INFORMATICA, 1995, Vol. 6, No. 1, 85–92

STRONGLY SELFGUESSING FUZZY CLASSIFIERS

Sergey VATLIN

Belarusian State University of Informatics and Radioelectronics
220027 Minsk, P. Brovka St. 6, Republic of Belarus

Abstract. The problem of construction of the fuzzy classification models (fuzzy classifiers) with high generalization ability is discussed. The strong selfguessing property of fuzzy classificational models is introduced and examined. It is proved that this characteristic doesn't form a full system of restrictions, i.e., for the unambiguous detection of the most valid fuzzy classifier (among the set of fuzzy classifiers agreed with arbitrary learning set) it is necessary to use additional "regularizing" restrictions.

Key words: fuzzy classifier, learning set, strongly selfguessing fuzzy classifier, generalized training sequence, generalized expansion of learning set, full system of restrictions.

1. Introduction. Classificational models for transformation of information (classifiers) are used in unformalized fields of science and technology. On the basis of such models it can be determined with the objects of what set (from the prefixed collections of sets $\Omega = (\omega_1, \omega_2, \dots, \omega_I)$) we operate in the present time.

One of the directions in the development of classificational models is connected with the theory of fuzzy sets. As L. Zadeh notes, the essential connection between such models and the fuzzy sets theory is based on the fact that most of real sets (classes) are fuzzy by their nature, i.e., transition from membership to not membership to these classes is rather gradual than discontinuous (Zadeh, 1977).

Let G denote a fuzzy classificational algorithm of information transformation (fuzzy classifier). We'll assume that it's inputs in an arbitrary discrete time $i = 0, 1, 2, \dots$ can be coded by the elements $x_i \in R^m$, and it's outputs – by the elements μ from the M_Ω set so that $\mu \in M_\Omega \Leftrightarrow \mu : \ \Omega \rightarrow [0, 1]$.

If $i = 1$, then this classifier takes a set consisting of one input-output pare

$\langle x_0, \mu_{x_0}\rangle$ together with an arbitrary input element $x_1 \in R^m$ and maps it to a fixed output element – μ_{x_1}. If $i = 2$, it takes a set consisting of two input-output pairs $\langle x_0, \mu_{x_0}\rangle$, $\langle x_1, \mu_{x_1}\rangle$ together with an arbitrary input elements $x_2 \in R^m$ and maps it to a fixed output element $\mu_{x_2} \in M_\Omega$ and so on.

Therefore in a general case an m-dimensional fuzzy classifier G is a countable infinite set of functional mappings (functions) $g\{i\}$ of the following kind

$$g\{i\}\colon R^{im} \times M_\Omega^i \times R^m \to M_\Omega, \quad i \in I \subset N^0 \tag{1}$$
$$(\text{for } i = 0 \text{ we put } g\{0\}\colon\ R^m \to M_\Omega).$$

Further on the elements of M_Ω set will be referred as responses or reactions of $g\{i\}$ functions. The last R^m entry in the argument of $g\{i\}$ will be called a question and each of the i $(R^m \times M_\Omega)$ spaces – a datum space.

We also assume that every $g\{i\}$, $i \in I$ belongs to a fixed parametrical family of functions $H\{i, \alpha\}$ where α is an unknown vector of parameters.

Let a subset $\Theta_x = \{x_1, x_2, \ldots, x_n\}$, $x_i \in R^m$, $\forall i = \overline{0.n}$ for all elements of which we know the true values of the responses for functions $g\{i\} \to \mu_{x_i}$, $\mu_{x_i} \in M_\Omega$ be available. This subset of elements together with the corresponding reactions will be called the learning set (Wolpert, 1989) and denoted by $\Theta = (x_0, y_0,\ x_1, y_1, \ldots,\ x_n, y_n)$, where $y_i = \mu_{x_i}$, $\forall i = \overline{0.n}$ and n is power of learning set.

Let Θ be a fixed learning set of n power. The fuzzy classifier G will be called an agreed with Θ if and only if the following equalities are obeyed

$$g\{n+1\}(x_0, y_0,\ x_1, y_1,\ \ldots,\ x_n, y_n;\ x_i) \equiv \mu_{x_i}, \quad \forall i = \overline{0.n}. \tag{2}$$

Relationship (2) requires from algorithm G a correct classification of all elements x which belong to the learning set Θ. At the same time, the fulfillment of such condition must not obligely mean that G will classify elements which don't belong to Θ_x correctly. In order to reach a high efficiency of G for the elements mentioned it's necessary to parametrizate the $g\{i\}$ function so that it should maximally correspond to the peculiarity of the problem solving (Barnard, 1991).

One of the effective methods of revealing of an "optimal" parametrization of the ordinary (not fuzzy) classificational models are connected with the generation of a family of strongly selfguessing (for a fixed learning set) classifiers (Wolpert, 1989).

On the informal level a classifier is said to be selfguessing for a given set Θ if it obeys the property that when taught with a subset of the learning set it recognizers (guesses) the elements of the rest part of this set. At the same time the strongly selfguessing classifier is a selfguessing classifier for which all the elements "inside" the fixed learning set are tantamount to elements "outside" this set.

In this paper we present and investigate the following formal interpretation of the last property in terms of fuzzy classificational models.

2. Basic concepts and definitions. Let $\Theta \subset R^{mn} \times M_{\Omega}^{n}$ be a learning set of n power and G – a fixed classifier agreed with Θ.

Let's assume that all considered further classifiers obey the properties of countinuity, invariancy (under permutation of the datum spaces), reproducibility, upward compatibility and are single-valued. The rigorous definitions of these properties are adduced in the Appendix 1.

Let's define an arbitrary combination $\omega_{\Theta} = (x_0, y_0,\ x_1, y_1,\ \ldots,\ x_{i-1}, y_{i-1};\ x, y)$, $i \in \{1, 2, \ldots, n\}$ belonging to Θ elements of data spaces, an arbitrary question $x \in R^m$ and a response of corresponding $g^1\{i\} \in G^1$ ($G^1 = [g^1\{i\}(\cdot)]_{i \in I}$ is a fixed particular fuzzy classifier) as a generalized training sequence (g.t.s.). The number $k = i - 1$ we'll call a range of a such a sequence.

Let's consider three types of generalized expansions of the learning set Θ.

A generalized initial expansion of Θ we specify by defining a set $W_{\Theta}^{H} = \bigcup\limits_{k} W_{\Theta,k}^{H}$, where $W_{\Theta,k}^{H}$ is a collection of all g.t.s. of k range, $k \in \{0, 1, \ldots, n\}$.

Let W_{Θ}^{H} be a generalized initial expansion of learning set Θ. Let's fix an arbitrary family of the functional mappings $\Pi = \{\tau_i\},\ i \in I$ so that

$$\tau_i\colon\ R^{im} \times M_{\Omega}^{i} \times R^{m} \times M_{\Omega} \to R^{im} \times M_{\Omega}^{i} \times R^{m} \times M_{\Omega},\ \forall i \in I,$$

and all τ_i obey the properties 1–5 adduced in the Appendix 1.

We'll call the generalized full expansion of the learning set Θ a W_{Θ}^{Π} set, including into itself those and only those g.t.s. which can be received from any $\omega_{\Theta}^{H} \in W_{\Theta}^{H}$ when applying to it a corresponding mapping $\tau_i \in \Pi$.

For an arbitrary element $\omega_{\Theta}^{\Pi} \in W_{\Theta}^{\Pi}$ we specify a set of its generalized permutational expansion $W = \{\omega_l\}_{l=0}^{i}$, $i \in \{0, 1, \ldots, n\}$ with the help of the following rule: if $\omega_{\Theta}^{\Pi} = (x_0, y_0,\ x_1, y_1,\ \ldots,\ x_i, y_i;\ x, y)$, where

$x_j \in \Theta,\ \forall j = \overline{0.i};\ x \in R^m$ then

$$\omega_0 = (x, y,\ x_1, y_1,\ \ldots,\ x_i, y_i;\ x_0, y_0),$$
$$\omega_1 = (x_0, y_0,\ x, y,\ \ldots,\ x_i, y_i;\ x_1, y_1),$$
$$\ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots$$
$$\omega_i = (x_0, y_0,\ x_1, y_1,\ \ldots,\ x, y;\ x_i, y_i).$$

Let's fix a set of all g.t.s. received as a result of building up of the generalized permutational expansion for all the elements from W^{Π}_{Θ}. We will denote this set as $\widehat{W}^{\Pi}_{\Theta}$ and call it a generalized permutation expansion of learning set Θ (under the family of $\Pi = \{\tau_i\}_{i \in I}$ mappings).

Let K_Θ be one of the generalized expansions of Θ. We'll say that G is consistent with K_Θ if and only if for all of $g\{i\} \in G,\ i = \overline{0.n}$ and for all $\omega_\Theta \in K_{\Theta,i}$ such that $\omega_\Theta = (x_0, y_0,\ x_1, y_1,\ \ldots,\ x_i, y_i;\ x, y)$ the following relation is obeyed

$$g\{i+1\}(x_0, y_0,\ x_1, y_1,\ \ldots,\ x_i, y_i;\ x) = y.$$

The m-dimensional fuzzy classifier $G = [g\{i\}(\cdot)]^n_{i \in I}$ will be referred as a strongly selfguessing (for the fixed learning set Θ) if and only if G is consistent with $\widehat{W}^{\Pi}_{\Theta}$.

3. Character of constraints, imposed on a set of fuzzy classifiers by the requirement of strong selfguessing. Let Θ be an arbitrary learning set of n power. We'll denote by $SSg(\Theta)$ a set of strongly selfguessing (for Θ) fuzzy classifiers.

In Vatlin (1993) it was shown that in solving of the most practical problems the average efficiency of such classifiers is much higher then the corresponding efficiency of those classificational models, which don't belong to the $SSg(\Theta)$ set. In connection with this circumstance the question about character of constraints, induced on the fuzzy classifiers set by the property of being selfguessing arises.

In some way the following theorem gives the answer to this question.

Theorem (A fuzzy analog of the corresponding assertion from Wolpert (1989). *The relationship* $\text{Card}(SSg(\Theta)) = 1$ *doesn't obeyed for all* Θ.

Proof. Let's preliminary formulate the following auxiliary assertion.

Assertion. Let $G = [g\{i\}(\cdot)]_{i\in I}^{n}$ be a fuzzy classifier, agreed with the learning set (of n power) Θ. Let assume that there is a continuous functional mapping (function) f obeying the following conditions:

1) any subset of the ordered pares (x_i, y_i) from Θ belong to Γ, where Γ is graph of f function;

2) for an arbitrary combination $\omega = (x_0, y_0,\ x_1, y_1,\ \dots,\ x_i, y_i;\ x, y)$ of elements from $R^{im} \times M_{\Omega}^{i}$, R^{m} and M_{Ω} correspondingly, if $\omega \subset \Gamma$ than $g\{i+1\}(x_0, y_0,\ x_1, y_1,\ \dots,\ x_i, y_i;\ x) = y,\ \forall i \in \{0, 1, \dots, n\}$.

Then G is a strongly selfguessing (for the learning set Θ) fuzzy classifier.

The correctness of this assertion directly follows from the definition of strong selfguesseness. (Under the circumstances the values of functions $g^1\{i\} \in G^1$ are immediately determined by function $f(x)$).

Let's now consider a fixed learning set Θ_1 of n power. Let also denote by f an arbitrary continuous mapping (function) which passes through the all pairs (x_i, y_i) of elements from Θ_1.

We'll further describe the procedure for designing (an operating with Θ_1) fuzzy classifier $G^* = [g\{i\}(\cdot)]_{i\in I}^{n}$ such that for G^* and $\Theta_1 f$ is a function, obeying the conditions of the assertion.

Let V be an arbitrary learning set of n power, $V = \{(x_i, v_i)\}_{i=0}^{n}$. We'll create a new learning set V' under rule: $V' = \{(x_i, v_i - f(x_i))\}_{i=0}^{n}$, where $f(x_i) = y_i, \forall i = \overline{0.n}$. Let's also denote $V_i' \overset{\text{det}}{=} V' - \{(x_{i+j}, v_{i+j} - f(x_{i+j}))\}_{j=1}^{n-i}$, $\forall i = \overline{0.n-1}$.

Let $G_M = [g\{i\}(\cdot)]_{i\in I}^{n}$ be a fuzzy MB-classifier operating on V' (the rigorous definition of MB-classifier is adduced in the Appendix 2).

Let required fuzzy classifier $G^* = [g^*\{i\}(\cdot)]_{i\in I}$ will be created such that

$$g^*\{i+1\}(V_i, x) = g\{i+1\}(V_i', x) + f(x),\ \forall i = \overline{0.n-1},\ \forall x \in R^m,$$
$$g^*\{n+1\}(V, x) = g\{n+1\}(V', x) + f(x),\ \forall x \in R^m.$$

The direct check shows that if all functions $g\{i\} \in G_M$ obeys the properties 1–5 adduced in the Appendix 1 and $f(x)$ is continuous, then all functions $g^*\{i\}$ will be obeyed the above properties too.

If the all elements of $V_i,\ i = \overline{0.n-1}$ and V "lie" on $f(x)$, then $g\{i+1\}(V_i', x)$ and $g\{n+1\}(V', x)$ are identically equal to zero and

$$g^*\{i+1\}(V_i, x) \equiv f(x), \quad \forall i = \overline{0.n-1},\ \forall x \in R^m, \tag{3}$$
$$g^*\{n+1\}(V, x) \equiv f(x), \quad \forall x \in R^m. \tag{4}$$

The relationship (3)–(4) guarantee the fulfillment of the conditions 1–2 of the assertion. In that G^* is dependent on $f(x)$ and $f(x)$ is not uniquely defined we see that the relationship $\mathrm{Card}(SSg(\Theta)) = 1$ doesn't obeyed for any Θ.

4. Conclusion. Given theorem shows that the requirement of strong self-guessing property of fuzzy classifiers doesn't form a full system of restrictions, i.e., for the unambiguous detection of most valid classifier (among the prefixed set of fuzzy classifiers agreed with arbitrary learning set) it is necessary to attract additional restrictions (additional hypothesis) (Wolpert, 1994).

Appendix 1.

1. The fuzzy classifier (f.c.) $G = [g\{i\}(\cdot)]_{i\in I}$ will be called continuous if and only if every $g\{i\}$ is continuous in all its domain.
2. The f.c. $G = [g\{i\}(\cdot)]_{i\in I}$ will be called invariant under permutation if and only if every $g\{i\} \in G$ is invariant under all possible permutation of datum spaces.
3. The f.c. $G = [g\{i\}(\cdot)]_{i\in I}$ will be referred as single-valued if and only if for every $g\{i\}$ from the fact that two fixed elements of its datum spaces have equal values of R^m-entries it always follows that the mentioned element have equal values of M_Ω-entries.
4. The f.c. $G = [g\{i\}(\cdot)]_{i\in I}^{n}$ will be referred as reproducing (for the learning set Θ, power $\Theta = n$) if and only if for every $g\{i\} \in G,\ i < n$ in the case when the value of the question coincides with the value of R^m-entry for an element from datum space belonging to Θ, the reaction of $g\{i\}$ must coincide with the value of the M_Ω-entry of mentioned element.
5. The f.c. $G = [g\{i\}(\cdot)]_{i\in I}$ will be referred as upward compatibiling if and only if for every $g\{i\}$: $R^{im} \times M_\Omega^i \times R^m \to M_\Omega$ in the case when the values of elements of two datum spaces coincide with each other, the reaction of $g\{i\}$ must coincide with the reaction $g\{i-1\}$ working on Cartezian product $R^{(i-1)m} \times M_\Omega^{(i-1)} \times R^m$ that differs from the initial one by removing of a datum space which includes equal elements.

Appendix 2.

Let Θ be an arbitrary learning set of n power, $\Theta = (x_0, y_0,\ x_1, y_1,\ \ldots,\ x_n, y_n)$, $x_i \in R^m$, $y_i \equiv \mu_{x_i}$, $\forall i = \overline{0.n}$.

Let's consider a finite set of function $g\{i\}$ such that

$$\mu_{x_i} = g\{i\}(x_0, y_0,\ x_1, y_1,\ \ldots,\ x_{i-1}, y_{i-1};\ x_i),$$

and all $g\{i\},\ i \in \{0, 1, \ldots, n\}$ obey the properties 1–5 adduced in the Appendix 1.

For every $j = 0, 1, \ldots$ we'll create a $g\{n+j\}$ function under the following rule:

$$\begin{aligned} g\{n+j\}(&x_0, y_0,\ x_1, y_1,\ \ldots,\ x_n, y_n,\ x_{n+1}, y_{n+1}, \ldots, \\ &x_{n+j-1}, y_{n+j-1};\ x_{n+j}) \\ &= \left[\sum_{i=0}^{n+j-1} (y_i/d(x_i, x))\right] \Big/ \left[\sum_{i=0}^{n+j} (1/d(x_i, x))\right]; \end{aligned} \tag{5}$$

where d – Euclidean metric.

Let also assume that if for some x_{i_1} and x_{i_2}, $x_{i_1} = x_{i_2}$ when $i_1 \neq i_2$, than the corresponding them summand in the right part of expression (5) will be allowed one and only one time.

The classifier created in this manner we'll call MB-fuzzy classifier.

The direct check shows that MB-classifier is agreed with an arbitrary learning set Θ.

REFERENCES

Barnard, E. (1991). Performance and generalization of the classification figure of merit criterion function. *IEEE Trans. on Neural Networks,* **2**(2), 322–325.

Vatlin, S.I. (1993). Selfguessing fuzzy classifiers. *Informatica,* **4**(3–4), 406–413.

Wolpert, D. (1989). *Generalization, Surface-Fitting and Network Structures.* Ph. D. Thesis, Univ. of California, CA.

Wolpert, D. (1994). *The Limitations of the PAC Framework for Addressing Real-World Generalization.* Private communication.

Zadeh, L. (1977). Fuzzy set and their application to pattern classification and cluster analysis. In J. Van Ryzin (Ed.), *Classification and Clustering.* Academic Press, New York. pp. 198–233.

Received December 1994

S. Vatlin received the Degree of candidate of Technical Science from the Minsk Radio Engineering Institute (MREI) in 1990. He is an associate professor at the Computer Science Department, MREI. His interests include mathematical theory of classification, concept learning and pattern recognition.

GRIEŽTAS SAVISPĖJAMUMAS NERYŠKIUOSE KLASIFIKATORIUOSE

Sergey VATLIN

Straipsnyje yra nagrinėjama neryškių klasifikacinių modelių problema, tokių modelių griežto savispėjamumo savybė. Pagrindinis darbo rezultatas yra tai, kad įrodyta teorema, kuri tvirtina, kad griežto savispėjamumo savybė nesudaro pilnos ribojimų aibės nedviprasmiškam geriausio klasifikatoriaus suradimui. Analizuojamos praktinės pagrįstų neryškių klasifikatorių konstravimo problemos.