

Joshua Huang

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/5648769/publications.pdf>

Version: 2024-02-01

94
papers

3,207
citations

304368

22
h-index

161609

54
g-index

95
all docs

95
docs citations

95
times ranked

2570
citing authors

#	ARTICLE	IF	CITATIONS
1	Automated variable weighting in k-means type clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27, 657-668.	9.7	624
2	An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data. IEEE Transactions on Knowledge and Data Engineering, 2007, 19, 1026-1041.	4.0	489
3	Big data analytics on Apache Spark. International Journal of Data Science and Analytics, 2016, 1, 145-164.	2.4	244
4	Agglomerative Fuzzy K-Means Clustering Algorithm with Selection of Number of Clusters. IEEE Transactions on Knowledge and Data Engineering, 2008, 20, 1519-1534.	4.0	217
5	Stratified sampling for feature subspace selection in random forests for high dimensional data. Pattern Recognition, 2013, 46, 769-787.	5.1	127
6	Fuzzy nonlinear regression analysis using a random weight network. Information Sciences, 2016, 364-365, 222-240.	4.0	108
7	Joint Optimization of Energy Consumption and Latency in Mobile Edge Computing for Internet of Things. IEEE Internet of Things Journal, 2019, 6, 4791-4803.	5.5	96
8	Knowledge-based vector space model for text clustering. Knowledge and Information Systems, 2010, 25, 35-55.	2.1	78
9	Incremental density-based ensemble clustering over evolving data streams. Neurocomputing, 2016, 191, 34-43.	3.5	73
10	I-nice: A new approach for identifying the number of clusters and initial cluster centres. Information Sciences, 2018, 466, 129-151.	4.0	70
11	Random Sample Partition: A Distributed Data Model for Big Data Analysis. IEEE Transactions on Industrial Informatics, 2019, 15, 5846-5854.	7.2	66
12	Semi-supervised Feature Selection via Rescaled Linear Regression. , 2017, , .		65
13	Unbiased Feature Selection in Learning Random Forests for High-Dimensional Data. Scientific World Journal, The, 2015, 2015, 1-18.	0.8	60
14	Semi-supervised Aspect-level Sentiment Classification Model based on Variational Autoencoder. Knowledge-Based Systems, 2019, 171, 81-92.	4.0	58
15	Stratified feature sampling method for ensemble clustering of high dimensional data. Pattern Recognition, 2015, 48, 3688-3702.	5.1	50
16	PurTreeClust: A Clustering Algorithm for Customer Segmentation from Massive Customer Transaction Data. IEEE Transactions on Knowledge and Data Engineering, 2018, 30, 559-572.	4.0	44
17	Variable Weighting in Fuzzy k-Means Clustering to Determine the Number of Clusters. IEEE Transactions on Knowledge and Data Engineering, 2020, 32, 1838-1853.	4.0	35
18	A Data Cube Model for Prediction-Based Web Prefetching. Journal of Intelligent Information Systems, 2003, 20, 11-30.	2.8	32

#	ARTICLE	IF	CITATIONS
19	Random weight network-based fuzzy nonlinear regression for trapezoidal fuzzy number data. Applied Soft Computing Journal, 2018, 70, 959-979.	4.1	29
20	An efficient random forests algorithm for high dimensional data classification. Advances in Data Analysis and Classification, 2018, 12, 953-972.	0.9	28
21	Distributed Data Strategies to Support Large-Scale Data Analysis Across Geo-Distributed Data Centers. IEEE Access, 2020, 8, 178526-178538.	2.6	28
22	Scalable Normalized Cut with Improved Spectral Rotation. , 2017, , .		28
23	Adaptive discriminant analysis for semi-supervised feature selection. Information Sciences, 2021, 566, 178-194.	4.0	25
24	Segmentation of Factories on Electricity Consumption Behaviors Using Load Profile Data. IEEE Access, 2016, 4, 8394-8406.	2.6	24
25	Enhanced Balanced Min Cut. International Journal of Computer Vision, 2020, 128, 1982-1995.	10.9	22
26	Recommending high-utility search engine queries via a query-recommending model. Neurocomputing, 2015, 167, 195-208.	3.5	21
27	Two-level quantile regression forests for bias correction in range prediction. Machine Learning, 2015, 101, 325-343.	3.4	21
28	An Algorithm for Clustering Categorical Data With Set-Valued Features. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29, 4593-4606.	7.2	21
29	Variational Autoencoder-Based Dimensionality Reduction for High-Dimensional Small-Sample Data Classification. International Journal of Computational Intelligence and Applications, 2020, 19, .	0.6	21
30	Exploring and cleaning big data with random sample data blocks. Journal of Big Data, 2019, 6, .	6.9	18
31	A distributed data management system to support large-scale data analysis. Journal of Systems and Software, 2019, 148, 105-115.	3.3	18
32	An Asymptotic Ensemble Learning Framework for Big Data Analysis. IEEE Access, 2019, 7, 3675-3693.	2.6	17
33	Fuzzy K-Means with Variable Weighting in High Dimensional Data Analysis. , 2008, , .		16
34	Trend analysis of categorical data streams with a concept change method. Information Sciences, 2014, 276, 160-173.	4.0	15
35	A Robust k -Means Clustering Algorithm Based on Observation Point Mechanism. Complexity, 2020, 2020, 1-11.	0.9	14
36	An effective content-based event recommendation model. Multimedia Tools and Applications, 2021, 80, 16599-16618.	2.6	14

#	ARTICLE	IF	CITATIONS
37	A Two-Stage Data Processing Algorithm to Generate Random Sample Partitions for Big Data Analysis. Lecture Notes in Computer Science, 2018, , 347-364.	1.0	13
38	Generate pairwise constraints from unlabeled data for semi-supervised clustering. Data and Knowledge Engineering, 2019, 123, 101715.	2.1	12
39	Machine Learning-Based Multi-Layer Multi-Hop Transmission Scheme for Dense Networks. IEEE Communications Letters, 2019, 23, 2238-2242.	2.5	12
40	Service selection in dynamic demand-driven Web services. , 2004, , .		11
41	Automatic Transaction Compensation for Reliable Grid Applications. Journal of Computer Science and Technology, 2006, 21, 529-536.	0.9	11
42	Recent advances in multiple criteria decision making techniques. International Journal of Machine Learning and Cybernetics, 2016, , 1.	2.3	11
43	A fuzzy SV- k -modes algorithm for clustering categorical data with set-valued attributes. Applied Mathematics and Computation, 2017, 295, 1-15.	1.4	11
44	TWCC: Automated Two-way Subspace Weighting Partitional Co-Clustering. Pattern Recognition, 2018, 76, 404-415.	5.1	11
45	Improved I-nice clustering algorithm based on density peaks mechanism. Information Sciences, 2021, 548, 177-190.	4.0	11
46	Learning classifier system ensemble and compact rule set. Connection Science, 2007, 19, 321-337.	1.8	10
47	Creating synthetic minority class samples based on autoencoder extreme learning machine. Pattern Recognition, 2022, 121, 108191.	5.1	10
48	CPLP: An algorithm for tracking the changes of power consumption patterns in load profile data over time. Information Sciences, 2018, 429, 332-348.	4.0	9
49	Machine Learning Based Dynamic Cooperative Transmission Framework for IoUT Networks. , 2019, , .		9
50	Activeness and Loyalty Analysis in Event-Based Social Networks. Entropy, 2020, 22, 119.	1.1	9
51	Novel kernel density estimator based on ensemble unbiased cross-validation. Information Sciences, 2021, 581, 327-344.	4.0	9
52	PurTreeClust: A purchase tree clustering algorithm for large-scale customer transaction data. , 2016, , .		8
53	k -mw-modes: An algorithm for clustering categorical matrix-object data. Applied Soft Computing Journal, 2017, 57, 605-614.	4.1	8
54	C3C: A New Static Content-Based Three-Level Web Cache. IEEE Access, 2019, 7, 11796-11808.	2.6	8

#	ARTICLE	IF	CITATIONS
55	A novel dependency-oriented mixed-attribute data classification method. Expert Systems With Applications, 2022, 199, 116782.	4.4	8
56	CNP-based Implementation of Service-oriented Workflow Mapping in SHGWMS. World Wide Web, 2008, 11, 199-226.	2.7	7
57	Cluster Survival Model of Concept Drift in Load Profile Data. IEEE Access, 2018, 6, 51269-51285.	2.6	6
58	RRPlib: A spark library for representing HDFS blocks as a set of random sample data blocks. Science of Computer Programming, 2019, 184, 102301.	1.5	6
59	Novel electricity pattern identification system based on improved I-nice algorithm. Computers and Industrial Engineering, 2020, 150, 106908.	3.4	6
60	Clustering of SNP Data with Application to Genomics. , 2006, , .		5
61	Bias-corrected Quantile Regression Forests for high-dimensional data. , 2014, , .		5
62	Query ranking model for search engine query recommendation. International Journal of Machine Learning and Cybernetics, 2017, 8, 1019-1038.	2.3	5
63	A New Location-Based Topic Model for Event Attendees Recommendation. , 2019, , .		5
64	A new approach to solve opinion dynamics on complex networks. Expert Systems With Applications, 2020, 145, 113132.	4.4	5
65	MMCo-Clus – An Evolutionary Co-clustering Algorithm for Gene Selection. IEEE Transactions on Knowledge and Data Engineering, 2022, 34, 4371-4384.	4.0	5
66	Integrating constraints to support legally flexible business processes. Information Systems Frontiers, 2011, 13, 171-189.	4.1	4
67	A frequency-based gene selection method with random forests for gene data analysis. , 2016, , .		4
68	I-Sampling: A New Block-Based Sampling Method for Large-Scale Dataset. , 2017, , .		4
69	Random Sample Partition-Based Clustering Ensemble Algorithm for Big Data. , 2021, , .		4
70	Recommending High Utility Queries via Query-Reformulation Graph. Mathematical Problems in Engineering, 2015, 2015, 1-14.	0.6	3
71	A Hierarchical Gamma Mixture Model-Based Method for Classification of High-Dimensional Data. Entropy, 2019, 21, 906.	1.1	3
72	Improved blog clustering through automated weighting of text blocks. , 2009, , .		2

#	ARTICLE	IF	CITATIONS
73	Ensemble subspace clustering of text data using two-level features. International Journal of Machine Learning and Cybernetics, 2017, 8, 1751-1766.	2.3	2
74	Selection of diverse features with a diverse regularization. Pattern Recognition, 2021, 120, 108154.	5.1	2
75	A data warehousing and data mining framework for web usage management. Communications in Information and Systems, 2004, 4, 301-324.	0.3	2
76	A Hybrid Method to Measure Distribution Consistency of Mixed-Attribute Datasets. IEEE Transactions on Artificial Intelligence, 2023, 4, 182-196.	3.4	2
77	Bayesian Attribute Bagging-Based Extreme Learning Machine for High-Dimensional Classification and Regression. ACM Transactions on Intelligent Systems and Technology, 2022, 13, 1-26.	2.9	2
78	RSP-Hist: Approximate Histograms for Big Data Exploration on Hadoop Clusters. , 2021, , .		2
79	Adaptive scheduling for shared window joins over data streams. Frontiers of Computer Science, 2007, 1, 468-477.	0.6	1
80	A Changing Window Approach to Exploring Gene Expression Patterns. , 2008, , .		1
81	An improved random forest approach for detection of hidden web search interfaces. , 2008, , .		1
82	Latent Feature Group Learning for High-Dimensional Data Clustering. Information (Switzerland), 2019, 10, 208.	1.7	1
83	Long and Short Term Risk Control for Online Portfolio Selection. Lecture Notes in Computer Science, 2020, , 472-480.	1.0	1
84	FINC: An Efficient and Effective Optimization Method for Normalized Cut. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, PP, 1-1.	9.7	1
85	A Two-Stage Missing Value Imputation Method Based on Autoencoder Neural Network. , 2021, , .		1
86	Random vector functional link network with subspace-based local connections. Applied Intelligence, 0, , 1.	3.3	1
87	Wireless Network Slice Assignment With Incremental Random Vector Functional Link Network. IEEE Transactions on Network Science and Engineering, 2023, 10, 1283-1296.	4.1	1
88	CLUSTERING AND CLUSTER VALIDATION IN DATA MINING. , 2003, , .		0
89	SLF4SS: Facilitating Flexible Services Selection. , 2006, , .		0
90	A Rough Set Method for Supplier Performance Evaluation in E-Business. , 2007, , .		0

#	ARTICLE	IF	CITATIONS
91	Clustering Ensembles Based on Probability Density Function Estimation. , 2020, , .		0
92	Observation Points-Based Particle Swarm Optimization Algorithm. , 2020, , .		0
93	A Compressed Hidden Naive Bayesian Classifier. , 2021, , .		0
94	Observation points classifier ensemble for high-dimensional imbalanced classification. CAAI Transactions on Intelligence Technology, 2023, 8, 500-517.	3.4	0