Université de Tunis
Institut Supérieur de Gestion
Ecole Doctorale Sciences de Gestion

DEALING WITH UNCERTAINTY IN CLUSTERING USING POSSIBILITY THEORY: POSSIBILISTIC K-MODES

Mémoire en vue de l'obtention du Mastère en

Informatique Appliquée à la Gestion

Présenté par

Asma AMMAR

Sous la direction de

M. Zied ELOUEDI (LARODEC, ISG Tunis)

Mastère de Recherche / ISG / Juin 2011
Dealing with Uncertainty in Clustering using Possibility Theory: Possibilistic K-Modes
AMMAR Asma

Résumé:

Le travail présenté dans ce rapport étudie le problème du regroupement des données où règne l'incertitude. Lors de la modélisation des connaissances dans les domaines de la vie réelle, plusieurs pièces d'information incertaines peuvent exister. Faire face à cette incertitude, en particulier, à travers l'utilisation des méthodes de clustering peut être d'un grand intérêt étant donné qu'il permet une meilleure prise de décision. Dans ce rapport, nous combinons la méthode k-modes avec la théorie des possibilités afin d'obtenir une nouvelle approche de clustering pour les données catégoriques incertaines. Plus précisément, nous développons la méthode dite le k-modes possibiliste (PKM) pour traiter les valeurs des attributs incertaines des objets où l'incertitude est présentée à travers des distributions de possibilité. Nous avons aussi proposé une amélioration du PKM, notée PKM-SIM, par l'introduction au niveau de ses paramètres d'une méthode pour sélectionner les k modes initiaux. Afin d'évaluer le PKM et le PKM-SIM, nous avons effectué des simulations sur des bases de données réelles. Les résultats expérimentaux montrent une bonne performance.

Mots-clés:

Clustering, théorie de possibilités, incertitude, données catégoriques, méthode k-modes.

Abstract:

The work presented in this report investigates the problem of clustering data pervaded by uncertainty. When modeling knowledge in real world fields, several uncertain pieces of information are found. Dealing with this uncertainty, in particular, using clustering methods can be of great interest since it helps to make a better decision. In this report, we combine the k-modes method within the possibility theory in order to obtain a new clustering approach for uncertain categorical data; more precisely we develop the so-called possibilistic k-modes method (PKM) allowing to deal with uncertain attribute values of objects where uncertainty is presented through possibility distributions. We have also proposed an improvement of the PKM, denoted by PKM-SIM, by introducing in its parameters a new method to select the k initial modes. In order to evaluate the PKM and the PKM-SIM, we have performed simulations using real databases. Experimental results show good performance.

Keywords:

Clustering, possibility theory, uncertainty, categorical data, k-modes method.