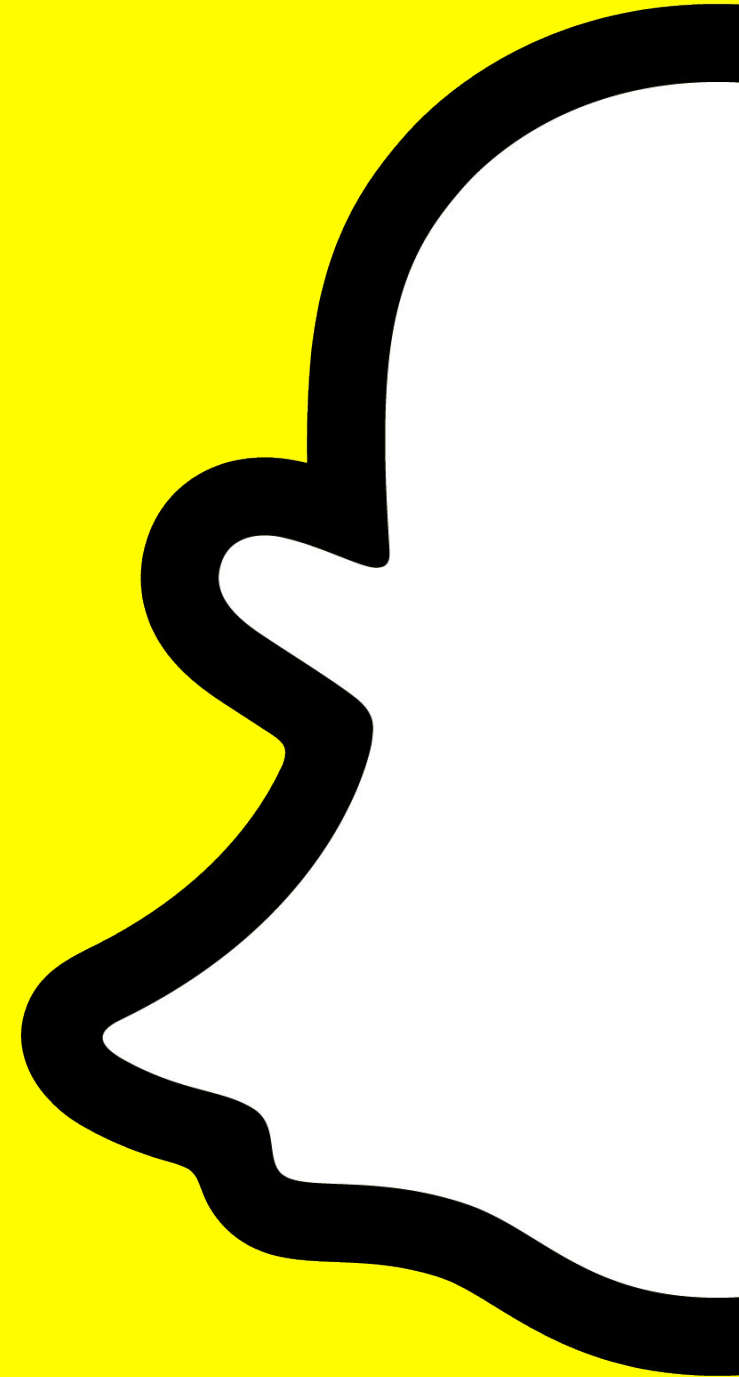


The Importance of "In the Moment" Knowledge

Contextualizing NLP Models with Time,
Location and Weather Information

Francesco Barbieri

Snap Research, User Modeling and Personalization

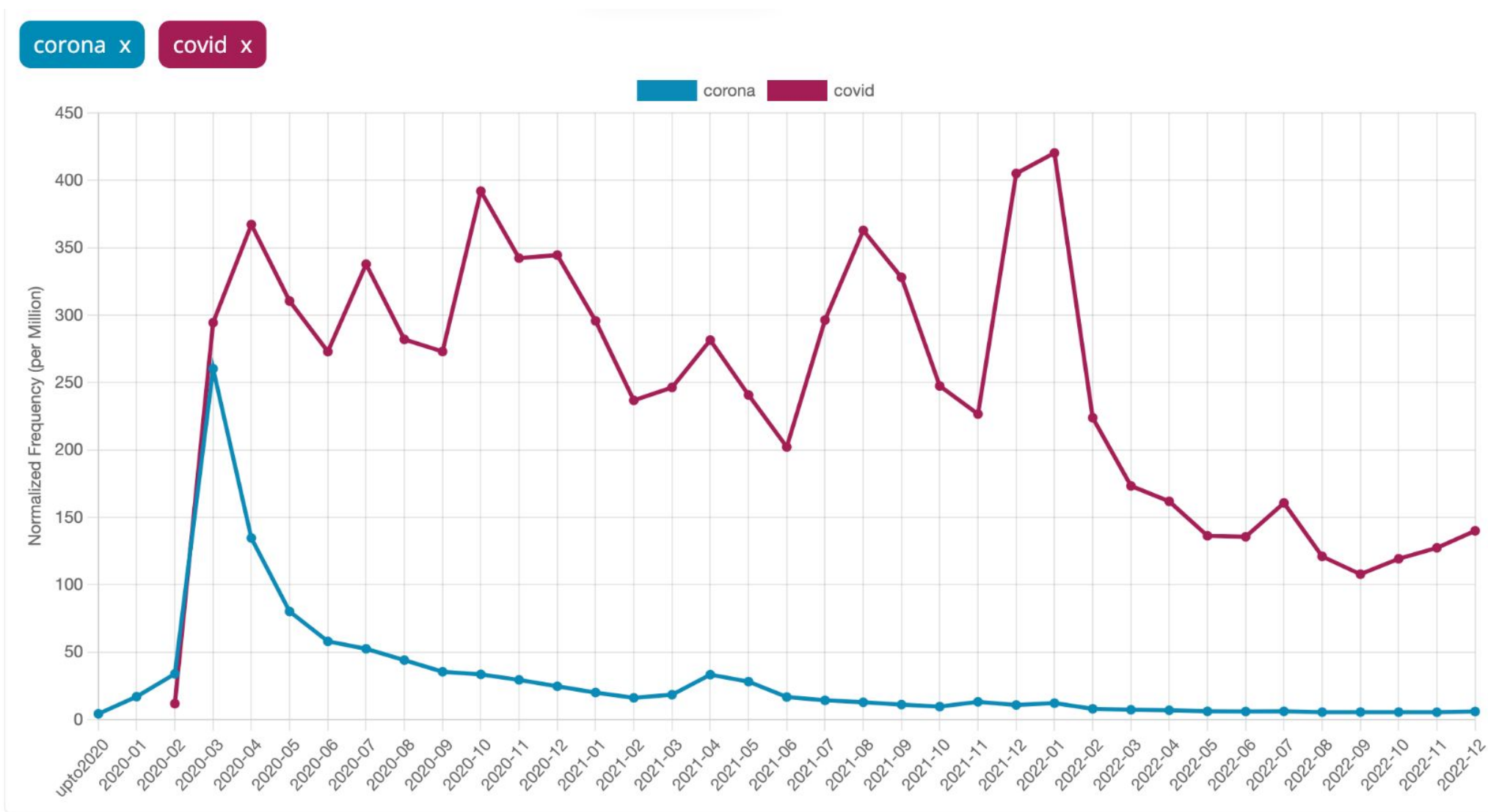




Who won the world cup in 2022?



I apologize, but as an AI language model, my knowledge is up-to-date only until September 2021, and I do not have information on events that occurred after that date. The FIFA World Cup in 2022 took place in Qatar, but I cannot provide the winner or any other details about the tournament. To find out the winner of the 2022 FIFA World Cup, I suggest checking the latest news or conducting an online search for the most recent information.



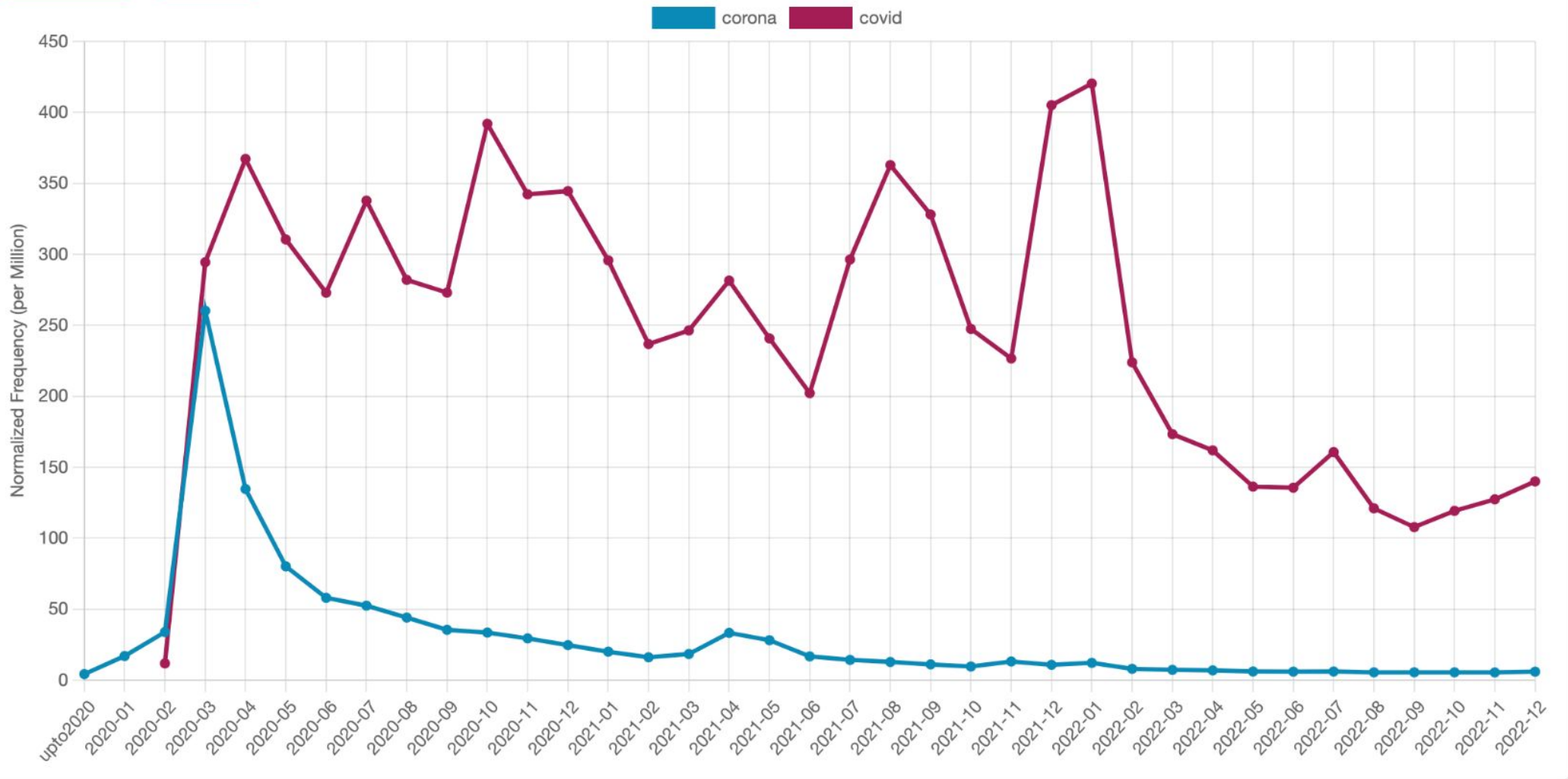


- Frequency
- Meaning Shift
- Sentiment
- Topic

Type a term to get its usage frequency on twitter (you can compare up to five terms)

 Add

corona x covid x



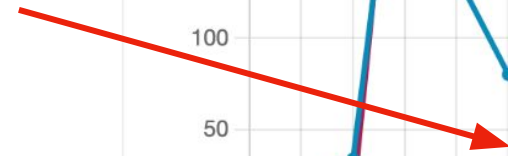
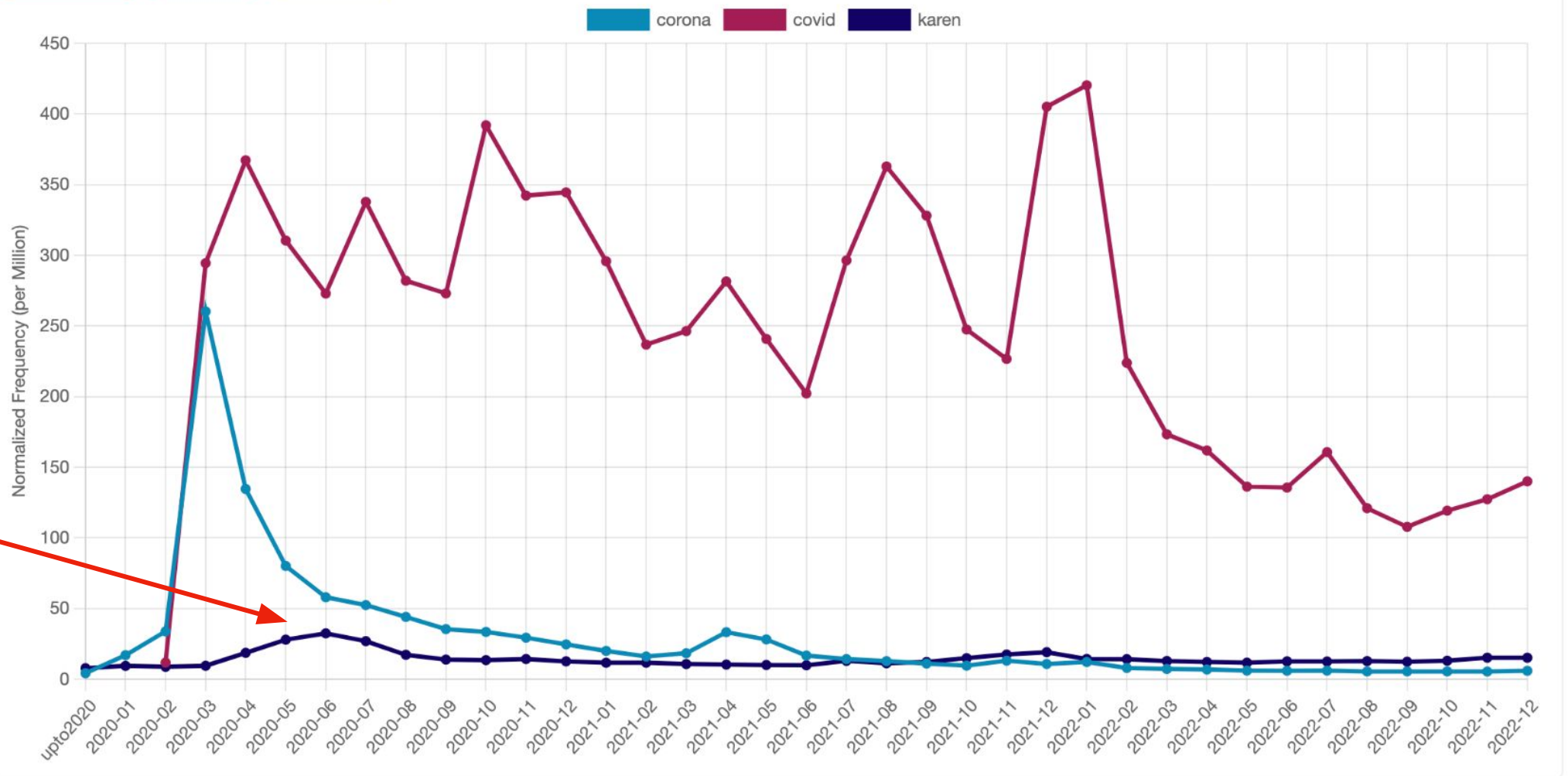


- Frequency
- Meaning Shift
- Sentiment
- Topic

Type a term to get its usage frequency on twitter (you can compare up to five terms)

Enter... Add

corona x covid x karen x





- Frequency
- Meaning Shift
- Sentiment
- Topic

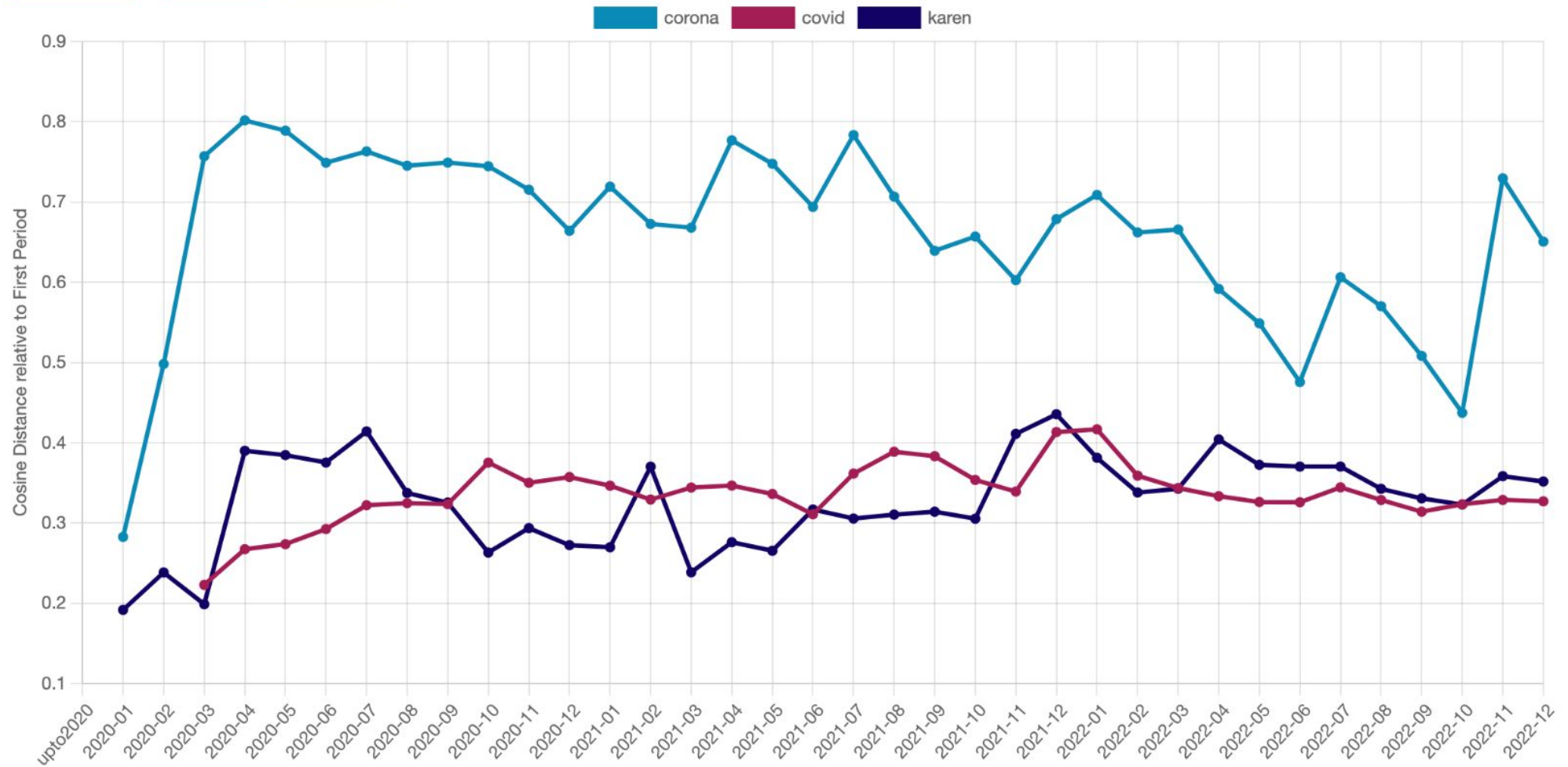
Type a term to get its meaning shift on twitter (you can compare up to five terms)

Add

corona x

covid x

karen x





- Frequency
- Meaning Shift
- Sentiment
- Topic

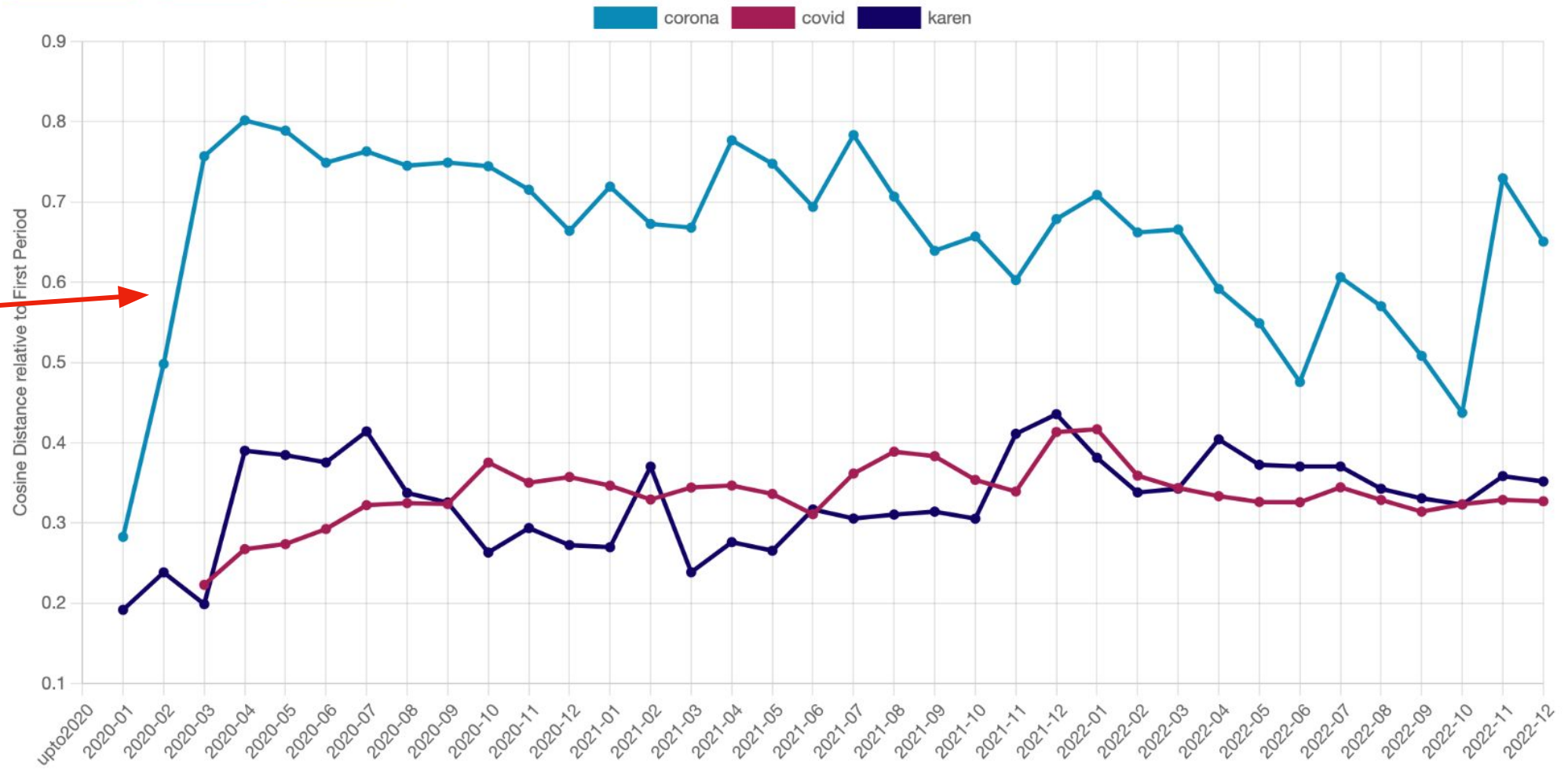
Type a term to get its meaning shift on twitter (you can compare up to five terms)

Add

corona x

covid x

karen x





- Frequency
- Meaning Shift
- Sentiment
- Topic

Type a term to get its meaning shift on twitter (you can compare up to five terms)

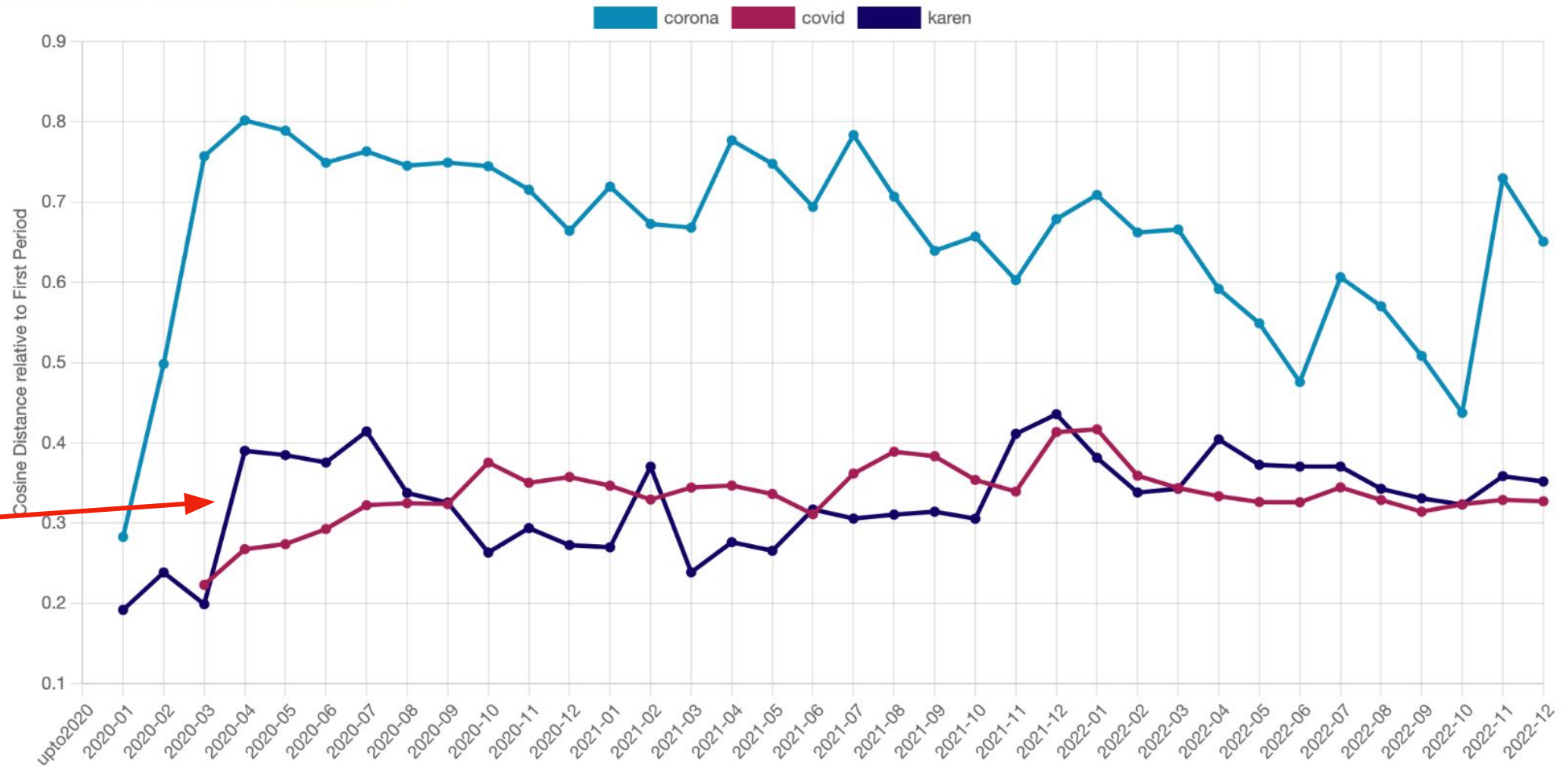
Enter...

Add

corona x

covid x

karen x





Time LMs

TimeLMs: Diachronic Language Models from Twitter

Daniel Loureiro*[♣], Francesco Barbieri*[♣],

Leonardo Neves[♣], Luis Espinosa Anke[◇], Jose Camacho-Collados[◇]

[♣] LIAAD - INESC TEC, University of Porto, Portugal

[♣] Snap Inc., Santa Monica, California, USA

[◇] Cardiff NLP, School of Computer Science and Informatics, Cardiff University, UK

[♣] daniel.b.loureiro@inesctec.pt, [♣] {fbarbieri, lneves}@snap.com,

[◇] {espinosa-ankel, camachocolladosj}@cardiff.ac.uk

- TimeLMs is a set of language models specialized on diachronic Twitter data.
- A continual learning strategy for Twitter-based language models' capacity to deal with future and out-of-distribution tweets
- We train LM models on new data every 3 months

Model Name	# Tweets	Last Date
twitter-roberta-base-2019-90m	90.26M	2019-12-31
twitter-roberta-base-mar2020	94.46M	2020-03-31
twitter-roberta-base-jun2020	98.66M	2020-06-30
twitter-roberta-base-sep2020	102.86M	2020-09-30
twitter-roberta-base-dec2020	107.06M	2020-12-31
twitter-roberta-base-mar2021	111.26M	2021-03-31
twitter-roberta-base-jun2021	115.46M	2021-06-30
twitter-roberta-base-sep2021	119.66M	2021-09-30
twitter-roberta-base-dec2021	123.86M	2021-12-31
twitter-roberta-base-2021-124m	123.86M	2021-12-31
twitter-roberta-base-mar2022	128.06M	2022-03-31
twitter-roberta-base-jun2022	132.26M	2022-06-30
twitter-roberta-base-mar2022-15M-incr	138.86M	2022-03-31
twitter-roberta-base-jun2022-15M-incr	153.86M	2022-06-30
twitter-roberta-base-sep2022	168.86M	2022-09-30



Time LMs

Models	2020-Q1	2020-Q2	2020-Q3	2020-Q4	2021-Q1	2021-Q2	2021-Q3	2021-Q4	Change
2019-90M	4.823	4.936	4.936	4.928	5.093	5.179	5.273	5.362	N/A
2020-Q1	4.521	4.625	4.699	4.692	4.862	4.952	5.043	5.140	-
2020-Q2	4.441	4.439	4.548	4.554	4.716	4.801	4.902	5.005	-4.01%
2020-Q3	4.534	4.525	4.450	4.487	4.652	4.738	4.831	4.945	-2.15%
2020-Q4	4.533	4.524	4.429	4.361	4.571	4.672	4.763	4.859	-2.81%
2021-Q1	4.509	4.499	4.399	4.334	4.439	4.574	4.668	4.767	-2.89%
2021-Q2	4.499	4.481	4.376	4.319	4.411	4.445	4.570	4.675	-2.83%
2021-Q3	4.471	4.455	4.335	4.280	4.366	4.394	4.422	4.565	-3.26%
2021-Q4	4.467	4.455	4.330	4.263	4.351	4.381	4.402	4.463	-2.24%
2021-124M	4.319	4.297	4.279	4.219	4.322	4.361	4.404	4.489	N/A

Pseudo-perplexity results (lower is better) of all models in the Twitter test sets sampled from different quarters (each quarter correspond to three months. Q1: Jan-Mar; Q2: Apr-Jun; Q3: Jul-Sep; Q4: Oct-Dec). The last column reports difference in pseudo-perplexity, comparing the value obtained for each quarter's test set, between the model trained on the previous quarter and the model updated with data from that same quarter



Time LMs

Models	2020-Q1	2020-Q2	2020-Q3	2020-Q4	2021-Q1	2021-Q2	2021-Q3	2021-Q4	Change
2019-90M	4.823	4.936	4.936	4.928	5.093	5.179	5.273	5.362	N/A
2020-Q1	4.521	4.625	4.699	4.692	4.862	4.952	5.043	5.140	-
2020-Q2	4.441	4.439	4.548	4.554	4.716	4.801	4.902	5.005	-4.01%
2020-Q3	4.534	4.525	4.450	4.487	4.652	4.738	4.831	4.945	-2.15%
2020-Q4	4.533	4.524	4.429	4.361	4.571	4.672	4.763	4.859	-2.81%
2021-Q1	4.509	4.499	4.399	4.334	4.439	4.574	4.668	4.767	-2.89%
2021-Q2	4.499	4.481	4.376	4.319	4.411	4.445	4.570	4.675	-2.83%
2021-Q3	4.471	4.455	4.335	4.280	4.366	4.394	4.422	4.565	-3.26%
2021-Q4	4.467	4.455	4.330	4.263	4.351	4.381	4.402	4.463	-2.24%
2021-124M	4.319	4.297	4.279	4.219	4.322	4.361	4.404	4.489	N/A

Pseudo-perplexity results (lower is better) of all models in the Twitter test sets sampled from different quarters (each quarter correspond to three months. Q1: Jan-Mar; Q2: Apr-Jun; Q3: Jul-Sep; Q4: Oct-Dec). The last column reports difference in pseudo-perplexity, comparing the value obtained for each quarter's test set, between the model trained on the previous quarter and the model updated with data from that same quarter



Time LMs

Models	2020-Q1	2020-Q2	2020-Q3	2020-Q4	2021-Q1	2021-Q2	2021-Q3	2021-Q4	Change
2019-90M	4.823	4.936	4.936	4.928	5.093	5.179	5.273	5.362	N/A
2020-Q1	4.521	4.625	4.699	4.692	4.862	4.952	5.043	5.140	-
2020-Q2	4.441	4.439	4.548	4.554	4.716	4.801	4.902	5.005	-4.01%
2020-Q3	4.534	4.525	4.450	4.487	4.652	4.738	4.831	4.945	-2.15%
2020-Q4	4.533	4.524	4.429	4.361	4.571	4.672	4.763	4.859	-2.81%
2021-Q1	4.509	4.499	4.399	4.334	4.439	4.574	4.668	4.767	-2.89%
2021-Q2	4.499	4.481	4.376	4.319	4.411	4.445	4.570	4.675	-2.83%
2021-Q3	4.471	4.455	4.335	4.280	4.366	4.394	4.422	4.565	-3.26%
2021-Q4	4.467	4.455	4.330	4.263	4.351	4.381	4.402	4.463	-2.24%
2021-124M	4.319	4.297	4.279	4.219	4.322	4.361	4.404	4.489	N/A

Pseudo-perplexity results (lower is better) of all models in the Twitter test sets sampled from different quarters (each quarter correspond to three months. Q1: Jan-Mar; Q2: Apr-Jun; Q3: Jul-Sep; Q4: Oct-Dec). The last column reports difference in pseudo-perplexity, comparing the value obtained for each quarter's test set, between the model trained on the previous quarter and the model updated with data from that same quarter



Time LMs

Models	2020-Q1	2020-Q2	2020-Q3	2020-Q4	2021-Q1	2021-Q2	2021-Q3	2021-Q4	Change
2019-90M	4.823	4.936	4.936	4.928	5.093	5.179	5.273	5.362	N/A
2020-Q1	4.521	4.625	4.699	4.692	4.862	4.952	5.043	5.140	-
2020-Q2	<u>4.441</u>	<u>4.439</u>	4.548	4.554	4.716	4.801	4.902	5.005	-4.01%
2020-Q3	4.534	4.525	<u>4.450</u>	4.487	4.652	4.738	4.831	4.945	-2.15%
2020-Q4	4.533	4.524	4.429	<u>4.361</u>	4.571	4.672	4.763	4.859	-2.81%
2021-Q1	4.509	4.499	4.399	4.334	<u>4.439</u>	4.574	4.668	4.767	-2.89%
2021-Q2	4.499	4.481	4.376	4.319	4.411	4.445	4.570	4.675	-2.83%
2021-Q3	4.471	4.455	4.335	4.280	4.366	4.394	4.422	4.565	-3.26%
2021-Q4	4.467	4.455	<u>4.330</u>	<u>4.263</u>	<u>4.351</u>	<u>4.381</u>	<u>4.402</u>	<u>4.463</u>	-2.24%
2021-124M	4.319	4.297	4.279	4.219	4.322	4.361	4.404	4.489	N/A

Pseudo-perplexity results (lower is better) of all models in the Twitter test sets sampled from different quarters (each quarter correspond to three months. Q1: Jan-Mar; Q2: Apr-Jun; Q3: Jul-Sep; Q4: Oct-Dec). The last column reports difference in pseudo-perplexity, comparing the value obtained for each quarter's test set, between the model trained on the previous quarter and the model updated with data from that same quarter

Time LMs



Model	So glad I'm <mask> vaccinated.	I keep forgetting to bring a <mask>.	Looking forward to watching <mask> Game tonight!
2020-Q1	not getting self	bag purse charger	the The this
2020-Q2	not getting fully	mask bag purse	The the End
2020-Q3	not getting fully	mask bag purse	the The End
2020-Q4	not getting fully	bag purse charger	the The End
2021-Q1	getting not fully	purse charger bag	the The End
2021-Q2	fully getting not	bag charger lighter	the The this
2021-Q3	fully getting not	charger bag purse	the The This
2021-Q4	fully getting not	bag lighter charger	Squid the The

Time LMs



Model	So glad I'm <mask> vaccinated.	I keep forgetting to bring a <mask>.	Looking forward to watching <mask> Game tonight!
2020-Q1	not getting self	bag purse charger	the The this
2020-Q2	not getting fully	mask bag purse	The the End
2020-Q3	not getting fully	mask bag purse	the The End
2020-Q4	not getting fully	bag purse charger	the The End
2021-Q1	getting not fully	purse charger bag	the The End
2021-Q2	fully getting not	bag charger lighter	the The this
2021-Q3	fully getting not	charger bag purse	the The This
2021-Q4	fully getting not	bag lighter charger	Squid the The

Time LMs



Model	So glad I'm <mask> vaccinated.	I keep forgetting to bring a <mask>.	Looking forward to watching <mask> Game tonight!
2020-Q1	not getting self	bag purse charger	the The this
2020-Q2	not getting fully	mask bag purse	The the End
2020-Q3	not getting fully	mask bag purse	the The End
2020-Q4	not getting fully	bag purse charger	the The End
2021-Q1	getting not fully	purse charger bag	the The End
2021-Q2	fully getting not	bag charger lighter	the The this
2021-Q3	fully getting not	charger bag purse	the The This
2021-Q4	fully getting not	bag lighter charger	Squid the The

Time LMs



Model	So glad I'm <mask> vaccinated.	I keep forgetting to bring a <mask>.	Looking forward to watching <mask> Game tonight!
2020-Q1	not getting self	bag purse charger	the The this
2020-Q2	not getting fully	mask bag purse	The the End
2020-Q3	not getting fully	mask bag purse	the The End
2020-Q4	not getting fully	bag purse charger	the The End
2021-Q1	getting not fully	purse charger bag	the The End
2021-Q2	fully getting not	bag charger lighter	the The this
2021-Q3	fully getting not	charger bag purse	the The This
2021-Q4	fully getting not	bag lighter charger	Squid the The





TempoWiC

TempoWiC: An Evaluation Benchmark for Detecting Meaning Shift in Social Media

Daniel Loureiro[◇]

Aminette D'Souza^{*◇}, Areej Nasser Muhajab^{*◇}, Isabella A. White^{*◇}, Gabriel Wong^{*◇}
Luis Espinosa Anke^{◇♡}, Leonardo Neves[♣], Francesco Barbieri[♣], Jose Camacho-Collados[◇]

[◇] Cardiff NLP, School of Computer Science and Informatics, Cardiff University, UK

[♣] Snap Inc., Santa Monica, California, USA; [♡] AMPLYFI, UK

[◇] boucanovaloureirod@cardiff.ac.uk, cardiffnlp.contact@gmail.com

Label	Target	Context-1	Context-2
F	bed	There's a lot of trash on the <u>bed</u> of the river	I keep a glass of water next to my <u>bed</u> when I sleep
F	land	The pilot managed to <u>land</u> the airplane safely	The enemy <u>landed</u> several of our aircrafts
F	justify	<u>Justify</u> the margins	The end <u>justifies</u> the means
T	beat	We <u>beat</u> the competition	Agassi <u>beat</u> Becker in the tennis championship
T	air	<u>Air</u> pollution	Open a window and let in some <u>air</u>
T	window	The expanded <u>window</u> will give us time to catch the thieves	You have a two-hour <u>window</u> of clear weather to finish working on the lawn



TempoWiC

TempoWiC: An Evaluation Benchmark for Detecting Meaning Shift in Social Media

Daniel Loureiro[◇]

Aminette D'Souza^{*◇}, Areej Nasser Muhajab^{*◇}, Isabella A. White^{*◇}, Gabriel Wong^{*◇}
Luis Espinosa Anke^{◇♡}, Leonardo Neves[♣], Francesco Barbieri[♣], Jose Camacho-Collados[◇]

[◇] Cardiff NLP, School of Computer Science and Informatics, Cardiff University, UK

[♣] Snap Inc., Santa Monica, California, USA; [♡] AMPLYFI, UK

[◇] boucanovaloureirod@cardiff.ac.uk, cardiffnlp.contact@gmail.com

Label	Target	Context-1	Context-2
F	bed	There's a lot of trash on the <u>bed</u> of the river	I keep a glass of water next to my <u>bed</u> when I sleep
F	land	The pilot managed to <u>land</u> the airplane safely	The enemy <u>landed</u> several of our aircrafts
F	justify	<u>Justify</u> the margins	The end <u>justifies</u> the means
T	beat	We <u>beat</u> the competition	Agassi <u>beat</u> Becker in the tennis championship
T	air	<u>Air</u> pollution	Open a window and let in some <u>air</u>
T	window	The expanded <u>window</u> will give us time to catch the thieves	You have a two-hour <u>window</u> of clear weather to finish working on the lawn



TempoWiC

TempoWiC: An Evaluation Benchmark for Detecting Meaning Shift in Social Media

Daniel Loureiro[◇]

Aminette D'Souza^{*◇}, Areej Nasser Muhajab^{*◇}, Isabella A. White^{*◇}, Gabriel Wong^{*◇}
Luis Espinosa Anke^{◇♡}, Leonardo Neves[♣], Francesco Barbieri[♣], Jose Camacho-Collados[◇]

[◇] Cardiff NLP, School of Computer Science and Informatics, Cardiff University, UK

[♣] Snap Inc., Santa Monica, California, USA; [♡] AMPLYFI, UK

[◇] boucanovaloureirod@cardiff.ac.uk, cardiffnlp.contact@gmail.com

Label Target Context-1

Context-2

F Corona I love to drink corona

I hate corona and lockdown

T Corona I love to drink corona

A cold corona at the beach



TempoWiC

1. Select trending terms
2. Take tweet examples from before and after trending date
3. Annotate the examples to see if there was a meaning shift

	Word	# Instances (% Diff. Meaning)	Trending Date	Agreement (Krippendorf's α)
Train	frisk	99 (54%)	11/2/2020	0.718
	pogrom	99 (5%)	25/2/2020	0.482
	containment	100 (33%)	12/3/2020	0.274
	virus	96 (48%)	12/3/2020	0.254
	epicenter	100 (71%)	14/3/2020	0.124
	ventilator	99 (17%)	27/3/2020	0.541
	villager	100 (64%)	10/4/2020	0.546
	turnip	100 (95%)	10/5/2020	0.316
	bunker	98 (61%)	1/6/2020	0.408
	mask	99 (76%)	14/7/2020	0.255
	teargas	98 (3%)	18/7/2020	0.786
	paternity	100 (22%)	30/7/2020	0.289
	entanglement	99 (89%)	1/8/2020	0.623
	folklore	82 (92%)	3/8/2020	0.917
parasol	100 (85%)	2/9/2020	0.446	
Validation	impostor	99 (76%)	23/9/2020	0.544
	lotte	98 (43%)	27/9/2020	0.514
	recount	100 (28%)	6/11/2020	0.682
	primo	100 (77%)	9/11/2020	0.528
Test	milker	99 (50%)	4/3/2021	0.699
	moxie	97 (83%)	5/3/2021	0.755
	unlabeled	100 (90%)	10/3/2021	0.711
	pyre	100 (32%)	27/4/2021	0.243
	gaza	100 (60%)	15/5/2021	0.749
	ido	91 (83%)	27/5/2021	0.712
	airdrop	99 (40%)	6/6/2021	0.918
	bullpen	99 (9%)	16/6/2021	0.388
	crt	100 (68%)	26/6/2021	0.867
	monet	98 (94%)	8/7/2021	1.000
	burnham	100 (16%)	1/8/2021	0.964
	delta	100 (100%)	11/8/2021	1.000
	gala	100 (46%)	14/9/2021	0.498
	launchpad	99 (81%)	17/9/2021	0.558
vanguard	99 (95%)	21/9/2021	0.421	

Table 2: Details all the words included in TempoWiC. Maximum pairwise agreement is reported.

TempoWiC

1. Select trending terms
2. Take tweet examples from before and after trending date
3. Annotate the examples to see if there was a meaning shift



	Word	# Instances (% Diff. Meaning)	Trending Date	Agreement (Krippendorf's α)
Train	frisk	99 (54%)	11/2/2020	0.718
	pogrom	99 (5%)	25/2/2020	0.482
	containment	100 (33%)	12/3/2020	0.274
	virus	96 (48%)	12/3/2020	0.254
	epicenter	100 (71%)	14/3/2020	0.124
	ventilator	99 (17%)	27/3/2020	0.541
	villager	100 (64%)	10/4/2020	0.546
	turnip	100 (95%)	10/5/2020	0.316
	bunker	98 (61%)	17/6/2020	0.408
	mask	99 (76%)	14/7/2020	0.255
	teargas	98 (3%)	18/7/2020	0.786
	paternity	100 (22%)	30/7/2020	0.289
	entanglement	99 (89%)	1/8/2020	0.623
	folklore	82 (92%)	3/8/2020	0.917
	parasol	100 (85%)	2/9/2020	0.446
Validation	impostor	99 (76%)	23/9/2020	0.544
	lotte	98 (43%)	27/9/2020	0.514
	recount	100 (28%)	6/11/2020	0.682
	primo	100 (77%)	9/11/2020	0.528
Test	milker	99 (50%)	4/3/2021	0.699
	moxie	97 (83%)	5/3/2021	0.755
	unlabeled	100 (90%)	10/3/2021	0.711
	pyre	100 (32%)	27/4/2021	0.243
	gaza	100 (60%)	15/5/2021	0.749
	ido	91 (83%)	27/5/2021	0.712
	airdrop	99 (40%)	6/6/2021	0.918
	bullpen	99 (9%)	16/6/2021	0.388
	crt	100 (68%)	26/6/2021	0.867
	monet	98 (94%)	8/7/2021	1.000
	burnham	100 (16%)	1/8/2021	0.964
delta	100 (100%)	11/8/2021	1.000	
gala	100 (46%)	14/9/2021	0.498	
launchpad	99 (81%)	17/9/2021	0.558	
vanguard	99 (95%)	21/9/2021	0.421	



Table 2: Details all the words included in TempoWiC. Maximum pairwise agreement is reported.



TempoWiC

	Model	Accuracy
Fine-tuning	RoBERTa-base	66.89%
	RoBERTa-large	66.49%
	TimeLMs-2019-90M	66.46%
	TimeLMs-2021-124M	65.04%
	BERTweet-base	61.46%
	BERTweet-large	<u>67.93%</u>
Similarity	RoBERTa-base	67.96%
	RoBERTa-large	72.98%
	TimeLMs-2019-90M	<u>74.07%</u>
	TimeLMs-2021-124M	71.01%
	BERTweet-base	69.45%
	BERTweet-large	69.18%



TempoWiC

	Model	Accuracy
Fine-tuning	RoBERTa-base	66.89%
	RoBERTa-large	66.49%
	TimeLMs-2019-90M	66.46%
	TimeLMs-2021-124M	65.04%
	BERTweet-base	61.46%
	BERTweet-large	<u>67.93%</u>
Similarity	RoBERTa-base	67.96%
	RoBERTa-large	72.98%
	TimeLMs-2019-90M	<u>74.07%</u>
	TimeLMs-2021-124M	71.01%
	BERTweet-base	69.45%
	BERTweet-large	69.18%



Topic Classification

Twitter Topic Classification

Dimosthenis Antypas*, **Asahi Ushio***, **Jose Camacho-Collados**

Cardiff NLP, School of Computer Science and Informatics, Cardiff University, United Kingdom

{AntypasD, UshioA, CamachoColladosJ}@cardiff.ac.uk

Leonardo Neves, **Vítor Silva**, **Francesco Barbieri**

Snap Inc., Santa Monica, CA, United States

{lneves, vsilvasousa, fbarbieri}@snap.com

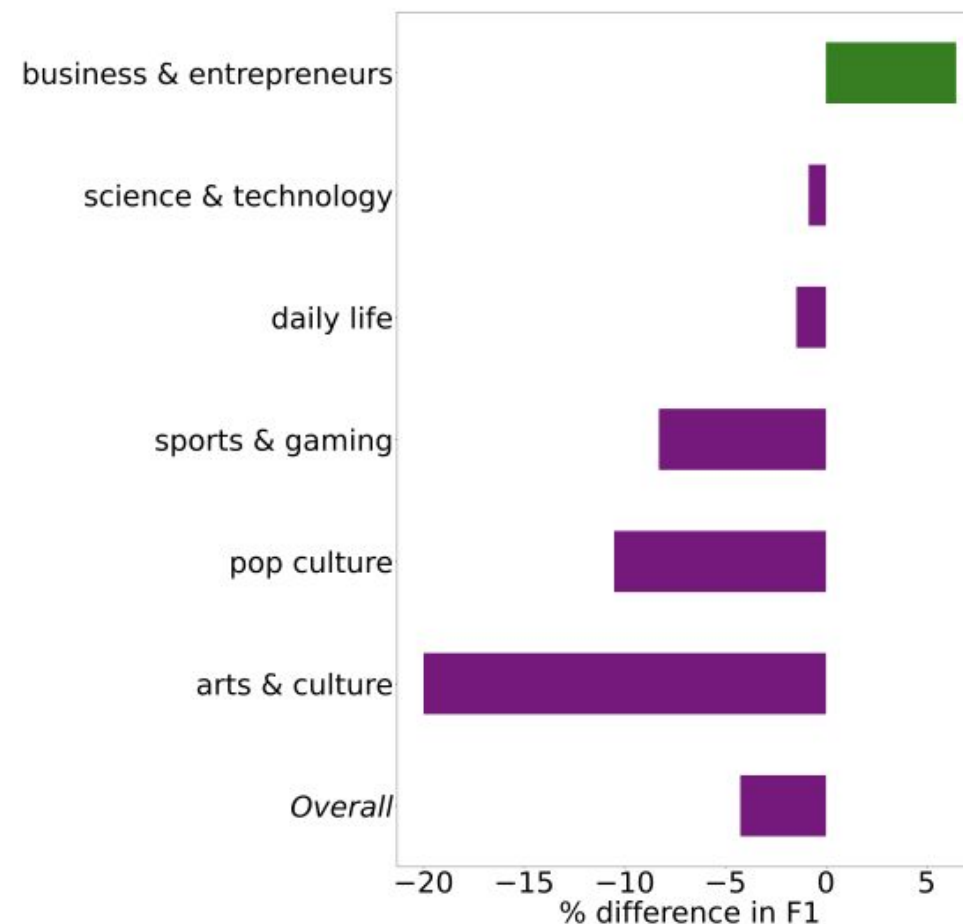


Figure 4: Relative (%) differences in F1 scores when *TimeLM-19* is trained in a temporal and in a random setting for the single-label setting. Negative values indicate that when using the temporal split the model’s performance decreases.



Topic Classification

Twitter Topic Classification

Dimosthenis Antypas*, **Asahi Ushio***, **Jose Camacho-Collados**

Cardiff NLP, School of Computer Science and Informatics, Cardiff University, United Kingdom

{AntypasD, UshioA, CamachoColladosJ}@cardiff.ac.uk

Leonardo Neves, **Vítor Silva**, **Francesco Barbieri**

Snap Inc., Santa Monica, CA, United States

{lneves, vsilvasousa, fbarbieri}@snap.com

Class	TimeLM-19		TimeLM-21	
	temp	rand	temp	rand
arts & culture	21.3	39.1	35.4	44.8
business & entrepreneurs	58.6	55.3	56.3	54.0
celebrity & pop culture	48.6	47.8	46.4	57.6
diaries & daily life	44.5	51.2	44.7	49.8
family	46.4	55.2	53.1	56.2
fashion & style	66.4	75.2	67.2	75.2
film tv & video	66.1	72.2	65.4	70.6
fitness & health	55.7	42.2	58.6	52.6
food & dining	75.4	70.7	80.4	71.6
gaming	64.6	69.2	64.8	71.2
learning & educational	49.3	47.1	48.9	47.0
music	88.1	87.8	86.9	88.2
news & social concern	84.4	86.2	84.5	85.0
other hobbies	27.7	30.3	31.1	26.2
relationships	35.3	51.6	44.5	54.0
samples avg	74.3	75.2	74.7	75.2
science & technology	50.5	56.0	50.2	52.1
sports	95.6	94.8	95.2	94.8
travel & adventure	57.1	56.0	52.2	54.7
youth & student life	50.4	43.6	50.8	51.0
macro avg	57.2	59.6	58.8	60.9



Topic Classification

Twitter Topic Classification

Dimosthenis Antypas*, **Asahi Ushio***, **Jose Camacho-Collados**

Cardiff NLP, School of Computer Science and Informatics, Cardiff University, United Kingdom

{AntypasD, UshioA, CamachoColladosJ}@cardiff.ac.uk

Leonardo Neves, **Vitor Silva**, **Francesco Barbieri**

Snap Inc., Santa Monica, CA, United States

{lneves, vsilvasousa, fbarbieri}@snap.com

Class	TimeLM-19		TimeLM-21	
	temp	rand	temp	rand
arts & culture	21.3	39.1	35.4	44.8
business & entrepreneurs	58.6	55.3	56.3	54.0
celebrity & pop culture	48.6	47.8	46.4	57.6
diaries & daily life	44.5	51.2	44.7	49.8
family	46.4	55.2	53.1	56.2
fashion & style	66.4	75.2	67.2	75.2
film tv & video	66.1	72.2	65.4	70.6
fitness & health	55.7	42.2	58.6	52.6
food & dining	75.4	70.7	80.4	71.6
gaming	64.6	69.2	64.8	71.2
learning & educational	49.3	47.1	48.9	47.0
music	88.1	87.8	86.9	88.2
news & social concern	84.4	86.2	84.5	85.0
other hobbies	27.7	30.3	31.1	26.2
relationships	35.3	51.6	44.5	54.0
samples avg	74.3	75.2	74.7	75.2
science & technology	50.5	56.0	50.2	52.1
sports	95.6	94.8	95.2	94.8
travel & adventure	57.1	56.0	52.2	54.7
youth & student life	50.4	43.6	50.8	51.0
macro avg	57.2	59.6	58.8	60.9



Topic Classification

Twitter Topic Classification

Dimosthenis Antypas*, **Asahi Ushio***, **Jose Camacho-Collados**

Cardiff NLP, School of Computer Science and Informatics, Cardiff University, United Kingdom

{AntypasD, UshioA, CamachoColladosJ}@cardiff.ac.uk

Leonardo Neves, **Vítor Silva**, **Francesco Barbieri**

Snap Inc., Santa Monica, CA, United States

{lneves, vsilvasousa, fbarbieri}@snap.com

Class	TimeLM-19		TimeLM-21	
	temp	rand	temp	rand
arts & culture	21.3	39.1	35.4	44.8
business & entrepreneurs	58.6	55.3	56.3	54.0
celebrity & pop culture	48.6	47.8	46.4	57.6
diaries & daily life	44.5	51.2	44.7	49.8
family	46.4	55.2	53.1	56.2
fashion & style	66.4	75.2	67.2	75.2
film tv & video	66.1	72.2	65.4	70.6
fitness & health	55.7	42.2	58.6	52.6
food & dining	75.4	70.7	80.4	71.6
gaming	64.6	69.2	64.8	71.2
learning & educational	49.3	47.1	48.9	47.0
music	88.1	87.8	86.9	88.2
news & social concern	84.4	86.2	84.5	85.0
other hobbies	27.7	30.3	31.1	26.2
relationships	35.3	51.6	44.5	54.0
samples avg	74.3	75.2	74.7	75.2
science & technology	50.5	56.0	50.2	52.1
sports	95.6	94.8	95.2	94.8
travel & adventure	57.1	56.0	52.2	54.7

macro avg	57.2	59.6	58.8	60.9
------------------	-------------	-------------	-------------	-------------



NER

Named Entity Recognition in Twitter: A Dataset and Analysis on Short-Term Temporal Shifts

Asahi Ushio¹, Leonardo Neves², Vítor Silva², Francesco Barbieri², Jose Camacho-Collados¹

¹Cardiff NLP, School of Computer Science and Informatics, Cardiff University, United Kingdom

{UshioA, CamachoColladosJ}@cardiff.ac.uk

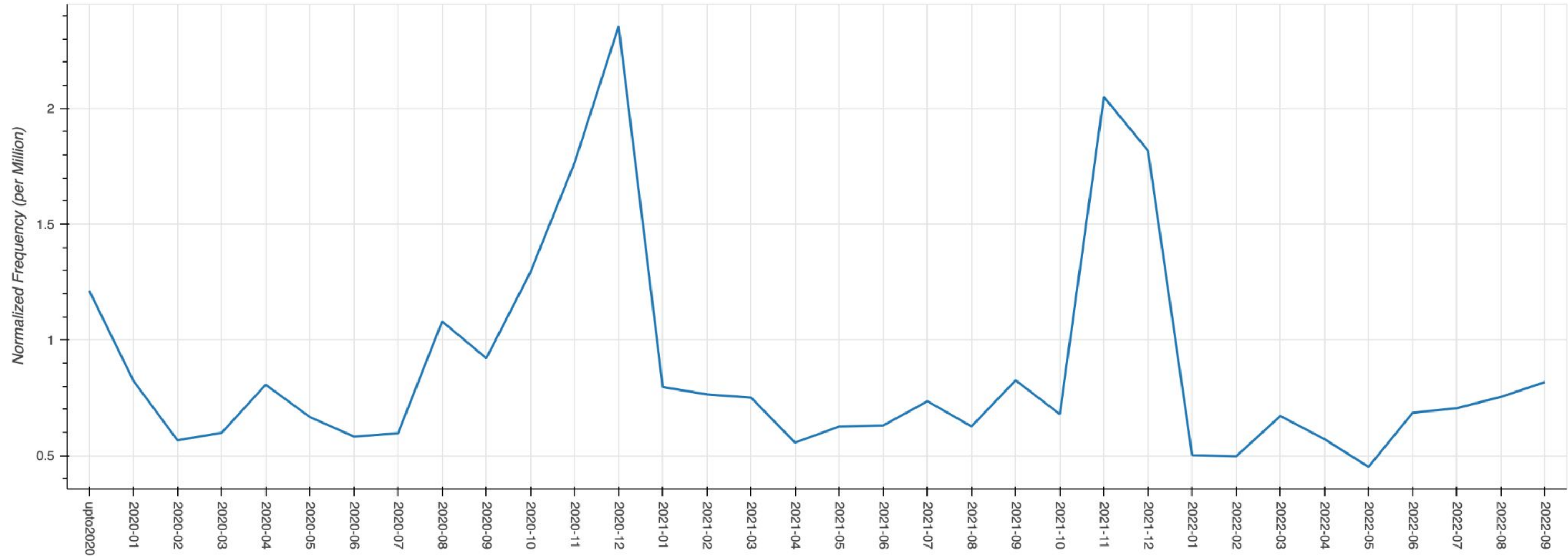
²Snap Inc., Santa Monica, CA, United States

{lneves, vsilvasousa, fbarbieri}@snap.com

Model	Macro F1 2021 / 2020
TimeLM ₂₀₁₉	59.3 / 61.1
TimeLM ₂₀₂₀	58.3 / 60.3
TimeLM ₂₀₂₁	59.5 / 61.1

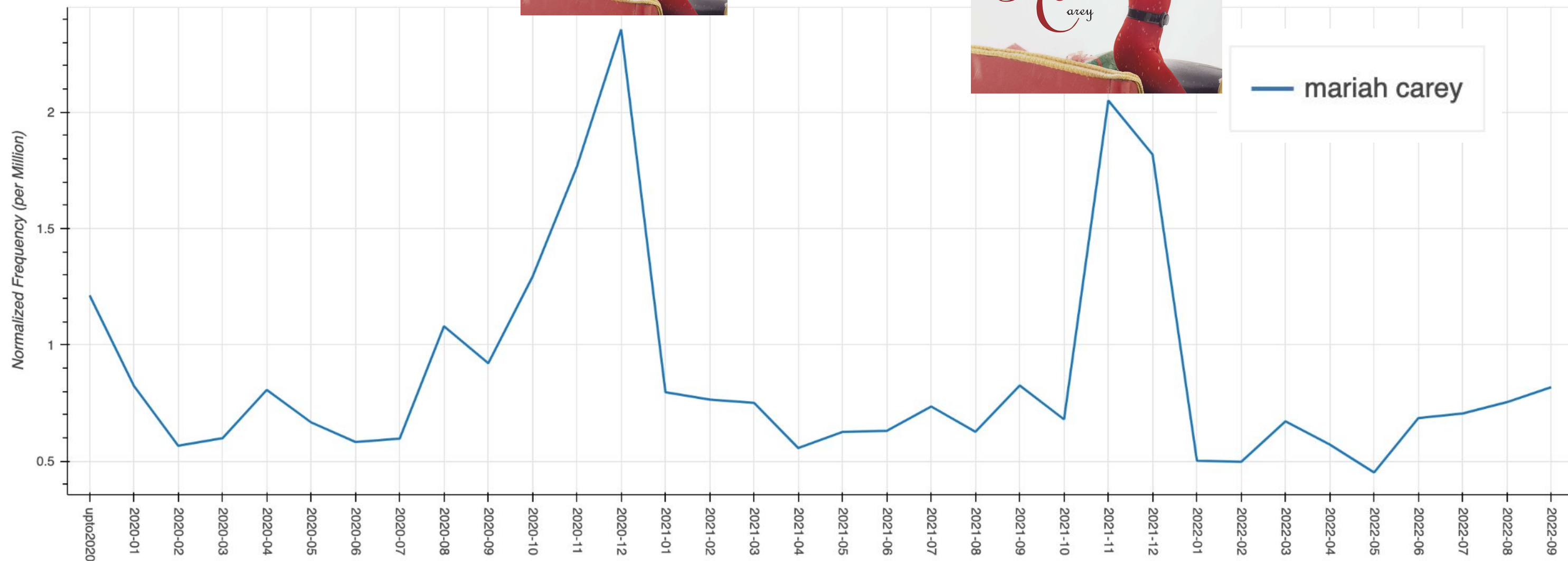


Can we use recurrent patterns of time?



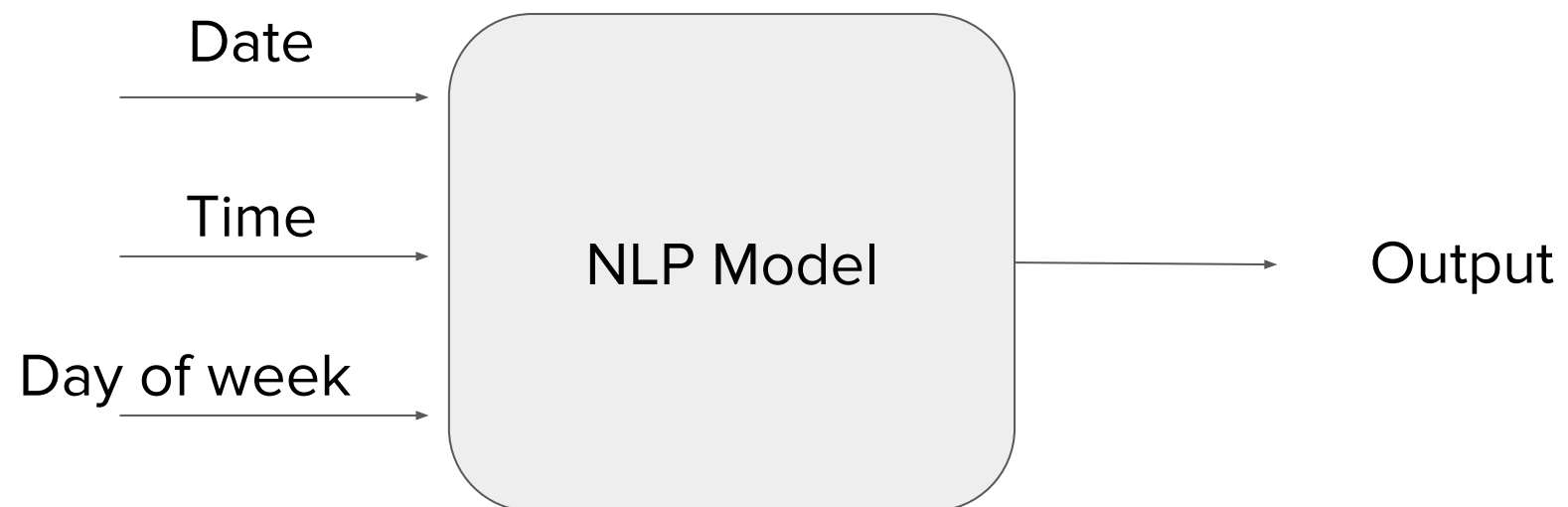


Can we use rec patterns






Can we use recurrent patterns of time?





Roger Federer  @rogerfederer · Mar 15



Can't wait to visit [#Chicago](#) for the first time next week. What shall I do while I am in town? [#LaverCup](#) 

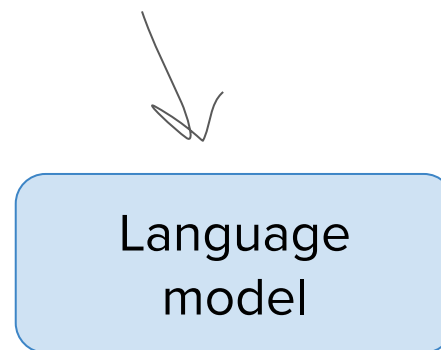


Roger Federer @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup



Can't wait to visit #Chicago for the first time next Week. What shall I do while I am in town?
#LaverCup <mask>





Roger Federer @rogerfederer · Mar 15

Can't wait to visit [#Chicago](#) for the first time next week. What shall I do while I am in town? [#LaverCup](#)

Can't wait to visit [#Chicago](#) for the first time next Week. What shall I do while I am in town?
[#LaverCup](#) **<mask>**

8 am
Wed
March

Language model





Hour

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

Months

jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec	



Seasons

Exploring Emoji Usage and Prediction Through a Temporal Variation Lens

Francesco Barbieri* Luís Marujo♡ Pradeep Karuturi♡

William Brendel♡ Horacio Saggion*

* Large Scale Text Understanding Systems Lab, TALN, UPF, Barcelona, Spain

♡ Snap Inc. Research, Venice, California, USA

RANK	SP	SU	AU	WI
1	❤️	❤️	❤️	❤️
2	😍	😍	😂	😂
3	😂	😂	😍	😍
4	🔥	🔥	🔥	🔥
5	💕	☀️	💕	🎄
6	☀️	😎	100	❄️
7	😎	💕	🌟	100
8	👏	💙	👏	💕
9	100	🌟	💙	😄
10	💙	👏	🎄	👏
11	🌟	🇺🇸	😄	😘
12	😄	100	🎉	😎
13	🎉	🎉	🎃	🎉
14	😘	💪	😎	🌟
15	💪	😄	😘	💪

Seasons

Comparing 10 Nearest Neighbours of the same emoji
on different models (seasons)









Spring										
Summer										
Autumn										
Winter										

Seasons

Comparing 10 Nearest Neighbours of the same emoji
on different models (seasons)



Spring										
Summer										
Autumn										
Winter										

Seasons

Comparing 10 Nearest Neighbours of the same emoji
on different models (seasons)



Spring										
Summer										
Autumn										
Winter										



Roger Federer @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup



Can't wait to visit #Chicago for the first time next Week. What shall I do while I am in town?
#LaverCup <mask>

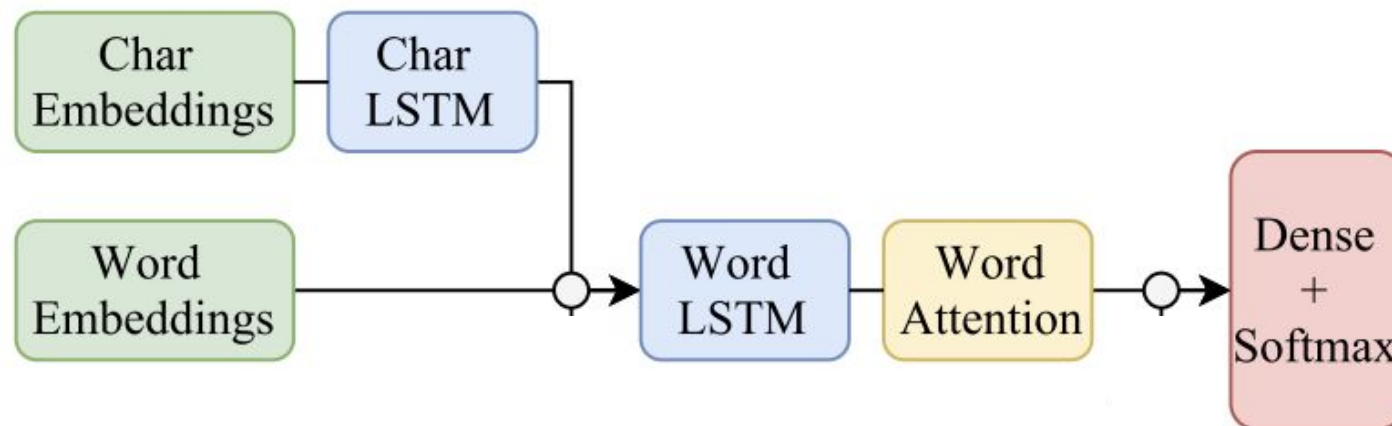
8 am
Wed
March

Language model





Predict emojis using time information



Exploring Emoji Usage and Prediction Through a Temporal Variation Lens

Francesco Barbieri[♣] Lu s Marujo[♡] Pradeep Karuturi[♡]

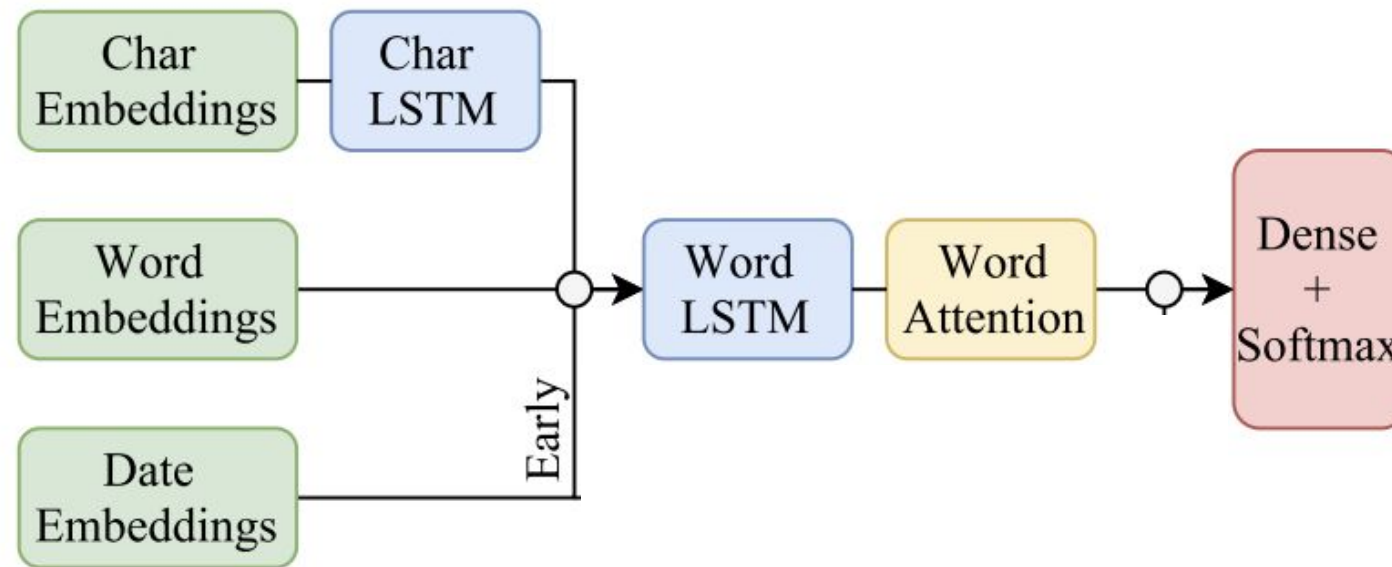
William Brendel[♡] Horacio Saggion[♣]

[♣] Large Scale Text Understanding Systems Lab, TALN, UPF, Barcelona, Spain

[♡] Snap Inc. Research, Venice, California, USA



Predict emojis using time information



Exploring Emoji Usage and Prediction Through a Temporal Variation Lens

Francesco Barbieri[♣] Luís Marujo[♡] Pradeep Karuturi[♡]

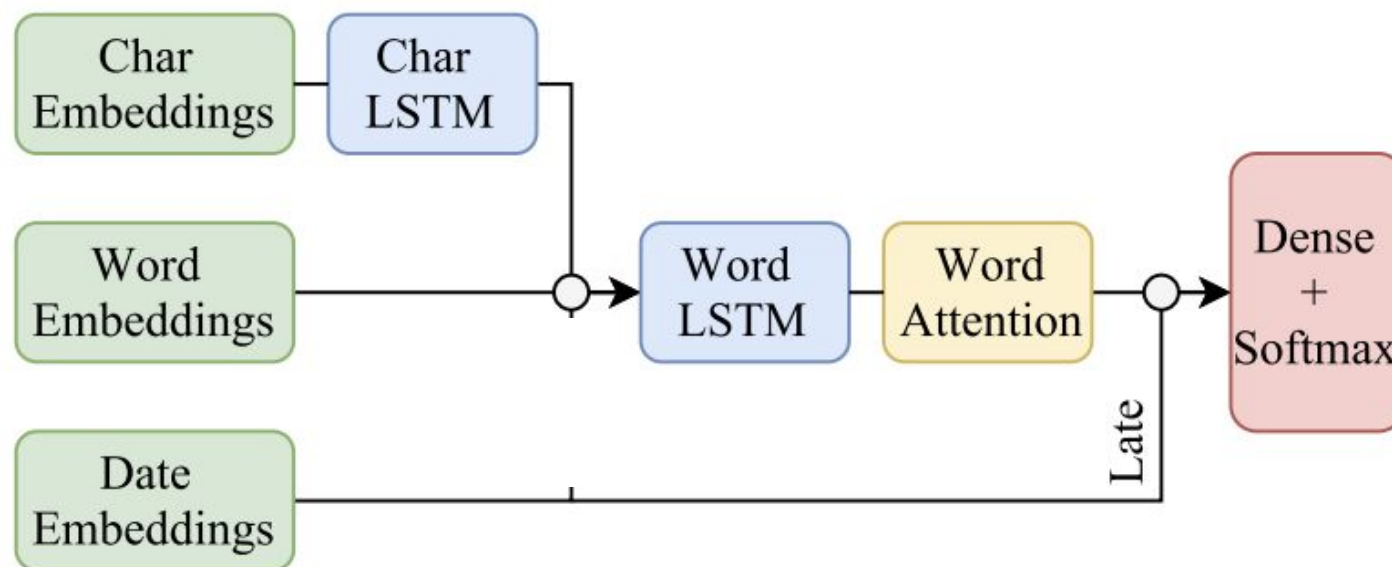
William Brendel[♡] Horacio Saggion[♣]

[♣] Large Scale Text Understanding Systems Lab, TALN, UPF, Barcelona, Spain

[♡] Snap Inc. Research, Venice, California, USA



Predict emojis using time information



Exploring Emoji Usage and Prediction Through a Temporal Variation Lens

Francesco Barbieri[♣] Lu s Marujo[ ] Pradeep Karuturi[ ]

William Brendel[ ] Horacio Saggion[♣]

[♣] Large Scale Text Understanding Systems Lab, TALN, UPF, Barcelona, Spain

[ ] Snap Inc. Research, Venice, California, USA



Predict emojis using time information

Table 5: Results for the three models: without date (W/O), *Early* date fusion, and *Late* date fusion. Precision, Recall, F1, accuracy at 1, 5, 10, and Coverage Error.

	P	R	F1	a@1	a@5	a@10	CE
W/O	21.97	23.22	21.89	23.13	38.22	45.70	44.29
Early	22.10	23.43	22.06	23.33	38.55	46.29	42.59
Late	21.83	23.00	21.63	22.91	37.85	45.62	43.91

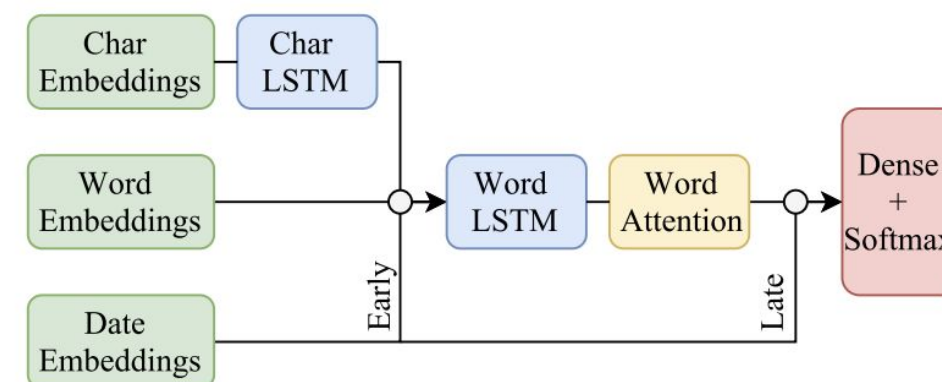
Exploring Emoji Usage and Prediction Through a Temporal Variation Lens

Francesco Barbieri* Lu s Marujo* Pradeep Karuturi*

William Brendel* Horacio Saggion*

* Large Scale Text Understanding Systems Lab, TALN, UPF, Barcelona, Spain

* Snap Inc. Research, Venice, California, USA





Predict emojis using time information

Table 5: Results for the three models: without date (W/O), *Early* date fusion, and *Late* date fusion. Precision, Recall, F1, accuracy at 1, 5, 10, and Coverage Error.

	P	R	F1	a@1	a@5	a@10	CE
W/O	21.97	23.22	21.89	23.13	38.22	45.70	44.29
Early	22.10	23.43	22.06	23.33	38.55	46.29	42.59
Late	21.83	23.00	21.63	22.91	37.85	45.62	43.91

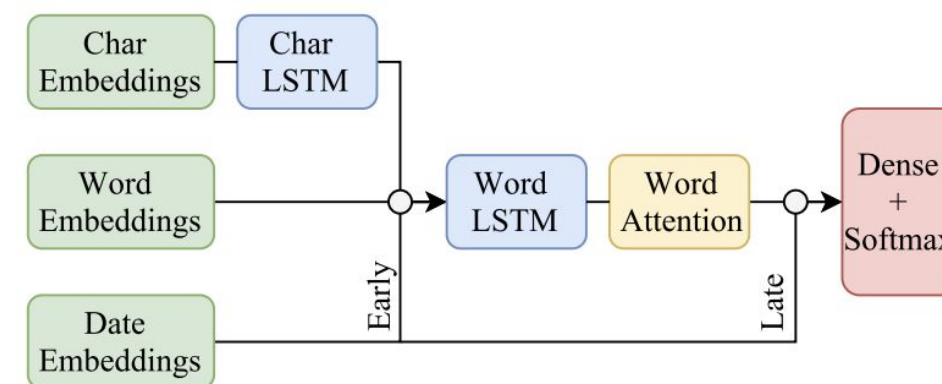
Exploring Emoji Usage and Prediction Through a Temporal Variation Lens

Francesco Barbieri* Luís Marujo[♡] Pradeep Karuturi[♡]

William Brendel[♡] Horacio Saggion*

* Large Scale Text Understanding Systems Lab, TALN, UPF, Barcelona, Spain

[♡] Snap Inc. Research, Venice, California, USA





Predict emojis using time information

Table 5: Results for the three models: without date (W/O), *Early* date fusion, and *Late* date fusion. Precision, Recall, F1, accuracy at 1, 5, 10, and Coverage Error.

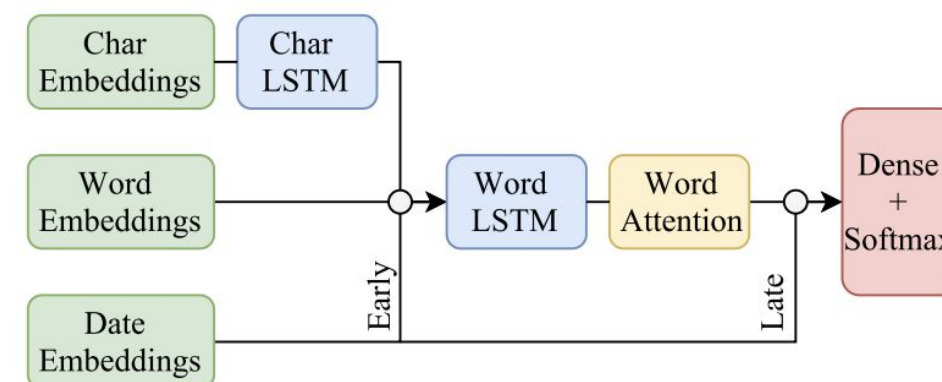
	P	R	F1	a@1	a@5	a@10	CE
W/O	21.97	23.22	21.89	23.13	38.22	45.70	44.29
Early	22.10	23.43	22.06	23.33	38.55	46.29	42.59
Late	21.83	23.00	21.63	22.91	37.85	45.62	43.91

Emoji										
W/O	0.54	0.4	0.15	0.34	0.40	0.4	0.22	0.19	0.30	0.31
Early	0.63	0.11	0.22	0.41	0.47	0.10	0.27	0.24	0.35	0.36
Emoji										
W/O	0.44	0.30	0.36	0.50	0.10	0.12	0.15	0.18	0.22	0.11
Early	0.49	0.34	0.40	0.54	0.14	0.16	0.19	0.22	0.26	0.15

Exploring Emoji Usage and Prediction Through a Temporal Variation Lens

Francesco Barbieri* Luís Marujo[♡] Pradeep Karuturi[♡]
William Brendel[♡] Horacio Saggion*

* Large Scale Text Understanding Systems Lab, TALN, UPF, Barcelona, Spain
[♡] Snap Inc. Research, Venice, California, USA





Roger Federer @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup



Can't wait to visit #Chicago for the first time next Week. What shall I do while I am in town?
#LaverCup

8 am
Wed
March

Language model

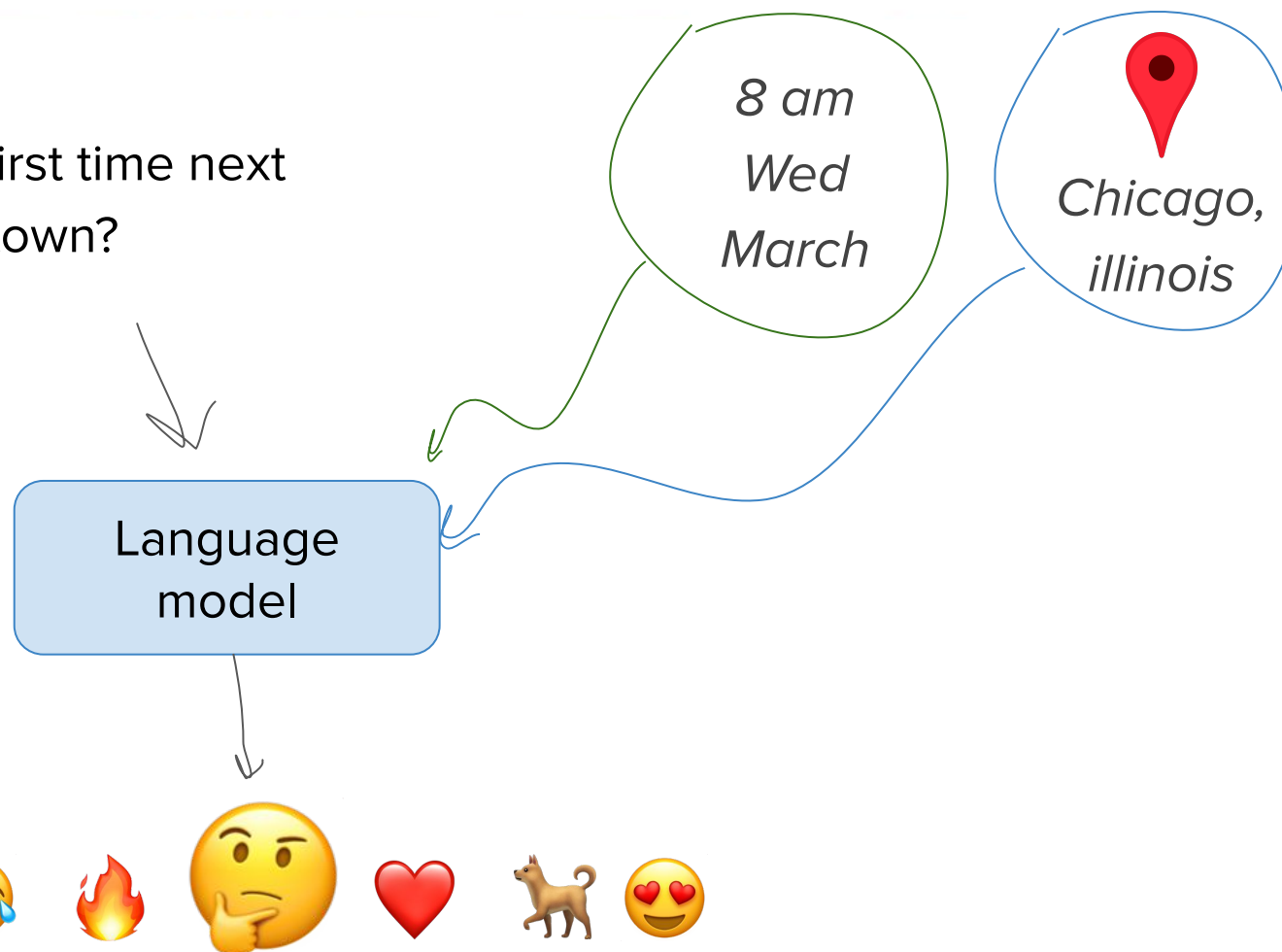




Roger Federer @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup

Can't wait to visit #Chicago for the first time next Week. What shall I do while I am in town? #LaverCup





Why is **location** important?

		Albuquerque
		Atlanta
		Boston
		Charlotte
		Chicago
		Dallas
		Denver
		Detroit
		Houston
		Indianapolis
		Jacksonville
		Kansas City
		Las Vegas
		Los Angeles
		Miami
		Minneapolis
		Nashville
		New York
		Philadelphia
		Phoenix
		Pittsburgh
		Portland
<td></td> <td>Saint Louis</td>		Saint Louis
		San Antonio
		San Diego
		San Francisco



Why is **location** important?

How Cosmopolitan Are Emojis?

Exploring Emojis Usage and Meaning over Different Languages
with Distributional Semantics

Francesco Barbieri
Universitat Pompeu Fabra
Barcelona, Spain
francesco.barbieri@upf.edu

German Kruszewski
University of Trento
Trento, Italy
germank@gmail.com

Francesco Ronzano
Universitat Pompeu Fabra
Barcelona, Spain
francesco.ronzano@upf.edu

Horacio Saggion
Universitat Pompeu Fabra
Barcelona, Spain
horacio.saggion@upf.edu



Why is **location** important?

How Cosmopolitan Are Emojis?

Exploring Emojis Usage and Meaning over Different Languages
with Distributional Semantics

Francesco Barbieri
Universitat Pompeu Fabra
Barcelona, Spain
francesco.barbieri@upf.edu

German Kruszewski
University of Trento
Trento, Italy
germank@gmail.com

Francesco Ronzano
Universitat Pompeu Fabra
Barcelona, Spain
francesco.ronzano@upf.edu

Horacio Saggion
Universitat Pompeu Fabra
Barcelona, Spain
horacio.saggion@upf.edu



NN from Model trained on
tweets from **USA**

- God
- praise
- finally
- thank
- hands



Why is **location** important?

How Cosmopolitan Are Emojis?

Exploring Emojis Usage and Meaning over Different Languages
with Distributional Semantics

Francesco Barbieri
Universitat Pompeu Fabra
Barcelona, Spain
francesco.barbieri@upf.edu

German Kruszewski
University of Trento
Trento, Italy
germank@gmail.com

Francesco Ronzano
Universitat Pompeu Fabra
Barcelona, Spain
francesco.ronzano@upf.edu

Horacio Saggion
Universitat Pompeu Fabra
Barcelona, Spain
horacio.saggion@upf.edu



NN from Model trained on
tweets from **USA**

- God
- praise
- finally
- thank
- hands

NN from Model trained
on tweets from **Spain**

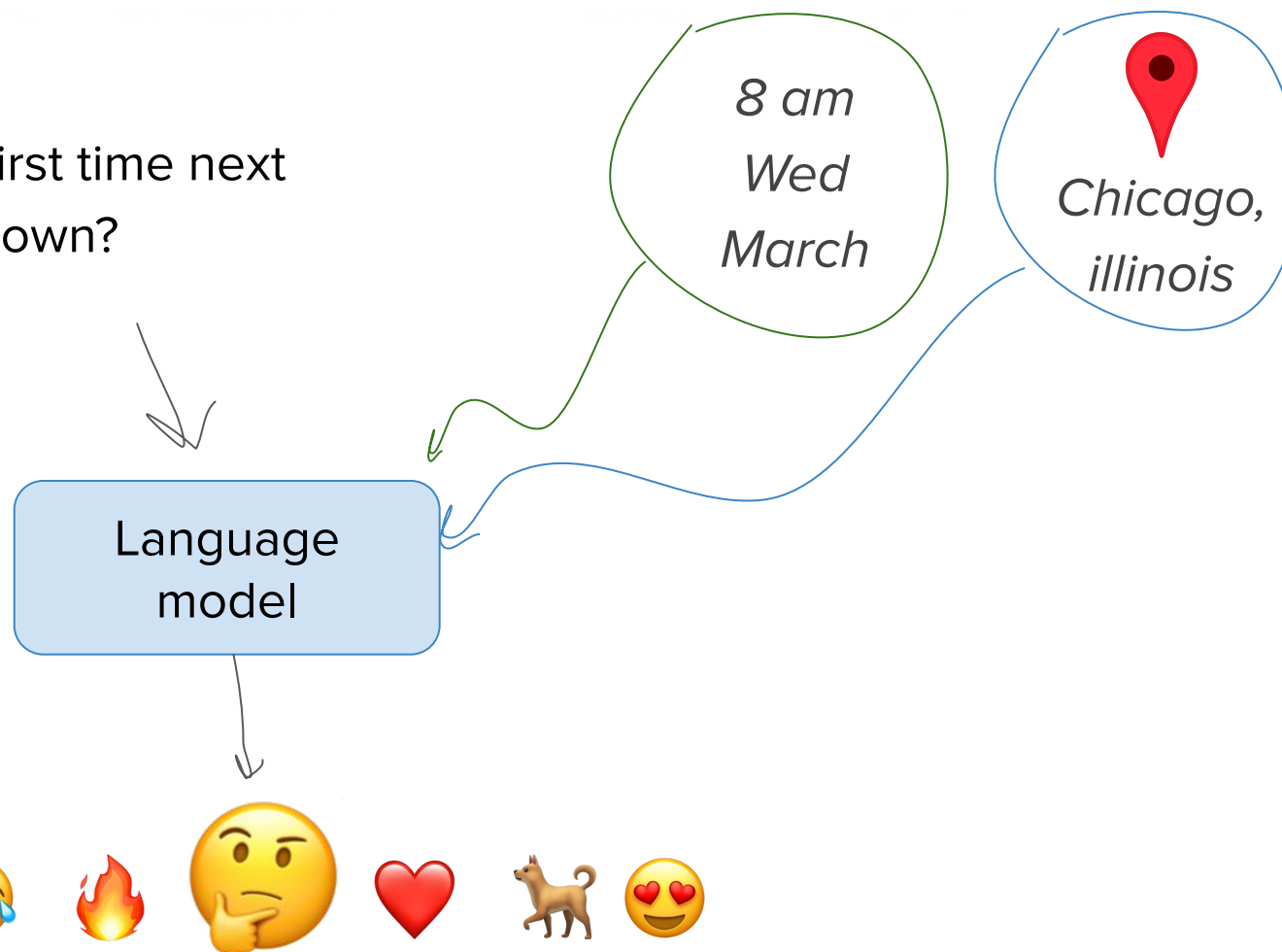
- Ibiza
- festival
- hands
- amnesia
- viernes



Roger Federer @rogerfederer · Mar 15

Can't wait to visit [#Chicago](#) for the first time next week. What shall I do while I am in town? [#LaverCup](#)

Can't wait to visit [#Chicago](#) for the first time next Week. What shall I do while I am in town? [#LaverCup](#)

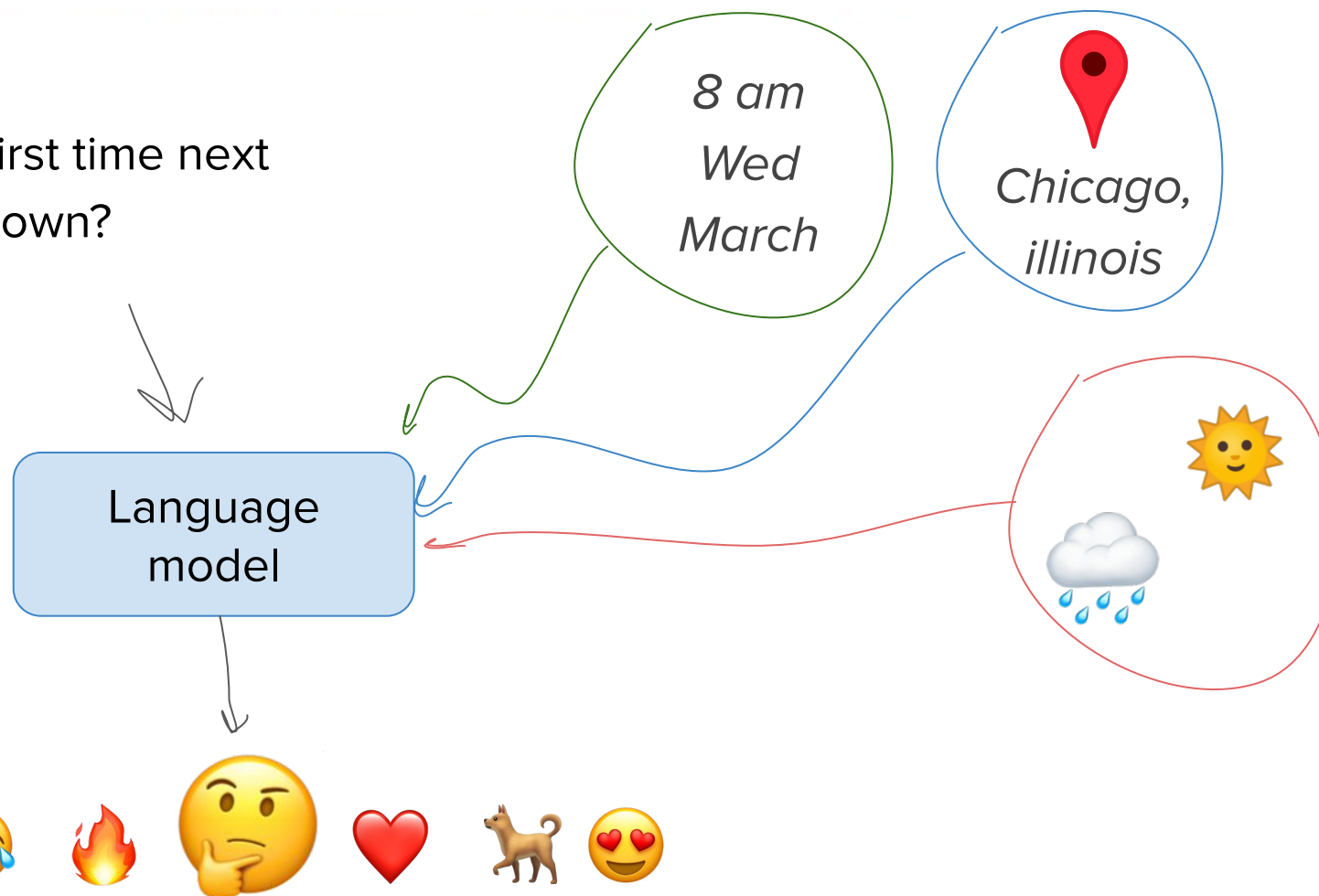




Roger Federer @rogerfederer · Mar 15

Can't wait to visit #Chicago for the first time next week. What shall I do while I am in town? #LaverCup

Can't wait to visit #Chicago for the first time next Week. What shall I do while I am in town? #LaverCup

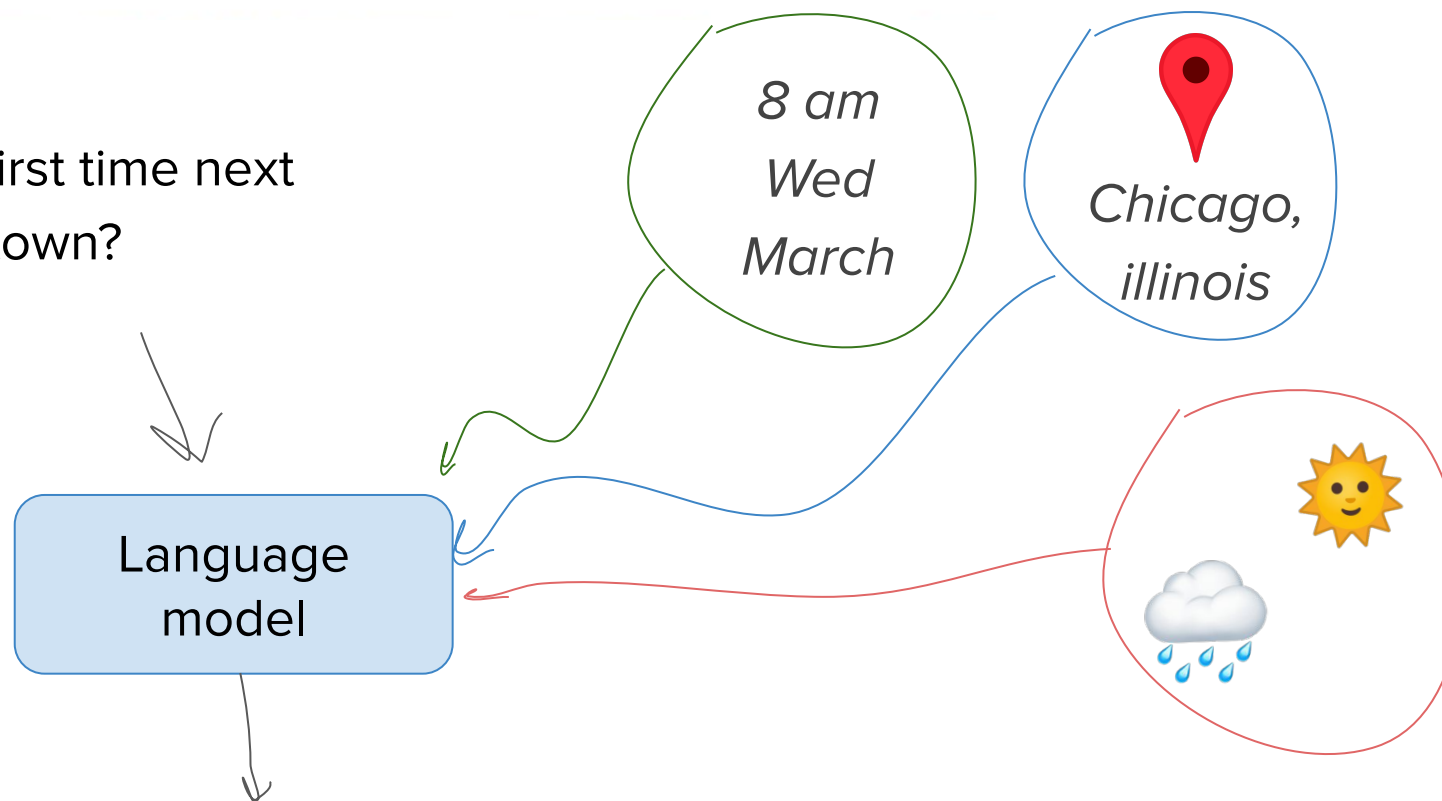




Roger Federer @rogerfederer · Mar 15

Can't wait to visit [#Chicago](#) for the first time next week. What shall I do while I am in town? [#LaverCup](#)

Can't wait to visit [#Chicago](#) for the first time next Week. What shall I do while I am in town? [#LaverCup](#)



Positive, Neutral, Negative

Sunshine with a Chance of Smiles: How Does Weather Impact Sentiment on Social Media?

**Julie Jiang,^{1,2*} Nils Murrugara-Llerena,² Maarten W. Bos,² Yozen Liu,² Neil Shah,²
Leonardo Neves,² Francesco Barbieri²**

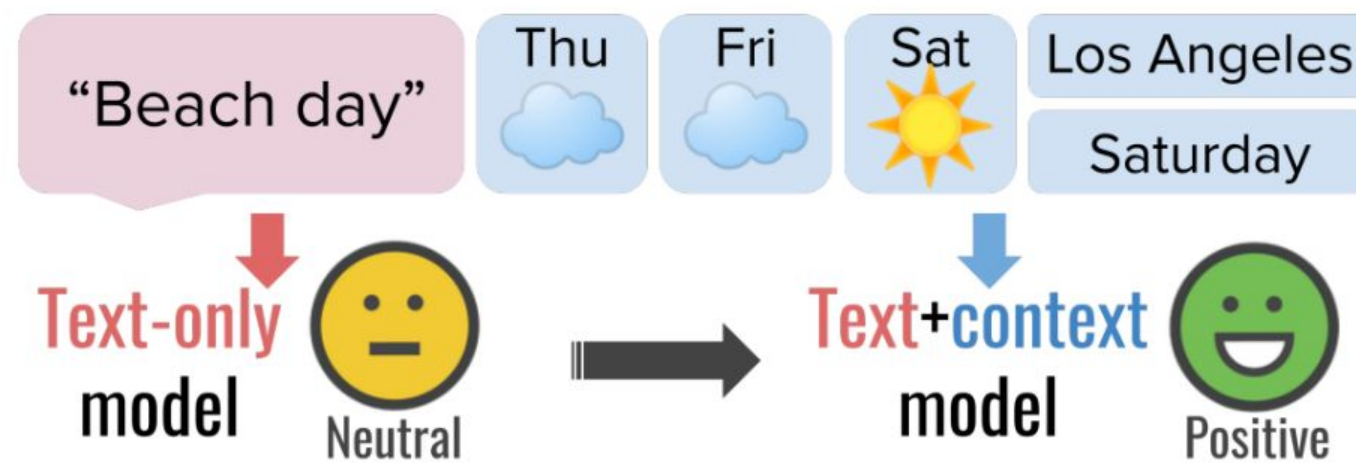
¹ Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292, USA

² Snap Inc., Santa Monica, CA 90405, USA

juliej@isi.edu, {nmurrugarraller, maarten, yliu2, nshah, lneves, fbarbieri}@snap.com

Abstract

The environment we are in can affect our mood and behavior. One environmental factor is weather, which is linked to sentiment as expressed on social media. However, less is known about how integrating changes in weather, along with time and location contextual cues, can improve sentiment detection and understanding. In this paper, we explore the effects of three contextual features—weather, location, and



“Beach day”



state-of-the-art
sentiment classifier:
?

“Beach day”



state-of-the-art
sentiment classifier:
NEUTRAL

“Beach day”



state-of-the-art
sentiment classifier:
NEUTRAL

Los Angeles

SATURDAY

+

Weather is
AMAZING



Contextually-aware classifier:
POSITIVE

Background

Weather impacts our **mood** and **behavior**

(e.g., Howard & Hoffman, 1984)

- Sun => happy
- Too hot => aggressive
- Too humid => irritable
- Rain => depression

It can predict...

- The stock market (Chang et al., 2008)
- The housing market (Hu & Lee, 2020)
- Crime rate (Chen et al., 2015)
- Dating prospects (Guéguen, 2013)



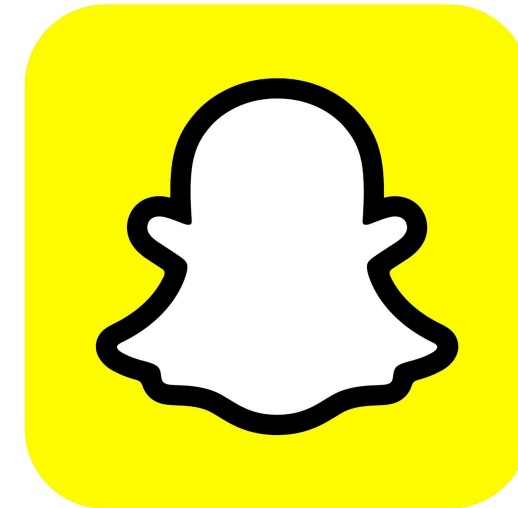
Research Questions

RQ1 MODELING: Does weather **improve** sentiment detection?

RQ2 ANALYSIS: How does weather **impact** sentiment?

Snapchat data

- 8M **Public Snapchat Stories** from 2020
 - 3K annotated
- **Textual feature:** the caption
- **Contextual features:**
location, time, and weather
 - Current weather
 - Historical weather



WEATHER DATA



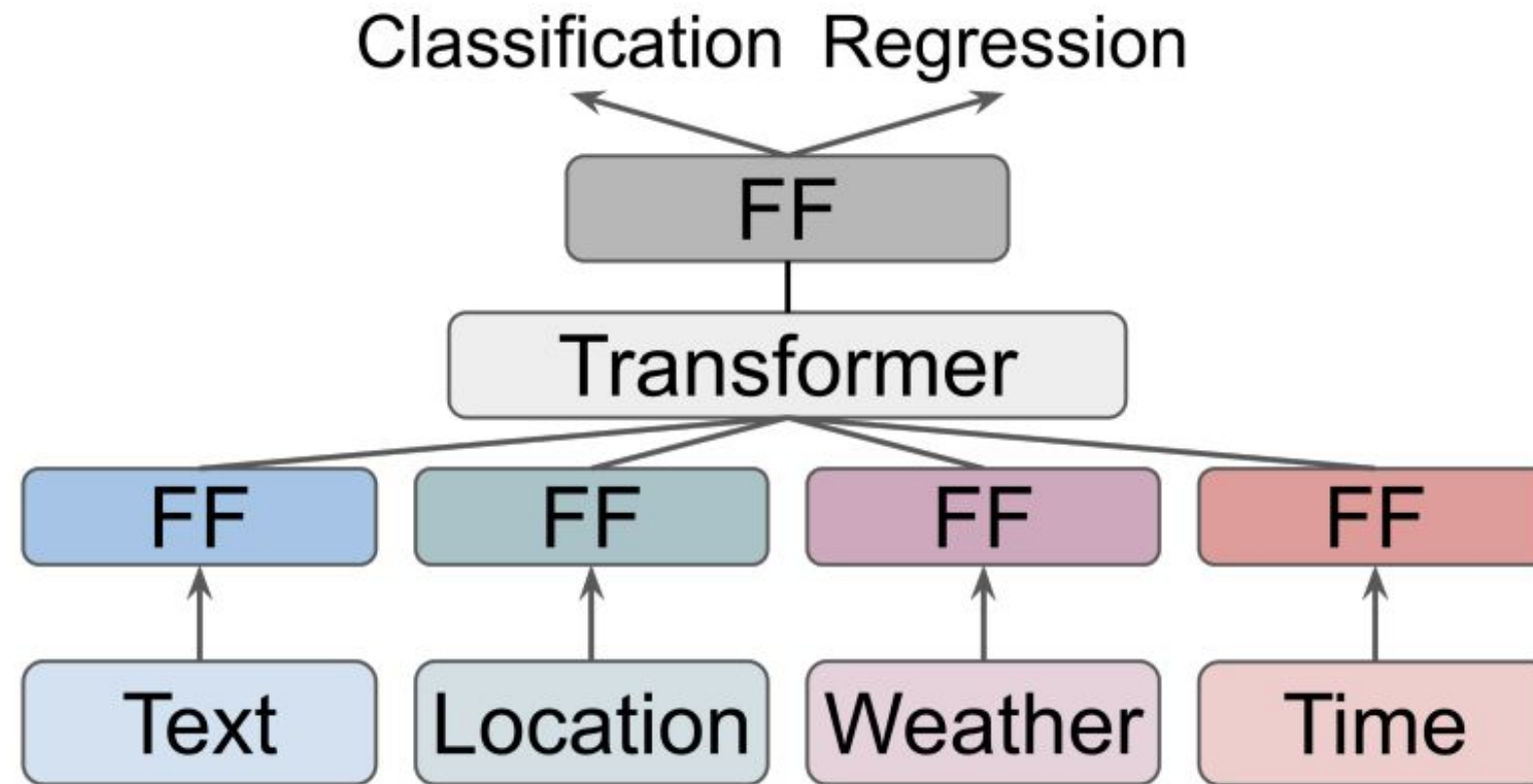
Historically averaged weather prior to the day of the Snap



RQ1 MODELING

Does weather **improve** sentiment detection?

RQ1: Model



RQ1: Result

	Score	%
LANGUAGE ONLY		
RoBERTa-base	65.07	
Snap-RoBERTa (SR)	74.23	
LANGUAGE+CONTEXT		
		<i>(from SR)</i>
SR+Weather+Location+Time	76.64	3.2%*

$\frac{1}{2}$ F1-Macro + $\frac{1}{2}$ Pearson
Classification *Regression*

* *significant improvement (P<0.05)*

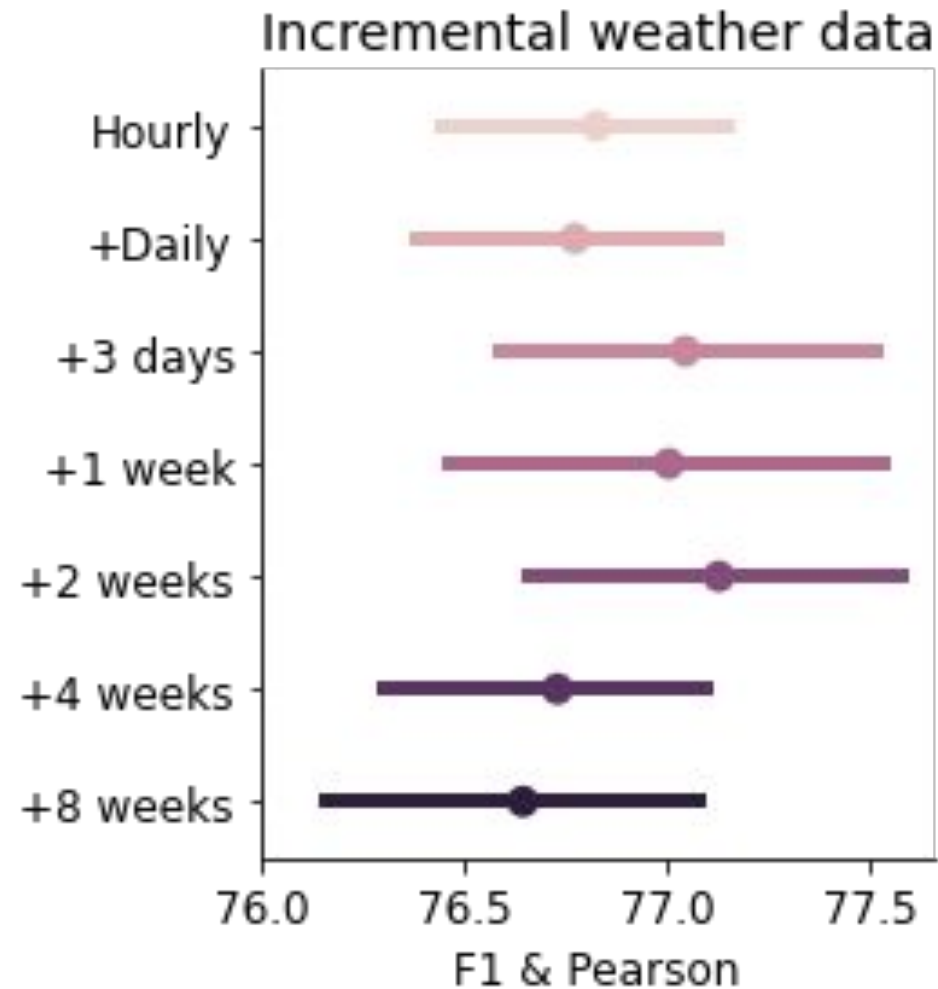
RQ1: Result

	Score	%
LANGUAGE ONLY		
RoBERTa-base	65.07	
Snap-RoBERTa (SR)	74.23	
LANGUAGE+CONTEXT <i>(from SR)</i>		
SR+Weather+Location+Time	76.64	3.2%*
CONTEXT ABLATION <i>(from SR)</i>		
SR+Weather	76.46	3.0%*
SR+Time	75.85	2.2%*
SR+Location	75.77	2.1%*

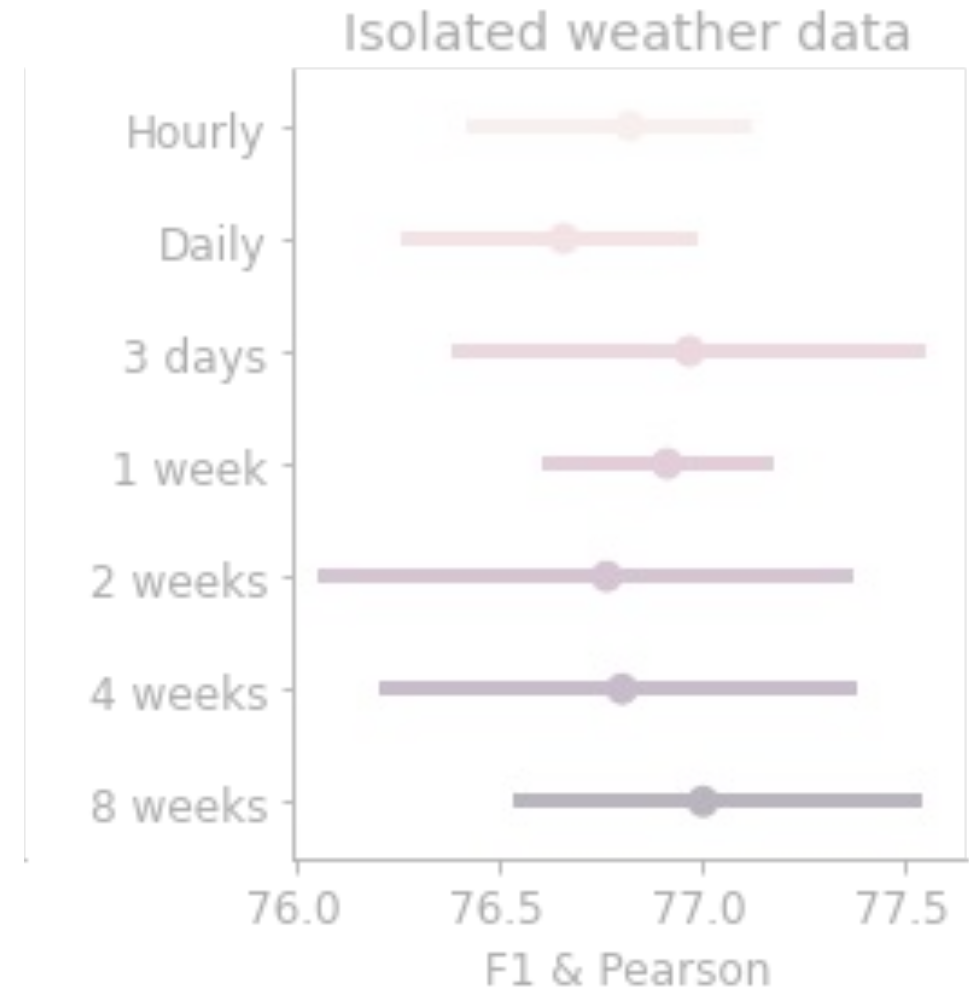
$\frac{1}{2}$ F1-Macro + $\frac{1}{2}$ Pearson
Classification *Regression*

* *significant improvement (P<0.05)*

Weather ablation

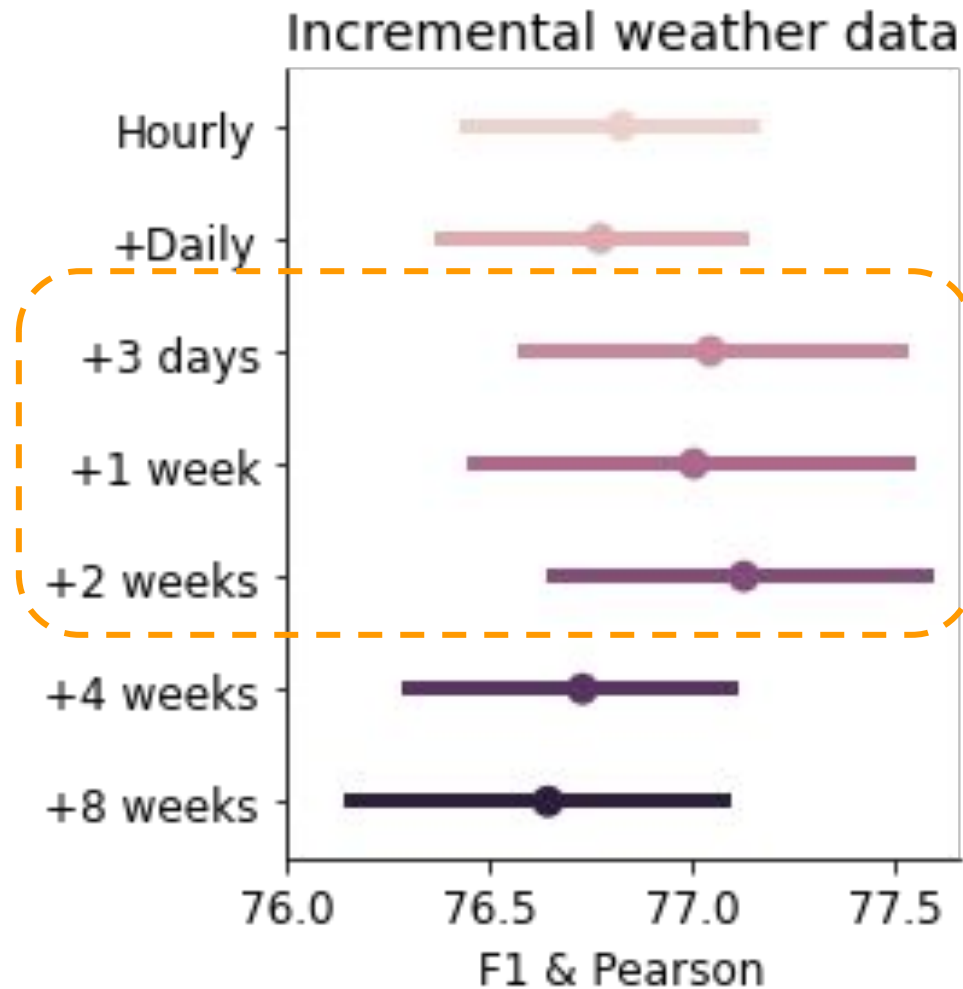


How much historical weather info is needed for good performance?

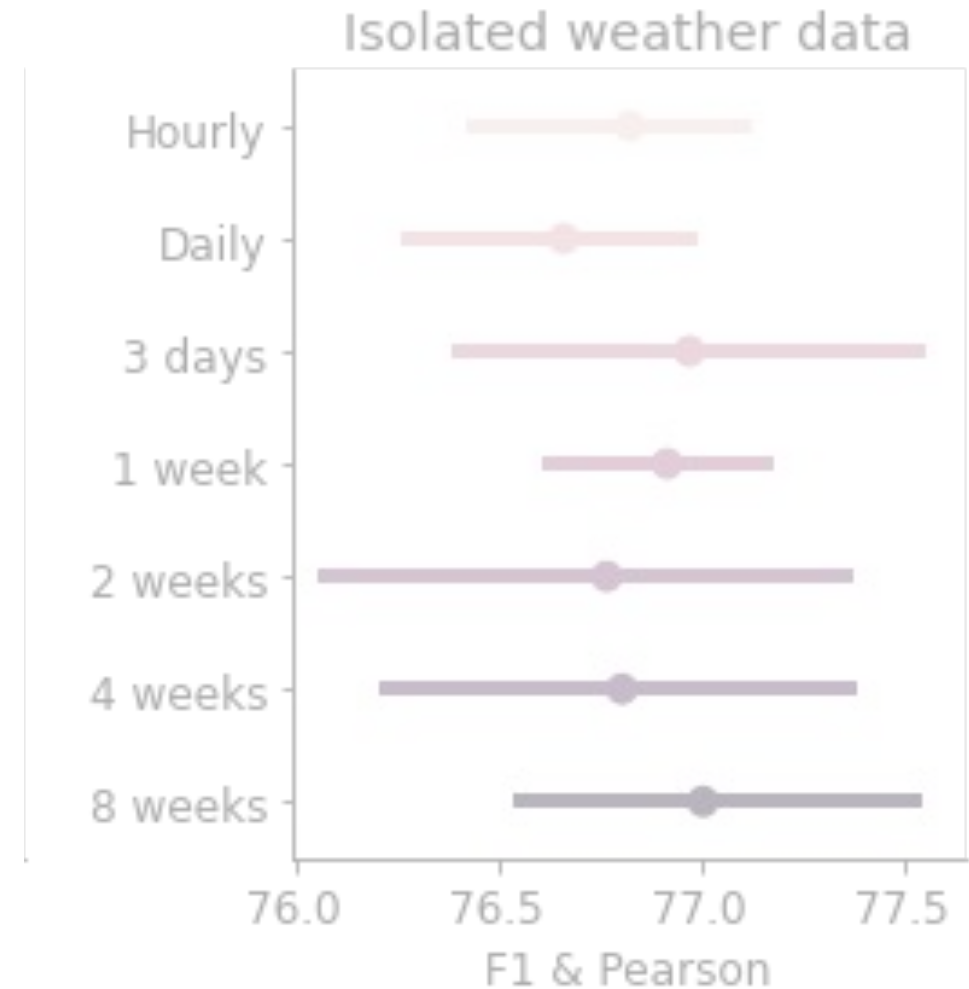


Weather from which timeframe is the most important?

Weather ablation

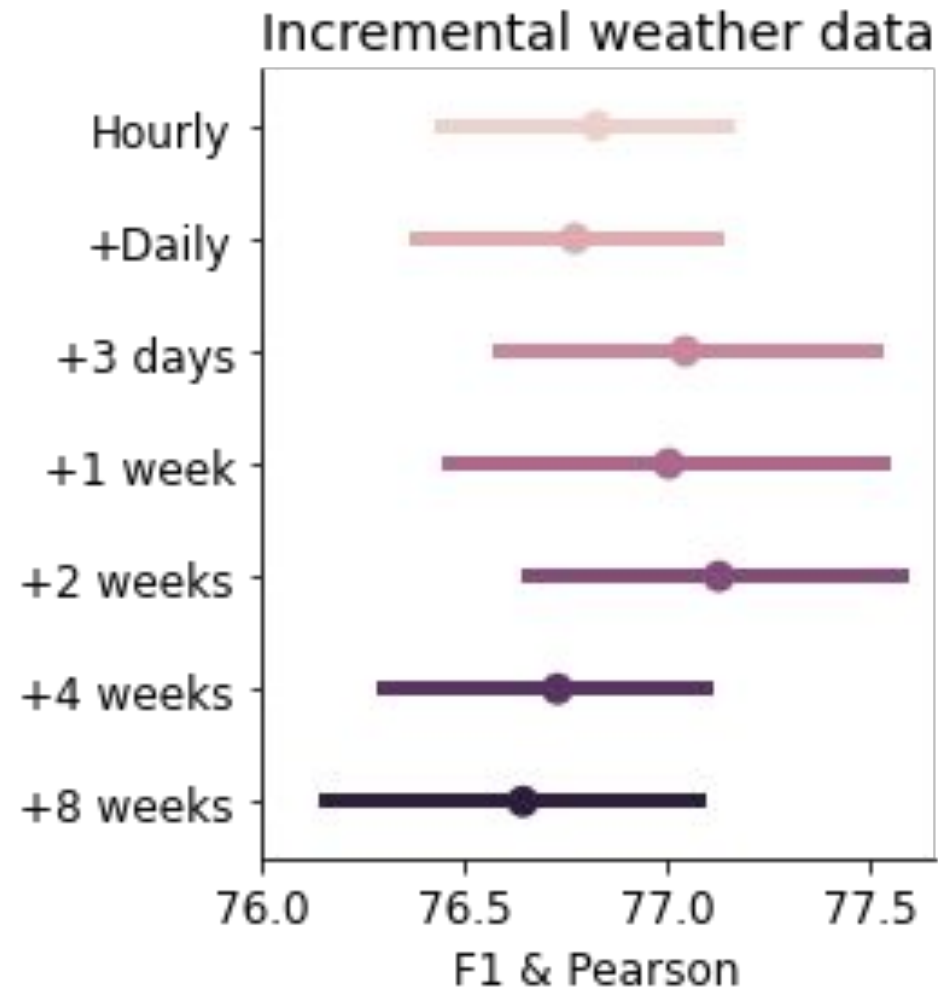


How much historical weather info is needed for good performance?

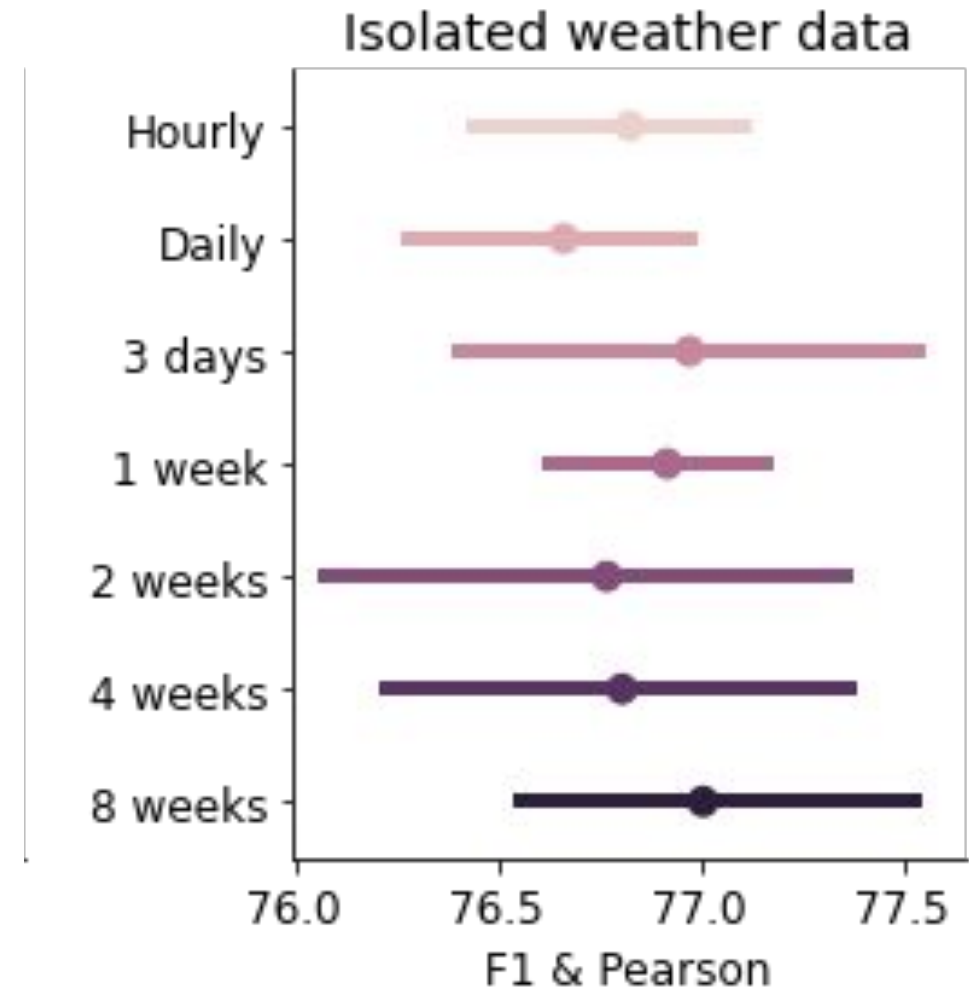


Weather from which timeframe is the most important?

Weather ablation

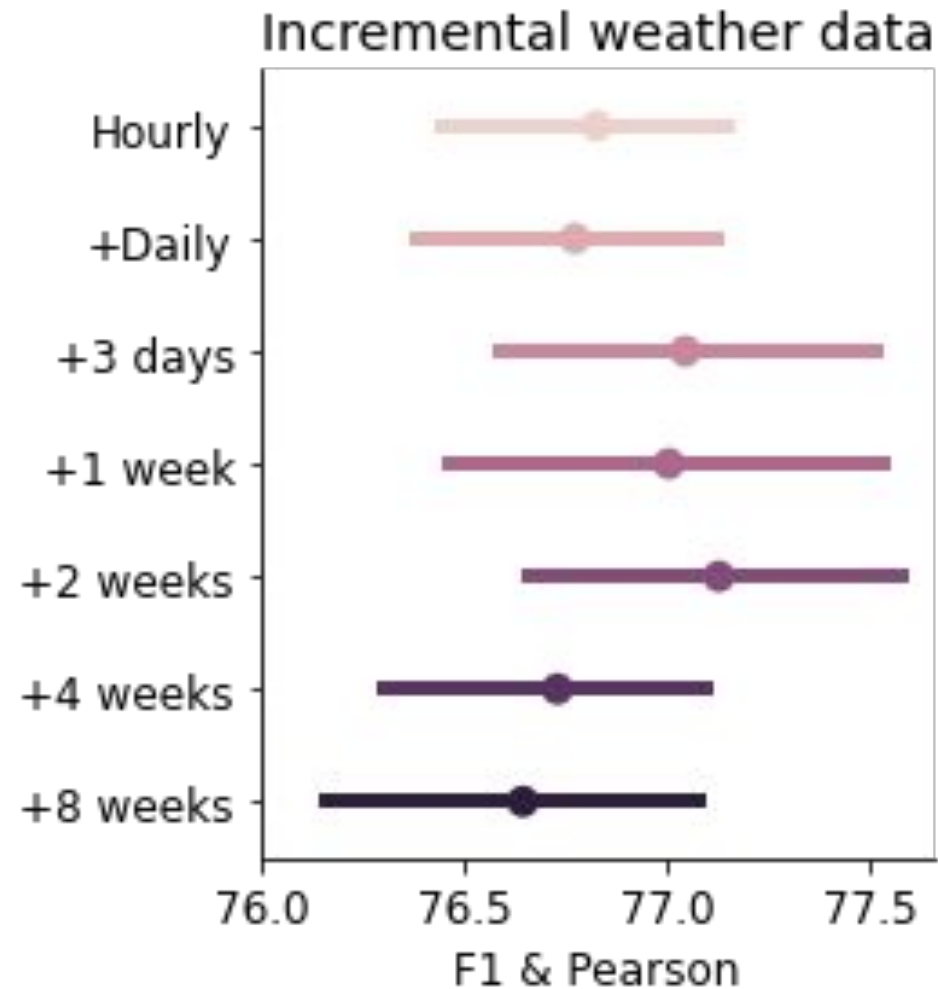


How much historical weather info is needed for good performance?

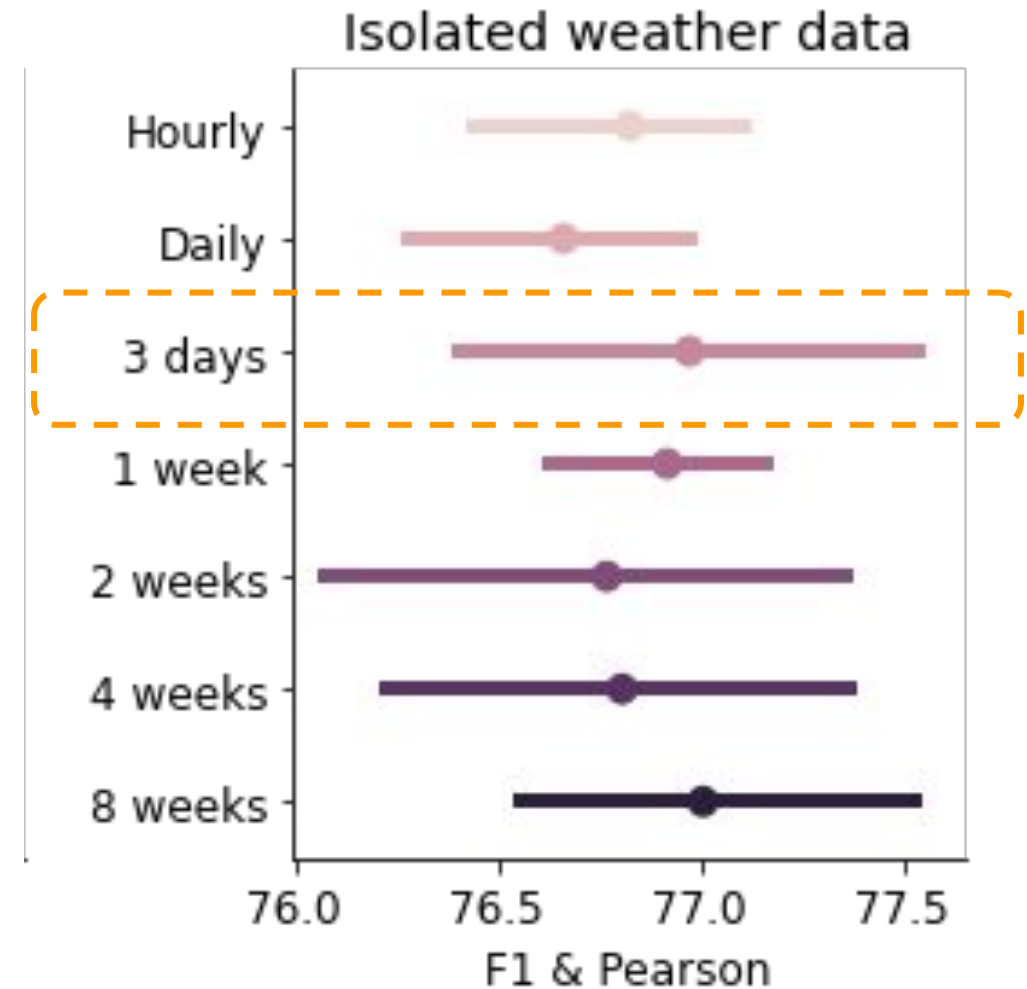


Weather from which timeframe is the most important?

Weather ablation



How much historical weather info is needed for good performance?



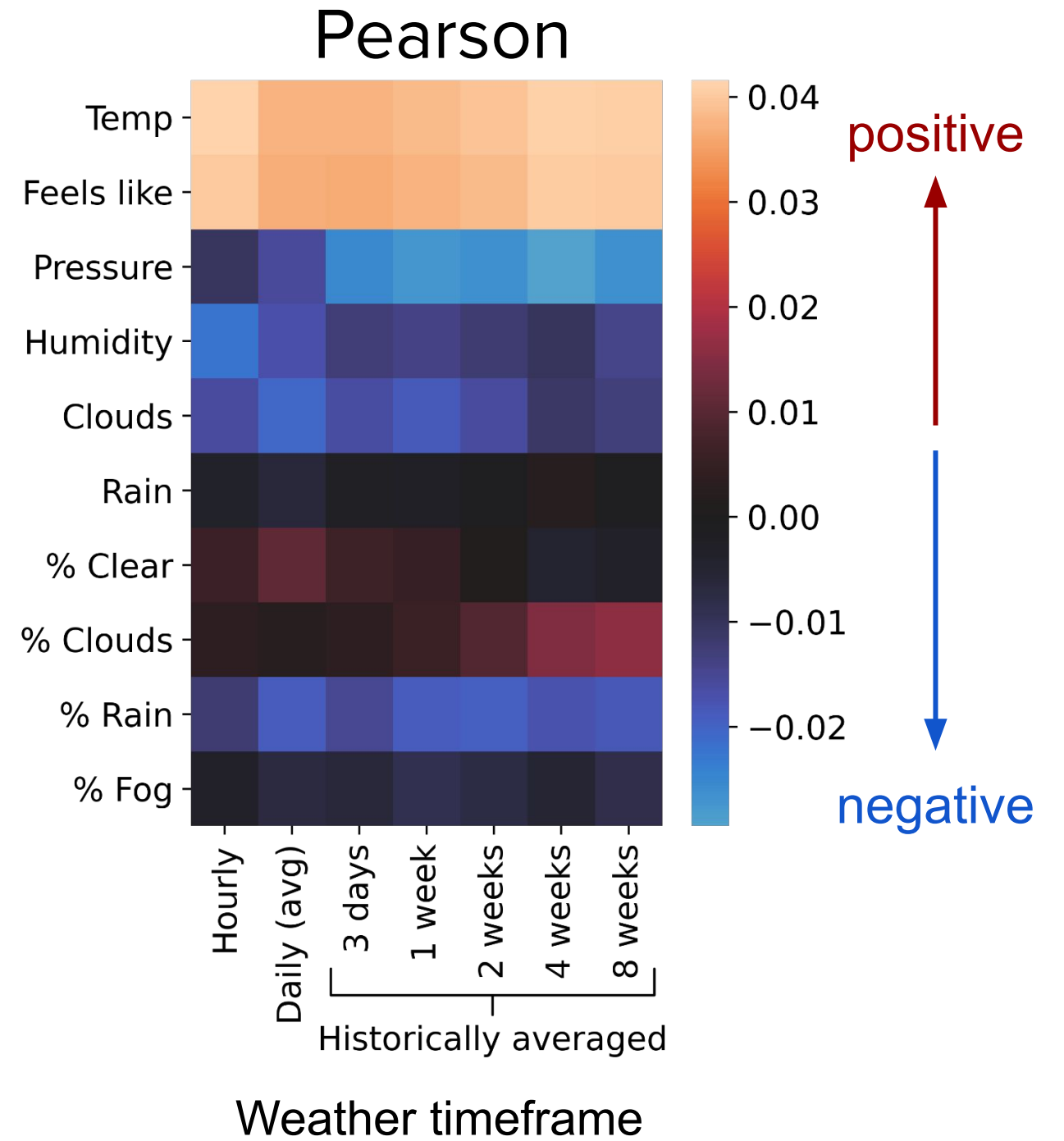
Weather from which timeframe is the most important?



RQ2 Analysis

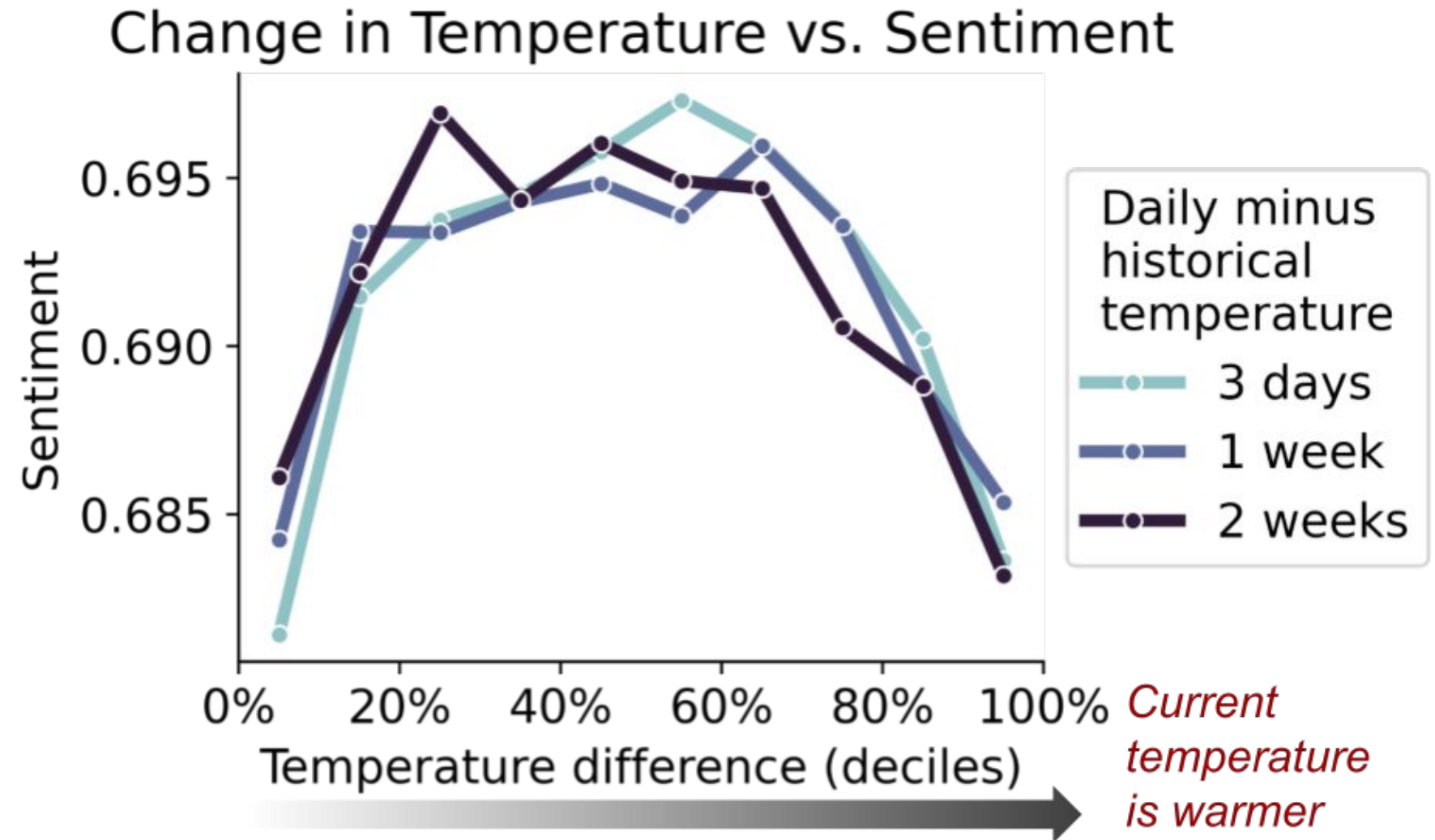
How does weather **impact** sentiment?

RQ2: Weather-induced sentiment



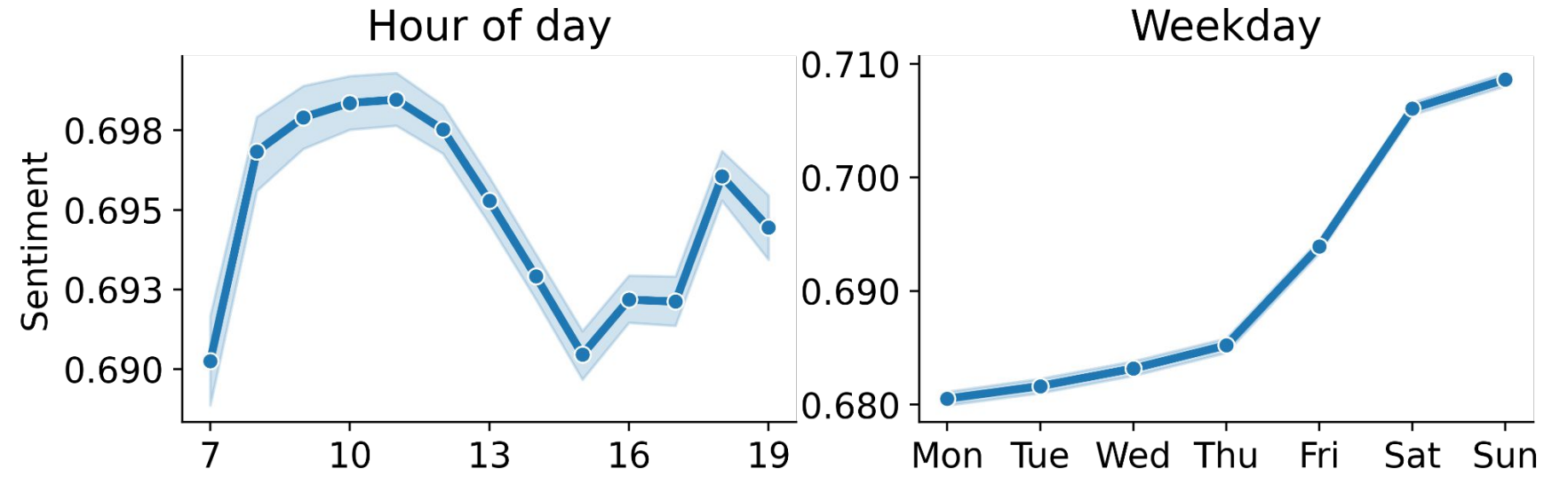
Weather-induced sentiment

Sentiment declines both when the current temperature is too **hot** and too **cold** compared to previously.



Time and sentiment

Time and Temperature vs. Sentiment



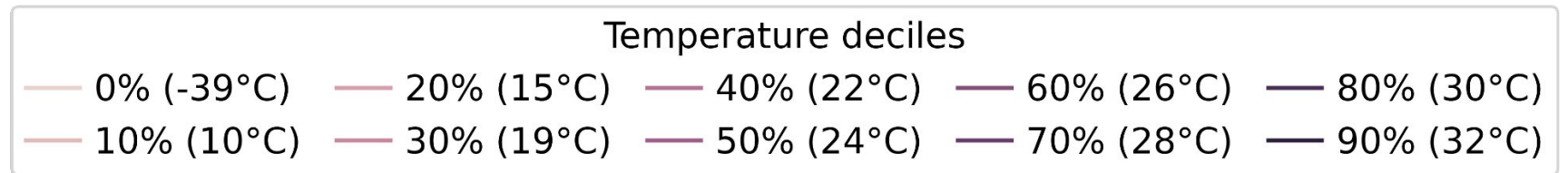
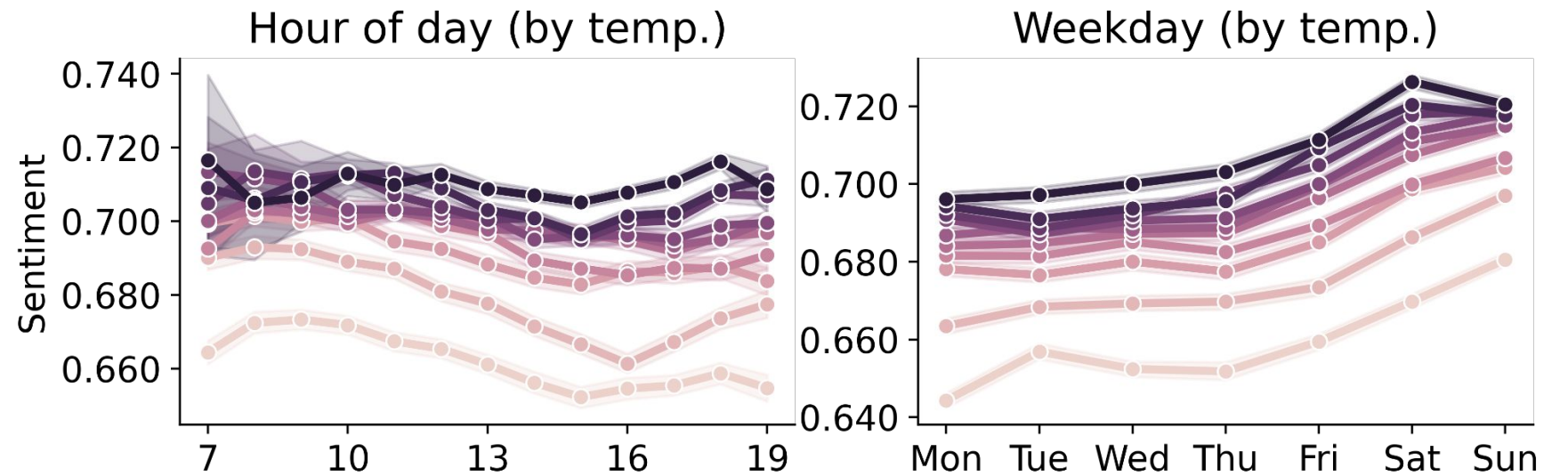
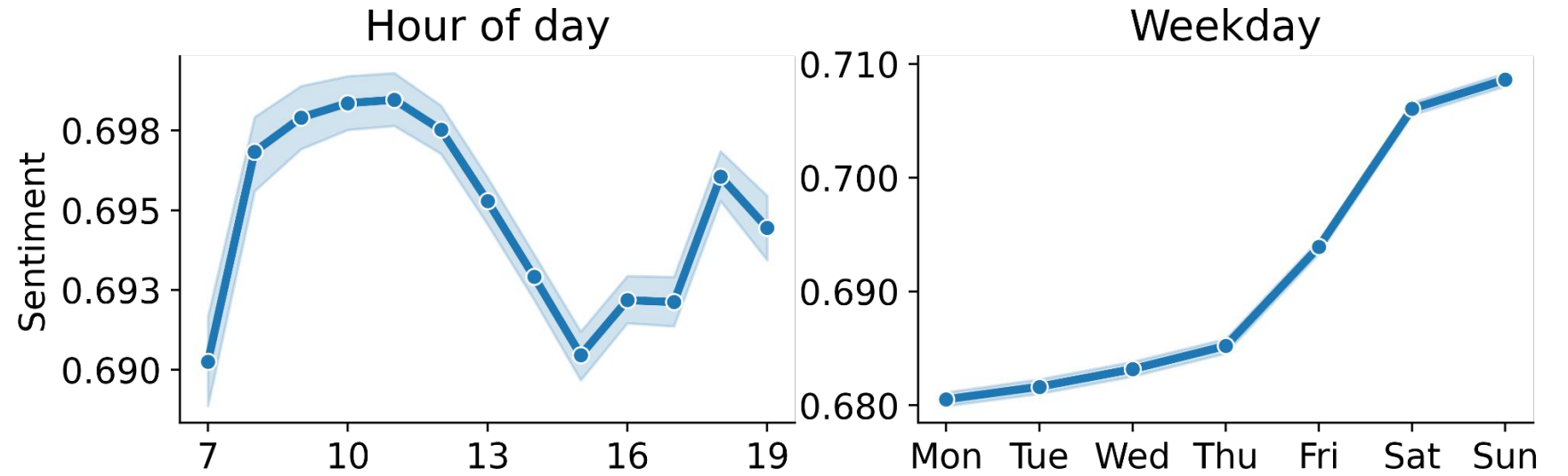
**AFTERNOON
SLUMP**



Time and sentiment

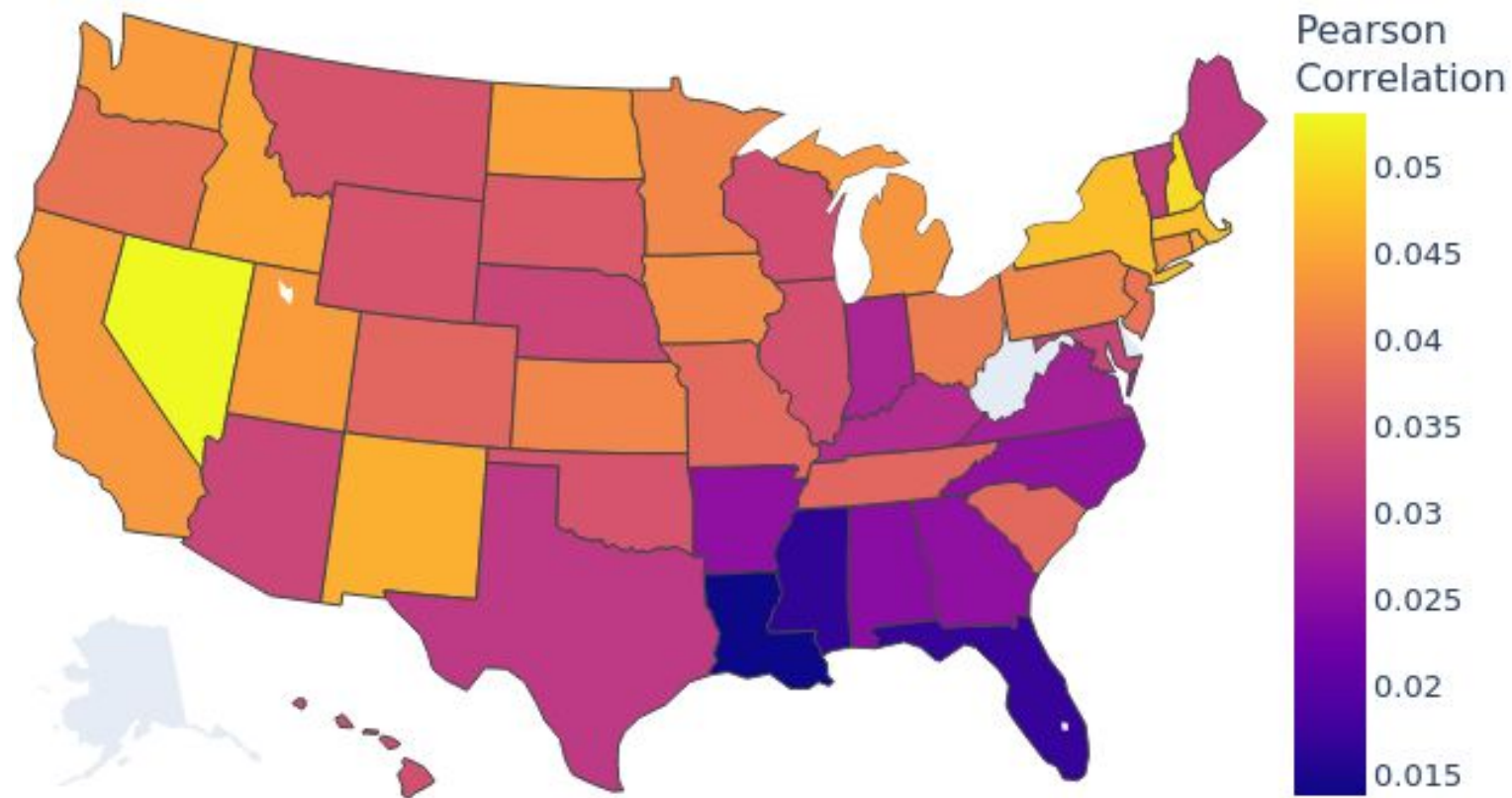
Sentiment is **consistently higher** when the weather is **warmer** controlling for time

Time and Temperature vs. Sentiment



RQ2: Location and sentiment

Pearson corr. of temperature and sentiment by state



LA has the **biggest** decline in expressed sentiment due to temperature drops and rainfall

Conclusion

RQ1 MODELING:

- Contextual factors **improve** language modeling
- **Weather** is the most important contextual factor

RQ2 ANALYSIS:

- Weather is significantly **correlated** with expressed sentiment
- Mood sensitivity to changes in weather depends on **location** but does not vary much with **time**

Implications

- Large-scale empirical proof of weather's impact on expressed sentiment
- Weather's effects on expressed sentiment appears to be **implicit** rather than explicit



Collaborators

- Julie Jiang
- Maarten W. Bos
- Yozen Liu
- Neil Shah
- Leonardo Neves
- Nils Murrugarra-Llerena
- Vitor Silva
- Luis Marujo
- Pradeep Karuturi
- William Brendel
- German Kruszewski
- Francesco Ronzano
- Horacio Saggion
- Luis Espinosa-Anke
- Jose Camacho
- Daniel Loureiro



Snap research at KDD

[CARL-G: Clustering-Accelerated Representation Learning on Graphs](#)

*William Shiao, Uday Saini, **Yozen Liu**, **Tong Zhao**, **Neil Shah**, Evangelos Papalexakis*

We propose a new framework for graph self-supervised learning by adapting clustering validation indices as loss functions, with over 79x training speedup and no performance degradation.

[Semi-supervised Graph Imbalanced Regression](#)

*Gang Liu, **Tong Zhao**, Eric Inae, Tengfei Luo, Meng Jiang*

We propose a semi-supervised framework for graph regression tasks, which uses pseudo-labeling and latent space augmentation to achieve better data balance and reduce model bias, with promising results in 7 benchmarks.

[Sketch-based Anomaly Detection in Streaming Graphs](#)

*Siddharth Bhatia, Mohit Wadhwa, Kenji Kawaguchi, **Neil Shah**, Philip Yu, Bryan Hooi*

We propose a first-of-its-kind constant-time and constant-space approach for detecting graph anomalies in the streaming setting using higher-order sketching.

[Balancing Approach for Causal Inference at Scale](#)

Sicheng Lin, Meng Xu, Xi Zhang, Shih-Kang Chao, Ying-Kai Huang, Xiaolin Shi

We present two scalable algorithms for balancing approaches to solve causal inference problems at scale of 10 million units, which are deployed in an end-to-end system at Snap and significantly reduce both bias and variance in causal effect estimation.