ISSN: 2366-1313

A Research on Doppler Impact prediction using Supervised Machine Learning

[1]S. AKHILA, [2]R.VEERA REDDY

[1,2]Assistant Professor, Dept of CSE, Megha institute of engineering and Technology for women, Ghatkesar (T.S)

Abstract: The Indian Space Research Organisation (ISRO) has created a system for using the Indian constellations for navigation called NavIC. Four geosynchronous satellites and three geostationary satellites make up the NavIC satellite constellation. There are several external influences that might impair navigation. One cause of tracking mistakes caused by geostationary satellites is Doppler collision. It takes place in the space between the IRNSS 1C-1G, 1C-1F, and 1F-1G geostationary satellites. Doppler Collision (DC) period is seen when the relative Doppler frequency of satellites is smaller than the bandwidth of the code tracking loop. DC hinders the NavIC system's ability to pinpoint precise locations. When it comes to DC interference, the 1C-1G geostationary satellite pair is the worst hit. Predicting DC using machine learning techniques is a powerful tool for mitigating its effects on positional accuracy. Relative Doppler, satellite location, satellite velocity, length of occurrence, and relative Doppler are all taken into account to make predictions. Linear regression, the Random Forest regressor, and the K-Nearest Neighbours (KNN) regressor are three supervised machine learning algorithms utilised in this prediction process. The Doppler Collision was best predicted by the random forest regressor.

Keywords: **Geo stationary satellites, Doppler Collision, Relative Doppler, satellite velocity, satellite position, Linear regression.**

I. INTRODUCTION

ISRO has created a regional satellite-based navigation system called Navigation with Indian Constellation (NavIC), which is now fully functional. It offers precise location services in India, reaching out to customers within a radius of 1,500 km from the country's borders. There are a total of seven satellites in the NavIC network, four of which are Geo synchronous and the other three of which are geostationary. Currently, IRNSS 1C, 1F, and 1G are in geostationary orbit, whereas IRNSS 1B, 1D, 1E, and 1I are in geosynchronous orbit.

The two types of services offered by a satellite navigation system are the Standard Positioning Services (SPS) and the Restricted Services (RS). The L1, S1, and L5 bands (1575.42MHz,

ISSN: 2366-1313

2492.028MHz, and 1176.45MHz) are used by the NavIC system. The NavIC architecture is made up of the "ground segment," "space segment," and the "user segment." of the NCRC lab of the Electrical and Electronics Engineering department at Chaitanya Bharathi Institute of Technology (CBIT), in Gandipet, Hyderabad, an IGS (IRNSS-GPS-SBAS) receiver has been installed. The CDMA concept is at the heart of the IGS receiver.

Doppler collisions result from the correlation of signals from different satellites.

As a result, CDMA systems have code measuring problems [4]. Because their satellites are in medium Earth orbit with identical Doppler frequencies and the period of Doppler Collision will be substantially brief [3], the Doppler Collision is not seen as a concern for other navigation systems like GALILEO, GLONASS, and GPS. Machine learning methods are utilised to forecast the occurrence of Doppler Collisions in the future. Predicting the future and comprehending the present is where machine learning really shines. The ability to automatically learn and improve based on experience, with no further programming required, is made possible by machine learning.

II. LITERATURE SURVEY

While most studies in sentiment analysis have focused on product analysis—that is, analysing people's opinions on products like phones, hotels, or movies—other studies have extended sentiment analysis to a deeper level by extracting people's overall opinion on product features (Bhadane et al., 2015; Guha et al., 2015; Hu & Liu, 2004; Kiritchenko et al., 2014a; Nguyen et al., 2017; Potdar et As a result, sentiment analysis has become more common in business settings, with marketing firms increasingly use it to gauge consumer reaction to new offerings, assess the efficacy of existing ones, identify market trends, and more.

Sentiment analysis is so useful that even government organisations use it to gauge the safety of the country. For instance, the United States government spent a hefty $2.4 million on a sentiment analysis initiative that tracks internet activity (New York Times, 2006). Predicting the outcome of an election is one of the most significant applications of sentiment analysis. It aids politicians in gauging their support from the public and learning more about their constituents. For instance, a well-known Massachusetts firm called Crimson Hexagon employed sentiment analysis to gauge public opinion on the Gulf of Mexico oil disaster. The data showed that those who lived closer to the gulf were less

ISSN: 2366-1313

likely to attribute blame. Instead, they concentrated on aid work (New York Times, 2010).

Since the year 2000, research into sentiment analysis has exploded, making it one of the most dynamic areas of study in natural language processing. Summary of people's thoughts and experiences has shown to be quite valuable for recommender systems, businesses, and editorial sites. Existing studies in sentiment analysis have used a variety of approaches to perform sentiment analysis, including natural language processing (Yi et al., 2003), machine learning (Pang & Lee 2004; Pang et al., 2002; Reis et al., 2019), an unsupervised lexicon-based approach (Hu & Liu, 2004; Kim & Hovy, 2004; Liu, 2010; Taboada et al., 2010; Turney, 2002), Over the last decade, researchers have investigated several facets of sentiment analysis, including its subjective vs objective nature, its polarity, and its ability to categorise emotions like happiness, fear, and sorrow. The methods were surveyed in Pang and Lee's (2008) and Liu and Zhang's (2012) studies. Pang and Lee (2008) and Owoputi et al. (2013) investigated numerous different features and their efficacy. These included post, topic-based features, negation handling, and many more.

In the past, academics have approached the sentiment analysis problem as binary classification, or dividing a text into positive and negative categories (Hu & Liu, 2004; Kim & Hovy; Turney, 2002). Fewer publications from the past (Pang & Lee, 2004; Wilson et al., 2005) went beyond simple binary categorization to include subjective and objective measures. In other words, we first determine if a piece of text is subjective or objective, and then we determine whether the subjective text is polar (positive or negative) or agnostic (neutral).

When classifying degrees of subjectivity, Pang and Lee (2004) only looked at negative and positive categories. Wilson et al. (2005), on the other hand, also thought of the neutral class. Since Pang et al. (2002) employed machine learning methods (SVM and NB) to categorise texts, Pang and Lee (2004) suggested that using lexical elements alone is inadequate to produce reliable prediction of sentiment. However, by eliminating objective clauses, they saw a rise in polarity classification accuracy. The team increased their score significantly, from 82.8 to 86.4.

It follows that studies in sentiment analysis have been conducted at several levels, from the document level (Pang & Lee, 2008) to the determination of phrase and

ISSN: 2366-1313

word polarity (Esuli & Sebestiani 2006; Hatzivassiloglou & Mckeown 1997). In addition, most researchers have historically concentrated on sentiment analysis of longer, more structured texts like blog posts, movie reviews (Pang & Lee, 2004; Pang et al., 2002), product reviews, and reviews of the services offered by businesses (Siqueira &Barros, 2010).

Kiritchenko et al. (2014a), who took part in the SemEval-2014 shared task on aspect-level sentiment analysis, have produced a remarkable contribution in the area of sentiment analysis. They suggested using an in-house sequence tagger to identify feature terms (e.g., camera, battery life) and supervised classifiers (SVM) to identify category terms (e.g., all food items can be grouped into one category food) and sentiment towards feature terms and categories. When it came to classifying aspects and gauging opinions on classes, they were unrivalled. They placed third in recognising aspect words and first in gauging how people feel about such terms in the context of laptops.

In 2016, the research team of Potdar et al. released a programme called SAMIKSHA (review bot) that creates a factual summation of public reviews on a product, allowing consumers to gain a sense of the general public's opinion of the product before making a purchase. Features of products may be rated on average with the help of the suggested tool.

III. IMPLE M E N T A T IO N OF SUPERV I SED ML ALGORITHMS

Future Doppler Collision occurrences are predicted using Supervised Machine Learning. Supervised machine learning requires labelled data for model training. In order to train the model, it is necessary to have knowledge about the object's characteristics as well as labels associated with those features [8]. In order to foretell Doppler Collision occurrences, three supervised machine learning techniques based on regression are used. The approach is based on regression theory and is used to make predictions about a continuous output given a discrete input. The relative Doppler is provided as an output variable, with the satellite location of 1C-1G, its velocities, and occurrence period (seconds) as multiple predictor variables. Linear Regression is the first method employed in a regression setting. The goal of a linear regression model is to establish a connection between the variables you feed it and the results you get back. We refer to the input variables as independent variables and the output

ISSN: 2366-1313

variable, which is being predicted, as the dependent variable [9]. Since there is more than one independent variable, this model is known as a multiple regression. This model seeks to establish a connection between a number of variables in order to identify the line that most closely represents the observed data [10]. Random Forest Regression is the second regression-based technique; it makes use of the predictions of a forest of decision trees [11]. Due to the size and complexity of the data set, a single decision tree cannot be used for training and testing the model. Different decision trees are given access to the data in the form of subsets. The most accurate estimate is obtained by averaging the outputs of the several decision trees.

K-Nearest Neighbours, the third regression-based technique, makes predictions by looking at pairs of neighbouring locations. In order to forecast a label, this method follows its origin to the K-nearest point. Prediction of a continuous value is made by averaging K observations. The training set error is used to determine the optimal K value for the model [12].

In regression-based models, the fit is evaluated using root mean square error (RMSE). The formula for root-mean-square error is:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

y_i - i^{th} Actual value
$\hat{y}_i$ - i^{th} Predicted value
n - Total data values

The deviation of predicted values from actual values gives an absolute number. The RMSE value is considered small when the predicted values are closer to actual values. For better prediction, the RMSE value is generally chosen close to zero. Using RMSE value, regression-based algorithms can be compared and helps to select the best regression model.

IV. RESULTS & DISCUSSION

CBIT's IGS receiver provides the raw data used to train and forecast Doppler Collision. The IRNSS L5 band data was collected on March 17, 2018, and is being used to make predictions about the Doppler Collision.

Table 1: DC event period for IRNSS 1C-1G

IRNSS Satellite pair	DC event periods			
	DC start time1 (IST)	*Time period (IST)*	*DC start time 2 (IST)*	*Time period (IST)*
1C-1G	8:37:24	1:03:21	19:29:50	1:17:15

Machine learning methods from the Scikit-learn toolkit have been used in this study. When doing data analysis, 80 percent of the data is used for training and 20 percent for testing. Parameters in a linear

ISSN: 2366-1313

regression model are negatively correlated; when the independent variable's value rises, the dependent variable's value falls. As a result, the Linear Regression model does not accurately forecast the values, as shown by the Mean Squared Error (RMSE) of 1.18 for the first Doppler Collision period and RMSE of 3.58 for the second.

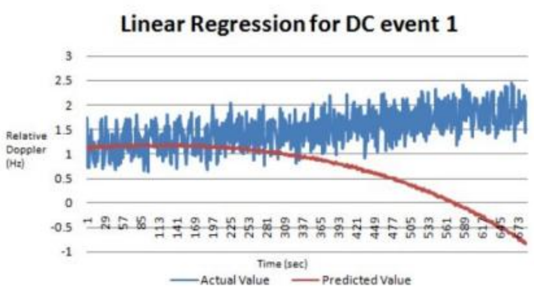

Figure1: Results for Doppler event1 using linear regression

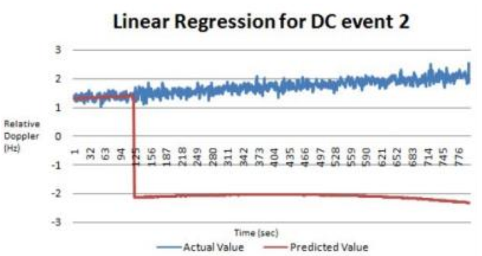

Figure2: Results for Doppler event2 using linear regression

Random Forest model improves its accuracy by using averaging technique for predicting Doppler Collision. The RMSE value for first Doppler Collision period is 0.56 and for second Doppler Collision period is 0.47. When the RMSE value is close to 0, it means that it is suitable for predicting the Doppler Collision period.

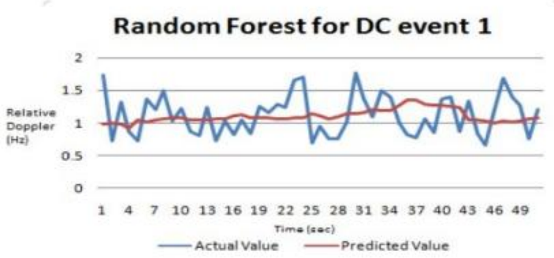

Figure3: Results for Doppler event1 using Random Forest

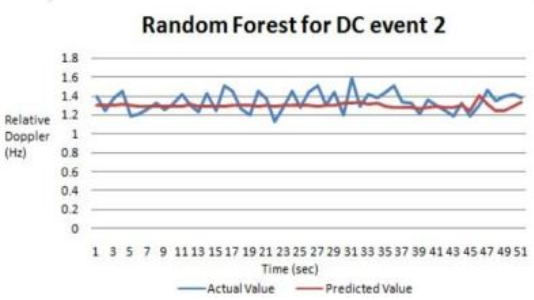

Figure4: Results for Doppler event2 using Random Forest

Due to the class imbalance in the input parameters, the K-Nearest Neighbours model always produces the same relative Doppler. A class imbalance occurs when there are fewer observations in one group of data (the positives) compared to another group of data (the negatives). For the first period of Doppler Collision, the RMSE value is 0.61, and for the second period, it is 0.46. K-Nearest Neighbours has an issue with class imbalance, therefore even if the RMSE value is near to 0, it cannot be utilised for forecasting Doppler Collision.

ISSN: 2366-1313

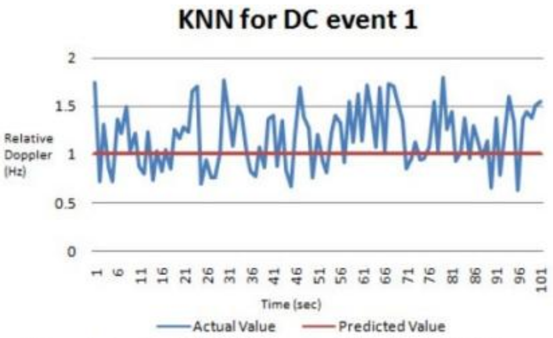

Figure5: Results for Doppler event1 using K-Nearest Neighbors

V. CONCLUSION

Analysis of three different methods for forecasting Doppler collisions indicated that Random Forest Regression had the lowest root-mean-squared error (RMSE) value. For the first period of Doppler Collision, the root-mean-squared error (RMSE) of the Linear Regression model is 1.18, whereas for the second period of Doppler Collision, it is 3.58. Using a Random Forest regressor model, the RMSE values for the first and second Doppler Collision periods are 0.56 and 0.47, respectively. The root-mean-squared error (RMSE) of the K-Nearest Neighbours regressor model is 0.61 for the first Doppler Collision period and 0.46 for the second. A Doppler collision between 1C and 1G takes place after 1 hour, 3 minutes, and 21 seconds during event 1, and after 1 hour, 17 minutes, and 15 seconds during event 2. Time and the location and speed of the satellites are found to have a bearing on the relative Doppler effect. To take this further, the RMSE value may be enhanced by using digital filters in conjunction with Machine Learning algorithms.

ACKNOWLEDGMENT

The authors would like to acknowledge Space Application Centre (SAC), ISRO Ahmedabad, for the provision of IGS receiver to NCRC laboratory in the department of ECE, CBIT, Hyderabad and Dr. A.D Sarma, Director R&D, for his technical support.

REFERENCES

[1] Asghar Tabatabaei Balaei, Dennis M Akos, “Doppler Collision and cross correlation interference in GPS receivers”, IGNSS Symposium 2009, December 1 to December 3, 2009.

[2] Vimal Kumar Bhandari, Kyle O’Keefe, "Characterization of Doppler collision and its impact on carrier phase ambiguity resolution using geostationary satellites", GPS Solutions, Volume 21, Issue 4, Page number 1707, October 2017.

[3] “What is Doppler Collision and is it a problem in GNSS?”, Inside GNSS, January 18,2016.

[4] Vijayalakshmi K, Yoshitha G, Sahithi B, Sathish P, “Preliminary Analysis of Doppler Collision Occurrence in various Kinematic Conditions for NavIC system”.

ISSN: 2366-1313

[5] Sathish P, Krishna Reddy D, "Predictive Data Optimization of Doppler Collision Events for NavIC system".

[6] Elliott, D. Kalpan, "Understanding GPS Principles and Applications", Artech House Publishers, Boston & London 1996. ISBN 0-0890006-793

[7] IRNSS "Signal in Space Interface Control Document for Standard Positioning Service", version 1.0, ISRO, June 2014.

[8] Amanpreet Singh, Narina Thakur, Aakanksha Sharma, "A review of supervised machine learning algorithms".

[9] H. Shakouri, G. R. Nadimi, F. Ghaderi, " Fuzzy linear regression models with absolute errors and optimum uncertainty".

[10] Khushbu Kumari, Suniti Yadav, "Linear regression analysis study", Journal of the Practice of Cardiovascular Sciences, Volume 4, Issue 1, January to April 2018.

[11] Gerard Biau, "Analysis of a random forest model", Journal of Machine Learning Research, Volume 13, Page number 1063, December 2012.

[12] Kashvi Taunk, Sanjukta De, Srishti Verma, Aleena Sweta Padma, "A Brief Review of Nearest Neighbour Algorithm for Learning and Classification"