

Large Scale Gaussian Processes in Benthic Habitat Mapping

Research Methods - INFO5993 Assignment 3

Justin Ting, 430203826

April 2016

1 Abstract

(0.5 column = 0.25 page)

Earth's oceans cover 70% of its surface, but only less than 10% of the Earth's oceans have been explored to date¹. There have been increasing efforts over the past few decades to map out these unexplored areas to monitor marine ecosystems to be able to track the state of them over time for management, preservation, etc. purposes. The process used is called benthic habitat mapping, which is the predicting of what exists at the bottom of a body of water. Most recent studies looking to create benthic habitat maps share some basic key steps - acoustic data is used to estimate properties about the surface of the water, which are then mapped to, using machine learning algorithms, *in situ* data such as still images, videos, or samples of the area in question. It is the relationship which is inferred between the different data sets inferred using machine learning techniques that varies between studies.

2 Introduction

(1 page)

3 Problem Statement

(1 columns = 0.5 page)

Much of the research done in benthic habitat mapping generates deterministic maps using untuned (as-is) machine learning techniques and implementations. This is potentially an issue (although the same can be said of most problems using such machine learning methods) because the resultant maps make broad conclusions and assumptions across the different features used to create them, such as the direct relationship between the topography of an area of benthos and what resides there (CITE), among others which are contested between different studies. It is arguably impossible to decisively confirm or deny these relationships without actually exploring a majority of the entire ocean - the expeditions for which would be prohibitively expensive.

In the absence of such confirmation, we want to both quantify the uncertainty of the maps generated and be able to do so without explicitly assuming relationships between data features. Using a method that achieves this and that can state the certainty of its predictions, it would allow augmentation of experts' decision making to be better informed, and allowing management bodies to more efficiently use their resources in a more focussed manner. Moreover, deterministic methods have generally produced mixed results, while the more state of the art techniques have been limited in the data they could use due to computational constraints, pointing towards a need to be able to bring more complex techniques together with all available data to create more accurate maps than have been previously possible. This would allow us to better monitor, assess ma-

¹Oceanservice.noaa.gov. (2016). How much of the ocean have we explored?. [online] Available at: <http://oceanservice.noaa.gov/facts/exploration.html>

rine habitats, and be able to assess human impact which will allow decisions to be made on steps to take in terms of preservation, etc.

4 Algorithm/Solution

(1 page)

To attempt to improve this problem, we use Gaussian processes for the classification process - and as a result, obtain a distribution of probable habitats for a given area rather than a deterministic result for a given space, and can thus state our certainty of a particular classification for an area of the benthos. Moreover, because the high resolution data used were images collected using automated vehicles that were then pre-clustered in Steinberg et al. (2011) using Variational Dirichlet Processes, the classifications themselves are also represented as continuous probabilities.

However, there is an inherent limitation in the use of Gaussian processes in that its computational complexity is $O(n^3)$ - meaning that for our relatively 'small' training dataset of 16,000, naive use of GPs are already computationally infeasible - let alone our unclassified dataset of 7 million (for which we wish to create a habitat map for) that only possesses corresponding low resolution (acoustic) data.

To explain Gaussian processes in simple terms - first consider a single variable that fits a Gaussian distribution, commonly known as the bell curve, or normal distribution. If we generalise this single dimension to infinitely many, where any linear combination of components is normally distributed, then this is what we call a Gaussian Process.

Compared to more commonly used supervised classification algorithms which require modeling around the chosen features, Gaussian processes instead look at the joint distribution across all features. However, they require a matrix inversion step with $O(n^3)$ complexity - this does not scale well beyond at most several thousand datapoints. To overcome this and be able to utilise as much of the available data as possible, we use approximation methods.

While a number of studies have undertaken com-

paring techniques in sparse GP regression, there are fewer for classification, as the stochastic nature of labels inherently steers away from being Gaussian itself. The two differ in implementation whereby in classification, the function that is otherwise used in regression has to pass its predictions through a squashing function to bound the results between (0,1), resulting in a non-Gaussian distribution. A study by Kuss and Rasmussen (2005) found that Expectation Propagation was superior to the previously used Laplace's method for projecting classes back to an approximated Gaussian distribution. It is using Expectation Propagation that we will be able to use approximated sparse GP regression techniques and apply them to classification.

5 Results

(1 page) To compare the performance of our Gaussian Process (GP) classifier using varying sparse approximate GP methods, we want to obtain benchmarks by getting the F-scores and accuracies using deterministic machine learning algorithms. The first table represents the original 24 granular classes, whereas the second one is using the 5 aggregated classes, which are a summarisation of the more specific 24. This was done in collaboration with an expert who grouped the original unsupervised clusterings of habitat classes into similar groups, where the aggregated classes provide an obvious advantage in performance.

24 Habitat classes

Algorithm	F1 score	Accuracy
KNN (5)	0.33278	0.62459
Logistic Regression	0.20285	0.682705
Random Forest	0.20283	0.68258
SVM	0.20284	0.68270

5 Aggregated habitat classes

Algorithm	F1 score	Accuracy
KNN (5)	0.61347	0.749214
Logistic Regression	0.50381	0.814503
Random Forest	0.50932	0.819253
SVM	0.50290	0.818273

These next two set of results are using Gaussian processes with the granular and aggregated habi-

tat classes, respectively, with different methods applied to minimise the original $O(n^3)$ computational complexity as much as possible. With the exception of the subset of data (SoD) benchmark, the rest are wrapped with Expectation Propagation

24 Habitat classes

Approximation method	F1 score	Accuracy
Subset of Data	0.72948	0.84106
SR	0.79258	0.87151
PP	0.81755	0.89162
FIC	0.83291	0.91418

5 Aggregated habitat classes

Approximation method	F1 score	Accuracy
Subset of Data	0.84281	0.91968
SR	0.88954	0.93014
PP	0.91758	0.95417
FIC	0.92856	0.96281

Key:

SR - subset of regressors ; FIC - fully independent conditional ; PP - projected process

6 Analysis, Discussion

Approximated sparse Gaussian Processes were used to classify benthic habitats into habitat classes, and a number of ways in which to perform the approximation were tested. All of them were found to perform considerably better than the more popular and widely used methods of SVMs, Random Forests, Logistic Regression, and kNN. Of the approximation techniques, fully independent conditionaals (FITC) performed the best.

The first thing to take note of is the performance of the algorithms when classifying the aggregated classes, compared to the more granular ones. Although the reason is fairly obvious, it is worth pointing out that this is due to the variational dirichlet processes generating habitat classes that are very similar (in expert opinion) - and once consolidated, better represented the segregation subject matter experts would deem more appropriate.

Another interesting point is that in [Bender et al. \(2012\)](#), only subset of data is used in their probabilistic least target squares classification, similarly using approximated Gaussian Processes, and yet achieved over 98% accuracy. Given the distribution of data and habitat boundaries in the dataset between the two studies, Scott Reef and O'Hara Bluff respectively, the former has less distinct boundaries, with different habitats overlapping with the next quite often, whereas O'Hara Bluff has a wider variety of unique and distinct habitats from their adjacent ones, with clearer boundaries.

Moreover, (**TODO insert here**) has found that different techniques have worked better on certain datasets. Thus, to get a more comprehensive performance review of the techniques explored in this paper, a range of datasets from entirely different habitats need to be analysed in order to get a more holistic picture on how the tested techniques work in the arena of benthic habitat mapping in general.

7 Conclusion

(0.5 column = 0.25 page)

8 References

References

- Asher Bender, Stefan B., Williams, and Oscar Pizarro. Classification with probabilistic targets. 2012.
- Malte Kuss and Carl Edward Rasmussen. Assessing approximate inference for binary gaussian process classification. *Max Planck Institute for Biological Cybernetics*, 38, 2005.
- D. Steinberg, A. Friedman, O. Pizarro, Williams, and S.B. A bayesian nonparametric approach to clustering data from underwater robotic surveys. International Symposium on Robotics Research, 2011.