Dance Hit Song Science

[a]Dorien Herremans, [b]David Martens, and [a]Kenneth Sörensen

[a]*ANT/OR, University of Antwerp Operations Research Group*
[b]*Applied Data Mining Research Group, University of Antwerp*
Prinsstraat 13, B-2000 Antwerp
{dorien.herremans,david.martens,kenneth.sorensen}@uantwerpen.be
[a]http://antor.ua.ac.be
[b]http://www.applieddatamining.com

Abstract. With annual investments of several billions of dollars worldwide, record companies can benefit tremendously by gaining insight into what actually makes a *hit* song. This question is tackled in this research by focussing on the *dance* hit song problem prediction problem. A database of dance hit songs from 1985 until 2013 is built, including basic musical features, as well as more advanced features that capture a temporal aspect. Different classifiers are used to build and test dance hit prediction models. The resulting model has a good performance when predicting whether a song is a "top 10" dance hit versus a lower listed position.

Keywords: classification, hit song science, hit prediction, music information retrieval

1 Introduction

Record companies invest billions of dollars in new talent each year. They would benefit tremendously from gaining insight into what actually makes a *hit* song. This idea is the main drive behind the new research field referred to as "Hit song science" which Pachet [9] define as "an emerging field of science that aims at predicting the success of songs before they are released on the market". Many music information retrieval (MIR) systems have been developed for a range of different purposes such as automatic classification per genre [12] and composer [5], Yet, as it appears, the use of MIR systems for hit prediction remains relatively unexplored. Dhanaraj and Logan [2] explored the use of support vector machines (SVM) and boosting classifiers to distinguish top 1 hits from other songs in various styles based on acoustic and lyric-based features. Pachet and Roy [10] tried to develop an accurate classification model for low, medium or high popularity based on acoustic and human features but were not succesful. They suggest that the acoustic features they used are not informative enough to be used for aesthetic judgements. Borg and Hokkanen [1] draw similar conclusions as Pachet and Roy [10] when trying to predict the popularity of music videos based on their YouTube view count by training support vector machines but

In MML 2013: International Workshop on Machine Learning and Music, ECML/PKDD, Prague, 2013.

were not successful. The experiment by Ni et al. [8] has more optimistic results when predicting if a song would reach a top 5 position on the UK top 40 singles chart compared to a top 30-40 position.

In this research accurate models are built to predict if a song is a top 10 dance hit or not. For this purpose, a dataset of dance hits including some unique audio features is compiled. Based on this data different efficient models are built and compared. To the authors' knowledge, no previous research has been done on the dance hit prediction problem.

2 Dataset

A dataset of dance *hit listings* was created based on two singles dance hit archives available online: the Official Charts Company and Billboard[1] . The hit listings were parsed with the Open Source Java html parser library JSoup and resulted in a dataset of 21,692 instances with five features: song title, artist, position, peak position and date. In a second step, the hit listings were mapped to their *musical features*. These features were obtained from The Echo Nest[2] with the Open Source java client library jEN. Data was retrieved from 3,452 out of 4,120 unique songs in the hit list data base. The 668 songs with missing data were removed from the dataset. The extracted features can be divided into three categories. The first category is *meta-information*, which is descriptive information about the song, often not related to the audio signal itself. This information was discarded when building the classification models. In this way, the model can work with unknown songs, based purely on audio signals. Secondly, *basic features* from The Echo Nest Analyzer [6] are extracted. These include duration, tempo, time signature, mode, key, loudness, etc. A final category of features was added to incorporate the *temporal aspect* of two features offered by the Analyzer. The first feature is *Timbre*, a 12-dimensional vector which captures different aspects of the tone colour for each segment of a song. The second one is *beatdiff*, which represents the time difference between subsequent beats. Based on Schindler and Rauber [11], the statistical moments along with some extra descriptive statistics were used to capture the temporal aspect of *timbre* and *beatdiff* throughout a song: mean, variance, skewness, kurtosis, standard deviation, 80th percentile, min, max, range and median. The resulting dataset contains 139 usable features.

A Google motion chart[3] as well as more classical 2-dimensional plots were used to visualize the time series data for a number of features. This revealed an evolution of the characteristics of top 10 dance hits over time. A rising trend could be detected for the loudness, tempo and 1st aspect of timbre (brightness) among others.

---

[1] `officialcharts.com` and `billboard.com`

[2] `echonest.com`

[3] Interactive motion chart available at `http://antor.ua.ac.be/dance`

3 Dance Hit Prediction

3.1 Preprocessing

Three datasets were made with each a different gap between the two classes, with data from 2009 until 2013. In the first dataset (D1), hits are considered to be songs with a peak position in the top 10. Non-hits are those that only reached a peak position of at most 30. In the second dataset (D2), non-hits can have a peak position of 20. To form the third dataset (D3), the original dataset is split in two at position 20, without a gap.

3.2 Classification Techniques

Five classification techniques were used to build dance hit classification models. The first two models (C4.5 decision tree and RIPPER ruleset) can be considered as the easiest to understand classification models due to their linguistic nature [7]. The other three models (naieve Bayes, logistic regression, support vector machines) focus on accurate prediction. The last technique, support vector machines (SVM) was tested with several settings for the complexity parameter C with both RBF and polynomial kernel.

3.3 Results

Both the decision tree and the ruleset show that time differences between beats and the third timbre vector are important features for classification. An experiment with 10-fold cross validation (10CV) is done on all three datasets, both with and without feature selection of the input data (IS). Because the datasets are not entirely balanced, the AUC is a more suited measure. A Wilcoxon signed-rank test is conducted to compare the performance of the models with the best performing model. Table 1 shows that the overall best technique is logistic regression, closely followed by naive Bayes and that results obtained on the datasets with input selection are generally better. The ROC (see Figure 1) confirms that logistic regression clearly scores better than a random classification. As expected, the best results can be obtained using the dataset with the biggest gap, namely D1. These results are confirmed by a second experiment with an out-of-time test set based on D1 (IS). For a detailed explanation of the applied techniques and their performance, the authors refer to the full paper of this research [4].

4 Conclusion

A model was built that can successfully predict if a dance song is going to be a top 10 hit versus a lower positioned dance song. Data from two hit charts was gathered and mapped to standard audio features provided by The Echo Nest, as well as more advanced features that capture the temporal aspect. In future research, this model will be incorporated in an online application were users

AUC	D1		D2		D3	
	-	IS	-	IS	-	IS
C4.5	*0.51*	0.58	*0.56*	0.55	0.52	*0.55*
RIPPER	*0.54*	*0.6*	0.58	*0.55*	0.54	0.54
Naive Bayes	**0.64**	**0.67**	**0.64**	**0.65**	**0.61**	0.62
Log. regr.n	**0.65**	**0.68**	**0.66**	**0.63**	**0.64**	**0.65**
SVM (Polyn.)	0.57	**0.6**	**0.65**	0.6	0.57	**0.61**
SVM (RBF)	0.58	*0.6*	**0.63**	0.6	0.57	**0.6**

$p < 0.01$: italic, $p > 0.05$: bold.

Table 1. Results with 10-fold validation (AUC).

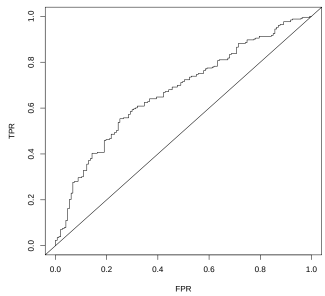

Fig. 1. ROC for Logistic regression

can upload their audio data and get the probability of it being a hit. Other ideas for future research include an expansion to other music styles or improving the accuracy of the existing model by including more features such as lyrics, social network information and others. Another idea for expansion would be to see if the model could be included in an optimisation function of an automatic composition system [3].

References

[1] N. Borg and G. Hokkanen. What makes for a hit pop song? what makes for a pop song? 2011. URL `http://cs229.stanford.edu/proj2011/BorgHokkanen-WhatMakesForAHitPopSong.pdf`.

[2] R. Dhanaraj and B. Logan. Automatic prediction of hit songs. In *Proceedings of the International Conference on Music Information Retrieval*, pages 488–91, 2005.

[3] D Herremans and K Sörensen. Composing fifth species counterpoint music with a variable neighborhood search algorithm. *Expert Systems with Applications*, 40: 6427–6437.

[4] D. Herremans, D. Martens, and K. Sörensen. Dance hit song prediction. (subm).

[5] D. Herremans, K. Sörensen, and D. Martens. Classification and generation of composer specific music. *Working paper - University of Antwerp*, 2013.

[6] T. Jehan and D. DesRoches. *EchoNest Analyzer Documentation*, 2012. URL `developer.echonest.com/docs/v4/_static/AnalyzeDocumentation.pdf`.

[7] D. Martens. Building acceptable classification models for financial engineering applications. *SIGKDD Explorations*, 10(2):30–31, 2008.

[8] Y. Ni, R. Santos-Rodríguez, M. McVicar, and T. De Bie. Hit song science once again a science? *4th International Workshop on Machine Learning and Music, Spain*, 2011.

[9] F. Pachet. *Music Data Mining: Hit song science.* CRCPress, pages 305–326, 2012.

[10] F. Pachet and P. Roy. Hit song science is not yet a science. In *Proc. of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 355–360, 2008.

[11] A. Schindler and A. Rauber. Capturing the temporal domain in echonest features for improved classification effectiveness. *Proc. Adaptive Multimedia Retrieval*, 2012.

[12] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *Speech and Audio Processing, IEEE transactions on*, 10(5):293–302, 2002.