

Vers un cadre évolutif de classification non supervisée

Mohamed Charouel*, Minyar Sassi-Hidri** Mohamed Ali Zoghlami***

Université Tunis El Manar
Ecole Nationale d'Ingénieurs de Tunis
BP. 37 le Belvédère 1002 Tunis, Tunisie
{*mohamed.charouel, **minyar.sassi}@enit.rnu.tn, ***ma.zoghlami@gmail.com

Résumé. La classification non supervisée (clustering) évolutive surpasse généralement par celle statique en produisant des groupes de données (clusters) qui reflètent les tendances à long terme tout en étant robuste aux variations à court terme. Dans ce travail, nous présentons un cadre différent pour le clustering évolutif d'une manière incrémentale par un suivi précis des variables de proximité temporelles entre les objets suivis par un clustering statique ordinaire.

1 Introduction

Dans de nombreuses applications pratiques de gestion de clusters, résultat d'une opération de classification non supervisée (clustering), les objets à classer évoluent dans le temps. Le but est donc d'obtenir des clusters optimaux à chaque pas de temps (Falkowski et al. (2006)).

Tang et al. (2008) et Zhang et al. (2009) ont proposés des méthodes de clustering évolutif dans le but de produire des clusters optimaux qui reflètent les dérives à long terme dans les objets tout en étant robuste aux variations à court terme. Chi et al. (2009) ont développé cette idée en proposant deux cadres évolutifs pour le clustering spectral : *PCQ* (Preserving Cluster Quality) et *PCM* (Preserving Cluster Membership). Les deux cadres ont été proposés afin d'optimiser la modification de la fonction de coût proposée initialement par Chakrabarti et al. (2006).

Notre travail adapte le principe d'incrémentabilité afin de le généraliser à un ensemble d'algorithmes de clustering. Le cadre proposé consiste à estimer les états de données à l'aide des proximités qui sont à la fois actuels et passés. Puis, il effectue un clustering statique sur les estimations de ces états. Ce cadre de suivi de clustering évolutif d'une manière incrémentale a été utilisé pour étendre une variété d'algorithmes de clustering statiques tels que les C-Moyennes Floues (*CMF*) de Bezdek (1984), les *k-moyennes* de Mac-Queen (1967) et les approches spectrales de clustering présentées par Filippone et al. (2008).

Le reste de ce papier est organisé comme suit : en section 2, nous présentons le cadre évolutif du clustering. La section 3 présente les résultats d'expérimentation du cadre proposé sur une variété différente d'algorithmes de clustering statiques. La section 4 conclue le papier et présente les travaux futures.