So what?
Applications of word embeddings for CSS

Bennett Kleinberg

Department of Security and Crime Science

University College London

bkleinberg.net

bennett.kleinberg@ucl.ac.uk

Problem-led research

- Computational Social Science research is led by problems/questions

Method follows problem

- But: you also need to understand the methods that could solve some problems
- Today: **<u>word embeddings</u>**

<u>So which problems can I actually solve with “word embeddings”?</u>

So what? – General approach

- High-level idea: create a numerical representation of your text input
- BUT: isn't the same done with matrices of POS tags, named entities, CfG?
- Or the LIWC?

Language tasks and word embeddings

- Two classes of applications: **prediction tasks vs. <u>“language” tasks</u>**
- Prediction tasks: using embeddings as features in (supervised) ML
- Language tasks: ″understanding″ processes that are reflected in language

- Core idea: proximity in the vector space ≈ similarity in meaning

Note: NLP’s Clever Hans moment (https://thegradient.pub/nlps-clever-hans-moment-has-arrived/)

Example: # prediction with tweet2vec

https://arxiv.org/abs/1605.03481

Tweet2Vec: Character-Based Distributed Representations for Social Media

Bhuwan Dhingra[1], **Zhong Zhou**[2], **Dylan Fitzpatrick**[1,2]
Michael Muehl[1] and **William W. Cohen**[1]
[1]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
[2]Heinz College, Carnegie Mellon University, Pittsburgh, PA, USA
{bdhingra,djfitzpa,mmuehl}@andrew.cmu.edu
zhongzhou@cmu.edu wcohen@cs.cmu.edu

- Social media text problems: "Informal language, spelling errors, abbreviations, and special characters"
- Character-based embeddings model
- Compared with word model

Tweets	Word model baseline	*tweet2vec*
ninety-one degrees. ☀❤😎	#initialsofsomeone.. #nw #gameofthrones	#summer **#loveit** #sun
self-cooked scramble egg. yum!! !url	#music #cheap #cute	**#yummy** #food #foodporn
can't sleeeeeeep	#gameofthrones #heartbreaker	#tired **#insomnia**
oklahoma!!!!!!!!!!! champions!!!!!	#initialsofsomeone.. #nw #lrt	**#wcws** #sooners #ou
7 % of battery . iphones die too quick .	#help #power #money #s	#fml #apple #bbl **#thestruggle**
i have the cutest nephew in the world !url	#nephew **#cute** #family	#socute **#cute** #puppy

Table 1: Examples of top predictions from the models. The correct hashtag(s) if detected are in bold.

Example: Stereotype quantification

- "word embeddings as a quantitative lens through which to study historical trends – specifically trends in the gender and ethnic stereotypes in the 20th and 21st centuries in the United States"
- Focus on gender, race, occupation
- Outcome metric: *relative norm difference*
- Google Books embeddings per decade in the 1900s

Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes

Nikhil Garg
Stanford University
nkgarg@stanford.edu

Londa Schiebinger
Stanford University
schieb@stanford.edu

Dan Jurafsky
Stanford University
jurafsky@stanford.edu

James Zou
Stanford University
jamesz@stanford.edu

https://arxiv.org/abs/1711.08412

Hispanic	Asian	White
housekeeper	professor	smith
mason	official	blacksmith
artist	secretary	surveyor
janitor	conductor	sheriff
dancer	physicist	weaver
mechanic	scientist	administrator
photographer	chemist	mason
baker	tailor	statistician
cashier	accountant	clergy
driver	engineer	photographer

(c) The top ten occupations most closely associated with each ethnic group in the Google News embedding.

Example: Stereotype quantification

Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes

Nikhil Garg
Stanford University
nkgarg@stanford.edu

Londa Schiebinger
Stanford University
schieb@stanford.edu

Dan Jurafsky
Stanford University
jurafsky@stanford.edu

James Zou
Stanford University
jamesz@stanford.edu

Bigger picture:
Quantification of historically qualitative problems!

1910	1950	1990
charming	delicate	maternal
placid	sweet	morbid
delicate	charming	artificial
passionate	transparent	physical
sweet	placid	caring
dreamy	childish	emotional
indulgent	soft	protective
playful	colorless	attractive
mellow	tasteless	soft
sentimental	agreeable	tidy

(a) Top adjectives associated with women in 1910, 1950, and 1990 by relative norm difference in the COHA embedding.

1910	1950	1990
irresponsible	disorganized	inhibited
envious	outrageous	passive
barbaric	pompous	dissolute
aggressive	unstable	haughty
transparent	effeminate	complacent
monstrous	unprincipled	forceful
hateful	venomous	fixed
cruel	disobedient	active
greedy	predatory	sensitive
bizarre	boisterous	hearty

(b) Top Asian (vs White) Adjectives in 1910, 1950, and 1990 by relative norm difference in the COHA embedding.

Examples: Soft skills in job ads

- Aim: role of soft skills in job ads (salary, gender)
- Job ads dataset + crowdsourcing of soft skills
- Problem: dimension reduction of collected soft skills
- E.g. “team-worker”, “ability to work in team”, “able to work in team”, “good in a team”
- Needed: numerical representation that is sensitive to semantic relationships

Responsible team players wanted: an analysis of soft skill requirements in job advertisements

Federica Calanca, Luiza Sayfullina, Lara Minkus, Claudia Wagner & Eric Malmi

EPJ Data Science **8**, Article number: 13 (2019) | Download Citation

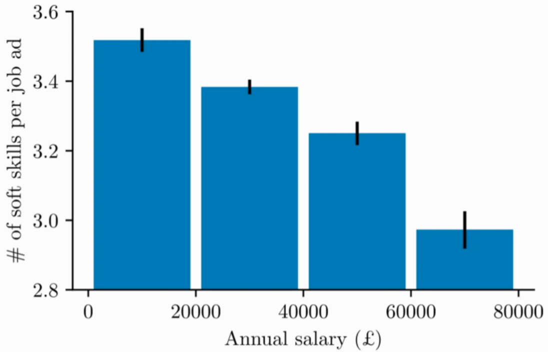
948 soft skills
→ Averaged embeddings
→ Clustering on vector repres.
→ 190 final clusters

Examples: Soft skills in job ads

Responsible team players wanted: an analysis of soft skill requirements in job advertisements

Federica Calanca, Luiza Sayfullina, Lara Minkus, Claudia Wagner & Eric Malmi

EPJ Data Science **8**, Article number: 13 (2019) | Download Citation



Low-paid job ads contain, on average, more soft skills than high-paid job ads. Bars show the average number of soft skills for the ads in four different salary groups, and the error bars indicate the 95% confidence intervals obtained via bootstrap re-sampling with replacement

Skill cluster	*r*	Count
maturity	11.9**	112
delegation skills	10.2**	53
team building skills	9.8*	50
strategic planning	9.1**	608
ability to work in a fastpaced environment	8.0*	51
leadership	7.4**	4743
constructive feedback	6.9*	74
proposal writing	6.2*	84
ability to improve skills	6.0**	108
discretion	5.7	309
results driven	4.9**	541
presentation skills	4.5**	1464
telephone skills	−7.3**	227
polite	−5.9**	339
dynamic person	−5.2	70
dedication	−4.6**	467
friendly personality	−4.6	97
listening skills	−4.3**	355
punctual	−4.1*	248
ability to identify problems	−3.1	132
calm	−2.8*	787
professional manner	−2.6**	2303
willingness to learn	−2.2**	1652
time management	−1.8	2149

**$p<0.01$, *$p<0.05$.
The *Count* column shows the denominator from Eq. (2), which can roughly be interpreted as the sample size. Only skill clusters with *Count* ≥ 50 are shown.

Example: Drill music & gang violence

Note: this is work in progress

Is UK drill music really behind London's wave of violent crime?

It's a menacing strain of hip-hop with a powerful presence on social media, but does drill reflect or drive crime?

Example: Drill music & gang violence

	Drill music lyrics	UK rap music lyrics
% unknown (10k word list)	18.64 (4.46)	14.88 (8.17)
Top 10 unknown words (10k word list comparison)	"tryna”, “bro”, “nigga[s]”, “mans”, “ting”, “opp[s]”, “splash”, “gyal”, “cah”, “skeng[s]”	“nigga[s]”, “tryna”, “ting”, “mans”, “nah”, “bitches”, “fam”, “imma”, “gyal”, “bro”
% unknown (300k word list)	7.38 (3.17)	5.99 (7.09)
Top 10 unknown words (300k word list comparison)	“whos”, “du”, “hella”, “4s”, “nuff”, “yute[s]”, “bruck”, “x2”, “skrr”, “mandem”	“mic”, “mandem”, “mcs”, “fuckin”, “cuh”, “blud”, “lil”, “cah”, “ayy”, “oi”

Table 2: Out-of-vocabulary statistics and terms for both corpora.

Problem:
How can we study a niche language if we don’t understand it?

Idea: vector neighbours could reveal the meaning 😎

Spoiler alert: it doesn’t work! 🤦‍♂️

Example: Drill music & gang violence

	840B GloVe	Drill GloVe	UK rap GloVe
ting	tings, ying, ling, wei, yu	mad, man, back, like, get	rivals, watch, easter, tobz, shiner
opp	Opp, ind, blk, det, opportunity	block, see, boy, get, man	snaps, arse, rae, cha, selling
splash	splashes, splashing, Splash, splashed, lime	ching, dip, wet, man, gang	seen, tilt, backs, boon, switching
gyal	yuh, bwoy, nuh, mek, likkle	want, love, said, take, ah	asses, notre, girlies, members, patten
skeng	n.i.m.	whip, 4, free, bae, beat	soccer, chuck, nicest, suggest, legit
4s	4s, 3gs, 2s, 4S, 6s	leave, scene, gon, yeah, left	hustlers, slang, paigans, braces, rattling
nuff	Nuff, nuf, sayin, nah, yah	dem, man, vehicle, chat, give	boil, predictable, slept, dane, tumble
yute	likkle, gwaan, bwoy, bredda, gwan	basic, fish, one, deals, dat	chill, canon, switchblade, motha, dehydrate
bruck	mixit, wiid, besa, adham, bluck	dada, married, whoosh, mixtape, side	ripped, march, haircut, code, investigating
mandem	wasteman, tingz, bruva, eitha, hommies	broski, slapped, feds, problem, locked	woe, approaches, sake, crew, sermon

Table 3: Closest neighbours in cosine distance of OOV terms for different GloVe models.

Example: Drill music & gang violence

	External definitions	Evaluation
ting	Pistol [hand ting][b], gyal or sometimes gun[c], an ethnic Caribbean way of saying thing[d]	"Thing" not found as a neighbour in any model; No gun-related terms found in any of the models
opp	Enemies [b], opponent, opposition[c],A member of an opposite gang[d]	No (rival) gang-related terms found.
splash	Stab [b,c], To splash someone, is uk street slang to stab someone with a knife[d]	No stabbing or violence-related terms in the GloVe models, mostly variations of "splash" itself
gyal	girl, lady, female person[c], a girl[d]	Antonym ("bwoys") found as a neighbor in GloVe 42B and 840B models; "want" and "love" in drill model may describe verbs associated with girls; Synonym "girlies" neighbour in UK rap model
skeng	knife, gun, weapon[b], gun[c], A knife/blade weapon[d]	No weapon-related terms found as neighbours.
4s	44mag or referring to 410[c] (a shotgun)	No weapon-related terms as neighbors. Some terms may relate to committing a shooting ("leave", "scene", "left") in drill corpus
nuff	short for "enough", but often "enough" is used to mean "lots of"[d]	Synonyms ("nuff, "nough") found in 42B GloVe
yute	young person or young people on the street[b]; Jamaican slang for youth, young person or child[d]	Terms similarly referring to people ("gyal", "bwoy", "bredda") in 840B models
bruck	sawn-off shotgun [bruckshot][b], broke, damaged;	No weapon-related terms found as neighbours.
mandem	A good friend, in a group of boys[a], multiple guys[c], Man Dem is the term used for a group of friends or homies. This term originated from the carribean but found its way to the streets of the uk where it is widely used.[d]	Terms similarly referring to (a group of) friends in all models: "bredrin", "bruva", "hommies", "broski", "crew"

Table 4: Evaluation of similarity findings against external evidence.

Conclusions:

- Word embeddings a lot of need data (dah!)
- Niche languages ≠ "a lot of data"
- Evaluation of semantic neighbours is inherently qualitative!

Other examples

Unsupervised word embeddings capture latent knowledge from materials science literature

Vahe Tshitoyan, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A. Persson, Gerbrand Ceder & Anubhav Jain

Nature **571**, 95–98 (2019) | Download Citation

https://www.nature.com/articles/s41586-019-1335-8

Deep Learning and Word Embeddings for Tweet Classification for Crisis Response

Reem ALRashdi
Department of Computer Science; Department of Computer Science and Software Engineering
University of York; University of Hail
York, United Kingdom; Hail, Saudi Arabia
Rmma502@york.ac.uk; r.alreshede@uoh.edu.as

Simon O'Keefe
Department of Computer Science
University of York
York, United Kingdom
Simon.okeefe@york.ac.uk

https://arxiv.org/ftp/arxiv/papers/1903/1903.11024.pdf

ATTACK2VEC: Leveraging Temporal Word Embeddings to Understand the Evolution of Cyberattacks

Yun Shen
Symantec Research Labs
yun_shen@symantec.com

Gianluca Stringhini
Boston University
gian@bu.edu

https://arxiv.org/pdf/1905.12590.pdf

More than Bags of Words: Sentiment Analysis with Word Embeddings

Elena Rudkowsky, Martin Haselmayer, Matthias Wastian, Marcelo Jenny, Štefan Emrich & Michael Sedlmair

Pages 140-157 | Published online: 10 Apr 2018

https://www.tandfonline.com/doi/full/10.1080/19312458.2018.1455817

New ways to answer (old) questions

- **Gender bias** in the wording of calls for funding

Gender contributes to personal research funding success in The Netherlands

Check for updates

Romy van der Lee and Naomi Ellemers

PNAS October 6, 2015 112 (40) 12349-12353; first published September 21, 2015 https://doi.org/10.1073/pnas.1510159112

- Role for word embeddings: examine the proximity to "winning attributes"
- Also: potential complementary method for IATs

New ways to answer (old) questions

- Gender bias in the wording of calls for funding

Gender contributes to personal research funding success in The Netherlands

Romy van der Lee and Naomi Ellemers

PNAS October 6, 2015 112 (40) 12349-12353; first published September 21, 2015 https://doi.org/10.1073/pnas.1510159112

Dutch research funding, gender bias, and Simpson's paradox

Casper J. Albers

PNAS December 15, 2015 112 (50) E6828-E6829; first published December 3, 2015 https://doi.org/10.1073/pnas.1518936112

No evidence that gender contributes to personal research funding success in The Netherlands: A reaction to van der Lee and Ellemers

Beate Volker and Wouter Steenbeek

PNAS December 22, 2015 112 (51) E7036-E7037; first published December 8, 2015 https://doi.org/10.1073/pnas.1519046112

New ways to answer (old) questions

- **Implicit associations**: the IAT

Implicit Association Test

Next, you will use the 'E' and 'I' computer keys to categorize items into groups as fast as you can. These are the four groups and the items that belong to each:

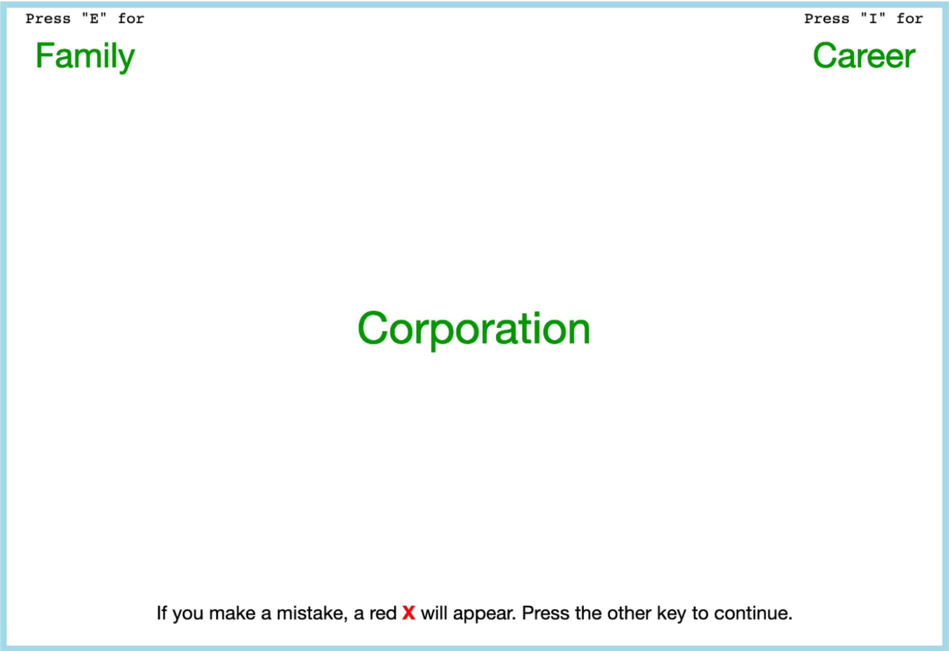
Category	Items
Male	Ben, Paul, Daniel, John, Jeffrey
Female	Rebecca, Michelle, Emily, Julia, Anna
Career	Career, Corporation, Salary, Office, Professional, Management, Business
Family	Wedding, Marriage, Parents, Relatives, Family, Home, Children

There are seven parts. The instructions change for each part. Pay attention!

Continue

New ways to answer (old) questions

- Implicit associations: the IAT (https://implicit.harvard.edu/implicit/Study?tid=-1)



New ways to answer (old) questions

- Implicit associations: the IAT
- Typical finding: faster response times for congruent trials ("John + Career") than for incongruent trials ("John + Parent")

- Question for research: do these implicit associations manifest themselves in language?
- If so: what does this mean?

New ways to answer (old) questions

- **Measuring (and detecting) attitudes** beyond sentiment analysis
- Sentiment analysis mostly context-free

Example: YouTube vlog transcripts

"[…] and **Brexit** will be the major challenge for the young generation."

- Word embeddings: find neighbours of Brexit, and include them in the sentiment analysis
- Premise: semantically similar words hold a similar sentiment value.

Questions?