Combining Soft Computing and Statistical Methods to Improve Data Analysis Solutions

IC0702

Start date: 01/04/2008

End date: 31/03/2012

Year: 4

Christian Borgelt

Chair

European Centre for Soft Computing, Spain

COST is supported by the EU Framework Programme

ESF provides the COST Office through a European Commission contract

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.
- **Objective of the Action**
  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

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

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

3. 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: organized several exchange events, including a conference with a focus on the topics of the action: SMPS 2010)
- Execute short term scientific missions, preferably between a statistics host and a soft computing visitor or vice versa.
  (successfully achieved: greatly exceeded the projected number of STSMs)
- Organize training schools to disseminate statistical and soft computing knowledge and successful joint applications.
  (cancelled for 2010/2011 due to initial budget problems)
- Apply for new research projects in which statisticians and soft computing researchers work together and develop new data analysis methods.
  (still struggling with this objective: progress to this goal is slow and no real leader emerges – however, we are continuing our efforts)

Significant Highlights in Science or Networking (1/2)

- Conference **Soft Methods in Probability and Statistics (SMPS 2010),** Sep. 28 – Oct. 1, 2010
  - Scope: approaches that allow for more flexible modeling of imprecision, uncertainty, vagueness and ignorance in probability and statistics
  - Venue: European Centre for Soft Computing, Mieres, and University of Oviedo, Oviedo, Spain
  - 95 papers in proceedings (Springer), 24 session, 4 plenary talks, 120 participants from 17 countries
  - Great feedback from participants
  - http://smps2010.cost-ic0702.org/

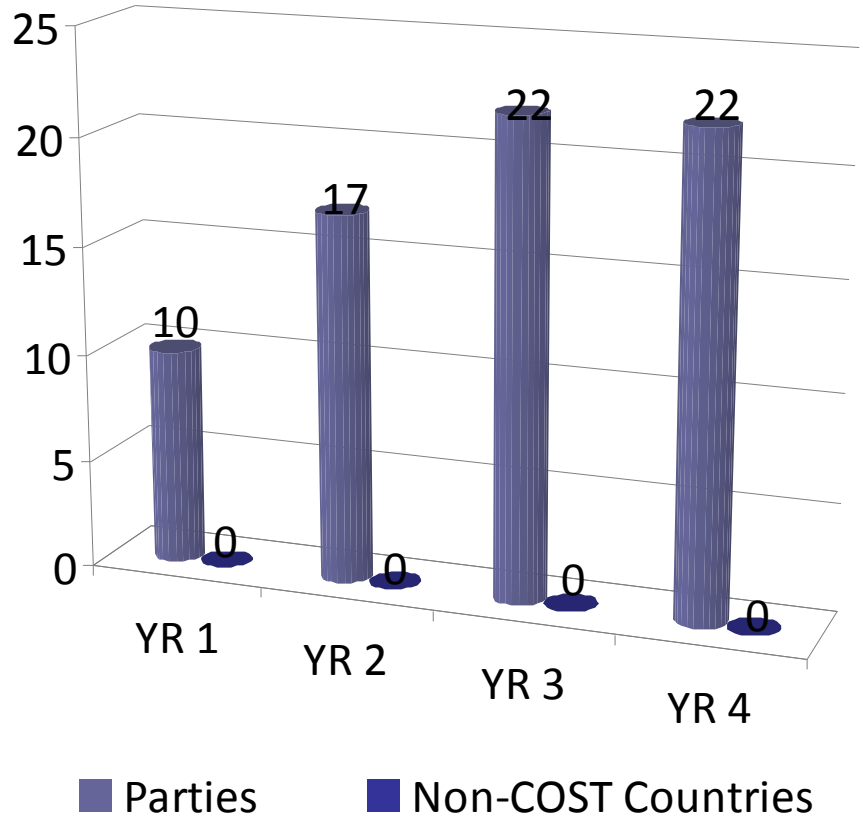
Significant Highlights in Science or Networking (2/2)

- Due to the initial budget issues, we installed several measures to minimize costs (only one day MC/work group meeting, limit on travel expenses).
- When the budget was increased later, these cuts could not be undone, leaving us with excess money for STSMs.
- By specially motivating people, the 15 STSMs projected (and reported on in the monitoring progress report) could be increased to 27.
- 12 STSM visitors are women, 24 are early stage researchers.
- A large number lasted 6 weeks or longer and yielded substantial results in the form of research paper submissions.

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 very slow:
  - Exchange about the research areas of the working group members took longer than expected,
  - Current topics in FP7 ICT calls do not fit the scope of the action well (digital libraries is closest, but difficult to match properly),
  - Finding a topic in an FP7 application domain (e.g. health care, environmental issues etc.) is challenging, because for any such topic only few action members have work experiences,
  - No project leader emerged.
- Second core challenge: maintain and improve the success of the short-term scientific missions.

Action Parties



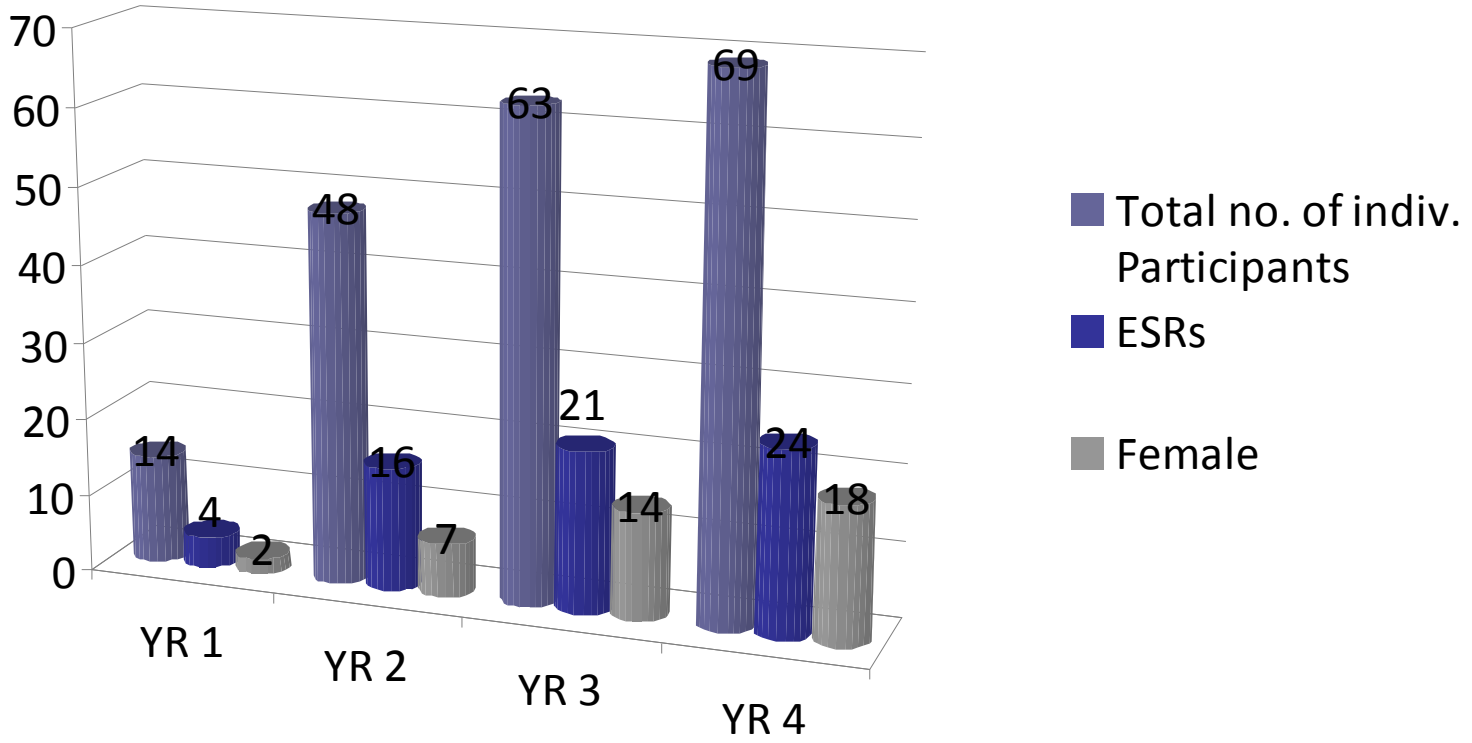
Grant Holder:

European Centre for Soft Computing

Raul del Coso

Spain

Action participants



Use of COST Instruments

Activity (No.)	Year 1	Year 2	Year 3	Year 4
MC/WG Meetings	2(3)	2	2	(2)
STSMs	9	18	27	(12)
Training Schools	1	1	0	1
Workshops or Conferences	1	1	1	1
Joint Publications	12	35	50	40