Methodological Issues in Humanities Computing

HITL Models, Impossibility Results, and Interactive Computing

K. Nielbo
knielbo.github.io

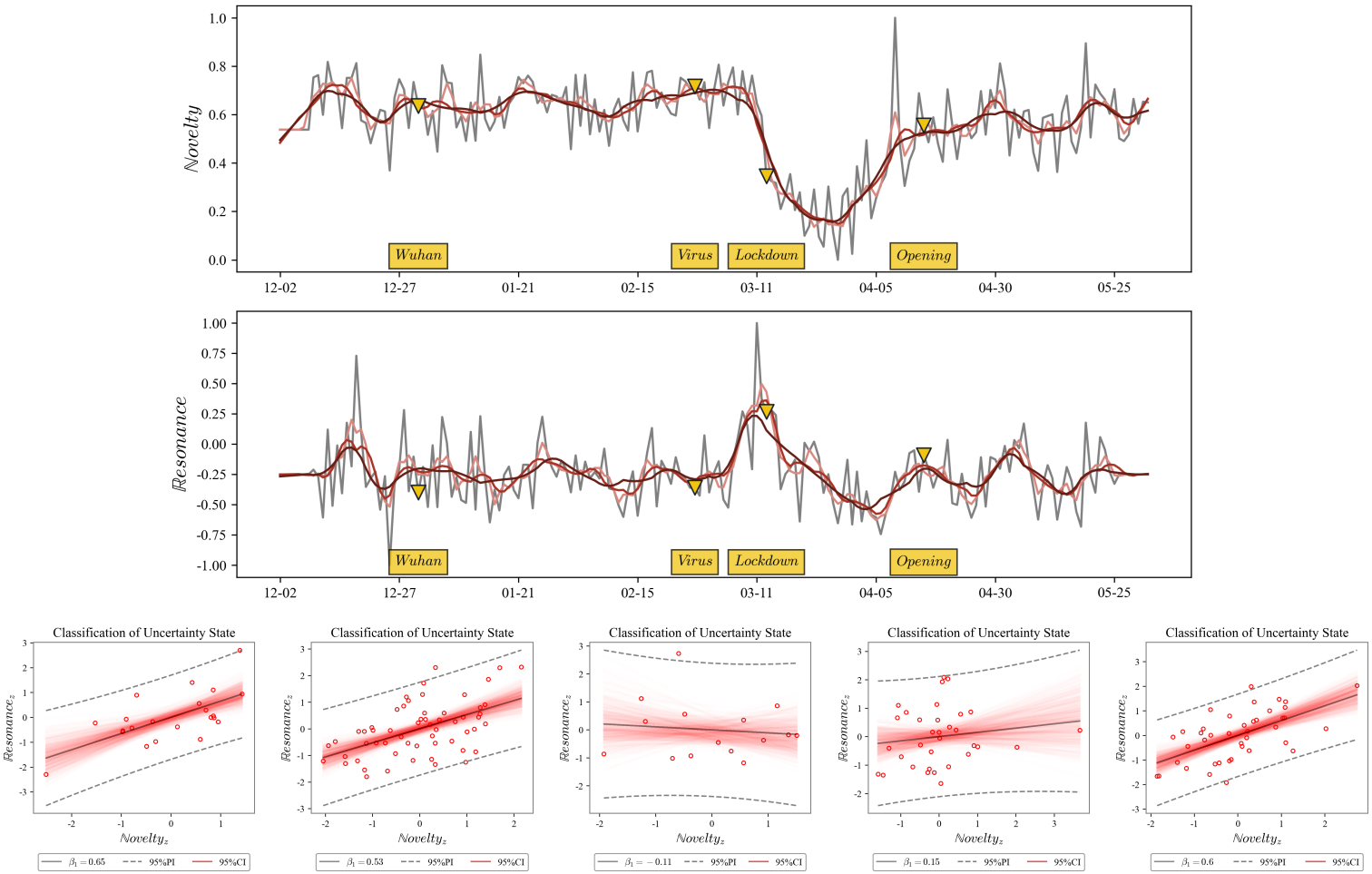
Center for Humanities Computing Aarhus|chcaa.io
Aarhus University, Denmark

CENTER FOR HUMANITIES
COMPUTING AARHUS

outline

1. Research Problems
   - News Information Decoupling
   - FabulaNET
2. Methodological Issues
   - Human-in-the-Loop Models
   - Impossibility Results
   - Interactive Computing

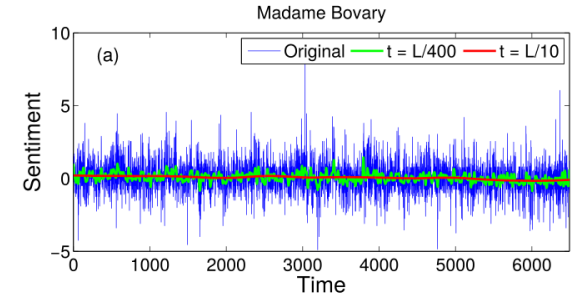
News Information Decoupling

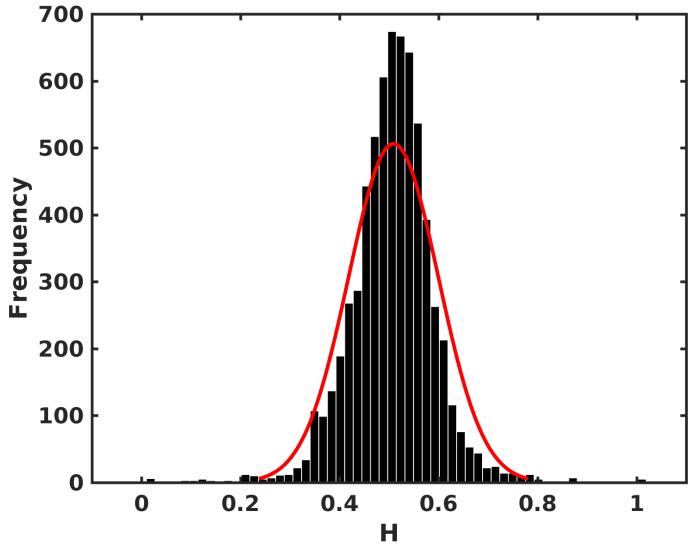


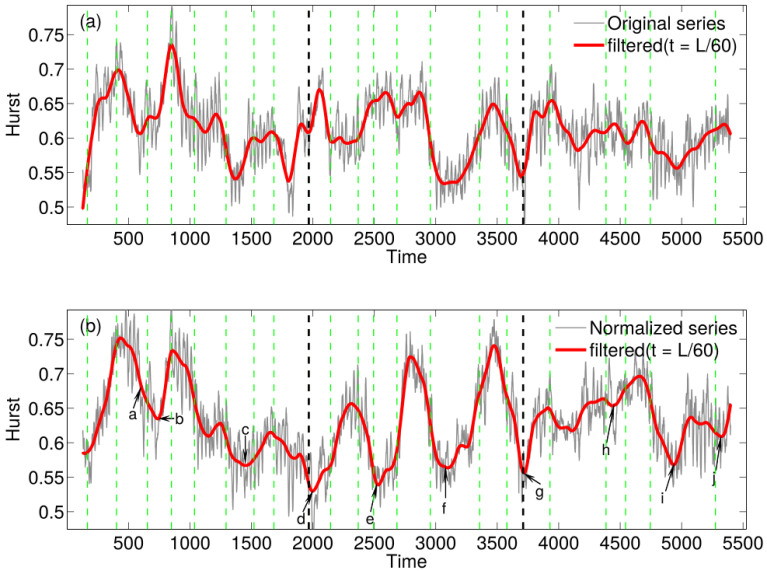
front pages from danish legacy print media *politiken* during Covid-19 phase 1

K. L. Nielbo, R. B. Baglini, P. B. Vahlstrup, K. C. Enevoldsen, A. Bechmann, and A. Roepstorff (2021) "News Information Decoupling: An Information Signature of Catastrophes in Legacy News Media," arXiv:2101.02956 [cs]

Research Problems
News Information Decoupling
FabulaNET
Methodological Issues
Human-in-the-Loop Models
Impossibility Results
Interactive Computing







FabulaNET

- global H of story arcs provide an index of narrative coherence
- narrative coherence: $0.5 < H \leq 1.0$
- local H detects changes in the narrative
- optimal narratives balances coherence to maximize reader motivation

---

Q. Hu, B. Liu, M. R. Thomsen, J. Gao, and K. L. Nielbo (2020) "Dynamic evolution of sentiments in Never Let Me Go: Insights from multifractal theory and its implications for literary analysis," DSH, DOI: 10.1093/fqz092.

Human-in-the-Loop Models

as task complexity increases, a need for (operational approaches to) leveraging human intelligence in the development of learning algorithms has become apparent

Type	Human Involvement	Resources	Relevance
Out-of-the-loop	not required	low	low
On-the-loop	checking	medium	medium↓
In-the-loop	required	high	medium↑

WHEN

algorithms are not understanding the input

data input is interpreted incorrectly

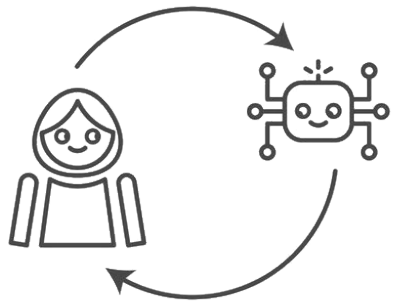
algorithms do not know how to perform the task

to make models more accurate

cost of errors is too high in development

data is rare or not available

THEN



HITL Models

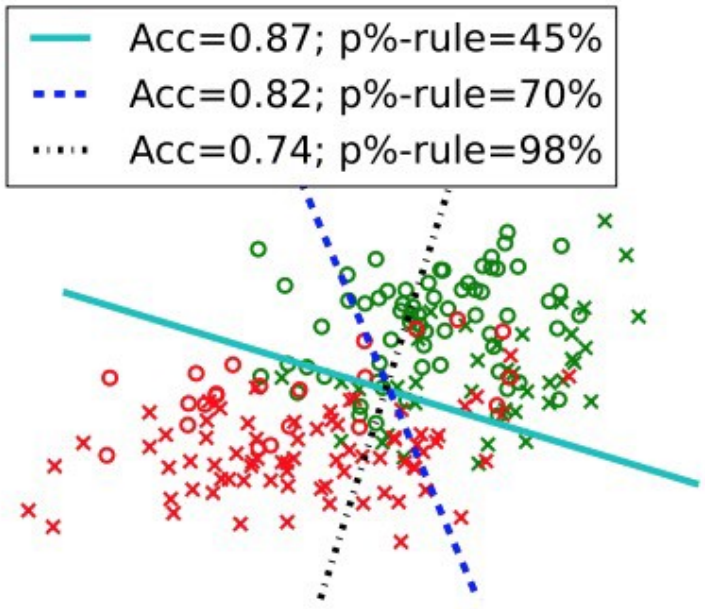
Impossibility Result

Assume differing base rates, $Pr_a(Y = 1) \neq Pr_b(Y = 1)$, and an imperfect learning algorithm, $C \neq Y$, then you cannot simultaneously achieve:

Precision parity $Pr_a(Y = 1 \mid C = 1) = Pr_b(Y = 1 \mid C = 1)$

True positive parity $Pr_a(C = 1 \mid Y = 1) = Pr_b(C = 1 \mid Y = 1)$

False positive parity $Pr_a(C = 1 \mid Y = 0) = Pr_b(C = 1 \mid Y = 0)$

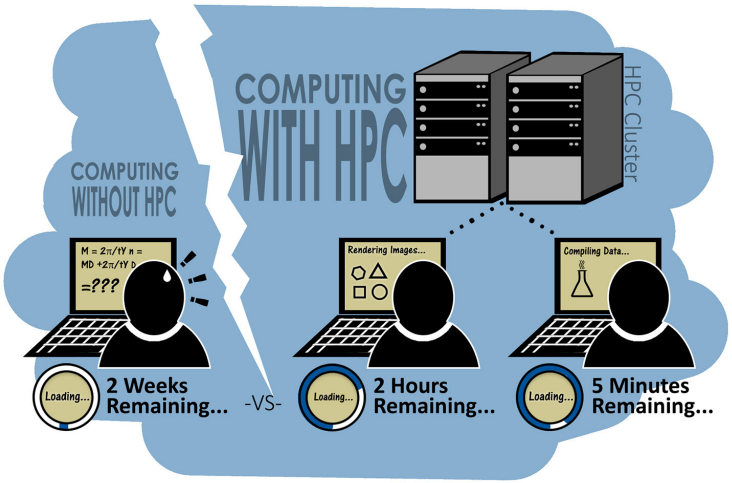


`soft breakdown of performance` (precision, recall) in response to ethical and legal requirements, e.g., demographic parity at level x 'costs' $pr = y$

Kleinberg J., S. Mullainathan, & M. Raghavan (2016), Inherent Trade-Offs in the Fair Determination of Risk Scores, arXiv:1609.05807

Research Problems
News Information Decoupling
FabulaNET
Methodological Issues
Human-in-the-Loop Models
Impossibility Results
Interactive Computing

Interactive Computing



"Efficient use of computing platforms for running applications quickly"

- *speed* matters because we do not want to wait too long
- *efficient use* means more resources for all of us

TRADITIONAL USERS

By default HPC *is not for all*: "you must learn to crawl before you can walk"

1. write code base for project
2. formulate batch script
3. submit to job queue

NEW USERS

New technologies (Big Data, AI) have created new users

- easy & interactive access
- collaborative development
- code and data sharing

for exploration, experimentation, and debugging in a sandbox-like environment

Research Problems
News Information Decoupling
FabulaNET
Methodological Issues
Human-in-the-Loop Models
Impossibility Results
Interactive Computing

Research Problems
News Information Decoupling
FabulaNET
Methodological Issues
Human-in-the-Loop Models
Impossibility Results
Interactive Computing

THANKS

kln@au.dk
knielbo.github.io
chcaa.io

SLIDES

knielbo.github.io/files/kln_methissues.pdf