THE JOURNAL OF PREVENTIVE MEDICINE
2003; 11 (1): 47-52

MYCOBACTERIUM CLASSIFICATION USING MULTIVALUED LOGIC

Lucian Vasile Boiculese[1], Olivia Simona Dorneanu[2]

1. Medical Informatics and Biostatistics Department, 2.Microbiology Department
University of Medicine and Pharmacy "Gr.T.Popa"

Abstract. Usually people think and learn creating similarities between facts, processes and events. The multiple value (fuzzy) logic expresses how we reason with vague rules. A fuzzy system is made up of rules that maps inputs to the outputs. Multivalued logic systems cope with the lack of information related to the belonging to a multitude. This type of incertitude is present in biological processes. Fuzzy systems are suited in classifications, modeling and controlling tasks with applications in microbiology. This paper describes classification of mycobacteria by two methods: fuzzy logic and classical selection (using a metric distance).
Key-words: ***Classification, mycobacteria, fuzzy logic, metric distance***

Rezumat. De obicei oamenii gândesc şi învaţă făcând asemănări între fapte, procese şi evenimente. Logica multivalorică (fuzzy) exprimă cum gândim folosind reguli vagi. Un sistem fuzzy este format din reguli care creează legături între intrări şi ieşiri. Sistemele logice multivalorice lucrează bine cu lipsa de informaţie datorată apartenenţei la o mulţime. Acest tip de incertitudine este prezent în procesele biologice. Sistemele fuzzy sunt adecvate clasificărilor, modelărilor şi proceselor de control cu aplicaţii în domeniul microbiologiei. În această lucrare este prezentată clasificarea micobacteriilor prin două metode: logica fuzzy şi metoda clasică (bazată pe o distanţă metrică).
Cuvinte-cheie: Clasificare, micobacterii, logica fuzzy, distanţa metrică

INTRODUCTION

Tuberculosis is a public health problem worldwide, but mainly in developing countries. With the increasing number of immune compromised hosts, etiological spectrum of mycobacteriosis is diversifying in Romania, too.

Accurate identification of mycobacteria is mandatory for the adequate treatment of the infections they have produced, since atypical mycobacteria are resistant to some of the first-line antituberculosis drugs (1). Rapid and accurate identification of mycobacteria is possible using their classification based on the chemical composition of the cell wall.

We have chosen two methods for determining the mycobacteria type:

1) The fuzzy system is governed by a number of rules that must be settled by the user (at a first glance this must be equal with the number of classes) (2,3). The patterns for the inputs (all numeric values) and outputs were deduced from experimental results. The backpropagation algorithm is applied for training (4,5).

2) The classic technique is based on the measurement of a distance

between the input and the class center. The class center is deduced by averaging the characteristics of elements, which define that class.

The standard classification technique depends on the used metric distance, while for the fuzzy system the number of rules must be carefully chosen because there is a possibility of failing in learning the classes' features.

The aim of this paper is to obtain good mycobacteria classification based on the percentage of the chemical compounds in their cell wall.

MATERIAL AND METHODS

There were 33 chemicals possibly present in the mycobacterium cell wall in our experiments.

For example, *Mycobacterium avium-intracellulare* contains 13 chemical compounds, as shown in the table 1.

Table 1. Composition of *Mycobacterium avium-intracellulare*

NO.	*INDEX*	*FEATURE NAME*	*%*
1	9	14 : 0 FAME	4.03
2	12	15:0 FAME	0.59
3	15	16 : 1 w9c FAME	0.93
4	16	16:1 w7c FAME	1.48
5	17	16 :1 w6c FAME	6.74
6	19	16 : 0 FAME	35.11
7	25	18 : 1 w9c FAME	18.65
8	27	18 : 0 FAME	4.18
9	28	10Me-18 : 0 TBSA FAME	13.3
10	30	20 : FAME	0.52
11	31	SUMMED FEATURE # 1	0.69
12	32	SUMMED FEATURE # 2	2.24
13	33	SUMMED FEATURE # 3	11.12

There are 33 inputs (chemical compounds), meaning that the classification must be made in a 33 dimensional space. The system output value must conclude the class number, which the mycobacterium belongs to. There are two possibilities of creating this system. One is to have only one output that deduces a number equal with the class number, while the other method consists in creating a system with c outputs (where c represent the number of classes). In the second method, for each class only one output must be activated for a mycobacterium type (this method was tested in our experiments).

The classic grouping method consists in computing a metric distance between the input and the class center. The minimum distance found specifies that the similitude between the input and that class is maximum. Thus, it is logical to decide that the sample contains a mycobacterium from that class.

In order to classify the mycobacteria, the mean percent composition of each substance and for each group (class) is computed. The mean value illustrates the feature of each class and represents the class center.

A metric distance between the input and the class center is computed. The input is denoted by: $x = (x_1, x_2, x_3, \ldots, x_n)$, were $n = 33$.

The class center is obtained by averaging the chemical compounds` percentage:

$$Cj = \left(c_1^j, c_2^j, \ \ldots \ , c_n^j\right)$$

Denote $d(x,y)$ a metric distance between x and y. One way to compute the distance is the Euclidean formula:

$$D_x^j = \sqrt{\sum_{i=1}^{n}\left(xi - c_i^j\right)^2}$$

, where D_x^j stands for the distance between the input x and the j class; x_i is the i input; j represents the class number; c_i^j depicts the i center of the class j.

This distance reflects the discrepancy between the input and the particular class. The exponential function with a negative distance exponent is applied in order to reflect the similarity between the class and the input. The next formula depicts the similarity level:

$$S_j = e^{-D_x^j}$$

, Sj reflects the similarity level between the x input and the j class.

For the mycobacterium class selection, the Sj values between the input and each class center are computed. Finally, it is deduced that the present mycobacterium is the one with the maximum similarity level.

The other classification method is created using a fuzzy system. Fuzzy systems and neural networks have much in common as usually a fuzzy system can be created by a neural network and a neural network can be simulated by a fuzzy system (6,7).

The fuzzy system is used in order to check its ability and advantages in classifications and because the experimental values are sometimes close enough to be confused.

The fuzzy system is made up of independent cells of rules and each rule has 33 inputs and only one output. We realized a neuro-fuzzy system because the implementation of a fuzzy system can be made with a neural network. This way it is possible to apply the known optimization methods, developed for neural networks. The used membership function must be differentiable and that's why the Gaussian type is chosen.

Each rule is characterized by the inputs, with their means and standard deviations. The optimization process is applied on the membership functions' characteristics: the center and the dispersion of the Gaussian function. The number of outputs is equal with the number of classes. For each class only one output must be activated in a certain moment if the input matches that class. This way the system output number of neurons is settled equal with the number of *Mycobacterium* species.

The backpropagation algorithm with momentum element is applied for training the neural network (8).

In order to set up the network it is necessary to define the number of

rules for each network cell. This number must be carefully chosen, as it is very important for the network functionality.
The entire system is depicted in figure 1.

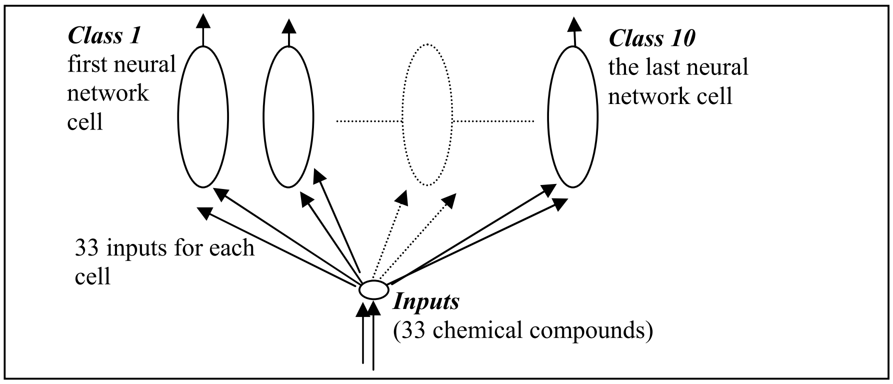

Fig. 1 Neural network system classification

RESULTS AND DISCUSSION

The chemical composition of the mycobacterial cell wall can be analyzed by gas chromatography (GC) and may be used for the classification and identification of mycobacteria. Fatty acid methyl esters (FAME) resulted from the mycobacterial cell wall fatty acids during sample processing (mycobacterial culture) are isolated by GC, analyzed further by mass-spectrometry and identified through their specific retention times (9). The resulted FAME patterns are the base for the classification of mycobacteria.
The number of specific chemical compounds is 33 in the first experiment and the number of classes is 10. These numbers differ depending on the characteristics of the studied problem.
The classification tests were applied on several mycobacteria. The diagnosis accuracy is an important task. Therefore the next four similar mycobacteria were chosen to be classified:

1) *Mycobacterium chelonae* subsp. *abcessus*,
2) *Mycobacterium chelonae* subsp. *chelonae*,
3) *Mycobacterium fortuitum* GC subgroup A,
4) *Mycobacterium fortuitum* GC subgroup B.

A 10% variation on the chemical composition of these mycobacteria was considered. As this variation increases, the classes may overlap and the system may fail in determining the mycobacterium species.
The fuzzy system was trained till the misclassification was less than 5%. Thus 1000 epochs of training were necessary for learning the classes' characteristics. To test the system in more difficult conditions we have increased the percent of variation on the chemical composition to 15%. The

training process was repeated for the new data. The system managed again to correctly identify the mycobacteria, with less than 5% misclassifications after more training. There were needed about 1500 training epochs. The number of training epochs depends on the start set point that is randomly chosen.

The second system that uses a metric distance was less accurate than the fuzzy system. The variation of the chemical composition affects the classification operation (with this system, 7% misclassifications were achieved).

This result was expected as it is known that fuzzy systems are suited to model nonlinearities. The fuzzy system complexity may be enlarged (adding more rules to the system) in order to cope with more difficult real situations.

CONCLUSIONS

The two methods presented for mycobacteria classification produced good results.

One advantage of multivalued system is the ability of learning from examples. Thus, after learning the features of the classes, the neural fuzzy classifier was very good in performing its task, overrunning the metric distance method. The possibility of increasing the number of rules and readapting on the new situations brings a second advantage on the fuzzy systems.

Backpropagation learning method has the drawback of failing in local minims during training. That's why the training process must be restarted trying a new optimization.

As it was anticipated, the classical technique, based on the metric distance, was easier to implement but less accurate in classification providing acceptable results, also.

REFERENCES

1. Young DB, Duncan K: *Prospects for new interventions in the treatment and prevention of mycobacterial disease.* Annu. Rev. Microbiol. 1995; 49: 641-673.
2. Pensiero S, Accardo AP, Vlahinic S, Salvetat ML: *Neural network for detecting keratoconus using corneal topography.* Proceedings of the 8th International IMEKO Conference on Measurement in Clinical Medicine-Croatia 1998; 6.7-6.9.
3. Wang LX: *Adaptive Fuzzy Systems and Control - Design and Stability Analysis,* Prentice Hall , 1994.
4. Sofrom E, Bizon N, Ionit S, Raducu R: *Fuzzy systems for intelligent control.* In: Fuzzy System Control. All Press, 1998, 35-92 (in Romanian).
5. Widrow B, Stearns SD: Adaptive *Signal processing,* Prentice Hall, 1985.
6. Wieland F, Aliev F: *Neuro-Fuzzy-Genetic Adaptive Control System*. Fourth European Congress on Inteligent Techniques and soft Computing, Achen Germany, 1996, 245-250.
7. Witold P: *Theoretical developments in the construction of fuzzy controllers*. In: Fuzzy Control and Fuzzy Systems. Taunton, Somerset, England: Research Study Press, 1989, 141-166
8. Truta AM, Slate M: *Neural nets activity from neuroinformatics theory viewpoint*. Scientific Bulletin of Polytechnic University Timişoara 2002; 47 (61): 105-110.

9. Parez JJ, Fauville-Dufaux M, Dossogne JL, de Hoffmann E, Pouthier F: *Faster identification on mycobacteria using gas liquid and thin layer chromatography*. Eur. J. Clin. Microbiol. Infect. Dis. 1994; 13 (9): 717-725.