Kernel methods for vessel trajectories

de Vries, G.K.D.

Citation for published version (APA):
INTRODUCTION

In the past decade, tracking of people and objects in geographical space has become ubiquitous. Smart phones have GPS sensors, cars are equipped with navigation systems and vessels are carrying special transponders to transmit information called AIS. All of this tracking data can easily be stored, generating a type of data called the moving object trajectory.

In most cases, these moving object trajectories are tracked and stored as is, i.e. they have no attached labels indicating, for instance, what the person or object is doing. Adding semantics to the tracks can help end-users or operators to interpret the data. For instance, tracking passengers in an airport and grouping these into different passenger types, such as tourist and business man, can greatly help in optimizing walking route lay-outs.

Another example is in maritime safety and security, where vessels can be tracked with GPS or radar. Trajectories made by fishing vessels are totally different from the paths generated by tankers or cargo ships, which follow regular routes. Identifying whether a trajectory belongs to one or the other class can help in determining whether a vessel exhibits unwanted behavior and thus should be further investigated.

Consider all the vessel movements that occur around a major port, like the busy harbors of Singapore and Rotterdam. Grouping these movements into clusters of similar behavior can help to get an overview of the general movement patterns. This overview can assist the operator to better spot irregular movements. Performing tasks like those mentioned above on vessel trajectories is the subject of study in this thesis.

One option to perform these tasks of assigning and discovering labels and classes for vessel trajectories is to design algorithms for them by hand. This is time consuming and difficult, and made more complicated by the inherent sensor noise in the data. However, we can also use automatic techniques from the fields of machine learning and data-mining that work with available historical movement data directly. This is the approach that we take in this thesis.

1.1 APPROACH

The tasks of assigning semantic labels to vessel trajectories are handled using techniques from machine learning and data-mining. We consider clustering of trajectories to discover groups of similar move-
The goal of this thesis is to provide automatic techniques to label vessel trajectories with semantic labels, such as the type of vessel or whether it shows strange behavior. The vessel trajectory data have a number of properties that make it interesting to apply automatic methods. First, there is often a large amount of data per trajectory. And, second, these trajectories are temporal, and hence sequential, in nature and commonly of different length.

In general, we investigate the following research question.
How do we, effectively and efficiently, assign semantic labels to vessel trajectory data?

We answer this question by applying machine learning techniques to vessel trajectories. By using classification, types can be assigned. Clustering is used for grouping vessels into similar behavior classes. Outlier detection allows for discovering abnormal behavior among regular trajectories. The algorithms that we use are based on kernel methods. While answering the above question by applying machine learning, we pose a number of more specific research questions.

The first question addresses the problem that each trajectory is based on a lot of samples, and these samples are often redundant.

1. How do we reduce the large volume and redundancy of vessel trajectories, and make them suitable for higher level reasoning?

This question is answered in Chapter 3 by looking at a trajectory compression technique. We adapt this technique in such a way that we generate representations of trajectories that can be used in higher level reasoning, for which we give a number of examples. This compression is also the first step in applying machine learning algorithms to vessel trajectories.

Secondly, we consider the issue of how to perform typical machine learning tasks with trajectories. The sequential and non-fixed length nature of trajectories suggests a similarity measure, as opposed to feature based, approach to do the tasks of clustering, classification and outlier detection. More specifically we answer the question:

2a. How do we compare vessel trajectories to be able to apply machine learning techniques?

In Chapters 4 and 5 we investigate this question by defining a common set of similarity measures from the fields of time-series analysis and computational geometry and use these in clustering, classification and outlier detection. We test these measures in a kernel method framework and research both their traditional and kernelized versions.

To further investigate the use of trajectory compression we also look at the following question.

2b. What is the influence of applying trajectory compression on comparing vessel trajectories for machine learning?

This question is answered in Chapter 4 by applying a subset of the similarity measures that is sensitive to compression to uncompressed and compressed trajectory data for a range of compression settings.

Most types of moving object trajectories occur in a space with semantics. In contrast to, for instance, handwritten digits or eye movements, our trajectories occur in a geographical world filled with places and regions. Utilizing this type of information can enhance machine learning and leads to the following question.
3. How do we incorporate geographical domain knowledge to improve machine learning for vessel trajectories?

We answer this question in Chapter 6 by defining similarity measures that can handle both raw trajectory information and conceptual domain knowledge simultaneously. We show how this measure can improve classification performance over using just raw trajectories or just geographical domain knowledge and give clustering results that illustrate the discovery of concepts that combine raw trajectories and domain knowledge.

1.3 Poseidon Project

The research presented in this thesis has been carried out in the context of the Poseidon project. The aim of this project was to develop a Maritime Safety and Security (MSS) system. To create such a system, research had to be done on integration and testing of the components of such a system, and also on the data that such an MMS system provides and needs to analyze. For the research in this thesis we worked on interpreting and analyzing the vessel trajectory data created by the MSS system and linking this data to higher level concepts and labels. We did this by creating solutions for the three tasks of clustering, classification and outlier detection, introduced above.

A summary of the research and achievements of the entire Poseidon project, including our own, is to be published in the forthcoming book by van de Laar et al. (to appear).

1.4 Outline

The rest of this thesis is structured as follows.

In Chapter 2 we give preliminary definitions and concepts that are used throughout the rest of the thesis. We give a brief introduction to machine learning and kernel methods. Then we give an overview of the field of spatio-temporal data-mining, in which the research in this thesis is situated. We end with introducing the datasets that we used in this thesis and the Poseidon project.

We extend an existing trajectory compression technique in Chapter 3 so that it works better with vessel trajectory data, which we show with experimental results. We also illustrate how this compression can be used as the first step in higher level reasoning.

Alignment similarity measures are the subject of Chapter 4. We study two common measures, and their kernel versions. The presented experiments investigate which measures perform best in the three tasks of clustering, classification and outlier detection. Furthermore, we specifically consider what effect the compression method

http://www.esi.nl/poseidon
from the previous chapter has on task performance using alignment measures.

In Chapter 5 we study a family of integral over time based similarity measures. For these measures we give corresponding positive semi-definite kernel versions, some of which did not exist previously in the literature. In the experimental section we provide results on which measures perform best and the difference with alignment techniques.

Chapter 6 investigates the use of geographical domain knowledge in the machine learning tasks. This knowledge is incorporated in the alignment similarity measures.

Finally, Chapter 7 revisits the research questions posed in the Introduction and answers them.