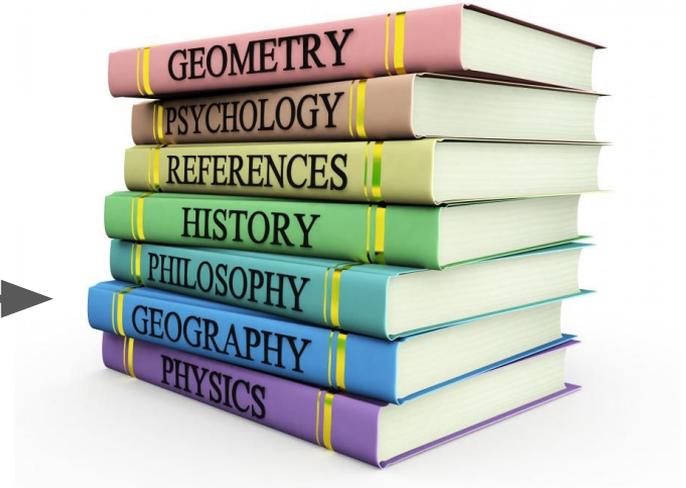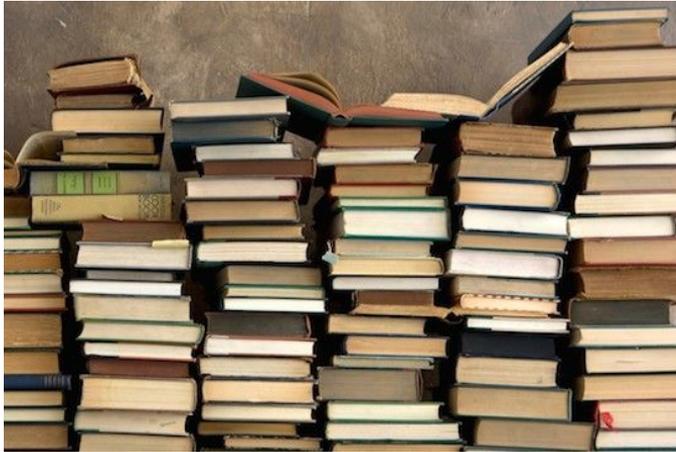# Neural Variational Author Topic Process
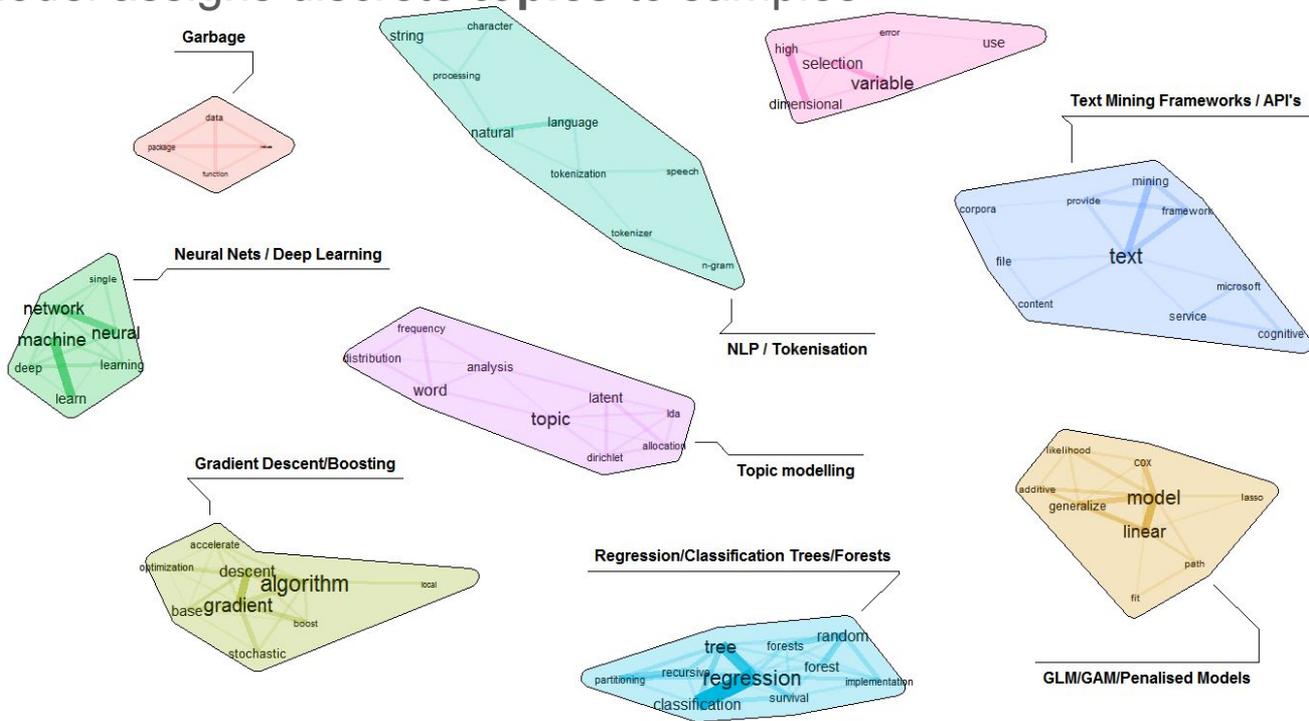
Ulzee An
1/25

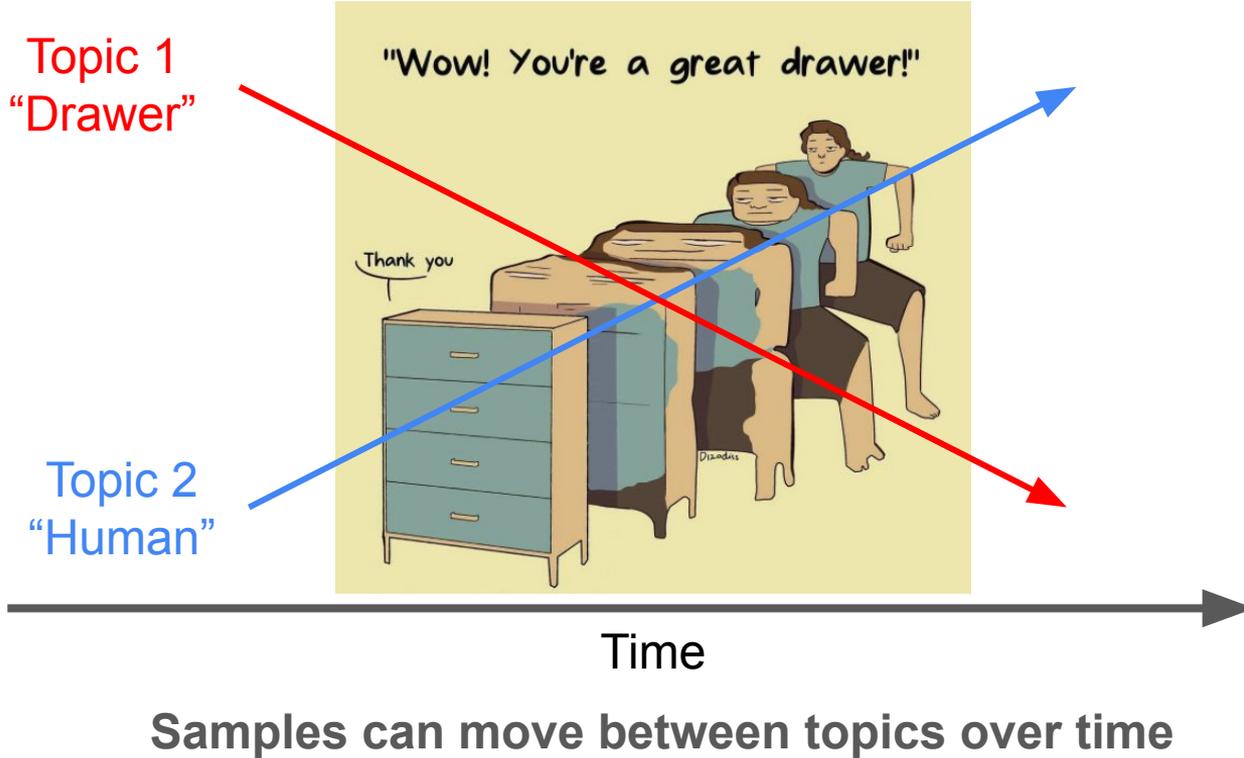# We often find "topics" in unstructured data

Topic models allow a categorization of unlabelled data

- The model assigns discrete **topics** to samples

"Topics of ML"

Topic membership can change over time

Topic 1 "Drawer"

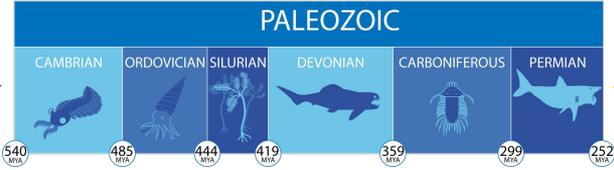Topic 2 "Human"

Time

Samples can move between topics over time

Existing models use time or authorship, but not both

"By time" (Dynamic Topic Model)

No structure in data
(LDA)

By author (Author Topic Model)
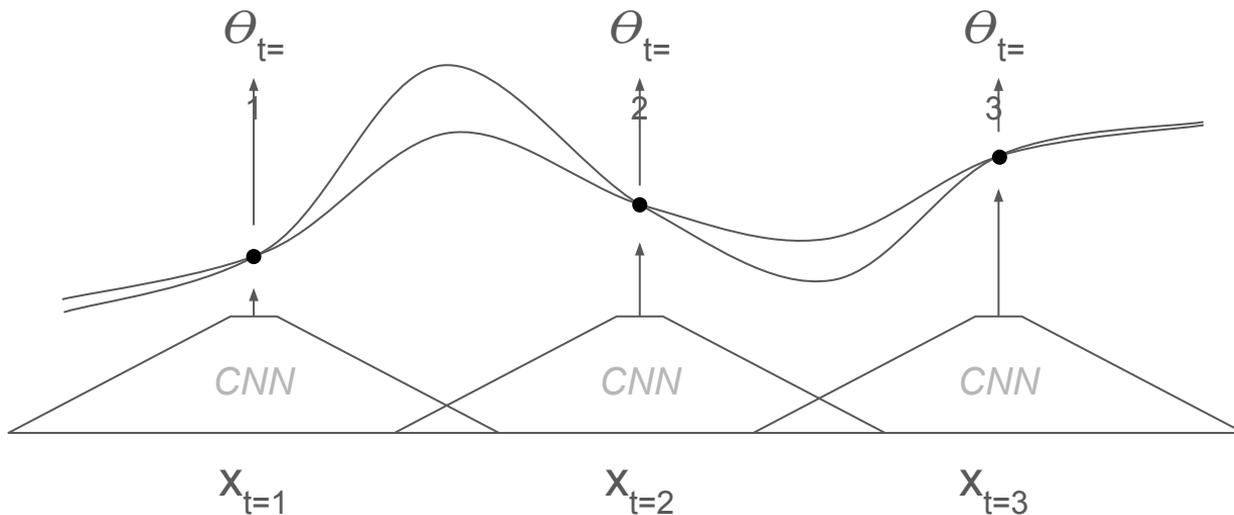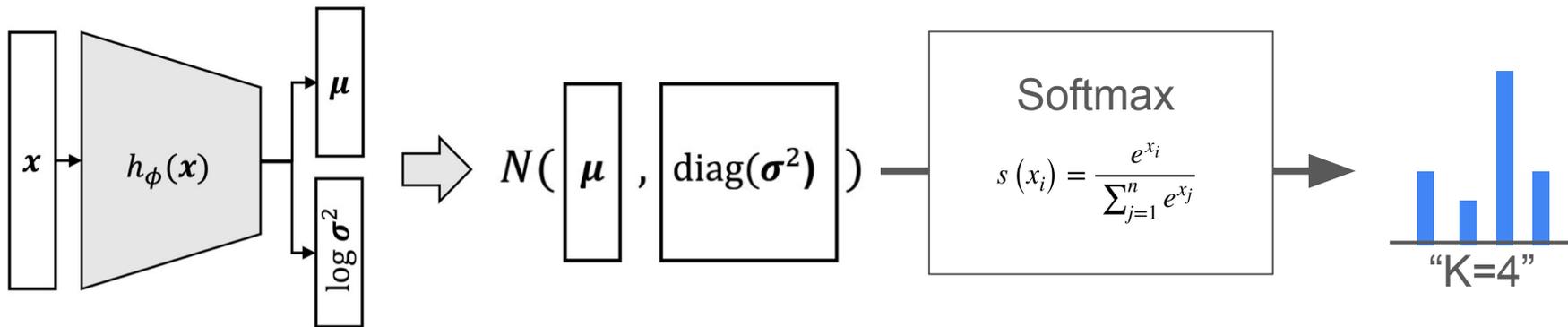
Author
x
Time?
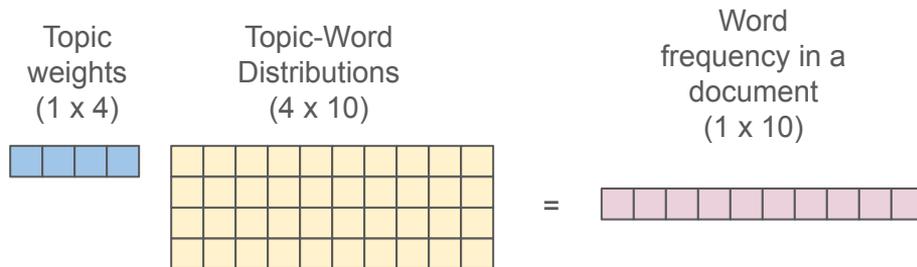
Our contribution

Author Topic Process (ATP)

- Relate an author's topic proportions over time through a Gaussian Process
- Common pitfall: observations are infrequent and irregular over time
  - Topic proportion dynamics are learned in an interpolated latent space
- A type of Neural Topic Model (NTM)

Inference of ATP using a Variational Auto Encoder

$$N(\boxed{\mu}, \boxed{\text{diag}(\sigma^2)})$$

Softmax

$$s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

"K=4"

Topic decoder

Topic weights (1 x 4)

Topic-Word Distributions (4 x 10)

Word frequency in a document (1 x 10)

=

Experiments on UN Debates & ACL Papers

Datasets with author & time: We organized two popular datasets in the literature in terms of papers written by authors over time

	YRS	OBS	AU	DOC	V
UN	46	5	199	7,507	1,690
ACL	20	1	1,763	37,144	2,725

Reconstruction task: Complete second half of document given first (Compl) or generative next document given previous (Next).
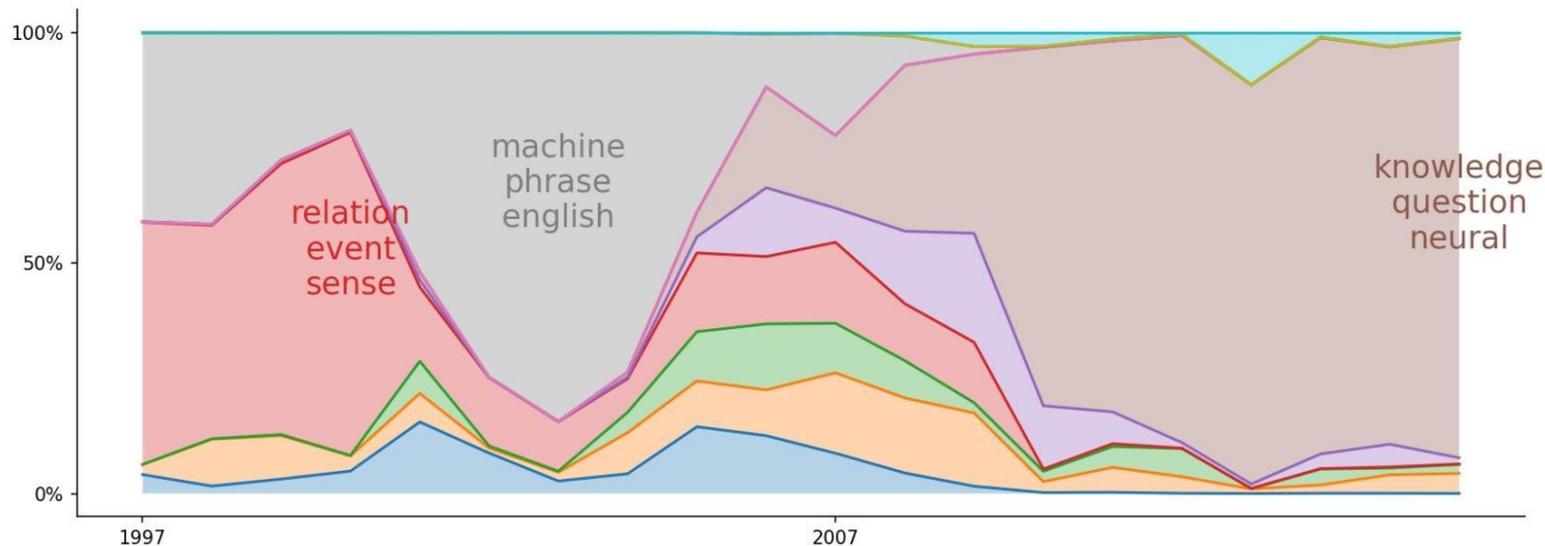
"Deep" LDA

LDA over time

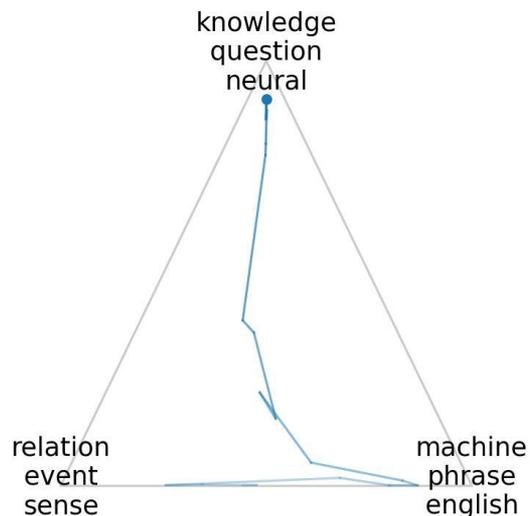|	UN DEBATES		ACL PAPERS	
	COMPL	NEXT	COMPL	NEXT
GSM	1007	1127	1159	1374
DTM	1176	1305	1497	1542
ATP	1032	1012	1141	1274

Perplexity = exp(NLL per word)

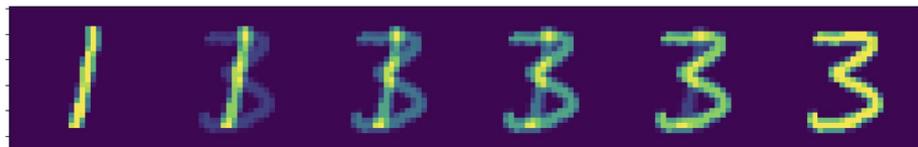# Example of topics inferred for a country (UN Debates)
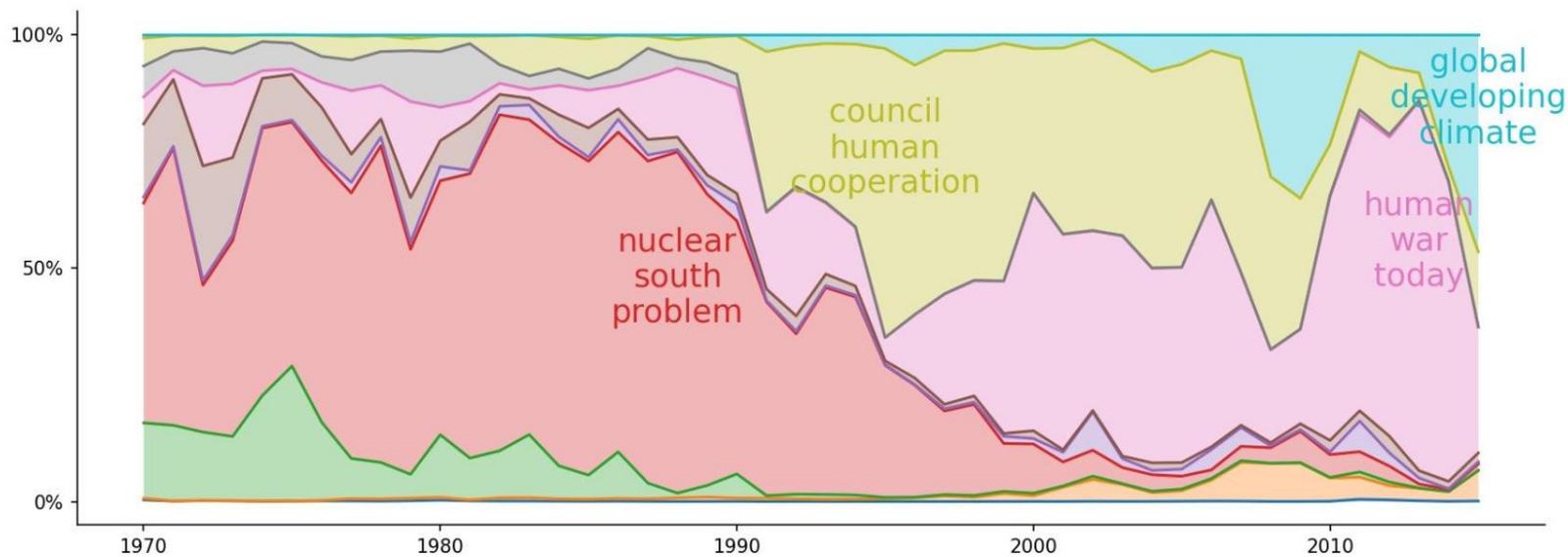
Example of topics inferred for an author (ACL Papers)

Future directions of ATP

- Can ATP help use discover meaningful trajectories in Eg. disease history? and other modalities (images)
- The use of word embeddings & document embeddings instead of bag-of-words (BOW) representations

"Topic 1"   "Topic 2"

Thank you!

Professor Sriram
Sankararaman

Computer Science Department
Department of Human Genetics
Department of Computational Medicine
Bioinformatics Interdepartmental Graduate
Program

Shuai Zhang
Amazon
AWS

Contact:
ulzee@cs.ucla.edu

**SCIENCE HUB FOR HUMANITY
AND ARTIFICIAL INTELLIGENCE**
UCLA   amazon