DECISION HELP SYSTEM SUPPORTED DATA-MINING METHOD

Péter Tamás, Norbert Szakály
Budapest University of Technology and Economics Mechanical engineering Faculty Department of
Mechatronics, Optics and Engineering Informatics
tamas@mogi.bme.hu

Abstract

Database of scoliosis examination stores a lot of data of patients. The exact numbers and the written opinion of doctors are useable in decision about the resolution of screening process. The moiré method and results of depth scanning are pictures. The conclusions are subjective. The paper is about a data mining based method for qualification of case to help of diagnosis.

Keywords: scoliosis, screening, moiré method, POTSI, computer aided decision

1. Introduction

The aim of the project was to develop a teachable system which is able to store all data of examinations and it is able to suggest a conclusion or a diagnosis to doctors based on the database stored cases.

2. Methods

The heart of the system a database and selection of the nearest data and the interpolation.

2.1 Database of the system

The method of decision making is the known data mining method named n-nearest neighbor[1] and the interpolation technics.[2] Data of examination is stored in database and upon the vector of actual parameters of examination is compared with all of the stored exams data vector and select the k nearest examinations. The nearest examination that where the difference between the stored data vectors and the actual data vector is the minimal one. Conclusion can be defined on the subset with k element, where the known conclusions interpolated with a polynomial as the function of the known parameters. The actual parameters are substituted into the interpolated function value will be the suggested conclusion.

There is a frame system teachable with measured data and accepted decisions. The aim to find the closest cases to actual examination data set (green in the *Figure 1*.). Based this data the system evaluates the conclusion upon the known cases (red in *Figure 1*.).
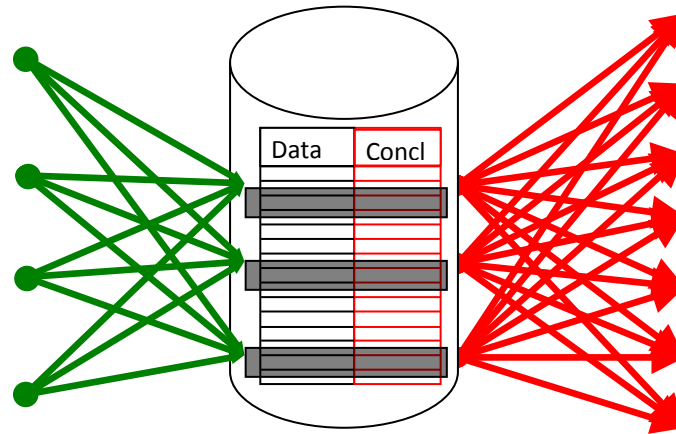
Figure 1. Schema of the system

2.2 Working of the system

The data of all examination are stored in numerical form in database. Concluded data are numerical too. The applied method is a multidimensional interpolation. The first step is to define the set of closest data with help of a query. Let us imagine the decision data of the known cases as the multidimensional function of the known data!

2.3 The developed mathematical background

Data of examinations are defined as elements of intervals. Limits of intervals defined by the possible maximum and minimum of these values $[x_i^{min}, x_i^{max}]$ where $i=1,...N$ and N is the number of known and accepted cases. That means data are in an interval of N dimension.

$$[x_1^{min}, x_1^{max}] \; x \; [x_2^{min}, x_2^{max}] x....x[x_N^{min}, x_N^{max}] = X \tag{1}$$

Similarly the accepted conclusions are in an interval of P dimension.

$$[y_1^{min}, y_1^{max}]x[y_2^{min}, y_2^{max}]x....x[y_P^{min}, y_P^{max}] = Y \tag{2}$$

The basis of work is to find the function φ mapping the known data to known conclusions.

$$\Phi : X \rightarrow Y \; ; \; \Phi(\underline{x}) = \underline{y}; \qquad\qquad \underline{x} \in X \text{ and } \underline{y} \in Y \tag{3}$$

If ɸ function is known we can substitute the questionable examination data in it and the suggestion will be the function values. Because of numerical stability it is worth to project all interval into [0,1]. Let the α_I and β_j the normalizer functions.

$$\alpha_i : [x_i^{min}, x_i^{max}] \rightarrow [0,1] \; ; \; \alpha_i(\underline{x}) = \underline{\xi} \; ;$$
$$\beta_j : [y_j^{min}, y_j^{max}] \rightarrow [0,1] \; ; \; \beta_j(\underline{y}) = \underline{\eta} \; ;$$

$i=1\dots N$ and $j=1\dots P$
$\underline{x} \in X$ and $\underline{y} \in Y$ 　　(4)
$\underline{\xi} \in [0,1]^N$ and $\underline{\eta} \in [0,1]^P$

Using the normalized data we search for the ɸ* function. It can be computed from (3) and (4).

$$\phi^* : [0,1]^N \rightarrow [0,1]^P \; ; \; \phi^*(\underline{\xi}) = \underline{\eta} \; ;$$

$\underline{\xi} \in [0,1]^N$ and $\underline{\eta} \in [0,1]^P$ 　　(5)

The following shows the definition of ɸ* function.

In the first step we search for the function in third order:

$$\phi^*_j(\underline{x}^*) = \sum_{i=1}^{N}(a_{i,j}\xi_i^3 + b_{i,j}\xi_i^2 + c_{i,j}\xi_i + d_{i,j})$$

$j = 1\dots P$ 　　(6)

We can select the $4*N$ closest records to the questionable one based on the Euclidian norm. The interpolation function of ϕ^* is defined with the $4*N$ equations for $4*N$ unknown $a_{i,j}$, $b_{i,j}$, $c_{i,j}$, $d_{i,j}$ coefficient. If the equation system is solvable then we are ready. We can substitute the questionable data end we get the suggestions.

If it is not possible to solve the equation system (there is no enough data or there is a singularity), than the ϕ^* function can be searched for the following form

$$\phi^*_j(\underline{x}^*) = \sum_{i=1}^{N}(a_{i,j}\xi_i^2 + b_{i,j}\xi_i + c_{i,j})$$

$j = 1\dots P$ 　　(7)

In this case we have to select the closest $3*N$ records from the database. If the interpolation ϕ^* function in second order successful, the coefficients $a_{i,j}$, $b_{i,j}$, $c_{i,j}$ can be defined we are ready and ϕ^* function is defined.

If it was not successful because of there is no enough data or there is a singularity, than we search for the ϕ^* function upon the $2*N$ closest records in form of (8).

$$\phi^*_j(\underline{x}^*) = \sum_{i=1}^{N}(a_{i,j}\xi_i + b_{i,j}) \qquad\qquad j = 1...P \qquad (8)$$

When there are *2∗N* closest records and it is successful to compute $a_{i,j}$, $b_{i,j}$ coefficients (*i=1...N*), then ϕ^* defines ϕ.

If it was unsuccessful then it is enough to search the closest record to the actual data. The conclusion part of it will define ϕ^* function and the ϕ function and the suggestion.

3. Used data

There are two special examination included in the system.

3.1 A POTSI (POsterior TrunkSymmetry Index)

Upon POTSI - posterior trunk symmetry index[3] we are able to conclude the deformity of back surface and so the curvature of vertebral in the vertical plane. There are eight points on the back we have to define. It is easy to do however on a photo or on the moiré picture. In a symmetric back surface the POTSI index is zero. The higher POTSI index is the worse. The index under 27.5 it is normal. Upon the experiences POTSI classification is good for asymmetry, but it show a lot of cases of the scoliosis to be healthy. The experiences shows that the POTSI is the function of Cobb angle in 44.6% and the rotation in 26.8%. The POTSI is defined by equation of (9) based on *Figure 2*.

$$POTSI = FAI + HAI \qquad\qquad (9)$$

where definitions of FAI: is the Frontal Asymmetry Index and HAI: Height Asymmetry Index There are the follows:
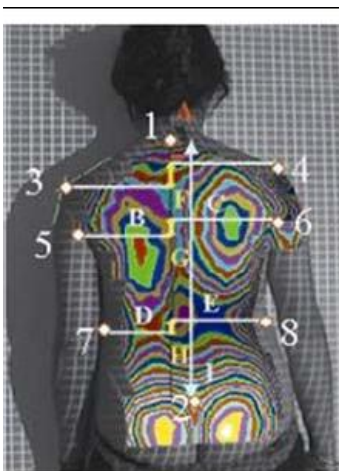

Figure 2. Definition of POTSI

The eight anatomic points are:

1. C7 vertebra
2. upon point of back-end
3. intersection point of tangent of left shoulder and tangent of left hand
4. intersection point of tangent of right shoulder and tangent of right hand
5. left armpit
6. right armpit
7. left waist point
8. right waist point

Upon these

_I_: distance between **_1_**. and **_2_**. points

_A_: length of vertical projection of **_I_** (this is the reference line of following calculations.

_B_: distance of point *5*. from *A* section

_C_: distance of point *6*. from *A* section

_D_: distance of point *7*. from *A* section

_E_: distance of point *8*. from *A* section

_F_: vertical projection of distance from *3*. to *4*.

_G_: vertical projection of distance from *5*. to *6*.

_H_: vertical projection of distance from *7*. to *8*.

From these

$$FAI = FAI_{C7} + FAI_{armpit} + FAI_{waist} \qquad (10)$$

where

$$FAI_{C7} = A*100/I \qquad (11)$$

and

$$FAI_{armpit} = |B\text{-}C|*100/I \qquad (12)$$

and

$$FAI_{waist} = |D\text{-}E|*100/I \qquad (13)$$

and

$$HAI = HAI_{shoulder} + HAI_{armpit} + HAI_{waist} \qquad (14)$$

where

$$HAI_{shoulder} = F*100/I \qquad (15)$$

$$HAI_{armpit} = G*100/I \qquad (16)$$

$$HAI_{waist} = H*100/I \qquad (17)$$

3.2 Prediag method

There was a survey made in frame of Gerinco2 project with more than 300 examinations of students. Upon the test doctors created a new estimation method named prediag.[4]
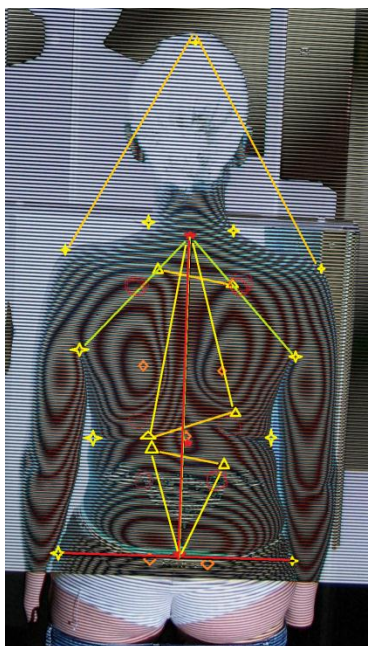
Figure 3. The estimation data

On *Figure 3* the definitions of estimation data are shown. The main contour points are signed with four sided yellow starts.

KFT	maximum height of head,
KNYB, KNYJ	contour points on the neck,
KVB, KVJ	contour on the shoulder,
KHB, KHJ	contour points on the armpits,
KDB, KDJ	contour points on the waist,
KTB, KTJ	contour points on the hams.

The special main points are signed by red pentangular.

C7	vertebra C7
S1	spinal process of vertebra
L	the lower point of the lumbal sector

Other point on the moiré picture signed by yellow triangle

ISOTHPB, ISOTHPJ	proximal iso-lines on left and right thoracic region
ISOTHDB, ISOTHDJ	distal iso-lines on left and right thoracic region
ISOLPB, ISOLPJ	proximal iso-lines on left and right lumbar region

Estimated centrum points of iso curves are signed by orange rectangles. t

BDC	Left dorsal centrum
JDC	Right dorsal centrum
LC	Lumbar centrum

BGC	Left gluteal centrum
JGC	Right gluteal centrum (same as L)

From the above mentioned the healthy case when the C7-S1 line is the same than the VL - vertical line (the sagittal plane) (Vertical Line) and the KTB-KTJ line is the same as the (HL-horizontal line). The measure of compensation is the angle of C7-S1 and the vertical line.

4. Results, Processing of moiré picture

In the software system there is a panel of moire picture (*Figure 4*). On the panel we are able to show and to analyze the stored moire picture.
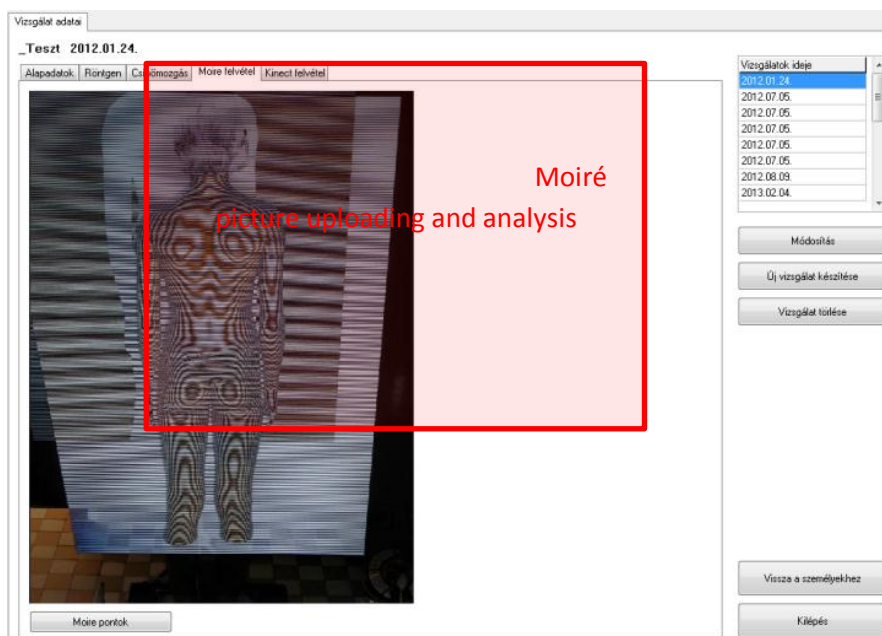
Figure 4. Moiré panel

If we have a new examination or we want to modify an old one we show the dialog of moire examination (*Figure 4.*). There is the possibility to zoom and re-zoom the picture by buttons. With

![POTSI] button we select the POTSI examination. After selection the button ![icon] and the selection one element on list of *Figure 5.* (Hungarian software) we can select a distance to define in the picture *Figure 6*.
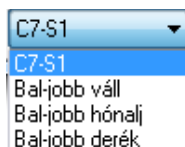

Figure 5. Selection of the POTSI distance

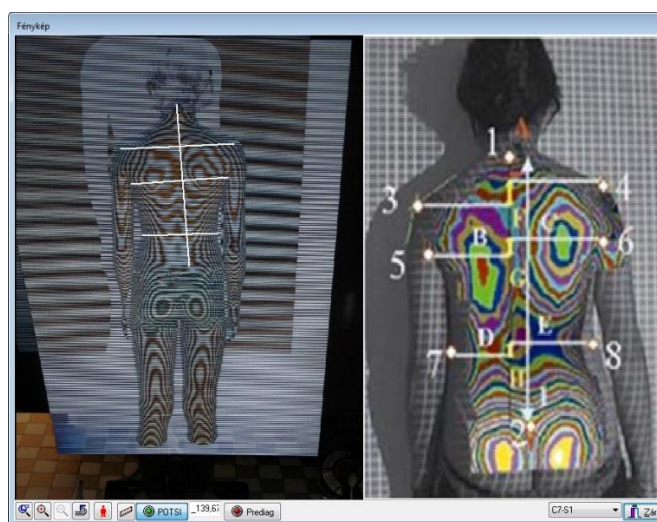After defining the POTSI panel shows the POTSI value.

Figure 6. The POTSI panel

If we choose the [Prediag] button we will get the Prediag dialog (*Figure 7.*)
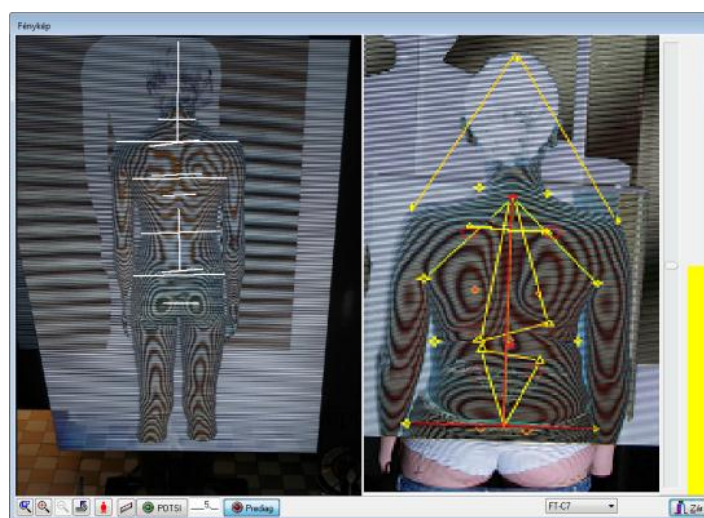

Figure 7. The Prediag panel

Like in the case demonstrated above we can choose a distance by a list shown in *Figure 8.*
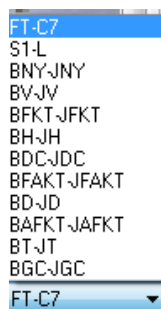
Figure 8. The distances of Prediag

In the case of new examination or modified examination in the right scrollbar of the panel doctor is able to qualify the patient. The healthy case is colored green, the problematic is yellow and the dangerous is red. The system helps to decide in way written in 2. In the case of old examinations the system automatically shows the qualification of the case.

5. Discussion

The presented method worked well in test phase. The next step is the verification in real screening processes. Tuning of setting parameters can be made after the processing of results.

REFERENCES

1. Abonyi J. Adatbányászat a hatékonyság eszköze , ComputerBooks Kiadó Budapest 2006.

2. Dr. Bajcsay P. Numerikus analízis, Tankönyvkiadó, Budapest, 1978

3. Minguez MF, Buendia M, Cibrián RM, Salvador R, Laguia M, Martin A, Gomar F. Quantifier variables of the back surface deformity obtained with a noninvasive structured light method: evaluation of their usefulness in idiopathic scoliosis diagnosis. Eur Spine J 2007;16:73–82.

4. Report of Gerinco2 project