Int. J. Bio-Medical Computing, 21 (1987) 7–12
Elsevier Scientific Publishers Ireland Ltd.

A GENERAL ALGORITHM FOR ASSISTING MEDICAL DIAGNOSIS*

J.C. TOHÁ, W. PALMA, R. MARIANGEL and R. PONCE

Biofísica, Departamento de Física, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Av. Blanco Eucalada 2008, Casilla 487-3, Santiago (Chile)

(Accepted October 30th, 1986)

An algorithm for assisting medical diagnosis is designed; this algorithm is highly reliable and can be given general use. The method is based on the comparison of the column vector of the symptoms of a given patient with each of the columns of disease symptoms which appear on a matrix.

On this matrix, the symptoms have been arranged according to the hierarchic order of each of the symptoms, which facilitates the creation of a file.

Moreover, this system undertakes a simultaneous comparison of all the diseases considered in the differential diagnosis, in comparison to the alternate procedures in which the patient's symptoms are compared in separate succession with each of the vectors of the disease under study.

Keywords: Medical diagnosis computation; Algorithm for medical diagnosis; Software for differential diagnosis

Introduction

In a previous paper we described a simple algorithm for assisting medical diagnosis (Tohá *et al.*, 1986). The basic idea supporting this algorithm was the existence of a hierarchy of values in the symptoms that characterize an illness or a patient, and that those values, expressed in terms of vectors, could be compared.

In the above-mentioned procedure a matrix was constructed with the vectors that grouped the symptoms taken into consideration in the differential diagnosis, and then the vectors were grouped into a ring in which similar diseases appeared next to each other on the matrix. Furthermore, the grouping of the symptoms on the respective vectors or columns corresponded to the hierarchy of the patient's symptoms.

In the next stage a successive one-to-one replacement of each disease vector for the patient vectors was carried out and the results were then evaluated to find out which of the replacements induced the smallest perturbation on the matrix; this being the most probable diagnosis.

---

*This work was partially supported by Departamento de Investigación y Bibliotecas, Universidad de Chile.

0020-7101/87/$03.50
Printed and Published in Ireland

In this paper we simplify and generalize this algorithm (Argyle, 1980), as is shown in the examples that follow. At this point, the diseases appear on the matrix according to similarity, but do not form a ring, because in the similarity ring not always the most similar vectors become immediate neighbours. Therefore, the evaluation of the matrix is simplified and, in order to avoid any ambiguity in its results, it is no longer necessary to construct it according to the hierarchy of the patient's symptoms; it allows for the use of the symptoms vector, of the diseases, which facilitates the creation of a filing system for future reference.

On the other hand, in this paper we compare this method where the most probable disease is identified through the simultaneous analysis of the patient vector against all the disease vectors considered in the matrix, and an alternative approach of successive one-to-one comparisons of the patient's symptoms vector with each of the diseases included in the differential diagnosis, (Ledley and Lusted, 1969; Fagot Largeault, 1985).

Example 1

In this case the patient's symptoms vector follows – in general terms – the hierarchic distribution of the symptoms present in at least one of the diseases considered in the differential diagnosis. In this simple example, let *A, B, C* and *D* be the diseases selected for the differential diagnosis and *a, b, c* and *d* the symptoms considered.

The hierarchic values of the patient's symptoms vector $= A'$ are as follows:

$$a' = 48; b' = 32; c' = 22; d' = 8.$$

The matrix of the diseases considered in the diagnosis has the following configuration, where the symptoms follow the hierarchy of disease *A*.

Symptoms	Diseases			
	A	*B*	*C*	*D*
a	50	30	10	10
b	30	50	30	20
c	20	20	50	30
d	10	10	20	50

We can evaluate the distribution of elements in this matrix through the calculation of an σ value, defined as

$$\sigma = \frac{\sum_{k=-n}^{n} k^2 \cdot Dk}{\sum_{k=-n}^{n} Dk} \qquad k \neq 0, 1$$

where k is the lesser distance between each diagonal and the main diagonal; Dk is the sum of the elements of each diagonal (in such a way that the more distant diagonals have fewer elements participating in their sum). In the example: $\sigma = 1.85864$. The procedure is repeated, using the hierarchic arrangement corresponding to each of the diseases under comparison and observing which set of diseases generates the lowest σ value.

The matrix shows the lowest σ value will be the most sensitive one for analyzing the diagnosis. In the example, the lowest σ values obtained for the 4 hierarchic arrangements of the diseases studied are the following: For disease A (**ABCD**) $\sigma = 1.85864$; B (**BACD**) $\sigma = 1.94001$; C (**CDEA**) $\sigma = 2.02709$; D (**DCBA**) $\sigma = 1.85864$. The lowest value (1.85864) corresponds to the following arrangements of diseases A or D: **A–B–C–D** or **D–C–B–A**.

In one of these matrix patterns (in this case A–B–C–D), we carry out a successive substitution of each disease vector value for the values of the patient vector. The σ value obtained after the substitution of vectors **A**, **B**, **C** and **D** for the patient vector was: 1.85374, 1.87131, 1.94188 and 2.13201. Therefore, the most probable diagnosis corresponds to disease A, which when substituted gives the lowest σ value.

The ratio: min. σ diseases/min. σ substitution: 1.85864/1.85374 = 1.00264 shows the degree of coincidence existing between disease A and patient (A').

Example 2

In this example, we worked with the same diseases as in the previous example: A, B, C and D, but in this instance the patient, B in this case, is an atypical patient with a symptomatic hierarchy represented by the following vector: $\mathbf{b}' = 40$, $\mathbf{c}' = 35$, $\mathbf{d}' = 25$ and $\mathbf{a}' = 10$. Just as in the previous case, we selected the matrix configuration: **A–B–C–D** corresponding to the hierarchy of disease A, because it has the lowest σ value.

After making the successive substitution of the disease columns for the patient-vector, we obtained the following replacement values for C, B, D and A: 1.87326, 1.90454, 1.93071 and 2.01585. This means that disease C corresponds to the most probable diagnosis for patient B'. The ratio min. σ disease/σ substitution $C = 1.85864/1.87326 = 0.99220$ shows the degree of correspondence existing between patient B' and the assigned C disease.

Example 1: Analyzed according to the method of successive comparison of the patient's symptoms vector and each disease vector. In this case, the evaluation of the degree of similarity between the patient-vector and each of the disease vectors is carried out by analyzing both vectors and their hierarchies, in pairs, and then by adding the results of the respective modules of differences in magnitude of similar symptoms and multiplying them by the differences of positions of those symptoms appearing on both vectors (sum of moments), as follows:

Hierarchic vector of the symptoms disease *A*:	Hierarchic vector of the symptoms patient *A′*:	Diff. in values	+1X moments	Diff. in position
a - - - - - 50	**a′** - - - 48	**a** - - - **a′** ($\lvert 2\rvert$ + 1)	X	($\lvert 0\rvert$ + 1) = 3
b - - - - - 30	**b′** - - - 32	**b** - - - **b′** ($\lvert 2\rvert$ + 1)	X	($\lvert 0\rvert$ + 1) = 3
c - - - - - 20	**c′** - - - 22	**c** - - - **c′** ($\lvert 2\rvert$ + 1)	X	($\lvert 0\rvert$ + 1) = 3
d - - - - - 10	**d′** - - - 8	**d** - - - **d′** ($\lvert 2\rvert$ + 1)	X	($\lvert 0\rvert$ + 1) = 3
		Total sum of moments value:		12

Similarly, the comparison of the ordered vectors of diseases *B, C* and *D* with the ordered patient vector, gives the following sum of moments values: 82, 272 and 372.

In the above evaluation the patient-vector is more closely related to disease *A*, followed by diseases *B, C* and *D* in exactly the same order as seen in the matrix substitution method.

Example 2: Analyzed according to the method of successive comparison of the patient vector with each of the disease vectors. In this case patient *B′* has a special hierarchy of symptoms that are different from each of the diseases against which it is compared, by means of a vector-to-vector comparison. The differences found were: 61, 138, 148 and 250, when patient *B′*s vectors was compared with diseases *C, B, D* and *A*. Once again the order obtained corresponds to the order resulting from the procedure of disease matrix substitution.

To illustrate this, we show the calculation of the comparison of the patient vector **B′** with disease vector **C**.

Hierarchy of symptoms of disease *C*	Hierarchy of symptoms of patient *B′*	Diff. in in values	+1X moments	Diff. in position
c - - - - - 50	*b′* - - - 40	*c* - - *c′* ($\lvert 15\rvert$ + 1)	X	($\lvert 1\rvert$ + 1) = 32
b - - - - - 30	*c′* - - - 35	*b* - - *b′* ($\lvert 10\rvert$ + 1)	X	($\lvert 1\rvert$ + 1) = 22
d - - - - - 20	*d′* - - - 25	*d* - - *d′* ($\lvert 5\rvert$ + 1)	X	($\lvert 0\rvert$ + 1) = 6
a - - - - - 10	*a′* - - - 10	*a* - - *a′* ($\lvert 0\rvert$ + 1)	X	($\lvert 0\rvert$ + 1) = 1
		Total sum of moments:		61

Analysis of the degree of similarity of the group of disease considered in the differential diagnosis

We have seen that matrix $A-B-C-D$, which follows the hierarchy of disease A, and that matrix $D-C-B-A$ which follows the hierarchy of disease D, show the lowest σ value (1.85864). This σ value can be compared with an ideal value that corresponds to an arrangement in which the components of each disease column are shifted vertically so as to minimize the diagonal variation in each independent column. In this case the ratio σ ideal matrix/σ $A-B-C-D$ matrix $= 1$ shows the homogeneity of the group of diseases included in the differential diagnosis.

For the successive comparison of each disease with each one of the others, we obtain an arrangement that groups together the most similar diseases. This arrangement coincides with the similarity arrangement obtained in the disease matrix method. The arrangement $A-B-C-D$ has the following sum of the differences of moments: 86, 220 and 158, corresponding to $A-B$, $B-C$ and $C-D$.

Discussion

In this paper we describe a simplified form of the algorithm for assisting medical diagnosis published previously. The calculation of the σ value of a disease matrix is now reduced to the sum of the moments of its diagonals, without having to take into account the number of elements they contain. With this procedure the matrix structure formed for the lowest σ value corresponds to an arrangement where the most similar diseases become neighbours in a similarity frame without resulting in a similarity ring.

The advantages of this new method (where the order of the columns in the matrix coincides strictly with the highest similarity of the immediate neighbours) is that we can faithfully substitute each disease for the patient vector, and obtain a diagnosis by working with a matrix in which the hierarchy of symptoms corresponds to one of the diseases studied.

In our previous paper, where the disease matrix grouped diseases in a similarity ring, if we wanted to avoid any ambiguity in the results of the substitution (especially in the case of atypical patients), it was recommended to arrange the disease matrix according to the patient's symptomatic hierarchy. This made it extremely difficult to develop a file of previous similar diseases which is required in a differential diagnosis. On the other hand, both the calculation of the σ value and a successive one-to-one comparison give results which are notably coincidental where diagnosis and arrangement are concerned.

In any case, the use of the matrix calculation method permits a simultaneous comparison of the patient vector with all the diseases involved, and facilitates a speedy computer evaluation, storage of data and the organization of software that groups similar diseases together in such a way that the physician can rapidly check a diagnosis by substituting the patient's vector on a frame of similar diseases that have already been programmed.

References

Argyle, E., 1980, A similarity ring for amino acids based on their evolutionary substitution rates, *Orig Life,* 10, 357–361.

Fagot Largeault, A., 1985, La simulation de raisonement médical, *La Recherche,* 170, 1176–1187.

Ledley, R.S. and Lusted, L.B., 1959, Reasoning Foundation of Medical Diagnosis, *Science,* 130, 9–21.

Tohá, J.C., Obando, G., Soto, M.A. and Sepúlveda, A., 1986, A simple algorithm for assisting medical diagnosis. *Int J Biomed Comput,* 19, 289–293.