Inference and Representation Fall 2017 Final Project Proposal

Zhi-Wei Li   Alexander Rich   Anselm Rothe

Topic models have been used successfully to capture structure in large text corpora and make these corpora more human-understandable (Blei, Ng & Jordan 2003). However, traditional topic models to not capture the temporal ordering of documents, and the way topics change over time. Dynamic topic models address this issue by allowing topic mixing proportions and the distribution of words in topics to change as a function of time (Blei & Lafferty 2006). In this project, we will apply dynamic topic modeling to 35 years of the Proceedings of the Cognitive Science Society to understand and visualize how the research topics of the society have changed over its history.

Data Set

The Proceedings of the Cognitive Science Society are a representative corpus of the ongoing research in the cognitive science community, and are available online for the 3rd through 38th annual conferences (1981-2016). While we have not yet counted the documents in the corpus, we expect it to be on the order of 10,000 documents, with each document about 6 pdf pages long. Older proceedings were scanned and processed using OCR, so there may be some issues with data quality for these earlier years. We will likely preprocess this data set using stemming and/or lemmatization techniques, and remove common words that do not carry interesting meaning.

Model

We will implement a dynamic topic model (DTM) as described by Blei and Lafferty (2006). As with traditional topic models, documents are assumed to be generated by a model in which topic proportions are first generated according to a Dirichlet distribution, words are assigned to topics according to a multinomial distribution based on the topic proportions, and then words are generated according to a multinomial distribution based on the assigned topic. In a dynamic topic model, topic proportions and word distributions within a topic are both allowed to drift over time according to a logistic normal distribution. The full graphical model is summarized in Figure 1. To fit the DTM to our data set, we will write the model as a directed graphical model using the Stan modeling language (Stan Development Team, 2017). We will then fit the model using Stan's built-in variational inference algorithm.

Evaluation

Following Blei and Lafferty, we will evaluate the model by fitting the model to all data up to a given year, and then testing its ability to predict the articles from that year. For comparison, we will also test traditional topic models trained on all previous years and on only the most recent previous year.
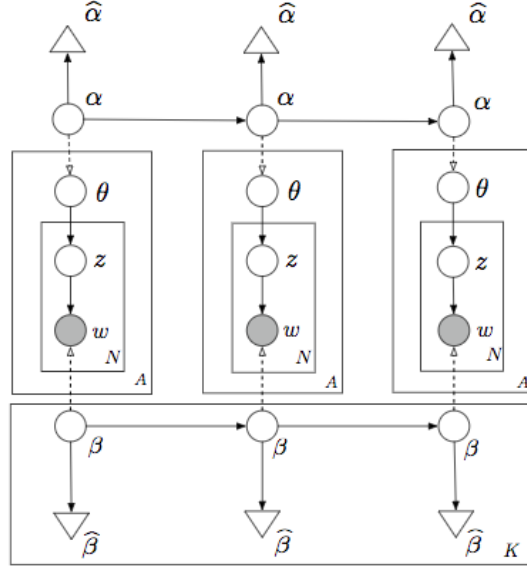
Figure 1: Full graphical model of a dynamic topic model (DTM).

Deliverables

Our goal is to use our model to understand how the field of cognitive science has changed over time, as attempted with traditional topic models by Cohen, Priva & Austerweil (2015) for the journal Cognition (which focuses more squarely on cognitive psychology). To achieve this, we plan to visualize the models' topics and their change over time in an understandable manner. Building off the work of Blei and Lafferty (see Figure 2), we will visualize each topic by showing its shifting popularity (i.e., proportion) over time, top words over time, and a selection of article titles typical of the topic over time. We will also perform analyses to reveal which topics have changed the most and least in popularity over time, and which have changed most and least in the distribution of words within the topics.

Acknowledgements

References

Blei, D., & Lafferty, J.D. (2006). Dynamic topic models. Proceedings of the 23rd international Conference on Machine Learning. ICML '06, 113-120.

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research. 993-1022.

Cohen Priva, U., & Austerweil, J.L. (2015). Analyzing the history of Cognition using Topic Models.

Stan Development Team. 2017. The Stan Core Library, Version 2.16.0. http://mc-stan.org

1881	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
force	motion	magnet	force	atom	ray	energy	energy	radiat	electron	electron	electron	state
energy	force	electric	magnet	theory	measure	radiat	radiat	energy	energy	energy	atom	energy
motion	magnet	measure	theory	electron	energy	measure	ray	electron	atom	particle	energy	electron
differ	energy	force	electric	energy	theory	energy	electron	measure	measure	field	structur	magnet
light	measure	theory	atom	measure	light	light	measure	ray	radiat	radiat	field	field
measure	differ	system	system	ray	wave	particle	atom	atom	field	model	model	atom
magnet	direct	motion	measure	electr	radiat	ray	particle	field	ray	atom	state	system
direct	line	line	line	line	atom	radiat	two	two	model	two	two	two
matter	result	point	energy	force	electric	point	light	particle	particle	ray	magnet	quantum
result	light	differ	body	value	value	theory	absorpt	observe	magnet	measure	ray	physic

"Atomic Physics"

1881 On Matter as a form of Energy
1892 Non-Euclidean Geometry
1900 On Kathode Rays and Some Related Phenomena
1917 ``Keep Your Eye on the Ball''
1920 The Arrangement of Atoms in Some Common Metals
1933 Studies in Nuclear Physics
1943 Aristotle, Newton, Einstein. II
1950 Instrumentation for Radioactivity
1965 Lasers
1975 Particle Physics: Evidence for Magnetic Monopole Obtained
1985 Fermilab Tests its Antiproton Factory
1999 Quantum Computing with Electrons Floating on Liquid Helium

1881	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
brain	movement	brain	movement	movement	stimulate	record	respons	response	respons	cell	cell	neuron
movement	eye	eye	brain	sound	muscle	nerve	record	stimulate	cell	neuron	channel	active
action	right	right	sound	muscle	sound	stimulate	stimulate	record	potential	response	neuron	brain
right	hand	movement	nerve	active	movement	response	nerve	condition	stimul	active	ca2	cell
eye	brain	left	active	nerve	response	muscle	muscle	active	neuron	brain	active	fig
hand	left	hand	muscle	stimulate	nerve	electrode	active	potential	active	stimul	brain	response
left	action	nerve	left	fiber	frequency	active	frequency	stimulus	nerve	muscle	receptor	channel
muscle	muscle	vision	eye	reaction	fiber	brain	electrode	nerve	eye	system	muscle	receptor
nerve	sound	sound	right	brain	active	fiber	potential	subject	record	nerve	respons	synapse
sound	experiment	muscle	nervous	response	brain	potential	study	eye	abstract	receptor	current	signal

"Neuroscience"

1887 Mental Science
1900 Hemianopsia in Migraine
1912 A Defence of the ``New Phrenology''
1921 The Synchronal Flashing of Fireflies
1932 Myoesthesis and Imageless Thought
1943 Acetylcholine and the Physiology of the Nervous System
1952 Brain Waves and Unit Discharge in Cerebral Cortex
1963 Errorless Discrimination Learning in the Pigeon
1974 Temporal Summation of Light by a Vertebrate Visual Receptor
1983 Hysteresis in the Force-Calcium Relation in Muscle
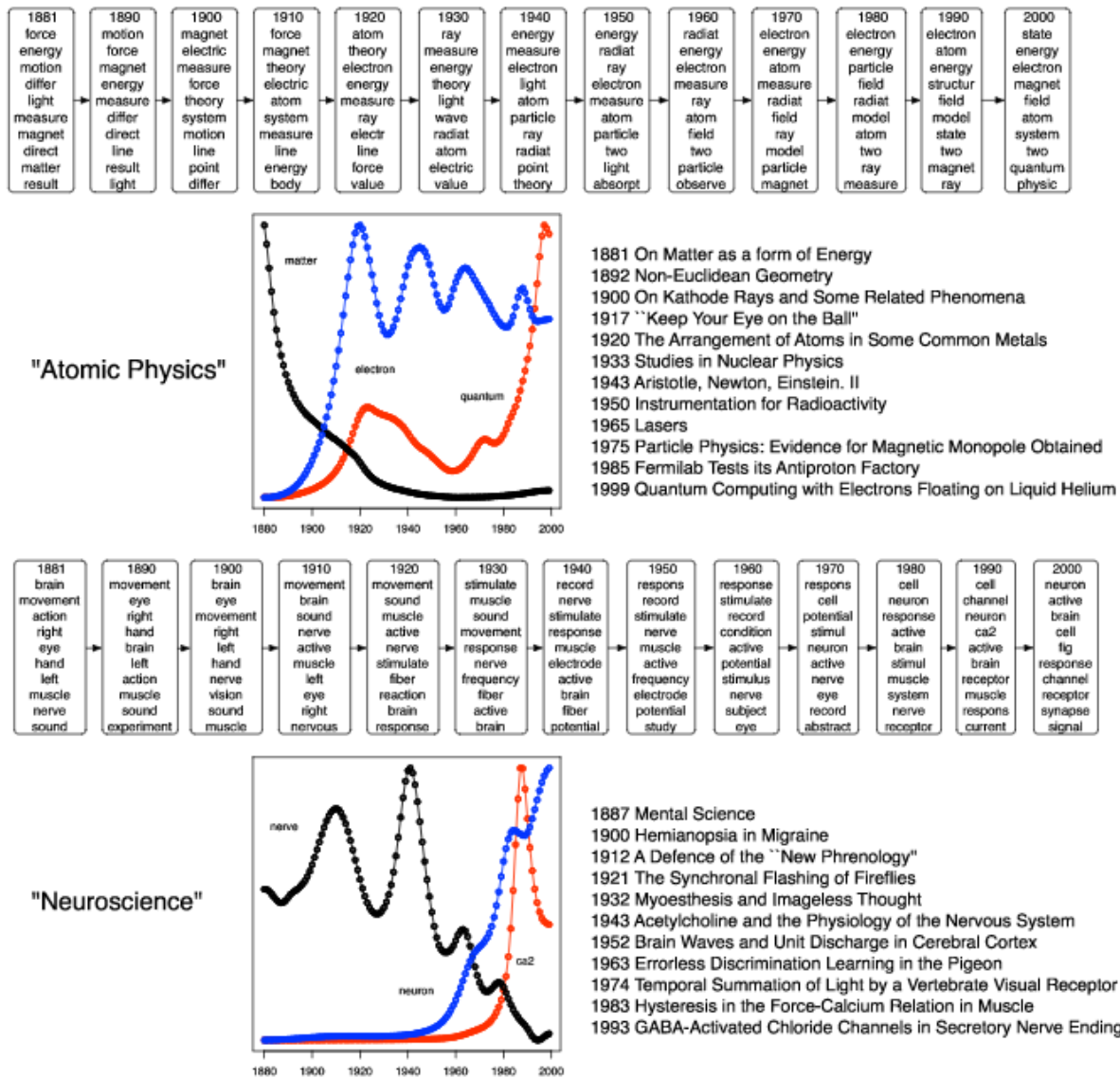1993 GABA-Activated Chloride Channels in Secretory Nerve Endings

Figure 2: Trends uncovered by the DTM in Blei and Lafferty (2006).