MCAST IICT Research Paper

Lee Pace, Frankie Inguanez, Ivan Briffa
Institute of Information & Communication Technology
Malta College or Arts, Science & Technology
Corradino Hill
Paola PLA 9032
{lee.pace.a100433, frankie.inguanez, ivan.briffa}@mcast.edu.mt

***Abstract*—Marine traffic is increasing as time goes by, so in order to solve future problems like to manage the marine traffic, construction (ex: wind farms, fish farms) or activities like sailing competitions. The project will create a map with Automatic Identification Data (AIS) and will create a history map of all the marine vessel navigation routes and with it can be displayed where the heavy routes are and possibly detect some illegal activities like fishing in protected areas.**

***Index Terms*—AIS data, AIS data Visualization**

I. Introduction

Various studies have pointed out the importance of AIS (Automatic Identification System) (Jiacai, P., Qingshan, J., Jinxing, H. and Zheping, S., 2012; Fiorini, M., Capata, A. and Bloisi, D.D., 2016; Perez, H.M., Chang, R., Billings, R. and Kosub, T.L., 2009; Olindersson, F. and Janson, C.E., 2015.; Mustaffa, M., Ahmat, N.H. and Ahmad, S., 2015.) and of Anomaly Detection (Mascaro, S., Nicholso, A.E. and Korb, K.B., 2014) for vessel traffic visualization. Since 2002 all passenger and large sea vessels are required to carry AIS onboard. With this system ships can make themselves visible to each other. This aids in collision avoidance and in taking records of ship maneuverability history (Mustaffa, M., Ahmat, N.H. and Ahmad, S., 2015.). With “AIS data” (Fiorini, M., Capata, A. and Bloisi, D.D., 2016) datasets are created so they can be used in various studies like visualizing the path history of the vessels.

II. Literature Review

Various studies have pointed out the importance of AIS (Automatic Identification System) (Jiacai, P., Qingshan, J., Jinxing, H. and Zheping, S., 2012; Fiorini, M., Capata, A. and Bloisi, D.D., 2016; Perez, H.M., Chang, R., Billings, R. and Kosub, T.L., 2009; Olindersson, F. and Janson, C.E., 2015.; Mustaffa, M., Ahmat, N.H. and Ahmad, S., 2015.) and of Anomaly Detection (Mascaro, S., Nicholso, A.E. and Korb, K.B., 2014) for vessel traffic visualization. Since 2002 all passenger and large sea vessels are required to carry AIS onboard. With this system ships can make themselves visible to each other. This aids in collision avoidance and in taking records of ship maneuverability history (Mustaffa, M., Ahmat, N.H. and Ahmad, S., 2015.). With “AIS data” (Fiorini, M., Capata, A. and Bloisi, D.D., 2016) datasets are created so they can be used in various studies like visualizing the path history of the vessels.

I. Automatic Identification System (AIS)

In case of Mustaffa, M., Ahmat, N.H. and Ahmad, S., (2015) they used several equipment like AIS receiver and antenna, AIS decoder and AIS analyzer. This was done so they can collect the AIS data for port Klang. The data was all recorded in a CSV file which later was analyzed using spreadsheets program.

Where incase of Fiorini, M., Capata, A. and Bloisi, D.D., (2016) they obtained the AIS raw data and a database management system (DBMS) was used to handle the millions of records.

In the AIS data contains all the information of the ship that was transmitted from the AIS transmitter. This data contains all the actions of the vessel made during the route like destination, speed, longitude and latitude which is transmitted every 2 to 10. Moreover every 6 minutes more information is sent like ship name, identification number, type of vessel, etc. Because of this the data must be filtered to key indexes so later “map construction and interactive map visualization” (Fiorini, M., Capata, A. and Bloisi, D.D., 2016) can be created.

In case of Jiacai, P., Qingshan, J., Jinxing, H. and Zheping, S., (2012) they created “AIS Data Visualization Model” with combines three features which are rate of encounter, rate of turn and speed acceleration which later is visualized on “Electronic Chart Display” and “Information System (ECDIS)”. This model uses various algorithms to process the data in each feature so later the longitude and latitude are found of visualized on a map.

II. Anomaly Detection

The Gaussian mixture models(GMMs) is a proven popular choice for representing normality in models of vessel behavior. As it name implies GMM is a combination of a multivariate Gaussian distributions and these distributions aim to summarize how the training data cluster and spread in the multi-dimensional space. (Mascaro, S., Nicholso, A.E. and Korb, K.B., 2014)

According to Mascaro, S., Nicholso, A.E. and Korb, K.B., (2014) models from Gaussian mixture models, support vector machines and neural networks have a disadvantage which they do not provide a transparent model for the human user. At Bayesian Network (BN) anomaly detection has 2 advantages over the others which is that they are easily understood by people who are not BN specialist and they allow for the straightforward incorporation of expert knowledge.

III. Research Methodology

Hypothesis/Research question Can a map be created showing marine traffic, so that with the growth for demand in the marine areas (e.g. offshore wind power plants) require some planning about the heavy marine traffic. With this project future allocation plans in marine about human and temporally activities can be placed in safe areas away from the heavy marine traffic.

1. What dataset is needed, and features are important for the creation of this prototype? 2. What types of techniques can be used for detection and recognition of vessels? 3. What evaluation techniques are needed to benchmark such a prototype?

First thing to begin with is to obtain an AIS data set from a company or by gathering it yourself. In my case the data set was taken from Marine Cadastre (Marine Cadastre, n.d.) Then the data set will be examined and after thoroughly went through it and understood all the information the cleaning process begins. This process consist of removing garbage data or empty cells in the data set and its important to do it because it can mess up with the algorithms and other the final result and also the machine will be wasting time going through data that is not important for the task. After the cleaning process is done for the data set, we can move on programming the model for the system. In this case the Gaussian Mixture Model (GMM) will be used which consist of multi-variate dimensional space. For the Gaussian Distribution model, the equation below will be used and it will be responsible for estimating the Gaussian distribution for each features. If a dataset is given $x(1), \ldots, x(m)(wherex(i)R^n$, here $n = 2$,were for each $(I = 1 \ldots n)$ the parameters mean and variance $(mu, sigma2)$ and like that the we can calculate the mean and variance of the array X. (Kumar, 2019)

$$\mu_j = \frac{1}{m} \sum_{i=1}^{m} x_j^{(i)}$$

$$\sigma_j^2 = \frac{1}{m} \sum_{i=1}^{m} (x_j^{(i)} - \mu_j)^2$$

Fig. 1. Gaussian Distribution Model

After finishing the distribution model, we can for the next step which is to find contextual anomalies in a time series data (Paul, 2019). And for the system to proceed forward we have to use K-Means clustering which KMeans uses Euclidean distance internally and uses this formula. (Paul, 2019)

Anomaly detection in python is going to be made possible through a k-NN classification method and also a python library called PyOD.

Another approach is to use power bi for anomaly/outlier detection which has a similar process to the python approach

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Fig. 2. Euclidean Equation

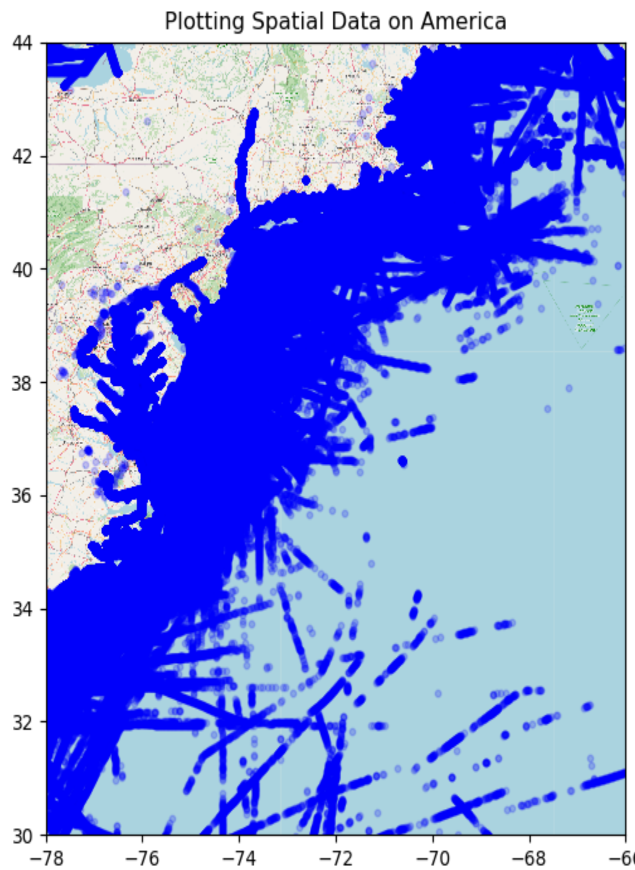

Fig. 3. Python Plot Map

where u get a data set goes through a cleaning process and after u know which features to use u begin plotting with the outlier detection scatter plot chart.

IV. Evaluation

There was a lot of data sets to choose and work with so no problem obtaining material to work with. First time I used phyton language and other software's like power bi gnu octave and pycharm. The Data gathered couldn't be really completed since the python program had some clash with the data set since I didn't know which features to keep and not to keep same with power bi, I have some confusion on how the anomaly detection in time series data exactly works

V. Conclusion

Can't be called a successful outcome since it's not complete and no actual data was pulled because it wasn't possible since

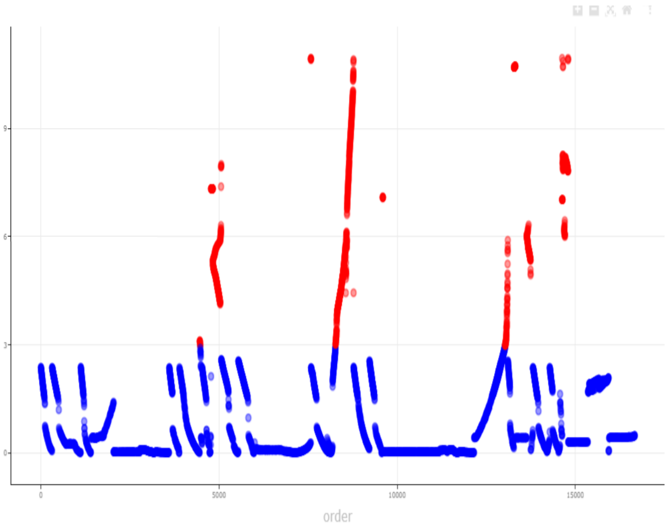

Fig. 4. Power Bi Scatter Plot

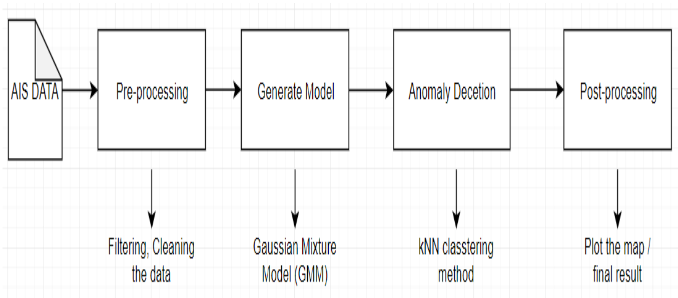

Fig. 5. Pipline

there were faults in the data set filtering/cleaning process and on the understanding of the time series algorithm

VI. References

Kumar, A. (2019, 2 10). TowardsDataScience. Retrieved from https://towardsdatascience.com/wondering-how-to-build-an-anomaly-detection-model-87d28e50309

Marine Cadastre. (n.d.). Retrieved from https://marinecadastre.gov/ais/

Paul, S. (2019, April 4). Floyd Hub. Retrieved from https://blog.floydhub.com/introduction-to-anomaly-detection-in-python/

Jiacai, P., Qingshan, J., Jinxing, H. and Zheping, S., 2012. An AIS data visualization model for assessing maritime traffic situation and its applications. Procedia Engineering, 29, pp.365-369.

Fiorini, M., Capata, A. and Bloisi, D.D., 2016. AIS data visualization for maritime spatial planning (MSP). International Journal of e-Navigation and Maritime Economy, 5, pp.45-60.

Perez, H.M., Chang, R., Billings, R. and Kosub, T.L., 2009, April. Automatic identification systems (AIS) data use in marine vessel emission estimation. In 18th Annual International Emission Inventory Conference (Vol. 14, p. e17).

Olindersson, F. and Janson, C.E., 2015. Development of a software to identify and analyse marine traffic situations. In International Conference on Marine Simulation and Ship Manoeuvrability (MARSIM), Newcastle, United Kingdom.

Mustaffa, M., Ahmat, N.H. and Ahmad, S., 2015. Mapping Vessel Path of Marine Traffic Density of Port Klang, Malaysia using Automatic Identification System (AIS) Data. International Journal of Science and Research (IJSR), 4(11), pp.245-248.

Mascaro, S., Nicholso, A.E. and Korb, K.B., 2014. Anomaly detection in vessel tracks using Bayesian networks. International Journal of Approximate Reasoning, 55(1), pp.84-98.

Sciberras, R. and Inguanez, F., 2019. Automated Traffic Congestion Estimation Via Public Video Feeds. mcast.

Y. Xu et al., "Dual-Mode Vehicle Motion Pattern Learning for High Performance Road Traffic Anomaly Detection," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, 2018, pp. 145-1457.

C. Pascoal, M. R. de Oliveira, R. Valadas, P. Filzmoser, P. Salvador and A. Pacheco, "Robust feature selection and robust PCA for internet traffic anomaly detection," 2012 Proceedings IEEE INFOCOM, Orlando, FL, 2012, pp. 1755-1763.

Kong, Xiangjie Song, Ximeng Xia, Feng Guo, Haochen Wang, Jinzhong Tolba, Amr. (2017). LoTAD: long-term traffic anomaly detection based on crowdsourced bus trajectory data. Worldwide Web. 10.1007/s11280-017-0487-4.