CEM_Charity.R

henrydambanemuya

2019-12-13

```
#!/usr/bin/env Rscript

# Install Packages
# install.packages("cem")

# Import Packages
library(cem)
```

```
## Loading required package: tcltk
```

```
## Loading required package: lattice
```

```
##
## How to use CEM? Type vignette("cem")
```

```
library(readr)
library(xtable)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
library(reshape)
```

```
##
## Attaching package: 'reshape'
```

```
## The following object is masked from 'package:dplyr':
##
##     rename
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:reshape':
##
##     expand
```

```r
# Set working directory
setwd("~/Documents/CRII/DonorsChoose")
# Import Data
charity <- read.csv("./Data/charity_cem.csv")

# Independent Variables (Crowd Features)
charity$NumContributors <- as.numeric(charity$NumContributors)
charity$CovInterEventTime <- as.numeric(charity$CovInterEventTime)
charity$CovContributionAmount <- as.numeric(charity$CovContributionAmount)
charity$TimeToFirstContribution..sec. <- as.numeric(charity$TimeToFirstContribution..se
c.)
charity$Duration..days. <- as.numeric(charity$Duration..days.)

# Pre-Treatment Control Variables (Platform Features)
charity$primary_focus_area <-  as.factor(charity$primary_focus_area)
charity$grade_level <-  as.factor(charity$grade_level)
charity$school_metro <-  as.factor(charity$school_metro)
charity$resource_type <-  as.factor(charity$resource_type)
charity$teacher_prefix <-  as.factor(charity$teacher_prefix)
charity$AmountRequested <-  as.numeric(charity$AmountRequested)

# Dependent Variable (Class Feature)
charity$Status <-  as.numeric(charity$Status)

# Compare class sizes

ct <- which(charity$Status==0)
tr <- which(charity$Status==1)

ntr <- length(tr)
nct <- length(ct)

table(charity$Status)
```

```
##
##      0     1
## 1295 1295
```

```
# Platform variables: pre-treatment covariates (not randomly assigned)
vars.platform <- c("school_metro", "primary_focus_area", "grade_level", "resource_type",
"teacher_prefix", "AmountRequested", 'Status')

# Focus on these pre-treatment covariates
# Compute L1 statistic, as well as several unidimensional measures of imbalance
imbalance(group = charity$Status, data = charity[vars.platform], drop = "Status")
```

```
## Warning in chisq.test(cbind(t1[keep], t2[keep])): Chi-squared approximation
## may be incorrect
```

```
##
## Multivariate Imbalance Measure: L1=0.652
## Percentage of local common support: LCS=19.0%
##
## Univariate Imbalance Measures:
##
##                      statistic   type          L1        min         25%
## school_metro        3.33662435 (Chi2) 0.024710425         NA          NA
## primary_focus_area 11.96926813 (Chi2) 0.042471042         NA          NA
## grade_level        28.19175619 (Chi2) 0.094980695         NA          NA
## resource_type       6.25028736 (Chi2) 0.035521236         NA          NA
## teacher_prefix     11.95526285 (Chi2) 0.061776062         NA          NA
## AmountRequested     0.01422197 (diff) 0.003861004 0.00696999 0.004803197
##                            50%         75%        max
## school_metro                NA          NA         NA
## primary_focus_area          NA          NA         NA
## grade_level                 NA          NA         NA
## resource_type               NA          NA         NA
## teacher_prefix              NA          NA         NA
## AmountRequested    0.008444728 0.007396346  0.4015056
```

```
# Automated Coarsening
mat <- cem(treatment = "Status", data = charity[vars.platform], drop = "Status", eval.im
balance = TRUE, keep.all = TRUE)
```

```
## Warning in chisq.test(cbind(t1[keep], t2[keep])): Chi-squared approximation
## may be incorrect
```

```
# mat # L1 Statistic

# Categorical variables levels
# levels(charity$school_metro)
# levels(charity$primary_focus_area)
# levels(charity$resource_type)
# levels(charity$grade_level)
# levels(charity$teacher_prefix)

# Numerical Variables
# table(charity$AmountRequested)

# qplot(data = charity, AmountRequested, geom = "histogram", binwidth=0.1)

# relax matches
tab <- relax.cem(mat, charity, depth = 1, perc = 0.3)
```
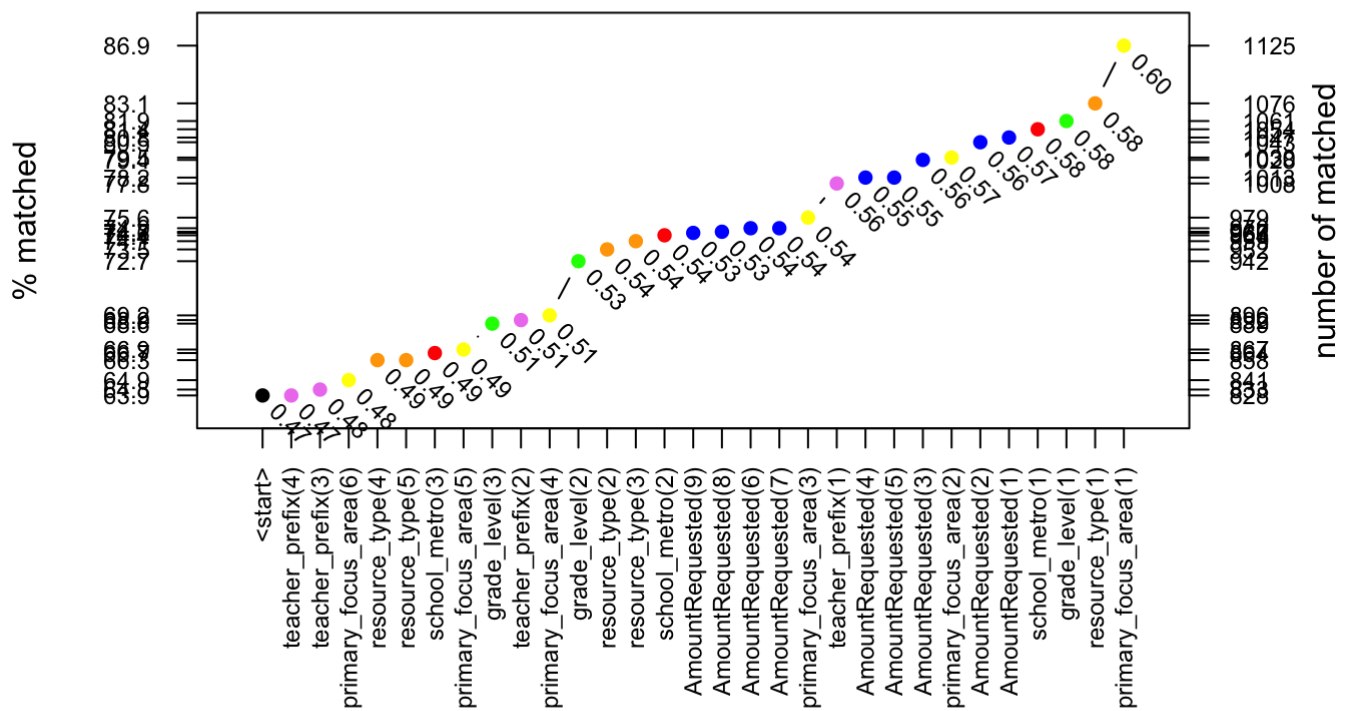
```
## Executing 30 different relaxations
##
   |
   |                                                               |    0%
   |
   |=============                                                  |   20%
   |
   |=========================                                     |   40%
   |
   |====================================                          |   60%
   |
   |==================================================            |   80%
   |
   |==============================================================| 100%
```

Pre-relax: 828 matched (63.9 %)


```
## SATT: Estimating the causal effect from cem output

# Appeal
att(mat, NumContributors ~ Status, data = charity, model="logit")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
##              G0    G1
## All        1295  1295
## Matched     851   828
## Unmatched   444   467
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 0.906094 (p.value=0.000157)
## 95% conf. interval: [0.436224, 1.375965]
```

```
# Momentum
att(mat, CovInterEventTime ~ Status, data = charity, model="logit")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
##                G0    G1
## All          1295  1295
## Matched       851   828
## Unmatched     444   467
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 0.819088 (p.value=0.000000)
## 95% conf. interval: [0.585535, 1.052641]
```

```
# Variation
att(mat, CovContributionAmount ~ Status, data = charity, model="logit")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
##                G0    G1
## All          1295  1295
## Matched       851   828
## Unmatched     444   467
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 0.513508 (p.value=0.000435)
## 95% conf. interval: [0.227428, 0.799588]
```

```
# Latency
att(mat, TimeToFirstContribution..sec. ~ Status, data = charity, model="logit")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
##                G0    G1
## All          1295  1295
## Matched       851   828
## Unmatched     444   467
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 0.048946 (p.value=0.624071)
## 95% conf. interval: [-0.146800, 0.244693]
```

```
# Engagement
att(mat, Duration..days. ~ Status, data = charity, model="logit")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
##              G0    G1
## All        1295 1295
## Matched     851   828
## Unmatched   444   467
##
## Logistic model on CEM matched data:
##
## SATT point estimate: -1.054396 (p.value=0.000000)
## 95% conf. interval: [-1.367365, -0.741426]
```

```
write.csv(mat$X, file = "./Data/cem_results.csv")
```