UNIVERSITÀ DI PISA

Dottorato di Ricerca in Ingegneria dell'Informazione

Data Mining and Soft Computing

Francisco Herrera

Research Group on Soft Computing and Information Intelligent Systems (SCI2S)
Dept. of Computer Science and A.I.
University of Granada, Spain

Email: herrera@decsai.ugr.es http://sci2s.ugr.es
http://decsai.ugr.es/~herrera

DECSAI
Universidad de Granada

Summary

1. Introduction to Data Mining and Knowledge Discovery
2. Data Preparation
3. Introduction to Prediction, Classification, Clustering and Association
4. Data Mining - From the Top 10 Algorithms to the New Challenges
5. Introduction to Soft Computing. Focusing our attention in Fuzzy Logic and Evolutionary Computation
6. Soft Computing Techniques in Data Mining: Fuzzy Data Mining and Knowledge Extraction based on Evolutionary Learning
7. Genetic Fuzzy Systems: State of the Art and New Trends
8. Some Advanced Topics I: Classification with Imbalanced Data Sets
9. Some Advanced Topics II: Subgroup Discovery
10. Some advanced Topics III: Data Complexity
11. Final talk: How must I Do my Experimental Study? Design of Experiments in Data Mining/Computational Intelligence. Using Non-parametric Tests. Some Cases of Study.

Outline

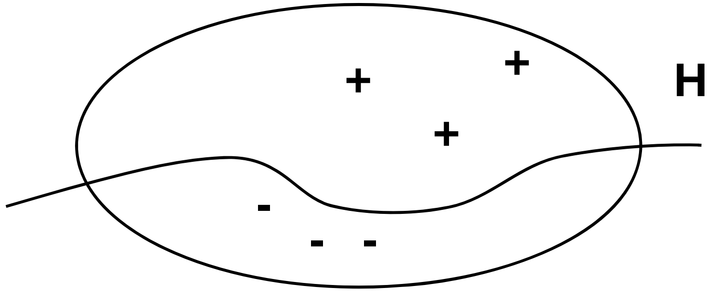
- Introduction
- Subgroup discovery
- Evaluation measures
- Data preprocessing and subgroup discovery
- A case of study: Fuzzy subgroup extraction in a marketing problem
- Concluding Remarks

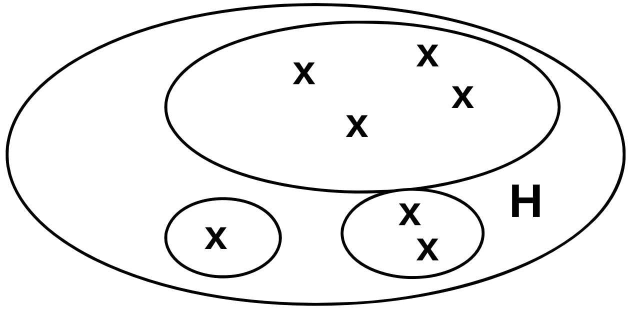
Outline

- Introduction
- Subgroup discovery
- Evaluation measures
- Data preprocessing and subgroup discovery
- A case of study: Fuzzy subgroup extraction in a marketing problem
- Concluding Remarks

Types of DM tasks

- **Predictive DM:**
  - Classification (learning of rulesets, decision trees, ...)
  - Prediction and estimation (regression)
  - Predictive relational DM (RDM, ILP)
- **Descriptive DM:**
  - description and summarization
  - dependency analysis (association rule learning)
  - discovery of properties and constraints
  - segmentation (clustering)
  - subgroup discovery
- **Text, Web and image analysis**





Predictive vs. descriptive induction

- **Predictive induction:** Inducing classifiers for solving classification and prediction tasks,
  - Classification rule learning, Decision tree learning, ...
  - Bayesian classifier, ANN, SVM, ...
  - Data analysis through hypothesis generation and testing
- **Descriptive induction:** Discovering interesting regularities in the data, uncovering patterns, ... for solving KDD tasks
  - Symbolic clustering, Association rule learning, Subgroup discovery, ...
  - Exploratory data analysis

Predictive vs. descriptive induction: A rule learning perspective

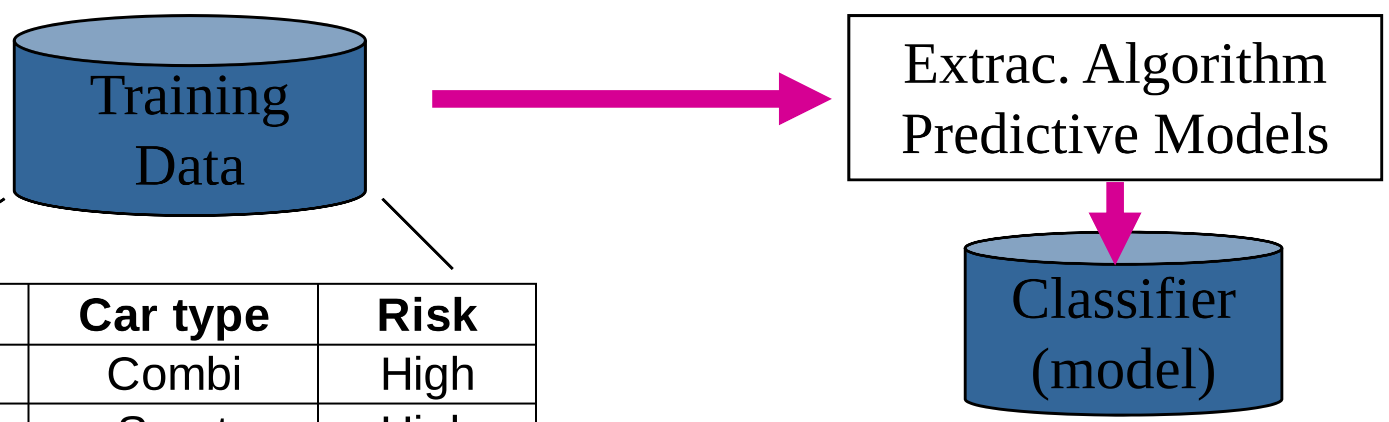
- **Predictive induction:** Induces **rulesets** acting as classifiers for solving classification and prediction tasks
- **Descriptive induction:** Discovers **individual rules** describing interesting regularities in the data

- **Therefore:** Different goals, different heuristics, different evaluation criteria

Predictive vs. descriptive induction: A rule learning perspective

•Prediction Models: Applied for inductive prediction and composed of rule sets used for classification.

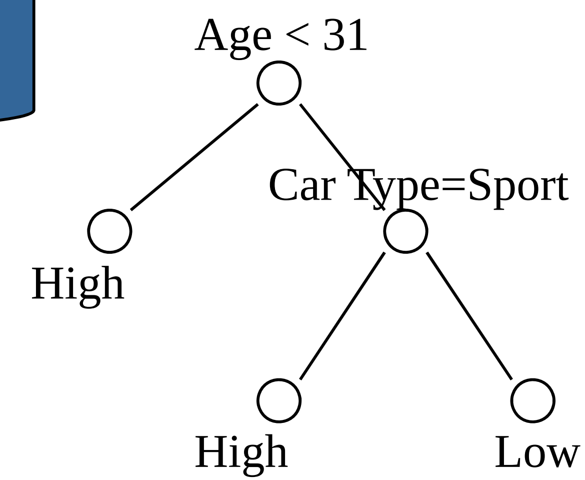
Kweku-Muata, Osei-Bryson, Evaluation of decision trees: a multicriteria approach. Computers and Operations Research, 31, MIT Press, 1993 -1945, 2004.



IND, S-Plus Trees, C4.5, CN2, FACT, QUEST, CART, OC1, LMDT, CAL5, T1…

Age	Car type	Risk
20	Combi	High
18	Sport	High
40	Sport	High
50	Familiar	Low
35	Minivan	Low
30	Combi	High
32	Familiar	Low
40	Combi	Low

if Age < 31
or Car Type = Sport
then Risk = High

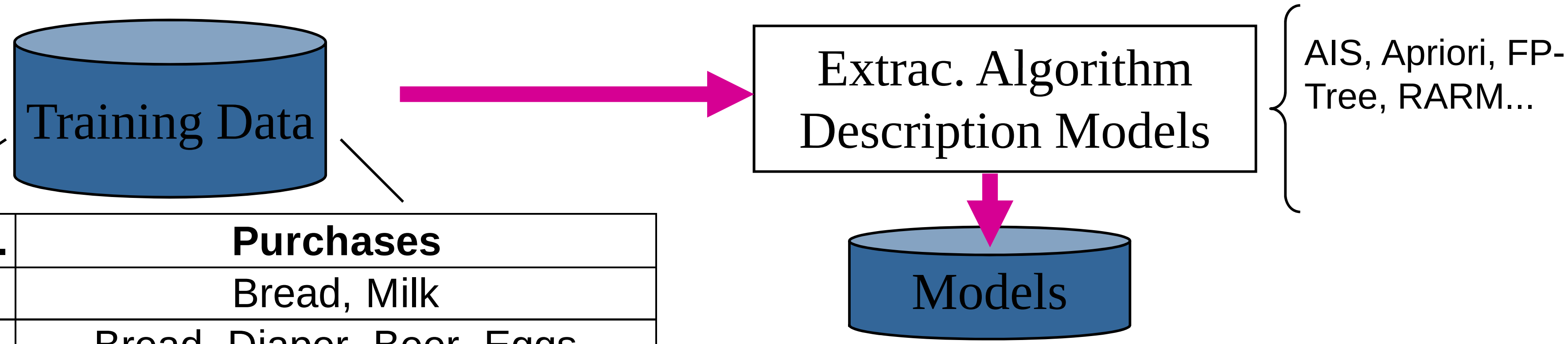


Quality Measures: Precision, Interpretability

Predictive vs. descriptive induction: A rule learning perspective

- Description Models: Applied for descriptive induction by searching for rules that define interesting patterns in data.

N. Lavrac, B.Kavsec, P. Flach, L. Todorovski, Subgroup Discovery with CN2-SD, Journal of Machine Learning Research, 5, 153-188, 2004.



Id. Trans.	Purchases
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Soft Drink
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Soft Drink

{Diaper} *→ {Beer},
{Milk, Bread} → {Eggs, Soft Drink},
{Beer, Bread} → {Milk}

* Implication means simultaneity, not causality

Quality measures: Confidence, Support

Supervised vs. unsupervised learning: A rule learning perspective

- **Supervised learning:** Rules are induced from labeled instances (training examples with class assignment) - usually used in **predictive induction**
- **Unsupervised learning:** Rules are induced from unlabeled instances (training examples with no class assignment) - usually used in **descriptive induction**
- **Exception: Subgroup discovery**

  Discovers **individual rules** describing interesting regularities in the data induced from **labeled** examples

Subgroups vs. classifiers

- Classifiers:
  - Classification rules aim at pure subgroups
  - A set of rules forms a domain model
- Subgroups:
  - Rules describing subgroups aim at significantly higher proportion of positives
  - Each rule is an independent chunk of knowledge

- Link: SD can be viewed as a form of cost-sensitive classification

Outline

- Introduction
- Subgroup discovery
- Evaluation measures
- Data preprocessing and subgroup discovery
- A case of study: Fuzzy subgroup extraction in a marketing problem
- Concluding Remarks

Subgroup Discovery

W. Klösgen , 1996:

"Given a population of individuals and a property of those individuals we are interested in, find population subgroups that are statistically 'most interesting', e.g., are as large as possible and have the most unusual statistical charasteristics with respect to the property of interest".

W. Klösgen, Explora: A multipattern and multistrategy discovery assistant, Advance in Knowledge Discovery and Data Mining, MIT Press, 249-271, 1996.

Subgroup Discovery

Task:

- Find subgroups of members of a population that exhibit interesting deviations from overall population behavior

Definition:

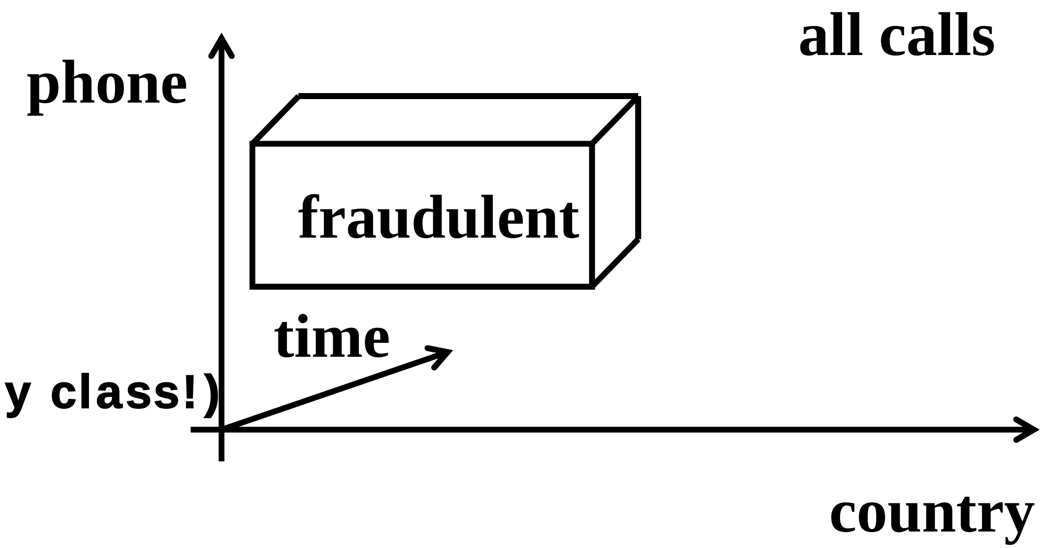
- A local pattern is interesting if it exhibits properties that deviate significantly from the properties that would be expected based on some prior knowledge.

Subgroup Discovery

Example: Fraud Detection

- Assume 100 % of all calls made to Australia from a mobile phone at night are fraudulent (total of 0.01% of all calls)
- but fraudulency does not otherwise depend on country, phone or time

not fraudulent



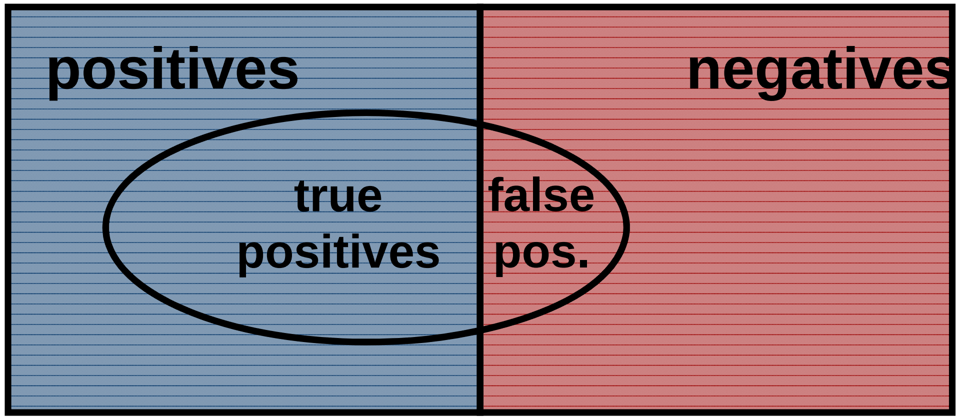
(Simply always answers majority class!)

Global model:
decision tree
Constructed with C4.5

local model:
Subgroup discovery

Subgroups vs. classifiers

- Classification rules aim at pure subgroups
- Subgroups aim at significantly higher (or different) proportion of positives
  - essentially the same as cost-sensitive classification
  - instead of *FNcost* we have *TPprofit*



Subgroup Discovery

Each *rule returns a prob*ability distribution, instead of class distribution in terms of the number of examples covered. Using this voting scheme the subgroups covering a small number of examples are not so heavily penalized when classifying a new example.

Weighted Relative Accuracy

WRAcc(Class←Condition)
= p(Condition)[p(Class | Condition) – p(Class)]
~ TPrate – FPrate

$$WRAcc(Cond \rightarrow Class) = \frac{n'(Cond)}{N'} \cdot \left(\frac{n'(Cond, Class)}{n'(Cond)} - \frac{n'(Class)}{N'} \right)$$

Subgroup Discovery: Example Apriori-SD

Fig. 1. Pseudocode of Apriori-SD algorithm

1. algorithm $APRIORI - SD(Examples, Classes, minSup, minConf, k)$
2. Ruleset= $APRIORI - C(Examples, Classes, minSup, minConf)$
   set all example weights of Examples to 1)
3. Majority= the majority class in Examples
4. Resultset= {}
5. Repeat
   6. BestRule= rule with the highest weighted relative accuracy
      in Ruleset.
   7. Resultset= Resultset $\bigcup$ BestRule
   8. Ruleset= Ruleset \ decrease the weights of examples covered
      by BestRule remove from Examples the examples covered more
      than k-times
9. until Examples={} or Ruleset={}
10. return Resultset= Resultset $\bigcup$ true $\rightarrow$ Majority

parameter k determines the threshold for covered example elimination in rule post-processing ensuring the convergence of the algorithm

Subgroup Discovery: Example Apriori-SD

Post-process: Rule subset selection by a weighted covering approach:

Take the best rule w.r.t. WRAcc
Decrease the weights of covered examples
Reorder the remaining rules and repeat until stopping criterion is satisfied
- **significance threshold**
- **WRAcc threshold**

CN2-SD: Adapting CN2 Rule Learning to Subgroup Discovery

- Weighted covering algorithm
- Weighted relative accuracy (WRAcc) search heuristics, with added example weights
- Probabilistic classification
- Evaluation with different interestingness measures

CN2

- Procedure CN2Unsorted(allExamples,Classes)
  - *RuleSet* ← { }
  - *For each Class in Classes*
    - *Generate rules with CN2ForOneClass (allExamples,Class)*
    - *Add rules to RuleSet*
  - Return RuleSet

- Procedure CN2ForOneClass(Examples,Class)
  - *Rules* ← { }
  - *Repeat*
    - *bestCondition* ←FindBestCondition*(Examples,Class)*
    - *If (bestCondition is not null) Then*
      - Add Rule 'If bestCondition then Class' to Rules and remove from Examples all the examples of the class 'Class' that are covered by bestCondition.
  - Until bestCondition is null
  - Return Rules

[P. Clark and T. Niblett, "The cn2 induction algorithm", *Machine Learning*, vol. 3, no. 4, pp. 261–283, Mar. 1989.

CN2-SD: CN2 Adaptations

- General-to-specific search (beam search) for best rules
- Rule quality measure:
  - CN2: Laplace: Acc(Class $\leftarrow$ Cond) =

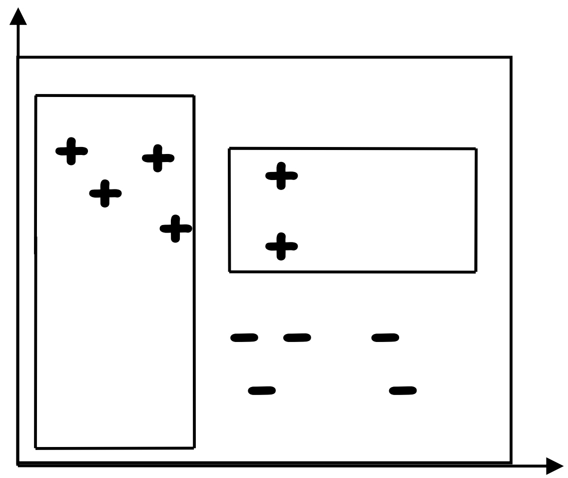
    = p(Class|Cond) = $(n_c+1)/(n_{rule}+k)$
  - CN2-SD: Weighted Relative Accuracy

    WRAcc(Class $\leftarrow$ Cond) =

    p(Cond) (p(Class|Cond) - p(Class))
- Weighted covering approach (example weights)
- Significance testing (likelihood ratio statistics)
- Output: Unordered rule sets (probabilistic classification)

CN2-SD: Weighted Covering

- Standard covering approach:
  covered examples are deleted from current training set
- Weighted covering approach:
  - weights assigned to examples
  - covered pos. examples are re-weighted: in all covering loop iterations, store count i how many times (with how many rules induced so far) a pos. example has been covered: w(e,i), w(e,0)=1



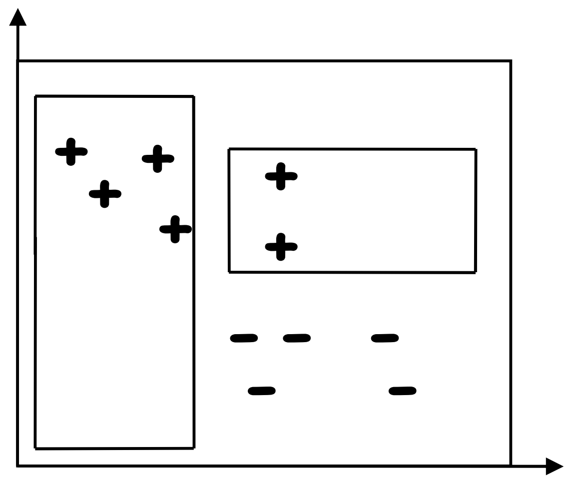
CN2-SD: Weighted Covering

- **Additive weights:** `w(e,i) = 1/(i+1)`

  `w(e,i)` – pos. example e being covered `i` times
- **Multiplicative weights:** `w(e,i) = gamma`^{i}, **`0<gamma<1`**

  note: **`gamma = 1`** ➔ find the same (first) rule again and again

  `gamma = 0` ➔ behaves as standard CN2



CN2-SD: Weighted WRAcc Search Heuristic

- **Weighted relative accuracy (WRAcc) search heuristics, with added example weights**

$$WRAcc(Cl \leftarrow Cond) = p(Cond)\ (p(Cl|Cond) - p(Cl))$$

increased coverage, decreased # of rules, approx. equal accuracy

CN2-SD: Weighted WRAcc Search Heuristic

- In WRAcc computation, probabilities are estimated with relative frequencies, adapt:

$$WRAcc(Cl \leftarrow Cond) = p(Cond)\ (p(Cl|Cond) - p(Cl)) =$$
$$n'(Cond)/N'\ (\ n'(Cl.Cond)/n'(Cond) - n'(Cl)/N')$$

  - N' : sum of weights of examples
  - n'(Cond) : sum of weights of all covered examples
  - n'(Cl.Cond) : sum of weights of all correctly covered examples

Probabilistic classification

- **A simplified example:**

`class=bird ← legs=2 & feathers=yes`
`[13,0]`

`class=elephant ← size=large & flies=no`
`[2,10]`

`class=bird ← beak=yes[20,0]`

[35,10]

Two-legged, feathered, large, non-flying animal with a beak?

bird !

Subgroup Discovery

Historical revision:

EXPLORA: All the learning process is performed by keeping all the information in only one table.

W. Klösgen, Explora: A multipattern and multistrategy discovery assistant, Advance in Knowledge Discovery and Data Mining, MIT Press, 249-271, 1996.

MIDOS: This algorithm extends the process to multirelational data bases.

S. Wrobel, An algorithm for multi-relational discovery of subgroups, Proceedigs of the 4th European Conference on Principles of Data Mining and Knowledge Discovery, Springer, 78 - 87, 1997.

EXPLORA and MIDOS use decision trees. Lately, separate-and-conquer models have been used, different than those used with trees, that allow to include non null intersections among rules.

Subgroup Discovery

Historical revision:

CN2-SD: Adaptation of CN2 by modifying the covering algorithm, the heuristic search, the probabilistic instance selection and the evaluation measures.

N. Lavrac, B.Kavsec, P. Flach, L. Todorovski, Subgroup Discovery with CN2-SD, Journal of Machine Learning Research, 5, 153-188, 2004.

Apriori-SD: Adaptation of Apriori-C by using the weighted relative success as the measure to assess the quality of the rules.

B. Kavsek, N.Lavrac, V. Jovanoski, Apriori-sd: Adapating association rule learning to subgroup discovery, Proceedings of the 5th International Symposium on Intelligent Data Analysis, Springer, 230 -241, 2003.

Kavsek, B., Lavrac, N., APRIORI-SD: Adapting association rule learning to subgroup discovery. Applied Artificial Intelligence, 20(7) (2006) 543-583.

Subgroup Discovery

Historical revision:

RSD: Adaptation of the relational rule learning to the problem and addition of weights to the example data.

Zelezny, F., Lavrac, N. (2006).Propositionalization-based relational subgroup discovery with RSD. Machine Learning, 62, 33-63.

SD-MAP: A quite efficient model based on rules, but it needs of a discretization on the domain of the variables.

Atzmueller, M., Puppe, F.. SD-Map - A Fast Algorithm for Exhaustive Subgroup Discovery. Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases, LNAI 4231, 6-13, 2006.

SDIGA: A model based on genetic algorithms to extract fuzzy rules in SD.

M.J. del Jesus, P. González, F. Herrera, M. Mesonero, Evolutionary Fuzzy Rule Induction Process for Subgroup Discovery: A Case Study in Marketing. *IEEE Transactions on Fuzzy Systems* 15:4 (2007) 578-592

Outline

- Introduction
- Subgroup discovery
- Evaluation measures
- Data preprocessing and subgroup discovery
- A case of study: Fuzzy subgroup extraction in a marketing problem
- Concluding Remarks

Evaluation Measures

Coverage:

$$COV = \frac{1}{n_R}\sum_{i=1}^{n_R} Cov(R_i) \qquad Cov(R_i) = p(Cond) = \frac{n(Cond)}{N}$$

Completeness:

$$Comp(R_i) = Comp(Cond_i \rightarrow Clase) = \frac{n(Cond_i, Clase)}{N}$$

$$COMP = \frac{1}{N}\sum_{Clase_j} n(Clase_j \underset{Cond_i \rightarrow Clase_j}{\vee} Cond_i)$$

Confidence:

$$Conf(R_i) = \frac{p(Class|Cond)}{p(Cond)} = \frac{n(Class, Cond)}{n(Cond)} \qquad CONF = \frac{1}{n_R}\sum_{i=1}^{n_R} Conf(R_i)$$

Evaluation Measures

Unusualness:

$$WRACC = \frac{1}{n_R}\sum_{i=1}^{n_R} WRAcc(R_i)$$

$$WRAcc(R_i) = p(Cond) \cdot$$

$$(p(Class|Cond) - p(Class)) =$$

$$\frac{n(Cond)}{N} \cdot \left(\frac{n(Class,Cond)}{n(Cond)} - \frac{n(Class)}{N}\right)$$

Significance:

$$Sig(R_i) = 2 \cdot \sum_j n(Class_j, Cond) \cdot log\frac{n(Class_j, Cond)}{n(Class_j)}$$

$$SIG = \frac{1}{n_R}\sum_{i=1}^{n_R} Sig(R_i)$$

Outline

- Introduction
- Subgroup discovery
- Evaluation measures
- Data preprocessing and subgroup discovery
- A case of study: Fuzzy subgroup extraction in a marketing problem
- Concluding Remarks

Data preprocessing and subgroup discovery

J.R. Cano, S. García, F. Herrera, Subgroup Discovery in Large Size Data Sets Preprocessed Using Stratified Instance Selection for Increasing the Presence of Minority Classes. ***Pattern Recognition Letters 29 (2008) 2156-2164, doi:10.1016/j.patrec.2008.08.001*****.**

J.R. Cano, F. Herrera, M. Lozano, S. García, Making CN2-SD Subgroup Discovery Algorithm scalable to Large Size Data Sets using Instance Selection. ***Expert Systems with Applications 35 (2008) 1949-1965, doi:10.1016/j.eswa.2007.08.083*****.**

Outline

- Introduction
- Subgroup discovery
- Evaluation measures
- Data preprocessing and subgroup discovery
- A case of study: Fuzzy subgroup extraction in a marketing problem
- Concluding Remarks

Case of Study

Fuzzy subgroup extraction in a marketing problem

M.J. del Jesus, P. González, F. Herrera, M. Mesonero, **Evolutionary Fuzzy Rule Induction Process for Subgroup Discovery: A Case Study in Marketing.** *IEEE Transactions on Fuzzy Systems* 15:4 (2007) 578-592

Motivation

- Trade fairs are a basic instrument in company marketing policies, especially in Industrial Marketing
  - They facilitate the attainment of commercial objectives
  - But also require a elevated investment and need some planning

- The available data are obtained in the Machinery and Tools biennial (Bilbao, March 2002)
  - 228 exhibitors
  - 104 variables (continuous and categorical)
  - Stand efficiency rated depending on the achievement of objectives

Motivation

Objective

Determine the relationship between the variables which describe aspects of trade fairs and the variable which measures the achievement of the objectives planned by the exhibitors

Solution

Evolutionary model for the descriptive induction of rules which describe subgroups, including a genetic algorithm in an iterative model to extract a variable number of fuzzy or crisp rules

An evolutionary approach to obtain descriptive fuzzy rules

- Key features of the proposal:
  - Descriptive rule induction algorithm; the extracted rules allow the expression of relationships between variables
  - The genetic representation of the solutions of a Genetic Algorithm is the most determining aspect of the characteristics of any proposal
    - Approaches “Chromosome = Rule” or “Chromosome = Set of rules”
    - The proposal follows the Iterative Rule Learning (IRL) approach, a kind of “Chromosome = Rule”
  - The consequent of the rules is prefixed to assure the extraction of rules for all the values of the target variable
  - The rules extracted are fuzzy rules to express the extracted knowledge in an understandable way, close to the expert

An evolutionary approach to obtain descriptive fuzzy rules

- Two components:
  - Iterative model of extraction of fuzzy rules
  - A hybrid Genetic Algorithm for the extraction of one fuzzy rule

An evolutionary approach to obtain descriptive fuzzy rules

```
START
    RuleSet ← Ø
    REPEAT
        Execute the GA obtaining rule R
        Post-processing of rule R (Local Search)
        RuleSet ← RuleSet + R
        Modify the set of examples
     WHILE confidence(R) ≥ minimum confidence and
           R represents new examples
END
```

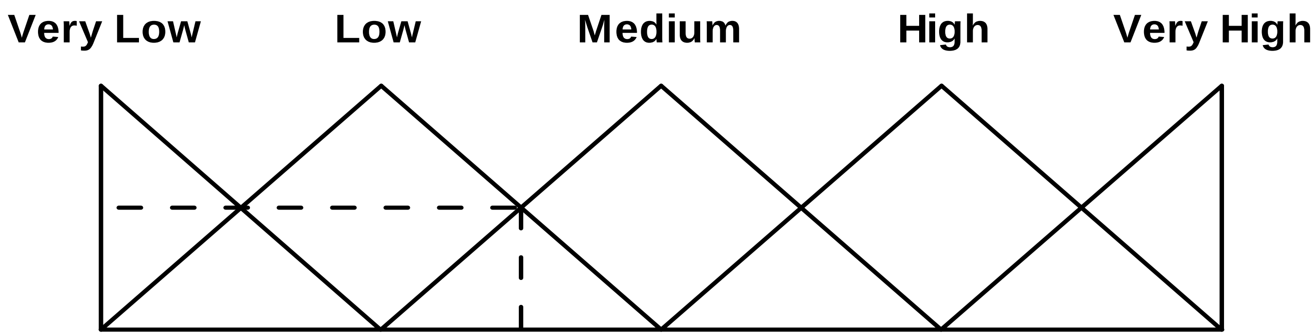
An evolutionary approach to obtain descriptive fuzzy rules

Elements of the proposal:

1. Chromosome representation
2. Fitness function
3. Genetic operators
4. Post-processing phase

An evolutionary approach to obtain descriptive fuzzy rules

Chromosome representation



IF *zone* is *centre* and *sector* is *accessories* and ... and *Bar* is *Yes*
THEN Efficiency is high

An evolutionary approach to obtain descriptive fuzzy rules

Fitness function

$$fitness\,(c) = \frac{\omega_1 \cdot \text{Support}\,(c) + \omega_2 \cdot \text{Interest}\,(c) + \omega_3 \cdot \text{Confidence}\,(c)}{\omega_1 + \omega_2 + \omega_3}$$

$$Confidence = \frac{\sum_{\forall examples \in class} \text{membership\_anteceden t\_subspace}}{\sum_{\forall examples} \text{membership\_anteceden t\_subspace}}$$

$$Support = \frac{\text{New\_covere d\_class\_ex amples}}{\text{Previusly\_ uncovered\_ class\_exam ples}}$$

$$Interest = 1 - \left(\frac{\sum_{i=1}^{n} Gain(A_i)}{n \cdot \log_2\left(\left|dom(G_k)\right|\right)} \right)$$

An evolutionary approach to obtain descriptive fuzzy rules

Post-processing

```
START
   Best_ Rule ← R; Best_support ← support(R);
   Better ← True
   REPEAT WHILE Better
      Better ← False
      FOR (i=1 to gene_number)
         R'i = R without considering variable i
         IF (support (R'i) >= support (R))
            Better ← True
            IF (support (R'i) > Best_support)
               Best_support ← support (R'i)
               Best_Rule ← R'i
      END FOR
   END WHILE
   IF (Better AND confidence(Best_Rule) >= min_conf)
      Return Best_Rule
   ELSE
        Return R
END
```

Experimentation

- Market dataset
  - Marketing experts have made a selection of variables, reducing the original set of 104 variables to a subset of 18 variables
  - The evolutionary rule induction algorithm has been applied to this subset of variables
- Parameters of the experimentation:
  - 5 runs for each value of the target variable
  - 100 individuals in the population of the Genetic Algorithm
  - 5000 maximum evaluations of individuals in each Genetic Algorithm run
  - Fitness function weights:
    - Support: 0.4
    - Confidence: 0.3
    - Interest: 0.3
  - Minimum confidence value: 0.6

Variables

Name	Description
Efficiency	Global efficiency for the stands
Zone	Geographic zone of the company
Sector	Sector to which the exhibitor belongs
Fair utility	Utility provided by the fairs
Annual fair number	Number of fairs participating annually as exhibitor
Written objectives	Existence of objectives for the BIEMH in writing
Previous promotion	Accomplishment of previous promotion to the fair
Promotion listings	Listings of clients to inform of the presence in the fair
...	

Experimentation

Class	Rule	Support	Confid.	Interest
Low	1	10,526	100,000	61,282
	2	13,158	100,000	60,663
	3	18,421	100,000	58,341
	4	7,895	100,000	58,248
	5	7,895	100,000	59,971
	6	5,263	100,000	57,806
	7	5,263	100,000	53,024

Experimentation

Class	Rule	Support	Confid.	Interest
Medium	1	10,811	100,000	59,112
	2	10,135	100,000	55,906
	3	6,081	100,000	58,062
	4	3,378	100,000	61,805
	5	6,081	100,000	59,567
	6	3,378	100,000	57,870
	7	4,730	100,000	59,923
	8	3,378	100,000	60,617
	9	2,027	100,000	60,929
	10	3,378	100,000	59,232
	11	95,946	64,840	62,340
	12	0,676	100,000	60,977

Experimentation

Class	Rule	Support	Confid.	Interest
High	1	4,762	100,000	62,110
	2	9,524	100,000	59,904
	3	11,905	100,000	59,045
	4	4,762	100,000	59,845
	5	7,143	100,000	60,580

Experimentation

- The algorithm induces a set of rules with a high confidence and interest level
  - The variables which intervene in the rules are variables with low information gain value, more surprising to the user and carrying more information
- The rule support, except for some rules, is low
  - The model induces, for this problem, specific rules which represent a small number of examples
- The knowledge discovered for each one of the target variable values is understandable by the user due to the use of Fuzzy Logic, and the small number of rules and conditions in the rule antecedents

Experimentation

- Analysis of the results from the point of view of marketing:
  - The companies obtain better results (high efficiency) if they write the objectives, present authentic innovations in the fair and come from the East zone (Catalonia and Levant)
  - The exhibitors obtain worse results if they are manufacturers of the North zone, belonging to the sectors of Deformation and Starting, which had not written objectives and had not made any effort to plan the promotion campaign before the event

Experimentation: Rules induced for “low” efficiency

1	IF Sector = Starting+Deformation AND Written objectives = No AND Previous promotion = No THEN Efficiency = Low
2	IF Written objectives = No AND Importance of present clients contacts = Low AND Quality of contacts= High AND Stand at entrance = No AND Near of stairs= No THEN Efficiency = Low
3	IF Zone = North AND Sector = Starting+Deformation AND Written objectives = No AND Telephone calls = Yes AND New features = Product improvement AND Stand at entrance = No THEN Efficiency = Low
4	IF Importance of contacts = Low AND Quality of contacts= Low THEN Efficiency = Low
5	IF Zone = East AND Written objectives = No AND Existence of promotion listings = No AND Importance of operations after the fair = High AND Stand at entrance = No AND Near of stairs= No THEN Efficiency = Low
6	IF Zone = North AND Fairs utility = Low AND Importance of contacts = Medium AND New features = Product improvement THEN Efficiency = Low
7	IF Sector = Starting+Deformation AND Promotion campaign monitoring = No AND Importance of present clients contacts = High AND Machinery demonstrations type = Sporadic operation AND Stewardesses = Yes THEN Efficiency = Low

Experimentation:
Rules induced for “low” efficiency

1	IF Zone = North and Fairs utility = Low AND Visitors number importance = Medium AND Stand at entrance = Yes THEN Efficiency = Medium
2	IF Zone = North AND Quality of contacts= High AND Telephone calls = Yes AND New features = "Catalogue" THEN Efficiency = Medium
3	IF Sector = Rest AND Importance of operations after the fair = Medium AND New features = Product improvement THEN Efficiency = Medium
4	IF Sector = Starting+Deformation AND Number of annual fairs = More than 11 THEN Efficiency = Medium
5	IF Previous promotion = Yes AND Visitors number importance = Low AND Stand at entrance = Yes THEN Efficiency = Medium
6	IF Sector = Rest AND Importance of operations after the fair = Low AND Visitors number importance = High THEN Efficiency = Medium
7	IF Zone = North AND Sector = Starting+Deformation AND Fairs utility = Low AND Previous promotion = Yes AND Quality of contacts= Medium THEN Efficiency = Medium
8	IF Quality of contacts= Medium AND Stewardesses = Yes THEN Efficiency = Medium
9	IF Previous promotion = No AND Quality of contacts= High AND Stand at entrance = Yes THEN Efficiency = Medium
10	IF Sector = Rest AND Importance of operations after the fair = Low AND Quality of contacts= Medium THEN Efficiency = Medium
11	IF Number of annual fairs = Less than 11 THEN Efficiency = Medium
12	IF Number of annual fairs = More than 11 AND Quality of contacts= Medium THEN Efficiency = Medium

Experimentation: Rules induced for “low” efficiency

1	IF Written objectives = Yes AND Stewardesses = No AND Stand at entrance = Yes AND Near of stairs= Yes THEN Efficiency = High
2	IF Sector = Rest AND Number of annual fairs = More than 11 AND New features = Authentic newness THEN Efficiency = High
3	IF Zone = East AND Sector = Rest AND Fairs utility = High AND Importance of contacts quality = High AND New features = Authentic newness THEN Efficiency = High
4	IF Zone = East AND Sector = Rest AND Number of annual fairs = Less than 11 AND Existence of promotion listings = Yes AND Importance of operations after the fair = High AND Quality of contacts= Medium AND Stand at entrance = No THEN Efficiency = High
5	IF Fairs utility = High AND Written objectives = Yes AND New features = Authentic newness AND Stand at entrance = No AND Near of stairs= No THEN Efficiency = High

Comments

- Fuzzy Logic allows the user to incorporate directly linguistic knowledge into the data mining process, to mix this knowledge with non-linguistic information and to treat appropriately incomplete data or data with noise

- The experiment carried out with the model proposed has determined a simple set of rules which use few variables and therefore has a simple structure. The information extracted is comprehensible to and usable by the final user

Outline

- Introduction
- Subgroup discovery
- Evaluation measures
- Data preprocessing and subgroup discovery
- A case of study: Fuzzy subgroup extraction in a marketing problem
- Concluding Remarks

Concluding Remarks

Subgroup discovery is a task at the intersection of predictive and descriptive induction.

Predictive vs. descriptive induction: Summary

- **Predictive induction:** Induces **rulesets** acting as classifiers for solving classification and prediction tasks
  - Rules are induced from labeled instances
- **Descriptive induction:** Discovers **individual rules** describing interesting regularities in the data
  - Rules are induced from unlabeled instances
- **Exception: Subgroup discovery**

  Discovers **individual rules** describing interesting regularities in the data induced from **labeled** examples

Summary

1. Introduction to Data Mining and Knowledge Discovery
2. Data Preparation
3. Introduction to Prediction, Classification, Clustering and Association
4. Data Mining - From the Top 10 Algorithms to the New Challenges
5. Introduction to Soft Computing. Focusing our attention in Fuzzy Logic and Evolutionary Computation
6. Soft Computing Techniques in Data Mining: Fuzzy Data Mining and Knowledge Extraction based on Evolutionary Learning
7. Genetic Fuzzy Systems: State of the Art and New Trends
8. Some Advanced Topics I: Classification with Imbalanced Data Sets
9. Some Advanced Topics II: Subgroup Discovery
10. Some advanced Topics III: Data Complexity
11. Final talk: How must I Do my Experimental Study? Design of Experiments in Data Mining/Computational Intelligence. Using Non-parametric Tests. Some Cases of Study.