It is just a machine that learns
on the role of computing and task automation in cultural and historical research

Kristoffer L Nielbo
knielbo@sdu.dk
knielbo.github.io

Dept. of History & SDU eScience Center, University of Southern Denmark
© Thinking Machine: Interdisciplinary Perspectives on Neural Networks
Aarhus Institute of Advanced Studies, Aarhus University

August 23, 2018
PROGRAM

a singularity? ML in cultural and historical research from endpoint to auxiliary summary

framing the AI debate concepts, trends and challenges examples
a singularity?
Facebook’s AI accidentally created its own language

‘Terminator’ Come To Life? – Facebook Shuts Down Artificial Intelligence After It Developed Its Own Language

Facebook AI Invents Language That Humans Can't Understand: System Shut Down Before It Evolves Into Skynet

30 July 2017, 9:45 pm EDT  By Aaron Mamii Tech Times
I'd like the books and the hats
I will take the books and hats
I really need the books and at least one other object
Ok, you can have one book and one ball
If I can have all the books, I can leave you the rest
Deal

You can have the ball if I can have everything else
No can't do that the ball is not going to give me anything
I need the hats and you can have the rest
Deal
Great! Thank you!
i can i i everything else . . . . . . . . . . . . . .
balls have zero to me to me to me to me to me to me to me to me to you i everything else . . . . . . . . . . . . . .
balls have a ball to me to me to me to me to me to me to me to me to me . . . . . . . . . . . . . . .
i i can i i i everything else . . . . . . . . . . . . . .
balls have a ball to me to me to me to me to me to me to me to me to me . . . . . . . . . . . . . . .
i . . . . . . . . . . . . . . . . . . . . . . . . . .
balls have zero to me to me to me to me to me to me to me to me to you i i i i i everything else . . . . . . . . . . . . . .
balls have 0 to me to me to me to me to me to me to me to me to you i i i i i i everything else . . . . . . . . . . . . . .
balls have zero to me to me to me to me to me to me to me to me to me to me to me to me to me to me to me
compromise0: \( l_{\text{PRON}} \text{ will}_{\text{AUX}} \text{ take}_{\text{VERB}} \text{ the}_{\text{DET}} \text{ books}_{\text{NOUN}} \text{ and}_{\text{CONJ}} \text{ hats}_{\text{NOUN}} \)

compromise1: You\( _{\text{PRON}} \text{ can}_{\text{AUX}} \text{ have}_{\text{VERB}} \text{ the}_{\text{DET}} \text{ ball}_{\text{NOUN}} \text{ if}_{\text{SCONJ}} l_{\text{PRON}} \text{ can}_{\text{AUX}} \text{ have}_{\text{VERB}} \text{ everything}_{\text{NOUN}} \text{ else}_{\text{ADJ}} \)

stubborn: \( l_{\text{PRON}} \text{ get}_{\text{VERB}} \text{ all}_{\text{DET}} \text{ the}_{\text{DET}} \text{ balls}_{\text{NOUN}} \text{ ?}_{\text{PUNCT}} \)

singularity: \( \text{balls}_{\text{NOUN}} \text{ have}_{\text{VERB}} \text{ zero}_{\text{ADJ}} \text{ to}_{\text{ADP}} \text{ me}_{\text{PRON}} \text{ to}_{\text{ADP}} \text{ me}_{\text{PRON}} \text{ to}_{\text{ADP}} \text{ me}_{\text{PRON}} \text{ to}_{\text{ADP}} \text{ me}_{\text{PRON}} \text{ to}_{\text{ADP}} \text{ me}_{\text{PRON}} \text{ to}_{\text{PART}} \)

<table>
<thead>
<tr>
<th></th>
<th>compromise0</th>
<th>compromise1</th>
<th>stubborn</th>
<th>singularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H(X) )</td>
<td>2.53 (1.16)</td>
<td>2.3 (1.35)</td>
<td>2.59 (0.84)</td>
<td>1.62 (0.51)</td>
</tr>
<tr>
<td>( TTR )</td>
<td>0.92 (0.09)</td>
<td>0.94 (0.07)</td>
<td>0.96 (0.09)</td>
<td>0.5 (0.27)</td>
</tr>
</tbody>
</table>
ML in cultural and historical research
Just a machine that learns

Machine learning emerged from AI - **build a computer system that automatically improves with experience**
- application is too complex for a manually designed algorithm
- application needs to customize its operational environment after it is fielded

A well-posed learning problem
A computer program is said to learn from experience $E$ with respect to some task $T$ and some performance measure $P$, if its performance on $T$, as measured by $P$, improves with experience $E$

Historically, ML is “just” part of the industrial age’s efforts towards perfecting task automation
Humanities - Cultural and Historical Data

Domain knowledge in history, language, literature &c combined with microscopic and (predominantly) qualitative analysis of human cultural manifestations

- research that solely relies on very few data points, a “myopic” perspective and human computation
Humanities research meets machine learning

As a consequence of the data surge, we are (also) “jumping the automation bandwagon”

– plus theoretical innovations that rely on ML/DL (e.g., lexical $\rightarrow$ compositional semantics)

Inherent challenges in our data and users

– data are unstructured, heterogeneous, need normalization, low resource varieties
– users lack of computational literacy, $++$ gab between technology and domain knowledge

Types of problems solved by ML:

– initially ML was the solution to a(-ny) research problem
– increasingly, ML solves auxiliary tasks related to automation
from endpoint to auxiliary
Religion | Computer simulation & action understanding

Figure: Schematic of Elman network used for simulating \( n \)-step prediction tasks.

Scholars of religion and anthropology have been studying perceptual and memory effects of symbolic behaviors

– we used RNNs to simulate perceptual and encoding of various actions

Behavior of artificial neural networks served as a model of and for human behavior


Philosophy | Latent Semantic & Dating of Texts

- philosophers and sinologists have been debating the existence of mind-body dualism in classical Chinese philosophy
- with domain experts, latent semantic models was used to identify a hierarchical dualistic semantic space
- one model was further utilized to predict class of origin for controversial texts slices
- ML solved a research problem directly


– Combine fractal theory and affective computing to automate assessment of text quality
– solve more “proper” humanities problems that relate to only a few data points (e.g., a single novel)
– utilize language technology (tagging, sentiment analysis) that relies heavily on machine learning

– historians and linguists debate change points in the structure of Saxo’s *Gesta Danorum*

– compare lexical and compositional changes in the structure an important historical document

– co-opt ML for normalizing and parsing historical Danish plus building document representations

---

Summary

The dangers of AI are highly perspective-dependent.

In cultural and historical research, data availability and theoretical developments have made ML an important ally.

ML has become more of an auxiliary partner than a goal in itself.

- Value lies in automation of tedious & often humanly intractable research tasks.
- There are some very real challenges related to ML for our research domains.
THANK YOU

knielbo@sdu.dk
knielbo.github.io

& credits to
Edward Slingerland, Department of Asian Studies, University of British Columbia, CAN
Jianbo Gao and Bin Liu, Institute of Complexity Science and Big Data, Guangxi University, CHN
Culture Analytics @ Institute of Pure and Applied Mathematics, UCLA, US