

Package ‘WordListsAnalytics’

May 6, 2024

Type Package

Title Multiple Data Analysis Tools for Property Listing Tasks

Version 0.2.2

Maintainer Sebastian Moreno <sebastian.moreno.araya@gmail.com>

Description Application to estimate statistical values using properties provided by a group of individuals to describe concepts using 'shiny'. It estimates the underlying distribution to generate new descriptive words Canessa et al. (2023) <[doi:10.3758/s13428-022-01811-w](https://doi.org/10.3758/s13428-022-01811-w)>, applies a new clustering model, and uses simulations to estimate the probability that two persons describe the same words based on their descriptions Canessa et al. (2022) <[doi:10.3758/s13428-022-02030-z](https://doi.org/10.3758/s13428-022-02030-z)>.

License GPL (>= 3)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.1

Imports ggplot2, readr, dplyr, reshape2, grDevices, stats, graphics

Depends shiny, R (>= 4.2.0)

Collate 'CPN_27.R' 'CPN_120.R' 'tab_upload_data.R' 'tab_estimations.R' 'tab_estimate_participants.R' 'tab_property_simulator.R' 'tab_pa_data.R' 'tab_pa_values.R' 'tab_cluster_image.R' 'PLT_ui.R' 'fun_generate_norms.R' 'fun_estimate_participant.R' 'fun_property_simulator.R' 'fun_pa_function.R' 'fun_threshold_graph.R' 'fun_cluster_image_function.R' 'PLT_server.R' 'PLT_app.R'

NeedsCompilation no

Author Sebastian Moreno [aut, cre],
Cristobal Heredia [aut],
Enrique Canessa [ths],
Sergio Chaigneau [ths]

Repository CRAN

Date/Publication 2024-05-06 14:20:02 UTC

R topics documented:

clusterImage	2
CPN_120	3
CPN_27	3
estimate_participant	4
generate_norms	5
WordListsAnalytics	5

Index	6
--------------	----------

clusterImage	<i>This function receives a property listing task, a given concept, and a threshold. It clusterizes the data according to the order of the listed properties. Given the mentioned properties of all users for a specific concept, the algorithm estimates a similarity among properties, based on the number of words mentioned between properties. For example, if the properties A and B are usually mentioned one after another, their similarity will be higher than the properties A and C which are usually not even mentioned together. The properties with low similarity to all other properties (below the user-defined threshold) are discarded from the plot.</i>
--------------	---

Description

This function receives a property listing task, a given concept, and a threshold. It clusterizes the data according to the order of the listed properties. Given the mentioned properties of all users for a specific concept, the algorithm estimates a similarity among properties, based on the number of words mentioned between properties. For example, if the properties A and B are usually mentioned one after another, their similarity will be higher than the properties A and C which are usually not even mentioned together. The properties with low similarity to all other properties (below the user-defined threshold) are discarded from the plot.

Usage

```
clusterImage(data, distThreshold, concept = NULL)
```

Arguments

data	Data frame with 3 columns: ID, Concept and Property
distThreshold	Distance value. It assign properties to specific cluster if their similarity is greater than distThreshold
concept	Text value. Clusters will only be generated with properties from this concept.

Value

List with 2 elements: ggplot2 plot and data frame with cluster information

Examples

```
data_cpn = data.frame(CPN_27)
threshold = 0.061
concept = "Ability"
cluster_data = clusterImage(data_cpn, threshold, concept)
```

CPN_120

*CPN Example data***Description**

The CPN120 dataset is a property listing task dataset over 120 concepts (60 concrete and 60 abstract). The dataset was generated from 221 voluntary Chilean university students (71% male, 28.5% female, average age = 23.7 years with s.d. = 6.2 years). Each participant listed up to 10 characteristics for each concept. The dataset had over 32,000 responses, which were categorized into valid and invalid, obtaining 31,864 valid responses.

Usage

```
data(CPN_120)
```

Format

A data frame with 31864 rows and 3 variables:

ID ID for original subject

Concept Concept asked

Property Property given by the subject ...

Source

Fondecyt proyect #1200139, Chilean government

CPN_27

*CPN Example data***Description**

The CPN27 dataset is a property listing task dataset over 27 abstract concepts. The dataset was generated from 100 voluntary Chilean university students (51% males, 49% females, mean age = 21.0 years with s.d. = 1.42 years). Each student listed features for 10 of the 27 concepts. The dataset had over 5,000 responses, which were sorted into valid and invalid, obtaining 4697 valid responses.

Usage

```
data(CPN_27)
```

Format

A data frame with 4618 rows and 3 variables:

ID ID for original subject

Concept Concept asked

Property Property given by the subject ...

Source

Fondecyt proyect #1200139, Chilean goverment

estimate_participant	<i>Estimate the number of people needed and expected number of unique properties for a determined coverage based on the estimated norms</i>
----------------------	---

Description

Estimate the number of people needed and expected number of unique properties for a determined coverage based on the estimated norms

Usage

```
estimate_participant(est_norms, target_cover)
```

Arguments

est_norms A data frame with the estimated norms (generated by generateNorms function)

target_cover Float between 0 and 1, corresponding to coverage (the fraction of the total incidence probabilities of the reported properties that are in the reference sample)

Value

A vector with the extra number of participant to achieve the specific coverage, and the estimate of the number of unique properties listed by the new amount of suggested people

Examples

```
data_cpn = data.frame(CPN_27)
estimated_norms = generate_norms(data_cpn)
estimated_norms = na.omit(estimated_norms)
estimate_participant(estimated_norms, 0.8)
```

generate_norms	<i>Calculate all the norms from a Conceptual properties</i>
----------------	---

Description

Calculate all the norms from a Conceptual properties

Usage

```
generate_norms(orig_data)
```

Arguments

orig_data Data frame with 3 columns: id, concept and properties

Value

Data frame with all the estimations of norms

Examples

```
data_test = data.frame(CPN_27)
generate_norms(data_test)
```

WordListsAnalytics	<i>PLT App function</i>
--------------------	-------------------------

Description

Main function of package. It executes a shiny application in local session. The user can load data, generate new descriptive words, apply a new clustering model, and use simulations to estimate the probability that two persons describe the same words based on their descriptions.

Usage

```
WordListsAnalytics()
```

Value

None (it executes a shiny application).

Examples

```
if(interactive()){
  WordListsAnalytics()
}
```

Index

* datasets

CPN_120, [3](#)

CPN_27, [3](#)

clusterImage, [2](#)

CPN_120, [3](#)

CPN_27, [3](#)

estimate_participant, [4](#)

generate_norms, [5](#)

WordListsAnalytics, [5](#)