Objective

- Investigate the blurry notion of “interpretability” within the Theory of Fuzzy Information Granulation
- Develop new contributions, both theoretical and algorithmic, for interpretable information granulation
Outline

1. Interpretability Issues in Fuzzy Information Granulation
2. Theoretical Contributions to Interpretability in Fuzzy Information Granulation
3. Algorithms for Deriving Interpretable Fuzzy Information Granules

Interpretability Issues in Fuzzy Information Granulation

Granular Computing
Interpretability Constraints for Fuzzy Information Granules
Granular Computing

- Granular Computing is an emerging computing paradigm of information processing.
  - Representation and processing of "information granules"
  - Information granulation: abstraction of data and derivation of knowledge from information
- Theory of Fuzzy Information Granulation: Granular Computing with Fuzzy Set Theory as formal underpinning
  - Fuzzy Information Granules are able to represent a kind of knowledge that is close to human percepts
  - Double layer of knowledge representation
    - Numerical level → knowledge-based inference
    - Symbolic level → knowledge communication

Knowledge Communication

- Communication is important when knowledge is acquired from observations to better understand the process that produced the samples;
- Knowledge communication needs a language of symbols to exist, i.e. a system consisting of a representation (symbols) along with metaphor and some kind of grammar;
- Metaphor is a relation between representation and semantics and is implicitly shared among all communicating actors.
Knowledge Base in Fuzzy Models

- Fuzzy models embody a knowledge base made of fuzzy information granules
  - Advantage: Semantics of fuzzy information granules expresses vagueness, similarly to concepts of the human mind
  - Problem: the association of linguistic terms to fuzzy information granules is not straightforward

Interpretable Knowledge Base

- Comprehensibility Postulate:
  - An interpretable knowledge base “should be [defined by] symbolic descriptions of given entities, semantically and structurally similar to those a human expert might produce observing the same entities.
  - Components of these descriptions should be comprehensible as single chunks of information, directly interpretable in natural language, and should relate quantitative and qualitative concepts in an integrated fashion” (Michalski, 1983)
  - An interpretable knowledge base should be represented by linguistic terms whose metaphor is shared among all users
Interpretability Constraints for Fuzzy Information Granules

- Constraint-based approach for interpretability
  - Crisp constraints
  - Fuzzy constraints
- Fuzzy modeling with interpretability constraints
- Issues:
  - Which constraints characterize ‘interpretability’?
  - What are psychological / semiotic / computational motivations behind each interpretability constraint?

A Survey on Interpretability Constraints: Objectives

- Homogenous description of the interpretability constraints
- Identification of potential different notions of interpretability;
- Critical review of interpretability constraints;
- Possibility of identifying new methods for interpretable information granulation.
**Constraints on Fuzzy Sets & FoC**

- Fuzzy Sets
  - Normality
  - Convexity
  - Unimodality
  - Continuity

- Frames of Cognition
  - Proper ordering
  - Justifiable number of elements
  - Distinguishability
  - Completeness (coverage)
  - Complementarity
  - Uniform Granulation
  - Leftmost/Rightmost fuzzy sets
  - Natural zero positioning
  - Error Free Reconstruction

**Constraints on Information Granules & Rules**

- Information Granules
  - Description Length
  - Attribute Correlation

- Rules
  - Rule length
  - High-order consequents
  - Improved consequents
Constraints on Fuzzy Models

- Knowledge representation and inference
  - Implication and Aggregation
  - Number of rules (compactness)
  - Number of firing rules
  - Shared fuzzy sets
  - Rule locality
  - Modus Ponens
  - Consistency
  - Model completeness
  - Number of variables
  - Granulated output

- Knowledge refinement
  - No position exchange
  - No sign change
  - Similarity preservation

Theoretical Contributions to Interpretability in Fuzzy Information Granulation

Distinguishability Quantification
Interface Optimality
Distinguishability Quantification

▲ Distinguishability is a common constraint applied to fuzzy sets defined on the same Universe of Discourse
▲ Intuitively:
  ➢ Distinguishable fuzzy set represent different concepts;
  ➢ Undistinguishable fuzzy set represent (almost) the same concept
▲ Distinguishable fuzzy sets can be assigned to different linguistic labels

Definition of Distinguishability

▲ Intuitive definition:
  ➢ Fuzzy relation between fuzzy sets
  ➢ Degree of inequality of two fuzzy sets:
    \[ D(A,B) = 1 - \mu(A \neq B) \]
▲ Disjoint fuzzy sets are completely distinguishable
▲ Other interpretability constraints require overlapping fuzzy sets (e.g. coverage, etc.)
▲ Distinguishability must be balanced
Measures for Distinguishability

**Similarity**
- Good mathematical definition for fuzzy equality between fuzzy sets
- Calculation of similarity is computationally expensive
  + Cardinality calculation is expensive for real universes of discourse
  + Redefinition of similarity in terms of fuzzy sets' parameters is applicable only to a small class of fuzzy sets
- Use of similarity:
  + In genetic, evolutive and affine massive search strategies
  + In a separate stage after model learning
  + Not used in on-line efficient learning schemes

**Possibility**
- Degree of overlapping of two fuzzy sets
- Clear semantics in terms of elastic-constraints satisfaction
- Can be re-defined in terms of fuzzy sets parameters for a wide class of fuzzy sets
  + Efficient calculation

\[
S(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}
\]
\[
\Pi(A,B) = \sup \min \left\{ \mu_1(x), \mu_2(x) \right\}
\]

Possibility and Distinguishability

- Is possibility a valid characterization of distinguishability? Generally, no!
  + The possibility of two intervals [0,1] and [1,2] is **one**, their similarity is **zero**.
  + The possibility of a subnormal fuzzy set with itself is **less than one**, its similarity is **one**.
- However, there is evidence that possibility can be interpreted as degree of overlap, hence a good candidate to quantify distinguishability.
  + Conditions of validity
Possibility and (Maximal) Similarity

Theorem

Let \( A \) and \( B \) two normal, continuous and convex fuzzy sets. Let \( p_A \) and \( p_B \) two prototypes of \( A \) and \( B \) resp. and \( x_\pi \) the point between the prototypes, with \( \Pi(x_\pi) = \pi \). Furthermore, assume that the membership function of \( A \) is convex in \([p_A, x_\pi]\) and the membership function of \( B \) is convex in \([x_\pi, p_B]\). Then, the similarity of \( A \) and \( B \) is upper-bounded by

\[
S(A, B) \leq \frac{2\pi}{r + 2\pi - r\pi}
\]

Comments to the Theorem

Under theorem hypothesis, the similarity between two fuzzy sets cannot exceed a threshold that is monotonically related to the possibility measure of the two fuzzy sets. Theorem hypothesis are usually met in fuzzy modeling with interpretability constraints. In this sense, the use of possibility is fully justified as a quantification of distinguishability in interpretable fuzzy modeling.
Possibility Reduction to Improve Distinguishability

Corollary

Let $\Phi$ be a transformation on fuzzy sets such that $B' = \Phi(B)$ preserves theorem’s hypothesis. Suppose that $\prod(A, B') < \prod(A, B)$.

Then the maximal similarity between $A$ and $B'$ is less than the maximal similarity between $A$ and $B$, provided that:

Comments to the corollary

The corollary justifies the adoption of the possibility measure in most model adaption processes that is aimed at improving distinguishability.

Transformations that verify the corollary hypothesis:

- Positive translations $B(x) = B(x-l), l > 0$
- Contractions $B'(x) = B(x)^k, k > 1$
- ...
Interface Optimality

- Fuzzy models
  - Input interface
  - Processing Module
  - Output interface
- Information transformation
  - Matching procedure
  - Same framework for input and output

Precise Representation

- Conservation of information in the conversion process from numerical to linguistic format of information.
- Realization of a one-to-one mapping of the external information into its internal representation.
- Precise representation of data in output interfaces is a necessary condition to assure a meaningful correspondence between the information coming from the processing module and the final numerical output.
  - Precise representation improves the understanding of the system's behavior, thus preserving the overall readability of the fuzzy system
**Optimality Definition**

\[ \mathfrak{A} : \mathbf{X} \subseteq \mathbb{R}^n \rightarrow [0, 1]^{|\mathfrak{A}|} \]

\[ \text{opt}(\mathfrak{A}, \mathbf{X}) = \exists \mathfrak{A}^{-1} \text{ s.t. } \forall x \in \mathbf{X} : \mathfrak{A}^{-1}(\mathfrak{A}(x)) = x \]

\[ \mathfrak{D} : [0, 1]^{|\mathfrak{D}|} \rightarrow \mathbb{R}^n \]

\[ \text{opt}'(\mathfrak{D}, [0, 1]^{|\mathfrak{D}|}) \iff \exists \mathfrak{D}^{-1} \forall x \in \mathbf{X} : \pi_y = \mathfrak{D}^{-1}(\mathfrak{D}(x)) \rightarrow \mathfrak{D}^{-1}(\mathfrak{D}(\pi_y)) = \pi_y \]

---

**Bi-monotonic fuzzy sets (2MFS)**

\[ A : \mathbf{X} \subseteq \mathbb{R} \rightarrow [0, 1] \]

2MFS\((A) \iff \exists p \in \mathbf{X} \text{ s.t.} : \]

- \( A(p) = 1 \),
- \( A_{|x \cap [p, +\infty)} \) strictly/loosely monotonic
- \( A_{|x \cap (-\infty, p]} \) strictly/loosely monotonic
Optimality conditions for Strictly 2MFS

Theorem

A one-dimensional fuzzy interface based on strictly bi-monotonic fuzzy sets is always optimal, provided that each input can be enclosed between two prototypes.

The condition is easy to fulfill (e.g. by defining the leftmost and the rightmost fuzzy sets).

Optimality Condition for Loosely 2MFS

Theorem

A one-dimensional fuzzy interface based on loosely bi-monotonic fuzzy sets is optimal if any ambiguity of a fuzzy set is removed by some other fuzzy set in the interface.
Remarks

- Strictly 2MFS are advisable in fuzzy modeling
  - Optimality easily guaranteed
- Loosely 2MFS must be carefully designed for interface optimality
  - Flat areas generate ambiguity
- Optimality of multidimensional interfaces depends on optimality of the one-dimensional projection

Optimality of Output Interfaces

- Output Interface
  \[ \mathcal{D} : [0,1]^n \rightarrow \mathbb{R} \]
- Mapping a high dimensional space to a lower dimensional space
- Optimality is a problem
  - None of the known defuzzification methods guarantee optimality, even in the restricted definition
  \[ \mathcal{D} : \mathcal{P}(X) \rightarrow \mathbb{R} \]
A Simple Model

1. If x is A Then y is A
2. If x is B Then y is B
3. If x is C Then y is C

Implication: MIN
Aggregation: MAX
Defuzzification
+ Centroid
+ Bisector
+ Min/Max/Mean of Maxima

Spurious Non-Linearities

centroid
bisector
mean of maxima
min of maxima
max of maxima
ideal
Optimality Degree

- OD in [0,1]
- OD = 1 : Optimal Output Interface
- The lower is OD, the farther is the behavior of the Output Interface from the ideal behavior

Definition of OD
- Punctual (for each x in X)
- Global (for the entire system)

Reversely Optimal Output Interface

- Given an Output Interface
  \[ \mathcal{O} : [0,1]^n \rightarrow \mathbb{R} \]
- A Related Input Interface (RII) is defined
  \[ \mathcal{\tilde{O}} : \mathbb{R} \rightarrow [0,1]^n \]
- In Mamdani Fuzzy Systems the RII is:
  \[ \mathcal{\tilde{O}}(y) = (\mu_{\tilde{u}_1}(y), \mu_{\tilde{u}_2}(y), \ldots, \mu_{\tilde{u}_n}(y)) \]
- The Output Interface is reversely optimal if the RII is optimal
  - Optimality of the RII is easy to verify
Optimality Degree Definition

\[ od(x) = \exp \left( -\| \pi_i - \bar{\pi} \|^2 \right) \]

\[ OD = \int_x \od(x) \, dx \int_x \, dx \]

Illustrative Example
Optimality Degree vs. Accuracy

\[ \frac{dx}{dt} = \frac{0.2x(t-17)}{1 + x(t-17)^3} - 0.1x(t) \]

<table>
<thead>
<tr>
<th>Defuzz. method</th>
<th>Optimality Degree</th>
<th>MSE (training)</th>
<th>MSE (test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centroid</td>
<td>0.4823</td>
<td>0.0243</td>
<td>0.0180</td>
</tr>
<tr>
<td>Bisector</td>
<td>0.4698</td>
<td>0.0265</td>
<td>0.0198</td>
</tr>
<tr>
<td>MOM/LOM/SOM</td>
<td>0.3832</td>
<td>0.0692</td>
<td>0.0596</td>
</tr>
</tbody>
</table>

Uses

- **Evaluation**
  - Quantification of output interfaces quality
  - Tool to support the choice of an appropriate defuzzification method

- **Optimization**
  - Combining different defuzzification methods to improve optimality degree
  - Devising new optimal interfaces
Algorithms for Deriving Interpretable Fuzzy Information Granules

- Gaussian Information Granulation
- Minkowski Information Granulation
- Information Granulation through Prediction Intervals
- Information Granulation through Double Clustering
- Refinement of Interpretable Fuzzy Information Granules through Neural Learning

Gaussian Information Granulation

- **Objective**: Generate interpretable fuzzy information granules from fuzzy clustering results
  - Information granules to represent fuzzy quantities
- **Expected results**
  - Efficient generation
  - Compact representation of granules
  - Interpretable representation
Proposed solution

- Gaussian fuzzy sets to represent granules
- Gaussian centers = cluster prototypes
- Gaussian widths determined automatically
  - Constrained quadratic programming problem resolution
  - Full exploitation of the partition matrix returned by the clustering process
  - Resulting granules are well separated

\[
\mu_{\omega,C} (\mathbf{x}) := \exp\left(- (\mathbf{x} - \omega)^T C (\mathbf{x} - \omega)^T \right)
\]

\[
C := \text{diag} \mathbf{c} = \text{diag}(c_1, c_2, \ldots, c_m), \ c_i > 0
\]

Optimization Strategy

- Ideal solution:

\[
\forall i : \mu_{\omega,C} (\mathbf{x}_i) = u_i
\]

and: \( \forall \mathbf{x} \neq 0 : \mathbf{x}^T \mathbf{C} \mathbf{x} > 0 \) (spd)

- Optimization objective:

minimize: \( f (\mathbf{c}) = \sum_{i=1}^{m} \left( \sum_{k=1}^{n} \tilde{x}_{ik}^2 c_i + \log u_i \right)^2 \)

subject to: \( \mathbf{c} > 0 \)

- Justification

\[
\varepsilon = \left( \frac{\log \mu_{\omega,C} (\mathbf{x}_i)}{u_i} \right)^2 \approx \left( \frac{\mu_{\omega,C} (\mathbf{x}_i)}{u_i} - 1 \right)^2
\]
Illustrative Example: Spatial Granulation

<table>
<thead>
<tr>
<th>Measure</th>
<th>FCM</th>
<th>Gaussian clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xie-Beni index</td>
<td>0.1656</td>
<td>0.2687</td>
</tr>
<tr>
<td>FLOPS</td>
<td>792M</td>
<td>14.1M</td>
</tr>
<tr>
<td>Time required</td>
<td>138.1 s (81 iterations)</td>
<td>14.7 s</td>
</tr>
<tr>
<td>Memory required</td>
<td>2,966,280 B</td>
<td>144 B</td>
</tr>
</tbody>
</table>

Illustrative Example: Auto MPG prediction

R1: IF weight is about 2000 KG AND year is about 1973 THEN MPG = 3 \times (\text{weight} \times \text{year})

R2: IF weight is about 2000 KG AND year is about 1974 THEN MPG = 2 \times (\text{weight} \times \text{year})

\[ \lambda_1(w, y) = \frac{w - 1800}{1000} \times \frac{y - 1973}{100} + 12 \times 1000 \]  
\[ \lambda_2(w, y) = \frac{w - 1800}{1000} \times \frac{y - 1974}{100} + 1000 \]

\[ \text{MSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2/n \]

<table>
<thead>
<tr>
<th>Method</th>
<th>3 rules</th>
<th>4 rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>trained</td>
<td>test</td>
<td>trained</td>
</tr>
<tr>
<td>Gaussian Granulation</td>
<td>2.92</td>
<td>2.18</td>
</tr>
<tr>
<td>ANFIS</td>
<td>2.97</td>
<td>2.95</td>
</tr>
<tr>
<td>FNN</td>
<td>2.96</td>
<td>2.90</td>
</tr>
<tr>
<td>SVM-N</td>
<td>2.91</td>
<td>2.95</td>
</tr>
</tbody>
</table>
Minkowski Information Granulation

- Objective: generate hyper-box shaped information granules
  - Information granules represented in terms of intervals (crisp or fuzzy)
- Approach: fuzzy clustering with Minkowski distance
  - Variation of the standard FCM
  - High order of Minkowski distance approximate Tchebychev distance, which provides for boxlike clusters

Minkowski FCM
Granule Definition

\[ [p]_\gamma = \left[ v_i - l_i^{(\gamma)}, v_i + s_i^{(\gamma)} \right] \]

\[ \phi_i (\gamma) = \sum_{m \in \gamma \text{\textbf{m}}} \left| \gamma - \text{\textbf{m}} \right| \]

\[ p = 2 \quad p = 4 \quad p = 6 \quad p = 50 \]

Granule Definition

\[ B_i = \left[ \tau_i - l_i^{(\mu)}, \tau_i + s_i^{(\mu)} \right] = [b^{-i}, b^{+i}] \]

\[ [0, 1]^n = \mathbb{R} \cup \bigcup_{i=1}^{n} B_i \]
Information Granulation through Prediction Intervals

- **Objective:** improve the interpretability of Takagi-Sugeno models, by providing uncertainty information about the predicted output.
- **Approach:** derive prediction intervals for output on the basis of the model accuracy.

### Prediction Intervals Derivation

\[ e(x) = \hat{y}(x) - \bar{y}(x) \]

\[ e_i = \frac{1}{N} \sum_{i=1}^{N} e_i \]

\[ P(\epsilon_{\text{new}} \in [L_a, U_a]) \geq 1 - \alpha \]

\[ P(\hat{y} \in [\hat{y} - U_a, \hat{y} - L_a]) \geq 1 - \alpha \]
Illustrative Example

The prediction problem consists in the estimation of amounts of chemical elements present after the combustion process in a thermoelectric generator.

- 32 inputs
- 22 outputs
- 54 examples
- Enel Produzione & Ricerca

Application: Gas Combustion Process

- The prediction problem consists in the estimation of amounts of chemical elements present after the combustion process in a thermoelectric generator.

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>Examples outside the prediction interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>1 (0.0%)</td>
</tr>
<tr>
<td>0.1%</td>
<td>1 (0.1%)</td>
</tr>
<tr>
<td>0.05%</td>
<td>0</td>
</tr>
</tbody>
</table>

IF \( x \in A \), THEN \( \hat{y} = \frac{1}{2} \left[ a_1 - U_{\alpha,A} - L_{\alpha,A} \right] \) (1 - \( \alpha \)) %
Information Granulation through Double Clustering

**Objective**: Generate interpretable fuzzy information granules from data that can be labelled with qualitative linguistic terms

**Approach**: Double Clustering
- First Clustering: find multidimensional relationships among data
- Second Clustering: aggregate multidimensional prototypes projected onto each dimension

---

Double Clustering Process
Implementations

- Fuzzy Double Clustering
  - \( (FCM + \text{Hierarchical}) \) clustering
  - Suited as wrapping procedure of existing FCM scheme

- Crisp Double Clustering
  - \( (LBG + \text{Hierarchical}) \) clustering
  - Memory saving

- DC Class
  - \( (LVQ + \text{automatic}) \) clustering
  - Suited for classification problems

- DC* (work in progress)
  - \( (LVQ + A^*) \) clustering
  - Suited for classification problems
  - Optimal granularity

  *Information granules with low dimension w.r.t. data*

Illustrative Example: Fuzzy Diagnosis with CDC

- WBC data
- 683 examples
  - Ten attributes + class
- 10-fold cross-validation

<table>
<thead>
<tr>
<th>Fuzzy sets per input</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.97%</td>
<td>1.66%</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>3.34%</td>
<td>1.66%</td>
<td>4.56%</td>
</tr>
<tr>
<td>4</td>
<td>3.22%</td>
<td>4.56%</td>
<td>4.56%</td>
</tr>
<tr>
<td>5</td>
<td>3.23%</td>
<td>2.62%</td>
<td>2.62%</td>
</tr>
<tr>
<td>6</td>
<td>4.42%</td>
<td>2.62%</td>
<td>4.71%</td>
</tr>
<tr>
<td>7</td>
<td>6.46%</td>
<td>2.88%</td>
<td>2.88%</td>
</tr>
<tr>
<td>8</td>
<td>4.41%</td>
<td>2.88%</td>
<td>4.12%</td>
</tr>
<tr>
<td>9</td>
<td>4.42%</td>
<td>2.72%</td>
<td>4.12%</td>
</tr>
<tr>
<td>10</td>
<td>5.44%</td>
<td>2.50%</td>
<td>2.97%</td>
</tr>
</tbody>
</table>

\( \text{Climat} \text{Temperature} \text{High} \) AND 
\( \text{Unintimidated} \text{Officer} \text{High} \) AND 
\( \text{Unintimidated} \text{Officer} \text{Low} \) AND 
\( \text{Small} \text{Vehilce} \text{Small} \) AND 
\( \text{Big} \text{Vehilce} \text{Large} \) AND 
\( \text{Small} \text{Cabin} \text{Small} \) AND 
\( \text{Big} \text{Cabin} \text{Large} \) AND 
\( \text{High} \text{Speed} \text{High} \) AND 
\( \text{Low} \text{Speed} \text{Low} \) AND 
\( \text{High} \text{Speed} \text{Low} \) AND 
\( \text{Low} \text{Speed} \text{High} \)

**THEN**
 \( \text{Cancer} \text{Benign} \) WITH DEGREE 0.99171,
 \( \text{Malign} \) WITH DEGREE 0.0082893.
**Fuzzy Diagnosis with DCClass**

- Average classification error: 3.975% on the test set
- Mean number of rules: 3.6

---

**Refinement of Fuzzy Information Granules through Neural Learning**

- **Objective**: define a neural architecture and a corresponding learning scheme for refining fuzzy information granules with interpretability protection
- **Approach**: change the space of parameters to be acquired through learning
  - Elements of this space always correspond to interpretable configurations
  - Interpretability preserved before, during and after learning
Parameter Space

\[ \Omega^0 = \left\{ \left( \omega_1, \omega_2, \ldots, \omega_m \right) \in \mathbb{R}^m : \omega_i \leq \omega_i^0 < \omega_i \right\} \]

\[ \mathcal{T}^0 = \left\{ \left( t_1^0, t_2^0, \ldots, t_m^0 \right) \in \mathbb{R}^{m-1} : t_i < t_i^0 < \ldots < t_i^0 < t_i^0 \right\} \]

\[ \dim \Omega = 2 \sum_{i=1}^m K_i \]

\[ \dim \Omega^* = \sum_{i=1}^m \dim \Omega^*(i) = \sum_{i=1}^m (K_i - 1) \]

Neuro-Fuzzy Architecture
Illustrative Example

Accuracy Results (WBC)

Table 9.1: Accuracy comparison of the proposed neuro-fuzzy model v.s. other approaches for the WBC dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ImmNet</td>
<td>97.1%</td>
<td>(1)</td>
</tr>
<tr>
<td>k-NN</td>
<td>97.8±0.12</td>
<td>(2)</td>
</tr>
<tr>
<td>Fisher LDA</td>
<td>96.2%</td>
<td>(2)</td>
</tr>
<tr>
<td>MLP with Hebb-prop</td>
<td>96.7%</td>
<td>(2)</td>
</tr>
<tr>
<td>LVQ</td>
<td>96.6%</td>
<td>(2)</td>
</tr>
<tr>
<td>Bayes (pairwise dep.)</td>
<td>96.4%</td>
<td>(2)</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>96.4%</td>
<td>(2)</td>
</tr>
<tr>
<td>DB-CART</td>
<td>96.2%</td>
<td>(2)</td>
</tr>
<tr>
<td>LDA</td>
<td>96.0%</td>
<td>(2)</td>
</tr>
<tr>
<td>LPC, ASL, ASR</td>
<td>94.4%–96.6%</td>
<td>(2)</td>
</tr>
<tr>
<td>CART</td>
<td>92.5%</td>
<td>(4)</td>
</tr>
<tr>
<td>Quadratic DA</td>
<td>94.5%</td>
<td>(2)</td>
</tr>
<tr>
<td>FCM, 15 rules</td>
<td>92.5%</td>
<td>(2)</td>
</tr>
<tr>
<td>SVM, 2 rules</td>
<td>92.5%±0.2%</td>
<td>(2)</td>
</tr>
<tr>
<td>NEPSCLASS-X, 3 rules, 5-6 foot</td>
<td>95.68%</td>
<td>(6)</td>
</tr>
<tr>
<td>Proposed model</td>
<td>96.05%</td>
<td>(6)</td>
</tr>
</tbody>
</table>

(1): (Jabbari and Kadirkamanathan, 1997)
(2): (Bash et al., 1998)
(3): (Stein and DeShazer, 1996)
(4): (Stein and Stein, 1990)
(5): (Stein and Kraus, 1999)
Conclusions & Future Research

- Different contributions can be unified in a single framework for interpretable fuzzy information granulation
  - HUGE
    - (Human Understandable Granule Extraction)
  - Matlab toolbox for interpretable information granulation
    - Wizard-based user interface
    - Extensibility & reuse

Conclusions & Future Research

- New research directions opened:
  - MOON
    - (Maximal Optimality degree Output iNterface)
  - DC*
    - (Double Clustering with A*)

- Fundamental issues on interpretability to be investigated
  - Cognitive, psychological and philosophical aspects
  - “System humanization”