



HAL
open science

Clustering analysis of railway driving missions with niching genetic algorithms

Amine Jaafar, Bruno Sareni, Xavier Roboam

► **To cite this version:**

Amine Jaafar, Bruno Sareni, Xavier Roboam. Clustering analysis of railway driving missions with niching genetic algorithms. 11th International Workshop on Optimization and Inverse Problems in Electromagnetism (OIPE 2010), Sep 2010, Sofia, Bulgaria. hal-03865575

HAL Id: hal-03865575

<https://hal.science/hal-03865575v1>

Submitted on 20 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CLUSTERING ANALYSIS OF RAILWAY DRIVING MISSIONS WITH NICHING GENETIC ALGORITHMS

Amine JAAFAR, Bruno SARENI, Xavier ROBOAM

Université de Toulouse, LAPLACE, UMR CNRS-INPT-UPS, site ENSEEIHT
2 rue Camichel, BP 7122, 31 071 Toulouse Cedex, France: E-mail: {name}@laplace.univ-tlse.fr

Abstract. A large number of applications requires classifying or grouping data into a set of categories or clusters. Most popular clustering techniques to achieve this objective are K-means clustering and hierarchical clustering. However, both of these methods necessitate the *a priori* setting of the cluster number. In this paper, a clustering method based on the use of a niching algorithm is presented, with the aims of finding the best compromise between the inter-cluster distance maximization and the intra-cluster distance maximization. This method is applied to a standard test problem and to the classification of railway driving missions.

Keywords: Clustering, K-means, niching genetic algorithms, railway locomotive, driving mission,

INTRODUCTION

The design of hybrid electrical locomotive requires taking into account at a system level couplings between architecture, sizing, energy management strategy and environment [1]. For such devices, the system environment is associated with specific power driving profiles that have to be provided by all energy sources. These profiles contain a large number of features related to the system sizing, efficiency and lifetime which can be represented by specific indicators (e.g. maximal and average mission powers, statistical indices associated with the mission power distribution). Integrating these indicators in the design process is not straightforward because of the important number of driving missions than can be imposed to the locomotive. For this purpose, cluster analysis methods can be useful in order to find the most “representative” driving missions. They can help the designer to identify different system configurations in compliance with the corresponding clusters.

CLUSTERING WITH NICHING GENETIC ALGORITHMS

Basic niching genetic algorithm for data clustering

Instead of K-means clustering methods [2] which require the *a priori* setting of the number of cluster for minimizing a similarity metrics (typically the sum of the distance of the points from their cluster centroids), the proposed algorithm aims at simultaneously finding the appropriate number of clusters and the correct classification of the data set. For that purpose, partition criteria (e.g. Calinski and Harabasz criterion, Silhouette index or Dunn’s function) are exploited [3]. These indices emphasize the compactness of intra-cluster and the isolation of inter-cluster. Thus, optimizing these indices consists in determining a tradeoff between the inter-cluster distance maximization and the intra-cluster distance minimization and consequently the best classification of a data set. Because of the multimodality feature of data clustering, the use of niching Genetic Algorithms (GAs) [4] is recommended to avoid premature convergence. In this work, we employ the restricted tournament selection [5]. The number of clusters and the corresponding cluster positions are coded in the chromosome of the individuals. However, the chromosome size is fixed containing the same number of cluster positions (i.e. the maximum number of clusters). The other genes (i.e. the following cluster positions) can be considered as “recessive”. They are not used in the individual decoding but participate to crossover and mutation operations (note that all genes including the number of clusters undergo crossover and mutation). The use of this coding strategy in association with niching increases population diversity, particularly the exploration of clustering configurations with variable number of clusters, which improves the algorithm efficiency.

Application to a test problem

A typical clustering benchmark is constructed in order to verify the effectiveness of the proposed algorithm. The problem is composed of a two-dimensional data set with 200 points and 12 non-overlapping clusters (see Fig. 1a). The density of the different clusters is non-uniform varying from 4 in the smallest cluster to 34 in the biggest. The niching GA with a population size of 100 is applied for maximizing the silhouette index relative to this data set. Results obtained from a typical run are shown in Fig. 1, especially the silhouette index (Fig. 1b) and the number of clusters (Fig. 1c) during generations. It can be seen from this figure that the niching GA quickly converges to the correct classification.

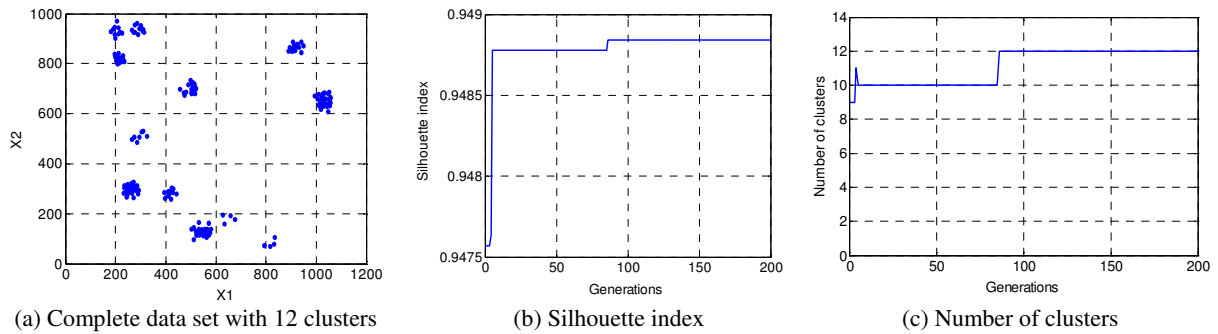


Figure 1. First benchmark with 13 clusters

CLUSTERING ANALYSIS OF RAILWAY DRIVING PROFILES

Design of hybrid supplies for electrical locomotives

The design of hybrid supplies for electrical locomotive requires the knowledge of the driving missions for sizing the system elements [1]. For this kind of hybrid devices, a primary energy source provides the average part of the mission. The rest of the mission (i.e. the fluctuant part) is devoted to a storage system (i.e. the auxiliary energy source). With this energy management strategy, the size of both energy sources depends on the average mission power P_{av} , on the maximal mission power P_{max} and on the global energy E_u devoted to the storage system. Consequently, the critical values of these indices have to be identified to design supply systems that fulfills all missions. However, this approach may lead to the oversizing of the sources, especially when a “critical” mission significantly differs from the others. Therefore, the use of clustering analysis can be interesting in order to find the most representative mission (i.e. the cluster centroid) from a set of missions. This can help the designer in the investigation of tradeoffs between the mission fulfillment and the energy supply sizing.

Clustering analysis of a set of railway driving profiles

A second benchmark consisting of a set of 105 driving missions is used for assessing the efficiency of the clustering method proposed in this paper. This set is composed of three subsets of missions devoted to three different railway systems: the BB 63000 locomotive, the BB 460000 locomotive and the auxiliary supply for TGV. All missions are represented in the feature space in Fig. 2a. The centroids of each subset are also indicated with a full back mark. We plot in Fig. 2b, the classification of the missions obtained from a niching GA run after 500 generations. These results will be discussed in details in the final paper. They confirm the good efficiency of the niching GA which globally finds the correct classification of these missions.

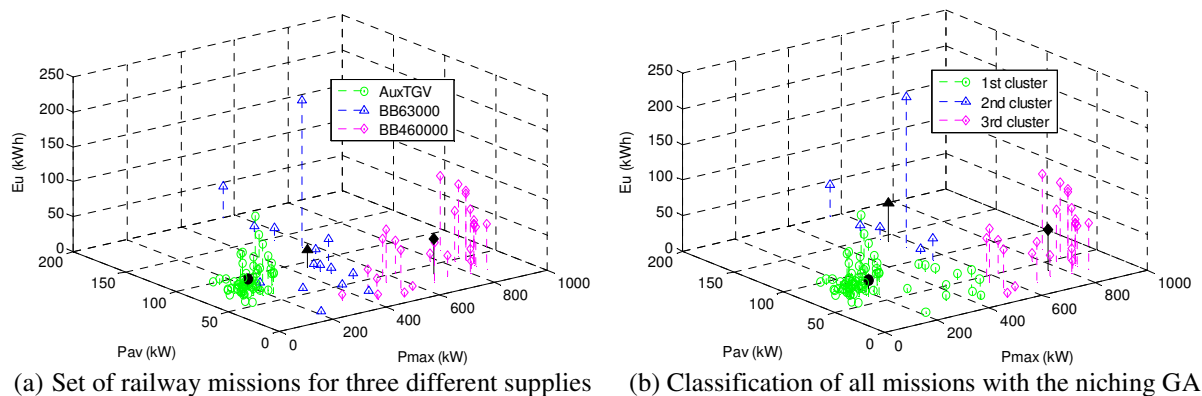


Figure 2. Second benchmark with three subsets of railway driving missions

REFERENCES

- [1] C. Akli, B. Sareni, X. Roboam, A. Jeunesse, “Integrated optimal design of a hybrid locomotive with multiobjective genetic algorithms”, selected paper from *OIPE’2008, International Journal of Applied Electromagnetics and Mechanics*, Vol. 4, n°3-4, 2009.
- [2] R. Xu, D.C. Wunsch II, ‘Survey of Clustering Algorithms’, *IEEE Trans. On Neural Networks*, Vol. 16, n°3, pp. 645-678, 2005.
- [3] W. Sheng., S. Swift, L. Zhang, X. Liu, “A weighted sum validity function for clustering with a hybrid niching genetic algorithm,” *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 35, no. 6, pp. 1156-1167, 2005
- [4] B. Sareni, L. Krähenbühl, “Fitness sharing and niching methods revisited”, *IEEE Trans. on Evolutionary Computation*, Vol. 2, n°3, pp.97-106, 1998.
- [5] G. Harik, “Finding Multimodal Solutions Using Restricted Tournament Selection”, *Proceedings of the Sixth International Conference on Genetic Algorithms*, San Francisco, CA: Morgan Kaufmann, pp. 24-31, 1995.