

Ninth International Conference

# IPMU 2002

Information Processing and Management of  
Uncertainty in Knowledge-based Systems

Annecy - France  
July 1-5, 2002

Proceedings  
Volume I



**Proceedings**

**Ninth International Conference**

**IPMU 2002**

**Information Processing and  
Management of Uncertainty  
in Knowledge-Based Systems**

**July 1-5, 2002  
Volume I**

**Organized by Université de Savoie, LISTIC - ESIA  
ANNECY, FRANCE**

© 2002, ESIA - Université de Savoie, France

All rights reserved. No part of this book may be reproduced, in any form or by any means, without the permission of the editor.

Editor: ESIA - Université de Savoie

Printed in France

ISBN Vol. 1: 2-9516453-1-7

Dépot légal - Juillet 2002

Imprimé par Gutenberg

To order this book, mail to [ipmu2002@univ-savoie.fr](mailto:ipmu2002@univ-savoie.fr)

## HONORARY PRESIDENT

Lotfi A. ZADEH (University of California at Berkeley)

## CHAIRPERSONS

Bernadette BOUCHON-MEUNIER (CNRS, Université Paris VI, France)

Laurent FOULLOY (Université de Savoie, France)

Ronald R. YAGER (Iona College, USA)

## ORGANIZING COMMITTEE

LISTIC-ESIA, Université de Savoie

G. Mauris (President), I. Alloui, E. Benoit, L. Berrah, S. Galichet

## INTERNATIONAL PROGRAM COMMITTEE

- |                            |                           |                         |
|----------------------------|---------------------------|-------------------------|
| J. Aczel (Canada)          | S. Guiasu (Canada)        | E. Ruspini (USA)        |
| J. Aguilar-Martin (France) | J. Gutiérrez-Rios (Spain) | E. Sanchez (France)     |
| A. Appriou (France)        | K. Hirota (Japan)         | R. Scozzafava (Italy)   |
| J. Baldwin (U.K.)          | J. Kacprzyk (Poland)      | G. Shafer (USA)         |
| H. Berenji (USA)           | A. Kandel (USA)           | P. Shenoy (USA)         |
| P. Besnard (France)        | E. Klement (Austria)      | R. Slowinski (Poland)   |
| J. Bezdek (USA)            | G. Klir (USA)             | P. Smets (Belgium)      |
| I. Bloch (France)          | T. Kohonen (Finland)      | M. Sugeno (Japan)       |
| P. Borne (France)          | R. Kruse (Germany)        | H. Thiele (Germany)     |
| P. Bosc (France)           | H. Larsen (Denmark)       | C. Tijus (France)       |
| G. Coletti (Italy)         | L. Magdalena (Spain)      | A. Titli (France)       |
| P. Diamond (Australia)     | C. Marsala (France)       | L. Travé-Massuyès (Fr.) |
| M. Delgado (Spain)         | H. Nguyen (USA)           | E. Trillas (Spain)      |
| T. Dencœux (France)        | S. Ovchinnikov (USA)      | I.B. Turksen (Canada)   |
| D. Dubois (France)         | Z. Pawlak (Poland)        | L. Valverde (Spain)     |
| F. Esteva (Spain)          | W. Pedrycz (Canada)       | A. Ventre (Italy)       |
| M. Fedrizzi (Italy)        | H. Prade (France)         | J. Verdegay (Spain)     |
| D. Fogel (USA)             | A. Ralescu (USA)          | M. Vila (Spain)         |
| F. Gomide (Brazil)         | D. Ralescu (USA)          | H. Zimmermann (Germany) |
| M. Grabisch (France)       | M. Rifqi (France)         |                         |
| S. Grossberg (USA)         | A. Rocha (Brazil)         |                         |

## ORGANIZERS OF SPECIAL SESSIONS

F. d'Alche-Buc  
J. Benois-Pineau  
I. Bloch  
G. Coletti  
R. Decaluwe  
G. De Tre  
E. Diday  
M. Grabisch

F. Herrera  
E. Hullermeier  
C. Marsala  
M.J. Martin-Bautista  
R. Mesiar  
E. Pap  
N. Perrot  
L. Polkowski

O. Pons  
A. Ralescu  
D. Ralescu  
M. Ramdani  
A.F. Rocha  
U. Sandler  
E. Szmidt  
C. Tijus  
X. Zeng

## REVIEWERS

(Other non-program committee members)

H. Akdag  
F. d'Alche-Buc  
A. Attoui  
C. Barret  
S. Benferhat  
J. Benois-Pineau  
G. Bordogna  
R. Boukezzoula  
S. Boverie  
J. Buckley  
L. Cauffriez  
C. Cayrol  
S. Cimpan  
R. Dapoigny  
B. De Baets  
M. Detyniecki  
L. De Campos  
G. De Cooman  
J. Fauqueur  
H. Fargier  
J. Fodor  
P.M. Frank  
C. Frelicot  
L. Gacogne  
S. Gentil  
M.A. Gil

L. Godo  
Y. Grandvalet  
T.M. Guerra  
Z. Guessoum  
M. Herbin  
F. Herrera  
E. Herrera-Viedma  
M.C. Jaulent  
P. Joly  
L. Kam  
L. Koczy  
V. Kreinovich  
L. Kuncheva  
P. Lambert  
J. Lawry  
E. Levrat  
L. Lietard  
R. Lopez De Mantaras  
D. Luzeaux  
J.L. Marichal  
R. Martin-Clouaire  
M. Masson  
R. Mesiar  
P.A. Monney  
S. Moral  
N. Mouaddib

V. Novak  
A. Nürnbergger  
F. Oquendo  
E. Pairel  
R. Palm  
E. Pap  
G. Pasi  
J.I. Pelaez  
I. Perfilieva  
O. Pivert  
L. Polkowski  
L. Ralaivola  
M. Rombaut  
L. Saitta  
S. Sandri  
H. Sawada  
A. Skowron  
E. Szmidt  
S. Termini  
V. Torra  
E. Trouve  
L. Ughetto  
D. Van De Ville  
F. Vernadat  
S. Zadrozny

## ACKNOWLEDGEMENTS

The organizers of the 9th IPMU Conference would like to thank all institutions and people without them this conference could not be held.

For their sponsorship

- Association Française pour l'Intelligence Artificielle (AFIA)
- European Society for Fuzzy Logic and Technology (EUSFLAT),
- GDR Information, Signal, Images et ViSion (ISIS)
- Institute of Electrical and Electronics Engineers (IEEE),
- International Fuzzy Systems Association (IFSA)
- World Organisation of Systems and Cybernetics (WOSC)

For their support

- Laboratoire d'Informatique de Paris 6 (LIP6)
- Ecole Supérieure d'Ingénieurs d'Annecy (ESIA)
- Université de Savoie
- Ministère de l'Education Nationale
- Ministère des Affaires Etrangères
- Centre National de la Recherche Scientifique (CNRS)
- Chambre de Commerce et d'Industrie (CCI)
- Communauté d'Agglomération d'Annecy (CAA)
- Conseil Général de la Haute-Savoie
- Région Rhône-Alpes

For their help in the organisation

- Pôle Mécatronique et Management de Haute-Savoie (THESAME)

For their help in preparing the conference

- The members of the International Program Committee
- The organizers of special sessions
- The members of the organizing committee
- The session chairpersons
- The referees
- and all those who helped

# Table of Contents

<b>Plenary Session I</b>	<b>1</b>
<b>Arthur P. DEMPSTER, Harvard University (USA)</b>	
Theory of belief functions: history and prospects	3
<b>Computing with words: Models and Applications - I</b>	<b>5</b>
Linguistic Modelling of Imperfect Spatial Information in a Fuzzy Database	7
<i>G. Bordogna, S. Chiesa, D. Geneletti</i>	
Fusion of Nuclear Safeguards Indicator Information Based on an Ordinal Linguistic Approach	15
<i>J. Liu, D. Ruan</i>	
An Information Retrieval System with Unbalanced Linguistic Information Based on the Linguistic 2-tuple Model	23
<i>F. Herrera, E. Herrera-Viedma, L. Martinez</i>	
<b>Networks</b>	<b>31</b>
Modified Algorithm for Fuzzy Bayesian Networks Inference	33
<i>J.F. Baldwin, E. Di Tomaso</i>	
Validation of Diagnostic Models Using Graphical Belief Networks	39
<i>O. Kipersztok</i>	
Computing Diagnoses With Higher Posterior Probability Using Bayesian Networks	45
<i>V. Delcroix, S. Piechowiak, J. Rodriguez</i>	
<b>Automatic Methods for Odor and Taste Recognition</b>	<b>53</b>
A fuzzy rule base for modelling red wine color in relation with winemaking parameters	55
<i>S. Guillaume, B. Charnomodic</i>	
Man-machine interaction for odor prediction	63
<i>M. Kissi, B. Bouchon-Meunier, M. Ramdani, D. Zakarya</i>	
Les nez électroniques, entre sensoriel et mesure	69
<i>P. Grenier</i>	
<b>Argumentation</b>	<b>75</b>
Detecting Conflict-Free Argumentative or Abductive Knowledge Bases	77
<i>R. Haenni</i>	
Gradual handling of contradiction in argumentation frameworks	83
<i>C. Cayrol, M.C. Lagasque</i>	

Knowledge management as a support for collective decision-making and argumentation processes	91
<i>J. Montmain, A. Akharraz, G. Mauris</i>	
<b>Learning and Mining</b>	<b>99</b>
Mining Implication-Based Fuzzy Association Rules in Databases	101
<i>E. Hullermeier</i>	
Acquiring an expert's obvious knowledge by using fuzzy repertory tables	109
<i>J.J. Castro-Schez, J.L. Castro, J.M. Zurita</i>	
Learning Rules from Multiple Instance Data: Issues and Algorithms	117
<i>Y. Chevalyere, N. Bredeche, J.D. Zucker</i>	
Mining Weighted Linguistic Association Rules	125
<i>S.L. Wang, T.P. Hong, M.J. Chiang</i>	
Continuous Method for Discovering Quantitative Association Rules in Databases	131
<i>A. Shragai, M. Schneider</i>	
Using PCA paradigm to mine functional data	139
<i>D. Clot, S. Bonnevey, M. Lamure</i>	
<b>Fuzzy Relations</b>	<b>147</b>
Normal forms for fuzzy relations and their contribution to universal approximation	149
<i>I. Perfilieva</i>	
Fuzzified properties of fuzzy relations	157
<i>J. Jacas, J. Recasens</i>	
Numerical representations of fuzzy relational systems	163
<i>S. Ovchinnikov</i>	
On Clustering and Fuzzy Relations Decomposition	167
<i>M. Wagenknecht, V. Schneider</i>	
Degrees of Contradiction in Fuzzy Sets Theory	171
<i>E. Castineira, S. Cubillo, S. Bellido</i>	
Inclusion relations and orthogonality relations over property systems: first-order characterization and modal analysis	177
<i>P. Balbiani</i>	
<b>Pattern Recognition</b>	<b>185</b>
Values for the fuzzy C-means classifier in change detection for remote sensing	187
<i>P. Deer, P. Eklund</i>	
A Hierarchical Linguistic Clustering Algorithm for Prototype Induction	195
<i>I. Gonzalez Rodriguez, J. Lawry, J.F. Baldwin</i>	



An Investigation into How ADABOOST Affects Classifier Diversity <i>C.A. Shipp, L.I. Kuncheva</i>	203
A statistic method to compare dynamic systems in the state space <i>P. Loonis, P. Franco</i>	209
Fuzzy Clustering Algorithm Based on the Maximum Penalized Likelihood for Latent Class Model <i>C.B. Chen, C.T. Lin, W.H. Wu</i>	217
A new fuzzy clustering technique based on pdf estimation <i>J. Cutrona, N. Bonnet, M. Herbin</i>	225
<b>Rough Sets Methods and Applications - I</b>	<b>233</b>
Learning Rules from Very Large Databases: A Rough Multiset Approach <i>C.C. Chan</i>	235
MLEM2: A New Algorithm for Rule Induction from Imperfect Data <i>J.W. Grzymala-Busse</i>	243
Approximation of Functions by Means of Fuzzy Rough Sets <i>M. Inuiguchi, T. Tanino</i>	251
Neighborhood Systems and Qualitative Fuzzy Sets <i>T.Y. Lin, S. Tsumoto</i>	259
On A-exact and A-rough sets in potentially infinite information systems <i>L. Polkowski, M. Polkowska</i>	267
Extraction of Hierarchical Decision Rules from Medical Databases using Rough Set Model <i>S. Tsumoto</i>	273
<b>Intelligent Information Systems - I</b>	<b>281</b>
A big-stepped probability approach for discovering default rules <i>S. Benferhat, D. Dubois, S. Lagrue, H. Prade</i>	283
A mechanism for deduction in a fuzzy relational database <i>I.J. Blanco, M.J. Martin-Bautista, O. Pons, M.A. Vila</i>	291
Closed Set Based Discovery of Maximal Covering Rules <i>M. Kryszkiewicz</i>	299
Using Object Relational Features to Build a Fuzzy DataBase Server <i>J.M. Medina, J. Galindo, F. Berzal, J.M. Serrano</i>	307
Flexible Unary Multidimensional Queries and their Combinations <i>A. Laurent, B. Bouchon-Meunier, A. Doucet</i>	315

<b>Modeling and Fuzziness</b>	<b>323</b>
Uncertainty, Type-2 Fuzzy Sets, and Footprints of Uncertainty	325
<i>J.M. Mendel</i>	
Fuzzy Prototypical Knowledge Discovery to Predict Information Systems Maintainability	333
<i>J.A. Olivas, M. Genero, M. Piattini, F.P. Romero</i>	
Combining Fuzzy Models Using Prototype Based Reasoning	341
<i>R.R. Yager</i>	
Interval Approximation of a Fuzzy Number and the Principle of Information Invariance	347
<i>P. Grzegorzewski</i>	
On Approximate Reasoning with Type-2 Fuzzy Sets	355
<i>H. Thiele</i>	
<b>Fuzzy Arithmetics and the Brain</b>	<b>363</b>
Fuzzy Arithmetic: An Overview	365
<i>F. Gomide</i>	
Mental processes of arithmetic calculation	369
<i>A.F. Rocha, F.T. Rocha</i>	
Implementing arithmetical knowledge in a distributed intelligent processing system	373
<i>E. Massad, A.F. Rocha</i>	
Difference of area of the brain for fuzzy and crisp calculations	377
<i>T. Yamanoi, M. Saito, M. Sugeno, E. Sanchez</i>	
<b>Non Classical Logic</b>	<b>383</b>
Models and submodels of fuzzy theories	385
<i>V. Novak</i>	
A new approach to completeness for multi-adjoint logic programming	391
<i>J. Medina, M. Ojeda-Aciego</i>	
Consistency degrees in fuzzy logics	399
<i>R. Horcik, M. Navara</i>	
Which fuzzy logics satisfy the compactness property	405
<i>P. Cintula, M. Navara</i>	
Multi-agent Based Method for Reactive Systems Formal Specification and Validation	411
<i>A. Hasbani</i>	
<b>Plenary Session II</b>	<b>417</b>
<b>Lotfi A. Zadeh, University of California, Berkeley (USA)</b>	
It is a Fundamental Limitation to Base Probability Theory on Bivalent Logic	419

<b>Decision Making</b>	<b>421</b>
Markov Decision Process for the Task Selection Problem in Uncertain Environments	423
<i>H. Hanna, A.I. Mouaddib</i>	
A Hybrid Fuzzy-Fractal Approach for Time Series Analysis and Its Applications to Plant Monitoring and Diagnosis	431
<i>O. Castillo, P. Melin</i>	
Managing Heterogeneous Information in Group Decision Making	439
<i>F. Herrera, L. Martinez, P. Sanchez, E. Herrera-Viedma</i>	
Batch Process and Market Entry/Exit Decision Making by Real Options under Exchange Rate Uncertainty	447
<i>C.T. Lin, C.R. Wu</i>	
Predicting Run Times of Applications Using Rough Sets	455
<i>S. Krishnaswamy, A. Zaslavsky, S.W. Loke</i>	
Decision Making for Discrete-Time American Options with Uncertainty of Stock Prices in Financial Engineering	463
<i>Y. Yoshida, M. Yasuda, J. Nakagami, M. Kurano</i>	
<b>Intelligent Systems for Video Processing</b>	<b>471</b>
Extraction of regions of interest based on motion activity for video retrieval with partial query	473
<i>R. Fablet, P. Boutheymy</i>	
Real time commercial detection using MPEG features	481
<i>N. Dimitrova, S. Jeannin, J. Nesvadba, T. Mcgee, L. Agnihotri, G. Mekenkamp</i>	
General Purpose Real-Time Object Tracking Using Hausdorff Transforms	487
<i>D. Vignon, B.C. Lovell, R.J. Andrews</i>	
Uncertainty of Motion Estimation and its Application to Video Processing	493
<i>D. Wang, A. Vincent</i>	
Adding the concept of video editing levels in shot segmentation	501
<i>R. Ruiloba, P. Joly</i>	
Comparing descriptions of multimedia data for simplification and integration	507
<i>N. Adami, M. Corvaglia, R. Leonardi</i>	
<b>Fuzzy Measures and Integrals - I</b>	<b>515</b>
Ordinal Integration	517
<i>G.A. Koshevoy</i>	
Almost-Measurability Induced by Fuzzy and Possibilistic Measures	521
<i>I. Kramosil</i>	
Interaction indices for games with forbidden coalitions	529
<i>C. Labreuche</i>	

Mobius transform and k-order additivity	535
<i>R. Mesiar</i>	
The n-step Choquet integral on finite spaces.	539
<i>Y. Narukawa, T. Murofushi</i>	
p-symmetric fuzzy measures	545
<i>P. Miranda, M. Grabisch</i>	
<b>Belief Functions</b>	<b>553</b>
Towards another logical interpretation of Theory of Evidence and a new combination rule	555
<i>L. Cholvy</i>	
Showing why measures of quantified beliefs are belief functions	563
<i>P. Smets</i>	
Clustering belief functions based on attracting and conflicting metalevel evidence	571
<i>J. Schubert</i>	
A Pre-Pruning Method in Belief Decision Trees	579
<i>Z. Elouedi, K. Mellouli, P. Smets</i>	
Combination of Belief Functions and Coarsening / Refinement	587
<i>M. Daniel</i>	
Dempster-Turksen Multi-Valued Mapping	595
<i>I.B. Turksen</i>	
<b>Fuzzy Sets and Possibility Theory in Machine Learning and Data Mining</b>	<b>601</b>
Learning Graphical Models by Extending Optimal Spanning Trees	603
<i>C. Borgelt, R. Kruse</i>	
Clustering of proximity data using belief functions	609
<i>T. Denoeux, M. Masson</i>	
Stratified induction in possibilistic logic	617
<i>D. Dubois, H. Prade, G. Richard, M. Serrurier</i>	
Discovering and Incrementally maintaining fuzzy association rules	625
<i>S. Ben Yahia</i>	
Data Mining and Computational Intelligence	633
<i>M. Last</i>	
<b>Defuzzification and Averaging of Fuzzy Sets</b>	<b>641</b>
Quantitative Revision of Quantitative Beliefs	643
<i>A. Ramer</i>	
Linear Non-additive Set-functions	651
<i>B. Bouchon-Meunier, R. Mesiar, D. Ralescu</i>	

Averaging procedures in defuzzification processes	659
<i>E. Roventa, T. Spiricu</i>	
Ordering and bounds in specific classes of aggregation operators	667
<i>T. Calvo, R. Mesiar</i>	
Vague Verbal Quantities in Transport Timetable	675
<i>M. Mares</i>	
Postmodernism, Cybernetics and Fuzzy Set Theory	681
<i>C.V. Negoita</i>	
<b>Applications of Soft Computing</b>	<b>687</b>
A probabilistic approach to analyse the evolutionary process in circuit design	689
<i>T. Kalganova, I. Baradavka</i>	
Managing Uncertainty in New Product Development	697
<i>G. Buyukozkan, O. Feyzioglu</i>	
Evolutionary techniques for mobile-robot office delivery service	705
<i>S. Badaloni, P. Bison</i>	
Some Attempts in Improving Cochlear Implanted Patients Performances: Modeling and Automatic Methods	711
<i>M. Costin, M. Zbancioc, A. Ciobanu, C. Berger Vachon</i>	
Global Optimization of Neural Networks	719
<i>G. Beliakov</i>	
Computer Assisted Diagnosis of Cardiac Heart Disease	727
<i>S. Zahan, R. Bogdan, M. Cremene</i>	
<b>Fuzzy Measures and Integrals - II</b>	<b>733</b>
Extension of coherent lower previsions to unbounded random variables	735
<i>M.C.M. Troffaes, G. De Cooman</i>	
Investigation of the class of non-additive measures by their subclasses	743
<i>E. Pap</i>	
Choquet integral representation and preference	747
<i>Y. Narukawa, T. Murofushi</i>	
The Möbius function on ordered structures and its application to capacities and integrals	755
<i>M. Grabisch</i>	
An axiomatic approach to the definition of the entropy of a discrete Choquet capacity	763
<i>I. Kojadinovic, J.L. Marichal, M. Roubens</i>	

<b>Fuzziness and Conditional Probability</b>	<b>769</b>
Reading the membership function in different ways	771
<i>S. Termini</i>	
A reflection on rationality, guessing and measuring	777
<i>E. Trillas, A. Pradera</i>	
Fuzzy sets and coherent conditional probability	785
<i>R. Scozzafava</i>	
Coherent conditional probability and possibility measures	793
<i>G. Coletti</i>	
Measures of fuzziness: a reinterpretation	801
<i>G. Nicotra, B. Vantaggi</i>	
How measure of uncertainty changes due to unreliability	809
<i>C. Bertoluzza, V. Doldi, G. Naval</i>	
<b>Evolutionary Techniques</b>	<b>815</b>
GPEG: Graph Partitioning using Evolutionary Games	817
<i>A. Cincotti, V. Cutello, F. Pappalardo, M. Pavone</i>	
A Multiobjective Genetic Algorithm for Feature Selection and Data Base Learning in Fuzzy-Rule Based Classification Systems	823
<i>O. Cordon, F. Herrera, M.J. Del Jesus, L. Magdalena, A.M. Sanchez, P. Villar Castro</i>	
What are the main parameters involved in the design of a Genetic Algorithm?	831
<i>I. Rojas, H. Pomares, J. Gonzalez, P.A. Castillo, O. Valenzuela</i>	
Evolutionary Algorithms As Function Optimizers	837
<i>A. Barkat</i>	
Evolutionary Algorithms for Multiperiod Arc Routing Problems	845
<i>P. Lacomme, C. Prins, W. Ramdane-Cherif</i>	
Evolutionary strategies to achieve knowledge in faded temporal fuzzy logic controllers in systems with noisy sensors	853
<i>M.A. Gadeo Martos, J.R. Velasco Perez</i>	
<b>Symbolic Data Analysis</b>	<b>859</b>
From Schweizer to Dempster: mixture decomposition of distributions by copulas in the symbolic data analysis framework	861
<i>E. Diday</i>	
Mixture decomposition of copulas and application to climatology	869
<i>M. Vrac, E. Diday, A. Chedin</i>	
A Top-Down Binary Tree Method for Symbolic Class Descriptions	877
<i>L.M. Mehdi, M. Vrac, S. Winsberg, E. Diday</i>	

Segmentation avec seuil aléatoire <i>J.P. Aboa Yapo, R. Emilion</i>	883
<b>Information Processing in Food Engineering Processes</b>	<b>891</b>
Probabilistic Determination of an Upper Bound for the Final Diacetyl Concentration at Early Stages of Beer Fermentation <i>I.C. Trelea, E. Latrille, S. Landaud, G. Corrieu</i>	893
Handling uncertainty in diagnosis using a combined interval based and a fuzzy logic based approach - Application to wastewater treatment <i>J.P. Steyer, J. Harmand, J.P. Delgenes</i>	899
Processing operator reasoning and its uncertainty to control cheese ripening: a fuzzy symbolic approach <i>I. Ioannou, L. Agioux, N. Perrot, G. Mauris, G. Trystram</i>	905
Fuzzy Methods for Ginseng Drying Control <i>V.J. Davidson, X. Li, R.B. Brown</i>	913
Symbolic Estimation of Food Odors using Fuzzy Techniques <i>A. Loutfi, P. Wide</i>	919
<b>Aggregation</b>	<b>927</b>
The Induced and the Limited Fuzzy Integral Operators <i>R.R. Yager</i>	929
Weighted Fuzzy R-Norm and its Mean Codeword Lengths <i>O. Parkash, P.K. Sharma</i>	937
Three new techniques of approximating aggregation operators from empirical data <i>G. Beliakov</i>	945
Usage of Fuzzy Aggregation for Perceptual Relevance Evaluation <i>A. Soria-Frisch, M. Koppen, T. Sy</i>	953
Propagation of Pertinence Indicator using Distance Models <i>J. Revault</i>	961
<b>Data Mining and Multimedia Systems</b>	<b>967</b>
Image segmentation in compressed domain by clustering methods with euclidean and p-adic metrics <i>J. Benois-Pineau, A. Khrennikov</i>	969
Histogram-Based Color Signatures for Image Indexing <i>S. Boughorbel, N. Boujemaa, C. Vertan</i>	977
A Fuzzy Theoretic Approach for Camera Motion Detection <i>R.S. Jadon, S. Chaudhury, K.K. Biswas</i>	985
Improving Clustering and Visualization of Multimedia Data Using Interactive User Feedback <i>A. Nurnberger, A. Klose</i>	993

Fuzzy Multimedia Mining Applied to Video News <i>M. Detyniecki, C. Marsala</i>	1001
<b>Uncertainty and Fuzziness Management in Evolutionary Processes</b>	<b>1009</b>
Software Uncertainty in General and in KBS Applications in Particular <i>M.M. Lehman, J.F. Ramil</i>	1011
Bacterial algorithm applied for fuzzy rule extraction <i>J. Botzheim, B. Hamori, L. Koczy, A.E. Ruano</i>	1021
Neuro-Fuzzy System for On-line Prediction of Diseases Evolution <i>A. Rotshtein, M. Posner, H. Rakytyanska</i>	1027
Fuzzy theory of evolutionary processes: Two types of dynamics <i>U. Sandler</i>	1035
<b>Plenary Session III</b> <b>Berthold SCHWEIZER, University of Massachusetts (USA)</b>	<b>1041</b>
Probabilistic measurement	1043
<b>Possibility and Imprecise Probability</b>	<b>1045</b>
Reasoning with Partially Ordered Information in a Possibilistic Logic Framework <i>S. Benferhat, S. Lagrue, O. Papini</i>	1047
On Interpretation of T-product Possibility Distributions <i>J. Vejnarova</i>	1053
Imprecise probabilities induced by multi-valued mappings <i>E. Miranda, G. De Cooman, I. Couso</i>	1061
Transforming Imprecise Probabilities into Partial Possibilities <i>P. Baroni, P. Vicig</i>	1069
Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities <i>D. Dubois, L. Foulloy, H. Prade, G. Mauris</i>	1077
<b>Mathematics</b>	<b>1085</b>
Fuzzy Control and t-Norm-Based Fuzzy Logic - Some Recent Results <i>S. Gottwald, V. Novak, I. Perfilieva</i>	1087
A Framework for Unification using Powersets of Terms <i>P. Eklund, M.A. Galan, J. Medina, M. Ojeda-Aciego, A. Valverde</i>	1095
A note on ergodic transformations on MV-algebras <i>B. Riecan</i>	1099
BZW algebras for an abstract approach to roughness and fuzziness <i>G. Cattaneo, D. Ciucci</i>	1103



A characterization of the lattice of fuzzy weak subalgebras of a partial algebra <i>W. Bartol, J. Casasnovas, F. Rossello</i>	1111
Fuzzy extension of interval-based temporal sub-algebras <i>S. Badaloni, M. Giacomini</i>	1119
<b>Computing with words: Models and Applications - II</b>	<b>1127</b>
Towards a Higher Level of Linguistic Granulation <i>R. John</i>	1129
Implementation of Linguistic Labels Based Group Decision-Making on the Web/Intranet <i>A. Marimin, D. Risnawati Kumala</i>	1135
Towards the construction of regression models for categorical variables <i>V. Torra, M. Ng</i>	1143
A linguistic decision-making method using fuzzy integral <i>C.T. Chen</i>	1149
Fusion of Multigranular Linguistic Information based on the 2-tuple Fuzzy Linguistic Representation Model <i>F. Herrera, L. Martinez, E. Herrera-Viedma, F. Chilclana</i>	1155
NPN Logic Based Reasoning With Linguistic Preferences <i>G. Devedzic</i>	1163
<b>Intelligent Information Systems - II</b>	<b>1169</b>
Multi-Objective Resource Selection in Distributed Information Retrieval <i>S. Wu, F. Crestani</i>	1171
Image Database Summarization with the SaintEtiQ System <i>R. Saint-Paul, G. Raschia, N. Mouaddib</i>	1179
Methods for Intelligent Information Access. A real experience <i>J. Cardenosa, L. Iraola, E. Tovar</i>	1187
Reducing term to term relationships in an extended Bayesian Network Retrieval model <i>L.M. De Campos, J.F. Huete Guadix, J.M. Fernandez-Luna</i>	1195
Efficient importance weighted aggregation between min and max <i>H.L. Larsen</i>	1203
Mining data about web usage: a Markov based modeling technique <i>S. Massa, P.P. Puliafito</i>	1209
<b>Plenary Session IV</b> <b>Walter KINTSCH, University of Colorado at Boulder (USA)</b>	<b>1217</b>
Predication	1219

<b>Computing with words: Models and Applications - III</b>	<b>1221</b>
Intelligent linguistic characterization and retrieval of textual documents: an Internet-based application	1223
<i>S. Zadrozny, K.A. Lawcewicz, J. Kacprzyk</i>	
A Linguistic Aggregation Operator Based on the Weak Majority Process	1231
<i>J.I. Pelaez, J.M. Dona</i>	
Using Likelog for Computing with Words	1237
<i>F. Arcelli, F. Formato</i>	
Linguistic Modelling Using a Semi-Naïve Bayes Framework	1243
<i>N.J. Randon, J. Lawry</i>	
Modeling majority opinion in multi-agent decision making	1251
<i>G. Pasi, R.R. Yager</i>	
<b>Probability</b>	<b>1259</b>
Detection of Independence Relations from Persegrams	1261
<i>R. Jirousek</i>	
Estimation of Probabilistic Dependencies of Processes in Nonlinear Systems	1269
<i>K. Chernyshov</i>	
Probabilistic Approach using MCMC for Indirect Measurement by Knowledge-Based Models - Applications to Thermophysics and Biology	1277
<i>H. Baili, G. Fleury</i>	
Estimating tetragram probabilities by using multiple data samples from a natural text. Case study: printed Romanian	1285
<i>A. Vlad, A. Mitrea, M. Mitrea</i>	
Weighted entropy : an optimal measure	1293
<i>R.P. Singh, R. Kumar, R. Gorshi</i>	
<b>Metaphor and Analogy - I</b>	<b>1301</b>
What is Metaphor? From Aristotle to Cognitive Science	1303
<i>E. Hamilton</i>	
Spontaneous Analogies: Analogies or Categorizations?	1309
<i>E. Sander</i>	
Reasoning: The Psychological Relevance of the Hofstadter's Copycat Model	1317
<i>B. Vivicorsi</i>	
Analogical Reasoning: Unity or Diversity?	1323
<i>E. Marmeche, N. Bonnardel, A. Didierjean</i>	
Analogy and Cognitive Ergonomics: Experimentation and Simulation	1329
<i>J. Barcenilla, C. Leproux, S. Poitrenaud</i>	

The role of context on the comprehension of referential metonymies <i>T. Baccino, S. Grossin</i>	1335
<b>Aggregation Operators</b>	<b>1341</b>
On Order Invariant Synthesizing Functions <i>J.L. Marichal</i>	1343
On Symmetric Pseudo-Additions and Pseudo-Multiplications: Is it Possible to Build Rings on $[-1,+1]$ ? <i>M. Grabisch, B. De Baets, J. Fodor</i>	1349
An overview of the definition of uninorm <i>M. Mas, M. Monserrat, J. Torrens</i>	1357
Symmetric aggregation functions on a finite chain <i>J. Martin Pelayo, G. Mayor, J. Suner</i>	1365
<b>Information Retrieval and Querying</b>	<b>1373</b>
Disambiguation translation in multilingual queries <i>S. Benferhat, M. Boughanem, C. Chrismont, N. Nassr, H. Prade</i>	1375
Improving meta-search by using query-weighting and numerical aggregation operators <i>M. Gomez, J.M. Abasolo</i>	1383
Fuzzy Degrees of Knowledge Integrity <i>J. Debenham</i>	1391
Information retrieval in heterogeneous XML knowledge bases <i>G. Menier, P.F. Marteau</i>	1399
Connexion de Galois floue : extension et application au textmining <i>C. Latiri Cherif, S. Elloumi, S. Ben Yahia</i>	1407
<b>Statistical Processing</b>	<b>1415</b>
A fuzzy method for handling measurement data gathered from channel electron multipliers: estimation of statistic noise and pile up phenomena <i>A. Bakhti, F. Khaber</i>	1417
An optimal rule for many outliers detection <i>A. Mahfoudi, J. Aguilar-Martin</i>	1423
Random Sampling and Uncertain Inference <i>H.E. Kyburg, C.M. Teng</i>	1429
A new adaptive kernel density estimation <i>M. Herbin, N. Bonnet, P. Vautrot</i>	1437
Towards an Efficient Density-Based Clustering of Spatial Data <i>A. Badee Salem, A. Ella Hassanien, T. El Areef, M.F. Khater</i>	1445

Boosting up Support Vector Machines in Density Estimation Problems <i>M. Gonzalez Mendoza, A. Titli, N. Hernandez-Gress</i>	1453
<b>Metaphor and Analogy - II</b>	<b>1459</b>
A computational theory of metaphor comprehension <i>W. Kintsch</i>	1461
The Metaphor between Categorization and Similarity <i>B. Pudelko, D. Legros</i>	1465
Metaphor and N400 ERPs <i>P. Cristini, C. Tijus</i>	1473
A Critical Analysis of Current Models of Analogy <i>T. Ripoll, J. Eynard</i>	1477
Le traitement des expressions idiomatiques - Intérêt d'un corpus et de l'Analyse Sémantique Latente <i>F. Pariollaud, G. Denhiere, J.C. Verstiggel</i>	1481
La Compréhension des Métaphores par les Enfants : Questions Développementales, Différentielles, Méthodologiques <i>C. Guerini</i>	1485
<b>Fuzzy Systems</b>	<b>1493</b>
Adjustment of parallel queuing processes by multi-agent control <i>R. Palm, T.A. Runkler</i>	1495
Fuzzy Graphs and Error-proof Keyboards <i>F.L. Luccio, A. Sgarro</i>	1503
Schedulability Analysis of Real-Time Systems under Uncertainty: Fuzzy Approach <i>A.P. Cucala, J. Villar</i>	1509
Perception-based Data Processing by Introducing n -dimensional Metric Space of Linguistic Variables <i>S. Ribaric, J. Drozdek</i>	1517
Effectiveness as Function of Information Uncertainties and Timeliness <i>P. Labbe</i>	1525
Calcul d'itinéraires flous dans les réseaux de transport de surface <i>A. Boulmakoul, R. Laurini, M. Janati</i>	1533
<b>Advances in Database Technology - I (Operators)</b>	<b>1539</b>
About difference operations on fuzzy bags <i>P. Bosc, D. Rocacher</i>	1541
On the comparison of aggregates over fuzzy sets <i>P. Bosc, O. Pivert, L. Lietard</i>	1547

Fuzzy Summarization of Data Using Fuzzy Cardinalities <i>P. Bosc, D. Dubois, O. Pivert, H. Prade, M. De Calmes</i>	1553
On knowledge-guided fuzzy aggregation <i>T. Andreassen</i>	1561
Compressing sets of objects by means of fuzzy object oriented structures <i>F. Berzal, N. Marin-Ruiz, O. Pons, M.A. Vila</i>	1569
<b>Dealing with Uncertainty in Spatial Representation and Reasoning</b>	<b>1577</b>
A model for lines with a broad boundary <i>E. Clementini</i>	1579
On the Representation of Fuzzy Spatial Relations in Robot Maps <i>I. Bloch, A. Saffiotti</i>	1587
Quantification of neurotransmission defects in functional imaging using information fusion: a prospective study <i>E. Frenoux, V. Barra, J.Y. Boire</i>	1595
Fuzzy Fusion of Different Space Representations in Mobile Robotics <i>G. Trainito, P. Bison</i>	1601
World Modelling with Local Uncertainty for Robot Navigation <i>H. Martinez Barbera, M.A. Zamora Izquierdo, A. Gomez Skarmeta</i>	1609
<b>Rough Sets Methods and Applications - II</b>	<b>1617</b>
Rough Sets, Bayes' Theorem and Flow Graphs <i>Z. Pawlak</i>	1619
An Application of Approximated Entropy Measures in Decision Tree Induction <i>S.H. Nguyen, H.S. Nguyen</i>	1625
Rough Set Based Feature Extraction for Medical Data <i>D. Slezak, P. Synak, J. Wroblewski</i>	1633
Generating approximate concepts for the UAV <i>M.S. Szczuka</i>	1639
Order-based genetic algorithms for extraction of approximate bayesian networks from data <i>D. Slezak, J. Wroblewski</i>	1645
<b>Reasoning</b>	<b>1653</b>
A similarity-based approach to deal with inconsistency in systems of fuzzy gradual rules <i>L. Godo, S. Sandri</i>	1655
Coherent Conditional Probability as a Tool for Default Reasoning <i>G. Coletti, R. Scozzafava, B. Vantaggi</i>	1663

Subjective Evidential Reasoning	1671
<i>A. Josang</i>	
Qualitative Reasoning with Statistical Knowledge Expressed in Natural Language	1679
<i>L. Garcia, M.Y. Khayata, D. Pacholczyk</i>	
A Distributed Self-stabilizing Approach to Defeat Status Computation in Argumentation	1687
<i>P. Baroni, M. Giacomin</i>	
<b>Advances in Database Technology - II (Applications and Perspectives)</b>	<b>1695</b>
Uncertainty in an OODB Modeled by Rough Sets	1697
<i>T. Beaubouef, F.E. Petry</i>	
Fuzzy Querying in Crisp and Fuzzy Relational Databases: An Overview	1705
<i>A. Rosado, J. Kacprzyk, R.A. Ribeiro, S. Zadrozny</i>	
Fuzzy and Uncertain Spatio-Temporal Database Models: A Constraint-Based Approach	1713
<i>G. De Tre, R. De Caluwe, A. Hallez, J. Verstraete</i>	
Contourline Based Modeling of Vague Regions	1721
<i>A. Hallez, J. Verstraete, G. De Tre, R. De Caluwe</i>	
<b>Plenary Session V</b>	<b>1727</b>
<b>Paul M. FRANK, University of Duisburg (Germany)</b>	
Handling Modelling Uncertainty in Fault Detection and Isolation Systems	1729
<b>Intelligent Information Processing</b>	<b>1747</b>
Characterization of approximate plane symmetries for 3D fuzzy objects	1749
<i>O. Colliot, I. Bloch, A.V. Tuzikov</i>	
Recognition of dynamic situations in the classroom using multi-modal sensor information	1757
<i>M. Minoh, S. Nishiguchi, H. Higashi</i>	
Intelligent Information Processing for User Modeling	1763
<i>A. Tzanavari, P. Paulson</i>	
Modeling the Perception of the Physical Effort in Manual Lifting Tasks	1771
<i>W. Tembe, A. Ralescu, S.S. Yeung, A. Genaidy</i>	
Linguistic Modeling of Physical Task Characteristics	1779
<i>S. Visa, A. Ralescu, S.S. Yeung, A. Genaidy</i>	
<b>Fuzzy Modeling and Control</b>	<b>1787</b>
A step towards conceptually improving Takagi-Sugeno's Approximation	1789
<i>S. Guadarrama, E. Trillas, J. Gutierrez, F. Fernandez</i>	
Mamdani and Takagi-Sugeno Inference: a Unified View	1795
<i>P.Y. Glorennec</i>	

Multi-Architecture Application Framework for Efficient Fuzzy Processing Using Interpolators <i>J.C. Crespo, J.J. Iglesias</i>	1803
Some conditions for non quadratic stability of a class of non linear systems <i>T.M. Guerra, L. Vermeiren, H. Tirmant</i>	1811
Estimation of the Validity of a Partial Model: Fuzzy and Multi-Models Approaches <i>F. Delmotte, T.M. Guerra</i>	1817
On-line Adaptive Fuzzy Controller in Real Time: Application to Water Supply System of a City <i>H. Pomares, I. Rojas, J. Gonzalez, M. Damas</i>	1825
<b>Intuitionistic Fuzzy Sets - Theory and Applications</b>	<b>1831</b>
On the Temporal Intuitionistic Fuzzy Sets <i>K.T. Atanassov</i>	1833
Intuitionistic Fuzzy Connectives Revisited <i>G. Deschrijver, C. Cornelis, E.E. Kerre</i>	1839
Soft Querying via Intuitionistic Fuzzy Sets <i>P. Grzegorzewski, E. Mrowka</i>	1845
Independence, Conditional Probability and the Bayes Formula for Intuitionistic Fuzzy Events <i>P. Grzegorzewski, E. Mrowka</i>	1851
Analysis of Agreement in Group of Experts via Distances Between Intuitionistic Fuzzy Preferences <i>E. Szmidi, J. Kacprzyk</i>	1859
Intuitionistic Fuzzy S-Implications <i>H. Bustince, E. Barrenechea, V. Mohedano</i>	1867
<b>Textile Applications</b>	<b>1873</b>
Using Internet Distributed Communication in Management of a Textile Research Project <i>P. Bastos, L. Amaral, M. Santos, T. Amorim, R. Vasconcelos</i>	1875
Using Clementine Data Mining System in the Process of Analysis of Cotton Fiber Properties <i>S. Dias, R. Vasconcelos, M. Santos, T. Amorim, L. Amaral</i>	1881
A short term forecasting system adapted to textile distribution <i>S. Thomassey, M. Happiette, J.M. Castelain</i>	1889
Reduction of the estimation uncertainty of the textile sales profiles by using a data-mining technique <i>J.J. Denimal, F. Boussu, O.W. Diallo</i>	1895
Modeling the relationship between subjective and objective fabric hand evaluations using fuzzy techniques <i>M. Sahnoun, X. Zeng, L. Koehl, M.A. Bueno, M. Renner</i>	1901

Measurement and Processing of Sewing Variables towards Real-Time Control and Off-Line Process Planning <i>H.M.T. Carvalho, A.M. Rocha, J. Monteiro</i>	1909
<b>Learning from Partially Labeled Samples</b>	<b>1917</b>
Handling uncertain labels in multiclass problems using belief decision trees <i>P. Vannoorenberghe, T. Denoëux</i>	1919
Text Classification from Positive and Unlabeled Examples <i>F. Denis, R. Gilleron, M. Tommasi</i>	1927
Logistic regression for partial labels <i>Y. Grandvalet</i>	1935
Contextual Learning and Retrieval in a stochastic Network <i>M. Spies</i>	1943
<b>Linguistic Methods</b>	<b>1951</b>
Binary Ordering-Based Modifiers <i>U. Bodenhofer</i>	1953
A Fuzzy Set Theoretical Approach to the Automatic Generation of Absenteism Analyses in Natural Language <i>N. Du Bois, M. De Cock, E.E. Kerre, R. Babuska</i>	1961
Hybridization of Components to Improve the Accuracy in Linguistic Fuzzy Modeling <i>R. Alcalá, J. Casillas, O. Cordon, F. Herrera</i>	1969
A Fuzzy Fusion Sensor of the Speed of a Railway Car During the Sliding of the Wheels, Including a Model of the ABS Braking Equipment <i>V. Balas</i>	1979
<b>Image Processing</b>	<b>1987</b>
Motion adaptive de-interlacing using fuzzy logic <i>D. Van De Ville, R. Van De Walle, W. Philips, I. Lemahieu</i>	1989
Une nouvelle approche connexionniste pour la segmentation d'images couleur <i>P. Biela Enberg, J.G. Postaire</i>	1997
Segmentation des images de cellules biologiques par une approche multifractale <i>N. Lassouaoui, L. Hamami</i>	2005
Approche exploratoire multirésolution basée sur le contenu d'une base d'images paléontologique <i>J. Landre, F. Truchetet</i>	2013



## Fuzzy Prototypical Knowledge Discovery to predict information systems maintainability

**José A. Olivas**

Department of Computer Science  
University of Castilla-La Mancha  
Paseo de la Universidad 4  
13071, Ciudad Real (Spain)  
joseangel.olivas@uclm.es

**Marcela Genero**

Department of Computer Science  
University of Castilla-La Mancha  
Paseo de la Universidad 4  
13071, Ciudad Real (Spain)  
marcela.genero@uclm.es

**Mario Piattini**

Department of Computer Science  
University of Castilla-La Mancha  
Paseo de la Universidad 4  
13071, Ciudad Real (Spain)  
mario.piattini@uclm.es

**Francisco P. Romero**

Soluziona Consultoría  
Soluziona Software Factory,  
c/ Pedro Muñoz 1,  
13071-Ciudad Real, (Spain)  
fromero@uf.isf.es

### Abstract

This contribution presents a new approach, based on Fuzzy Prototypical Knowledge Discovery, to predict the maintainability of class diagrams done using the Unified Modelling Language (UML), which has great importance on the final quality of object-oriented information systems (OOIS). The prediction model is built from metrics values obtained at the early phases of OOIS life-cycle. We will start with Fuzzy Prototypical Knowledge Discovery process (FPKD) for finding Fuzzy Deformable Prototypes of class diagram maintainability, and later we will predict a real case class diagram maintainability deforming the similar prototypes using the degree of compatibility with them.

**Keywords:** Fuzzy Prototypes, Data Mining, OOIS Maintainability, Class Diagrams.

### 1. Introduction

A great effort has been made in the field of software measurement for improving the quality of the OOIS [14], [18], [31], [11], but most of them pursue the goal of evaluating -by means of quantitative measures- the quality of the final product, i.e. the code or the advanced design. But, in Software Engineering it is widely accepted that the quality of OOIS is highly dependent on the decision taken early in the

development. So that, we believe that in order to get better quality OOIS we should focus on measuring internal quality characteristics of early artifacts, such as class diagrams, and based on those measurements thereby obtain early in the life-cycle a prediction model for external quality characteristics, such as for example maintainability, which is one of the major concerns of software developers and industries.

In response to the great demand for measures for measuring quality characteristics of class diagrams, such as maintainability [13], and after a thorough review of some of the existing OO measures that can be applied at a high level design stage [8], [16], [5], [17], we have proposed a set of measures for UML class diagram structural complexity [12]. As maintainability is an external quality characteristic that can be evaluated once a product is finished or nearly finished, we centre our work on measuring an internal quality characteristic, the structural complexity of class diagrams. Our idea is to use those measures to predict class diagram maintainability early in the OOIS development.

This paper is organised in the following way. In section 2 we will present a set of metrics for UML class diagram structural complexity. In section 3 we explain how we carried out a controlled experiment in order to evaluate if there is empirical evidence that UML class diagram structural complexity metrics are correlated with maintainability sub-

characteristics: such as understandability, analysability and modifiability (ISO, 1999). In section 4 we will use the empirical data for building fuzzy prototypes, that characterise UML class diagram maintainability, and based on those prototypes, in section 5, we will build a prediction model for UML class diagram maintainability. The last section summarises the paper, presents some conclusions and identifies further work.

## 2. Metrics for UML class diagram structural complexity

We only present here those metrics presented in [12], which can be applied at class diagram level as a whole (see table 1). These metrics measure the structural complexity of UML class diagrams due to the use of relationships, such as associations, generalisations, aggregations and dependencies. We also consider traditional metrics such as, the number of classes, the number of attributes, etc.

Table 1. Metrics for UML class diagram structural complexity

Metric name	Metric definition
NUMBER OF CLASSES (NC)	The total number of classes.
NUMBER OF ATTRIBUTES (NA)	The total number of attributes.
NUMBER OF METHODS (NM)	The total number of methods
NUMBER OF ASSOCIATIONS (NAssoc)	The total number of associations
NUMBER OF AGGREGATION (NAgg)	The total number of aggregation relationships within a class diagram (each whole-part pair in an aggregation relationship)
NUMBER OF DEPENDENCIES (NDep)	The total number of dependency relationships
NUMBER OF GENERALISATIONS (NGen)	The total number of generalisation relationships within a class diagram (each parent-child pair in a generalisation relationship)
NUMBER OF AGGREGATIONS HIERARCHIES (NAGGH)	The total number of aggregation hierarchies in a class diagram.
NUMBER OF GENERALISATIONS HIERARCHIES (NgenH)	The total number of generalisation hierarchies in a class diagram
MAXIMUM DIT	It is the maximum between the DIT value obtained for each class of the class diagram. The DIT value for a class within a generalisation hierarchy is the longest path from the class to the root of the hierarchy.
MAXIMUM HAGG	It is the maximum between the HAgg value obtained for each class of the class diagram. The HAgg value for a class within an

aggregation hierarchy is the longest path from the class to the leaves.

## 3. Empirical validation of the proposed metrics

In this section we describe an experiment we have carried out for empirically validating the proposed metrics (see section 2). We will only be able to draw conclusions about the relationship between the cause and the effect for which we stated a hypothesis (which we want to corroborate by means of experiments); if the experiment is properly set up.

To perform an experiment, several steps have to be taken and they have to be in a certain order.

### 3.1 Definition

As Wholin et al. [29], suggested, we follow the GQM template [1], [2], [3], [28], for goal definition.

### 3.2 Planning

After the definition of the experiment, the planning took place. The definition determines the foundation of the experiment -why the experiment is conducted- while the planning prepares for how the experiment is conducted.

#### 3.2.1 Context selection

The context of the experiment is a group related to the area of Software Engineering at the university, and hence the experiment is run-off line (not industrial software development), it is conducted by 7 teachers and 10 students enrolled in the final-year of Computer Science in the Department of Computer Science at the University of Castilla-La Mancha in Spain. All of the teachers belong to the Software Engineering area.

The experiment is specific since it is focused on UML class diagram structural complexity metrics. The ability to generalise from this specific context is further elaborated below when discussing threats to the experiment. The experiment addresses a real problem the correlation between metrics and maintainability sub-characteristics.

### 3.2.2 Hypothesis formulation

An important aspect of experiments is to know and to state in a clear and formal fashion what we intend to evaluate in the experiment. This leads us to the formulation of a hypothesis (or several hypotheses). We wish to test the hypothesis that there is a significant correlation between the current metric data set (NC, NA, NM, NAssoc, NAgg, NDep, NGen, NAggH, NGenH, MaxHAgg, MaxDIT) and the subject's rating of three maintainability sub-characteristics, such as understandability, analysability and modifiability.

### 3.2.3 Instrumentation

The objects were class diagrams done using UML.

The independent variable (UML class diagram structural complexity) was measured through the metrics, presented in section 2.

The dependent variables (understandability, analysability and modifiability) were measured according to subject's rating.

### 3.2.4 Validity evaluation

We will discuss the empirical study's various threats to validity and the way we attempted to alleviate them:

*Threats to Construct Validity.* We propose subjective metrics for measuring each of the dependent variables (maintainability sub-characteristics) based on the judgement of the subjects. As the subjects involved in this experiment have medium experience in UML class diagram design we think their ratings can be considered significant. The independent variables (each of the metrics proposed in section 2) that measure the structural complexity of class diagrams can also be considered constructively valid, because from a system theory point of view, a system is called complex if it is composed of many (different types of elements), with many (different types of) (dynamically changing) relationships between them [25].

*Threats to Internal Validity.* Seeing the results of the experiment we can conclude that empirical evidence of the existing relationship between the independent and the dependent variables exists. We have tackled different

aspects that could threaten the internal validity of the study, such as: differences among subjects, knowledge of the universe of discourse among class diagrams, accuracy of subject responses, learning effects, fatigue effects, persistence effects and subject motivation.

*Threats to External Validity.* Two threats to external validity have been identified which limit the ability to apply any such generalisation, and we have tried to alleviate them: materials and tasks, and subject selection. In general in order to extract a final conclusion that can be generalised, we need to replicate this experiment with a greater number of subjects, including practitioners. After doing replication we will have a cumulative body of knowledge; which will lead us to confirm if the presented metrics could really be used as early quality indicators, and could be used to predict class diagram maintainability.

## 3.3 Operation

### 3.3.1 Preparation

By the time the experiment was done all of the students had had two courses on Software Engineering, in which they learnt in depth how to build OO software using UML. All the selected professors had enough experience in the design and development of OOIS. Moreover, subjects were given an intensive training session before the experiment took place. The subjects were not aware of what aspects we intended to study. Neither they were aware of the actual hypothesis stated.

We prepared the material we had to give to the subjects, consisting of 28 class diagrams of the same universe of discourse, related to Bank Information Systems. Each diagram has a test enclosed which includes the description of maintainability sub-characteristics, such as: understandability, analysability, modifiability. Each subject has to rate each sub-characteristic using a scale consisting of seven linguistic labels. For example for understandability we proposed the following linguistic labels:

Extremely difficult to underst	Very difficult to underst	A bit difficult to underst	Neither difficult nor easy to underst	Quite easy to underst	Very easy to underst	Extremely easy to underst
--------------------------------------	------------------------------------	-------------------------------------	--	-----------------------------	----------------------------	---------------------------------

We also prepared a debriefing questionnaire. This questionnaire included (i) personal details and experience, (ii) opinions on the influence of different components of UML class diagrams,

such as: classes, attributes, associations, generalisations, etc... on their maintainability.

### 3.3.5 Execution

The subjects were given all the material described in the previous section. We explained to them how to carry out the experiment. We allowed one week to do the experiment, i.e., each subject had carry out the test alone, and could use unlimited time to solve it.

We collected all the data, including subjects' rating obtained from the responses of the experiment and the metrics values automatically calculated by means of a metric tool we had designed.

### 3.3.6 Data Validation

All tests were considered valid because all of the subjects have at least medium experience in building UML class diagrams and developing OOIS.

## 3.4 Analysis and Interpretation

Analyzing the Spearman's correlation coefficients, we can conclude that there exists a high correlation between most of the UML class diagram structural complexity metrics and the subject's rating of understandability, analysability and modifiability. We can deduce this due to the fact that almost all the metrics have a correlation greater than 0.7. NDep is the only one that has a lesser correlation. This fact should be studied in detail by carrying out further experimentation.

## 3.5 Presentation and package

The last activity is concerned with presenting and packaging of the findings. The diffusion of the experimental results and the way they are presented are relevant so that they are really put into use. Therefore we published our findings in this paper, and we are also planning to publish a lab package on the web for replication purposes.

## 4. A prediction model for UML class diagram maintainability

In this section we explain the steps involved in the Fuzzy Prototypical Knowledge Discovery (FPKD) process Olivas et al. [20], [21], [22],

which lead us to the construction of fuzzy prototypes (Zadeh, 1982) [30], that characterize the maintainability of UML class diagrams. The FPKD is a fuzzy extension of the traditional Knowledge Discovery in Databases (KDD) [9].

Zadeh [30] mentioned the classical prototype theories from the point of view of psychology, criticizing that these theories do not fit the function that a prototype should have. Zadeh's approach to what must be taken as a prototype is less intuitive than the conceptions of psychological theories but is more rational and closer to the meaning of a prototypical concept displayed in a more detailed examination. In our case, we have observed that Zadeh's idea suggests a concept that encompasses a set of prototypes, which represent the high, medium, or low compatibility of the samples with the concept A. "The prototype is not a single object or even a group of objects in A. Rather, it is a fuzzy schema for generating a set of objects which is roughly coextensive with A" [30].

The prototypes obtained from the FPKD, will form the foundation of the prediction model that allows us to predict class diagram maintainability. This approach is more representative than standard approaches, because the use of an isolated algorithm or method over-simplifies the complexity of the problem. Statistical methods or decision trees (ID3, C4.5, CART) are only classification processes, and it is very important to include a clustering model for finding some kinds of patterns in the initial chaos of data. The use of fuzzy schemas for representing these patterns allows us to achieve better and more understandable results.

### 4.1 The FPKD process

The FPKD process consist of different steps:

*Selection of the target data.* We have taken as a start set a relational database that contains 476 records (with 14 fields, 11 represent metric values, 3 represent maintainability sub-characteristics, understandability, analysability and modifiability respectively) obtained from the calculation of the metric values (for each class diagram) and the responses of the experiment given by the subjects.

*Preprocessing.* The Data-Cleaning was not necessary because we did not find any errors.

**Transformation.** This step was performed doing different tasks:

**Summarising subject responses.** We built a table with 28 records (one record for each class diagram) and 14 fields. The metric values were calculated measuring each diagram, and the values for each maintainability sub-characteristics were obtained aggregating subjects' ratings using their mean.

**Clustering by Repertory Grids.** In order to detect the relationships between the class diagrams, to obtaining those which are easy, medium or difficult to maintain (based on subject rates of each maintainability sub-characteristics), we have carried out a hierarchical clustering process by Repertory Grids, based on subject's rating for each diagram. The set of elements is composed of the 28 class diagrams, the constructions are the intervals of values of the subject's rating. The application of Repertory Grids Analysis Algorithm (hierarchical clustering algorithm) returns a graphic which reflects each prototype (easy, medium and difficult to maintain), and the class diagrams which pertain to them (see figure 1).

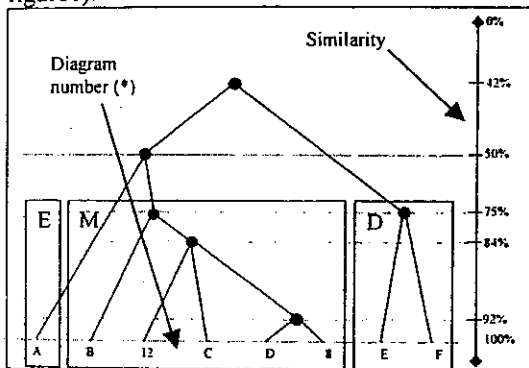


Figure 1. Clustering results (E: Easy to maintain, M: Medium to maintain, D: Difficult to maintain).

(\*) We have grouped some class diagrams assigning them one letter because they have 100% similarity (see Clustering Results in Fig.2)

**Data Mining.** The selected algorithm for the data mining process was summarise functions. Table 3 shows the parametric definition of the prototypes.

Table 3. Prototypes "Easy, Medium and Difficult to maintain"

	Understandability	Analisisability	Modifiability
<b>Difficult</b>			
Average	6	6	6
Maximum	6	6	7
Minimum	6	5	6
<b>Medium</b>			
Average	5	5	5
Maximum	5	6	5
Minimum	4	4	4
<b>Easy</b>			
Average	2	2	3
Maximum	3	3	3
Minimum	2	2	2

**Formal Representation** of conceptual prototypes. The prototypes have been represented as fuzzy numbers, which are going to allow us to obtain a degree of membership in the concept. In order to construct the prototypes (triangular fuzzy numbers) we only need to know their centerpoints ("center of the prototype"), which are obtained by normalising and aggregating the metric values corresponding to the class diagrams of each of the prototypes (see figure 2).

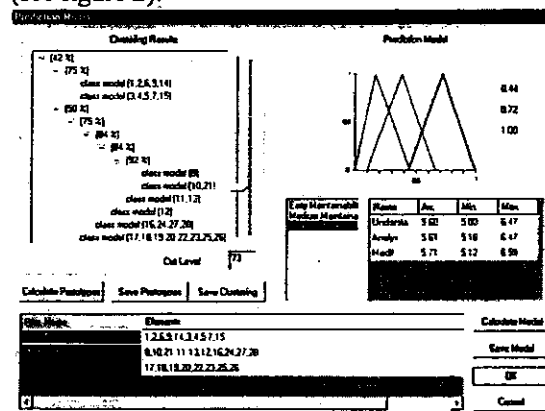


Figure 2. Representation of the prototypes.

## 5. Example of prediction of UML class diagram maintainability

Using Fuzzy Deformable Prototypes (Olivas et al. [20], [21], [22]), we can deform the prototypes with some degree of affinity to a new class diagram (change the values for describing exactly the new real situation), using a linear combination with the degrees of membership as coefficients of the prototypes vectors. We will give an example of how to deform the fuzzy prototypes found in section 4.1.

Given the following metric values corresponding to a new class diagram:

NC	NA	NM	NAssoc	NAgg	NAggH	NDep	NGenR	NGenH	MaxDIT	MaxHagg
21	30	70	10	6	2	3	20	5	2	3

And their normalised values:

NC	NA	NM	NAssoc	NAgg	NAggH	NDep	NGenR	NGenH	MaxDIT	MaxHagg
0.69	0.48	0.67	0.71	0.67	0.67	0.75	0.83	1	0.40	0.75

The final average is 0.65. The affinity with the prototypes is shown in figure 3.

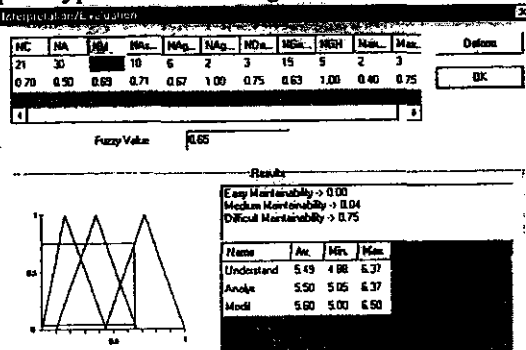


Figure 3. Affinity of the real case with the prototypes.

The prototypes with some degree of affinity for this new class diagram are "Difficult to maintain" and "Medium to maintain". Then, the prediction based on the deformation of these prototypes is (see also fig. 3):

	Understandability	Analysability	Modifiability
Average	5	5	6
Maximum	5	5	6
Minimum	6	5	6

## 6. Conclusions and future work

In this paper we have used Genero et al.'s metrics [12], with the objective of corroborating that there exists a great correlation between these metrics values and the maintainability of a class diagram, we have carried out a controlled experiment.

Also we have used a fuzzy extension of the traditional KDD process, the FPKD (Olivas et al. [20], [21], [22]), for building the maintainability prototypes which serve as the basis of the prediction model for the sub-characteristics that affect class diagram maintainability.

The prediction model was built using the FPKD process. This process was used not only in the software measurement area, but was also used

for different kinds of real problems, such as forest fire prediction [20], [22], financial analysis or medical diagnosis, obtaining satisfactory results.

We want to highlight that this is a first approach to predicting UML class diagram maintainability, we need "real data" about UML class diagram maintainability efforts, such as time spent in maintenance tasks in order to predict data that can be highly useful to software designers and developers.

Nevertheless, despite the promising nature of the obtained results, towards of seeking correct OO metrics applied at a high level design stage, we are aware that we need to do more metric validation, both empirical and theoretical in order to obtain conclusive evidence of the usefulness of the proposed metrics.

Pending is the theoretical validation of the proposed metrics using the DISTANCE framework proposed by Poels and Dedene [24], [25], which is in our knowledge the most appropriate for OO measurements.

In future work, we will also tackle the measurement of other quality factors like those proposed in the ISO 9126 [13], which not only addresses class diagrams, but also evaluates other UML diagrams, such as use-case diagrams, state diagrams, etc. To our knowledge, little work has been done towards measuring dynamic and functional models [25]. As is quoted in Brito e Abreu et al. [6], this is an area which lacks in depth investigation.

## Acknowledgements

This research is part of the DOLMEN project supported by CICYT (TIC 2000-1673-C06-06) and the CIPRESES project supported by CICYT (TIC 2000-1362-C02-02).

## References

- [1] Basili V. and Weiss D. (1984). A Methodology for Collecting Valid Software Engineering Data, *IEEE Transactions on Software Engineering*, 10, 728-738.
- [2] Basili V. and Rombach H. (1988). The TAME Project: Towards Improvement-Oriented Software Environments. *IEEE Transactions on Software Engineering* 14, 758-773.

- [24] Poels G. and Dedene G. (1999). DISTANCE: A Framework for Software Measure Construction, research report DTEW9937, Dept. Applied Economics, Katholieke Universiteit Leuven, Belgium, 46 p.
- [25] Poels G. and Dedene G. Measures for Assessing Dynamic Complexity Aspects of Object-Oriented Conceptual Schemes. In: *Proceedings of the 19<sup>th</sup> International Conference on Conceptual Modeling (ER 2000)*, Salt Lake City, (2000a), 499-512.
- [26] Poels G. and Dedene G. (2000b). Distance-based software measurement: necessary and sufficient properties for software measures. *Information and Software Technology*, 42(1), 35-46.
- [27] Schneidewind N. (1992). Methodology For Validating Software Metrics. *IEEE Transactions of Software Engineering*, 18(5), 410-422.
- [28] Van Solingen R. and Berghout E. (1999). The Goal/Question/Metric Method: *A practical guide for quality improvement of software development*. McGraw-Hill.
- [29] Wohlin C., Runeson P., Höst M., Ohlson M., Regnell B. and Wesslén A. (2000) *Experimentation in Software Engineering: An Introduction*. Kluwer Academic Publishers.
- [30] Zadeh L. (1982). A note on prototype set theory and fuzzy sets. *Cognition* 12, 291 - 297.
- [31] Zuse H. (1998). *A Framework of Software Measurement*. Berlin, Walter de Gruyter.