

# Soft Subspace Growing Neural Gas pour le Clustering de Flux de Données

Mohammed Oualid Attaoui\*, \*\*, Mustapha Lebbah\*, Nabil Keskes\*\*  
Hanene Azzag\*, Mohammed Ghesmoune\*

\* University of Paris 13, Sorbonne Paris City LIPN-UMR 7030 - CNRS,  
99, av. J-B Clément, 93430 Villetaneuse, France  
attaoui, mustapha.lebbah, azzag, ghesmoune@lipn.univ-paris13.fr

\*\* Higher School of Computer Science (ESI-SBA), LabRI Laboratory, Sidi Bel-Abbes, Algeria  
n.keskes@esi-sba.dz

**Résumé.** Le clustering de sous-espaces a été appliquée avec succès dans de nombreux domaines, son objectif est de détecter simultanément les clusters et les sous-espaces d'attributs d'origine dans lequel ces clusters existent. Un flux de données est une séquence massive de données venant en continu. Le clustering de ce type de données nécessite certaines restrictions de temps et de mémoire. Dans cet article, nous proposons une nouvelle méthode appelée S2G-Stream basée sur le clustering de flux de données et le clustering de sous-espaces souple. Des expériences sur des ensembles de données ont montré la capacité de S2G-Stream à détecter simultanément les meilleures attributs, sous-espaces et le meilleur clustering.

## 1 Introduction

Contrairement aux données traditionnelles qui sont immuables et statiques, un flux de données a ses propres caractéristiques : (1) il est infinie. (2) il est en évolution rapide et se produit en temps réel avec des exigences de réponse rapides. (3) un seul (ou peu) passage est possible à travers les données. (4) le stockage du flux de données est limité, seul un résumé des données peut être enregistré. Le clustering de flux de données est une technique qui effectue une analyse de cluster sur un flux de données et est capable de produire des résultats en temps réel. La capacité de traiter les données en un seul passage et de les résumer, tout en utilisant une mémoire limitée, est cruciale pour la clustering des flux de données. Subspace Clustering est une extension de la sélection d'attributs qui tente d'identifier les clusters dans différents sous-espaces du même jeu de données. En tant que sélection d'attributs, le subspace clustering nécessite à la fois une méthode de recherche et un critère d'évaluation. De plus, le clustering de sous-espaces doit restreindre d'une manière ou d'une autre la portée du critère d'évaluation afin de prendre en compte différents sous-espaces pour chaque cluster différent.

Dans le travail précédent Ouattara et al. (2013), les auteurs ont introduit 2S-SOM pour le subspace clustering basé sur les cartes auto-organisatrices SOM. Notre contribution principale dans cet article consiste à étendre le travail dans Ouattara et al. (2013) au clustering de flux de