Center for Text Analytic Methods in Legal Studies

• Kevin D. Ashley, School of Law, LRDC, ISP
• James M. Anderson, RAND Institute for Civil Justice
• Daqing He, Informatics and Networked Systems, SCI, ISP
• Rebecca Hwa, Computer Science, SCI, ISP
• Diane J. Litman, Computer Science, SCI, LRDC, ISP

Motivation

• Increasingly, legal texts seen as data:
  • Court decisions, statutes, and regulations.
  • New sources of legal data are available for research:
    • Harvard Caselaw Access Project
    • 6.7 million US cases
  • New higher resolution text analytic techniques
    • Transformer language models (e.g., BERT)
    • Extract case topics, legal areas, timeline events, legal tests, and outcomes
    • Identify bias in legal language
• Center will connect:
  • Pitt Law & RAND legal domain experts and
  • Pitt NLP/ML/text analytics researchers.

Project Description

• Identify socially relevant empirical legal hypotheses testable via text analytics.
• Adapt and apply text analytics to extended legal data sets to evaluate hypotheses.
• Based on results, publish papers and apply for external funding to sustain Center’s research.

Context

• Empirical legal studies have been constrained to manual text analyses.
• Pitt team experienced in developing / applying new text analytics:
  • Natural language processing
  • Machine learning
  • Applied to legal cases and statutes.
• Rand team and Pitt Law faculty experienced in identifying / evaluating socially relevant empirical legal hypotheses

Apply new analytic techniques to new sources of legal text data to evaluate legal hypotheses in ways not previously possible.

Project Deliverables

Step 1, Spring 2021: Identify candidate legal hypotheses
Step 2: Strategize NLP/ML approach
Step 3: Annotate data for supervised ML
Step 4: Process pipeline; evaluate hypothesis
Step 5: Develop pedagogical approach
Step 6 & 7: Submit papers & funding proposals

Potential Impact

• Focuses on important hypotheses.
• Social issues: racism, gender equality, immigration, public health, crime, or education.
• Engages current literature.
• Empirical legal research field awakening to methodological possibilities of text analytics.
• Pedagogical opportunities.
• Introduce them to technology and to socially relevant legal issues.
• Motivate / engage law and prelaw students.
• Help them learn to read legal decisions more effectively.
• Focal center for research on text analytics and legal studies in Pittsburgh

References


Acknowledgements

Thanks to the University of Pittsburgh School of Law, School of Computing and Information, and to the RAND Corp.