

Department of Computer Science
University of Hong Kong
Final Year Project

Project Plan

Author: GHANI Saad Muhammad Abdul
Date: September 29, 2019
Issue: 1.0

Table of Contents

| | |
|---|----------|
| 1. Project Background | 3 |
| 1.1. <i>Red tides: Causes and Effects</i> | 3 |
| 1.2. <i>Current detection systems</i> | 3 |
| 1.3. <i>Motivations</i> | 3 |
| 2. Project Objective | 4 |
| 2.1. <i>Classification of harmful algal</i> | 4 |
| 2.1.1. Testing new techniques to tackle class imbalance | 4 |
| 2.2. <i>Time-Series Analysis and Forecasting</i> | 4 |
| 3. Project Methodology | 4 |
| 3.1. <i>Dataset</i> | 4 |
| 3.2. <i>Classification</i> | 5 |
| 3.2.1. Convolutional Neural Network..... | 5 |
| 3.2.2. Transfer Learning | 5 |
| 3.3. <i>Time-Series Analysis and Forecasting</i> | 5 |
| 3.3.1. Long Short Term Memory Recurrent Neural Network..... | 5 |
| 4. Project Schedule and Milestones | 6 |
| 5. References | 7 |

1. Project Background

1.1. Red tides: Causes and Effects

Red Tides are the discoloration of water when there is rapid growth of colonies of algae. They are naturally occurring but human activities seem to increase their frequency and intensity. From discharge of nutrient-rich fertilizers to global warming, factors of nutrition, salinity and temperature all come into play and can be traced back to human activities.

Most red tides are harmless but some are toxic and are harmful to people, fish, shellfish, marine animals and birds. These red tides are known as Harmful Algal Blooms (HAB)[14] caused by harmful algae species. Though Red Tides or Algal Blooms are a common occurrence, HAB occur far less frequently[14]. Nonetheless, when HABs do occur they can potentially cause severe losses to the aquaculture industry locally and throughout the world[15].

1.2. Current detection systems

Traditionally, lab based methods would be used to discriminate algae species and detect red tides [16]. However, with advancing technology, off-site, time-consuming methods are being replaced by real-time, on-site techniques using miniature sensors and cameras. Classification of Algae species are being done mainly by on-board image processors which can store its predictions and can then later be fine tuned. Detecting Red Tides are done on a large scale by remote sensing of the surface pigment, reflectance or temperature. Satellite Imaging techniques such as CZCS[16] and SeaWiFS [16] is also used.

1.3. Motivations

Previous Early Detection Systems use indicators of the root cause of the problem – rapid growth of algae. This project aims to analyse the growth of particular algae species using microscopic images. Due to the high impact of Red Tides in the socio-economic sectors of our society, an early detection system which directly analysis the growth of algae would bring about the fastest results in warning us of a potential HAB.

2. Project Objective

2.1. Classification of harmful algal

The first objective of this project is to make a computer vision system which will detect and classify harmful algal species in microscopic images.

2.1.1. Testing new techniques to tackle class imbalance

Additionally, this project aims to test new techniques to tackle class imbalance. The nature of Algae databases are such that there is a large class imbalance. In one such database, 3 classes make up approximately 85% of all 3 million images.

2.2. Time-Series Analysis and Forecasting

Once classification is done to a reasonable accuracy, the project will begin analysing the growth of particular HAB causing species. First, trends will be taken out to do a feasibility study to see if forecasting is possible. Then a system will be created to model and forecast possible HAB outbreaks given the growth of particular algal species.

3. Project Methodology

3.1. Dataset

Finding the right dataset is a challenge as there are two domains of data – the list of harmful algal species and the dataset containing the said species and more.

The resources looked into were:

- IOC UNESCO Taxonomic Reference List [5]
- AFCD's HK Red Tide Information Network [13]

There are numerous public imaging datasets available which give annotated plankton images. The ones looked into are:

- EcoTaxa [1]
- COPEPOD Database [2]
- ICES Plankton Database [3]
- PlanktonNet [4]
- WHOI-Plankton Database [6]

The WHOI-Plankton Database was selected as it contained 14 harmful algal species listed in the above two resources.

3.2. Classification

3.2.1. Convolutional Neural Network

Convolutional Neural Networks [7] (CNNs) combine the benefit of convolution in image processing with the predictive powers and optimization techniques neural networks have to offer. A CNN is able to train and optimize its kernels to solve the problem at hand. Thus, the need for feature engineering has been removed which greatly simplifies the process. Furthermore, subsequent downsizing makes it easier for the CNN to train itself as it gets rid of much of the irrelevant data present in the picture.

3.2.2. Transfer Learning

Transfer Learning has been widely used on multiple domains such as NLP, sentiment analysis and image classification [11]. Current state-of-the-art plankton image classification models use Transfer Learning with a slight change in architecture along with feature engineering [12]. This project will make use of Transfer Learning from multiple domains of planktons to classify microscopic plankton images.

3.3. Time-Series Analysis and Forecasting

3.3.1. Long Short Term Memory Recurrent Neural Network

Recurrent Neural Networks (RNNs) are widely used on time-series data [9] to make predictions about the future given a certain trend of data. Furthermore, Long Short Term Memory (LSTM) have improved the accuracy and training time of RNNs [8]. LSTMs have also been used in the domain of time-series data and have outperformed all other techniques [10]. Thus, this project will make use of LSTMs to forecast red tides.

4. Project Schedule and Milestones

| Time | Milestone |
|-------------|---|
| 2018 | |
| September | <ul style="list-style-type: none"> • Research on project - background • Research on existing solutions • 29th – Deliverables of Phase 1 |
| October | <ul style="list-style-type: none"> • Research on other datasets • Analysis of possibilities of time-series data |
| November | <ul style="list-style-type: none"> • Improving accuracy of classification • Forecasting using Time-Series |
| December | <ul style="list-style-type: none"> • Evaluation of implementation • Iteration and improvement of results |
| 2019 | |
| January | <ul style="list-style-type: none"> • Exploration of alternate approaches • 7-11 – Project presentation |
| February | <ul style="list-style-type: none"> • Implementation of alternate approaches • 2 – Deliverables of Phase 2 |
| March | <ul style="list-style-type: none"> • Evaluation of alternate implementation • Testing and experimentation |
| April | <ul style="list-style-type: none"> • Final cleanup • Report writing • 19 – Deliverables of Phase 3 • 20-24 – Final Presentation |
| May | <ul style="list-style-type: none"> • 5 – Project exhibition |

5. References

1. Picheral M, Colin S, Irisson J-O (2017). EcoTaxa, a tool for the taxonomic classification of images. <http://ecotaxa.obs-vlfr.fr>.
2. <https://www.st.nmfs.noaa.gov/copepod/about/databases.html>
3. <http://www.ices.dk/marine-data/dataset-collections/Pages/Plankton.aspx>
4. <https://planktonnet.awi.de/#search>
5. <https://www.whoi.edu/website/redtide/species/by-name/>
6. Orenstein, E. C., Beijbom, O., Peacock, E. E., & Sosik, H. M. (2015). Whoi-plankton-a large scale fine grained visual recognition benchmark dataset for plankton classification. *arXiv preprint arXiv:1510.00745*.
7. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105)
8. Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.
9. Hüsken, M., & Stagge, P. (2003). Recurrent neural networks for time series classification. *Neurocomputing*, 50, 223-235.
10. F. Karim, S. Majumdar, H. Darabi and S. Chen, "LSTM Fully Convolutional Networks for Time Series Classification," in *IEEE Access*, vol. 6, pp. 1662-1669, 2018. doi: 10.1109/ACCESS.2017.2779939
11. Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359.
12. Dai, J., Yu, Z., Zheng, H., Zheng, B., & Wang, N. (2016, November). A hybrid convolutional neural network for plankton classification. In *Asian Conference on Computer Vision* (pp. 102-114). Springer, Cham.
13. <https://www.afcd.gov.hk/english/fisheries/hkredtide/database/database.html>
14. <https://www.afcd.gov.hk/english/fisheries/hkredtide/redtide.html>
15. <https://www.afcd.gov.hk/english/fisheries/hkredtide/redtide/red04.html>
16. Sellner, K. G., Doucette, G. J., & Kirkpatrick, G. J. (2003). Harmful algal blooms: causes, impacts and detection. *Journal of Industrial Microbiology and Biotechnology*, 30(7), 383-406.