Star Cluster Simulation

Neelesh C. A.
Dept. of Computer Science Engineering
PES University
Bengaluru, India
neeleshca26@gmail.com

Ninaad R. Rao
Dept. of Computer Science Engineering
PES University
Bengaluru, India
ninaadrrao@gmail.com

***Abstract*—Stars are mostly present as clusters rather than isolation and most of the stars are not visible to the naked eye or even very powerful telescopes. Various factors change the course of these clusters and the cluster may dissolve quickly or are present for a longer duration of time. We consider a data for a single such cluster. Some stars tend to remain in the cluster and some stars leave the cluster. We predict whether a particular star will tend to remain in the cluster or leave the cluster after certain amount of time. We focus on the non-astrophysical point of view for finding the position of the stars and focus mainly on how we try to implement different models that could solve the above mentioned problem.**

***Index Terms*—star cluster simulation, predictive analytics, exploratory data analytics**

I. Introduction

Depending on a variety of conditions, star clusters may dissolve quickly or be very long lived. The dynamical evolution of star clusters is a topic of very active research in astrophysics. Some popular models of star clusters are the direct N-body simulations [1, 2], where every star is represented by a point particle that interacts gravitationally with every other particle. This kind of simulation is computationally expensive, as it scales as O(N^2) where N is the number of particles in the simulated cluster.

An N body simulation is one where a dynamic system of particles are simulated, interacting with each other through physical forces, such as gravity in our case. The types of particles can range from stars in the universe to atoms in a gas cloud. A star cluster is a group of stars. There are two types of star clusters, globular clusters and open clusters. Globular clusters are large number of stars that are tightly bound by gravity while open clusters are more loose. We are going to work on a globular cluster. Since the lifetime of star clusters can be billions of years, actually observing them is not possible. Hence, there is a need to predict the star clusters at a particular period of time. If it is done for many star clusters, researchers can have better knowledge about the origins and evolution of the galaxy.

At present, star cluster simulations are done on supercomputers that have numerous GPUs and specialized hardware such as GRAPE. Most people do not have access to these resources, and hence the number of people that can work on them are not too many. If we're able to build a suitable model, it would advance the field by a significant amount.

The data set consists of snapshots of the cluster for 19 different time stamps and each time stamp is separated based on the standard mentioned in N-body units[4]. The data set contains information of the star such as the position along the x, y and z direction and also the velocities in the three directions. It also contains the mass of each star and the id of the star. For the purpose of the simulation, the mass of each star is assumed to be the same across the cluster, i.e. 1/64000. At time t=0(reference time), there are 64000 stars that are present and this number reduces as time passes and at t=19 units, there are 63970 stars. This shows that some of the stars have already left the cluster. We propose to build a model that finds whether a star will remain in the cluster or will leave the cluster at t=19 and further and we will test on t=19. The result of this experiment will also depend on various other factors such as collision between two stars which would drastically change all the attributes of a given star and whether it will remain in the cluster or not. For the purpose of prediction, we will not be considering such external factors and assume the trend is general. There is another way to split the testing and training data set . We can take a set of stars as the training set and train it over all the 19 time stamps. The testing data is the rest of the stars and their initial values, for which we predict their values over the rest of the time stamps. This can be used in a case when a star enters a cluster. Most of the researches have considered this for the field of astrophysics and we focus this data for the purpose of data analysis and predictive modelling.

The data set uses direct N-body simulation of a star cluster, hence other methods can be compared with it. Stars go out of the cluster when they reach escape velocity, and this happens when there are close encounters with other stars, i.e. the distance is small. On the scale of celestial bodies, this happens for a very small period of time, which would require far too many data points to accurately predict, hence high accuracy is not expected for this case. We can use a decision tree or SVM to classify a star as whether it stays in a cluster or whether it leaves it. The easiest to implement and most accurate solution is the computation of Newton's equation for gravitational force, but that would take a significant amount of time for our data set as there are 64000 stars at the start and they do not reduce by too much.

II. Literature Survey

Most of the previous work has been done considering only the aspects of astrophysics and not the analytics point of view.

Aarseth [2] presented an O(N^2) solution where every star is represented as an individual that could interact with every other particle. Here, N is the number of stars in the cluster to be considered. The main principle behind this computation is the Newtons law and any additional potential field if present. The problem is thus a set of non-linear second order differential equations relating the acceleration with the position of all the particles in the system. Once the conditions are specified, model can be constructed accurately. Though this solution could be highly accurate but the computation power is very high and the time it takes to find whether a particular star will remain in the cluster or not could take a really long time.

The center for Astrophysics and Super computing at Swinburne University of Technology[3], have built upon the work done by Aarseth and have done some collaboration which was aimed at producing realistic models of star cluster models. A semi-analytical criterion is utilized to detect and evolve stable hierarchical triple and quadruple systems that otherwise would prove extremely time-consuming by direct integration. An important aspect is that stellar and binary evolution are performed in step with the dynamical evolution so that interaction between these processes is modelled consistently. They have considered collision of particles for a star to reach escape velocity and leave the cluster and many other external factors that are too deep for a analytics point of view.

In one of the kaggle repository, an animation for the position of the cluster has been done depicting how the cluster changes with time. Some of the kaggle repositories have done basic stock taking such as finding the summary statistics for the cluster at different time intervals. Some other repositories have found which are the stars that have already escaped the cluster. None of the repositories mentioned above have predicted which is expected to leave the cluster after a certain time.

A binary classifying Deep Neural Network model was built to predict the stability of circumbinary planets in [5]. It is trained on a N-body simulations generated with a data set called REBOUND N-body integrator. This model has an accuracy of around 86 percent, and it does not go below 86 percent even when tightly surrounded by instability.

III. Problem Statement

There are two main questions to be addressed with respect to star clusters. One is, whether a star will remain in the cluster. And if it does leave, when it will do so. The other question is, to predict the position and by extension, the velocities of the stars at a particular point in time. Other details about the cluster like centre of mass of the cluster can be obtained by accurately answering the aforementioned questions. From the figures shown, it is clear that stars tend to change their position every time a collision happens and the figures shows the evolution of the same star cluster for 19 different time-frames. Some of the stars leave the cluster immediately while others might or might not leave i.e. they tend to move within the cluster or reach their escape velocity and leave the cluster.

A faster solution than the direct N-body simulation is the BarnesHut simulation/approximation [6]. The Barnes-Hut simulation considers a group of stars that are far away as one single body, using their centre of mass. This speeds up execution, as only one body has to be looked at, instead of each individual star. The stars that are close by are still looked at as individual entities. The space is divided into an octree, where the root node represents the whole space and its eight children represent the 8 quadrants. Once all the points are inserted, each node contains either 0 or 1 point. The parents contain the centre of mass and total mass of their children. Depending on the parameter that is set for what should be considered long range/short range, the accuracy/computation require varies.

There are other methods of simulation such as the particle mesh method and the fast multipole method. The use of machine learning and deep learning in this field is fairly recent [5]. There is an assumption that direct N-body simulation is the most accurate simulation available. But there can be no way to conclude that due to the observable data being so small. The equations that are used for the simulations may not be able to take into account every possible interaction between the objects, i.e. apart from gravity. A model might be able to recognize features that the formula might not have considered. Hence, the model might be able to create a better simulation, at least with respect to computational power vs accuracy.

IV. Analysis done so far

Since the simulation of this data set has considered the mass of all the stars to be equal, we may be able to use that for better predictions in this data set. Although it may not translate well to a realistic situation where the masses are different. Comparisons can be done between the two, i.e. one where mass is considered and one where it is not to see if there is any significant difference in our case.

For now, we have added velocity and distance to our columns as they are non linear in nature and may not be captured by the models. Some more parameters may be added to create better models.

All the columns of the data have a correlation of nearly 0 (apart from distance and velocity (those were computed additionally) that has a value of -0.45, which is to be expected). Although auto correlation is expected in such a case, the time stamps are at a large enough distance that their effect is uncertain. Taking these things into consideration, regression seems like it would be a suitable model. Logistic regression for predicting whether a star would stay in a cluster or not. Polynomial or Linear Multi-target regression for the position/velocity.

Basic stock taking such as finding the relation between the mean, median distance with respect to the time frame and mean, median velocity with respect to time frame has been done and the pattern is quite irregular to conclude anything. All methods for prediction will be tried to best fit and predict the data and many more analysis will be done to check for any other relation between the variables.

A 3D plot of the centre of mass of the cluster was plotted but although a slight trend was seen, the values did not change by any significant amount.

The data was examined for normality. The velocity in every direction was found out to be normal by examining the histogram, qqplot and the values of kurtosis and skewness. This was not the case for the positions. The histogram did not have any glaring outliers. But the qqplot was nearly horizontol and the kurtosis was extremely high.

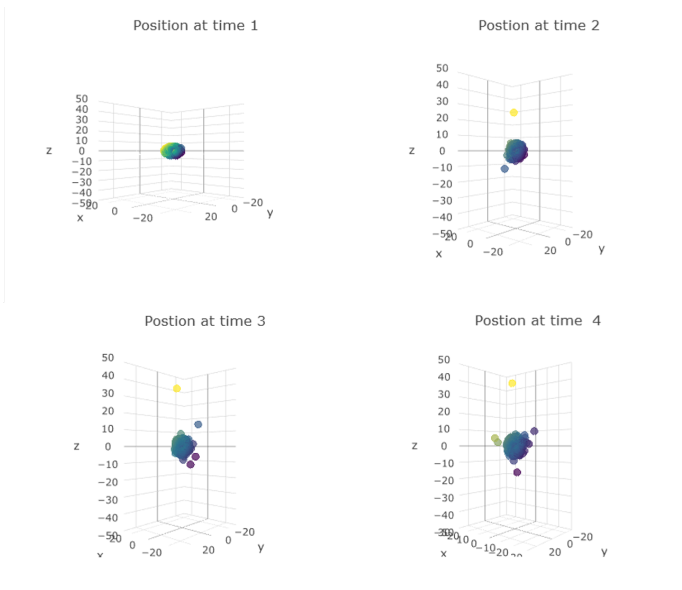

Fig. 1. The change in the position of stars at given timeframes

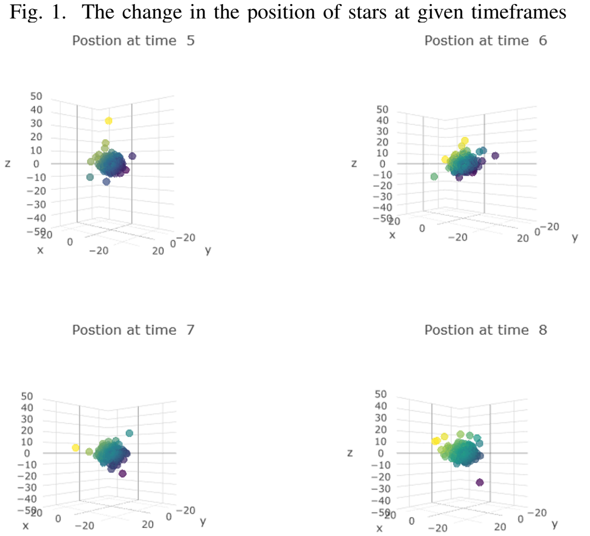

Fig. 2. The change in the position of stars at given timeframes

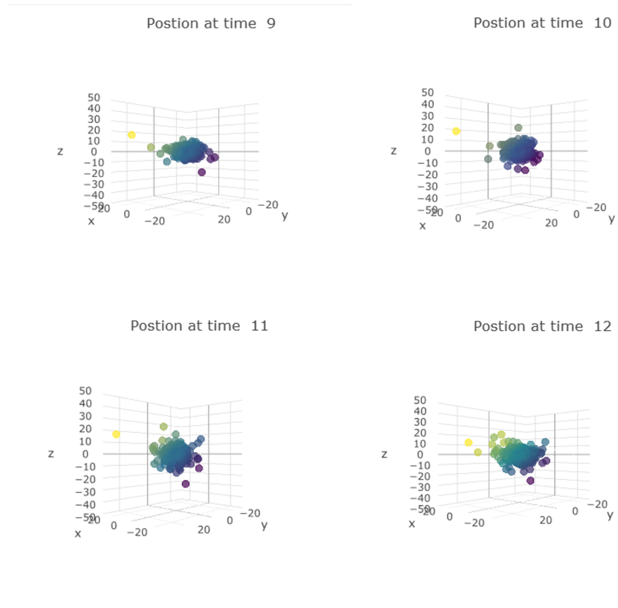

Fig. 3. The change in the position of stars at given timeframes

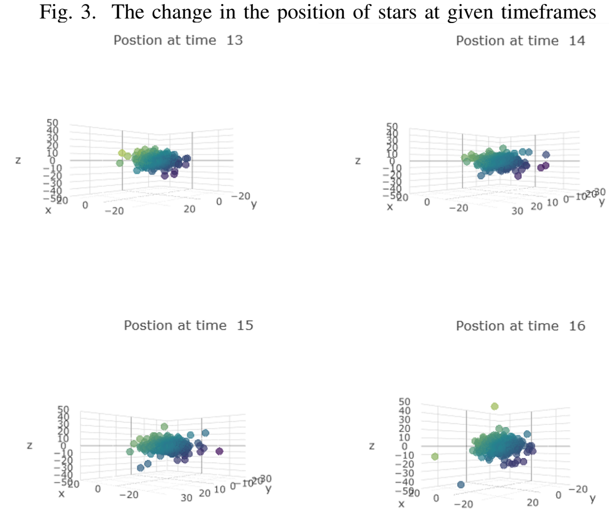

Fig. 4. The change in the position of stars at given timeframes

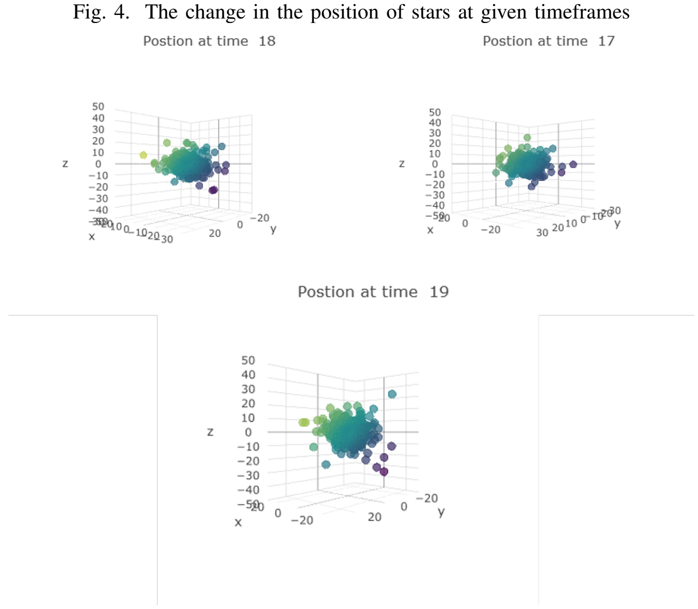

Fig. 5. The change in the position of stars at given timeframes

An attempt will be made to increase the accuracy of the prediction as compared to [5] which has an accuracy of 86 percent. Since different models could give different accuracy and the best will be kept.

REFERENCES

[1] Heggie, D., Hut, P. 2003, The Gravitational Million-Body Problem: A Multidisciplinary Approach to Star Cluster Dynamics Cambridge University Press, 2003

[2] Aarseth, S. J. 2003, Gravitational N-Body Simulations - Cambridge University Press, 2003

[3] The center for Astrophysics and Supercomputing at Swinburne University of Technology, N-body star cluster models, http://astronomy.swin.edu.au/ jhurley/nbody

[4] Heggie, D. C., Mathieu, R. D. 1986, Lecture Notes in Physics, Vol. 267, The Use of Supercomputers in Stellar Dynamics, Berlin, Springer

[5] Christopher Lam, David Kipping; A machine learns to predict the stability of circumbinary planets, Monthly Notices of the Royal Astronomical Society, Volume 476, Issue 4, 1 June 2018, Pages 56925697, https://doi.org/10.1093/mnras/sty022

[6] BarnesHut simulation/approximation https://www.nature.com/articles/324446a0