Package 'MAP'

January 20, 2025

Type Package

Title Multimodal Automated Phenotyping

Version 1.0.0

Description Electronic health records (EHR) linked with biorepositories are
a powerful platform for translational studies. A major bottleneck exists
in the ability to phenotype patients accurately and efficiently.
Towards that end, we developed an automated high-throughput
phenotyping method integrating International
Classification of Diseases (ICD) codes and narrative data extracted
using natural language processing (NLP). Specifically, our proposed method,
called MAP (Map Automated Phenotyping algorithm), fits an ensemble of latent
mixture models on aggregated ICD and NLP counts along with healthcare
utilization. The MAP algorithm yields a predicted probability of phenotype
for each patient and a threshold for classifying subjects with phenotype
yes/no (See Katherine P. Liao, et al. (2019) <doi:10.1093/jamia/ocz066>.).

URL https://celehs.github.io/MAP/

BugReports https://github.com/celehs/MAP/issues

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 3.4.0), flexmix (>= 2.3-14), Matrix(>= 1.2-10)

Imports magrittr

Suggests knitr, pROC, rmarkdown, testthat

VignetteBuilder knitr

NeedsCompilation no

RoxygenNote 7.3.2

Author Thomas Charlon [aut, cre] (<https://orcid.org/0000-0001-7497-0470>),
Chuan Hong [aut],
Jiehuan Sun [aut],
Katherine Liao [aut],
Sheng Yu [aut],

Tianxi Cai [aut],
PARSE Health [aut] (<https://parse-health.org>),
CELEHS [aut] (<https://celehs.hms.harvard.edu>)

Contents

MAP   *MAP algorithm*

Description

Main function to perform MAP algorithm to calculate predicted probabilities of positive phenotype
for each patient based on NLP and ICD counts adjusted for healthcare utilization. For large number
of patients (>50k) it may take very long to compute, so a subset_sample parameter is provided to
perform the fit on a subset of patients and project the remaining. The subset_sample_size controls
the maximum number of patients on which to perform the fit.

Usage

```
MAP(
  mat = NULL,
  note = NULL,
  yes.con = FALSE,
  full.output = FALSE,
  subset_sample = FALSE,
  subset_sample_size = 5000,
  verbose = TRUE
)
```

Arguments

mat	Count data (sparse matrix). One of the columns has to be ICD data with name being ICD.
note	Note count (sparse matrix) indicating healthcare utilization.
yes.con	A logical variable indicating if concomitant is desired. Not used for now.

full.output	A logical variable indicating if full outputs are desired.
subset_sample	Logical, perform fit on a subset of patients and project remaining.
subset_sample_size	
	If subset_sample TRUE, number of patients on which to perform the fit (default 50k).
verbose	Print model information

Value

Returns a list with following objects:

| scores | Indicates predicted probabilities. |
| cut.MAP | The cutoff value that can be used to derive binary phenotypes. |

References

High-throughput Multimodal Automated Phenotyping (MAP) with Application to PheWAS. Katherine P. Liao, Jiehuan Sun, Tianrun A. Cai, Nicholas Link, Chuan Hong, Jie Huang, Jennifer Huffman, Jessica Gronsbell, Yichi Zhang, Yuk-Lam Ho, Victor Castro, Vivian Gainer, Shawn Murphy, Christopher J. O'Donnell, J. Michael Gaziano, Kelly Cho, Peter Szolovits, Isaac Kohane, Sheng Yu, and Tianxi Cai with the VA Million Veteran Program (2019) <doi:10.1101/587436>.

Examples

```
## simulate data to test the algorithm
n = 400
ICD = c(rpois(n/4,10), rpois(n/4,1), rep(0,n/2) )
NLP = c(rpois(n/4,10), rpois(n/4,1), rep(0,n/2) )
mat = Matrix(data=cbind(ICD,NLP),sparse = TRUE)
note = Matrix(rpois(n,10)+5,ncol=1,sparse = TRUE)
res = MAP(mat = mat,  note=note)
head(res$scores)
res$cut.MAP
```

---

phecode.cuis.list   *MAP dictionary*

Description

MAP dictionary that maps phecode to CUIs

Usage

```
phecode.cuis.list
```

Format

A list of 1866

Examples

```
head(phecode.cuis.list)
tail(phecode.cuis.list)
```

%<>%	*Assignment pipe*

Description

Pipe an object forward into a function or call expression and update the 'lhs' object with the resulting value. Magrittr imported function, see details and examples in the magrittr package.

Arguments

lhs   An object which serves both as the initial value and as target.

rhs   a function call using the magrittr semantics.

Value

None, used to update the value of lhs.

%$%	*Exposition pipe*

Description

Expose the names in 'lhs' to the 'rhs' expression. Magrittr imported function, see details and examples in the magrittr package.

Arguments

lhs   A list, environment, or a data.frame.

rhs   An expression where the names in lhs is available.

Value

Result of rhs applied to one or several names of lhs.

%>%	*Pipe*

Description

Pipe an object forward into a function or call expression. Magrittr imported function, see details and examples in the magrittr package.

Arguments

lhs	A value or the magrittr placeholder.
rhs	A function call using the magrittr semantics.

Value

Result of rhs applied to lhs, see details in magrittr package.

Index