

政治学におけるテキスト分析の動向

機械学習、多言語化

September 27, 2019

Akitaka Matsuo (松尾 晃孝)

議会関連のテキストを中心として、政治学でのテキスト分析の応用を目的に応じてレビューしていく。

- ▶ Scaling メソッド
- ▶ Topicmodels
- ▶ 新しい展開
 - ▶ 機械学習の適用
 - ▶ クラウドワーカーの利用
 - ▶ 多言語への展開

Scaling から始まった

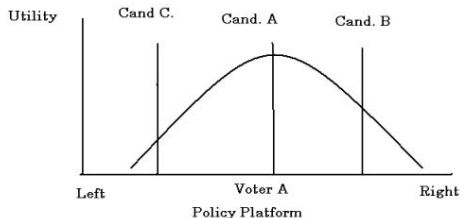
2000 年代のテキスト分析は、文章を低次元の空間に落とし込む Scaling が主流。具体的には、テキストから、政策空間におけるポジションや、イデオロギーポジションを推定する。

代表例

- ▶ Wordscores (Laver, Benoit and Garry, 2003)
- ▶ Wordfish (Slapin and Proksch, 2008)

なぜ?

- ▶ 空間投票理論との親和性
 - ▶ アクターのポジションさえ分かれば、適用可能な理論の蓄積があった
 - ▶ 点呼投票のある議会では、測定のメソッドが確立していた



政治学のコンピュータテキスト分析は、2000 年代から流行しですが、これは、Wordscores (Laver, Benoit and Garry, 2003) の開発によるところが多い。

- ▶ もともとは、Comparative Manifesto Project (CMP) のコーディングの信頼性に対する不満が開発の動機
- ▶ **セッティング**
 - ▶ Reference text を用意する（ポジションが既知かつ、極端なアクター、そして、そのポジションを代表するような文書）。
 - ▶ Reference text の単語使用の比較から、それぞれの単語のスコアを計算、他のアクターのポジションも、スコアの平均として計算
- ▶ コンピュータ・サイエンスへの言及はないものの、教師つき学習モデルとして考えられる

Wordscores の特徴

- ▶ Reference text に存在しない単語は分析から外れる
 - ▶ 短所: 他のアクターのポジションは中心によりがち (Lowe, 2008) (c.f. Martin and Vanberg (2008) による修正)
 - ▶ 長所: 文脈を固定しやすい
- ▶ ポジションが既知のアクターのいない場面では使えない

Wordscores の、このような特徴を踏まえつつ、Reference を必要としない、Wordfish (Slapin and Proksch, 2008) が開発される。

- ▶ 文書特徴行列のみを入力として、ポジションを一次元に落とし込む
- ▶ 単語のスコアと、ポジションを同時に推定
 - ▶ 点呼投票からポジションを推定するメソッド (W-Nominate (Poole and Rosenthal, 2000), Ideal (Clinton, Jackman and Rivers, 2004)) と類似。
- ▶ Computer Science への言及。教師なし学習としての位置づけ

Wordfish の特徴

- ▶ 教師なしなので、軸が何を意味するかは出てきた単語のリストによる (e.g. Curini, Hino and Osaka, 2018)
- ▶ トピックがバラバラなコーパスでは変な結果が出やすい (c.f. Wordshoal Lauderdale and Herzog (2016) による拡張)

Scaling が大きな関心を集めていたとはいえ、それ以外の手法に対する関心の高まりから、他のモデルも使われ始める。特に広く使われているのは topicmodels.

基本的なアイデア

- ▶ 文書のコーパスを、いくつかのトピックに分類する
- ▶ 教師なし学習モデルを使うことが多いが、教師つきモデルも存在する（ただしこの場合、普通の分類モデルとあまり変わらない）

トピックモデルはさまざまなヴァリエーションがあるが、基本は Latent Dirichlet Allocation (LDA) モデル (Blei, Ng and Jordan, 2012)

LDA モデルの特徴

- ▶ それぞれの文書をトピックの Dirichlet 分布だと考える
- ▶ 例えば、[トピック 1: 0.3, トピック 2: 0.3, トピック 3: 0.4]
- ▶ 各単語のスロットごとに、このトピック分布からトピックをまずドロウする
- ▶ トピックは単語の multinomial 分布なので、この分布から、単語をドロウする

LDA モデルを使うときの実際の手順

- ▶ トピックの数 k を決める
- ▶ 推定する
- ▶ 各トピックが何を言っているのかを決める
 - ▶ トピックで特徴的な単語を眺める (highest probability, frequency, exclusivity など)
 - ▶ 各トピックで高確率の文書を読む

Catalinac, Amy. 2016. "From Pork to Policy: The Rise of Programmatic Campaigning in Japanese Elections." *The Journal of Politics* 78(1):1–18.

- ▶ Model: LDA with 69 topics
 - ▶ Why?: "We fit the model with 69 topics because this was one of the lowest specifications that produced topics that were fine-grained enough to resemble our quantities of interest."

Moser, Scott and Andrew Reeves. 2014. "Taking the Leap: Voting, Rhetoric, and the Determinants of Electoral Reform." *Legislative Studies Quarterly* 39(4):467–502.

- ▶ Model: LDA with 10 topics
 - ▶ Why? "The 10-topic model specification presents the best fit for the data based on held-out perplexity."

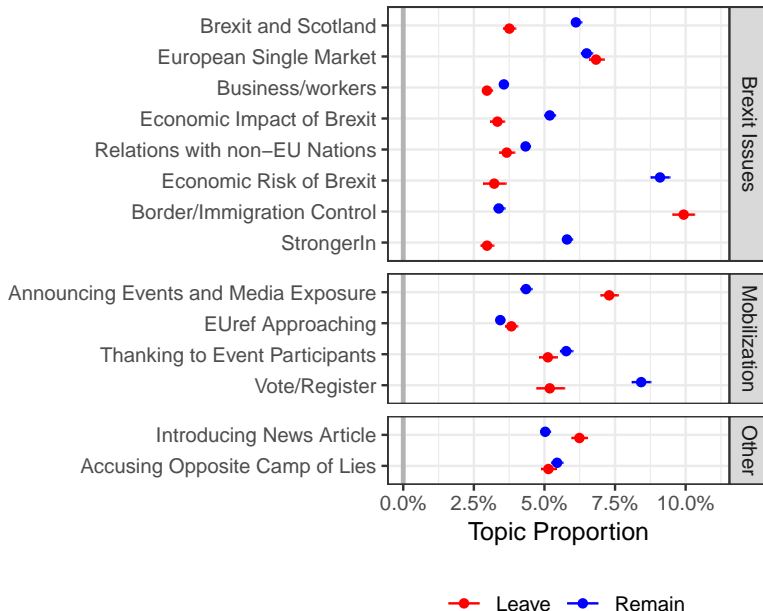
STM (Roberts, Stewart and Tingley, 2015; Roberts et al., 2014) は、トピックや単語のドローでは LDA にてているが、トピックに関連する変数を導入することで、説明という目的により適したモデル。

STM の特徴

- ▶ トピック間の相関を認める
- ▶ トピックの分布に影響を与える変数を使える (e.g. 時間、属性、性別)
- ▶ stm package が使いやすい (searchK, plot.stm)

昨年 of Polmeth Software Award 受賞。広く使われている (54 本の出版されたペーパー、<https://www.structuraltopicmodel.com/>)。Poltext カンファレンスでもかなりの数。

STM の使用例: Matsuo and Benoit (2019)



なぜこんなに流行っているのか？（特に STM）

- ▶ トピックの解釈により豊かなストーリーを作れる
- ▶ R パッケージで簡単に分析 (quanteda + stm/topicmodels)
- ▶ 政治学者は新しいもの好き

難しさ？

- ▶ トピックの解釈は、研究者次第 (Chang et al., 2009)
- ▶ トピックの数を一意に決める方法は存在しない
- ▶ モデルのアウトプットは初期値により、推定の度に大きく異なる
 - ▶ 結果の再現性の問題
 - ▶ model shopping の余地？
 - ▶ 推定する度に、トピックの内容をチェックしないといけない
 - ▶ searchK も例外ではない
- ▶ 解釈できないトピックをどうするのか？
 - ▶ 大きいkを使って少ししか使わないのは正当か？
- ▶ 必要ない場面で使いがち？(後で触れる)

近年の研究はコンピュータ・サイエンスの発展を取り入れたものが多い

- ▶ Machine Learning 特に classification method の利用
- ▶ クラウドワーカーを利用して Ground Truth データを作成して利用する
- ▶ 多言語比較

かなりざっくりとした、政治学との典型的な違い

	政治学	機械学習
目的	説明	予測
学習モデル	教師なし	教師つき
出力	色々（連続変数也多）	カテゴリー

- ▶ 典型的には、機械学習では、すでにカテゴリーがラベルされたデータが有り、それを元に、いかに予測の精度を上げていくのかという問題に取り組む
 - ▶ 変数の入れ替え
 - ▶ 新たな手法の開発
- ▶ 過学習の問題は cross validation で評価することにより回避

政治学もそういう方向に進むのか？

⇒ 恐らくはそうはならない

Peterson, Andrew and Arthur Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis* 26(01):120–128.

- ▶ 予測エラーを分極化の指標として利用するというアイディア
- ▶ 議会でのスピーチの内容を使って、発言者の政党所属を予測する
- ▶ 分極化が進むほど予測がしやすくなっていく

Slapin, Jonathan B. and Justin H. Kirkland. 2019. "The Sound of Rebellion: Voting Dissent and Legislative Speech in the UK House of Commons." *Legislative Studies Quarterly* (June):lsq.12251.

- ▶ 発言の内容から造反を予測できるのか
- ▶ 第一人称、単純なスピーチがある程度は造反を予測する

これらの研究は、ある程度、機械学習の作法に則っているが、予測の精度を上げること自体には主たる関心がない。それよりは、予測ができるかどうか、できるのであれば何がそれを決めるのか、という方向。これが政治学において、機械学習を利用した研究として価値のあるものと見られる可能性かもしれない。

クラウドワーカーを利用すること自体は、CS とは直接関係がないし、社会科学における利用は、サーベイ実験の利用しやすい被験者としての利用が中心 (e.g. Coppock, 2018; Berinsky, Huber and Lenz, 2012)。しかしそもそもは、機械学習用の学習データを Amazon が安く調達する目的で Amazon Mechanical Turk (MTurk) は作られた。

代表的なサービス

- ▶ MTurk
- ▶ Figure Eight (Crowdfunder)
- ▶ CrowdWorks (日本)

クラウドワーカーを利用してテキストの人的コーディングを行い、そこから何らかの知見を得るという研究も、それなりに見られる。機械学習の学習データとして使ったり、テキスト分析の validation のための ground truth として使ったりする (c.f. King, Lam and Roberts, 2017)。

Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver and Slava Mikhaylov. 2016. "Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data." *American Political Science Review* 110(02):278–295.

- ▶ Crowd Task: イギリスの選挙マニフェストを分単位に分割、文章を読ませ、政党のポジションを訊く
- ▶ CMP によるコーディングを再現できる
- ▶ 研究プロセスの透明性を確保できる

Benoit, Kenneth, Kevin Munger and Arthur Spirling. 2019. "Measuring and Explaining Political Sophistication through Textual Complexity." *American Journal of Political Science* 63(2):491–508.

- ▶ 政治的言説における「複雑性」を担保するものはなにか、という研究
- ▶ Crowd Task: 2つの短い言説を見せ、それを読ませどちらが複雑か訊く
- ▶ これを ground truth として、どのような要素が複雑だ、という判断につながるのかを検証

テキスト分析を比較政治学の一般的なテーマの研究に応用しようと考えるときに壁となるのが、異なる言語で書かれたテキストをどう分析するのか、という問題。何らかの形でブリッジしないといけないが、それにアプローチする研究も、盛んに行われている。

- ▶ 機械翻訳の利用の可能性 (Lucas et al., 2015)
- ▶ 公式の翻訳があるコーパスを利用する (e.g. 欧州議会)
- ▶ これがない場合には、基本的には、外部データを使った validation が必要になってくる
 - ▶ 専門家サーベイとの比較
 - ▶ 人的コーディング
 - ▶ 専門家
 - ▶ クラウド

もし、多言語の比較分析を自ら行いたいとなったときに、研究手順の参考になるのは以下。

de Vries, Erik, Martijn Schoonvelde and Gijs Schumacher. 2018. "No Longer Lost in Translation. Evidence that Google Translate Works for Comparative Bag-of-Words Text Applications." *Political Analysis* pp. 1–31.

- ▶ 公式の翻訳があるコーパスを利用した、方法論的研究
- ▶ europarl データセットを利用、一部を機械翻訳
- ▶ 機械翻訳したテキストでも、topicmodel の結果をかなり再現できる

Proksch, Sven Oliver, Will Lowe, Jens Wäckerle and Stuart Soroka. 2018. "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches." *Legislative Studies Quarterly* (September):1–35.

- ▶ 議会演説のセンチメントの比較研究
- ▶ 辞書ベースのセンチメント分析で、辞書を機械翻訳して、各国のテキストに適用。内閣提出法案に対する与野党議員のセンチメントの違いを分析。
 - ▶ 機械翻訳前後に、かなりの手間をかける（辞書の拡大、人的チェック）
- ▶ validation には europarl を利用（EUP, the State of the Union Debate）

最後に: テキスト分析をするときの一般的な Tips

- ▶ 文章が生成された経緯に関する知識の重要性。文脈に関する知識が研究の手段を変える
 - ▶ メソッドの選択 (新聞記事にトピックモデルは必要だろうか?)
 - ▶ 出てくる結果の解釈 (e.g. Wordfish のポジション推定)
 - ▶ 軸の意味は?
 - ▶ ポジションは発話者の真の意図を反映しているだろうか?
- ▶ テキストのメタデータの重要性
 - ▶ 議会のアジェンダ
 - ▶ ツイッターアカウントのユーザー情報 (e.g. スケーリングの方法としては、text も使えるけど、followers network (Barberá, 2015) も使える)
- ▶ コーパスのサイズ、各文書のサイズで使うメソッドを変える。例えば、
 - ▶ コーパスのサイズが相当大きくなないと、独自の word embedding モデルはあまり意味がない
 - ▶ 短い文書に関して、いくつかのトピックが想定できるのか考える
 - ▶ 辞書ベースの分析?

- Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23:76–91.
- Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver and Slava Mikhaylov. 2016. "Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data." *American Political Science Review* 110(02):278–295.
- Benoit, Kenneth, Kevin Munger and Arthur Spirling. 2019. "Measuring and Explaining Political Sophistication through Textual Complexity." *American Journal of Political Science* 63(2):491–508.
- Berinsky, Adam J., Gregory A. Huber and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(3):351–368.
- Blei, David M, Andrew Y Ng and Michael I Jordan. 2012. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3(4-5):993–1022.
- Catalinac, Amy. 2016. "From Pork to Policy: The Rise of Programmatic Campaigning in Japanese Elections." *The Journal of Politics* 78(1):1–18.
- Chang, Jonathan, Sean Gerrish, Chong Wang and David M Blei. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models." *Advances in Neural Information Processing Systems* 22 pp. 288—296.
- Clinton, Joshua D., Simon Jackman and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *The American Political Science Review* 98(2):355–370.
- Coppock, Alexander. 2018. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* (2015):1–16.
- Curini, Luigi, Airo Hino and Atsushi Osaka. 2018. "The Intensity of Government-Opposition Divide as Measured through Legislative Speeches and What We Can Learn from It: Analyses of Japanese Parliamentary Debates, 1953-2013." *Government and Opposition* pp. 1–18.
- de Vries, Erik, Martijn Schoonvelde and Gijs Schumacher. 2018. "No Longer Lost in Translation. Evidence that Google Translate Works for Comparative Bag-of-Words Text Applications." *Political Analysis* pp. 1–31.
- King, Gary, Patrick Lam and Margaret E. Roberts. 2017. "Computer-Assisted Keyword and Document Set Discovery from Unstructured Text." *American Journal of Political Science* pp. 1–17.

- Lauderdale, Benjamin E. and Alexander Herzog. 2016. "Measuring Political Positions from Legislative Speech." *Political Analysis* 24(03):374–394.
- Laver, Michael, Kenneth Benoit and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97(02):311–331.
- Lowe, Will. 2008. "Understanding Wordscores." *Political Analysis* 16:356–371.
- Lucas, C., R. a. Nielsen, Margaret E. Roberts, B. M. Stewart, A. Storer and D. Tingley. 2015. "Computer-Assisted Text Analysis for Comparative Politics." *Political Analysis* pp. 254–277.
- Martin, Lanny W. and Georg Vanberg. 2008. "A robust transformation procedure for interpreting political text." *Political Analysis* 16(1):93–100.
- Matsuo, Akitaka and Kenneth Benoit. 2019. "Political Messaging and Public Reception on Social Media during the Brexit Referendum ☒".
- Moser, Scott and Andrew Reeves. 2014. "Taking the Leap: Voting, Rhetoric, and the Determinants of Electoral Reform." *Legislative Studies Quarterly* 39(4):467–502.
- Peterson, Andrew and Arthur Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis* 26(01):120–128.
- Poole, Keith T. and Howard Rosenthal. 2000. *Congress : A Political-Economic History of Roll Call Voting*. Oxford University Press.
- Proksch, Sven Oliver, Will Lowe, Jens Wäckerle and Stuart Soroka. 2018. "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches." *Legislative Studies Quarterly* (September):1–35.
- Roberts, Margaret E., Brandon M. Stewart and Dustin Tingley. 2015. "stm: R Package for Structural Topic Models." *Journal of Statistical Software* VV.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082.
- Slapin, Jonathan B. and Justin H. Kirkland. 2019. "The Sound of Rebellion: Voting Dissent and Legislative Speech in the UK House of Commons." *Legislative Studies Quarterly* (June):lsq.12251.
- Slapin, Jonathan B. and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52(3):705–722.