Using Multimodal Data and AI to Dynamically Map Flood Risk

Lydia Bryan-Smith

University of Hull
Hull, United Kingdom, HU6 7RX
L.Bryan-Smith-2014@hull.ac.uk

Abstract

Classical measurements and modelling that underpin present flood warning and alert systems are based on fixed and spatially restricted static sensor networks. Computationally expensive physics-based simulations are often used that can't react in real-time to changes in environmental conditions. We want to explore contemporary artificial intelligence (AI) for predicting flood risk in real time by using a diverse range of data sources. By combining heterogeneous data sources, we aim to nowcast rapidly changing flood conditions and gain a greater understanding of urgent humanitarian needs.

Introduction

Flood prediction has traditionally been done with physics based models which can take a while to run (Declan Valters 2018) and are computationally expensive (Teng et al. 2017), yielding predictions based on outdated data.

Existing solutions to this problem usually rely on a single data source such as river levels (Le et al. 2019; Mioc et al. 2011) and drainage system sensors (Keung et al. 2018). Such systems predict sensor values in the future and rarely use these predictions as an input to a physics-based model, or indirectly identify which areas are most badly affected.

Other studies use both text and images from twitter (Said et al. 2020), but fail to account for additional data sources such as rainfall radar information. Those studies that do often predict rainfall in advance instead of something more direct such as water depths or flood risk (Agrawal et al. 2019).

To tackle these issues, an AI based approach is explored, with the goal of making use of a diverse range of data sources to inform real-time responses from emergency services to a flood situation.

This project looks at using rainfall radar data, a heightmap, an autoencoder, and a traditional flood-based simulation model (Declan Valters 2018) to train an AI to predict water depths. It also looks at analysing social media posts to see if flood-related tweets, including text and images, can be used further to aid information on the dynamically evolving flooding situation - e.g. visual or textual cues to water depth, suddenness of a flood or humanitarian situation in different areas.
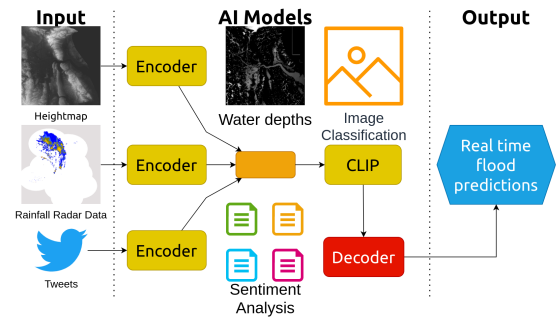
Figure 1: The architecture of the final system being developed (final model architecture subject to change).

The third and final aspect of the project is combining the above heterogeneous data sources to make a more robust single unified prediction of near-term flood risk, in that uncertainties in 1 modality can be compensated for by another.

Rainfall Radar for Flood Prediction

Taking rainfall radar data from the Met Office and a heightmap from the Ordnance Survey, a Convolutional Neural Network (CNN)-based autoencoder with multiple inputs is being developed to predict water depths in 2 dimensions. The model design was originally inspired by U-Nets (Ronneberger, Fischer, and Brox 2015) that were employed to predict rainfall radar in 2D (Agrawal et al. 2019), although since then the model has undergone significant revisions which were influenced by (Silberer and Lapata 2014).

To generate labels to train an autoencoder, the data is fed first through the traditional physics based model HAIL-CAESAR (Declan Valters 2018) (which is a headless implementation of CAESAR-Lisflood (Coulthard et al. 2013)), the output of which is used to generate labels to train the autoencoder. Generated labels are a 2D array of binary values - i.e. water / no water.

Social Media for Flood Severity Analysis

It has been observed that social media users are communicating for a range of purposes during disasters, from the status of public infrastructure to crowdsourcing volunteers

(Kankanamge et al. 2020). For this reason, tweets from twitter are also being analysed as part of the project. Example questions being asked are "Can flood severity be inferred from social media posts?" and "Can we classify images by the sentiment of the associated tweet?".

Answering such questions could help in understanding humanitarian needs in different locations. By identifying where is most badly affected, information about the help that is most urgently needed there can be used to direct available resources. Water depths could also be estimated here.

To answer these questions, posts from twitter were downloaded from various hashtags including more general ones such as `#floods` and `#flashfloods` - as well as hashtags for more specific flooding events, like `#StormChristoph`, and `#HuricaneEta`. A transformer encoder has been implemented to predict the sentiment (positive / negative) of the tweets downloaded - with the labels calculated from the emojis in the tweet text.

Using this model as a starting point, the above questions will be answered.

Fusion Model

It is planned to investigate fusion model approaches to handle the disparate data sources mentioned above in a single system to generate a single unified near-term prediction of flood risk. To do this, the CLIP model (Radford et al. 2021) will be used as a base model. It will then be extended by combining the available disparate data sources.

By geospatially analysing social media posts, tweets and their sentiments can be paired with nowcasted water depth predictions using CLIP to detect communities most immediately at risk or in most urgent need of assistance.

Project Timeline

In the first phase of the project, existing literature was reviewed to identify potential promising approaches, such as rainfall radar data from CEDA and social media from twitter. In the second phase, data was downloaded and parsed. This phase took much longer than expected due to the complex nature of the rainfall radar data.

The third phase began by implementing and training some initial models with mixed results. This phase continues to the time of typing, during which continuous experiments are being run to incrementally improve upon the initial model designs.

In the fourth phase, the disparate datasets mentioned above will be tied together into a single system (described above).

Finally, the project will be written up into the thesis itself.

Future Work

For the rainfall radar autoencoder model, once the implementation of the autoencoder is complete, is is planned to enhance the model by binning water depths (e.g. 0cm to 1cm, 1cm to 50cm, 50cm to 1m, 1m+) instead of predicting a binary label. Adding additional data layers to the model may also improve accuracy (e.g. satellite data).

Building on the work done so far, the available data sources and their associated models will be combined into a single system using CLIP (Radford et al. 2021) with the goal of improving situational awareness in flooding events. Additionally, other disparate data sources (e.g. satellite data) will also be explored.

References

Agrawal, S.; Barrington, L.; Bromberg, C.; Burge, J.; Gazen, C.; and Hickey, J. 2019. Machine Learning for Precipitation Nowcasting from Radar Images. *arXiv preprint arXiv:1912.12132*.

Coulthard, T. J.; Neal, J. C.; Bates, P. D.; Ramirez, J.; de Almeida, G. A. M.; and Hancock, G. R. 2013. Integrating the LISFLOOD-FP 2D hydrodynamic model with the CAESAR model: implications for modelling landscape evolution. *Earth Surface Processes and Landforms*, 38(15): 1897–1906.

Declan Valters. 2018. GitHub - dvalters/HAIL-CAESAR: The High-Performance Architecture-Independent LISFLOOD-CAESAR model of floodplain, river, and sediment dynamics. Available online: https://github.com/dvalters/HAIL-CAESAR [Accessed 07/04/2020].

Kankanamge, N.; Yigitcanlar, T.; Goonetilleke, A.; and Kamruzzaman, M. 2020. Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets. *International journal of disaster risk reduction*, 42: 101360.

Keung, K. L.; Lee, C. K. M.; Ng, K. K. H.; and Yeung, C. K. 2018. Smart City Application and Analysis: Real-time Urban Drainage Monitoring by IoT Sensors: A Case Study of Hong Kong. In *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 521–525.

Le, X.-H.; Ho, H. V.; Lee, G.; and Jung, S. 2019. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water*, 11(7): 1387.

Mioc, D.; Anton, F.; Nickerson, B.; Santos, M.; Adda, P.; Tienaah, T.; Ahmad, A.; Mezouaghi, M.; MacGillivray, E.; Morton, A.; and Tang, P. 2011. Flood Progression Modelling and Impact Analysis. In *Efficient Decision Support Systems - Practice and Challenges in Multidisciplinary Domains*, chapter 12. InTech.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *ICML 2021: 38th International Conference on Machine Learning*, 8748–8763.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Said, N.; Ahmad, K.; Gul, A.; Ahmad, N.; and Al-Fuqaha, A. I. 2020. Floods Detection in Twitter Text and Images. *MediaEval*.

Silberer, C.; and Lapata, M. 2014. Learning Grounded Meaning Representations with Autoencoders. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 721–732.

Teng, J.; Jakeman, A.; Vaze, J.; Croke, B.; Dutta, D.; and Kim, S. 2017. Flood inundation modelling. *Environmental Modelling and Software*, 90(90): 201–216.