

Future Ship Data Prediction Rules with C4.5 and SVM Decision Classifiers using Big Data

S. Sahana

PG Scholar

Department of CSE

University College of Engineering Trichy

(BIT Campus)

Trichy.

Mr. S. Anuvelavan

Assistant Professor

Department of CSE

University College of Engineering Trichy

(BIT Campus)

Trichy.

Abstract :-The advance changes in Maritime Situational Awareness, there causes an increase in maritime data. While there exists a large amount of data with variable quality, decision makers need reliable information about possible situations or threats. To address this requirement, it's tending to use huge information with final server storage capability. Hence a method to analyses the historical data for processing safe journey of ship so as to implement the predictive values of data. Our dataset indicate the Geo Space nature, which can be regenerate into compressed data by means of suitable classifier and trimming operation. The entire system deals with manipulation of Geo Space Dataset with a newly proposed ASVM algorithm that will be the resultant combination of two powerful classification algorithms, C4.5 and Support Vector Machine (SVM) to predict the future analysis of the marine vessels. The abstraction question possibility is employed for looking out the resultant information from the large information server.

Key Words:- Geo space dataset, C4.5 classification algorithm, SVM classification algorithm, prediction technique.

1. INTRODUCTION

As per the technology grows there causes an increase in the growth of data so that, the amount of data goes on increasing day by day. Hence, it is necessary to develop a proper technique for storing and analyzing those data. They are in need to extract and classify those data for our future work. The big data which is an advance of data mining is the most popular technique for performing and predicting information from the large database. In this paper, the marine connected dataset such as ship name, id, wind, pressure, humidity, speed, etc. are collected and stored in the server. The real data should be demonstrated and an objective of data comparison is done through most popular classification techniques [1]. It is that transfer within the server as huge information. Here it tends to use two styles of formula to analyses the worth of knowledge. The Associate in nursing formula specifically C4.5 is employed to predict the values of

the dataset. The final method of building a choice tree is as follows. Given a group of coaching information, apply a menstruation operate onto all attributes to search out a best rending attribute. Once the rending attribute is decided, the instance house is partitioned off into many elements. Among every partition, if all coaching instances belong to at least one single category, the formula terminates. In some case it may tend to fuse both online and offline data of vessel monitoring using some clustering and association rules [2]. The detection algorithm is used for random transformation [3]. It causes fusion of several data that we get from SAR data, AIS data and RADAR data for better surveillance of ship [4]. Ship surveillance is done by integration of multiple sensors, including space-borne SAR, AIS, optical, infrared and hyper-spectral sensors, etc., has the potential to achieve high accuracy and efficiency, which plays an important role in maintaining maritime surveillance data [5]. Gerard Margarit et al used fuzzy logic for classification of single pore SAR images data. It estimate the parametric vector using decision rules [6]. The classification of vessel types can be done by single-pass polarimetry synthetic aperture radar (SAR) [7]. Though the research found that many types of classification has been done by vessel or data of ships, an efficient method of classification is to be performed for better prediction and safe surveillance of ship.

2. MACHINE LEARNING TECHNIQUES

There are two types of machine learning technique namely supervised and unsupervised learning technique where it prefers through training data and test data. Classification comes under supervised technique which is based on processing prediction of data values. Data set is the main sources of classification task that contribute many numeric values which lead to predictive model. While there are many classification algorithm exit, the two powerful algorithms namely c4.5 and SVM were used in this paper.

2.1 C4.5 DECISION CLASSIFIER

The decision classifier is used widely in classification technique. It is a powerful and popular tool used for classification and prediction. Once a choice tree is made, classification rules may be simply generated, which might be used for categorizing of recent instances with unknown class labels. C4.5 may be a customary formula for inducement classification rules within the kind of call tree. As Associate in nursing extension of ID3, the default criteria of selecting rendering attributes in C4.5 at info gain quantitative relation [8]. Rather than using info gain as that in ID3, info gain quantitative relation avoids the bias of choosing attributes with several values. The decision tree is constructed by using C4.5 algorithm with training data set. It contains different classes with discrete values [9].

2.2 SUPPORT VECTOR MACHINE ALGORITHM

SVM, a powerful machine learning technique developed from statistical learning which makes significant achievements. It performs high accuracy of classification than the other data classification algorithm [10]. Text classification comes under supervised learning, where SVM combines with SSK mean clustering to perform minimum number of labeled data [11]. It performs on both the linear and nonlinear classification of input data, so that the hyper plane is constructed to classify those data. A new kernel based classification is done by SVM [12]. Thus the hyper plane may divide the n-data into two classes for binary classification of SVM. The classification is done by mapping the training data into higher dimensional feature [13].

2.3 ASVM ALGORITHM

The newly proposed algorithm is implemented by the resultant combination of both C4.5 and SVM algorithm. As C4.5 gives the result of some amount of gain values which shows the accuracy of the given dataset. Same as that SVM can calculate the accuracy of the given dataset. So as our work is to propose a newly designed ASVM algorithm in which it gives better accuracy that the other two algorithms.

3. ARCHITECTURAL DESIGN

Fig 1 shows Architecture of system, which explains that data are extracted from the ship which is a geo space data such as location, speed, temperature, humidity and etc. Thus collected data are the classified using C4.5 and SVM classification to find the accuracy of those data. The classified data are then upload to the virtual server. By using spatial query we can easily predict the best value in which the ship had travel and also can predict the future data in which the ship can travel in a safe way for further journey. Thus only by historical data we can predict the future data and we can conclude an assumption that by using the correct path and values.

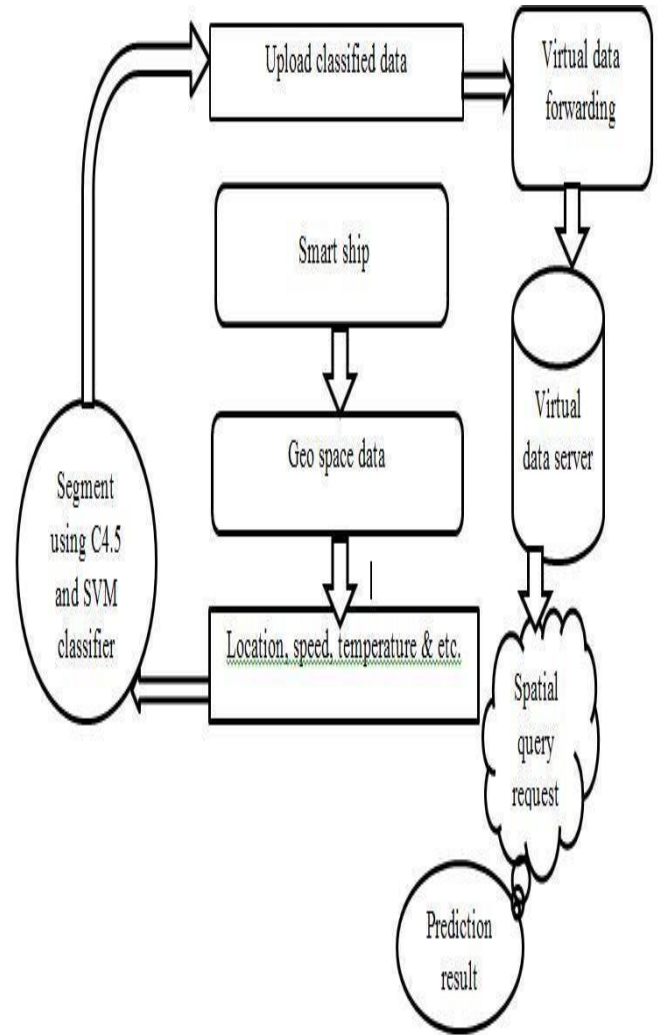


Fig 1 System Architectural diagram

4. RESULTS AND DISCUSSION

In this phase we classify the data using C4.5 algorithm that may give some of the accuracy values. Thus the by performing SVM it may show the better value than the C4.5 algorithm. By performing that algorithm we can get an ASVM results in more accurate. Hence for prediction ASVM is used for finding out the future values.

Fig 2 gives the results of classified data using c4.5. Thus it calculate the success ratio, failure ratio and gain for the given dataset.

The formula for calculating the gain ratio in c4.5 algorithm.

$$Info_x(T) = \sum ((|Ti| / |T|) \cdot Info(Ti))$$

$$Gain(X) = Info(T) - Info_x(T)$$

$$GainC4=(-1)*GainC4$$

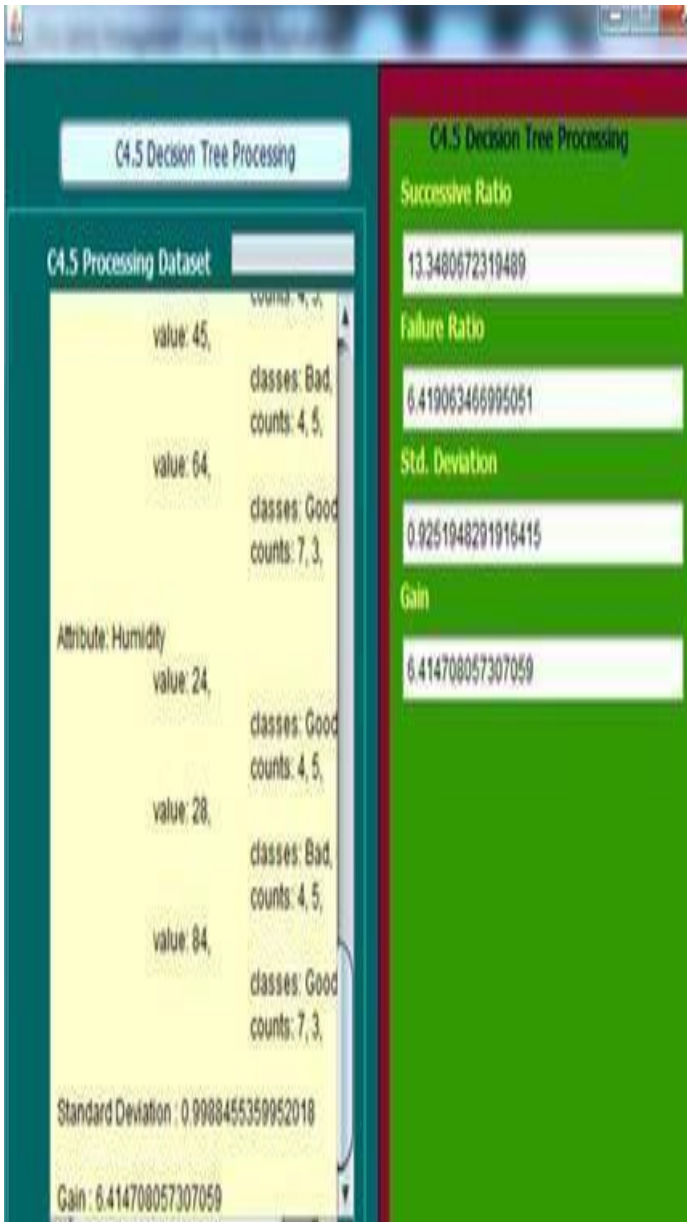


Fig 2 Implementation of C4.5 algorithm

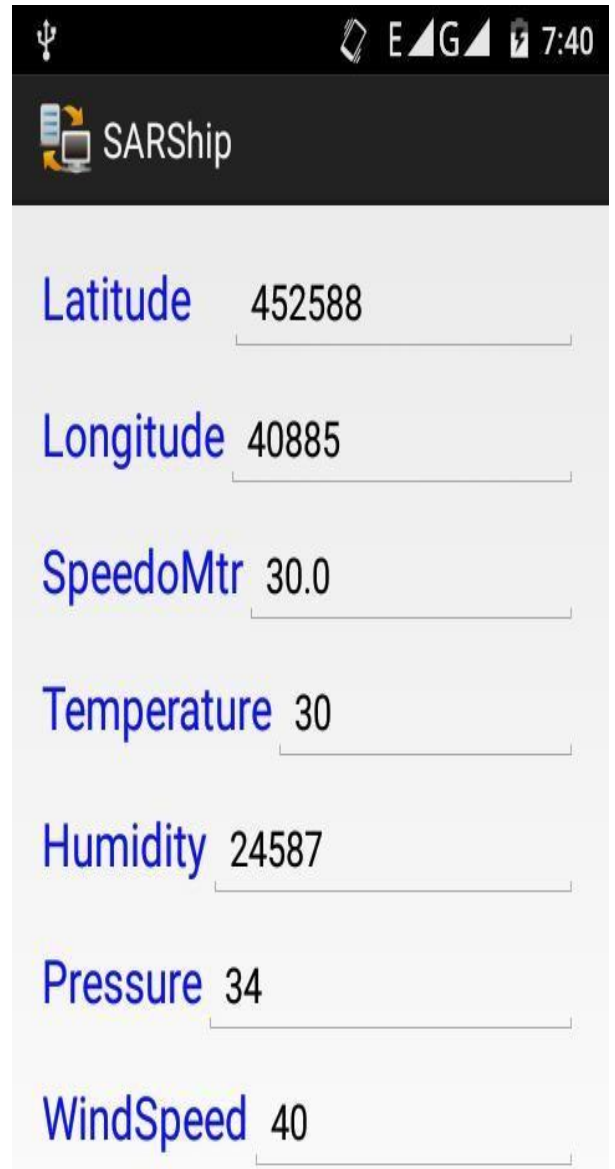


Fig 3 Mobile application for dynamic dataset

4.1 DATASET

The dynamic dataset is used for this classification hence it is a real time geo space data. Hence a mobile application is developed as a simulation for getting the dynamic dataset.

Thus the data can be sending to the system through Bluetooth by serial port connection. So that we can get the dynamic data as we get form the ships from SAR and AIS system.

Fig 3 shows the mobile application data which a dynamic one. As we cannot get the real time data from the satellite or RADAR, hence for assumption we are using the mobile application which is developed by our own. This application will gives the real time data such as location and all data related to ship that we get form the satellite. This data are then send to the pc via Bluetooth and create a dataset which is then stored in the server. Once the data set is created we started the process of classification using the decision classifier algorithm. After finishing the classification it will predicts the best and accurate values in which the ship had travel in a good condition.

Table 1. Experimental result on dynamic dataset

Various performance	C4.5	SVM	ASVM
Success ratio	13.348	14.56	15.987
Failure ratio	6.41906	7.654	8.342
Standard deviation	0.9251	1.0123	2.345
Gain	6.414	6.987	7.234
Mean	6.77654	7.18233	8.6062

The above table 1 shows the difference of the algorithm by performing its mean values. Hence it clarify that ASVM gives the best values of the other two algorithms. So we can use the ASVM algorithm for predicting the future values of ship related data.

5. CONCLUSION

This system has shown the relevance of maritime traffic knowledge on the SAR ship detection and self-reporting data. It is demonstrated through real data and an objective comparison with the most popular linear-propagation-based strategy that it is possible to increase the quality of correlation by exploiting knowledge of historical traffic patterns to project self-reported AIS observations to the SAR image acquisition time. In addition, the performance gain increases with the complexity of the routing system regulating the traffic in the area of interest. It can predict the best algorithm which is used to process the data. Hence by future prediction technique the decision maker can conclude the best values of the ship data.

6. FUTURE ENHANCEMENT

Now days the volume of data goes on increasing not only to any of the particular field. It depends upon the users that they collected and saved it for future purpose. Hence, it is important to process a system to store and retrieve the data in an easy and efficient manner. Thus the storage capacity of the data should be increase in order to maintain a large amount of data. The improvement in the algorithm technique leads a better way of prediction in maritime data.

REFERENCE

[1] Fabio Mazzarella, Michele Vespe, and Carlos Santamaria, SAR Ship etection and Self-Reporting Data Fusion Based on Traffic Knowledge "IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, VOL. 12, NO. 8, AUGUST 2015".

[2] F. Mazzarella et al., "Data fusion for wide-area maritime surveillance," in Proc. Workshop Moving Objects Sea , Brest, France, 2013, pp. 1-5.

[3] Maisondes Technologies, Place Georges Pompidou, An improvement of ship wake detection based on the radon transform Ph. Courmontagne ISEN, Laboratoire L2MP, 83000 Toulon, France Received 2 August 2004.

[4] Craig Carthel, Stefano Coraluppi, Raffaele Grasso, and Patrick Grigna, "Fusion of AIS, RADAR and SAR Data for Maritime Surveillance " NATO Undersea Research Centre Viale S. Bartolomeo 400 19138 La Spezia, Italy.

[5] Z. Zhao, K. Ji, X. Xing, H. Zou, and S. Zhou, "Ship surveillance by integration of space-borne SAR and AIS—Further research," J. Navigat., vol. 67, no. 2, pp. 295-309, Mar. 2014.

[6] G. Margarit and A. Tabasco, "Ship classification in single-pol SAR images based on fuzzy logic," IEEE Trans. Geosci. Remote Sens., vol. 49, no. 8, pp. 3129- 3138, Aug. 2011.

[7] Gerard Margarit, Member, IEEE, Jordi J. Mallorquí, Member, IEEE, and Xavier Fàbregas, Member, IEEE "Single-Pass Polarimetric SAR Interferometry for Vessel Classification" Gerard Margarit, Member, IEEE, Jordi J. Mallorquí, Member, IEEE, and Xavier Fàbregas, Member, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 45, NO. 11, NOVEMBER 2007.

[8] Seema sharma et al, "Classification Through Machine Learning Technique:C4.5 Algorithm based on Various Entropies"International Journal of Computer Applications (0975 – 8887)Volume 82–No 16, November 2013

[9] Salvatore Ruggieri, "Efficient c4.5" Dipartimento di informatica, Universitia di pisa corso italia 40 56125 pisa Italy.

[10] Durgesh k. Srivastava, Lekha bhambhu data classification using support vector machine Journal of Theoretical and Applied Information Technology © 2005 - 2009.

[11] Hongcan Yan Chen Lin, Bicheng Li, "A SVM-based Text Classification Method with SSK-means clustering algorithm" 2009 International Conference on Artificial Intelligence and Computational Intelligence.

[12] Ranzhe Jing Yong Zhang,"A View of Support Vector Machines Algorithm on Classification Problems" 2010 International Conference on Multimedia Communication.

[13] Z.Nematzadeh Balagatabi, H.Nematzadeh Balagatabi Comparison of Decision Tree and SVM Methods in Classification of Researcher's Cognitive Styles in Academic Environment, Indian,Journal of Automation and Artificial Intelligence Vol: 1 Issue: 1 January 2013 ISSN 2320 – 4001.