

University of Groningen

Scalable analysis and visualization of high-dimensional astronomical data sets

Ferdosi, Bilkis Jamal

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Ferdosi, B. J. (2011). *Scalable analysis and visualization of high-dimensional astronomical data sets*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Contents

1	Introduction	1
1.1	Illustrating the Problem	2
1.2	Goal of the Thesis	4
1.3	High-Dimensional Data Visualization	4
1.3.1	Volume Visualization	5
1.3.2	Scatter Plot Matrix (SPM)	6
1.3.3	Parallel Coordinate Plot (PCP)	7
1.3.4	Radial visualization (RadViz)	9
1.3.5	Tours	9
1.3.6	Interaction	10
1.4	Astronomical Tools in Use	11
1.5	Thesis Contribution and Organization	15
2	Comparison of Density Estimation Methods for Astronomical Datasets	17
2.1	Introduction	17
2.2	Density estimation methods	19
2.2.1	k-nearest neighbor method	19
2.2.2	Adaptive Epanechnikov kernel density estimation	20
2.2.3	Adaptive Gaussian kernel density estimation (DEDICA)	23
2.2.4	Delaunay Tessellation Field Estimator (DTFE)	24
2.2.5	Error measures	25
2.3	Datasets	26
2.3.1	Simulated datasets with known density fields	26
2.3.2	Astronomical datasets with unknown density fields	28
2.4	Results	32
2.4.1	Simulated Datasets	32
2.4.2	Application to SDSS datasets	35
2.5	Conclusions and Recommendations	40
2.5.1	Artificial and Simulated Datasets	40
2.5.2	SDSS Datasets	41

2.5.3	DEDICA	41
2.5.4	Computational Complexities	42
2.5.5	Recommendations	43
2.A	Appendix. Mock samples – selection effects	44
2.A.1	Edge Effects	45
2.B	Appendix. DTFE Monte Carlo sampling	48
3	Finding and Visualizing Relevant Subspaces for Clustering High-Dimensional Data Using Connected Morphological Operators	49
3.1	Introduction	50
3.2	Related Work	52
3.2.1	Subspace Clustering and Ranking	52
3.2.2	Visualization	53
3.3	Searching Relevant Subspaces for Clustering	53
3.3.1	Overview of the Method	53
3.3.2	Density Estimation	54
3.3.3	Connected Morphological Operators	55
3.3.4	Dynamics	56
3.3.5	Subspace Quality Criterion	57
3.3.6	Subspace Finding	58
3.4	Interactive Visual Subspace Exploration	60
3.5	Experiments and Results	61
3.5.1	Synthetic Data	62
3.5.2	Astronomical Data	63
3.6	Gene Expression Data	67
3.7	Summary and Future Plans	72
4	Visualizing High-Dimensional Structures by Dimension Ordering and Filtering using Subspace Analysis	73
4.1	Introduction	74
4.2	Related work	75
4.3	Dimension reordering	77
4.3.1	Overview of the method	77
4.3.2	Structure-based Full Ordering (SBF) for PCP	78
4.3.3	Structure-based Partial Ordering (SBP) for PCP	79
4.3.4	Structure-based Simple Ordering (SBS) for SPM	80
4.4	Visualization	80
4.4.1	Histogram-based Parallel Coordinate Plot	80
4.5	Experimental results	81
4.5.1	Datasets	81
4.5.2	Performance of the Methods	82
4.5.3	SBP method	86
4.5.4	SBS method	86

4.5.5	Comparison between PCP and SPM	88
4.5.6	Limitations	88
4.6	Summary and future work	90
5	Towards the Design of a Visual Analytics Tool for Astronomical Data on Large, Touch-Sensitive Displays	91
5.1	Introduction	91
5.2	Related Work	93
5.3	Background	94
5.3.1	Subspace Ranking for Clustering	94
5.3.2	Touch Interactions with 7 DOF	95
5.4	Design Issues and the Prototype	98
5.5	Summary and Conclusion	100
6	Concluding Remarks	103
6.1	Summary and Conclusion	103
6.2	Future Outlook	105
6.2.1	Visual Clustering	105
6.2.2	Subspace Ranking for Outlier Detection	105
6.2.3	Analyzing Astronomical Datasets in Co-located Collaborative Environment	105
	Samenvatting	107
	Publications	109
	Bibliography	111