

# The 1st Workshop on ROC Analysis in Artificial Intelligence (ROCAI-2004)

José Hernández-Orallo  
Univ. Politèc. de València  
Cami de Vera, S/N 46022  
València, Spain  
jorallo@dsic.upv.es

Cèsar Ferri  
Univ. Politèc. de València  
Cami de Vera, S/N 46022  
València, Spain  
cferri@dsic.upv.es

Nicolas Lachiche  
LSIIT 64, bld Brant, F-67400  
Illkirch-Graffenstaden  
France  
lachiche@lsiit.u-strasbg.fr

Peter Flach  
University of Bristol  
Woodland Road  
Bristol BS8 1UB, U.K  
Peter.Flach@bristol.ac.uk

## ABSTRACT

This short report includes a summary of the presentations and discussions held during the ROCAI-2004 workshop, as well as the workshop conclusions and the future agenda. ROCAI-2004 was held in Valencia, on August the 22nd, as part of the 16th European Conference on Artificial Intelligence, ECAI-2004, in Valencia, Spain.

## Keywords

Artificial Intelligence, ROC Analysis, Machine Learning

## 1. INTRODUCTION AND MOTIVATION

Receiver Operating Characteristic (ROC) analysis has been regularly used in medicine (radiology, diagnosis, ...) and psychology for many decades, as a powerful tool for cost/benefit analysis in decision making. It has been introduced relatively recently in several areas of artificial intelligence: machine learning, data mining, intelligent decision support and expert systems. In this context, ROC analysis provides techniques to select possibly optimal models and to discard suboptimal ones independently from (and prior to specifying) the cost context or the class distribution. Furthermore, the Area Under the ROC Curve (AUC) has been shown to be a better evaluation measure than accuracy in contexts with variable misclassification costs and/or imbalanced datasets. AUC is also the standard measure when using classifiers to rank examples, and, hence, is used in applications where ranking is crucial, such as campaign design, model combination, collaboration strategies, and co-learning.

Nevertheless, there are several open questions and limitations that hamper a broader use and applicability of ROC analysis. Its connections with other evaluation measures is not yet completely clarified, its incorporation in decision support and expert systems technology just envisaged, its use for improving the decisions of (communities of) intelligent agents unexplored, and its use in data mining hasn't yet reached its full potential. Among the limitations of ROC analysis, an important one, despite some recent progress, is its possible but difficult and computationally expensive extension to more than two classes.

The first motivation for a first workshop on ROC analysis was to have a first meeting to exchange ideas and advances in the fundamentals and applications of ROC analysis from the point of view of computer science. The second motivation was to broaden its scope and applicability to other areas of computer science, as well as getting attention from those fields in AI that could benefit as well from ROC analysis. And here we mean a broad view of ROC analysis in the sense of exploiting the ROC analysis

philosophy: a model can be analysed, previously to knowing its context of application, and we can even select and discard suboptimal ones.

The topics suggested in the call for papers were:

- Adaptation of classical learning methods for improving AUC instead of accuracy.
- Agent selection/ranking in multiagent systems.
- Alternative ROC representations.
- Alternatives to AUC measures (such as AUC\*).
- Analysis of the performance of learners or decision systems based on AUC.
- Applications of ROC Analysis in Expert Systems.
- Classifier Evaluation.
- Co-learning (collaborative learning) and ROC Analysis.
- Consensus and Collaborative/Distributed Decision Making.
- Constraint Satisfaction Methods for ROC Analysis.
- Cost-sensitive Learning.
- Decision Networks and Probabilistic Reasoning with ROC Analysis.
- Expert Systems, Knowledge Bases and ROC Analysis.
- Inductive Logic Programming and ROC Analysis.
- Multi-class ROC on top of binary ROC.
- Oversampling and ROC analysis.
- Precision & Recall measures in Information Retrieval.
- Ranking Actions. Applications in Planning and Robotics with variable contexts.
- Reinforcement Learning and ROC analysis.
- ROC Analysis for Model Building and Modification.
- ROC Analysis for Descriptive Data Mining (association rules, subgroup discovery).
- Soft classifiers and Probability Estimators.
- Software Packages and Efficient Implementations (Convex Hull).
- Use of ROC analysis and measures as fitness and selection criteria in evolutionary techniques.
- Working with Imbalanced datasets.

The program committee was selected to cope with this broad and varied topic list, and was composed of 11 people, representing 10 countries: Austria, Belgium, Canada, France, Germany, Italy,

Spain, The Netherlands, UK and USA. We had submissions from 7 different countries: Canada, France, Germany, Slovenia, Spain, UK and USA. Each paper was reviewed by at least two program committee members. Reviewers' scores aggregated and the 11 papers with highest scores were selected for presentation at the workshop. These papers were published in a workshop notes volume, edited by the workshop organisers and the ECAI organisation.

## 2. ORGANISATION AND SESSIONS

The technical program of ROCAI-2004 consisted of one invited talk, eleven paper presentations, one invited workshop presentation and a round table, organised in several sessions during this one-day workshop. The topics of these sessions covered a wide spectrum on ROC Analysis theory and applications, including new ideas and results, recent developments, and new research directions

The workshop started with a short welcome and had an average audience of 20 attendees. Then, the technical program started with an invited talk by Peter Flach on "Recent advances in Machine Learning applications of ROC analysis", shortly revisiting some ROC analysis basics and then moving to recent advances in the interpretation of several evaluation measures (ISO-accuracy lines, decision tree splitting criteria...) and new advances in calibration and the notion of deriving better models from existing models using ROC analysis.

The paper presentation started with a paper from Chris Drummond and Robert C. Holte on "What ROC Curves Can't Do (and Cost Curves Can)", which discussed an alternative way of analysing and comparing classifiers, based on Cost Curves, where each line represents a classifier (instead of points in classical ROC analysis). Robert showed several examples where Cost curves could be advantageous over classical ROC Curves representation. The paper raised significant interest and comments from the attendees.

### Session 1

Robert Holte presented the first paper of the session: "What ROC Curves Can't Do (and Cost Curves Can)" by Chris Drummond and Robert C. Holte. In this work the authors show some of the limitations of classical ROC curves, and how these problems are solved with their alternative cost curves.

The next presentation was "Confidence Bands for ROC Curves: Methods and an Empirical Study" by Sofus A. Macskassy and Foster Provost. Sofus A. Macskassy explained some techniques for generating and evaluating confidence bands on ROC curves. The authors showed the convenience of the techniques with some experiments.

Alexandru Niculescu-Mizil presented "An Empirical Analysis of the Relationship Between AUC and 8 Standard Supervised Learning Performance Criteria" by Rich Caruana and Alexandru Niculescu-Mizil. In this work the authors studied by a thorough experimental evaluation the behaviour of 9 different performance metrics, including AUC.

### Session 2

The first speaker of this session was Jerome Aze, with the paper "Interestingness measures based on supervised learning. Application to terminology extraction" by Mathieu Roche, Jerome Aze, Yves Kodratoff and Michele Sebag. This work shows how

the evolutionary algorithm ROGER can be used to term extraction, formalised here as a supervised learning task. Some experiments with real-world datasets exhibit that the approach significantly improves on the standard statistical and IR-related criteria on the frequent terms.

The following paper was "ROC Optimisation of Legacy Critical Systems" by Jonathan E. Fieldsend and Richard M. Everson. It was presented by Jonathan E. Fieldsend. This work shows how ROC analysis can be employed for the tuning of critical systems as a multiobjective optimisation problem. Specifically, they apply this methodology to the (Short Term Conflict Alert) STCA system. This system is employed to give warning of potential breaches in air proximity by aircraft.

Alain Rakotomamonjy presented his paper "Optimizing Area Under Roc Curve with SVMs". This work proposes a SVMs based algorithm for maximising the Area Under Roc curve measure. The experiments of the paper prove that the proposed method can actually maximize AUC and that the performance is comparable to ensemble methods such as RankBoost.

### Session 3

The Session started with the paper "Cautious Classifiers" by Cèsar Ferri, José Hernández-Orallo, presented by José Hernández-Orallo. This paper discusses how to evaluate a cautious classifier, a classifier able to abstain where it has not enough confidence to make a decision. Some issues as performance measures or ROC analysis with cautious classifiers are addressed.

The second paper of the session was "Precision and Recall Optimisation for Information Access Tasks" by Michelle J. Fisher and Jonathan E. Fieldsend and Richard M. Everson presented by Michelle J. Fisher. The authors investigate the use of multi-objective evolutionary algorithms for finding model parameters that optimise precision and recall measures. The proposed method is compared to other techniques on some information access experiments.

Tobias Sing presented "Learning Mixtures of Localized Rules by Maximizing the Area Under the ROC Curve" by Tobias Sing and Niko Beerenwinkel and Thomas Lengauer. The authors introduce a model class for statistical learning which is based on mixtures of propositional rules. Model inference consists of repeated iteration through a sequence of three steps: First, a new rule is mined from a resampled data set using the apriori algorithm. Next, the localisation information for the rule is computed. Finally, the weights of all rules in the mixture model are re-optimised simultaneously. This weight optimization is done using the area under the ROC curve rather than the error rate as the objective function.

### Session 4

Nicola Lama and Laura Antolini gave a presentation on EWADP 2004 (1st European Workshop on Assessment of Diagnostic Performance) and the Biopattern network, and discussed where ROC analysis was more useful for these application areas. The presentation raised significant interest. They announced a workshop for the next year about prognostic instead of diagnostic performance (EWAPP).

The session followed with the work "ROC Analysis of Example Weighting in Subgroup Discovery" by Branko Kavsek, Nada Lavrac, Ljupco Todorovski presented by Branko Kavsek. This

paper presents two new ways of example weighting for subgroup discovery. The authors implemented them in APRIORI-SD and studied their behaviour both theoretically by means of ROC analysis and practically by application to a real-life data set.

The last paper was "An Empirical Evaluation of Supervised Learning for ROC Area" by Rich Caruana and Alexandru Niculescu-Mizil. The paper was presented by Alexandru Niculescu-Mizil. The work contains an empirical comparison of the AUC performance of seven supervised learning methods. The paper also presents an ensemble selection method that yields even better AUC than the rest of methods. Ensembles are built with forward stepwise selection, the model that maximizes ensemble AUC performance being added at each step.

### 3. CONCLUSION AND DISCUSSION

The workshop concluded with the general impression that it succeeded as a first specific workshop on ROC analysis in a computer science scenario. First, it was a success mainly because of the number and quality of submissions. Secondly, we got attention to a workshop on ROC analysis, which is, per se, very particular, but in this case even more specific, since it was focussed to computer science. Thirdly, participants were really active during the workshop, with many questions and discussions.

The open discussion centred around these three issues: 1) Most relevant topics, open questions, 2) Broadening the scope, and 3) Continuation.

About the most relevant topics treated during the workshop or that needed to be tackled next, some of them were suggested: first, the idea of a "ROC analysis" software repository was suggested, warmly welcome by the rest of attendees. Some people volunteered to start working on that. Another very important issue raised by some participants was the multiclass extension of ROC analysis. We had different views on this, maybe because many application areas only deal with 2-class problems or just because multiclass ROC analysis is very difficult or a combination of both. Finally, another important point suggested was the statistical validation and confidence bands, a topic that was covered by some ROCAI-2004 papers.

One of the motivations of the workshop was broadening the scope and application areas of ROC analysis. It was suggested that for most of the AI topics where ROC analysis could be applied (agent selection/ranking, expert systems and knowledge bases, consensus and collaboration in multi-agent systems, constraint satisfaction, planning and robotics, control, reinforcement learning, etc.) we didn't quite succeed to get submissions about this. Everybody agreed that this doesn't mean we shouldn't go on trying to

popularise ROC analysis in other areas of computer science. On the contrary, this could be viewed as a further reason to continue with this scope broadening. Some attendees suggested to connect to the use of ROC analysis in "Pattern recognition", which, jointly to the use in machine learning and data mining, could help disseminate this into other areas of artificial intelligence.

Finally, there was a consensus that we should try to continue with the workshop. Most people also thought that the idea of associating to a big conference was good to broaden the workshop and ease the organisation. The organisers suggested three possibilities for 2005: ICML, IJCAI and ECML/PKDD and other participants did also mention other possibilities.

The Workshop's information and selected papers can be found on the Web at <http://www.dsic.upv.es/~flip/ROCAI2004/>

## 4. OTHER INFORMATION

### 4.1 Program Committee

- Hendrik Blockeel, K.U.Leuven, Belgium.
- Stephan Dreiseitl FHS Hagenberg, Austria.
- Tom Fawcett, HP Labs, Palo Alto, CA, USA.
- Cèsar Ferri, Technical University of Valencia, Spain.
- Peter Flach, University of Bristol, UK.
- Johannes Fürnkranz, University of Vienna, Austria.
- José Hernández-Orallo, Technical University of Valencia, Spain.
- Nicolas Lachiche, University of Strasbourg, France.
- Charles Ling, University of Western Ontario, London, Ontario, Canada
- Maarten van Someren, University of Amsterdam, The Netherlands.
- Francesco Tortorella, University of Cassino, Italy

### 4.2 Organising Committee

- Cèsar Ferri, Technical University of Valencia, Spain.
- Peter Flach, University of Bristol, UK.
- José Hernández-Orallo, Technical University of Valencia, Spain.
- Nicolas Lachiche, University of Strasbourg, France.