International Journal of Future Generation Communication and Networking
Vol.14, No. 1, (2021), pp. 4053 – 4059

Hit Song Prediction: Using Ant Colony Optimization for Feature Selection

Prashant Giridhar Shambharkar[1], Abhijeet Singh[2], Akash Yadav[3]

[#]*Department of Computer Science & Engineering, Delhi Technological University*

[1]prashant.shambharkar@dtu.ac.in

[2]abhijeets809@gmail.com

[3]akashyadav92133@gmail.com

Abstract

Hit Song Science is a binary classification problem which aims at anticipating the success of a song before its release. This is particularly useful to identify talented singers, musicians, lyricists etc. before they get contracts from Music labels. In this work, we considered the Hit Song Science problem as a classification problem and attempted to solve it using a feature selection technique known as Ant colony Optimization to identify most prominent and definitive features and improve upon the accuracy of previous work done on this problem statement. We have used Spotify Web API to extract acoustic features of almost 6000 songs accumulated from online repositories. We test multiple classification models such as XG Boost, Random Forest etc. on our dataset. The results demonstrate that, as compared to existing approaches, our approach selects a minimal number of features and achieves better performance.

Keywords— *Machine Learning, Swarm Intelligence, Hit Song Prediction, Ant Colony Optimization, Feature Selection*

I. INTRODUCTION

Recently, Hit Song Science has emerged as one of the most important and active research topics in Music Information Retrieval and this research is particularly relevant to record labels. Record and Music labels can use this model to determine the songs to produce and devise an effective advertisement and publicity campaign to generate good revenues and greater return on investment. In this work we have considered the following definitions for a Hit and Non-Hit Song:

Hit Song: If a song is listed on the Billboard Top 100 chart at any given moment then it is referred to as a Hit Song.

Non-Hit Song: If a song has not been listed on the Billboard Top 100 chart at any given moment, then it is referred to as a Non-Hit Song.

The hit song prediction task has been approached in various ways such as using only acoustic features extracted from the audio such as acousticness, loudness, key, rhythm, mode etc., using external features like popularity, date of release etc. gathered from various music streaming applications such as Spotify, using data gathered from social media websites such as Twitter to access the current trending music across the globe, and treating this problem solely as a Natural Language Processing task which uses the lyric feature of the song only.

The procedure of selection of a minimal number of features that enhance the performance of machine learning models in terms of accuracy, precision etc. is referred as Feature Selection (FS). The basic assumption in FS techniques is that the dataset must have irrelevant or redundant features. The subgroup of optimal features is produced by a heuristic search procedure which employs a search space that contains a number of states. Each state defines a candidate subset to be evaluated further. Search strategy and Search starting point are primary factors considered and evaluated for feature subset generation. Forward Selection, Bi-directional and Backward Elimination are the search strategies currently utilized in FS tasks. In forward selection, appropriate features are repetitively appended on an empty subset (search start point) to obtain optimal subset whereas in backward elimination, the optimal subset is generated after repetitively eliminating irrelevant features from a complete set of features (search start point). In bidirectional selection, search begins with both the sides that append and remove features concurrently. Search may begin with randomly selected features to avoid local optimum region. There are three approaches for feature subset selection, wrapper, filter and hybrid.

In this research, Ant Colony Optimization was incorporated for feature selection. ACO is a swarm intelligence method prompted from the behaviour of scout ants searching for resources around the ant colony. The algorithm helps in the identification of distinguished and definitive internal and external acoustic features in this task. This research is a first of its kind to integrate swarm intelligence in Hit Song Science.

According to the authors' knowledge, no prior research has employed swarm intelligence algorithms like ACO on the hit song prediction problem. The remainder of the paper is laid out as follows: Section 2 discusses the related work on hit song science and feature selection approaches using ACO. In Section 3, we discuss the ACO approach. In Section 4, we give a brief description of the dataset. The experiment setup and results are discussed in Section 5. Sections 6 draw conclusions and address shortcomings as well as future scope.

II. Literature Review

A. *Hit Song Science*

Machine Learning is one of the most used approaches in search of an accurate and precise solution to the Hit Song prediction task. In this paragraph we review the previous research work done on Hit Song Science. One of the earliest researches done on HSS by Dhanraj and Logan [1] analysed each track's acoustic features as well as lyric data and to discern hits from non-hits, regular classifiers such as SVM and boosting classifiers were used. The study produced exemplary results showing the importance of lyric feature more than acoustic feature in identification of hit songs and identified new ways to approach the task however it did not discuss much about the dataset, feature set and experimental setup details. In 2008, Pachet and Roy [2] in their research suggested that the music features used in music analysis during that time period were not informative enough and concluded that the previous study produced by Dhanraj and Logan were based out of spurious data or biased experiments. Bischoff et al. [3] work in 2009 introduced a framework for forecasting the success of tracks by using social activity and annotations without considering inherent track characteristics. To assess the prospects of the track they directly applied data mined from a social music network. Another research conducted by Borg and Hokkanen [4] (2011) proposed a method of predicting potential success of a song based on their YouTube video count but the experiments were unsuccessful and the authors drew inference analogous to Pachet and Roy. In 2011, Ni et al. [5] conducted experiments on a dataset consisting of the UK top 40 singles chart and predicted that if a song that was lower in the standings at the time reached the top 5 spots in the UK top 40 singles chart. They were able to achieve promising results and insisted that HSS is an active scientific problem statement. However, the research lacked proper information about the dataset and experimental setup. Research conducted by Fan and Casey [6] in 2013 used The Echo Nest to obtain the audio features of the track and then employed two Machine Learning algorithms, Linear Regression and Support Vector Machine, following which they compared the results of the Chinese hit songs with UK hit songs and concluded that Chinese hit song prediction had better accuracy than the UK hit song prediction. In 2014 Herremans et al. [7] predicted if a track would become a top 10 hit or slide to a lower position. They were able to achieve positive results which they possibly accounted for using advanced temporal features and using songs of the same decade. In 2015 Jang Hyuk Lee et al. [8] predicted the music popularity based on proposed popularity benchmarks such as Max, Debut etc. computed on the basis of long term real-world ranking data of the track listed on Billboard Hot 100 charts. The authors analysed the growth pattern of the songs on Billboard charts. The results indicated that each popularity metric exhibited a distinct distribution. In 2017, Yu et al. [9] constructed a siamese convolutional neural network (CNN) model that considered the HSS as a ranking problem rather than a classification or regression problem. They have used a A/B sampling technique to alleviate the data imbalance and were able to achieve much higher accuracies than the baseline regression model. In 2019, Zangerle et al. [10] used a combination of high- and low-level acoustic features and made use of wide and deep neural networks. The authors concluded that the coalition of low- and high-level acoustic attributes in a network architecture delivers better performance. Recently, Herremans and Bergmans [11] (2020) incorporated both the audio features and early adopter data for prediction and were able to anticipate upcoming top 20 hit songs accurately. The early adopter data was generated by analysing the ways in which the popularity of songs is increased by sharing of the songs on social media. Another recent research conducted by Raza and Nanath [12] in 2020 combined the technical parameters and the sentiment analysis of the lyrics to generate the dataset and concluded that Hit Song Science is still not a data science activity. Martín-Gutiérrez et al. [13] (2020) created a dataset named SpotGenTrack as well as developed an end-to-end deep learning architecture and achieved an optimal solution to predict Hit/Non-Hit song as compared to previous studies.

B. Ant Colony Optimisation

Swarm Intelligence algorithms are a class of algorithms inspired from insects (Particle Swarm, Ant Colony), birds (bird flock, chicken), mammals (wolf pack, cat), bacteria (superbug) etc. and they've been used to choose features recently. In this paragraph we review the previous research work done on Ant Colony Optimization. Ant System (AS) was one of the first implementations of ACO by Dorigo [14] in the 1990s which was later transformed to Ant Colony System (ACS) that was utilized to unravel the Traveling Salesman Problem (TSP). In 2005, Al-Ani [15] introduced an algorithm incorporating both filter and wrapper-based approaches in ACO. The author used Mutual Information Evaluation Function (MIEF) which is a filter method to evaluate local importance of a particular feature and Updated Selection Measure (USM) which is a wrapper tool for assessing the potential of subsets. Some of the extensions of the ACO algorithm for FS require a predetermined number of features to be selected while other extensions employ some evaluation function for termination criterion. In the methods based on evaluation function, the search and selection of features is halted as soon as the stopping criterion is satisfied by the current feature subset in consideration. Therefore, in these kinds of methods the ants do not consider all the features while traversing through the graph and have a finite number of steps. To overcome this problem, Binary Ant Colony Optimization was first presented by Touhidi et al [16] in 2007.

III. Ant Colony Optimisation

Inspired from the real-life foraging behavior of ants, Ant Colony Optimization technique is a population-based metaheuristic technique introduced by Dorigo et al. [14] in the 1990's. It is a time efficient algorithm used to optimize difficult problems such as the Traveling Salesman Problem, vehicle routing, data mining, telecommunications networks etc. to find an approximate optimal solution. Ants are eusocial insects who communicate with each other using sound, touch and pheromone. Pheromone is an organic chemical compound secreted by ants on their trail to seek and search resources. When an ant discovers a food source it secretes pheromones back and forth on the route from the ant nest, this increases the probability of the next batch of worker ants to choose this path over others and the next ants increases the density of pheromones on this trail by secreting its own pheromone creating a positive feedback loop. The deposition of pheromones on ideal routes i.e. shortest path to a generous amount of resources gradually increases because each ant stochastically chooses to follow a direction having abundance of pheromone. Pheromone evaporation is a critical feature of ACO which iteratively reduces the pheromone density on longer routes thus increasing the probability of ants choosing the shortest route to reach food. This prevents the algorithm from converging to suboptimal subregions and when the maximum number of the ants follow the same route the algorithm converges to an optimal solution. The ACO algorithm investigates a greater range of options than greedy heuristic due to its stochastic choice based on local heuristic facts and artificial pheromone trails.

A. ACO For Feature Selection

The task of feature selection is formulated into a graph problem which is ideal for the application of Ant Colony Optimization. The main idea of ACO is to determine the shortest route between the ant nest and the food source. In this algorithm each attribute is represented as a vertex in the graph and all the nodes are linked to one another. In this work we have used a modified Binary Ant Colony Optimization Algorithm (BACO) inspired from [17][18]. In this modified BACO each feature node is subdivided into two sub-nodes ("1" to select and "0" to deselect the feature). Hence, we have 2 possible numbers of states for each feature therefore the number of possible links considering just two features is 4.

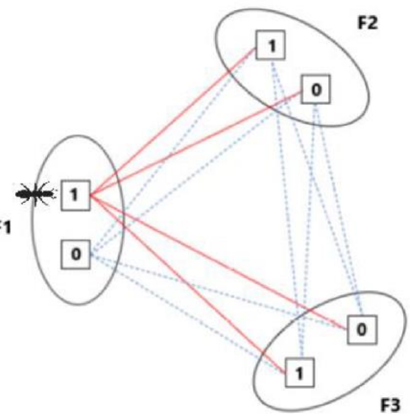

Fig. 1 Ant at sub-node “1” of Feature 1 analyzing the next edge to select (represented with red bidirectional edges)

Initially, a specified number of ants are generated and placed on random positions of the graph following which the journey of the ant starts. Ants on each node selects the next unvisited node based on a stochastic value computed using local heuristic information and artificial pheromone trails for every link between nodes.

Suppose an ant is at a feature node Fi,x (i = 1,2,3,.. n; x = 0/1) and has to travel to another unvisited feature then to choose a path to a feature Fj,y (j all unvisited features, y = 0/1) we define a probabilistic function to choose the same. The probability function P, denoting the possibility of transition of an ant at Fi,x to select the link to Fj,y is computed by combining the pheromone density τ and heuristic value η of the link. The probability of the ant k at sub-node Fi,x to choose the edge (i,j) at time t is :

$$P_{ix,jy}^{k} = \frac{\tau_{ix,jy}^{\alpha}\eta_{ix,jy}^{\beta}}{\sum_{l} \tau_{ix,l0}^{\alpha}\eta_{ix,l0}^{\beta} + \sum_{l} \tau_{ix,l1}^{\alpha}\eta_{ix,l1}^{\beta}}$$

where l, j is an unvisited attribute. α and β are the relative weights to give preference to pheromone density and heuristic value. P evaluates to 0 for visited nodes.

Correlation is a bivariate analysis that computes the intensity of connection between two variables and the direction of the relationship. In statistics different types of methods have been developed to compute correlation between two variables including Spearman correlation, Pearson correlation, Kendall rank correlation, etc. The heuristic information in this work is computed using the Pearson correlation coefficient matrix generated for every pair of attributes.

This process is completed when the ant has visited all the features. The iteration is completed when the process is computed for all the ants. After visiting all the features, each and every ant constitutes a binary string of the length equal to the number of attributes representing '1' if the corresponding feature was selected and '0' if it was deselected. The local best ant is selected over every iteration and this ant is then compared with the global best ant to select the best possible ant with optimal accuracy. In this work only the best ant (the ant that represents the optimal subset of features generating the classifier's lowest mean square error) is permitted to lay pheromones on the path it has traversed. In this type of ACO each ant visits all the features thus overcoming the drawback of specifying the number of attributes to be selected prior to the execution of the algorithm. The pheromone levels are decreased after every iteration to decrease the popularity of paths leading to suboptimal and mediocre solutions. The pheromone levels on each path are proportional to the optimality of the solutions.

IV. Dataset And Features

The dataset consisted of thousands of western pop songs of a single decade i.e. from 2010 to 2020 collected from multiple online music repositories. To assess the popularity of the songs the Billboard Top 100 chart was used. The Billboard magazine publishes one of the most authentic music industry uniform album charts in the United States for songs every week. The songs are ranked on the Billboard charts using a metric that takes into account radio airplay audience impressions, retail and digital sales info, and streaming activity from online music outlets. In the next step the unique songs on Billboard Top 100 charts for the time period of 2010 to 2020 were collected and a combined dataset with songs from online repositories was created. Subsequently, a balanced dataset of approximately 6000 songs was created by removing duplicate songs and building an equivalent group of non-hits and hit tracks. The tracks were labelled as Hit songs ("1") if they ever appeared on the billboard Top 100 chart and Non-hit ("0") otherwise. The acoustic features and external features of the songs were extracted using the Spotify Web API which is currently one of the most popular online music streaming applications. The Spotify Web API was used to gather 18 features for every audio track, of which we selected 15 features such as Danceability, Valence, Key, Mode, Speechiness etc. for our analysis.

V. Experimental Setup And Results

To investigate the performance of our proposed approach for the hit song binary classification task we conducted a series of experiments. The experiments were conducted on the dataset described in section 4 of this paper on a machine with 2.4 GHz quad-core Intel i5 processor with 8 GB of RAM. All the data points of the dataset were scaled to [0,1]. To evaluate the feature subset represented by the ant, Random Forest Classifier was used using five fold cross validation. The initial values for different parameters that were used in the ACO experiments are presented in Table 1.

TABLE I
Initial Value of Parameters

Parameter	Parameter Description	Value
m	Number of Ants	20
Iteration	Number of Iterations	50
α	Pheromone Importance Factor	0.8
β	Heuristic Importance Factor	0.2
ρ	Pheromone Evaporation Rate	0.05

Random Forest, Support Vector Machine, XGBoost and Logistic Regression are the machine learning models employed to draw differences between performance of models using all the features and other models using the features selected from the ACO algorithm and the results are presented in Table II. In this work we primarily focused on the accuracy of the results. Moreover, precision and recall were also considered as false positives, which may be more dreadful because investment in non-hit songs would cost the music labels. The Random Forest model outperforms the other classification model in this work with an accuracy of 81%. It is also worth mentioning that all the classification models exhibited improvement in accuracy when integrating features selected using the ACO algorithm. Random Forest and XGBoost models showed the most improvement with almost 9.5% increment in accuracy.

TABLE II
COMPARISON OF RESULTS OF DIFFERENT CLASSIFICATION MODELS

Classification	Without ACO			With ACO		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
SVM	0.76	0.78	0.77	0.80	0.80	0.80
Random Forest	0.74	0.78	0.74	0.81	0.82	0.82
Logistic Regression	0.75	0.77	0.76	0.79	0.79	0.79
XG- Boost	0.73	0.77	0.74	0.80	0.81	0.81

VI. CONCLUSION AND FUTURE SCOPE

The application of a swarm intelligence algorithm such as Ant Colony Optimization adds a new dimension to Hit Song Science research and Music Information Retrieval research. The experiment results show distinct improved performance of our approach over past algorithms used on this problem statement. Our method generates an optimal subset of features and achieves better classification accuracies. We acknowledge the fact that we have worked upon a relatively smaller dataset which consisted of data points partial towards Western Music. Moreover, the Billboard Top 100 charts are mostly used for the ranking of western commercial music restricted to the United States market only. We have not considered the lyric features of the audio in this work. The research work can be extended by using other swarm intelligence algorithms such Particle Swarm Optimization, Genetic Algorithm, Bee Colony Optimization etc. and generating a comparative study between different algorithms

REFERENCES

[1] Ruth Dhanaraj and Beth Logan. Automatic prediction of hit songs. pages 488–491, 01 2005.
[2] Francois Pachet and Pierre Roy. Hit song science is not yet a science. pages 355–360, 01 2008.
[3] Kerstin Bischoff, Claudiu S Firan, Mihai Georgescu, Wolfgang Nejdl, and Raluca Paiu. Social knowledge-driven music hit prediction. In the International Conference on Advanced Data Mining and Applications, pages 43–54. Springer, 2009.
[4] Nicholas Borg and George Hokkanen. What makes for a hit pop song? what makes for a pop song. Unpublished thesis, Stanford University, California, USA, 2011.
[5] Yizhao Ni, Raul Santos-Rodriguez, Matt Mcvicar, and Tijl De Bie. Hit song science once again a science. In the 4th International Workshop on Machine Learning and Music. Citeseer, 2011.
[6] Jianyu Fan and Michael Casey. Study of Chinese and UK hit songs prediction. In Proceedings of the International Symposium on Computer Music Multidisciplinary Research, pages 640–652, 2013.
[7] Dorien Herremans, David Martens, and Kenneth Sörensen. Dance hit song prediction. Journal of New Music Research, 43(3):291–302, 2014.
[8] Junghyuk Lee and Jong-Seok Lee. Music popularity: Metrics, characteristics, and audio-based prediction. IEEE Transactions on Multimedia, 20(11):3173–3182, 2018.
[9] Lang-Chi Yu, Yi-Hsuan Yang, Yun-Ning Hung, and Yi-An Chen. Hit song prediction for pop music by siamese cnn with ranking loss. arXiv preprintarXiv:1710.10814, 2017.
[10] Eva Zangerle, Michael Vötter, Ramona Huber, and Yi-Hsuan Yang. Hit song prediction: Leveraging low-and high-level audio features. In IZMIR, pages 319–326, 2019.
[11] Dorien Herremans and Tom Bergmans. Hit song prediction based on early adopter data and audio fea-tures. arXiv preprint arXiv:2010.09489, 2020.
[12] A. H. Raza and K. Nanath, "Predicting a Hit Song with Machine Learning: Is there an a priori secret formula?," 2020 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA), 2020, pp. 111-116, doi: 10.1109/DATABIA50434.2020.9190613.
[13] D. Martín-Gutiérrez, G. Hernández Peñaloza, A. Belmonte-Hernández and F. Álvarez García, "A Multimodal End-to-End Deep Learning Architecture for Music Popularity Prediction," in IEEE Access, vol. 8, pp. 39361-39374, 2020, doi: 10.1109/ACCESS.2020.2976033.
[14] Marco Dorigo, Vittorio Maniezzo, and Alberto Colorni. Ant system: optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 26(1):29–41,1996.
[15] Ahmed Al-Ani. Feature subset selection using ant colony optimization. International journal of computational intelligence, 2005.
[16] Touhidi, H Nezamabadi-pour, and S Saryazdi. Feature selection using binary ant algorithm. In First joint congress on fuzzy and intelligent systems, 2007.
[17] Shima Kashef and Hossein Nezamabadi-pour. A new feature selection algorithm based on binary ant colony optimization. In The 5th Conference on In-formation and Knowledge Technology, pages 50–54. IEEE, 2013.
[18] Shima Kashef and Hossein Nezamabadi-pour. Advanced aco algorithm for feature subset selection. Neurocomputing, 147:271–279, 2015.

ISSN: 2233-7853 IJFGCN