

Big Data Meets Big Water: Analytics of the AIS Ship Tracking Data

Stan Matwin

Institute for Big Data Analytics

Dalhousie University

Halifax, NS, Canada B3H 1W5 Canada

Email: stan@cs.dal.ca

IN THIS presentation we will argue that Big Data technologies can contribute in an important way to an unprecedented breakthrough in the understanding of oceans as a factor in climate change, in transportation, and in supplying humanity with its important food component.

Oceans cover almost 70% of the surface of the earth, and supply at least 15% of animal protein intake for 4.5 billion people. At the same time, ocean are an area of human interest that undergoes currently a massive infusion of information technology. As a result, ocean data and its challenges will become a fertile ground for data science.

After briefly introducing Big Data, we will argue that many of the emerging data sources focused on oceans are Big Data. We will use the global Automatic Identification System as an example. We will introduce the AIS system and characterize it quantitatively. We will then illustrate some of the Big Data projects under way in the Institute for Big Data Analytics, Dalhousie University. In particular, we will focus on the analysis of fishing ship trajectories available through AIS data, and will show how this analysis can lead in the future to

unprecedented quality of estimates of the fish intake by global fisheries.

We will discuss AIS data management, data preprocessing techniques, data segmentation, data representation, and data modeling (point-wise and geometrically). We will demo a specific implementation of our data management solution. We will present our early experiences with some of the basic classification tasks (ship kind classification, fishing gear classification, fishing-non fishing classification) using Markovian approaches, standard data exploration approaches, and classifier induction approaches. We will also show how alternative methods from Natural Language Processing can assist in the same task. We will also discuss early results and challenges with the use of Deep Learning methods (e.g. Long Short Term Memory) on the AIS data.

Finally we will discuss the ongoing efforts in data integration, particularly in standardization of ocean data metadata under way as an IODE and Ocean Data Integration Project. International Oceanographic Data and Information Exchange.