

COURBOTREE : UNE METHODE DE CLASSIFICATION DE COURBES APPLIQUEE AU LOAD PROFILING

Véronique Stéphan

EDF R&D

1, av. du Général de Gaulle

92 141 Clamart Cédex

Veronique.Stephan@edf.fr

Résumé

Depuis le 1^{er} Juillet 2004, l'ensemble des clients industriels et professionnels peuvent choisir leur fournisseur d'électricité. Pour la majeure partie de ces clients, EDF ne dispose pas des courbes de consommation électrique mais seulement d'index qui permettent de calculer le volume total consommé entre 2 index consécutifs. Un profilage (load profiling), c'est-à-dire une estimation de la courbe de consommation électrique des clients, point par point sur cette période, est alors nécessaire. Celui-ci peut être réalisé en prenant en compte les connaissances métiers et par l'analyse des données récoltées à partir d'un échantillon de clients télérelevés. Ainsi, face à ces nouveaux enjeux commerciaux, des besoins spécifiques en techniques classificatoires appliquées aux courbes apparaissent. Nous proposons dans cet article une méthode de classification de courbes, appelée Courbotree, s'intégrant dans une démarche plus globale de load profiling. Cette méthode repose sur les techniques d'arbres de régression multivariée. Elle répond à un double objectif de classification et de prédiction de courbes. Dans le contexte multivarié, la construction de l'arbre est similaire à celle des méthodes AID et CART. La seule différence réside dans le choix du critère de coupure qui est celui de l'inertie calculée sur les composantes des courbes. A partir d'un échantillon de clients, sur lesquels on dispose d'une part de leurs caractéristiques et d'autre part d'une courbe de consommation, Courbotree fournit une classification de ces courbes directement interprétable par l'utilisateur en terme de règles d'affectation métiers. Elle permet ainsi de profiler tout nouveau client selon ses valeurs observées sur les variables explicatives.

Mots-Clés : classification de courbes, load profiling, arbre de régression, méthode AID, classification divisive.

Abstract

Since July 2004, industrial and professional sector can choose their electric power supplier. For the most part of these customers, load curves aren't enabled and EDF just has a total load consumption over a large period. A load profiling approach, which consists in estimating customer load curve, is therefore necessary. This technique may be performed taking into account business knowledge and analyzing data from telemetered customers. Faced to new marketing challenges, dedicated needs in clustering techniques appear. This article details a curves clustering technique, called Courbotree, integrated in a global load profiling approach. We present how applying multivariate regression trees meets these aims. These ones answer both problems of clustering and prediction applied to curves. Considering multivariate regression trees, tree building is performed in a similar way than AID and CART methods do. The main difference lies in the cutting criterion which is based on inertia computed on the curve components. Applied to a sample of customers, which are described by a load curve and a set of characteristics, Courbotree offers both a curve sample partition and a set of identification rules to affect a curve profile to a new individual according to his explanatory values.

Key-Words : curves partitioning, regression trees, AID method, divisive partitionning.