma.

1. INTRODUCTION

Concerning the medical units efficiency, our intention, is the measurement of medical practice value (regarded as outputs) with respect to the value of all types of expenses (management expenses, working expenses, overhead expenses, sundry charges) (regarded as inputs) for a group of general or specialized medical units (DMUs) (hospitals, medical offices, etc.). The Medical Units EFficiency (MUEF) [1] problem can be represented as a multi criteria decision making (MCDM) problem. In this context, the DMUs can be considered to be the alternatives, the outputs could be viewed as criteria or attributes that we would like to maximize, and the inputs can be considered to be the attributes that we would like to minimize. Among other things, MCDM problems refer to performing ranking decisions over several available alternatives that are characterized by multiple conflicting attributes (inputs vs. outputs).

Generally, all quantitative attribute ratings or values for a decision or evaluation problem are normalized to eliminate computational problems caused by differing measurement units/scales. Attribute normalization is essential for many compensatory MCDM methods. Normalization aims at obtaining comparable scales, which allow inter-and intra-attribute comparisons.

Subsequent to attributes selection and weighting, another important step in order to solve a MCDM problem is the evaluation or quantification (i.e. the assignment of numerical values to qualitative data) of the state of a qualitative attribute at each alternative.

Traditionally problems with several competing criteria were reformulated by using one criterion or scalar objective function and the multicriterion nature of the original problem was more or less hidden. One popular approach is to combine all the criteria into one scalar objective function. Another popular approach is to choose one of the criteria as the objective function and transform the others into constraints. These techniques may look reasonable but they prove to have several shortcomings.

A Pareto optimal set is regarded as the mathematical solution to a multicriterion problem. In continuous problems the number of Pareto optimal solutions is usually infinite. Only in relatively simple cases the entire Pareto optimal set can be determined analytically. In a typical problem we must be satisfied with obtaining enough Pareto optima to cover the minimal set in the criteria space properly. This computed subset of Pareto optima could be called as a representative Pareto optimal set and its quality can be judged for example by its

ability to cover the whole minimal set evenly. In some problems, however, the cost of generating just one Pareto optimum may become so high that the designer can afford only a few Pareto optimal solutions. Before performing a numerical optimization we must select a generation strategy which guarantees that only Pareto optima are obtained. It is also useful to know which generation techniques cannot reach all Pareto optima and why not.

The most frequently used methods for generating Pareto optima are based on the idea of replacing the multicriterion problem by a parametrized scalar problem. Typically, each parameter combination corresponds to one Pareto optimum and by varying their values it is possible to generate all or the part of the Pareto optimal set.

There are some particular situations for which Pareto dominance cannot be applied while considering the problem as multiobjective without reducing it to a single objective one. For instance, in the situation in which all solution are nondominated we cannot say that one is better that the other by simply applying Pareto dominance definition.

In this paper we will analyze such a multiobjective optimization problem from the medical domain for which Pareto dominance alone cannot decide which solution is the best.

2. PROBLEM STUDIED

Bronchial asthma and associated allergies Asthma (bronchial asthma) is an inflammatory disorder of the airways, characterized by periodic attacks of wheezing, shortness of breath, chest tightness, and coughing. Asthma is a disease in which inflammation of the airways causes airflow into and out of the lungs to be restricted. When an asthma attack occurs, the muscles of the bronchial tree become tight and the lining of the air passages swells, reducing airflow and producing the characteristic wheezing sound. Mucus production is increased. Bronchial asthma is a public health problem with gradually increasing importance, affecting more than 100 million individuals worldwide and found independently of the level of development of the country. Factors related to lifestyle and the environment form the basis for the increase in the prevalence of the disease [3]. Most people with asthma have periodic wheezing attacks separated by symptom free periods. Some asthmatics have chronic shortness of breath with episodes of increased shortness of breath. Other asthmatics may have cough as their predominant symptom. Asthma attacks can last minutes to days and can become dangerous if the airflow becomes severely restricted.

In sensitive individuals, asthma symptoms can be triggered by inhaled allergens (allergy triggers), such as pet dander, dust mites, cockroach allergens, molds, or pollens. Asthma symptoms can also be triggered by respiratory infections, exercise, cold air, tobacco smoke and other pollutants, stress, food, or drug allergies. Aspirin and other non-steroidal anti-inflammatory medications (NSAIDS) provoke asthma in some patients.

Bronchial asthma occurs following the interaction between genes and environment, but none of these factors are enough for the disease to express itself.

The goal is to establish a hierarchy of the risk factors for BA and the onset of asthma in phenotype. Experiments were done on a group of patients which have been exposed to risk factors which can be grouped into two categories such as: genetic factors and environmental factors.

Genetic risk factors are:

- Maternal transmission of BA and atopies;
- Mother's BA;
- BA to antecessors;
- All allergies to antecessors (without BA);
- Urticaria to antecessors;
- Father's BA;
- Eczema to antecessors;
- Rhinitis to antecessors.

Environment risk factors are:

- Life in the city;
- House dust mite;
- House environment (crowd);
- Smoking in the family;

- Traffic pollution;
- Industrial pollution;
- Smoking mother.

The hierarchy has to be established by taking into account six criteria representing BA and the allergies which are present in the phenotype of the patients such as:

- rhinitis,
- conjunctivitis,
- eczema,
- urticaria
- the asthma onset coefficient.

Table 1. Risk factors for Bronchial asthma**.**

Risk factors		Criteria					
		BA	Conjunctivitis	Rhinitis	Urticaria	Eczema	CDA
1.	House dust mite	81.3	0	21.9	25	12.5	0.218448
2.	Father's BA	84.6	23.1	15.4	15.4	15.4	0.144254
3.	Eczema to antecessors	80	40	40	20	0	-0.22581
4.	Smoking mother	80.6	6.5	9.7	19.4	6.5	0.214712
5.	Traffic pollution	80	3.2	9.7	9.7	12.9	0.338986
6.	House environment (crowd)	73.9	5.8	15.9	23.2	10.1	0.231362
7.	Smoking of other members of the family	71.7	7.5	11.3	17	7.5	0.228511
8.	Industrial pollution	64	8	24	28	8	0.148545
9.	Mother's BA	77.3	0	4.5	9.1	0	0.169509
10.	BA to antecessors	66.7	2.4	12.2	22	6.1	0.087415
11.	Life in the city	65.6	6	15.8	22.6	7.5	0.02771
12.	Maternal transmission of BA and atopies	67.3	0	12.7	16.4	7.3	-0.01284
13.	Rhinitis to antecessors	65	6.7	18.3	15	8.3	-0.19996
14.	All allergies to antecessors (without BA)	60.4	4.4	16.5	17.6	5.5	-0.16459
15	Urticaria to antecessors	53.2	3.2	11.3	21	3.2	-0.19033

We mention that the evaluation of each criteria of each risk factor has been done by the percentage of the patients who had BA or/and other allergies and have been exposed to one of the risk factors. Asthma onset coefficient CDA represents an estimate of the risk regarding the BA onset in phenotype. CDA has been established through VDR representing the average age of the BA onset in patients exposed to a certain (risk) factor and VDF representing the average age of the BA onset in patients non-exposed to the corresponding (risk) factor. The formula for obtaining CDA is given by:

$$\text{CDA} = \frac{VDF - VDR}{\max(VDF, VDR)}.$$

Asthma onset coefficient CDA has values between -1 and 1.

The values of all criteria for the considered risk factor averaged for all the studied patients are given in Table 1. All objectives are to be maximized.

3. PROPOSED APPROACH

We used an Evolutionary Algorithm (EA) [6], [7] for solving this problem. We formulated a special way to initialize the population based on some existing information about the risk factors, which must be ranked. For initializing the population in a way which will help in finding the final solution, partially randomly generated individuals were used.

Each chromosome is represented as a string whose length is equal to the number of units, which must be ranked (risk factors for Bronchial Asthma in our case). Instead of filling the positions in the chromosomes with random generated values between 1 and the number of units, the following procedure is followed.

1. Using Pareto dominance, all the nondominated units (risk factors respectively) are kept in a separate set (and not considered in the initial set anymore). Let us denote this set by S_1 the size of this set by s_1. Therefore, first s_1 positions in the chromosome consist of a permutation of the units (indices) from the set S_1 whose ordering is randomly generated.
2. The procedure described above is repeated until all the levels of nondominated units are selected. This means, from the remaining units, the nondominated solutions are again selected. These units are nondominated only in respect with the current content of the units set.

The minimum number of such levels can be 1 (which means all the solutions are nondominated) and can be maximum units set size. All the units which were previously removed from the initial set will dominate these units. Consequently, chromosome positions between s_1+1 and s_2 will consist of random permutations of the units (indices) which are dominated by a single other unit, positions between s_2+1 and s_3 will consist of permutations of the units which are dominated by 2 units, and so on.

From the initialization of a chromosome, it is obvious that one gene (which represents a risk factor in our case) cannot be dominated by a successive gene. Two consecutive genes can represent either nondominated items or the second one is dominated by the first one. For establishing a hierarchy when all the items are nondominated and from a multiobjective perspective, the fitness of a chromosome is computed as follows:

- Set the value of fitness function to 0.
- For each pair of genes (i; i + 1) we will calculate the number of objectives for which i is better than i + 1.
- If this number is less than number of objectives/2 (in the case the number of objectives is and even number) or it is less than integer portion of number of objectives/2 + 1 (in the case the number of objectives is an odd number) then the value of objective function will increase by 1.

Our objective is to obtain the value 0 for the fitness function. The evolutionary scheme adopted in this research is very much similar to the classical evolutionary algorithm, which uses only mutation as a genetic operator. Each chromosome from the population is affected by mutation. Two genes are randomly chosen from each segment of the chromosome and their values are exchanged between them. The parent and offspring are directly compared. The one with a smaller fitness will be kept into the population of the new generation.

4. EXPERIMENTAL RESULTS

The evolutionary technique described in the previous section is applied for ranking the risk factors for Bronchial asthma. Results obtained by our approach are compared with the results obtained by applying weighted sum approach ([[8]], [[9]]). Parameters used by the evolutionary approach are:

- Population size: 100
- Number of generations: 100

Several hierarchies of risk factors are obtained. All the solutions are nondominated between them. Some of the solutions obtained in the final generation are given below:

8 7 14 5 4 2 1 11 3 13 4 10 12 15 9

3 6 2 1 8 13 14 5 7 11 4 19 12 15 9

1 8 6 2 14 5 13 4 3 7 11 12 10 15 9

8 1 4 3 2 5 6 7 13 11 14 12 10 15 9

1 6 2 14 5 7 13 4 3 8 11 10 12 15 9

The evolution of the fitness function is illustrated in Figure 1.

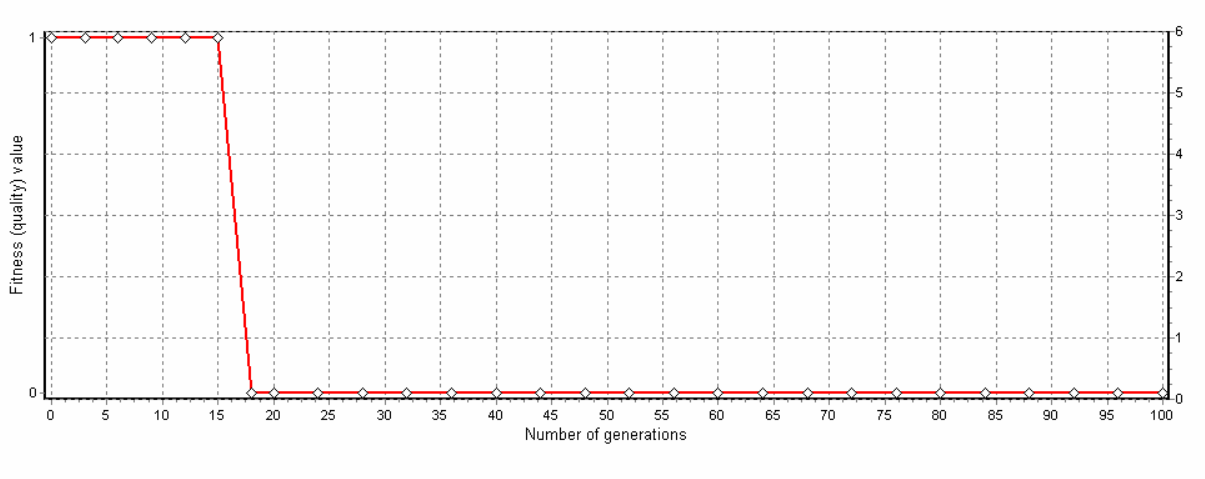

Figure 1. Evolution of the fitness function for the evolutionary approach.

Table 2. Results obtained by weighted sum method for ranking the Bronchial asthma risk factors

Risk factors		Criteria						Score
		BA	**Conjunctivitis**	**Rhinitis**	**Urticaria**	**Eczema**	**CDA**	
1.	House dust mite	81.3	0	21.9	25	12.5	0.218448	**50.74**
2.	Father's BA	84.6	23.1	15.4	15.4	15.4	0.144254	**50.23**
3.	Eczema to antecessors	80	40	40	20	0	-0.22581	**48.79**
4.	Smoking mother	80.6	6.5	9.7	19.4	6.5	0.214712	**47.61**
5.	Traffic pollution	80	3.2	9.7	9.7	12.9	0.338986	**46.71**
6.	House environment (crowd)	73.9	5.8	15.9	23.2	10.1	0.231362	**46.11**
7.	Smoking of other members of the family	71.7	7.5	11.3	17	7.5	0.228511	**43.12**
8.	Industrial pollution	64	8	24	28	8	0.148545	**42.42**
9.	Mother's BA	77.3	0	4.5	9.1	0	0.169509	**42.32**
10.	BA to antecessors	66.7	2.4	12.2	22	6.1	0.087415	**40.27**
11.	Life in the city	65.6	6	15.8	22.6	7.5	0.02771	**40.05**
12.	Maternal transmission of BA and atopies	67.3	0	12.7	16.4	7.3	-0.01284	**38.75**
13.	Rhinitis to antecessors	65	6.7	18.3	15	8.3	-0.19996	**36.89**
14.	All allergies to antecessors (without BA)	60.4	4.4	16.5	17.6	5.5	-0.16459	**34.87**
15	Urticaria to antecessors	53.2	3.2	11.3	21	3.2	-0.19033	**30.85**
weights		50	4	13	19	6	8	

Results obtained for weighted sum method for ranking the risk factors for Bronchial Asthma are taken from [[5]]. The weight of each criterion has been evaluated following its frequency in the patient phenotype. The frequencies of the allergies of the patient have been mentioned. BA was evaluated after the frequency of the patients with BA alone. The sum of frequencies is 114, 3% which represents 100 points from the total weight. Thus has been set the weight of each criterion. BA and BA onset have been initially evaluated

together, with a total weight of 58 points, and then they have been evaluated after the importance considered to represents the BA onset in the evaluation of the risk disease and have been granted 8 points. Results obtained by applying weighted sum method are presented in Table 2.

The risk factor 9 is dominated with respect to all objectives by three other risk factors (1, 4 and 5), risk factor 15 is dominated by two risk factors (6 and 8), risk factors 10 and 12 are dominated by a single risk factor (6). This means, in the resulting hierarchy, factor 9 must be on the last position.

Factors 15, 10 and 12 are also dominated and must be in one of the last positions. But this is not evident from the hierarchy obtained by the weighted sum approach. All these domination relations are taken into account by the evolutionary method and several alternatives of risk factors ranking are obtained at the end of the search process.

In the results obtained by the evolutionary approach for the second case study, genetic factors are of a very high importance (as shown in other medical studies such as [5], [3], [2]). Also, hierarchy obtained by the evolutionary approach shows the same thing. Experiment results show a very fact convergence of the evolutionary approach.

5. CONCLUSIONS

By combining all the objectives in a single objective function (and transforming the multiobjective optimization problem into a single objective problem) at one application of the algorithm, at most one solution can be obtained. In order to obtain multiple solutions, we have to apply the algorithm several times. Even then, we cannot be sure that all solutions are different. Running time required is another disadvantage of the weighted sum approach. This is the case for our case study: Bronchial Asthma risk factors ranking.

The evolutionary approach obtains several solutions in one run. Also, the dominance concept (which is more than standard Pareto dominance relationship) is playing an important role in the final hierarchy. Both genetic and environmental factors represent a risk for the Bronchial Asthma and their influence differs from a patient to another, from a world region to another, etc. Genetic susceptibility is both context dependent and developmentally regulated, and ignoring the environmental context will miss many important associations and clues to pathogenesis. That's why a right classification of the risk factors is very important in control and prevention of Bronchial Asthma.

The evolutionary approach proposed inn this research is very flexible and can be applied to any problem of this kind. No additional information about the problem is required (like different weights in the case of weighted sum method).

REFERENCES

[1] Burke, W., Fesinmeyer, M., Reed, K., Hampson, L., Carlsten C., Family history as a predictor of asthma risk, American Journal of Preventive Medicine, 24(2), pp. 160-169, 2003

[2] Cengizlier, M.R., Misirlioglu E.D., Evaluation of risk factors in patients diagnosed with bronchial asthma, Allergology and Immunopathology (Madr), 34(1), pp. 4-9, 2006.

[3] Cimoca, Gh., A simple algorithm for comparing Hospital units efficiency, Appl. Med. Inf., no.1-2, (2001), 3-7

[4] Cimoca, Gh., Dollinger, R., Specific quantitative and qualitative attributes for medical ranking/evaluation applications, Appl. Med. Inf., no.1-2, (2001), 8-18

[5] Gherman, M., Moldoveanu, M., Tigan, S. Hierarchy of risk factors in bronchial asthma. Applied Medical Informatics, 14(1-2), pp. 35-41.

[6] Goldberg, D.E. Genetic algorithms in search, optimization and machine learning. Addison Wesley, Reading, MA, 1989

[7] Holland, J. Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975

[8] Kim, I.Y., de Weck, O.L. Adaptive weighted-sum method for bi-objective optimization: Pareto front generation. Structural and Multidisciplinary Optimization, 29, pp. 149-158, 2005.

[9] Zadeh, L. Optimality and Non-Scalar-Valued Performance Criteria. IEEE Transaction on Automation Control 8, 59-60, 1963.