# Can you judge a music album by its cover?

Petar Petrovski[1], Anna Lisa Gentile[1]

Data and Web Science Group, University of Mannheim, Germany
{petar,annalisa}@informatik.uni-mannheim.de

**Abstract.** In this work we explore the potential role of music album cover arts for the task of predicting the overall rating of music albums and we investigate if one can judge a music album by its cover alone. We present the results of our participation to the Linked Data Mining Challenge at the Know@LOD 2016 Workshop, which suggest that the the cover album alone might not be sufficient for the rating prediction task.

**Keywords:** Classification, Image Embeddings, DBpedia

## 1 Introduction

Music is not just about the music. Granted, the listening experience is certainly the most essential element, but one can not ignore the role of cover art in making a brilliant album. In the literature several methods have been proposed to represent music albums not only with information directly extracted from the musical content [2], but also e.g. from entities and relations extracted from unstructured text sources [1], user interaction evidence [6]. With respect to the role of visual clues, Lībeks and Turnbull [5] have investigated how the image of an artist can play a role in how we judge a music album. They proposed a method to predict music genre tags based on image analysis, with the purpose to assess similarity amongst artists to inform the music discovery process. In our work we investigate if there is any correlation between certain patterns in the album cover and the overall rating for the music album as a whole. Our question therefore is "can one judge a music album by its cover, after all?".

We present results of our participation[1] to the Linked Data Mining Challenge[2] at the Know@LOD 2016 Workshop. The task of the challenge consists of predicting the rating of music albums, exploiting any available data from the Linked Open Data cloud. The dataset of the challenge is obtained from Metacritic[3], which collects critics' reviews on musical works. Specifically, the organisers sampled a number of musical albums from the website and labeled

[1] Submission Petar P. Grd at http://ldmc.informatik.uni-mannheim.de/
[2] http://knowalod2016.informatik.uni-mannheim.de/en/linked-data-mining-challenge/
[3] http://www.metacritic.com/

each album as "good" when the critics' score for it is greater than 80 and "bad" when lesser than 60.

To answer our question, we learn a SVM model that classifies music albums either as "good" or "bad" and we train the model only using the cover art of each album. The feature extraction from the cover art is based on image embeddings.

## 2 Classification of music albums

Our proposed approach for album classification consists of three main steps. Given a collection of music albums, we first obtain the image of their cover art. Then, using off-the-shelf tools we obtain a feature vector representation of the images. We then learn a classifier to label each album as "good" or "bad", only exploiting the feature space obtained from its cover art. Figure 1 depicts the proposed pipeline.

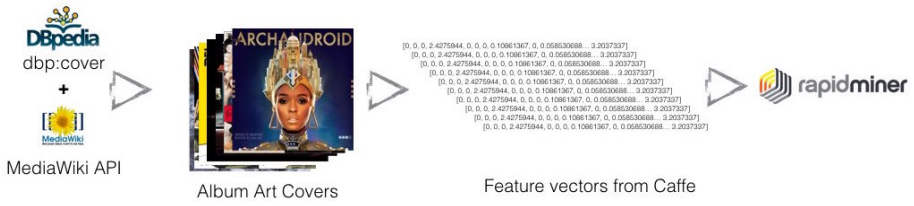


Fig. 1: Pipeline for the classification of music albums using their cover art.

### 2.1 Dataset

In this work we use the Know@LOD 2016 challenge dataset. It consists of 1,600 music albums. Each item provides:

- album name
- artist name
- album release date
- DBpedia[4] URI for the album
- the classification of the album as "good" or "bad".

The organisers split the dataset as 80% (1,280 instances) for training, and 20% (320 instances) for the test.
In our experiment we deliberately only want to exploit the cover art of the albums, therefore we only use only the DBpedia URIs of the albums to obtain their cover images. First, by using Rapidminer LOD extension [7], we retrieve the dbp:cover[5] property for each album. The property contains the path to the

[4] http://dbpedia.org
[5] The prefix dbp: stands for the namespace http://dbpedia.org/property/.

image of the cover art. Then, by using the Mediawiki API[6], we download all the images.

The resulting image set consists of 1558 images, with 2 images missing (the path obtained from dbp:cover did not correspond to any Wikipedia resource).

The dataset is available at https://github.com/petrovskip/know-lod2016, together with the extracted feature vectors and process used (explained in section 3).

### 2.2 Classification approach.

We learn a SVM model that classifies music albums either as "good" or "bad". Starting from our image set, we use image embeddings to obtain a feature space.

*Feature set.* We use the Caffe deep learning framework [3] to obtain image embeddings. Together with deep learning algorithms, Caffe also provides a collection of reference models, which are ready to use for certain labelling tasks. Specifically, we used the *bvlc* model[7] by Krizhevsky et al. [4]. It consists of a neural network of five convolutional layers, and three fully-connected layers. The model has been trained on the ImageNet collection, a dataset of 1.2 million labelled images from the ILSVRC2012 challenge[8]. Each layer of the model provides a representation of the image features, with associated weights. The last layer is the one that outputs the labels for each image. This model outputs as labels 1000 different classes, according to the training ImageNet collection. Since we are not interested in these 1000 labels, but we only want to classify images as "good" or "bad", we classify our image set with this model, but we only use the output layer before the last (the second fully-connected layer), from which we obtain (i) images features and their weighting. The resulting feature vector has a length of 4,096. These features represent visual components of the images, e.g. colours, shapes etc.

*Learning process* We use all obtained features (without any fine-tuning) to train the C-SVM classifier (a wrapper implementation of libSVM classifier on Rapidminer) with a linear kernel and the default parameters.

## 3 Experiment

We evaluated our approach using 10-fold cross validation on the training set and we obtained an accuracy of 58.03%. The resulting confusion matrix is presented in Table 1. The accuracy of the test set reported by the challenge system is of 60.3125%.

The low accuracy of our approach seems to suggest that the image alone is not a good predictor of the overall rating of the musical album. Nevertheless it would

[6] https://www.mediawiki.org/wiki/API:Main_page
[7] bvlc_reference_caffenet from caffe.berkeleyvision.org
[8] http://image-net.org/challenges/LSVRC/2012/

Table 1: Confusion matrix.

| | true good | true bad | ***class precision*** |
|---|---|---|---|
| pred. good | 385 | 282 | 57.72% |
| pred. bad | 254 | 356 | 58.36% |
| ***class recall*** | 60.25% | 55.80% | |

be interesting to investigate if fine-tuning of the features or the combination pf other factors could lead to better results.

## 4 Conclusion

This paper presents our submission to the Linked Data Mining Challenge at the Know@LOD 2016 workshop. We proposed an approach that classifies music albums into "good" or "bad" based solely on their cover art. We trained a SVM classifier with the feature vector calculated from the album's cover art to solve the prediction problem of music album classification. While our approach has some interesting results, our experiment hints that only using album covers as features is not the best fit for the task of the challenge.

## References


1. Bolshakov, I., Gelbukh, A.: Natural Language Processing and Information Systems. Natural Language Processing and Information Systems 1959, 103 – 114 (2001), http://www.springerlink.com/index/10.1007/3-540-45399-7
2. Horsburgh, B., Craw, S., Massie, S.: Learning pseudo-tags to augment sparse tagging in hybrid music recommender systems. Artificial Intelligence 219, 25–39 (2015), http://dx.doi.org/10.1016/j.artint.2014.11.004
3. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the ACM International Conference on Multimedia. pp. 675–678. ACM (2014)
4. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
5. Lībeks, J., Turnbull, D.: You can judge an artist by an album cover: Using images for music annotation. Multimedia, IEEE (99), 1–1 (2011)
6. Pichl, M., Zangerle, E., Specht, G.: Combining spotify and twitter data for generating a recent and public dataset for music recommendation. CEUR Workshop Proceedings 1313, 35–40 (2014)
7. Ristoski, P., Bizer, C., Paulheim, H.: Mining the web of linked data with rapidminer. Web Semantics: Science, Services and Agents on the World Wide Web 35, 142–151 (2015)