**Prior applications of topic models: Examples from the social sciences and law**

One can find many papers in the literature that analyze textual information successfully. Being successful means that these analyses gain important insights on important substantive issues – in diverse fields such as economics, business, sociology and the law. The study of media slant by Gentzkow and Shapiro (2010), the studies mentioned in the review paper by Gentzkow, Kelly and Taddy (2017), and the contributions to textual analysis in sociology and political science by Grimmer and Stewart (2013) and Wilkerson and Casas (2016) are excellent examples.

Many applications of topic modelling to the study of legal texts can be found, and their contributions are briefly summarized in this section. We mention these studies so that readers can appreciate the broad scope and usefulness of these contributions.

Young (2013) applies topic modeling to 19,000 (front) pages of U.S. newspapers published between 1866 and 1884, with the goal of learning about the discourse during the ratification of the Fourteenth Amendment of the U.S. Constitution. Young uses the resulting topics to assess the degree of media attention that is given to constitutional issues during this critical period.

Rice (2018) looks at U.S. Supreme Court Opinions between 1803 and 2010, and extracts through topic modeling the subject matters that these opinions address. He compares each opinion’s resulting topics with the “subject matter of controversy” metric that is used in the Supreme Court Data Base. Manual coding usually highlights only one dimension and assigns to each document a single mutually-exclusive topic. The advantage of topic modeling is that it permits more topics than the manual coding and also assigns multiple topics to an opinion.

Leibon et. al. (2018) fit a topic model with 100 topics to all (almost 22,000) Supreme Court opinions between 1951 and 2002.

Livermore, Riddell and Rockmore (2017) report results of a quantitative analysis of Supreme Court opinions that measures the degree to which the content of the Court’s opinions conforms to (or departs from) the judicial genre. Using federal appellate court opinions as a baseline, they use topic modeling to estimate the degree of semantic distinctiveness of Supreme Court opinions and track changes in that distinctiveness during the second half of the twentieth century.

Law and Ginsburg (2017) use topic models to analyze a corpus of 615 constitutions, drawn from different nations. They relate the presence of found topics to covariates such as the age, the region, and the legal family of the constitution. Law (2016) applies topic models to the study of archetypes in the preambles of 171 national constitutions.

Rockmore et. al. (2017) study the cultural evolution of national constitutions by looking at a corpus of 591 national constitutions enacted between 1789 and 2008. They obtain valuable information on the topics of these constitutions and trace their topics' time diffusion.

Quinn et al (2010) use topic models to identify important issues discussed in the U.S. Senate between 1997 and 2004.

Macey and Mitts (2014) estimate topic models on the full text of 9,380 judicial opinions dealing with certain corporate issues.

Carter, Brown and Rahmani (2016) apply topic modelling to judgments of the High Court of Australia from 1903 to 2015. The resulting topic models assess the High Court's judicial workload and the shifting make-up of its legal subject matter over time.

Hagen et al. (2015) apply LDA topic modeling to 2,850 petition documents (of at most 800 characters) on the “We the People (WtP)” platform established by the Obama administration, and explore how well the 39 categories provided by the White House represent citizen petitions. Their results imply that topic modeling has the potential to enable the interpretation of large quantities of text through a largely automated process. They also study the effects of stemming on the quality of the topic models by comparing three stemming strategies: no stemming, minimal stemming including pluralization and suffix removal (stripping –ed and –ing from verbs, and –s from nouns), and aggressive stemming (using the Porter stemmer). Topic quality (which was obtained by manually investigating the cohesiveness of the eight most-frequent words in each topic) indicates a slight advantage of the Porter stemmer.

Livermore, Eidelman and Grom (2018) carry out computational text analysis of nearly three million public comments that were received by administrative agencies during the Obama administration. They illustrate that computational text analysis methods show great promise for both researchers and policymakers who are interested in understanding and improving regulatory decision-making.

Ruhl, Nay and Gilligan (2017) study presidential direct actions. Their data running from 1936 to 1999 includes 3933 documents from 11 presidents (meta-variables). They consider a 20 topic model. The dominant words for each topic, in conjunction with the manual reading of the documents that are most-highly related to the topic, allows them to affix to each topic a content label and a coherence rating.

**References**

Carter DJ, Brown J, Rahmani A (2016). Reading The High Court At A Distance: Topic Modelling The Legal Subject Matter And Judicial Activity Of The High Court Of Australia. University of New South Wales Law Review 39(4), 1300-1354.

Gentzkow M, Kelly BT, Taddy M (2017). Text as Data. Working paper. <https://web.stanford.edu/~gentzkow/research/text-as-data.pdf>

Gentzkow M, Shapiro JM (2010). What drives media slant? Evidence from U.S. daily newspapers. Econometrica 78(1): 35–72.

Grimmer J, Stewart BM (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis 21: 267–297.

Hagen L, Uzuner O, Kotfila C, Harrison T., Lamanna D (2015). Understanding Citizens' Direct Policy Suggestions to the Federal Government: A Natural Language Processing and Topic Modeling Approach. Proceedings of the Annual Hawaii International Conference on System Sciences System Sciences DO 10.1109/HICSS.2015.257.

Law DS (2016). Constitutional Archetypes. Texas Law Review 95: 153-243.

Law DS, Ginsburg T (2017). Constitutional Dialects: The Language of Transnational Legal Orders. Working paper

Leibon G, Livermore M, Harder R, Riddell A, and Rockmore D (2018). Bending the Law: Geometric Tools for Quantifying Influence in the Multinetwork of Legal Opinions. Artificial Intelligence and Law 26(2): 145-167.

Livermore MA, Eidelman V, Grom B (2018). Computationally Assisted Regulatory Participation. Notre Dame Law Review 93: 977-1034.

Livermore MA, Riddell AB, Rockmore DN (2017). The Supreme Court and the Judicial Genre. Arizona Law Review 59: 837-901.

Macey J, Mitts J (2014). Finding Order in the Morass: The Three Real Justifications for Piercing the Corporate Veil. Cornell Law Review 100(1): 99-155.

Quinn KM, Monroe BL, Colaresi M, Crespin MH, Radev DR (2010). How to analyze political attention with minimal assumptions and costs. American Journal of Political Science 54(1): 209–228.

Rice D (2018). Measuring the Issue Content of Supreme Court Opinions. The Journal of Law & Courts, forthcoming.

Rockmore DN, Fang C, Foti NJ, Ginsburg T, Krakauer DC (2018). The cultural evolution of national constitutions. Journal of the Association for Information Science and Technology 69(3): 483-494.

Ruhl JB, Nay J, Gilligan JM (2018). Topic Modeling the President: Conventional and Computational Methods. George Washington Law Review 86: 1243–1315.

Wilkerson J, Casas A (2017). Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges. Annual Review of Political Science 20:529-544

Young DT (2013). How Do You Measure a Constitutional Moment? Using Algorithmic Topic Modeling to Evaluate Bruce Ackerman’s Theory of Constitutional Change, Yale Law Journal 122(7): 1990-2054.

**Website Materials**

* Chapter9RCodeTerritorialPapers1.docx: R-Code for the analysis in Section 9.3.1 (Territorial Papers)
* Chapter9RCodeTerritorialPapers2.docx: R-Code for the analysis in Section 9.3.1 (Territorial Papers)
* datacomb.csv: Raw data for Territorial Papers
* Chapter9RCode39Congress.docx: R-Code for the analysis in Section 9.3.2 (39th Congress)
* PrelimData.RData: Data for the speeches of the 39th Congress