Humor in Word Embeddings: Cockamamie Gobbledegook for Nincompoops

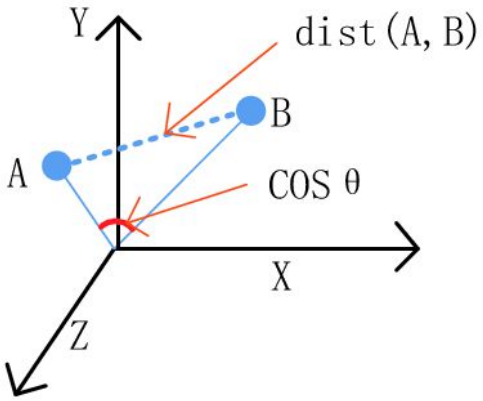
Limor Gultchin, Genevieve Patterson, Nancy Baym, Nathaniel Swinger, and Adam Tauman Kalai

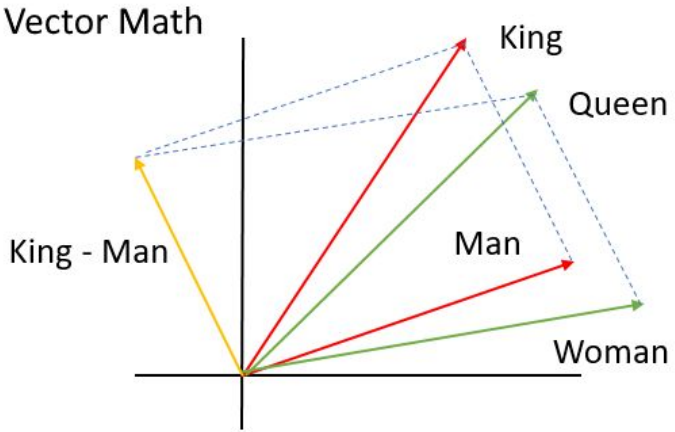
https://github.com/limorigu/Cockamamie-Gobbledegook

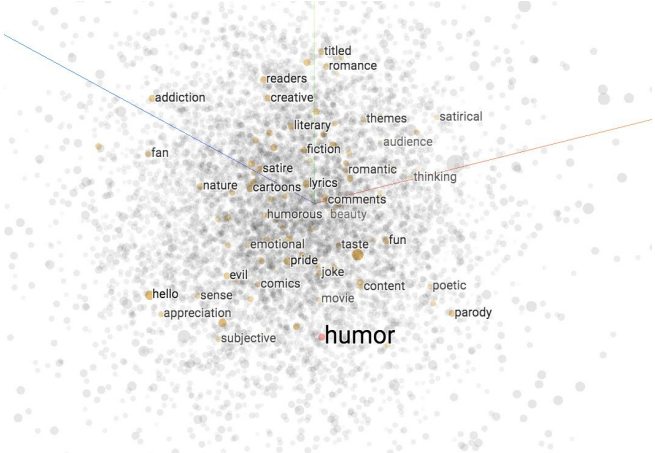
Toolkit: Word embeddings

- Allowing us to relate words to each other, similarity is defined by distance to neighboring vectors. Each word is represented as a high dimensional vector (learned by a pre-trained neural network on some corpora)

- Similarity between words can then be captured and computed via cosine similarity, and even define logical analogies







TensorFlow embedding projector

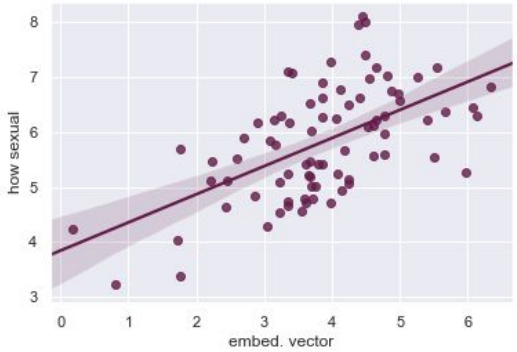
Original Data collection (From Mechanical Turk)

Additional dimensions of data:
Humor theory features

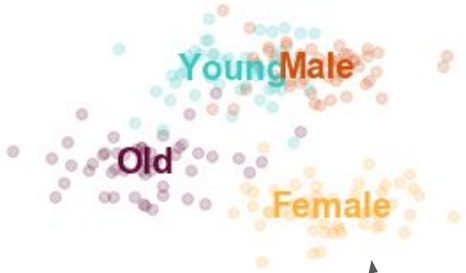
	Is ‘yadda yadda’...	Yes/No
1	Funny sounding	✓
2	Juxtaposition	✓
3	Sexual	✗

Our approach

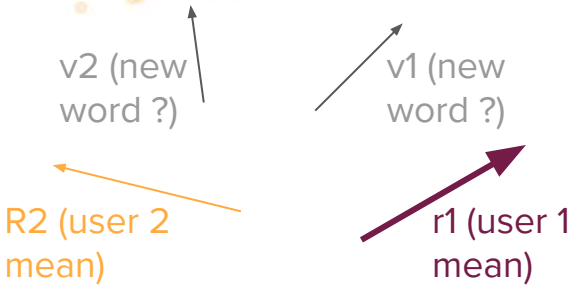
1. Can we use word embeddings to capture humor theories and a humor direction?



2. Can we identify different senses of humor across demographic groups?

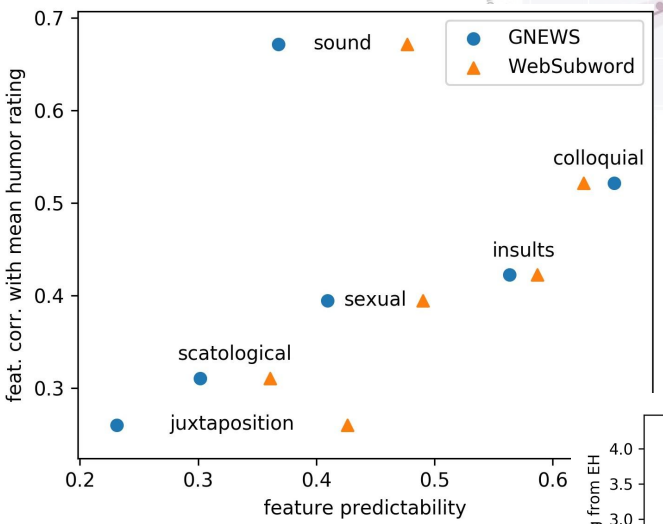


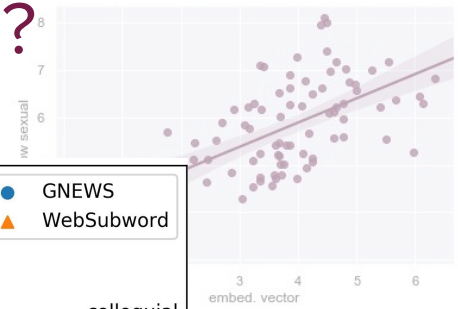
3. Can we define individual sense of humor and predict users' taste?

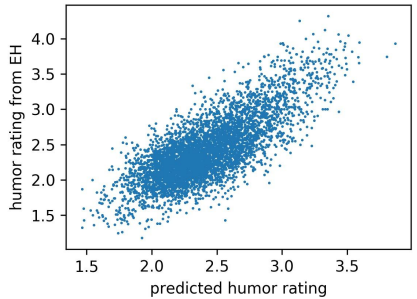


1 Can we use word embeddings to capture humor theories and identify a ‘humor direction’?

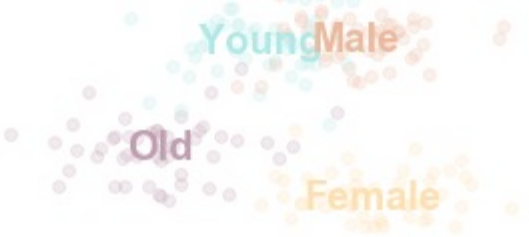
- Ridge regression to predict theory rating (average of 8 users) from word embedding vector for each word (90-10% train/test split)
- Correlation b/w predictions and actual ratings
- ‘Predictability score’=mean over 1,000 runs







2 Can we identify different senses of humor across demographics?



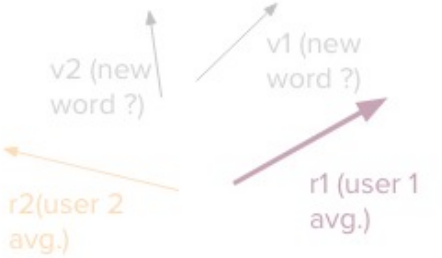
- K-means clustering of individuals' average vector of 36 favourite words
- Demographics of each cluster uncovered later
- 'Most characteristic' word for cluster defined as

$$C_i(w) = freq(\text{users in cluster } i \text{ found } w \text{ funny}) - freq(\text{all users found } w \text{ funny})$$

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
word 1	gobbledegook	tootsies	clusterfuck	whakapapa	dickheads
word 2	kerfuffle	fufu	batshit	codswallop	twat
word 3	hullaballoo	squeegee	crapola	dabbawalas	cocking
word 4	razzmatazz	doohickey	apeshit	pooja	titties
word 5	gazumped	weenies	fugly	spermatogenesis	asshattery
word 6	boondoggle	muumuu	wanker	diktats	nutted
word 7	galumphing	thingies	schmuck	annulus	dong
word 8	skedaddle	wigwams	arseholes	chokecherry	wanker
word 9	guffawing	weaner	dickheads	piggeries	cockling
word 10	bamboozle	peewee	douchebaggery	viagogo	pussyfooting
sound	**1.11**	1.02	0.97	1.02	0.90
scatological	0.80	0.99	**1.15**	0.89	1.14
colloqial	0.95	1.00	**1.14**	0.87	1.02
insults	0.86	0.90	**1.23**	0.84	1.12
juxtaposition	0.89	0.86	0.99	1.10	**1.13**
sexual	0.81	0.91	0.99	1.00	**1.25**
female %	**70.3%***	57.5%	53.8%	52.4%	**35.2%***
mean age	38.5	37.4	**42.3***	37.2	34.8*

3 Can we define individual senses of humor and predict users' taste?

- Define a mean vector of words rated funny for each user
- 'Know-your-audience' test, match unseen words to the right individual (see formula)
- Compute accuracy score



$$\|\vec{r}_1 - \vec{v}_1\|^2 + \|\vec{r}_2 - \vec{v}_2\|^2 < \|\vec{r}_1 - \vec{v}_2\|^2 + \|\vec{r}_2 - \vec{v}_1\|^2$$

$$0 < (\vec{r}_1 - \vec{r}_2) \cdot (\vec{v}_1 - \vec{v}_2)$$

Know-your-audience test	Success rate
Easy: disjoint sets of 35 training words.	78.1%
Normal: 35 training words.	68.2%
Hard: 5 training words.	65.0%