2016-05-06

Machine Learning in Benthic Habitat Mapping

Machine Learning in Benthic Habitat Mapping

Justin Ting — Honours Student

Simon O'Callaghan — NICTA Researcher

NICTA

SIT

Introduction

- ▶ Less than 10% of the world's oceans are mapped compared
- ▶ To map more of the world's oceans at a high resolution, we employ benthic habitat mapping techniques

to 99% of Earth's topology mapped (low resolution)¹

 Marine habitat mapping cuts across marine biology, geology, hydrography, oceanography, geophysics (Brown et al., 2011), along with habitat mapping

2016-05-06

- 0.5 minutes
- What is benthic habitat mapping?
- Benthic ecological region at the lowest level of a body of water

Introduction

Less than 10% of the world's oceans are mapped compared

to 99% of Earth's topology mapped (low resolution)2 To map more of the world's oceans at a high resolution, we employ benthic habitat mapping techniques Marine habitat mapping cuts across marine biology, geology hydrography, oceanography, geophysics (Brown et al., 2011),

1http://www.wired.com/2009/06/nasa-satellite-maps-99-of-earths-

- benthos things related to the benthic layer • Habitat mapping - based on a small amount of high resolution data,

and a considerably larger amount of lower resolution data, a relationship is created to correlate the data in overlapping regions,

- which is then projected to the regions without the high resolution data to create a 'habitat map' • The high resolution data is generally actual sediment samples or
- organism samples at the benthos • Low resolution data is generally some sort of acoustic data representing basic properties of the seafloor

http://www.wired.com/2009/06/nasa-satellite-maps-99-of-earthstopography/

⁻Introduction

Problem Statement

- ► Much research in benthic habitat mapping generates deterministic maps using as-is machine learning techniques/implementations
- ► We need to be able to monitor marine habitats on a large scale to assess human impact over time to be able to make informed management decisions

2016-05-06

Problem Statement

 Much research in benthic habitat mapping generates deterministic maps using as-is machine learning techniques/implementations

Problem Statement

 We need to be able to monitor marine habitats on a large scale to assess human impact over time to be able to make informed management decisions

1 minute

Solution - Overview

- ► We will use Gaussian Processes to obtain distributions of likely habitat maps
- Probabilistic approach allows us to state certainty about a particular mapped area
- ▶ Use of Gaussian Processes requires a matrix inversion step with $O(n^3)$ complexity attempt to overcome this by making our covariance matrix sparse and hence the inversion step computationally feasible for large datasets

2016-05-06

Solution - Overview

- We will use Gaussian Processes to obtain distributions of likely habitat maps
- Probabilistic approach allows us to state certainty about a particular mapped area

Solution - Overview

 Use of Gaussian Processes requires a matrix inversion step with O(π²) complexity - attempt to overcome this by making our covariance matrix sparse and hence the inversion step computationally feasible for large datasets

2.5 minutes

Solution - Gaussian Processes

TODO - a few slides providing an overview of Gaussian Processes here (may not include in the 10 minute presentation if time doesn't allow - will time once slides have been written out)

2016-05-06

Solution - Gaussian Processes

TODO - a few slides providing an overview of Gaussian Processes here (may not include in the 10 minute presentation if time doesn't

allow - will time once slides have been written out)

Solution - Gaussian Processes

Results

We would like to compare the results obtainable from use of Gaussian Processes in generating habitat maps to those of more naive methods. Note that measurements are taken by averaging over 10-fold cross validations.

This first set of results is using 24 separate habitat classes as the labels.

Algorithm	F1 score	Accuracy
KNN (5)	0.04823	0.14192
Logistic Regression	0.00900	0.17119
Random Forest	0.04960	0.15240
SVM	0.00890	0.17119

2016-05-06

☐ Results

We would like to compare the results obtainable from use of Gaussian Processes in generating habitat maps to those of more naive methods. Note that measurements are taken by averaging

This first set of results is using 24 separate habitat classes as the

orithm	F1 score	Accura
	0.04823	
gistic Regression	0.00900	0.1711
ndom Forest	0.04950	0.1524
M	0.00890	0.1711

score	Accuracy
04823	0.14192
00900	0.17119
04950	0.15240

Results

This second set of results is using a higher level mapping of the habitats created in collaboration with an expert - bringing the number of classes down to 5. The accuracy increases notably when the aggregation of visually matching habitat classes is performed.

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

2016-05-06

Results

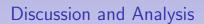
sults

This second set of results is using a higher level mapping of the habitats created in collaboration with an expert - bringing the number of classes down to 5. The accuracy increases notably when the aggregation of visually matching habitat classes is

lgorithm	F1 score
NN (5)	0.33278
ogistic Regression	0.20285
andom Forest	0.20283
VM	0.20284

re Accuracy 78 0.62459 85 0.682705 83 0.68258 94 0.68270 These final two set of results are using G

2 minutes for all results



2016-05-06

__Discussion and Analysis

• stuff

Discussion and Analysis

▶ stuff

Bibiolgraphy

Craig Brown, Stephen J Smith, and Peter Lawton. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine, Coastal and Shelf Science*, 92, 2011.

90-92-0-Bibiolgraphy

Craig Brown, Stephen J Smith, and Peter Lawton. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the sealloor using acoustic techniques. Estuarine, Coastal and Shelf Science, 92, 2011.

Bibiolgraphy

2 minutes