Combining Soft Computing and Statistical Methods to Improve Data Analysis Solutions

IC0702

Start date: 31/03/2008

End date: 30/03/2012

Year: 1

Christian Borgelt

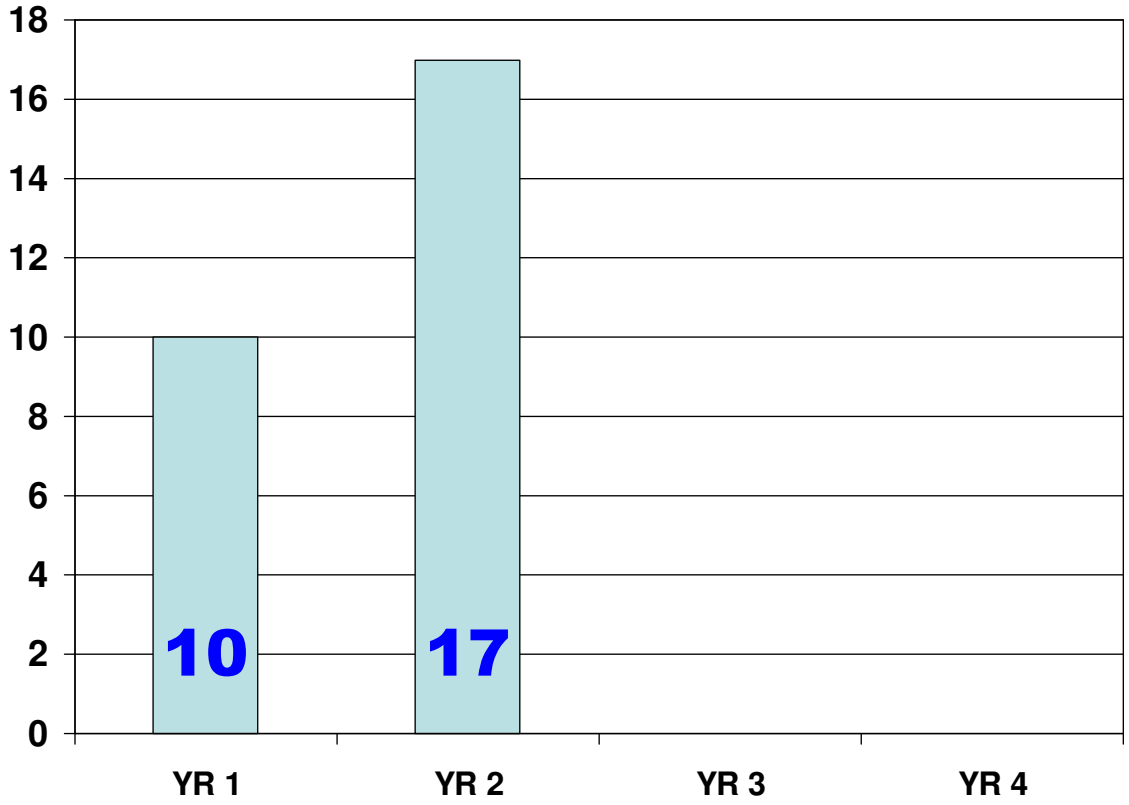
Chair

European Centre for Soft Computing, Spain

COST is supported by
the EU RTD Framework Programme

ESF provides the
COST Office through an EC contract

Action Parties



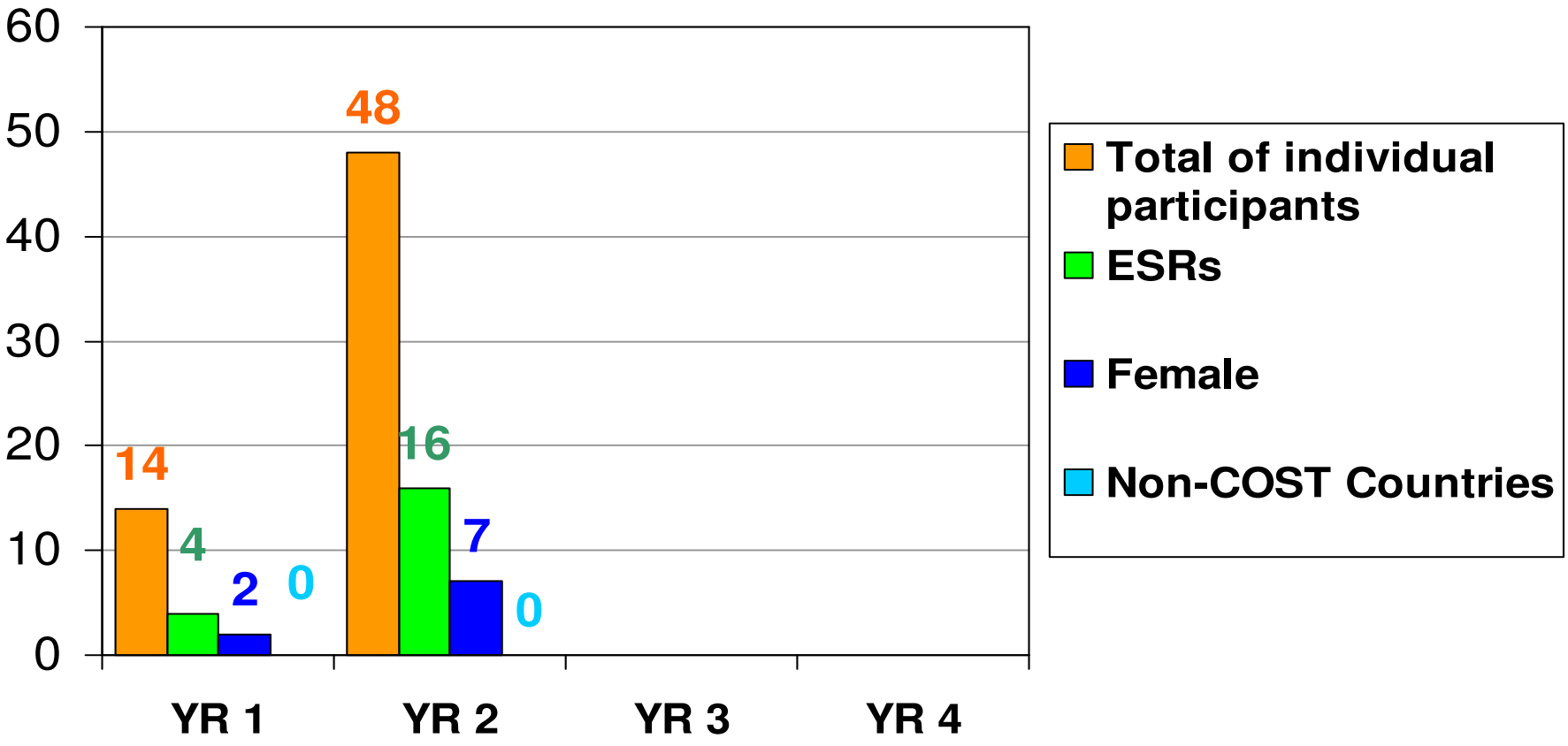
Grant Holder:

European Centre for Soft Computing

Raul Del Coso

Spain

Action participants



Use of COST instruments

	YR 1	YR 2	YR 3	YR 4
No. of MC / WG meetings	2 (3)	(2)		
No. of STSMs	9	(15)		
No. of workshops / conferences	1	1		
No. of joint publications	0 (8)	(12-15)		
No. of training schools	1	1		
GASG (activities)	posters, website	posters, website		

Scientific context and objectives (1/2)

- Soft Computing, as an engineering science, and Statistics, as a classical branch of mathematics, emphasize different aspects of data analysis
- **Soft Computing**
  focuses on quickly obtaining working solutions that meet the needs arising in applications.
- **Statistics**
  focuses on establishing objective conclusions by rigorously analyzing all possible situations.
- Improving the currently rather limited dialogue and interaction between these areas can lead to new and improved data analysis methods.

Scientific context and objectives (2/2)

- **Soft Computing Methods and Models**
  - often lack sound mathematical foundations,
  - rarely make their underlying assumptions explicit, thus impeding a reliable transfer to new applications.
  - are seldom checked rigorously and monitored w.r.t. performance and robustness,
  - can rarely be generalized or easily transferred
- **Statistical Methods and Models**
  - tend to focus on models, the mathematical properties of which are easy to analyze,
  - constrain the set of eligible models, thus perhaps ruling out the most suitable or promising ones,
  - can be difficult to understand or apply for a non-mathematician.

Working groups

- Working Group A
  Model Selection and Validation
  - Statistical validation and monitoring of Soft Computing models
  - Model selection and validation for neural networks

- Working Group B
  Bio-inspired Metaheuristics
  - Metaheuristics (evolutionary algorithms in particular) as Estimators
  - Estimation of distribution algorithms

- Working Group C
  Statistics with Imperfect and Incomplete Data
  - Statistics with fuzzy data (estimation and regression)
  - Psychological versus statistical complexity

Results vs. Objectives

- Gather researchers from soft computing and statistics in working groups, stimulate exchange and discussion.
  (successfully achieved: presentations triggered animated discussions)
- Execute short term scientific missions, preferably between a statistics host and a soft computing visitor or vice versa.
  (more or less achieved: bridging the gap is difficult)
- Organize training schools to disseminate statistical and soft computing knowledge and successful joint applications.
  (successfully achieved: spring school and summer course)
- Apply for new research projects in which statisticians and soft computing researchers work togetherand develop new data analysis methods.
  (not achieved yet: progress to this goal is slower than hoped for)

Significant highlights (1/2)

- Spring School "**Reasoning and Decision Making under Uncertainty and Imprecision**"
  May 18-20, 2009
  European Centre for Soft Computing, Mieres, Spain
  http://www.cost-ic0702.org/springschool/
- Topics: Soft Computing, Statistics, Logic, Decision Theory
- 15 high-level lecturers from 8 countries,
  > 35 participants (early stage researchers) from 10 countries.
- Organized jointly with COST Actions IC0602 and IC0801.
- Initiated a closer cooperation between COST Actions IC0702, IC0602 and IC0801, which will be continued.

Significant highlights (2/2)

- Only one short-term scientific mission less than planned.
- Topics cover mainly three research areas:
  - information retrieval
    novelty detection, summarization, topic detection/tracking
  - statistics with fuzzy data
    hypothesis tests with imperfect data, linear regression with imprecise response, robust fuzzy regression
  - knowledge discovery in complex and structured domains
    stream mining, change mining, spatial data analysis, feature selection
- For already completed STSMs draft research papers are available and will be submitted in the course of this year.
- Noteworthy STSM: “A Determination Coefficient for a Linear Regression Model with Imprecise Response“

Challenges

- We planned to achieve or at least get closer to the goal of compiling and submitting a research proposal in the area of soft computing and statistics.
- Progress towards this goal has been slower than hoped for, because the exchange about the research areas of the working group members took longer than expected.
- A subgroup has formed in working group C, which works towards a project application in fuzzy statistics. The challenge is to actually reach a project proposal.
- Second core challenge: maintain and improve the success of the short-term scientific missions.