

TextEffect with R

[Code ▼](#)

Ryan Leung

[Hide](#)

```
library(glmnet)
library(texteffect)
```

[Hide](#)

```
library(readr)
prev_tf_matrix <- read_csv("C:/Users/aruba/Documents/SOCI40133/Homework-Notebooks-2024-Winter/week 7/prev_tf_matrix.csv")
```

```
New names:
* `` -> ...1
Rows: 4188 Columns: 1645
-- Column specification -----
Delimiter: ","
dbl (1645): ...1, 20, 21, 3rr, abide, ability, able, absolute...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

[Hide](#)

```
Y <- prev_tf_matrix$outcome_variable
X <- prev_tf_matrix[,-ncol(prev_tf_matrix)]
train.ind <- sample(1:nrow(X), size = 0.5*nrow(X), replace = FALSE)
```

[Hide](#)

```
# Fit an sIBP on the training data
sibp.search <- sibp_param_search(X, Y, K = 5, alphas = c(2,3,4), sigmasq.ns = c(0.6, 0.8, 1), iterations=1, train.ind = train.ind)
```

```
[1] 2
[1] 0.6
[1] 0.8
[1] 1
[1] 3
[1] 0.6
[1] 0.8
[1] 1
[1] 4
[1] 0.6
[1] 0.8
[1] 1
```

[Hide](#)

```
sibp_rank_runs(sibp.search, X, 10)
```

	alpha <dbl>	sigmasq.n <dbl>	iter <dbl>	exclu <dbl>
7	4	0.6	1	119.34659
9	4	1.0	1	106.12568
8	4	0.8	1	51.65482
6	3	1.0	1	43.73557
2	2	0.8	1	38.63475
1	2	0.6	1	35.34216
5	3	0.8	1	33.70035
4	3	0.6	1	25.67178
3	2	1.0	1	-71.89198
9 rows				

[Hide](#)

```
sibp_top_words(sibp.search[["3"]][["0.6"]][[1]], colnames(X), 10, verbose = TRUE)
```

```

[1] "Frequency of treatments: "
[1] 1236.0610 161.9882 119.9989 183.9891 106.0000
[1] "Relation between top words and treatments"
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.03901764 0.8618673 0.9815107 0.8420143 1.1391208
[2,] 0.03590121 0.8407664 0.9613068 0.7008482 1.1322244
[3,] 0.03126353 0.7865005 0.8210622 0.7002333 1.0845342
[4,] 0.02813294 0.7266219 0.8158471 0.6393465 1.0205249
[5,] 0.02657503 0.7235296 0.7707338 0.6335171 0.9862209
[6,] 0.02588912 0.7067829 0.7638483 0.6160248 0.9688858
      [,1]      [,2]      [,3]      [,4]
[1,] "blank"      "say"      "verifiable"    "subject"
[2,] "precede"     "source"    "term"          "article"
[3,] "to"          "way"       "check"         "think"
[4,] "yeah"        "original"  "bank"          "like"
[5,] "dead"        "time"      "economic"      "fact"
[6,] "wwe"         "week"      "usage"         "political"
[7,] "vandalize"   "find"      "historian"     "relationship"
[8,] "unsigned"    "possibly"  "spanish"       "have"
[9,] "match"       "reliable"  "mention"       "know"
[10,] "constitution" "look"      "understanding" "consider"
      [,5]
[1,] "people"
[2,] "make"
[3,] "article"
[4,] "example"
[5,] "point"
[6,] "scholarly"
[7,] "talk"
[8,] "university"
[9,] "correctly"
[10,] "faith"

```

[Hide](#)

```

# Plotting the AMCE Validation on the test set
sibp.fit_1 <- sibp.search[["3"]][["0.6"]][[1]]
amce<-sibp_amce(sibp.fit_1, X, Y)
amce

```

	x <int>	effect <dbl>	L <dbl>	U <dbl>
Intercept	1	0.041817426	0.01146535	0.072169506
Z1	2	-0.078050141	-0.12096864	-0.035131646
Z2	3	-0.118717455	-0.22795694	-0.009477969
Z3	4	0.046260859	-0.10512950	0.197651215
Z4	5	0.001494957	-0.12933710	0.132327014

	x <int>	effect <dbl>	L <dbl>	U <dbl>
Z5	6	-0.057419386	-0.31441898	0.199580208
6 rows				

Hide

```
sibp_amce_plot(amce)
```

