Verbal analysis in multicriteria risk factors environment for construction expert based allergy diagnostic and treatment software

SLIESORAITYTE IEVA
Faculty of Fundamental Science
Vilnius Gediminas Technical University
Sauletekio al. 11, LT-10223 Vilnius
LITHUANIA

DUBAKIENE RUTA
Allergology Centre
Vilnius University
Antakalnio 124, LT-10200 Vilnius
LITHUANIA

USTINOVICIUS LEONAS
Faculty of Civil Engineering
Vilnius Gediminas Technical University
Sauletekio al. 11, LT-10223 Vilnius
LITHUANIA

PODVEZKO VALENTINAS
Faculty of Civil Engineering
Vilnius Gediminas Technical University
Sauletekio al. 11, LT-10223 Vilnius
LITHUANIA

SLIESORAITIENE VIKTORIJA
Faculty of Mechanics
Vilnius Gediminas Technical University
J.Basanavičiaus str.28, LT-03224 Vilnius
LITHUANIA

Abstract: - The quantity of dramatically increased figures of patients suffering from allergic diseases is observed world over. Predicting the development of allergic processes or assessing various strategies (sensitization – non sensitization – disease and strategic potential of particular object), while expert decision analysis is the reasonable one. Our research focuses on the tool and software which applied for screening of various clinical symptoms as markers of sensitization. The present paper describes a feasibility study of using verbal analysis for determining a diagnostic and treatment course of allergic diseases, depending on a particular patient's personality and risk factors environment. This integrated approach could present potential of targeted treatment scheme creation for particular patient considering the subject individualities. Our created tool has the potentiality to identify risk level for allergy assessment and define the proper treatment schema with high sensitivity considering positive and negative predictive values for particular subject with apropos allergy.

Key-Words: - Allergy, Multicriteria risk factors environment, Decision analysis

1 Introduction

The quantity of dramatically increased figures of patients suffering from allergic diseases is observed world over during the past decade [1], and allergic diseases have become a social problem affecting medical costs and quality of life [2] In determining the causes of this increase and aiming for the adequate prevention approach it is important to formulate the proper multicriteria risk factors environment for pathological process development concerning the particular subject individualities. Primary causes are those considered to induce allergic disease in a non-sensitised person, while secondary ones are those that trigger symptoms in people who are already sensitized [3].

Predicting the development of allergic processes or assessing various strategies, sensitization – non sensitization – disease and strategic potential of particular objects, expertise is reasonable one. In multicriteria environment it is hardly possible to achieve without resorting to special techniques [4-5]. In medical aspect analytical verbal decision methods are most expedient for multicriteria classification. However, expert attitudes concerning a particular problem often differ considerably and may even be conflicting. Making an expertise-based decision, it is necessary to determine the concordance degree of expert solution and multiple criteria evaluation method should be applied [6-9].

Our research focuses on computer science in medicine, particularly on the tool and software which applied for screening of various clinical symptoms as markers of sensitization and correlation with allergy process. The present paper describes a feasibility study of using verbal analysis for determining a diagnostic and treatment tool course of allergic diseases, depending on a particular patient's personality and risk factors environment. Proposed a susceptible marker – selectable tool predicts the development of allergic disease and can be applied to characterize development of sensitization leading to allergy concerning the multicriteria risk factors environment and creation of treatment scheme for particular subject as following.

2 Problem Formulation

Classification is a very important aspect in decision making. Classes in decision making are determined by the particular parameters, i.e. the efficiency of technical and technological decisions concerning the subject individualities. Many different methods for solving multicriteria classification problems are widely known. ORCLASS, as an ordinary classification, was one of the first methods designed to solve these kinds of problems. More recent methods such as DIFCLASS, CLARA and CYCLE have been applied for multicriteria expert analysis [10-11]. Formalizing our problem step by step analysis was applied:

1. G – a feature, corresponding to the target criterion (i.e. allergy treatment scheme).
2. $K = \{K_1, K_2, ..., K_N\}$ – a set of criteria, used to assess each alternative (course of treatment).
3. $S_q = \{k_1^q, ..., k_{w_q}^q\}$ – for q=1,..., N – a set of verbal estimates on the scale of criterion K_q, w_q – a number of estimates for criterion K^q; estimates in S_q are ordered based on increasing intensity of the feature G.
4. $Y = S_1 \times ... \times S_N$ – a space of the alternative features to be classified. Each alternative is described by a set of estimates obtained by using criteria $K_1, ..., K_N$ and can be presented as a vector $y \in Y$, where $y = (y_1, y_2, ..., y_N)$, y_q is an index of estimate from set S_q.
5. $C = \{C_1, ..., C_M\}$ – a set of decision classes, ordered based on the increasing intensity of feature G.
6. A binary relation of strict dominance was introduced. While relation is anti-reflexive and anti-symmetric and the transitive one. It may be also useful to consider a reflexive, anti-symmetric, transitive binary relation of weak dominance Q. The goal of decision making preferences to create imaginary *F*: $Y \to \{Y_i\}, i = 1, ..., M$, where Y_i – a set of vector estimations belonging to class C_i, satisfying the condition of consistency: $\forall x, y \in Y : x \in Y_i, y \in Y_j, (x, y) \in P \Rightarrow i \geq j$ (1).

2.1 Comparative assessment of multicriteria risk factors environment

A great number of multicriteria evaluation methods while the main principle of multicriteria evaluation – determination of weights (significances) of complex criteria, i.e. method known as Simple Additive Weighting (SAW) was introduced [7, 9]. The values of the criterion S_j for the j-th alternative combining the normalized values of $\tilde{r}_{ij}$ and weights ω_i of criteria R_i are calculated:

$$S_j = \sum_{i=1}^{m} \omega_i \tilde{r}_{ij} \quad (j = 1,...,n) \ (2).$$

The values of the criteria R_i are assumed to be positive, while all particular criteria should be maximizing. Minimizing criteria were transformed into maximizing ones prior to normalization:

$$\hat{r}_{ij} = \frac{\min_j r_{ij}}{r_{ij}} \quad (i = 1,...,m; j = 1,...,n) \ (3),$$

where the lowest positive criterion values are transformed into a maximizing value equal to one.

More complex multicriteria methods often rely on the method criterion S_j of *SAW*. Thus, one of two criteria of the compromising method *VIKOR* [12] matches the criterion S_j, using, however, another form of normalization (data transformation). The method of complex proportional evaluation suggested by the authors [1], when applying it to identify the pathological allergy process development concerning the particular subject individualities was as following:

$$Z_j^* = S_{+j} + \frac{S_{-\max} S_{-\min}}{S_{-j}} \ (4), \text{ where } S_{-\max} = \max_j S_{-j}.$$

2.2 Concordance coefficient for apropos alternative design formation

Expert evaluation results are presented in the matrix $E = \|e_{ij}\| \ (i = 1,...,m; j = 1,...,r)$, where m is the number of the criteria compared, and r is the number of experts. In evaluating the criteria, experts use various methods as well as scales of measurement, e.g. units, percentage, parts of one, ten-point score or pairwise comparison scale suggested by T. Saaty. It should be noted, however, that, in calculating the concordance coefficient, expert ranking of the criteria was applied. The concordance coefficient was described by M.Kendall [13]. The coefficient is related to the sum of ranks of a particular criterion elicited from all experts:

$$e_i = \sum_{j=1}^{r} e_{ij} \quad (i = 1,...,m) \ (5),$$

when it is associated with the sum S (variance analogue):

$$S = \sum_{i=1}^{m} (e_i - \overline{e})^2 \ (6),$$

The general mean $\overline{e}$ was calculated:

$$\overline{e} = \frac{\sum_{i=1}^{m} e_i}{m} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{r} e_{ij}}{m} \ (7).$$

The sum of ranks of m criteria assigned by r experts was as follows:

$$\sum_{i=1}^{m} e_i = \frac{1}{2} rm(m+1) \ (8),$$

and the general mean:

$$\overline{e} = \frac{1}{2} r(m+1) \ (9),$$

it depends only on the values m and r being independent of the concordance level. Basing ourselves the sum of natural numbers and their squares, one can prove that, in ideal case, the value S calculated by:

$$S_{\max} = \sum_{i=1}^{m} (ri - \frac{1}{2} r(m+1))^2 = \frac{r^2 m(m^2-1)}{12} \ (10),$$

This is the largest possible S value when the solutions of experts are in good agreement. An opposite case would be when the estimates are absolutely different, i.e., if all ranks from one to m are used and the sum of ranks of each criterion is the same, matching the mean value of the ranks. In this case, S is equal to zero, though this result is extremely rare and may be treated as a purely theoretical or boundary case. If we denote by S the actual deviation of the sum of squares of the criteria mean values from the general mean is calculated, then the concordance coefficient may be expressed by the relationship between the calculated S and the largest $S_{\max}$:

$$W = \frac{12S}{r^2 m(m^2-1)} \ (11).$$

If the solutions of experts are in agreement, the value of the concordance coefficient W is approaching one, if they differ considerably,

W is about zero. The concordance coefficient used for practical purposes, if its boundary value, showing that expert estimates are still in agreement, is determined.

In biomedical approach we are encountered with cases when two or more criteria are similar and it is hardly possible to give the priority to any of them. Such criteria are referred to as tied. The same rank is assigned to all of them, which is, in fact, the arithmetical mean of their ranks. It has been proved [13] that, in this case, the concordance coefficient was as following:

$$W = \frac{12S}{r^2 m(m^2-1) - r\sum_{j=1}^{r} T_j} \quad (12).$$

While the tied ranks indicator T_j of the j-th expert is calculated by the formula:

$$T_j = \sum_{k=1}^{H_j} (t_k^3 - t_k) \quad (13),$$

where H_j number of equal ranks for the j-th expert, t_k k-th number of equal tied ranks of the group.

2.3 Analytical verbal decision methods for classification of alternatives

In this chapter some most frequently used verbal ordinal classification methods are considered. All these methods belong to Verbal Decision Analysis group and have the following common features [11]. Attribute scale is based on verbal description not changed in the process of solution, when verbal evaluation is not converted into the numerical form or score. An interactive classification procedure is performed in steps, where the DM is offered an object of analysis (allergy treatment schema). A project is presented as a small set of rankings. The DM is familiar with this type of description; therefore one can make the classification based on particular expertise and intuition.

When the DM has decided to refer a project to a particular class, the decisions are ranked on the dominance basis. This provides the information about other classes of projects related with it by the relationship of dominance. Thus, an indirect classification of all the projects can be made based on a single decision of the DM. A set of projects dominating over a considered project are referred to as domination cone. A great number of projects have been classified many times. This ensures error – free classification.

If the DM makes an error, violating this principle, one is informed about conflicting decision on the screen and is prompted to adjust it. The comprehensive classification may be obtained for various numbers of the DM decisions and phases in an interactive operation.

The efficiency of multicriteria classification technique is determined based on the number of questions to DM needed to make the classification. This approach is justified because it takes into consideration the cost of the DM's time and the need for minimizing classification expenses.

2.4 Verbal analysis for allergy diagnostic and treatment scheme

This technique allows the classification to be developed in a series of successive steps, checking the conflicting information and arriving at a general method of solution. The method described takes into account the possibilities and limitations of the human data processing system [14]. Let us consider metric $\rho(x,y)$ in discrete space Y defined as:

$$\rho(x,y) = \sum_{q=1}^{N} |x_q - y_q| \quad (14).$$

Let us denote by $\rho(\vec{0}, y)$, i. e. the sum of vector's components, the index of vector $y \in Y$ (written as $\|y\|$). For vectors $x, y \in Y$ such that $(x, y) \in P$, let us consider a set:

$$\Lambda(x,y) = \{v \in Y | (x,v) \in Q(v,y) \in Q\} \quad (15),$$

that is a set of vectors weakly dominating y and weakly dominated by x. Having denoted y' =(1, …, 1), y''=(w_1, ..., w_N), we can see that $\Lambda(y'', y')$ matches the entire space Y. We also introduce a set

$$L(x,y) = \left\{ v \in \Lambda(x,y) \middle| \|v\| = \frac{\|x\| + \|y\|}{2} \right\} \quad (16),$$

that is vectors from $\Lambda(x,y)$ set equidistant from x and y (here and further division is done without remainder). We will need numerical functions $C^U(x)$ и $C^L(x)$ defined on Y, which are respectively equal to the highest and lowest class number allowable for x, that is a class for x not violating the condition of consistency. Let us consider vector x to be classified and belonging to class C_k, if the following condition is valid for x:

$C^U(x)=C^L(x)=k$ (17).

Let us define the procedure $S(x)$ (spreading by dominance). It is assumed that the class of x is known: $x \in Y_k$ (that is $C^U(x)=C^L(x)=k$). Therefore, for all $y \in Y$ such as $(x,y) \in P$ and $C^U(y) > k$ function $C^U(y)$ is redefined so that $C^U(y) = k$. Similarly, for all $z \in Y$, such as $(z, x) \in P$ and $C^L(z)$<k, function $C^L(z)$ is redefined so that $C^L(z)$=k.

2.4.1 Basic mechanism of the algorithm

Let us denote by $D(a, b)$ a procedure of classification on $\Lambda(a,b)$ set using the idea of dynamic construction of chains linking vectors a and b. It is assumed that $(a,b) \in P$ and classes of vectors a and b are known: $a \in Y_k, b \in Y_l$. The algorithm is as follows:

1) For each vector $x \in L(a, b)$ the steps 2-4 are made.
2) If a class for x is unknown ($C^L(x) < C^U(x)$) then x is presented to the DM for classification. Suppose that $x \in Y_r$. The spreading by dominance S(x) is being done. The condition of consistency is being checked.
3) If $r<k$ and $(a,x) \in P$, then perform $D(a, x)$.
4) If $r<k$ and $(x,b) \in P$, then perform $D(x, b)$.

In classifying vector x at the second step, the DM can make a mistake, and a pair of vectors $x, y \in Y$ violating the consistency condition (3) appears. Procedure R of resolving contradictions consists in the following. Let us denote the set of vectors explicitly classified by DM as E. So, while E contains a pair of vectors violating (3), such a pair is presented to DM with a proposition to change a class for one or two vectors. Then, functions C^U and C^L are redefined to their initial values and spreading by dominance $S(v)$ is done for each $v \in E$.

Generally speaking, the parameters of the algorithm including the number of questions to DM depend on the choice of vector x in the first step. The following heuristics is proposed: among all not yet classified vectors from $L(a, b)$ set the object, which explicitly dominates a maximum number of unclassified vectors is chosen. That is, one chooses the vector:

$$x^* = \arg \max_{x \in L(a,b)} \left| \{ y \in Y | (x,y) \in P \text{ or } (y,x) \in P, \rho(x,y) = 1, C^L(y) < C^U(y) \} \right| \quad (18).$$

2.4.2 CYCLE Algorithm features

Statement 1. At the end of CYCLE algorithm the space Y will be fully classified, that is $\forall y \in Y \; C^L(y)=C^U(y)$.

Lemma 1. For any $x, y \in Y$, such as $(y, x) \in P$, and for any chain $\Re = \langle x,...,y \rangle$, cardinality of the set $\Re \cap L(y,x)$ equals to 1.

Statement 2. A classification made with the help of the CYCLE algorithm is consistent, implying that a condition of consistency is satisfied.

3 Problem Solution

3.1 Classification of allergy treatment schema and description of the main criteria

A method of verbal classification used to analyze the diagnosis and treatment allergy origin diseases concerning mulricriteria risk factors environment (Fig. 1). A list of medicinal schema including particular medicine and dosage used for treating particular patient was determined.

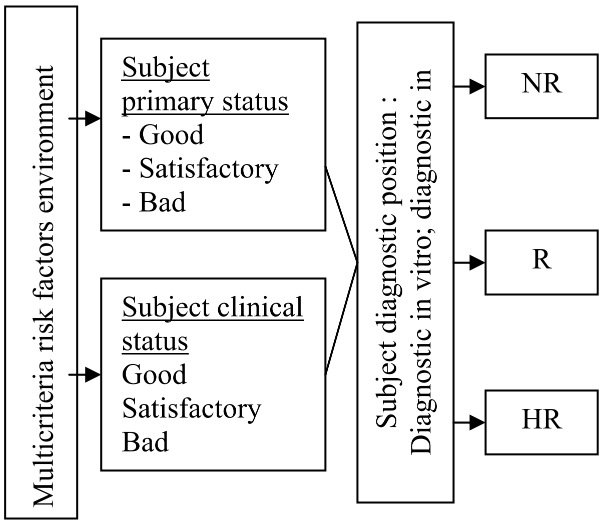

Fig.1. General scheme for apropos diagnostic and treatment scheme estimation.

After a series of iterations carried out under methodological control of consultants from Vilnius University Centre of Allergology, the following final decision classes were chosen. All classes are hierarchically structured. They include a list of medicines to be used for treating a particular allergy with the following recommendations given for each medicine usage taking into account patient individualities supported through the risk factors environment:

1. Highly recommended: this scheme of treatment is highly needed for treating the patient with particular allergy.

2. Recommended: this medicine may be used for treating the patient.
3. Not recommended: this medicine should not be used for treating the patient.

The hierarchical structure of classes depends on the following criteria: subject's primary status, subject's clinical status, subject's diagnostic position. While constructing the apropos treatment schema for particular subject concerning the mulricriteria risk factors environment the subject primary status, subject clinical status and subject diagnostic position was introduced. A more detailed description of the above groups is given below.

1. *Criteria `Subject primary status`* is the derived value from the data set of the one's health condition a priori allergy, i.e. subject age and gender, sensitization and allergy anamnesis, concomitant disease, family anamnesis of allergic diseases, the season and the exposure of potential allergens in apropos region.
2. *Criteria `Subject clinical status`* is the derived value from the data set of the one's symtomatoallergic character, i.e. particular clinical symptoms, the character of onset, generalized and allergy specific organs disturbance. Subject diagnostic position is composed of the recorded data concerning the diagnostic possibilities for particular subject, i.e. diagnostic in vivo with specific skin test and specific allergens and diagnostic in vitro concerning the specific IgE.
3. *Criteria `Subject diagnostic position`* employed for determining the diagnostic schema, i.e. while subject is at the very first visit and there are now any additional procedures performed, there is a possibility to chose of diagnostic procedures in vitro or/either in vitro for particular patient according individual data set. Subject diagnostic position employed for final classification purposes support determining the treatment schema for allergic patient.

The results obtained should be thoroughly validated. The DM can determine the effectiveness of a course of treatment on the basis of the available information. It should be noted, that only the criteria of the first hierarchical level may be used. Having difficulties in assessment, the DM can use more detailed data from the second level. The possibility of using the information from the second hierarchical level exists even for some first level criteria.

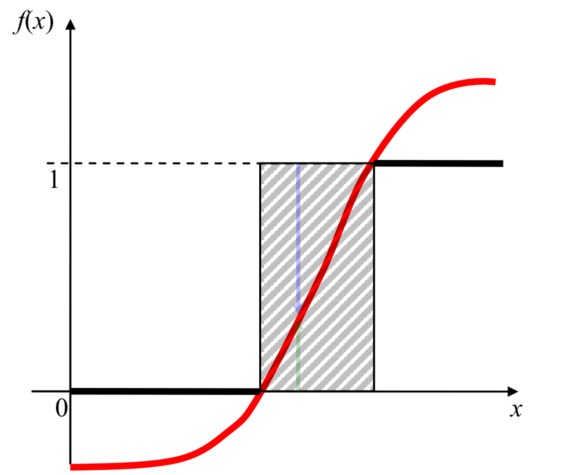

Fig.2. Allergy risk evaluation applying SVM approximated feature space.

Algorithm was verified for 1000 particular experimental patient cases, applying support vector machine (SVM) training/test features selected checkerboard feature selection algorithm was employed of . Minkovski error function was adapted for SVM error elucidation (Fig.2) On SVM basis was created the automatic tool for ROP risk evaluation for premature neonates while the error is equal to 0.1 percent.

4 Conclusion

The comparative analysis while applying the idea of the DM dynamic chain construction allows toget a nearly optimal algorithm by asking the minimum number of questions needed to build a comprehensive classification of optimal diagnostic and treatment schema for particular patient concerning the allergy risk factors environment.

The developed method was validated by solving actual problems of selecting the best alternatives of the available courses of treatment for particular patients.

Our created tool has the potentiality to identify risk level for allergy assessment and define the development process of sensitization – non sensitization – disease and strategic potential of particular objects with high sensitivity considering positive and negative predictive values for particular subject upon the subject risk factors environment.

References:

[1]Mannino DM et al., Surveillance for asthma-United States, 1980–1999. *MMWR CDC Surveill Summ,* Vol. 51, No 1, 2002, pp.1-13.

[2]American Academy of Allergy, Asthma and Immunology (AAAAI). *The Allergy Report: Science Based Findings on the Diagnosis & Treatment of Allergic Disorders*, 1996-2001.

[3] Kemp AS. Allergy prevention – what we thought we knew. *MJA*, Vol. 178, 2003, pp. 254-255.

[4]Sliesoraityte I, Lukosevičius A, Sliesoraitiene V. Corneal thickness factor and artificial intelligent control for intraocular pressure estimation. *Elektronics and electrotechnics,* Vol. 3, No 59, 2005, pp. 37-41.

[5]Ustinovichius L., Balcevich R., Kochin D., Sliesoraityte I. The use of verbal classification for determining the course of medical treatment by medicinal herbs. *AIME:* LNAI 3581, 2005, pp. 276-285.

[6]Ginevicius R., Podvezko V., Mikelis D, Quantitative Evaluation of Economic and Social Development of Lithuanian Regions, *Economics, Research papers,* Vilnius University, ISSN 1392–1258, Vol. 65, 2004, pp.67 – 81.

[7] Zavadskas EK., Peldschus F, Kaklauskas A. *Multiple criteria evaluation of projects in construction*, Vilnius: Technika, 1994.

[8] Ustinovičius L., Zavadskas EK. *Assessment of Investment profitability in construction from technological perspectives*, Vilnius: Technika, 2004.

[9] Hwang CL., Yoon KS, *Multiple Attribute Decision Making Methods and Applications*, Berlin, Heidelberg, New York: Springer – Verlag, 1981.

[10]Larichev O., Kochin D., Kortnev A., Decision support system for classification of a finite set of multicriteria alternatives. *Decision Support Systems.* Vol.33, 2002, pp. 13-21.

[11]Larichev O., Leonov A.: Method CYCLE of serial classification of multicriteria alternatives. *The report of the Russian Academy of sciences*, December, 2000.

[12] Opricovic S., Tzeng GH, Compromise solution by MCDM methods: A comparative analysis of VIKON and TOPSIS, *European Journal of Operational Research,* No. 156, 2004, pp.445 – 455.

[13]Kendall M., *Rank correlation methods*, London: Griffin, 1970.

[14]Solso P., *Cognitive psychology*. Trivola, Moscow, 1996.