Towards Predicting the Popularity of Music Artists

Fabian Jetzinger, Florian Huemer, Markus Schedl

Johannes Kepler University
Department of Computational Perception
Linz, Austria
markus.schedl@jku.at,
http://www.cp.jku.at/people/schedl

Abstract. This paper explores the possibility of predicting a music artist's future popularity, quantified by how often their tracks are listened to in the past, on a daily basis. Using the LFM-1b dataset of listening histories by Last.fm users, we investigated three regression techniques to predict the amount of listening events an artist will generate per day. To this end, we adopt linear regression, support vector machines, and neural networks to create, analyze, and optimize predictions, which we finally visualize for easy exploration.

Keywords: music, regression, popularity prediction

1 Introduction

How will a music artist's popularity evolve over the next month? This question lays at the basis for our work. The creation of accurate predictions of an artist's future popularity offers a multitude of possible applications, such as in music recommendation systems or as a decision guidance for investors or music labels. To the best of our knowledge, such popularity prediction experiments in the music domain have only been conducted on rather small datasets, exploiting only content features or peer-to-peer networks, the latter having faced a substantial decrease in usage during the last few years, due to the emergence of streaming services like Spotify, Apple Music, or Last.fm. In this paper, in contrast, we exploit a large-scale dataset (LFM-1b) of more than a billion user-generated listening events. Our goal was the generation of predictions and the subsequent continuous optimization of the predictive algorithms to increase accuracy. In this comparative study, we relate the results achieved with different regression approaches to determine which method generates the most accurate predictions.

2 Related Work

Similar work has already been undertaken by various researchers. Staying within the topic of music, one highly relevant work is [8], in which Pachet and Roy use content-based features for popularity prediction of music items, with limited success though. Dhanaraj et al. [3] try to predict hit songs based on extracted acoustic and lyrical information. They ultimately conclude that their lyrics-based features produced slightly more accurate results. Similarly, Herremans et al. [2] explore the prediction of dance hit songs based on several different classifiers and a database of dance hit songs from 1985 to 2013. Furthermore, Ni et al. [7] investigate the prediction of hit songs based on the UK top 40 charts of the last 50 years, with the aim to distinguish songs with their peak positions within the top 5 from songs which peak in the top 30 to 40. Using web sources instead of audio, Schedl et al. [10] determine country-specific popularity of music artists. They investigate search engine playcounts, popularity derived from Twitter, from shared folders in the peer-to-peer network Gnutella, and from Last.fm playcounts. Their conclusion is that these sources are largely inhomogeneous and yield to different popularity scores. Koenigstein and Shavitt [6] try to forecast the Billboard charts based on search queries issued within Gnutella. They show that a songs popularity in the network highly correlates with its ranking in the Billboard charts.

In the multimedia domain, Bandari et al. [1] predict the popularity of news items prior to their release to the public, achieving an overall accuracy of 84%. Yu et al. [11] explore the effect that Twitter contributions have on the amount of views a YouTube video receives over a certain time span, differentiating between sudden increases in viewcount, named “Jumps”, and the initial viewcount a video receives shortly after its upload, named “Early”.

3 Experiments and Results

3.1 Dataset

The LFM-1b dataset [9] used in our work contains information on users, artists, tracks, and listening events. The dataset contains more than 1 billion listening events for more than 3 million individual artists. Listening events, which constitute the main building block for our experiments, are defined by a specific date and time and the corresponding information about track and user. We considered in our experiments the top 100 artists according to number of total listening events to ensure a sufficient amount of data.

Before we were able to start working on the actual predictions, we first had to aggregate the LFM-1b data and transfer it into a suitable database structure (using SQLite[1]) as we were interested in the total number of listening events per artist per day, rather than the raw data contained in the LFM-1b dataset’s [9] listening events file.

3.2 Experimental Setup

To generate a prediction, we use a certain number of past days, which can be specified individually for each experiment. Each value in the feature vector

[1] https://www.sqlite.org

constitutes the number of listening events the specified artist accumulated that day, i.e., over all the artist's tracks. Based on these feature vectors, the goal of the algorithm is to calculate a single value representing the amount of listening events the artist would receive a certain number of days after the last known value. More formally, we use a time series $LE^a_{t_0 \cdots t_N}$ of listening events for artist a, starting at day 0 up to day N, where N is the same number for all artists. We then train different regressors to predict $LE^a_{t_{N+1} \cdots t_{N+M}}$, where M is the time period to forecast, in days.

We investigate variants of linear regression, support vector machines, and neural networks, as provided in the scikit-learn[2] Python package, for our regression task and measure accuracy in terms of the R^2 metric.

Linear Regression As a fairly simple but efficient algorithm, linear regression represented our first approach to create predictions. We were not expecting this method to generate accurate results, instead viewing it as a first step towards further optimization. We did, however, quickly realize that with fairly little optimization, the results achieved with linear regression already appeared to be promisingly accurate, as fairly early tests already achieved an average R^2 value of 84%.

Support Vector Machines We next investigated epsilon-support vector regression [4], which is based on a more sophisticated methodology than linear regression and allows for a more complex range of options concerning the optimization of the algorithm to the specific task at hand. In typical classification problems, support vector machines perform a non-linear transformation on the data, allowing the model to separate the classes more easily. In regression use-cases, such as ours, a line of best fit is calculated instead and the parameter ϵ is introduced as a tolerance range, hence the name epsilon-support vector regression. The algorithm's behavior is strongly dependent on the specified kernel, which is represented by different mathematical functions. For our purposes, we assessed linear, radial basis function (rbf), and polynomial kernels (poly). Overall, using the linear kernel yielded similar results to linear regression, with average R^2 scores of around 83%. When using the kernels rbf and poly, further fine-tuning can be made via the parameters ϵ, C, and γ. Epsilon determines the size of the tolerance range for data that significantly deviates from the calculated model. The tolerance penalty C specifies how harsh data outside this tolerance range should be penalized. Finally, γ determines the intensity of the influence a single data point can have on the overall model. We continuously tweaked these parameters by hand, constantly analyzing the results and comparing them to previously achieved ones. We achieved the best results when using the rbf kernel with $\gamma = 0.00001$, $\epsilon = 1.0$, and $C = 625$, which accomplished an average R^2 value of over 86%.

[2] `http://scikit-learn.org`

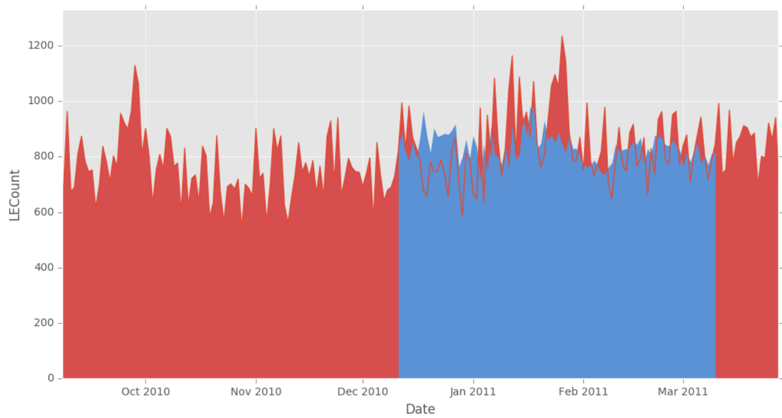

Fig. 1. Prediction over 90 days for Metallica using *linear regression*, December 2010

Neural Networks Artificial neural networks are an advanced machine learning methodology that tries to solve problems based on a layered structure of nodes. We used feed-forward neural networks [5], where each node of a layer is connected to each node in the layer above and below it. As the main purpose of our work with neural networks was getting the most accurate predictions based on a specified number of listening events, the most essential part was the configuration of the network itself. We used a sliding window approach to train our neural network where the window size is 120 days. To determined the best solver, we compared the accuracy of all solvers provided by scikit-learn (sgd, adam, lbfgs) and eventually determined that the lbfgs solver was best suited for the amount of data available in our dataset.The second configuration step was to choose the activation function, for which we used a linear model due to accuracy and consistency of the achieved results. Lastly, we determined the amount of layers and nodes. We chose one hidden layer and increased the number of nodes until further change produced no noticeable differences in results and ended up with 120 input nodes, 16 hidden nodes and one output node. Our best results achieved with neural networks in terms of the R^2 score were around 91%.

3.3 Results and Discussion

As illustrated in Figures 1 and 2 for Metallica, using respectively linear regression and neural networks, the achieved results all appear to be fairly plausible predictions, regardless of the applied algorithm. Red areas represent the true evolution of listening events, blue areas the predictions. Naturally, the longer the predicted time span, the less accurate the achieved results are. Additionally, the amount of available data is a strong limiting factor, meaning that predictions for a well-known artist are usually significantly more accurate than those generated

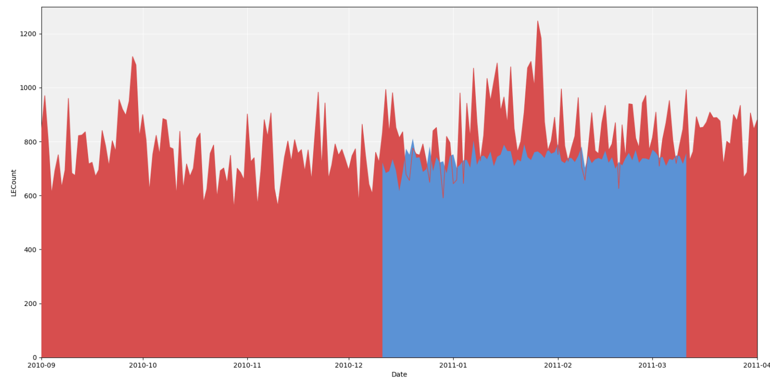

Fig. 2. Prediction over 90 days for Metallica using *neural networks*, December 2010

for an underground band. Achieved results did, however, also strongly depend on the chosen time span. Approaching average R^2 scores of slightly over 93%, some of our best results using linear regression were attained with The Beatles in 2012. Overall, we achieved the highest R^2 scores for artists like The Beatles, Metallica, and Pink Floyd, which we ascribe to the fact that these artists were already well established and fairly popular throughout the time span our data covered. For these artists, all of our applied methods reached average R^2 values of 89% to 94%, with the best scoring predictions lying within a time span of 2012 to 2014. Naturally, we found that artists which exhibit significant jumps or spikes in popularity were much harder to create accurate predictions for. For example, when trying to predict the popularity of Daft Punk in 2013, average R^2 scores of our support vector machine algorithm dropped to around 43%, while linear regression scores sank to 39%.

4 Conclusions

In conclusion, we find that all three regression techniques generate surprisingly accurate results when predicting well established artists, e.g., The Beatles (R^2 of 89% using linear regression) or Metallica (94% using support vector machines). Each technique does, however, possess certain advantages and disadvantages. Linear regression is fairly simple and quick to implement and understand and exceeded our expectations in regards to its accuracy, but is most likely still not the best suited option for real life applications of such problems due to its simplicity. Support vector machines offered slightly higher accuracy and more consistency over artists than linear regression, but performance quickly became a limiting factor when using a larger number of features or predicting a longer time span. Neural networks, on the other hand, probably constitute the best option in our eyes as they allow to use a large number of features (preceding days), which boosted the achieved accuracy, and were also able to generate adequate predictions further into the future.

For future work, we contemplate many ways in which the predictive algorithms could be improved. One of the most obvious and probably also most effective approaches would be to take recent album releases into account when creating predictions. Another idea would be observing social media activity pertaining to specific artists.

Acknowledgments

This research is supported by the Austrian Science Fund (FWF): P25655.

References

1. R. Bandari, S. Asur, and B. A. Huberman. The Pulse of News in Social Media: Forecasting Popularity. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM)*, Dublin, Ireland, June 2012.
2. K. S. D. Herremans, D. Martens. Dance hit song prediction. *Journal of New Music Research*, 43(3):291–302, 2014.
3. R. Dhanaraj and B. Logan. Automatic prediction of hit songs. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, London, UK, September 2005.
4. H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. N. Vapnik. Support Vector Regression Machines. In *Advances in Neural Information Processing Systems 9*, pages 155–161, Cambridge, MA, USA, 1996. MIT Press.
5. G. E. Hinton. Connectionist learning procedures. *Artificial Intelligence*, 40:185–234, 1989.
6. N. Koenigstein and Y. Shavitt. Song Ranking Based on Piracy in Peer-to-Peer Networks. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, October 2009.
7. Y. Ni, R. Santos-Rodríguez, M. McVicar, and T. D. Bie. Hit song science once again a science? In *4th International Workshop on Machine Learning and Music*, Sierra Nevada, Spain, December 2011.
8. F. Pachet and P. Roy. Hit Song Science is Not Yet a Science. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR)*, Philadelphia, PA, USA, September 2008.
9. M. Schedl. The LFM-1b Dataset for Music Retrieval and Recommendation. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR)*, New York, USA, June 2016.
10. M. Schedl, T. Pohle, N. Koenigstein, and P. Knees. What's Hot? Estimating Country-Specific Artist Popularity. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, the Netherlands, August 2010.
11. H. Yu, L. Xie, and S. Sanner. Twitter-driven YouTube Views: Beyond Individual Influencers. In *Proceedings of the ACM International Conference on Multimedia*, Orlando, Florida, USA, November 2014.