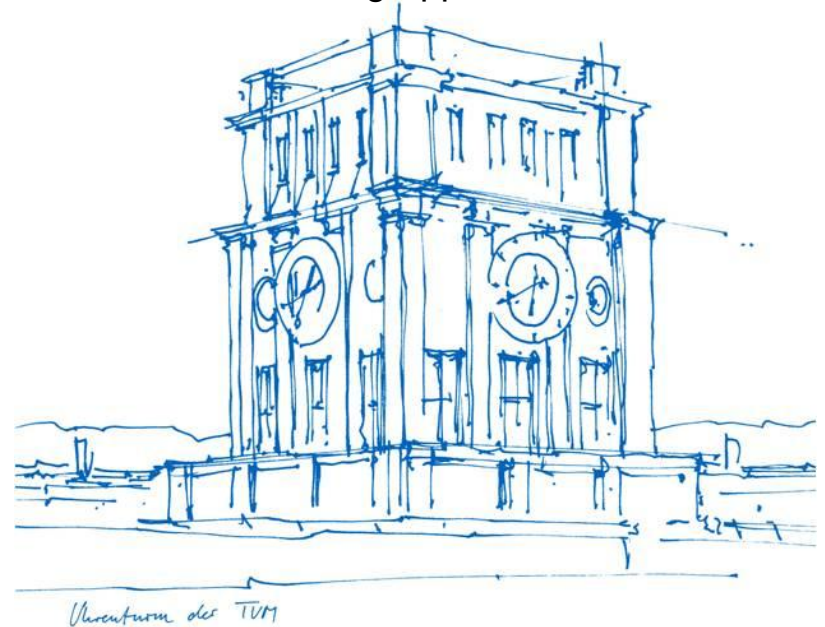


Opinion Mining Lab Group 1.3

Topic: Weakly Supervised Aspect Extraction Using a Student-Teacher Co-Training Approach

Group members: Jingpei Wu, Ke Xin Chen, Kevin George

17.05.2021

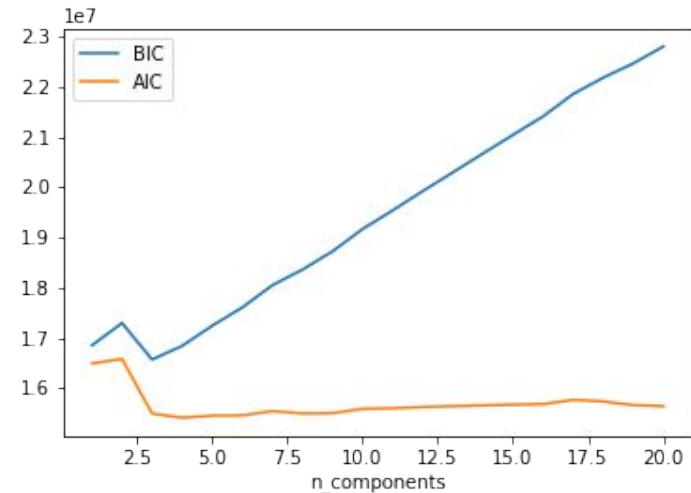
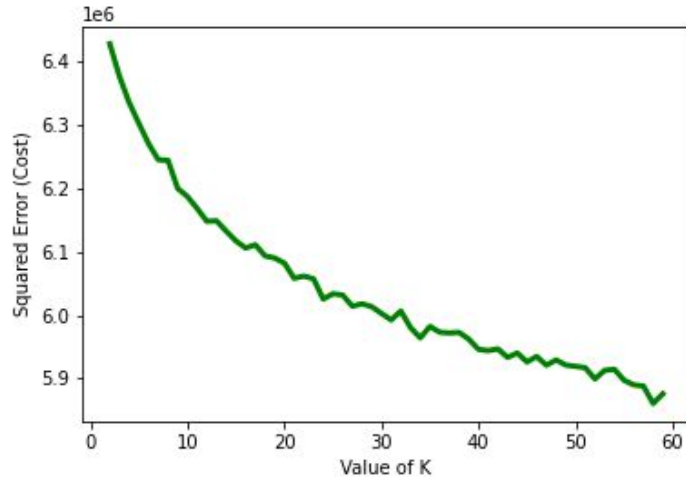


For the past two weeks

- Get familiar with the annotated dataset
- Fixed preprocessing steps
- Different approaches to get seed words
 - Update k-means method
 - Applying NMF with both biased and unbiased comments
 - Applying clarity scoring function with annotated dataset

k-means: continuation from last week

- remove tf-1 words
- use articles and comments for fine-tuning
- number of K: elbow method and BIC/AIC



NMF - with/without stop words

topic 0: food, grow, store, healthy, process, health, shop ,price, label, real

topic 1: organic, pesticide, use, farming, produce, conventional, mean, certify, farmer

topic 2: thank, share post info information great article love god **sharing**, **sign**

topic 3: buy, local, product, stop, farmer, egg, store, market, anymore, brand

topic 4: people, know, make, like, use, say, think, need, time, **just**, **pron**

topic 5: eat, meat, animal, stop, chicken, healthy, egg, diet, feed, vegan

topic 6: yes, vote, right, oh, course, love, hell, answer, question, **agree**, **boycott**

topic 7: good, thing, taste, news, luck, point, way, idea, bad, **feel**, **job**

topic 8: milk, cow, raw, drink, almond, dairy, baby, love, calf, human

topic 9: gmo, label, monsanto, corn product, non, crop, seed, free, **labeling**, **gm**os

Black words - in common

Orange words - stop words removed

Red words - stop words kept

NMF - nouns only

0: food, health, store, label, supply, price, industry, consumer, company, ingredient, {choice, chemical, junk, country, world}
 1: people, money, thing, world, lot, government, problem, country, health, fact, {company, corporation, choice, reason, day}
 2: milk, cow, dairy, baby, almond, calf, human, goat, hormone, reason, {cheese, drink, cream, soy, kid}
 3: farmer, pesticide, farm, year, crop, thing, chemical, plant, market, seed, {soil, water, produce, use, health}
 4: product, company, label, store, ingredient, consumer, brand, price, dairy, money, {market, quality, business, grocery, oil}
 5: thank, info, article, information, post, comment, link, answer, share, work, {point, question, idea, list, head}
 6: way, money, thing, company, life, day, world, animal, business, nature, {body, government, point, science, change}
 7: time, store, year, day, money, life, grocery, week, article, lot, {comment, price, thing, change, period}
 8: meat, animal, chicken, egg, feed, dairy, factory, beef, grass, cow, {antibiotic, eat, diet, vegan, human}
 9: http, com, www, article, org, study, https, href, title, target, {blank, link, youtube, facebook, html}

Red words - existed in the previous results

Black words - only existed in this result

Clarity Scoring Function - aspects from annotated dataset

	0	1	2	3	4	5	6	7	8	9
general	gmo	farming	store	organic	all	india	definition	online	farm	seed
price	price	expensive	cost	more	pay	money	demand	higher	extra	charge
taste	taste	better	flavor	difference	tasty	delicious	fruit	sweeter	flavour	beef
nutritional quality/freshness/appearance	nutrient	nutritional	quality	nutritious	vitamin	antioxidant	higher	nutrition	milk	more
safety	safer	safe	safety	dangerous	human	cancer	consensus	test	are	no
healthiness	health	healthier	healthy	benefit	food	body	study	disease	powder	evidence
chemicals/pesticides	pesticide	chemical	fertilizer	synthetic	use	residue	used	herbicide	toxic	are
label	label	certified	certification	usda	marketing	labeled	ingredient	organic	certifying	labeling
origin, source	china	brand	strictly	versus	orchardists	cultivable	country	origin	sold	west
local	local	locally	burger	fresh	allotment	push	buy	degrades	season	travel
availability	online	sale	shop	bangalore	supplier	location	delhi	find	market	store
environment	environment	soil	farming	environmental	environmentally	sustainable	destructive	equally	water	the
animal welfare	animal	cow	chicken	hen	feed	cage	fed	access	outdoors	antibiotic
productivity	yield	crop	land	acre	population	le	farming	impact	grow	economic

Clarity Scoring Function - stop words removed

	0	1	2	3	4	5	6	7	8	9
general	gmo	farming	store	organic	india	definition	online	farm	seed	start
price	price	expensive	cost	pay	money	demand	higher	extra	charge	afford
taste	taste	better	flavor	difference	tasty	fruit	delicious	sweeter	flavour	beef
nutritional quality/freshness/appearance	nutrient	nutritional	quality	nutritious	vitamin	antioxidant	higher	nutrition	milk	tomato
safety	safer	safe	safety	dangerous	human	cancer	consensus	test	harmful	cause
healthiness	health	healthier	healthy	benefit	food	body	study	disease	powder	evidence
chemicals/pesticides	pesticide	chemical	fertilizer	synthetic	use	residue	used	herbicide	toxic	approved
label	label	certified	certification	usda	marketing	labeled	organic	ingredient	certifying	labeling
origin, source	china	brand	strictly	versus	country	orchardists	cultivable	origin	sold	mile
local	local	locally	burger	fresh	buy	push	allotment	degrades	season	travel
availability	online	sale	shop	bangalore	supplier	location	delhi	market	store	agro
environment	environment	soil	farming	environmental	environmentally	sustainable	destructive	equally	water	better
animal welfare	animal	cow	chicken	hen	feed	cage	fed	access	outdoors	antibiotic
productivity	yield	crop	land	acre	population	le	farming	impact	grow	economic

Current Challenges

- How to merge seed words from three different approach.
- Issue with k-means

Plans for the next two weeks

- Select seed words for training process
- Start building student-teacher modules
- Apply the approaches to the german dataset