Multiobjective Optimization of Co-clustering Ensembles

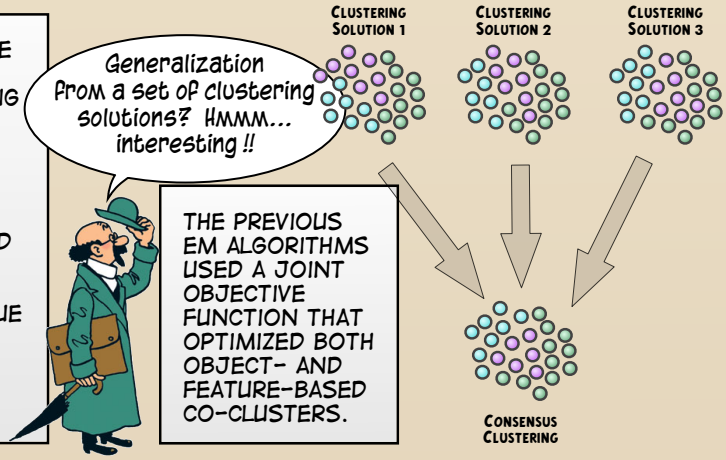
Francesco Gullo Yahoo Research, gullo@yahoo-inc.com
AKM Khaled A. Talukder George Mason Univ., atalukde@gmu.edu
Sean Luke George Mason Univ., sean@cs.gmu.edu
Carlotta Domeniconi George Mason Univ., cdomenic@gmu.edu
Andrea Tagarelli Univ. of Calabria, tagarelli@deis.unical.it

Summary: We compare two preference-based multiobjective evolutionary algorithms against two expectation-maximization (*EM*) methods for finding optimal co-clustering ensembles.

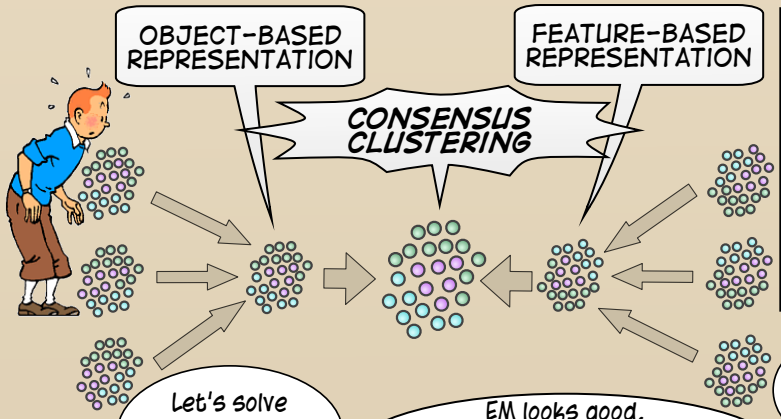
Previously, an *EM*-like algorithm (CB-PCE) was developed, along with a faster approximation algorithm (FCB-PCE). The new ***multiobjective*** method optimizes two aspects of co-clustering solutions: the ***object***-based and ***feature***-based representations of their data points.

Co-clustering: An unsupervised machine learning technique to identify object groups related by similar feature values.

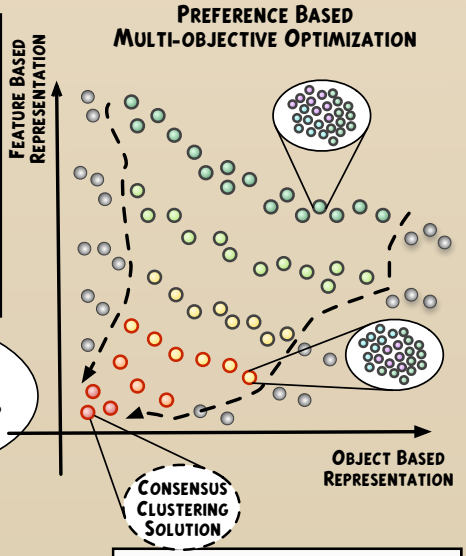
Clustering ensembles: Generalizing a clustering from a set of clustering solutions.

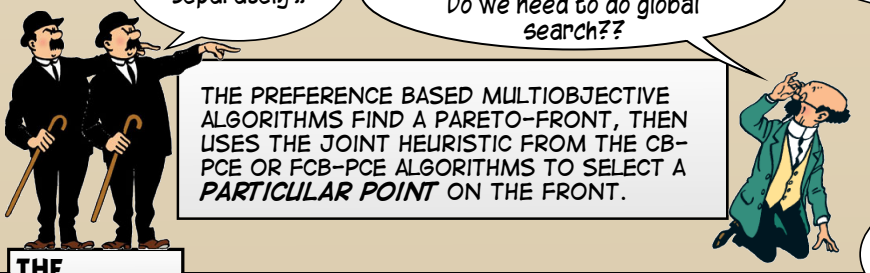


The previous EM algorithms used a joint objective function that optimized both object- and feature-based co-clusters.



In this study, we developed two preference-based multiobjective evolutionary algorithms that each used two objectives as ***separate*** goals, unlike the EM algorithms.





The preference based multiobjective algorithms find a Pareto-front, then uses the joint heuristic from the CB-PCE or FCB-PCE algorithms to select a ***particular point*** on the front.

The preference based multiobjective algorithms ***outperformed*** their respective EM methods on simpler datasets but ***underperformed*** on more complex datasets.

A possible reason: the search space may ***not*** have local optima (we don't know yet for sure)! And EM, essentially a hillclimber, would be expected to perform best in this situation.

The results...

	Algorithm	Θ_f	P-value	Θ_o	P-value	Θ_{of}	P-value
Iris	MOEA-CB-PCE	**0.5564**	(> 99.8%)	0.0813		**0.7348**	(> 95.0%)
	CB-PCE	0.2332		**0.4702**	(> 99.8%)	0.6902	
	MOEA-FCB-PCE	**0.5444**	(> 99.8%)	0.1826	≈	**0.2027**	(> 95.0%)
	FCB-PCE	0.2002		0.2655	≈	0.1805	
Wine	MOEA-CB-PCE	**0.5945**	(> 99.8%)	0.1943		**0.7748**	(> 99.8%)
	CB-PCE	0.1142		**0.4119**	(> 99.0%)	0.3402	
	MOEA-FCB-PCE	**0.6244**	(> 99.8%)	0.1001		0.2381	≈
	FCB-PCE	0.1463		**0.2518**	(> 99.8%)	0.2355	≈
Glass	MOEA-CB-PCE	**0.7464**	(> 99.8%)	0.0834		**0.9048**	(> 99.8%)
	CB-PCE	0.1302		**0.4702**	(> 99.8%)	0.1203	
	MOEA-FCB-PCE	**0.6244**	(> 95.0%)	0.1026		0.2397	≈
	FCB-PCE	0.4639		**0.4525**	(> 99.8%)	0.1193	≈
E. Coli	MOEA-CB-PCE	-0.003		-0.0013		0.011	
	CB-PCE	**0.0046**	(> 99.8%)	**0.0894**	(> 99.8%)	**0.0881**	(> 99.8%)
	MOEA-FCB-PCE	-0.003		-0.0009		0.019	
	FCB-PCE	**0.0043**	(> 99.8%)	**0.105**	(> 99.8%)	**0.112**	(> 99.8%)
TraceData	MOEA-CB-PCE	-0.00241		-0.0492		-0.0587	
	CB-PCE	**0.018**	(> 99.8%)	**0.2347**	(> 99.8%)	**0.2493**	(> 99.8%)
	MOEA-FCB-PCE	-0.00246		-0.04956		-0.0591	
	FCB-PCE	**0.0177**	(> 99.8%)	**0.1854**	(> 99.8%)	**0.247**	(> 99.8%)

George Mason University
NSF
Università della Calabria
Yahoo!