Before big data: Using a small data study of politics and law to examine and assess two methods of big data analysis

Pertti Ahonen
University of Helsinki, Finland
pertti.ahonen@helsinki.fi

* I addressed first: An expert audience interested in what digital sociological methods might contribute
* Next: A private sector foundation that kindly provided funding for a 2015–2017 project
* Now also: Social and political scientists and others interested in critical theories of digital phenomena, including the digital methods of analysis
* Later: Referees writing their opinions during evolving rounds of competitively calls for projects

1 Whom did and will I address?

* A first application of digital textual analysis:
  * Latent trait scaling (an unsupervised version)
  * Topic modeling (with latent Dirichlet allocation, LDA)
* The funded project shall use further methods:
  * Classifier techniques to construct better discourse quality indexes (DQIs)
  * Sentiment analysis a.k.a. opinion mining
* To get published has presupposed:
  * Framing experiences on methods use with critical social and political theories of the digital itself

2 What have I done and why?

* The technical approach has been becoming an approach of social and political theories to digital methods and other digital phenomena
* Technically, the scope of the relevant methods has extended, and also will comprise methods development
* Presentationally, mathematics has been pushed somewhat towards the background
* Baseline comparisons between digital and traditional methods must have been included

3 What has happened during the process?

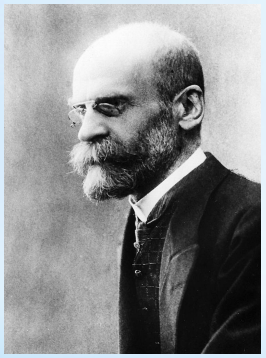
$y_{ij} \sim \text{Poisson}\ (\lambda_{ij})$ (1)

$\lambda_{ij} = \exp\ (\alpha_i + \psi_j + \beta_j \times \omega_i)$ (2)

$\beta \sim \text{Dirichlet}\ (\delta)$ (3)

$\theta \sim \text{Dirichlet}\ (\alpha)$ (4)

$z_i \sim \text{Multinomial}\ (\theta)$ (5)





3 (Continued) Supplementing formulas … with ideas of people who think

* It was not necessary to abandon social and political research after all and become a data scientist
* Mainstreaming the digital methods poses both opportunities (e.g., doing things better and new things) and threats (e.g., doing few better things, and irritating the entrenched mainstream)
* To get funded and published, a balance need be found between what is fashionable enough to win funding for innovative projects on the one hand, and becoming a mere technician on the other

4 Implications for theory and practice

* The possible gearing of ”X” (here, membership in the R club of program libraries) to the benefit of ”Y” (personal digital social research)
* Positive side effects: e.g., shedding new light on certain established themes and methods of political science research
* But few social and political scientists are likely to become foremost innovators of digital methods and techniques

5 What do others benefit, and what we/I cannot do?

* Please ask at pertti.ahonen@helsinki.fi
  * Information on the root study from which it all started by and large, Appendix 1
  * Information on the project indicated in this presentation, Appendix 2

Available on request for limited supply

We expect big data methods to make rational contributions by means of helping generate research results that are not inferior to those attained in other ways but are possibly better, or hard or impossible to generate in those other ways. Those who apply these methods may also aspire to use them to augment the available arsenal of research methods, offer surrogates for existing research designs, and re-orient research. Moreover, we can critically examine the direct and indirect societal and political effects of the institutionalization of big data methods. To reach its first objective, this article elaborates in its final section conclusions on how big data methods, not only by means of their 'social life' but also by their 'political life', influence the institutionalization of social research with special reference to political science research. To advance towards its conclusions, the article first pursues a second objective, re-examining a comparative 'intermediate data' study of budgetary legislation in thirteen countries to draw conclusions concerning the augmentation of the arsenal of research methods, the surrogation of existing research designs, and the re-orientation of research.

Appendix 1. Abstract of the root study in its final form

Abstract. The project *Digital humanities of public policy-making* addresses an audience of political science and data science scholars and other scholars, political, policy, management and other practitioners, citizens and their organizations, and businesses offering digital applications. During 2015–2017, the project team will contribute to digital humanities research by means of developing selected new digital humanities methods, applying existing methods in new ways using research material not examined in those ways before, and, in synthesizing the project results, framing the digital humanities in their political and social contexts. The project motivation derives from a desire to advance the appropriation of developments of big data and open data within social research without comprising the achievements of cutting-edge data science on the one hand, and without cutting ties with mainstream political and social research. The project expects to make advances by means of developing discourse quality indexes (DQIs); using big data methods to examine jointly political party programs and government political master programs in the longer term; using big data methods to examine selected policy documents and organizational strategy documents within a given field of inquiry; examining an 'intermediate data' set of policy evaluation documents; and appropriately framing the digital humanities in some of their political and social contexts of application and influence. The project will tap the richness of methods including the construction of classifiers to establish better DQIs; extending the applications of the big data methods of latent trait scaling; topic modeling; and carrying out sentiment analysis a.k.a. opinion mining. The project team expects its research results to enable drawing conclusions both as concerns the opportunities and the limitations of the digital humanities in social research, including the practical applications of this research. The team also expects that its findings will help better specify the place of the digital humanities as concerns theory of political and social research and as concerns the contribution of digital humanities research to practical mastery, healthy criticism and constant improvement of the digital humanities and their methods. The project has entered a rapidly evolving field of research, which emphasizes the importance of vigilance to be able to make suitable adjustments to the orientation of the project during its course.

Appendix 2. Abstract of the funded project

Thinking about my boating scenery of next summer, thank you!