
Knowledge Integration in Deep Clustering

Nguyen Viet Dung Nghiem^{*1}, Christel Vrain^{†1}, and Thi-Bich-Hanh Dao^{‡1}

¹Laboratoire d'Informatique Fondamentale d'Orléans – Université d'Orléans : EA4022, Institut National des Sciences Appliquées - Centre Val de Loire : EA4022, Institut National des Sciences Appliquées – France

Résumé

Constrained clustering that integrates knowledge in the form of constraints in a clustering process has been studied for more than two decades. Popular clustering algorithms such as K-means, spectral clustering, and recent deep clustering already have their constrained versions, but they still have several drawbacks. A common problem is the lack of expressiveness in the form of constraints. The systems are often targeted to solve some specific types of constraints but cannot reinforce complex knowledge. We propose a deep learning framework that can integrate knowledge presented in any logical formula. In this framework, based on logical formulas, we define two versions of loss computed on the outputs of the deep clustering system. Therefore, our method can integrate general knowledge while having the advantage of being independent of the neural architecture (as long as it can integrate a new form of loss). Experiments are conducted to compare our unified framework to the state-of-the-art systems for together/apart and triplet constraints in terms of computational cost and clustering quality. We show that our system can achieve comparable results with these systems (tailored for these specific constraints) while being flexible to integrate and to learn from high-level domain constraints. To show the flexibility of our method, we consider implication constraints, and we introduce a new constraint called m-clusters Group Constraint.

Mots-Clés: constrained clustering, deep clustering, weighted model counting

*Intervenant

†Auteur correspondant: Christel.Vrain@univ-orleans.fr

‡Auteur correspondant: thi-bich-hanh.dao@univ-orleans.fr