Using Social Media To Enhance Emergency Situation Awareness

SAPIENZA
UNIVERSITÀ DI ROMA

Engineering in Computer Science
Visual analytics, ay 2020/2021
09/06/2021

The Team

- David Guzman -> 1909580
- Hamza Bouzidi -> 1909250

Motivations

Reference paper:

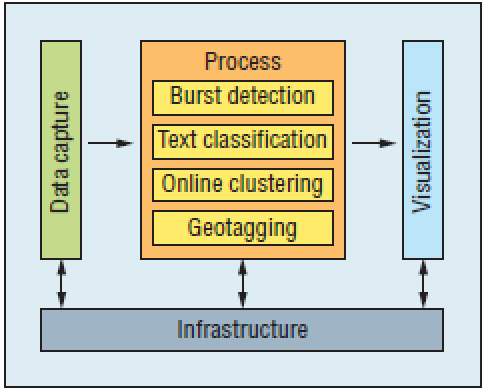
J. Yin, A. Lampert, M. Cameron, B. Robinson and R. Power, "***Using Social Media to Enhance Emergency Situation Awareness***," in IEEE Intelligent Systems, vol. 27, no. 6, pp. 52-59, Nov.-Dec. 2012, doi: 10.1109/MIS.2012.6.

Motivation and goals:

NLP, ML, Data Mining

tweets during crisis →

- + ubiquity
- + communication rapidity
- + cross-platform accessibility
- + real-time nature



→ **knowledge** for emergency response

- ➔ emergency management
- ➔ crisis coordination
- ➔ decision making

> SITUATION AWARENESS

(perception + comprehension + projection)

Online Clustering: Description

Dataset and ground truth:

- .csv extracted from **CrisisNLP**
- human-labeled tweets
- 3000 tweets related to natural disasters
- Earthquake, Hurricane, Volcano, MERS, Typhon, Cyclone, Airplane disaster

Assumptions and simplifications:

- No time distance between tweets.
- No prefiltering of unimportant tweets (No usage of burst-detection)

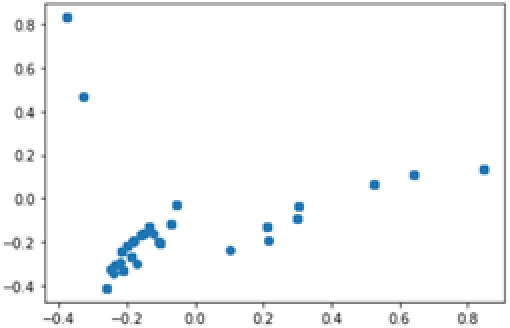
About the model:

- Text preprocessing (tokenization, stop words, remove of frequent words, etc)
- Tf-idf representation of the tweets using *tf-idf vectorizer*
- Online incremental clustering
- Similarity measures: *cosine* and *jaccard coefficient*

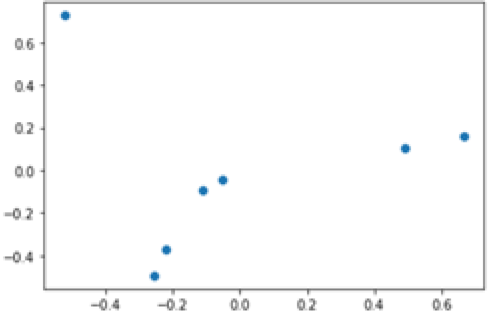
Evaluation metrics:

- *Clustering quality* through comparison with offline clustering
- We know in advance the number of labels on dataset
- Clustering quality using the *Silhouette score*

Online Clustering: results

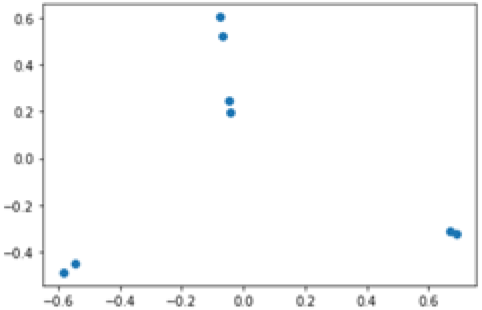


Tweets



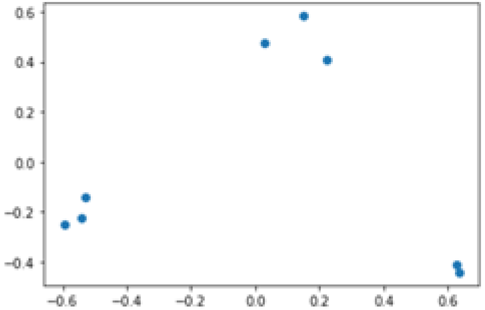
K-means

Silhouette score = 0.75



Cosine

Silhouette score = 0.59

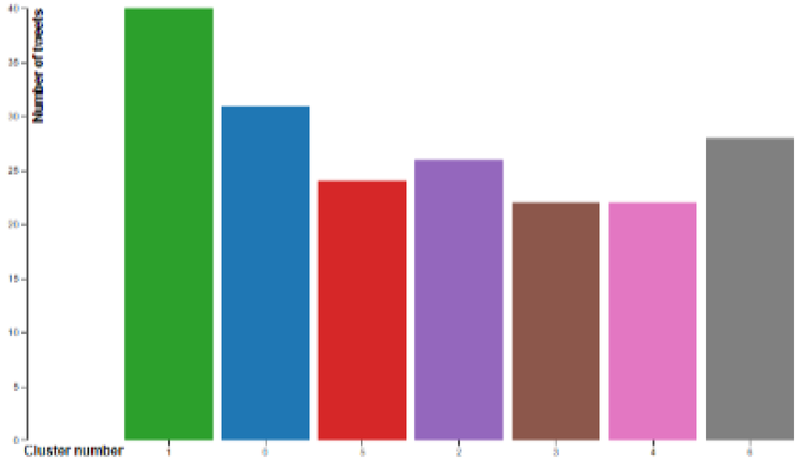


Jaccard

Silhouette score = 0.64

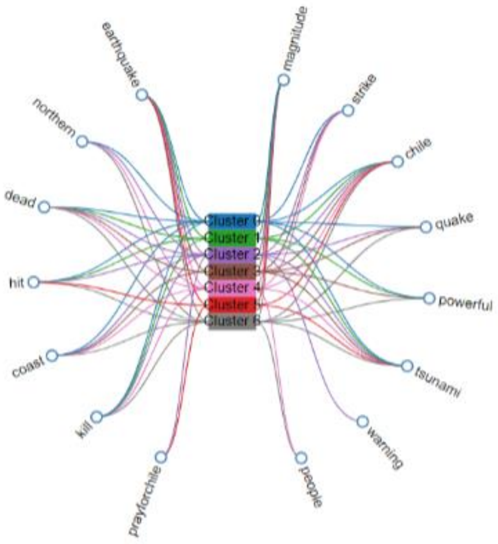
Visualizations: bar plot

- Visualized using vertical bars
- Clear way to reveal tendencies
- Multiple volumes demonstrate differences between each bar
- The bars identify clusters with most relevant information
- shows the relationship between a numeric and a categoric variable.



Visualizations: concept map

- Visualize meaningful relationships among clusters
- Each node in the map contains an important keyword.
- Bidirectional flow of information (concepts <->cluster)

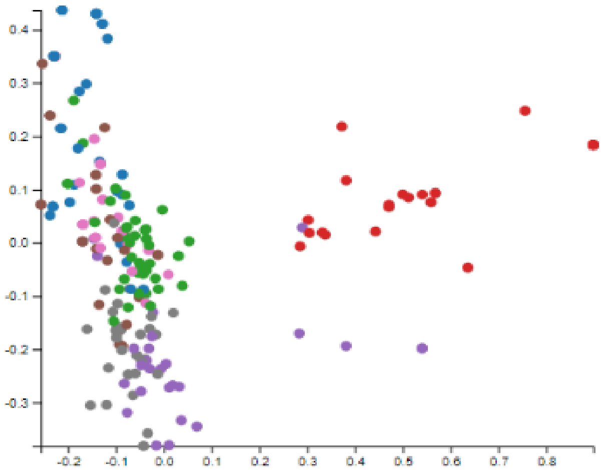


Concept Map: How it was created

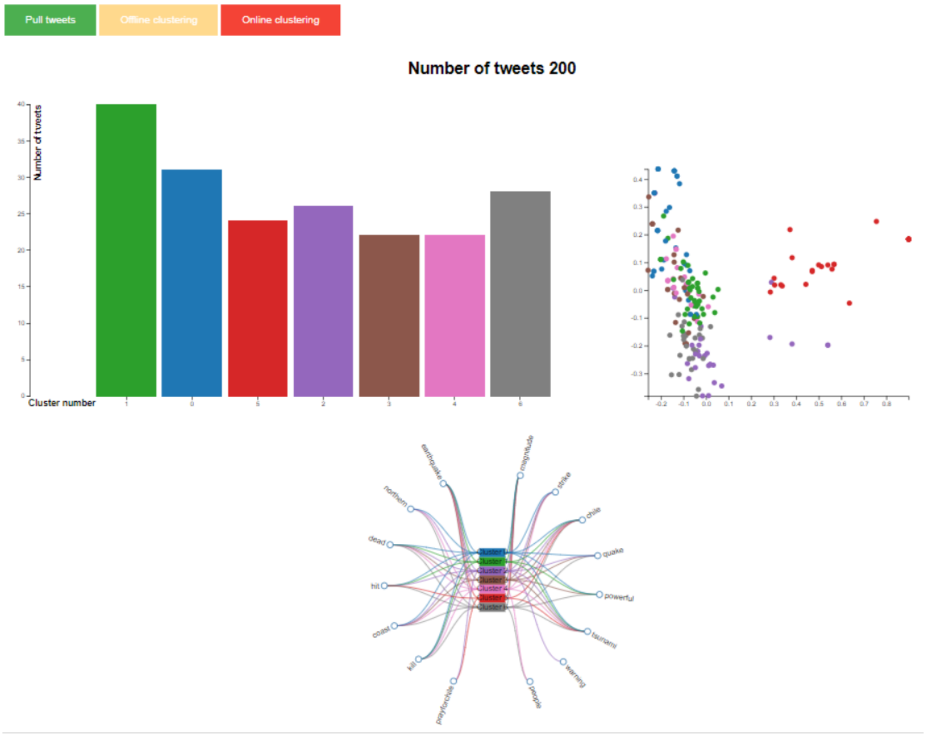
- Select a drawing medium
- Establish a main concept (in this case the total categories of clusters)
- Identify related concepts (important keywords related to the clusters)
- Organize shapes and lines
- Fine-tune the map

Visualizations: 2D scatter plot

- Comparison on behaviour between the two clusters
- Each cluster is plotted using a different color.
- PCA projection on a 2D space
- Clean and stylish option for a reduced number of labels.



Visualizations: final interface



Conclusions

- ➔ The TF-IDF representation must be used and tuned carefully in order to not remove relevant information.
- ➔ Quality of clusters is a key aspect when evaluating the *efficiency* of clustering algorithm and also a research field.
- ➔ We found concept map as one of the key on the information visualization
- ➔ PCA was a good tool for a comparison between the two algorithms used in this work.
- ➔ A concept map helps to illustrate a set of meaningful propositions about a topic, in our case it showed the relationships between the most important keywords of the dataset, and the clusters they belong to.

Future works and improvements

- ➔ Consider the semantic of hashtags: split hashtags considering as features both the hashtag as it is, and splitted
  (ex: `#PrayForChile` generates the features ‘`prayforchile`’, ‘`pray`’, ‘`for`’, ‘`chile`’, you’ll increase the vocabulary size, but you’ll get more information)
- ➔ Consider also photos posted with the tweets for visualization and related
- ➔ Include time distance for clustering tweets and visualization
- ➔ Combine burst detection with clustering algorithm for event detection and grouping

Thanks

for your

attention!