Unknown Knowledge Mining: An Incremental Fuzzy Approach

Presented to

Dr. Ali Jaoua

By: Md. Abdullah Al Mamun (201610594)

Seeba AbdulKader (sa1606892)

Mays Alshaikhli (200753008)

Outline

- Introduction
- Database Description
- Methodology Adopted
- Live Demo
- Results
- Comparative Analysis
- Conclusion

Introduction

- **Machine learning** is discovering knowledge from databases in order to classify new patterns in an efficient way.
  - Relational database –class labels and attributes
  - Analyze the data, build a classifier and decides upon the class assignment.(IT-THEN)
  - Fuzzy set theory used for dealing with imprecise and uncertain data.
- **Fuzzy incremental production rule (FIPR) system** is a rule that generated imprecise and uncertain IF-THEN rules from data record.
- **Incremental learning** aims at enhancement of knowledge.
- **The aim of our project is to develop self adaptive algorithm, which involves improving performance over time.**

Advantages of Fuzzy classifier

- Classification behavior can be easily understood by human users (expressed in linguistic forms) by checking carefully the fuzzy if-then rules in the fuzzy classifier.

- Nonlinearity in classification, which leads to high generalization ability of fuzzy rule-based classifiers while its classification behavior is linguistically understood.

Database Description

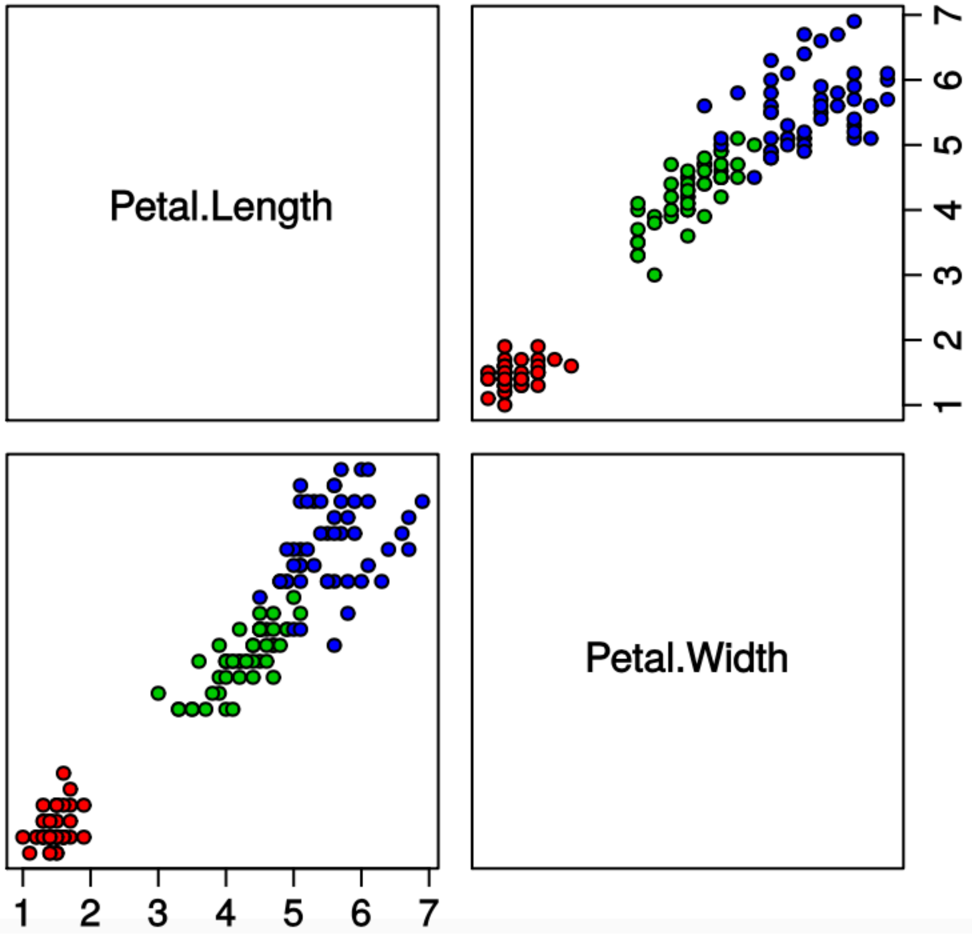
- The experiments are done on the IRIS database.
- It is 150 records with 3 classes (SENTOSA, VERSICOLOR, VERGINICA)
- Each class contains fifty patterns
- Each pattern is described by four numerical attributes
  (Petal Length, Petal Width, Sepal Length, Sepal Width)
- We can use 6 machine learning methods to classify the new coming flower
  (PVM, IPR, FIPR, CART, ID3, SIPINA)
  - Our project implements FIPR .
  - Comparative analysis was done with new methods
    - CNN
    - RNN

Dataset

- Train Data = 30 of each
- Validate = 10 of each
- Test = 10 of each

Petal Length	Petal Width	Flower Name
1.4	0.2	I. Setosa
4	1	I. Versicolor
6.6	2.1	I. Verginica

Classification



86	4.5	Iris-versicolor
107	4.5	Iris-virginica
55	4.6	Iris-versicolor
59	4.6	Iris-versicolor
92	4.6	Iris-versicolor
51	4.7	Iris-versicolor
57	4.7	Iris-versicolor
64	4.7	Iris-versicolor
74	4.7	Iris-versicolor
87	4.7	Iris-versicolor
71	4.8	Iris-versicolor
77	4.8	Iris-versicolor
127	4.8	Iris-virginica
139	4.8	Iris-virginica
53	4.9	Iris-versicolor
73	4.9	Iris-versicolor
122	4.9	Iris-virginica
124	4.9	Iris-virginica
128	4.9	Iris-virginica
78	5	Iris-versicolor
114	5	Iris-virginica

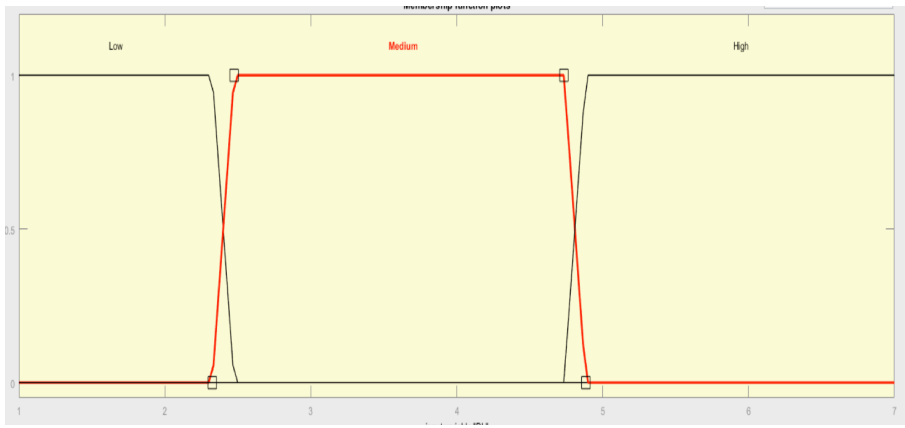
Methodology

- Fuzzification.
- Membership function: Petal Length
- Alpha cut = 0.15
- Variables
  - Input: low, medium, high
  - Output: C1, C2, C3

Fuzzification

- *Step 1: Linguistic representation*

Petal Length	Class
1.5	C1
4.4	C2
4.8	C2
5.7	C3
6.1	C3



PL	pl1	pl2	pl3	C
1.5	1	0	0	C1
4.4	0	1	0	C2
4.8	0	0.86	0.14	C2
5.7	0	0	1	C3
6.1	0	0	1	C3

Fuzzyfication

Step 2: Rectangular Decomposition of a Fuzzy Binary Relation

Obj	pl1	pl2	pl3	C
O1	1	0	0	C1
O2	0	1	0	C2
O3	0	0.86	0.14	C2
O4	0	0	1	C3
O5	0	0	1	C3

$\propto = 0.15$

Obj	pl1	pl2	pl3	C
O1	1	0	0	C1
O2	0	1	0	C2
O3	0	1	0	C2
O4	0	0	1	C3
O5	0	0	1	C3

RE1: {(O1),(pl1)}

O1* → *Pl1

RE2: {(O2,O3),(pl2)}

O2* → *pl2

O3* → *pl2

RE3: {(O4,O5),(pl3)}

O4* → *pl3

O5* → *pl3

Minimal coverage of $R_{\propto}$($\propto$=0.15): C_v ={RE1,RE2,RE3}

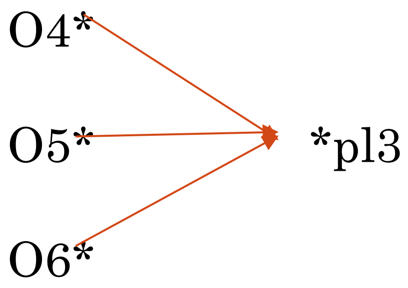
Fuzzyfication

- ***Step 3: Incremental Learning***

Object	PL
6	5.2

Object	pl1	pl2	pl3	C
6	0	0	1	C3

RE3: {(O4,O5,O6),(pl3)}



Algorithm

Begin

1. Let R_α 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 CV_k)

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

End **While**

End **FOR**

End

Demonstration

Live Demo

Versicolor	
PL	PW
53	78
73	71
78	
84	

Virginica	
PL	PW
107	135
127	120
139	134
	130
	107

Results and Discussion

Method s	Error Rate	Accuracy	Complexity	Knowledge Representation
CNN	0.12	0.88	$O(\sum_{l=1}^{d} n_{l\,-\,1} s_l^2\ n_l\, m l^2)$ l = index of layers d = depth of layers n_l = number of filters m_l = spatial size	Multi-layer
RNN	0.09	0.91		Multi-layer
FIPR	0.01	99.0	n^2 (c+d)^2 c = number of attributes d = number of patterns n=cd	Linguistic

Conclusion

- Classifications of new objects are done using FIPR, the accuracy approved a lot when compared to other methods, the time complexity proves that it's with acceptable in compared to existing systems.

- The FIPR maintain the knowledge base with minimal overhead.

- The rule that extracted from the databases with FIPR system handles imprecise and uncertain IF-THEN rules in a linguistic terms.