Unsupervised Extractive Summarization of Emotion Triggers

Tiberiu Sosea[1], Hongli Zhan[2], Junyi Jessy Li[2], and Cornelia Caragea[1]

[1]University of Illinois Chicago,[2]The University of Texas at Austin

Contributions

- We introduce CovidET-EXT, a manually annotated benchmark dataset for the task of emotion detection and extractive trigger summarization in a Covid-19 Reddit posts.
- We propose Emotion-Aware PageRank, a variation of PageRank that combines a language understanding module and external emotion knowledge to generate emotion-specific extractive summaries.
- We carry out a comprehensive set of experiments using numerous baselines to evaluate the performance on CovidET-EXT and show that our proposed EAP significantly outperforms strong baselines.

Emotion Detection and Extractive Trigger Summarization

Emotion Triggers in a Reddit Post

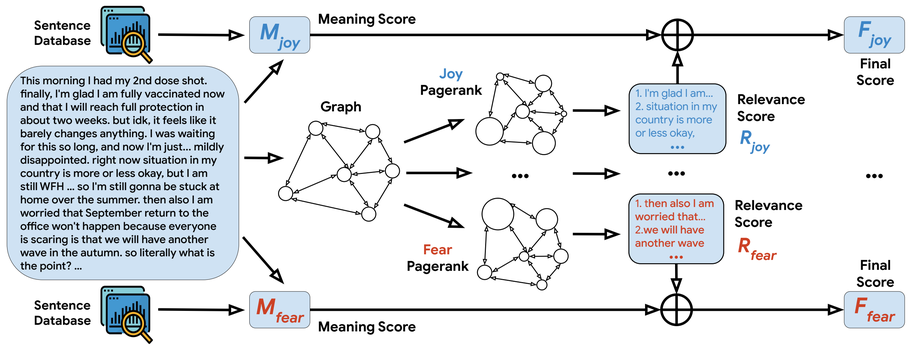
1. It finally happened.
2. Took an older relative for her first Pfizer dose.
3. Not that many people showed up so all accompanying family members were offered the shot and give papers for a second dose.
4. I wasnt due to get my shot for the next couple of months and have had some scares, many for which Ive gotten support from you lovely awesome people.
5. I now have a whole different perspective on my governments organization.
6. They have a good oiled vaccination machine. **Trust**
7. They just need more doses.
8. Best part is I was told I could come back for my second dose whenever my relative was scheduled to get Hers. **Joy**
9. I have a lil arm pain.
10. But its the slightest arm pain when moving it past certain angles.
11. Ive noticed that a lil blood drop shows on my vaccination spot (took a shower half an hour later). **Fear**
12. Is this normal?

Figure: An example post from CovidET-EXT annotated with emotion triggers. The highlighted sentences represent triggers of the tagged emotions.

Annotation and Characteristics

- Given a post from CovidET annotated with an emotion e, we ask annotators to highlight sentences in the post that best describe the trigger for e.
- We recruit both undergraduate students (in a Linguistics department) as well as pre-qualified crowd workers (from the Amazon Mechanical Turk) for this task.
- We monitor the annotation quality and work with the annotators during the full process. Similar to CovidET, the test set is annotated by undergraduate students.
- We follow the benchmark setup in [Zhan et al., 2022] with $1,200$ examples for training, 285 examples for validation, and 398 examples for testing.
- We validate the annotation quality of extractive summaries of emotion triggers thoroughly in CovidET-EXT through inspections from third-party validators on the Amazon Mechanical Turk crowdsourcing platform.
- The average Fleiss' kappa is 0.89 across all the emotions in CovidET-EXT, indicating substantial agreement among our annotators.
- We emphasize that the focus of this work is the **unsupervised setup**. However, we hope that CovidET-EXT can spur further research into both supervised and unsupervised methods, hence we maintain the splits in [Zhan et al., 2022].

Unsupervised Emotion-Aware PageRank (EAP)



Algorithm Description

- Run a word-level biased pagerank for each emotion and compute relevance scores for every word w_i:

$$\mathcal{R}_e(w_i) = \lambda \sum_{k=1}^{|V|} \beta(w_k, w_i)\mathcal{R}_e(w_k) + (1-\lambda)\frac{i_e(w_i)}{N} \quad (1)$$

- Use the NRC Emotion Intensity lexicon to bias the pagerank to assign higher relevance scores to emotion-intensive words.
- Score a sentence s_i by aggregating the Pagerank scores of all the words in the sentence:

$$R_e(s_i) = \frac{\sum_{w_j \in s_i} R_e(w_j)}{|s_i|} \quad (2)$$

- Use a language model to generate embeddings for each sentence in a post and score each sentence based on how semantically close it is to other relevant input sentences:

$$M_e(s_i) = \frac{\sum_{s \in \mathcal{T}} sim(\mathbf{s_i}, \mathbf{s}) * \mathcal{R}_e(s)}{|\mathcal{T}|} \quad (3)$$

- Combine the scores obtained from the biased pagerank and those obtained using similarity into a single score:

$$\mathcal{F}_e(s_i) = \mathcal{R}_e(s_i) * M_e(s_i) \quad (4)$$

- Use the $\mathcal{F}_e(s_i)$ score to generate summaries using high-score input sentences.

Main Results

	ANGER		DISGUST		FEAR		JOY		SADNESS		TRUST		ANTICIPATION		AVG	
	R-2	R-L	R-2	R-L	R-2	R-L	R-2	R-L	R-2	R-L	R-2	R-L	R-2	R-L	R-2	R-L
1-SENT	0.174	0.240	0.095	0.170	0.202	0.256	0.119	0.179	0.110	0.177	0.189	0.236	0.160	0.220	0.149	0.211
3-SENT	0.301	0.315	0.196	0.253	0.322	0.343	0.273	0.310	0.239	0.292	0.248	0.279	0.263	0.307	0.258	0.288
PacSum	0.308	0.314	0.210	0.218	0.327	0.331	0.276	0.282	0.287	0.304	0.225	0.234	0.283	0.295	0.273	0.282
PreSumm	0.306	0.312	0.219	0.221	0.332	0.335	0.268	0.274	0.295	0.317	0.222	0.227	0.284	0.291	0.275	0.282
TextRank	0.296	0.301	0.236	0.235	0.319	0.326	0.272	0.276	0.286	0.306	0.225	0.231	0.218	0.221	0.264	0.270
EmoLex	0.213	0.260	0.218	0.256	0.309	0.341	0.218	0.252	0.301	0.331	0.176	0.203	0.207	0.242	0.234	0.269
EmoIntensity	0.307	0.322	0.269	0.281	0.342	0.355	0.222	0.235	0.329	0.341	0.227	0.242	0.295	0.310	0.284	0.298
BERT-GoEmo	0.247	0.264	0.232	0.237	0.296	0.312	0.221	0.247	0.314	0.321	0.201	0.204	0.247	0.225	0.253	0.258
EAP	**0.324**†	**0.348**†	**0.285**†	**0.296**†	**0.364**†	**0.373**†	**0.285**†	**0.319**†	**0.348**†	**0.354**†	**0.258**†	**0.291**†	**0.319**†	**0.324**†	**0.309**†	**0.325**†

We explore:

- Emotion-agnostic baselines such as PacSum [Zheng and Lapata, 2019], PreSum [Liu and Lapata, 2019] and TextRank [Mihalcea and Tarau, 2004].
- Emotion-specific baselines such as using the EmoLex [Mohammad and Turney, 2013] and EmoIntensity [Mohammad, 2018] lexicon to generate summaries, as well as using an emotion detection model trained on Reddit text to extract sentences.
- **EAP consistently yields the highest results both in terms of Rouge-2 and Rouge-L.**

Ablation Study

	ANGER		DISGUST		FEAR		JOY		SADNESS		TRUST		ANTICIPATION		AVG	
	R-2	R-L	R-2	R-L	R-2	R-L	R-2	R-L	R-2	R-L	R-2	R-L	R-2	R-L	R-2	R-L
EAP	**0.324**	**0.348**	**0.285**	**0.296**	**0.364**	**0.373**	**0.285**	**0.268**	**0.348**	**0.354**	**0.239**	**0.264**	**0.319**	**0.324**	**0.309**	**0.318**
-int	0.317	0.336	0.274	0.282	0.353	0.362	0.276	0.261	0.339	0.347	0.231	0.252	0.312	0.317	0.300	0.308
-sim	0.314	0.332	0.277	0.284	0.351	0.360	0.272	0.260	0.340	0.342	0.232	0.254	0.311	0.31	0.299	0.306
-int -sim	0.300	0.316	0.263	0.275	0.341	0.353	0.261	0.253	0.325	0.339	0.224	0.247	0.308	0.309	0.28	0.298

- We perform a thorough ablation study to tease apart and analyze the components lead to the success of Emotion-aware Pagerank.
- We remove the emotion intensity component of our method $\longrightarrow$ *-int*.
- We remove the language understanding module $\longrightarrow$ *-sim*.
- Removing emotion intensity leads to a performance degradation, as well as removing the language understanding module, emphasizing that both similarity and intensity are core components of EAP and both consistently contribute to the success of EAP.

Anecdotal Evidence

[this morning$^{0.212}$ I$^{0.145}$ had my 2nd dose$^{0.028}$ shot$^{0.075}$.]$^{0.074}$ [finally$^{0.054}$, I$^{0.145}$ 'm glad$^{0.865}$ I$^{0.145}$ am fully$^{0.142}$ vaccinated$^{0.215}$ now and that]$^{0.463}$ [I$^{0.145}$ will reach$^{0.206}$ full protection$^{0.531}$ in about two weeks$^{0.105}$, but idk]$^{0.217}$ [it feels$^{0.153}$ like$^{0.363}$ it barely$^{0.06}$ changes$^{0.095}$ anything]$^{0.187}$ [I was waiting$^{0.121}$ for this so long$^{0.143}$, and now I$^{0.145}$ 'm just... mildly$^{0.185}$ disappointed$^{0.003}$.]$^{0.093}$ [right$^{0.174}$ now situation$^{0.118}$ in my country$^{0.121}$ is more$^{0.147}$ or less$^{0.105}$ okay$^{0.163}$]$^{0.231}$, [but I$^{0.145}$ am still WFH$^{0.164}$]$^{0.251}$ [and my office$^{0.121}$ sent us official$^{0.142}$ info$^{0.111}$ that]$^{0.142}$ [we$^{0.131}$ can start$^{0.152}$ coming$^{0.123}$ back to the office$^{0.121}$ only since September$^{0.118}$]$^{0.246}$ [so I$^{0.145}$ 'm still gonna be stuck$^{0.054}$ at home$^{0.104}$ over the summer$^{0.131}$]$^{0.076}$. [then also I$^{0.145}$ am worried$^{0.074}$ that September$^{0.118}$ return$^{0.142}$ to the office$^{0.121}$ won't happen$^{0.103}$]$^{0.054}$ [because everyone$^{0.131}$ is scaring$^{0.052}$ is that we will have another$^{0.117}$ wave$^{0.186}$ in the autumn$^{0.128}$]$^{0.095}$. [so literally$^{0.118}$ what is the point$^{0.142}$]$^{0.131}$

[this morning$^{0.196}$ I$^{0.132}$ had my 2nd dose$^{0.073}$ shot$^{0.113}$.]$^{0.136}$ [finally$^{0.051}$, I$^{0.132}$ 'm glad$^{0.004}$ I$^{0.132}$ am fully$^{0.094}$ vaccinated$^{0.108}$ now and that]$^{0.055}$ [I$^{0.132}$ will reach$^{0.141}$ full protection$^{0.076}$ in about two weeks$^{0.103}$. but idk,]$^{0.134}$ [it feels$^{0.125}$ like$^{0.100}$ it barely$^{0.315}$ changes$^{0.169}$ anything$^{0.147}$]$^{0.148}$ [I$^{0.132}$ was waiting$^{0.196}$ for this so long, and now I$^{0.132}$ 'm just... mildly$^{0.075}$ disappointed$^{0.241}$]$^{0.185}$ [right$^{0.105}$ now situation$^{0.143}$ in my country$^{0.105}$ is more$^{0.122}$ or less$^{0.162}$ okay$^{0.104}$]$^{0.142}$, [but I^{132} am still WFH$^{0.151}$]$^{0.152}$ [and my office$^{0.117}$ sent us official$^{0.123}$ info$^{0.113}$ that]$^{0.105}$ [we can start$^{0.106}$ coming$^{0.121}$ back to the office$^{0.123}$ only since September$^{0.152}$.]$^{0.142}$ [so I$^{0.132}$ 'm still gonna be stuck$^{0.386}$ at home$^{0.164}$ over the summer$^{0.126}$]$^{0.216}$. [then also I$^{0.132}$ am worried$^{0.523}$ that September$^{0.152}$ return$^{0.107}$ to the office$^{0.111}$ won't happen$^{0.142}$]$^{0.599}$ [because everyone$^{0.116}$ is scaring$^{0.387}$ is that we will have another$^{0.175}$ wave$^{0.221}$ in the autumn$^{0.131}$]$^{0.372}$. [so literally$^{0.101}$ what is the point$^{0.121}$]$^{0.155}$

Figure: Word-level Emotion-Aware PageRank scores and sentence-level meaning scores for the joy (Upper Box) and fear (Lower Box) emotions. The term relevance score is superscripted to each word (i.e., w^{score}), while the meaning score of sentences is superscripted at the end of the sentence (i.e., $[.]^{score}$).

- We indicate for each word both its relevance for joy and for fear.
- The scores produced by our model are very relevant. *protection* has a very large value for joy of 0.531 and a very small value of 0.076 for *fear*.
- The similarity scores are also accurate. *glad I am fully vaccinated* has a score for joy of 0.463, 9 times as large of the score of the same sentence for fear.

Conclusion and Future Work

- We introduce CovidET-EXT, a new benchmark dataset composed of $1,883$ Reddit posts annotated for the task emotion detection and extractive trigger summarization in the context of the COVID-19 pandemic.
- Emotion-Aware Pagerank approach yields strong results on our datasets, consistently outperforming prior work in an unsupervised learning context.
- In the future, we plan to study abstractive trigger summarization from a zero-shot perspective to remove the requirement for labeled data.

Acknowledgements: This research is supported by NSF.

https://github.com/tsosea2/CovidET-EXT