Package 'MAP'

April 1, 2019

Type Package

Title Multimodal Automated Phenotyping

Version 0.1.3

Author Jiehuan Sun [aut, cre], Katherine P. Liao[aut], Sheng Yu [aut], Tianxi Cai [aut]

Maintainer Jiehuan Sun <jiehuan.sun@gmail.com>

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:10.1101/587436>.).

License GPL-2

Encoding UTF-8

LazyData true

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

Suggests knitr, rmarkdown

NeedsCompilation no

RoxygenNote 6.1.1

Repository CRAN

Date/Publication 2019-04-01 08:20:09 UTC

R topics documented:

```
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.

Usage

```
MAP(mat = NULL, note = NULL, yes.con = FALSE, full.output = FALSE)
```

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 health 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.

Value

Returns a list with following objects:

scores	indicates predicted probabilities.
cut.MAP	the cutoff value that can be used to derive binary phenotype.

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
```

Index