Grimmer et al.
Fit model and apply topically…

Will Lowe

MZES, University of Mannheim

Topic models

Topic models

    assume documents are *admixtures of topics* (Blei et al.)

    offer a *probability model* for traditional content analysis

except that in this application

    "Even though press releases are usually written with one focused point, we model each document as a mixture of multiple topics."

because

    we capture the "asides [and] nuances"

    if *not,* then common words would "make topics look less distinct"

Rephrase

Would make topics look less distinct *than they are,* so

Q What *are* topics?

A1 *Not* spatial or personal

> "We also remove representative names and major cities [. . .] since we do not want the clustering portion of the model [...] to cluster the documents according to these representative-specific topics."

A2 *Maybe* temporal

> "The temporal dynamics [. . .] also correspond to real events"

Rephrase

Q What *are* topics?

A3 Whatever makes word counts independently distributed after you condition on it in the generative structure

A4 Whatever partitions experts *like*
(think internally similar, find distinguishable, would label similarly)

The dilemma: statistical horn

"statistical guidance alone can cause researchers to select models that are substantively sub-optimal"

"Chang et al. (2009) show that a model's performance on predictive tasks is often negatively correlated with a model's substantive fit."

Topic models "predict well the contents of new documents [. . .] But *social scientists are rarely interested in prediction*"[1]

Wallach: different priors, different consequences

_____

[1] "This is not to say that social scientists are not interested in prediction"

The dilemma: substantive horn

"social scientists *do not have strong prior beliefs* over this structure of political debate"

We do have some *candidate concepts*: the agenda, the secondary policy preference dimensions, talking points, frame markers?

Fortunately we know a good topic when we see it. . .

"Human evaluators may not be able to replicate topics before hand, but evaluate models highly after the fact."

So what are these human annotators *doing* exactly?

And are they reliable at it?

What are raters doing?

"asking humans to form impressions about the quality of topics [may] *inject error* into the model selection process."

What is error here?

"Asking a user to repeat this task [...] places a *significant cognitive burden* [...] This is likely to induce *satisficing* as researchers focus on only a few topics or fail to identify inconsistencies in their model"

Model inconsistency? Is satisficing an error or a fact about categorization?

- You have 10 (20,60) minutes. Find the topics!
- Here are some words, can you identify the category?

The psychology of categorization

Ad hoc category:

'things you would grab in the event of a fire in your house'

Engagement with the (rational) psychology literature on catgorization, e.g.

Griffiths et al. (2007) Unifying rational models of categorization via the hierarchical Dirichlet process'

(Anderson, Chater and Tenenbaum meet Blei, Jordan and Mimo)

Results and pictures

In the absence of clearly interpretable evaluation criteria, model *visualisation* is vital

Cluster proportion data, or count samples from it, are a reduced representation of the full word count matrices, but may have reduced rank spatial representations themselves

- correspondence analysis of topics
- canonical correlation analysis with Nominate scores etc. as covariate

to *show* the sorts of results in Table 4,

and put items and speakers on the same plot

Polarization

What could "artificial polarization" be?

We *know* there is

    selection bias on roll-call votes

    selection bias on distribution of topics spoken and written about

but what is the 'unbiased' topic choice distribution relative to which yours is artifically polarized?

Finally

"Social scientists do not have strong prior beliefs over [the] structure of political debate"

So we ought to get some...

Happiness is never having to say

"performance on predictive tasks is often negatively correlated with a model's substantive fit"

Discussion of

Burton: Assessing gender differences with automated and semi-automated text analytics.

Gendered Language

Looking for "gender-specific characteristics in language *beyond* gender preferences in topics and genre"

so, we expect gender effects at two levels:

> choice of *topic and genre*

> choice of *words,* conditional on topic and genre

The theory:

> "much of politics is driven by natural language"

> "words operate as framing devices"

The data

Articles from Campaigns and Elections magazine, 'text of books', and 'an edited volume'

    US edition of C&E?

    sampled or filtered?

    which book and volume?

Chosen because

    "the putative goal of writing – in this case, winning elections – would seem gender-neutral"

Conditioning on *topic* is a very good idea (does this do it?). And don't forget *genre…*

Operationalisation

Aggregating texts by the same author probably biases the result (unmeasured confounder: topic and genre)

Would you write an article and a C&E piece the same way?

also unnecessary.

Working with word rates directly can be problematic

very small numbers and variance rises with magnitude

Working with different size categories (Appendix 2) causes similar problems

Model assumptions

Some of your models, e.g. C4.5, are inherently interactive but your statistics, e.g. the T-tests in Table 1 are not

Theory says sex/gender causes lexical variation, so why not fit the models that way? e.g.

Poisson / multinomial regressions of (category or word) counts on sex

would give you theoretically relevant quantities of interest.

Unlike scaling and topic models the variable of interest (sex) is observed

Rethinking Appendix 1

Make a picture, and sort by gender bias on the word

Turn

Word	Sex	Rate
Woman	F	0.0092036531
	M	0.0032312909

into

1.048 $\approx$ 2.8 times more likely to be generated by female
(0 equally balanced)

Looking at the data: KWIC