INFORMATICA, 1997, Vol. 8, No. 3, 599–605

IMPROVEMENT OF THE CONSISTENT AND STRONGLY SELFGUESSING FUZZY CLASSIFIERS

Sergei VATLIN

Belarussian State University of Informatics and Radioelectronics
P. Brovka 6, 220027 Minsk, Belarus
E-mail: vatlin@micro.rei.minsk.by

Abstract. Let G_0 and G_1 be arbitrary fuzzy classifiers (Vatlin, 1993). We say that G_1 improves G_0 iff the performance of G_1 is more than G_0 one. We also introduced the concepts of consistent and strongly selfguessing fuzzy classifiers. The criterion of strong selfguessing is formulated. The theorems on the conditions of probabilistic improvement of consistent and monotonic improvement of strongly selfguessing fuzzy classifiers are proved.

Key words: machine learning algorithm, fuzzy classifier, probabilistic and monotonic improvement of fuzzy classifiers.

1. Introduction. This paper conciders the problem of fuzzy classification learning which represents a particular case of a more general problem of machine learning (Michalski *et al.*, 1983). The latter problem can be formulated as follows.

We have an input space X and an output space Y. There is an unknown function from X to Y that will be referred further as the target function.

It is given a set of n samples of the target function (the learning set). The problem is to use the learning set to guess (identify) the approximation of the target function (the hypothesis function). An algorithm that produces a hypothesis function is called a machine learning algorithm. Some examples of such algorithms are back-propagated neural nets (McClelland and Rumelhart, 1986) and Holland's classifier system (Holland, 1975).

The procedure of construction of an acceptable hypothesis function represents a sequential process rather than a single act. In order that this sequential process converge, it is necessary that the following approximation improves the preceding ones. Thus the problem of improvement of the machine learning algorithm arose (Vatlin, 1996).

The problem of fuzzy classification learning arises in a situation where Y is a set of fuzzy labels. In this case, a machine learning algorithm will be called as a fuzzy classifier (Vatlin, 1994).

One of the most interesting types of fuzzy classifiers is the strongly selfguessing fuzzy classifiers, which reproduce the target function behaviour based on information contained in a subset of the learning set θ. Some illustrative examples of such classifiers can be found (Vatlin, 1994; Vatlin, 1995).

This paper contains certain theoretical results concerning the problem of improvement of consistent and strongly selfguessing fuzzy classifiers. Some points concerning a practical employment of these results will be dealt with in the next paper.

2. The consistent and strongly selfguessing fuzzy classifiers. Let X ($\operatorname{Card} X = m$) and L ($\operatorname{Card} L < m$) be fixed sets of objects of unspecified nature. Let's denote by Y a family of all possible continuous functions translating L into a real-valued interval $[0,1]$ ($\mu \in Y \iff \mu\colon L \to [0,1]$ and μ is a continuous function). The sets X and Y will be called further as sets of initial and finite symbols in the problem of fuzzy classification.

Let f be a target function from X to Y, $f\colon X \to Y$. Let's denote by θ a collection of ordered pairs from $X \times Y$ such that $\theta = \{(x_i, y_i)\}_{i=1}^{n}$, $x_i \in X$, $y_i = f(x_i)$, $\forall i = \overline{1,n}$. The set θ will be referred further as a learning set (for a function f) and the number n – as the power of learning set θ.

Let G be an algorithm to take θ into the hypothesis function $h_{G\theta}$ from X to Y, $h_{G\theta} = G(\theta)$, $h_{G\theta}\colon X \to Y$.

We say that G is *consistent with the fixed learning set* θ (Vatlin, 1994) iff

$$h_{G\theta}(x_i) = y_i, \quad \forall i = \overline{1,n}.$$

The fuzzy classifier G, consistent with the learning set θ, will be called *strongly selfguessing for* θ (Vatlin, 1995) iff:

a) G consistent with any subset $\theta' \subseteq \theta$ such that $\operatorname{Card}\theta' = n' \geqslant n_0 - 1$ (n_0 – fixed parameter characterizes the structure of X);

b) $h_{G\theta'}(x) = h_{G\theta}(x)$, $\forall x \in X$.

Theorem 1 (recursive criterion of strong selfguessing). *The fuzzy classifier G is a strongly selfguessing one for θ iff G is a strongly selfguessing fuzzy classifier for any subset $\theta'' \subseteq \theta$ such that* $\operatorname{Card}\theta'' = n'' \geqslant n_0$.

Proof. Necessity. For any learning set $\widehat{\theta}''$ such that $\widehat{\theta}'' \subset \theta''$ and $\operatorname{Card}\widehat{\theta}'' = \widehat{n}'' \geqslant n_0 - 1$, we have $\widehat{\theta}'' \subset \theta$ and $\operatorname{Card}\widehat{\theta}'' = \widehat{n}'' \geqslant n_0 - 1$. Consequently, G is consistent with $\widehat{\theta}''$.

For any learning sets θ'' ($\operatorname{Card}\theta'' = n'' \geqslant n_0 - 1$) and $\widehat{\theta}''$ ($\operatorname{Card}\widehat{\theta}'' = n'' \geqslant n_0 - 1$) such that

$$\widehat{\theta}'' \subseteq \theta'' \subseteq \theta, \tag{$*$}$$

the following relations are obeyed

$$\begin{aligned} h_{G\hat{\theta}''}(x) &= h_{G\theta}(x), \quad \forall x \in X, \\ h_{G\theta''}(x) &= h_{G\theta}(x), \quad \forall x \in X. \end{aligned} \tag{$**$}$$

From $(*)$ and $(**)$ it follows that

$$h_{G\hat{\theta}''}(x) = h_{G\theta''}(x), \quad \forall x \in X.$$

The necessity is proved.

Sufficiency immediately follows from the theorem's condition.

Let θ be a fixed learning set. Let's denote by $SSg(\theta)$ and $Con(\theta)$ the sets of all possible strongly selfguessing (for θ) and consistent with θ fuzzy classifiers.

Theorem 2.

$$SSg(\theta) \subset Con(\theta). \tag{1}$$

Proof. Directly follows from the definitions of consistent and strongly selfguessing fuzzy classifiers.

3. Probabilistic improvement of the consistent fuzzy classifiers. Let M_G be a collection of such learning sets from $2^{X\times Y}$ that a fixed fuzzy classifier G is strongly selfguessing for its an arbitrary element. Also let θ_1, $\operatorname{Card}\theta_1 = n_1$ and $\theta_2 = \theta_1 \cup \{(x, y)\}$ be arbitrary elements from M_G.

Let's assume that any element from θ_i gets into θ_i', $\operatorname{Card}\theta_i' = n_i'$, $i = \overline{1,2}$, with equal probability. Then the next theorem holds.

Theorem 3. *The probability of event* $\{h_{G\theta_1}(x) = f(x)\}$ *is equal to 1.*

Proof. Directly follows from the previous definitions.

For an arbitrary learning set θ and fixed X and Y, let's denote by $\{G_{XY}(\theta)\}$ a set of all possible fuzzy classifiers, operating with θ (we'll omit lower indexes and write down $\{G(\theta)\}$ in the situations where it's clear what X and Y we speak about).

Let's denote by F_θ an arbitrary mapping of $\{G(\theta)\}$ into itself $(F_\theta\colon \{G(\theta)\} \to \{G(\theta)\})$, and by $\{L(G(\theta))\}$ – a collection of all possible mappings F_θ.

Also let $K = \{\overline{\theta}\}$ correspond to an arbitrary collection of learning sets from $2^{X\times Y}$ and $\overline{\theta}$, $\overline{\theta}_1 = \overline{\theta} \cup \{(x,y)\}$ – to arbitrary elements from K.

Let's denote by G_0 and G_1 fixed elements from $\{G(\overline{\theta})\}$. We say that G_1 *probabilistically improves* G_0 *on the element* $x \in X$, and put down this fact as $G_1 = \overline{F}_{\overline{\theta}}(G_0)$ iff

$$P\left(h_{G_1\overline{\theta}}(x) = f(x)\right) > P\left(h_{G_0\overline{\theta}}(x) = f(x)\right).$$

For fixed $\overline{\theta}$ and G_0, let's denote by $\{H(G_0(\overline{\theta}))\}$ a set of all possible mappings $\overline{F}_{\overline{\theta}}$.

Let G be a fixed fuzzy classifier and θ_1 ($\operatorname{Card}\theta_1 = n_1$), $\theta_2 = \theta_1 \cup \{(x,y)\}$ be the arbitrary elements from M_G.

Theorem 4. *For an arbitrary fuzzy classifier* $G_0 \in Con(\theta_1)$, *there exists* $\overline{F}_{\overline{\theta}} \in \{H\left(G_0(\theta_1)\right)\}$ *such that* $G = \overline{F}_{\theta_1}(G_0)$.

Proof. Under Theorem 4 conditions we have

$$P\left(h_{G_0\Theta_1}(x) = f(x)\right) = 1/N, \tag{2}$$

where $N = B^p$, $B = \operatorname{Card} L$, p is the number of possible scales of membership $x \in X$ to a fixed class from collection L.

By virtue of Theorem 3 we have

$$P\left(h_{G_0\theta_1}(x) = f(x)\right) = 1. \tag{3}$$

It follows from (2) and (3) that

$$G = \overline{F}_{\theta_1}(G_0),$$

for some $\overline{F}_{\theta_1} \in \{H\left(G_0(\theta_1)\right)\}$.

Thus, Theorem 4 is proved.

4. The structure of the monotonic improvement of strongly selfguessing fuzzy classifiers. Let θ correspond to an arbitrary learning set and G_0 and

G_1 to the fixed elements from $\{G(\theta)\}$. We assume that G_1 *monotonically improves* G_0 and put down this fact as $G_1 = \widetilde{F}_\theta(G_0)$ iff the following relations are satisfied:

I. $\exists x \in X$ such that $h_{G_0\theta}(x_p) \neq f(x_p)$ but $h_{G_1\theta}(x_p) = f(x_p)$;

II. $\forall x_k \in X,\ h_{G_1\theta}(x_k) \neq f(x_k) \Rightarrow h_{G_0\theta}(x_k) \neq f(x_k)$.

Under fixed θ and G_0, let's denote by $\{L(G_0(\theta))\}$ a set of all possible mappings $\widetilde{F}_\theta$.

Let M_{G_0} be a collection of learning sets from $2^{X\times Y}$ such that a fixed fuzzy classifier G_0 is strongly selfguessing for its arbitrary element. Also let θ_1 and θ_2 be arbitrary elements from M_{G_0} such that $\theta_1 \subseteq \theta_2$.

Let's consider two fuzzy classifiers G_{10} and G_{20} for which the following relationships are obeyed:

$$G_{10} = \widetilde{F}_{\theta_1}(G_0), \quad h_{G_{10}\theta_1} = G_{10}(\theta_1);$$
$$G_{20} = \widetilde{F}_{\theta_2}(G_0), \quad h_{G_{20}\theta_2} = G_{20}(\theta_2).$$

Then the following theorem takes place.

Theorem 5. *There exist* $\widetilde{F}'_{\theta_2} \in \{L(G_0(\theta_2))\}$ *and* $\widetilde{F}'_{\theta_1} \in \{L(G_0(\theta_1))\}$ *such that*

$$G'_{20} = \widetilde{F}'_{\theta_2}(G_0),$$
$$G'_{10} = \widetilde{F}'_{\theta_1}(G_0),$$

and

$$h_{G_{10}\theta_1}(x) = h_{G'_{20}\theta_2}(x), \quad \forall x \in X,$$
$$h_{G_{20}\theta_2}(x) = h_{G'_{10}\theta_1}(x), \quad \forall x \in X.$$

Proof. Directly follows from the definitions of monotonic improvement of fuzzy classifiers and strong selfguessing.

5. Conclusion. It was previously established in (Vatlin, 1993-1996) that in solving many practical and model problems the efficiency of the strongly self-guessing fuzzy classifiers is much higher then the efficiency of fuzzy classifiers which don't belong to $SSg(\theta)$.

However, due to the famous induction paradoxes (Goodman, 1955) such behaviour of the strongly selfguessing fuzzy classifiers can't be universal.

That is why it is important to elucidate conditions under which the joint use of strong selfguessing and monotonic improvement principles will bring us to the guaranteed results.

One of such conditions is given by Theorem 4 revealing restrictions under which a strongly selfguessing fuzzy classifier could be a probabilistic improvement of a consistent fuzzy classifier.

Theorem 5 gives other such conditions. It shows that a monotonic improvement of a strongly selfguessing fuzzy classifier G_0, operating with the subset θ_1, will lead us to a monotonic improvement of G_0 in the situation, where it is operating with an arbitrary learning set θ_2 such that $\theta_2 \supseteq \theta_1$, $\theta_i \in M_{G_0}$, $i = \overline{1,2}$.

Theorem 3 shows that pay for the mentioned improvements may be high since the move from the right to the left in inclusion (1) cannot be performed automatically.

Theorem 1 specifies the practical methods for construction of the (approximately) strongly selfguessing fuzzy classifiers.

Acknowledgment. The author would like to thank the anonymous reviewer who helped to improve the quality and presentation of the paper.

REFERENCES

Goodman, N. (1955). *Fact, Fiction and Forecast.* Harvard University Press, Cambridge. 394 pp.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems.* Michigan University Press, Ann Arbor. 248 pp.

McClelland, J.L., and D.E. Rumelhart (1986). *Parallel Distributed Processing.* V. 1–2, MIT Press, Cambridge. 547 pp. 611 pp.

Michalski, R.S., J.G. Carbonnell and T.M. Mitchell (1983). *Machine Learning: An Artificial Intelligence Approach.* M. Kaufmann, Palo Alto, CA. 482 pp.

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

Vatlin, S.I. (1994). *The monotonic improvement principle for fuzzy classifiers.* Belarusian state university of informatics and radioelectronics, *Report BSUIR-94-19,* Minsk.

Vatlin, S.I. (1995). Strongly selfguessing fuzzy classifiers. *Informatica,* **6**(1), 85–93.

Vatlin, S.I. (1996). The covariant monotonic improvement of fuzzy classifiers is impossible. *Neural Network World,* **6**(3), 401–406.

Received July 1997

S. Vatlin received the degree of candidate of Technical Science from the Minsk Radio Engineering Institute (MREI) in 1990. At present he is an associate professor at the Computer Engineering Department Belarussian State University of Informatics and Radioelectronics (BSUIR). His research interests include mathematical theory of classification, neural networks and pattern recognition.

SUDERINTŲJŲ IR GRIEŽTAI SAVISPĖJANČIŲJŲ NERYŠKIŲ KLASIFIKATORIŲ PAGERINIMAS

Sergej VATLIN

Straipsnyje nagrinėjami suderintieji ir griežtai savispėjantieji neryškūs klasifikatoriai bei jų tikimybinio arba monotoninio pagerinimo galimybė. Suformuluotas rekursyvus griežto savispėjamumo kriterijus, leidžiantis praktiškai konstruoti griežtai savispėjančius neryškius klasifikatorius. Apibrėžtos tikimybinio ir monotoninio pagerinamumo sąvokos, įgalinančios palyginti neryškių klasifikatorių efektyvumą. Nurodytos sąlygos, nusakančios, kada pasirinktąjį suderintąjį (savispėjantįjį) neryškų klasifikatorių galima tikimybiškai (monotoniškai) pagerinti.