Course Sequence Analysis

Tim Ransom Randi Sims

2023-01-27

Contents

Chapter 1

midfield

1.1 Exploring the tables

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
##
## Attaching package: 'magrittr'
##
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

1.2 collection of utility functions

```
did_student_graduate <- function(mcid) {
  return(degree %>% filter(mcid == mcid) %>% nrow() > 0)
}
```

1.3 Pulling student course sequences

```
# convert to tibble
course <- tibble(course) %>% select(mcid, abbrev, number, term_course) %>% nest_by(mcid
```

1.4 Visualizing a student's course sequence

TODO: visualize a single students path to graduation

Chapter 2

modeling course sequences

We've already gotten our course sequences, lets use them to train a model!

2.1 Random forest model

```
library(caret)
```

```
## Loading required package: ggplot2
```

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

```
# https://topepo.github.io/caret/train-models-by-tag.html#random-forest
```

https://stackoverflow.com/questions/57939453/building-a-randomforest-with-caret

Chapter 3

model visualization

good science comm uses visuals