Text Translation, Classification Writing Shiny apps

Session # 5

Text Analytics for Batch 12

CBA @ ISB

Sudhir Voleti

1

Session Plan

- Writing a shiny app
  - Cluster-an via kmeans
- Language Translation in Py
  - Textblob+Translation
- Intro to Supervised Text Classification in Py
- Text Summarization in R
- Course Wrap-up
  - Summary and Recap

2

2

Writing a simple shiny app

3

Writing a shiny app

- Based on popular demand, I present to you a small workshop on writing a simple shiny app.
- Recall the kmeans cluster-an exercise we did previously? Let's now code a small shiny app that:
  - to take a matrix as input,
  - run kmeans on it,
  - select the optimal number of clusters from a scree plot,
  - view a summary table of cluster means, and
  - view the resulting clustering assignment.
- Open file '*shiny demo*'. In parallel, open *ui.R* and click on 'Run App'.
- How easy or difficult was it? Thoughts? Comments?

4

4

Language Translation

Using Textblob in Py

5

Language Detection & Translation

- Open file '*language detection n translation.ipynb*' in Jupyter
- Install textblob and execute the notebook's ops.
- Now answer the following Qs:
- 1. What modules and main funcs did we invoke?
- 2. Describe any current use-case in your org where you could use this.
- 3. Implications? Extensions? Possibilities? Limitations?

Open file 'Regular NLP in textblob'in Jupyter

6

6

Other textblob functionality in NLP

- Open file '*Regular NLP in textblob.ipynb*' in Jupyter
- Execute the nb and answer the following Qs:
- 1. What modules and main funcs were salient and noteworthy?
- 2. What is 'subjectivity' assessment? Did it work well?
- 3. How well does spellcheck do its job?
- 4. Describe any current use-case in your org where you could use this.

7

7

Introduction to Supervised

Text Classification

8

A Simple Example with Sentiment-An in Text

- Evaluate the 'overall sentiment' of these tweets:

- 1. "The Da Vinci Code book is just awesome."
- 2. "da vinci code was a terrible movie."
- 3. "I liked the Da Vinci Code but it ultimately didn't seem to hold it's own."

- Analyze HOW you arrived at your conclusions → What variables/ features/ keywords triggered a decision from you?

- Can we 'teach' a machine to learn similarly…
- … *without* having to *program* explicitly every exception and every contingency that might arise?

9

9

How Does Supervised Learning Work?

- User presents pre-coded / classified data to the machine.

- Learning algorithm creates a "behavioral model", and adjusts behavior given function parameters.
  - Starts first with averages: [On the average] A means the code is +1.
  - Then adds interactions: [On the average] A and B imply a stronger probability of +1.
  - Then adds exceptions: However, if A, B and C are present, then it is likely to be 0.
  - And more combinations of these internal decision rules …
  - Think of func parameters as a set of [regression] weights or coefficients.

- Software then classifies data the computer has never seen ('holdout data')
  - Think of this as a regression prediction made about Y based on Xs.

10

10

How Does Supervised Learning Work? Example

- Recall our dummy example. Suppose they're pre-coded as below:

- 1. "The Da Vinci Code book is just awesome." [+1]
- 2. "da vinci code was a terrible movie." [0]
- 3. "I liked the Da Vinci Code but it ultimately didn't seem to hold it's own." [+1]

- Machine will start by inferring that "da vinci code", "book", "just", "awesome" etc are all predictors for [+1] after statement 1.

- Then, after statement 2, will discard/ de-weigh "da vinci code" as a predictor; pickup "terrible", "movie" as predictors for [0].

- And so on.

Open file 'text_classification_Amazon_Reviews'

11

11

Recap: Supervised Text Classification Example

- What did we do just now?

- What main libraries and functions did we use?

- What are the main inputs to and outputs from this supervised classification exercise?

- Extension, improvement, combination possibilities with other things we've learned so far?

12

12

Text Summarization

13

Text Summarization in R and Py

- What is 'text summarization'? Why care about it?
- Open file '*textrank summzrizn.Rmd*' and run it.
- What inputs and outputs did you see? What libraries and modules did we invoke?
- How does textrank method choose key sentences?
- The *sumy* module in py provides a variety of text summarization methods, not just textrank.

14

14

Course Wrap-up

15

Recap: Course Plan - I

- Session 1: Introduction to Text Analysis
  - Regex Primer, Tokenization procedures
  - Simple Bag-of-words (BOW) representation
  - Token-Document Matrix (TDM) structures
  - Intro to *tidytext, NLTK* in Py
  - *Stringdist* introduction
- Session 2: Basic Sentiment Analysis
  - Recap and Exercise
  - Intro to Sentiment-an
  - Sentiment scoring schemes in R (tidytext)
  - Sentiment-An in Py
  - Cluster-an with text data (stringdist and k-means)

16

16

Recap: Course Plan - II

- Session 3: Vector-Space modeling
  - Factor-An primer
  - Latent Topic Modeling (LTM) primer and Simulation
  - Latent Dirichlet Allocations (LDA) notes
  - LDAtuning to find optimal #topics
- Session 4: Basic Natural Language Processing (NLP)
  - Word and sentence annotations (*Spacy*)
  - Parts of Speech (POS) tagging, chunking (*Spacy*)
  - Named Entity Recognition (NER) with *Spacy*
  - Chunking and Phrase Extractions (*UDpipe*)
- Session 5: Text Translation, Classification, miscell
  - Elementary text classifiers, Examples, Text summarization
  - Course summary, recap, wrap-up.

17

17

Goodbye and Goodluck.

Shukriya aur Alvida

18