First Approach toward Online Evolution of Association Rules with Learning Classifier Systems

Albert Orriols-Puig[1]

Jorge Casillas[2]

Ester Bernadó-Mansilla[1]

[1]Research Group in Intelligent Systems
Enginyeria i Arquitectura La Salle, Ramon Llull University

[2]Dept. Computer Science and Artificial Intelligence
University of Granada

Framework

- **Michigan-style LCSs are reaching maturity**
  - First successful implementations (Wilson, 1995; Wilson, 1998)
    - Many other derivations YCS, UCS, XCSF, and many others
  - Applications in important domains
    - ***Data mining*** (Bernadó et al, 02; Wilson, 02; Bacardit & Butz, 04; Butz, 06; Orriols et al, 2008)
    - Function approximation (Wilson, 2002; Butz et al. 2008)
    - Reinforcement Learning (Lanzi et al, 2006, Butz et al. 2006)
    - Clustering (Tamee et al., 2006; 2007)
  - Theoretical analyses for design (Butz et al., 2004; 2007; Drugowitsch, 2008)

Motivation

- **Which types of problems have we shown to be able to solve?**
  - **Data mining** (supervised learning)
    - Classification
    - Function approximation
  - **Reinforcement learning**
  - **Data streaming**: *Hot area in the ML community*
    - **Learn online from stream of examples**
      - **Supervised learning**
      - **Unsupervised learning**
    - Two edited books last year: (Aggarwal, 2007; Gama, 2007)
    - Decision tree able to learn online in JMRL: (Nuñez et. al, 2007)

Aim

- **General purpose**

Mine association rules online from streams of examples,
adapting the model to association drift

- **Why?**
  - Many industrial processes generate streams unlabeled data
    - Marketing problem
    - Modeling Spanish power consumption
  - LCS seems to provide a natural support to mine data online
  - There are not many approaches to this problem

- **Particular scope of the present work?**

First proposal of an architecture able to mine association rules from streams of data

Outline

1. Mining Association Rules
2. Description of CSar
3. Experimental Methodology
4. Results
5. Conclusions and Further Work

Mining Association Rules

- **Goal:** Extract rules that denote relationships among variables with high *support* and *confidence* from data sets

Market basket transactions (Data set)

N.	Items
1	*diapers*, *bread*, *ham*, *cheese*, jam, beer
2	*beer, diapers*, monkfish, sausage
3	salmon, *beer*, *bread*, *diapers, ham, cheese*
	...

Rule set with rules <u>IF</u> X <u>THEN</u> Y

<u>IF</u> diapers <u>THEN</u> beer

<u>IF</u> bread <u>and</u> ham <u>THEN</u> cheese

...

- **Rules evaluation:** *support* and *confidence*

$$supp(R) = \frac{supp(X \cup Y)}{|T|}$$

$$conf(R) = \frac{supp(X \cup Y)}{supp(X)}$$

Approaches to Mine Association Rules

- **First approaches on categorical data**
  - AIS (Agrawal, 1993)
  - Apriori (Agrawal & Srikant, 1994)
  - Many others
  - Limitation: Only categorical data

Many of them, two-phase alg.:

1. **Obtain frequent item sets**
2. **Create rules from these item sets**

- **Quantitative Association Rules Mining**
  - Discretize and apply Apriori (Srikant & Agrawal, 1996; Fukuda et al, 1996)
  - Interval-based Association Rules (Mata et al., 2002)
  - Fuzzy association rules (Yager, 1995; Hong, 2001)
- **Online association rule mining** (Wang et al., 2007)

Outline

1. Mining Association Rules
2. Description of CSar
   - Rule Representation
   - Process Organization
3. Experimental Methodology
4. Results
5. Conclusions and Further Work

Description of CSar

- **Rule representation**
  - Problems with ℓ attributes

    IF x_i **is** v_i **and** … **and** x_j **is** v_j **THEN** x_k **is** v_k

  - Antecedent formed by *a conjunction of ℓ_a variables* ($0 < \ell_a < \ell$)
  - Consequent contains a *single variable*
  - Each variable is defined by
    - Nominal attributes → A single value
    - Continuous attributes → An interval of maximum size **[ℓ_i, u_i];**
      - **Restriction:** $u_i - \ell_i <$ maxInt
  - Example:

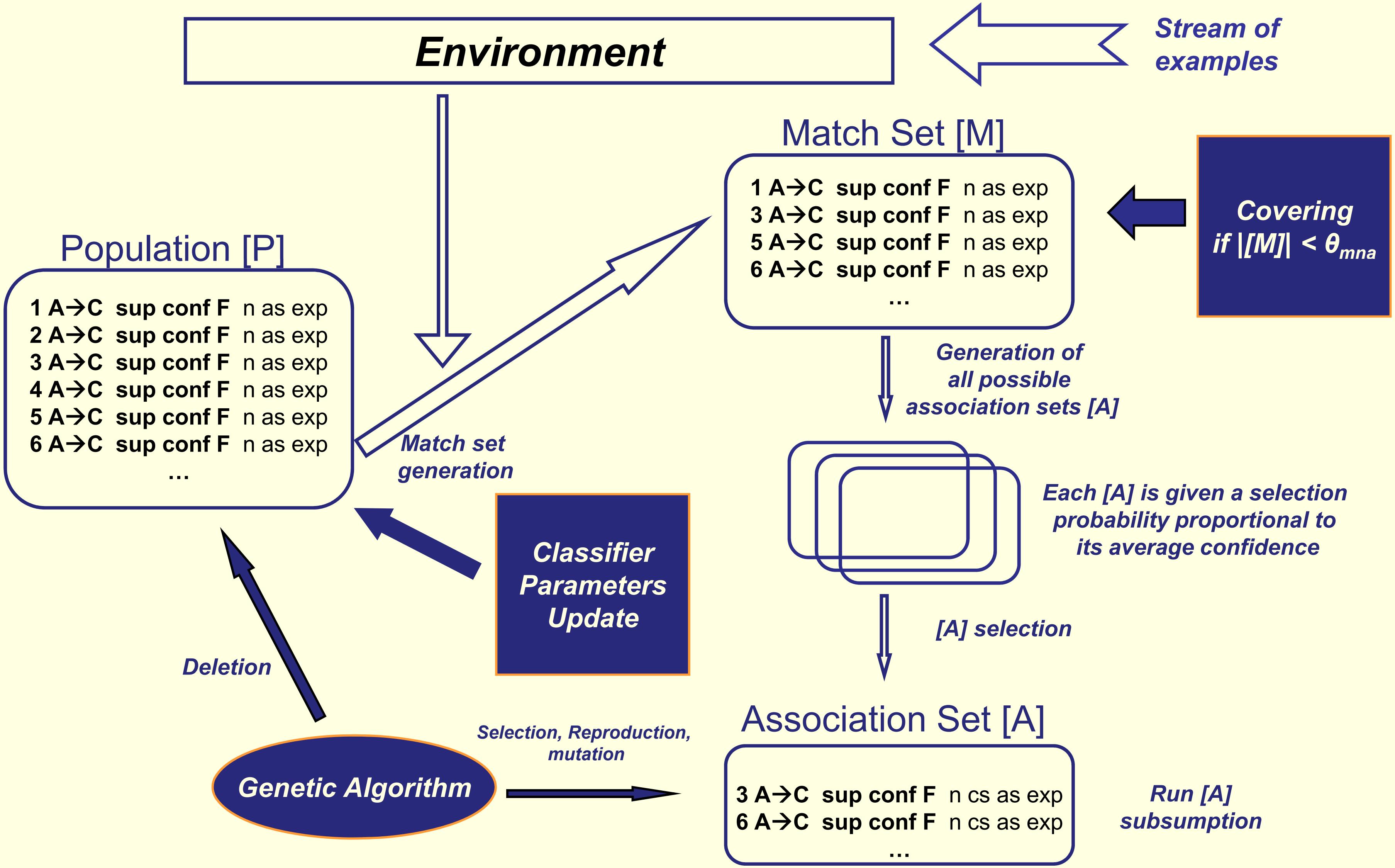
    IF age in [20,25] **and** height in [5,6] **and** sport=yes **THEN** weight **is** [160, 170]

Description of CSar

Classifier Parameters

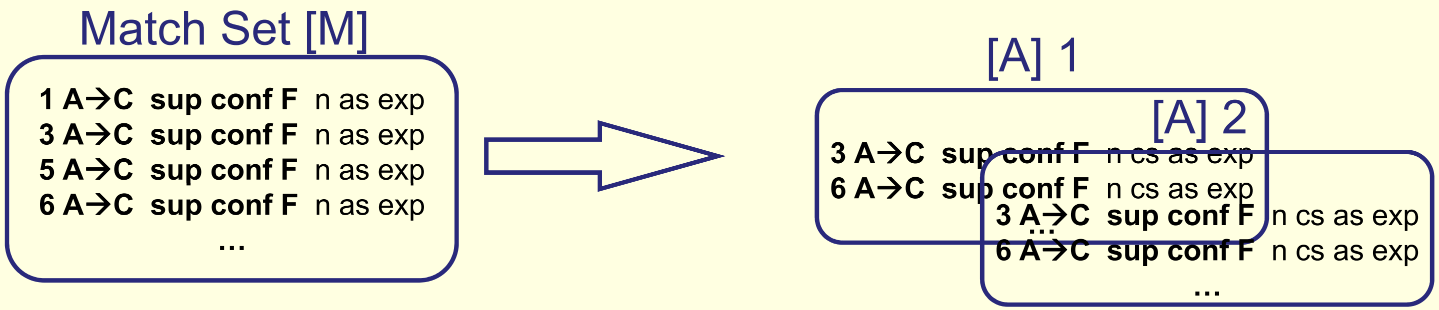
1. Support (supp): occurring frequency of the rule $$supp(R) = \frac{supp(X \cup Y)}{|T|}$$
2. Confidence (conf): strength of the implication $$conf(R) = \frac{supp(X \cup Y)}{supp(X)}$$
3. Fitness (F): computed from support and confidence $$F = (conf \cdot supp)^{\nu}$$
4. Association set size (as): average size of the association sets the rule participates in
5. Numerosity (n): copies of the rule
6. Experience (exp): number of examples matched by the rule antecedent

Description of CSar



Description of CSar

- **Creation of association set candidates**



- **Goal:**
  - Organize the classifiers of [M] in different [A]
  - Similar classifiers should participate in the same [A]
- **Strategies:**
  - Antecedent grouping
  - Consequent grouping

Description of CSar

- **Antecedent grouping:**
  - Put in [A] the rules with the same variables in the antecedent
  - Underlying idea: Rules with the same antecedent may express similar conclusions
- **Consequent grouping:**
  - For nominal rep., put in [A] rules with same consequent and values
  - For interval-based rep., put overlapped rules in the same [A]
  - Underlying idea: Rules that describe the same range of values of the same concept, may have similar antecedents.

Outline

1. Mining Association Rules
2. Description of CSar
3. Experimental Methodology
4. Results
5. Conclusions and Further Work

Experimental Methodology

- **First experiment**:
  - Compare CSar with Apriori (Agrawal & Srikant, 1994)
  - Data set with nominal values: the *zoo* problem
    - 15 binary attributes and 2 categorical attributes
- **Second experiment**
  - Analyze CSar in problems with real-valued attributes
  - The Wisconsin breast-cancer diagnosis problem
    - 9 continuous attributes and 1 nominal attribute
- **CSar configuration**

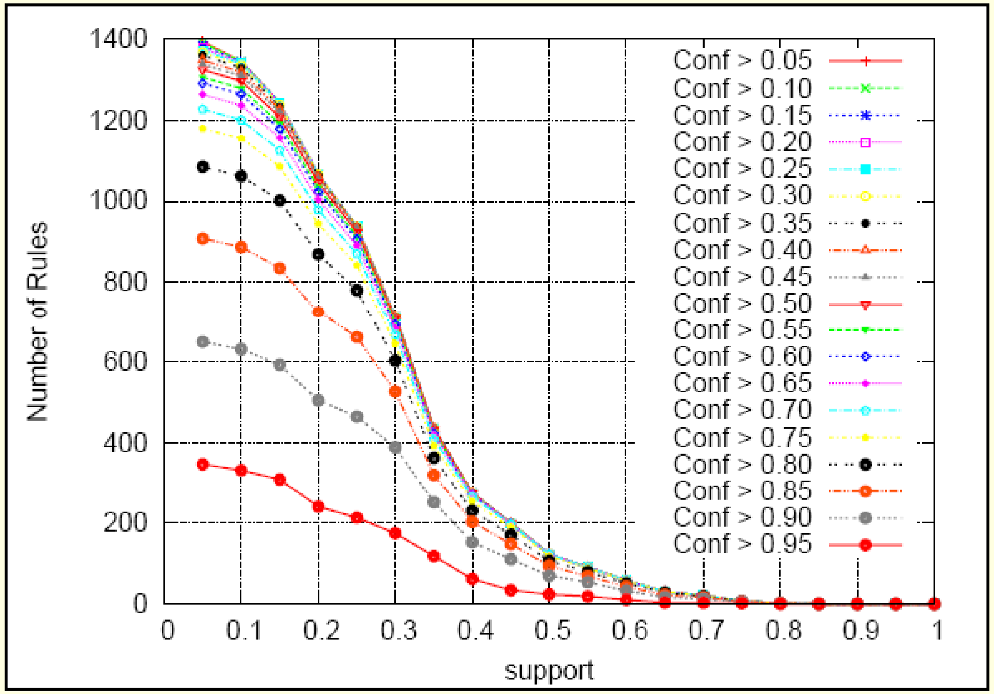
  Iterations = 100,000, popSize = 6,400, θ_{GA} = 50, v = 10, θ_{del} = 10 …
- **Evaluation method:** Count the number of rules for a certain support and confidence

Outline

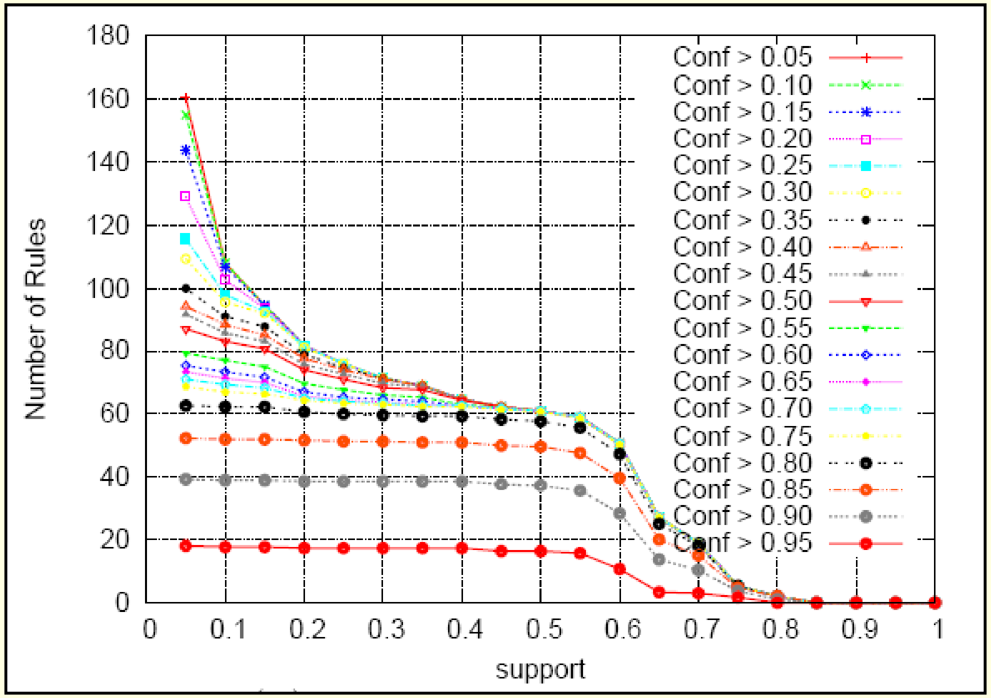
1. Mining Association Rules
2. Description of CSar
3. Experimental Methodology
4. Results
5. Conclusions and Further Work

First experiment: the *Zoo* Problem

Antecedent grouping



Consequent grouping



- **Antecedent grouping maintains more niches and so creates more rules**
- **Rules with high confidence and support are approximately the same in both cases**

First Experiment: the *Zoo* Problem

		Confidence								
		antecedent grouping			consequent grouping			A-priori		
		0.4	0.6	0.8	0.4	0.6	0.8	0.4	0.6	0.8
Support	0.40	274.7	270.7	230.3	64.7	62.7	59.3	2613.0	2514.0	2070.0
	0.50	122.7	122.7	106.0	61.0	61.0	57.7	530.0	523.0	399.0
	0.60	57.7	57.7	50.7	50.7	50.7	47.3	118.0	118.0	93.0
	0.70	21.0	21.0	19.0	19.0	19.0	18.0	30.0	30.0	27.0
	0.80	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	0.90	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

- The three learners discover the same number of rules of high confidence and support.
- As the support and confidence decreases
  - *A-priori* creates a larger number of rules than *Csar*
  - *CSar with antecedent grouping* evolves more rules than *Csar with consequent grouping*

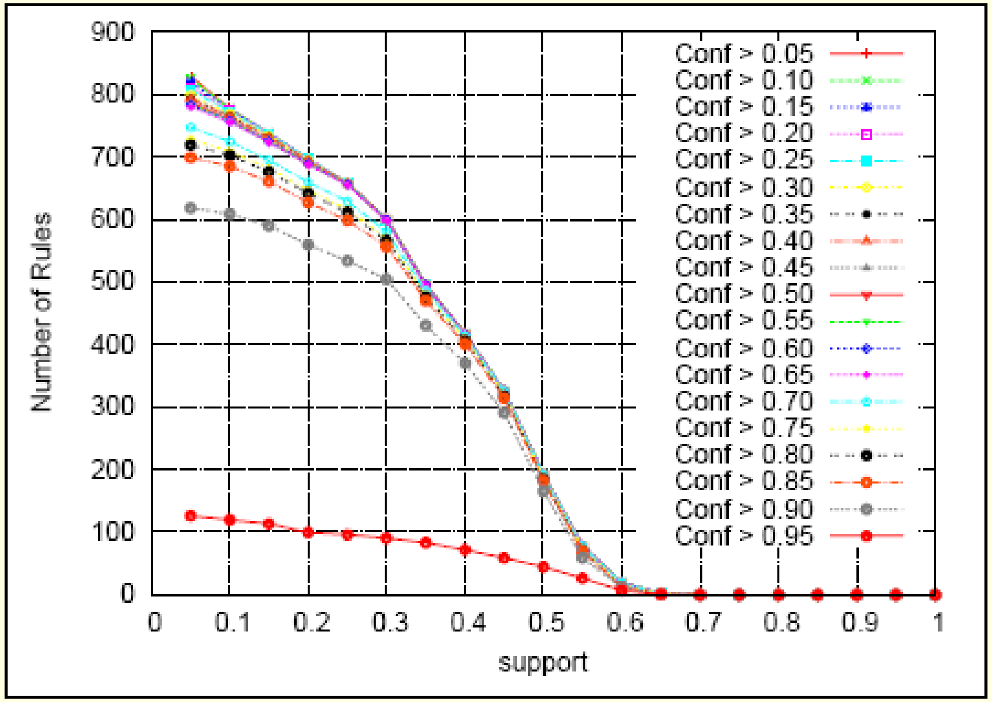
Second Experiment: Continuous Data

- We want to analyze:
  - Whether CSar can extract a population of interval-based rules with high support and confidence
  - Differences between consequent and antecedent grouping
  - The effect of varying the maximum size allowed for an interval
- Apriori not included since it only deals with categorical data
- Which experiments?
  - Move maximum interval to {0.1, 0.9}
  - Use the same configuration as in experiment 1

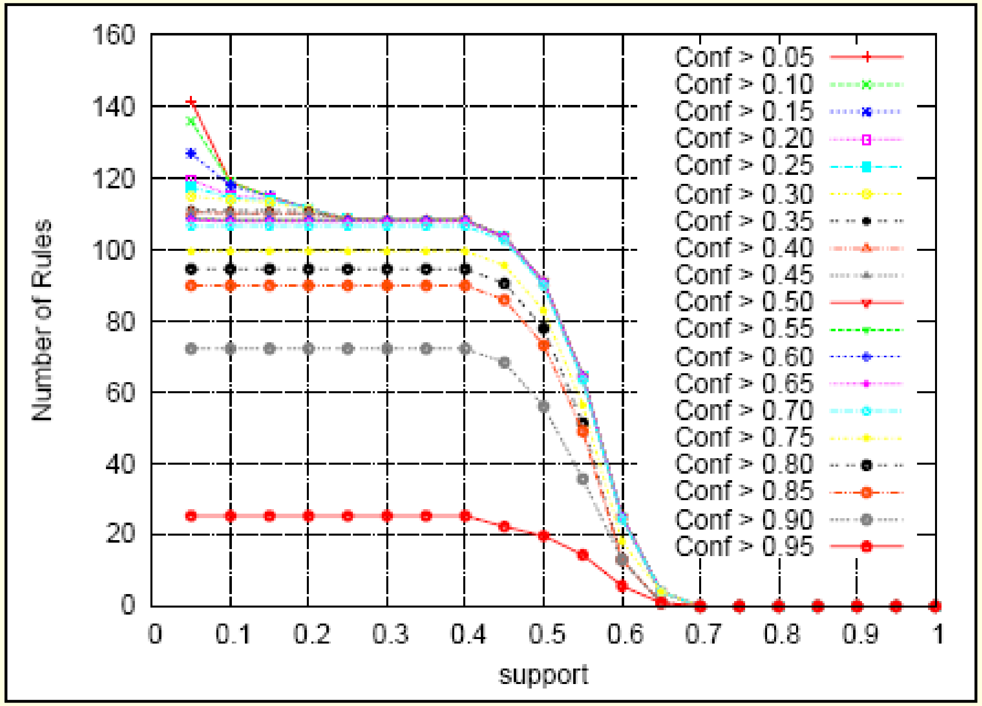
Second Experiment: Continuous Data

MAXIMUM INTERVAL SIZE = 0.1

Antecedent grouping



Consequent grouping

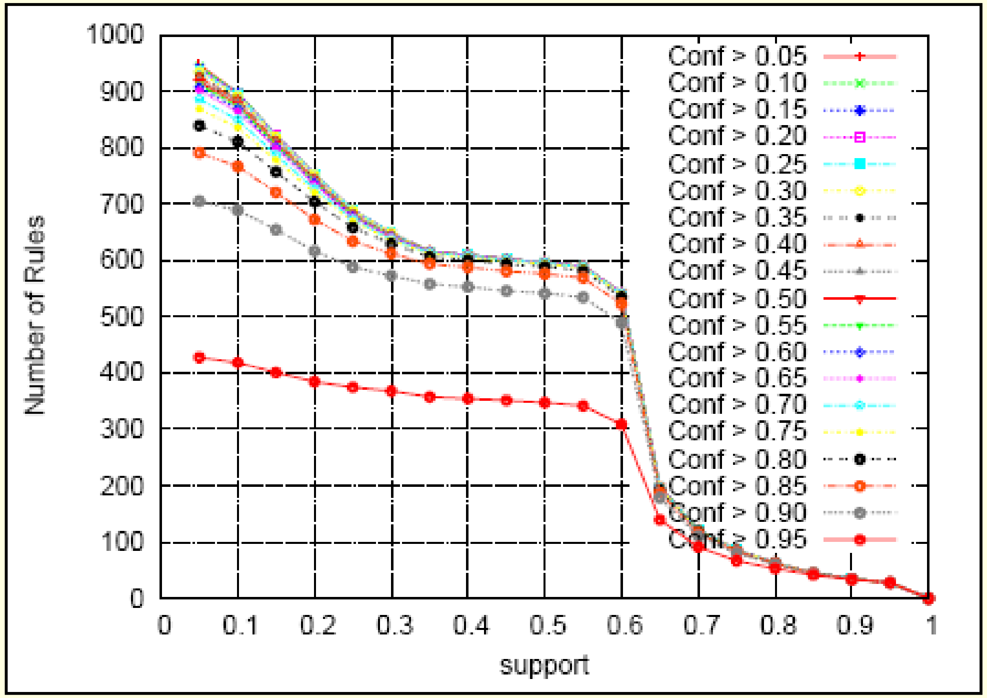


- **Antecedent grouping maintains more niches and so creates more rules**
- **Rules with high confidence and support are approximately the same in both cases**

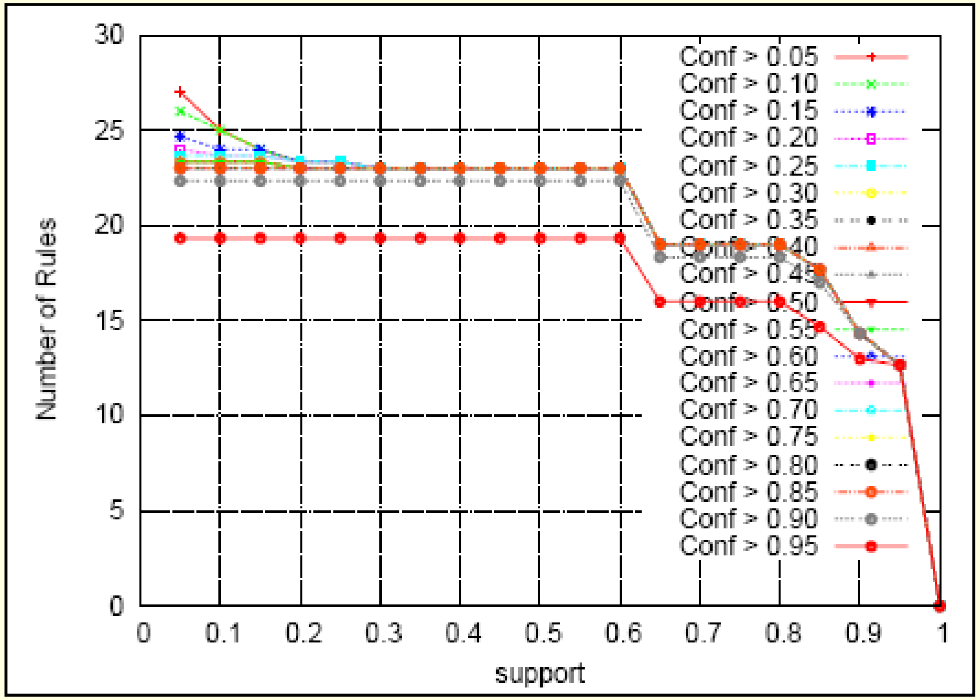
Second Experiment: Continuous Data

MAXIMUM INTERVAL SIZE = 0.9

Antecedent grouping



Consequent grouping



- **As the maximum size of the interval increases both configurations can learn rules with more support and confidence**

- **The number of rules in the consequent grouping reduces considerably, since subsumption applies and subsumes rules in favor of the most general ones**

Outline

1. Mining Association Rules
2. Description of CSar
3. Experimental Methodology
4. Results
5. Conclusions and Further Work

Conclusions

- Michigan-style LCS can extract association rules
  - Similar number of rules with high support and confidence to this of Apriori
  - Besides, CSar can extract quantitative association rules
- Niching is crucial to evolve different association rules
  - Antecedent grouping yield more rules when support and confidence are low than consequent grouping
  - Both groupings resulted in a similar number of rules with high support and confidence
- But still much needs to be analyzed

Further Work

- Compare CSar to other quantitative association rule miners
- Studies on the population size required to sustain associations
- Add fuzzy representation
  - Deal more effectively with the problem of maximum interval size
- Mining changing environments
  - Can CSar adapt quickly to *association changes*?
  - Can we *mine large data streams* with fixed population size and *get the most relevant associations*

First Approach toward On-line Evolution of Association Rules with Learning Classifier Systems

Albert Orriols-Puig[1]

Jorge Casillas[2]

Ester Bernadó-Mansilla[1]

[1]Research Group in Intelligent Systems
Enginyeria i Arquitectura La Salle, Ramon Llull University

[2]Dept. Computer Science and Artificial Intelligence
University of Granada