CLASSIFICATION VIA PROTOTYPE MATCHING

Kaoru Hirota
College of Engineering,Dept.of Instrument and Control Eng.
Hosei University Kajino-cho 3-7-2 Koganei-city Tokyo 184
Japan

and

Witold Pedrycz
Dept.of Automatic Control and Computer Sci.Silesian Technical University 44-100 Gliwice,Poland

Introductory remarks.Basic notation

In many situations of pattern recognition in a fuzzy environment [1], [2] , [4] or rule-based expert systems we are faced with matching fuzzy or nonfuzzy data with the antecedents of the rules.Several approaches have been already realized in the setting of fuzzy sets;cf.matching imprecise objects[2].It is particularly attractive to consider a situation of pattern classification where features of the patterns taken into account are described in a linguistic fashion by assigning them fuzzy sets-fuzzy labels.

Given a finite collection of prototypes of the classes specified by the appropriate fuzzy sets(or more precisely fuzzy relations)we are to classify any input pattern described by numerical or linguistic quantities This is made by the use of one of matching indices known in fuzzy sets e.g.coming from possibility theory.Denote by $\omega_1, \omega_2, \dots, \omega_c$ the classes discussed.The input pattern is classified according to self evident formula:denote it as belonging to the class in which the specified matching index attains maximum.Moreover the value of this index provides an information to which extent the pattern belongs to the class.Thus we get more valuable information than this is possible in Boolean case,viz.by only"yes-no" evaluation.

While discussing practical problems two aspects the matching procedures cannot be neglected and deserve a special attention:

i.consideration of different grades of importance of the features of the pattern space,

ii.consideration of uncertainty attached to the hypothesis emitted by the

classification mechanism previously introduced.
Both the problems will be discussed in foregoing sections.
At the moment let us establish some notation used further on. As usually denote by capital letters fuzzy sets and fuzzy relations. As standard $\wedge$ and $\vee$ will be used for min and max, respectively. The prototypes representing the classes will be denoted by $\underline{A}_1, \underline{A}_2, \ldots, \underline{A}_c$. They are expressed in cartesian product of feature spaces $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m$, $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_m$, thus $A: \mathcal{X} \to [0,1]$. And, every coordinate of the prototype form a fuzzy set defined in the coordinate of $\mathcal{X}$,

$$\underline{A}_i = A_{i1} \times A_{i2} \times \ldots \times A_{im} \qquad (1)$$

where cartesian product yields the following membership function,

$$\underline{A}_i(\underline{a}) = \min(A_{i1}(a_1), A_{i2}(a_2), \ldots, A_{im}(a_m)) \qquad (2)$$

In general matching any input fuzzy information $\underline{X}$ relies on matching each of its coordinate with the corresponding coordinate of the prototype. Thus denting by $M(.|.)$ the matching operator we get

$$M(\underline{X}|\underline{A}_i) = [M(X_1|A_{i1})\ M(X_2|A_{i2}) \ldots M(X_m|A_{im})]^T \qquad (3)$$

where $\underline{X}$ constitues the cartesian product,

$$\underline{X} = X_1 \times X_2 \times \ldots \times X_m \ .$$

Matching algorithms

As mentioned earlier the given pattern $\underline{X}$ is matched against $\underline{A}_i$ coordinate-wise, thus $M(\underline{X}|\underline{A}_i)$ is the vector as shown in (3). So, quite clearly, the result of matching forms a vector of real numbers lying in $[0,1]$ interval indicating to which extent the fuzzy sets X_j and A_{ij}, $j=1,2,\ldots,m$ coincide. Afterwards $M(\underline{X}|\underline{A}_i)$ is scalarized taking minimum of $M(X_j|A_{ij})$'s i.e.

$$M'(\underline{X}|\underline{A}_i) = \min_{1 \le j \le m} M(X_j|A_{ij}) \qquad (4)$$

(this is usually used way of combination). The result of the matching procedure (4) can be conveniently treated as a grade of belongingness of $\underline{X}$ to the class ω_i:

a grade of belongingness of $\underline{X}$ to ω_i $= M'(\underline{X}|\underline{A}_i)$, $i=1,2,\ldots,c$.

The higher the value of $M'(\underline{X}|\underline{A}_i)$, the better a fit of the pattern to the i-th class. Finally, a class assignment is performed according to a self-evident rule:

$$\max M'(\underline{X}|\underline{A}_i) = M'(\underline{X}|\underline{A}_{i_0}) \Rightarrow \underline{X} \in \omega_{i_0} \ ,$$

where "$\in$" is read in a context of fuzzy sets viz. with a strength of membership just given by $M'(\underline{X}|\underline{A}_{i_0})$. Fuzzy sets offer much more flexible interpretation than a simple 0-1 valuation. Beside continuous membership assignment, we can speak about judgement separability, namely distinguish the cases in which exactly one class membership dominates the remaining classes. To introduce it in a numerical fashion let us normalize all the grades of membership such they sum up to unity. Obviously it is easy to perform: replace just $M'(\underline{X}|\underline{A}_i)$ by $M'(\underline{X}|\underline{A}_i)/\sum_{k=1}^{c} M'(\underline{X}|\underline{A}_k)$. Then as an index of class separability for the specified $\underline{X}$ take the following **expression**

$$\xi = 1 - c^c \prod_{i=1}^{c} M'(\underline{X}|\underline{A}_i) / \sum_{k=1}^{c} M'(\underline{X}|\underline{A}_k) \qquad (5)$$

Notice in the worst case we have all the normalized grades of membership to the classes equal to $1/c$, therefore $\xi = 0.0$. On the other hand if exactly one dominated class membership has been noticed, $M'(\underline{X}|\underline{A}_i) = 1$, and $M'(\underline{X}|\underline{A}_j) = 0$ for $j \neq i$, $j = 1,2,\dots,c$, then ξ attains 1.0.

Until now we have not mentioned any form of the matching formula (3). To a certain extent it is a matter of taste and is application-oriented. One of those being in a common use is the possibility measure originated from the early papers on possibility theory[8],

$$M(X_j|A_{ij}) = \pi(X_j|A_{ij}) = \sup_{x \in \mathbb{X}_j} [X_j(x) \wedge A_{ij}(x)] \qquad (6)$$

Perhaps one can speculate whether it possesses significant discriminant power. For instance the possibility measure in such the form as given above yields the same value for two fuzzy sets X_j diverse in **their** character:

1. X_j - a singleton such that $X_j(x) = \delta(x, x_0)$, such the value of the possibility measure is equal to $\sup_{x \in \mathbb{X}_j} A_{ij}(x) = A_{ij}(x_0)$
2. a constant fuzzy set X_j with the value of the membership function equal to 1, thus we have $\pi(X_j|A_{ij}) = A_{ij}(x_0)$.

In both examples we get the same value of the matching grade equal to $A_{ij}(x_0)$. This would be a bit misleading: in the first example we match exactly defined input pattern against the linguistic category of the prototype of the class; in the second we deal with the pattern which the j-th coordinate is viewed as completely unknown. Notice the fuzzy set with the constant membership function equal to 1.0. is sought as modelling the term "unknown" conveying no relevant information. Nevertheless the proper definition of the matching expression is of a technical nature and a slight modification would help to resolve the abovementioned difficulty.

Finally performing scalarization we get,

$$M'(\underline{X} \mid \underline{A}_i) = \bigwedge_{j=1}^{m} \pi(X_j \mid A_{ij}) \tag{7}$$

and, in general, the result of matching can emit the following hypothesis:
-the pattern $\underline{X}$ is ω_i to the degree equal to $M'(\underline{X}|\underline{A}_i)$

A logical background supporting the use of such the aggregation procedure comes from a pessimistic view:matching is good to such a degree that is equal to the lowest degree of matching among the coordinates of $\underline{X}$ and $\underline{A}_i$. Unfortunatelly all the coordinates are evaluated uniformly.This would form a significant shorcoming in a wide group of application.For a certain analogy recall a classical pattern recognition problem in which we deal with the matching prototypes and input patterns in m-dimensional feature space($= \mathbb{R}^m$).The essence of matching relies on computation respective distance between the points in this space.Only in a simplest situation one can treat the Euclidean distance;usually the Mahalanobis distance is accepted that is especially useful to handle different grades of importance of the features.

Thus we will try to improve the matching formula to make it more applicable to be powerful in situations various coordinates of $\underline{X}$ have different values of importance (different discrimination power).Moreover the proposal should be realistic;namely it should give a constructive way to determine the hierarchy of the features (coordinates of $\underline{X}$).Therefore it is sometimes doubtful to accept a simplified version of the problem solution such that

$$M'(\underline{X}|\underline{A}_i) = \bigwedge_{j=1}^{m} (\pi(X_j|A_{ij}))^{w_j} \tag{8}$$

w_j being properly chosen weight factors,or forming a linear combination

$$M'(\underline{X}|\underline{A}_i) = \sum_{j=1}^{m} \pi(X_j \mid A_{ij})\, w_j \Big/ \sum_{j=1}^{m} w_j \tag{9}$$

Notice that we have to possess more refinement knowledge on the environment in which the classification is performed that allows us to find out a reasonable basis to accept the special form of the aggregation function (8) or (9).

The weights of the features can be achieved by applying Saaty's priority theory[6].In this way the features are collected in pairs and shown to the expert in this field in which the classifier will be used in future.The essence of the experiment deals with pairwise feature comparison,the expert is requested to estimate the ration w_i/w_j denoted by r_{ij} in a specified scale,usually of 7-10 point range.If the i-th feature is completely prefered over the j-th one,then r_{ij} is equal to the highest score viz. r_{ij}= 7 or 10 .If both the features are sought as equal then r_{ij}=is put to 1.0 .Moreover a reciprocal relationship holds,namely

$r_{ij}=1/r_{ji}$. Possessing the results of pairwise comparison, the weight vector can be obtained by solving an appropriate optimization problem: for details consult e.g. [6], see also an approach where the scores r_{ij} are fuzzy in their nature (triangular fuzzy numbers) [5]. Having at our disposal the vector $\underline{w} = [w_1\ w_2\ \dots\ w_m]^T$ we can try to perform mechanisms of weighting the results of matching $M'(X_j|A_{ij})$. It is plausible to accept here a fuzzy measure approach ($g_\lambda(\cdot)$). Bear in mind a monotonicity property $g(C) \leq g(D)$, for $C \subset D$ being two subsets of the features $\mathbb{F}=\{f_1, f_2 \dots f_m\}$ can be interpreted as a measure of "certainty" of object matching (the more features we have already matched, the higher the grade of certainty on the measure of fit is reached). According to the theory of fuzzy measures and integrals worked out by Sugeno [7] we obtain

$$M'(\underline{X}|\underline{A}_i) = \int_{\mathbb{F}} M'(X_j|A_{ij}) \circ g_\lambda(\cdot) \qquad (10)$$

with $g_\lambda(\cdot)$ being a λ-fuzzy measure. For practical purposes $g_\lambda(\cdot)$ is calculated on the basis of the vector $\underline{w}$ already experimentally collected.

Measuring certainty of the hypothesis of class assignment

Having a global view at the investigations performed until now as well as the remarks contained in the previous section, the hypothesis generated cifies a degree of membership of the pattern to the given class. Noticing what has been told about certainty of such the hypothesis we can formulate the following question: Is there any un certainty tied with the hypothesis? Consider two examples:

a. input information dealt with the pattern is nonfuzzy one, thus the matching yields the grade of membership to the j_0-th class equal to $A_{j_0}(x_0)$,

b. the input pattern is specified by the fuzzy set such that $X(x) = A_{j_0}(x_0)$ for any $x \in [\alpha, \beta]$ and 0, otherwise such that $x_0 \in [\alpha, \beta]$. Here we get the same value of matching as before applying the possibility measure.

However the two situations are extremaly different. In the first we play with precise input information, in the second observe the information is quite fuzzy. Thus it is wothwhile to have the result of matching enriched by a certainty value or uncertainty value attached to the grade of membership accordingly,

certainty{ $\underline{X}$ belongs to ω_i with the grade of membership equal to

$$M'(\underline{X}|\underline{A}_i)\} = \lambda,$$

or

uncertainty $\underline{X}$ belongs to ω_i with the grade of membership equal to μ

In situation a. λ should be put to 1, while for b. should be significantly

lower. Thus the question arises how the value of may be conveniently obtained.
More deeper insight into the problem addressed moves us to more refinement local look at the matching stage. Previously we have concentrated our attention on a scalar index of matching. To get a quantitative expression for uncertainty tied with the hypothesis generated let us perform pointwise matching. More formally in sequel assume the fuzzy relation $\underline{X}$ has been defined in a finite number of elements of $\mathcal{X}$ we get,

$$\underline{A}(x_1,x_2,\ldots,x_n) \equiv A_i(x_1,x_2,\ldots,x_n) = [\,\underline{X}(x_1,x_2,\ldots,x_n) \rightarrow \underline{A}_i(x_1,x_2,\ldots,x_n)] \wedge [\,\underline{A}_i(x_1,x_2,\ldots,x_n) \rightarrow \underline{X}(x_1,x_2,\ldots,x_n)] \tag{11}$$

$x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, \ldots, x_n \in \mathcal{X}_n$. Calculating the values of probabilities of occurence of different grades of equality,

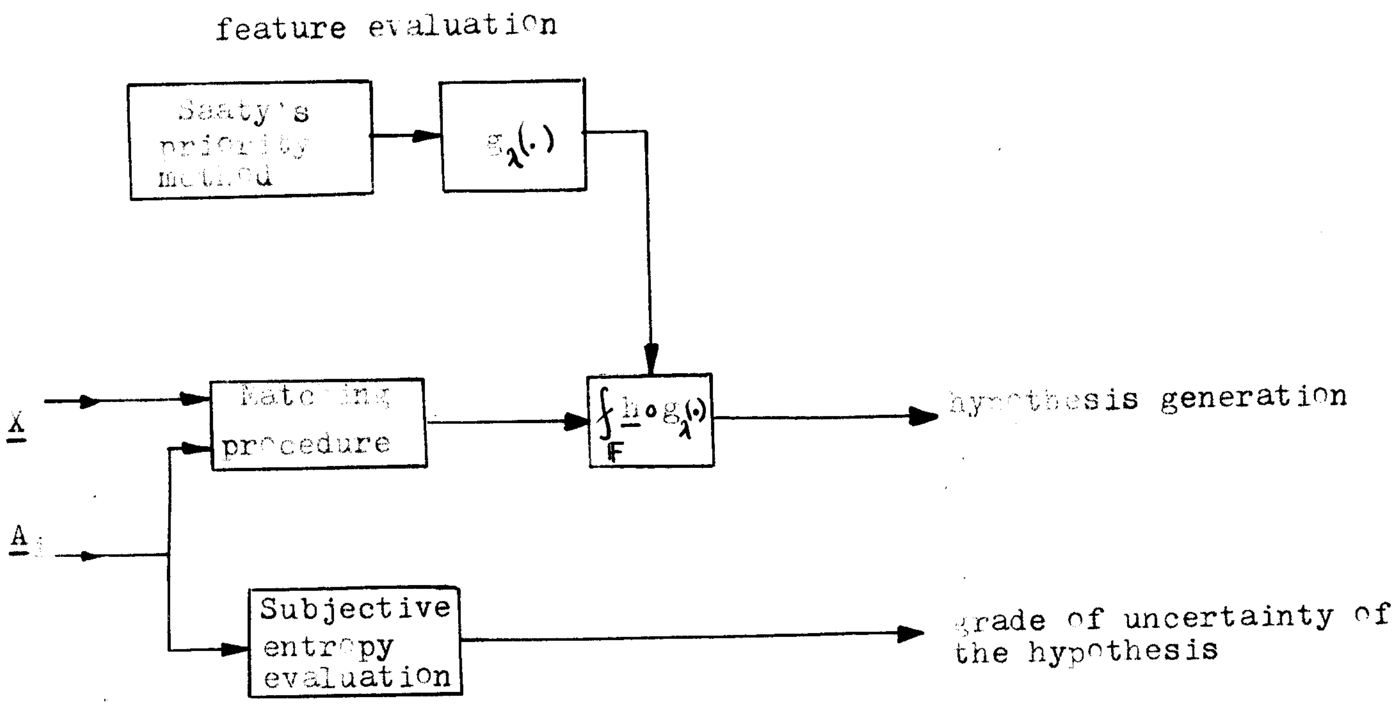
$$p_k = \frac{\text{card}\{x_1,x_2,\ldots,x_n \mid \underline{X}(x_1,x_2,\ldots,x_n) \equiv \underline{A}_i(x_1,x_2,\ldots,x_n) = a_k\}}{\text{card}(\mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_n)}$$

$a_k \in [0,1]$, $k=1,2,\ldots,K$,
one arrives a to a notion of subjective entropy defined as in[3],

$$H_i = -\sum_{k=1}^{K}(a_k p_k \log_2 a_k p_k + (1-a_k)p_k \log_2(1-a_k)p_k) \tag{12}$$

Having the problem of maximization of the subjective entropy already solved, see[3], we normalize H_i obtaining the required grade of uncertainty $u = H_i/H_{max}$, or grade of certainty $\lambda = 1-u$.
The introduced proposal is attractive from at least two points of view,
- firstly one can get a reasonable view on uncertainty of the hypothesis,
- secondly one may control an emission of such the hypothesis that has an acceptable level of uncertainty with still significant grade of membership. Therefore the statements about the hypothesis are significantly modified. Instead determining the membership to the given class we may specify a grade of membership to a given logical combination of the classes in general taking the form $\psi(\omega_1, \omega_2, \ldots, \omega_c)$. For instance $\psi(\omega_1,\omega_2,\ldots,\omega_c) = \omega_1$ or ω_2 that is a weaker form of the class assignment $\psi(\omega_1,\omega_2,\ldots,\omega_c) = \omega_1$, or $\psi(\omega_1,\omega_2,\ldots,\omega_c) = \omega_1$ and(not ω_2 and not ω_3 and not .... and not ω_c) that is a stronger judgement in comparison to the original one expressing only the grade of membership to the class ω_1.

A complete system for pattern classification with the blocks disscussed until now is displayed in the figure presented below.



Conclusions

We have discussed a scheme of pattern classification that makes use of procedures of pattern matching. It is presented how the matching can be realized and in which way the weights of the features are obtained and aggregated. Moreover the grades of uncertainty tied with the hypothesis generated are presented; it is also underlined the way in which the hypothesis dealing with several classes are evaluated.

References

1. J.C.Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, N.Y. 1981.
2. M.Cayrol, H.Farreny, H.Prade, Fuzzy pattern matching, Kybernetes, 11, 1982, 103-116.
3. K.Hirota, Ambiguity based on the concept of subjective entropy. In: Fuzzy Information and Decision Processes M.M.Gupta, E.Sanchez eds. North Holland, Amsterdam, 1982, 29-40.
4. K.Hirota, W.Pedrycz, Characterization of fuzzy clustering algorithms in terms of entropy of probabilistic sets, Pattern Recognition Letters, 2, 1984, 213-216.
5. P.J.M.van Laarhoven, W.Pedrycz, A fuzzy extension of Saaty's priority theory, Fuzzy Sets and Systems, 11, 1983, 229-242.

6.Th.L.Saaty,The Analytic Hierarchy Process,Planning,Priority Setting, Resource Allocation,McGraw Hill N.Y.1980.
7.M.Sugeno,Theory of fuzzy integral and its applications.Ph.D.thesis, Tokyo Institute of Technology,1974.
8.L.A.Zadeh,Fuzzy sets as a basis for a theory of possibility,Fuzzy Sets and Systems,1,1978,3-28.