# Classification and Clustering in the Belief Function Framework

Zied Elouedi

LARODEC, Université de Tunis, ISG Tunis, Tunisia
zied.elouedi@gmx.fr

BFTA Spring School 20-24 May 2013
Carthage, Tunisia






# Outline

1 Introduction

# Outline


1. Introduction

2. Classification
   - Belief Decision Trees
   - Belief Rough Set Classifier

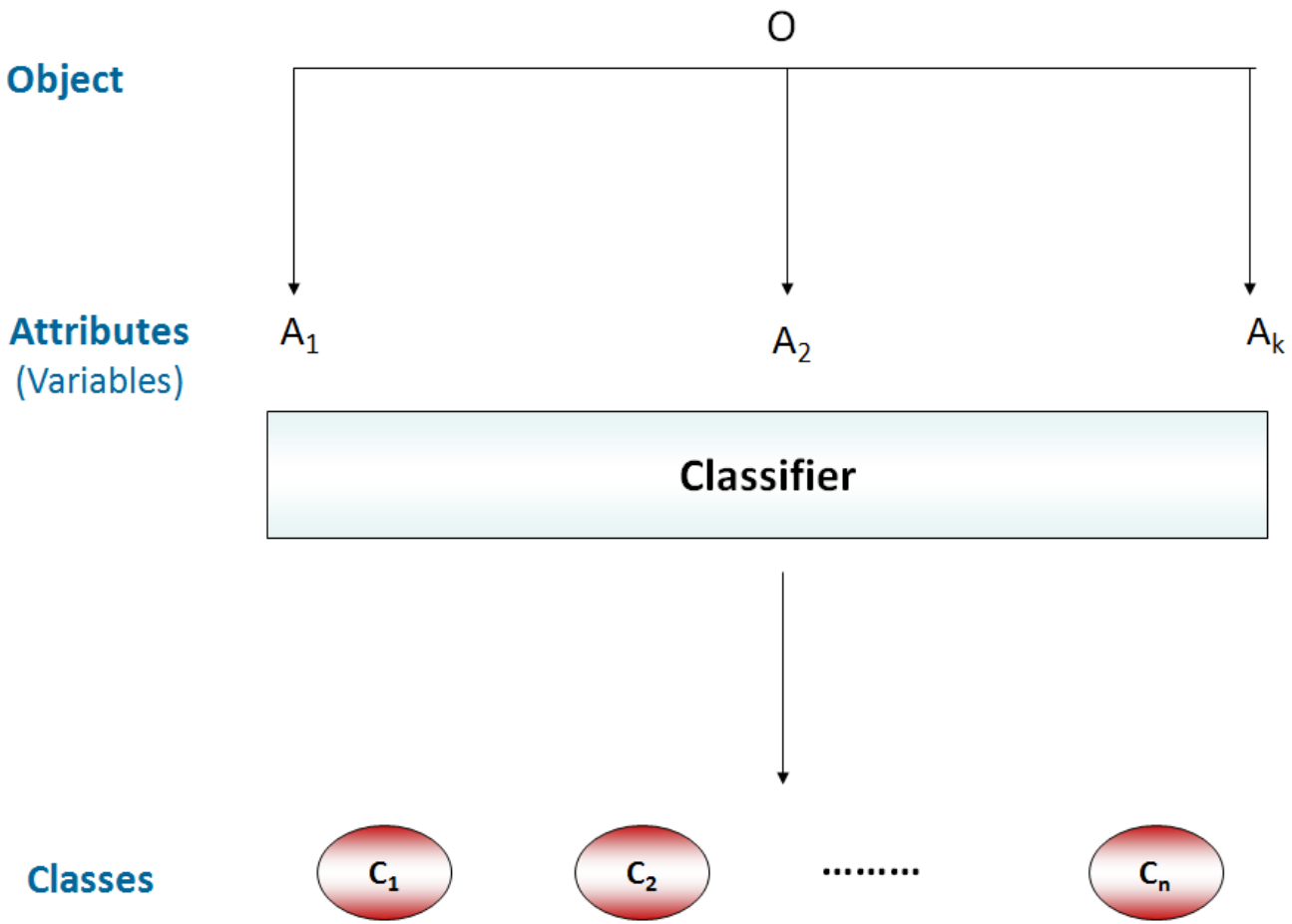
# Outline


1. Introduction
2. Classification
   - Belief Decision Trees
   - Belief Rough Set Classifier
3. Clustering
   - Belief K-Modes
   - Dynamic Belief K-Modes

# Outline


1. Introduction
2. Classification
   - Belief Decision Trees
   - Belief Rough Set Classifier
3. Clustering
   - Belief K-Modes
   - Dynamic Belief K-Modes
4. Conclusion

# Definition


O
Object
Attributes
(Variables)
$A_1$
$A_2$
$A_k$
Classifier
Classes
$C_1$
$C_2$
.....
$C_n$

# Supervised versus Unsupervised

## Supervised classification

Classes are known a priori.

# Supervised versus Unsupervised

## Supervised classification

Classes are known a priori.

- Decision trees,
- K-nearest neighbors,
- Neural networks,
- Naive Bayes,
- Rough set classifier,
- etc.

# Supervised versus Unsupervised

## Unsupervised classification: Clustering

Classes are unknown a priori.

## Unsupervised classification: Clustering

Classes are unknown a priori.

- K-means,
- K-modes,
- Dbscan,
- etc.

# Applications

- Credit risk assessment,

# Applications

- Credit risk assessment,
- Medical diagnosis,

# Applications

- Credit risk assessment,
- Medical diagnosis,
- Finance,

# Applications

- Credit risk assessment,
- Medical diagnosis,
- Finance,
- Marketing,

# Applications

- Credit risk assessment,
- Medical diagnosis,
- Finance,
- Marketing,
- Intrusion detection,

# Applications

- Credit risk assessment,
- Medical diagnosis,
- Finance,
- Marketing,
- Intrusion detection,
- Fraud detection,

# Applications

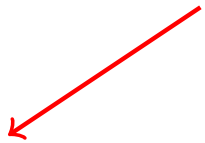
- Credit risk assessment,
- Medical diagnosis,
- Finance,
- Marketing,
- Intrusion detection,
- Fraud detection,
- etc,.

# Problem

Problems in handling uncertainty and imprecision

# Problem

Problems in handling uncertainty and imprecision



# Problem

Problems in handling uncertainty and imprecision



## Training phase

Uncertainty in parameters

## Classification phase

- Missing or uncertain attribue values
- Belongness to several classes or several clusters

# Problem

Problems in handling uncertainty and imprecision

## Training phase

Uncertainty in parameters

## Classification phase

- Missing or uncertain attribue values
- Belongness to several classes or several clusters

$\Rightarrow$ Use of the belief function theory.

# Why?

The belief function theory offers:

# Why?

The belief function theory offers:

- The ability to treat subjective and personal judgements.

# Why?

The belief function theory offers:

- The ability to treat subjective and personal judgements.
- The expression of partial beliefs.

# Why?

The belief function theory offers:

- The ability to treat subjective and personal judgements.
- The expression of partial beliefs.
- The representation of partial or total ignorance.

# Why?

The belief function theory offers:

- The ability to treat subjective and personal judgements.
- The expression of partial beliefs.
- The representation of partial or total ignorance.
- The combination of evidence from several information sources.

# Why?

The belief function theory offers:

- The ability to treat subjective and personal judgements.
- The expression of partial beliefs.
- The representation of partial or total ignorance.
- The combination of evidence from several information sources.
- The adaptation to reasoning based systems.

# *Belief Decision Trees*

# Decision trees

Decision trees are classification techniques used in datamining:

# Decision trees

Decision trees are classification techniques used in datamining:

- Capability to handle complex problems.

# Decision trees

Decision trees are classification techniques used in datamining:

- Capability to handle complex problems.
- Ability to express knowledge.

# Decision trees

Decision trees are classification techniques used in datamining:

- Capability to handle complex problems.
- Ability to express knowledge.
- Formalism easy to interpret.

# Decision trees

Decision trees are classification techniques used in datamining:

- Capability to handle complex problems.
- Ability to express knowledge.
- Formalism easy to interpret.
- Participation of experts in the production of rules.

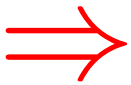
# Training set

| Income | Property | Unpaid-credit | Class |
|---|---|---|---|
| High | Greater | Yes | $C_1$ |
| Average | Less | No | $C_1$ |
| High | Greater | Yes | $C_1$ |
| Average | Greater | Yes | $C_2$ |
| Low | Less | Yes | $C_2$ |
| No | Less | Yes | $C_3$ |
| High | Greater | No | $C_1$ |
| Average | Less | Yes | $C_1$ |

C1: good clients.
C2: moderate clients.
C3: bad clients.

# Algorithms

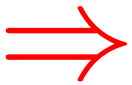
Top Down Induction of Decision Trees (TDIDT)



Divide to conquer (Descendant Induction)

# Algorithms

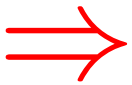
Top Down Induction of Decision Trees (TDIDT)

 Divide to conquer (Descendant Induction)

- ID3 (Quinlan, 1979),

# Algorithms

Top Down Induction of Decision Trees (TDIDT)

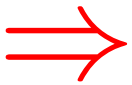


Divide to conquer (Descendant Induction)

- ID3 (Quinlan, 1979),
- CART (Breiman et al., 1984),

# Algorithms

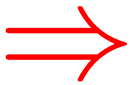
Top Down Induction of Decision Trees (TDIDT)

 Divide to conquer (Descendant Induction)

- ID3 (Quinlan, 1979),
- CART (Breiman et al., 1984),
- ASSISTANT (Bratko, 1984),

# Algorithms

Top Down Induction of Decision Trees (TDIDT)

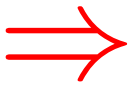


Divide to conquer (Descendant Induction)

- ID3 (Quinlan, 1979),
- CART (Breiman et al., 1984),
- ASSISTANT (Bratko, 1984),
- C4.5 (Quinlan, 1993),

# Algorithms

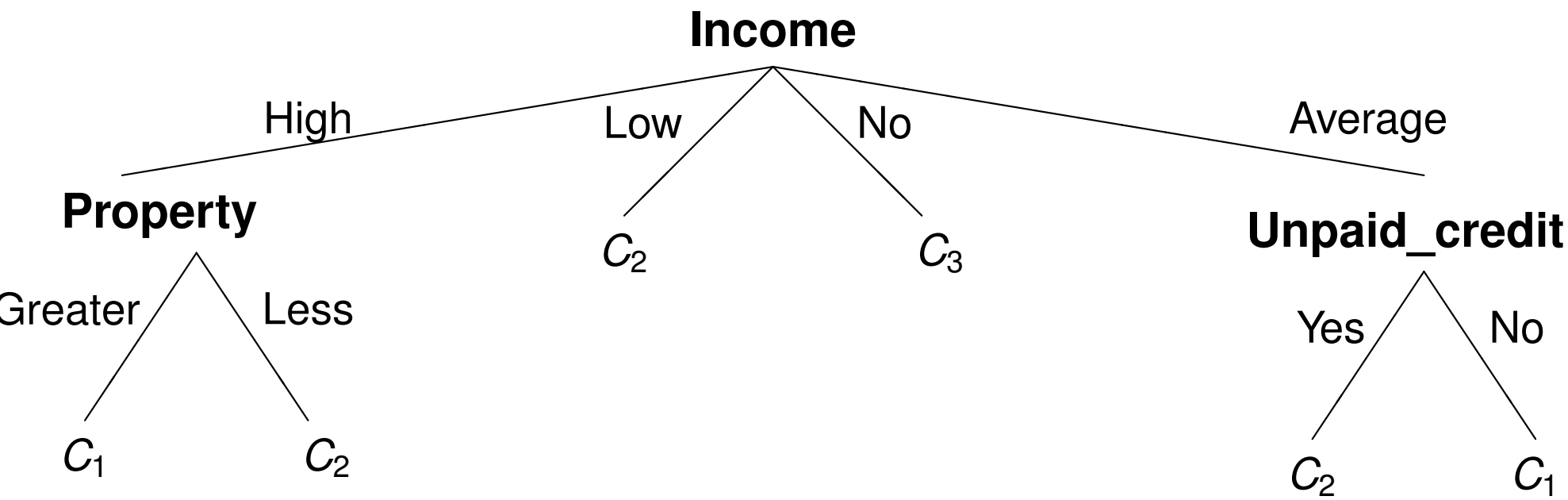
Top Down Induction of Decision Trees (TDIDT)



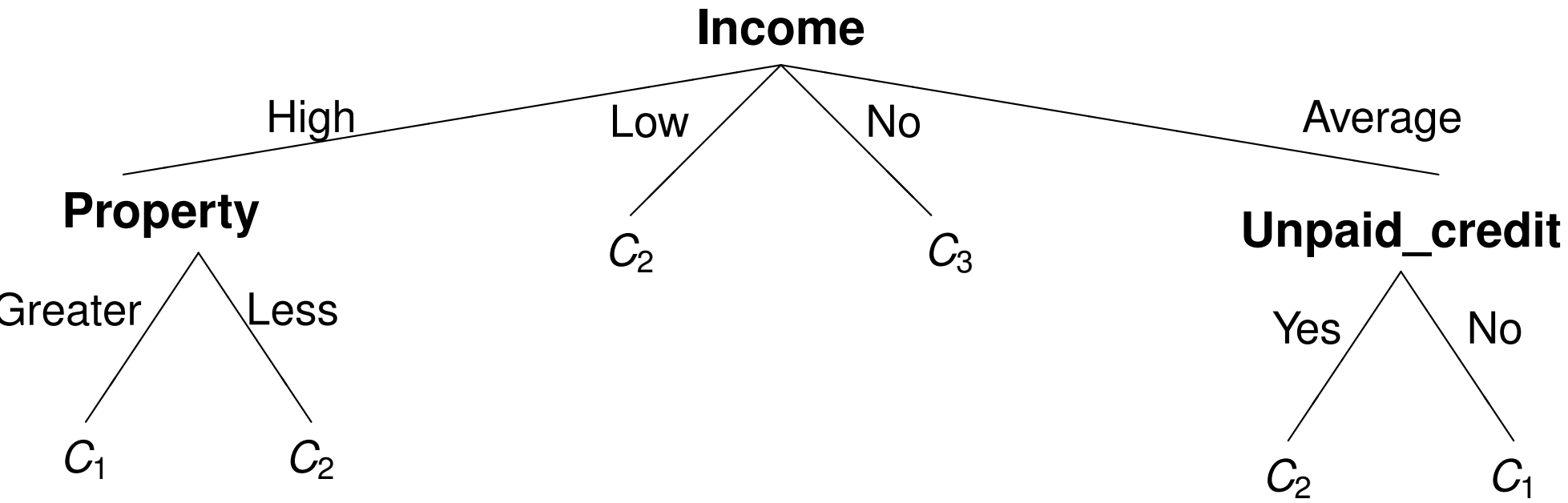
Divide to conquer (Descendant Induction)

- ID3 (Quinlan, 1979),
- CART (Breiman et al., 1984),
- ASSISTANT (Bratko, 1984),
- C4.5 (Quinlan, 1993),
- etc.

# Representation

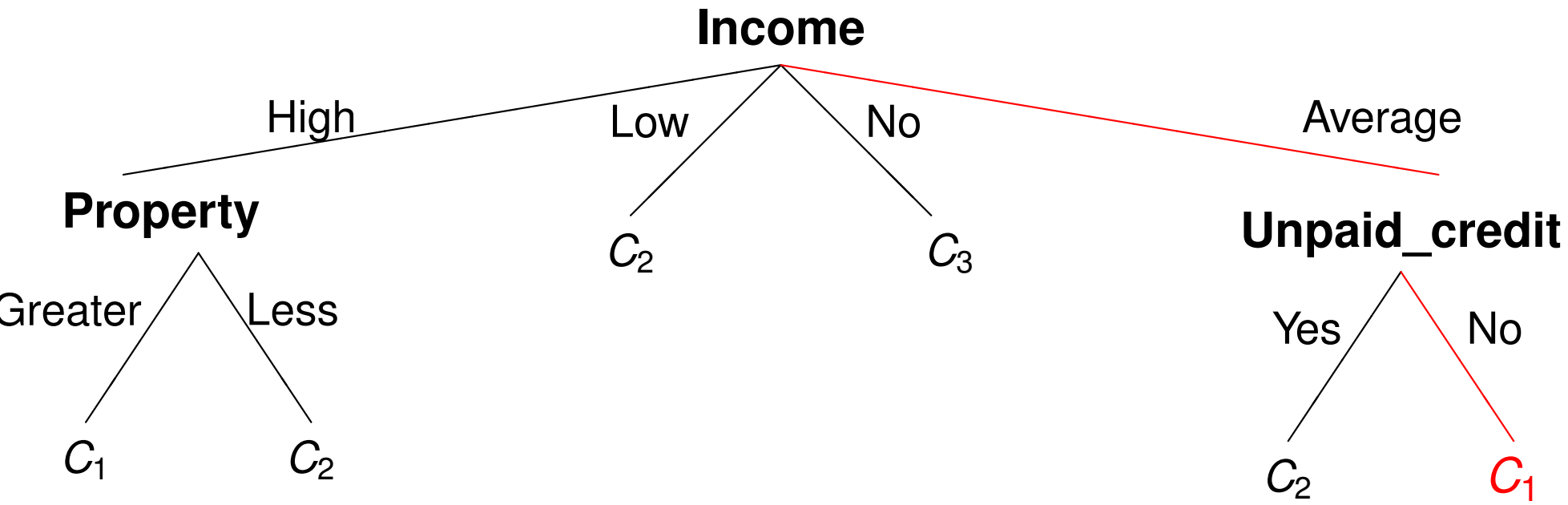


# How to classify?



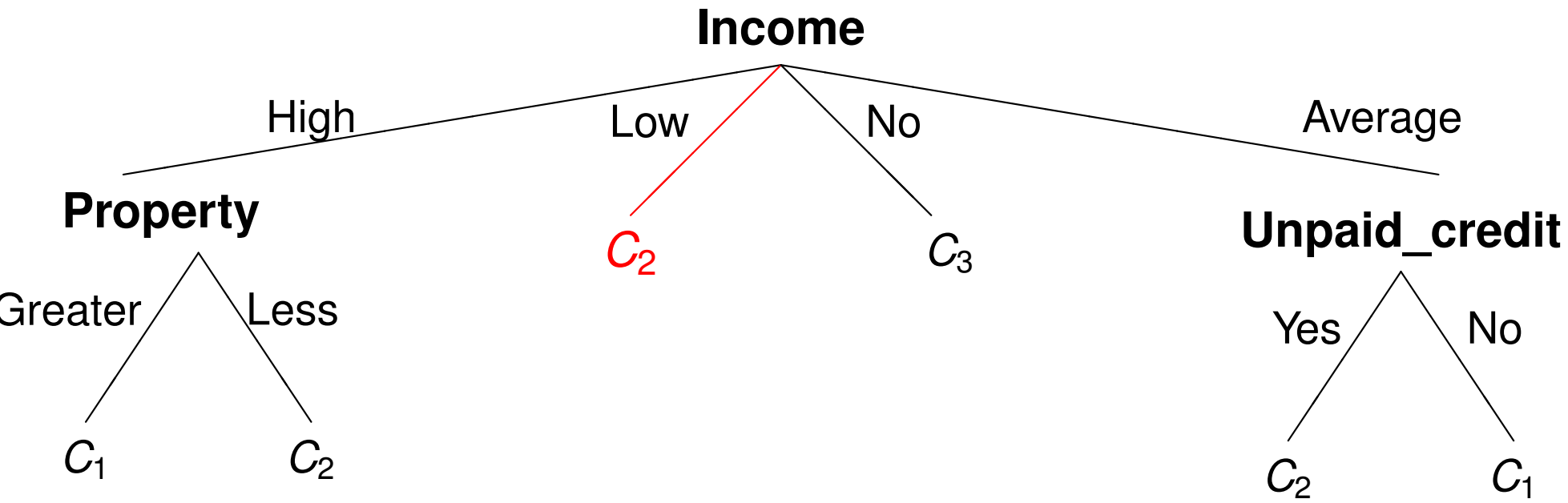

| | | | |
|---|---|---|---|
| Average | Greater | No | ? |
| Low | Greater | Yes | ? |

# How to classify?




| | | | |
|---|---|---|---|
| Average | Greater | No | $C_1$ |
| Low | Greater | Yes | ? |

# How to classify?




| | | | |
|---|---|---|---|
| Average | Greater | No | $C_1$ |
| Low | Greater | Yes | $C_2$ |

## Problems in handling uncertainty

## Problems in handling uncertainty

- Attribute values.
- Instances' classes.

# Belief decision trees

## Objective

Decision trees + Belief function theory

## Objective

Decision trees + Belief function theory

$\Rightarrow$ Belief Decision Trees (BDT)

## Objective

Decision trees + Belief function theory

$\Rightarrow$ Belief Decision Trees (BDT)

- Dealing with uncertainty in classification parameters.

# Training set

- The value of each attribute is known with certainty.
- Training instances' classes are uncertain.

# Training set

- The value of each attribute is known with certainty.
- Training instances' classes are uncertain.

$\Rightarrow$ bba for each instance

# Training set

- The value of each attribute is known with certainty.
- Training instances' classes are uncertain.

$\Rightarrow$ bba for each instance

$\Downarrow$

Opinions-beliefs about the actual value of the class for each training instance.

# Example

| Income | Property | Unpaid-credit | Class |
|---|---|---|---|
| High | Greater | Yes | $m^{\Theta}\{l_1\}$ |
| Average | Less | No | $m^{\Theta}\{l_2\}$ |
| High | Greater | Yes | $m^{\Theta}\{l_3\}$ |
| Average | Greater | Yes | $m^{\Theta}\{l_4\}$ |
| Low | Less | Yes | $m^{\Theta}\{l_5\}$ |
| No | Less | Yes | $m^{\Theta}\{l_6\}$ |
| High | Greater | No | $m^{\Theta}\{l_7\}$ |
| Average | Less | Yes | $m^{\Theta}\{l_8\}$ |

# Example

| Income | Property | Unpaid-credit | Class |
|---|---|---|---|
| High | Greater | Yes | $m^{\Theta}\{I_1\}$ |
| Average | Less | No | $m^{\Theta}\{I_2\}$ |
| High | Greater | Yes | $m^{\Theta}\{I_3\}$ |
| Average | Greater | Yes | $m^{\Theta}\{I_4\}$ |
| Low | Less | Yes | $m^{\Theta}\{I_5\}$ |
| No | Less | Yes | $m^{\Theta}\{I_6\}$ |
| High | Greater | No | $m^{\Theta}\{I_7\}$ |
| Average | Less | Yes | $m^{\Theta}\{I_8\}$ |

$$m^{\Theta}\{I_5\}(C_3) = 0.7;\ m^{\Theta}\{I_5\}(C_2 \cup C_3) = 0.2;\ m^{\Theta}\{I_5\}(\Theta) = 0.1$$

# Special cases

- Certainty: Certain *bba*
  $m^{\Theta}\{I_j\}(C_1) = 1$

# Special cases

- Certainty: Certain *bba*
  $m^{\Theta}\{I_j\}(C_1) = 1$
- Total ignorance: Vacuous *bba*
  $m^{\Theta}\{I_j\}(\Theta) = 1$

# Special cases

- Certainty: Certain *bba*
  $m^{\Theta}\{I_j\}(C_1) = 1$
- Total ignorance: Vacuous *bba*
  $m^{\Theta}\{I_j\}(\Theta) = 1$
- Imprecision: Categorical *bba*
  $m^{\Theta}\{I_j\}(C_1 \cup C_3) = 1;$

# Special cases

- Certainty: Certain *bba*
  $m^{\Theta}\{I_j\}(C_1) = 1$
- Total ignorance: Vacuous *bba*
  $m^{\Theta}\{I_j\}(\Theta) = 1$
- Imprecision: Categorical *bba*
  $m^{\Theta}\{I_j\}(C_1 \cup C_3) = 1;$
- Probabilistic: Bayesian *bba*
  $m^{\Theta}\{I_j\}(C_1) = 0.6;\ m^{\Theta}\{I_j\}(C_3) = 0.4$

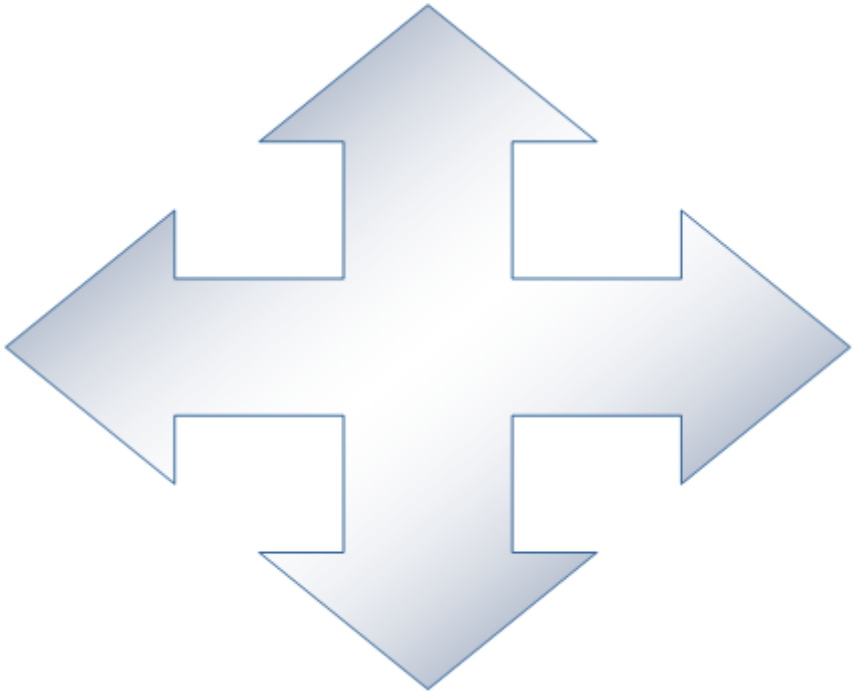
# Special cases

- Certainty: Certain *bba*
  $m^{\Theta}\{I_j\}(C_1) = 1$
- Total ignorance: Vacuous *bba*
  $m^{\Theta}\{I_j\}(\Theta) = 1$
- Imprecision: Categorical *bba*
  $m^{\Theta}\{I_j\}(C_1 \cup C_3) = 1$;
- Probabilistic: Bayesian *bba*
  $m^{\Theta}\{I_j\}(C_1) = 0.6$; $m^{\Theta}\{I_j\}(C_3) = 0.4$
- Possibilistic: Consonant *bba*
  $m^{\Theta}\{I_j\}(C_1) = 0.6$; $m^{\Theta}\{I_j\}(C_1 \cup C_2) = 0.3$; $m^{\Theta}\{I_j\}(\Theta) = 0.1$

# Parameters

**Root**
- **Test attribute**

**Decision nodes**
- **Tests on attributes**



**Leaves**
- **bba's**

**Branches**
- **Attribute values**

# Building phase

Attribute selection measure

# Building phase

**Attribute selection measure**
Measure the discriminative power of each attribute over classes

# Building phase

**Attribute selection measure**

Measure the discriminative power of each attribute over classes

**Partitioning strategy**

# Building phase

**Attribute selection measure**
Measure the discriminative power of each attribute over classes

**Partitioning strategy**
Divide the current training set into subsets according to the selected test attribute

# Building phase

**Attribute selection measure**
Measure the discriminative power of each attribute over classes

**Partitioning strategy**
Divide the current training set into subsets according to the selected test attribute

**Stopping criteria**

# Building phase

**Attribute selection measure**
Measure the discriminative power of each attribute over classes

**Partitioning strategy**
Divide the current training set into subsets according to the selected test attribute

**Stopping criteria**
Determine whether or not a training set should be further divided

# Building phase

**Attribute selection measure**
Measure the discriminative power of each attribute over classes

**Partitioning strategy**
Divide the current training set into subsets according to the selected test attribute

**Stopping criteria**
Determine whether or not a training set should be further divided

**Structure of leaves**

# Building phase

**Attribute selection measure**
Measure the discriminative power of each attribute over classes

**Partitioning strategy**
Divide the current training set into subsets according to the selected test attribute

**Stopping criteria**
Determine whether or not a training set should be further divided

**Structure of leaves**
Labeld by what?

# Attribute selection measures

## Averaging approach

## Averaging approach

Adaptation of the gain ratio to the uncertain context

## Averaging approach

Adaptation of the gain ratio to the uncertain context



Extension of the Quinlan algorithm (C4.5)

# Attribute selection measures

## Averaging approach

Adaptation of the gain ratio to the uncertain context



Extension of the Quinlan algorithm (C4.5)

## Conjunctive approach

# Attribute selection measures

## Averaging approach

Adaptation of the gain ratio to the uncertain context



Extension of the Quinlan algorithm (C4.5)

## Conjunctive approach

Based on the ideas behind the Transferable Belief Model (TBM)

# Attribute selection measures

## Averaging approach

Adaptation of the gain ratio to the uncertain context



Extension of the Quinlan algorithm (C4.5)

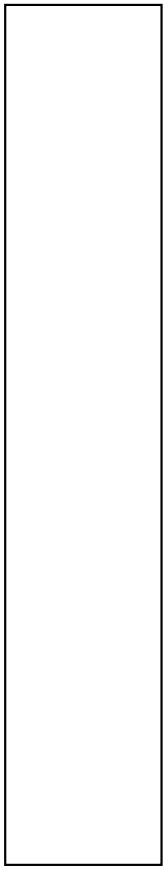
## Conjunctive approach

Based on the ideas behind the Transferable Belief Model (TBM)

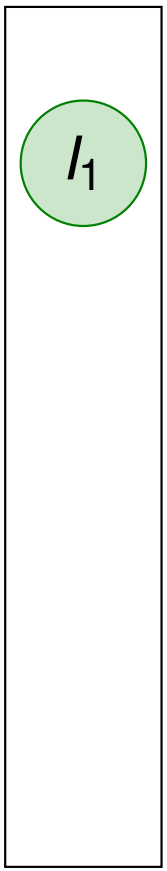


Using distances between instances

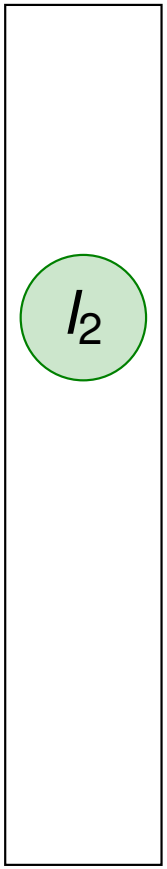
# Averaging approach



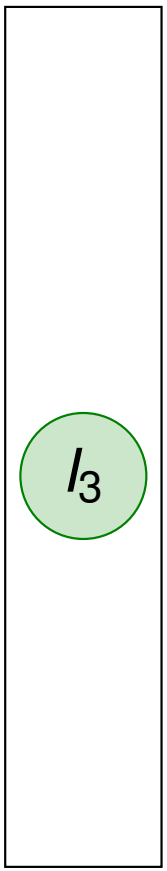
# Averaging approach



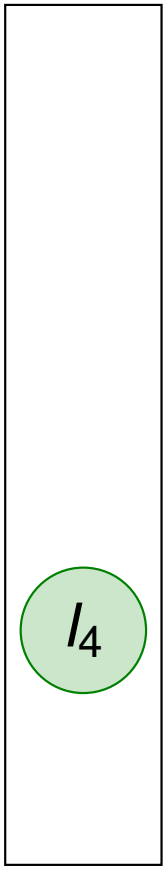
# Averaging approach



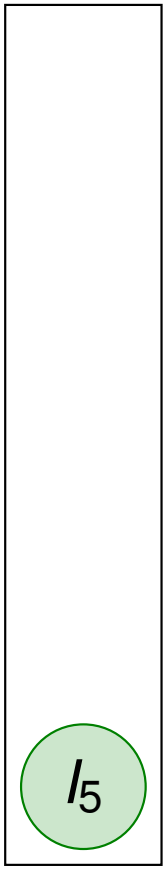
# Averaging approach



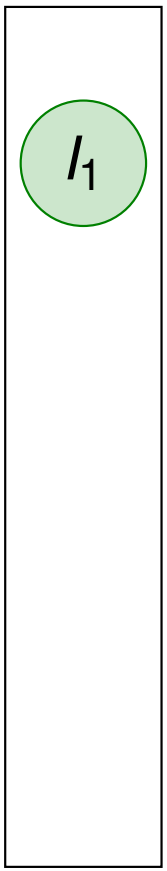
# Averaging approach



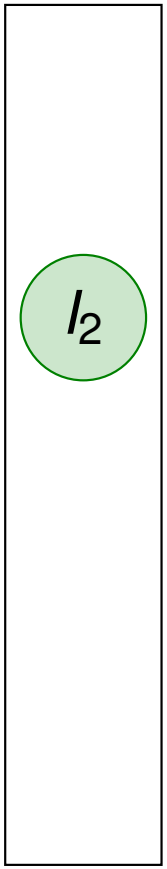
# Averaging approach



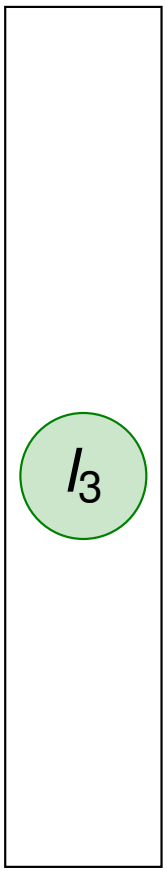
# Averaging approach



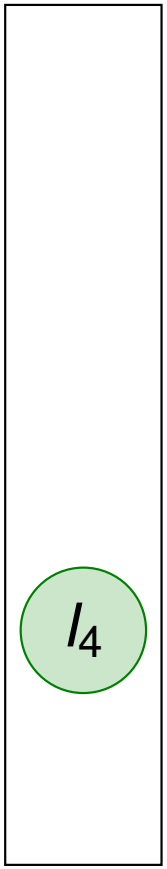
# Averaging approach



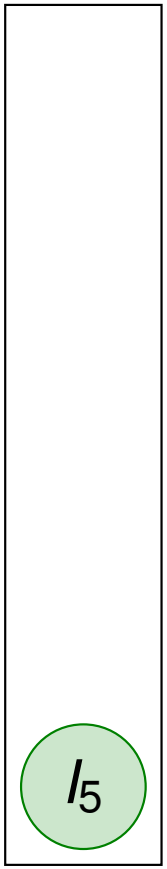
# Averaging approach



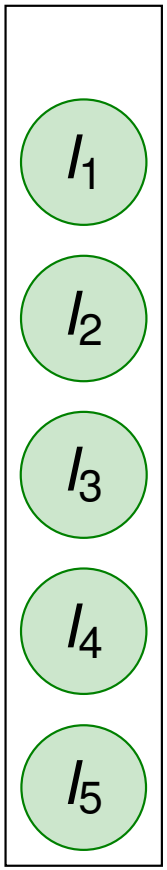
# Averaging approach



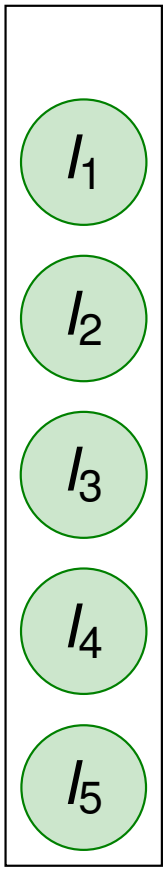
# Averaging approach



# Averaging approach

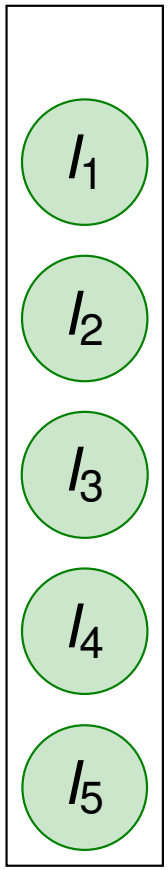


# Averaging approach




- Compute the average bba $m^{\Theta}\{T\}$.

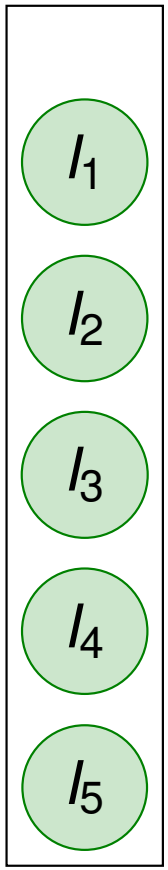
# Averaging approach




- Compute the average bba $m^{\Theta}\{T\}$.

  $\Longrightarrow$ Beliefs on the actual class of a training instance.

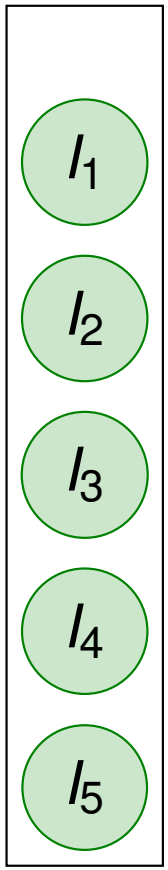
# Averaging approach




- Compute the average bba $m^{\Theta}\{T\}$.

  $\Longrightarrow$ Beliefs on the actual class of a training instance.
- Compute the average pignistic probability $BetP^{\Theta}\{T\}$.

# Averaging approach



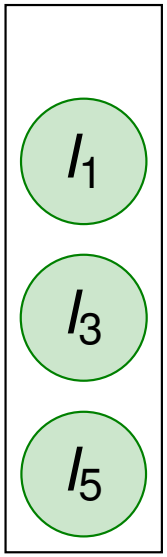

- Compute the average bba $m^{\Theta}\{T\}$.

  $\Longrightarrow$ Beliefs on the actual class of a training instance.
- Compute the average pignistic probability $BetP^{\Theta}\{T\}$.

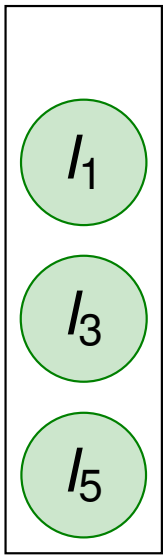
$\Downarrow$

- $Info(T)$: The entropy of the class distribution in $T$.

# Averaging approach

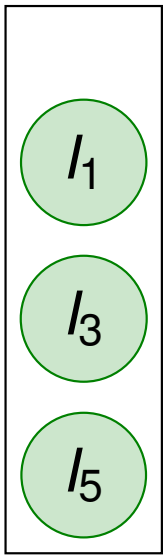


# Averaging approach




- For each value $v$ of the attribute $A$.

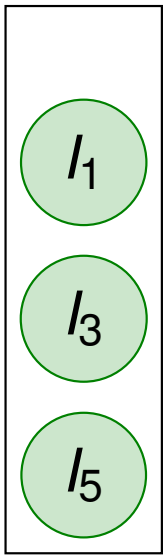
# Averaging approach




- For each value $v$ of the attribute $A$.

  $\Rightarrow$ Build the set of objects $T_v^A$.

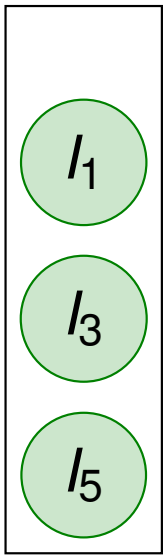
# Averaging approach




- For each value $v$ of the attribute $A$.

  $\Longrightarrow$ Build the set of objects $T_v^A$.
- Compute the average bba $m^{\Theta}\{T_v^A\}$.

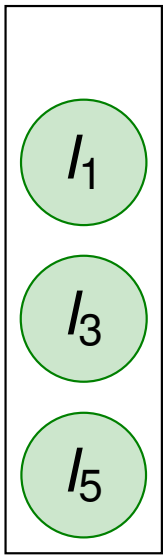
# Averaging approach




- For each value $v$ of the attribute $A$.

  $\Longrightarrow$ Build the set of objects $T_v^A$.
- Compute the average bba $m^{\Theta}\{T_v^A\}$.
- Compute the average pignistic probability $BetP^{\Theta}\{T_v^A\}$

# Averaging approach




- For each value $v$ of the attribute $A$.

  $\Longrightarrow$ Build the set of objects $T_v^A$.
- Compute the average bba $m^{\Theta}\{T_v^A\}$.
- Compute the average pignistic probability $BetP^{\Theta}\{T_v^A\}$

$\Downarrow$

- $Info(T_v^A)$: The entropy of the class distribution in $T_v^A$.

# Averaging approach

For each attribute

- Compute $Info_A(T)$ the weighed sum of the different $Info(T_v^A)$.

# Averaging approach

For each attribute

- Compute $Info_A(T)$ the weighed sum of the different $Info(T_v^A)$.
- Compute the information gain of the attribute

$\Longrightarrow Gain(T, A) = Info(T) - Info_A(T)$

# Averaging approach

For each attribute

- Compute $Info_A(T)$ the weighed sum of the different $Info(T_v^A)$.
- Compute the information gain of the attribute

$$\Rightarrow Gain(T, A) = Info(T) - Info_A(T)$$

- Compute the gain ratio of the attribute

$$\Rightarrow Gain\ ratio(T, A) = \frac{Gain(T,A)}{Split\ \ Info(T,A)}$$

# Averaging approach

For each attribute

- Compute $Info_A(T)$ the weighed sum of the different $Info(T_v^A)$.
- Compute the information gain of the attribute

$$\Rightarrow Gain(T, A) = Info(T) - Info_A(T)$$

- Compute the gain ratio of the attribute

$$\Rightarrow Gain\ ratio(T, A) = \frac{Gain(T,A)}{Split\ \ Info(T,A)}$$

$$\Downarrow$$

Choose the attribute with the highest gain ratio.

# Conjunctive approach

# Conjunctive approach

- Development of a distance between bba's:

# Conjunctive approach

- Development of a distance between bba's:

  All objects in a leaf are close to each other.

# Conjunctive approach

- Development of a distance between bba's:

  All objects in a leaf are close to each other.

- Minimize the intra-group distance.

# Conjunctive approach

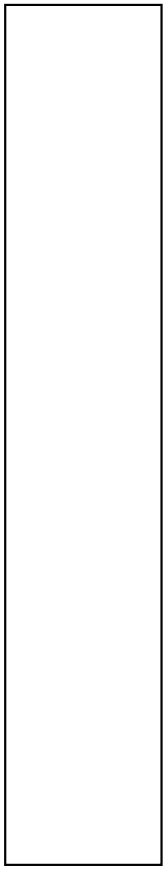
- Development of a distance between bba's:

  All objects in a leaf are close to each other.

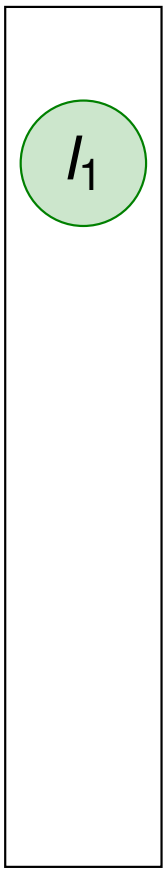
- Minimize the intra-group distance.

- Maximize the inter-group distance.

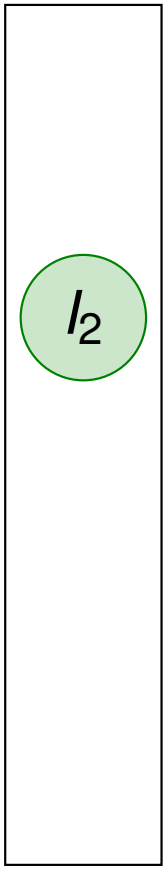
# Conjunctive approach



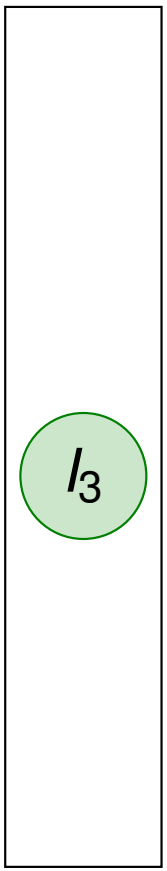
# Conjunctive approach



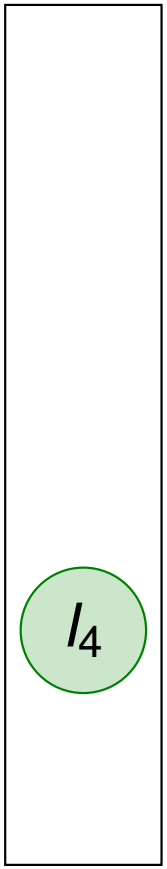
# Conjunctive approach



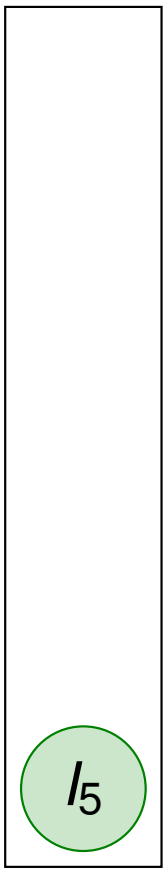
# Conjunctive approach



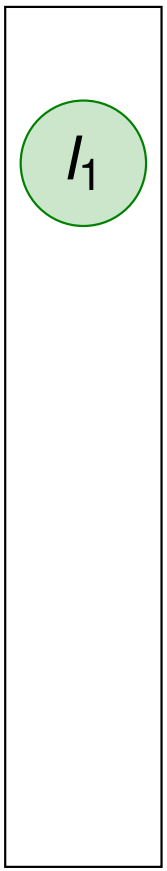
# Conjunctive approach



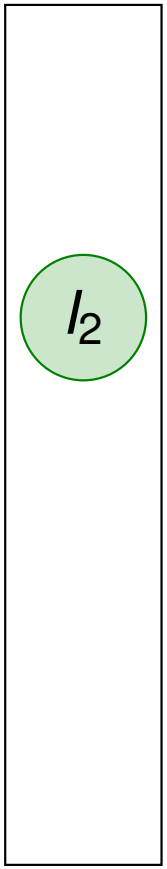
# Conjunctive approach



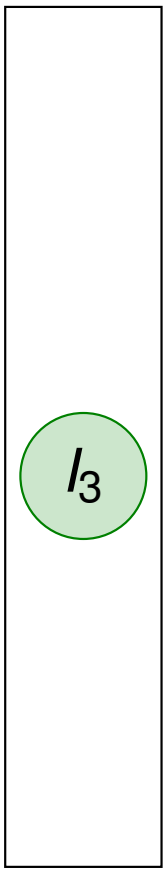
# Conjunctive approach



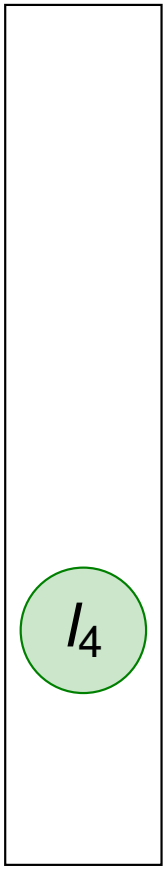
# Conjunctive approach



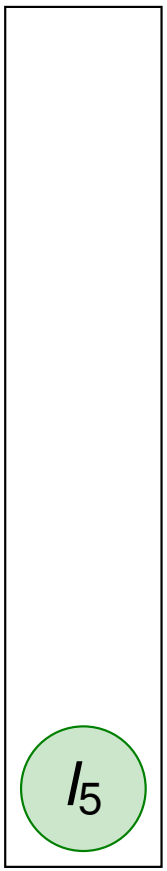
# Conjunctive approach



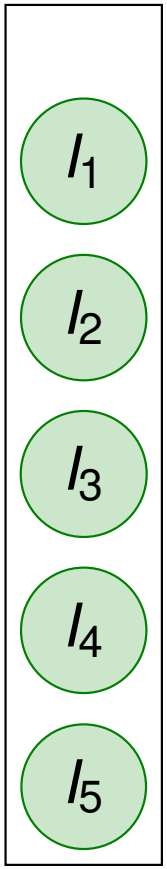
# Conjunctive approach



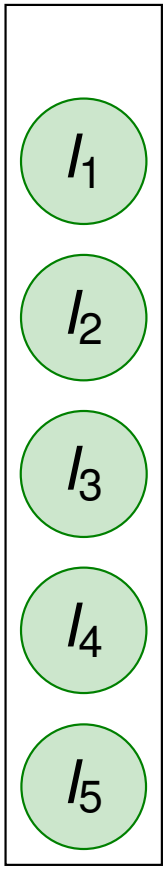
# Conjunctive approach



# Conjunctive approach

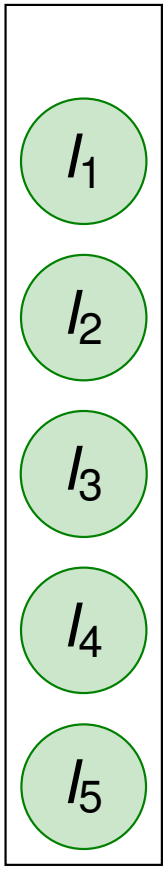


# Conjunctive approach




- Compute the joint bba $m^{\Theta}\{T\}$.

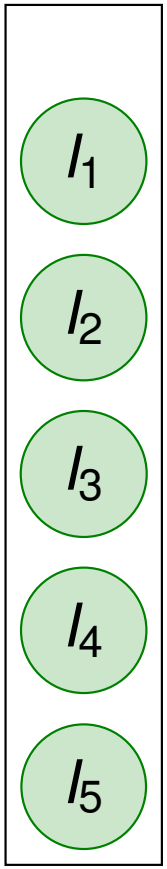
# Conjunctive approach




- Compute the joint bba $m^{\Theta}\{T\}$.

  $\Rightarrow$ Beliefs on the actual class of a training instance taking into account of what all training instances jointly express.

# Conjunctive approach



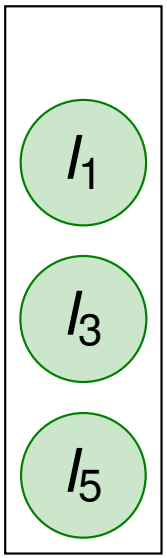

- Compute the joint bba $m^{\Theta}\{T\}$.

  $\Rightarrow$ Beliefs on the actual class of a training instance taking into account of what all training instances jointly express.

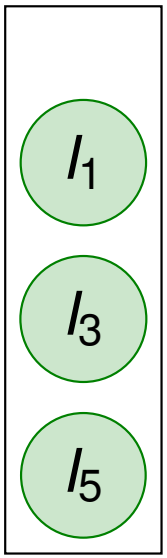
$\Downarrow$

- Compute the sum of distances $SumD(T)$ separating instances in $T$.

# Conjunctive approach

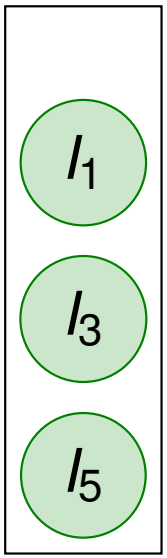


# Conjunctive approach




- For each value $v$ of the attribute $A$.

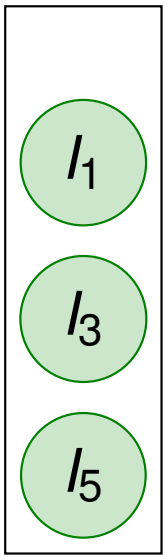
# Conjunctive approach




- For each value $v$ of the attribute $A$.

  $\Rightarrow$ Build the set of objects $T_v^A$.

# Conjunctive approach

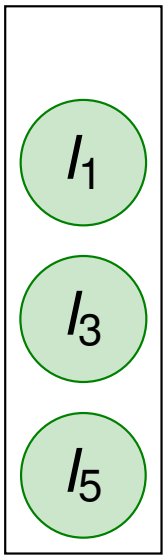



- For each value $v$ of the attribute $A$.

  $\Rightarrow$ Build the set of objects $T_v^A$.

- Compute the joint bba $m^{\Theta}\{T\}$.

# Conjunctive approach




- For each value $v$ of the attribute $A$.

  $\Longrightarrow$ Build the set of objects $T_v^A$.

- Compute the joint bba $m^{\Theta}\{T\}$.



- Compute the sum of distances $SumD(T_v^A)$ separating instances in $T_v^A$ .

# Conjunctive approach

For each attribute

- Compute $SumD_A(T)$ the weighed sum of the different $SumD(T_v^A)$.

# Conjunctive approach

For each attribute

- Compute $SumD_A(T)$ the weighed sum of the different $SumD(T_v^A)$.
- Compute the diff measure of the attribute

$\Longrightarrow Diff(S, A_k) = SumD(S) - SumD_A(S)$

# Conjunctive approach

For each attribute

- Compute $SumD_A(T)$ the weighed sum of the different $SumD(T_v^A)$.
- Compute the diff measure of the attribute

$$\Rightarrow Diff(S, A_k) = SumD(S) - SumD_A(S)$$

- Compute the diff ratio of the attribute

$$\Rightarrow Diff\ Ratio(S, A) = \frac{Diff(S,A)}{Split\ Info(S,A)}$$

$$\Downarrow$$

Choose the attribute with the highest diff ratio.

# Partitioning strategy

## Symbolic attributes

# Partitioning strategy

Symbolic attributes



Creation of an edge for each attribute value

# Partitioning strategy

Symbolic attributes



Creation of an edge for each attribute value



Construction of several sub decision trees

# Stopping criteria

1. The treated node includes instances having the same bba.

# Stopping criteria

1. The treated node includes instances having the same bba.
2. There is no further attribute to test.

# Stopping criteria

1. The treated node includes instances having the same bba.
2. There is no further attribute to test.
3. The node includes only one object.

# Stopping criteria

1. The treated node includes instances having the same bba.
2. There is no further attribute to test.
3. The node includes only one object.
4. The value of the chosen attribute selection measure of the remaining attributes are less or equal than zero.

# Structure of a leaf

Leaf $\Rightarrow$ bba on classes

# Structure of a leaf

Leaf $\Rightarrow$ bba on classes

One object attached to the leaf

$\Downarrow$

Leaf's bba = object's bba

# Structure of a leaf

Leaf $\Rightarrow$ bba on classes

One object attached to the leaf

$\Downarrow$

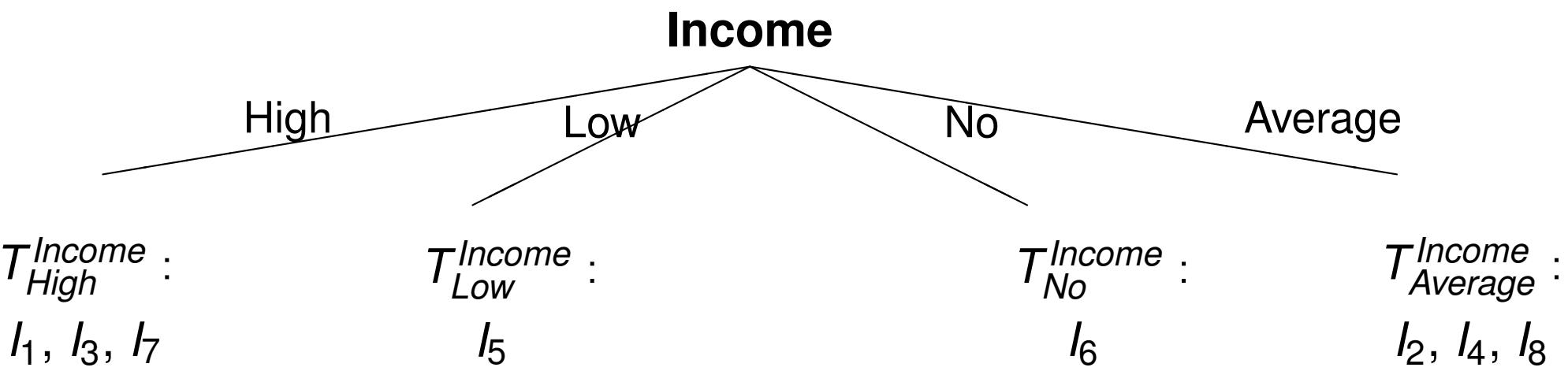
Leaf's bba = object's bba

Several objects are attached to the leaf

$\Downarrow$

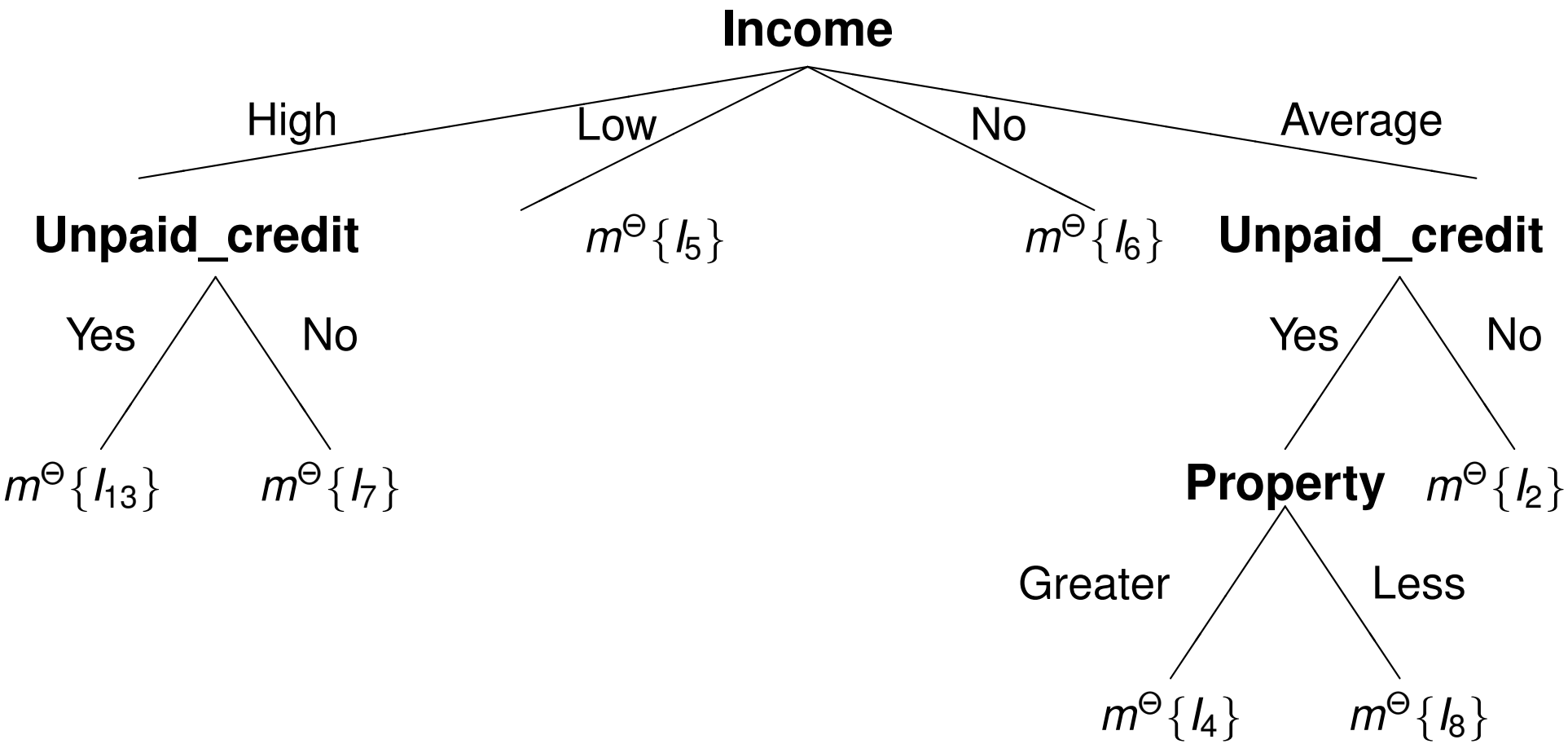
Leaf's bba = the average bba or the joint bba

# Building phase




First generated belief decision tree

# Building phase




Final belief decision tree

# Standard classification

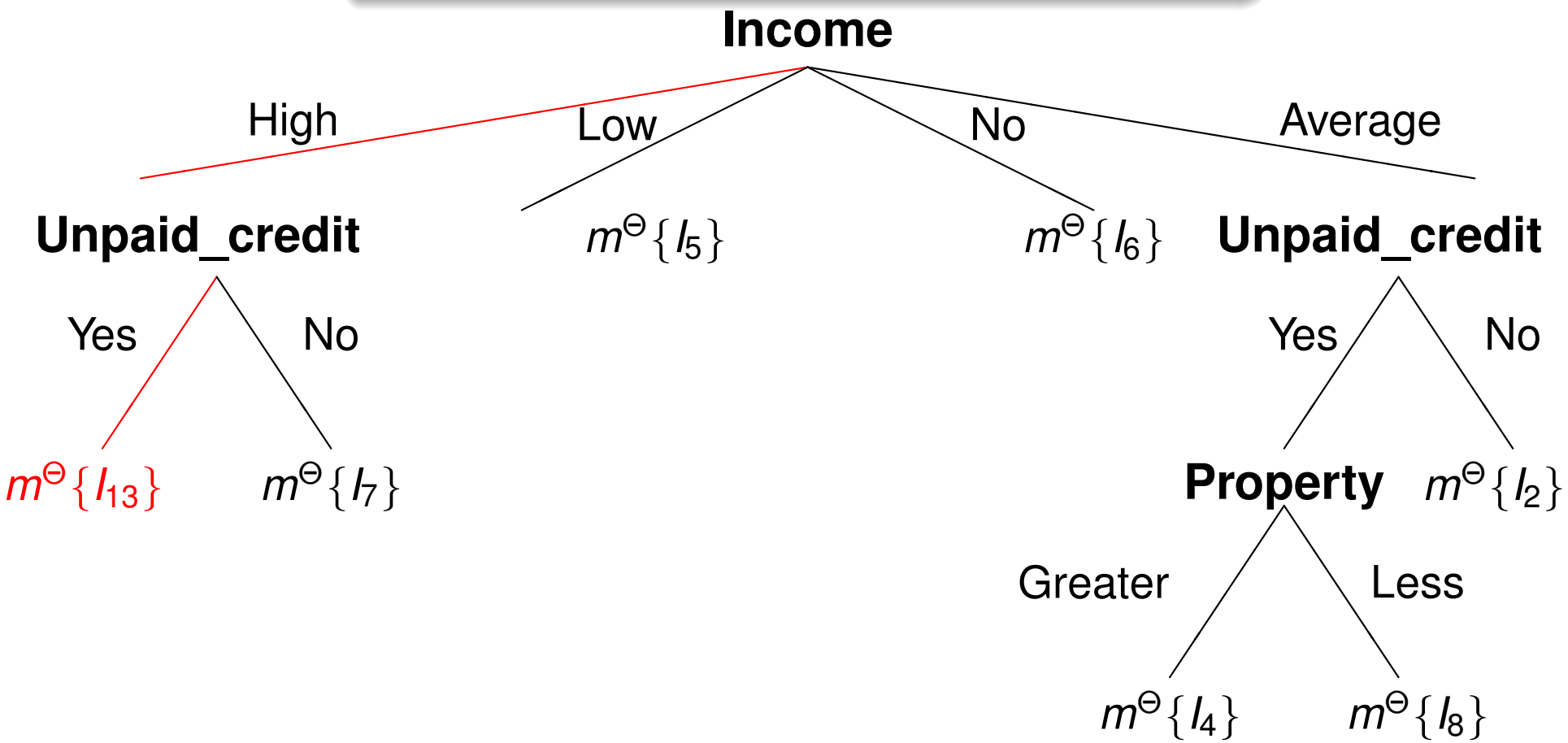
**Attribute values are known with certainty**

- Income: High
- Property: Greater
- Unpaid_credit: Yes

# Standard classification

**Attribute values are known with certainty**

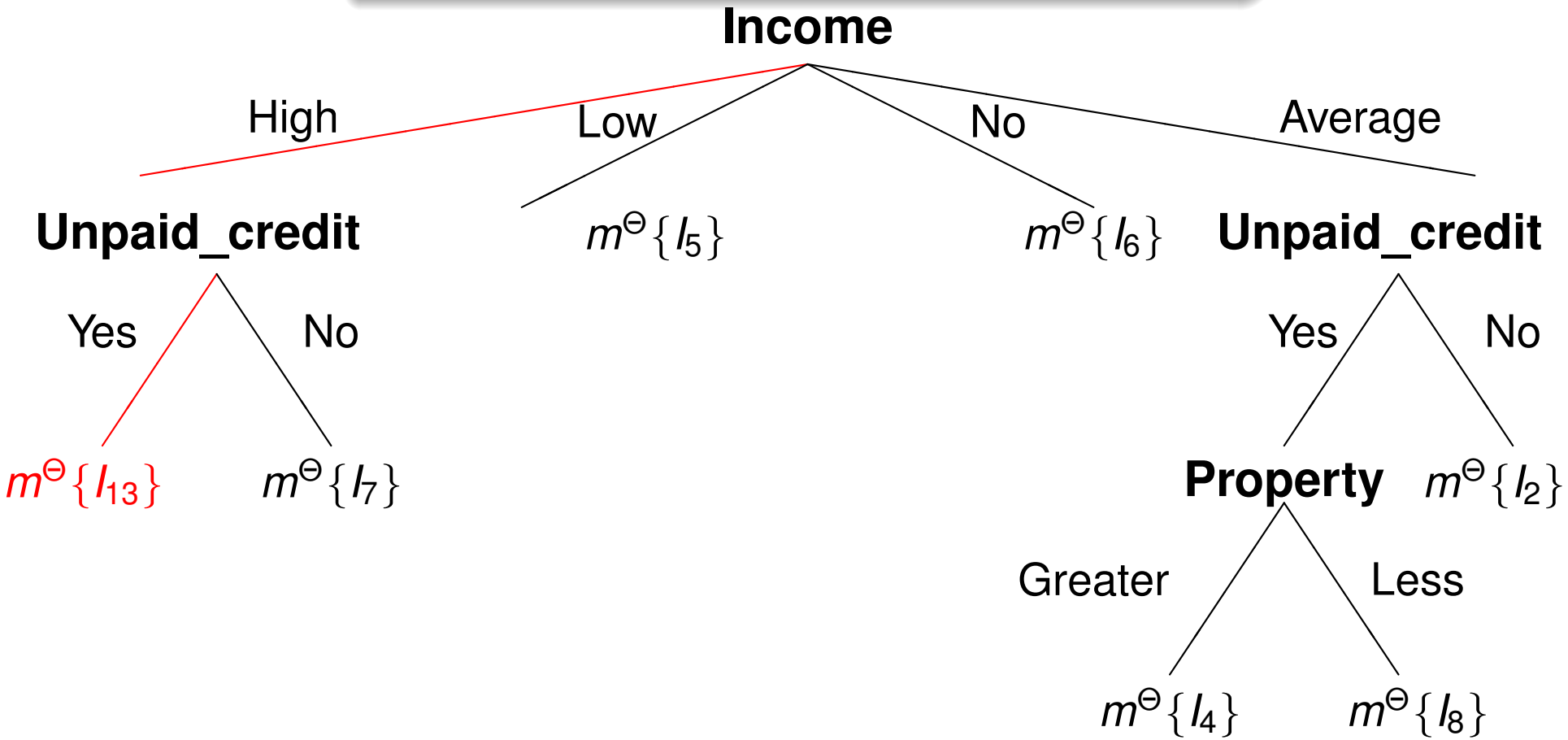
- Income: High
- Property: Greater
- Unpaid_credit: Yes



# Standard classification

$m^{\Theta}\{I_{13}\}(C_1) = 0.65$; $m^{\Theta}\{I_{13}\}(\Theta) = 0.35$;

- $C_1$: 0.76
- $C_2$: 0.12
- $C_3$: 0.12



# Disjunctive case

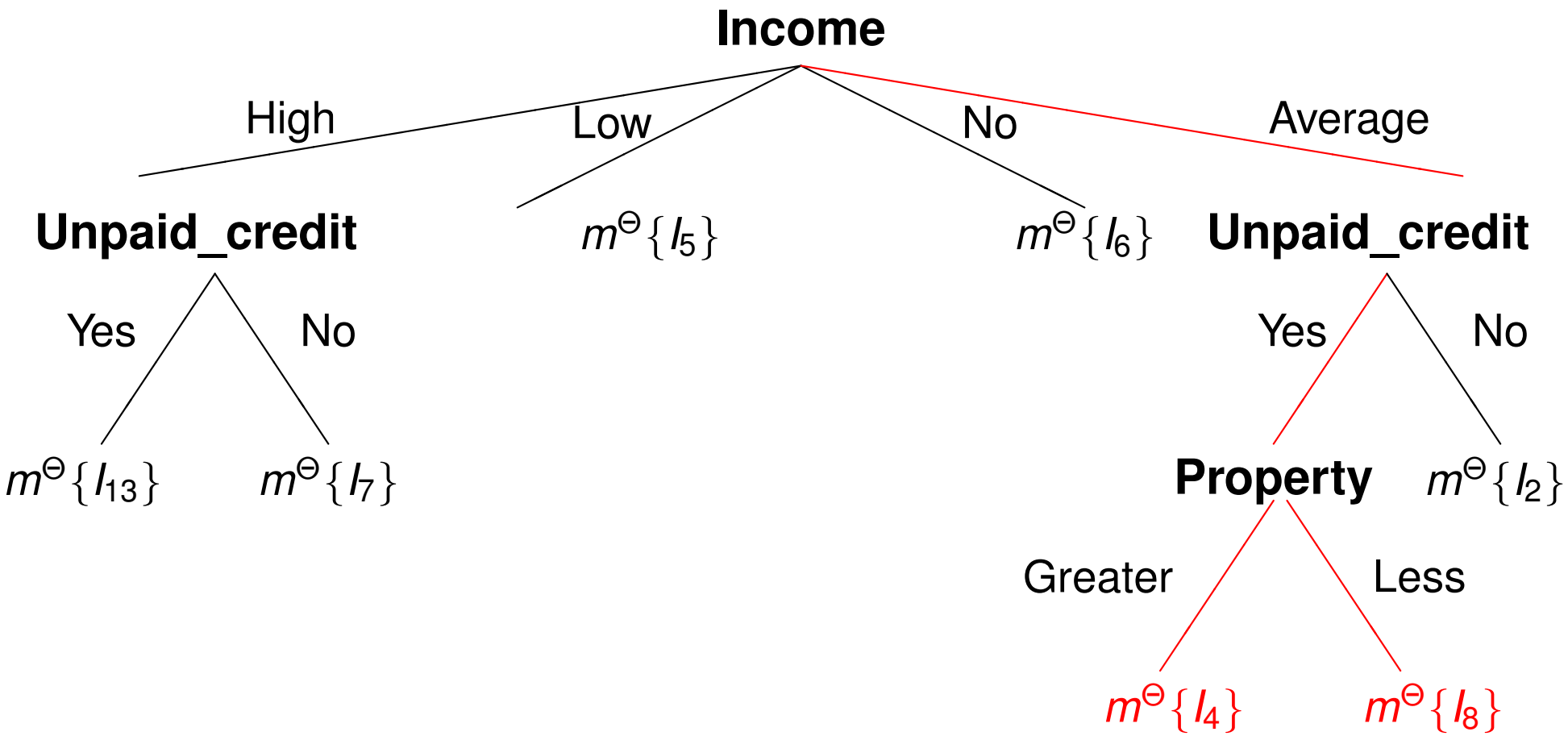
## Disjunctive values for some attributes

- Income: Average
- Property: Greater or Less
- Unpaid_credit: Yes

# Disjunctive case

**Disjunctive values for some attributes**

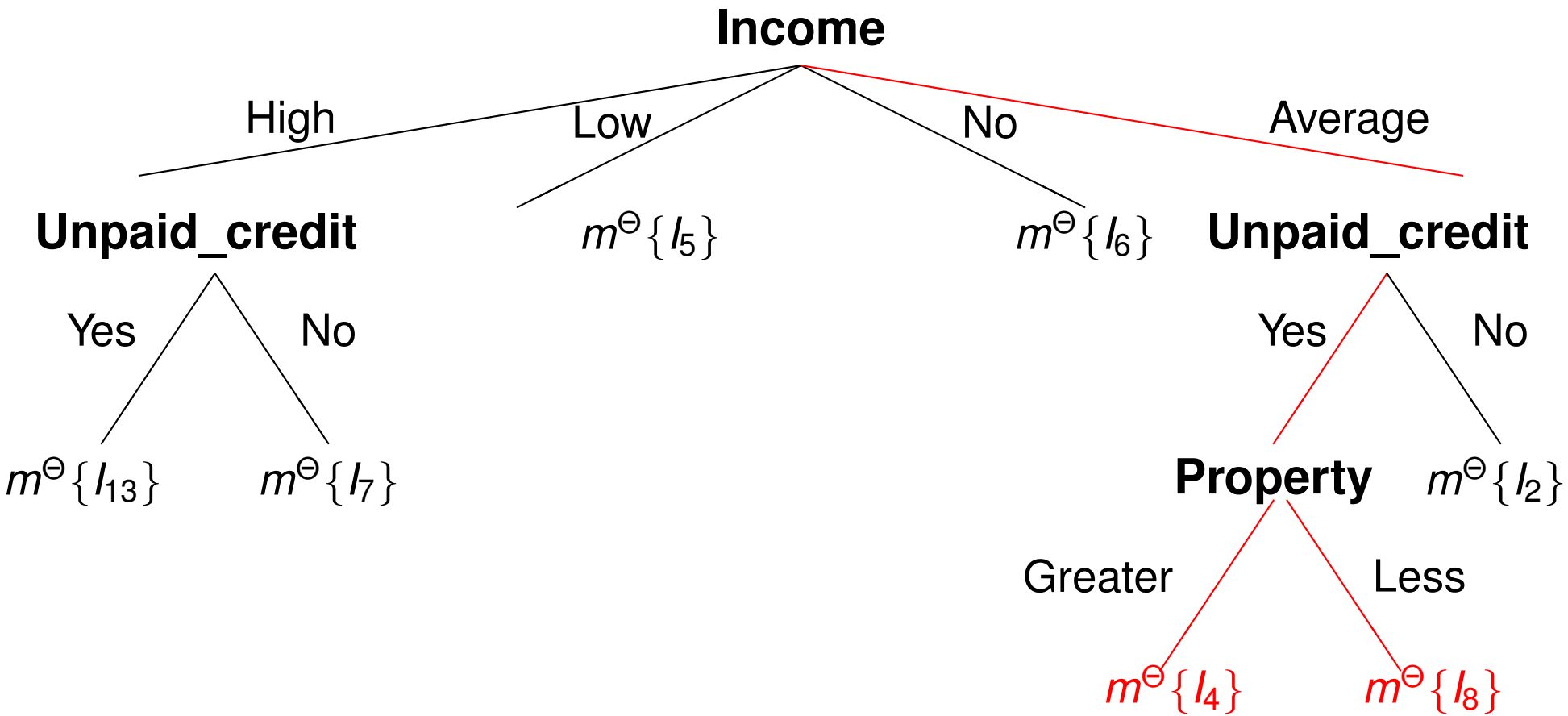
- Income: Average
- Property: Greater or Less
- Unpaid_credit: Yes



# Disjunctive case

$m^{\Theta}\{I\}(C_2) = 0.24$; $m^{\Theta}\{I\}(C_3) = 0.12$; $m^{\Theta}\{I\}(C_2 \cup C_3) = 0.36$; $m^{\Theta}\{I\}(\Theta) = 0.28$;

- $C_1$: 0.1
- $C_2$: 0.51
- $C_3$: 0.39



# Genral case

## Uncertain value for some attributes

$\Rightarrow$ bba for each attribute

- Income: Average
- Property: $m^{Property}(\{Greater\}) = 0.4$; $m^{Property}(\{Less\}) = 0.3$; $m^{Property}(\Theta_{Property}) = 0.3$;
- Unpaid_credit: Yes

# Genral case

**Uncertain value for some attributes**

$\Rightarrow$ bba for each attribute

- Income: Average
- Property: $m^{Property}(\{Greater\}) = 0.4$; $m^{Property}(\{Less\}) = 0.3$; $m^{Property}(\Theta_{Property}) = 0.3$;
- Unpaid_credit: Yes

- Compute the bba on the different combination of attribute values $m^A = m^{Income\uparrow A} \bigcirc\!\!\!\!\cap\, m^{Property\uparrow A} \bigcirc\!\!\!\!\cap\, m^{Unpaid_credit\uparrow A}$ such that:
  $m^A(\{(Average, Greater, Yes)\}) = 0.4$;
  $m^A(\{(Average, Less, Yes)\}) = 0.3$;
  $m^A(\{Average\} \times \Theta_{Property} \times \{Yes\}) = 0.3$;

# Genral case

## Uncertain value for some attributes

$\Rightarrow$ bba for each attribute

- Income: Average
- Property: $m^{Property}(\{Greater\}) = 0.4$;
  $m^{Property}(\{Less\}) = 0.3$; $m^{Property}(\Theta_{Property}) = 0.3$;
- Unpaid_credit: Yes

- Compute the bba on the different combination of attribute values
  $m^A = m^{Income\uparrow A} \bigcap m^{Property\uparrow A} \bigcap m^{Unpaid_credit\uparrow A}$ such that:
  $m^A(\{(Average, Greater, Yes)\}) = 0.4$;
  $m^A(\{(Average, Less, Yes)\}) = 0.3$;
  $m^A(\{Average\} \times \Theta_{Property} \times \{Yes\}) = 0.3$;
- For each focal element, compute $bel^{\Theta}[X]$:
  - If these paths lead to one leaf $\Rightarrow$ bel's leaf

# Genral case

**Uncertain value for some attributes**

$\Rightarrow$ bba for each attribute

- Income: Average
- Property: $m^{Property}(\{Greater\}) = 0.4$; $m^{Property}(\{Less\}) = 0.3$; $m^{Property}(\Theta_{Property}) = 0.3$;
- Unpaid_credit: Yes

- Compute the bba on the different combination of attribute values $m^A = m^{Income\uparrow A} \bigcap m^{Property\uparrow A} \bigcap m^{Unpaid_credit\uparrow A}$ such that:
  $m^A(\{(Average, Greater, Yes)\}) = 0.4$;
  $m^A(\{(Average, Less, Yes)\}) = 0.3$;
  $m^A(\{Average\} \times \Theta_{Property} \times \{Yes\}) = 0.3$;
- For each focal element, compute $bel^{\Theta}[X]$:
  - If these paths lead to one leaf $\Rightarrow$ bel's leaf
  - If these paths lead to distinct leaves $\Rightarrow$ disjunctive combination of bel's leaves

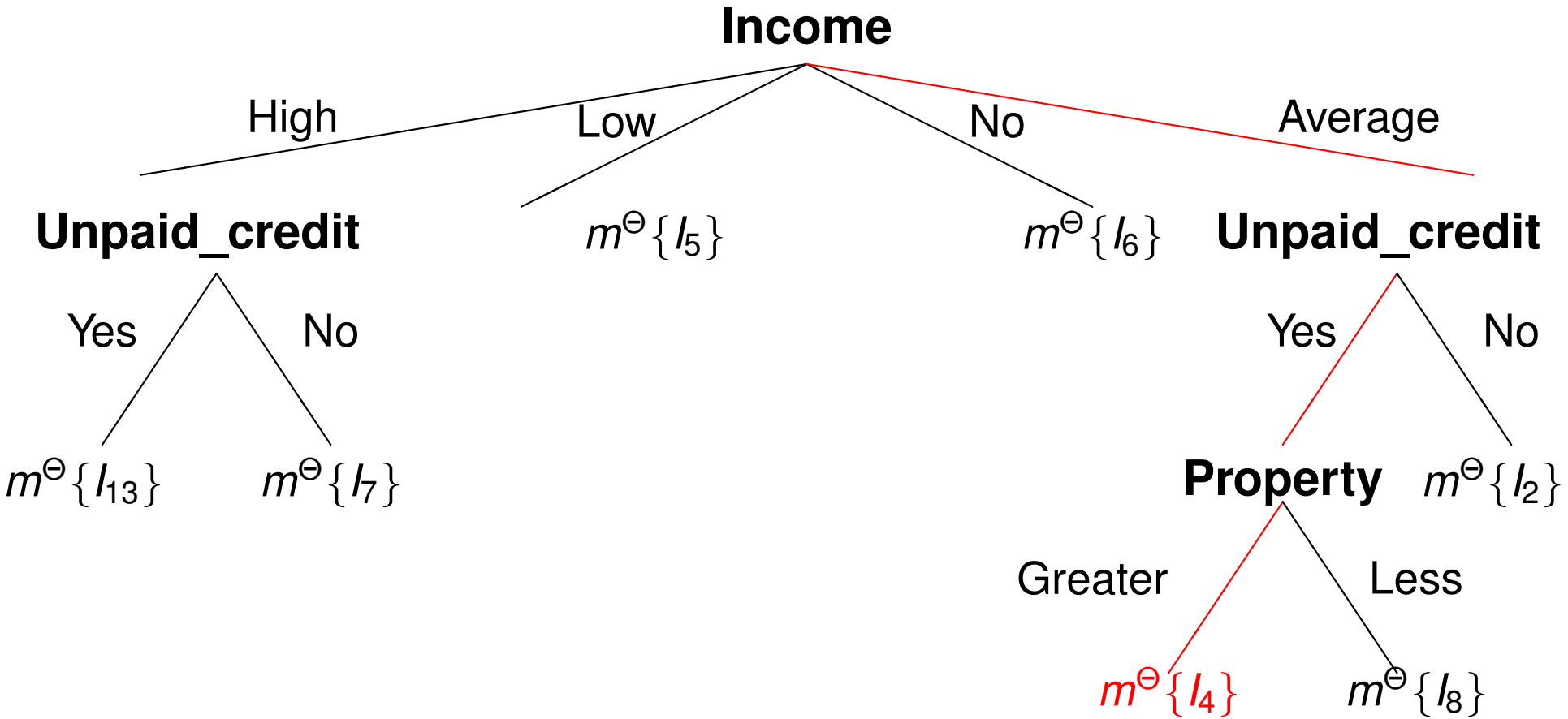
# General case

$m^A(\{(Average, Greater, Yes)\}) = 0.4;$

$$bel^{\Theta}[m^A](C) = \sum_{x \subseteq \Theta_A} m^A(x) bel^{\Theta}[x](C) \text{ for } C \subseteq \Theta$$

$m^A(\{(Average, Greater, Yes)\}) = 0.4;$

$$bel^{\Theta}[m^A](C) = \sum_{x \subseteq \Theta_A} m^A(x) bel^{\Theta}[x](C) \text{ for } C \subseteq \Theta$$



# General case

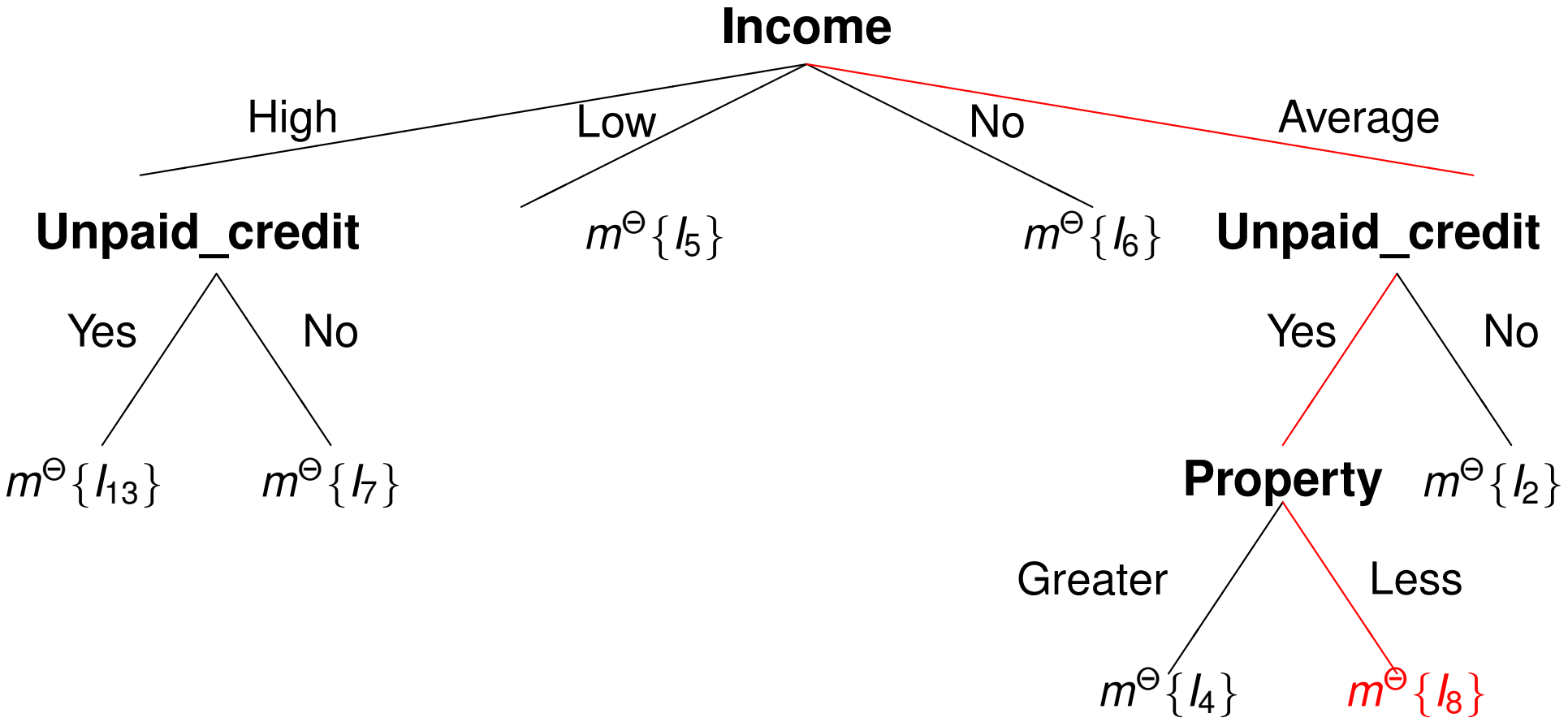
$m^A(\{(Average, Less, Yes)\}) = 0.3;$

$$bel^{\Theta}[m^A](C) = \sum_{x \subseteq \Theta_A} m^A(x) bel^{\Theta}[x](C) \text{ for } C \subseteq \Theta$$

# General case

$m^A(\{(Average, Less, Yes)\}) = 0.3;$

$$bel^{\Theta}[m^A](C) = \sum_{x \subseteq \Theta_A} m^A(x) bel^{\Theta}[x](C) \text{ for } C \subseteq \Theta$$



# General case

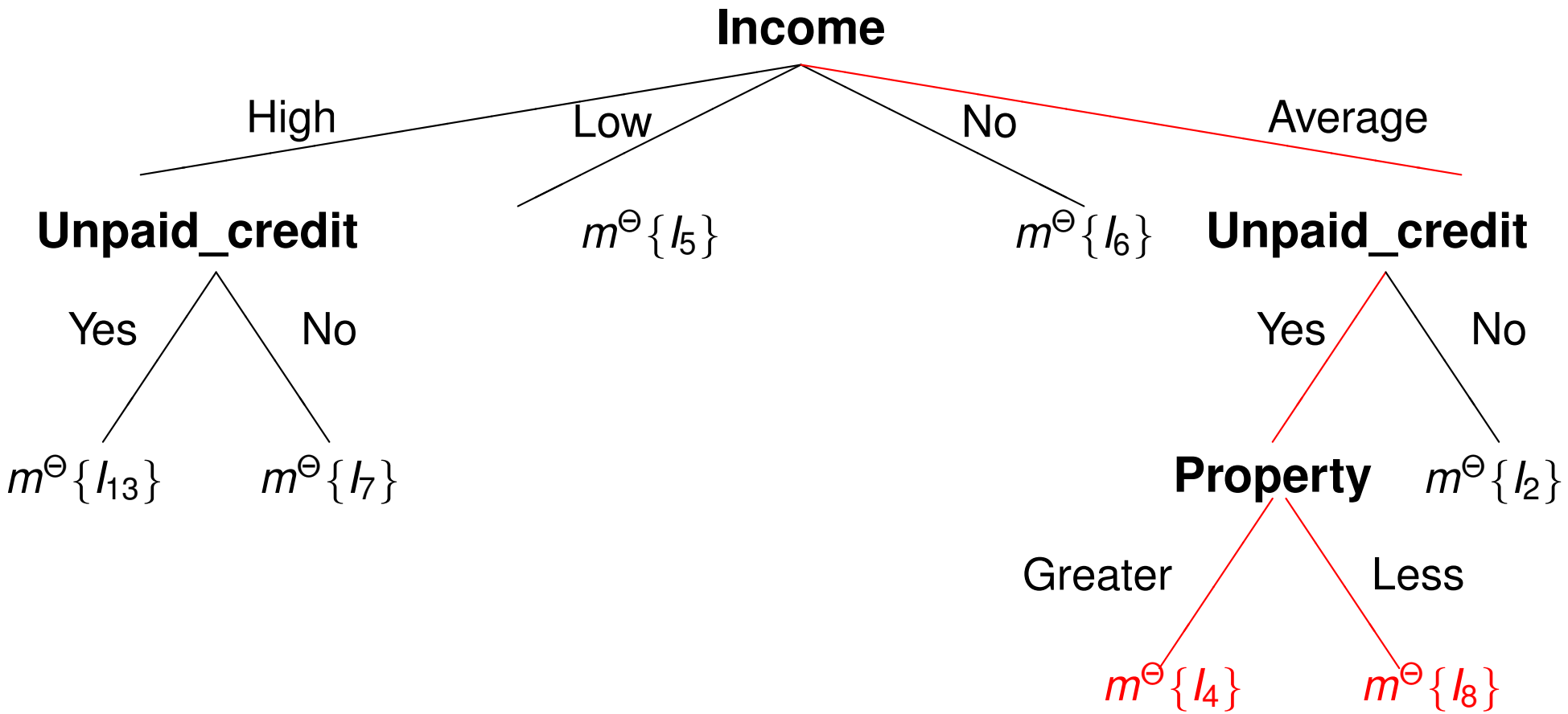
$m^A(\{Average\} \times \Theta_{Property} \times \{Yes\}) = 0.3;$

$$bel^{\Theta}[m^A](C) = \sum_{x \subseteq \Theta_A} m^A(x) bel^{\Theta}[x](C) \text{ for } C \subseteq \Theta$$

# General case

$m^A(\{Average\} \times \Theta_{Property} \times \{Yes\}) = 0.3;$

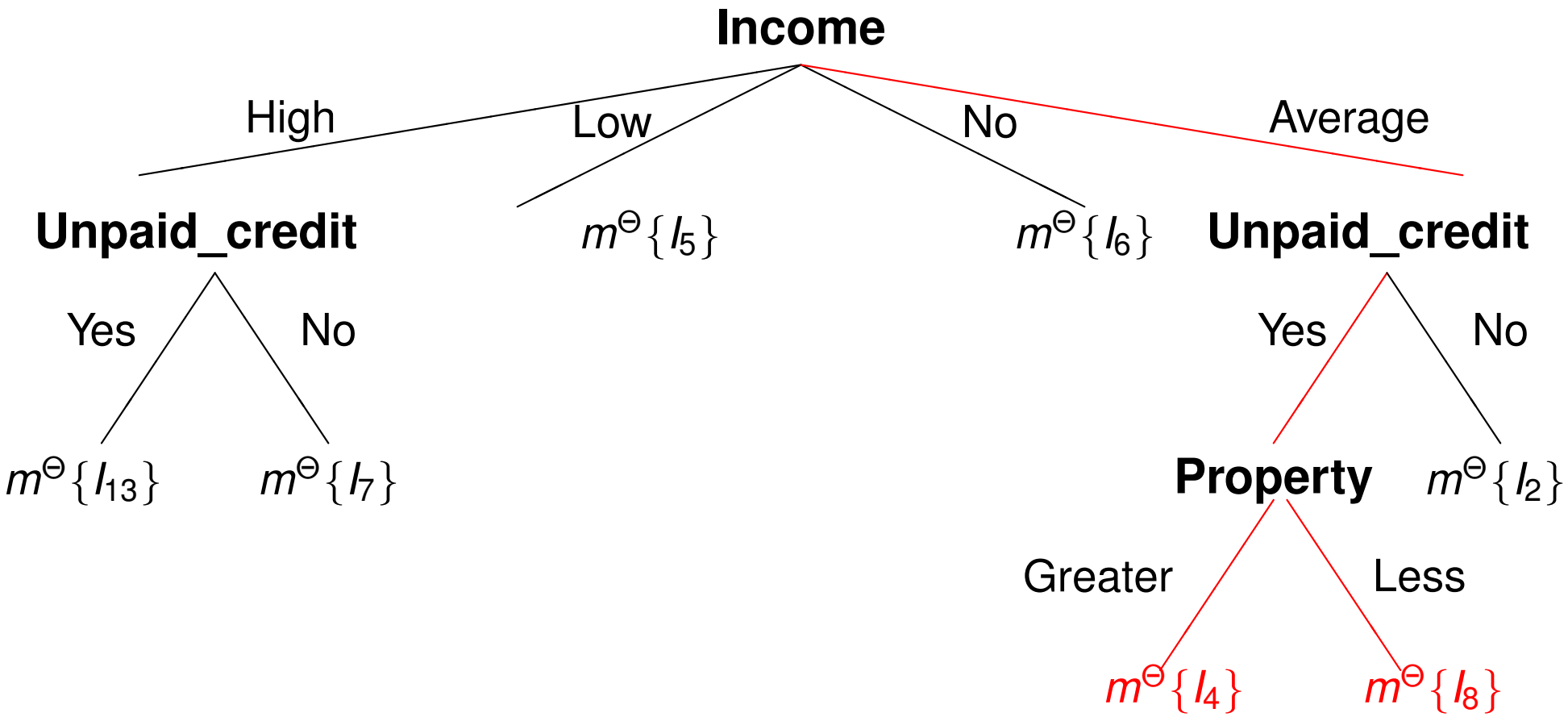
$$bel^{\Theta}[m^A](C) = \sum_{x \subseteq \Theta_A} m^A(x) bel^{\Theta}[x](C) \text{ for } C \subseteq \Theta$$



# General case

$m^{\Theta}[m^{A}](C_2) = 0.43$; $m^{\Theta}[m^{A}](C_3) = 0.28$; $m^{\Theta}[m^{A}](C_2 \cup C_3) = 0.11$; $m^{\Theta}[m^{A}](\Theta) = 0.18$;

- $C_1$: 0.06
- $C_2$: 0.55
- $C_3$: 0.39



# Pruning

## Why?

Problem of overfitting:

- Too many branches
- Tree difficult to interpret
- Poor accuracy

# Pruning

## Why?

Problem of overfitting:

- Too many branches
- Tree difficult to interpret
- Poor accuracy

## Aims

- Reduce the size
- Increase classfication accuracy

# Pre-pruning BDT

## Pre-pruning

Stop growing the tree before it perfectly classifies the training set.

# Pre-pruning BDT

## Pre-pruning

Stop growing the tree before it perfectly classifies the training set.

$\Rightarrow$ Halt tree construction early.

# Pre-pruning BDT

## Pre-pruning

Stop growing the tree before it perfectly classifies the training set.

$\Rightarrow$ Halt tree construction early.

## Improving stopping rules of BDT

- Discunting factor
- Number of objects in a decision node

# Pre-pruning BDT

## Pre-pruning

Stop growing the tree before it perfectly classifies the training set.

$\Rightarrow$ Halt tree construction early.

## Improving stopping rules of BDT

- Discunting factor
- Number of objects in a decision node

$\Rightarrow$ Discounting the node's bba taking into account the number of objects in the node.

# Pre-pruning BDT

## Pre-pruning

Stop growing the tree before it perfectly classifies the training set.

$\Rightarrow$ Halt tree construction early.

## Improving stopping rules of BDT

- Discunting factor
- Number of objects in a decision node

$\Rightarrow$ Discounting the node's bba taking into account the number of objects in the node.

$\Downarrow$

Preserve or improve the quality of the classifier.

# Post-pruning BDT

## Post-pruning

Grow the full tree, allow it to overfit the data and then post-prune it.

## Post-pruning

Grow the full tree, allow it to overfit the data and then post-prune it.

## Standard Post-pruning methods

- MCCP: Minimal Cost-Complexity Pruning (Breiman et al, 1984).
- CVP: Critical Value Pruning (Mingers, 1987).
- REP: Reduced Error Pruning (Quinlan, 1987).
- MEP: Minimum Error Pruning (Niblett et Bratko, 1986).
- PEP: Pessimistic Error Pruning (Quinlan, 1987).
- EBP: Error Based Pruning (Quinlan, 1993).

# Post-pruning BDT based on MCCP

## Why?

- Accurate
- Good tree size
- Series Pruned trees

# Post-pruning BDT based on MCCP

## Why?

- Accurate
- Good tree size
- Series Pruned trees

## MCCP in averaging approach

Error rate based on BetP.

# Post-pruning BDT based on MCCP

## Why?

- Accurate
- Good tree size
- Series Pruned trees

## MCCP in averaging approach

Error rate based on BetP.

## MCCP in conjunctive approach

Error rate based on distance.

BDT without pruning

# BDT and improvements

BDT without pruning

BDT with pre-pruning

# BDT and improvements

BDT without pruning

BDT with pre-pruning

BDT with post-pruning

# Evaluation criteria

- The Percent of Correct Classification: *PCC*

$$PCC = \frac{\text{number of well classified instances}}{\text{number of classified instances}}$$

# Evaluation criteria

- The Percent of Correct Classification: *PCC*

$$PCC = \frac{\text{number of well classified instances}}{\text{number of classified instances}}$$

- The distance criterion: *dist_crit*

$$\begin{aligned} dist_crit(I_j) &= Distance(BetP^{\Theta}\{L\}, C(I_j)) \\ &= \sum_{i=1}^{n} (BetP^{\Theta}\{L\}(C_i) - \delta_{j,i})^2 \end{aligned}$$

where the real class of the testing instance $I_j$ is $C(I_j)$, and $\delta_{j,i} = 1$ if $C(I_j) = C_i$ and 0 otherwise.

# Evaluation criteria

- The Percent of Correct Classification: *PCC*

$$PCC = \frac{\text{number of well classified instances}}{\text{number of classified instances}}$$

- The distance criterion: *dist_crit*

$$\begin{aligned} dist_crit(I_j) &= Distance(BetP^{\Theta}\{L\}, C(I_j)) \\ &= \sum_{i=1}^{n} (BetP^{\Theta}\{L\}(C_i) - \delta_{j,i})^2 \end{aligned}$$

where the real class of the testing instance $I_j$ is $C(I_j)$, and $\delta_{j,i} = 1$ if $C(I_j) = C_i$ and 0 otherwise.

- The size: the number of decision nodes and leaves in the belief decision tree.

# UCI repository databases

| Database | #Instances | #Attributes | #Classes |
|---|---|---|---|
| W. Breast Cancer | 690 | 8 | 2 |
| Balance Scale | 625 | 4 | 3 |
| C. Voting records | 497 | 16 | 2 |
| Zoo | 101 | 17 | 7 |
| Nursery | 12960 | 8 | 3 |
| Solar Flares | 1389 | 10 | 2 |
| Lung Cancer | 32 | 56 | 3 |
| Hayes-Roth | 160 | 5 | 3 |
| Car Evaluation | 1728 | 6 | 4 |
| Lymphography | 148 | 18 | 4 |
| Spect Heart | 267 | 22 | 2 |
| Tic-Tac-Toe Endgame | 958 | 9 | 2 |

# UCI repository databases

| Database | #Instances | #Attributes | #Classes |
|---|---|---|---|
| W. Breast Cancer | 690 | 8 | 2 |
| Balance Scale | 625 | 4 | 3 |
| C. Voting records | 497 | 16 | 2 |
| Zoo | 101 | 17 | 7 |
| Nursery | 12960 | 8 | 3 |
| Solar Flares | 1389 | 10 | 2 |
| Lung Cancer | 32 | 56 | 3 |
| Hayes-Roth | 160 | 5 | 3 |
| Car Evaluation | 1728 | 6 | 4 |
| Lymphography | 148 | 18 | 4 |
| Spect Heart | 267 | 22 | 2 |
| Tic-Tac-Toe Endgame | 958 | 9 | 2 |

# Certain case

| Database | PCC bef.Prun | PCC aft.Pr.Prun | PCC aft.Pt.Prun |
|---|---|---|---|
| Zoo | 85.21% | 86.74% | 90.08% |
| Nursery | 93.66% | 94.14% | 96.07% |

# Certain case

| Database | PCC bef.Prun | PCC aft.Pr.Prun | PCC aft.Pt.Prun |
|---|---|---|---|
| Zoo | 85.21% | 86.74% | 90.08% |
| Nursery | 93.66% | 94.14% | 96.07% |

| Database | Dist bef.Prun | Dist aft.PrPrun | Dist aft.Pt Prun |
|---|---|---|---|
| Zoo | 0.31 | 0.29 | 0.26 |
| Nursery | 0.26 | 0.25 | 0.23 |

# Certain case

| Database | PCC bef.Prun | PCC aft.Pr.Prun | PCC aft.Pt.Prun |
|---|---|---|---|
| Zoo | 85.21% | 86.74% | 90.08% |
| Nursery | 93.66% | 94.14% | 96.07% |

| Database | Dist bef.Prun | Dist aft.PrPrun | Dist aft.Pt Prun |
|---|---|---|---|
| Zoo | 0.31 | 0.29 | 0.26 |
| Nursery | 0.26 | 0.25 | 0.23 |

| Database | Size bef.Prun | Size aft.Pr.Prun | Size aft.Pt.Prun |
|---|---|---|---|
| Zoo | 89 | 53 | 47 |
| Nursery | 398 | 261 | 176 |

# Uncertain case
# Mean PCC

## Zoo Dataset

| Deg. of unc. | PCC.bef Pru_ave | PCC.bef Pru_conj | PCC.aft Pr.Pru_ave | PCC.aft Pr.Pru_conj | PCC.aft Pt.Pru_ave | PCC.aft Pt.Pru_conj |
|---|---|---|---|---|---|---|
| low | 85.09% | 85.12% | 86.78% | 86.81% | 90.65% | 91.94% |
| middle | 84.42% | 84.61% | 85.92% | 86.23% | 89.74% | 91.36% |
| high | 83.26% | 83.43% | 86.05% | 86.11% | 90.02% | 91.41% |
| Mean | 84.25% | 84.38% | 86.25% | 86.38% | 90.13% | 91.57% |

# Uncertain case
# Mean PCC

## Zoo Dataset

| Deg. of unc. | PCC.bef Pru_ave | PCC.bef Pru_conj | PCC.aft Pr.Pru_ave | PCC.aft Pr.Pru_conj | PCC.aft Pt.Pru_ave | PCC.aft Pt.Pru_conj |
|---|---|---|---|---|---|---|
| low | 85.09% | 85.12% | 86.78% | 86.81% | 90.65% | 91.94% |
| middle | 84.42% | 84.61% | 85.92% | 86.23% | 89.74% | 91.36% |
| high | 83.26% | 83.43% | 86.05% | 86.11% | 90.02% | 91.41% |
| Mean | 84.25% | 84.38% | 86.25% | 86.38% | 90.13% | 91.57% |

## Nursery Dataset

| Deg. of unc. | PCC.bef Pru_ave | PCC.bef Pru_conj | PCC.aft Pr.Pru_ave | PCC.aft Pr.Pru_conj | PCC.aft Pt.Pru_ave | PCC.aft Pt.Pru_conj |
|---|---|---|---|---|---|---|
| low | 93.69% | 93.72% | 94.58% | 94.73% | 95.68% | 95.84% |
| middle | 93.01% | 93.18% | 94.19% | 94.26% | 94.92% | 95.13% |
| high | 92.27% | 92.29% | 93.65% | 93.71% | 94.73% | 95.11% |
| Mean | 92.99% | 93.06% | 94.14% | 94.23% | 95.11% | 95.36% |

## Zoo Dataset

| Deg. of unc. | Dist.bef Pru_ave | Dist.bef Pru_conj | Dist.aft Pr.Pru_ave | Dist.aft Pr.Pru_conj | Dist.aft Pt.Pru_ave | Dist.aft Pt.Pru_conj |
|---|---|---|---|---|---|---|
| low | 0.32 | 0.31 | 0.3 | 0.3 | 0.24 | 0.25 |
| middle | 0.32 | 0.3 | 0.29 | 0.28 | 0.25 | 0.25 |
| high | 0.33 | 0.32 | 0.31 | 0.3 | 0.26 | 0.24 |
| Mean | 0.32 | 0.31 | 0.3 | 0.29 | 0.25 | 0.24 |

# Uncertain case
# Mean distance

## Zoo Dataset

| Deg. of unc. | Dist.bef Pru_ave | Dist.bef Pru_conj | Dist.aft Pr.Pru_ave | Dist.aft Pr.Pru_conj | Dist.aft Pt.Pru_ave | Dist.aft Pt.Pru_conj |
|---|---|---|---|---|---|---|
| low | 0.32 | 0.31 | 0.3 | 0.3 | 0.24 | 0.25 |
| middle | 0.32 | 0.3 | 0.29 | 0.28 | 0.25 | 0.25 |
| high | 0.33 | 0.32 | 0.31 | 0.3 | 0.26 | 0.24 |
| Mean | 0.32 | 0.31 | 0.3 | 0.29 | 0.25 | 0.24 |

## Nursery Dataset

| Deg. of unc. | Dist.bef Pru_ave | Dist.bef Pru_conj | Dist.aft Pr.Pru_ave | Dist.aft Pr.Pru_conj | Dist.aft Pt.Pru_ave | Dist.aft Pt.Pru_conj |
|---|---|---|---|---|---|---|
| low | 0.27 | 0.27 | 0.25 | 0.22 | 0.2 | 0.2 |
| middle | 0.28 | 0.27 | 0.25 | 0.24 | 0.21 | 0.21 |
| high | 0.28 | 0.26 | 0.25 | 0.24 | 0.22 | 0.21 |
| Mean | 0.27 | 0.25 | 0.25 | 0.23 | 0.21 | 0.2 |

## Zoo Dataset

| Deg. of unc. | Size bef Pru_ave | Size bef Pru_conj | Size aft Pr.Pru_ave | Size aft Pr.Pru_conj | Size aft Pt.Pru_ave | Size aft Pt.Pru_conj |
|---|---|---|---|---|---|---|
| low | 95 | 97 | 63 | 69 | 57 | 61 |
| middle | 98 | 102 | 71 | 73 | 58 | 57 |
| high | 113 | 106 | 76 | 75 | 58 | 59 |
| Mean | 102 | 101 | 70 | 72 | 57 | 59 |

# Uncertain case
# Mean size

## Zoo Dataset

| Deg. of unc. | Size bef Pru_ave | Size bef Pru_conj | Size aft Pr.Pru_ave | Size aft Pr.Pru_conj | Size aft Pt.Pru_ave | Size aft Pt.Pru_conj |
|---|---|---|---|---|---|---|
| low | 95 | 97 | 63 | 69 | 57 | 61 |
| middle | 98 | 102 | 71 | 73 | 58 | 57 |
| high | 113 | 106 | 76 | 75 | 58 | 59 |
| Mean | 102 | 101 | 70 | 72 | 57 | 59 |

## Nursery Dataset

| Deg. of unc. | Size.bef Pru_ave | Size.bef Pru_conj | Size.aft Pr.Pru_ave | Size.aft Pr.Pru_conj | Size.aft Pt.Pru_ave | Size.aft Pt.Pru_conj |
|---|---|---|---|---|---|---|
| low | 456 | 463 | 224 | 230 | 178 | 192 |
| middle | 453 | 461 | 206 | 213 | 171 | 184 |
| high | 471 | 469 | 237 | 235 | 166 | 187 |
| Mean | 460 | 464 | 222 | 226 | 171 | 187 |

# References


- Zied Elouedi, Khaled Mellouli, Philippe Smets
  Belief decision trees: theoretical foundations
  International Journal of Approximate Reasoning, IJAR 28, 91-124, 2001.

- Zied Elouedi, Khaled Mellouli, Philippe Smets
  A pre-pruning method in belief decision trees
  IPMU 2002, Vol I, 579-586, 2002.

- Salsabil Trabelsi, Zied Elouedi, Khaled Mellouli
  Pruning belief decision tree methods in averaging and conjunctive approaches
  International Journal of Approximate Reasoning, IJAR, 46, 568-595, 2007.

# References for related works


- Patrick Vannoorenberghe, Thierry Denoeux
  Handling uncertain labels in multiclass problems using belief decision trees
  IPMU 2002, 1916-1926, 2002.
- Patrick Vannoorenberghe
  On aggregating belief decision trees
  Information Fusion 5 (2) 179-188, 2004.
- Huawei Guo, Wenkang Shi, Feng Du
  EDTs: Evidential Decision Trees
  Fuzzy Systems and Knowledge Discovery, LNCS 3614 398-407, 2005
- John Ross Quinlan
  Decision trees as probabilistic classifiers
  The Fourth International on Machine Learning, Morgan Kaufman, 31-37, 1987.
- Yufei Yuan, Michael J. Shawb
  Induction of fuzzy decision trees
  Fuzzy Sets and Systems, FSS 69, 125-139, 1995.
- Ilyes Jenhani, Nahla Ben Amor, Zied Elouedi
  Decision trees as possibilistic classifiers
  International Journal of Approximate Reasoning, IJAR, 48(3), 784-807, 2008.
- etc.

# *Belief Rough Set Classifier*

# Rough set theory

Rough set theory constitutes a sound basis for data mining as a tool to:

# Rough set theory

Rough set theory constitutes a sound basis for data mining as a tool to:

- Discretization.

# Rough set theory

Rough set theory constitutes a sound basis for data mining as a tool to:

- Discretization.
- Feature selection.

# Rough set theory

Rough set theory constitutes a sound basis for data mining as a tool to:

- Discretization.
- Feature selection.
- Decision rule generation.

# Rough set theory

Rough set theory constitutes a sound basis for data mining as a tool to:

- Discretization.
- Feature selection.
- Decision rule generation.
- Classification.

# Rough set classifier

Rough set classifier is a successful technique of classification.

# Rough set classifier

Rough set classifier is a successful technique of classification.

- Generated models take the form of "IF-THEN" rules.

# Rough set classifier

Rough set classifier is a successful technique of classification.

- Generated models take the form of "IF-THEN" rules.
- Data reduction before generation rules.

# Rough set classifier

Rough set classifier is a successful technique of classification.

- Generated models take the form of "IF-THEN" rules.
- Data reduction before generation rules.
- Rapid without many iterations.

# Rough set classifier

Rough set classifier is a successful technique of classification.

- Generated models take the form of "IF-THEN" rules.
- Data reduction before generation rules.
- Rapid without many iterations.
- Optimal and minimal set of generating rules.

- Information Table and Decision Table

# Basics

- Information Table and Decision Table
- Indiscernibility relation

# Basics

- Information Table and Decision Table
- Indiscernibility relation
- Set approximation

# Basics

- Information Table and Decision Table
- Indiscernibility relation
- Set approximation
- Reduct and Core

# Basics

- Information Table and Decision Table
- Indiscernibility relation
- Set approximation
- Reduct and Core
- Reduct value and Core value

# Basics

- Information Table and Decision Table
- Indiscernibility relation
- Set approximation
- Reduct and Core
- Reduct value and Core value
- Decision rules

# Information Table

$A = (U, C)$

$U$ is a non-empty, finite set of objects.

$C$ is a non-empty, finite set of attributes.

| U | Headache | Muscle-pain | Temperature |
|---|---|---|---|
| $o_1$ | no | yes | high |
| $o_2$ | yes | no | high |
| $o_3$ | yes | yes | very high |
| $o_4$ | no | yes | normal |
| $o_5$ | yes | no | high |
| $o_6$ | no | yes | very high |

# Decision Table

$DS = (U, C \cup \{d\})$

The elements of C are called the condition attributes.

$d \notin C$ is the decision attribute.

| U | Headache | Muscle-pain | Temperature | Flu |
|---|---|---|---|---|
| $o_1$ | no | yes | high | yes |
| $o_2$ | yes | no | high | yes |
| $o_3$ | yes | yes | very high | yes |
| $o_4$ | no | yes | normal | no |
| $o_5$ | yes | no | high | no |
| $o_6$ | no | yes | very high | yes |

# Decision Table

$DS = (U, C \cup \{d\})$

The elements of C are called the condition attributes.

$d \notin C$ is the decision attribute.

| U | Headache | Muscle-pain | Temperature | Flu |
|---|---|---|---|---|
| $o_1$ | no | yes | high | yes |
| $o_2$ | yes | no | high | yes |
| $o_3$ | yes | yes | very high | yes |
| $o_4$ | no | yes | normal | no |
| $o_5$ | yes | no | high | no |
| $o_6$ | no | yes | very high | yes |

## Issues in the decision table

- Same objects may be represented several times.
- Some of the attributes may be superfluous.
- Some of attribute values may be superfluous.

# Indiscernibility relation

The indiscernibility relation is based on a subset of the condition attributes $B$

$$IND_B = U/B = \{[o_j]_B | o_j \in U)\}$$
where $[o_j]_B = \{o_i | B(o_i) = B(o_j)\}$

# Indiscernibility relation

The indiscernibility relation is based on a subset of the condition attributes $B$

$$IND_B = U/B = \{[o_j]_B | o_j \in U)\}$$
where $[o_j]_B = \{o_i | B(o_i) = B(o_j)\}$

| U | Headache | Muscle-pain | Temperature | Flu |
|---|---|---|---|---|
| $o_1$ | no | yes | high | yes |
| $o_2$ | yes | no | high | yes |
| $o_3$ | yes | yes | very high | yes |
| $o_4$ | no | yes | normal | no |
| $o_5$ | yes | no | high | no |
| $o_6$ | no | yes | very high | yes |

# Indiscernibility relation

The indiscernibility relation is based on a subset of the condition attributes $B$

$$IND_B = U/B = \{[o_j]_B | o_j \in U)\}$$
where $[o_j]_B = \{o_i | B(o_i) = B(o_j)\}$

| U | Headache | Muscle-pain | Temperature | Flu |
|---|---|---|---|---|
| $o_1$ | no | yes | high | yes |
| $o_2$ | yes | no | high | yes |
| $o_3$ | yes | yes | very high | yes |
| $o_4$ | no | yes | normal | no |
| $o_5$ | yes | no | high | no |
| $o_6$ | no | yes | very high | yes |

$IND_{\{Headache\}} = \{\{o_2, o_3, o_5\}, \{o_1, o_4, o_6\}\}$

# Indiscernibility relation

The indiscernibility relation is based on a subset of the condition attributes $B$

$$IND_B = U/B = \{[o_j]_B | o_j \in U)\}$$
where $[o_j]_B = \{o_i | B(o_i) = B(o_j)\}$

| U | Headache | Muscle-pain | Temperature | Flu |
|---|---|---|---|---|
| $o_1$ | no | yes | high | yes |
| $o_2$ | yes | no | high | yes |
| $o_3$ | yes | yes | very high | yes |
| $o_4$ | no | yes | normal | no |
| $o_5$ | yes | no | high | no |
| $o_6$ | no | yes | very high | yes |

$IND_{\{Headache\}} = \{\{o_2, o_3, o_5\}, \{o_1, o_4, o_6\}\}$
$IND_{\{Headache, Muscle-pain\}} = \{\{o_1, o_4, o_6, \}, \{o_2, o_5\}, \{o_3\}\}$

# Indiscernibility relation

The indiscernibility relation is based on a subset of the condition attributes $B$

$$IND_B = U/B = \{[o_j]_B | o_j \in U)\}$$
where $[o_j]_B = \{o_i | B(o_i) = B(o_j)\}$

| U | Headache | Muscle-pain | Temperature | Flu |
|---|---|---|---|---|
| $o_1$ | no | yes | high | yes |
| $o_2$ | yes | no | high | yes |
| $o_3$ | yes | yes | very high | yes |
| $o_4$ | no | yes | normal | no |
| $o_5$ | yes | no | high | no |
| $o_6$ | no | yes | very high | yes |

$IND_{\{Headache\}} = \{\{o_2, o_3, o_5\}, \{o_1, o_4, o_6\}\}$
$IND_{\{Headache, Muscle-pain\}} = \{\{o_1, o_4, o_6, \}, \{o_2, o_5\}, \{o_3\}\}$
$IND_{\{C\}} = \{\{o_1\}, \{o_2, o_5\}, \{o_3\}, \{o_4\}, \{o_6\}\}$

# Set approximation

Let $B \subseteq C$ and $X \subseteq U$.

# Set approximation

Let $B \subseteq C$ and $X \subseteq U$.

- The lower approximation : $\underline{B}(X) = \{o_j | [o_j]_B \subseteq X\}$

# Set approximation

Let $B \subseteq C$ and $X \subseteq U$.

- The lower approximation : $\underline{B}(X) = \{o_j | [o_j]_B \subseteq X\}$
- The upper approximation : $\bar{B}(X) = \{o_j | [o_j]_B \cap X \neq \emptyset\}$

# Set approximation

Let $B \subseteq C$ and $X \subseteq U$.

- The lower approximation : $\underline{\mathrm{B}}(X) = \{o_j | [o_j]_B \subseteq X\}$
- The upper approximation : $\bar{\mathrm{B}}(X) = \{o_j | [o_j]_B \cap X \neq \emptyset\}$
- The Boundary region : $BN_B(X) = \bar{\mathrm{B}}(\mathrm{X}) - \underline{\mathrm{B}}(\mathrm{X})$

# Set approximation

Let $B \subseteq C$ and $X \subseteq U$.

- The lower approximation : $\underline{\mathrm{B}}(X) = \{o_j | [o_j]_B \subseteq X\}$
- The upper approximation : $\bar{\mathrm{B}}(X) = \{o_j | [o_j]_B \cap X \neq \emptyset\}$
- The Boundary region : $BN_B(X) = \bar{\mathrm{B}}(\mathrm{X}) - \underline{\mathrm{B}}(\mathrm{X})$

A set is said to be rough if its boundary region is non-empty.

# Set approximation

Let $B \subseteq C$ and $X \subseteq U$.

- The lower approximation : $\underline{B}(X) = \{o_j | [o_j]_B \subseteq X\}$
- The upper approximation : $\bar{B}(X) = \{o_j | [o_j]_B \cap X \neq \emptyset\}$
- The Boundary region : $BN_B(X) = \bar{B}(X) - \underline{B}(X)$

A set is said to be rough if its boundary region is non-empty.

| U | Headache | Muscle-pain | Temperature | Flu |
|---|---|---|---|---|
| $o_1$ | no | yes | high | yes |
| $o_2$ | yes | no | high | yes |
| $o_3$ | yes | yes | very high | yes |
| $o_4$ | no | yes | normal | no |
| $o_5$ | yes | no | high | no |
| $o_6$ | no | yes | very high | yes |

# Set approximation

Let $B \subseteq C$ and $X \subseteq U$.

- The lower approximation : $\underline{B}(X) = \{o_j | [o_j]_B \subseteq X\}$
- The upper approximation : $\bar{B}(X) = \{o_j | [o_j]_B \cap X \neq \emptyset\}$
- The Boundary region : $BN_B(X) = \bar{B}(X) - \underline{B}(X)$

A set is said to be rough if its boundary region is non-empty.

| U | Headache | Muscle-pain | Temperature | Flu |
|---|---|---|---|---|
| $o_1$ | no | yes | high | yes |
| $o_2$ | yes | no | high | yes |
| $o_3$ | yes | yes | very high | yes |
| $o_4$ | no | yes | normal | no |
| $o_5$ | yes | no | high | no |
| $o_6$ | no | yes | very high | yes |

Let $X_1 = \{o_j | Flu(o_j) = \text{yes}\}$

- $\underline{C}(X_1) = \{o_1, o_3, o_6\}$
- $\bar{C}(X_1) = \{o_1, o_2, o_3, o_5, o_6\}$
- $BN_C(X_1) = \{o_2, o_5\}$

# Set approximation

Let $B \subseteq C$ and $X \subseteq U$.

- The lower approximation : $\underline{B}(X) = \{o_j | [o_j]_B \subseteq X\}$
- The upper approximation : $\bar{B}(X) = \{o_j | [o_j]_B \cap X \neq \emptyset\}$
- The Boundary region : $BN_B(X) = \bar{B}(X) - \underline{B}(X)$

A set is said to be rough if its boundary region is non-empty.

| U | Headache | Muscle-pain | Temperature | Flu |
|---|---|---|---|---|
| $o_1$ | no | yes | high | yes |
| $o_2$ | yes | no | high | yes |
| $o_3$ | yes | yes | very high | yes |
| $o_4$ | no | yes | normal | no |
| $o_5$ | yes | no | high | no |
| $o_6$ | no | yes | very high | yes |

Let $X_1 = \{o_j | Flu(o_j) = \text{yes}\}$

- $\underline{C}(X_1) = \{o_1, o_3, o_6\}$
- $\bar{C}(X_1) = \{o_1, o_2, o_3, o_5, o_6\}$
- $BN_C(X_1) = \{o_2, o_5\}$

Let $X_2 = \{o_j | Flu(o_j) = \text{no}\}$

- $\underline{C}(X_2) = \{o_4\}$
- $\bar{C}(X_2) = \{o_2, o_4, o_5\}$
- $BN_C(X_2) = \{o_2, o_5\}$

# Reduct and core

## Reduct

It is a minimal subset of attributes B that enables the same classification of elements as the whole set of attributes C does.



The attributes, not in reduct, are considered as superfluous.

# Reduct and core

## Reduct

It is a minimal subset of attributes B that enables the same classification of elements as the whole set of attributes C does.



The attributes, not in reduct, are considered as superfluous.

## Core

- It is a set of indispensable attributes.
- It is included in every reduct.



The attributes in core cannot be removed from the decision table.

# Reduct and core

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | no | high | yes |
| $o_2$ | yes | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | no | normal | no |
| $o_5$ | yes | high | no |
| $o_6$ | no | very high | yes |

# Reduct and core

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | no | high | yes |
| $o_2$ | yes | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | no | normal | no |
| $o_5$ | yes | high | no |
| $o_6$ | no | very high | yes |

Reduct 2 = {Muscle-pain, Temperature}

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | high | yes |
| $o_2$ | no | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | yes | normal | no |
| $o_5$ | no | high | no |
| $o_6$ | yes | very high | yes |

# Reduct and core

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | no | high | yes |
| $o_2$ | yes | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | no | normal | no |
| $o_5$ | yes | high | no |
| $o_6$ | no | very high | yes |

Reduct 2 = {Muscle-pain, Temperature}

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | high | yes |
| $o_2$ | no | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | yes | normal | no |
| $o_5$ | no | high | no |
| $o_6$ | yes | very high | yes |

$$Core = \{Temperature\}$$

# Reduct value and core value

## Reduct value

It is a minimal subset of attributes $B$ from $C$ for the object $o_j$ that preserves the partitioning of the universe.



The attribute values, not in reduct values, can be removed.

# Reduct value and core value

## Reduct value

It is a minimal subset of attributes $B$ from $C$ for the object $o_j$ that preserves the partitioning of the universe.



The attribute values, not in reduct values, can be removed.

## Core value

- It is the set of indispensable attribute values for the object $o_j$.
- It is included in every reduct value.

# Reduct value and core value

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | no | high | yes |
| $o_2$ | yes | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | no | normal | no |
| $o_5$ | yes | high | no |
| $o_6$ | no | very high | yes |

Reduct 2 = {Muscle-pain, Temperature}

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | high | yes |
| $o_2$ | no | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | yes | normal | no |
| $o_5$ | no | high | no |
| $o_6$ | yes | very high | yes |

# Reduct value and core value

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | no | high | yes |
| $o_2$ | yes | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | no | normal | no |
| $o_5$ | yes | high | no |
| $o_6$ | no | very high | yes |

Reduct values 1

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | no | high | yes |
| $o_2$ | yes | high | yes |
| $o_3$ | - | very high | yes |
| $o_4$ | - | normal | no |
| $o_5$ | yes | high | no |
| $o_6$ | - | very high | yes |

Reduct 2 = {Muscle-pain, Temperature}

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | high | yes |
| $o_2$ | no | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | yes | normal | no |
| $o_5$ | no | high | no |
| $o_6$ | yes | very high | yes |

# Reduct value and core value

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | no | high | yes |
| $o_2$ | yes | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | no | normal | no |
| $o_5$ | yes | high | no |
| $o_6$ | no | very high | yes |

Reduct 2 = {Muscle-pain, Temperature}

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | high | yes |
| $o_2$ | no | high | yes |
| $o_3$ | yes | very high | yes |
| $o_4$ | yes | normal | no |
| $o_5$ | no | high | no |
| $o_6$ | yes | very high | yes |

Reduct values 1

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | no | high | yes |
| $o_2$ | yes | high | yes |
| $o_3$ | - | very high | yes |
| $o_4$ | - | normal | no |
| $o_5$ | yes | high | no |
| $o_6$ | - | very high | yes |

Reduct values 2

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | high | yes |
| $o_2$ | no | high | yes |
| $o_3$ | - | very high | yes |
| $o_4$ | - | normal | no |
| $o_5$ | no | high | no |
| $o_6$ | - | very high | yes |

Reduct values 1

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | no | high | yes |
| $o_2$ | yes | high | yes |
| $o_3$, $o_6$ | - | very high | yes |
| $o_4$ | - | normal | no |
| $o_5$ | yes | high | no |

# Decision rules

Reduct values 1

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | no | high | yes |
| $o_2$ | yes | high | yes |
| $o_3$, $o_6$ | - | very high | yes |
| $o_4$ | - | normal | no |
| $o_5$ | yes | high | no |

- If Headache=no and Temperature=high then Flu=yes
- If Headache=yes and Temperature=high then Flu=yes S=1/2
- If Temperature=very high then Flu= yes
- If Temperature=normal then Flu= no

## Reduct values 2

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | high | yes |
| $o_2$ | no | high | yes |
| $o_3$, $o_6$ | - | very high | yes |
| $o_4$ | - | normal | no |
| $o_5$ | no | high | no |

# Decision rules

## Reduct values 2

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | high | yes |
| $o_2$ | no | high | yes |
| $o_3$, $o_6$ | - | very high | yes |
| $o_4$ | - | normal | no |
| $o_5$ | no | high | no |

- If Muscle-pain=yes and Temperature=high then Flu=yes
- If Muscle-pain=no and Temperature=high then Flu=yes S=1/2
- If Temperature=very high then Flu= yes
- If Temperature=normal then Flu= no

# Motivation

Belief Decision Tree (BDT): is a supervised classification technique from uncertain environment under belief function framework.

# Motivation

Belief Decision Tree (BDT): is a supervised classification technique from uncertain environment under belief function framework.

- Very large trees
- In some cases weak PCC

# Motivation

Belief Decision Tree (BDT): is a supervised classification technique from uncertain environment under belief function framework.

- Very large trees
- In some cases weak PCC

- Pre-pruning BDT
- Post-pruning BDT

# Motivation

Belief Decision Tree (BDT): is a supervised classification technique from uncertain environment under belief function framework.

- Very large trees
- In some cases weak PCC

- Pre-pruning BDT
- Post-pruning BDT



## Two cost tasks:

- Construction of the belief decision tree
- Pruning of the belief decision tree

# Belief rough set classifier

## Objective

Rough Set Theory + Belief function theory

## Objective

Rough Set Theory + Belief function theory

$\Rightarrow$ Belief Rough Set Classifier

## Objective

Rough Set Theory + Belief function theory

$\Rightarrow$ Belief Rough Set Classifier

- Dealing with uncertainty in objects' classes in the training phase.

# Uncertain Decision Table

## Uncertain decision table

| U | Headache | Muscle-pain | Temperature | Flu |
|---|---|---|---|---|
| $o_1$ | yes | yes | very high | $m_1(\{yes\}) = 0.95 \quad m_1(\{no\}) = 0.05$ |
| $o_2$ | yes | no | high | $m_2(\{no\}) = 1$ |
| $o_3$ | yes | yes | normal | $m_3(\{yes\}) = 0.5 \quad m_3(\Theta) = 0.5$ |
| $o_4$ | no | yes | normal | $m_4(\{no\}) = 0.6 \quad m_4(\Theta) = 0.4$ |
| $o_5$ | no | yes | normal | $m_5(\{no\}) = 1$ |
| $o_6$ | yes | no | high | $m_6(\{no\}) = 0.95 \quad m_6(\Theta) = 0.05$ |
| $o_7$ | no | yes | very high | $m_7(\{yes\}) = 1$ |
| $o_8$ | no | yes | high | $m_8(\{yes\}) = 0.9 \; m_8(\Theta) = 0.1$ |

# Tolerance relation

- Indiscernibility relation for the condition attributes is the same as in the certain case.
- Indiscernibility relation for the decision attribute is not the same.

# Tolerance relation

- Indiscernibility relation for the condition attributes is the same as in the certain case.
- Indiscernibility relation for the decision attribute is not the same.

- For every uncertain decision value $ud_i$.

$$X_i = \{o_j | dist(m, m_j) < 1 - threshold\}, such\ that\ m(\{ud_i\}) = 1$$

- Tolerance relation for the decision attribute

$$Tol_{\{ud\}} = U/\{ud\} = \{X_i | ud_i \in \Theta\}$$

# Tolerance relation

| $m_j$ | $dist(m_j, m)$ such that m({yes})=1 | $dist(m_j, m)$ such that m({no})=1 |
|---|---|---|
| $m_1$ | $0.08 < 0.9$ | $0.92 > 0.9$ |
| $m_2$ | $1.00 > 0.9$ | $0.00 < 0.9$ |
| $m_3$ | $0.66 < 0.9$ | $0.34 < 0.9$ |
| $m_4$ | $0.84 < 0.9$ | $0.26 < 0.9$ |
| $m_5$ | $1.00 > 0.9$ | $0.00 < 0.9$ |
| $m_6$ | $0.93 > 0.9$ | $0.07 < 0.9$ |
| $m_7$ | $0.00 < 0.9$ | $1.00 > 0.9$ |
| $m_8$ | $0.08 < 0.9$ | $0.92 > 0.9$ |

# Tolerance relation

| $m_j$ | $dist(m_j, m)$ such that m({yes})=1 | $dist(m_j, m)$ such that m({no})=1 |
|---|---|---|
| $m_1$ | $0.08 < 0.9$ | $0.92 > 0.9$ |
| $m_2$ | $1.00 > 0.9$ | $0.00 < 0.9$ |
| $m_3$ | $0.66 < 0.9$ | $0.34 < 0.9$ |
| $m_4$ | $0.84 < 0.9$ | $0.26 < 0.9$ |
| $m_5$ | $1.00 > 0.9$ | $0.00 < 0.9$ |
| $m_6$ | $0.93 > 0.9$ | $0.07 < 0.9$ |
| $m_7$ | $0.00 < 0.9$ | $1.00 > 0.9$ |
| $m_8$ | $0.08 < 0.9$ | $0.92 > 0.9$ |

$X_1 = \{o_1, o_3, o_4, o_7, o_8\}$
$X_2 = \{o_2, o_3, o_4, o_5, o_6\}$
$Tol_{\{ud\}} = U/\{ud\} = \{\{o_1, o_3, o_4, o_7, o_8\}, \{o_2, o_3, o_4, o_5, o_6\}\}$

# Set approximation

- To compute the new lower and upper approximations:
  1. For each equivalence class from $U/C$, combine their bba using mean operator.
  2. For each equivalence class from $U/ud$, compute the lower and upper approximations:

$$\underline{C}(X_i) = \{o_j | [o_j]_C \subseteq X_i \textit{ and } \textit{dist}(\bar{m}_{[o_j]_C}, m) \leq \textit{threshold}\}$$

$$\bar{C}(X_i) = \{o_j | [o_j]_C \cap X_i \neq \emptyset\}$$

# Set approximation

| $\bar{m}_{[o_j]_C}$ | $dist(\bar{m}_{[o_j]_C}, m)$ *such that* $m(\{yes\}) = 1)$ |
|---|---|
| $m_1$ | $0.07 < 0.1$ |
| $m_3$ | $0.66 > 0.1$ |
| $m_7$ | $0.00 < 0.1$ |
| $m_8$ | $0.07 < 0.1$ |
| $\bar{m}_{[o_j]_C}$ | $dist(\bar{m}_{[o_j]_C}, m)$ *such that* $m(\{no\}) = 1)$ |
| $m_{2,6}$ | $0.02 < 0.1$ |
| $m_3$ | $0.34 > 0.1$ |
| $m_{4,5}$ | $0.26 > 0.1$ |

**$X_1 = \{o_1, o_3, o_4, o_7, o_8\}$**
**$ud_1 = yes$**

$\underline{C}(X_1)=\{o_1, o_7, o_8\}$

$\bar{C}(X_1)=\{o_1, o_3, o_4, o_5, o_7, o_8\}$

$BN_C(X_1)=\{o_3, o_4, o_5\}$

**$X_2 = \{o_2, o_3, o_4, o_5, o_6\}$**
**$ud_2 = no$**

$\underline{C}(X_2)=\{o_2, o_6\}$

$\bar{C}(X_2)=\{o_2, o_3, o_4, o_5, o_6\}$

$BN_C(X_2)=\{o_3, o_4, o_5\}$

# Set approximation

| $\bar{m}_{[o_j]_C}$ | $dist(\bar{m}_{[o_j]_C}, m)$ *such that* $m(\{yes\}) = 1)$ |
|---|---|
| $m_1$ | $0.07 < 0.1$ |
| $m_3$ | $0.66 > 0.1$ |
| $m_7$ | $0.00 < 0.1$ |
| $m_8$ | $0.07 < 0.1$ |
| $\bar{m}_{[o_j]_C}$ | $dist(\bar{m}_{[o_j]_C}, m)$ *such that* $m(\{no\}) = 1)$ |
| $m_{2,6}$ | $0.02 < 0.1$ |
| $m_3$ | $0.34 > 0.1$ |
| $m_{4,5}$ | $0.26 > 0.1$ |

**$X_1 = \{o_1, o_3, o_4, o_7, o_8\}$**
**$ud_1 = yes$**

$\underline{C}(X_1)=\{o_1, o_7, o_8\}$

$\bar{C}(X_1)=\{o_1, o_3, o_4, o_5, o_7, o_8\}$

$BN_C(X_1)=\{o_3, o_4, o_5\}$

**$X_2 = \{o_2, o_3, o_4, o_5, o_6\}$**
**$ud_2 = no$**

$\underline{C}(X_2)=\{o_2, o_6\}$

$\bar{C}(X_2)=\{o_2, o_3, o_4, o_5, o_6\}$

$BN_C(X_2)=\{o_3, o_4, o_5\}$

Uncertain decision attribute, Flu, is rough.

# Reduct and core

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | yes | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | no | normal | $m_4$ |
| $o_5$ | no | normal | $m_5$ |
| $o_6$ | yes | high | $m_6$ |
| $o_7$ | no | very high | $m_7$ |
| $o_8$ | no | high | $m_8$ |

# Reduct and core

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | yes | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | no | normal | $m_4$ |
| $o_5$ | no | normal | $m_5$ |
| $o_6$ | yes | high | $m_6$ |
| $o_7$ | no | very high | $m_7$ |
| $o_8$ | no | high | $m_8$ |

Reduct 2 = {Muscle-pain, Temperature}

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | no | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | yes | normal | $m_4$ |
| $o_5$ | yes | normal | $m_5$ |
| $o_6$ | no | high | $m_6$ |
| $o_7$ | yes | very high | $m_7$ |
| $o_8$ | yes | high | $m_8$ |

# Reduct and core

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | yes | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | no | normal | $m_4$ |
| $o_5$ | no | normal | $m_5$ |
| $o_6$ | yes | high | $m_6$ |
| $o_7$ | no | very high | $m_7$ |
| $o_8$ | no | high | $m_8$ |

Reduct 2 = {Muscle-pain, Temperature}

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | no | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | yes | normal | $m_4$ |
| $o_5$ | yes | normal | $m_5$ |
| $o_6$ | no | high | $m_6$ |
| $o_7$ | yes | very high | $m_7$ |
| $o_8$ | yes | high | $m_8$ |

$$Core = \{Temperature\}$$

# Reduct value and core value

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | yes | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | no | normal | $m_4$ |
| $o_5$ | no | normal | $m_5$ |
| $o_6$ | yes | high | $m_6$ |
| $o_7$ | no | very high | $m_7$ |
| $o_8$ | no | high | $m_8$ |

Reduct 2 = {Muscle-pain, Temperature}

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | no | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | yes | normal | $m_4$ |
| $o_5$ | yes | normal | $m_5$ |
| $o_6$ | no | high | $m_6$ |
| $o_7$ | yes | very high | $m_7$ |
| $o_8$ | yes | high | $m_8$ |

# Reduct value and core value

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | yes | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | no | normal | $m_4$ |
| $o_5$ | no | normal | $m_5$ |
| $o_6$ | yes | high | $m_6$ |
| $o_7$ | no | very high | $m_7$ |
| $o_8$ | no | high | $m_8$ |

Reduct 2 = {Muscle-pain, Temperature}

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | no | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | yes | normal | $m_4$ |
| $o_5$ | yes | normal | $m_5$ |
| $o_6$ | no | high | $m_6$ |
| $o_7$ | yes | very high | $m_7$ |
| $o_8$ | yes | high | $m_8$ |

Reduct values 1

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | - | very high | $m_1$ |
| $o_2$ | - | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | no | normal | $m_4$ |
| $o_5$ | no | normal | $m_5$ |
| $o_6$ | - | high | $m_6$ |
| $o_7$ | - | very high | $m_7$ |
| $o_8$ | - | high | $m_8$ |

# Reduct value and core value

Reduct 1 = {Headache, Temperature}

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | yes | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | no | normal | $m_4$ |
| $o_5$ | no | normal | $m_5$ |
| $o_6$ | yes | high | $m_6$ |
| $o_7$ | no | very high | $m_7$ |
| $o_8$ | no | high | $m_8$ |

Reduct 2 = {Muscle-pain, Temperature}

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | yes | very high | $m_1$ |
| $o_2$ | no | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | yes | normal | $m_4$ |
| $o_5$ | yes | normal | $m_5$ |
| $o_6$ | no | high | $m_6$ |
| $o_7$ | yes | very high | $m_7$ |
| $o_8$ | yes | high | $m_8$ |

Reduct values 1

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$ | - | very high | $m_1$ |
| $o_2$ | - | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | no | normal | $m_4$ |
| $o_5$ | no | normal | $m_5$ |
| $o_6$ | - | high | $m_6$ |
| $o_7$ | - | very high | $m_7$ |
| $o_8$ | - | high | $m_8$ |

Reduct values 2

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$ | - | very high | $m_1$ |
| $o_2$ | - | high | $m_2$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$ | yes | normal | $m_4$ |
| $o_5$ | yes | normal | $m_5$ |
| $o_6$ | - | high | $m_6$ |
| $o_7$ | - | very high | $m_7$ |
| $o_8$ | - | high | $m_8$ |

# Decision rules

## Reduct values 1

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$, $o_7$ | - | very high | $m_{1,7}$ |
| $o_2$, $o_6$, $o_8$ | - | high | $m_{2,6,8}$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$, $o_5$ | no | normal | $m_{4,5}$ |

# Decision rules

## Reduct values 1

| U | Headache | Temperature | Flu |
|---|---|---|---|
| $o_1$, $o_7$ | - | very high | $m_{1,7}$ |
| $o_2$, $o_6$, $o_8$ | - | high | $m_{2,6,8}$ |
| $o_3$ | yes | normal | $m_3$ |
| $o_4$, $o_5$ | no | normal | $m_{4,5}$ |

- If Temperature=very high then $m_{1,7}$
- If Temperature=high then $m_{2,6,8}$
- If Headache=yes and Temperature=normal then $m_3$
- If Headache=no and Temperature=high then $m_{4,5}$

## Reduct values 2

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1, o_7$ | - | very high | $m_{1,7}$ |
| $o_2, o_6, o_8$ | - | high | $m_{2,6,8}$ |
| $o_3, o_4, o_5$ | yes | normal | $m_{3,4,5}$ |

## Reduct values 2

| U | Muscle-pain | Temperature | Flu |
|---|---|---|---|
| $o_1$, $o_7$ | - | very high | $m_{1,7}$ |
| $o_2$, $o_6$, $o_8$ | - | high | $m_{2,6,8}$ |
| $o_3$, $o_4$, $o_5$ | yes | normal | $m_{3,4,5}$ |

- If Temperature=very high then $m_{1,7}$
- If Temperature=high then $m_{2,6,8}$
- If Muscle-pain=yes and Temperature=normal then $m_{3,4,5}$

# Belief rough set classifier

## BRSC

Only based on the redefined basic concepts of rough sets under uncertainty especially reduct and reduct value.

# Belief rough set classifier

## BRSC

Only based on the redefined basic concepts of rough sets under uncertainty especially reduct and reduct value.

## BRSC-GDT

Derived from the hybrid system GDT-RS which is a combination of the Generalization Distribution Table (GDT) and the Rough Sets (RS) methodology.

# Belief rough set classifier

## BRSC

1. Elimination of the superfluous condition attributes.
2. Elimination of the redundant objects.
3. Elimination of the superfluous condition attribute values.
4. Generation of the decision rules.

# Belief rough set classifier

## BRSC

1. Elimination of the superfluous condition attributes.
2. Elimination of the redundant objects.
3. Elimination of the superfluous condition attribute values.
4. Generation of the decision rules.

## BRSC-GDT

1. Creation of the GDT.
2. Definition of the compound objects.
3. Elimination of the contradictory compound objects.
4. Minimal description length of decision rule.
5. Selection of best rules.

# Heuristic attribute selection method

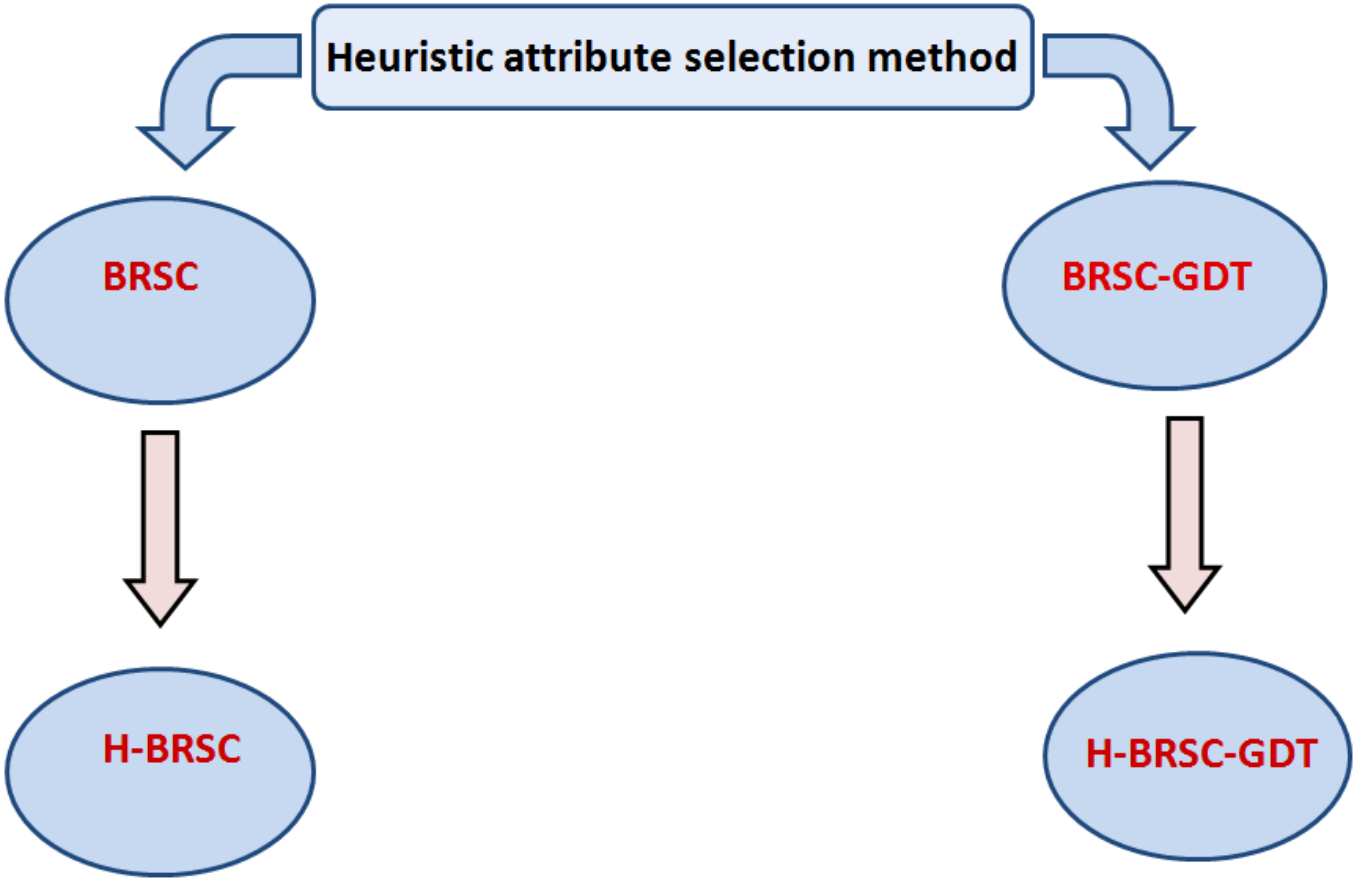
- Finding all possible reducts from uncertain decision table is NP-hard complexity problem.
- The solution is to apply heuristic attribute selection method able to select the relevant features from our uncertain data without costly calculation.
- We choose the heuristic proposed by (Zhong et al., 2001) to adapt it in our context.

# Heuristic attribute selection method

- Finding all possible reducts from uncertain decision table is NP-hard complexity problem.
- The solution is to apply heuristic attribute selection method able to select the relevant features from our uncertain data without costly calculation.
- We choose the heuristic proposed by (Zhong et al., 2001) to adapt it in our context.

- It is fast and generates only one reduct.
- It is better than the other heuristics at trying to avoid local optimum.
- It holds more flexibility with threshold.

# Heuristic attribute selection method



# Dynamic attribute selection method

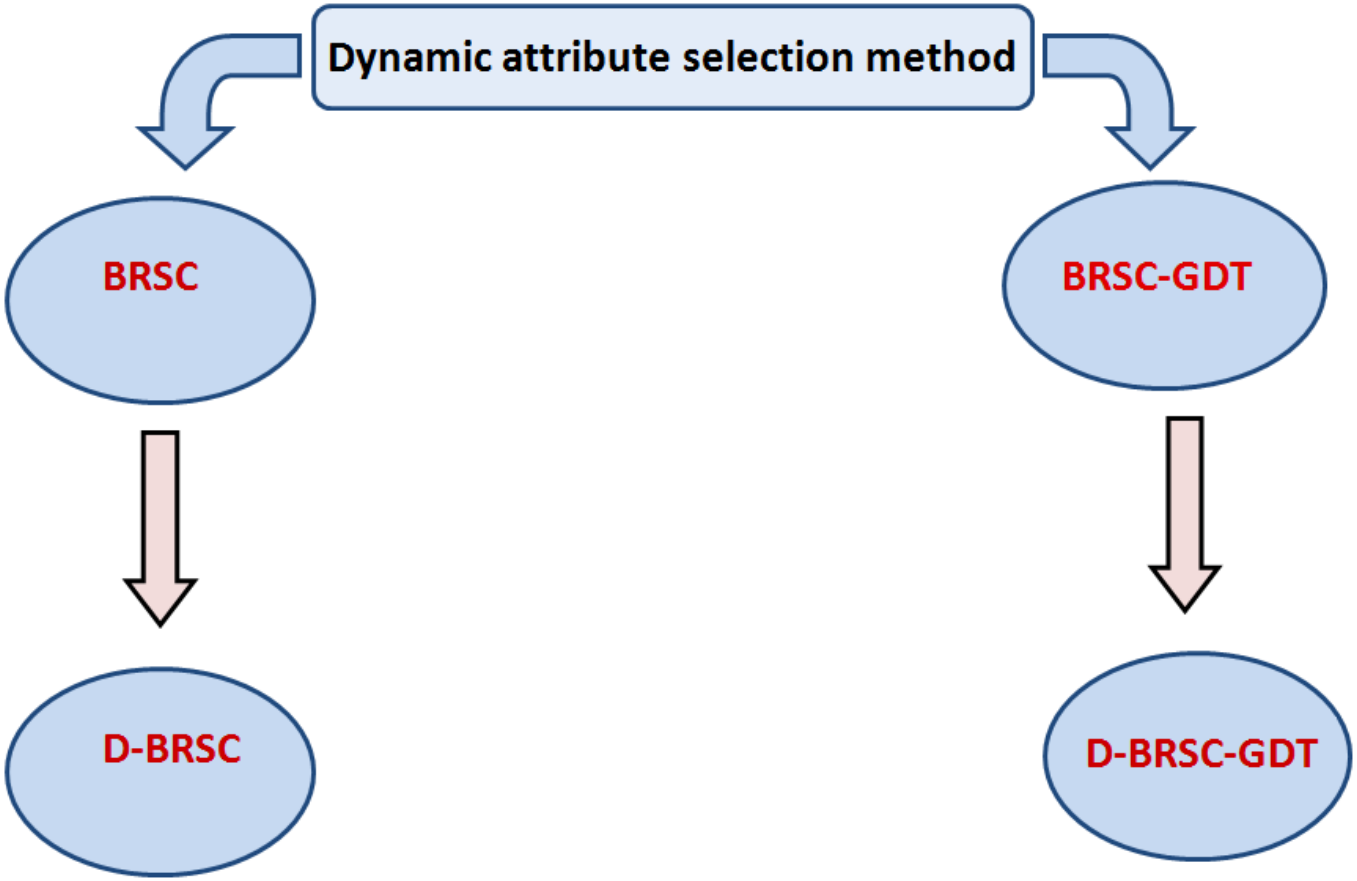
- Computing reduct from uncertain and noisy data makes the results unstable and sensitive to the sample data.
- We have generalized the concept of 'dynamic reduct' proposed by (Bazan et al., 1994) in the new context.
- These reducts are in some sense the most stable reducts.
- They are the most frequently appearing in sub-decision tables created by random samples of a given decision system.

# Dynamic attribute selection method

- Computing reduct from uncertain and noisy data makes the results unstable and sensitive to the sample data.
- We have generalized the concept of 'dynamic reduct' proposed by (Bazan et al., 1994) in the new context.
- These reducts are in some sense the most stable reducts.
- They are the most frequently appearing in sub-decision tables created by random samples of a given decision system.

- Dynamic reducts can lead to better performance in very large datasets.
- They provide the ability to accommodate noisy data.
- The rules generated by means of dynamic reduct are better to classify unseen cases.

# Dynamic attribute selection method



# Experimentations

- Tested classifiers based on rough sets:

BRSC

BRSC-GDT

# Experimentations

- Tested classifiers based on rough sets:

BRSC

BRSC-GDT

HBRSC

HBRSC-GDT

# Experimentations

- Tested classifiers based on rough sets:

BRSC

BRSC-GDT

HBRSC

HBRSC-GDT

DBRSC

DBRSC-GDT

# Experimentations

- Tested classifiers based on rough sets:

BRSC

BRSC-GDT

HBRSC

HBRSC-GDT

DBRSC

DBRSC-GDT

- Comparing with a similar classifier:

Pruned BDT

# Evaluation criteria

- To evaluate the construction phase:

  Time requirements (seconds)

# Evaluation criteria

- To evaluate the construction phase:

  Time requirements
  (seconds)

- To evaluate the classification phase:

  Size
  (Number of decision rules)

# Evaluation criteria

- To evaluate the construction phase:

  Time requirements (seconds)

- To evaluate the classification phase:

  Size (Number of decision rules)

  Classification accuracy (PCC)

# UCI repository databases

| Database | #Instances | #Attributes | #Classes |
|---|---|---|---|
| W. Breast Cancer | 690 | 8 | 2 |
| Balance Scale | 625 | 4 | 3 |
| C. Voting records | 497 | 16 | 2 |
| Zoo | 101 | 17 | 7 |
| Nursery | 12960 | 8 | 3 |
| Solar Flares | 1389 | 10 | 2 |
| Lung Cancer | 32 | 56 | 3 |
| Hayes-Roth | 160 | 5 | 3 |
| Car Evaluation | 1728 | 6 | 4 |
| Lymphography | 148 | 18 | 4 |
| Spect Heart | 267 | 22 | 2 |
| Tic-Tac-Toe Endgame | 958 | 9 | 2 |

# Time requirement (seconds)

| Bases | H-BRSC-GDT | H-BRSC | BRSC-GDT | BRSC | Pruned BDT | D-BRSC-GDT | D-BRSC |
|---|---|---|---|---|---|---|---|
| W. Breast Cancer | 41 | 66 | 136 | 154 | 156 | 381 | 393 |
| Balance Scale | 35 | 42 | 126 | 129 | 139 | 282 | 297 |
| C. Voting records | 27 | 68 | 106 | 110 | 117 | 224 | 235 |
| Zoo | 34 | 39 | 96 | 101 | 103 | 207 | 221 |
| Nursery | 127 | 191 | 356 | 380 | 386 | 867 | 1106 |
| Solar Flares | 104 | 108 | 146 | 157 | 160 | 391 | 412 |
| Lung Cancer | 15 | 29 | 42 | 48 | 56 | 174 | 187 |
| Hayes-Roth | 28 | 34 | 82 | 91 | 93 | 198 | 205 |
| Car Evaluation | 105 | 133 | 156 | 178 | 189 | 418 | 426 |
| Lymphography | 43 | 65 | 97 | 102 | 108 | 209 | 217 |
| Spect Heart | 52 | 71 | 89 | 109 | 111 | 226 | 233 |
| Tic-Tac-Toe | 98 | 107 | 127 | 139 | 149 | 293 | 312 |

# Time requirement (seconds)

| Bases | H-BRSC-GDT | H-BRSC | BRSC-GDT | BRSC | Pruned BDT | D-BRSC-GDT | D-BRSC |
|---|---|---|---|---|---|---|---|
| W. Breast Cancer | 41 | 66 | 136 | 154 | 156 | 381 | 393 |
| Balance Scale | 35 | 42 | 126 | 129 | 139 | 282 | 297 |
| C. Voting records | 27 | 68 | 106 | 110 | 117 | 224 | 235 |
| Zoo | 34 | 39 | 96 | 101 | 103 | 207 | 221 |
| Nursery | 127 | 191 | 356 | 380 | 386 | 867 | 1106 |
| Solar Flares | 104 | 108 | 146 | 157 | 160 | 391 | 412 |
| Lung Cancer | 15 | 29 | 42 | 48 | 56 | 174 | 187 |
| Hayes-Roth | 28 | 34 | 82 | 91 | 93 | 198 | 205 |
| Car Evaluation | 105 | 133 | 156 | 178 | 189 | 418 | 426 |
| Lymphography | 43 | 65 | 97 | 102 | 108 | 209 | 217 |
| Spect Heart | 52 | 71 | 89 | 109 | 111 | 226 | 233 |
| Tic-Tac-Toe | 98 | 107 | 127 | 139 | 149 | 293 | 312 |

Use H-BRSC-GDT for quick model building.

# Size (generated decision rules)

| Bases | BRSC-GDT | H-BRSC-GDT | D-BRSC-GDT | Pruned BDT | BRSC | H-BRSC | D-BRSC |
|---|---|---|---|---|---|---|---|
| W. Breast Cancer | 38 | 39 | 41 | 45 | 50 | 51 | 52 |
| Balance Scale | 48 | 50 | 51 | 51 | 59 | 61 | 63 |
| C. Voting records | 39 | 40 | 40 | 42 | 51 | 53 | 54 |
| Zoo | 26 | 28 | 30 | 32 | 38 | 42 | 44 |
| Nursery | 188 | 191 | 194 | 202 | 227 | 229 | 233 |
| Solar Flares | 101 | 112 | 117 | 121 | 110 | 127 | 129 |
| Lung Cancer | 21 | 22 | 24 | 24 | 27 | 28 | 28 |
| Hayes-Roth | 29 | 30 | 32 | 33 | 39 | 42 | 44 |
| Car Evaluation | 154 | 157 | 161 | 170 | 174 | 177 | 179 |
| Lymphography | 41 | 43 | 44 | 49 | 57 | 59 | 59 |
| Spect Heart | 40 | 41 | 42 | 45 | 53 | 54 | 54 |
| Tic-Tac-Toe | 106 | 109 | 110 | 115 | 123 | 124 | 125 |

# Size (generated decision rules)

| Bases | BRSC-GDT | H-BRSC-GDT | D-BRSC-GDT | Pruned BDT | BRSC | H-BRSC | D-BRSC |
|---|---|---|---|---|---|---|---|
| W. Breast Cancer | 38 | 39 | 41 | 45 | 50 | 51 | 52 |
| Balance Scale | 48 | 50 | 51 | 51 | 59 | 61 | 63 |
| C. Voting records | 39 | 40 | 40 | 42 | 51 | 53 | 54 |
| Zoo | 26 | 28 | 30 | 32 | 38 | 42 | 44 |
| Nursery | 188 | 191 | 194 | 202 | 227 | 229 | 233 |
| Solar Flares | 101 | 112 | 117 | 121 | 110 | 127 | 129 |
| Lung Cancer | 21 | 22 | 24 | 24 | 27 | 28 | 28 |
| Hayes-Roth | 29 | 30 | 32 | 33 | 39 | 42 | 44 |
| Car Evaluation | 154 | 157 | 161 | 170 | 174 | 177 | 179 |
| Lymphography | 41 | 43 | 44 | 49 | 57 | 59 | 59 |
| Spect Heart | 40 | 41 | 42 | 45 | 53 | 54 | 54 |
| Tic-Tac-Toe | 106 | 109 | 110 | 115 | 123 | 124 | 125 |

Use BRSC-GDT for quick decisison making.

# Classification accuracy (PCC %)

| Bases | D-BRSC | D-BRSC-GDT | BRSC | BRSC-GDT | H-BRSC | H-BRSC-GDT | Pruned BDT |
|---|---|---|---|---|---|---|---|
| W. Breast Cancer | 89.78 | 86.68 | 86.32 | 84.60 | 84.24 | 83.47 | 83.04 |
| Balance Scale | 87.47 | 86.49 | 85.28 | 84.29 | 83.66 | 80.54 | 78.18 |
| C. Voting records | 98.77 | 98.60 | 98.46 | 97.81 | 97.46 | 97.27 | 98.07 |
| Zoo | 96.45 | 96.31 | 95.17 | 94.55 | 94.23 | 93.39 | 91.69 |
| Nursery | 98.47 | 98.36 | 97.42 | 97.36 | 96.23 | 95.98 | 95.53 |
| Solar Flares | 92.66 | 91.78 | 90.75 | 89.63 | 87.73 | 86.64 | 85.74 |
| Lung Cancer | 81.63 | 80.14 | 79.43 | 78.52 | 77.45 | 76.51 | 74.51 |
| Hayes-Roth | 98.45 | 98.28 | 98.21 | 97.52 | 95.47 | 94.35 | 83.28 |
| Car Evaluation | 85.36 | 84.56 | 83.90 | 83.39 | 82.63 | 81.37 | 73.29 |
| Lymphography | 86.20 | 86.07 | 85.18 | 84.66 | 84.51 | 84.12 | 79.17 |
| Spect Heart | 89.35 | 88.37 | 87.32 | 86.67 | 86.43 | 85.34 | 83.16 |
| Tic-Tac-Toe | 88.34 | 87.49 | 87.39 | 86.34 | 85.77 | 85.35 | 83.75 |

# Classification accuracy (PCC %)

| Bases | D-BRSC | D-BRSC-GDT | BRSC | BRSC-GDT | H-BRSC | H-BRSC-GDT | Pruned BDT |
|---|---|---|---|---|---|---|---|
| W. Breast Cancer | 89.78 | 86.68 | 86.32 | 84.60 | 84.24 | 83.47 | 83.04 |
| Balance Scale | 87.47 | 86.49 | 85.28 | 84.29 | 83.66 | 80.54 | 78.18 |
| C. Voting records | 98.77 | 98.60 | 98.46 | 97.81 | 97.46 | 97.27 | 98.07 |
| Zoo | 96.45 | 96.31 | 95.17 | 94.55 | 94.23 | 93.39 | 91.69 |
| Nursery | 98.47 | 98.36 | 97.42 | 97.36 | 96.23 | 95.98 | 95.53 |
| Solar Flares | 92.66 | 91.78 | 90.75 | 89.63 | 87.73 | 86.64 | 85.74 |
| Lung Cancer | 81.63 | 80.14 | 79.43 | 78.52 | 77.45 | 76.51 | 74.51 |
| Hayes-Roth | 98.45 | 98.28 | 98.21 | 97.52 | 95.47 | 94.35 | 83.28 |
| Car Evaluation | 85.36 | 84.56 | 83.90 | 83.39 | 82.63 | 81.37 | 73.29 |
| Lymphography | 86.20 | 86.07 | 85.18 | 84.66 | 84.51 | 84.12 | 79.17 |
| Spect Heart | 89.35 | 88.37 | 87.32 | 86.67 | 86.43 | 85.34 | 83.16 |
| Tic-Tac-Toe | 88.34 | 87.49 | 87.39 | 86.34 | 85.77 | 85.35 | 83.75 |

Use D-BRSC for more accurate decision making.

# Web usage mining database

- This database was obtained from web access logs of the introductory computing science course at Saint Mary's University.
- The course is 'Introduction to Computing Science and Programming' offered in the first term of the first year.
- The uncertainty appears only in the decision attribute and is handled by the belief function theory.
- Instead of using crisp assignment of a visit to one of the three clusters, a previous study (Lingras and West, 2004) associated a basic belief assignment (bba).

| Database | #Instances | #Attributes | #Classes |
|---|---|---|---|
| Web usage mining | 7965 | 5 | 3 |

# Web usage mining database

| Approaches | Time requirement (seconds) | Size | | PCC (%) | |
|---|---|---|---|---|---|
| | | certain case | uncer case | certain case | uncer case |
| Pruned BDT | 188 | 37 | 39 | 84.07 | 85.12 |
| BRSC | 157 | 41 | 46 | 85.16 | 89.63 |
| BRSC-GDT | 143 | 32 | 35 | 84.92 | 88.46 |
| H-BRSC | 127 | 43 | 47 | 84.91 | 87.84 |
| H-BRSC-GDT | 113 | 35 | 37 | 84.65 | 87.26 |
| D-BRSC | 332 | 43 | 47 | 86.24 | 91.63 |
| D-BRSC-GDT | 308 | 35 | 37 | 85.24 | 90.53 |

# Web usage mining database

| Approaches | Time requirement (seconds) | Size | | PCC (%) | |
|---|---|---|---|---|---|
| | | certain case | uncer case | certain case | uncer case |
| Pruned BDT | 188 | 37 | 39 | 84.07 | 85.12 |
| BRSC | 157 | 41 | 46 | 85.16 | 89.63 |
| BRSC-GDT | 143 | 32 | 35 | 84.92 | 88.46 |
| H-BRSC | 127 | 43 | 47 | 84.91 | 87.84 |
| H-BRSC-GDT | 113 | 35 | 37 | 84.65 | 87.26 |
| D-BRSC | 332 | 43 | 47 | 86.24 | 91.63 |
| D-BRSC-GDT | 308 | 35 | 37 | 85.24 | 90.53 |

# References


- Salsabil Trabelsi, Zied Elouedi
  Heuristic method for attribute selection from partially uncertain data using rough sets
  International Journal of General Systems, IJGS, Volume 39, Issue 3, 271-290, 2010.

- Salsabil Trabelsi, Zied Elouedi, Pawan Lingras
  Classification with dynamic reducts and belief functions
  Transactions on Rough Sets, TRS, 66 (14), 202-233, 2011.

- Salsabil Trabelsi, Zied Elouedi, Pawan Lingras
  Classification systems based on rough sets under the belief function framework
  International Journal of Approximate Reasoning, IJAR, 52(9), 1409-1432, 2011.

# References for related works


- Pawan Lingras
  Plausibilistic rule extraction from incomplete databases using non-transitive rough set model
  Workshop on Rough Set and Database Mining, ACM, 1995.

- Juzhen Dong, Ning Zhong, Setsuo Ohsuga
  Probabilistic rough induction: The GDT-RS methodology and algorithms
  Foundations of Intelligent Systems, 621-629, 1999.

- Hideo Tanaka, Peijun Guo
  Possibilistic data analysis and its similarity to rough sets
  Data mining, rough sets and granular computing, 518-536, 2002.

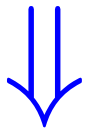
- Alpna Singh, Aruna Tiwari, Sujata Naegi
  Building a Rule-Based Classifier - Fuzzy Rough Set Approach Based Classifier
  Swarm, Evolutionary, and Memetic Computing, 550-558, 2011.

- etc.

# *Belief K-Modes*

# Definition

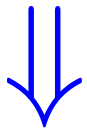
Clustering techniques are unsupervised classification techniques.



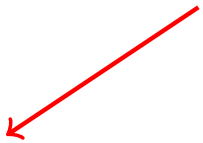
Distinguishing groups of similar objects

# Definition

Clustering techniques are unsupervised classification techniques.

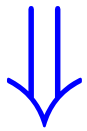


Distinguishing groups of similar objects

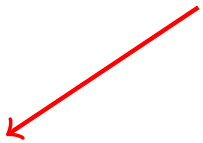


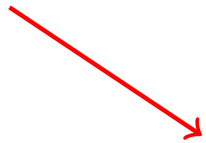
# Definition

Clustering techniques are unsupervised classification techniques.



Distinguishing groups of similar objects





High intra-cluster similarity

Low inter-cluster similarity

# Standard K-Modes (*SKM*)

- Proposed by Huang (1998)
- A K-means extension for categorical data

# Standard K-Modes (*SKM*)

- Proposed by Huang (1998)
- A K-means extension for categorical data

| **Standard K-modes method** | **Standard K-means method** |
|---|---|
| Mode initialization | |
| Random mode initialization | |
| Cluster representative definition | |
| Most frequent attribute values | Means attribute values |
| Distance measure | |
| Simple matching distance | Euclidean distance |

# Limitations

- Management of only certain data.
- Random mode initialization impact.
- Unstability due to the non-uniqueness mode problem.

# Belief K-modes

## Objective

Standard K-modes + Belief function theory

## Objective

Standard K-modes + Belief function theory

$\Rightarrow$ Belief K-Modes (BKM)

## Objective

Standard K-modes + Belief function theory

$\Rightarrow$ Belief K-Modes (BKM)

- Dealing with uncertainty in attribute values.

# Structure of the training set

- Training instances' attribute values are uncertain.



bba for attribute



Beliefs about the actual value of the attributes for each training instance.

## Construction phase

Take into account the uncertainty expressed by the means of bba.

- How to compute the cluster mode ?
- Looking for one distance measure based on bba concept.

# Belief K-modes (BKM)

## Construction phase

Take into account the uncertainty expressed by the means of bba.

- How to compute the cluster mode ?
- Looking for one distance measure based on bba concept.

## Classification phase

New object to classify.

# Cluster mode

## For each attribute value

Compute the mean mass of the attribute value.



$m_{v_j}$ : the relative bbm of category $v_j$ of attribute $A_j$ within the cluster $C$.



Belief about the value $v_j$ of attribute $A_j$ of the cluster mode.

## Distance measures based on pignistic transformation

No bijection between bba's and the obtained pignistic probabilities.

## Distance measures based on pignistic transformation

No bijection between bba's and the obtained pignistic probabilities.

## Distance measures between bba's

Jousselme et al. distance (2001).

$\Rightarrow$ Adaptation of the distance measure developed by Jousselme et al.

# *SKM* versus *BKM* methods

| **Standard K-modes method** (*SKM*) | **Belief K-modes method** (*BKM*) |
|---|---|
| Mode initialization | |
| Random mode initialization | |
| Cluster modes definition | |
| Most frequent attribute values | bba's means |
| Distance measure | |
| Simple matching distance | Belief distance (Jousselme et al., 2001) |

## Limitation

- Dependency of the results on the random mode initialization step.

## Limitation

- Dependency of the results on the random mode initialization step.

## Objective

- Propose a novel initialization method handling the randomness problem.

# Max-dissimilarity based initialization method

## *BKM*: Local minimum problem

Different random mode initialization

# Max-dissimilarity based initialization method

## *BKM*: Local minimum problem

Different random mode initialization

$\Rightarrow$ Distinct cluster partition results.

# Max-dissimilarity based initialization method

## *BKM*: Local minimum problem

Different random mode initialization

$\Rightarrow$ Distinct cluster partition results.

## Solution

Non-random selection:

Choosing deterministically $k$ objects from the data set giving the maximum of dissimilarity measure between them.

## *BKM*: Local minimum problem

Different random mode initialization

$\Longrightarrow$ Distinct cluster partition results.

## Solution

Non-random selection:

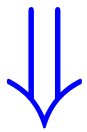
Choosing deterministically $k$ objects from the data set giving the maximum of dissimilarity measure between them.

$\Longrightarrow$ The $k$ objects which are the most far/the most dissimilar ones.

Initial mode selection

# Principle

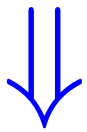
Initial mode selection



- Random-based approaches
- Frequency-based approaches

# Principle

Initial mode selection



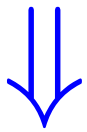
Not appropriate: Local minimum solution



Frequency-based approaches

# Principle

Initial mode selection





Random-based approaches

Not appropriate: bba's not categorical values

# Algorithm

Step 1: Initialization phase

# Algorithm

Step 1: Initialization phase

Step 2: Remaining modes selection

# Algorithm

## Step 1: Initialization phase

1. Select two objects with the maximum of dissimilarity.

# Algorithm

## Step 1: Initialization phase

1. Select two objects with the maximum of dissimilarity.
2. Assign these objects to the two first cluster modes.

## Step 2: Remaining modes selection

1. Compute the distance sum of the remaining objects to already chosen ones.

## Step 2: Remaining modes selection

1. Compute the distance sum of the remaining objects to already chosen ones.
2. Assign the object with the maximum of distance to the following mode.

## Step 2: Remaining modes selection

1. Compute the distance sum of the remaining objects to already chosen ones.
2. Assign the object with the maximum of distance to the following mode.
3. Repeat till all $k$ modes are selected.

# So..

- Ensures the initial cluster modes selection in uncertain context.

# So..

- Ensures the initial cluster modes selection in uncertain context.
- Even in the certain case, it ensures the initial cluster modes selection through certain bba.

# So..

- Ensures the initial cluster modes selection in uncertain context.
- Even in the certain case, it ensures the initial cluster modes selection through certain bba.
- Solves the non-uniqueness mode problem.

# Experimental strategy

## Notations

1. $BKM_{Init1}$ : Total random selection of $k$ modes.

# Experimental strategy

## Notations

1. $BKM_{Init1}$: Total random selection of $k$ modes.
2. $BKM_{Init2}$: The $k$ initial modes are the $k$ first distinct items.

# Experimental strategy

## Notations

1. $BKM_{Init1}$: Total random selection of $k$ modes.
2. $BKM_{Init2}$: The $k$ initial modes are the $k$ first distinct items.
3. $BKM_{Init3}$: The developed max-dissimilarity based initialization approach.

# Experimental strategy

## Notations

1. $BKM_{Init1}$: Total random selection of $k$ modes.
2. $BKM_{Init2}$: The $k$ initial modes are the $k$ first distinct items.
3. $BKM_{Init3}$: The developed max-dissimilarity based initialization approach.

- Test our proposal $BKM_{Init3}$.

# Experimental strategy

## Notations

1. $BKM_{Init1}$: Total random selection of $k$ modes.
2. $BKM_{Init2}$: The $k$ initial modes are the $k$ first distinct items.
3. $BKM_{Init3}$: The developed max-dissimilarity based initialization approach.

- Test our proposal $BKM_{Init3}$.
- Compare to random based initialization approaches results ($BKM_{Init1}$ and $BKM_{Init2}$).

# Categorical UCI databases

| Databases | #instances | #attributes | #classes |
|---|---|---|---|
| Breast cancer Wisconsin (BCW) | 699 | 9 | 2 |
| Car evaluation (CE) | 1728 | 6 | 4 |
| Congressional voting records (CVR) | 435 | 16 | 2 |
| Hayes-Roth (HR) | 160 | 5 | 3 |
| Lenses (LS) | 24 | 4 | 3 |
| Mushroom (MM) | 8124 | 22 | 2 |
| Nursery (NY) | 12960 | 8 | 5 |
| Spect heart (SH) | 267 | 22 | 2 |
| Tic-Tac-Toe Endgame (TE) | 958 | 9 | 2 |
| Zoo (ZO) | 101 | 17 | 7 |

]

# Evaluation criteria

## Clustering accuracy (%)

$$ACC = \frac{\sum_{l=1}^{k} a_l}{n} * 100$$

# Evaluation criteria

## Clustering accuracy (%)

$$ACC = \frac{\sum_{l=1}^{k} a_l}{n} * 100$$

## Time indicators

- Running time (seconds)
- Iteration number

$\Rightarrow$ Needed to reach a final stable partition

| Databases | $BKM_{Init1}$ | $BKM_{Init2}$ | $BKM_{Init3}$ |
|---|---|---|---|
| Breast Cancer Wisconsin (BCW) | 79.78 | 80.10 | 80.25 |
| Car evaluation (CE) | 81.72 | 82.05 | 82.90 |
| Congressional voting records (CVR) | 88.77 | 88.95 | 89.05 |
| Hayes-Roth (HR) | 77.50 | 77.65 | 78.45 |
| Lenses (LS) | 88.56 | 89.20 | 89.90 |
| Mushroom (MM) | 77.25 | 76.95 | 79.65 |
| Nursery (NY) | 88.23 | 90.25 | 92.95 |
| Spect heart (SH) | 86.83 | 86.90 | 89.51 |
| Tic-Tac-Toe Endgame (TE) | 71.30 | 72.15 | 75.45 |
| Zoo (ZO) | 74.23 | 74.16 | 75.58 |

# Certain case

## Clustering accuracy (*ACC* (%))

| Databases | $BKM_{Init1}$ | $BKM_{Init2}$ | $BKM_{Init3}$ |
|---|---|---|---|
| Breast Cancer Wisconsin (BCW) | 79.78 | 80.10 | 80.25 |
| Car evaluation (CE) | 81.72 | 82.05 | 82.90 |
| Congressional voting records (CVR) | 88.77 | 88.95 | 89.05 |
| Hayes-Roth (HR) | 77.50 | 77.65 | 78.45 |
| Lenses (LS) | 88.56 | 89.20 | 89.90 |
| Mushroom (MM) | 77.25 | 76.95 | 79.65 |
| Nursery (NY) | 88.23 | 90.25 | 92.95 |
| Spect heart (SH) | 86.83 | 86.90 | 89.51 |
| Tic-Tac-Toe Endgame (TE) | 71.30 | 72.15 | 75.45 |
| Zoo (ZO) | 74.23 | 74.16 | 75.58 |

| Databases | $BKM_{Init1}$ | | $BKM_{Init2}$ | | $BKM_{Init3}$ | |
|---|---|---|---|---|---|---|
| | Time(s) | Iter | Time(s) | Iter | Time(s) | Iter |
| (BCW) | 348 | 6 | 331 | 7 | 305 | 7 |
| (CE) | 452 | 10 | 468 | 11 | 411 | 8 |
| (CVR) | 311 | 9 | 312 | 10 | 268 | 7 |
| (HR) | 213 | 10 | 186 | 11 | 154 | 9 |
| (LS) | 63 | 6 | 59 | 5 | 49 | 5 |
| (MM) | 540 | 10 | 578 | 11 | 408 | 8 |
| (NY) | 648 | 12 | 641 | 12 | 521 | 11 |
| (SH) | 199 | 8 | 200 | 8 | 187 | 7 |
| (TE) | 297 | 11 | 285 | 11 | 209 | 7 |
| (ZO) | 294 | 10 | 266 | 8 | 161 | 6 |

| Databases | $BKM_{Init1}$ | | $BKM_{Init2}$ | | $BKM_{Init3}$ | |
|---|---|---|---|---|---|---|
| | Time(s) | Iter | Time(s) | Iter | Time(s) | Iter |
| (BCW) | 348 | 6 | 331 | 7 | 305 | 7 |
| (CE) | 452 | 10 | 468 | 11 | 411 | 8 |
| (CVR) | 311 | 9 | 312 | 10 | 268 | 7 |
| (HR) | 213 | 10 | 186 | 11 | 154 | 9 |
| (LS) | 63 | 6 | 59 | 5 | 49 | 5 |
| (MM) | 540 | 10 | 578 | 11 | 408 | 8 |
| (NY) | 648 | 12 | 641 | 12 | 521 | 11 |
| (SH) | 199 | 8 | 200 | 8 | 187 | 7 |
| (TE) | 297 | 11 | 285 | 11 | 209 | 7 |
| (ZO) | 294 | 10 | 266 | 8 | 161 | 6 |

| Databases | $BKM_{Init1}$ | | $BKM_{Init2}$ | | $BKM_{Init3}$ | |
|---|---|---|---|---|---|---|
| | Time(s) | Iter | Time(s) | Iter | Time(s) | Iter |
| (BCW) | 348 | 6 | 331 | 7 | 305 | 7 |
| (CE) | 452 | 10 | 468 | 11 | 411 | 8 |
| (CVR) | 311 | 9 | 312 | 10 | 268 | 7 |
| (HR) | 213 | 10 | 186 | 11 | 154 | 9 |
| (LS) | 63 | 6 | 59 | 5 | 49 | 5 |
| (MM) | 540 | 10 | 578 | 11 | 408 | 8 |
| (NY) | 648 | 12 | 641 | 12 | 521 | 11 |
| (SH) | 199 | 8 | 200 | 8 | 187 | 7 |
| (TE) | 297 | 11 | 285 | 11 | 209 | 7 |
| (ZO) | 294 | 10 | 266 | 8 | 161 | 6 |

| Databases | $BKM_{Init1}$ | | $BKM_{Init2}$ | | $BKM_{Init3}$ | |
|---|---|---|---|---|---|---|
| | Time(s) | Iter | Time(s) | Iter | Time(s) | Iter |
| (BCW) | 348 | 6 | 331 | 7 | 305 | 7 |
| (CE) | 452 | 10 | 468 | 11 | 411 | 8 |
| (CVR) | 311 | 9 | 312 | 10 | 268 | 7 |
| (HR) | 213 | 10 | 186 | 11 | 154 | 9 |
| (LS) | 63 | 6 | 59 | 5 | 49 | 5 |
| (MM) | 540 | 10 | 578 | 11 | 408 | 8 |
| (NY) | 648 | 12 | 641 | 12 | 521 | 11 |
| (SH) | 199 | 8 | 200 | 8 | 187 | 7 |
| (TE) | 297 | 11 | 285 | 11 | 209 | 7 |
| (ZO) | 294 | 10 | 266 | 8 | 161 | 6 |

# Uncertain case

## Different levels of uncertainty degrees

- Mean accuracy values (%)
- Mean running time (s)
- Mean iteration number

# Uncertain case

## Clustering accuracy (Mean *ACC* (%))

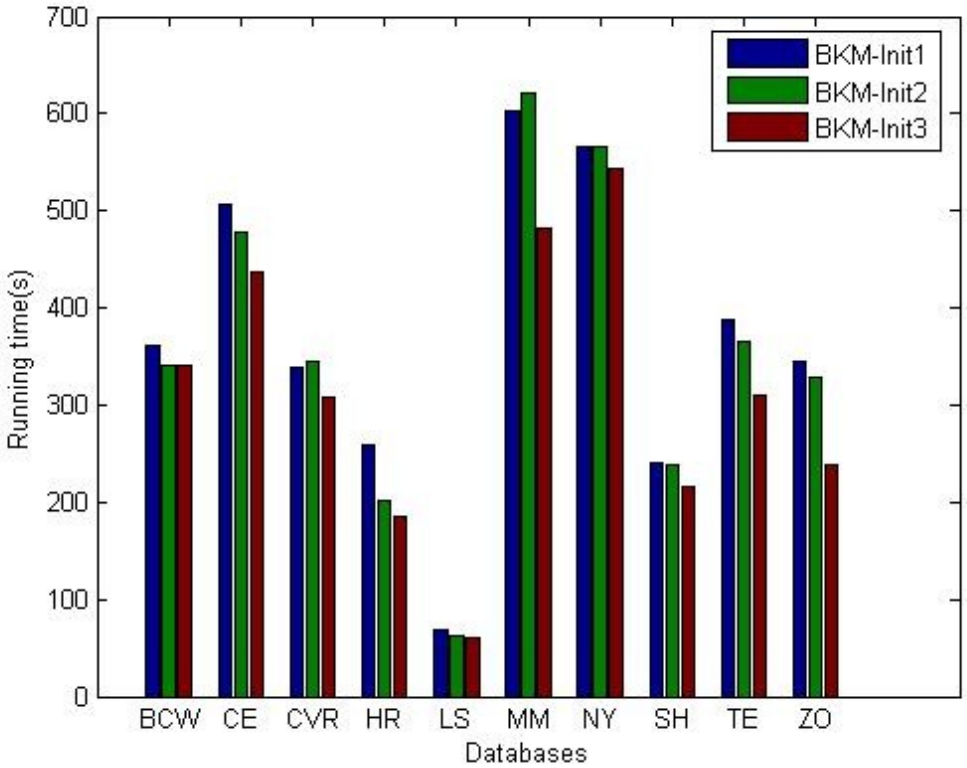
| Databases | $BKM_{Init1}$ | $BKM_{Init2}$ | $BKM_{Init3}$ |
|---|---|---|---|
| BCW | 79.06 | 80.18 | 81.66 |
| CE | 80.00 | 81.08 | 83.05 |
| CVR | 86.86 | 86.54 | 87.60 |
| HR | 77.18 | 77.23 | 78.47 |
| LS | 87.06 | 87.61 | 88.60 |
| MM | 77.33 | 77.47 | 81.86 |
| NY | 87.40 | 89.26 | 91.06 |
| SH | 85.33 | 85.55 | 88.18 |
| TE | 72.88 | 73.75 | 76.71 |
| ZO | 74.76 | 75.04 | 76.11 |

# Uncertain case

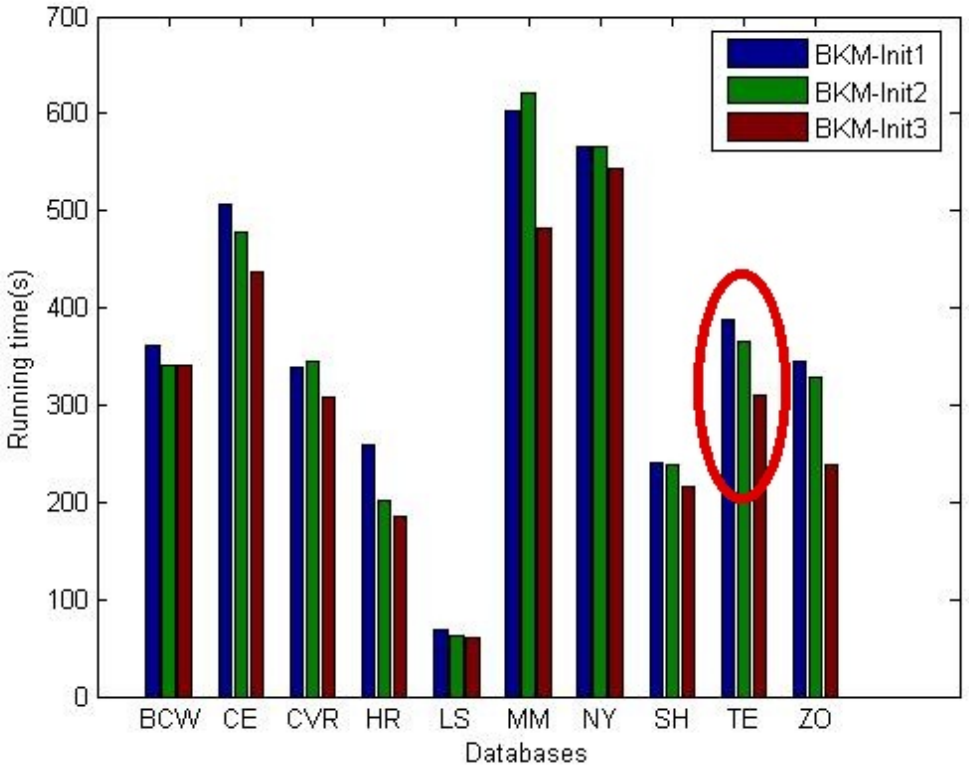
## Clustering accuracy (Mean *ACC* (%))

| Databases | $BKM_{Init1}$ | $BKM_{Init2}$ | $BKM_{Init3}$ |
|---|---|---|---|
| BCW | 79.06 | 80.18 | 81.66 |
| CE | 80.00 | 81.08 | 83.05 |
| CVR | 86.86 | 86.54 | 87.60 |
| HR | 77.18 | 77.23 | 78.47 |
| LS | 87.06 | 87.61 | 88.60 |
| MM | 77.33 | 77.47 | 81.86 |
| NY | 87.40 | 89.26 | 91.06 |
| SH | 85.33 | 85.55 | 88.18 |
| TE | 72.88 | 73.75 | 76.71 |
| ZO | 74.76 | 75.04 | 76.11 |

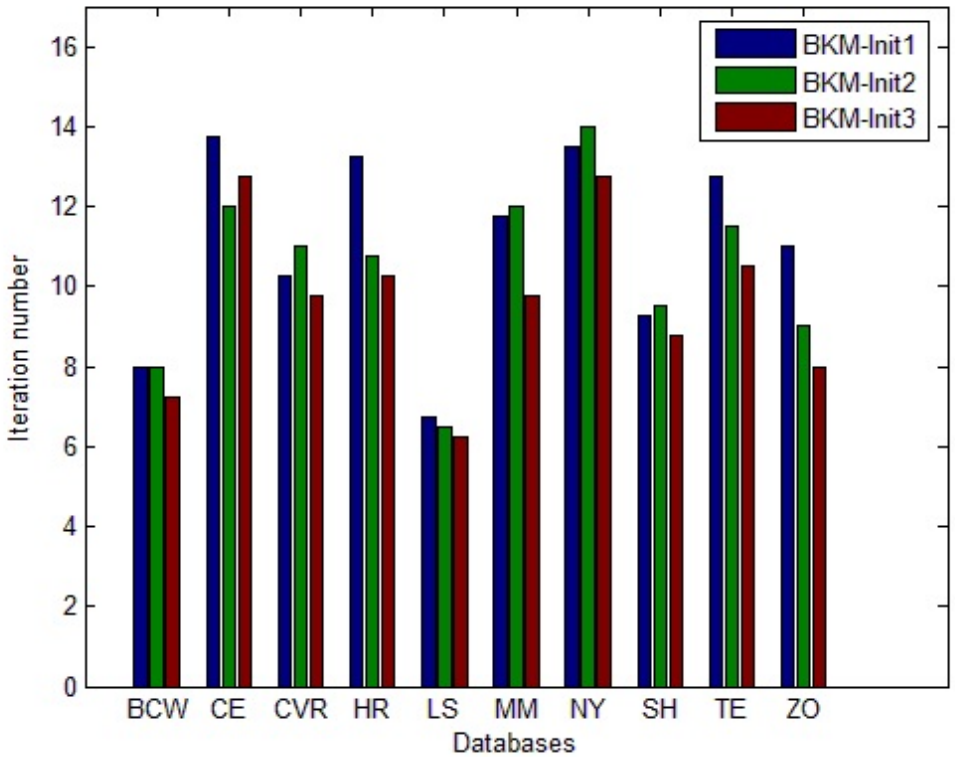
**Running time(s)**



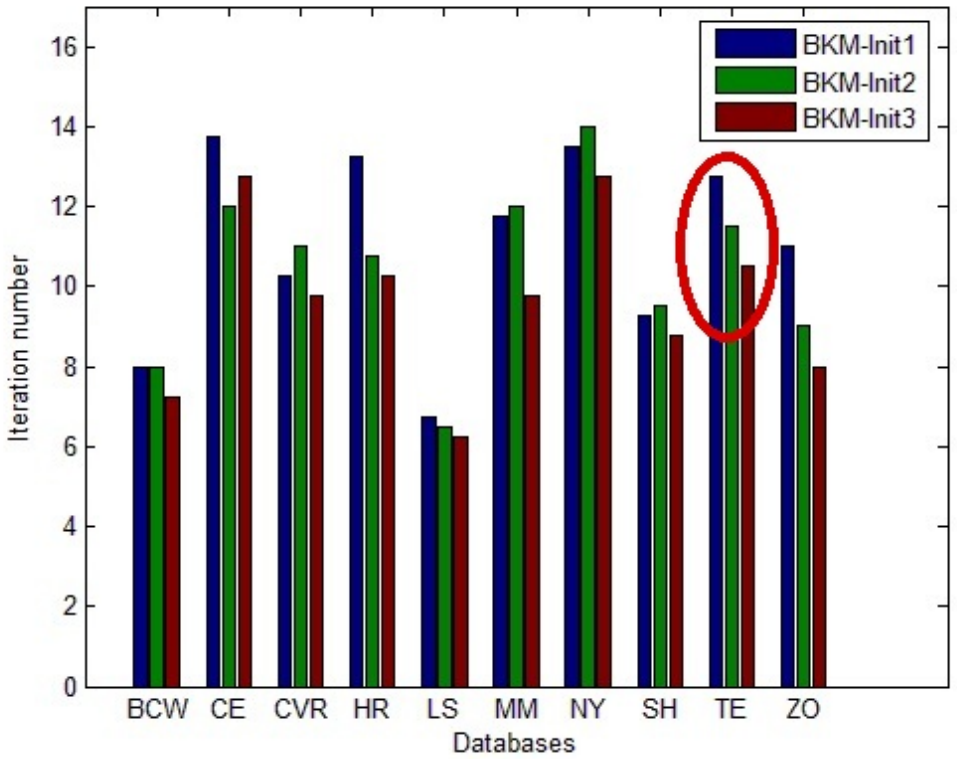
**Running time(s)**



**Iteration number**



**Iteration number**



# *Dynamic Belief K-Modes*

## Clustering parameters

- Cluster number
- Feature number
- Item number

## Clustering parameters

- Cluster number
- Feature number
- Item number

Entirely available at the beginning of the process and do not change over time

## Clustering parameters

- Cluster number
- Feature number
- Item number

Entirely available at the beginning of the process and do not change over time

$\Rightarrow$ Static Belief Clustering

## Clustering parameters

- Cluster number
- Feature number
- Item number

Entirely available at the beginning of the process and do not change over time

$\Rightarrow$ Static Belief Clustering

Dynamically detected or modified.

## Clustering parameters

- Cluster number
- Feature number
- Item number

Entirely available at the beginning of the process and do not change over time $\Rightarrow$ Static Belief Clustering

Dynamically detected or modified. $\Rightarrow$ Dynamic Belief Clustering

# Why dynamic cluster number?

Supermarket customer segmentation in three groups

# Why dynamic cluster number?

Supermarket customer segmentation in three groups



Thereafter, these groups not well represent the actual requirements of this customers

# Why dynamic cluster number?

Supermarket customer segmentation in three groups



Thereafter, these groups not well represent the actual requirements of this customers



Update dynamically the initial partition

# Why dynamic cluster number?

Supermarket customer segmentation in three groups



Thereafter, these groups not well represent the actual requirements of this customers



Update dynamically the initial partition





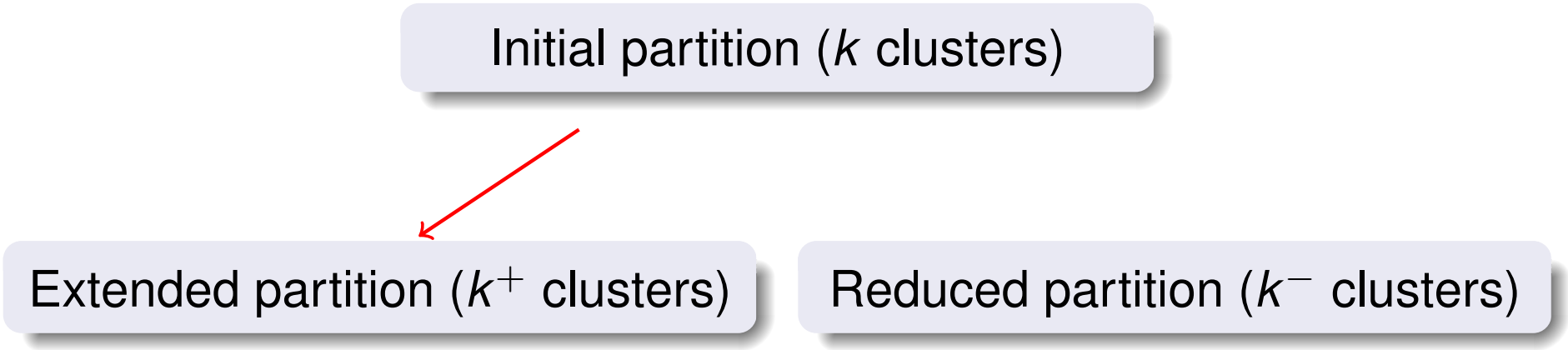
Incremental method

Decremental method

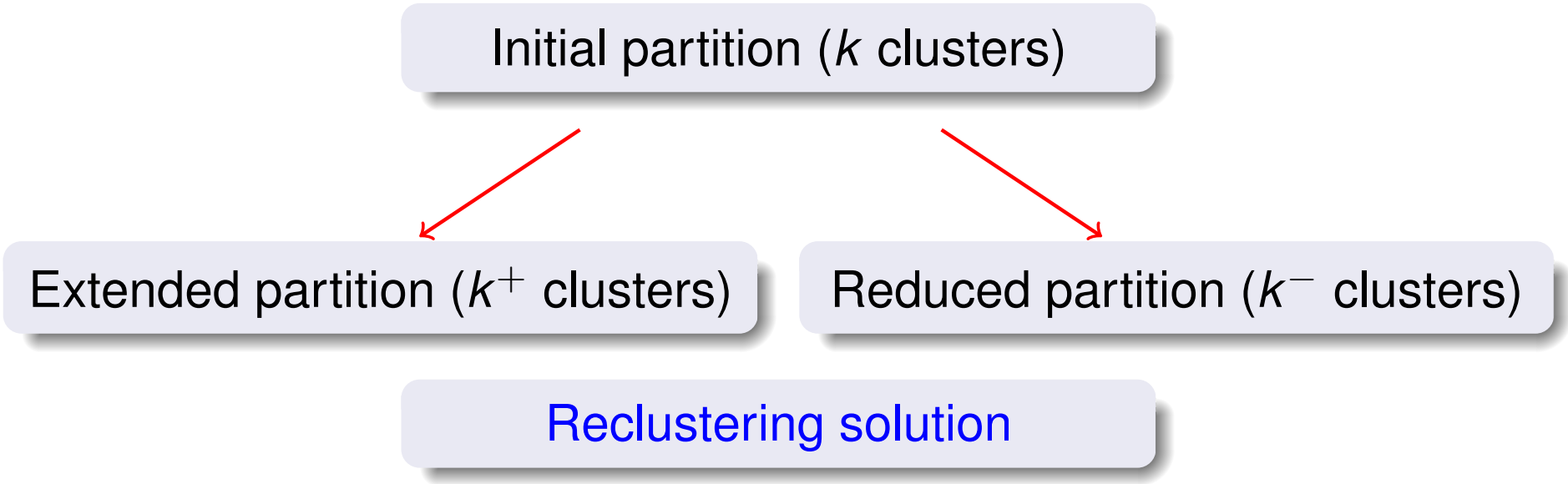
# Principle

Initial partition ($k$ clusters)

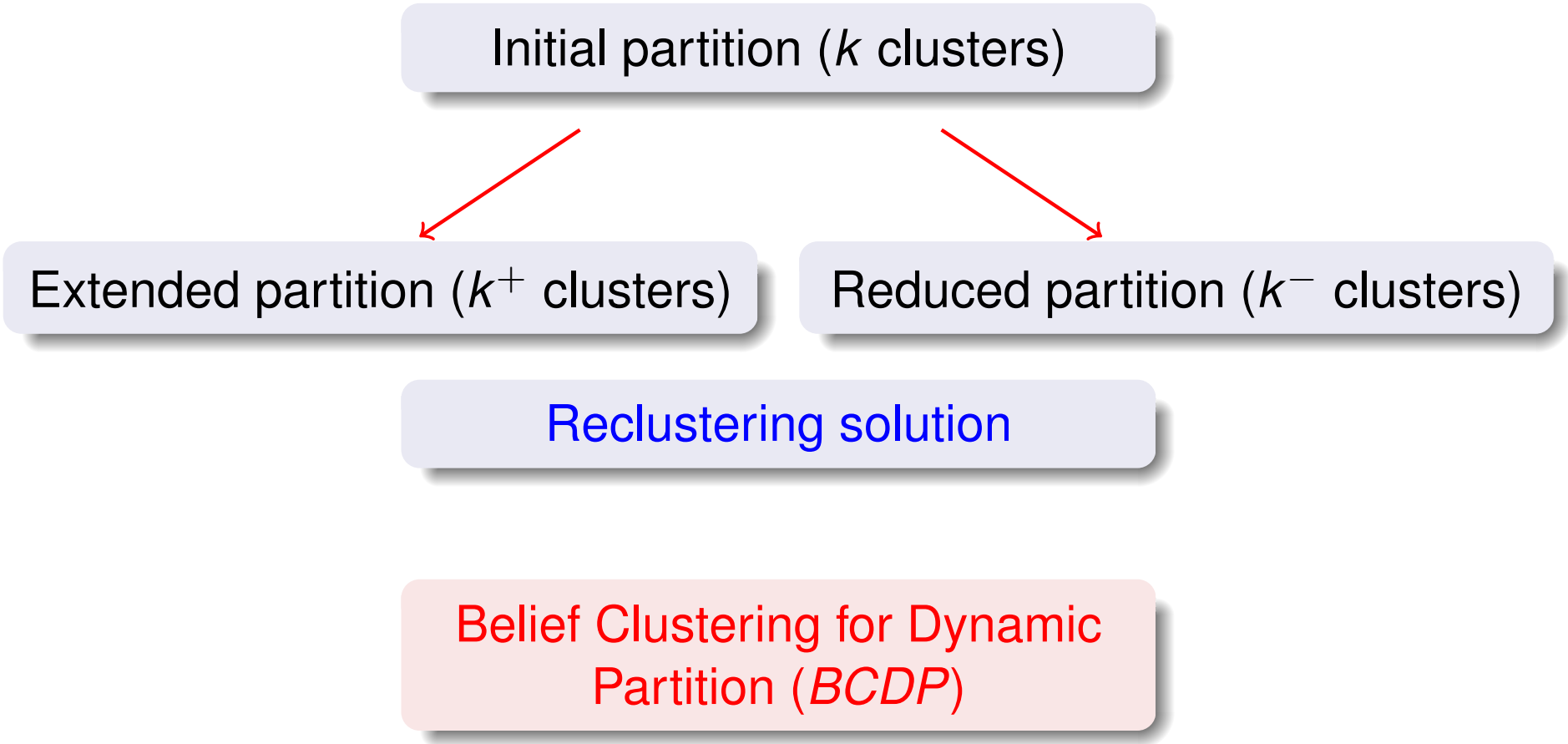
# Principle


Initial partition ($k$ clusters)
Extended partition ($k^+$ clusters)
Reduced partition ($k^-$ clusters)

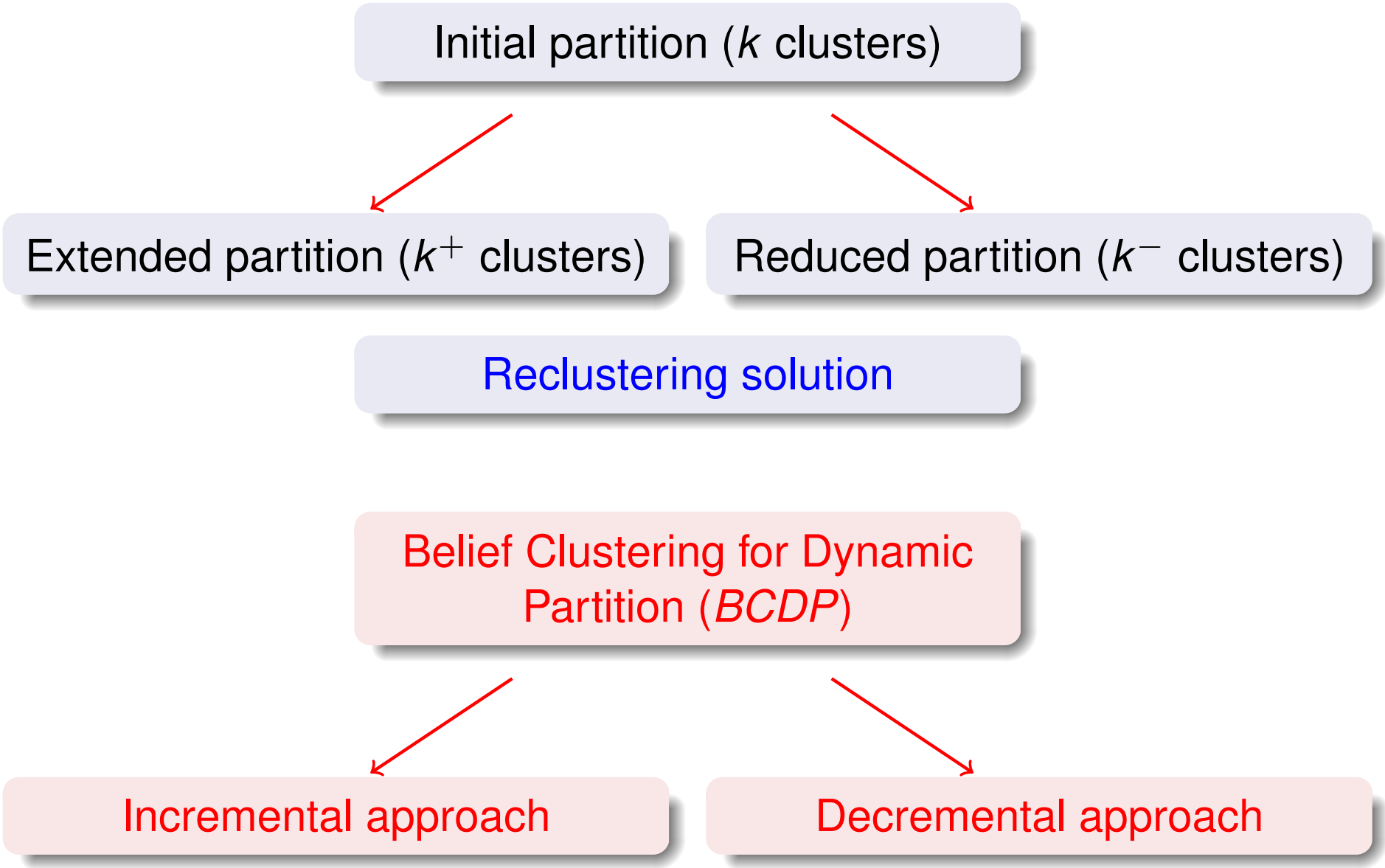
# Principle


Initial partition ($k$ clusters)
Extended partition ($k^+$ clusters)
Reduced partition ($k^-$ clusters)
Reclustering solution

# Principle


Initial partition ($k$ clusters)
Extended partition ($k^+$ clusters)
Reduced partition ($k^-$ clusters)
Reclustering solution
Belief Clustering for Dynamic Partition (*BCDP*)

# Principle


Initial partition ($k$ clusters)
Extended partition ($k^+$ clusters)
Reduced partition ($k^-$ clusters)
Reclustering solution
Belief Clustering for Dynamic Partition (*BCDP*)
Incremental approach
Decremental approach

Two families of dynamic belief clustering approaches

*Cluster cohesion based approaches*

*Cluster separation based approaches*

Two families of dynamic belief clustering approaches

*Cluster cohesion based approaches*



Minimization problem

$$Min \sum_{l=1}^{k} \sum_{i=1}^{n_l} D(X_i, Q_l)$$

*Cluster separation based approaches*



Maximization problem

$$Max \sum_{l,r=1,r>l}^{k} (\sum_{i=1}^{n_l} D(X_i, Q_r) + \sum_{u=1}^{n_r} D(X_u, Q_l))$$

## Cluster cohesion

The distance between all objects of one cluster (or objects versus mode).

$\Rightarrow$ The intra-cluster distance/dissimilarity.

## Cluster cohesion

The distance between all objects of one cluster (or objects versus mode).

$\Rightarrow$ The intra-cluster distance/dissimilarity.

## Cluster separation

The distance between all objects of two clusters (or between all objects of one cluster and the mode of the other cluster).

$\Rightarrow$ The inter-cluster distance/dissimilarity.

- Splitting and merging operators.
- Belief distance measures

## Intra-cluster distance

The distance between all objects of one cluster and its mode.

- Splitting and merging operators.
- Belief distance measures

## Intra-cluster distance

The distance between all objects of one cluster and its mode.

## Inter-cluster distance

The distance between two clusters by considering for one cluster all its objects, while for the other one only its mode.

- Splitting and merging operators.
- Belief distance measures

**Intra-cluster distance**

The distance between all objects of one cluster and its mode.

**Inter-cluster distance**

The distance between two clusters by considering for one cluster all its objects, while for the other one only its mode.

**Inter-object distance**

The objects' similarity/dissimilarity measure.

## Step 1: Initialization phase

- Selection of one existing cluster (maximal cohesion).
- Creation of new empty cluster.

## Step 1: Initialization phase

- Selection of one existing cluster (maximal cohesion).
- Creation of new empty cluster.

## Step 2: Updating phase

- Assignment of the most dissimilar object(s) of cluster with the maximal cohesion to the cluster with the minimal cohesion.
- Repeat till no improvement of total cluster cohesion or *NoMaxIter* is reached.

## Step 1: Initialization phase

- Test all $k$ splitting possibilities.
- Preserve the solution which provides the maximum of separation improvement.

## Step 1: Initialization phase

- Test all $k$ splitting possibilities.
- Preserve the solution which provides the maximum of separation improvement.

## Step 2: Updating phase

- Assignment of the most dissimilar object(s) of cluster with the maximal cohesion to the cluster with the minimal cohesion.
- Repeat till no improvement of total separation or *NoMaxIter* is reached.

## Experimental strategy

Partition of only $k-1$ clusters (*BKM*)

$\Longrightarrow$ Update this partition to attain $k$ clusters (*IKBKM*).

## Experimental strategy

Partition of only $k-1$ clusters (*BKM*)

$\Longrightarrow$ Update this partition to attain $k$ clusters (*IKBKM*).

## Certain case results

Hayes-Roth (HR) database:
*ACC* (%): 84.75%, 90.91% and 89.97% respectively to *BKM*, $IKBKM_{Min}$ and $IKBKM_{Max}$.
Running time (s): 148, 118 and 129.
Iteration number: 8, 5 and 6.

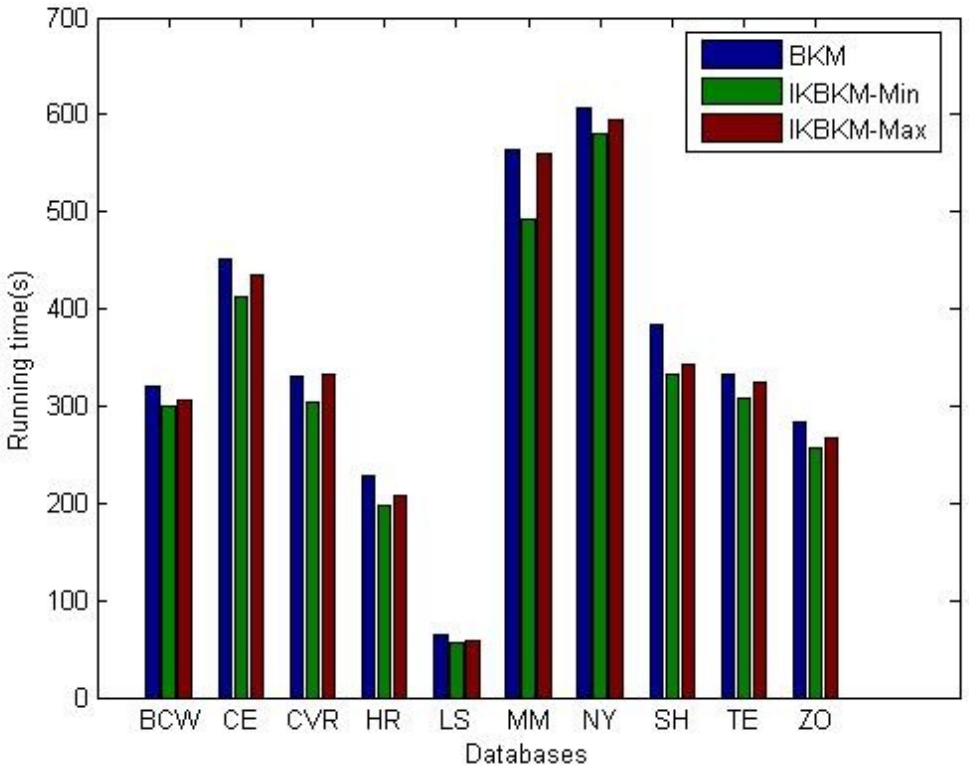
| Databases | *BKM* | $IKBKM_{Min}$ | $IKBKM_{Max}$ |
|---|---|---|---|
| BCW | 81.74 | 87.37 | 84.80 |
| CE | 82.16 | 90.20 | 85.66 |
| CVR | 82.83 | 88.65 | 85.46 |
| HR | 81.43 | 88.71 | 84.39 |
| LS | 80.48 | 87.09 | 83.87 |
| MM | 79.20 | 86.93 | 83.06 |
| NY | 79.30 | 86.46 | 82.91 |
| SH | 78.69 | 85.12 | 81.36 |
| TE | 77.71 | 85.43 | 81.65 |
| ZO | 80.96 | 86.48 | 84.08 |

# Incremental versions

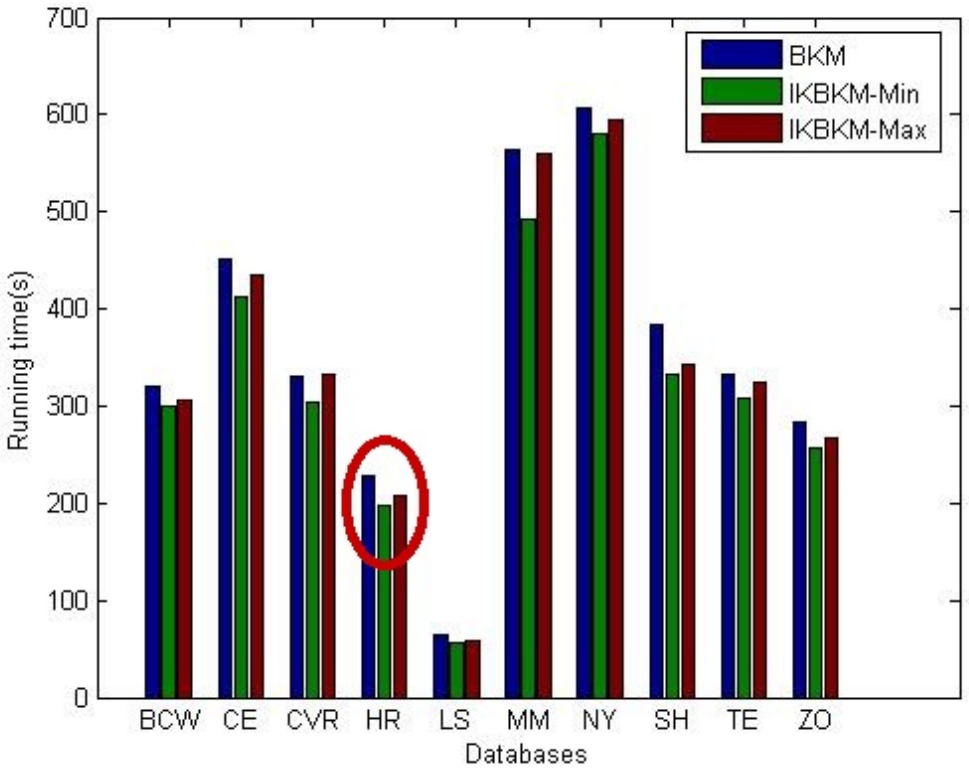
## Experimentations: Uncertain case (Mean *ACC* (%))

| Databases | *BKM* | $IKBKM_{Min}$ | $IKBKM_{Max}$ |
|---|---|---|---|
| BCW | 81.74 | 87.37 | 84.80 |
| CE | 82.16 | 90.20 | 85.66 |
| CVR | 82.83 | 88.65 | 85.46 |
| HR | 81.43 | 88.71 | 84.39 |
| LS | 80.48 | 87.09 | 83.87 |
| MM | 79.20 | 86.93 | 83.06 |
| NY | 79.30 | 86.46 | 82.91 |
| SH | 78.69 | 85.12 | 81.36 |
| TE | 77.71 | 85.43 | 81.65 |
| ZO | 80.96 | 86.48 | 84.08 |

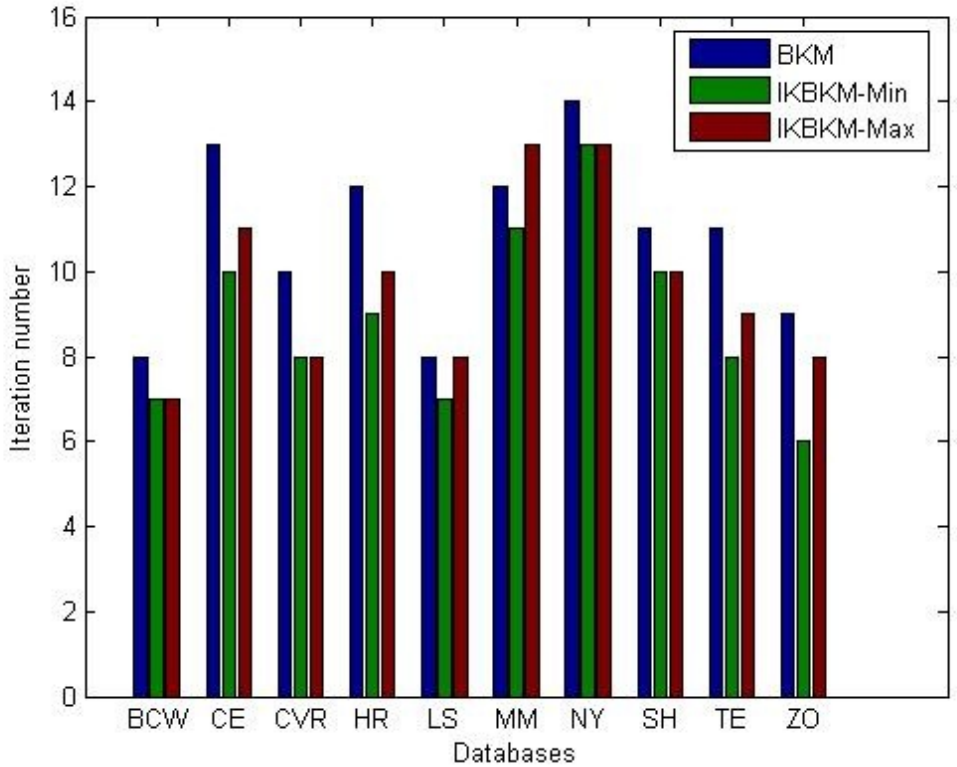
## Running time(s)



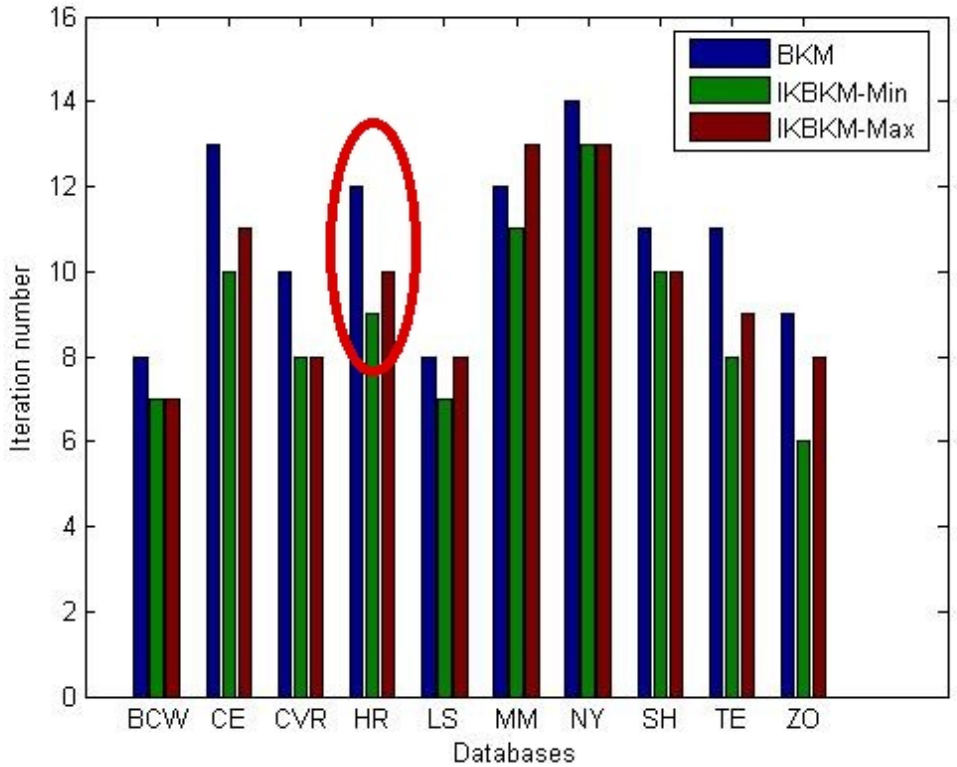
## Running time(s)



## Iteration number



## Iteration number



## Step 1: Initialization phase

- Test all merging possibilities $C_k^2$.
- Preserve the solution which provides the best reduction of the cluster cohesion.

### Step 1: Initialization phase

- Test all merging possibilities $C_k^2$.
- Preserve the solution which provides the best reduction of the cluster cohesion.

### Step 2: Updating phase

- Assignment of the most dissimilar object(s) of cluster with the maximal cohesion to the cluster with the minimal cohesion.
- Repeat till no improvement of total cluster cohesion or *NoMaxIter* is reached.

## Step 1: Initialization phase

- Derive the two clusters giving the maximum of similarity.

## Step 1: Initialization phase

- Derive the two clusters giving the maximum of similarity.

## Step 2: Updating phase

- Assignment of the most dissimilar object(s) of cluster with the maximal cohesion to the cluster with the minimal cohesion.
- Repeat till no improvement of total separation or *NoMaxIter* is reached.

## Experimental strategy

Partition of $k+1$ clusters (*BKM*)

$\Longrightarrow$ Eliminate one cluster and attain only $k$ (*DKBKM*).

## Certain case results

Zoo (ZO) database:
*ACC* (%): 85.07%, 88.46% and 90.37% respectively to *BKM*, $DKBKM_{Min}$ and $DKBKM_{Max}$.
Running time (s): 174, 158 and 147.
Iteration number: 6, 5 and 5.

## Experimental strategy

Partition of $k+1$ clusters (*BKM*)

$\Longrightarrow$ Eliminate one cluster and attain only $k$ (*DKBKM*).

## Certain case results

Zoo (ZO) database:
*ACC* (%): 85.07%, 88.46% and 90.37% respectively to *BKM*, $DKBKM_{Min}$ and $DKBKM_{Max}$.
Running time (s): 174, 158 and 147.
Iteration number: 6, 5 and 5.

# Decremental versions

## Experimentations: Uncertain case (Mean *ACC* (%))

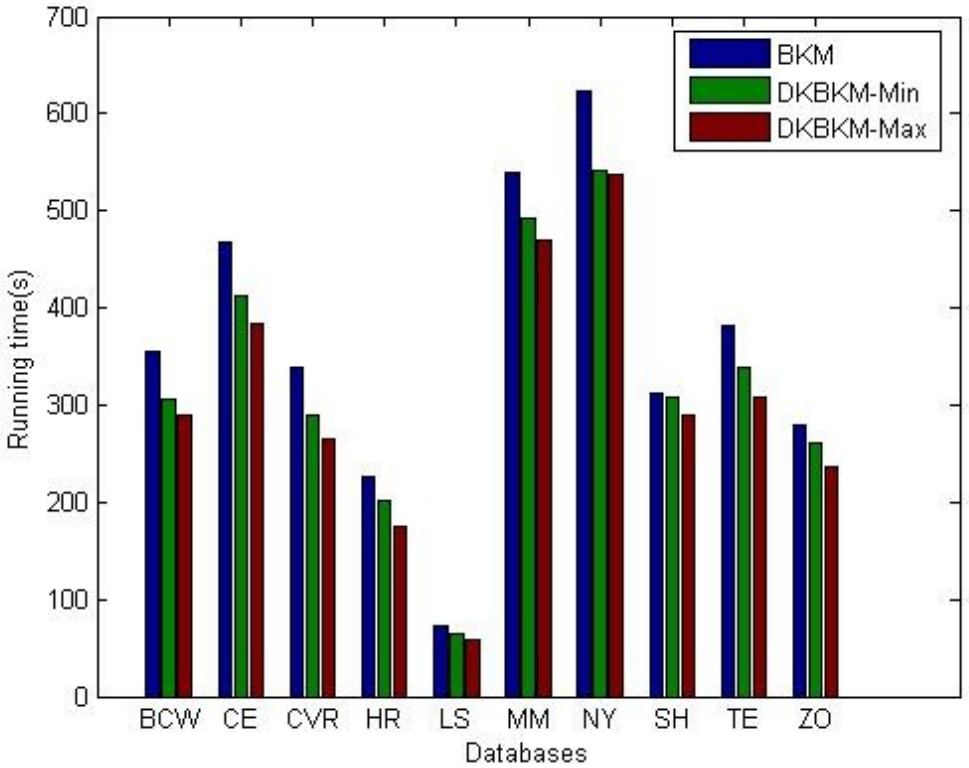
| Databases | *BKM* | $DKBKM_{Min}$ | $DKBKM_{Max}$ |
|---|---|---|---|
| BCW | 81.74 | 84.54 | 87.25 |
| CE | 82.16 | 86.49 | 90.48 |
| CVR | 82.83 | 85.42 | 88.99 |
| HR | 81.43 | 84.21 | 88.07 |
| LS | 80.48 | 84.21 | 87.69 |
| MM | 79.20 | 82.54 | 86.29 |
| NY | 79.30 | 82.89 | 86.01 |
| SH | 78.69 | 81.91 | 85.35 |
| TE | 77.71 | 81.82 | 85.52 |
| ZO | 80.96 | 83.31 | 86.14 |

# Decremental versions

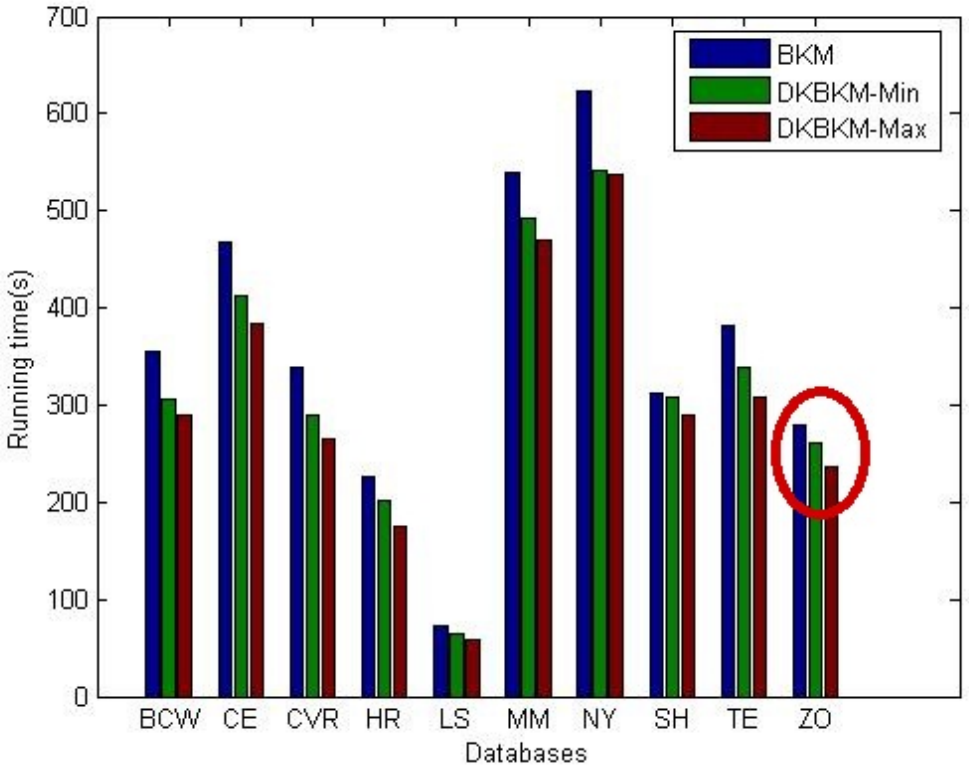
## Experimentations: Uncertain case (Mean *ACC* (%))

| Databases | *BKM* | $DKBKM_{Min}$ | $DKBKM_{Max}$ |
|---|---|---|---|
| BCW | 81.74 | 84.54 | 87.25 |
| CE | 82.16 | 86.49 | 90.48 |
| CVR | 82.83 | 85.42 | 88.99 |
| HR | 81.43 | 84.21 | 88.07 |
| LS | 80.48 | 84.21 | 87.69 |
| MM | 79.20 | 82.54 | 86.29 |
| NY | 79.30 | 82.89 | 86.01 |
| SH | 78.69 | 81.91 | 85.35 |
| TE | 77.71 | 81.82 | 85.52 |
| ZO | 80.96 | 83.31 | 86.14 |

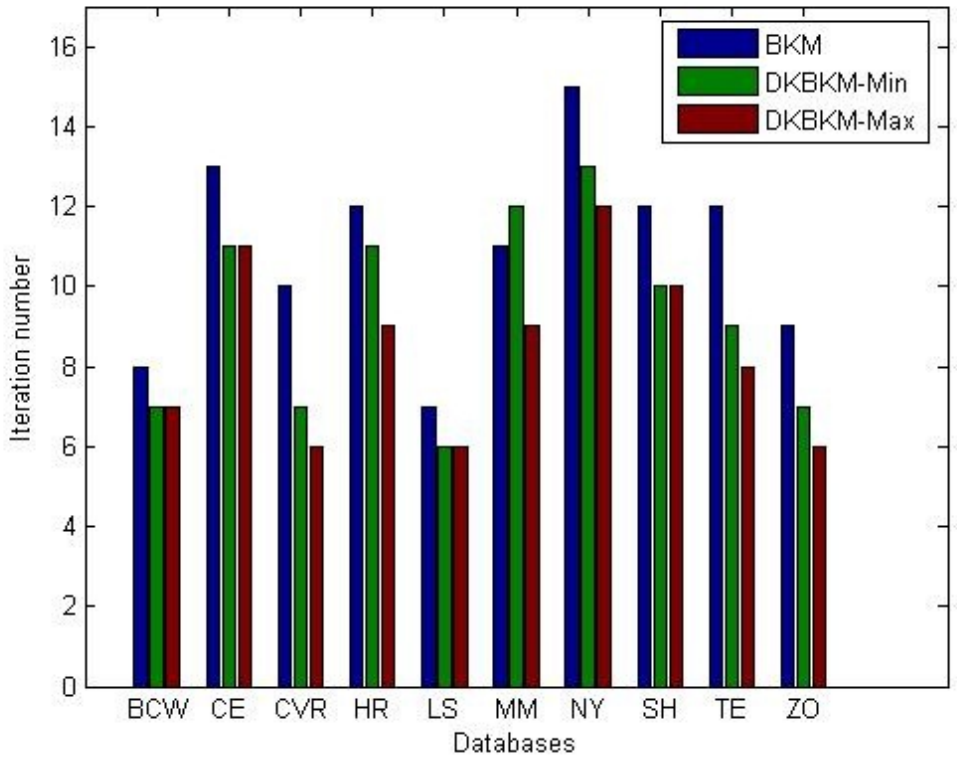
## Running time(s)



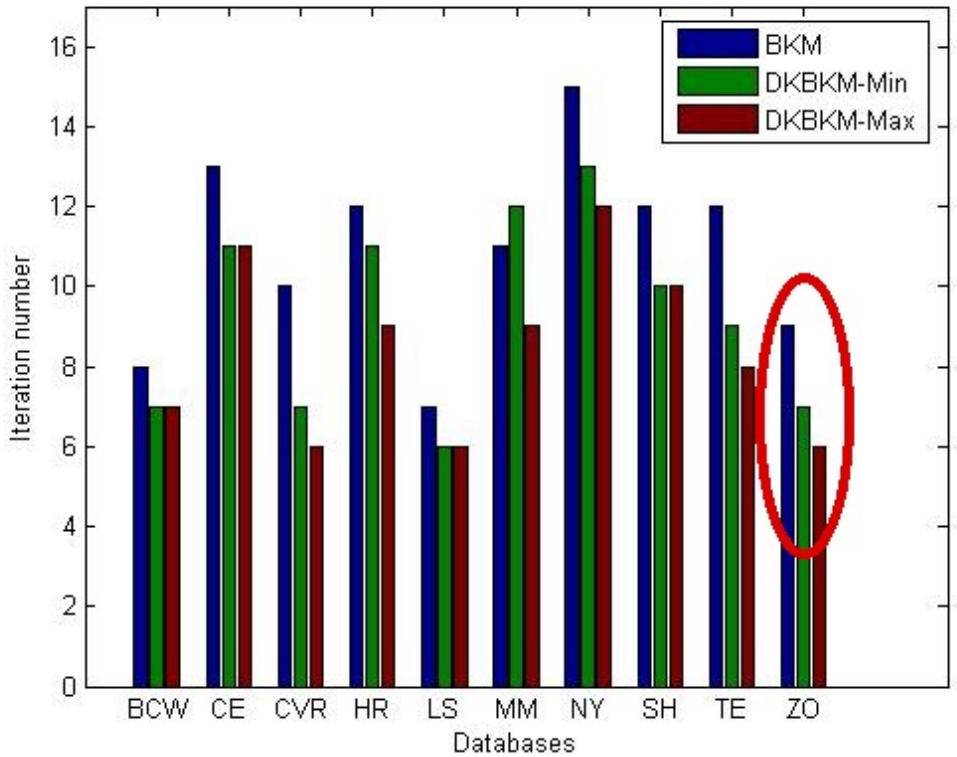
## Running time(s)



## Iteration number



## Iteration number



Belief Clustering for Dynamic Data sets Features
(*BCDDF*)

# Dynamic Belief K-Modes

Belief Clustering for Dynamic Data sets Features
(*BCDDF*)

Belief Clustering for Dynamic Data sets Items
(*BCDDI*)

# References


- Sarra Ben Hariz, Zied Elouedi
  IK-BKM: An incremental clustering approach based on intra-cluster distance
  AICSSA'10, 1-8, 2010.

- Sarra Ben Hariz, Zied Elouedi
  DK-BKM: Decremental K belief k-modes method
  SUM'10, 84-97, 2010.

- Sarra Ben Hariz, Zied Elouedi, Khaled Mellouli
  Ranking-based feature selection method for dynamic belief clustering
  ICAIS'11, 308-319, 2011.

- Sarra Ben Hariz, Zied Elouedi
  New Dynamic Clustering Approaches within Belief Function Framework
  Intelligent Data Analysis, 18(2), 2014.

# References for related works


- Thierry Denoeux, Marie-Hélène Masson
  EVCLUS: Evidential Clustering of Proximity Data
  IEEE Transactions on Systems, Man and Cybernetics - Part B, 34(1):95-109, 2004

- Violaine Antoine, Bejamin Quost, Marie-Hélène Masson, Thierry Deneoux
  CECM: Constrained Evidential C-Means algorithm
  Computational Statistics and Data Analysis, 56(4), 894-914, 2012

- Zhiwei Huang, M. Ng
  A fuzzy K-modes algorithm for clustering categorical data
  IEEE Transaction on Fuzzy Systems, 7(4):446-452, 1999.

- Asma Ammar, Zied Elouedi
  A new possibilistic clustering method: The possibilistic k-modes
  AI*IA 2011, 413-419, 2011.

- etc.

# Conclusion

- Belief decision trees,
- Belief rough set classifier,
- Belief K-nearest neighbors,
- Belief Neural Network classifier,
- Belief clustering (Belief K-modes, Dynamic Belief K-modes, Evidential c-means, etc),
- Credal EM,
- etc.