Package 'workflowsets'

October 12, 2022

Version 1.0.0 **Description** A workflow is a combination of a model and preprocessors (e.g, a formula, recipe, etc.) (Kuhn and Silge (2021) https://www.tmwr.org/). In order to try different combinations of these, an object can be created that contains many workflows. There are functions to create workflows en masse as well as training them and visualizing the results. License MIT + file LICENSE URL https://github.com/tidymodels/workflowsets, https://workflowsets.tidymodels.org BugReports https://github.com/tidymodels/workflowsets/issues **Depends** R (>= 3.4)Imports cli, dplyr (>= 1.0.0), generics (>= 0.1.2), ggplot2, glue, hardhat (>= 1.2.0), lifecycle (>= 1.0.0), parsnip (>= 1.0.0), pillar (>= 1.7.0), prettyunits, purrr, rlang, rsample (>= 0.0.9), stats, tibble (>= 3.1.0), tidyr, tune (>= 1.0.0), vctrs, withr, workflows (>= 1.0.0) Suggests covr, dials (>= 0.1.0), kknn, knitr, modeldata, recipes (>= 1.0.0), rmarkdown, spelling, testthat (>= 3.0.0), yardstick (>= 1.0.0) VignetteBuilder knitr Config/Needs/website discrim, rpart, mda, klaR, earth, tidymodels, tidyverse/tidytemplate Config/testthat/edition 3 **Encoding UTF-8** Language en-US LazyData true RoxygenNote 7.2.0 NeedsCompilation no

Title Create a Collection of 'tidymodels' Workflows

as_workflow_set

Author Max Kuhn [aut] (https://orcid.org/0000-0001-2676-2107), Simon Couch [aut, cre] (https://orcid.org/0000-0001-5676-5107), RStudio [cph, fnd]

Maintainer Simon Couch <simonpatrickcouch@gmail.com>

Repository CRAN

Date/Publication 2022-07-12 23:20:01 UTC

R topics documented:

workflow set	
update_workflow_model	
two_class_set	14
rank_results	
pull_workflow_set_result	
option_add	
leave_var_out_formulas	
extract_workflow_set_result	
collect_metrics.workflow_set	
autoplot.workflow_set	
as_workflow_set	2

Description

Use existing objects to create a workflow set. A list of objects that are either simple workflows or objects that have class "tune_results is converted into a workflow set.

Usage

```
as_workflow_set(...)
```

Arguments

One or more named objects. Names should be unique and the objects should have at least one of the following classes: workflow, iteration_results, tune_results_results_or tune_results_ele-

tune_results, resample_results, or tune_race. Each tune_results element should also contain the original workflow (accomplished using the save_workflow

option in the control function).

Value

A workflow set. Note that the option column will not reflect the options that were used to create each object.

Examples

```
# ------
# Existing results
# Use the already worked example to show how to add tuned
# objects to a workflow set
results <- two_class_res %>% purrr::pluck("result")
names(results) <- two_class_res$wflow_id</pre>
# These are all objects that have been resampled or tuned:
purrr::map_chr(results, ~ class(.x)[1])
# Use rlang's !!! operator to splice in the elements of the list
new_set <- as_workflow_set(!!!results)</pre>
# ------
# Make a set from unfit workflows
library(parsnip)
library(workflows)
lr_spec <- logistic_reg()</pre>
main_effects <-</pre>
 workflow() %>%
 add_model(lr_spec) %>%
 add_formula(Class ~ .)
interactions <-
 workflow() %>%
 add_model(lr_spec) %>%
 add_formula(Class ~ (.)^2)
as_workflow_set(main = main_effects, int = interactions)
```

autoplot.workflow_set Plot the results of a workflow set

Description

This autoplot() method can performance metrics that have been ranked using a metric. It can also run autoplot() on the individual results (per wflow_id).

Usage

```
## S3 method for class 'workflow_set'
autoplot(
  object,
  rank_metric = NULL,
  metric = NULL,
  id = "workflow_set",
  select_best = FALSE,
  std_errs = qnorm(0.95),
  ...
)
```

Arguments

object A workflow_set whose elements have results.

rank_metric A character string for which metric should be used to rank the results. If none

is given, the first metric in the metric set is used (after filtering by the metric

option).

metric A character vector for which metrics (apart from rank_metric) to be included

in the visualization.

id A character string for what to plot. If a value of "workflow_set" is used, the

results of each model (and sub-model) are ordered and plotted. Alternatively, a value of the workflow set's wflow_id can be given and the autoplot() method

is executed on that workflow's results.

select_best A logical; should the results only contain the numerically best submodel per

workflow?

std_errs The number of standard errors to plot (if the standard error exists).

... Other options to pass to autoplot().

Details

This function is intended to produce a default plot to visualize helpful information across all possible applications of a workflow set. A more appropriate plot for your specific analysis can be created by calling rank_results() and using standard ggplot2 code for plotting.

The x-axis is the workflow rank in the set (a value of one being the best) versus the performance metric(s) on the y-axis. With multiple metrics, there will be facets for each metric.

If multiple resamples are used, confidence bounds are shown for each result (90% confidence, by default).

Value

A ggplot object.

Examples

```
autoplot(two_class_res)
autoplot(two_class_res, select_best = TRUE)
autoplot(two_class_res, id = "yj_trans_cart", metric = "roc_auc")
```

collect_metrics.workflow_set

Obtain and format results produced by tuning functions for workflow sets

Description

Return a tibble of performance metrics for all models or submodels.

Usage

```
## S3 method for class 'workflow_set'
collect_metrics(x, summarize = TRUE, ...)
## S3 method for class 'workflow_set'
collect_predictions(
    x,
    summarize = TRUE,
    parameters = NULL,
    select_best = FALSE,
    metric = NULL,
    ...
)
```

Arguments

x	A workflow_set object where all workflows have been evaluated.
summarize	A logical for whether the performance estimates should be summarized via the mean (over resamples) or the raw performance values (per resample) should be returned along with the resampling identifiers. When collecting predictions, these are averaged if multiple assessment sets contain the same row.
•••	Not currently used.
parameters	An optional tibble of tuning parameter values that can be used to filter the predicted values before processing. This tibble should only have columns for each tuning parameter identifier (e.g. "my_param" if tune("my_param") was used).
select_best	A single logical for whether the numerically best results are retained. If TRUE, the parameters argument is ignored.
metric	A character string for the metric that is used for select_best.

6 comment_add

Details

When applied to a workflow set, the metrics and predictions that are returned do not contain the actual tuning parameter columns and values (unlike when these collect functions are run on other objects). The reason is that workflow sets can contain different types of models or models with different tuning parameters.

If the columns are needed, there are two options. First, the .config column can be used to merge the tuning parameter columns into an appropriate object. Alternatively, the map() function can be used to get the metrics from the original objects (see the example below).

Value

A tibble.

See Also

```
tune::collect_metrics(), rank_results()
```

Examples

comment_add

Add annotations and comments for workflows

Description

comment_add() can be used to log important information about the workflow or its results as you work. Comments can be appended or removed.

Usage

```
comment_add(x, id, ..., append = TRUE, collapse = "\n")
comment_get(x, id)
comment_reset(x, id)
comment_print(x, id = NULL, ...)
```

Arguments

Χ	A workflow set
id	A single character string for a value in the wflow_id column. For comment_print(), id can be a vector or NULL (and this indicates that all comments are printed).
	One or more character strings.
append	A logical value to determine if the new comment should be added to the existing values.
collapse	A character string that separates the comments.

Value

comment_add() and comment_reset() return an updated workflow set. comment_get() returns a
character string. comment_print() returns NULL invisibly.

Examples

```
two_class_set %>% comment_get("none_cart")

new_set <-
   two_class_set %>%
   comment_add("none_cart", "What does 'cart' stand for\u2753") %>%
   comment_add("none_cart", "Classification And Regression Trees.")

comment_print(new_set)

new_set %>% comment_get("none_cart")

new_set %>%
   comment_reset("none_cart") %>%
   comment_get("none_cart")
```

```
extract_workflow_set_result
```

Extract elements of workflow sets

Description

These functions extract various elements from a workflow set object. If they do not exist yet, an error is thrown.

- extract_preprocessor() returns the formula, recipe, or variable expressions used for preprocessing.
- extract_spec_parsnip() returns the parsnip model specification.
- extract_fit_parsnip() returns the parsnip model fit object.
- extract_fit_engine() returns the engine specific fit embedded within a parsnip model fit. For example, when using parsnip::linear_reg() with the "lm" engine, this returns the underlying lm object.
- extract_mold() returns the preprocessed "mold" object returned from hardhat::mold(). It contains information about the preprocessing, including either the prepped recipe, the formula terms object, or variable selectors.
- extract_recipe() returns the recipe. The estimated argument specifies whether the fitted or original recipe is returned.
- extract_workflow_set_result() returns the results of workflow_map() for a particular workflow.
- extract_workflow() returns the workflow object. The workflow will not have been estimated.

Usage

```
extract_workflow_set_result(x, id, ...)

## S3 method for class 'workflow_set'
extract_workflow(x, id, ...)

## S3 method for class 'workflow_set'
extract_spec_parsnip(x, id, ...)

## S3 method for class 'workflow_set'
extract_recipe(x, id, ..., estimated = TRUE)

## S3 method for class 'workflow_set'
extract_fit_parsnip(x, id, ...)

## S3 method for class 'workflow_set'
extract_fit_engine(x, id, ...)

## S3 method for class 'workflow_set'
extract_mold(x, id, ...)

## S3 method for class 'workflow_set'
extract_preprocessor(x, id, ...)
```

leave_var_out_formulas 9

```
## S3 method for class 'workflow_set'
extract_parameter_set_dials(x, id, ...)
## S3 method for class 'workflow_set'
extract_parameter_dials(x, id, parameter, ...)
```

Arguments

x A workflow set.

id A single character string for a workflow ID.

... Other options (not currently used).

estimated A logical for whether the original (unfit) recipe or the fitted recipe should be

returned.

parameter A single string for the parameter ID.

Details

These functions supersede the pull_*() functions (e.g., extract_workflow_set_result()).

Value

The extracted value from the object, x, as described in the description section.

Examples

```
library(tune)
extract_workflow_set_result(two_class_res, "none_cart")
extract_workflow(two_class_res, "none_cart")
```

```
leave_var_out_formulas
```

Create formulas without each predictor

Description

From an initial model formula, create a list of formulas that exclude each predictor.

Usage

```
leave_var_out_formulas(formula, data, full_model = TRUE, ...)
```

10 option_add

Arguments

formula A model formula that contains at least two predictors.

A data frame.

A logical; should the list include the original formula?

Options to pass to stats::model.frame()

Details

The new formulas obey the hierarchy rule so that interactions without main effects are not included (unless the original formula contains such terms).

Factor predictors are left as-is (i.e., no indicator variables are created).

Value

A named list of formulas

See Also

```
workflow_set()
```

Examples

```
data(penguins, package = "modeldata")
leave_var_out_formulas(
  bill_length_mm ~ .,
  data = penguins
)
leave_var_out_formulas(
  bill_length_mm ~ (island + sex)^2 + flipper_length_mm,
  data = penguins
)
leave_var_out_formulas(
  bill_length_mm ~ (island + sex)^2 + flipper_length_mm +
        I(flipper_length_mm^2),
  data = penguins
)
```

option_add

Add and edit options saved in a workflow set

Description

The option column controls options for the functions that are used to *evaluate* the workflow set, such as tune::fit_resamples() or tune::tune_grid(). Examples of common options to set for these functions include param_info and grid.

These functions are helpful for manipulating the information in the option column.

option_add 11

Usage

```
option_add(x, ..., id = NULL, strict = FALSE)
option_remove(x, ...)
option_add_parameters(x, id = NULL, strict = FALSE)
```

Arguments

X	A workflow set.
•••	A list of named options to pass to the tune_*() functions (e.g. tune::tune_grid()) or tune::fit_resamples(). For option_remove() this can be a series of unquoted option names.
id	A character string of one or more values from the wflow_id column that indicates which options to update. By default, all workflows are updated.
strict	A logical; show execution stop if existing options are being replaced?

Details

option_add() is used to update all of the options in a workflow set.

option_remove() will eliminate specific options across rows.

option_add_parameters() adds a parameter object to the option column (if parameters are being tuned).

Note that executing a function on the workflow set, such as tune_grid(), will add any options given to that function to the option column.

These functions do *not* control options for the individual workflows, such as the recipe blueprint. When creating a workflow manually, use workflows::add_model() or workflows::add_recipe() to specify. extra options. To alter these in a workflow set, use update_workflow_model() or update_workflow_recipe().

Value

An updated workflow set.

Examples

```
two_class_set %>%
  option_add(grid = 10)

two_class_set %>%
  option_add(grid = 10) %>%
  option_add(grid = 50, id = "none_cart")

library(tune)
two_class_set %>%
  option_add_parameters()
```

option_list

Make a classed list of options

Description

This function returns a named list with an extra class of "workflow_set_options" that has corresponding formatting methods for printing inside of tibbles.

Usage

```
option_list(...)
```

Arguments

... A set of named options (or nothing)

Value

A classed list.

Examples

```
option_list(a = 1, b = 2)
option_list()
```

```
pull_workflow_set_result
```

Extract elements from a workflow set

Description

[Soft-deprecated]

Usage

```
pull_workflow_set_result(x, id)
pull_workflow(x, id)
```

Arguments

x A workflow set.

id A single character string for a workflow ID.

rank_results 13

Details

pull_workflow_set_result() retrieves the results of workflow_map() for a particular workflow while pull_workflow() extracts the unfitted workflow from the info column.

The extract_workflow_set_result() and extract_workflow() functions should be used instead of these functions.

Value

pull_workflow_set_result() produces a tune_result or resample_results object. pull_workflow()
returns an unfit workflow object.

Examples

```
library(tune)
pull_workflow_set_result(two_class_res, "none_cart")
pull_workflow(two_class_res, "none_cart")
```

rank_results

Rank the results by a metric

Description

This function sorts the results by a specific performance metric.

Usage

```
rank_results(x, rank_metric = NULL, select_best = FALSE)
```

Arguments

x A workflow set that has all results.

rank_metric A character string for a metric.

select_best A logical; should the results only contain the numerically best submodel per

workflow.

Details

If some models have the exact same performance, rank(value, ties.method = "random") is used (with a reproducible seed) so that all ranks are integers.

No columns are returned for the tuning parameters since they are likely to be different (or not exist) for some models. The wflow_id and .config columns can be used to determine the corresponding parameter values.

two_class_set

Value

A tibble with columns: wflow_id, .config, .metric, mean, std_err, n, preprocessor, model, and rank.

Examples

```
rank_results(chi_features_res)
rank_results(chi_features_res, select_best = TRUE)
rank_results(chi_features_res, rank_metric = "rsq")
```

two_class_set

Example Data Sets

Description

Example Data Sets

Details

Example workflow sets and associated model fits.

two_class_set and two_class_res were generated using the data in the package file example-data/two-class-set.R chi_features_set and chi_features_res were generated using the data in the package file example-data/chi-features-res.R. It is meant to approximate the sequence of models built in Section 1.3 of Kuhn and Johnson (2019).

Value

Workflow sets.

References

Max Kuhn and Kjell Johnson (2019) Feature Engineering and Selection, https://bookdown.org/max/FES/a-more-complex-example.html

Examples

```
data(two_class_set)
two_class_set
```

update_workflow_model Update components of a workflow within a workflow set

Description

Workflows can take special arguments for the recipe (e.g. a blueprint) or a model (e.g. a special formula). However, when creating a workflow set, there is no way to specify these extra components. update_workflow_model() and update_workflow_recipe() allow users to set these values *after* the workflow set is initially created. They are analogous to workflows::add_model() or workflows::add_recipe().

Usage

```
update_workflow_model(x, id, spec, formula = NULL)
update_workflow_recipe(x, id, recipe, blueprint = NULL)
```

Arguments

x A workflow set.

id A single character string from the wflow_id column indicating which workflow

to update.

spec A parsnip model specification.

formula An optional formula override to specify the terms of the model. Typically, the

terms are extracted from the formula or recipe preprocessing methods. However, some models (like survival and bayesian models) use the formula not to preprocess, but to specify the structure of the model. In those cases, a formula specifying the model structure must be passed unchanged into the model call

itself. This argument is used for those purposes.

recipe A recipe created using recipes::recipe()

blueprint A hardhat blueprint used for fine tuning the preprocessing.

If NULL, hardhat::default_recipe_blueprint() is used.

Note that preprocessing done here is separate from preprocessing that might be

done automatically by the underlying model.

Examples

```
library(parsnip)
new_mod <-
    decision_tree() %>%
    set_engine("rpart", method = "anova") %>%
    set_mode("classification")

new_set <- update_workflow_model(two_class_res, "none_cart", spec = new_mod)
new_set

extract_workflow(new_set, id = "none_cart")</pre>
```

16 workflow_map

workflow_map

Process a series of workflows

Description

workflow_map() will execute the same function across the workflows in the set. The various tune_*() functions can be used as well as tune::fit_resamples().

Usage

```
workflow_map(
  object,
  fn = "tune_grid",
  verbose = FALSE,
  seed = sample.int(10^4, 1),
  ...
)
```

Arguments

object A workflow set.

fn The function to run. Acceptable values are: tune::tune_grid(), tune::tune_bayes(),

tune::fit_resamples(), finetune::tune_race_anova(), finetune::tune_race_win_loss(),

or finetune::tune_sim_anneal().

verbose A logical for logging progress.

A single integer that is set prior to each function execution.

Options to pass to the modeling function. See details below.

Details

When passing options, anything passed in the ... will be combined with any values in the option column. The values in ... will override that column's values and the new options are added to the options column.

Any failures in execution result in the corresponding row of results to contain a try-error object.

In cases where a model has no tuning parameters is mapped to one of the tuning functions, tune::fit_resamples() will be used instead and a warning is issued if verbose = TRUE.

If a workflow required packages that are not installed, a message is printed and workflow_map() continues with the next workflow (if any).

Value

An updated workflow set. The option column will be updated with any options for the tune package functions given to workflow_map(). Also, the results will be added to the result column. If the computations for a workflow fail, an try-catch object will be saved in place of the results (without stopping execution).

workflow_set 17

See Also

```
workflow_set(), as_workflow_set(), extract_workflow_set_result()
```

Examples

```
# An example of processed results
chi_features_res

# Code examples at
if (interactive()) {
   system.file("example-data", package = "workflowsets")
}
```

workflow_set

Generate a set of workflow objects from preprocessing and model objects

Description

Generate a set of workflow objects from preprocessing and model objects

Usage

```
workflow_set(preproc, models, cross = TRUE, case_weights = NULL)
```

Arguments

preproc	A list (preferably named) with preprocessing objects: formulas, recipes, or workflows::workflow_variables().
models	A list (preferably named) of parsnip model specifications.
cross	A logical: should all combinations of the preprocessors and models be used to create the workflows? If FALSE, the length of preproc and models should be equal.
case_weights	A single unquoted column name specifying the case weights for the models. This must be a classed case weights column, as determined by hardhat::is_case_weights(). See the "Case weights" section below for more information.

Details

The preprocessors that can be combined with the model objects can be one or more of:

- A traditional R formula.
- A recipe definition (un-prepared) via recipes::recipe().
- A selectors object created by workflows::workflow_variables().

Since preproc is a named list column, any combination of these can be used in that argument (i.e., preproc can be mixed types).

18 workflow_set

Value

A tibble with extra class 'workflow_set'. A new set includes four columns (but others can be added):

- wflow_id contains character strings for the preprocessor/workflow combination. These can be changed but must be unique.
- info is a list column with tibbles containing more specific information, including any comments added using comment_add(). This tibble also contains the workflow object (which can be easily retrieved using extract_workflow()).
- option is a list column that will include a list of optional arguments passed to the functions from the tune package. They can be added manually via option_add() or automatically when options are passed to workflow_map().
- result is a list column that will contain any objects produced when workflow_map() is used.

Case weights

The case_weights argument can be passed as a single unquoted column name identifying the data column giving model case weights. For each workflow in the workflow set using an engine that supports case weights, the case weights will be added with workflows::add_case_weights(). workflow_set() will warn if any of the workflows specify an engine that does not support case weights—and ignore the case weights argument for those workflows—but will not fail.

Read more about case weights in the tidymodels at ?parsnip::case_weights.

See Also

```
workflow_map(), comment_add(), option_add(), as_workflow_set()
```

Examples

```
library(workflowsets)
library(workflows)
library(modeldata)
library(recipes)
library(parsnip)
library(dplyr)
library(rsample)
library(tune)
library(yardstick)
 ______
data(cells)
cells <- cells %>% dplyr::select(-case)
set.seed(1)
val_set <- validation_split(cells)</pre>
basic_recipe <-</pre>
 recipe(class ~ ., data = cells) %>%
```

workflow_set 19

```
step_YeoJohnson(all_predictors()) %>%
 step_normalize(all_predictors())
pca_recipe <-</pre>
 basic_recipe %>%
 step_pca(all_predictors(), num_comp = tune())
ss_recipe <-
 basic_recipe %>%
 step_spatialsign(all_predictors())
# ------
knn_mod <-
 nearest_neighbor(neighbors = tune(), weight_func = tune()) %>%
 set_engine("kknn") %>%
 set_mode("classification")
1r_mod <-
 logistic_reg() %>%
 set_engine("glm")
# -----
preproc <- list(none = basic_recipe, pca = pca_recipe, sp_sign = ss_recipe)</pre>
models <- list(knn = knn_mod, logistic = lr_mod)</pre>
cell_set <- workflow_set(preproc, models, cross = TRUE)</pre>
cell_set
# -----
# Using variables and formulas
# Select predictors by their names
channels <- paste0("ch_", 1:4)</pre>
preproc <- purrr::map(channels, ~ workflow_variables(class, c(contains(!!.x))))</pre>
names(preproc) <- channels</pre>
preproc$everything <- class \sim .
preproc
cell_set_by_group <- workflow_set(preproc, models["logistic"])</pre>
cell_set_by_group
```

Index

```
* datasets
                                               extract_spec_parsnip.workflow_set
    two_class_set, 14
                                                       (extract_workflow_set_result),
as_workflow_set, 2
                                               extract_workflow(), 13, 18
as_workflow_set(), 17, 18
                                               extract_workflow.workflow_set
autoplot.workflow_set, 3
                                                       (extract_workflow_set_result),
chi_features_res (two_class_set), 14
                                               extract_workflow_set_result, 7
chi_features_set (two_class_set), 14
                                               extract_workflow_set_result(), 9, 13, 17
collect_metrics.workflow_set, 5
collect_predictions.workflow_set
                                               hardhat::default_recipe_blueprint(),
        (collect_metrics.workflow_set),
                                                       15
                                               hardhat::is_case_weights(), 17
comment_add, 6
                                               hardhat::mold(), 8
comment_add(), 18
comment_get (comment_add), 6
                                               leave_var_out_formulas, 9
comment_print (comment_add), 6
                                               option_add, 10
comment_reset (comment_add), 6
                                               option_add(), 18
extract_fit_engine.workflow_set
                                               option_add_parameters (option_add), 10
        (extract_workflow_set_result),
                                               option_list, 12
                                               option_remove (option_add), 10
extract_fit_parsnip.workflow_set
        (extract_workflow_set_result),
                                               parsnip::linear_reg(), 8
                                               pull_workflow
extract_mold.workflow_set
                                                        (pull_workflow_set_result), 12
        (extract_workflow_set_result),
                                               pull_workflow_set_result, 12
extract_parameter_dials.workflow_set
                                               rank_results, 13
        (extract_workflow_set_result),
                                               rank_results(), 4, 6
                                               recipes::recipe(), 15, 17
extract_parameter_set_dials.workflow_set
        (extract_workflow_set_result),
                                               stats::model.frame(), 10
extract_preprocessor.workflow_set
                                               tune::collect_metrics(), 6
        (extract_workflow_set_result),
                                               tune::fit_resamples(), 10, 11, 16
                                               tune::tune_bayes(), 16
extract_recipe.workflow_set
                                               tune::tune_grid(), 10, 11, 16
        (extract_workflow_set_result),
                                               two_class_res (two_class_set), 14
                                               two_class_set, 14
```

INDEX 21