# Unknown Knowledge Mining: An Incremental Fuzzy Approach

Abdullah Al Mamun, Ali Jaoua, Mays Al-Shaikhli, Seeba Abdulkader, Somaya Al-maadeed
Department of Computer Science & Engineering, Qatar University
Doha, Qatar
mm1610594@qu.edu.qa

***Abstract*— Machine learning methods are an integral part of artificial intelligence, concerned with the design and development of algorithms that allow computers to evolve behaviors based on experimental data. Machine learning deals with discovering knowledge from databases by analyzing the data to build a classifier and classify new patterns is an efficient way. This paper is based on the rule induction approach which overcomes the demerits of the decision tree approach. The paper makes use of the fuzzy incremental production rule (FIPR) which is a rule induction system that generates imprecise and uncertain IF-THEN rules from data records. They are well suited for adaptive learning and knowledge discovery. An incremental maintenance of the knowledge base with a minimal overhead can be realized using FIPR. This project deals with classifying a new object to any one of the existing classes using FIPR and update the database with the new item. Further when another object is to be classified it does the knowledge discovery from the updated database.**

***Keywords—FIPR; Machine Learning; style; Incremental;***

## I. Introduction

Machine learning is programming computers to optimize a performance criterion using empirical data or experience. It is one of the best existing methods to solve decision making problems. The success of machine learning system also depends on the algorithms that control the search to find and build the knowledge structures. The learning algorithms should extract useful information from existing database. Pattern recognition, speech recognition, face recognition, etc. are some of the major applications of machine learning. The previous methods adopted for these applications were based on supervised learning which makes use of decision trees to create rules to predict the output for future inputs. Although such rules are easier to understand and simple to implement there are many exemptions that are not shielded by the rule and hence more error prone. We make use of FIPR to tide over the shortcomings of supervised learning.

In this field, the input is a relational database. Each tuple in the database is an object that consists of a record of values of two attributes. An endogenous attribute, which is the class label and some exogenous attributes, which are the attributes used to define the class assignment. To predict the class of an unresolved object, we need to extract a set of IF-THEN rules from the resolved objects i.e. objects for which we know the exact class [1] This problem becomes more complex when we want to get rid of the exact representation of knowledge and to deal with imprecise or inexact information. So, fuzzy set theory that has brought many tools for handling inexact knowledge represented by linguistic terms becomes the motivating choice. Fuzzy set theory provides a strict mathematical framework in which unclear conceptual phenomena can be precisely and thoroughly studied. It can also be regarded as a modelling language, well fit for circumstances in which fuzzy relations, criteria, and phenomena exist [2].

Most of the researchers in machine learning and knowledge discovery make use of fuzzy set theory. Several methods such as IPR, PVM, CART, ID3, SIPINA and FIPR are used for machine learning. An imprecise and indeterminate knowledge acquisition method based on the Fuzzy Incremental Production Rule (FIPR) is presented in this paper. The recurring problems of classic learning can be solved by using this FIPR method. In this expert system, fuzzy encoded crisp value of the input data allows linguistic representation of knowledge which can handle uncertain rules. Besides, with minimal overhead, it permits the incremental maintenance of the discovered knowledge, and arranges the knowledge base by inducing meta-knowledge rules [1]. Incremental learning enables the system to extrapolate rules repeatedly as new data becomes available without forgetting the previously resolved ones and without referring at any time to the earlier used data. Hence, the system becomes self-adaptive and the acquired knowledge becomes self-corrective. As new data arrives, new rules may be framed and prevailing ones modified allowing the system to evolve over time.

In this project, the input is a relational database called iris database which consists of 3 different categories of flowers. When a new object comes, the classifier should match it to any one of the existing patterns and update the database. Our objective is to develop an algorithm to keep the error probability in classification to the minimum.

## II. Related Work

[3] presents a new tactic for learning fuzzy rules in an incremental way which constitutes three steps: incremental clustering, rule generation and rule base optimization. During each step of the fuzzy classifier construction, the incrementality aspect is taken care of. The experiments revealed that as new data batches arrive, new rules may be discovered and existing ones may be updated or partially removed. In [4], an incremental learning based multi-objective fuzzy clustering for categorical data is projected. For this purpose, a multi-objective modified differential evolution based fuzzy clustering algorithm is developed that yields a set of optimal clustering solutions called pareto-optimal solutions, by optimizing two

contradicting objectives simultaneously. Consequently, through incremental learning using a well-known supervised classifier, called random forest classifier, the final solution is evolved from the ensemble pareto-optimal solutions. In [5], a new technique to database summarization is presented through a hierarchical conceptual clustering algorithm called SAINTETIQ that incrementally builds a summary hierarchy from database records. To handle the imprecise information and improve the robustness of this process, the fuzzy set-based presentation of the input data is used. Also, to make a highly intelligible summury and synthesis, a user defined dictionary is provided.

[6] presents an incremental learning algorithm within the framework of a fuzzy intelligent system which is based on priority values attached to fuzzy rules that are generated based on the fuzzy belief values of the fuzzy rule derived from the training data. The algorithm can detect and recover from incorrect knowledge once new knowledge is available; it will not lose the valuable knowledge generated from the old data while it attempts to learn from new data; and it offers a mechanism allowing to stress on knowledge learnt from the new data. This algorithm has been implemented efficiently in fuzzy intelligent systems. [7] introduces Learn++, an algorithm for incremental training of neural network (NN) pattern classifiers which allows supervised NN paradigms, such as the multilayer perceptron (MLP), to accommodate new data, including examples that correspond to previously unseen classes. The algorithm does not require access to previously used data during subsequent incremental learning sessions, but it does not forget previously acquired knowledge. The outputs of the resulting classifiers are combined using a weighted majority voting procedure.

The improvement of fuzzy inference and neuro learning is integrated in fuzzy neural network (FNN) which has been manipulated by numerous scholars. FNN follows the style of human inference and natural language representation that helps process the network quickly and parallel. FNNs become a famous tool to solve the various real-life problems due to its accurate learning ability.

## III. Basic Background

Classical set theory allows the membership of elements in the set in binary terms, a bivalent condition i.e. an element either belongs to or does not belong to a set. Fuzzy sets are sets whose elements have degrees of membership. Fuzzy set theory permits gradual assessment of membership of elements in a set, described with the aid of a membership function valued in the real interval [0,1]. With fuzzy sets, an element belongs to a set with certain degree of certainty.

In this section, we explain the fundamental notions used to build a crisp relation from a fuzzy relation and to cover the crisp relation by a set of rectangles [1].

Let O be the set of objects and L be the set of properties (linguistic values) in a fuzzy relation.

**Definition 1**. A fuzzy binary relation on the universe U = O x P, is a function R : O × P ~ [0, 1]. For o∈O, p ∈ P, the value R(o, p) ∈ [0,1] represents the degree to which o and p are related under R.

**Definition 2**. Let α ∈ [0, 1]. The α-cut of R, denoted Rα, is a crisp binary relation such that, for all (o,p) ∈ 0 × P, Rα(o,p) = 1, if R(o,p) ≥ α. Otherwise, Rα (o,p) = o.

**Definition 3**. For any crisp binary relation Rα (as an or α-cut), we associate the following definitions. The domain of Rα is defined by Dom(Rα) = {e|∃e':(e,e' ) ⊆Rα}; The codomain of Rα is defined by Cod(Rα)= {e' | ∃e: (e, e') ⊆Rα }; The cardinality of Rα is defined by Card(Rα) = number of pairs in Rα..

**Definition 4**. A rectangle of Rα is a cartesian product of two sets (A, B) such as A ⊆O, B⊆P and A ×B⊆R.

**Definition 5**. A rectangle (A, B) is said to be maximal if A × B ⊆ A' × B' ⊆ Rα ⇒ A = A' and B = B'.

**Definition 6**. The maximal rectangle containing (a, b) is said to be optimal if it maximizes the gain function. The gain function of a rectangle RE = (A, B) is given by:
g(RE) = Card(A) * Card(B) - [Card(A) + Card(B)].

**Definition 7**. The coverage of Rα is defined as a set of optimal rectangles CV = { REI, RE2,..., REn } of Rα such as any element (a, b) of Rα is included in at least one rectangle of CV.

**Definition 8**. A coverage CV = {REI, RE2,..., REn} of a relation Rα is minimal if it is composed by a minimal number of optimal rectangles.

In our project, we chose the α-cut of the fuzzy relation as 0.15 to find the crisp relation. Then we found the minimal coverage of the crisp relation by a technique called formal concept analysis (by grouping the ones to form a minimal set of optimal rectangles). For data clustering, a very straight forward and well defined method is offered by formal concept analysis. Here, we minimize information representation by selecting only "optimal concepts". We assume that data may be converted to a binary relation as a subset of the product of a set of objects and a set of properties [8].

## IV. Methodology

The method of FIPR approach consist of three steps: representation of data to transform the linguistics value from relational dataset to binary relations by determining the membership functions, Rule extraction from the optimal rectangle to obtain the minimal coverage rectangular and FIPR engine will describe the system of FIPR briefly.

### A. Representation of Data

In Human nature, the human reasoning is build based on vague notations which are decision rules. In our Iris dataset, it's interesting to know if specific plant is "VERSICOLOR" and has a high sepal width, but it's not if we know that sepal width is 2.5 cm or 2.8 cm. To transform the attributes to an understanding modal the "Modal representation problem" will solve this type of well-known patterns. This modal is a pre-processing step that will model each linguistic value by an interval, the major difficulty is to choose appropriate threshold to get the optimal number of intervals that will decompose into three values "LOW, MEDIUM and HIGH".

In our project, the user will define their membership function for each attribute to transform the relational dataset into fuzzy binary relation R. For example, fuzzification of five relational objects in Iris dataset with one attribute which is petal length, the user will choose appropriate membership function to cover the attributes by three value "LOW, MEDIUM and HIGH". For example, the system will transform 4.8 of O3 using the membership function to "O3 equals 4.8 cm" by the expression "O3 has a medium petal length of degree 0.86 and high petal length of degree 0.14 (Fig. 1).

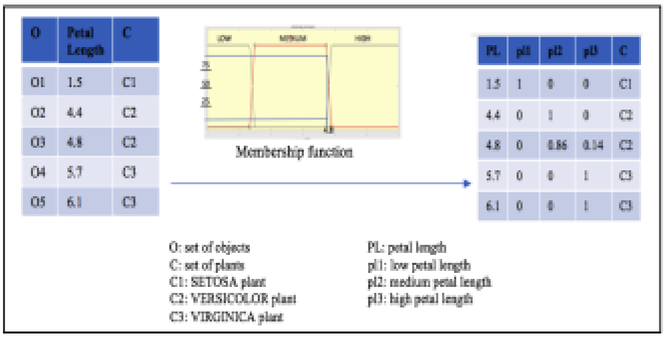

Fig. 1. Data representation of a relational database to fuzzy database

Algorithm 1 describes how to generate a triangular membership functions using the clusters [3].

### *B. Rule Extraction*

In representation of data step, we proposed how to transform from relational dataset to fuzzy dataset (binary relation) R in O x P, where O (set of objects) and P (set of properties) using a membership function chosen by the user. With the modal representation of dataset, the optimal rectangle RE= (A, B) it is a Cartesian product between A (subset of objects) which is the domain of the RE and B (subset of properties) which us the codomain of RE.

The benefits of the Cartesian product are to assure that all objects of the domain A are verify with all properties of the codomain B. Hence we will not

**Algorithm 1** : Generation of triangular functions

**for all** features **do**
- Project the clusters $(l_j, v_j, r_j)$ onto the current feature axis $i$. Let this projection correspond to partition $j$ given by $A_{ij}$ which is defined by $(l_{ij}, m_{ij}, r_{ij})$ corresponding to the left, middle, and right endpoints of the triangular function.
- Sort the projected clusters in an ascending order according to their first endpoint $l_j$.

**for all** sorted clusters **do**
- Compute the breakpoints of the triangular function $A_{ij} = (l_{ij}, m_{ij}, r_{ij})$ for the current cluster $(l_j, v_j, r_j)$ as follows:
The projected center $v_j$ corresponds to $m_{ij}$.
**if** the current cluster is the first on the feature axis, **then**
$l_{i1} = l_j$
**else**
**if** $r_j < l_{j+1}$ // j+1 means next cluster on the axis **then**
$l_{ij} = \frac{r_j + l_{j+1}}{2}$
$l_{i,j+1} = r_j$
**else**
$r_{ij} = r_j$
$l_{i,j+1} = l_{j+1}$
**end if**
**end if**
**end for**
**end for**

find any other linguistic value that will is controlled by all the objects of A, similarly to the objects that will not find any objects controlled by all the linguistic values, this will describe the definition of the maximal rectangle.

Every optimal rectangle RE of R can be describe class of objects, which will verify all the properties of the B (codomain) of RE. In general, for each optimal rectangle RE can assign a class by categorizing the properties. From the characteristics of the optimal rectangle, each RE of the fuzzy relation R (binary) equals to the data presentation of the production rule and the codomain of RE contain the rule and domain cardinality of RE that will used in case of conflict rule problem. Since the optimal rectangle RE is NP hard complete its hard to searching for the optimal rectangle, it is same difficulty of extract the rule from the dataset, to deduce the minimal coverage of optimal rectangles of R from the database

In our project, (Fig.2) describes how to generate the minimal coverage $C_v$ looked from a rectangular decomposition of fuzzy binary relation with $\alpha$ = 0.15 of fuzzy binary to obtain a crisp binary relation one in (Iris database). By looking to Algorithm 2, it describes how to cover the crisp binary relation by minimal coverage of optimal rectangles. Let $R_\alpha$ is the crisp relation that need to include then divide the $R_\alpha$ into $p$ packages from $p_1$, $p_2$ …, $p_n$. Each sub-relation is a subset of crisp relation which consist of one or more pairs. The idea is built in steps the minimal coverage of $R_\alpha$ .First, we will represent the crisp relation related to its package (e.g. R1 = P1 by CVI). In the *kth* step, let $R_{k-1} = P1 \cup P2 \cup \ldots \cup P_{k-1}$ and let $CV_{k-1}$ be its minimal coverage of $R_\alpha$ , the seek is to build the minimal coverage $CV_k$ of $R_k = R_{k-1} \cup P_k$ using only the initial coverage $CV_{k-1}$ and the package $P_k$.

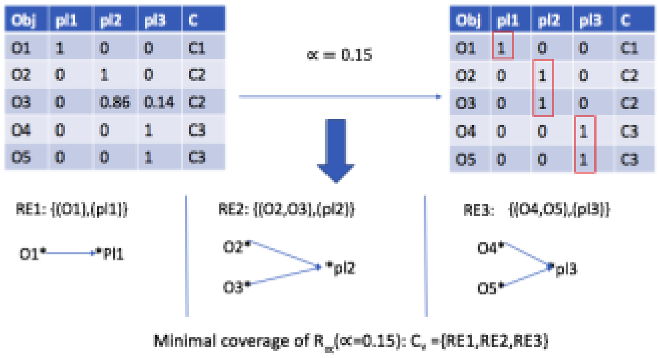


Fig. 2. Rectangular decomposition of a fuzzy binary relation

In the *pth* step, we obtain a set of optimal rectangles covering the relation $R_{\alpha}$. Which is also prove about the usefulness of Incremental fuzzy [1].

**Algorithm** 2- Generate optimal rectangles covering by R
Begin
1. Let $R_x$ be partitioned into $p$ packages $P_1, \ldots, P_p$.
2. **FOR** $k = 1$ **to** $p$ **DO**
2.1. Sort the couples of $P_k$ by the value of the gain function in a decreasing order
2.2. **While** $(P_k \neq \emptyset)$ **Do**
- Select a couple $(a, b)$ in $P_k$ by the sorted order
- Search PR: the pseudo-rectangle containing $(a, b)$ within $R_k = \mathrm{CV}_{k-1} \cup P_k$
- Search RE: the optimal rectangle containing $(a, b)$ within PR
- $\mathrm{CV}_k = (\mathrm{CV}_{k-1} - \{r \in \mathrm{CV}_{k-1} / r \subseteq \mathrm{RE}\}) \cup \{\mathrm{RE}\}$. (Delete all the redundant rectangles from $\mathrm{CV}_k$)

$\mathrm{P}_k = \mathrm{P}_k - \{(X, Y) \in P_k / (X, Y) \in \mathrm{RE}\}$
End **While**
End **FOR**
End

After determining the minimal coverage of $R_{\alpha}$, if there is new object comes to test, the system will perform the incremental step of rectangular decomposition algorithm. The system will assign the new object to the suitable coverage by applying fuzzy method to the values to decide to which rectangle belongs to and update it. (Fig.3) explains the addition of the new coming object O6 to the coverage $C_v$ of Incremental method.

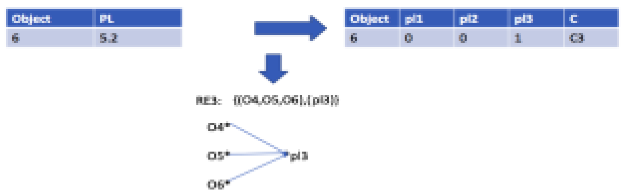


Fig. 3. Adding new object to the coverage $C_v$

(Table 1) explains how to extract the rule associated (IF-THEN) rules from an optimal rectangle, objects that we already know it belongs to which class to predict the class of a new coming objects (unresolved objects).

| The optimal rectangle | The rule associated |
|---|---|
| O1*→*Pl1<br>RE1 belongs to class C1 | **IF** low petal length **THEN** "SETOSA" plant |
| O2*, O3* → *pl2<br>RE2 belongs to class C2 | **IF** medium petal length **THEN** "VERSICOLOR" plant |
| O4*, O5*, O6* → *pl3<br>RE3 belongs to class C3 | **IF** high petal length **THEN** "VIRGINICA" plant |

Table 1. Example of imprecise production rules extracted from Fig.2, Fig.3.

## V. FIPR ENGINE

The FIPR system use the incremental fuzzy method to generate the new objects, with a query from a user to know the type of plant that is associated to a certain attribute, for example if the petal length = 1.5 cm as shown in Fig.1, so what is the class of the 1.5 cm petal length? The engine of the FIPR will transform the petal length of 1.5 cm into a fuzzy fact by selecting set of rules, then will move forward to evaluation of fuzzy conditions. By handling several predictions, and use the evidence combination if there are more than two rules have the same predictions, the user will be referred to. For (r1 and r2) rules have common facts can be solved as contributing evidence. The final solution of Cr can be expressed from the evidence. For instance [1,11]:

| | |
|---|---|
| **rules** | r1 : IF A1 THEN C<br>r2 : IF A2 THEN C |
| **facts** | A'1 , A'2 |
| **conclusion** | $C_r$ obtained from C' with CF1 and C'' with CF2 |
| where, | |
| r1, r2 | : rule codes |
| A1, A2 | : antecedent propositions |
| C | : consequent proposition |
| C', C'' | : conclusions from (r1, A1') and (r2, A2')<br>C' = C'', since C is non fuzzy |
| CF1, CF2 | : certainty factors of the conclusion |

The overall factor CFr for Cr are grouped from the respective conclusions of C′ and C″ the two rules are generated and combined every time by using the formula of:

$$\mathrm{CFr} = \mathrm{CF1} + \mathrm{CF2} - (\mathrm{CF1} * \mathrm{CF2}).$$

## VI. EXPERIMENTATION

Raw data is processed to find the min alpha-cut so that we can reduce the error rate. Initially, petal-length is considered as the only feature in this experimentation. Once we find optimal alpha-cut, we created fuzzy membership function in using fuzzy type-2 logic controller library. The raw data then feed into fuzzy membership function to evaluate the value of three different petal lengths such as pl1, pl2, and pl3. Now, fuzzification result is ready to train our model. The main code is written in Java. Total data is divided into two parts such as train data and validation data. The amount of data distribution is as 60% training data and 40% validation data. The next important step

*Table 2 Machine learning methods time complexities and error rate comparisons*

| | Methods | Error rate | Accuracy | Complexity | Knowledge representation |
|---|---|---|---|---|---|
| Rule Induction methods | PVM | 0.040 | 0.96 | $\sum_{k=1}^{d} (n * d)^{2^{k-1}}$ | Symbolic |
| | IPR | 0.70 | 0.30 | $n^2 * (c + d)^2$ | Symbolic |
| | FIPR | 0.01 | 0.99 | $n^2 * (c + d)^2$ | Linguistic |
| | | | | | |
| Decision Tree methods | CART | 0.047 | 0.95 | $c * d^{11}$ | Tree |
| | ID3 | 0.040 | 0.96 | $c^2 * d^{11}$ | Tree |
| | SIPINA | 0.040 | 0.96 | $c^2 * d^2$ | Graph |
| | CNN | 0.12 | 0.88 | $O(\sum_{l=1}^{d} nl - 1^{sl^2} * nl\ ml^2)$ | Multi-layer |
| | RNN | 0.09 | 0.91 | - | Multi-layer |

is to run the model and train using training dataset. Once training is finish with ten separate epoch (run), the model is mature enough to test with validation data set. After validation process, the machine learning result is compiled in validation dataset in a CSV file. From this data, we can calculate the accuracy and error rate.

*CNN:* a feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex [14].
*RNN:* a neural network where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior [15].

## VII. Comparative Analysis and Evaluation

The experiments are done in the best-known Iris dataset to find the instance recognition. It is one fifty instances, which includes three classes of 50 instances in each and each class refers to an Iris plant. The three plants classes are SETOSA, VERSICOLOR and VIRGINICA. SETOSA class is linearly separable from remaining two, while VERSICOLOR and VIRGINICA are not linearly separable from each other. Each instance described by 4 attributes which is Sepal length in cm, Sepal width in cm, Petal length in cm and Petal width in cm. To predict the class of Iris plant of a new coming object we can use 8 methods of machine learning which is:

*ID3:* This method uses a chi-square test to test if there is a relationship between two categorical data, it depends on the confidence level of the statistical test [9].
*CART:* This method generates several classification and regression decision trees based on random samples from the dataset. The predictions are made for the objects that not included in the training data [10].
*FIPR:* This system allows fuzzy incremental production rule to generates imprecise IF-THEN rule from the full dataset [1].
*PVM:* This method called Predictive Value Maximization method is a heuristic search procedure through the space of variables to find the best combination of test for deciding [12].
*IRP:* This method used to transform a relation database to a binary database using incremental rule production by covering minimal coverage of rectangles [11].
*SIPINA:* This tool package includes several machine learning methods to develop the classification of the database [13].

In this paragraph, we will study the complexity analysis of the precision measure of the machine learning methods by letting d (number of instances), c (number of attributes) and n is the binary relation in dataset (n=c*d).

By using predictive value maximization rule method (PVM), first we select an arithmetic operator for each attribute using the median [O(c*d)]. Then, we calculate the percentage of the predictive value for each attribute and find the local max [$O(c*d)^2$]. Finally, create a table containing rules of size one, by choosing a threshold some rules will remove and the remaining rules will produce rules of size two [$O(c*d)^2$]. Same steps will be evaluated again using a threshold to remove some rules of size two $O[(c*d)^2*d^2)]$, so the general complexity of kth iterations is $O\ [(c * d)^{2^{k-1}} * d^{2^{k-1}}\ ]$. As it is explained before that n=c*d, we can express the general complexity by $O\left[\sum_{k=1}^{c} (n * d)^{2^k}\right]$ as shown the PVM method is an exponential complexity rule which is hard to implement it to big datasets.

In the incremental rule production (IRP) method, the time complexity differs where analyze the minimal coverage rectangular generation algorithm. The function which uses to calls called (Rect-Optimal), the complexity of this function is O(n) by looking for the optimal rectangle (a,b), the complexity cost will be $O[n^2*(c+d)^2]$.

In the fuzzy incremental production rule (FIPR), the time complexity depends on the incremental step that's generates on rectangular decomposition algorithm to find the minimal coverage of the rectangles which cost O(n) and look for the optimal rectangle (a,b), the complexity cost will be $O[n^2*(c+d)^2]$.

In the classification and regression decision tree (CART) method, the number of tree nodes around $O(d^2)$. By adding new node, it will build according c attributes at the most of d times. With each time (instances) will evaluate the tree $[O(c*d^2)]$ but need $O(c*d^4)$ operations.

The perform the pruning function, first, by using a method of evaluation the leaving one-out, by constructing d trees. With every tree, will test the k nodes ($k=d^2$). With every node pruned, will take $O(d^2)$ evaluations and $O(d^2)$ comparisons. Which makes $O(c*d^{11})$. Then, by evaluating k*d will a test sample tree to choose the optimal coverage $[O(d^5)]$. Finally, from the last tree will extract the final solution tree. For K pruning the complexity is $[O(c*d^{10})]$. For the interactive system for interrogation process (SIPINA) method done for non-tree decisions which gives the human experts the chance of visiting the leaves of the tree with a time complexity $n^{2k}(c+d)2$ [1,11].

Table 2 explains the error rate of the machine learning methods that mentioned previously. We note that the effect of fuzzy incremental on knowledge representation which produces a better error rate than all other methods which gives a more effective comparative to FIPR which induces rules based on linguistic terms written.

## VIII. CONCLUSION

In this paper, we presented the incremental fuzzy production rule method. This method generates the rule by IF-THEN rules which differs from any decision method and many production rules, FIPR method have a polynomial complexity which proves the efficient of this algorithm in the machine learning sector. The FIPR method handles the imprecise and uncertain rule which applied to classifies the tasks per their type of plant on Iris data. The precision measures proved that incremental fuzzy theory improve the precision of presentation from the database. By studying other methods for incremental machine learning and experienced the FIPR we conclude that there is a good rate improvement than others by extracting the rules using a specific membership functions and by fixing the value of alpha-cut to be the minimal value that should appear in the fuzzy relation. The FIPR futures allow us to implement the system to the largest applications.

## REFERENCES

[1] M. Maddouri, S. Elloumi, and A. Jaoua, “An Incremental Learning for Imprecise and Uncertain Knowledge Discovery”, Journal of Information Sciences, Elsevier N.H. 109 (1998) 149-164.

[2] H. Zimmermann, “Fuzzy set theory”, Journal of John & Sons Inc., 2(2010).

[3] A. Bouchachia and R. Mittermeir, “Towards incremental fuzzy classifiers”, Soft Comput (2007) 11: 193–207 DOI 10.1007/s00500-006-0077-3

[4] Saha, Indrajit, and Ujjwal Maulik. "Incremental learning based multiobjective fuzzy clustering for categorical data." Information Sciences 267 (2014): 35-57.

[5] Raschia, Guillaume, and Noureddine Mouaddib. "SAINTETIQ: a fuzzy set-based approach to database summarization." Fuzzy sets and systems 129.2 (2002): 137-162.M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989

[6] Murphey, Yi Lu, and Tie Qi Chen. "Incremental learning in a fuzzy intelligent system." Proceedings of the 16th international joint conference on Artificial intelligence-Volume 2. Morgan Kaufmann Publishers Inc., 1999.

[7] Polikar, Robi, et al. “Learn++: An incremental learning algorithm for supervised neural networks." IEEE transactions on systems, man, and cybernetics, part C (applications and reviews) 31.4 (2001): 497-508.

[8] A. Hasnah, A. Jaoua, J. Jaam , T. Mosaid, F. Hassan, H. Saleh, F. Abdullah, H. Cherif, and Hassan Mohamed. “CONCEPTUAL DATA CLASSIFICATION: APPLICATION FOR TEXT SUMMARIZATION”

[9] J. Quinian, “Introduction of Decision Trees”, Machine Learning. 1 (1998) 81-106.

[10] M. Petersen, A. Tolver, L. Husted and T. Pihl, “Repeated measurements of blood lactate concentration as a prognostic marker in horses with acute colitis evaluated with classification and regression trees (CART) and random forest analysis”, Journal of The Veterinary. 213 (2016) 18-23.

[11] M. Maddouri and A.Jaoua, “Incremental Rule Production: Towards a Uniform Approach for Knowledge Organization”, Gordon and Breach Science Publishers. (1996) 295-304.

[12] M. Weiss, R. Galen and P. Tadepalli, “Maximizing the predictive value of Producion Rules”, Rutgers University, Department of Computer Science. NJ 08904, Ohio

[13] W. Abramowicz and J. Zurada, “Book: Knowledge Discovery for Business Information Systems”, Kluwer Acadmin Publishers. (2002) eBook ISBN: 0-306—46991-X.

[14] Wikipedia webisite [URL: https://en.wikipedia.org/wiki/Convolutional_neural_network]

[15] Wikipedia website [URL: https://en.wikipedia.org/wiki/Recurrent_neural_network]