Intuitionistic Fuzzy Sets and Incremental Concept Formation in Artificial Intelligence

Blaga Nikolova Iordanova

Department of Information and Computer Science,
Technical University ,9010 Varna, Bulgaria
Phone (359-52) 444-982 ,Fax (359-52) 871 910

ABSTRACT

This article describes an improvement of BIOR, an Incremental Model of Concept Formation in Artificial Intelligence, published in the articles [11]. With a given Data-Base or a Learning Set of Observations as a sequential instances presented by their associated descriptions, the task of BIOR is to acquire incrementally without any advice from a tutor a Knowledge Base organized in a Hierarchy of Concepts. One of the problems in acquisition of Hierarchy of Concepts is that some initial sequence of instances could direct wrongly the Learning Process resulting in non-optimal Knowledge structure or with poor predictive ability. Two bi-directional operators have been introduced to reverse the effect of previous learning when new instances should suggest the need for combining concepts if the distinction between them fares poorly or for splitting the node into its children nodes if the parent node has less predictive ability than its children. In spite of the existence of these 'merge' and 'split' operators and their help for recovering the Hierarchy of Concepts the problem of keeping the right direction of the Learning could reach a reasonable solution to some extent. In fact two main problems exist in the Learning and Classification Process. One of the problems is to what extent various two real-attribute values of two instances could build a primary conceptual node through the 'GENERALIZE' Process. The second problem is to what extent a real-attribute value of a new instance with a build-up conceptual node could differ from the mean value of the relevant conceptual attribute so that the attribute value should match enough the concept description. The article proposes an approach for solving the problems in an automatic way with building Intuitinistic Fuzzy Sets [12], [13] over the Sets of real-attribute values in each conceptual description during the process of Learning.

1. INTRODUCTION

With a given learning set of instances the task of Incremental Concept Formation in Artificial Intelligence is to discover regularities between the parameter values of the learning instances and to devise the instances into categories organized in a Hierarchy of Concepts without any class information.

The first part of the article deals with some details necessary for the description of a new approach for keeping the right direction of the Learning through Fuzzy Logic Control.

1.1.REPRESENTATION OF THE LEARNING SET OF INSTANCES

Similar to EPAM [2] BIOR represents instances as a conjunction of attribute-value pairs, along with an optional ordered list of component objects. Each component object is in turn described as a conjunction of attribute-value pairs with its own optional components, and so forth.

More formally the set of instances of objects or events could be presented as follows:

$$\Sigma = \{\sigma_{j_0} / j_0 \in Z^+\},$$

where:

σ_{j_0} is an instance, Z^+ is a set of the positive integers.

An instance could be represented by recursive formula:

$$\sigma_{j_k} = S_{j_k}, \{ \sigma_{j(k+1)} / j_{(k+1)} \in Z^+\},$$

where:

$k \in Z^+$ is a level of recursion,

$S_{j_k} = \{< a_i , v_{i,j} > / a_i \in A_{j_k}, v_{i,j} \in V_{a_i,j_k}\}$,

where:

$A_{j_k} = \{ a_i / i \in Z^+ \}$ is a set of attribute names,

$V_{a_i,j_k} = \{ v_{a_i,j} / j \in Z^+ \}$ is a set of a_i - attribute values.

An instance σ_{j_k} is described by a set S_{j_k} of <'attribute' = 'value'> pairs and a set of optional component objects $\sigma_{j(k+1)}$ described by the same formula in a higher level of recursion. When k = 0 the formula describes the primary object representing an instance in the learning set.

Each component object is described by its own set $S_{j(k+1)}$ of <'attribute'='value'> pairs and a set of optional component objects $\sigma_{j(k+2)}$.

BIOR deals with nominal (symbolic), numeric and real-valued attributes as well. In addition nominal attributes could take more than one value.

1.2. REPRESENTATION OF THE ACQUIRED KNOWLEDGE.

BIOR organizes knowledge into a concept hierarchy like Lebowitz's [6,8] UNIMEM. The higher node in the hierarchy represents a more general concept and its children represent more specific ones and so forth. The hierarchy consists of concept nodes and links. Both terminal and non-terminal nodes have a concept description, a set of instances and a set of child-nodes associated with them. The system stores only unaccounted features of the instances in the sets of instances. Each concept description consists of a conjunction of attribute-value pairs and each value has an associated predictability score. The system allows each of the node links to specify the results of multiple features (attribute-value pairs). Each feature on the link has an associated predictiveness score. Multiple indices to a child-node provide constructing non-disjoint hierarchy. And here the similarity ends.

One important difference is that some link tests involve examining the value of an attribute , whereas others involve examining the category of the sub-object, which could itself be learned. Finally, the concept hierarchy includes sub-hierarchies of concepts that summarize and organize the knowledge about the component objects of the instances and each sub-hierarchy could include sub-hierarchies and so on until no more component object category of component objects have to be learned.

Formally a description of the concept node G_k could be presented as follows:

$$G_k = (C_k , D_k, L_k, M_k, S_k, \{G_{k,j} / 0 \leq j \leq m\}),$$

where:

$m \in Z^+$

$0 \leq k \leq n, n \in Z^+$

$C_k = \{ < a_i, v_{i,j} > / a_i \in A , v_{i,j} \in V_{a_i} , i \in Z^+, j \in Z^+\},$

$L_k = \{ < b_i ,v_{i,j} > / b_i \in A , v_{i,j} \in V_{b_i} , i \in Z^+, j \in Z^+\},$

$\{G_{k,j} / 0 \leq j \leq m, m \in Z^+\}$ is a set of Child-Nodes of the parent node G_k.

C_k is a set of < 'attribute' = 'value' > pairs describing the concept associated with the concept node G_k.

L_k is a set of < 'attribute' = 'value' > pairs serving as link indices from the parent node to the concept node G_k.

S_k is a set of instances associated with the concept node G_k.

There exists simple correspondence between sets C_k and D_k - a set of predictability scores associated with each one of the < 'attribute' = 'value' > pairs in the concept description C_k .

$$D_k=\{ P(< a_i,v_{i,j} >, C_k) / < a_i,v_{i,j} > \in C_k, i \in Z^+, j \in Z^+\},$$

The simple correspondence between sets C_k and D_k could be expressed as C_k--f--> D_k.

A set of predictiveness scores M_k associated with the set of the link indices L_k to the node G_k is as follows:

$$M_k =\{ P(C_k, < a_i ,v_{i,j} >) / < a_i ,V_{i,j} > \in L_k, i \in Z^+, j \in Z^+\},$$

Between sets L_k and M_k there exists a simple correspondence L_k--f--> M_k as well.

The UNIMEM' measures of predictiveness and predictability are informal and have no clear semantics. To improve this shortcoming in BIOR these scores have a formal grounding in probability theory. The conditional probability of the value with a given membership in the class $P(<a_i,v_{i,j}>,C_k)$ is its predictability and the conditional probability of membership in the class for the given value is its predictiveness, $P(C_k,<a_i,v_{i,j}>)$.

1.3. REPRESENTATION OF REAL-VALUED ATTRIBUTES

The introduction of real-value attributes requires an extension of representation of concepts in BIOR different from UNIMEM's. BIOR uses CLASSIT' [10] notion of storing a continuous normal distribution (bell-shaped curve) for each real-value attribute occurring in a concept. The continuous normal distribution is expressed in terms of a mean (average) value μ and standard deviation δ.

1.4. EVALUATION FUNCTION

BIOR uses mixed Evaluation Function called Category Utility of COBWEB [3-5] and CLASSIT [10] to control its classification and learning behavior. The Evaluation Function includes discrete conditional probabilities for symbolic attributes and variances for real valued attributes. The Category Utility is used to consider the choice between the conceptual nodes and sub-hierarchy of concepts for descending through the hierarchy in attempting to classify a new instance. The Evaluation Function is used as well to consider the choice between the learning processes: 'Creating a new Concept' using the 'GENERALIZE' operator, 'Finding a suitable Concept that classifies the new Instance' implementing 'MERGE' or 'SPLIT' or 'DELETE' operator and the process 'Storing the new Instance ' for creating a new disjunctive conceptual node in the future.

A detailed example explaining learning and classification processes as well as evaluating and pruning as a part of learning is given in the article [11].

The second part of the article deals with a method for subsidiary control of the process 'Creating a new Concept' and when refining an existing conceptual node in response to learning a new information given by a new instance during the process 'Classifying a new Instance' .

The method described below will prevent the Learning from going astray to some extent. This method will reduce the needs of merging and splitting the nodes and in this way will reduce a needed time for Learning, and will give the possibility of ranging by priority the concepts classifying an instance as well.

2. IMPLEMENTATION OF THE THEORY OF INTUITIONISTIC FUZZY SETS IN INCREMENTAL CONCEPT FORMATION

The learning is going astray mainly in two significant manner. The combination of initial sequence of instances could be irrelevant and their attribute values could provoke creating primary very general conceptual nodes whose attribute values have less predictive abilities than its children nodes could have or could provoke creating very specific conceptual nodes whose attribute values distinct poorly. In both cases the next instances will certainly suggest the need of recovering the hierarchy of concepts. In the first case the 'SPLIT 'operator is used to split the general node into its more specific ones. In the second one the recovering could be made by the 'MERGE' operator which will combine the conceptual nodes that distinct poorly. In both cases some time is lost.

Let us introduce two system parameters. The first one will not allow building a concept descriptions with real-valued attributes which vary greatly. This parameter Δ will define a 'maximum possible deviation' depending on the specific attribute area of the description. It could be expressed as a percent of a mean value μ. This parameter will decrease the possibility for generating a concept nodes from instances parameter values of which differ a lot and will decrease certainly operations of splitting nodes into its children nodes in the future. There exist of course a danger of creating a conceptual nodes which will need to be combined in turn with the similar ones through the 'MERGE' operator if this system parameter Δ is given to be extremely little.

We will consider an example. For the simplicity the concepts will be described by one attribute only . An attribute 'high' is given with values in the range [50.5, 220.5]. If the first three instances have values respectively 60.6 , 180.1 and 110.35 . From the first two instances a conceptual node G_1 with associated concept description C_1 will be generated .

$$C_1=\{<\mu_{hight,1}=120.35\ ,\ \delta_{hight,1}=59{,}75>\}.$$

The fourth instance with an attribute value 200.4 will provoke splitting the node G_1 into its two children nodes G_{11} and G_{12} . The concept descriptions C_{11} and C_{12} correspond to the nodes G_{11} and G_{12} respectively .

$$C_{11}=\{<\mu_{hight,11}=190.25,\ \delta_{hight,11}=10.15>\},$$
$$C_{12}=\{<\mu_{hight,11}=85.475,\ \delta_{hight,11}=24.875>\}.$$

A definition of a system parameter $\Delta_{high} = 40.5$ for the given attribute will prevent Learning from going astray. The primary conceptual node G_1 will not be generalized and therefore no need would exist to split G_1 into G_{11} and G_{12}.

The second system parameter τ defines to what extent an attribute value of an instance could differ from the mean value of the relevant conceptual attribute so that it could be considered matching the concept. This parameter τ gives the positive and negative boundaries or thresholds of the interval $[\mu-\tau\delta, \mu+\tau\delta]$ around the mean value. If the attribute value of an instance $v_{i,j}$ is within this interval then this attribute value belongs to the conceptual ones with a degree of affiliation. The parameter τ called 'Threshold of certainty' is expressed as a percent of the deviation δ.

This parameter defines the minimum boundary intervals of the conceptual attribute values $[\mu -\delta, \mu - \tau\delta]$ and $[\mu+\tau\delta, \mu+\delta]$ as well. If an attribute value of an instance is in this boundary intervals it means that this attribute value matches with a degree of undeterminateness to the conceptual ones.

If an attribute value $v_{i,j}$ of an instance is greater than $(\mu + \delta)$ or less than $(\mu - \delta)$, then it doesn't belong to the conceptual ones with a degree of non-affiliation.

For each conceptual node and for each conceptual real-valued attribute are created Intuitionistic Fuzzy Sets (IFS) from the attribute values of the instances matching the concept description.

Let $\beta_{a_i,k}$ (1) be an IFS of the values $v_{i,j} \in Va_i$ of the attribute $a_i \in A_k$, where $A_k \subseteq A$ is a set of conceptual attributes of C_k and Va_i is the set of the attribute a_i values. It is obvious that $\beta_{a_i,k}$ is a sub-set of the set Va_i (2).

(1) $\beta_{a_i,k} = \{< v_{i,j}, \Theta\beta_{a_i,k}(v_{i,j}), \Gamma\beta_{a_i,k}(v_{i,j}) > / \ v_{i,j} \in Va_i,\ j \in Z^+\}$,

where:

$i \in Z^+$,

$\Theta\beta_{a_i,k}(v_{i,j}) : Va_i \text{ -> } [0,1]$,

$\Gamma\beta_{a_i,k}(v_{i,j}) : Va_i \text{ -> } [0,1]$.

$$(2) \quad \beta_{a_i,k} \subseteq Va_i$$

The functions $\Theta\beta_{a_i,k}(v_{i,j})$ and $\Gamma\beta_{a_i,k}(v_{i,j})$ determine the degree of the affiliation and the degree of the non-affiliation respectively of the value $v_{i,j} \in Va_i$ to the set $\beta_{a_i,k}$. For each value $v_{i,j} \in \beta_{a_i,k}$ the inequality (3) remains valid

$$(3) \quad 1 \leq \Theta\beta_{a_i,k}(v_{i,j}) + \Gamma\beta_{a_i,k}(v_{i,j}) \leq 1$$

The function $\Pi\beta_{a_i,k}(v_{i,j})$ (4) defines the degree of undeterminateness of the $v_{i,j}$ affiliation to the set $\beta_{a_i,k}$.

$$(4) \quad \Pi\beta_{a_i,k}(v_{i,j}) = 1 - \Theta\beta_{a_i,k}(v_{i,j}) - \Gamma\beta_{a_i,k}(v_{i,j})$$

For the deduction of the function $\Theta\beta_{a_i,k}(v_{i,j})$ and $\Gamma\beta_{a_i,k}(v_{i,j})$ will be used the approach of Bayes. For each real attribute value $v_{i,j} \in Va_i$ two hypothesis H_1 and H_2 exist.

H_1= ' The value $v_{i,j}$ meets the requirements for the concept description C_k '
H_2= ' The value $v_{i,j}$ doesn't meet the requirements for the concept description C_k '

There exists the evidence E_1 in favour of the hypothesis H_1 and correspondingly E_2 in favour of H_2.

E_1= ' The value $v_{i,j}$ of the attribute a_i is in the interval of certain affiliation (5) or in the case a new conceptual node will be generated, the standard deviation is less or equal to the maximum possible deviation Δ (6)'

$$(5) \quad v_{i,j} \in [\ \mu - \tau\delta,\ \mu + \tau\delta\ ]$$

$$(6)\quad \delta \le \Delta\mu$$

where:

Δ - is the system parameter 'maximum possible deviation'

τ - is the system parameter ' threshold of certainty'.

E_2 = ' The value $v_{i,j}$ of the attribute a_i is in the interval of non-affiliation (7)

$$(7)\quad v_{i,j} \le (\mu - \delta) \quad \text{or} \quad v_{i,j} \ge (\mu + \delta)$$

The degree of the affiliation of the value $v_{i,j}$ to the set $\beta_{aj,k}$ could be fixed using the formula of Bayes (8).

$$(8)\quad \Theta\beta_{a_i,k}(\, v_{i,j}\,) = P(H_1{:}E_1) = (P(E_1{:}H_1)\; P(H_1))\;/\;P(E_1)$$

where:

$P(H_1)$ is a priori probability for the hypothesis H_1.

$P(H_1{:}E_1)$ is a posterior probability for the H_1 with the evidence of E_1.

$P(E_1{:}H_1)$ is the probability of the evidence E_1 given the truth of the hypothesis H_1.

After calculation of the $P(H_1{:}E_1)$ the primary $P(H_1)$ could be forgotten and the probability $P(H_1{:}E_1)$ be used as a renewed value of the probability $P(H_1)$. $P(H_1)$ being precised in the course of the work it follows that $P(E_1)$ could be made more precise as well (9).

$$(9)\qquad P(E_1) = \sum_{s=1}^{n} P(E_1{:}H_s)\;/\;P(H_s)$$

The probability of the evidence E_1 is the probability of each one of the situations, which could bring to the succession of the evidence E_1 multiplied by the probability for the same situation. The situations in this case are the conceptual descriptions which requirements are met by the value $v_{i,j}$ of the attribute a_i.

Therefore

$P(H_k)$ is equivalent to $P(C_k)$ and

$P(E_1{:}H_k)$ is equivalent to

$P(v_{i,j} \in [\mu_{a_i,k} - \tau_{a_i}\delta_{a_i,k}\,,\, \mu_{a_i,k} + \tau_{a_i}\delta_{a_i,k}] \mid C_k)$, where $\delta_{a_i,k} \le \Delta_{a_i}\mu_{a_i,k}$

Consequently:

$$\Theta\beta_{a_i,k}(v_{i,j}) = P(C_k \mid v_{i,j} \in [\mu_{a_i,k} - \tau_{a_i}\delta_{a_i,k}\,,\, \mu_{a_i,k} + \tau_{a_i}\delta_{a_i,k}]\,) =$$

$$= \frac{P(v_{i,j} \in [\,\mu_{a_i,k} - \tau_{a_i}\delta_{a_i,k}\,,\, \mu_{a_i,k} + \tau_{a_i}\delta_{a_i,k}\,] \mid C_k)\; P(C_k)}{\sum_{s=1}^{n} P(v_{i,j} \in [\mu_{a_i,s} - \tau_{a_i}\delta_{a_i,s}\,,\, \mu_{a_i,s} + \tau_{a_i}\delta_{a_i,s}] \mid C_s)\; P(C_s)}$$

where:

$\delta_{a_i,k} \leq \Delta_{a_i} \mu_{a_i,k}$

$v_{i,j} \in V_{a_i}$ is j-value of the attribute $a_i \in A$

$\mu_{a_i,k}$ is mean or average value of the conceptual attribute a_i of the concept C_k.

Δ_{a_i} and τ_{a_i} are the system parameters for the a_i attribute

$\delta_{a_i,k}$ is a standard deviation of the attribute a_i values on the concept description C_k.

In the similar way one could deduce the formula for calculation of the degree of the non-affiliation of the value $v_{i,j}$ to the IFS $\beta_{a_i,k} \subseteq V_{a_i}$.

$$\Gamma_{\beta_{a_i,k}}(v_{i,j}) = P(H_2{:}E_2) = (P(E_2{:}H_2)\, P(H_2)) / P(E_2)$$

where:

$P(H_2)$ is equivalent to $P(\overline{C_k}) = 1 - P(C_k)$.

$P(H_2{:}E_2)$ is equivalent to $P(\overline{C_k} \mid v_{i,j} \notin [\mu_{a_i,k} - \tau_{a_i}\delta_{a_i,k}, \mu_{a_i,k} + \tau_{a_i}\delta_{a_i,k}])$

$P(E_2{:}H_2)$ is equivalent to $P(v_{i,j} \notin [\mu_{a_i,k} - \tau_{a_i}\delta_{a_i,k}, \mu_{a_i,k} + \tau_{a_i}\delta_{a_i,k}] \mid \overline{C_k})$

$$P(E_2) = \sum_{s=1}^{n} P(v_{i,j} \notin [\mu_{a_i,s} - \tau_{a_i}\delta_{a_i,s}, \mu_{a_i,s} + \tau_{a_i}\delta_{a_i,s}] \mid C_s)\, P(C_s)$$

The probability of the evidence E_2 is calculated by the summation of the probabilities for the value $v_{i,j}$ to be not in the interval $[\mu_{a_i,s}-\tau_{a_i}\delta_{a_i,s}, \mu_{a_i,s}+\tau_{a_i}\delta_{a_i,s}]$ with the affiliation of the instance to the concept C_s, multiplied by the probability of the concept C_s =/= C_k.

Therefore:

$$\Gamma_{\beta_{a_i,k}}(v_{i,j}) = P(\overline{C_k} \mid v_{i,j} \notin [\mu_{a_i,k} - \tau_{a_i}\delta_{a_i,k}, \mu_{a_i,k} + \tau_{a_i}\delta_{a_i,k}]) = \frac{P(v_{i,j} \notin [\mu_{a_i,k} - \tau_{a_i}\delta_{a_i,k}, \mu_{a_i,k} + \tau_{a_i}\delta_{a_i,k}] \mid \overline{C_k})(1 - P(C_k))}{\sum_{s=1}^{n} P(v_{i,j} \notin [\mu_{a_i,s} - \tau_{a_i}\delta_{a_i,s}, \mu_{a_i,s} + \tau_{a_i}\delta_{a_i,s}] \mid C_s)\, P(C_s)}$$

where: C_s=/= C_k

Obviously the function fixing the degree of undeterminateness will be:

$$\Pi_{\beta_{a_i,k}}(v_{i,j}) = 1 - \Theta_{\beta_{a_i,k}}(v_{i,j}) - \Gamma_{\beta_{a_i,k}}(v_{i,j})$$

The overlapping concept descriptions classifying an instance are ranging in order of priority by the degree of the affiliation to the concept (the function $\Theta_{\beta_{a_i,k}}(v_{i,j})$) and the degree of undeterminateness (the function $\Pi_{\beta_{a_i,k}}(v_{i,j})$) of the attribute value $v_{i,j} \in \beta_{a_i,k}$ of the instance.

3. SUMMARY

This system parameters help Learning to become less dependent on the initial not relevant instances whose attribute values could provoke the creation of conceptual nodes which in the future will result in wasting time for splitting or merging the nodes. In this aspect these system parameters could prevent to some extent the Learning from going astray and in this way could help to gain time. The Learning remains unsupervised and the system makes decision about the number of concepts without any class information. The system parameters give the possibility to evaluate the overlapping conceptual nodes classifying an instance and to rang them in order of priority as well.

The system parameter 'Maximum possible deviation ' will decrease the creation of the nodes for which the next instances will suggest the need of splitting into its children nodes. This parameter must be settled carefully with respect to the interval possible values of the attribute because it may bring the Learning to the another extremity of creating much conceptual nodes which differ poorly and with a narrow normal distribution and in the future they must be merged with another similar nodes. The other system parameter called 'Threshold of certainty' will stimulate creation of the overlapping conceptual nodes and will decrease the need of 'split' to recover the hierarchy of concepts. If the thresholds are very near to the mean value it may be possible to stimulate creation of the redundant conceptual descriptions that overlap very much and also to increase the need of 'merge' operations to find an optimal hierarchy of concepts able to classify the new instance. These two system parameters must be settled carefully in respect to the described area with each real-valued attribute.

REFERENCES

1. Feigenbaum, E.A., The simulation of verbal learning behavior, in: E. A. Feigenbaum and J. Feldman (Eds.), Computers and Thought *McGrow_Hill, New York, 1963.
2. Feigenbaum, E., A. and Simon , H., EPAM-like models of recognition and learning, Cognitive Sci. 8 (1984) 305-336,
3. Fischer,D.,A hierarchical conceptual clustering algorithm.,Tech. Rept.No.85-21,Department of Information and Computer Scince,University of California,Irvine,CA(1984).
4. Fischer,D., Knowledge acquisition via incremental conceptual clustering. Mach.Learning 2 (1987) 139-172
5. Fischer,D., Knowledge acquisition via incremental conceptual clustering ,Tech. Rept. No. 87-22 (Doctoral Dissertation),Department of Information and Computer Science, University of California ,Irvine ,CA (1987)
6. Lebowitz,M.,Generalization from natural language text, Cognition Sci. 7 (1983) 1-40
7. Lebowitz, M., Concept learning in rich input domain :Generalization based memory, in : R.S. Michalski,J.,G. Carbonell and T.M. Mitchell (Eds) , Machine Learning :An Artificial Intelligence Approach 2 (Morgan Kaufmann ,Los Altos CA,1986).
8. Lebowitz,M., Experiments with incremental concept formation: UNIMEM ,Mach. Learnning 2 (1987) 103 - 138
9. Everitt ,B., Cluster Analysis (Heinemann Educational,London,1974).
10. John H. Genneri,Pat Langley and Doug Fischer ,Models of Incremental Concept Formation ,Department of Information and Computer Science, University of California,CA 92717,U.S.A.
11 Blaga Nikolova Iordanova, Incremental Model of Concept Formation. In Proceedings of the Conference of the ACMBUL, Varna, 1993.
12. Atanassov K. Intuitionistic fuzzy sets. Fuzzy sets and Systems, Vol. 20 (1986), No. 1, 87-96.
13. Atanassov K. More on intuitionistic fuzzy sets. Fuzzy sets and systems, Vol. 33, 1989, No. 1,37-46.