

Invited Talk

Constraint Programming for Constrained Clustering

Christel Vrain

Université d'Orléans
INSA Centre Val de Loire, LIFO EA 4022
45067 Orléans, France
`christel.vrain@univ-orleans.fr`

Abstract. Several works have shown the interest of declarative frameworks, such as Constraint Programming, SAT, Integer Linear Programming, for Data Mining [9,10,12,8,14,15]. Relying on Constraint Programming (CP) has several advantages: first, its inherent declarativity allows to easily model constraints on the Data Mining problem at hand, second, CP has the ability either to enumerate all the solutions or to find an optimal solution, in case of an optimization problem. Moreover, it relies on constraint propagation, which often allows to efficiently prune the search space.

I will mainly focus on constrained clustering [16,7]: user constraints are put on the solutions, thus allowing to get a clustering closer to the one expected by the user.

After giving some backgrounds on CP, I will present the seminal work of [9] on CP for itemset mining, its extension to k-pattern set mining under constraints [13] and its application to conceptual clustering.

The talk will then mainly be focused on distance-based constrained clustering. I will show how we have modeled this task in CP [1,4,3], the difficulties we have had to face, the solutions we have developed [2,5]. The interest of relying on CP will be illustrated through several extensions [4,6,11].

The work on distance-based constrained clustering in CP is a joint work with T.B.H. Dao and K.C. Duong.

References

1. Thi-Bich-Hanh Dao, Khanh-Chuong Duong, and Christel Vrain. A Declarative Framework for Constrained Clustering. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pages 419–434, 2013.
2. Thi-Bich-Hanh Dao, Khanh-Chuong Duong, and Christel Vrain. A Filtering Algorithm for Constrained Clustering with Within-Cluster Sum of Dissimilarities

- Criterion. In *Proceedings of the 25th International Conference on Tools with Artificial Intelligence*, pages 1060–1067, 2013.
3. Thi-Bich-Hanh Dao, Khanh-Chuong Duong, and Christel Vrain. Clustering sous contraintes en ppc. *Revue d'Intelligence Artificielle*, 28(5):523–545, 2014.
 4. Thi-Bich-Hanh Dao, Khanh-Chuong Duong, and Christel Vrain. Constrained clustering by constraint programming. *Artificial Intelligence Journal*, 2015.
 5. Thi-Bich-Hanh Dao, Khanh-Chuong Duong, and Christel Vrain. Constrained minimum sum of squares clustering by constraint programming. In *Principles and Practice of Constraint Programming - 21st International Conference, CP 2015, Cork, Ireland, August 31 - September 4, 2015, Proceedings*, pages 557–573, 2015.
 6. Thi-Bich-Hanh Dao, Christel Vrain, Khanh-Chuong Duong, and Ian Davidson. A framework for actionable clustering using constraint programming. In *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016)*, pages 453–461, 2016.
 7. Ian Davidson and S. S. Ravi. The Complexity of Non-hierarchical Clustering with Instance and Cluster Level Constraints. *Data Mining Knowledge Discovery*, 14(1):25–61, 2007.
 8. Ian Davidson, S. S. Ravi, and Leonid Shamis. A SAT-based Framework for Efficient Constrained Clustering. In *Proceedings of the 10th SIAM International Conference on Data Mining*, pages 94–105, 2010.
 9. Luc De Raedt, Tias Guns, and Siegfried Nijssen. Constraint programming for itemset mining. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 204–212, 2008.
 10. Luc De Raedt, Tias Guns, and Siegfried Nijssen. Constraint Programming for Data Mining and Machine Learning. In *Proc. of the 24th AAAI Conference on Artificial Intelligence*, 2010.
 11. Tias Guns, Thi-Bich-Hanh Dao, Christel Vrain, and Khanh-Chuong Duong. Repetitive branch-and-bound using constraint programming for constrained minimum sum-of-squares clustering. In *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016)*, pages 462–470, 2016.
 12. Tias Guns, Siegfried Nijssen, and Luc De Raedt. Itemset mining: A constraint programming perspective. *Artificial Intelligence*, 175:1951–1983, 2011.
 13. Tias Guns, Siegfried Nijssen, and Luc De Raedt. k-Pattern set mining under constraints. *IEEE Transactions on Knowledge and Data Engineering*, 25(2):402–418, 2013.
 14. Marianne Mueller and Stefan Kramer. Integer Linear Programming Models for Constrained Clustering. In *Proceedings of the 13th International Conference on Discovery Science*, pages 159–173, 2010.
 15. Jean-Philippe Métivier, Patrice Boizumault, Bruno Cremilleux, Medhi Khiari, and Samir Loudni. A constraint-based language for declarative pattern discovery. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 1112–1119, Dec 2011.
 16. Kiri Wagstaff and Claire Cardie. Clustering with instance-level constraints. In *Proceedings of the 17th International Conference on Machine Learning*, pages 1103–1110, 2000.