A comparison of supervised learning classifiers for oil spill detection

A.S.Y.Chiu
[Andy's TU email]@student.tudelft.nl

T.P.van Helden
[Thomas's TU email]@student.tudelft.nl

S.S.Jahanshahi
S.S.Jahanshahi@student.tudelft.nl

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

I. Introduction

Marine oil Spill is a form of human made environmental pollution usually occur during transportation of oil, drilling platforms or tankers [12]or during the maintenance on oil exploration sites in the ocean. One of the most destructive damages caused by oil spills in the history happened in the Gulf of Mexico as the oil spread out in the ocean by the explosion on the drilling platform affected marine ecosystem and wild life[2].

Detection of oil spills can help to control environmental risks and prevent incalculable damages by it. Synthetic aperture radar(RAS) image is useful to detect oil spills spots in the ocean. SAR image has a high resolution, wide area coverage and moreover images can be taken day and night under any weather conditions. This enables oil spill investigators to monitor ocean 24 hours a day[4].

There are various automated systems proposed in order to detect spills spots in SAR images. These systems analyzes the SAR images, assigns the probability of dark spots and proposes an algorithm to classify the dark images in to the oil spills and look alike shapes [10, 3, 7, 6].

In machine learning classification of oil spills from lookalike in SAR images is of highest importance. Main purpose of this article is to investigate three popular supervised learning classifiers in general. Later on more specific research is done on qualitative comparison between these classifiers in order to derive the criteria in choosing algorithm/s and classifier/s that would be most suitable for detection of oil spills in SAR images.

II. Oil spills summary

Oil spill detection can be done with several methods. There are different types of data that can be used. In the field of images even there are plenty of options. Aircrafts can create images across various spectra of light with different cameras. However, in this paper we focus on Synthetic Aperture Radar. SAR creates radar images with good spatial resolution and at a great distance. The main alternative is Side-Looking Airborne Radar, which is older but significantly cheaper.

These images are created through radio waves. Radio waves are sent from the SAR device to an area. That area reflects radio waves in a certain way, which is measured through the delay in time the wave takes to travel back. When an ocean is hit by radio waves, that energy is reflected in a certain way. Oil reflect that energy in a different way. A three-dimensional array of scene elements is defined which will represent the volume of space within which targets exist. Each element of the array is a cubical voxel representing the probability (a "density") of a reflective surface being at that location in space. This voxel density represents a difference in material (oil or water).

But there are many factors that cause problems. Wind, weather, fresh water and organisms in the water are all factors to take into account when reading these images. To have better classification of oil spills within SAR images, preprocessing is done. First the general there is a general quality assessment. Afterwards they look at speckle removal, noise removal. They also flag areas in the image which might interfere with the classification process. These areas include shorelines and land, high or low wind areas and but also algea infestations and seaweed dense areas.

M. Fingas, C. Brown (2014), "Review of oil spill remote sensing", pp14-17.
I. Keramitsoglou, C. Cartalis, C. T. Kiranoudis (2004), "Automatic identification of oil spills on satelite images", pp642-643
Wikipedia on "Synthetic Aperture Radar", will find appropriate source.

III. Classifier summary

I. What is supervised learning

II. Support Vector Machines

III. Decision trees

IV. Multi Layer Perceptrons

IV. Comparison of classifiers

I. SVM Vs DT

It is impossible to fully compare Support Vector Machines and Decision Trees. Both have advantages and disadvantages, each serving it's own purpose. One such difference is that DTs are faster when handling new dataset, compared to SVM. This is because of the complexity of SVMs' arithmatic computations, where DTs only need to follow a logical path in a tree. A more interesting difference is that SVM has higher accuracy in general as was observed in numerous studies[1].

II. SVM Vs MLP

- Comparison between MLP and SVM is done using same dataset. MLP accuracy in classifying the ECG(Electrocardiography) signals, is more accurate than other ANN. MLP with back propogation(BP) training algorithm suffer from slow convergance to local minima, on the other hand SVM classifier with (K-A) training do not trap in local minima points therefore they are faster than ANN. [8]

- In high dimensional data, SVM accomplishes better accuracy compare to MLP, this is when new kernel function proposed for SVM. The reason behind is that MLP needs more hidden units for tested data set and becomes more complex when the dimension of data set increases wheras SVM complexity does not depend on dimension of data set. SVM are efficient on optimal seperation of unseen data points.[11]

- SVM works better than MLP for the offline signiture recognition(within finite database), The comparison is done between the identification rate(increment of 20% for SVM) and for training time needed. The superiority of SVM is because of its generalization ability in high dimensional space[5].

- SVM outperforms to MLP, in Wind speed prediction. The comparison is done using a wind speed dataset that covers a 12 years between 1970-1982. Dataset is devided into three parts: traing, test and validation sets. The output result shows the SVM outperform MLP on all orders resulting in the lower MSE(MLP is 0.0090 while it is 0.0078)[9].

III. DT Vs MLP

V. Final words or conclusion

References

[1] M Arun Kumar and Madan Gopal. A hybrid svm based decision tree. *Pattern Recognition*, 43(12):3977–3987, 2010.

[2] Barry Bozeman. The 2010 {BP} gulf of mexico oil spill: Implications for theory of organizational disaster. *Technology in Society*, 33(3âĂŞ4):244 – 252, 2011.

[3] Camilla Brekke and Anne HS Solberg. Classifiers and confidence estimation for oil spill detection in envisat asar images. *Geoscience and Remote Sensing Letters, IEEE*, 5(1):65–69, 2008.

[4] Lena Chang, Z.S. Tang, S.H. Chang, and Yang-Lang Chang. A region-based {GLRT} detection of oil spills in {SAR} images. *Pattern Recognition Letters*, 29(14):1915 – 1923, 2008.

[5] E. Frias-Martinez, A. Sanchez, and J. Velez. Support vector machines versus multi-layer perceptrons for efficient off-line signature recognition. *Engineering Applications of Artificial Intelligence*, 19(6):693 – 704, 2006. Special Section on Innovative Production Machines and Systems (I*PROMS).

[6] Yue Guo and Heng Zhen Zhang. Oil spill detection using synthetic aperture radar images and feature selection in shape space. *International Journal of Applied Earth Observation and Geoinformation*, 30(0):146 – 157, 2014.

[7] Iphigenia Keramitsoglou, Constantinos Cartalis, and Chris T. Kiranoudis. Automatic identification of oil spills on satellite images. *Environmental Modelling and Software*, 21(5):640 – 652, 2006.

[8] Majid Moavenian and Hamid Khorrami. A qualitative comparison of artificial neural networks and support vector machines in {ECG} arrhythmias classification. *Expert Systems with Applications*, 37(4):3088 – 3093, 2010.

[9] M.A. Mohandes, T.O. Halawani, S. Rehman, and Ahmed A. Hussain. Support vector machines for wind speed prediction. *Renewable Energy*, 29(6):939 – 947, 2004.

[10] Linlin Xu, Jonathan Li, and Alexander Brenning. A comparative study of different classification techniques for marine oil spill identification using radarsat-1 imagery. *Remote Sensing of Environment*, 141(0):14 – 23, 2014.

[11] E.A. Zanaty. Support vector machines (svms) versus multilayer perception (mlp) in data classification. *Egyptian Informatics Journal*, 13(3):177 – 183, 2012.

[12] Peng Zhang, Ruijun Sun, Linke Ge, Zhen Wang, Hong Chen, and Ziwei Yao. Compensation for the damages arising from oil spill incidents: Legislation infrastructure and characteristics of the chinese regime. *Estuarine, Coastal and Shelf Science*, 140(0):76 – 82, 2014.