

Research Article

Building Expert Medical Prognostic Systems Using Voronoi Diagram

Maria A. Ivanchuk¹ and Igor V. Malyk²

¹Department of Biological Physics and Medical Informatics, Bukovinian State Medical University, Kobylianska Street 42, Chernivtsi 58000, Ukraine

²Department of the System Analysis and Insurance and Financial Mathematics, Yuriy Fedkovych Chernivtsi National University, Universitetska Street 12, Chernivtsi 58012, Ukraine

Correspondence should be addressed to Maria A. Ivanchuk; mgracia@ukr.net

Received 27 June 2014; Revised 14 December 2014; Accepted 1 February 2015

Academic Editor: Neeraj Mittal

Copyright © 2015 M. A. Ivanchuk and I. V. Malyk. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The method of building expert systems for medical prediction of severity in patients is purposed. The method is based on using Voronoi diagrams. Examples of using the method are described in the paper.

1. Introduction

Development of mathematical approaches for prediction in medicine was developed by Fisher, the father of the linear discriminant analysis [1]. Currently, there are many approaches to solving this problem such as cluster analysis [2], the construction of predictive tables [3], image recognition, and linear programming. Cluster analysis is commonly used for solving the tasks of medical prediction. The aim of a cluster analysis is to partition a given set of data or objects into clusters. This partition should have the homogeneity within the clusters and heterogeneity between clusters [4]. But cluster analysis has a significant disadvantage. It refers to the methods without teacher.

We purpose the method of building the prognostic system, which uses available information for teaching the expert system.

Let us have two sets of points $A = \{a_i = (a_i^1, a_i^2, \dots, a_i^n), i = \overline{1, m_A}\}$ and $B = \{b_i = (b_i^1, b_i^2, \dots, b_i^n), i = \overline{1, m_B}\}$ in Euclidean space R^n , where m is the number of points in the set. Set A is the training sample, which includes the patients with severity; set B is the training sample, which includes the patients without severity. There are n parameters (factors which affect the severity) known for each patient. Our task is to separate the space R^n into two half-spaces R_A^n (patients

with severity) and R_B^n (patients without severity) that for each point $x_i \in R^n$ determine its belonging to one of the half-spaces with predetermined significance level α . We will check the expert system using the control sample $Z = \{z_i = (z_i^1, z_i^2, \dots, z_i^n), i = \overline{1, m_Z}\}$.

We must use a smaller number of parameters to obtain the largest plausibility value for expert system.

2. Methods

We will use the Voronoi diagram [5] to solve the task of expert system's building. Let us concatenate sets A and B and build Voronoi diagram for the set $D = A \cup B$. Let us have a point $d_i \in D, i = \overline{1, m_A + m_B}$. Let $d_i \in A$. The same reasoning will be for $d_i \in B$.

Let V_i be Voronoi polygon for point d_i . The points for which Voronoi polygons have neighboring facets with polygon V_i will be called *point d_i 's nearest neighbors*. The set of all point d_i 's nearest neighbors will be denoted by NV_i .

Point d_i will be called *the internal point* of set A if all its nearest neighbors belong to set A

$$\forall x \in NV_i : x \in A. \quad (1)$$

There are the following cases for Voronoi polygon V_i .

- (1) All point d_i 's nearest neighborhoods belong to set A

$$\forall x \in NV_i : x \in A. \quad (2)$$

In this case point d_i is the internal point of set A .

- (2) All point d_i 's nearest neighborhoods belong to set B

$$\forall x \in NV_i : x \in B. \quad (3)$$

In this case point d_i is the outlier of set A .

- (3) There are points belonging to set A and the points belong to set B among point d_i 's nearest neighborhoods. In this case point d_i is the boundary point of set A or point d_i with one or several neighborhoods being the outliers of set A .

If there is a way from point d_i to any internal point of set A passing only through the points of set A , point d_i is the *boundary point* of set A . In a different case point d_i is the *outlier* of set A .

Let us eject outliers from set A (patients with severity) and build new Voronoi diagram. The diagram separates space R^n into two half-spaces R_A^n (patients with severity) and R_B^n (patients without severity). We will assign patient z from control set Z to the patients with severity if point z is in the Voronoi polygon of any point of set A ; in the different case we will assign patient z to the set of patients without severity.

Let us sort parameters according to Kulbak's information measure [3] and build Voronoi diagrams for different space dimension $n_i = 2, \dots, n$. To find the best of expert systems we will use Zagoruiko's likelihood measure [6].

Let point $z \in Z$. Euclidian distance between point z and the nearest point $a_i \in A$ is equal to $R(z, a_i)$. Euclidian distance between point z and the nearest point $b_j \in B$ is equal to $R(z, b_j)$. Than similarity signed measure (charge) of point z and set A is

$$F_{A/B}(z) = \frac{R(z, b_j) - R(z, a_i)}{R(z, a_i) + R(z, b_j)}. \quad (4)$$

Similarity measure $F_{A/B}(z)$ may range from -1 to 1 . Point z is assigned as the part of set A if $F_{A/B}(z) > 0$. The high value of the measure $F_{A/B}(z)$ indicates the high similarity between point z and set A .

Let Z_A be the set of control group patients with severity; Z_B is the set of control group patients without severity, $Z = Z_A \cup Z_B$.

We denote Z_{A^+} is the set of patients of Z_A which were correctly assigned as the patients with severity; Z_{A^-} is the set of patients of Z_A which were incorrectly assigned as the patients without severity (underdiagnosis cases); Z_{B^+} is the set of patients of Z_B which were correctly assigned as the patients without severity; Z_{B^-} is the set of patients of Z_B which were incorrectly assigned as the patients with severity (overdiagnosis cases).

Then $F_{Z_{A^+}}$ is the sum of similarity measures of the points of set Z_{A^+} and set A :

$$F_{Z_{A^+}} = \sum_{i=1}^{m_{Z_{A^+}}} F_{A/B}(z_i), \quad z_i \in Z_{A^+}, \quad (5)$$

$F_{Z_{B^+}}$ is the sum of similarity measures of the points of set Z_{B^+} and set B :

$$F_{Z_{B^+}} = \sum_{i=1}^{m_{Z_{B^+}}} F_{B/A}(z_i), \quad z_i \in Z_{B^+}, \quad (6)$$

$F_{Z_{A^-}}$ is the sum of similarity measures of the points of set Z_{A^-} and set B :

$$F_{Z_{A^-}} = \sum_{i=1}^{m_{Z_{A^-}}} F_{B/A}(z_i), \quad z_i \in Z_{A^-}, \quad (7)$$

and $F_{Z_{B^-}}$ is the sum of similarity measures of the points of set Z_{B^-} and set A :

$$F_{Z_{B^-}} = \sum_{i=1}^{m_{Z_{B^-}}} F_{A/B}(z_i), \quad z_i \in Z_{B^-}. \quad (8)$$

The likelihood measure of the expert system is

$$F = \frac{(F_{Z_{A^+}} + F_{Z_{B^+}})}{m_{Z_{A^+}} + m_{Z_{B^+}}} - \frac{(F_{Z_{A^-}} + F_{Z_{B^-}})}{m_{Z_{A^-}} + m_{Z_{B^-}}}, \quad (9)$$

if the aim of the expert system is the differential diagnostic of two similar diagnoses.

And

$$F = \frac{F_{Z_{A^+}}}{m_{Z_{A^+}}} - \frac{F_{Z_{A^-}}}{m_{Z_{A^-}}}, \quad (10)$$

if the aim of the expert system is finding the patients with severity.

We will use the Akaike information criterion [7]

$$AIC = 2n - 2 \ln(F) \quad (11)$$

to find the optimal ratio of the likelihood of model and the quantity of using parameters. The best expert system is the system with the least value of AIC. In other words, the best expert system is the system which uses the least number of parameters to have the greatest likelihood.

Let us formulate the following.

Algorithm for Modelling the Prognostic System

- (1) Sort parameters using Kulbak's information measure [3].
- (2) For each $n_i = 2, \dots, n$,
 - (i) build Voronoi diagram for the set $D = A \cup B$ in the space R^{n_i} ;
 - (ii) eject the outliers of set A ;
 - (iii) build the new Voronoi diagram;
 - (iv) if there are no outliers in set A go to step (v); else return to step (ii);
 - (v) check the system on the control set Z ;
 - (vi) calculate the likelihood measure F of the expert system;

TABLE 1: Building the expert system for abdominal surgery patients.

n	Outliers	F	AIC	Underdiagnosis cases	Overdiagnosis cases
2	1 (1.4%)	0.575	3.108	1 (16%)	2 (40%)
3	0 (0%)	0.389	4.889	1 (16%)	0 (0%)
4	0 (0%)	0.352	6.091	1 (16%)	0 (0%)
5	0 (0%)	0.627	5.933	2 (30%)	0 (0%)
6	0 (0%)	0.639	6.897	2 (30%)	0 (0%)
7	0 (0%)	0.684	7.761	2 (30%)	0 (0%)
8	0 (0%)	0.611	8.987	2 (30%)	1 (20%)

The best expert system was built using $n = 2$ parameters.

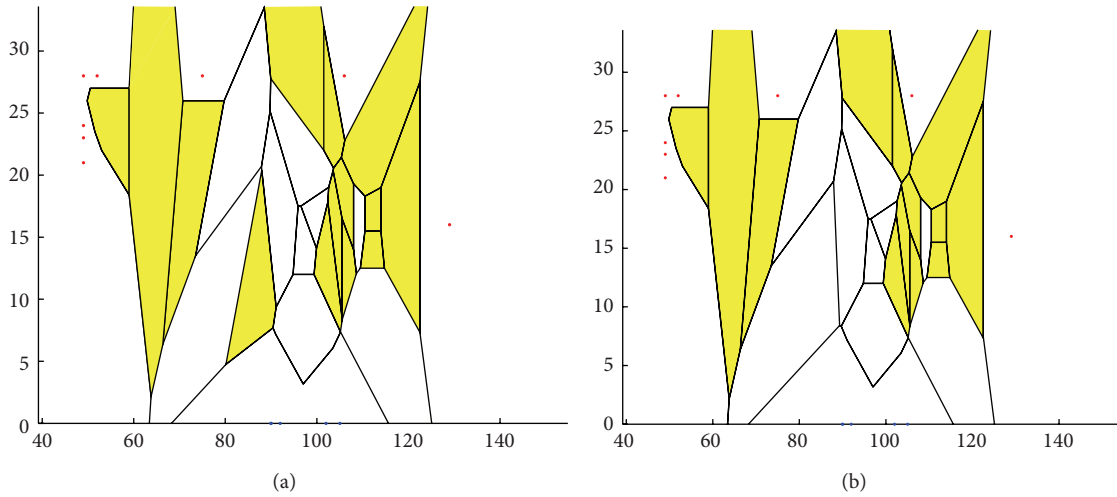


FIGURE 1: Voronoi diagram for abdominal surgery patients before (a) and after (b) ejecting the outliers, $n = 2$.

(vii) calculate the Akaike information criterion AIC_{n_i} .

(3) Find n^* : $AIC_{n^*} = \min_{n_i=2, \dots, n}(AIC_{n_i})$.

(4) As the expert system use Voronoi diagram in the space R^{n^*} . If point $z \in R_A^{n^*}$, assign it to set A. If point $z \in R_B^{n^*}$, assign it to set B.

Let us calculate the complexity of the algorithm. There are several ways to find Voronoi diagrams, one of which is known as Fortune’s algorithm [8]. Its complexity is $O(m \log m)$, where $m = m_1 + m_2$. The complexity of finding the likelihood measure F is $O(m_Z \log m)$, because we must find the similarity measure $F_{A/B}(z)$ for each point $z \in Z$. Step (2) is repeated $n - 1$ times and the complexity of the algorithm is $O((n - 1)m \log m)$. Since $n \ll m$, the total complexity of the algorithm is $O(m \log m)$.

3. Results

3.1. *The Expert System of Predicting the Presence of Severity in Abdominal Surgery Patients.* We built expert system using 8 parameters. Training sample consists of 28 patients with severity and 15 patients without severity. Control test consists of 8 patients with severity and 5 patients without severity. The level of significance was $\alpha = 0,01$. In this case the aim of

the expert system is finding the patients with severity; therefore we use formula (10) to find likelihood measure.

Voronoi diagram for $n = 2$ is represented on Figure 1. The results of calculations are represented in Table 1.

3.2. Researching the Anthropological Parameters in Teenagers.

We built expert system using 8 parameters. Training sample consists of 38 girls and 14 boys. Control test consists of 8 girls and 5 boys. The level of significance was $\alpha = 0,01$. In this case the aim of the expert system is the differential diagnostic of two similar diagnoses; therefore we use formula (9) to find likelihood measure.

Voronoi diagram for $n = 2$ is represented on Figure 2. The results of calculations are represented in Table 2.

4. Conclusions

The method of building expert medical prognostic systems is brought forward. The method is based on building Voronoi diagram in Euclidean spaces of different dimensions. The resulting expert systems are checked on the test samples. The expert system with the least value of the Akaike information criterion is accepted as the best system.

The described method is applied in practice to predict the presence of severity in abdominal surgery patients and gives 84% correct results for the patients with severity from

TABLE 2: Building the expert system for researching the anthropological parameters in teenagers.

n	Outliers	F	AIC	Errors in the control sample
2	2 (4.2%)	0.0271	9.216	2 (15.4%)
3	0 (0%)	0.1147	7.331	2 (15.4%)
4	0 (0%)	0.1007	8.591	2 (15.4%)
5	0 (0%)	0.1212	9.221	2 (15.4%)
6	0 (0%)	0.2587	8.704	3 (23%)
7	0 (0%)	0.2121	10.101	3 (23%)
8	0 (0%)	0.1946	11.274	3 (23%)

The best expert system was built using $n = 3$ parameters.

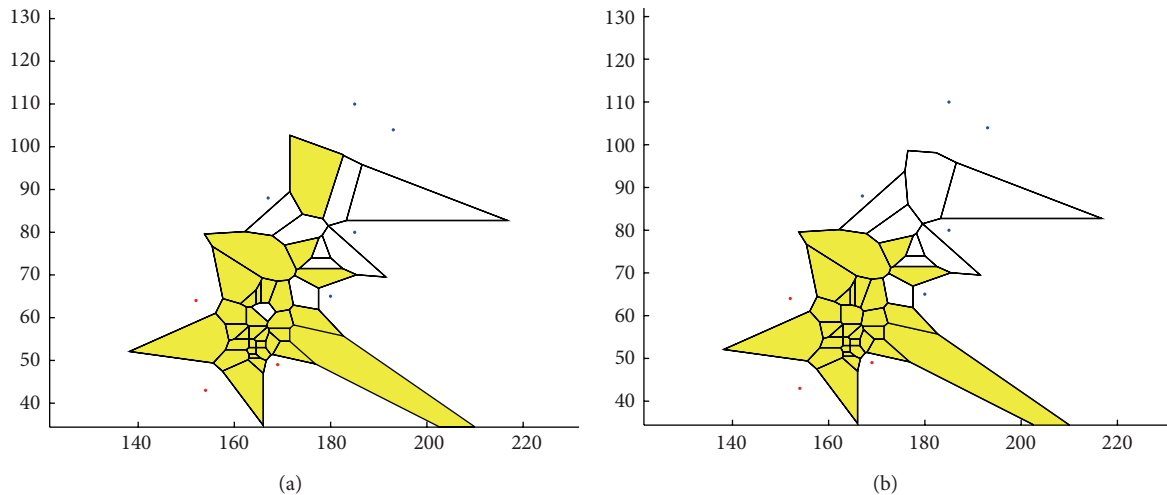


FIGURE 2: Voronoi diagram for anthropological parameters in teenagers before (a) and after (b) ejecting the outliers, $n = 2$.

the control sample. The expert system for researching the anthropological parameters in teenagers with 84.6% correct results was built, using the introduced method.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] R. A. Fisher, *Contributions to Mathematical Statistics*, John Wiley & Sons, New York, NY, USA, 1950.
- [2] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*, John Wiley & Sons, 5th edition, 2011.
- [3] E. V. Gubler and A. A. Genkin, *Primenenie Neparаметрических Критериев Статистики в Медико-Биологических Исследованиях (Using the Nonparametric Statistic Criteria in Medical and Biological Researches)*, Leningrad, Russia, 1973.
- [4] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, John Wiley & Sons, New York, NY, USA, 1999.
- [5] F. Preparata and M. Shamos, *Computational Geometry—An Introduction*, Springer, 1985.
- [6] N. G. Zagoruiko, I. A. Borisova, V. V. Dyubanov, and O. A. Kutnenko, "A quantitative measure of compactness and similarity in a competitive space," *Journal of Applied and Industrial Mathematics*, vol. 5, no. 1, pp. 144–154, 2011.
- [7] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [8] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry*, Springer, 2000.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

