Package 'bigml'

May 20, 2015

Type Package

Title Bindings for the BigML API

Version 0.1.2

Date 2015-05-08

Description The 'bigml' package contains bindings for the BigML API. The package includes methods that provide straightforward access to basic API functionality, as well as methods that accommodate idiomatic R data types and concepts.

License LGPL-3

URL https://github.com/bigmlcom/bigml-r

BugReports https://github.com/bigmlcom/bigml-r/issues

Imports RJSONIO, RCurl, plyr

Collate 'bigml-internal.R' 'formEncodeURL.R' 'bigml-package.R' 'createDataset.R' 'createModel.R' 'createPrediction.R' 'createSource.R' 'getDataset.R' 'getModel.R' 'getPrediction.R' 'getSource.R' 'listDatasets.R' 'listModels.R' 'listSources.R' 'quickDataset.R' 'quickModel.R' 'quickPrediction.R' 'quickSource.R' 'setCredentials.R' 'deleteResource.R'

NeedsCompilation no

Author Leon Hwang [cre, aut]

Maintainer Leon Hwang <hwang@bigml.com>

Repository CRAN

Date/Publication 2015-05-20 01:18:53

R topics documented:

`bigml-package` *R bindings for BigML API*

Description

Package:	bigml
Type:	Package
Version:	0.1-1
Date:	20012-04-30
License:	GPL (>= 2)
LazyLoad:	yes

Details

A set of methods that enable straightforward usage of the BigML API. The methods use R idioms and native datatypes where appropriate, while also providing access to more conventional API usage.

Author(s)

Leon Hwang <hwang@bigml.com>

Examples

```
## Not run:
# set default credentials
```

```
setCredentials('username', 'key')
model = quickModel(iris, 'Species')
quickPrediction(model, c(Petal.Width=0.2, Petal.Length=1.4))

# use specific credentials
quickPrediction(model, c(Petal.Width=0.2, Petal.Length=1.4),
    username='someuser', api_key='somekey')

# list most recent sources
 listSources()

# specify limit and offset
 listModels(limit=15,offset=300)

# specify filter criteria
 listDatasets(size__gt=1048576)

## End(Not run)
```

`createDataset` *Creating BigML Datasets*

Description

Creating BigML Datasets

Usage

```
createDataset(source_id, field_ids = NULL, name = NULL, size = NULL, ...)
```

Arguments

`source_id`	The relevant source id used to build the dataset.
`field_ids`	A list of field ids and field properties. See example.
`name`	The name for the dataset.
`size`	The amount (in bytes) of the source to use for creating the dataset.
`...`	Arbitrary named arguments that are passed on to `formEncodeURL` in order to create form-encoded URL options.

Details

This function needs to use id information from existing R resources. See the references for more details.

Value

category	numeric
code	numeric
content_type	character
created	character
credits	numeric
description	character
fields	data.frame (or list if flatten=FALSE)
file_name	character
md5	character
name	character
number_of_datasets	numeric
number_of_models	numeric
number_of_predictions	numeric
private	logical
resource	character
size	numeric
source_parser	list
status	list
tags	AsIs
type	numeric
updated	character

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/datasets

See Also

Other dataset methods: getDataset; listDatasets; quickDataset

Examples

```
## Not run:
# simple create dataset example
createDataset("source/1")
# configure a number of different parameters
createDataset("source/2", field_ids=c('000001'), name='test', size=10)

## End(Not run)
```

`createModel` *Creating BigML Models*

Description

Creating BigML Models

Usage

```
createModel(dataset_id, input_field_ids = NULL, name = NULL,
  objective_field_ids = NULL, range = NULL, ...)
```

Arguments

`dataset_id`	the relevant dataset_id used to create the model.
`input_field_ids`	a vector of field ids to use for training.
`name`	the name to give to the model.
`objective_field_ids`	a vector of objective fields used for training.
`range`	a vector of two values that define a range of instances from the dataset to train on.
`...`	Arbitrary named arguments that are passed on to `formEncodeURL` in order to create form-encoded URL options.

Details

This function needs to use id information from existing R resources. See the references for more details.

Value

model_return

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/models

https://bigml.com/developers/datasets

See Also

Other model methods: getModel; listModels; quickModel

Examples

```
## Not run:
# simple example
m1 = createModel("dataset/1")
# configure a number of different parameters
m2 = createModel("dataset/2", input_field_ids=c('000001'),
objective_field_ids='000003', name='test', range = c(10,1000))

## End(Not run)
```

---

createPrediction *Creating BigML Predictions*

Description

Creating BigML Predictions

Usage

```
createPrediction(model_id, input_field_ids, name = NULL,
  prediction_only = TRUE, ...)
```

Arguments

model_id	character string; the model id
input_field_ids	a list of input field ids and values to make a prediction for (see example).
name	character string; The given name for the prediction.
prediction_only	logical: Indicating whether the prediction should be returned as a simple value, or if the full response object should be returned.
...	Arbitrary named arguments that are passed on to formEncodeURL in order to create form-encoded URL options.

Details

This function needs to use id information from existing R resources. See the references for more details.

Value

atomic character or numeric value if prediction_only is TRUE, else return:

category	numeric
code	numeric
created	character
credits	numeric
dataset	character
dataset_status	logical
description	character
fields	list
input_data	numeric
locale	character
model	character
model_status	logical
name	character
objective_fields	character
prediction	character
prediction_path	list
private	logical
resource	character
source	character
source_status	logical
status	list
tags	AsIs
updated	character

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/predictions

See Also

Other prediction methods: getPrediction; quickPrediction

Examples

```
## Not run:
# simple example
m1 = createPrediction("model/1",
input_field_ids = c('000001'='somevalue', '000002'=9999))
# configure a number of different parameters
m2 = createPrediction("model/2",
input_field_ids = c('000001'='somevalue', '000002'=9999),
name='new prediction')

## End(Not run)
```

`createSource` *Creating BigML Sources*

Description

Creating BigML Sources

Usage

```
createSource(file_name, name = basename(file_name), header = TRUE,
  locale = "en-US", missing_tokens = c("NA"),
  quote = "\"", separator = ",", trim = TRUE, flatten = TRUE, ...)
```

Arguments

`file_name`	A string giving a file location
`name`	A string specifying the name of the source
`header`	logical; TRUE if data contains name information, false otherwise.
`locale`	A string giving the locale (defaults to en-US).
`missing_tokens`	A vector of character strings that will be used to specify missing values in a file name.
`quote`	A string specifying the quoting character used.
`separator`	the separator character used when a file name is specified.
`trim`	A logical value indicating whether white space should be trimmed.
`flatten`	A logical value indicating whether or not the returned field objects should be "flattened" into a data frame.
`...`	Arbitrary named arguments that are passed on to `formEncodeURL` in order to create form-encoded URL options.

Details

createSource

Value

category	numeric
code	numeric
content_type	character
created	character
credits	numeric
description	character
fields	data.frame (or list if flatten=FALSE)
file_name	character
md5	character
name	character
number_of_datasets	numeric
number_of_models	numeric
number_of_predictions	numeric
private	logical
resource	character
size	numeric
source_parser	list
status	list
tags	AsIs
type	numeric
updated	character

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/sources

See Also

Other source methods: getSource; listSources; quickSource

Examples

```
## Not run:
# simple example
m1 = createSource("/tmp/iris.csv")


## End(Not run)
```

`deleteResource` *Deleting BigML Resources*

Description

Deleting BigML Resources

Usage

```
deleteResource(resource_id, ...)
```

Arguments

`resource_id`	the resource to delete.
`...`	Arbitrary named arguments that are passed on to `formEncodeURL` in order to create form-encoded URL options.

Details

This function deletes bigml resources referenced by their resource id.

Value

TRUE if successful, FALSE otherwise.

Author(s)

Leon Hwang <hwang@bigml.com>

Examples

```
## Not run:
# replace with your valid credentials:
deleteResource("source/1")

## End(Not run)
```

formEncodeURL *A simple function to turn named arguments into a form-encoded string*

Description

A simple function to turn named arguments into a form-encoded string

Usage

```
formEncodeURL(a, ...)
```

Arguments

a	something
...	arbitrary named arguments that will become part of a form-encoded url.

Details

This function is called in every BigML API function. It helps build the URL that requests are forwarded to. It automatically adds any default user and api key settings specified by setCredentials. However, it also can be used to access advanced options that are otherwise undocumented here. For instance, it's possible to filter and/or sort on a number of different api requests, using a number of different fields (e.g., see the documentation on listing and sorting datasets.) Other usage includes specifying username and api_key for individual API requests; or limit or offset parameters useful for paging through list requests. Finally, it's possible to enable a simple debug mode by passing debug=TRUE. This will print the url request string to the screen, along with any posted json objects.

Value

form-encoded string result

Author(s)

Leon Hwang <hwang@bigml.com>

Examples

```
## Not run:
formEncodeURL(username="user1", api_key="test", limit=100, debug=TRUE)
# "?username=user1&api_key=test&limit=100&debug=TRUE"

## End(Not run)
```

`getDataset` *Retrieving a BigML Dataset*

Description

Retrieving a BigML Dataset

Usage

```
getDataset(source_id, include_overview = TRUE, ...)
```

Arguments

`source_id`	A string giving the name of the source id.
`include_overview`	A logical value indicating whether to provide a simple data frame overview of fields.
`...`	Arbitrary named arguments that are passed on to `formEncodeURL` in order to create form-encoded URL options.

Details

This function needs to use id information from existing R resources. See the references for more details.

Value

`category`	numeric
`code`	numeric
`content_type`	character
`created`	character
`credits`	numeric
`description`	character
`fields`	data.frame (or list if flatten=FALSE)
`file_name`	character
`md5`	character
`name`	character
`number_of_datasets`	numeric
`number_of_models`	numeric
`number_of_predictions`	numeric
`private`	logical

resource	character
size	numeric
source_parser	list
status	list
tags	AsIs
type	numeric
updated	character

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/datasets

See Also

Other dataset methods: createDataset; listDatasets; quickDataset

---

getModel *Retrieving a BigML Model*

Description

Retrieving a BigML Model

Usage

```
getModel(model_id, ...)
```

Arguments

model_id	A string giving the model id.
...	Arbitrary named arguments that are passed on to formEncodeURL in order to create form-encoded URL options.

Details

This function needs to use id information from existing R resources. See the references for more details.

Value

category	numeric
code	numeric
columns	numeric
created	character
credits	numeric
dataset	character
dataset_status	logical
description	character
input_fields	character
locale	character
max_columns	numeric
max_rows	numeric
model	list
name	character
number_of_predictions	numeric
objective_fields	character
private	logical
range	numeric
resource	character
rows	numeric
size	numeric
source	character
source_status	logical
status	list
tags	AsIs
updated	character

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/models

See Also

Other model methods: createModel; listModels; quickModel

getPrediction *Retrieving a BigML Prediction*

Description

Retrieving a BigML Prediction

Usage

```
getPrediction(prediction_id, ...)
```

Arguments

prediction_id	the id of the prediction resource.
...	Arbitrary named arguments that are passed on to formEncodeURL in order to create form-encoded URL options.

Details

This function needs to use id information from existing R resources. See the references for more details.

Value

atomic character or numeric value if prediction_only is TRUE, else return:

category	numeric
code	numeric
created	character
credits	numeric
dataset	character
dataset_status	logical
description	character
fields	list
input_data	numeric
locale	character
model	character
model_status	logical
name	character
objective_fields	character
prediction	character

prediction_path	list
private	logical
resource	character
source	character
source_status	logical
status	list
tags	AsIs
updated	character

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/predictions

See Also

Other prediction methods: createPrediction; quickPrediction

---

getSource *Retrieving a BigML Source*

Description

Retrieving a BigML Source

Usage

```
getSource(source_id, flatten = TRUE)
```

Arguments

source_id	A character value giving the name of the source.
flatten	A logical value indicating whether to flatten the response into a data frame.
...	Arbitrary named arguments that are passed on to formEncodeURL in order to create form-encoded URL options.

Details

This function needs to use id information from existing R resources. See the references for more details.

Value

category	numeric
code	numeric
content_type	character
created	character
credits	numeric
description	character
fields	data.frame (or list if flatten=FALSE)
file_name	character
md5	character
name	character
number_of_datasets	numeric
number_of_models	numeric
number_of_predictions	numeric
private	logical
resource	character
size	numeric
source_parser	list
status	list
tags	AsIs
type	numeric
updated	character

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/sources

See Also

Other source methods: createSource; listSources; quickSource

`listDatasets` *Listing BigML Datasets*

Description

Listing BigML Datasets

Usage

```
listDatasets(flatten = TRUE, datasets_only = TRUE, ...)
```

Arguments

`flatten`	A logical value indicating whether to flatten the response into a dataframe.
`datasets_only`	A logical value indicating whether to only return the data frame of field information (only valid if `flatten` is `TRUE`).
`...`	Arbitrary named arguments that are passed on to `formEncodeURL` in order to create form-encoded URL options.

Value

If flatten is TRUE, and datasets_only = TRUE a data frame of:

`category`	numeric
`code`	numeric
`columns`	numeric
`created`	character
`credits`	numeric
`description`	character
`locale`	character
`name`	character
`number_of_models`	numeric
`number_of_predictions`	numeric
`private`	logical
`resource`	character
`rows`	numeric
`size`	numeric
`source`	character
`source_status`	logical
`status.bytes`	numeric

status.code	numeric
status.elapsed	numeric
status.message	character
status.serialized_rows	numeric
updated	character

If flatten is TRUE and datasets_only = FALSE a list of:

meta	list
datasets	data.frame
fields	data.frame

If flatten is FALSE a list of:

meta	list
objects	list

see references for more details

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/datasets

See Also

Other dataset methods: createDataset; getDataset; quickDataset

---

listModels *Listing BigML Datasets*

Description

Listing BigML Datasets

Usage

```
listModels(flatten = TRUE, models_only = TRUE, ...)
```

Arguments

`flatten`	A logical value indicating whether to flatten the response into a data frame.
`models_only`	A logical value indicating whether to only return the data frame of model information (only valid if `flatten` is `TRUE`).
`...`	Arbitrary named arguments that are passed on to `formEncodeURL` in order to create form-encoded URL options.

Value

If flatten is TRUE, and models_only = TRUE a data frame of:

`category`	numeric
`code`	numeric
`columns`	numeric
`created`	character
`credits`	numeric
`dataset`	character
`dataset_status`	logical
`description`	character
`locale`	character
`max_columns`	numeric
`max_rows`	numeric
`name`	character
`number_of_predictions`	numeric
`objective_fields`	character
`private`	logical
`resource`	character
`rows`	numeric
`size`	numeric
`source`	character
`source_status`	logical
`updated`	character

If flatten is TRUE and models_only = FALSE a list of:

`meta`	list
`models`	data.frame

If flatten is FALSE a list of:

`meta`	list
`objects`	list

see references for more details

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/models

See Also

Other model methods: createModel; getModel; quickModel

---

listSources *Listing BigML Sources*

Description

Listing BigML Sources

Usage

```
listSources(flatten = TRUE, sources_only = TRUE, ...)
```

Arguments

flatten	A logical value indicating whether to flatten the response into a dataframe.
sources_only	A logical value indicating whether to only return the data frame of source information (only valid if flatten is TRUE).
...	Arbitrary named arguments that are passed on to formEncodeURL in order to create form-encoded URL options.

Value

If flatten is TRUE, and sources_only = TRUE a data frame of:

category	numeric
code	numeric
content_type	factor
created	factor
credits	numeric
description	factor
file_name	factor
md5	factor
name	factor
number_of_datasets	numeric

number_of_models
: numeric

number_of_predictions
: numeric

private
: logical

resource
: factor

size
: numeric

source_parser.header
: logical

source_parser.locale
: factor

source_parser.missing_tokens
: factor

source_parser.quote
: factor

source_parser.separator
: factor

source_parser.trim
: logical

status.code
: numeric

status.elapsed
: numeric

status.message
: factor

type
: numeric

updated
: factor

If flatten is TRUE and sources_only = FALSE a list of:

meta
: list

sources
: data.frame

fields
: data.frame

If flatten is FALSE a list of:

meta
: list

objects
: list

see references for more details

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/sources

See Also

Other source methods: createSource; getSource; quickSource

quickDataset *Quickly Creating BigML Datasets*

Description

Quickly Creating BigML Datasets

Usage

```
quickDataset(data, fields = names(data),
  name = paste(deparse(substitute(data)), "'s dataset", sep = ""),
  size = NULL, ...)
```

Arguments

data	A matrix or data frame containing data to upload to bigml.
fields	A vector of names in data that should be used for creating the dataset.
name	A string giving the name for the dataset.
size	A numeric value giving the amount (in bytes) of the source to use.
...	Arbitrary named arguments that are passed on to formEncodeURL in order to create form-encoded URL options.

Details

quickDataset will take its "data" dataframe argument and attempt to create an equivalent BigML dataset using quickSource. R "numeric" class fields will become "numeric" fields in the BigML dataset. R "character" class fields become "text" fields. R "factor" fields become "categorical" fields. However, if there are too many factors, BigML may convert the field to text. It is possible to specify the fields to include using the fields argument. This can be a a simple list of names that were present in the data argument. See references for more details.

Value

category	numeric
code	numeric
content_type	character
created	character
credits	numeric
description	character
fields	data.frame (or list if flatten=FALSE)
file_name	character
md5	character
name	character

number_of_datasets
: numeric

number_of_models
: numeric

number_of_predictions
: numeric

private
: logical

resource
: character

size
: numeric

source_parser
: list

status
: list

tags
: AsIs

type
: numeric

updated
: character

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/datasets

See Also

Other dataset methods: createDataset; getDataset; listDatasets

Other quick methods: quickModel; quickPrediction; quickSource

Examples

```
## Not run:
# simple example
iris.d = quickDataset(iris)
# configure a number of different parameters
iris.d2 = quickDataset(iris, fields = c('Species', 'Sepal.length'),
name='test', size=10000)

## End(Not run)
```

quickModel *Quickly Creating BigML Models*

Description

Quickly Creating BigML Models

Usage

```
quickModel(data, input_fields = names(data),
  objective_fields = tail(names(data), n = 1),
  name = paste(deparse(substitute(data)), "'s model", sep = ""),
  range = NULL, ...)
```

Arguments

data	A matrix or data frame containing data to upload to bigml.
input_fields	A vector of string names to use for training.
objective_fields	A single string value to use as an objective field (objective_fields is plural for future use).
name	A string giving the name of the model.
range	A two element numeric vector that defines a range over the dataset in which to train on.
...	Arbitrary named arguments that are passed on to formEncodeURL in order to create form-encoded URL options.

Details

quickModel will take its "data" dataframe argument and attempt to create a dataset using quickDataset. It is possible to specify the input_fields and objective_fields using the simple names from the data argument.

Value

category	numeric
code	numeric
columns	numeric
created	character
credits	numeric
dataset	character
dataset_status	logical
description	character

input_fields	character
locale	character
max_columns	numeric
max_rows	numeric
model	list
name	character
number_of_predictions	numeric
objective_fields	character
private	logical
range	numeric
resource	character
rows	numeric
size	numeric
source	character
source_status	logical
status	list
tags	AsIs
updated	character

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/models

See Also

Other model methods: createModel; getModel; listModels

Other quick methods: quickDataset; quickPrediction; quickSource

quickPrediction *Quickly Creating BigML Predictions*

Description

Quickly Creating BigML Predictions

Usage

```
quickPrediction(model, values, name = NULL, prediction_only = TRUE, ...)
```

Arguments

model	A character string or response object containing a valid model id value.
values	A named vector or list of elements to retrieve a prediction for
name	A string giving the name of the prediction.
prediction_only	if TRUE, only the predicted value is returned. Otherwise, the full API response is returned.
...	Arbitrary named arguments that are passed on to formEncodeURL in order to create form-encoded URL options.

Details

quickPrediction can operate on a model id string, or a model response object from an earlier request. The values are a list of named elements that are used as input.

Value

atomic character or numeric value if prediction_only is TRUE, else return:

category	numeric
code	numeric
created	character
credits	numeric
dataset	character
dataset_status	logical
description	character
fields	list
input_data	numeric
locale	character
model	character
model_status	logical

name	character
objective_fields	character
prediction	character
prediction_path	list
private	logical
resource	character
source	character
source_status	logical
status	list
tags	AsIs
updated	character

A numeric or string value giving the prediction.

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/predictions

See Also

Other prediction methods: createPrediction; getPrediction

Other quick methods: quickDataset; quickModel; quickSource

Examples

```
## Not run:
quickPrediction("model/1", list(Sepal.Width=3.5, Petal.Length=1.4))
# 'setosa'

## End(Not run)
```

quickSource *Quickly Creating BigML Sources*

Description

Quickly Creating BigML Sources

Usage

```
quickSource(data, name = deparse(substitute(data)),
  header = !is.null(names(data)), locale = "en-US",
  missing_tokens = c("NA"),
  quote = "\"", trim = TRUE, flatten = TRUE, ...)
```

Arguments

data	A matrix or data frame containing data to upload to bigml.
name	A string giving the name of the source.
header	A logical value indicating whether to use the first row of data as a header row.
locale	A string indicating the desired locale.
missing_tokens	A vector listing strings that should be treated as missing.
quote	A string giving the quote character to use.
trim	A logical value indicating whether to trim white space.
flatten	A logical value indicating whether to flatten the response into a data frame.
...	Arbitrary named arguments that are passed on to formEncodeURL in order to create form-encoded URL options.

Details

quickSource will take its "data" dataframe argument and attempt to create an equivalent BigML source. It does this by converting the dataframe to a csv file, compressing it, and uploading it directly to BigML. Generally, it's better to use quickDataset, since this method attempts to preserve any type information in the data frame.

Value

category	numeric
code	numeric
content_type	character
created	character
credits	numeric
description	character
fields	data.frame (or list if flatten=FALSE)

file_name	character
md5	character
name	character
number_of_datasets	numeric
number_of_models	numeric
number_of_predictions	numeric
private	logical
resource	character
size	numeric
source_parser	list
status	list
tags	AsIs
type	numeric
updated	character

Note

It is not currently possible to retrieve the original file from BigML, but it is possible to delete it.

Author(s)

Leon Hwang <hwang@bigml.com>

References

https://bigml.com/developers/sources

See Also

Other quick methods: quickDataset; quickModel; quickPrediction

Other source methods: createSource; getSource; listSources

`setCredentials` *Set BigML API authentication credentials*

Description

Set BigML API authentication credentials

Usage

```
setCredentials(username, api_key)
```

Arguments

`username`	use the given username for all subsequent API requests
`api_key`	use the given api key for all subsequent API requests

Details

This function sets default username and api_key information for subsequent BigML API access calls. The relevent username and key are stored in the R system environment variables. So, it's also possible to set these variables by setting BIGMLUSER and BIGMLAPIKEY in an .Renviron file.

Author(s)

Leon Hwang <hwang@bigml.com>

Examples

```
## Not run:
# replace with your valid credentials:
setCredentials('some_username', 'some_key')

## End(Not run)
```

Index