SVEUČILIŠTE U ZAGREBU

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Music Genre Classification Based on Song Lyrics

Rafael Josip Penić, Franko Pandžić

Zagreb, February 2021.

Content

About the problem

Music genre classification is a very complex and hard task even for humans. Main obstacle is most commonly a very strong similarity between some genres like, for example, Pop and R&B. Music genre classification models usually work with songs audio but in our project we tried to use songs lyrics to classify the song into the right genre.

It is to be expected that songs from the same genre will more or less have the same themes. To give an example, country songs are usually about home and land. Metal songs, on the other hand, mostly have darker themes like death and life.

This problem has been very well researched by many NLP experts. Currently best results have been achieved by Alexandros Tsaptsinos who used a hierarchical attention network which gave fantastic results.

Dataset

In our project we used Metrolyrics dataset which contains more than 380 000 labeled songs. This dataset consists of songs produced between 1970 and 2016. Each dataset entry carries information about song and artist name, genre and lyrics. Originally, this dataset also contained spanish songs and this would be very problematic for our model but luckily, we found a filtered out version of the dataset.

Besides spanish songs, we also had to filter out songs without lyrics since they are of no significance for us. We also had to “clean” lyrics so they don’t contain parts which mark the end of the chorus and similar. Dataset can be downloaded here.

In the image below you can see the most used words for each genre in the dataset.

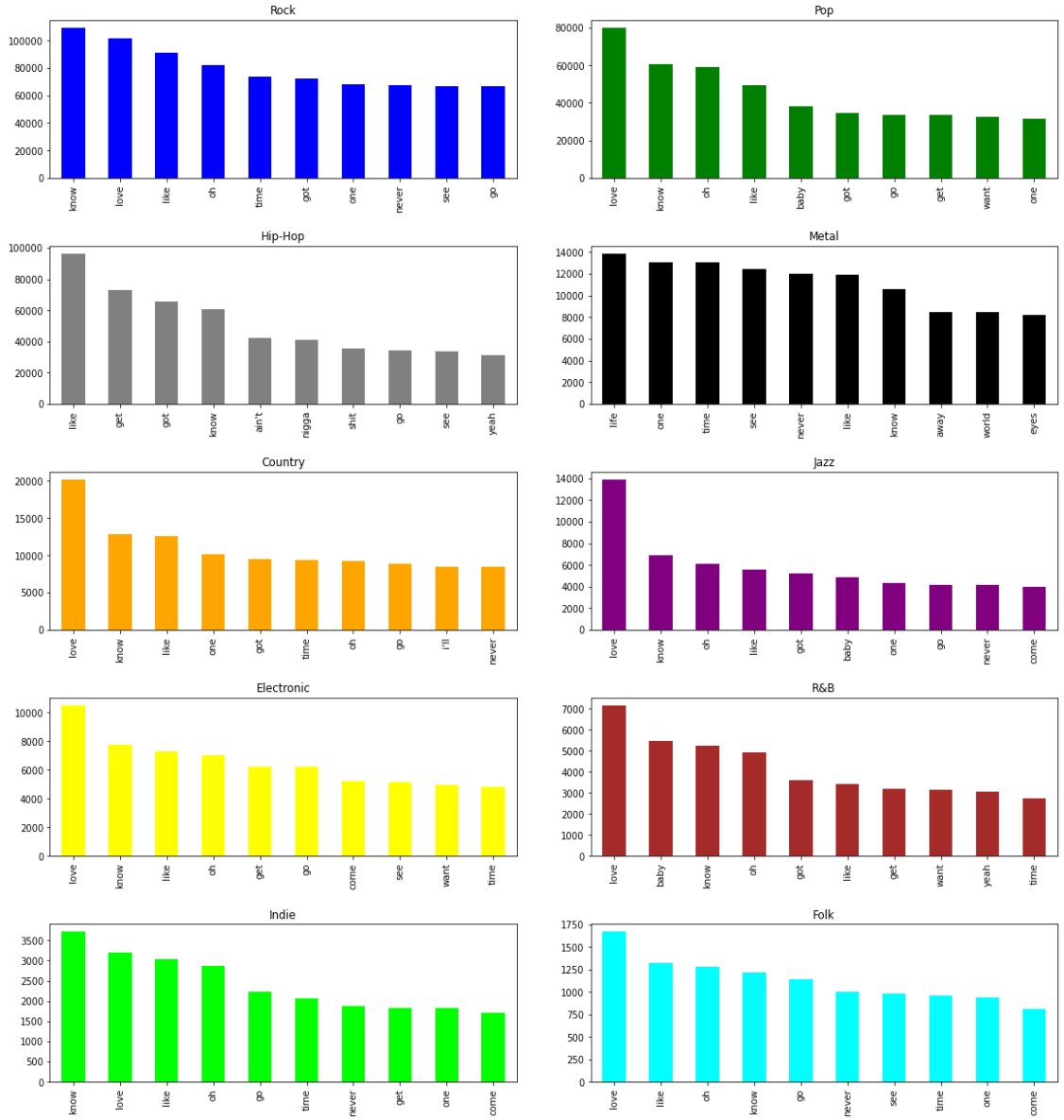

Figure 1: Most common words for each genre (frequencies)

Hierarchical Attention Network

As mentioned before, currently the best solution for the lyrics genre classification problem is Hierarchical Attention Network (HAN) [1]. Because of that we decided to implement such a model.

Main idea of the HAN model is simple: first read the context of each line individually and then combine these contexts to get a “full” context of the song. But how exactly does the HAN do that? It does it by using two layers of bidirectional gated recurrent unit (GRU) [7] followed by the attention. First layer is then used to “read” the relationship between words in the line and the second one is used to get the full context by combining all line contexts. In the end, we use a linear layer to do a simple genre classification.

For more information, check papers [1] and [2].

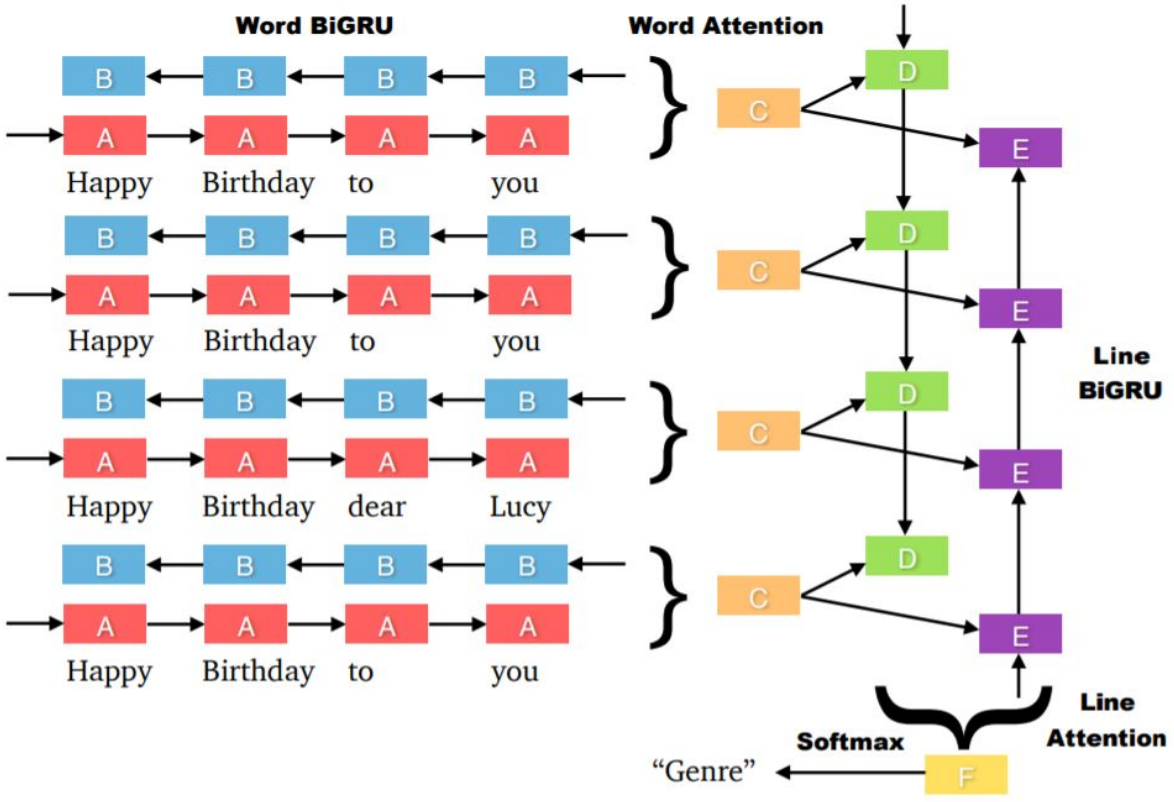

Figure 2: Hierarchical Attention Network structure (image taken from [1]), A, B - hidden state vectors (words context), C- line vector, D, E - hidden states vector (lines context), F - song vector

Training and results

Before training, from the dataset we removed a few genres that rarely appeared. In the end we ended up with the following genres: Pop, Hip-Hop, Rock, Metal, Country and Jazz. Another thing we noticed is that the amount of rock songs in the dataset is pretty high in comparison to other genres.To prevent the model from “preferring” the rock genre too much, we dropped half of the rock songs.

We trained the HAN with ADAM optimizator. Learning rate was 0.0003. Size of the mini-batch was 16. Number of training epochs was 12.

For word representation, we used pre-trained 100-dimensional GloVe representations. They can be downloaded here.

Results are presented on tables and images below.

true, prediction (%)	**Pop**	**Hip-Hop**	**Rock**	**Metal**	**Country**	**Jazz**
Pop	51.862	7.482	27.529	2.566	5.016	5.545
Hip-Hop	9.184	78.259	7.356	3.221	0.914	1.066
Rock	22.417	3.194	51.349	9.54	8.742	4.758
Metal	5.144	1.662	21.419	70.43	0.739	0.606
Country	12.092	1.062	25.286	1.103	48.529	11.928
Jazz	26.68	0.395	16.996	0.988	12.846	42.095

Table 1: Confusion matrix; row = true label, column = predicted label, cell (row = i, column = j) = percentage of labels ‘i’ predicted as label ‘j’

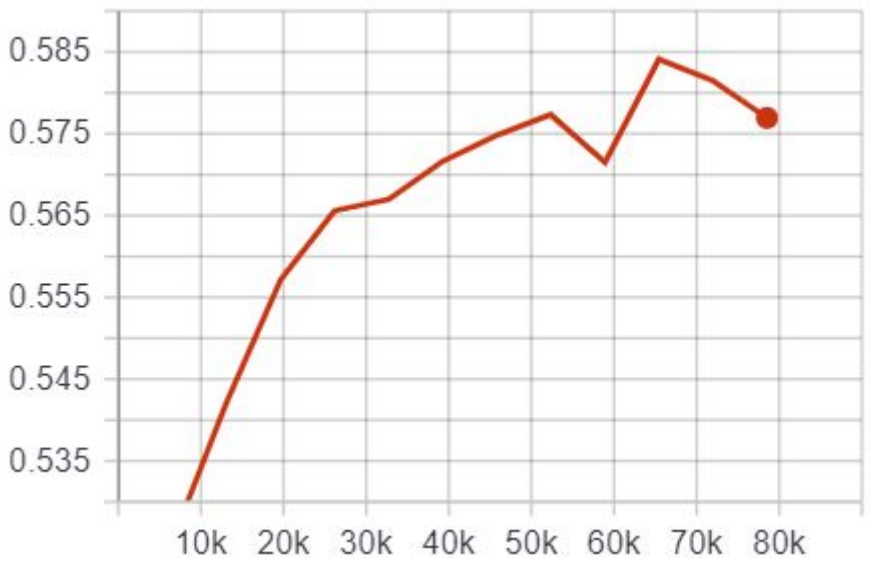

Figure 3: Accuracy on validation set (12 epochs)

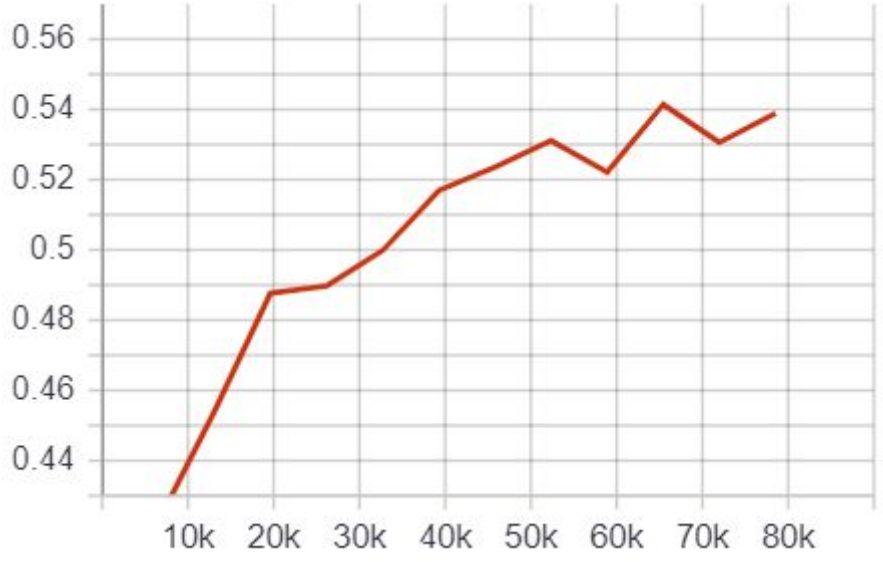

Figure 4: F1 metric (macro) on validation set (12 epochs)

Accuracy on the test set (from which the confusion matrix was constructed) is 57.62% and F1-macro is 53%.

Table 1 shows confusion matrix. Each cell of the matrix contains a percentage of the examples with the true label represented by the row that are predicted as the label represented by the column. For example, 7.482% Pop songs have been wrongly labeled as a Hip-Hop song. From this matrix, we can also see which genres “overlap”. We see that there is a decent amount of Pop lyrics predicted as Rock and vice versa. Additionally, it is clear that Hip-Hop is a very distinct genre that doesn’t overlap with other genres.

Figure 3 and 4 graphically show how accuracy and F1-macro “behaved” through the training (metrics were recorded at the end of each epoch).

Conclusion

All in all, the HAN model gave expected results. It is obvious that it is quite possible to classify songs into genres by only their lyrics. Only problem is that certain genres often write about similar themes and that confuses the model during the training.

Another thing we have to ask ourselves is if there is any way to improve this model or are there maybe some other models that are better but still unused on the genre classification problem.

First of all, we could try extending the HAN model. Currently we only have two layers in the model. We could try adding another one so we also “read” the context of lines in a certain song segment and in the end we combine contexts of song segments and not lines.

Secondly, we could always try using BERT [4] which is currently very popular. BERT is a self-supervised model so we would have to modify our data pre-processing a bit. For example, BERT uses NSP (Next Sentence Prediction) task where the model has to determine whether the sentences really come one after another or not. Once the model is trained on pre-train tasks, it is fine tuned on the real problem (in our case: genre classification).

Literature

1. Lyrics-Based Music Genre Classification using a Hierarchical Attention Network; Alexandros Tsaptsinos; 2017.; link
2. Hierarchical Attention Networks for Document Classification; Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, Eduard Hovy; 2016.; link
3. Hierarchical attention networks for information extraction from cancer pathology reports; Shang Gao, Michael T Young, John X Qiu, Hong-Jun Yoon, James B Christian, Paul A Fearn, Georgia D Tourassi, Arvind Ramnthan; 2018.; link
4. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding; Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova; 2019; link
5. https://medium.com/analytics-vidhya/hierarchical-attention-networks-d220318cf87e
6. https://humboldt-wi.github.io/blog/research/information_systems_1819/group5_han/
7. https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21