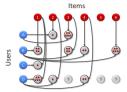
Reduced Rank Factor Modeling



Ayan Acharya

October 8, 2015

Matrices and Tensors are Ubiquitous



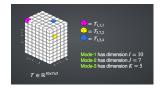


Figure: Movie Recommendation

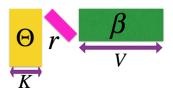
Figure: EHR Data



Figure: Social Network

Background

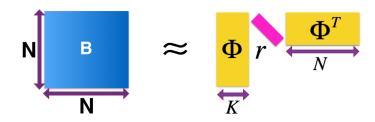
00000



•
$$y_{dw} \approx \sum_{k=1}^{K} r_k \theta_{dk} \beta_{wk}$$
.

Background

00000

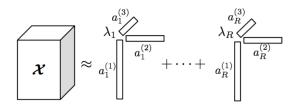


•
$$b_{nm} \approx \sum_{k=1}^{K} r_k \phi_{nk} \phi_{mk}$$
.

Tensor Decomposition

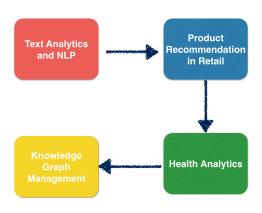
Background

00000



• $x_{d_1d_2d_3} \approx \sum_{k=1}^{K} \lambda_k a_{d_1k}^{(1)} a_{d_2k}^{(2)} a_{d_2k}^{(3)}$.

Talk Outline



Latent Semantic Analysis

- Probabilistic Latent Semantic Analysis
- Latent Dirichlet Allocation (topic models)

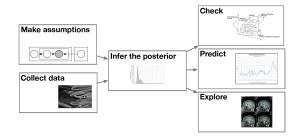


Figure: Objectives of Topic Models

Discovery of Topics (Insights?)

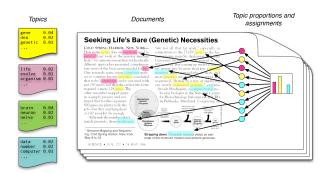


Figure: Assignment of an article to topics

Visualization of Topics

 Data: Collection of Science from 1990–2000, 17K documents, 11M words, 20K unique terms

1	2	3	4	5	
dna	protein	water	says	mantle	
gene	cell	climate	researchers	high	
sequence	cells	atmospheric	new	earth	
genes	proteins	temperature	university	pressure	
sequences	receptor	global	just	seismic	
human	fig	surface	science	crust	
genome	binding	ocean	lke	temperature	
genetic	activity	carbon	work	earths	
analysis	activation	atmosphere	first	lower earthouakes	
two	kinase	changes	changes years		
6	7	8	9	10	
end	time	materials	dna	disease	
article	data	surface	rna	cancer	
start	two	high	transcription	patients	
science	model	structure	protein	human	
readers	fo fo	temperature	site	gene	
service	system.	molecules	binding	medical	
news	number	chemical	sequence	studies	
card	-	molecular	proteins	drug	
circle		**	specific	normal	
letters	_	uniemity	sequences	drugs	
11	12	13	14	15	
vears	species	protein	cells	space	
million	evolution	structure	cell	solar	
million ago	evolution population	structure proteins	cell virus	solar observations	
million ago age	evolution population evolutionary	structure proteins two	cell virus hiv	solar observations earth	
million ago age university	evolution population evolutionary university	structure proteins two amino	cell virus hiv infection	solar observations earth stars	
fmillion ago age university north	evolution population evolutionary university populations	structure proteins two amino binding	cell virus hiv infection immune	solar observations earth stars university	
fmillion ago age university north early	évolution population evolutionary university populations natural	structure proteins two amino binding acid	cell virus hiv infection immune human	solar observations earth stars university mass	
fmillion ago age university north early fig	évolution population evolutionary university populations natural studies	structure proteins two amino binding acid residues	cell virus hiv infection immune human antigen	solar observations earth stars university mass sun	
million ago age university north earry fig evidence	evolution population evolutionary university populations natural studies genesis	structure proteins two amino binding acid residues molecular	virus hiv Infection immune human antigen infected	solar observations earth stars university mass sun astronomers	
fmillion ago age university north early fig	évolution population evolutionary university populations natural studies	structure proteins two amino binding acid residues	cell virus hiv infection immune human antigen	solar observations earth stars university mass sun	
million ago age university north early fig evidence record	evolution population evolutionary university populations natural stades genetic today 17	structure proteins two amino binding acid residues molecular structural	Cell Virus hiv Infection Immune human antigen Infected wrai	solar observations earth stars university mass sun astronomers telescope 20	
million ago age university north early fig evidence record 16 fax	evolution population evolutionary university populations natural studies geneti tompy 17 Cells	structure proteins two amino binding acid residues molecular structural 18 energy	Cell Virus hiv Infection Immune human antigen infected wrat 19 research	solar observations earth stars university mass sun satroomers telescope 20 neurons	
million ago age age university north early fig evidence record 16 fax manager	evolution population evolutionary university populations natural states period tomp 17 Cells Cell	structure proteins two amino binding acid residues molecular structural 18 energy electron	Cell Virus hiv Infection Immune human antigen Infected vrail 19 research science	solar observations earth stars university mass sun astronomen telescope 20 neurons brain	
million ago age university north early tg evidence record 16 fax manager science	evolution population evolutionary university populations natural statist period 17 Cells cell gene	structure proteins two amino binding acid residues molecular structural 18 energy electron state	Cell virus hiv infection immune human antigen infected virai 19 research science national	solar observations earth stars university mass sun astronomens telescope 20 neurons brain cetts	
million ago age age university north early ing evidence record 16 fax manager science aaas	evolution population evolutionary university populations natural states genetic torque of the color of the co	structure proteins two amino binding acid residues molecular structural 18 energy electron state light	Cell Virus hiv infection immune human artigen infected viral 19 research science national scientific	solar observations earth stars university mass su actoriones statescope 20 neurons brain cets activity	
million ago age university north early tg evidence record 16 fax manager science aaas advertising	evolution population evolutionary university populations and states and state	structure proteins two amino binding acid residues residu	cell virus hiv infection immune human antigen infected viral 19 research science national scientific somitific	solar observations earth stars university un	
million ago ago ago university north porth ing evidence record 16 fax manager science aaas advertising sales	evolution population evolutionary university populations stated s	structure proteins two amino binding acid residues notecular structural 18 energy electron state light quantum physics	cell virus hiv infection immune human aritigen infected viral 19 research science national scientific solvette row	solar observations earth stars university mass sun accordens telescope 20 neurons brain cets activity fig channels	
million ago ago ago university north early early early existence record 16 fax manager sclears advertising sales member	evolution population evolutionary university populations exclusionary university populations exclusions exclusions exclusions trackets wrong 17 Cells Cell Gene genes expression development mutant	structure proteins two amino binding acid residues residu	cell virus hiv infection immune human infected viral 19 research science scientific scientific scientific scientific scientific scientific scientific scientific	solar observations earth stars university mass sections to the section of the sec	
million ago ago ago university north porth ing evidence record 16 fax manager science aaas advertising sales member recruitment	evolution population evolutionary university populations stated s	structure proteins two amino binding acid residues residu	cell virus hiv infection immune human antigen infected viral 19 research science national scientific	solar observations earth stars university mass sun acronome telescope 20 neurons brain cells activity fig channels university cortes	
million ago ago ago university north early early early existence record 16 fax manager sclears advertising sales member	evolution population evolutionary university populations exclusionary university populations exclusions exclusions exclusions trackets wrong 17 Cells Cell Gene genes expression development mutant	structure proteins two amino binding acid residues residu	cell virus hiv infection immune human infected viral 19 research science scientific scientific scientific scientific scientific scientific scientific scientific	solar observations earth stars university mass sections to the section of the sec	

Domain Feedback into Topic Modeling

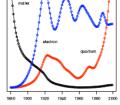
Operation	Meaning
Must-Link (school,college)	\forall topics t , $P(school t) \approx P(college t)$
Cannot-Link (school,cure)	no topic t has $P(\frac{school}{t})$ and
	P(cure t) both high

split	 [go school into college] vs [cancer free cure well] → Must-Link among words for each concept → Cannot-Link between words from different concepts
merge	[love marry together boyfriend] in one topic [married boyfriend engaged wedding] in another → Must-Link among concept words
isolate	 the year in 2008 in many wish topics → Must-Link among words to be isolated → Cannot-Link vs other Top N words for each topic

Figure: Domain Feedback

What if the documents change with time?





"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

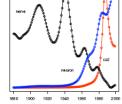
1999 Quantum Computing with Electrons Floating on Liquid Helium

Figure: Evolution of Topic 1

What if the documents change with time?



"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

1963 Erroriess 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: Evolution of Topic 2

Neural Word Embedding: Word2Vec (Deep Learning!)

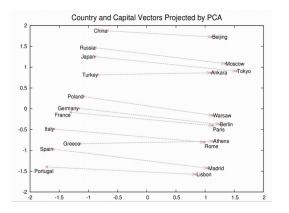


Figure: Illustration of Word2Vec

Illustration of Word Embedding ..

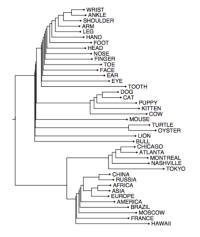


Figure: Illustration of Word Embedding

Illustration of Word Embedding

Nearest words to frog: 1. frogs 2. toad

3. litoria 4. leptodactylidae 5. rana 6. lizard



Figure: Illustration of Word Embedding

Recommender Systems

- Collaborative filtering
- Content-based filtering
- Content-based collaborative filtering

Musical instruments (Amazon)			Video games (Amazon)						
drums	strings	wind	microphones	software	fantasy	nintendo	windows	ea/sports	accessories
cartridge	0	reeds	mic	software	fantasy	mario	sims	drm	cable
sticks	violin	harmonica	microphone	interface	rpg	ds	flight	ea	controller
strings	strap	cream	stand	midi	battle	nintendo	windows	spore	cables
snare	neck	reed	mics	windows	tomb	psp	xp	creature	ps3
stylus	capo	harp	wireless	drivers	raider	wii	install	nba	batteries
cymbals	tune	fog	microphones	inputs	final	gamecube	expansion	football	sonic
mute	guitars	mouthpiece	condenser	usb	battles	memory	program	nhl	headset
heads	picks	bruce	battery	computer	starcraft	wrestling	software	basketball	wireless
these	bridge	harmonicas	filter	mp3	characters	metroid	mac	madden	controllers
daddario	tuner	harps	stands	program	ff	smackdown	sim	hockey	component

Figure: Users' purchase pattern described by latent groups

What if the user-item interaction changes with time?

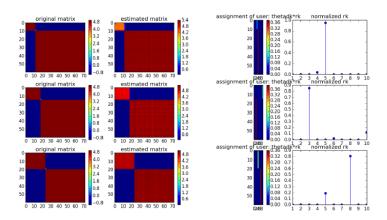


Figure: Static Recommendation

What if the user-item interaction changes with time?

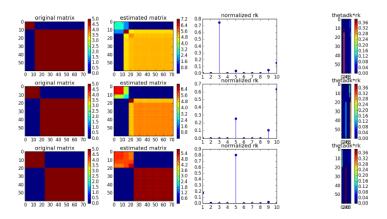


Figure: Dynamic Recommendation

Motivation – Manual Phenotyping of EHR Data

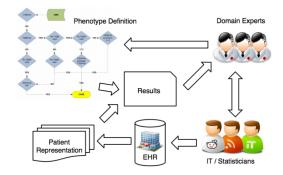


Figure: Manual Phenotyping of EHR Data

Automated Phenotyping

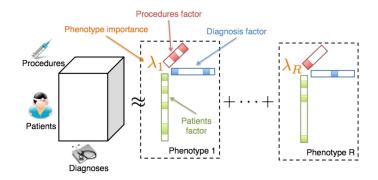


Figure: Automated Phenotyping of EHR Data

Phenotypes Discovered

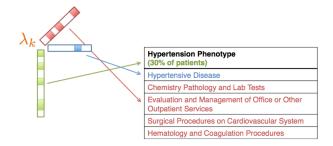


Figure: Phenotypes Discovered

What if the patients' condition changes?

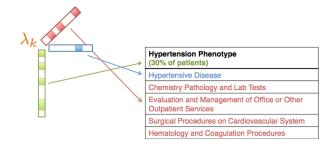


Figure: Phenotypes Discovered

Social Network Analysis

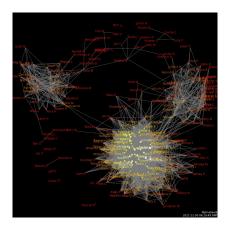


Figure: Social Network as a Graph

Eigen Decomposition

What if the users' association changes with time?

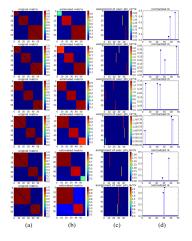


Figure: Static Modeling

What if the users' association changes with time?

Eigen Decomposition

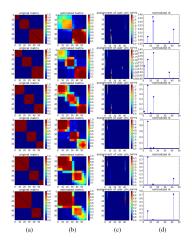


Figure: Dynamic Modeling

Knowledge Graph as Rank-reduced Tensor

- Significant reduction in storage
- Reduction in retrieval time/complexity

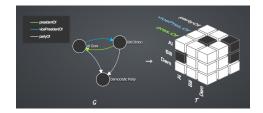


Figure: Knowledge Graph as Tensor

Conclusion

- Same mathematical tool used in solving problems in multiple domains
- Customization of a tool requires domain knowledge.
- Automated model selection is difficult.
- ML is not a magic, it requires a human to create one.

Questions?





Conclusion O