Bridging the gap between supervised classification & unsupervised topic modelling for social-media assisted crisis management

McGill | Environment and Climate Change Canada

Big Picture: Finetuned Tweet Embeddings (FTE)

Motivation: While crisis tweet classification predict useful labels (that may not all generalize to new crises), topic models discover event specific topics (that may not all be useful). Can we bridge this gap?
Method: Cluster BERT embeddings learned from supervised tweet classification (FTE) into topics; then extract keywords using tf-idf and attention; measure quality of clusters/keywords with human eval.
Results: In a novel snow storm crisis event, relevant classes from the supervised training are preserved, and novel event-specific topics are discovered in an unsupervised way; automatic/human evaluation show that FTE improves over topic models and vanilla BERT embeds.
Takeaways: Clustering representations learned from supervised classification can help adapt to domains with overlapping and different latent classes. Human involvement is key to ensure model outputs are aligned with stakeholder needs, especially in crisis management.

Interpretable Topics for Crisis Managers

Based on annotator interpretations of extracted keywords, FTE clustering preserved relevant classes from supervised training, and discovered novel event-specific topics.

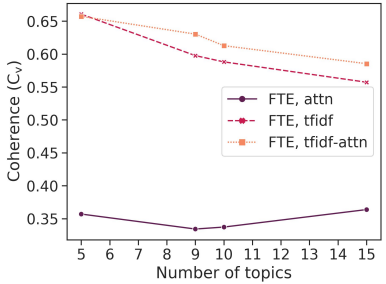
Model	Topic 1	2	3	4	5	6	7	8	9
FTE	reporting	ivyparkxadidas	outage	assistance	prayer	blowingsnow	trapped	monster	bread
	monster	mood	campus	assist	praying	alert	stranded	meteorologist	song
	snowiest	song	widening	troop	pray	advisory	hydrant	drifting	coffee
	recorded	blackswan	advisory	volunteer	wish	caution	ambulance	perspective	milk
	peak	le	reported	providing	wishing	advised	dead	stormofthecentury	feelin
	temperature	snowdoor	impassable	relief	humanity	stormsurge	garbage	mood	pin
	cloudy	perspective	remaining	aid	brave	wreckhouse	rescue	snowdrift	enjoying
	reported	ode	thousand	request	surviving	surge	permitted	climate	laugh
	equivalent	music	reporting	offering	loved	drifting	body	windy	favorite
	meteorologist	adidasxivypark	suspended	rescue	kindness	avoid	helped	snowdoor	girl
	Weather related	Unrelated information	Power outages	Donation + volunteer	Sympathy + support	Caution + advice	Trapped people		

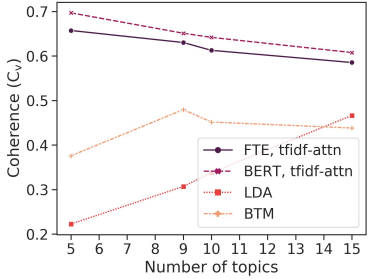
Evaluation

Traditional evaluation methods are useful in selecting hyperparameters. However, human evaluation is essential to ensure a model is valuable to crisis managers.

Automatic Evaluation

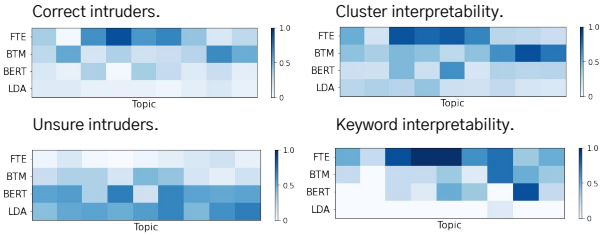
Combining tf-idf and attention for keyword extraction improved coherence in FTE. FTE performed better than traditional topic models, similar to vanilla BERT embeds.





Human Evaluation

With high agreement, annotators found FTE produces more interpretable, coherent, and useful topics in the context of crisis management. Results for vanilla BERT embeds show limitations of automatic evaluation and need for human-centered ML.



Keyword Evaluation scores averaged across topics, number of topics with average scores greater than 0.5.

Score	Average Score BTM	Average Score FTE	Topic Count BTM	Topic Count FTE	Fleiss' κ BTM	Fleiss' κ FTE
Interpretability	31.94	**65.28**	1	**5**	15.01	**17.97**
Usefulness	27.78	**59.72**	1	**5**	12.36	**21.55**

Cluster Evaluation scores averaged across top- ics, number of topics with average scores greater than 0.5

Score	Average Score BTM	Average Score FTE	Topic Count BTM	Topic Count FTE	Fleiss' κ BTM	Fleiss' κ FTE
Interpretability	50.28	**51.53**	3	**4**	11.05	**23.45**
Usefulness	45.46	**46.11**	3	**5**	**21.82**	21.60
Correct Intruders	35.28	**44.17**	2	**4**	25.78	**31.50**
Unknown Intruders	26.39	**8.89**	0	0	-	-