CIENCIAS DE LA COMPUTACIÓN
UNIVERSIDAD DE CHILE

Clustering: Casos de Estudio

Hernán Sarmiento

Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020)

Unsupervised User Stance Detection on Twitter

Kareem Darwish,[1] Peter Stefanov,[2] Michaël Aupetit,[1] Preslav Nakov[1]
[1]Qatar Computing Research Institute
Hamad bin Khalifa University, Doha, Qatar
[2]Faculty of Mathematics and Informatics,
Sofia University "St. Kliment Ohridski", Sofia, Bulgaria
{kdarwish, maupetit, pnakov}@hbku.edu.qa, p.stefanov@hotmail.com

Abstract

We present a highly effective unsupervised framework for detecting the stance of prolific Twitter users with respect to controversial topics. In particular, we use dimensionality reduction to project users onto a low-dimensional space, followed by clustering, which allows us to find core users that are representative of the different stances. Our framework has three major advantages over pre-existing methods, which are based on supervised or semi-supervised classification. First, we do not require any prior labeling of users: instead, we create clusters, which are much easier to label manually afterwards, e.g., in a matter of seconds or minutes instead of hours. Second, there is no need for domain- or topic-level knowledge either to specify the relevant stances (labels) or to conduct the actual labeling. Third, our framework is robust in the face of data skewness, e.g., when some users or some stances have

In either case, some form of initial manual labeling of tens or hundreds of users is performed, followed by user-level supervised classification or label propagation based on the user accounts and the tweets that they retweet and/or the hashtags that they use (Magdy et al. 2016; Pennacchiotti and Popescu 2011a; Wong et al. 2013).

Retweets and hashtags can enable such classification as they capture homophily and social influence (DellaPosta, Shi, and Macy 2015; Magdy et al. 2016), both of which are phenomena that are readily apparent in social media. With homophily, similarly minded users are inclined to create social networks, and members of such networks exert social influence on one another, leading to more homogeneity within the groups. Thus, members of homophilous groups tend to share similar stances on various topics (Garimella

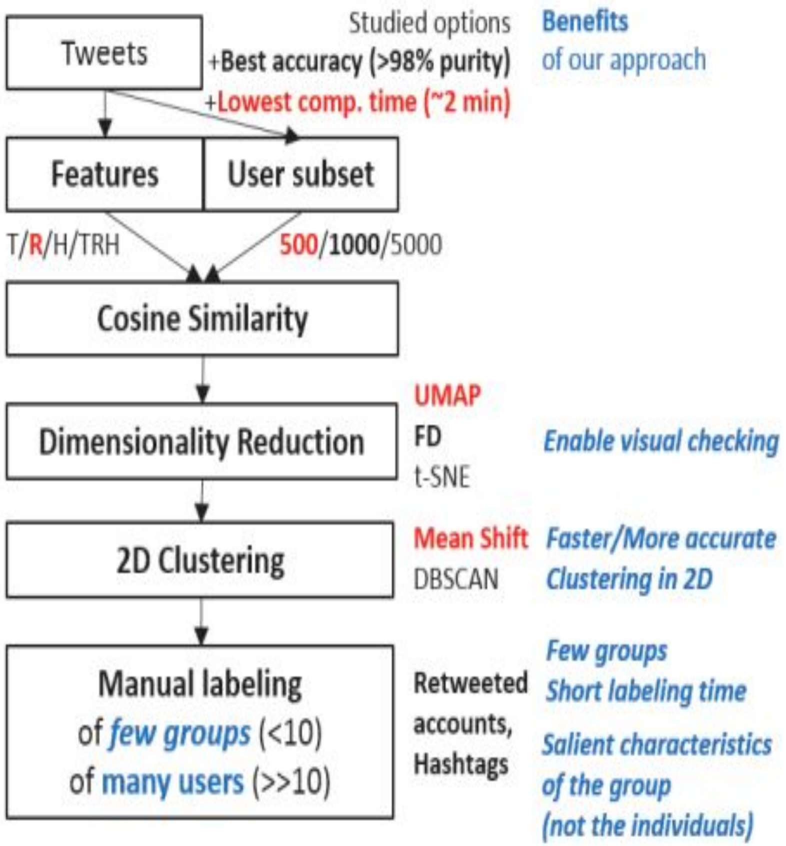
Motivación

- Usuarios tienden a alinearse y compartir ideas con otros que poseen pensamiento similares (homofilia -> filter bubble)
- Estos fenómenos gatillan la inclusión de comunidades polarizadas, las cuales existen (permanentemente) pero se ven mayormente dislumbradas durante eventos controversiales, sobre todo en redes sociales
- ¿Cómo se detecta la polarización en estas plataformas?

Motivación

- Desafíos al utilizar métodos **supervisados:**
  - Se requiere una colección inicial de usuarios etiquetados (posturas)
  - Lo anterior requiere experticia y conocimiento del dominio
  - Se invierte tiempo y es costoso (1-2 horas / 50-100 usuarios)
  - La distribución de usuarios podría no ser uniforme, generando sesgos de acuerdo al contenido que publican

Propuesta



Tweets
Studied options
+Best accuracy (>98% purity)
+Lowest comp. time (~2 min)
Benefits
of our approach
Features
User subset
T/R/H/TRH
500/1000/5000
Cosine Similarity
Dimensionality Reduction
UMAP
FD
t-SNE
Enable visual checking
2D Clustering
Mean Shift
DBSCAN
Faster/More accurate
Clustering in 2D
Manual labeling
of few groups (<10)
of many users (>>10)
Retweeted
accounts,
Hashtags
Few groups
Short labeling time
Salient characteristics
of the group
(not the individuals)

Feature selection

- (re) tweetear similares mensajes
- Hashtags que los usuarios utililizan
- Cuentas que los usuarios retweetearon

- Se evalúan individualmente y en conjunto estas features aplicando similitude coseno sobre el vector obtenido

Dimensionality reduction

- El vector de características es reducido a una dimensionalidad más pequeña luego de aplican la similitud coseno.

- Se utilizan los siguientes métodos para esta tarea:
  - Fruchterman and Reingold (FD)
  - t-distributed stochastic neighbor embedding (t-SNE)
  - Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP)

Clustering

- Utilizando el vector de características en 2 dimensiones, se aplican las siguientes técnicas de clustering (basado en densidades) individualmente:
  - DBSCAN
  - Mean shift

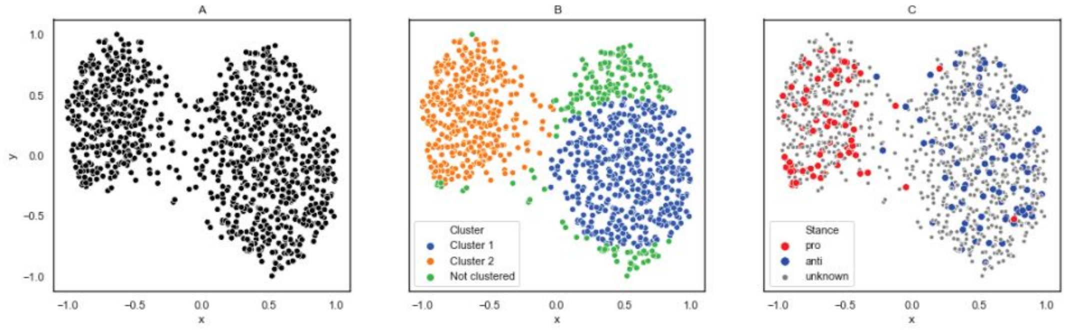


Figure 2: Successful setup: Plot (A) illustrates how user vectors get embedded by UMAP in two dimensions, Plot (B) presents the clusters Mean Shift produces for them, and Plot (C) shows the users' true labels.

Experimental setup

- Se escogen aleatoriamente usuarios para cada dataset estudio. Tamaños de 50k, 100k, 250k y 1M
- Se utilizan los 3 métodos de reducción de dimensionalidad, en su mayoría con parámetros por defecto
- Se usa DBSCAN y MeanShift con iguales parámetros en todos los experimentos

Set	# of Users	Feature(s)	Dim Reduce	Peak Detect	Avg. Purity	Avg. # of Clusters	Avg. Cluster Size	Avg. Recall
100k	500	R	FD	Mean Shift	**90.1**	2.0	100.9	40.4
		R	UMAP	Mean Shift	86.6	2.5	125.4	50.2
		TRH	UMAP	Mean Shift	85.5	2.0	145.9	**58.4**
	1,000	R	UMAP	Mean Shift	**90.5**	2.9	196.1	39.2
		TRH	UMAP	Mean Shift	88.3	2.3	305.8	**61.2**
250k	500	R	FD	Mean Shift	**98.7**	2.5	171.3	68.6
		R	UMAP	Mean Shift	98.5	2.1	179.9	**72.0**
		TRH	UMAP	Mean Shift	94.4	2.3	165.3	66.2
	1,000	R	FD	Mean Shift	**99.1**	2.3	353.5	70.6
		R	UMAP	Mean Shift	98.8	2.1	359.2	**71.8**
		TRH	UMAP	Mean Shift	97.9	2.5	355.5	71.2
	5,000	R	FD	Mean Shift	**98.8**	2.1	1,264.3	50.6
		R	UMAP	Mean Shift	98.6	2.4	1,322.2	52.8
		TRH	UMAP	Mean Shift	97.9	2.7	1,872.4	**74.8**
1M	500	R	FD	Mean Shift	**99.0**	2.6	180.4	72.2
		R	t-SNE	Mean Shift	94.9	2.1	165.1	66.0
		R	UMAP	Mean Shift	97.5	2.6	179.8	72.0
		T	UMAP	Mean Shift	98.0	2.0	162.3	65.0
		TRH	t-SNE	Mean Shift	91.7	2.3	171.3	68.6
		TRH	UMAP	Mean Shift	98.9	2.3	186.5	**74.6**
	1,000	R	FD	Mean Shift	**99.4**	2.1	366.7	73.4
		R	t-SNE	Mean Shift	94.6	2.0	309.9	62.0
		R	UMAP	DBSCAN	84.4	2.2	403.1	80.6
		R	UMAP	Mean Shift	98.9	2.7	369.5	73.8
		T	t-SNE	Mean Shift	92.7	2.0	307.7	61.6
		T	UMAP	Mean Shift	98.6	2.0	349.8	70.0
		TRH	FD	Mean Shift	95.7	2.1	326.3	65.2
		TRH	t-SNE	Mean Shift	96.0	2.1	348.1	69.6
		TRH	UMAP	DBSCAN	81.7	2.0	415.1	**83.0**
		TRH	UMAP	Mean Shift	98.7	2.7	366.8	73.4
	5,000	R	FD	Mean Shift	**99.6**	2.3	1,971.5	78.8
		R	UMAP	Mean Shift	99.3	2.5	1,965.2	78.6
		T	t-SNE	Mean Shift	97.8	2.0	1,795.0	71.8
		T	UMAP	Mean Shift	99.2	2.1	1,869.3	74.8
		TRH	FD	Mean Shift	99.1	2.0	1,838.8	73.6
		TRH	UMAP	DBSCAN	93.2	2.2	2,180.6	**87.2**
		TRH	UMAP	Mean Shift	99.4	2.3	1,980.7	79.2

Trump Dataset

Cluster 0 (Left-leaning)			Cluster 1 (Right-leaning)		
RT	Description	score	RT	Description	score
@TeaPainUSA	Faithful Foot Soldier of the #Resistance	98.5	@realDonaldTrump	45th President of the United States	95.4
@PalmerReport	Palmer Report: Followed by Obama. Blocked by Donald Trump Jr	69.8	@DonaldJTrumpJr	EVP of Development & Acquisitions The @Trump Org	72.4
@kylegriffin1	Producer. MSNBC's @TheLastWord.	66.5	@mitchellvii	(pro-Trump) Host of YourVoice™ America	47.9
@maddow	rachel.msnbc.com	39.5	@ScottPresler	spent 2 years to defeat Hillary. I'm voting for Trump	33.0
@tribelaw	(anti-Trump Harvard faculty)	32.0	@JackPosobiec	OANN Host. Christian. Conservative.	32.5
Hashtag	Description	score	Hashtag	Description	score
VoteBlue	Vote Dem	12	Fakenews		18.5
VoteBlueToSaveAmerica	Vote Dem	11	Democrats	-	15.5
AMJoy	program on MSNBC	5	LDtPoll	Lou Dobbs (Fox news) poll	12.0
TakeItBack	Democratic slogan	4	msm	main stream media	11.0
Hitler	controvercy over the term "nationalist"	3	FakeBombGate	claiming bombing is fake	11.0

Erdoğan Dataset

Cluster 0 (anti-Erdoğan)			Cluster 1 (pro-Erdoğan)		
RT	Description	score	RT	Description	score
@vekilince	(Muhammem Inci – presidential candidate)	149.6	@06melihgokcek	(Ibrahim Melih Gokcek – ex. Governer of Ankara)	64.9
@cumhuriyetgzt	(Cumhuriyet newspaper)	104.0	@GizliArsivTR	(anti-Feto/PKK account)	54.0
@gazetesozcu	(Sozcu newspaper)	82.5	@UstAkilOyunlari	(Pro-Erdoğan conspiracy theorist)	49.7
@kacsaatoldunet	(popular anti-Erdoğan account)	80.0	@medyaadami	(Freelance journalist)	42.0
@tgmcelebi	(Mehmet Ali Celebi – leading CHP member)	65.8	@Malazgirt_Ruhu		37.0
Hashtag	Description	score	Hashtag	Description	score
tamam	enough (anti-Erdoğan)	49.0	VakitTürkiyeVakti	AKP slogan "It is Turkey time"	42.7
Muharremİncee	Muharrem İnce – presidential candidate	43.5	iyikiErdoanVar	Great that Erdoğan is around	20.0
demirtaş	Selahattin Demirtaş – presidential candidate	12.0	tatanka	Inci's book of poetry	19.0
KılıçdaroğluNeSöyledi	"what did Kılıçdaroğlu (CHP party leader) say"	11.0	HazırızTürkiye	Turkey: We're Ready (AKP slogan)	17.7
mersin	place for Inci rally	11.0	katilHDPKK	Killer PKK (Kurdish group)	17.0

Crisis Communication: A Comparative Study of Communication Patterns Across Crisis Events in Social Media

Hernan Sarmiento
Department of Computer Science
University of Chile
IMFD Chile
hsarmien@dcc.uchile.cl

Barbara Poblete
Department of Computer Science
University of Chile
IMFD Chile
bpoblete@dcc.uchile.cl

ABSTRACT

Valuable and timely information about crisis situations such as natural disasters, can be rapidly obtained from user-generated content in social media. This has created an emergent research field that has focused mostly on the problem of filtering and classifying potentially relevant messages during emergency situations. However, we believe important insight can be gained from studying online communications during disasters at a more comprehensive level. In this sense, a higher-level analysis could allow us to understand if there are collective patterns associated to certain characteristics of events. Following this motivation, we present a novel comparative analysis of 41 real-world crisis events. This analysis is based on textual and linguistic features of social media messages shared during these crises. For our comparison we considered hazard categories (i.e., human-induced and natural crises) as well as subcategories (i.e., intentional, accidental and so forth). Among other things, our results show that using only a small set of textual features, we can

1 INTRODUCTION

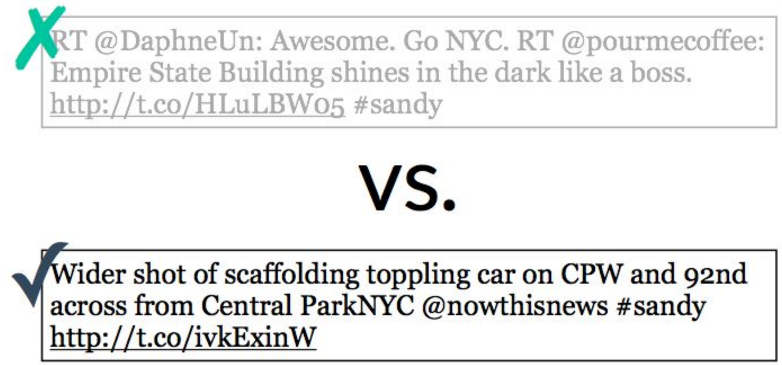
Disasters can be defined as social events that harm people and destroy property, damaging social communities as a result [23]. Extreme events, such as these, can have long-lasting effects, redefining cultural identities and highlighting collective needs [32]. Thus, even affecting socio-political structures to the extent of triggering changes in political and institutional laws to decrease the vulnerability of the population to such events. Hence, mechanisms to improve the understanding of these events are crucial to explain why they do or do not have a certain impact [2].

The increase in online user interactions, in the form of *conversations* that take place in online social networking platforms, has created a major source of real-time data about physical-world events, specially crisis situations. One of these platforms is Twitter[1], a widely adopted microblogging service where users post short messages (called *tweets*),[2] that is well-known for facilitating rapid information propagation to its users [34]. An example, among many,

Motivación

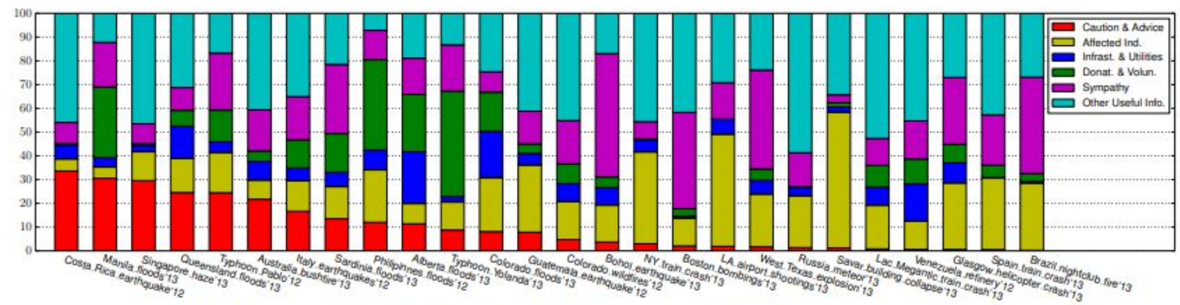
- Durante situaciones de crisis, usuarios tienden a compartir mensajes en redes sociales, incrementando la frecuencia de la red durante estos periodos

- Sin embargo, toda esta información no necesariamente es útil, por lo que vemos mucho ruido en estos mensajes

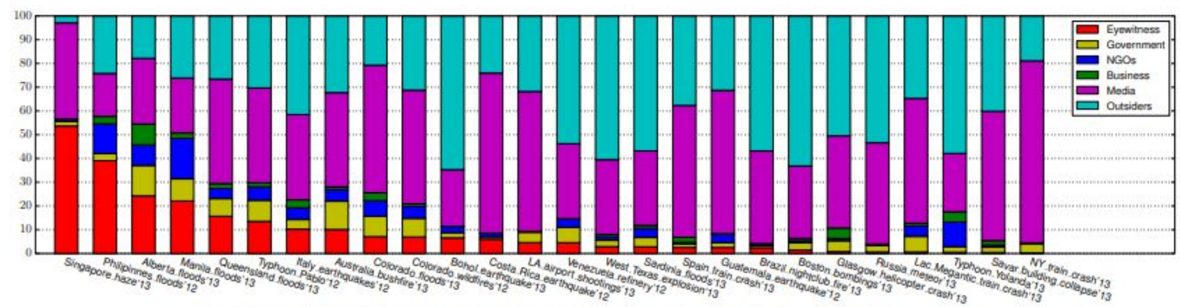


Motivación

- Identificar y caracterizar estos mensajes ayuda, principalmente, oficinas de emergencia, gobierno, prensa y público en general, entender cómo son los desastres en redes sociales.



(a) Distribution of information types, sorted by descending proportion of *caution and advice* tweets.



(b) Distribution of information sources, sorted by descending proportion of *eyewitness* tweets.

Figure 1. Distributions of information types and sources (best seen in color)

Propuesta

- Estudiar mensajes (relevantes) de redes sociales publicados durante crisis, desde un punto de vista del lenguaje de estos.
- El objetivo es encontrar patrones transversales a través de los eventos, permitiendo agruparlos en tipos de desastres similares

Table 2: Number of English messages and events by category, subcategory and type of crisis. Symbols (*) and (**) correspond to human-induced and natural crises respectively.

Hazard Category	No. Events	No. Messages
Human-induced	12	12,439
Natural	29	27,658
Hazard Subcategory	**No. Events**	**No. Messages**
Accidental (*)	7	5,502
Climatological (**)	3	2,305
Epidemic (**)	2	1,919
Geophysical (**)	8	5,903
Hydrological (**)	7	8,699
Intentional (*)	5	6,937
Meteorological (**)	8	8,459
Others	1	373
Hazard Type	**No. Events**	**No. Messages**
Bombing (*)	2	3,025
Bombing/Shooting (*)	1	1,437
Building collapse (*)	1	411
Crash (*)	1	449
Cyclone (**)	1	543
Derailment (*)	3	975
Earthquake (**)	3	5,903
Explosion (*)	1	3,569
Fire(*)	1	98
Flood (**)	7	8,699
Hurricane (**)	3	3,639
Meteorite (**)	1	373
Shooting (*)	3	2,475
Tornado (**)	1	2,428
Typhoon (**)	3	1,849
Viral disease (**)	2	1,919
Wildfire (**)	3	2,305

Feature extraction

- Extraemos 54 características de los mensajes (tweets):
  - Linguistir features (LF): usamos el *Linguistic Inquiry and Word Count (LIWC),* el cual permite categorizar palabras en ciertas categorías lingüísticas y sicológicas. Computamos la frecuencia relativa de uso de cada categoría por evento.
  - Twitter content features (TCF): usamos TweetNLP para contar la ocurrencia de menciones de usuarios, URL y hashtags
  - Entity-Based Features (EBF): usando spaCy, extraemos entidades del texto

Crisis similarity

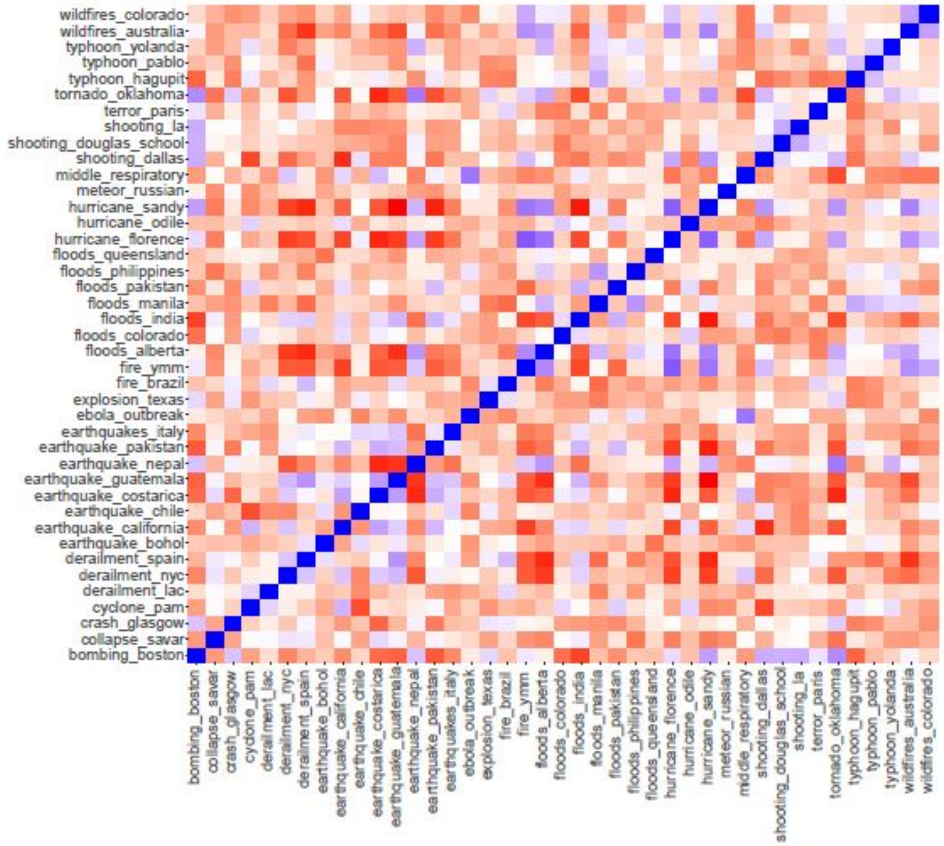


Figure 1: (Best viewed in color) Cosine similarity considering 54 textual features to represent crisis events. Blue and red colors mean high and low similarity values respectively

Crisis Evaluation

- Agrupar los eventos (y mensajes) de acuerdo a la categoría o tipo de crisis que ellos tienen.

- Dado que el vector de características era de tamaño considerable, aplicamos t-test para saber si los promedios de una variable son diferentes entre 2 grupos de crisis.

- Mantenemos las features en las cuales encontramos diferencias significativas y para luego aplicar clustering jerárquico

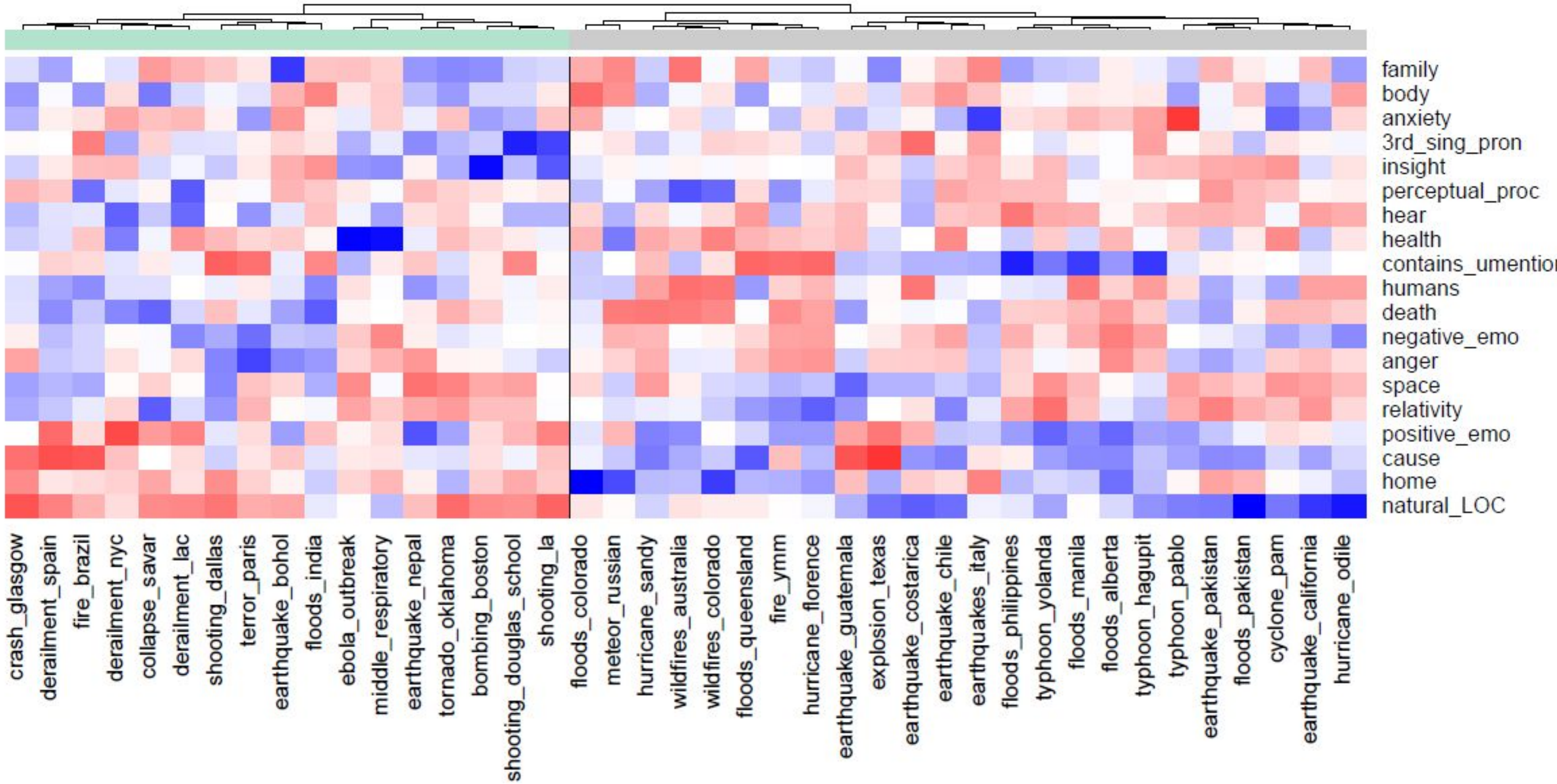


Figure 3: (Best viewed in color) Dendrogram obtained by hierarchical agglomerative clustering of crisis events, cut at two clusters finding a clear separation between ***human-induced and natural events*** (green and gray clusters, respectively). Rows represent features and columns crisis events. Blue and red cells indicate high and low values, respectively.

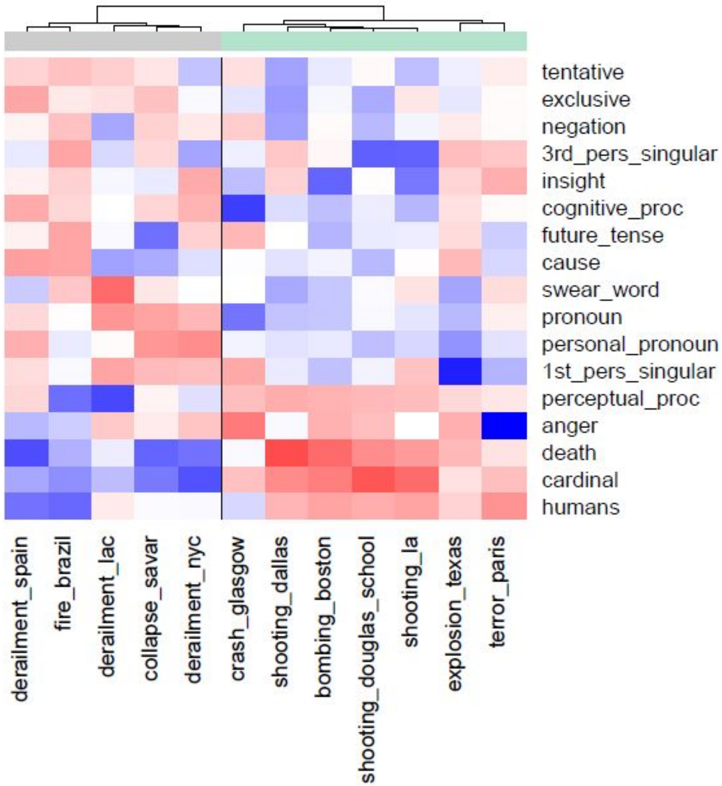


Figure 4: (Best viewed in color) Dendrogram obtained by hierarchical agglomerative clustering of crisis events, cut at two clusters finding a clear separation between *accidental and intentional human-induced events* (gray and green clusters, respectively). Rows represent features and columns crisis events.

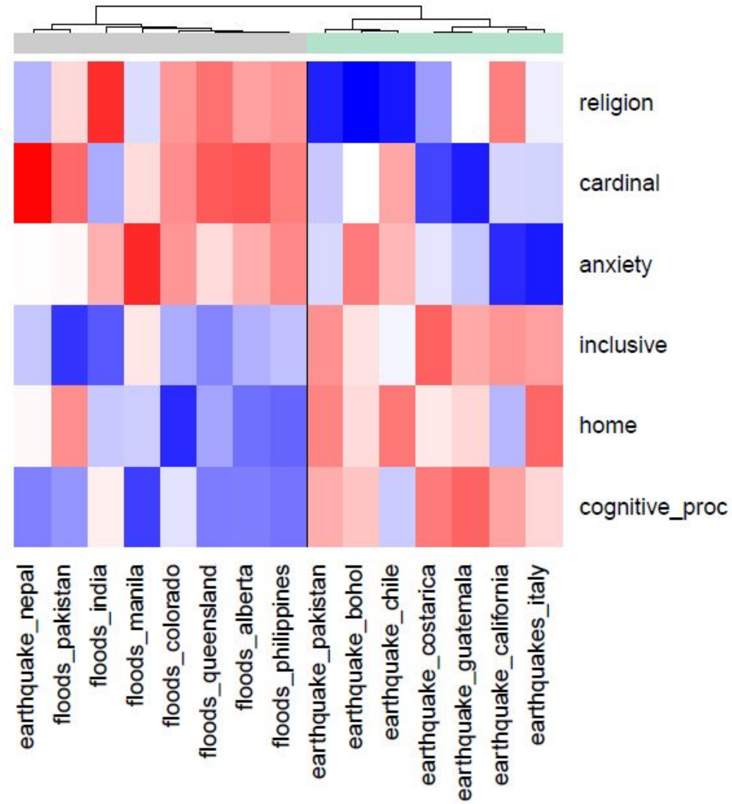


Figure 5: (Best viewed in color) Dendrogram obtained by hierarchical agglomerative clustering of crisis events, cut at two clusters finding a clear separation between *geophysical and hydrological natural events* (green and gray clusters, respectively). Rows represent features and columns crisis events.

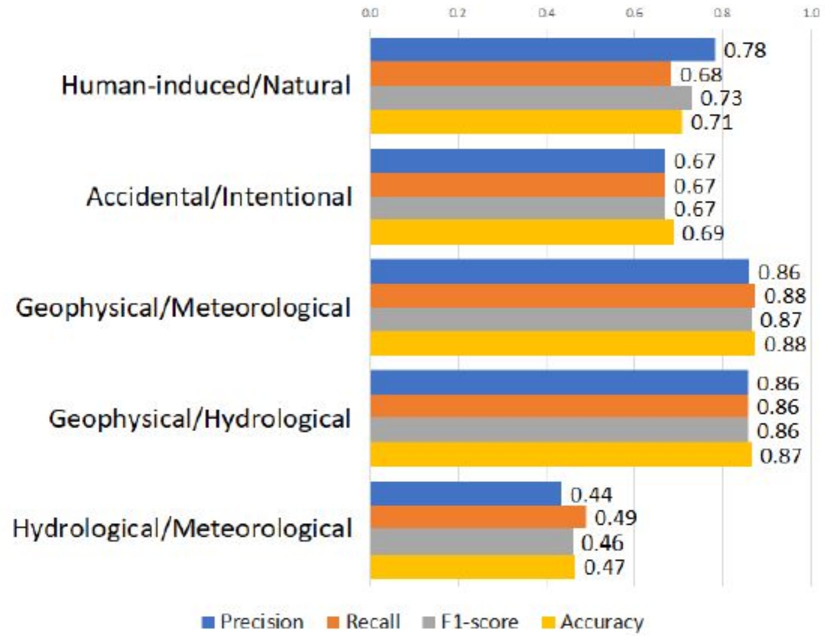


Figure 2: Crisis evaluation results using hierarchical agglomerative clustering techniques. Each evaluation compares two types of hazard categories or hazard subcategories.

Table 3: List of differences found for hazard categories and subcategories. Each row represents a comparison between two hazard categories or subcategories, represented as columns in the table.

Human-induced event tweets contain (1) more *family*-related and *humans*-related terms. (2) more *body*-related and *health*-related terms. (3) more terms about *perceptual* processes. (4) more *negative* emotions, *anger*-related, *anxiety*-related and *death*-related terms terms.	**Natural event tweets contain** (1) more mentions of natural places (*natural_LOC* feature) (2) more *home*-related terms. (3) more *positive emotions.* (4) more *cause*-related terms.
Accidental event tweets more (1) more *humans*-related terms. (2) more *anger*-related terms. (3) more *death*-related terms. (4) more terms about *perceptual* processes.	**Intentional event tweets more** (1) more *pronouns* and *future tenses* and *swear words.* (2) more *insight*-related, *negation*-related and *cause*-related terms.
Geophysical event tweets contain (1) more *anxiety*-related terms. (2) more *cardinal* numbers.	**Hydrological event tweets contain** (1) more terms about *cognitive* processes and *inclusive*-related terms. (2) more *home*-related terms.
Geophysical event tweets more (1) more *death*-related terms. (2) more terms about locations (*GPE* and *FAC* features). (3) more *cardinal* numbers.	**Meteorological event tweets contain** (1) more *pronouns, future tenses, prepositions* and *functional words.* (2) more *home*-related terms. (3) more terms about *cognitive* processes, *inclusive*-related and *tentative*-related terms.
Hydrological event tweets (1) Clustering metrics do not support significant differences.	**Meteorological event tweets** (1) Clustering metrics do not support significant differences.