The DataMiner Manager Web Interface

Giancarlo Panichi, Gianpaolo Coro Istituto di Scienza e Tecnologie dell'Informazione A. Faedo {panichi,coro}@isti.cnr.it

In this document we describe the DataMiner Manager Web interface that allows interacting with the gCube DataMiner service.

Overview

DataMiner is a cross-usage service that provides users and services with tools for performing data mining operations. Specifically, it offers a unique access to perform data mining and statistical operations on heterogeneous data, which may reside either at client side, in the form of comma-separated values files, or be remotely hosted, possibly in a database. The DataMiner service is able to take inputs and execute the operation requested by a client or a user, by invoking the most suited computational facility from a set of available computational resources. Executions can run either on multi-core machines or on different computational platforms, such as D4Science and other different private and commercial Cloud providers.

DataMiner



Figure 1: DataMiner Manager portlet. Main interface.

DataSpace

The DataSpace section of the Web interface shows the overall inputs and outputs data of the computations. It is the main entry point to upload new inputs for the computations.

D	ataMiner		🗶 go ba	ack	Access to Data Spa	o the 💦	Execute an Experiment	Check the Computations	
Er,	Output Data Sets 🛛 🔄 In	put Data	Sets						
	🖁 Download 🥥 Delete	😂 Refre	ish						
	Name	•	Created	computation_id	data_description	data_type	operator_name	VRE	Â
	OccurrencePointsT c858-4b21-9918- 965b53f2655b].csv	11 Jul 2016	11:09 AM	DBSCAN_ID_e363 c858-4b21-9918- 965b53f2655b	OccurrencePointsT	i text/csv	DBSCAN	/gcube/preprod/pre	1
	hcaf_filtered_1.csv 09e3-4977-82e3- 16c1da3621c6].csv	08 Jul	04:08 PM	BIOCLIMATE_HCA 09e3-4977-82e3- 16c1da3621c6	HCAF_Table_List	text/csv	BIOCLIMATE_HC	A /gcube/preprod/pre	J
	hcaf_filtered_2.csv 2253-4061-a2d6- a557857f6806.csv_ 7340-4c98-83c7- 288fec73550e].csv	23 Jur 2016	n 03:56 PM	BIOCLIMATE_HCA 7340-4c98-83c7- 288fec73550e	HCAF_Table_List	text/csv	BIOCLIMATE_HC	A /gcube/devsec/dev/ 🚺	1
	hcaf_filtered_1.csv 2253-4061-a2d6- a557857f6806.csv 7340-4c98-83c7- 288fec73550e].csv	23 Jur 2016	1 03:56 PM	BIOCLIMATE_HCA 7340-4c98-83c7- 288fec73550e	HCAF_Table_List	text/csv	BIOCLIMATE_HC	A /gcube/devsec/dev/ 🚺	J
	OccurrencePointsT 170e-4089-8942- a09e80484147].csv	23 Ju 2016	1 03:56 PM	DBSCAN_ID_5755 170e-4089-8942- a09e80484147	OccurrencePointsT	text/csv	DBSCAN	/gcube/devsec/dev	J

Figure 2: DataMiner Manager portlet. DataSpace panel.

New files can be added by selecting the Input Data Sets tab and using drag and drop from Desktop computer. The files will be automatically saved on the cloud storage named D4Science e-Infrastructure Workspace.

D	ata M iner	🗶 go b	ack	Access to Data Spa	o the ace	Execute an Experiment	Check the Computations
	Output Data Sets	Input Data Sets					
	Download 🥥 Delete	👶 Refresh					
	Name	 Created 			data_type		VRE
	OccurrencePoint 93cf-4134- 82ae- 2b300780300c].c	11 Jul 04:27 PM 2016	DBSCAN_ID_109 93cf-4134-82ae- 2b300780300c	OccurrencePoints t	ext/csv	EUZE: EUZE: EAIE: AN acultureProduc	/gcube/devNext/N
	OccurrencePoint c858-4b21- 9918- 965b53f2655b].c:	11 Jul 11:09 AM 2016	D-620p 9918- 965b53f2655b	files	here	DBSCAN	
	hcaf_filtered_1.cs 09e3-4977- 82e3- 16c1da3621c6].c	08 Jul 04:08 PM 2016	BIOCLIMATE_HC 09e3-4977- 82e3- 16c1da3621c6	HCAF_Tage_List	ext/csv	BIOCLIMATE_H	
	hcaf_filtered_2.cs 2253-4061- a2d6- a557857f6806.cs 7340-4c98- 83c7- 288fec73550e].cs	23 Jun 03:56 PM 2016	BIOCLIMATE_HC 7340-4c98- 83c7- 288fec73550e	HCAF_Table_List t	ext/csv	BIOCLIMATE_H	

Figure 3: DataMiner Manager portlet. Input data importing.

Execute an Experiment

In the *Execute an Experiments* section, a list of algorithms grouped by category can be found. By clicking on one of the algorithms categories on the left hand side, a list of algorithms belonging to the category appears. By clicking on the arrow next to the algorithm description, the parameters of the selected algorithm are displayed. The parameters must be filled and the *Start computation* button should be pressed in order to start the computation.

DataMiner	🕊 go back		6	Access to the Data Space	8	Execute an Experiment	Check the Computations	
Operators	•	C Operator	Computations Exe	scution				
🗟 🛃 📉 🕹 User Perspec	ctive +	Tools: 🙁 Re	emove All Operators					
BAYESIAN METHODS (4)	<u> </u>				•		Í.	
CHARTS (4)	<u> </u>	Dbscan						
CLASSIFICATION (1)	•	•						
CLIMATE (3)	<u> </u>	A clustering algorithm for real valued vectors that relies on the density-based spatial clustering of applications with noise (DBSCAN) algorithm. A maximum of 4000 points is allowed.						
CORRELATION ANALYSIS (1)	•	Parameters						
DATA CLUSTERING (1)	Ŧ	Occurrence	PointsTable:	🔄 Select Data Set	Occurrence Po following this to encoding]	oints Table. Max 4000 po emplate: (GENERIC) A (ints [a http link to a table in UTF-8 encoding generic comma separated csv file in UTF-8	
Dbscan A clustering algorithm for real valued vectors that relies the density-based spatial clustering of applications with noise (DBSCAN) algorithm. A maximum o	s on h	FeaturesCol	umnNames:			column Names columns from C	for the features [a sequence of names of occurrencePointsTable separated by]	
DATABASES (9)	▶							
EGIP (4)	+							
FILTERING (1)	-	0	Deinte Chusteril alta-la	Select table from param	eter OccurrencePo	sintsTable name of the resulting disl	tribution	
Gro Processing (10)	_	Occurrences	Points cluster cabel.	String Value	1000	tane of the resulting dis		
CEO PROCESSING (10)		epsilon:		10	DBSca	an epsilon parameter		
Maps (5)	<u> </u>			Integer Value				
OBIS OBSERVATIONS SPECIES DATA (3)	•	min_points:		1	DBSca	an minimum points para	meter (identifies outliers)	
OBIS OBSERVATIONS TRENDS (4)	•			Integer Value				
OCCURRENCES (10)	<u>▶</u>	Start C	Computation					

Figure 4: DataMiner Manager portlet. Executing an experiment.

By clicking on the Computations Execution tab the computations results are accessed.

DataMiner ((go back	Access to the Data Space Recute an Check the Computations
Operators		Coperator Computations Execution
😨 👌 📉 🕹 User Perspec	tive +	Tools: 😣 Remove All
BAYESIAN METHODS (4)	Ĵ	
CHARTS (4)	•	Computation of Dbscan Created the id is 1090/df3-93cf-4134-82ae-2b300780300c flink]
CLASSIFICATION (1)	•	Equivalent Get Request: Show
CLIMATE (3)	•	
CORRELATION ANALYSIS (1)	•	Computation Complete
DATA CLUSTERING (1)	Ŧ	The computation Dbscan finished.
Dbscan		The algorithm produced Multiple Results.
A clustering algorithm for real valued vectors that relies the density-based spatial clustering of applications with	on	Filee
noise (DBSCAN) algorithm. A maximum o		Cutput cluster table [a http link to a table in UTF-8 ecoding following this template: (CLUSTER) http://goo.gl/PnKhhb]
DATABASES (9)	•	Download File
EGIP (4)	•	
FILTERING (1)	+	
GEO PROCESSING (10)	•	
MAPS (5)	•	
OBIS OBSERVATIONS SPECIES DATA (3)	•	
OBIS OBSERVATIONS TRENDS (4)	•	
OCCURRENCES (10)	<u>+</u>	

Figure 5: DataMiner Manager portlet. View of a computation result.

By clicking on the Show button, the "Equivalent Http Get Request" is retrieved. This Http link can be pasted in a Web browser or used by a program client to execute the same experiment and retrieve the output in Web Processing Service format. The Http link is itself a Web Processing Service compliant call.



Figure 6: DataMiner Manager portlet. Equivalent Http Get Request.

Check the Computations

Once an algorithm has been executed, a user can disconnect from the portal and check for the computation completion after a while. In order to check the status of a computation, the "Check the computations" button should be pressed. The list of computations along with their status appears.

Da ta M iner		«	go back		Access to the Data Space	Exe Exp	cute an eriment	Check the Computations	
Eq.	List of Computations								
0	Show 🔓 Download	I 🥥 Delete 🛭 🗇 Re	fresh 🗦 Resubmit	Cancel					
	Name	 Created 	operator_name	start_date	end_date	status	execution_platfo	VRE	Ê
-	DBSCAN_ID_e3 c858-4b21- 9918- 965b53f2655b	11 Jul 11:09 AM 2016	DBSCAN	11/07/2016 11:09:17	11/07/2016 11:09:20	completed	LOCAL	/gcube/preprod/p 🚺)
	BIOCLIMATE_H 09e3-4977- 82e3- 16c1da3621c6	08 Jul 04:08 PM 2016	BIOCLIMATE_H	08/07/2016 16:08:38	08/07/2016 16:08:43	completed	LOCAL	/gcube/preprod/p 🚺	1
	TRAJECTORY_ f8bb-4b86- 9450- c7a2cf05bab0	07 Jul 01:04 PM 2016	TRAJECTORY_I	07/07/2016 13:04:17	07/07/2016 13:04:24	completed	LOCAL	/gcube/devsec/d)
8 ^{3E}	TRAJECTORY_ 3393-49a1- 83b3- b7bb478b5080	07 Jul 12:48 PM 2016	TRAJECTORY_I	07/07/2016 12:48:54	-	error	LOCAL	/gcube/devsec/d	1
836	TRAJECTORY_ cd2e-44d7- 8b67- 0f2029764010	07 Jul 12:31 PM 2016	TRAJECTORY_I	07/07/2016 12:31:07	-	error	LOCAL	/gcube/devsec/d	1

Figure 7: DataMiner Manager portlet. "Check the Computations" panel.

By clicking on a computation and then on the Show button, the complete provenance information is retrieved.

DataMiner	g o back	Access to the Data Space	Execute an Experiment	Check the Computations
🔄 List of Computations 📔	C DBSCAN_ID_e363e31c-c858-4b21-9918-965b	53f2655b 🗵		
	Computation Report of E	DBSCAN_ID_e363e31c-c858-4b21	-9918-965b53f2655b	Î
Output Result				
test1:	test1_[DBSCAN_ID_e363e31c	-c858-4b21-9918-965b53f2655b].csv		
Input Parameters				
epsilon:	10			
OccurrencePointsClust	erLabel: test1			
FeaturesColumnNames	s: csquarecode			
min_points:	1			
OccurrencePointsTable	CccurrencePointsTable_[DBSC	CAN_ID_e363e31c-c858-4b21-9918-965t	b53f2655b].csv	
Computation Details				
Start Date:	11/07/2016 11:09:17			
End Date:	11/07/2016 11:09:20			
Status:	completed			
VRE:	/gcube/preprod/preVRE			
Onerator Details				*

Figure 8: DataMiner Manager portlet. Computation provenance view.

By double-clicking on a computation folder, the set of inputs and outputs of the computation can be retrieved, along with an XML description of the Provenance information, following the PROV-O standard.