

Finding Event Location Based On The User Generated Tweets

COMP5800. Social Computing
UMASS Lowell

Amin Majdi
01924731

Introduction & Motivation

- Problem definition:
 - Detecting the event location and the groups connected to the events
- Importance:
 - Safety
 - Business
- Who will benefit:
 - Public safety
 - Big wise companies

Dataset , Evaluation Method & Method

- Data collection:
 - Tweepy:
 - Russia-Ukraine War , Easter , Ramadan
- Auxiliary databases:
 - Cities and Countries , Country Codes , Language code , Flags , Country specific domain
- Data preparation:
 - Metadata Introduction and name-spacing:
 - Place, Location, User's Name, User's screen Name, User's screen Description , Tweet's full text, tweet's Language, Website Domain
 - Label Check Methods
 - Check City or Country , Flags , Domain suffix

Dataset , Evaluation Method & Method

- Data preparation:

- Labeling:

- Priority Action:

- Place (label is available)
 - User's Location(city\country check)
 - User's Name(city\country check, Flag check)
 - User's screen Name(city\country check, Flag check)
 - User's screen Description (city\country check, Flag check)
 - Tweet's full text (Flag check)
 - WE DID NOT USE (LANGUAGE,DOMAIN)

- Classification:

- Initial class numbers: 252 (# of countries)
 - R-U-War (20 classes) , Ramadan (20 classes), Easter (10 classes)
 - %50- %50 Train and test (shuffled)
 - Using VW (mainly base features)

Results and Insights

Event	Top Countries who talked about the event
R-U-War	'united states', 'ukraine', 'russia', 'united kingdom', 'india', 'norway', 'japan', 'germany', 'czech republic', 'italy', 'netherlands', 'france', 'spain', 'philippines', 'finland', 'canada', 'poland', 'turkey', 'tonga'
Ramadan	'saudi arabia', 'united states', 'pakistan', 'kuwait', 'egypt', 'emirates', 'indonesia', 'algeria', 'india', 'united kingdom', 'oman', 'nigeria', 'france', 'netherlands', 'yemen', 'turkey', 'norway', 'bahrain', 'sudan'
Easter	'united states', 'ukraine', 'united kingdom', 'greece', 'norway', 'czech republic', 'netherlands', 'timor-leste', 'ethiopia'

- Tonga?
- Andora?
- And

- Future works:
 - Multi label
 - weight
 - Tokenization
 - Noun

Event	# class	Train loss	Test loss	Added features	Data size	Thoughts (problem)
R-U-War	20	0.12	0.17	quadratic	2732	Acceptable
Ramadan	20	0.25	0.24	quadratic	1438	Arabic
Easter	10	0.36	0.18	Quadratic, ngram, hinge	746	10class (others)

References

<https://simplemaps.com/data/world-cities>

<https://datahub.io/core/country-list#resource-data>

<https://developer.twitter.com/en/docs/twitter-for-websites/supported-languages>

<https://www.worldstandards.eu/other/tlds/>

<https://emojipedia.org/flags/>