# Feature subset selection by genetic algorithms and estimation of distribution algorithms -: A case study in the survival of cirrhotic patients treated with TIPS

Inza, I Merino, M Larrañaga, P Quiroga, J Sierra, B Girala, M

The transjugular intrahepatic portosystemic shunt (TIPS) is an interventional treatment for cirrhotic patients with portal hypertension. In the light of our medical staff's experience, the consequences of TIPS are not homogeneous for all the patients and a subgroup dies in the first 6 months after TIPS placement. Actually, there is no risk indicator to identify this subgroup of patients before treatment. An investigation for predicting the survival of cirrhotic patients treated with TIPS is carried out using a clinical database with 107 cases and 77 attributes. Four supervised machine learning classifiers are applied to discriminate between both subgroups of patients. The application of several feature subset selection (FSS) techniques has significantly improved the predictive accuracy of these classifiers and considerably reduced the amount of attributes in the classification models. Among FSS techniques, FSS-TREE, a new randomized algorithm inspired on the new EDA (estimation of distribution algorithm) paradigm has obtained the best average accuracy results for each classifier. (C) 2001 Elsevier Science B.V. All rights reserved.

[Paper](https://www.webofscience.com/api/gateway?GWVersion=2&SrcApp=edas&SrcAuth=WosAPI&KeyUT=WOS:000171913100004&DestLinkType=FullRecord&DestApp=WOS_CPL)

# Feature subset selection by Bayesian networks:: a comparison with genetic and sequential algorithms

Inza, I Larrañaga, P Sierra, B

In this paper we perform a comparison among FSS-EBNA, a randomized, population-based and evolutionary algorithm, and two genetic and other two sequential search approaches in the well-known feature subset selection (FSS) problem. In FSS-EBNA, the FSS problem, stated as a search problem, uses the estimation of Bayesian network algorithm (EBNA) search engine, an algorithm within the estimation of distribution algorithm (EDA) approach. The EDA paradigm is born from the roots of the genetic algorithm (GA) community in order to explicitly discover the relationships among the features of the problem and not disrupt them by genetic recombination operators. The EDA paradigm avoids the use of recombination operators and it guarantees the evolution of the population of solutions and the discovery of these relationships by the factorization of the probability distribution of best individuals in each generation of the search. In EBNA, this factorization is carried out by a Bayesian network induced by a cheap local search mechanism. FSS-EBNA can be seen as a hybrid Soft Computing system, a synergistic combination of probabilistic and evolutionary computing to solve the FSS task. Promising results on a set of real Data Mining domains are achieved by FSS-EBNA in the comparison respect to well-known genetic and sequential search algorithms. (C) 2001 Elsevier Science Inc. All rights reserved.

[Paper](https://www.webofscience.com/api/gateway?GWVersion=2&SrcApp=edas&SrcAuth=WosAPI&KeyUT=WOS:000169988900004&DestLinkType=FullRecord&DestApp=WOS_CPL)

# Feature selection for high-dimensional remote sensing data by Maximum Entropy Principle based optimization

Yu, SX Scheunders, P

For high-dimensional remote sensing data, the appropriate selection of features has a significant effect on the cost and accuracy of an automated classifier. In this paper, a method for feature selection by Estimation of Maximum Entropy Principle Algorithm, is presented. This method based on the EDA (Estimation of Distribution Algorithm) paradigm, avoids the use of crossover and mutation operators to evolve the populations, in contrast to Genetic Algorithms. It is combined with an approximate application of the Maximum Entropy Principle as the models for representing the probability distribution of a set of candidate solution in the feature selection problem, using the application of automatic learning methods to induce the right distribution model in each generation. Computational comparison is made between EDA in combination with Bayesian networks and EDA in combination with Maximum Entropy Principle. Experiments are performed on AVIRIS dataset.

[Paper](https://www.webofscience.com/api/gateway?GWVersion=2&SrcApp=edas&SrcAuth=WosAPI&KeyUT=WOS:000177297201067&DestLinkType=FullRecord&DestApp=WOS_CPL)

# Prototype selection and feature subset selection by estimation of distribution algorithms.: A case study in the survival of cirrhotic patients treated with TIPS

Sierra, B Lazkano, E Inza, I Merino, M Larrañaga, P Quiroga, J

The Transjugular Intrahepatic Portosystemic Shunt (TIPS) is an interventional treatment for cirrhotic patients with portal hypertension. In the light of our medical staffs experience, the consequences of TIPS are not homogeneous for all the patients and a subgroup dies in the first six months after TIPS placement. An investigation for predicting the conduct of cirrhotic patients treated with TIPS is carried out using a clinical database with 107 cases and 77 attributes. We have applied a new Estimation of Distribution Algorithms based approach in order to perform a Prototype and Feature Subset Selection to improve the classification accuracy obtained using all the variables and all the cases. Used paradigms are K-Nearest Neighbours, Artificial Neural Networks and Classification Trees.

[Paper](https://www.webofscience.com/api/gateway?GWVersion=2&SrcApp=edas&SrcAuth=WosAPI&KeyUT=WOS:000175009500003&DestLinkType=FullRecord&DestApp=WOS_CPL)

# Continuous estimation of distribution algorithms with probabilistic principal component analysis

Cho, DY Zhang, BT

Recently, many evolutionary algorithms have been studied to build and use an probability distribution model of the population for optimization problems. Most of these methods tried to represent explicitly the relationship between variables in the problem with factorization techniques or the graphical model such as Bayesian or Gaussian network. Thus enormous computational cost is required for constructing those models when the problem size is large. In this paper, we propose new estimation of distribution algorithm by using probabilistic principal component analysis (PPCA) which can explains the high order interactions with the latent variables. Since there are no explicit search procedures for the probability density structure, it is possible to rapidly estimate the distribution and readily sample the new individuals from it. Our experimental results support that presented estimation of distribution algorithms with PPCA can find good solutions more efficiently than other EDAs for the continuous spaces.

[Paper](https://www.webofscience.com/api/gateway?GWVersion=2&SrcApp=edas&SrcAuth=WosAPI&KeyUT=WOS:000172263900074&DestLinkType=FullRecord&DestApp=WOS_CPL)