Improved Orbit Prediction using Gradient Boost Regression Tree

Akshita Saxena, Satyakam Baraha, and Ajit Kumar Sahoo *Department of Electronics and Communication Engineering, National Institute of Technology, Rourkela, Odisha, India*
akshi.saxena8@gmail.com, satyakambaraha93@gmail.com, ajitsahoo@nitrkl.ac.in

***Abstract*—Orbit prediction is crucial and important for satellite tracking. To improve prediction, a person must be well equipped with knowledge of the earth's gravitational pull, atmospheric drag, radiation pressures, basic manoeuvring objects, and other information. Thus, orbit prediction has advanced in physics-based models. Most of the time, the above-mentioned information is not publicly available. Data related to satellites is kept with the respective space organizations. Using this concept, the proposed approach employs gradient boost regression trees (GBRT) with two-line element (TLE) data and, when compared to recently developed machine learning techniques such as artificial neural networks (ANN), support vector machines (SVM), and Gaussian processes (GP), it provides improved orbit prediction accuracies in terms of position and velocity. Further, the proposed method avoids the overfitting issue and shows better approximation ability. The simulations are carried out for a total of six resident space objects in low earth orbit, medium earth orbit, and sun synchronous orbit.**

***Index Terms*—Two line element, GBRT, machine learning, orbit prediction.**

I. Introduction

There is a rapid escalation of collision alarms between resident space objects (RSOs). The number of RSOs are also increasing rapidly [1]. Developing health monitoring for orbit prediction is the biggest challenge for space situational awareness (SSA). The 2009 collision between Cosmos 2251 and Iridium 33 demonstrates the necessity for extremely precise forecasting skills [2]. Satellite tracking is heavily dependent on physics models, and accuracy is somehow compromised as there is a scarcity of necessary data. Data like atmospheric drag, radiation pressure, gravitational pull, and basic information about manoeuvring objects. Using machine learning techniques, the current study on tracking management strategy is based on predictive models. The National Aeronautics and Space Administration (NASA) and the North American Aerospace Defense Command (NORAD) gather TLE data using radar to transmit a set of orbital elements. Encoded TLE can provide an accurate description of the satellite's orbit around Earth.

Machine learning (ML) methods have been used with great success for satellite orbit prediction. By studying complex patterns, predictions can be made for unobserved data. Researchers in this field have used various ML techniques. One of the most common and well-known methods used is SVM [3]. Similarly, SVM- and TLE-based methods have been used in [4] to improve trajectory prediction accuracy. In [5], authors made use of SVM in order to increase the accuracy of trajectory prediction by concentrating on the learning process associated with predictive modelling rather than on non-mechanical errors. Several prediction steps or phases, such as measurements, predictions, and evaluation, are captured by a model. But if the number of propagation days increases, there is a significant decrease in SVM performance. In separate studies, two different ML approaches, including ANN [6] and GP [7], have been used to get better predictions. All algorithms have unique approximations. In this study, the proposed method is implemented using GBRT in order to forecast the satellite orbit by utilizing TLE data from various LEO, MEO, and SSO satellites. The data is taken from the space track [8]. The trajectories are predicted by taking six Kepler TLE parameters that describe orbital angles. The position and velocity vector calculations are utilized for satellite orbital navigation. The simulations show that the effectiveness of prediction has been enhanced by employing GBRT. When compared with the state-of-the-art approaches, the proposed method gives better results, thereby minimizing the error to a great extent.

The remaining portion of the paper is structured as follows: The related works on orbit prediction are briefly summarized in Section II. In Section III, the proposed technique is described, including the design of training as well as targeted variables and a brief description of the algorithm. The simulations in Section IV compare the four ML algorithms with the proposed method, followed by a depth analysis. Finally, the conclusion and future work are mentioned in Section V.

II. Related Works

In this section, we dive into orbit prediction techniques and simultaneously investigate their merits and drawbacks. In recent years, there has been a significant amount of research on predicting satellite orbits using historical data.

A. Simplified General Perturbations-4 (SGP4) Model

The simplified general models are a collection of five mathematical models (SGP, SDP, SGP4, SDP4, SGP8, and SDP8) for calculating the orbital state vectors of RSOs with regard to the earth-centered inertial (ECI) coordinate system. The above model groups are commonly known as SGP4. The SGP4 model was developed by Lane in 1965, and it came into operation in the 1970s [4]. It is an advanced space surveillance system with space object inventory missions, inventory data

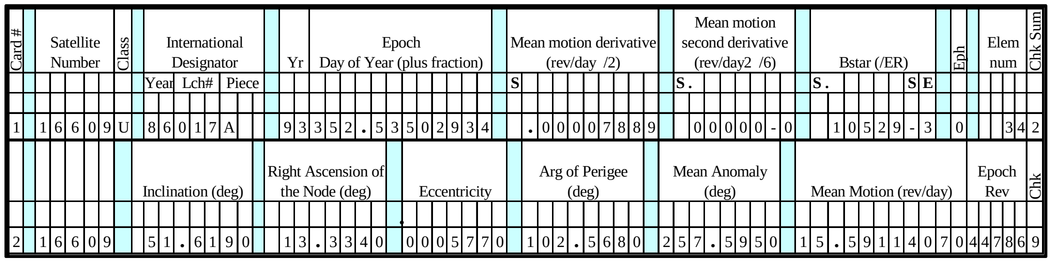

Fig. 1: TLE Data Format.

maintenance, tracking capabilities, and updates for space objects and orbital elements. This model needs to be fed data in TLE format, which gives the most accurate results when used with the SGP4 model [9]. Data from TLE is recognised as the most comprehensive cataloguing system of space objects, with information updated every 1–2 days for probable objects. Critical targets, on the other hand, are refreshed 2–3 times daily. TLE data sets are provided by NORAD and NASA [8]. However, increased propagation duration reduces the precision of the SGP4 model [4]. This is because the validity of TLEs is restricted to a certain range. Thus, the current system of the SGP4 model using TLE is not sufficient and requires appropriate prediction techniques [10].

B. Prediction Techniques

After physics-based models, ML techniques are used in the orbital propagation model. The training methods used are neural networks [6], SVM [11], and Kalman filter [7]. Using perturbation theory, neural networks enhance the positioning precision and velocity of interplanetary objects. The results suggested that the combination of these methods reduced the positioning inaccuracies. It was demonstrated that the use of training methods is effective and appropriate for the trajectory propagation model. The Kalman filter was implemented with an emphasis on data mining and the extraction of unknown power information, whereas the extended Kalman filter (EKF) computed the reproduction of a trajectory [7], thereby increasing the positional accuracies. SVM has shown exceptional capacity to enhance the precision of trajectory prediction [12]. Although SVM performances improve with sufficient data, processing little or huge volumes of data is beyond the capabilities of this system [13]. Hence, we must use other methods to solve this problem.

III. Proposed Method

Studies show that researchers may not be able to make predictions based on physics, but they are able to use historical data and learning techniques to obtain information about space objects. This section explains the TLE dataset, the proposed GBRT scheme, and the dataflow as mentioned in the following subsections.

A. Dataset

The TLE dataset is used in this paper. It is a standard format to describe the orbits and trajectories of satellites. With a suitable forecasting formula, the state at any point in the past or future can be estimated with some degree of accuracy. TLE data is a format for transmitting the components of a single set of encoded trajectories that represents a spacecraft's orbital position around the Earth. The TLE catalogues can be obtained from Celestrak [14]. Fig. 1 depicts the TLE of RSO number 16609. The TLE data format includes the following fields: satellite number, epoch year and day when TLE was extracted, drag term coefficient known as 'Bstar', inclination, which means orbital tilt, right ascension of node, shape of orbit in terms of eccentricity, orientation of ellipse in plane known as argument of perigee, and position of orbiting body.

B. Gradient Boost Regression Tree

Boosting is a modeling technique that builds a stronger classifier out of a weaker one. This is done by modeling the set of weak models. Initially, a model is created based on training data. Then that model is built in such a way that it corrects the error in the first model. This approach is repeated until the entire training dataset predicts the correct output or the maximum number of models has been added.

Gradient boosting is the technique through which the predictor rectifies faults in the prior model. Every predictor receives its training by having the residual error from the model that came before it serve as a label. As shown in Fig. 2, there are total N number of trees. The first tree is trained using

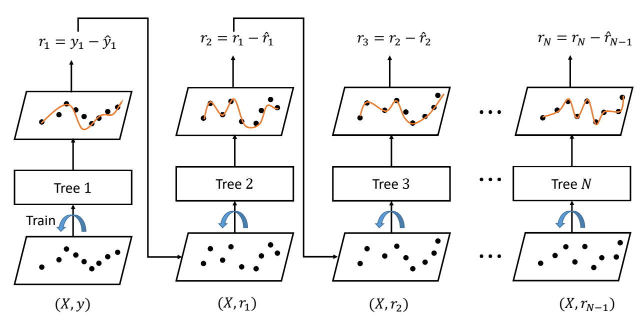

Fig. 2: Gradient Boost Regression Tree.

feature matrix X and label y. The prediction $\hat{y}_1$ is used to determine the residual error on the initial training set. Likewise, the subsequent trees are trained with the help of feature matrix X and the tree residual errors of the preceding trees. The predicted results are then utilized to evaluate the next remainder. The process repeats itself until we get nearly correct output or all trees forming the ensemble have been trained. Shrinkage describes the forecast for each tree in an ensemble that is decreased after being multiplied by a learning rate (η). The learning rate is between 0 and 1. The learning rate must increase the total estimate for the model to perform well which is set only once. Each tree makes a prediction, and the final prediction is provided by the formula below:

$$y_{\text{pred}} = y_1 + \eta r_1 + \eta r_2 + \cdots + \eta r_N. \quad (1)$$

Majorly,η in range of 0.1 to 0.3 gives best results.In case of base model(XGBoost), default value is set at $\eta = 0.3$.

C. Data Flow

The general structure of the model that has been proposed in this paper is shown in Fig. 3. The individual components and processes are described in their respective sections.

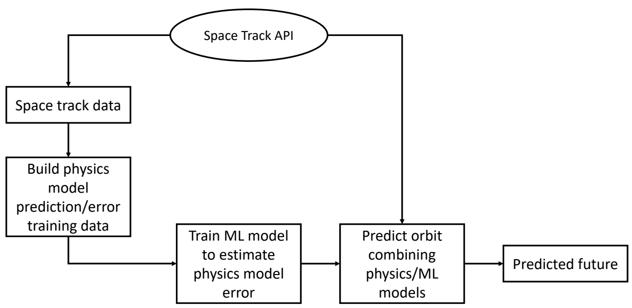

Fig. 3: Algorithm of GBRT.

D. Methodology

Orbital data is utilized from the space track website and application programming interface (API). The data is presented in the TLE format, which has fixed dimensions and has Keplerian orbital elements of RSOs at a particular point in time. The TLE data is passed, and then the position (r) and velocity (v) of the orbital state are calculated. Given an orbit data point for an ASO, we find all the orbit data points for that ASO that are within n days after the given data points. We then create a physics model starting at the given orbit data point and propagate the orbit data point and propagate the orbit to all the data points that are within n days in the future. We use the orbit data points as ground truth to determine the error in the propagation of the physics model. The training set builder requires the poliastro astrodynamics library to build a training data set including predictions and errors produced by a physical model. GBRT machine learning model is trained to estimate their prediction error. The training set is built as follows:

1) Provide information of all simulated RSOs as specified in TABLE I.
2) Find all the data points for the RSOs within n days from the given data point.
3) It builds a physics model starting at a given trajectory data point and propagates it to all points with n days in the future.

TABLE I: Information of simulated RSOs.

RSO #.	NORAD ID	SIMULATED RSO	ORBIT TYPE
1.	27944	LARETS	Sun Synchronous Orbit (SSO)
2	22824	STELLA	Sun Synchronous Orbit (SSO)
3.	39452	SWARM - A	Low Earth Orbit (LEO)
4.	16908	AJISAI	Low Earth Orbit (LEO)
5.	20026	ETALON -2	Medium Earth Orbit (MEO)
6.	19751	ETALON-1	Medium Earth Orbit (MEO)

4) Use orbital data points as ground truth to check for errors when propagating the physics model.

The ML module estimates the error allowed by the physical model in trajectory prediction by constructing a gradient boost regression tree using XG Boost as a base model. Consequently, the physical trajectory model is combined with the machine learning model by replacing the physics-predicted state vector with the amount of errors predicted by the ML model.

IV. Experimental Results

This section describes the experimental design of the proposed study. The TLE data are extracted from six satellites, i.e., two from each LEO, MEO, and SSO. The US government provides an API to download data sets. An account needs to be created on Space-Track for this purpose. The raw data is fetched, and extract, transform, and load (ETL) operations are performed on it. Every satellite has its own NORAD ID, and these IDs are used to fetch the data. After that, an ML training data set is built, and an ML model is trained to predict the values. The proposed approach is a gradient-boosted regression tree, which is executed via the XG Boost package. For training the models, 80% of the data is utilized, and the remaining 20% of the data is used for model evaluation. The proposed method is implemented using jupyter notebook platform, Python programming language, and the XG Boost package to execute the GBRT algorithm.

A. Quantitative Evaluation

Different metrics are used in the field of orbit prediction for quantitative evaluation. Some authors, such as Peng and Bai [5] preferred a single performance metric that they introduced in previous work, while others used common metrics such as root mean square error, mean square error, and mean absolute error. An intuitive and easy to use metric is the performance metric introduced in [5] as follows:

$$\mathrm{P_{ML}}(\text{in } \%) = \frac{\text{sum of residual error}}{\text{total error}} \times 100 \quad (2)$$

The value of $\mathrm{P_{ML}}$ should ideally be as low as possible for optimal performance.

B. Comparison and Evaluation of proposed ML model with other models

This research applies the GBRT technique to TLE data in order to forecast the orbital position (P_x, P_y, P_z) and velocity (V_x, V_y, V_z) vectors of a spacecraft. Visualization the satellite position is shown in Fig. 4. The performance metric of positional and velocity error components are calculated using (2) and presented in TABLE II. Considering the SSO satellites,

TABLE II: Performance metrics of position and velocity components of RSOs.

RSO #	P_{ML} (e_x)				P_{ML} (e_y)				P_{ML} (e_z)				P_{ML} (e_{vx})				P_{ML} (e_{vy})				P_{ML} (e_{vz})			
	SVM	ANN	GP	GBRT	SVM	ANN	GP	GBRT	SVM	ANN	GP	GBRT	SVM	ANN	GP	GBRT	SVM	ANN	GP	GBRT	SVM	ANN	GP	GBRT
1.	76.7	10.9	18.5	**3.2**	32	4.3	6.4	**1.8**	78.4	86.8	99	**17.6**	53.9	53.9	51	**23**	38.8	32.2	35	**7.1**	79.4	116	115	**53**
2.	76.7	**19.5**	33.8	43	27	**4.7**	4.9	5.3	92.3	81.8	104	**4.3**	75.2	74.3	69.5	**1.6**	52.4	42.9	43.7	**1.2**	92	97.6	102	**11.2**
3.	80	11.2	16.1	**1.8**	30.2	**5.6**	5.7	5.7	85.8	105	76	**1.7**	60.4	36.7	41.6	**3.7**	51.7	29.1	28.2	**5.9**	90.7	107	94	**1.1**
4.	98.6	76.5	46.3	**20.9**	72.4	**16**	30.9	18.7	**77.9**	108.5	78.6	96	82.9	147	115	**47.5**	71	76	59	**49**	84	105	76	**36**
5.	99.6	151	57	**8.6**	96	26	**17.3**	31.8	92.2	38.4	**28.7**	31	99.6	70.1	**27.6**	32	97.8	19.6	8.2	**4.6**	101	69	42	**12.8**
6.	104.3	72.6	**63**	87	100	**46**	40.7	47	99.8	41.1	50.8	**36**	101	47.4	**30.9**	33	99	49	**4.6**	7.3	98.5	61	37.5	**11.5**

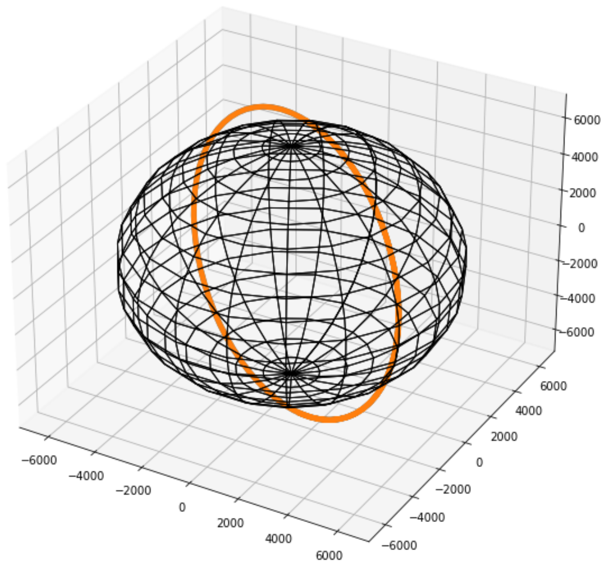

Fig. 4: Visulaization of orbit.

in the case of 'LARETS', the minimum value of P_{ML} is tabulated under GBRT. The positional errors range from 1.8 to 17.6, whereas the velocity errors range from 7.1 to 53. In case of other methods, these errors range from 6.4 to 116. Similar results are obtained when the value of P_{ML} is compared for the 'STELLA' satellite. As previously mentioned, the P_{ML} value must be low for better performance, and the proposed model has the minimum error. In the case of LEO satellites, the P_{ML} values of 'AJISAI' are taken into consideration. The P_{ML} of velocity vector is much better than positional vectors for the proposed approach. The GBRT and ANN give minimum error. The GBRT gives the lowest velocity error ranging from 36 to 49 while the other models range from 71 to 147.

The MEO satellite results vary a lot. These satellites are situated at a higher altitude, and adequate results are not captured. A large difference in error can be seen. For 'ETALON-2' and 'ETALON-1', the P_{ML} values are mentioned in TABLE II. The SSO shows much better results when compared to the LEO and MEO. The ANNs are usually easy to overfit, though they show better approximation capacity. The SVM [12] handles overfitting issues, but it cannot surpass ANN [5] and GBRT. The GP showed neutral output [15]. The proposed model displays improved performance by showing minimum errors in most of the scenarios, hence can be considered better as compared to other methods.

V. Conclusion and Future Work

This research contributes to the management of satellite tracking and monitoring by improving trajectory prediction in order to prevent collisions with debris or loss. The SSO satellites give much better results. The proposed method gives minimum P_{ML} values when compared with the other methods. The LEO satellites performance degrades to some extent as the input learning variable does not have up-to-date information regarding target variables. The MEO satellites remain at a much higher altitude. Atmospheric drag being weak will not give adequate results. The comparison is still performed in TABLE II. Most of these RSOs are with zero eccentricity. Non-zero eccentricity implies that orbits are slightly elliptical. The proposed ML approach using TLE data is capable of learning orbital trajectories from previous data. This hypothesis can be a significant improvement over current physics-based orbital prediction. The future work includes extending ML learning to many more RSOs and exploring the further drawbacks and limitations of these approaches by using a dictionary learning based approach.

References

[1] P. D. Anz-Meador, "Orbital debris quarterly news," *Orbital Debris Quarterly News (ODQN)*, vol. 24, no. JSC-E-DAA-TN77633, 2020.

[2] T. Kelso *et al.*, "Analysis of the iridium 33-cosmos 2251 collision," *Advances in the Astronautical Sciences*, vol. 135, no. 2, pp. 1099–1112, 2009.

[3] H. Peng and X. Bai, "Comparative evaluation of three machine learning algorithms on improving orbit prediction accuracy," *Astrodynamics*, vol. 3, pp. 325–343, 2019.

[4] D. Vallado and P. Crawford, "Sgp4 orbit determination," in *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, p. 6770, 2008.

[5] H. Peng and X. Bai, "Improving orbit prediction accuracy through supervised machine learning," *Advances in Space Research*, vol. 61, no. 10, pp. 2628–2646, 2018.

[6] H. Peng and X. Bai, "Artificial neural network–based machine learning approach to improve orbit prediction accuracy," *Journal of Spacecraft and Rockets*, vol. 55, no. 5, pp. 1248–1260, 2018.

[7] H.-N. Shou, "Orbit propagation and determination of low earth orbit satellites," *International Journal of Antennas and Propagation*, vol. 2014, 2014.

[8] "Space-Track.org." https://www.space-track.org/. Accessed: 2023-03-06.

[9] D. A. Vallado and P. J. Cefola, "Two-line element sets-practice and use," in *63rd International Astronautical Congress, Naples, Italy*, pp. 1–14, 2012.

[10] Z. Y.-C. Liu, S. Tarlow, M. Akbar, Q. Donnellan, and D. Senkow, "Improved orbital propagator integrated with sgp4 and machine learning," 2021.

[11] H. Peng and X. Bai, "Machine learning approach to improve satellite orbit prediction accuracy using publicly available data," *The Journal of the astronautical sciences*, vol. 67, no. 2, pp. 762–793, 2020.

[12] H. Peng and X. Bai, "Exploring capability of support vector machine for improving satellite orbit prediction accuracy," *Journal of Aerospace Information Systems*, vol. 15, no. 6, pp. 366–381, 2018.

[13] H. Peng and X. Bai, "Limits of machine learning approach on improving orbit prediction accuracy using support vector machine," in *Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference*, p. 15, 2017.

[14] "CelesTrak Since 1985." https://celestrak.org/. Accessed: 2023-03-06.

[15] H. Peng and X. Bai, "Obtain confidence interval for the machine learning approach to improve orbit prediction accuracy," in *AAS/AIAA Astrodynamics Specialist Conference, 2018*, pp. 2131–2147, Univelt Inc., 2018.