

A survey of density based clustering algorithms

Panthadeep BHATTACHARJEE, Pinaki MITRA

Frontiers of Computer Science, DOI: [10.1007/s11704-019-9059-3](https://doi.org/10.1007/s11704-019-9059-3)

Current Problem and original ideas

Problem: Lack of extensive study on Density based clustering algorithms(DBCLAs).

Existing ideas: Surveys on DBCLAs covered selected algorithms focusing on limited aspects.

Survey paper (year)	No. of DBCLAs studied	Common properties	Classification of DBCLAs	Conceptual dependency	Applications	Future directions
[13](2011)	5	×	×	×	✓	✓
[14](2011)	3	×	×	×	×	×
[15](2013)	7	×	×	×	×	×
[16](2014)	3	×	×	×	×	×
[17](2014)	3	×	×	×	×	×
[18](2014)	6	×	×	×	✓	✓
[19](2016)	2	×	×	×	×	×
Our survey of DBCLAs	32	✓	✓	✓	✓	✓

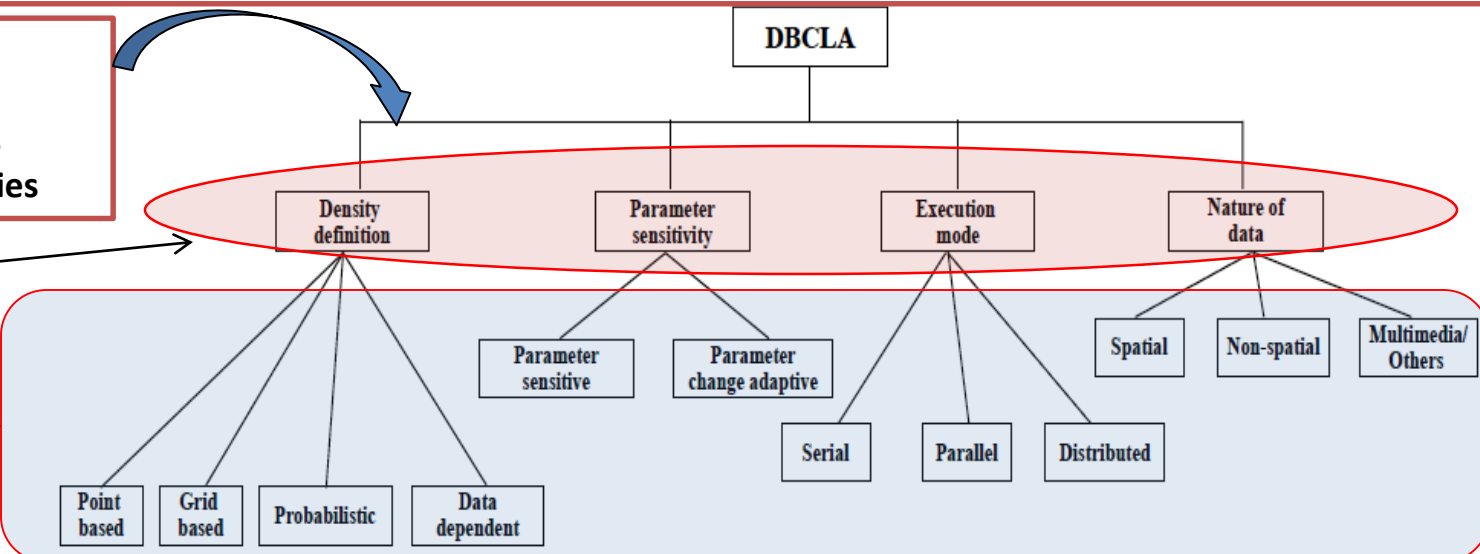
Refer Citation in paper

Our key contribution(s):

Our survey covers 32 major DBCLAs since 1996, which is the maximum number of DBCLAs that have ever been studied in a single survey ;Draw comparisons with previous studies on DBCLAs; Comprehensive classification of DBCLAs.

Contribution:
Provide strategy for classification of DBCLAs under different categories

Categories for classification
Classification heads for DBCLAs under each category



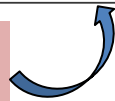
Main Results/Conclusions

Category	Class	#DBCLAs	#Total DBCLAs	Percentage of DBCLAs
Density definition	Point based	23	32	71.875 %
	Grid based	8		25 %
	Probabilistic	7		21.875 %
	Data dependent	2		6.25 %
Parameter sensitivity	Parameter sensitive	28		87.5 %
	Parameter change adaptive	6		18.75 %
Execution mode	Serial	27		84.375 %
	Parallel	4		12.5 %
	Distributed	3		9.375 %
Nature of data	Spatial	25		78.125 %
	Non-spatial	6		18.75 %
	Multimedia/others	5		15.625 %

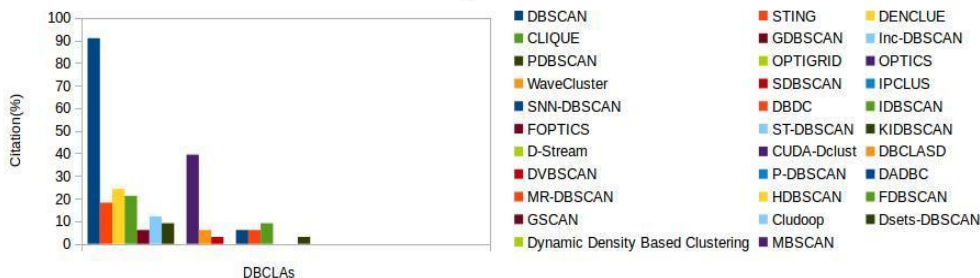
Major Conclusions:

1. One of the most comprehensive surveys with a wide variety of DBCLAs(32 algorithms) being studied since 1996.
2. DBSCAN remains one of the most influential DBCLAs with a large number of DBCLAs inheriting its clustering method.
3. Most DBCLAs are parameter sensitive, execute serially and rely on point based density computation.
4. The share of multi-class DBCLAs is at its highest while using a variety of techniques in defining the density.

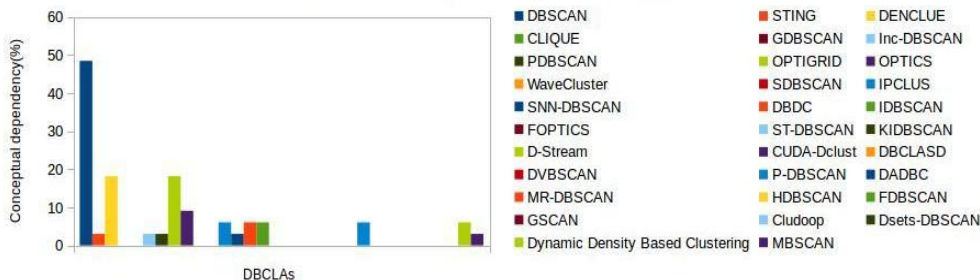
Percentage of DBCLAs in each class



Citation Percentage Variation of DBCLAs



Conceptual Dependency Percentage Variation of DBCLAs



Percentage of DBCLAs in multiple classes



Category	DBCLA	#Classes	Class name	Percentage of multi-class DBCLAs
Density definition	DENCLUE	2	Grid based, Probabilistic	21.875 %
	OPTIGRID	2	Grid based, Probabilistic	
	IPCLUS	2	Grid based, Probabilistic	
	SNN-DBSCAN	2	Point based, Data dependent	
	DBCLASD	2	Grid based, Probabilistic	
	GSCAN	2	Grid based, Probabilistic	
	MBSCAN	2	Point based, Data dependent	
Parameter sensitivity				
Execution mode	GSCAN	2	Serial, Parallel	3.125 %
Nature of data	GDBSCAN	2	Spatial, Non-spatial	9.375 %
	ST-DBSCAN	2	Spatial, Non-spatial	
	HDBSCAN	2	Spatial, Multimedia/others	