TREC 2019 Incident Streams Overview

RICHARD MCCREADIE,[1] IAN SOBOROFF[2] AND CODY BUNTAIN[3]

[1]UNIVERSITY OF GLASGOW, [2]NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY,
[3]NEW JERSEY INSTITUTE OF TECHNOLOGY

Background

Internationally, civil protection, police forces and emergency response agencies are under increasing pressure to more quickly and effectively respond to emergency situations.

- 50,000 people per-year on average die during natural disasters internationally





Situational Awareness

The corner-stone that forms the basis of successful response actions is ***situational awareness***

- Situational awareness is derived from accurate knowledge of what is occurring at the current moment (the operational picture)

Command and Control centre staff need an accurate and complete operational picture to:

- Choose effective actions to remedy the situation
- Take preventative steps to avoid further loss of life/damage.

To build an operational picture, command and control centre operators receive updates from local responders and members of the affected public, as well as other services (e.g. weather)




Social Media

The mass adoption of mobile internet-enabled devices paired with wide-spread use of social media platforms for communication and coordination has created new ways for the public on-the-ground to contact response services.

FEMA

"Social media is an important part of the whole community approach because it helps facilitate the vital two-way communication between emergency management agencies and the public."

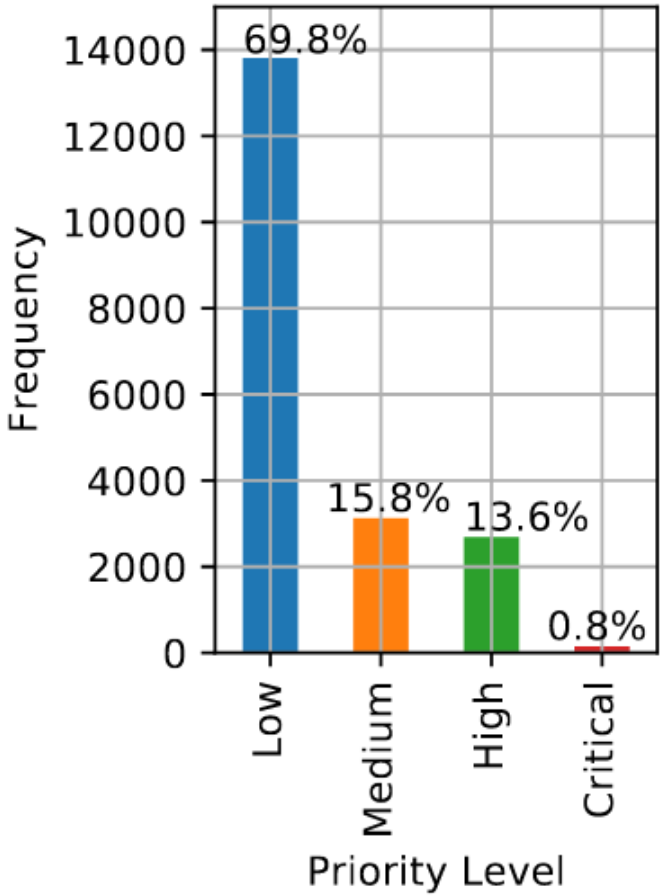
Craig Fugate (FEMA Administrator, 2009-2017)

Is there **Actionable Information** out there?

We performed a study based on last-year's labelled tweets (TREC-IS 2018) across 15 events to find out

We found that:

- Around **1% of information was marked as 'Critical'** (a response officer needs to see this immediately)
  - E.g. 'ACTalliance local partners see transitional shelters blown away. Need to assess damage and respond asap'
- Around **14% of information was marked as 'High'** (a response officer should see this within 30 minutes
  - 'Borongan, Eastern Samar's communication lines and power &water lines are still down. via @EdwinSevidal #RubyPH'



McCreadie R., Buntain, C. and Soboroff, I. (2019) TREC Incident Streams: Finding Actionable Information on Social Media. In: 16th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2019)

Who is Monitoring Social Media?

In many regions, social media is not monitored by response services

- Lack of manpower
- Considered 'too risky' to support due to lack of effective tools

In some regions, social media is monitored by volunteers groups (Europe) or by public information officers (U.S.)

- But have to manually find information using their personal accounts

Incident Streams Task

AIMS AND USE-CASE

Use Case

Before formally defining a task we needed to consider what sort of social media monitoring technology we are targeting

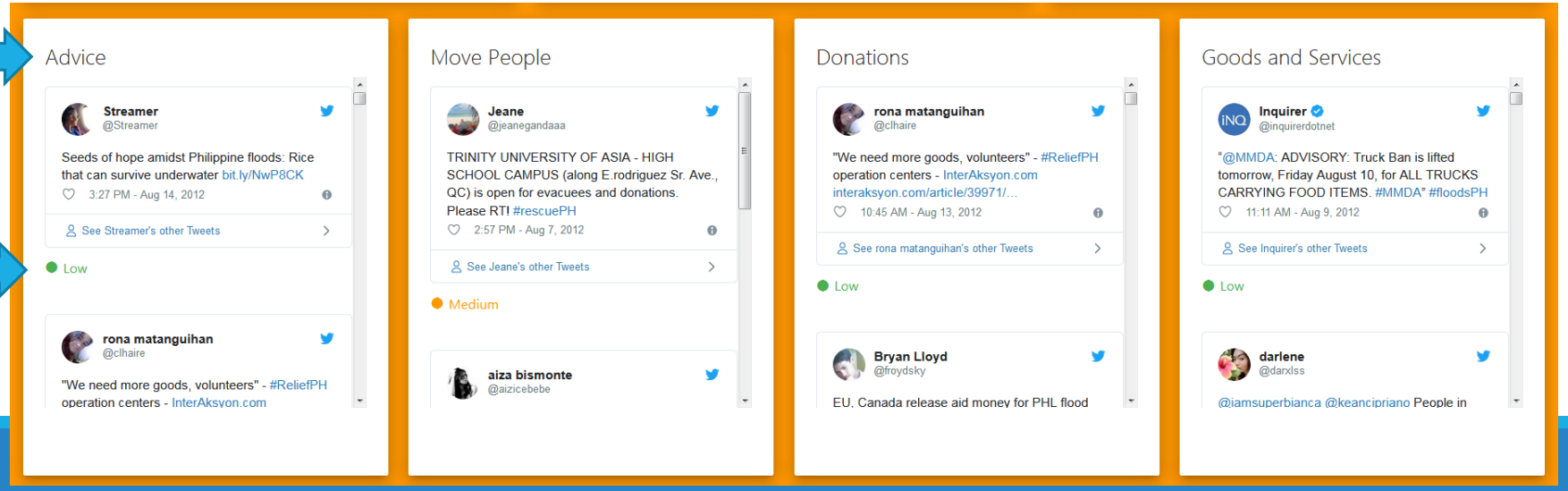
- What sort of tool would help improve situational awareness?

Our goal is to get informative information to the right response officer in a timely fashion

- Some types of information are useful, while others are not
- Some information is more important/critical than others

Information Types

Priority Level



2019 Task

Develop a system that process a real-time stream of social media posts during an event, categorizing that content based on an pre-defined ontology of user information needs and assigning priority levels for each.

- The system needs to assign information types to each tweet in the stream
- The system needs to assign a priority score to each tweet, representing how important it is for a response officer should see that tweet immediately

High-Level Information Type	Description
Request-GoodsServices	The user is asking for a particular service or physical good.
Request-SearchAndRescue	The user is requesting a rescue (for themselves or others)
Request-InformationWanted	The user is requesting information
CallToAction-Volunteer	The user is asking people to volunteer to help the response effort
CallToAction-Donations	The user is asking people to donate goods/money
CallToAction-MovePeople	The user is asking people to leave an area or go to another area
Report-FirstPartyObservation	The user is giving an eye-witness account
Report-ThirdPartyObservation	The user is reporting a information from someone else
Report-Weather	The user is providing a weather report (current or forecast)
Report-EmergingThreats	The user is reporting a potential problem that may cause future loss of life or damage
Report-MultimediaShare	The user is sharing images or video
Report-ServiceAvailable	The user is reporting that someone is providing a service
Report-Factoid	The user is relating some facts, typically numerical
Report-Official	An official report by a government or public safety representative
Report-CleanUp	A report of the clean up after the event
Report-Hashtags	Reporting which hashtags correspond to each event
Report-News*	The post providing/linking to continuous coverage of the event
Report-NewSubEvent*	The user is reporting a new occurrence that public safety officers need to respond to.
Report-Location*	The post contains information about the user or observation location.
Other-Advice	The author is providing some advice to the public
Other-Sentiment	The post is expressing some sentiment about the event
Other-Discussion	Users are discussing the event
Other-Irrelevant	The post is irrelevant, contains no information
Other-ContextualInformation*	The post is generic news, e.g. reporting that the event occurred
Other-OriginalEvent*	The Responder already knows this information

* – modified for 2019-A and 2019-B editions

Information Type Ontology

Editions, Events and Tweets

DATASETS

Event Types

From previous data challenges that have looked at emergency related data, we knew that systems often want to customise for particular event types

- E.g. for terrorist attacks we might want to identify what the terrorists are armed with

So we restricted the task to only six event types

- **Wildfire, Earthquake, Flood, Typhoon/Hurricane, Bombing, Shooting**
- Participants systems are assumed to have been told the event type (they don't need to figure that out automatically)

Editions

2019 is the second year the Incident Streams track has run

- Participants are able to use any datasets from previous editions

We ran two editions of the track in 2019

- **2019-A**, with submission in **June**
- **2019-B**, With submission in **September**

For each edition, we select a set of emergency-related events and crawl tweets for them

Dataset	Identifier	Event Name	Event Type	Source
2018 Training	TRECIS-CTIT-H-Training-001	2012 Colorado wildfires	wildfire	CrisisLexT26 Olteanu et al. 2015
	TRECIS-CTIT-H-Training-002	2012 Costa Rica Earthquake	earthquake	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Training-003	2013 Colorado Floods	flood	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Training-004	2012 Typhoon Pablo	typhoon/hurricane	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Training-005	2013 LA Airport Shooting	shooting	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Training-006	2013 West Texas Explosion	bombing	CrisisLexT26 (Olteanu et al. 2015)
2019-A Training	TRECIS-CTIT-H-Test-007	2012 Guatemala earthquake	earthquake	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Test-008	2012 Italy earthquakes	earthquake	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Test-009	2012 Philipinnes floods	flood	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Test-010	2013 Alberta floods	flood	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Test-011	2013 Australia bushfire	wildfire	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Test-012	2013 Boston bombings	bombing	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Test-013	2013 Manila floods	flood	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Test-014	2013 Queensland floods	flood	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Test-015	2013 Typhoon Yolanda	typhoon	CrisisLexT26 (Olteanu et al. 2015)
	TRECIS-CTIT-H-Test-016	2011 Joplin tornado	typhoon	CrisisNLP Resource #2 (Imran et al. 2013)
	TRECIS-CTIT-H-Test-017	2014 Chile Earthquake	earthquake	CrisisNLP Resource #1 (Imran et al. 2016)
	TRECIS-CTIT-H-Test-018	2014 Typhoon Hagupit	typhoon	CrisisNLP Resource #1 (Imran et al. 2016)
	TRECIS-CTIT-H-Test-019	2015 Nepal Earthquake	earthquake	CrisisNLP Resource #1 (Imran et al. 2016)
	TRECIS-CTIT-H-Test-020	2018 FL School Shooting	shooting	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-021	2015 Paris attacks	bombing	Collected via the GNIP service
2019-B Training	TRECIS-CTIT-H-Test-022	2019 Choco Flood	flood	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-023	2019 Andover Fire	wildfire	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-024	2014 California Earthquake	earthquake	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-025	2013 Bohol Earthquake	earthquake	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-026	2018 Florence Hurricane	typhoon	Donated by participant group
	TRECIS-CTIT-H-Test-027	2017 Dallas Shooting	shooting	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-028	2016 Fort McMurray Wildfire	wildfire	Crawled by the Organizers via Twitter API
2019-B Testing	TRECIS-CTIT-H-Test-029	2019 Alberta Wildfires	wildfire	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-030	2019 Cyclone Kenneth	typhoon	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-031	2019 Luzon earthquake	earthquake	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-032	2019 STEM School Highlands Ranch shooting	shooting	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-033	2019 Durban Easter floods	flood	Crawled by the Organizers via Twitter API
	TRECIS-CTIT-H-Test-034	2019 Poway synagogue shooting	shooting	Crawled by the Organizers via Twitter API

Tweet Filtering

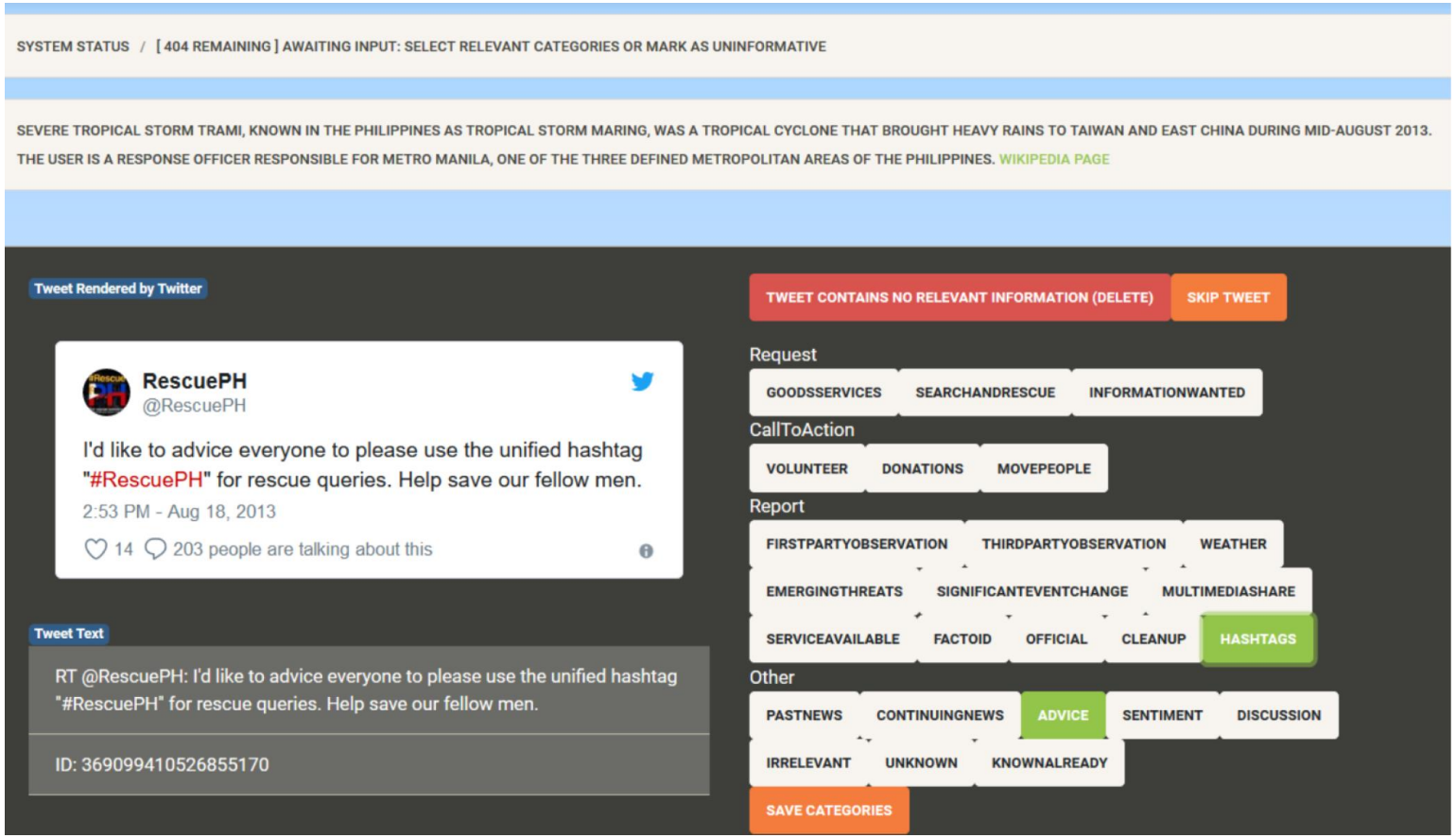
The social-media footprint for an event can vary greatly dependant on factors such as the event impact, international visibility and social media adoption in the affected region

- For this reason, the number of tweets collected for an event can range from 100's to millions

TREC does not have the resources (or time) to manually label all tweets for an event unless that event is small

- Hence, we pre-filter all of the events that we use
- In particular, we cluster tweets by their text and then select on tweet per cluster to be assessed

TRECIS-CTIT-H-Test-007	2012 Guatemala earthquake	178
TRECIS-CTIT-H-Test-008	2012 Italy earthquakes	118
TRECIS-CTIT-H-Test-009	2012 Philipinnes floods	480
TRECIS-CTIT-H-Test-010	2013 Alberta floods	739
TRECIS-CTIT-H-Test-011	2013 Australia bushfire	710
TRECIS-CTIT-H-Test-012	2013 Boston bombings	543
TRECIS-CTIT-H-Test-013	2013 Manila floods	443
TRECIS-CTIT-H-Test-014	2013 Queensland floods	744
TRECIS-CTIT-H-Test-015	2013 Typhoon Yolanda	629
TRECIS-CTIT-H-Test-016	2011 Joplin tornado	152
TRECIS-CTIT-H-Test-017	2014 Chile Earthquake	321
TRECIS-CTIT-H-Test-018	2014 Typhoon Hagupit	6,696
TRECIS-CTIT-H-Test-019	2015 Nepal Earthquake	7,301
TRECIS-CTIT-H-Test-020	2018 FL School Shooting	1,118
TRECIS-CTIT-H-Test-021	2015 Paris attacks	2,066
TRECIS-CTIT-H-Test-022	2019 Choco Flood	854
TRECIS-CTIT-H-Test-023	2019 Andover Fire	375
TRECIS-CTIT-H-Test-024	2014 California Earthquake	128
TRECIS-CTIT-H-Test-025	2013 Bohol Earthquake	646
TRECIS-CTIT-H-Test-026	2018 Florence Hurricane	2,500
TRECIS-CTIT-H-Test-027	2017 Dallas Shooting	2,500
TRECIS-CTIT-H-Test-028	2016 Fort McMurray Wildfire	2,500
TRECIS-CTIT-H-Test-029	2019 Alberta Wildfires	2,500
TRECIS-CTIT-H-Test-030	2019 Cyclone Kenneth	2,500
TRECIS-CTIT-H-Test-031	2019 Luzon earthquake	2,500
TRECIS-CTIT-H-Test-032	2019 STEM School Highlands Ranch shooting	2,500
TRECIS-CTIT-H-Test-033	2019 Durban Easter floods	2,500
TRECIS-CTIT-H-Test-034	2019 Poway synagogue shooting	2,500



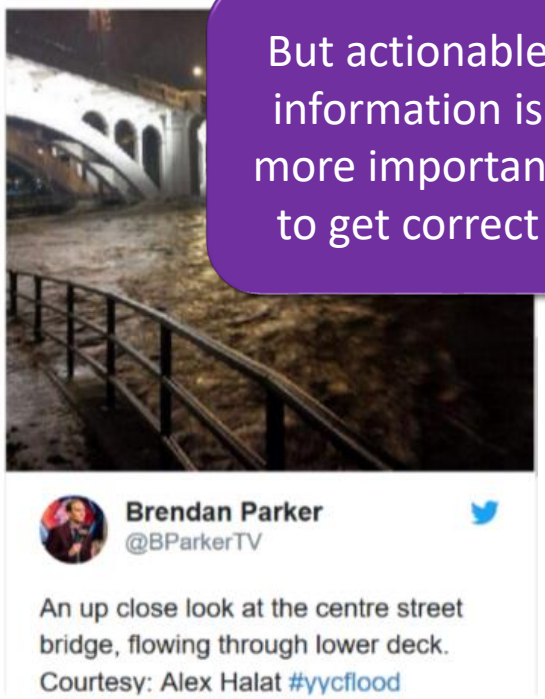
Test Data Labeling Interface

Evaluation

METRICS

What are we aiming to measure?

Tweet



But actionable information is more important to get correct

Ground Truth

- Multimedia Share
- Emerging Threat
- Third Party Observation

Priority: 0.75 (High)

Participant System

- Multimedia Share
- News
- First Party Observation
- Emerging Threat

Priority: 0.69 (Medium)

How accurate are the categories?

The false positives are less serious errors (in moderation)

Is the priority level accurate?

Distinguishing Actionable Information

Given that what we truly care about is systems ability to expose actionable information to the response officer, we need a way for metrics to capture this

For the purposes of evaluation, we primarily define actionability at category-level, i.e. we consider the following categories as actionable, based on our previous analysis

- `Request-GoodsServices',
- `Request-SearchAndRescue',
- `CallToAction-MovePeople',
- `Report-EmergingThreats',
- `Report-NewSubEvent'
- `Report-ServiceAvailable'

Actionable Categories

McCreadie R., Buntain, C. and Soboroff, I. (2019) TREC Incident Streams: Finding Actionable Information on Social Media. In: 16th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2019)

Metrics

Information Feed

Evaluates information type categorization performance

- Accuracy, All
- Positive F1, All
- Positive F1, Actionable

Prioritization

Calculates error between the human and system tweet priority scores

Accuracy, All

- RMSE, All
- RMSE, Actionable

Alerting

NEW!

A combination metric that incorporates both prioritization performance and categorization performance

- Accumulated Alert Worth, All
- Accumulated Alert Worth, High Priority Only

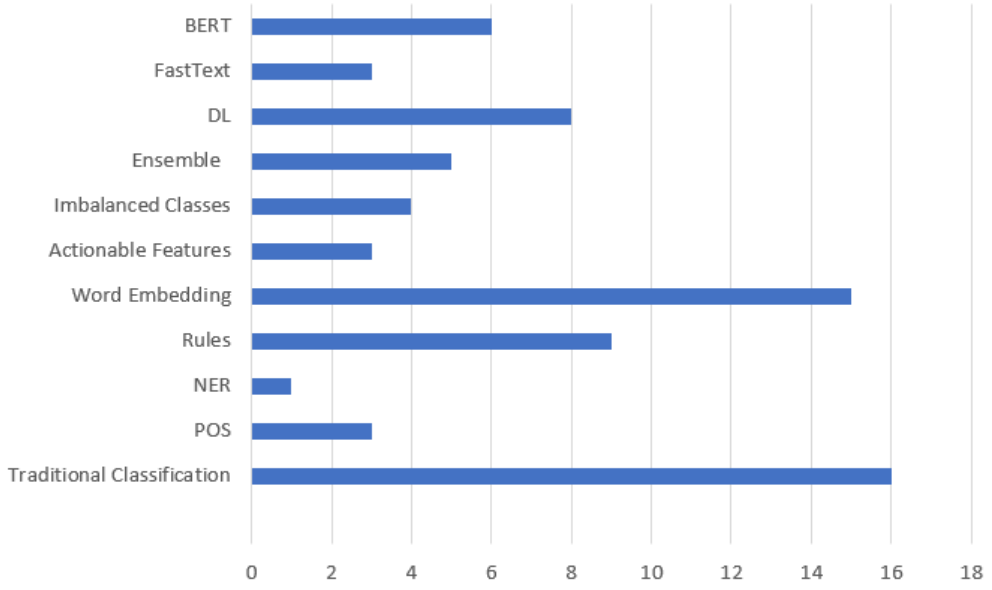
Observations and Results

PARTICIPANT PERFORMANCES

Participation

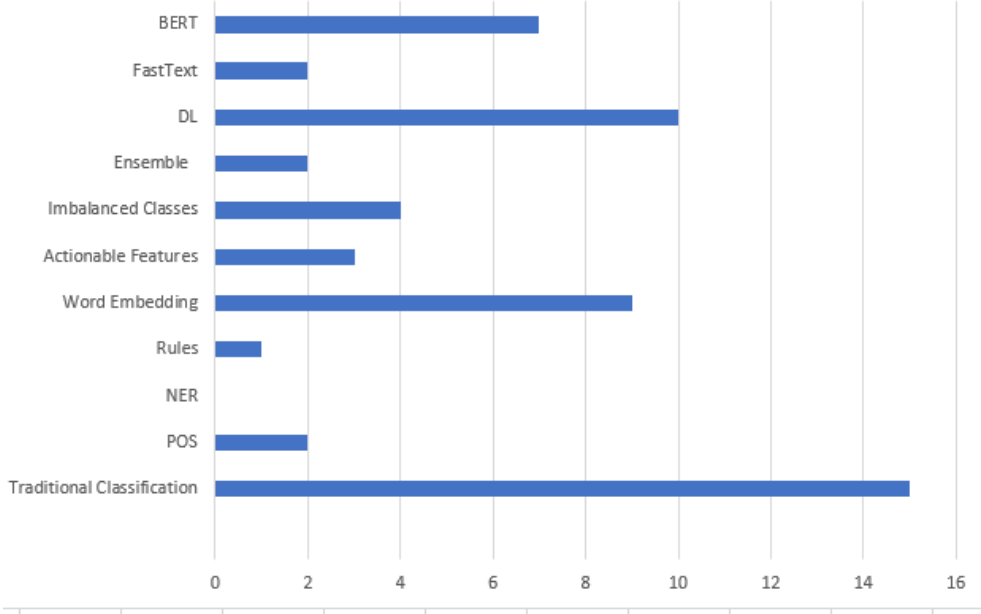
2019-A

- 10 participating groups
- 35 runs



2019-B

- 10 participating groups
- 32 runs



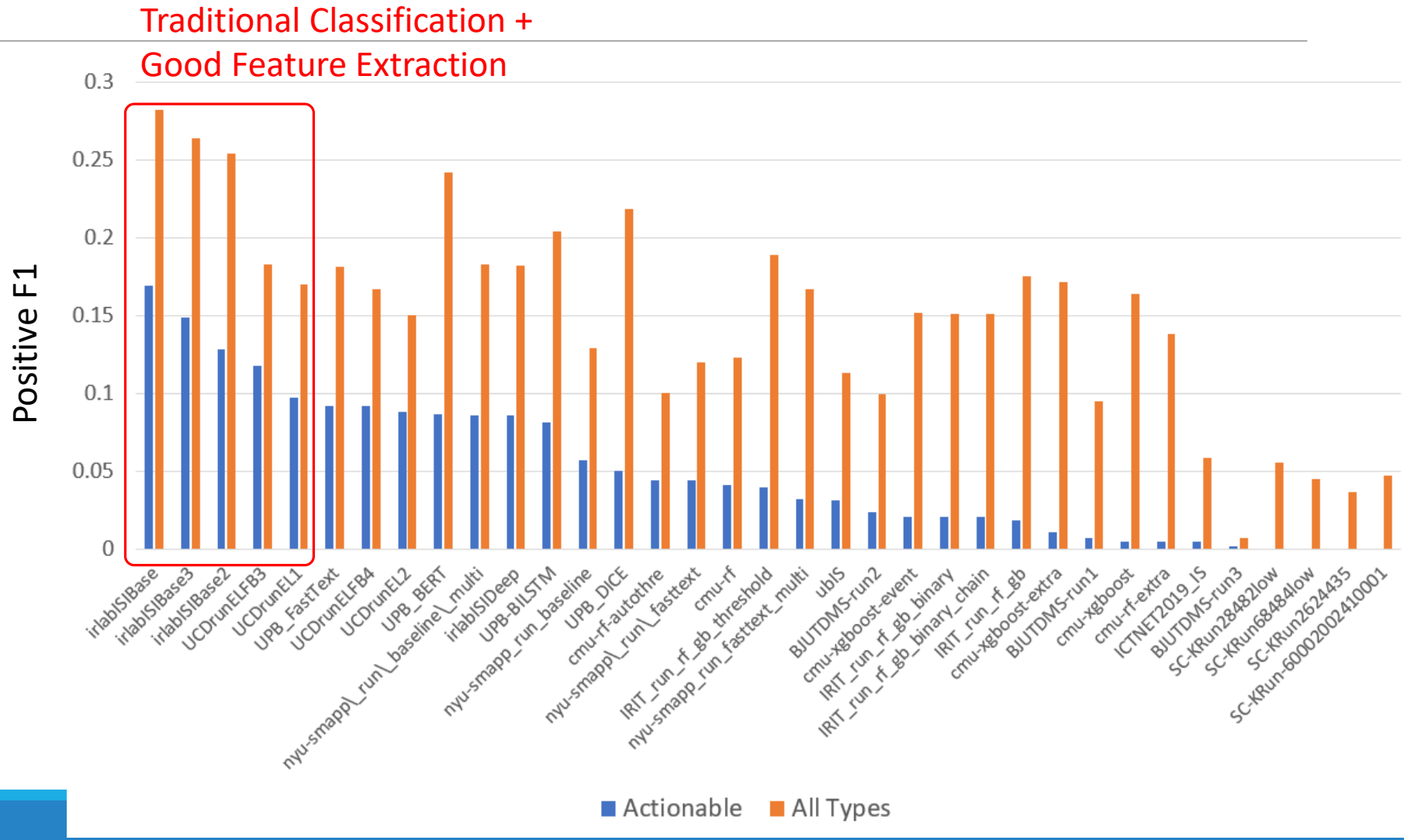
Information Feed

2019-A

Positive F1

All types vs.
Actionable types

Sorted by
Actionable



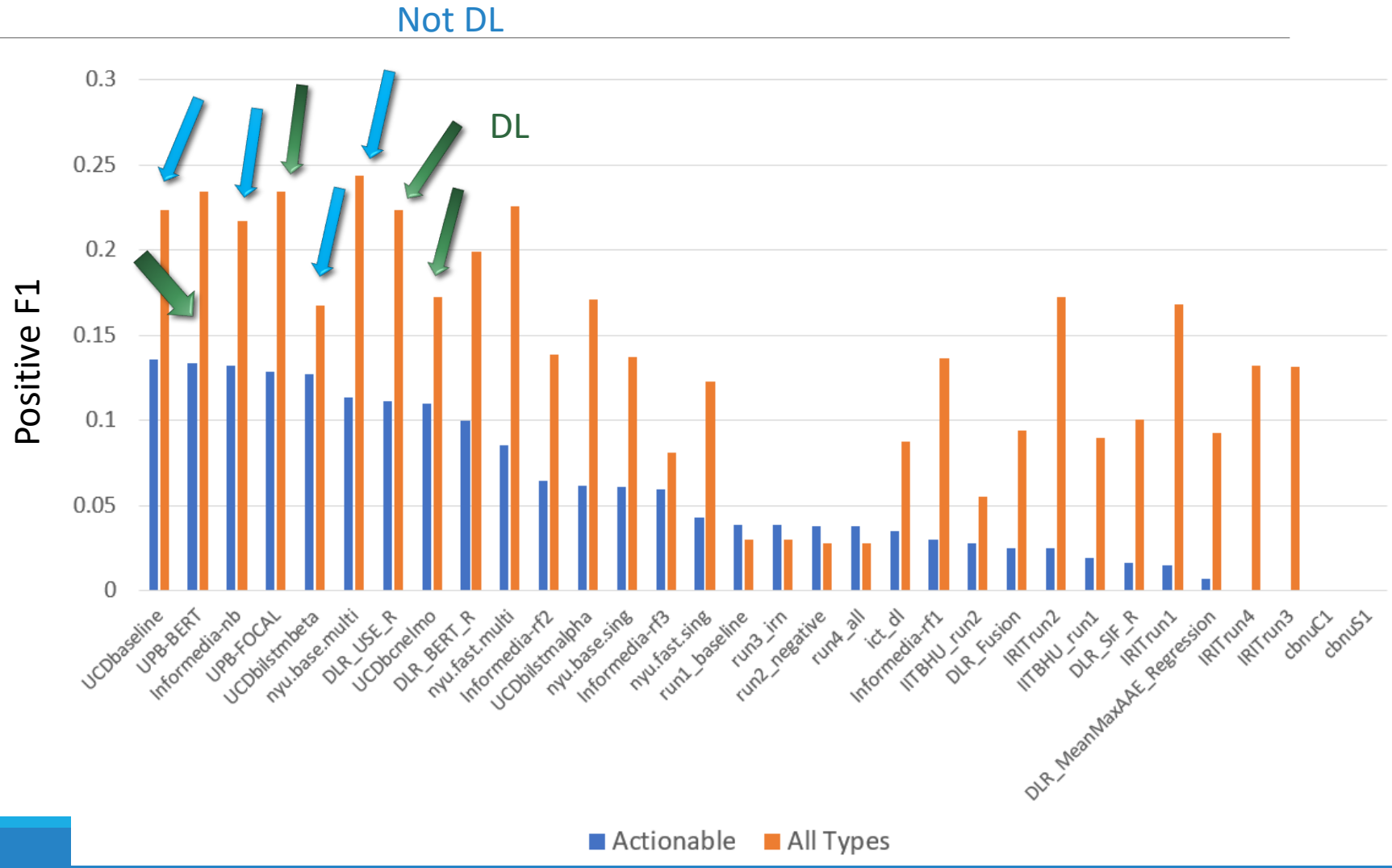
Information Feed

2019-B

Positive F1

All types vs.
Actionable types

Sorted by
Actionable

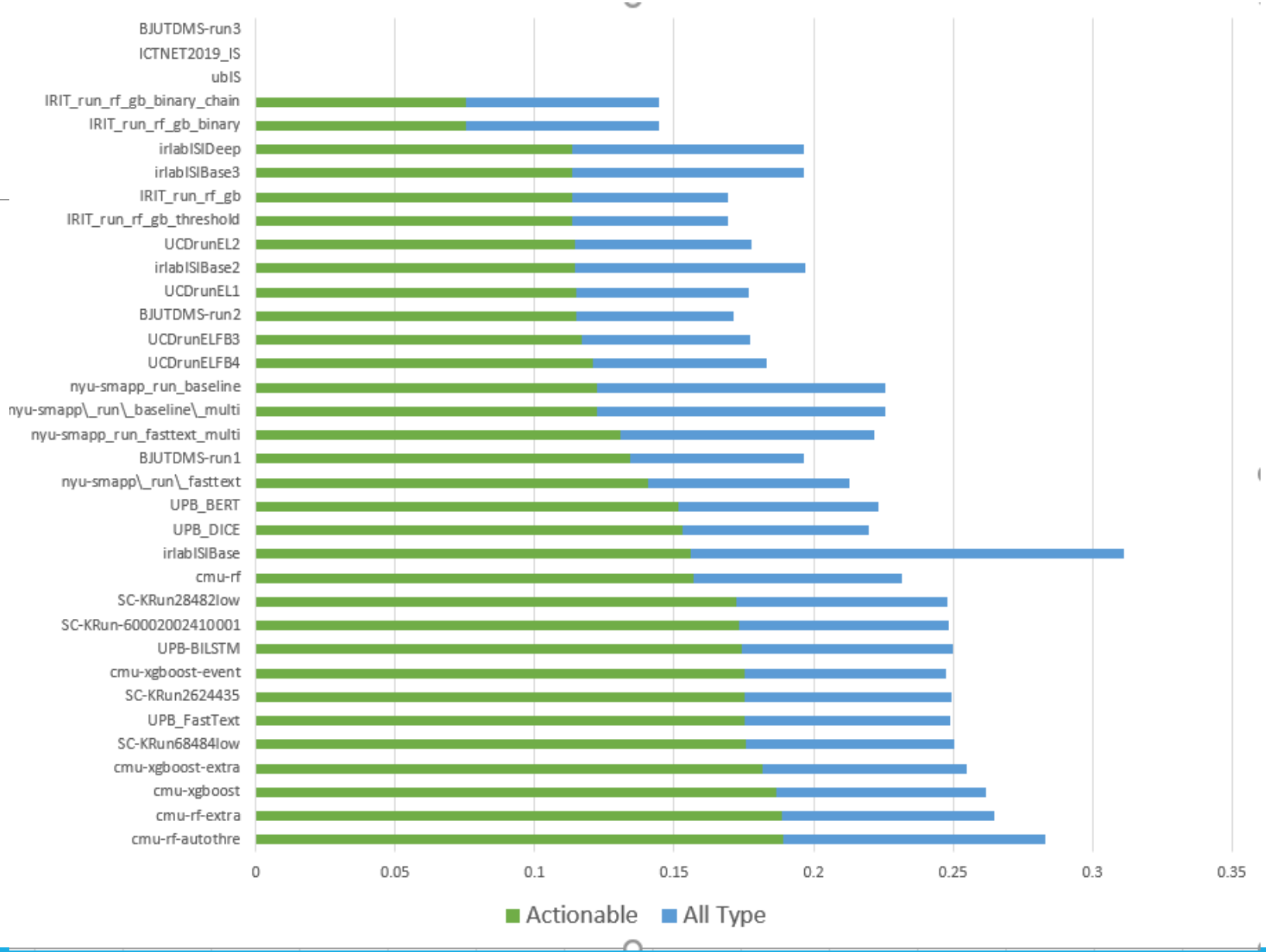


Prioritization
2019-A

Prioritization RMSE

Actionable (green), All Types (blue)

Sorted by Actionable

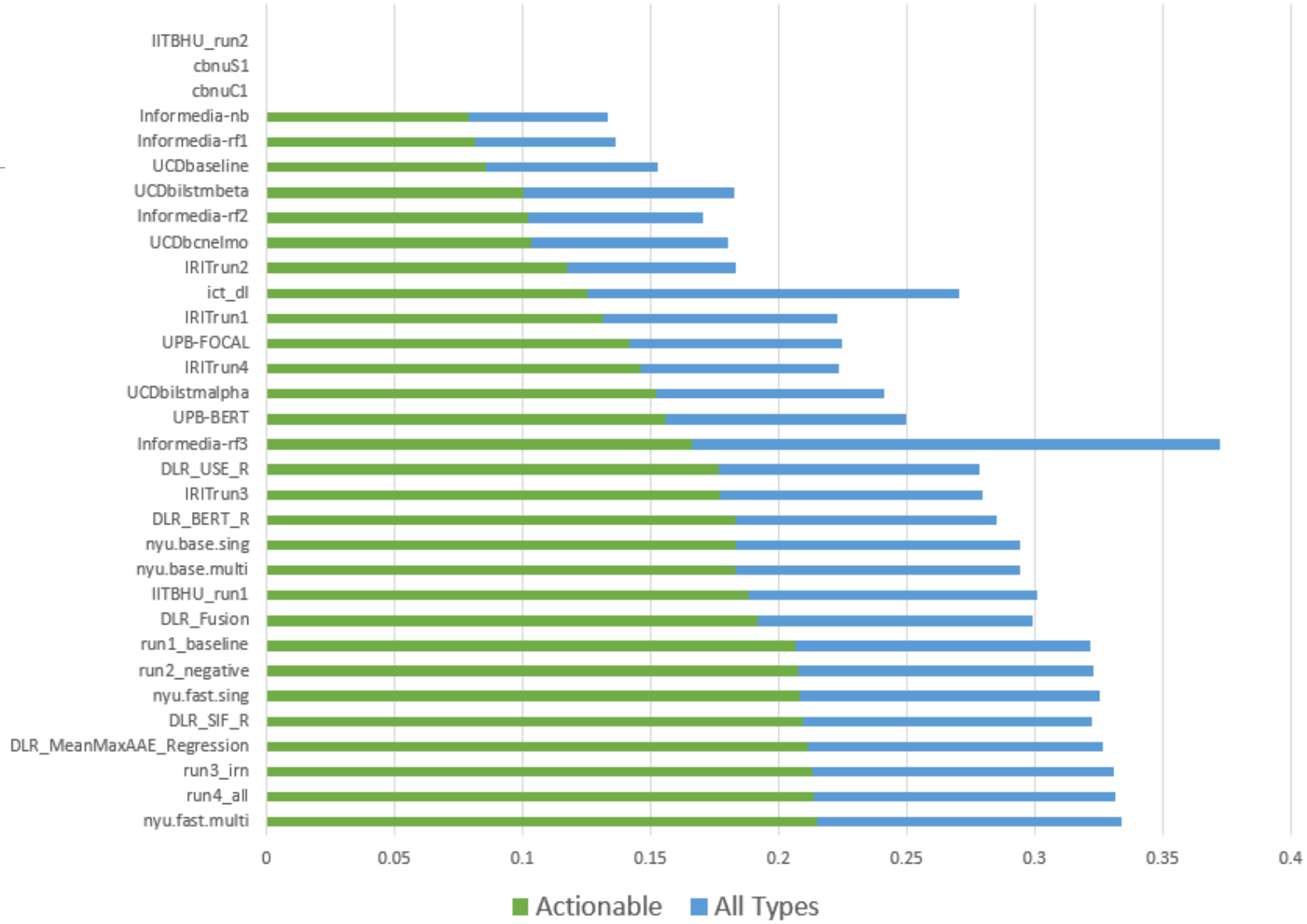


Prioritization
2019-B

Prioritization RMSE

Actionable (green), All Types (blue)

Sorted by Actionable



Summary

We now have run TREC-IS for three editions, 2018, 2019-A, 2019-B

- We have produced labelled Twitter datasets covering 34 events containing around 40,000 labelled tweets for 25 information types in total
- Demonstrated that there is valuable actionable information out there to find!

For 2019, we refined the metrics to focus more on actionable information types, as well as connecting metrics to particular emergency response use-cases

Participant systems are making progress, exploring different categorization techniques particularly in making better use of deep learning and BERT

- However, deep learning is not a magic wand here, as actionable information types have relatively few instances that can be used for training
- Having strong fundamental feature extraction and methods for tackling unbalanced data are critical factors here

There is still significant work needed before we have effective systems for finding actionable information from social media