Jeffrey Yan
LING 413
Final Project and Extra Credit

Term Project: Topic Models for Healthcare

Problem #1:

Gender	Sentiment (count and %)		Total (count and %)
	Positive	Negative	
Female	2953, 61.44%	1853, 38.56%	4806, 100%
Male	10616, 67.99%	4999, 32.01%	15615, 100%

The length of the smallest review was 0 tokens, and the length of the largest review was 4948 tokens. The average length of reviews was 341 tokens.

Problem #2:

No.	Questions	RateMD corpus	Healthcare company's corpus (i.e., reference corpus)
1	What is the language variety of the corpus (i.e., genre)?	Reviews written by RateMD users	Reviews written by patients of the company's clinics
2	What is the size of the corpus?	20,421 reviews	500,000 reviews
3	What meta-data is provided with the reviews?	Doctor's name, gender, clinic location, specialization, review sentiment	Doctor's name, gender, clinic location; review sentiment
4	What socio-demographic information is provided about the patients who wrote the reviews?	None	Gender, age, economic and educational status
5	Is the corpus balanced along the meta-data	No; there are more reviews for male doctors, and more	No (the dimensions are not uniformly distributed; they

	dimensions considered? (look only at sentiment and gender)	positive reviews than negative reviews	exhibit a natural distribution)

One important disadvantage is that the reviews by RateMD users are anonymous, so there is no information about the patients who wrote the reviews. In fact, there is no proof that the review writers were patients. Assuming that the large majority of reviews are legitimate, this dataset can still be very useful due to its large size.

Task #2a

When cleaning the corpus, I chose to remove punctuation and numbers, lowercase everything, and remove stopwords. I also removed “doctor” and “dr” from the result, since those words appeared often and wouldn’t be a satisfactory result.

Without lemmatization

Without lemmatization, pre-processing took 0.5 seconds. The term dictionary contained 31,087 tokens.

With Set 1, the runtime was 2m57s. With Set 2, the runtime was 3m7s.

My results for Set 1, as well as an attempt at a manual label for each topic, are below:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Ambiguous	Ambiguous	Surgery	Children	Good service
would told never go back get see said even didnt	manner bedside patel good cardiologist tooth date listener heart annoyed	surgery pain surgeon cancer life procedure performed went back years	daughter son us child baby pregnancy children delivered old first	staff great feel made experience office professional friendly comfortable amazing

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Ambiguous	Recommended	Appointment	Treatment	Ambiguous
time always best great years	recommend would highly anyone excellent	office staff wait time appointment	medical treatment patient patients care	pediatrician weight stuff gyn pulled

patients patient staff feel like	staff caring knowledgeable care helpful	waiting rude get phone minutes	health condition years problem doctors	migraines theyre card environment welcoming

There were a few vague topics such as Topic 1. For Topic 6, it feels like it could be positive, but there were not enough positive words for me to make a decision.

My results for Set 2:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Good service	High blood pressure	Recommended	Recovery	Ambiguous
smart lucky brain punctual average hearing forget considerate non cure	mother high blood pressure risk highest attack responded medicare denied	manner would bedside recommend feel great made anyone staff first	surgery pain went hospital procedure back cancer months vegas performed	years best life ever one doctors many patel ive would

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Ambiguous	Ambiguous	Good service	Ambiguous	Surgery
would told never office back get see said even time	billing super respectful annoyed prepared stupid price single corrected offers	staff recommend caring knowledgeable excellent great highly always helpful professional	time patients always wait like patient get really good never	surgery surgeon results happy knee excellent performed satisfied done procedure

Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Bad service	Ambiguous	Ambiguous	Ambiguous	Costs

medical md clinic mean terrible thyroid agree practice across cold	ear drugs referrals w med diabetes reschedule anyway trained correctly	skin finding u anxiety lee pm dermatologist depression incredibly disorder	top guy broken ankle turn house stating whats demeanor nasty	insurance bill company teeth pay polite receive oral group smile

Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
Treatment	Children	Ambiguous	Bad service	Ambiguous
daughter treatment diagnosis therapy condition options medicine pediatrician treatments conditions	son great feel children us makes kids comfortable knows love	god earth dealing bless town small retire switch delivering everytime	staff office rude front patients service good unprofessional worst poor	eye brown weight watch expert eyes quack smith loss please

I was not sure about labeling some topics even though I felt like there might be a loose correlation between words. Topic 20 had three eye-related words, but I didn't feel like it was enough.

Noun lemmatization

With noun lemmatization using WordNetLemmatizer, pre-processing took 3.1 seconds.
The term dictionary contained 28,950 tokens.
With Set 1, the runtime was 2m56s. With Set 2, the runtime was 3m6s.

My results for Set 1, as well as an attempt at a manual label for each topic:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Ambiguous	Ambiguous	Recommended	Ambiguous	Appointment
life saved dentist tooth teeth dismissed	informative father trained root whenever cavity	staff recommend would great highly excellent	manner bedside side bed sense sinus	office time never staff get see

uncaring glyman car pap	hormone avoided approachable rx	caring knowledgeable professional helpful	humor utmost recomend le	patient would appointment call

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Great service	Surgery	Ambiguous	Pregnancy	Family
time patient always care best take year feel great like	surgery surgeon procedure went result cancer performed breast done day	pain problem year told back would went month insurance said	baby pregnancy first pregnant medical child second life woman made	u daughter husband son heart child year hospital old mother

It was difficult to manually label many topics. For example, Topic 1 had both negative words (“uncaring”) and positive words (“saved”). I was not sure if I should label some topics as “ambiguous” if there were only about three words that seemed to fit. So, even though I labeled a few as “ambiguous”, the terms under that topic were not completely irrelevant.

For 20 topics:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Pregnancy	Great service	Fun	Recommended	Ambiguous
baby pregnancy delivered birth first made child ob high teeth	time feel always question take make patient care like great	sense dermatologist humor diabetes oral considerate afford complex mayo cured	recommend would highly anyone definitely friend knowledgeable wonderful recommended extremely	mark surgical smith gp reviewed removal shoe repair addiction nature

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Ambiguous	Appointment	Treatment	Ambiguous	Great service
year life husband son many old ago saved since month	time wait appointment hour patient see always minute waiting get	treatment patient problem medical condition health diagnosis concern issue year	told said would back went didnt never time see could	care excellent physician patient family caring compassionate knowledgeable staff friend

Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Ambiguous	Surgery	Bad service	Ambiguous	Ambiguous
breast cancer face skin look result nose removed sinus antibiotic	surgery surgeon pain performed procedure went back knee great foot	mother front rude attitude desk tooth unprofessional away arrogant horrible	procedure hospital charge letter stated would completely caused also insurance	ever best ive one seen worst met far life illness

Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
Ambiguous	Ambiguous	Good service	Staff	Good service
u daughter child ear funny cardiologist annoyed	know dont patient care help go want	manner bedside good great pediatrician bed parent	office staff call patient rude insurance never	staff helpful great friendly office professional knowledgeable

dangerous husband born	doesnt like pain	kid listener allergy	get phone would	caring always knowledgable

These results seemed to have more relevant results, especially related to good service. So, I think that Set 2 with 20 topics generates better topics.

Verb Lemmatization

Verb lemmatization using WordNetLemmatizer took 3.4 seconds. The term dictionary contained 27,234 tokens.
With Set 1, the runtime was 2m45s. With Set 2, the runtime was 3m16s.

My results for Set 1, as well as an attempt at a manual label for each topic:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Dentist	Recommended	Ambiguous	Treatment	Emotions
dentist teeth mom weight root tube wisdom gain sincere cavity	care staff great time recommend best always take years would	go tell would husband say hospital daughter baby could look	medical treatment patient condition diagnose test diagnosis patients treat health	informative luck disorder happier expectations utmost augmentation reccomend emotional strongly

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Surgery	Medication	Appointment	Costs	Good service
surgery surgeon procedure perform recommend would pain glyman result knee	pain help prescribe vegas meds cause medication back life drug	time get office go see call wait staff never tell	insurance pay bill charge company cost cover receive ear service	manner bedside outstanding manners timely genuine respectful ankle lack focus

These topics could be manually labeled effectively, except for Topic 3.

My results for Set 2:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Pain	Ambiguous	Ambiguous	Ambiguous	Recommended
pain vegas suffer back help severe las complain cause meds	heart mother father cardiologist hospital speak family brain die medical	go get tell call office see time say never wait	poor lack md communicate retire routine improve lee communication doubt	recommend would make highly feel anyone comfortable staff great go

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Ambiguous	Good service	Treatment	Good service	Family
prescribe review drug sign comment look read medication bad negative	staff office time helpful wait always great friendly good son	explain treatment procedure detail options procedures result receive experience give	care excellent manner bedside patients patient physician knowledgeable great staff	us daughter wife nose old husband pediatrician earth ms bear

Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
Ambiguous	Surgery	Thankful	Thankful	Service
surgery work breast go result back years surgeon	surgery hospital surgeon go perform husband eye procedure	cancer god skin thank daughters thyroid scar remove	life save say side cant enough doc years	question answer ask time concern baby call always

look pain	months days	bless incredible	bed thank	rush phone

Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
Good	Ambiguous	Ambiguous	Costs	Bad experience
best care know ever time take like see really always	young glad spot face w assistants house whats botox wall	years care time see take patient many find health issue	insurance bill pay company cover charge extra cost receive accept	rude ever worst arrogant experience awful horrible terrible attitude condescend

For k=20, the topics were slightly easier to label, and there seemed to be many with a general narrative.

Problem #3:
Lemmatization does not take much time compared to the runtime of the LDA models, so at least for this amount of data, the runtime is not an issue. Compared to no lemmatization, I did not see a significant difference in the goodness of topics for k=10, although verb lemmatization may have been slightly better for both k=10 and k=20. The comparison is complicated by manual labeling of each topic, which was very difficult because there is often no clear answer. So, if I labeled everything again, I would likely end up with different results. There were very rarely topics that were as clear as Topic 20 for verb lemmatization, where almost all words are negative. It may be possible that adding more words to the list of stopwords could change this outcome.

Task#2b
The runtime of ccLDA for 10 topics and 2000 iterations was 36 minutes, far longer than the previous runtimes. The results were impressive, with the positive reviews (Collections 0 and 2) clearly showing positive terms, and the negative reviews (Collections 1 and 3) having mostly negative terms. For that reason, these results were better than the ones I had obtained in Task #2a. The overall top 10 words were not always as clear, such as in Topic 2, so those were much noisier. This might be fixed in the pre-processing with more stopwords and names removed. The results are in the table below:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
patients	like	dreher	years	surgery

dreher care patient medical health one physician ever practice doctors knowledge well needs person find medicine treatment seems may -Collection 0 caring best excellent compassionate cares truly knowledgeable listens kind difficult personable competent outstanding genuinely intelligent rare physician lucky wonderful thoughtful -Collection 1 unprofessional poor lack license malpractice appears zero impatient inappropriate careless decision patronizing ego lawsuit abuse files revoked disrespectful physical confidentiality -Collection 2 best caring cares truly listens well compassionate	dreher dont go know im ever want doesnt going say one never child people see think baby even need -Collection 0 always love shes best amazing delivered awesome children pregnancy obgyn baby fantastic kids knows sweet honest pregnancies top high delivery -Collection 1 worst rude horrible shes money attitude insensitive dog weight yelled basically lose whatsoever disrespectful hodges rudest afraid hurry zyrtec nahra -Collection 2 hes love best always great wonderful amazing	would staff recommend manner bedside anyone helpful good extremely experience great family friendly patient caring care ever best nice -Collection 0 highly great knowledgeable excellent wonderful recommend professional friends caring understanding smart personable amazing recommended jacobson pleasant gentle courteous informative reassuring -Collection 1 rude worst uncaring horrible stories rawls lacks judgemental horror changing sarcastic nasty lacked futrell bonakdar abortion misdiagnoses snippy anger blau -Collection 2 great highly excellent caring knowledgeable kind recommend	doctors dreher problem ive seen one problems issues found many several find since times treatment ever specialist seeing help -Collection 0 years best great ive shes family love knowledgable well helped seeing listens many concerns gyn drive addressed obgyns overall weiss -Collection 1 cold awful test shes worse condescending arrogant wrong incompetent basic gyno sometimes refuses harwer bother false questioned cutlip sesslar swamy -Collection 2 dreher years recommend many always would problems	pain dreher would procedure back went surgeon done left one months breast two removed look performed knee second work -Collection 0 skin dermatologist happy pleased cancer lee treatments friend recently beautiful bonakdar dermatologists skilled dermatology javan happier melanoma sister implants annual -Collection 1 skin woman cancer routine please insisted mri gmitter fibroids suppose cost pulled supposed yearly iud levy hanging kolb responsibility formal -Collection 2 surgery surgeon amazing performed recommend recovery happy

always	really	wonderful	best	highly
excellent	years	professional	anyone	shakiba
physician	family	knowledgable	recommended	excellent
kind	children	thorough	highly	results
outstanding	delivered	personable	care	surgeries
knowledgeable	son	manner	patient	rivadeneira
trust	knows	compassionate	extremely	job
top	awesome	pleasant	thorough	schwartz
takes	ive	friends	friends	thank
understanding	daughter	concerns	helped	ago
really	guy	definitely	seeing	procedure
family	kids	efficient	say	plastic
dreher	obgyn	awesome	need	explained
-Collection 3	-Collection 3	-Collection 3	-Collection 3	-Collection 3
doesnt	hes	rude	worse	surgery
man	guy	terrible	worst	never
lack	money	worst	go	still
license	horrible	arrogant	failed	nothing
opinion	dont	poor	dismissive	worse
instead	bad	uncaring	opinion	bad
life	worst	horrible	response	mistake
unprofessional	stay	awful	arrogant	fix
beware	man	condescending	something	spine
board	wrong	unhelpful	idea	away
alone	please	cold	agree	spinal
lacks	believe	attitude	seeking	side
seems	life	enemy	disappointed	mri
ego	jerk	whatsoever	cost	xrays
rudest	heard	inconsiderate	someone	cut
total	basically	wish	expensive	bone
towards	maybe	abrupt	notes	causing
angry	place	lazy	favor	disc
zero	treat	shame	pompous	show
comes	drug	warn	wasted	nerve

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
time	office	dreher	dreher	didnt
wait	staff	us	time	back
office	call	husband	questions	went
appointment	insurance	hospital	feel	told
see	get	life	made	said
get	phone	son	like	would
minutes	never	medication	felt	see
waiting	even	daughter	concerns	even
room	calls	help	answer	could
dreher	rude	old	sure	came
long	new	could	never	got
staff	back	care	really	asked
never	called	years	experience	first
seen	results	mother	make	go
hours	would	year	also	called
patients	service	would	always	get
hour	patient	meds	care	wanted
one	make	heart	listen	wasnt
even	appointment	took	rushed	blood
late	medical	take	explain	another
-Collection 0	-Collection 0	-Collection 0	-Collection 0	-Collection 0
always	dreher	life	takes	really
helpful	staff	helped	caring	time
great	efficient	thank	thorough	took
thorough	helpful	saved	feel	found
takes	mccormack	kolb	comfortable	good
knowledgeable	able	grateful	professional	everything
spends	nice	thanks	friendly	wish

friendly	hollie	daughter	kind	right
wonderful	use	pediatrician	makes	hope
sometimes	working	expertise	ease	personally
doctors	satisfied	incredible	easy	beyond
lot	appreciate	daughters	explains	glad
needed	clean	miller	explained	weight
need	taylor	world	attentive	knew
worth	calls	wonderful	warm	barry
though	continue	impressed	listened	sure
willing	welcoming	woman	trust	little
quickly	available	today	confident	bit
within	busy	wife	answered	plan
answers	glazer	women	punctual	hamersley
-Collection 1	-Collection 1	-Collection 1	-Collection 1	-Collection 1
waited	refused	let	unhelpful	told
unprofessional	terrible	ignored	waste	said
hour	stay	pills	seem	pap
money	week	report	angry	refill
pm	later	nasty	disappointed	control
left	unless	cough	impersonal	I
unfortunately	cancelled	fluid	acted	asked
avoid	cash	misdiagnosed	dismissive	desk
neither	months	spinal	speaking	medicine
dermatologist	melissa	needless	lost	pill
unacceptable	disorganized	mri	whitney	pelvic
reschedule	paid	mistakes	supposedly	records
apology	requested	quack	probst	birth
dirty	leaving	means	unfriendly	seconds
arnold	hollie	fine	uninterested	diet
terms	code	file	thorough	despite
finally	messed	legal	value	ridiculous
flu	yu	lazy	another	obgyn
unorganized	kristensen	stated	bad	mccormack
products	welch	warning	gliksman	wasted
-Collection 2	-Collection 2	-Collection 2	-Collection 2	-Collection 2
time	dreher	life	staff	first
dreher	staff	thank	feel	like
always	office	pain	comfortable	one
takes	good	years	great	better
questions	great	saved	friendly	took
good	helpful	helped	takes	right
nice	care	grateful	professional	time
doctors	see	man	ease	able
patient	friendly	cancer	everything	day
patients	work	treatment	helpful	could
also	personally	best	wonderful	hospital
needed	doc	god	makes	husband
easy	person	back	answered	way
sometimes	everyone	treated	experience	going
helpful	nurse	referred	explained	every
knowledgeable	nurses	today	thorough	much
lot	clean	free	explains	sure
need	people	quality	took	well
worth	impressed	knowledge	everyone	problem
wait	needs	thanks	well	listened
-Collection 3	-Collection 3	-Collection 3	-Collection 3	-Collection 3
another	unprofessional	pain	didnt	told
hours	called	man	rude	said
left	paid	refused	like	asked
room	refused	even	seemed	mri
waited	money	told	waste	test
hour	wife	wrong	unprofessional	nurse
back	company	misdiagnosed	acted	telling
minutes	incompetent	er	interested	tried
got	sent	med	seem	stated
finally	pay	pills	difficult	report
reschedule	unless	though	return	later
anything	months	ignored	cold	already
thing	charged	caused	became	neurologist

trying walked dirty obviously pm paperwork explain	avoid billing correct script reason elsewhere billed	addict body second turned records meds complained	seems little uncomfortable comment wasting communication behavior	comp insisted saying letter medicine complete thats

The runtime for ccLDA for 20 topics and 2000 iterations was 40m45s. These results were noticeably not as good as the previous output, such as in Topic 1 Collection 0 where the results are very vague. However, the terms under each collection were mostly very good, with most of them being positive for positive reviews, and negative for negative reviews. The results are in the following table:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
dreher ever one ive doctors seen years far never met since many dentist going absolutely seeing times gone around come -Collection 0 best mclaughlin remembers gentle piper thoughtful notch elle de la kaye adore et doyle trs en specialty shah que je -Collection 1 worst flu mistake swamy	time questions dreher concerns answer always patient listen explain ask things question everything rushed answers answered issues talk understand took -Collection 0 takes caring thorough listens explains spends extra attentive punctual thoroughly loved gives eubanks wonderfully intelligent perfect easy types allen kakoulas -Collection 1 waste uncaring impersonal hodges	feel made like make felt dreher sure really makes going much well never better im making feeling way every understand -Collection 0 amazing comfortable great ease easy listened earth taylor polite incredible young fortunate soviero awesome salzman understands truly smiling dreher wonderful -Collection 1 worst probst sesslar privacy	go never see like else need want even someone would know back dont work wouldnt people treated get anything way -Collection 0 love kind sweet koster amazing beyond miss weiss area little compared upbeat previous kemp searched yu gyns beer busy montgomery -Collection 1 unprofessional rude stories careless	dreher care patient physician years family professional well doctors health many found issues practice kind find extremely several primary compassionate -Collection 0 knowledgeable thorough personable compassionate knowledgable smart recommended intelligent pleasure woman considerate understanding approachable responsive thoughtful use block diagnostician greatly ages -Collection 1 dismissive rawls handle indifferent

gyno	incredibly	messed	barclay	ira
gum	preached	offensive	please	practitioner
fault	unresponsive	littleton	hudson	sarcastic
uninterested	leopold	barkan	inappropriate	michaelson
mouth	adofo	encounter	ten	exploring
mis	hallway	frustrating	messed	january
fitzgerald	assuming	another	idiot	liability
tonya	results	julia	inexperienced	warranted
excess	unkind	mahon	joykutty	inaccurate
thinks	demeaning	customers	falling	lying
infact	hell	ptsd	helpless	factory
could	incomplete	dnc	disgrace	guinea
presciption	overheard	kulkarni	caterson	un
chipped	compensated	runny	recommend	interrogated
eskam	unprofessionalism	men	abusive	alnief
harwer	insufficient	obviously	hateful	roomdid
-Collection 2	-Collection 2	-Collection 2	-Collection 2	-Collection 2
best	takes	feel	dreher	caring
hes	always	comfortable	anyone	excellent
gentle	caring	staff	would	knowledgeable
always	listens	great	wonderful	compassionate
fantastic	questions	wonderful	kind	thorough
ive	knowledgeable	ease	caring	kind
need	explains	everything	friend	also
amazing	took	love	humor	extremely
teeth	kind	happy	cant	skilled
drs	answers	easy	sense	outstanding
professional	thorough	hands	treats	exceptional
dental	answered	confident	trust	top
davis	explain	explained	everyone	intelligent
job	spends	makes	respect	warm
smile	thoroughly	calm	whole	lucky
understanding	available	taken	good	considerate
incredible	professional	visits	else	notch
newman	gives	appreciated	helps	fortunate
respectful	detail	explaining	anywhere	personable
man	every	positive	love	diagnostician
-Collection 3	-Collection 3	-Collection 3	-Collection 3	-Collection 3
worst	rude	bad	rude	opinion
stay	doesnt	worse	horrible	another
away	waste	didnt	condescending	ego
rudest	seemed	arrogant	didnt	mckay
well	interested	problem	please	impossible
lies	wasted	jerk	nasty	negligence
sloppy	wasting	cold	shouldnt	levine
reputation	response	mistake	enemy	oxygen
disability	annoyed	bother	yelled	disrespect
sucks	defensive	sarcastic	told	downright
temper	distracted	listen	word	catheter
simple	seem	ok	horribly	notified
difficulty	argumentative	communication	rudely	partner
cavities	internet	creepy	unless	turns
misfortune	snow	condescending	loud	unwilling
type	sign	humiliated	report	factory
inappropriate	wo	attempted	mean	disease
sedated	business	prepared	hateful	language
unfounded	smoke	gun	repeated	unreliable
relative	intake	vulnerable	badly	outpatient

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
time wait appointment	staff office dreher	call get called	test tests results	first baby went

room minutes see waiting long hours hour waited visit late office times seen another get exam least -Collection 0 knowledgeable always sometimes wonderful worth quickly quick within addressed rawls yearly available emergencies checkup quite mind though seeing kirsten efficiently -Collection 1 unprofessional unhelpful worse chart arnold yearly unfriendly vegan whitaker barry sojourner pelvic canceled diet dingy showed apologies shell slice shouldnt -Collection 2 time patient dreher see get wait always need long	nice rude always front nurse great experience extremely friendly good desk service knowledgeable also practice needs helpful -Collection 0 friendly helpful professional pleasant efficient courteous jacobson welcoming prompt machado knowlegeable fantastic easy organized midwife hill accommodate samuels accomodating wagner -Collection 1 rude incompetent hollie disorganized nasty pincus nahra gliksman gulati blackburn glazer petok wedlow slightly strictly snippy cap worthless dated neither -Collection 2 helpful friendly professional knowledgeable everyone experience pleasant courteous efficient	office appointment phone back calls day new days next return appt never nurse see even make would -Collection 0 always personally great promptly glazer fit wonderful kim sick allan though booked wedlow smile croft jensen female hope alexander cool -Collection 1 terrible desk requested oh ignored elsewhere excuse werner insensitive accepting didnt serious derm nath asking clerk mammogram impatient bruscato skip -Collection 2 dreher staff day office one right personally quickly see	blood diagnosis problem saw diagnosed opinion condition went needed second specialist found wrong symptoms treatment another took -Collection 0 caring warm explained able rare terrific impressed detailed suggestions provided schoenberg improvement carefully small morgan style walker pipan hamilton checks -Collection 1 awful pap condescending yu lost wasted stated lazy tap goldhaber cutlip dosage doshi smears steer varhola cdc misinformed means untreated -Collection 2 treatment problem thorough problems condition recommended explained doctors aviv	pregnancy pregnant going would weeks also due saw months even got birth everything second high every experience -Collection 0 delivered obgyn pregnancy babies helped supportive beautiful pregnancies hamersley calm deliver evans fantastic delivery throughout takoudes hope encouraging feigel adofo -Collection 1 told I died stay suppose levy iud placed smell pull easily besson unsympathetic bc depo dose table mistreatment weeks dilated -Collection 2 amazing delivered shakiba dreher obgyn helped wonderful first well

needed	personable	follow	diagnosed	healthy
worth	kind	issue	cancer	happy
sometimes	polite	home	kakani	pregnancy
little	knowledgable	hospital	immediately	awesome
doctors	overall	kind	options	brown
spent	tye	work	treated	around
every	super	night	diagnosis	husband
though	informative	etc	several	little
much	prompt	used	found	throughout
within	nurses	following	knew	mind
drive	punctual	de	impressed	pregnancies
-Collection 3	-Collection 3	-Collection 3	-Collection 3	-Collection 3
back	rude	told	done	told
finally	unprofessional	asked	nothing	blood
got	incompetent	go	chart	apparently
didnt	unhelpful	stated	dismissive	mess
obviously	uncaring	still	requested	signs
walked	dirty	unprofessional	refused	alone
ask	disorganized	refused	failed	hit
showed	unfriendly	letter	send	threatened
place	training	saying	arrogant	insensitive
reschedule	hippa	telling	idea	jabara
gross	unpleasant	manager	mris	dx
consultation	unresponsive	goodman	tried	iui
appeared	inconsiderate	canceled	incorrect	back
slow	type	apt	advised	induce
acted	late	needless	insisted	moruzzi
martinez	employees	practitioner	grua	ivey
outrageous	trash	paperwork	prostate	listen
waited	okay	med	frustrated	steroids
wife	clueless	stand	incompetence	allergic
hallway	system	screamed	fatigue	mistake

Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
insurance	surgery	would	told	patients
office	pain	recommend	said	care
visit	back	dreher	didnt	patient
money	hospital	anyone	went	medical
medical	went	surgery	back	dreher
pay	months	procedure	asked	needs
would	knee	results	got	health
even	two	happy	could	practice
refused	left	definitely	came	cares
bill	surgeon	breast	even	knowledge
billing	still	experience	left	trust
work	neck	look	see	concern
paid	able	done	wanted	compassion
company	later	looking	took	best
records	weeks	great	another	treatment
service	surgeries	surgeon	go	personal
could	year	work	never	person
information	foot	extremely	wasnt	respect
told	performed	well	right	medicine
cost	much	recommended	gave	many
-Collection 0	-Collection 0	-Collection 0	-Collection 0	-Collection 0
pleased	excellent	highly	found	caring
mccormack	awesome	professional	skin	excellent
trust	friend	family	happy	truly
hollie	skilled	friends	dermatologist	cares
knowledgeable	arnold	everyone	lee	outstanding
continue	nicely	reassuring	glad	satisfied
awesome	healing	lot	barry	provides
plan	painless	javan	dermatologists	genuinely

brilliant issue informative confidence sarah cigna til holistic tattoo superb real monique -Collection 1 worst license melissa supposed kristensen welch straw nexium neither incorrectly refuses revoked refund note errors lawsuit notify oath besides vanessa -Collection 2 dreher care staff pleased satisfied received without team every happy clinic attention continue services necessary service impressed procedures unlike detailed -Collection 3 wife dollars lied paid charged refund fraud request refuses sorry dishonest accident explanation thousands	levy littleton reccomend docs colyer ratner pros cons mohs barclay system understand -Collection 1 hysterectomy iv fluid kolb experiencing appalled urine vaginal vagina bags discover woo watching contacted hired chastised intimidate liters discuss rely -Collection 2 surgery surgeon recovery performed first free ago recommended replacement fusion hand thank surgeries job surgeons also yoon team outcome orthopedic -Collection 3 mri cut comp nothing fusion disc please bone er nerve pa screws xrays denied	happier understanding confidence expert recommendations chun fantastic cooper mary conservative dallas jones -Collection 1 disrespect ruined pezzello bonakdar autistic nausea heartless toxicity puss wessman scam snappy intake torment honey k dermatology disfigured traumatized jawline -Collection 2 highly recommend staff amazing surgeon pleased friends excellent results happier nose plastic expectations augmentation greco needing natural talented clark surgeries -Collection 3 awful nose fix mistakes correct scar worse mistake revision favor uneven tip crooked stomach	bonakdar dermatology melanoma treatments trehan products laser acne ensure sun annual spots -Collection 1 rude horrible refused seconds mccormack beware gmitter dirty medication pill enemy loud blamed insisted harwer angry upset wed disorganized werent -Collection 2 done procedure could surgery rivadeneira better schwartz found exactly met everything skin glad days concerned schlessinger made brain detail surgical -Collection 3 said get kept nurse upset man throat brain cold door reason thats let dismissed	exceptional received harwer puts mile kindness stays difficult genuine matter hammack mensah -Collection 1 lack zero rudest confidentiality pathetic files whatsoever unpleasant patronizing futrell paycheck screamed trubish krosi woman instructed wing carried genital discriminated -Collection 2 well cares truly listens physician compassionate family outstanding understanding excellent field beyond attentive physicians genuinely expertise caring extra genuine mile -Collection 3 lack man license attitude unprofessional board empathy pompous revoked retire choi impatient lacks public

dropped	messed	botched	buy	anger
exact	screwed	promised	stupid	danger
demanded	complaining	lipson	reply	privacy
dare	complained	button	lie	provides
minute	second	schlessinger	cruel	verbally
big	ignored	chin	raised	wonder

Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
life	help	dreher	us	dont
dreher	medication	good	dreher	know
years	problem	manner	husband	doesnt
one	treatment	bedside	son	like
would	get	job	daughter	im
enough	meds	person	old	think
could	problems	experience	year	people
say	medications	really	hospital	get
help	find	seems	family	want
every	take	talk	child	take
way	better	doc	mother	say
cannot	medicine	side	going	going
still	also	knowledgeable	took	good
cant	prescribed	kind	care	cant
cancer	pain	problems	also	tell
much	seeing	best	heart	thats
believe	taking	helpful	wife	reviews
doctors	without	skills	well	bad
god	treat	caring	never	doc
thank	give	bed	see	wants
-Collection 0	-Collection 0	-Collection 0	-Collection 0	-Collection 0
kolb	years	great	love	shes
thank	really	excellent	great	ive
saved	helped	wonderful	wonderful	nice
grateful	listens	understanding	children	knows
needed	compassionate	listener	kids	appreciate
thankful	willing	lisa	thanks	honest
expertise	helps	straightforward	pediatrician	stuff
implants	whole	accurate	miller	straight
women	natural	nononsense	loves	forward
fountis	welch	andrews	fabulous	approach
drive	proactive	importantly	born	bit
illness	nancy	regularly	fantastic	nonsense
gulati	steen	carol	nahra	blum
blessed	emails	stein	gentle	retires
closely	insightful	sincere	lucky	arent
angel	nahas	communications	team	busy
weight	peers	role	blessing	surprised
reddoch	lukawski	gyno	anna	wright
breast	humor	appreciation	ray	respectful
explant	scheinost	fitting	amorapanth	condescending
-Collection 1	-Collection 1	-Collection 1	-Collection 1	-Collection 1
woman	cough	poor	cold	shes
whitney	malpractice	taylor	incompetence	money
garrett	mistakes	judgemental	prescribed	avoid
overweight	unknowledgeable	rudeness	throat	disrespectful
kelly	zyrtec	madoff	quack	yelled
ray	sells	step	similar	unfortunately
dey	sued	roland	breastfeeding	dollars
uba	confusing	insane	ramer	per
ending	ryan	randak	inserts	conclusion
pass	label	lunder	warning	value
worse	trinh	bluntly	apology	tag
susan	hurried	defensive	carla	furious
ford	duties	borderline	heads	unacceptable
truth	hound	accused	siudmak	means

messed	sporn	looked	ineffective	cancelling
meshe	caught	brained	teasley	samuels
germano	forbid	navarro	rakovshik	cancels
recorded	insurer	traditional	breastmilk	malkin
koster	interrogating	ratner	curry	unorganized
accusing	assumptions	authorities	decides	allen
-Collection 2	-Collection 2	-Collection 2	-Collection 2	-Collection 2
life	years	great	years	hes
thank	helped	excellent	always	good
man	dreher	wonderful	love	really
saved	many	manner	family	nice
grateful	able	knowledgable	children	knows
best	issues	awesome	great	like
thanks	willing	easy	wonderful	things
knowledge	seeing	knowledgeable	kids	guy
thankful	worked	smart	son	one
god	right	fantastic	loves	lot
better	others	listener	referred	say
confidence	listened	earth	pediatrician	make
quality	approach	personable	caring	work
brilliant	helping	fabulous	friends	sure
skill	gone	terrific	seen	forward
team	treatments	puts	since	honest
husband	hard	around	patel	busy
found	hope	internist	trust	stuff
forever	past	depth	born	straight
bless	since	bedside	members	important
-Collection 3	-Collection 3	-Collection 3	-Collection 3	-Collection 3
man	drug	poor	months	hes
arrogant	simply	horrible	died	guy
avoid	addict	terrible	infection	money
angry	believe	uncaring	father	wrong
stop	head	unhappy	turned	meds
ruined	another	run	flu	elsewhere
beware	supposed	whatsoever	supposed	quack
zero	legal	appears	chest	touch
malpractice	real	abrupt	er	basically
shocked	mg	demeaning	dad	horrible
excuse	iv	impersonal	yelling	hed
add	basically	disinterested	ended	ignorant
oath	smith	einbinder	spoke	act
warning	self	vet	caused	copy
nightmare	floor	adjei	misdiagnosed	hell
sued	failure	timson	questioned	hot
x	spouse	behavioral	wheelchair	luck
convinced	sports	wannabe	nasal	violation
actions	hell	staffing	careless	shame
witness	including	impolite	insulting	cuz

Extra-credit problem:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
fisher	worth	vegas	uncaring	couple
unfriendly	spend	las	earth	dad
diagnostician	skilled	recomend	end	calm
complain	walked	comeau	throat	certified
root	paid	turn	foot	non
understands	ordered	stent	mean	retired
trained	father	responsive	ent	till
average	pediatrician	students	husbands	arthritis

mental sympathetic	personally comments	greatful single	insensitive unless	credit pusher

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
disorder theres aloof avoided texas rushing hasnt reccommend calming weakness	staff great time office would good best recommend always care	dental price helpfull class pains feet lift trustworthy language drop	highest appropriate superb gp effective meticulous thompson twenty west contacting	punctual brown efficient informative tooth mind considerate gentle ankle alive

For 20 topics:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
greatest dismissed jerk plenty theres j definately lived till allows	satisfied good intelligent willing punctual shows children ease conditions spends	extra mile puts charge whenever price samples request reads staffing	informative dental involved smith lets gp liver notch upper medicines	vegas las foot skin half mark including body comforting allow

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
broken root understands environment trained sharp pains break pas signed	fisher manor pulled faith hernia unwilling advance tummy obnoxious injured	spot manage nightmare proved greatful payment impersonal tonsils appeared judgmental	smart quite throat nose receptionist agree current customer tells manager	staff great time office would best recommend always care years

Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
recomend solve students hurried medicaid slow individuals food perhaps valued	rushes ankle ego partner trustworthy mayo greedy impression taylor accomodating	lee dad father ms sucks miller recieved feet thompson license	pediatrician pushed class happens texas feelings accommodating injuries friday permanent	considerate migraines sincere annoyed via sleep forget hysterectomy sympathetic impatient

Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
town difficult whats end refill idea zero provider stars school	ensure bladder losing meticulous qualified professionals language adult relative contacting	cosmetic leaves reschedule dermatologist page talented knowlegable cancelled ears rudest	son appt kids breast listened check meds due children fantastic	visiting changes saver cyst welcoming aloof importantly professionally encouraging recommends

The output isn't totally irrelevant, but it does seem like many topics are noisier than without tf-idf. So, the output topics from Task #2a appear to be better.

b) In my opinion, tf-idf might be useful on its own for topic modeling, but with LDA, there is no need to give weights to words because it does not just take the top 10 most frequent words.