

DATA ANALYSIS WITH TOPIC MODELS FOR COMMUNICATIONS RESEARCHERS



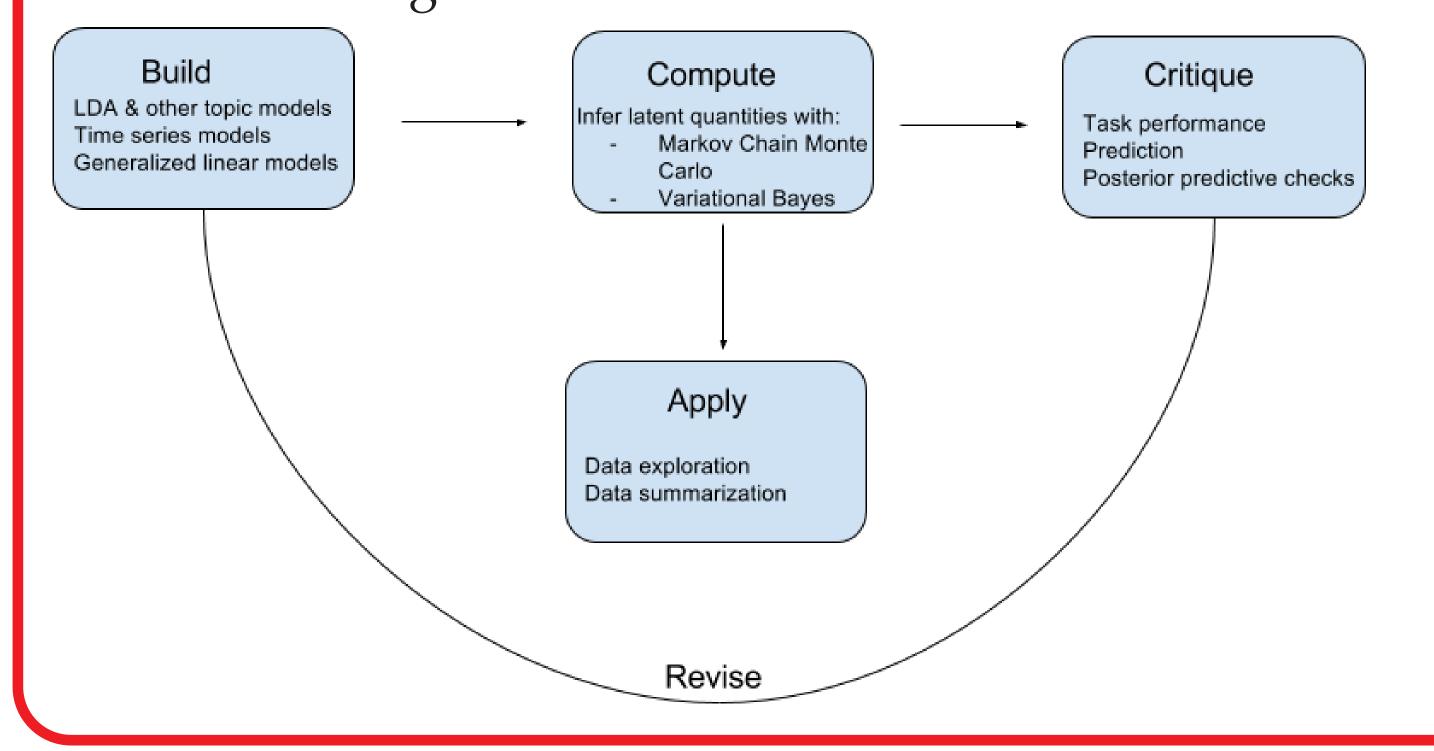
Frederick Boehm (fred.boehm@wisc.edu), Department of Statistics, University of Wisconsin-Madison

ABSTRACT

We introduce topic modeling as a tool when analyzing textual data. We illustrate our methods with analyses of New York Times transcripts and tweets from three days in March 2016. We argue that such analyses will be useful in mass communications and journalism research. These methods are especially useful for identifying topics, or themes, in large collections of texts, when reading each piece individually is impractical.

INTRODUCTION

Social media users flood us with tweets, status updates, and blog posts. Data analysis with topic models enable researchers to identify themes, or topics, in a collection of texts. We present results from separate analyses of 1) New York Times articles from March 19, 20, and 21, 2016 and 2) a collection of tweets (from Twitter) from the same three days. Our manuscript contains computing code to reproduce our findings.



METHODS

- 1. Fit latent Dirichlet allocation (LDA) models, separately, to 1) our collection of tweets and 2) our collection of print media articles from the New York Times
- 2. Visualized the topic modeling results with word clouds
- 3. Assessed resulting topics for coherence

RESULTS

album tonight house official fly star tonight radio fans rihanna de drake bieber today added band love march listen di priceless songft tour justinplay list watch live night np la stage reaction book playing soundcloud

Figure 1: Word cloud for one topic (from a 20-topic model) of tweets from March 19, 20, and 21, 2016.

demichelis basketball goals games home
years mufc back chelsea
day top live league year
won fans state
season st win half iowa good
final play games home
time good
time good
leicester cityman west
football great team today players
player points rashford
marchmadness
manchester premier

Figure 2: Word cloud for one topic (from a 20-topic model) of tweets from March 19, 20, and 21, 2016.

broadway company
plays ballet song movies
songs theater story
festival life movie ilmplay season
role
york musical
love director show musical
shows set show musical
shows set show drama album
bowie played drama album
dance record
performance altman
series

Figure 3: Word cloud for one topic (from a 20-topic model) of New York Times articles from March 19, 20, and 21, 2016.

norwegian crew franklin
coast oceanislands pacific
snow arcticship inuit
small baalsrud food
found baalsrud food
kelen Water hawaii william
miles island cruise
windfeet men ships lake
canoe genz huth di boatrivertrip joel
expedition frozen
baalsruds cabin

Figure 4: Word cloud for one topic (from a 20-topic model) of New York Times articles from March 19, 20, and 21, 2016.

DISCUSSION

REFERENCES

[1] David M Blei. Build, compute, critique, repeat: Data analysis with latent variable models. *Annual Review of Statistics and Its Application*, 1:203–232, 2014.