OPTIMAL DECISION TREES FOR INTERPRETABLE CLUSTERING WITH CONSTRAINTS

POUYA SHATI, ELDAN COHEN, SHEILA MCILRAITH

CP'23 – Doctoral Program

Originally appeared at IJCAI-23

UNIVERSITY OF TORONTO

VECTOR INSTITUTE

Summary

The constrained clustering task employs limited amounts of supervision formulated as constraints, to incorporate task-specific knowledge and improve accuracy [1].

Decision trees are compact yet accurate forms of solutions for interpretable machine learning, particularly classification [2].

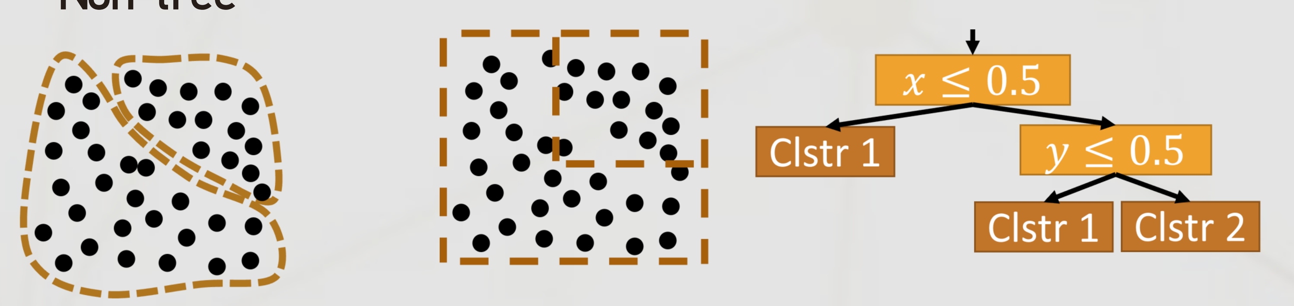
The state-of-the-art approaches for decision tree clustering do not support constraints or provide optimality guarantee.

Our approach is a MaxSAT-based encoding of decision tree clustering which supports pairwise constraints and optimizes an approximation of a well-known bi-criteria.

Experimental results show our approach is able to learn good quality solutions in short amounts of time. Decision tree clustering outperforms its non-tree counterpart and fits well with the bi-criteria and the utilization of constraints.

Problem Definition

Tree Clustering



Objectives

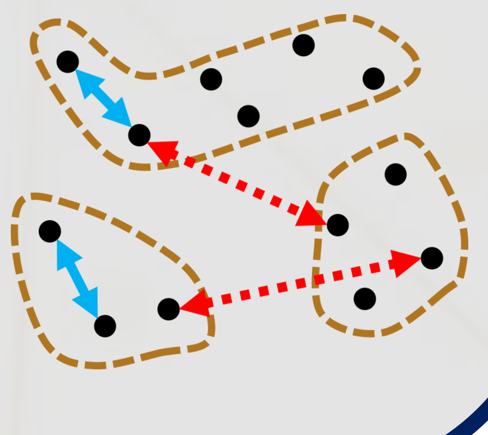
Minimize maximum diameter (MD) within clusters

Maximize minimum split (MS) between clusters

Constraints

Must-links: pairs that should be in the same cluster

Cannot-links: pairs that should be in different cluster



Encoding

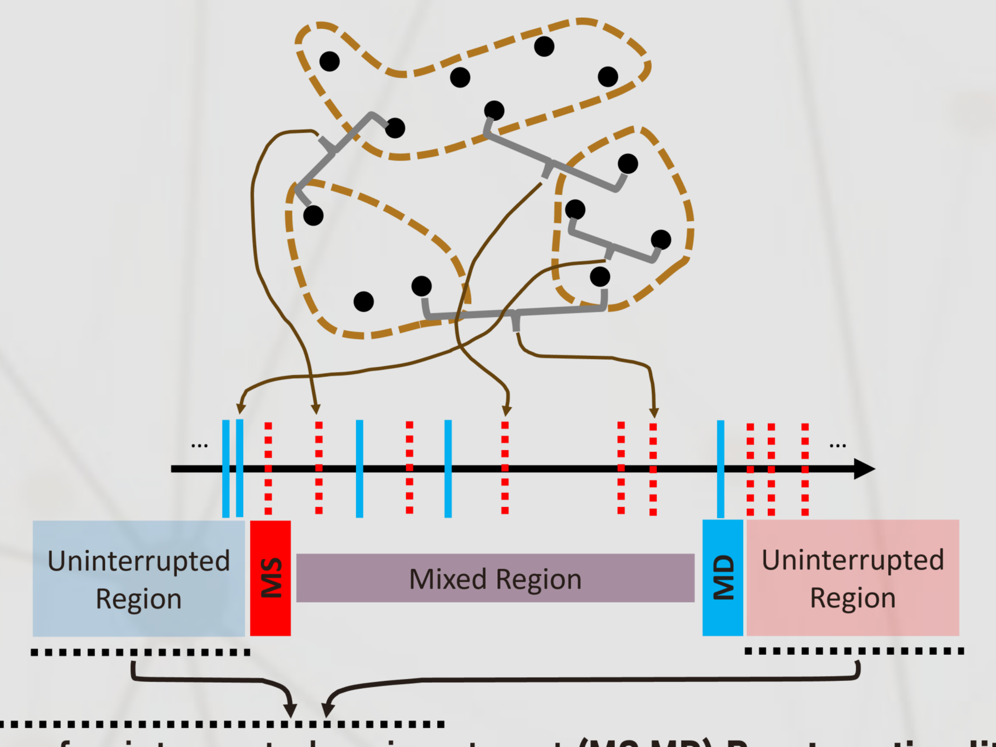
Based on our previous work on decision tree classifiers [3] with direct encoding of numerical features

Objectives

Our objectives involve sorting distances of pairs

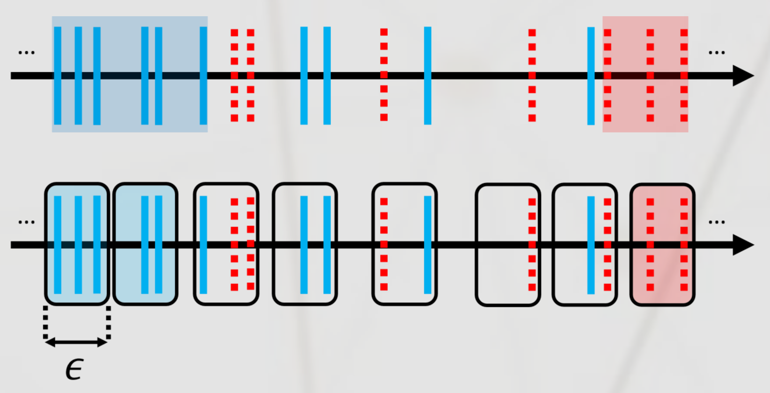
Given a clustering, each pair belongs to same (|) / different (⋮) clusters

Minimum split and maximum diameter are points along the axis



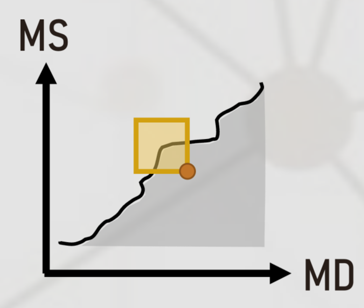
Optimize the sum of uninterrupted regions to get **(MS,MD) Pareto optimality**

Approximation



There are a quadratic number of pairs

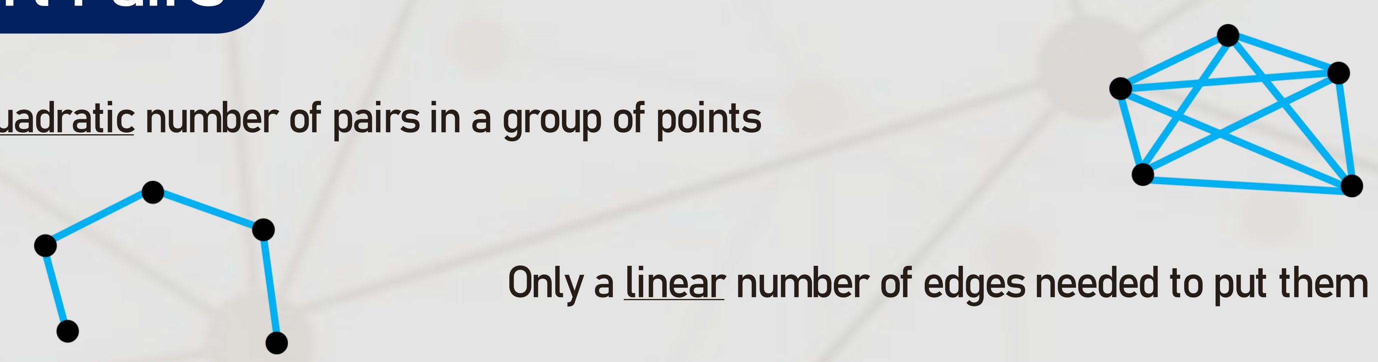
Use distance classes instead of individual pairs



Pareto-optimal in number of classes → Pareto-optimal in number of pairs

Smart Pairs

A quadratic number of pairs in a group of points



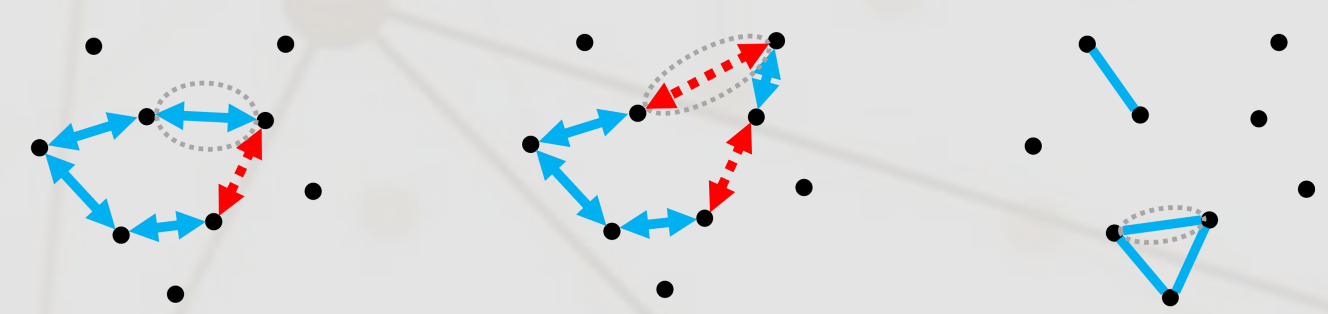
Only a linear number of edges needed to put them in the same cluster

	Constraints	Conditional
Force to be in the same cluster	Must-link	MS obj.
Force to be in different clusters	Cannot-link	MD obj.

Smart Pairs →

Detect **redundant** edges

Detect **infeasible** edges



Experiments

Setup

Baselines:
- Constrained Clustering (non-tree, max diameter only)
- Mixed Integer Optimization [4]

Datasets:
- 7 real datasets from the UCI repository
- 4 synthetic datasets from FCPS

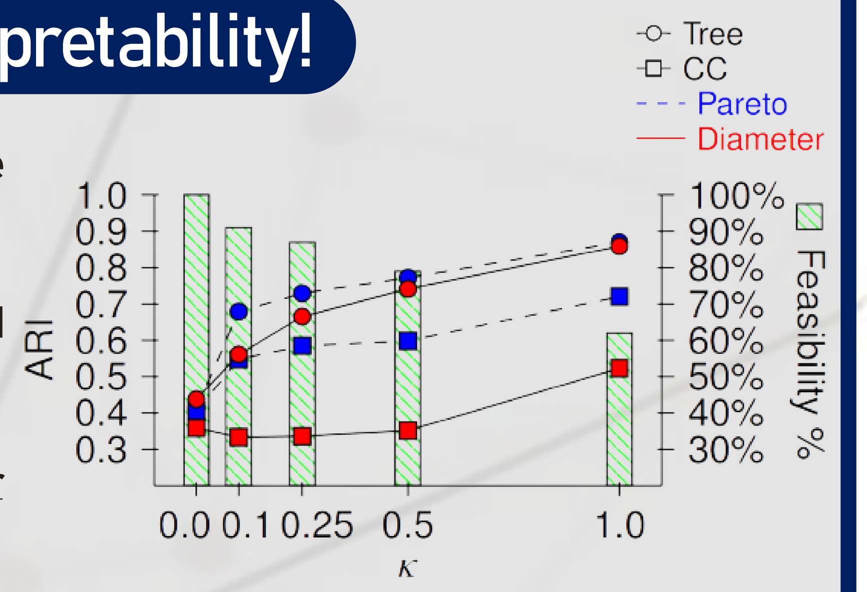
Solver:
- Loandra with 30 minutes time limit

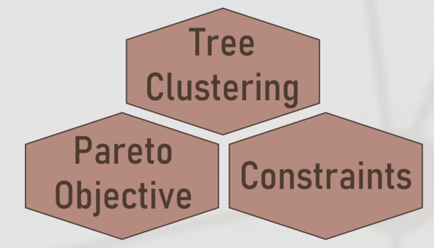
Better Score + Interpretability!

High quality interpretable solutions in short time

A trade-off between quality and feasibility

Higher accuracy despite lower objective value





The 3 main aspects improve the solution and complement each other

Better performance!

Adding smart pairs and approximation
- shortens runtime
- reduces number of clauses
- but maintains the accuracy

Dataset	Setting	ARI	Time (s)	# Clauses
Spam	SP & $\epsilon = 0.1$	Inf.	151.6	3M
	$\epsilon = 0.1$	Inf.	332.7	24M
	$\epsilon = 0$	Unk.	864.0	69M

Future Work

Investigate strategies or tools for exploring the Pareto front and selecting the most promising solution

Address the infeasibility issue, e.g., by converting hard clustering constraints into soft constraints that are encouraged rather than required to be satisfied

[1] Kiri Wagstaff and Claire Cardie, ICML-2000

[2] Alexey Ignatiev, Joao Marques-Silva, Nina Narodytska, and Peter J Stuckey, IJCAI-2021

[3] Pouya Shati, Eldan Cohen, and Sheila McIlraith, CP-2021

[4] Dimitris Bertsimas, Agni Orfanoudaki, Holly Wiberg, Machine Learning, 2021