European Centre for Soft Computing

Modeling Highly Interpretable Fuzzy Systems

José M. Alonso
jose.alonso@softcomputing.es

Index

Introduction
Motivation

Comprehensible Intelligent Systems are on demand

- **Humanistic systems:** *Those systems whose behavior is strongly influenced by human judgment, perception or emotions* (Zadeh, 1975)
- **Decision support systems:** Medicine, Economics, Robotics, etc.

Fuzzy Systems (Zadeh 1965, Mamdani 1974)

- Universal Approximators (Accuracy)
  $\Rightarrow$ System identification
- Semantic expressivity (Interpretability)
  $\Rightarrow$ Knowledge extraction and representation

Introduction

Accuracy vs. Interpretability

Accuracy

- How similar are the outputs of the model and the real system ?

Interpretability

- Comprehensibility, intelligibility, transparency, understandability, readability, etc.
- Is the model (description and behavior) understandable (to a human) ?
  - **Description ⇒ System structure readability (transparency)**
  - **Explanation ⇒ System comprehensibility (understandability)**

Introduction
Accuracy vs. Interpretability (History)

Interpretability - Accuracy (Fuzzy Logic)

- [1965] Fuzzy Logic (Zadeh)
- [1965 – 1990] **Interpretability** (I)
  - Simple linguistic rules with high interpretability
  - Expert knowledge
- [1990 – 2000] **Accuracy** (A)
  - Complicated fuzzy rules with high accuracy
  - Induced knowledge
- [2000 – 2010] **I-A Trade-off**
  - Simple linguistic rules with high accuracy
  - Expert + Induced knowledge ?
  - **Characterizing and assessing Interpretability**
  - **Looking for useful Interpretability indices**

Introduction

Regarding Interpretability in terms of complexity

Interpretability-Accuracy Trade-off

- Contradictory goals: **Looking for a good compromise** (multi-objective optimization techniques)
- *As the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance become almost mutually exclusive characteristics. The closer one looks at a real-world problem, the fuzzier becomes its solution* **(Principle of incompatibility, Zadeh 1973)**

Introduction
Regarding Interpretability in terms of complexity

Fuzzy Modeling (FM)

- Linguistic Fuzzy Modeling (LFM)
  1. Maximizing Interpretability
  2. Improving Accuracy
- Precise Fuzzy Modeling (PFM)
  1. Maximizing Accuracy
  2. Improving Interpretability
- **Model Refining**
  1. **Extending the modeling process**
     (new algorithms for learning partitions and rules)
  2. **Extending the model structure**
     (linguistics modifiers, weights, exceptions, etc.)
- **How to characterize and evaluate Interpretability ?**

Index

Characterization of Interpretability

Definitions (I)

- Assuming T and S are formal theories, T is said to be interpretable in S $\Leftrightarrow$ the language of T can be translated into the language of S in such a way that S proves the **translation** of every theorem of T **(Tarski 1953)**
- Interpretability means the possibility to **estimate** the system's behavior by **reading** and **understanding** its rule base **(Bodenhofer and Bauer 2003)**

Characterization of Interpretability

Definitions (II)

- Assessing Interpretability of a FS $\equiv$ measuring the complexity of making the **translation** from L (model **description** based on **FL**) to L' (model **explanation** based on **NL**) **(Mencar et al. 2005)**
- Comprehensibility Postulate **(Michalski 1983)**
  + Notion of Cointension **(Zadeh 2005)**
  $\Rightarrow$ Understandability of patterns **(Mencar et al. 2007)**
  $\Rightarrow$ Cointensive Interpretability **(Mencar et al. 2009)**

Interpretability must be a central point in system modeling

- **PNL** $\Rightarrow$ Precisiated Natural Language
- **CWW** $\Rightarrow$ Computing with Words
- **HCC** $\Rightarrow$ Human Centric Computing

Conceptual Framework (I)

Two different points of view

1. **Description (system structure readability)**
   - Number of variables, rules, linguistic terms, etc.
   - Regarding interpretability in terms of complexity:
     $\Rightarrow$ The more compact KB, the simpler its understanding
     $\Rightarrow$ Lower complexity means higher interpretability
2. **Explanation (system comprehension)**
   - Inference level (fuzzy operators, rules fired at the same time, etc.)
   - Regarding interpretability in terms of complexity:
     $\Rightarrow$ The more compact KB, the more rules fired at the same time
     $\Rightarrow$ Lower complexity means lower interpretability
   - Contradictory goals
   - **Cointensive Interpretability - Logical View (Mencar et al. 2009)**

Conceptual Framework (II)

Interpretability of Fuzzy Rule–Based Systems

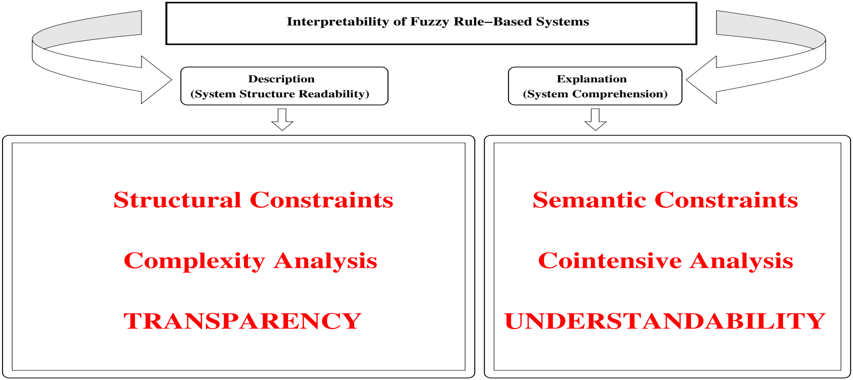
Description (System Structure Readability)

Levels		
High–Level	**Rule Base**	–> Number of rules –> Rule length (total number of premises)
	Linguistic Fuzzy Rules	–> Number of inputs by rule –> High–order consequents
	Linguistic Propositions	–> Description length –> Attribute correlation –> Elementary linguistic terms –> Composite (OR / NOT) linguistic terms –> Linguistic modifiers
	Linguistic Partitions	–> Global semantics (shared fuzzy sets) –> Matching meaninful linguistic terms and fuzzy sets (problem context dependant)
	Fuzzy Partitions	–> A justifiable number of fuzzy sets (magical number: seven, plus or minus two) –> Ordering –> Distinguishability (overlapping) –> Coverage (completeness) –> Complimentarity –> Uniform granulation –> Leftmost / rightmost –> Natural zero positioning
Low–Level	**Fuzzy Sets**	–> Normalization –> Convexity –> Unimodality (Prototype)

Explanation (System Comprehension)

–> Inference mechanism and rule aggregation (FITA / FATI)
–> Conjunction operator
–> Disjunction operator
–> Defuzzification (output interpretation)
–> Number of simultaneously fired rules
–> Consistency
–> Completeness
–> Rule structure (Mamdani / TSK, weights, exceptions)
–> Number of variables (inputs / outputs)
–> Rule interpretation (conjunctive / implicative)
–> Modus Ponens / Tollens
–> Negation
–> Antonyms
⋮

Conceptual Framework (II)



Index

Starting point (I)

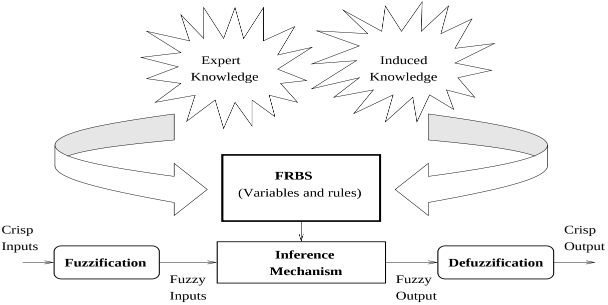
Proposal

- **HILK: Highly Interpretable Linguistic Knowledge** [IJIS 2008]
- Looking for a good **interpretability-accuracy trade-off** when modeling **fuzzy rule-based classifiers** (FRBCs)
  - KBCT
    Knowledge Base Configuration Tool ⇒ freeSW
  - GUAJE
    A Java Environment for Generating Understandable and Accurate Fuzzy Models ⇒ freeSW toolbox
  - **HILKMO**
    Embedding HILK into a Multi-objective evolutionary algorithm

Starting point (II)

HILK (Highly Interpretable Linguistic Knowledge) [IJIS 2008]

- Expert + Induced Knowledge
  - **Partition design**
  - **Rule base learning**
  - **Knowledge base improvement**



Expert Knowledge

Advantages

- General knowledge
  ⇒ Experience, education and training, several disciplines
- Global view of the problem
  - Most influential variables
  - Universal rules (involving a few variables)
- High Interpretability

Drawbacks

- Expert knowledge acquisition is a hard task (bottleneck)
- Interaction between variables is difficult to formalize

Induced Knowledge

Advantages

- Automatic learning (knowledge discovery)
- Finding out interaction between variables
- High Accuracy

Drawbacks

- Specific knowledge
  - Rule generality depends on available data
  - Non-universal rules
- Interpretability depends on the learning technique
- Collecting representative data is expensive (time and money)

Expert + Induced Knowledge

Cooperation (Integration)

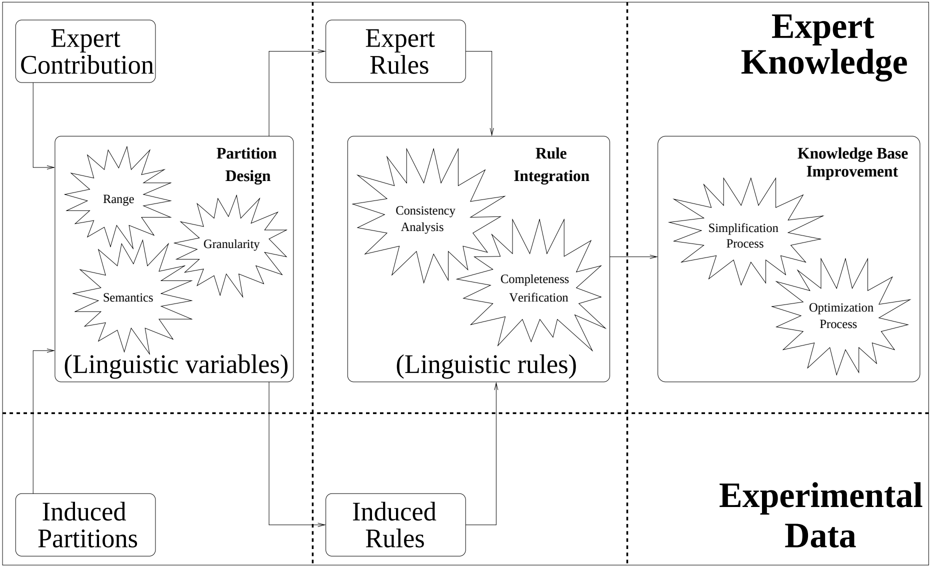
- Both kinds of knowledge convey complementary information
- Their combination is likely to yield compact and robust systems
- Several options:
  - First Expert $\Rightarrow$ Then Data
  - First Data $\Rightarrow$ Then Expert
  - **Iterative approach**

Expert + Induced Knowledge (Fuzzy Logic)

Cooperation (Framework)

- **Fuzzy Logic (FL) represents both kinds of knowledge under the same formalism**
  - Linguistic variables and rules (Mamdani)
  - Comparison at linguistic level
  - Automatic learning methods.
- **FL semantic expressivity is close to natural language**
  - FL favors the interpretability of the final model (but it is not enough to guarantee it)

HILK (expert + data)



KBCT (Knowledge Base Configuration Tool)

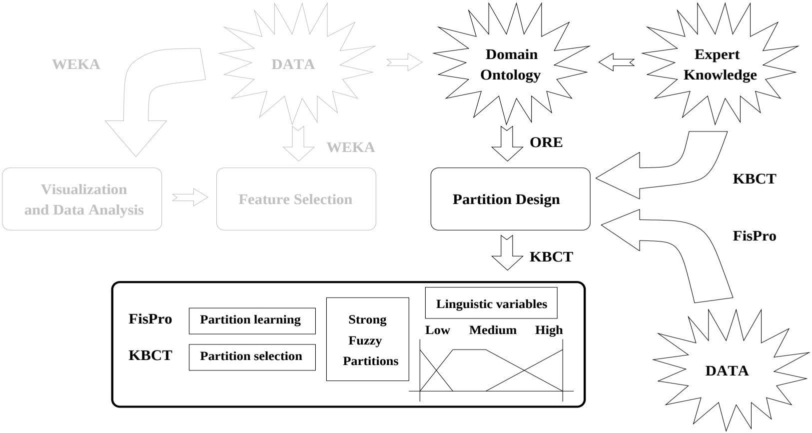
- Open-source (Free-software)
- Portable (Linux / Windows)
- User-friendly (Java Interface)
- Documentation
  - On-line documentation (HTML)
  - User Manual (PDF)
  - Java API
- KBCT
  $\Rightarrow$ FisPro C++ library
  $\Rightarrow$ Weka Java library
- Version 3 freely available at
  http://www.mat.upm.es/projects/advocate/kbct.htm

GUAJE (a Java Environment for Generating Understandable and Accurate Fuzzy Models)

- **KBCT (Knowledge Base Configuration Tool)**
- **FisPro (Fuzzy Inference System Professional)**
- ORE (Ontology Rule Editor)
- jMetal (Metaheuristic Algorithms in Java)
- Weka (Data Mining)
- Xfuzzy (Fuzzy Modeling)
- Matlab Fuzzy Toolbox

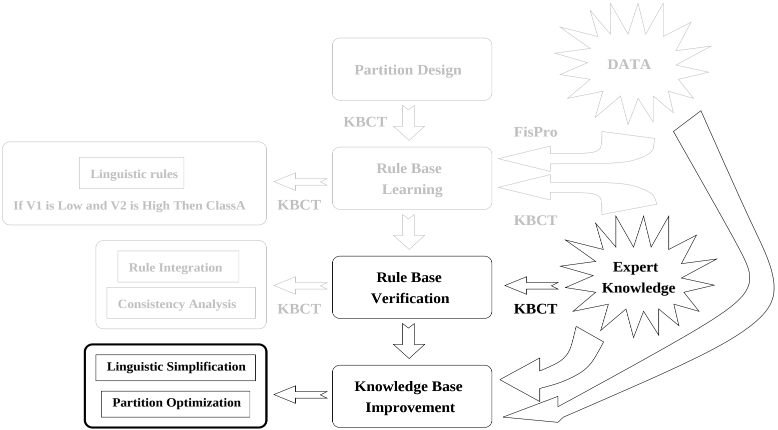
GUAJE

Combining several tools (Partition Design)



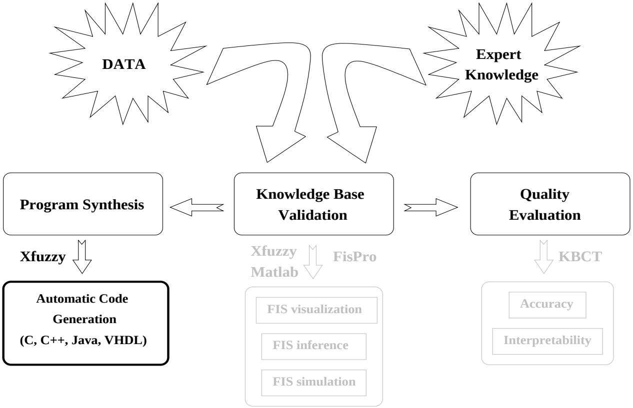
GUAJE

Combining several tools (RB verification and KB improvement)

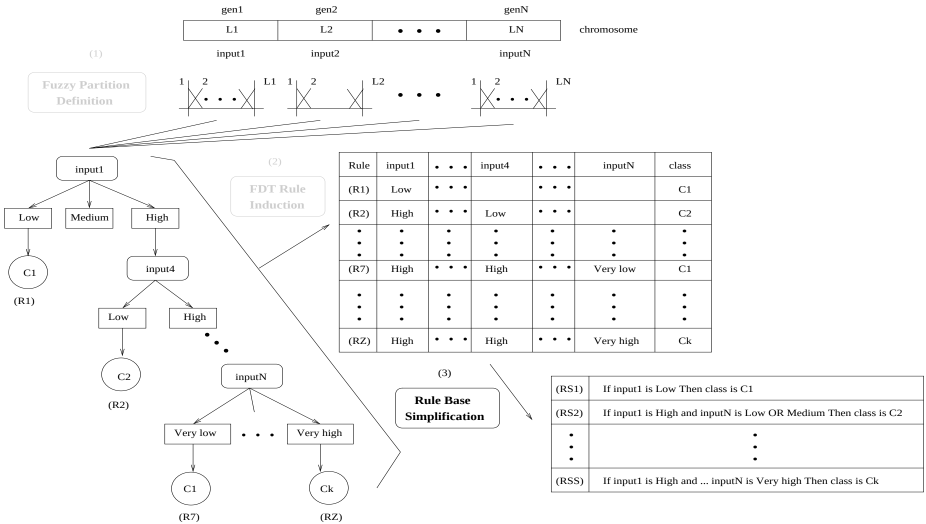


GUAJE

Combining several tools (KB validation)



HILKMO: Embedding HILK in a three-objective evolutionary algorithm



HILKMO: Three-objective evolutionary algorithm

NSGA-II

- Initial population randomly generated
- Binary tournament selection
- Two point crossover and Thrift mutation
- Pareto ranking with crowding distance measure
- Elitist replacement update procedure

Objective functions

- **Maximizing accuracy** (classification rate)
- Maximizing interpretability
  - **Maximizing readability** (rule base complexity)
  - **Maximizing comprehensibility** (average fired rules)

HILKMO: Experimentation (Problem description)

GLASS identification problem

Instances	Attributes	Classes	Class distribution
214	9	6	G1 (32.71%), G2 (35.51%), G3 (7.94%), G4 (6.074%), G5 (4.205%), G6 (13.561%)

- **Attributes**: RI, Na, Mg, Al, Si, K, Ca, Ba, Fe
- **UCI**: http://www.ics.uci.edu/-mlearn/MLSummary.html
- **5-fold cross-validation** (5CV): 80% (training) - 20% (test)

HILKMO: Experimentation (Parameter configuration)

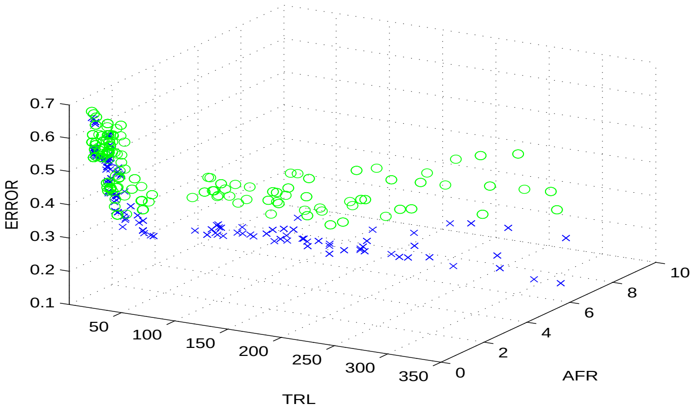
HILKMO - NSGAII

- Six runs for each training-test pair (6x5CV)
- P_s = 30 indviduals
- 12000 evaluations
- P_c = 0.6 (Two-point crossover)
- P_m = 0.1 (Thrift mutation)

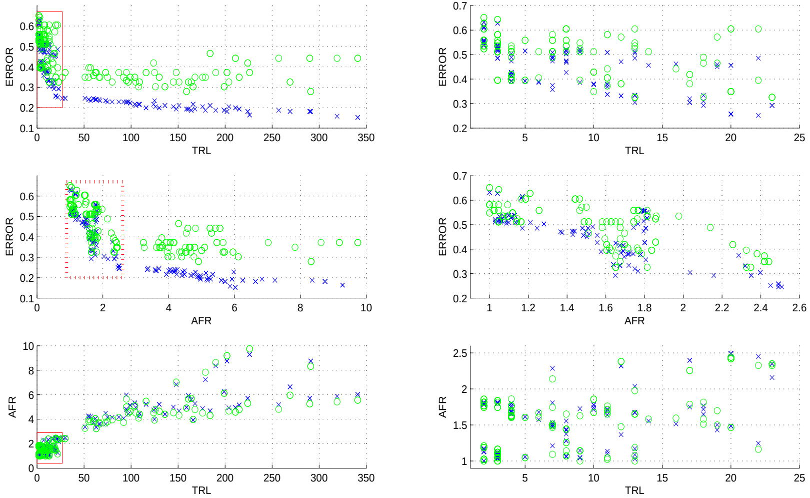
HILKMO - FRBC

- **Partition design**: Uniform SFPs for all inputs
- **Rule base definition**: pruned FDT (tolerance 30%)
- **Simplification**: TRL $<=$ 50

HILKMO: Aggregated Pareto Front Approximation
(P1 + ... + P30 - DS)



HILKMO: Aggregated Pareto Front Approximation
(Projection and Zoom)

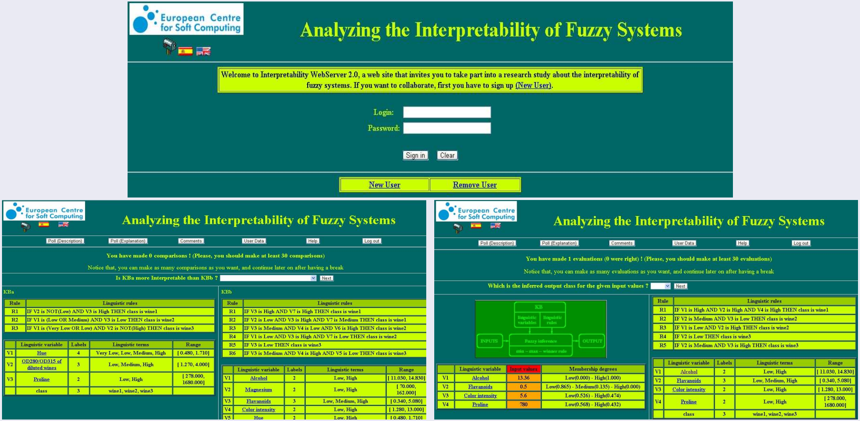


Index

GUAJEPOLL

Analyzing the Interpretability of Fuzzy Systems
https://apps.softcomputing.es/guajepoll/

Web poll



GUAJEPOLL

Analyzing the Interpretability of Fuzzy Systems
https://apps.softcomputing.es/guajepoll/

Description of the experiments

- **WINE problem**: 13 inputs and 1 output (3 classes)
- **HILK methodology** (IJIS 2008)
  $\Rightarrow$ six KBs of several sizes (with SFPs)
- **Comparing several interpretability indices**
- How to know which index is the best one ?
  $\Rightarrow$ ... asking people !
  $\Rightarrow$ The survey was addressed to
  fuzzy experts (50%) and **naive users (50%)**
  1. **How much interpretable are the analyzed KBs ?**
  2. **What is the best KB interpretability ranking ?**
  3. **What are the most relevant aspects when assessing interpretability ?**

GUAJEPOLL

Analyzing the Interpretability of Fuzzy Systems
https://apps.softcomputing.es/guajepoll/

Some preliminary conclusions (regarding readability)

- People get into **difficulties giving numerical indices**
- People find **much more natural to make approximate reasoning based on the use of linguistic terms** (*Highly interpretable*, *Moderately interpretable*, etc.)
- People feel much *more confidence setting rankings than giving numerical values*
- When two KBs are quite close regarding readability the final ranking choice depends in many subtle details, and as result, at the end there is a clearly subjective choice
- Because of this subjectivity there is a huge diversity of answers
- **Objectivity** (fair comparison) vs. **Subjectivity** (personalization)

Index

Publications

Theoretical and experimental analysis

PhD dissertation (October 2007)

- **Interpretable fuzzy systems modeling with cooperation between expert and induced knowledge**
  $\Rightarrow$ http://oa.upm.es/588/

Fuzzy system modeling

- A new index easily adaptable to the context of each problem by means of incorporating user's preferences and quality criteria (WCCI 2010)
- Multiobjective fuzzy system modeling
  $\Rightarrow$ interpretability vs. accuracy (GEFS 2010)
- Characterizing and measuring interpretability (IJAR 2009)
- HILK++: an enhanced version of HILK (SC 2010, ISDA09)

Publications

Theoretical and experimental analysis

Fuzzy system modeling (Knowledge extraction and representation)

- Ontology (ESTYLF 2008)
- Expert $\oplus$ Data (IJIS 2008)
- Consistency analysis (IJIS 2008)
- Accuracy improvement $\Rightarrow$ Optimization (FUZZ-IEEE 2007)
- Interpretability improvement $\Rightarrow$ Simplificationn (Mathware 2006)
- KBCT (FUZZIEEE 2004)

Publications

Theoretical and experimental analysis

Real-world applications

- **Human activity recognition fusing intensity of WiFi signal and accelerations** (WCCI 2010)
- **WiFi localization with robots** (ECSC-UAH, EUROCAST 2009)
- **An intelligent agent that analyzes data from medical devices for the management of Diabetes Mellitus patients** (ECSC-GBT, AIME 2009)
- **ADVOCATEII** $\Rightarrow$ **Avoiding undetectable obstacles by robots** (UPM-UAH, JRIS 2007)
- **Real-Time System for Monitoring Driver Vigilance** (UAH, IEEE Trans on ITS 2006)

Conclusions and Future Works

Conclusions

- Regarding **interpretability** in terms of **readability** (transparency) and **comprehensibility** (understandability)
- Quality-guided design of **Highly Interpretable Fuzzy Systems**
- Experimental analysis (web poll) **https://apps.softcomputing.es/guajepoll/**

Feature works (IFS ≡ Interpretable Fuzzy Systems)

- Organizing **Panel Session** and **Special Session** (July 21-22) during the IEEE WCCI 2010 (Barcelona, Spain)
- Editing a **Journal Special Issue** (Information Sciences, ELSEVIER)
- Writing a **co-authored book** (Authors: J. M. Alonso, C. Castiello, L. Magdalena, and C. Mencar)

THANKS FOR LISTENING !