

## The COST IC0702 Workshop at ERCIM 2008

The COST Action IC0702 organized a workshop consisting of three sessions — one for each work group — as part of the 1st Workshop of the ERCIM<sup>1</sup> Work Group on Computing and Statistics (ERCIM 2008), which was co-located with the 2nd Int. Workshop on Computational and Financial Econometrics (CFE 2008) in Neuchâtel, Switzerland. While the other sessions of ERCIM 2008 and CFE 2008 spanned three days, namely June 19 to 21, 2008, the three COST IC0702 sessions took place on Thursday, June 19, and Friday, June 20, 2008. In these sessions, members of the COST action work groups gave a total of 14 presentations about their recent research.

The purpose of this COST IC0702 workshop, which was the first of this action, was twofold: in the first place, the sessions provided an opportunity for an initial exchange about the research areas the work group members are currently working in. Secondly, the co-location with two workshops that are strongly oriented towards statistics made it possible to attract attention from statisticians to soft computing techniques and combinations of statistical and soft computing approaches.

Both purposes were successfully achieved. The exchange between the work group members brought them considerably closer together and sparked several ideas about potential future short-term scientific missions within the framework of COST Action IC0702. In addition, all of the three sessions, but particularly the first on “Statistics with Incomplete Data”, attracted participants of the other two workshops (ERCIM 2008 and CFE 2008), thus initiating an exchange beyond the boundaries of the COST Action.

In more detail, the three workshop sessions were the following:

### **Statistics with Incomplete Data**

Thursday, June 19, 2008, 10:12:15 hours  
(associated with Work Group C)

This session focussed on combinations of fuzzy approaches (as one of the core constituents of soft computing) with statistical methods. The presentations dealt with the goal of helping a user with interpreting results that were obtained with a statistics-based machine learning algorithm (first talk), the treatment of truncated and/or censored data in a financial prediction problem (second talk), the imprecise perception of values in the form of fuzzy sets and their statistical treatment in the framework of upper and lower probabilities (third talk), hypothesis testing for fuzzy random variables, in particular testing hypothesis about the mean value (fourth talk), and how fuzzy techniques may be used to improve the analysis of discrete-range real-valued variables (fifth talk).

Since the subjects in this session were, among those of the three COST sessions, the closest to statistics, it is not surprising that attendance from participants outside the COST Actions was largest in this session.

### **Probabilistic Methods in Learning Problems**

Thursday, June 19, 2008, 14:00:16:00 hours  
(associated with Work Group B)

---

<sup>1</sup>European Research Consortium for Informatics and Mathematics

In this session learning methods were studied, which mainly originated from machine learning and soft computing, and they were analysed and enhanced by probabilistic methods. The talks presented methods to generate probabilistic fuzzy systems, especially for modeling financial domains (first talk), approaches to select the relevant features for a prediction problem with mutual information (second talk), fuzzy approaches to text mining, with which it is tried to enhance statistical methods in order to achieve more interpretable results (third talk), a simulation study of several statistical approaches for estimating the prediction error of neural networks (fourth talk), and finally an application of neural networks and the closely related support vector machines to a financial prediction problem (fifth talk).

### **Intelligent Data Analysis**

Friday, June 20, 2008, 09:00-11:00 hours  
(associated with Work Group A)

The topics in this session were all arranged around soft classification and soft clustering, using both statistical and soft computing methods. The presentations addressed the question how to learn from data with soft class labels (that is, the outcome is not crisply known; first talk), how one may view the noise clustering approach of probabilistic fuzzy clustering as a kind of M-estimator (second talk), how feature selection for classification and prediction can be effectively achieved by a combination of ant colony algorithms and ANOVA methods (third talk), and how neural network techniques can be used to accelerate both fuzzy and probabilistic mixture model clustering (fourth talk).

All presentations were of high quality and were well received by the audience. Some sparked lively discussions, which were taken into the coffee breaks and led to plans of mutual visits to intensify the exchange (potential STSMs).

In addition to these sessions, two discussion meetings were held, in which administrative issues were discussed (Thursday, 19.06.2008, 16:30–17:20, and Friday, 20.06.2008, 08:00–08:55). Apart from technical issues, like how to apply for short-term scientific missions and the decision process, one of the main topics were possible future meeting dates and places. Since with this workshop we tried to bring soft computing approaches into a mainly statistics-oriented conference, the proposals focussed on the reverse: bringing statistical methods into soft computing conferences. Suggestions included a workshop of Work Group A at ESANN 2009 and of Work Group B or C at IFSA 2009, which will likely be pursued. In addition, a workshop independent of another conference could be held in November or December in Rotterdam, Netherlands. Furthermore, repeating this workshop at ERCIM 2009 in Cyprus was proposed due to the success of this workshop. Finally, combining at least one work group meeting with the next MC meeting in October 2008 in Warsaw, Poland, is a very convenient option. More concrete plans will be worked out by Email communication over the next weeks.