Learning Algorithms and Uncertain Variables in Knowledge-Based Pattern Recognition*

Z.Bubnicki
Institute of Control and Systems Engineering, Wroclaw University of Technology
Wyb. Wyspianskiego 27, 50-370 Wroclaw, POLAND
e-mail: zdzislaw.bubnicki@pwr.wroc.pl

Abstract

The paper is concerned with a class of intelligent recognition systems described by a relational knowledge representation with unknown parameters. The main idea consists in the combination of two approaches: application of uncertain variables and a learning process with *step by step* knowledge updating. The recursive recognition algorithm with a modification of a current expert's knowledge at each step is presented. An example illustrates the presented approach.

Keywords: uncertain variables, uncertain systems, learning systems, knowledge-based systems, recognition, intelligent systems

1 Introduction

The paper is concerned with a class of intelligent recognition systems described by a relational knowledge representation with unknown parameters. For decision making in such systems two different approaches have been proposed and developed [1,2]:

1. The application of uncertain variables described by certainty distributions given by an expert.
2. The application of a learning process consisting in *step by step* knowledge validation and updating.

The purpose of this paper is to present a new idea based on the combination of these two approaches. At each step of the learning process an expert's knowledge is modified according to the current result of the learning. Such an idea has been applied to uncertain decision systems [3]. The application to pattern recognition requires to take into account a specification of this problem. The uncertain variable is described by a certainty distribution given by an expert and characterizing his/her opinion on approximate values of the variable. In the description of the uncertain variable $\bar{x}$ we introduce a certainty distribution $h(x) = v(\bar{x} \cong x)$ where $\bar{x} \cong x$ means that "$\bar{x}$ is approximately equal to *x*" and v is a certainty index given by an expert. We introduce also a soft property $\bar{x} \,\tilde{\in}\, D$ which means that "$\bar{x}$ approximately belongs to the set *D*" or "an approximate value of $\bar{x}$ belongs to *D*", with the certainty index

$$v(\bar{x} \,\tilde{\in}\, D) = \max_{x \in D} h(x). \qquad (1)$$

The details concerning the uncertain variables and their applications may be found in [4,5] and in books [1,2].

2 Recognition problem based on uncertain variables

Let an object to be recognized or classified be characterized by a vector of features $u \in U$ which may be measured, and the index of a class j to which the objects belongs; $j \in \{1,2,...,M\} \stackrel{\Delta}{=} J$. The set of the objects may be described by a *relational knowledge representation* $R(u, j; x_j)$ where $x_j \in X_j$ is a vector of unknown parameters which is assumed to be a value of an uncertain variable $\bar{x}_j$ described by a certainty distribution $h_{xj}(x_j)$ given by an expert for all $j \in J$ .. The relation R is reduced to the sequence of sets

$$D_u(j; x_j) = \{u \in U : (u, j) \in R(u, j; x_j)\},\ j \in \overline{1, M}. \qquad (2)$$

Now we can formulate the recognition problem consisting in the determination of a class j maximizing the certainty index that j belongs to the set of all possible classes for the given u

$$D_j(x_1, ..., x_M) = \{j \in J : u \in D_u(j; x_j)\}.$$

Optimal recognition problem [2]: For the given $D_u(j; x_j)$, $h_{xj}(x_j)$ $(i \in \overline{1, M})$ and u (the result of measurement) one should find j^* maximizing

$$v[j \,\tilde{\in}\, D_j(\bar{x}_1, ..., \bar{x}_M)] \stackrel{\Delta}{=} v(j).$$

Using (1) and (2) we obtain

* This work was supported by Polish State Committee for Scientific Research under Grant No. 4 T11C 001 22

$$v(j)=v[u \,\tilde{\in}\, D_u(j;\bar{x}_j)]=v[\bar{x}_j \,\tilde{\in}\, D_{xj}(j)]=\max_{x_j \in D_{xj}(j)} h_{xj}(x_j) \quad (3)$$

where $D_{xj}(j)=\{x_j \in X_j : u \in D_u(j;x_j)\}$. (4)

To find j^* it is necessary to determine $v(j)$ according to (3) and to maximize it with respect to j. Such a way may be applied under the assumption that $\bar{x}_j$ and $\bar{x}_i$ are independent uncertain variables for $i \neq j$.

3 Learning process

Assume now the parameter x_j has the value $x_j = c_j$ and c_j is unknown. Assume also that we can use the learning sequence $(u_1, j_1), (u_2, j_2), \ldots, (u_n, j_n)$ where j_i is the result of the correct classification given by a trainer. Let us denote by $\bar{u}_{ji}$ the subsequence of u_i for which $j_i = j$, by $\hat{u}_{ji}$ the subsequence for which $j_i \neq j$, and introduce the following sets in X_j:

$$\overline{D}_{xj}(n)=\{x_j \in X_j : \text{each } \bar{u}_{ji} \in D_u(j;x_j)\}, \quad (5)$$

$$\hat{D}_{xj}(n)=\{x_j \in X_j : \text{each } \hat{u}_{ji} \in U - D_u(j;x_j)\}. \quad (6)$$

The set

$$\overline{D}_{xj}(n) \cap \hat{D}_{xj}(n) \stackrel{\Delta}{=} \Delta_{xj}(n)$$

may be proposed as the estimation of c_j [2]. The determination of $\Delta_{xj}(n)$ may be presented in the form of the following *recursive algorithm* for the fixed j:

If $j_n = j$ $(u_n = \bar{u}_{jn})$

1. **Knowledge validation** for $u_n = \bar{u}_{jn}$. Prove if

$$\bigwedge_{x_j \in \overline{D}_{xj}(n-1)} [u_n \in D_u(j;x_j)].$$

If yes then $\overline{D}_{xj}(n)=\overline{D}_{xj}(n-1)$. If not then one should determine new $\overline{D}_{xj}(n)$, i.e. update the knowledge.

2. **Knowledge updating** for $u_n = \bar{u}_{jn}$

$$\overline{D}_{xj}(n)=\{x_j \in \overline{D}_{xj}(n-1) : u_n \in D_u(j;x_j)\}.$$

If $j_n \neq j$ $(u_n = \hat{u}_{jn})$

3. **Knowledge validation** for $u_n = \hat{u}_{jn}$. Prove if

$$\bigwedge_{x_j \in \hat{D}_{xj}(n-1)} [u_n \in U - D_u(j;x_j)].$$

If yes then $\hat{D}_{xj}(n)=\overline{D}_{xj}(n-1)$. If not then one should determine new $\hat{D}_{xj}(n)$, i.e. update the knowledge.

4. **Knowledge updating** for $u_n = \hat{u}_{jn}$

$$\hat{D}_{xj}(n)=\{x_j \in \hat{D}_{xj}(n-1) : u_n = U - D_u(j;x_j)\}.$$

Put $\overline{D}_{xj}(n)=\overline{D}_{xj}(n-1)$ and $\Delta_{xj}(n)=\overline{D}_{xj}(n) \cap \hat{D}_{xj}(n)$. Without an *a priori* knowledge on x_j (considered in the next section), the result of the estimation may be used in the following way: Values x_{jn} are chosen randomly from $\Delta_{xj}(n)$ and put into $D_u(j;x_j)$ in place of x_j. Then $D_u(j;x_{jn})$ are used to determine j_n^* in a way described in the previous section.

4 Learning system based on current expert's knowledge

At the *n*-th step, the result of the learning process in the form of a set $\Delta_{xj}(n)$ may be used to present an expert's knowledge in the form of a certainty distribution $h_{jn}(x_j)$ such that $h_{jn}(x_j)=0$ for every $x_j \notin \Delta_{xj}(n)$. Thus, the expert formulates his / her current knowledge, using his / her experience and the current result of the learning process based on the knowledge of the external trainer. In particular, $h_{jn}(x_j)=h(x_j, b_{jn})$, i.e. the form of the certainty distribution is fixed, but the parameter b_{jn} (in general, b_{jn} is a vector of parameters) is currently adjusted. For example, if in one-dimensional case $\Delta_{xj}(n)=[x_{j\min,n}, x_{j\max,n}]$ and $h_{jn}(x_j)$ has a parabolic form presented in Fig. 1, then $b_{jn}=(d_{jn}, a_{jn})$ and

$$d_{jn}=\frac{x_{j\min,n}+x_{j\max,n}}{2}, \quad a_{jn}=\frac{x_{j\max,n}-x_{j\min,n}}{2}.$$

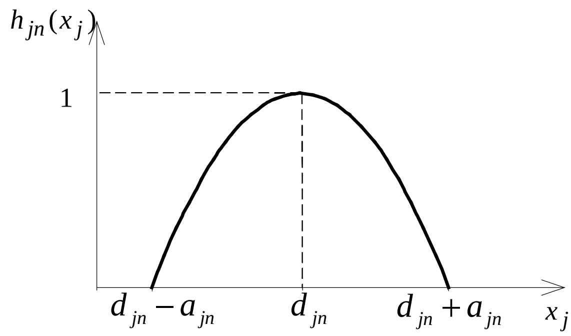

Fig. 1

For $h_{jn}(x_j)$ the next result of the recognition j_n^* may be determined in the way presented in Sec 2, i.e.

$$j_n^* = \arg\max_{j \in J} v_n(j) \qquad (7)$$

where $v_n(j)$ is determined by (3) with $h_{jn}(x_j)$ instead of $h_{xj}(x_j)$ and u_n instead of *u*. In general, as a result of the maximization (7) one may obtain a set of the classes $D_{jn} \in J$. For $h_j(x_j, b_{jn})$ we obtain the fixed form of the function $v(j, b_{jn})$:

$$v_n(j) = \max_{x_j \in D_{xj}(x_j)} h_j(x_j, b_{jn}) \stackrel{\Delta}{=} v(j, b_{jn})$$

and consequently, the fixed form of the final result, i.e. $j_n^* = j(b_{jn})$ or the set of the classes $D_{j,n+1} = D_j(b_{jn})$. The same approach may be presented in the case of C-uncertain variables, with v_c instead of v (see [3]).

The block scheme of the learning recognition system under consideration is presented in Fig. 2 where G is a generator of random variables for the random choosing of j_n^* from *D*. The blocks in the figure illustrate parts of the computer recognition system or parts of the program which may be used for computer simulations.

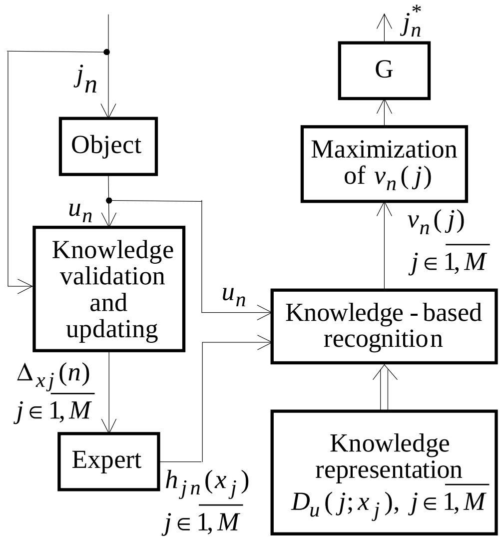

Fig. 2

5 Example

Let the sets $D_u(j;x_j)$ be described by the inequalities

$$x_j \le u^T u \le 2x_j, \quad j = 1,2,...,M.$$

Assume that the parameter x_j has the value $x_j = c_j > 0$ and c_j is unknown. It is easy to note that the sets $\overline{D}_{xj}(n)$ in (5) and $\hat{D}_{xj}(n)$ in (6) have now the forms of the following intervals:

$$\overline{D}_{xj}(n) = [\frac{1}{2}\max_i \overline{u}_{ji}^T \overline{u}_{ji},\ \min_i \overline{u}_{ji}^T \overline{u}_{ji}],$$

$$\hat{D}_{xj}(n) = (a_{jn}, \frac{1}{2} b_{jn})$$

where a_{jn} is the maximum value $\hat{u}_{ji}^T \hat{u}_{ji}$ less than $\min_i \overline{u}_i^T u_i$, and b_{jn} is the minimum value $\hat{u}_{ji}^T \hat{u}_{ji}$ greater than $\max_i \overline{u}_i^T u_i$. The estimation of c_j has the form $\Delta_{xj}(n) = \overline{D}_{xj}(n) \cap \hat{D}_{xj}(n)$ and may be determined by a recursive algorithm presented in Sec. 3. Assume that the certainty distributions $h_{jn}(x_j)$ has a parabolic form presented in Fig. 1. In this case the sets (4) for the given $u^T u$ are described by the inequality

$$\frac{1}{2} u^T u \le x_j \le u^T u \ .$$

Applying (3), one obtains $v_n(j)$ as a function of d_{jn} illustrated in Fig. 3 for $a_{jn} = 1$:

$$v_n(j) = \begin{cases} 0 & \text{for } d_{jn} \le \frac{1}{2} u^T u - a_{jn} \\ -(\frac{1}{2} u^T u - d_{jn})^2 + a_{jn} & \text{for } \frac{1}{2} u^T u - a_{jn} \le d_{jn} \le \frac{1}{2} u^T u \\ 1 & \text{for } \frac{1}{2} u^T u \le d_{jn} \le u^T u \\ -(u^T u - d_{jn})^2 + a_{jn} & \text{for } u^T u \le d_{jn} \le u^T u + a_{jn} \\ 0 & \text{for } d_{jn} \ge u^T u + a_{jn} \ . \end{cases} \qquad (8)$$

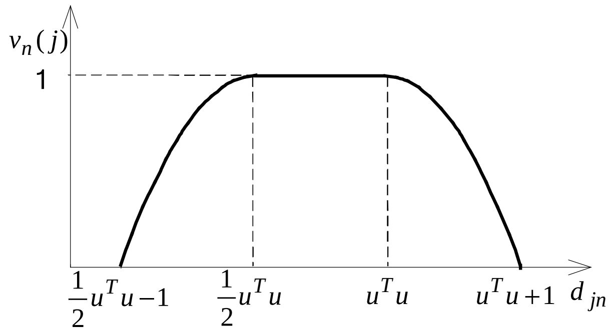

Fig 3.

For example, for $M = 3$, $u^T u = 5$, $d_{1n} = 2$, $d_{2n} = 5.2$, $d_{3n} = 6$ we obtain $v_n(1) = 0.75$, $v_n(2) = 0.96$ and $v_n(3) = 0$. Then $j_n^* = 2$, which means that for $u^T u = 5$ the certainty index that $j_n = 2$ belongs to the set of the possible classes has the maximum value equal to 0.96.

The recognition algorithm in the learning system is as follows:

1. Determine $x_{j,\min,n}$ and $x_{j,\max,n}$ using the estimation algorithm with the knowledge updating
2. Find d_{jn}, a_{jn} using the formulas in Sec.4
3. Determine $v_n(j)$ using (8) with $u = u_n$.
4. Find j_n^* maximizing $v_n(j)$.

6 Conclusions

For a class of the recognition systems under consideration, at each step of the learning process it is possible to use the current expert's knowledge based on uncertain variables. Numerical examples and simulations show that the using of the expert's knowledge during the learning process may improve the quality of the decisions. On the other hand, the algorithm is more complicated and requires using the correct classifications given by a trainer. The presented approach may be extended to other cases of uncertain recognition systems, for which the applications of uncertain variables and the learning algorithms were elaborated separately, especially for systems with three level uncertainty and for complex systems with a distributed knowledge [6,7,8,9].

References

[1] Bubnicki Z (2002), Uncertain logics, variables and systems. Springer-Verlag, Berlin, London, N. York

[2] Bubnicki Z (2004), Analysis and decision making in uncertain systems. Springer-Verlag, Berlin, London, N. York

[3] Bubnicki Z (2004), Application of uncertain variables to learning process in knowledge-based decision systems. Proceedings of the 9th International Symposium on Artificial Life and Robotics, Oita, Japan, Jan 28-30, 2004, Vol. 2, pp. 396-399

[4] Bubnicki Z (2001), Uncertain variables and their applications for a class of uncertain systems. International Journal of Systems Science 5:651-659

[5] Bubnicki Z (2001), Uncertain variables and their application to decision making. IEEE Trans. on SMC, Part A: Systems and Humans 6:587-596

[6] Bubnicki Z (2002), Application of uncertain variables to decision making in a class of distributed computer systems. Proceedings of 17th IFIP World Computer Congress, Montreal, Canada, Aug 25-30, 2002, Kluwer Academic Publishers, Norwell, MA, Vol. “Intelligent Information Processing”, pp. 261-264

[7] Bubnicki Z (2004), Application of uncertain variables and learning algorithms in a class of distributed knowledge systems. Proceedings of 18th IFIP World Computer Congress, Toulouse, France, Aug 22-27, 2004, Vol. “The Symposium on Professional Practice in AI”, pp. 111-120

[8] Bubnicki Z (2004), Uncertain variables and systems – new problems and results. Artificial Intelligence and Soft Computing – ICAISC 2004. Lecture Notes in Artificial Intelligence. Springer Verlag, Berlin, Vol. 3070, pp. 17-29

[9] Bubnicki Z (2005), Uncertain variables and their applications in knowledge-based decision systems. International Journal of Intelligent Systems (in press)