
LDA - Practical Exercise

04.10.2019

Task 1: Data

(30 Minutes)

1. Clone the repository from <https://github.com/kramerlab/CECAM>.
2. Have a look at the dataset in `rev_full.txt`.
3. Order the data by date.
4. Install gensim: `conda install -c conda-forge gensim`
5. Have a look at the gensim API to load the data and remove stop words <https://radimrehurek.com/gensim/apiref.html>.

Task 2: LDA

(30 Minutes)

1. Use gensims `ldamodel` <https://radimrehurek.com/gensim/models/ldamodel.html> and train it using the data.
2. Print out the topics, also evaluate the log perplexity during training.
 - (a) Try different numbers of topics.
 - (b) Try different hyperparameters.
 - (c) Try different numbers of iterations (100,500,1000,...).

Task 3: Online LDA

(60 Minutes)

1. Use the gensim `ldamodel` to implement the online LDA from the presentation:
 - (a) Divide the data into time slices.
 - (b) Train a separate model on each time slice.
 - (c) For each model use the parameters of the previous model as a prior. For the first model use a symmetric prior (0.01).
2. Run it on the data and print out the topics to see how each topic evolves over time.
 - (a) Try differently sized time slices.
 - (b) Try different hyperparameters.
 - (c) Try different numbers of iterations (1,10,50,...).