Artificial intelligence integration to crisis management. Part III. Event Detection and Uncertainty

Matthieu François

GreenAI U.P.P.A. x Altanoveo

November 21, 2022

Glossary

- **BNN** Bayesian Neural Network
- **BT** Breaking Ties
- **KL** Kullback-Leibler
- **NN** Neural Network
- **Oracle** Human annotator for Active Learning

1. Main Thread
2. Crisis & Information flow
3. Active Learning & Event Detection
4. Applications
5. Future Work

Uncertainty

Can a model be not self-confident?

MedIA Application

- Alpha deployment of MedIA
- Every Cwall's windows with sound can be transcribed

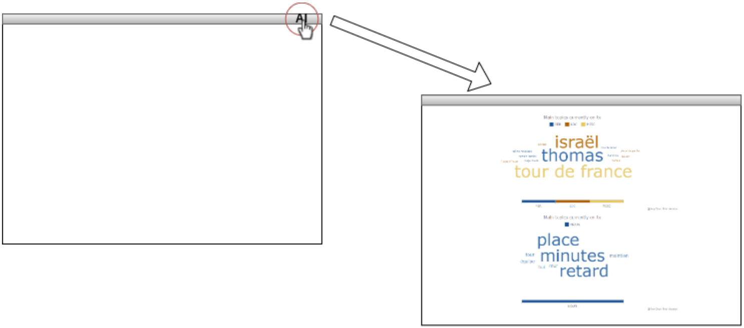


Figure 1: MedIA schema

Data flow and event detection

- Deal with big and miscellaneous data flow
- To have a model that can be quickly adapted to any crisis and new events

A Solution: Active Learning & Event Detection
Event Detection to detect and suggest new crisis related to the chosen one. Active Learning to improve classification results.

Basics

Possibilities

- Use **supervised / semi-supervised** classification with as small as possible dataset annotations to detect a defined subject
- Use **unsupervised** learning to detect new trend and suggest improvement for the classification

Active Learning for Text Classification

Aim reminder: *To have a model that can be quickly adapted to any crisis and new events.*
Problem: Model won't be accurate on new event during a crisis. It could be hard to improve by itself.
Introduce "Human in the loop" with Active Learning.
Query strategies are required **to reduce amount** and **identify data to annotate**.

Query strategies

Random strategy

Select random unlabelled data to annotate

Query strategies

Random strategy

Select random unlabelled data to annotate

Model strategy

A common used strategy is for Bayesian model and the weight distribution study. But we're working with pre-trained tranformers.

Query strategies

Random strategy

Select random unlabelled data to annotate

Model strategy

A common used strategy is for Bayesian model and the weight distribution study. But we're working with pre-trained tranformers.

Data strategy

Example: CoreSet Approach [Sener and Savarese, 2017], select homogeneous data batch regarding their position in a latent space (usually a layer of the NN)

Uncertainty Strategies

With uncertainty based strategies

- Epistemic Uncertainty: not enough information contained in train dataset
- Random Uncertainty: stochastic data

Uncertainty metrics

Some famous metrics

- Entropy [Roy and McCallum, 2001] [Gal and Ghahramani, 2016]
- Breaking Ties (BT) [Schröder et al., 2022]
- Least Confidence [Schröder et al., 2022]
- Contrastive Active Learning (CAL) [Schröder et al., 2022]
- Softmax variance using Dropout [Gal and Ghahramani, 2016]

Entropy

The more the Shanon Entropy is high the more the weight distribution is uniform.

Shanon Entropy

$$H_b(X) = -E[log_b P(X)] = \sum_{i=1}^{n} P_i log_b(\frac{1}{P_i}) = -\sum_{i=1}^{n} P_i log_b P_i \quad (1)$$

Breaking Ties

Predictions with the smallest difference between the 2 highest probabilities are considered the most uncertain.

BT metric

$$BT = argmin_{x_i}[P(yi = k_1^*|x_i) - P(y_i = k_2^*|x_i)] \tag{2}$$

Contrastive Active Learning

Select instances with the maximum mean Kullback-Leibler between prediction distribution and m nearest neighbors

CAL metric

$$CAL = argmax_{x_i}[\frac{1}{m}\sum_{j=1}^{m} KL(P(y_i|x_j^{knn})||P(y_i|x_i))] \quad (3)$$

Results

First model trained with 25 instances. Followed by 20 active learning iteration regarding uncertainty metric and label it to retrain the model using all data labeled so far.

Dataset	Model	Strategy	Acc.	Data Use
AGN	BERT	BT	0.904	0.4%
	BERT	passive (ours)	0.946	100.00%
	XLNet[1]	passive	0.955	100.00%
CR	BERT	LC	0.919	15.45%
	BERT	passive (ours)	0.925	100.00%
	HAC[2]	passive	0.889	100.00%
MR	BERT	PE, BT	0.857	0.547%
	BERT	passive (ours)	0.893	100.00%
	SimCSE[3]	passive	0.884	100.00%
SUBJ	BERT	LC	0.958	5.83%
	BERT	passive (ours)	0.969	100.00%
	AdaSent[4]	passive	0.955	100.00%
TREC-6	BERT	CA	0.968	9.55%
	BERT	passive (ours)	0.958	100.00%
	RCNN[5]	passive	0.962	100.00%

Figure 2: Best final accuracy compared to passive one [Schröder et al., 2022]

Softmax Variance using Dropout

Apply several random dropout on test set to evaluate uncertainty using softmax output variance.

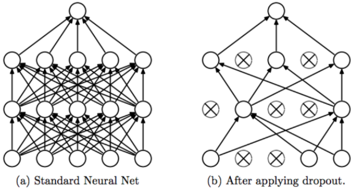
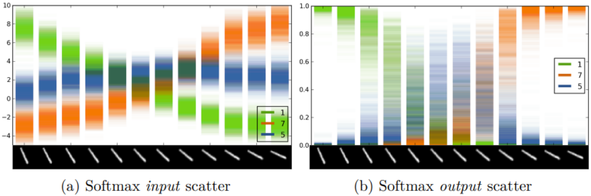


Figure 3: Dropout illustration on DNN



Event Detection

Definition

The goal of event detection is to find the first story discussing an event and then categorizing stories into topics [Rasouli et al., 2020]

- **Event** is a happening, which occurs at a specific time and place
- **Story** is a news article that delivers information about the occurrence of an event. Each document in an online news website is considered as a story.
- **Topic** is a set of news stories which are strongly relevant to an event. [Allan, 2002], [Yang et al., 1999]

Graph based detection

Common practices: First context filtering followed by co-occurrence graph on emerging and important words. [Katragadda et al., 2017]
An emerging word is $KL_t(word) > \alpha KL_{t-k}(word)$ with α fixed
An important word is $KL_t(word) > \beta$
With

$$KL = plog_2(\frac{p}{q}) \tag{4}$$

Graph based detection

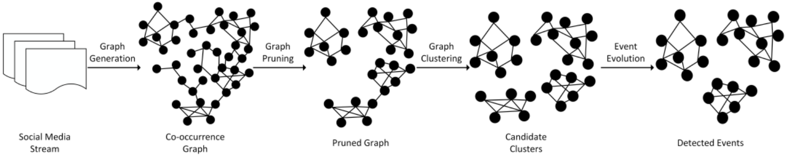


Figure 5: Workflow to detect events [Katragadda et al., 2017]

Difficulties

- Fact checking
- Social Network data format diversity

Context

To implement the uncertainty measure in an active learning framework we decided to use the **BLOOM** pre-trained model from BigScience. The choice of this transform was motivated by a small but efficient finetuning on pre-trained transformers models and multi-languages.

What is it?

> *BigScience is an open collaboration boot-strapped by HuggingFace, GENCI and IDRIS, and organised as a research workshop. This research workshop gathers academic, industrial and independent researchers from many affiliations and whose research interests span many fields of research across AI, NLP, social sciences, legal, ethics and public policy.*

Bloom: from **560m** to **176B** parameters. Similar to GPT3 (auto-regressive model for next token prediction), but it has been trained on 46 different languages and 13 programming languages.

Data Presentation

A french microblogging dataset about ecological crisis. [Kozlowski et al., 2020]

- 3 levels of annotation: relatedness, urgency and intentions to act.
- annotate data from 24 hours before the crisis to 72h after crisis
- scrapping using twitter api with keyword regex

Trained on **12,000 tweets** with **84% F1** score on binary labels using FlauBERT [Le et al., 2019]

Limits: Today, only **7,000 tweets** are still on the plateform.

Data exploration

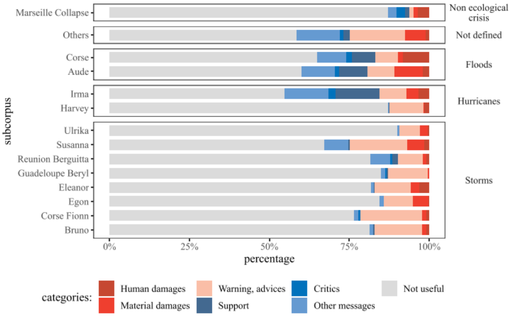


Figure 6: french ecological crisis dataset

Processing

We used the smallest (560m parameters) **BLOOM** version for binary classification.

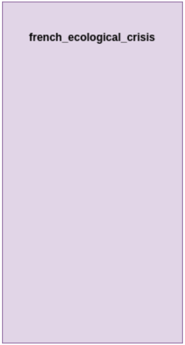


Figure 7: Training process

Processing

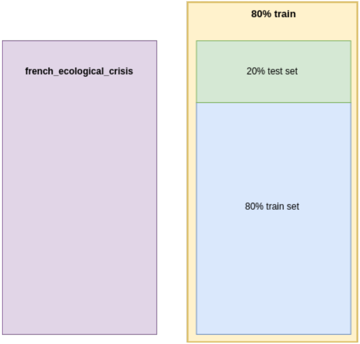


Figure 8: Training process

Processing

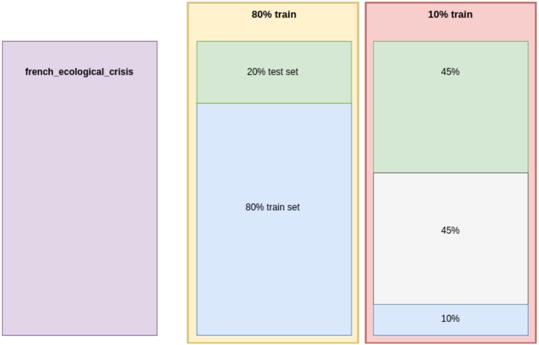


Figure 9: Training process

Processing

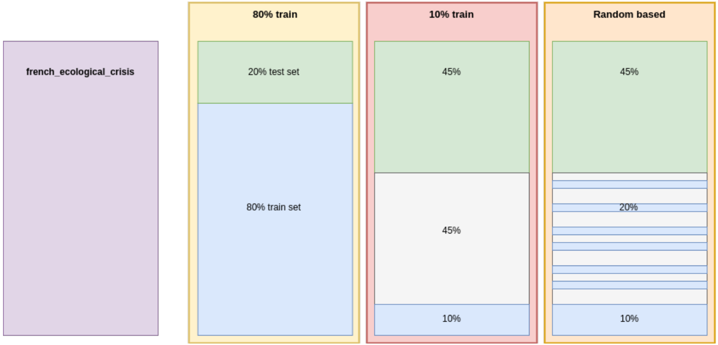


Figure 10: Training process

Processing

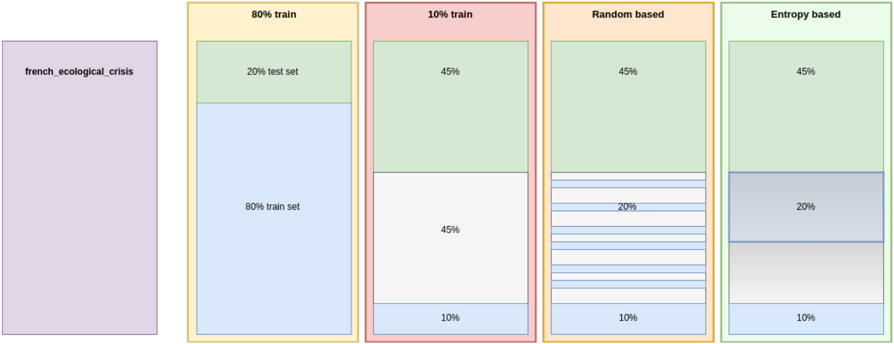


Figure 11: Training process

Some Results

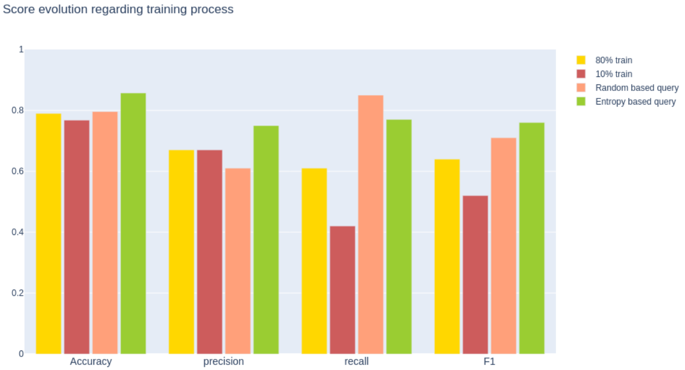


Figure 12: Score evolution on french_ecological_crisis dataset

Surprising data

Some tweets have strange annotations:

Figure 13: Susana annotated tweet

This tweet is considered as Useful and Urgent tweet regarding Susana storm.

Bloom Consumption

Estimating the carbon footprint of BLOOM, a 176b parameter language model [Luccioni et al., 2022]

- 118 days, 5 hours, 41 min training
- Consumption measure with CodeCarbon [Lacoste et al., 2019]
- 25t CO2 (vs 500t for GPT3) (50t CO2 in total)
- 433 MWh (vs 1,287MWh for GPT3)

1. Main Thread
2. Crisis & Information flow
3. Active Learning & Event Detection
4. Applications
5. Future Work

Future Work

- Continue evaluating metrics on bloom and compared to FlauBERT / mBERT / mGPT
- Data augmentation to avoid regex scraping biais?
- Consider diversity / uncertainty optimisation
- Consider model consumption

Thanks!

References I

[Allan, 2002] Allan, J. (2002).
Topic detection and tracking: event-based information organization, volume 12.
Springer Science & Business Media.

[Gal and Ghahramani, 2016] Gal, Y. and Ghahramani, Z. (2016).
Dropout as a bayesian approximation: Representing model uncertainty in deep learning.
pages 1050–1059.

[Katragadda et al., 2017] Katragadda, S., Benton, R., and Raghavan, V. (2017).
Framework for real-time event detection using multiple social media sources.

References II

[Kozlowski et al., 2020] Kozlowski, D., Lannelongue, E., Saudemont, F., Benamara, F., Mari, A., Moriceau, V., and Boumadane, A. (2020).
A three-level classification of french tweets in ecological crises.
Information Processing & Management, 57(5):102284.

[Lacoste et al., 2019] Lacoste, A., Luccioni, A., Schmidt, V., and Dandres, T. (2019).
Quantifying the carbon emissions of machine learning.
arXiv preprint arXiv:1910.09700.

References III

[Le et al., 2019] Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L., and Schwab, D. (2019).
Flaubert: Unsupervised language model pre-training for french.
arXiv preprint arXiv:1912.05372.

[Luccioni et al., 2022] Luccioni, A. S., Viguier, S., and Ligozat, A.-L. (2022).
Estimating the carbon footprint of bloom, a 176b parameter language model.
arXiv preprint arXiv:2211.02001.

References IV

[Rasouli et al., 2020] Rasouli, E., Zarifzadeh, S., and Rafsanjani, A. J. (2020).
Webkey: a graph-based method for event detection in web news.
Journal of Intelligent Information Systems, 54(3):585–604.

[Roy and McCallum, 2001] Roy, N. and McCallum, A. (2001).
Toward optimal active learning through sampling estimation of error reduction. int. conf. on machine learning.

[Schröder et al., 2022] Schröder, C., Niekler, A., and Potthast, M. (2022).
Revisiting uncertainty-based query strategies for active learning with transformers.
pages 2194–2203.

References V

[Sener and Savarese, 2017] Sener, O. and Savarese, S. (2017).
Active learning for convolutional neural networks: A core-set approach.
arXiv preprint arXiv:1708.00489.

[Yang et al., 1999] Yang, Y., Carbonell, J. G., Brown, R. D., Pierce, T., Archibald, B. T., and Liu, X. (1999).
Learning approaches for detecting and tracking news events.
IEEE Intelligent Systems and their Applications, 14(4):32–43.