

# DeepInsight 交互式可视化分析平台

算子库V1.0.2

发布日期：2018年10月12日

分级	算子分类	分类编号	算子库名称
标准版	数据源	(文本型)	数据源-文本型
		(关系型)	数据源-关系型
		(Nosql)	数据源-Nosql
		(保存还原)	数据源-保存还原
		(生成)	数据源-生成
	数据处理	(代码容器)	数据处理-代码容器
		(记录处理)	数据处理-记录
		(清洗处理)	数据处理-清洗
		(格式转换)	数据处理-格式转换
		(挖掘预处理)	数据处理-挖掘预处理
		(效果衡量)	数据处理-效果衡量
		(时间处理)	数据处理-时间
		(分箱处理)	数据处理-分箱
		(图关系处理)	数据处理-图关系
		(其它处理)	数据处理-其它
	数据挖掘	(相关分析)	数据挖掘-相关分析
		(分类算法)	数据挖掘-分类算法
		(回归算法)	数据挖掘-回归算法
		(降维算法)	数据挖掘-降维算法
		(聚类算法)	数据挖掘-聚类算法
		检验	数据挖掘-检验
		(时间序列)	数据挖掘-时间序列
		(分词)	数据挖掘-分词
		关联规则	数据挖掘-关联规则
		连通性分析	数据挖掘-连通性分析
	异常检测	(基于统计)	异常检测-基于统计
		(基于聚类)	异常检测-基于聚类
	特征提取	(频次特征)	特征提取-频次特征
	检索	ElasticSearch	检索-ElasticSearch
	流处理	流数据源	流处理-流数据源
		流框架	流处理-流框架
专业版	轨迹数据挖掘	(轨迹数据)	高级-轨迹数据挖掘-轨迹数据
	数据挖掘	(聚类算法)	高级-数据挖掘-聚类算法
		(分类算法)	高级-数据挖掘-分类算法
		(回归算法)	高级-数据挖掘-回归算法
	检索	ElasticSearch	高级-检索-ElasticSearch
	高级文本处理	(相似性)	高级-文本处理-相似性
		(特征)	高级-文本处理-特征
	异常检测	基于距离	高级-异常检测-基于距离
		(基于偏离)	高级-异常检测-基于偏离
	流处理	流数据源	高级-流处理-流数据源

算子库名称	序号	算子名称	说明文档	作者	ID	6088 版本号	6093 版本号	备注
数据源-文本型	1	文本目录数据源V2	文本目录数据源新版升级算子	何流	dd377635494a438a4125f1965e01	69	67	
	2	文本数据源V2	文本数据源V2.docx		b9e0cff9365e495a96ace978124d6a	36	36	
	3	Excel数据源	excel数据源.docx	高松, 宋波	2353	72	72	(csv, xls, xlsx, xlsxm)
	4	按目录读取excel文件	按目录读取excel文件.docx	宋波	fe3edac78942d6a935e01f19eabf	111	111	
数据源-关系型	1	数据库并行加载	数据库并行加载.docx	白圆	58b26f5b7549f4acad834b0b28d0	72	75	
	2	读取Access数据	读取Access数据.docx	朱航	fc34a434774a4f9bbc9d5894c24e	62	65	
	3	DF写入数据库	DF写入数据库.docx	白圆	e0f1397d0840099179eebcfca3f0	106	106	
	4	读关系型数据库	读关系型数据库.docx	高松	2354	7	11	( gbase, sqlserver, mysql, orac le, mongodb)
	5	读Hive数据源	读Hive数据源.docx	高松	2803	36	36	
数据源-Nosql	1	MongoDB并行加载	MongoDB并行加载.docx	白圆	f1e973712c4c74acc296cd009868	92	95	
	2	Hbase读取	HBase读取.docx	李先霖	6b53ae23784766b7c7b585777c7a	70	28	
	3	Hbase写入	HBase写入.docx	李先霖	032199447043c0a3b25ecd7c390e	21	8	
数据源-保存还原	1	读远程文件至HDFS	读远程文件至HDFS.docx	朱航	780e62842d46d8a0e81d6aee5350	63	63	(ftp, http)
	2	从HDFS还原DF	从HDFS还原DF.docx	金子添	2933	5	5	
	3	Mysql库表备份至HDFS	Mysql库表备份至HDFS.docx	朱航	66a6b811b3469ea5e46998da628c	37	37	
	4	DF保存到hdfs	DF保存到HDFS.docx		3018	16	16	
	5	DF保存到固定前缀HDFS	DF数据保存到固定前缀Hdfs.docx	马朋涛	2931	11	15	
数据源-生成	1	数据生成器	数据生成器.docx	马朋涛	322d06335e46df8475e8f5275a03	109	109	
数据处理-代码容器	1	Scala	Scala组件.docx	齐咏杰	2188	12	12	
	2	SQL	SQL.docx	齐咏杰	2182	12	12	
	3	Python	Python.docx	齐咏杰	5938c42ac84f1c91ff81934da38d	3	3	
	4	R	R算子使用文档.docx	项君	0c998ed5bc44d49adec48f80c524	8	8	
数据处理-记录	1	数据抽样	数据抽样.docx	宋波	2857	4	4	
	2	数据建模V2	数据建模v2-文档说明.docx	王韬	f9b337073d4c60838e0748388f2e	54	52	
	3	分组统计	分组统计.docx	高松	2806	18	20	
	4	数据表聚合或过滤	数据表聚合或过滤.docx	黄澍/王韬	2884	14	15	
	5	分组合并及统计	分组合并及统计算子.docx	何直洋	284025b29249ea9a1f3fa8ce288c	123	123	
	6	行列转换	行列转换.docx	李先霖	2828	5	6	
	7	变量计算	变量计算.docx	宋波 (bug修改)	2827	10	12	
	8	两表列合并	两表列合并.docx	高松	2816	31	32	
	9	三表列合并	三表列合并.docx	高松	2818	11	12	
	10	多表连接	多表连接.docx	李先霖	a74dfe52de44069d02865ffdb02c	30	30	
	11	交叉表	交叉表.docx	李先霖	2da81069b249179e23d637fb0fbe	14	14	
	12	平滑窗口表连接	平滑窗口表连接.docx	徐浩	2889	1	2	
	13	数据拆分	数据拆分.docx	冉超	2820	6	6	
	14	数据合并	数据合并.docx	高松	2810	4	17	
	15	数据字段插入	数据字段插入.docx	李先霖	45f46fb1c449118cbcbc8473b7f4	35	22	
	16	数据字段分裂V2	数据字段分裂.docx	黄澍/王韬	4003	16	18	
	17	数据重构	数据重构.docx	李先霖	2829	17	18	
	18	重新编码	重新编码.docx	高松	2826	4	4	
	19	列值映射V2	列值映射V2算子文档-代涛涛.docx	代涛涛	97470ea1574be4bf4f294c00ebe1	58	67	
	20	记录选择	记录选择.docx	史服宇	2370	44	48	
	21	记录排序	记录排序.docx	高松	2386	7	7	

算子库名称	序号	算子名称	说明文档	作者	ID	6088 版本号	6093 版本号	备注
数据处理-清洗	1	缺失值处理	缺失值处理.docx	陆访仪	2841	24	18	
	2	简单数据清洗	简单数据清洗.docx	彭火斌	635a8fe86b4344a0dcfdfb59045b	132	132	
数据处理-格式转换	1	DF转RDD	DF转RDD-代涛涛.docx	代涛涛	5d18bedeeb4d4090b06efe0339d0	13	13	
	2	RDD转DF	RDD转DF-代涛涛.docx	代涛涛	be2731a70e46da8a170b1238ece3	49	48	
	3	简单格式替换	简单格式替换.docx	马宇	2985	25	29	
	4	Gis位置转轨迹序列	Gis位置转轨迹序列.docx	朱航	ec67cf76004163bb9c0fe285e599	45	44	
	5	txt文件编码识别转换	txt文件编码识别转换算子-代涛涛.docx	代涛涛	5ea35e85534df39dd042bc384644	54	54	
数据处理-挖掘预处理	1	归一化预处理	归一化预处理算子.docx	马宇	8e1b8fe7404210968a33ab4f2c65	12	17	
	2	数值数据标准化	数值数据标准化.docx	徐浩	2893	3	4	
	3	误差值归并	误差值归并.docx	李先霖	348c4c6d9c450db61537dbc77918	1	10	
	4	数据集分离	数据集分离.docx	李先霖	b0b0a1670c45b08c4dc2dc9ff3ed	9	9	
	5	特征预处理	特征预处理.docx	徐浩	c8e8f43d2344fc9d76e92309194f	29	58	
数据处理-效果衡量	1	预测结果统计	预测结果统计.docx	李先霖	b8eedf80d54beeb98671a5ccce7	14	9	
	2	机器学习结果衡量	机器学习结果衡量.docx	李先霖	340baba5ac4f8caf48bceec4a29e5	10	12	
数据处理-时间	1	时间拆分	时间拆分.docx	马朋涛	2367	10	9	
	2	时间戳字符串转换	时间戳字符串转换.docx	李先霖	cc8c91867042a09b9d76cfb5eaf7	33	32	
	3	时间筛选	时间筛选.docx	徐浩	dd5e57831b4b5e8ee5f72c3eb409	8	10	
数据处理-分箱	1	数据分箱	数据分箱.docx	李先霖	2365	48	49	
	2	自然时间分箱	自然时间分箱.docx	徐浩	1de57cadab46f19961ec2a5bf34e	32	36	
	3	顺序时间分箱V2	顺序时间分箱v2.docx	徐浩	9b22cb70474ebd85005c9c83c87d	20	19	
	4	按时段分箱	按时段分箱.docx	徐浩	bcb6dc120b4913a0534319d9330a	9	10	
数据处理-图关系	1	关系数据转图形	关系数据转图形.docx	王韬	2972	1	1	
	2	图关系聚合	图关系聚合.docx	项博	2980	5	12	
	3	图基本信息统计	图基本信息统计.docx	项博	e359f30552477dab2092921a76a4	1	1	
	4	图属性关联	图属性关联.docx	项博	2982	2	4	
	5	图组装	图组装.docx	项博	5e2713fcb146d69cafb0bc9aec18	1	1	
	6	图度数计算	图度数计算.docx	项博	92b2-0425-44c5-866e-4d8f3f39	2	1	
	7	通联数据清洗	通联数据清洗.docx	李先霖	f709dfb42c4e9bae3af1eae4f786	17	17	
数据处理-其它	1	随机数	随机数.docx	宋波	2819	4	4	
	2	RowKey计算	RowKey计算.docx	李先霖	a03d8b41914d288c50c37b44ae4d	1	3	
数据挖掘-相关分析	1	均值分析	均值分析.docx	周洋生	2517	4	6	
	2	特征共现挖掘	特征共现挖掘.docx	项博	64239ddb314325b0a64abce110d0	17	17	
	3	频率分析	频率分析.docx	周洋生	2520	5	5	
	4	描述统计	描述统计算子.docx	潘咏霖	2575	1	1	
	5	相关性分析	相关性分析.docx	李先霖、金子添	2830	4	4	
	6	变量偏相关分析	变量偏相关分析.docx	李先霖、金子添	2831	4	4	
	7	典型相关分析	典型相关分析.docx	李先霖	0cc09754934f51be6ec0a8458977	3	5	
	8	MDS	待测算子-MDS-周洋生.docx	周洋生	99e8e806d448b887996dcfac4ad8	10	10	多维尺度分析
数据挖掘-分类算法	1	朴素贝叶斯训练	朴素贝叶斯-说明文档-周洋生.docx	周洋生	dc6ea54dbb4c2a8411a32aeb4ac3	14	4	(NB)
	2	朴素贝叶斯模型预测	朴素贝叶斯模型预测 说明文档.docx		0c0a65fb386a457c8683a8134beef2	20	20	
	3	KNN	KNN算子.docx	冉超	2536	7	7	(最邻近分类算法)
	4	二值逻辑回归训练	二值逻辑回归算子.docx	冉超	2533	4	5	
	5	多值逻辑回归训练	多值逻辑回归算子.docx	冉超	2535	5	6	
	6	决策树模型预测	决策树模型预测 说明文档.docx		a917e8b1c5564590851e61ea5a5f24	19	19	
	7	决策树模型训练	决策树模型训练算子.docx	徐浩	2566	16	15	

算子库名称	序号	算子名称	说明文档	作者	ID	6088 版本号	6093 版本号	备注
数据挖掘-分类算法	8	随机森林模型训练	随机森林模型训练算子.docx		2563	4	5	
	9	随机森林预测	随机森林模型预测算子.docx		2565	4	4	
	10	梯度森林模型训练	梯度森林模型训练算子.docx		2568	4	6	
	11	梯度森林模型预测	梯度森林模型预测 说明文档.docx		459b0a413c7a422eac554ba9a241a3	20	20	
	12	决策树分类模型预测	决策树分类模型预测.docx	徐浩	9dd58b60dc48d69e1e90823dba42	18	18	
	13	SVM模型预测	SVM模型预测算子.docx		2560	4	6	
数据挖掘-回归算法	14	SVM模型训练	SVM模型训练算子.docx		2558	5	5	
	1	标尺法线性回归预测	标尺法线性回归预测.docx	黄澍	2954	2	3	
	2	标尺法线性回归训练	标尺法线性回归训练.docx	黄澍	2953	2	8	
	3	线性回归模型预测	线性回归模型预测 说明文档.docx	石峰	dae4726bc84049943e2e65893edb	22	20	
	4	线性回归训练	线性回归训练算子.docx		2570	6	7	
	5	岭回归预测	岭回归算子.docx		2530	5	5	
	6	岭回归训练			2529	4	4	
	7	等渗回归模型训练	等渗回归模型训练算子.docx		2571	4	4	
	8	等渗回归模型预测	等渗回归模型预测算子.docx		2572	4	4	
	9	SVM模型训练	SVM模型训练算子.docx		2558	5	5	
	10	SVM模型预测	SVM模型预测算子.docx		2560	4	6	
	11	LASSO回归预测	Lasso回归算子.docx		2531	5	5	
数据挖掘-降维算法	12	LASSO回归训练			2532	4	4	
	1	主成分分析	主成分分析算子.docx	茆佳	3003	34	35	(PCA)
	2	LLE降维	LLE降维算子说明.docx	朱航	6ef4f329ec4b2a8c56654a63a44c	40	45	
	3	fastICA	fastICA.docx	徐浩	59ca31d34b4dd867196608bbdd0	3	3	
	4	拉普拉斯特征映射	拉普拉斯特征映射 说明文档.docx	彭火斌	041fe60b174e6d99707979fe6860	41	35	
数据挖掘-聚类算法	5	SVD奇异值分解	SVD奇异值分解.docx	金子添	2552	3	6	
	1	Kmeans	Kmeans.docx	周洋生	2508	24	24	
	2	K均值保留原数据列V2	K均值保留数据列算子V2-朱航.docx	朱航	5bcb72f8074cd6a371ecda78d1c0	50	67	
	3	Kmedoids	Kmedoids.docx	周洋生	00a5c4f1f048ebb1c2a295645998	7	7	
	4	Kprototypes训练	Kprototypes训练.docx	冉超	2633fedde74a02b31cc78b72d463	7	7	
	5	KModes训练	Kmodes训练.docx		ae3de34e3aad4197a344af587ff508	4	4	
	6	KModes预测	Kmodes预测.docx		9564efd58235486c9f01a1cd395f43	2	2	
	7	DBScan聚类算子	DBScan聚类算子.docx		7ff00debb9b941fb8addebdc2458c5	13	15	
	8	DBScan	DBScan.docx	周洋生	2513	8	8	
	9	C模糊均值聚类	C模糊均值聚类.docx	周洋生	2528	7	8	
	10	高斯矩阵	高斯矩阵.docx	周洋生	2523	16	17	
数据挖掘-检验	11	二分Kmeans	二分KMeans.docx	马朋涛	cbaeee0885460aa774bb3bcd0309	11	11	
	1	KruskalWallis检验	KruskalWallis检验.docx	周洋生	b0b928fb614cb3bab7f70902b06e	10	11	
	2	单样本F检验	单样本F检验算子说明.docx	朱航	063407c69b41cb80b27d7d4c9773	15	19	
	3	单样本Ks检验	单样本Ks检验.docx	马朋涛	41d9613ba24bc49114d7e9848e6f	22	23	
	4	单样本T检验	单样本T检验算子说明-朱航.docx	朱航	494fab59cb4dd281332b0330b2ee	25	27	
	5	二项式检验	二项式检验.docx	马朋涛	463cf5ab5b48b29c2b3a43b23c37	11	14	
	6	极端反应检验	极端反应检验-周洋生.docx	周洋生	c32bd1b96d400896f92ff6e9edd1	9	10	
	7	两独立样本检验	两独立样本检验-周洋生.docx	周洋生	3f5ebe4c224b43969a48a70f8a1d	35	45	
	8	两配对样本检验	两配对样本检验.docx	周洋生	1576c73af442a3b222852e8c5153	18	20	
	9	曼-惠特尼U检验	曼惠特尼U检验.docx	周洋生	c1442cc873429bbb1a1ccfa02874	9	10	
	10	配对双样本T检验	配对双样本T检验.docx	朱航	28a6f6c0df4182be54259d06532a	7	8	

算子库名称	序号	算子名称	说明文档	作者	ID	6088 版本号	6093 版本号	备注
数据挖掘-时间序列	11	双样本检验	双样本检验-周洋生.docx	周洋生	e316ff1e374eeba911f18676e49a	62	73	
	12	游程检验	游程检验.docx	马朋涛	b283e48100424f947b739400dc9d	11	14	
	1	ARIMA模型	ARIMA模型算子.docx	冉超	2722	8	8	
	2	X11季节调整模型	X11季节调整模型算子.docx	冉超	2726	3	3	
	3	时差相关分析	时差相关分析算子.docx	冉超	2551	5	5	
	4	误差修正模型	误差修正模型算子.docx	冉超	2545	10	10	
	5	协整检验	协整检验算子.docx	冉超	2544	7	7	
	6	移动平均法	移动平均法算子.docx	冉超	2539	7	7	
	7	指数平滑模型	指数平滑模型.docx	冉超	2540	9	9	
	8	向量自回归	向量自回归.docx	徐浩	43487d36d14a2cae88d6b2bb0621	24	29	
	9	单位根检验	单位根检验算子.docx	冉超	2542	3	4	
数据挖掘-分词	10	HMM训练	HMM_说明文档.docx	何直洋	8b1538e128493cb131026919c9fd	8	11	
	11	HMM预测		何直洋	e1ddbc9502403ea57e63f8de8edd	6	6	
	1	word分词	Word分词.docx	金子添	2554	53	55	
	2	复旦分词	复旦分词.docx	周洋生	2527	16	16	
数据挖掘-关联规则	3	NLPIR分词	NLPIR分词.docx	马朋涛	46c9fe7cf14869988a0658afe8bb	8	10	
	4	Hanlp分词及命名实体识别	Hanlp分词及命名实体识别.docx	何直洋	50f68208ff47caa358ffd6f64fef	113	116	
	1	FPGrowth频繁项集	FPGrowth频繁项集挖掘算子.docx	潘咏霖	2574	14	17	
	2	FPGrowth频繁项集关联规则	FPGrowth频繁项集规则挖掘算子.docx	潘咏霖	2573	6	10	
数据挖掘-连通性分析	3	Apriori频繁项集关联规则	Apriori频繁项集和关联规则_说明文档.docx	石峰	9fbf59da234a7c9c989bfbea849a	19	15	
	4	Prefixspan序列频繁项集	Prefixspan序列频繁项集.docx	王韬	eb54895f4540b39ca84c762af7c9	11	13	
	1	连通图分析	连通图分析.docx	项博	2988	2	2	
	2	网络节点分析	网络节点分析.docx	项博	2983	1	1	
异常检测-基于统计	3	子图抽取	目标子图抽取.docx	项博	2984	1	1	
	4	目标子图抽取	目标子图抽取.docx	项博	2975	2	2	
	1	置信区间法训练	置信区间法训练_改进1_说明文档.docx	黄澍	3024	8	20	
	2	置信区间法检测		黄澍	3025	6	8	
	3	iforest法训练	Iforest法训练与检测.docx	黄澍/金子添	2894	6	11	
异常检测-基于聚类	4	iforest检测		黄澍/金子添	3021	6	9	
	5	正态分布孤立点检测	正态分布孤立点检测.docx	马宇	77aedfd55d462b8db049ac4d15ab	11	11	
特征提取-频次特征	1	基于聚类的孤立点检测	基于聚类的孤立点检测.docx	马宇	690ecee0444d7d9af0c9d0a48084	27	35	
	2	基于密度的孤立点检测	基于密度的孤立点检测.docx	李先霖	2937a75ddf4efab935e0ec5852e2	7	8	
检索-ElasticSearch	1	单维频次特征提取	单维频次特征提取.docx	金子添	965a7070a144a1828c7727cbe6cf	3	5	
	2	多维频次特征提取	多维频次特征提取.docx	金子添	f655f9dd9a4f439d3e468916c3ce	4	6	
	3	区域频次特征提取	区域频次特征提取.docx	金子添	f6434d5b2e490db3d0c18304d254	1	14	
流处理-流数据源	1	ES插入	ES插入算子.docx	代涛涛	2994	6	15	
	5	DF插入ES	DF插入ES算子-代涛涛.docx	代涛涛	9500911e894f738615783e29bd54	34	46	
	7	ES列查询	ES列查询算子.docx	彭鹏	ff19e2b0f440d2809a4dd272a1a9	6	6	
	8	DF写入ES	DF写入ES算子-代涛涛.docx	代涛涛	b203f183b04722bdaa5650aba4ab	49	54	
	9	ElasticSearch全文搜索	ES全文搜索-朱航.docx	朱航	fd2f234ea9498784e645e46d07ec	20	35	
	10	ElasticSearch数据读入	ElasticSearch数据读入算子-代涛涛.docx	代涛涛	d0bc4c654f48949a1207dc3a207d	21	23	
流处理-流数据源	1	流Kafka源	Kafka流式数据源.docx	刘工	2763	1	1	
	2	流TCP源	TCP流式数据源.docx	刘工	2761	26	26	
	3	流Receiver源	Reciver流式数据源.docx	刘工	2767	5	5	

算子库名称	序号	算子名称	说明文档	作者	ID	6088 版本号	6093 版本号	备注
流处理-流框架	1	流变换	流变换.doc	刘工	2762	5	5	
	2	流处理框架节点	流基础组件.doc	刘工	2344	47	47	
	3	流窗口	流窗口.doc	刘工	2825	1	1	
	4	流合并	流合并.doc	刘工	2764	5	5	
	5	流启动	流基础组件.doc	刘工	2752	25	25	
	6	流上下文	流基础组件.doc	刘工	2347	26	26	
	7	流输出V2	流基础组件.doc	刘工	#N/A	#N/A	#N/A	
	8	流停止	流基础组件.doc	刘工	2686	12	12	
	9	流状态查看	流基础组件.doc	刘工	2863	6	6	
高级-轨迹数据挖掘-轨迹数据	1	Geo区域中心编码	Geo区域中心编码.docx	金子添	778b4bcfa14beb93003c6b52b3b3	9	11	
	2	多项式拟合	多项式拟合.docx	冉超	a71f4150574e2a9247b82aad56cf	6	6	
	3	Jin区域中心编码	Jin区域中心编码.docx	金子添	8662730e0b4eeba07ef65344da14	10	17	
	4	时空位置编码	时空位置编码.docx	冉超	2925	3	3	
	5	时空位置解码	时空位置解码.docx	冉超	2926	2	2	
高级-数据挖掘-聚类算法	1	谱聚类	谱聚类算子.docx	朱航	b2562b90bb4d54acafbf7cdea5b0	9	16	
	2	变色龙算法	变色龙算法.docx	李先霖	0ba93e04154d828cf3a27ff5f8de	5	5	
	3	OPTICS	OPTICS.docx	李先霖	d1f32960ff44b7a371d46968a3d8	2	3	
	4	Clara	clara.docx	彭火斌	2526	29	29	
	5	Clarans模型训练	Clarans-周洋生.docx	周洋生	38644593f842c0852fc2096598bc	12	12	
	6	Canopy优化预测	canopy优化-周洋生.docx	周洋生	bceb6a2e9b4470a4b549af21cd5a	4	4	
	7	KPrototypes预测	Kprototypes预测.docx	冉超	219c9cd6854793a4d04a9e775b5b	4	4	
	8	Canopy训练	Canopy.docx	彭火斌	210f834f5f4255a106270d8873e7	6	9	
	9	Canopy优化训练	canopy优化-周洋生.docx	周洋生	a7a025f0004af7b57841405645bd	9	9	
	10	Clarans模型预测	Clarans-周洋生.docx	周洋生	7164372ffb4d3986eac1cb6c3992	11	11	
	11	Canopy预测	Canopy.docx	彭火斌	51c6a4f8484109a127fc65de10fd	1	11	
	12	BIRCH	BIRCH.docx	彭火斌	1ef4708bea4364ba1d98d0cf9eb3	95	99	(利用层次方法的平衡迭代规约聚类)
	13	小波聚类	小波聚类 说明文档.docx	何直泽	21030c35a14ac0a6eacd8a97a19d	9	13	
高级-数据挖掘-分类算法	1	GDA	GDA.docx	徐浩	08a19d667146fd881c949ecb4b45	12	11	(高斯判别分析)
	2	probit回归	Probit回归.docx	徐浩	f1f774351d496389e048d54d82a5	13	15	
	3	多分类有序逻辑回归	多分类有序逻辑回归.docx	徐浩	dc0b67cd6541c3bea2ad03dcfae8	16	20	
高级-数据挖掘-回归算法	1	非线性最小二乘法	非线性最小二乘法.docx	李先霖	f9c348483942da91af05e38f4f76	6	6	
	2	Cox回归	COX回归.docx	李先霖	78effdca9b4e47acf170dd3e5b19	5	6	
高级-检索-ElasticSearch	1	HttpES插入	HttpES插入算子.docx	代涛涛	3001	11	14	
	2	HttpES滑动插入	HttpES滑动插入算子.docx	代涛涛	bc89aaf34546738ab4b211ab74c8	7	8	
	3	HttpES分区插入	HttpES分区插入算子-代涛涛.docx	代涛涛	de9995761a4ae5bd53207feaaafda	72	83	
	4	ESJava客户端插入	ESJavaClient写入算子-代涛涛.docx	代涛涛	9ecb42f10547f293a1e2e93e4e89	29	38	
	5	ES HTTP查询	ES HTTP查询算子.docx	彭鹏	95cf47d3a364e1f9ba9fe11cb6e64	65	65	
高级-文本处理-相似性	1	编辑距离	编辑距离算子.docx	莢佳	3002	33	35	
	2	TFIDF文本相似度	TFIDF文本相似度算子.docx	马朋涛	2da0df9a87478bb7300f43296ba5	13	21	
高级-文本处理-特征	1	关键词合并	关键词合并.docx	马朋涛	6bb90a38f94cc196f611cb84316c	13	13	
	2	关键词提取保留原表列	关键词提取保留原表列.docx	马朋涛	e75c60a67644228de41b2cfe2587	18	19	
	3	黑白名单提取	黑白名单提取.docx	马朋涛	d1b67d3a87429393fc7fcaae683d	25	24	
	4	文本自动摘要	文本自动摘要.docx	朱航	a02d20fbd0473fbff2146e63f6fb	17	17	
	5	关键词提取4RDD	关键词提取4RDD.docx	朱航	6fa0fb62d047e4bcf6212764d40e	24	24	

算子库名称	序号	算子名称	说明文档	作者	ID	6088 版本号	6093 版本号	备注
	6	元素连接检测通用	元素连接检测通用算子.docx	何直泽	00ab90a5b94da18c92fdf11c7e55	54	54	
	7	Word2Vec	Word2Vec算子.docx	何直泽	b716eba4742a3b3ecec77dd8110	15	15	
	8	TFIDF词频统计	TFIDF词频统计算子.docx	马朋涛	2f2285acba433dbe08a58525f5fa	9	14	
	9	词频统计	词频统计.docx	马朋涛	e8ef7b73f346ffbda4cf96570224	31	33	
高级-异常检测-基于距离	1	基于距离的孤立点检测	基于距离的孤立点检测算子_说明文档.docx	何直泽	2274ee7f1449bfb7517ad501fcb1	13	14	
高级-异常检测-基于偏离	1	序列异常技术	序列异常技术.docx	何直泽	a52f4fd88a4043b0df86e118cdd2	17	14	
	2	构造数据立方体	构造数据立方体算子.docx	何直泽	b94208c34240f68f45f7551998d0	7	7	
	3	OLAP孤立点检测	OLAP孤立点检测.docx	何直泽	6f1d5ff46c4ff99ed0a25aec66d8	15	18	
高级-流处理-流数据源	1	ICE算子探查	ICE算子探查.docx		2797	1	1	
	3	流ICE源	ICE流式数据源.docx	刘工	2794	2	3	
	4	流MQ源			5cae8f1407794e8b8a82e0292ed0a3	10	10	



算子库名称	序号	算子名称	说明文档	作者
数据源-领域	1	按目录读取二进制文件	54S按目录读取二进制文件.docx	宋波
	2	DBF文件解析	DBF文件解析.docx	马朋涛
	3	email解析	email解析.docx	马朋涛
	4	VISIO解析ASPOSE	VISIO解析ASPOSE.docx	马朋涛
	5	Neo4J读取	Neo4J读取.docx	朱航
	6	Neo4J写入	Neo4J写入算子说明.docx	朱航
	7	WindowsPE文件解析	WindowsPE文件解析.docx	代涛涛
	8	读取word_pdf文件	读取word_pdf文件.docx	朱航
	9	sql单文件解析	sql单文件解析.docx	代涛涛
	10	INI配置文件解析	INI配置文件解析.docx	马朋涛
	11	LDAP数据解析	ldap数据解析.docx	代涛涛
数据源-领域	1	关系型数据库规则读取	关系型数据库规则读取.docx	王韬
	2	日志规则数据源	日志规则数据源.docx	英佳
	3	特殊excel文件读取	特殊excel文件读取.docx	宋波
数据处理-记录-领域	1	分组合并	分组合并.docx	马朋涛
	2	表堆叠	表堆叠.docx	徐浩
数据处理-其他-领域	1	Byte数组提取内容	Byte数组提取内容.docx	朱航
异常检测-基于聚类-领域	1	文本异常检测	文本异常检测.docx	金子添
社交网络挖掘-社区发现-领域	1	Louvain社区聚类	Louvain社区聚类.docx	金子添
数据挖掘-轨迹数据-领域	1	伴随度计算	伴随度计算.docx	冉超
	2	轨迹规律航线	轨迹规律航线.docx	金子添/王韬
	3	轨迹规律区域	轨迹规律区域.docx	金子添/王韬
	4	轨迹区域转移统计	轨迹区域转移统计.docx	金子添/王韬
	5	轨迹停留点查找	轨迹停留点查找.docx	金子添/王韬
	6	轨迹异常点过滤	轨迹异常点过滤.docx	金子添/王韬
	7	区域选择	区域选择（说明文档）.docx	陈强伟
	8	查找轨迹区域活动ID号	查找轨迹区域活动ID号.docx	王韬
	9	区域圆分析	区域圆分析.docx	马朋涛
	10	航迹越界检测	航迹越界检测（说明文档）.docx	刘强
	11	盘旋分析_加方向	盘旋分析_加方向.docx	刘强
数据挖掘-深度学习-领域	1	卷积神经网络数据源	卷积神经网络数据源算子.docx	潘咏霖
	2	卷积神经网络训练	卷积神经网络训练算子.docx	潘咏霖
	3	卷积神经网络预测	卷积神经网络预测算子.docx	潘咏霖
	4	人工神经网络分类模型训练	人工神经网络分类算子.docx	潘咏霖
	5	人工神经网络分类模型预测		潘咏霖
	6	人工神经网络回归模型训练	人工神经网络回归算子.docx	潘咏霖
	7	人工神经网络回归模型预测		潘咏霖
	8	深度置信网络编解码变换	深度置信网络编解码算子.docx	潘咏霖
	9	深度置信网络编解码模型训练		潘咏霖
	10	深度置信网络降维模型训练	深度置信网络降维变换算子.docx	潘咏霖
	11	深度置信网络降维变化		潘咏霖
	12	循环神经网络分类模型训练	循环网络分类功能算子.docx	潘咏霖
	13	循环神经网络分类模型预测		潘咏霖
	14	循环神经网络回归模型训练	循环神经网络回归算子.docx	潘咏霖
	15	循环神经网络回归预测		潘咏霖
	16	自组织映射神经网络	自组织映射神经网络.docx	徐浩
数据挖掘-推荐-领域	1	ALS推荐	ALS.docx	史服宇