

LING 410X: Language as Data

Semester: Spring '18

Instructor: Sowmya Vajjala

Iowa State University, USA

8 March 2018

Class Outline

- ▶ Midterm feedback summary
- ▶ Question from Tuesday's class
- ▶ Discussion about your final project ideas
- ▶ Time for Assignment 4
- ▶ Reminder: Assignment 4 is due on 10th March 2018.

Mid-term feedback: summary -1

Delta for the course (for me)

- ▶ Meet in lab every time (Hmm, not in my control!)
- ▶ More help sessions like on 2/23 (I can try for one more)
- ▶ Windows computers and errors (Hmm!)
- ▶ Spend more time explaining concepts, discuss specifics of some functions (Okay, I will do this more often)

Mid-term feedback: summary -2

Delta for the student (for you)

- ▶ Practice more
- ▶ Participate more
- ▶ Read the book/read more about R functions

Last class' question

do.call

- ▶ It takes two arguments - a function (any function), and a list.
- ▶ What it does is - it applies that function on each element of a list, and returns output in the form of the first function
- ▶ Example 1: `list1 <- list(11,12,13,14)` and say:
 - (a) `do.call(cbind, list1)`
 - (b) `cbind(cbind(list1[[1]]), cbind(list1[[2]]), cbind(list1[[3]]), cbind(list1[[4]]))`(a) and (b) do the same thing.

do.call

- ▶ It takes two arguments - a function (any function), and a list.
- ▶ What it does is - it applies that function on each element of a list, and returns output in the form of the first function
- ▶ Example 1: `list1 <- list(11,12,13,14)` and say:
 - (a) `do.call(cbind, list1)`
 - (b) `cbind(cbind(list1[[1]]),cbind(list1[[2]]),cbind(list1[[3]]), cbind(list1[[4]]))`(a) and (b) do the same thing.
- ▶ It applies `cbind` on each element of the list, and binds all these `cbind` outputs together.
- ▶ This in essence becomes `rbind(list1)` for such a simple case.

do.call

The real use

- ▶ Let us say we have a embedded list:
`list2 = list(list(1,2,3), list(4,5,6), list(7,8,9))`
- ▶ Using `cbind` here:
(c) `do.call(cbind,list2)`
(d) `cbind(cbind(list2[[1]]),cbind(list2[[2]]),cbind(list2[[3]]))`
- ▶ (c) and (d) do the same thing, and gives us a matrix:
1 4 7
2 5 8
3 6 9

Discussion about final projects

Each person/team: please share with others what ideas you have for your final project addressing the following questions:

- ▶ what is the corpus you have in mind?
- ▶ how big is it? (how many texts/examples etc)
- ▶ Do you know how to get it in a format R can understand?
- ▶ what are the questions you want to address with that dataset?
- ▶ how will you address these questions?
- ▶ how would you evaluate what you did?

Discussion about final projects

Each person/team: please share with others what ideas you have for your final project addressing the following questions:

- ▶ what is the corpus you have in mind?
- ▶ how big is it? (how many texts/examples etc)
- ▶ Do you know how to get it in a format R can understand?
- ▶ what are the questions you want to address with that dataset?
- ▶ how will you address these questions?
- ▶ how would you evaluate what you did?

Remember: initial report is due on 7th April!

Assignment 4 practice & discussion

Working with csv files that have text columns

- ▶ What is a csv file?
- ▶ While reading in a csv file:

```
data_train <- read.csv("a4-train.csv",header = TRUE)
```

What is header = TRUE mean?

Working with csv files that have text columns

- ▶ What is a csv file?
- ▶ While reading in a csv file:

```
data_train <- read.csv("a4-train.csv",header = TRUE)
```

What is header = TRUE mean?

- ▶ How do I access individual columns in this?

Working with csv files that have text columns

- ▶ What is a csv file?
- ▶ While reading in a csv file:

```
data_train <- read.csv("a4-train.csv",header = TRUE)
```

What is header = TRUE mean?

- ▶ How do I access individual columns in this?
- ▶ So, how do I access only text column?

Working with csv files that have text columns

- ▶ What is a csv file?
- ▶ While reading in a csv file:

```
data_train <- read.csv("a4-train.csv",header = TRUE)
```

What is header = TRUE mean?

- ▶ How do I access individual columns in this?
- ▶ So, how do I access only text column?
- ▶ Let us say I do something like:

```
mega <- paste(mydata["text"], collapse = "")
```

What happens? How do I get answers for Q1 in A4?

Working with csv files that have text columns

- ▶ What is a csv file?

- ▶ While reading in a csv file:

```
data_train <- read.csv("a4-train.csv",header = TRUE)
```

What is header = TRUE mean?

- ▶ How do I access individual columns in this?

- ▶ So, how do I access only text column?

- ▶ Let us say I do something like:

```
mega <- paste(mydata["text"], collapse = "")
```

What happens? How do I get answers for Q1 in A4?

- ▶ stringsAsFactors=FALSE argument in read.csv

Remaining time

Assignment 4 practice & any other questions you have

Happy Spring break!!

Topic for the week after break: Topic Models.