# Assignment 3 - lyrics generation using RNNs

## .Intorduction

* + - 1. Dataset analysis

Sinai Tbd add analysis of the dataset + lyrics

* + - 1. Melody analysis

Sinai Tbd add analysis of the melody (size,len etc)

## Methodology

### 2.1 Lyrics preprocessing

### The preprocessing stage for the lyrics involves several key steps designed to standardize and streamline the input data for our model:

* **Lowercasing**: All lyrics are converted to lowercase to ensure uniformity and prevent case sensitivity from affecting the learning process.
* **Punctuation Removal**: All punctuation marks are removed from the lyrics. This simplifies the text data and helps in reducing the complexity of the model’s input space.
* **End of Sentence Marker**: An 'eos' (end of sentence) marker is added to signify the end of sentences, aiding the model in recognizing sentence boundaries and improving its predictive capabilities for lyric structure.
* **Preservation of '&'**: The ampersand ('&') is retained as it represents a line break in the lyrics. This inclusion is crucial as it allows the model to predict line breaks, thus enhancing its ability to generate lyrics that closely mimic the structure of real songs.

*The implementation details are outlined in the generate preprocessing\_lyrics*

*function within tools.py.*

### 2.2 Text embedding

We employed the word2vec module to convert words into 300-dimensional numeric vectors. This training occurred on the train dataset, allowing us to capture contextual representations of each term.

*The implementation details are outlined in main.py.*

### 2.3 Dataset procesing for LSTM network

For processing lyrics, we configure the input for the LSTM network as a sequence vector of size nnn. Each position in this vector represents a distinct timestep, analogous to a "time-line". In our model's context, each vector is a sentence where every position corresponds to a different word. The output (target) vector mirrors the input's size, with each position containing the subsequent word from the corresponding input vector position. The number of training vectors varies depending on the sequence length.

*The implementation details are outlined in the generate\_sequence function within tools.py.*

### 2.4 Melody procesing ( MIDI files)

We propose 2 approaches to handle the midi files;

#### 2.4.1 graph embedding

Utilizing the graph embedding technique described in Lisena's paper [1], we implement the MIDI2vec approach, which uses graph-based methods to represent MIDI files. This technique effectively captures musical elements such as tempo, time signature, programs, and notes. Specifically, it adapts the node2vec algorithm to generate embeddings that can predict musical genres and other metadata. This method is proven to deliver high accuracy when used with Feed-Forward Neural Networks, facilitating scalable and automated metadata tagging within symbolic music collections.

We applied an existing implementation from a GitHub repository [2] with specific parameters tailored for our needs:

* **For Computing Edgelist:**
  + -n = 100: This specifies that the algorithm considers 100 groups of simultaneous notes for each MIDI file.
* **For Converting to Vectors:**
  + walk\_length = 5: Determines the length of each walk per source.
  + –dimensions = 50: Sets the number of dimensions for the output vectors.

The remaining parameters were left as default, as specified in the referenced documentation.

Ultimately, this approach provided us with a 50-dimensional vector for each melody, capturing the essential characteristics needed for further analysis and processing.

*The implementation details are outlined in the midi2vector.py.*

#### 2.4.1 Sinai embedding

tbd

## 3. Model Architecture

### 3.1 train validation split:

We allocate 90% of our data to training and the remaining 10% to validation. Given the complex task of predicting and generating lyrics, we anticipate that the loss will primarily decrease during training. For validation, observing a significant reduction in loss is less likely due to the intricacies of the task. However, to gain a better understanding of the model’s performance on unseen data, we employ the perplexity metric during validation. This measure helps us evaluate the model's ability to predict the next word in the sequence effectively. We believe that a 10% validation split is sufficient for this specific application, as it provides a robust sample to assess the perplexity and overall performance of the model without requiring a larger subset, which may not yield proportionally more insights into model behavior.

*The implementation details are outlined in the main.py.*

### 3.2 Model Architecture

3.2.1 Naïve Concatenate Word and MIDI Vectors (Naïve)

From this point forward, the will be referred to as the **'naive'** model

We employ an LSTM model structured as follows:

1. **Concatenation**: The word vector and MIDI vector are concatenated to form a single input vector.
2. **LSTM Layer**: The concatenated vector is processed through an LSTM layer
3. **Fully Connected Layer**: Outputs from the LSTM are fed into a fully connected layer that matches the size of the vocabulary, producing the final output predictions.

*The implementation details are outlined in class* *LyricsGenerator within models.py.*

3.2.2 Merge Parallel Layers of MIDI and Word Vectors (Merge)

From this point forward, the will be referred to as the **‘merge’** model

1. **Dense Layer for MIDI**: The MIDI vector is first passed through a single dense layer with a ReLU activation function to create a densified representation.
2. **Concatenation**: The densified MIDI vector is then concatenated with the word vector to form the combined input.
3. **LSTM Layer**: as written before.
4. **Fully Connected Layer**: as written before.

*The implementation details are outlined in class* *MergeLyricsGenerator within models.py.*

To enhance model performance, we engage in hyper-parameter optimization, the details of which will be described in the final parameters section. During training, we monitor and record the training and validation loss, as well as the perplexity on both the train and test datasets. Additionally, to ensure meaningful tracking of the model's progress and capabilities, we generate and print lyrics at each epoch. This approach not only provides insights into the immediate performance of the model but also allows for adjustments based on real-time feedback from the generated outputs.

*The implementation details are outlined in the main.py.*

### 3.3 generated lyrics rules

To generate text with our model, we follow a pipeline that integrates both deterministic and stochastic elements:

1. Define the parameter: 1) max length – maximal lyrics length
2. Send the first word along with the chosen integrated MIDI embedding to the model
3. Instead of selecting the word with the highest probability, we make a probabilistic choice based on the entire distribution of output probabilities from the model. This approach allows for a more diverse and naturalistic text generation.
4. Append the selected word to the generated sentence
5. if the sequence exceeds 5 words, retain only the last 5 words and repeat steps 2-4 for continuation

The text generation process is halted under one of the following conditions:

* The sequence reaches the predefined maximum length.
* The model predicts the 'eos' (end of sentence) marker, indicating a natural conclusion of the lyrical passage.

*The implementation details are outlined in function generate\_text Within tools.py.*

## 5 Model Optimization

We evaluated two model configurations, 'Naïve' and 'Merge', combined with two different MIDI embedding methods: graph-based embedding and modified MIDI embedding.

**Hyperparameter Tuning**

The models were run with varying hyperparameters:

* **Learning Rate**: [0.0001, 0.001]
* **Sequence Length**: [3, 5]
* **Batch Size**: [16, 64]
* **Dropout**: [0.1,0.3]

Fixed parameters across all runs included:

* **Embedding Dimensions**: 300
* **Hidden Dimensions**: 40
* **LSTM Layers**: 2
* **Maximal Epochs**: 15
* **Maximum Sequence Generated**: 100
* **Weight Decay**: 1e-5

Data was split into 90% for training and 10% for testing in each iteration. In total, 32 different runs were conducted (8 parameter combinations across 4 model and embedding combinations). A more extensive grid search would be implemented if additional time and computational resources were available, potentially including variables such as dropout rates, the number of epochs, additional LSTM layers, and more granular entity handling for both words and MIDI files.

**Evaluation and Results**

Perplexity was used as the metric to determine the best configuration on the validation dataset. The two top-performing configurations, one from each model type, were:

|  |  |  |
| --- | --- | --- |
|  | Merge-Graph | Naive-Modified |
| Epochs | 15 | 7 |
| Learning Rate | 0.0001 | 0.001 |
| Sequence lenght | 5 | 5 |
| Batch size | 16 | 16 |
| Dropout | 0.1 | 0.1 |

The results showed that the Merge model with graph embedding outperformed the Naive model, indicating that introducing a learnable layer to the MIDI embedding was beneficial. Additionally, the graph embedding method achieved a lower perplexity score (154) compared to the modified embedding (174).

The full experiment results is found in reference [4]

Use tensorboard

## Generate Text Analysis

**Merged model + graph embedding**

**billy joel, honesty**

if for you & again when [chorus: & & don't ain't but comin' & & & i'll true girls love looking & & in your heart & and you have i could just & i'll just to go & & i have & to real value & & baby checkin' for learn & & love what someday & & by you can said to cause & & gone would & i sorry sandman music & it's just & to right to go you're here just will just to go & & i've go & by is me love never well

**cardigans lovefool**

dear we come & there i could see & & i how will live watch your go and i'm better christmas & & tv-funk in your city & to heal amadeus there so show so remember & & will just takes blue daddy! loving i it's much them one doin' & & that even & i icy up could spell devil in side & safe square beside anything & love maybe & to start & & all i those peacefully & & of listen & dawn she soulmates & love stand & like you could i &

**aqua barbie girl**

hi you love & & around & like bring by my without i taught to & & won't mistake he nightmare & & he used & & & & & whenever me & in happy he's and it taken & & & all & to say & & could left the friend & day & & the world love & all takes they're it & & & from devil's & love & it can you're i straight i & but the fiend wrap & i need & in the forgive & & love it's my needed i didn't to say

**blink 182 all the small things**

all & i had to say & & mirage & love at the texarkana will boy hey here a easy try & & ooh mission and mortal it really & & & & now have & you just daddy & love relax love & when man; & & to is no dining music & & york impressed & & what knows the nigga & of us she white a konk in headed & & more from you you're & can would to they sad & from am & & & day now & & make & now & i'm doubt

**Naive model + modified embedding**

**billy joel, honesty**

if & i'm the waitin' & & & love to be me. & & i've cryin' i attitude crack & i have & in along- & who ought & & don't slow in this shit do & & love i come do and says way & & & & love you we'll be this to songs your way again? you to are & the shot & & love & in my own & said their fantasic & of me & & you would it can't la & & & that i'm only place at me

**cardigans lovefool**

dear & & i can in. & & we'll hand from the night love love & i have & & & & love i'll be hot & i thing & for me & & you can right & & like your heard? & love opened & i have & & & come & and you could i name now & & my vain only when let gimme & do nowhere & & i much just and i will you'll are wanna & & only way i have like & & around & &

**aqua, barbie girl**

hi a leave i & & don't knows it yeah & & love & like you to time i say believe & & it i come & & i'm love & i'm thing & loved & i just can't go & & ever & i can blame? & from years flex & & & the world love soul remember & & in your voice his time & & the is left it for a country be-e-ell & & who will you're & there's & to know i all & you would & & & i have all her &

**blink 182, all the small things**

all and & i de & & & steal day i've paper the meet in is that says workin' by night? drought & & to alone you face only born girl will it's really na and 5 to say & i used & & could not mingled could guy oh nothin' & & moment love & i won't i always i have & you & hours mine prince always soon hey from all & rose my without & love i have & & love blues & at you little all & & love alright & when you or la &

6.2.1 Text output for each model: Same word

We used the best model : merge model with graph embedding, we used the words sunnu, life and red.

**Sunny**

* [billy joel, honesty] sunny desperado & & & & i've lick & and something love love & & live and would you & for a closer & sweet broken door i'm on and sparks they give & to put on free be arms and good to kiss & by the years we'll he's da & & up hurts you song what you're can't it come & well & & of my cold began & babies & & papers i eyes have & & i go & & killing out some all or i'm &
* [cardigans lovefool] sunny & this only only i've is well that look to be on & the save & at we in regular & & i yea love life & to time that listen again & & & & the miss have & & can she tonight and come are my that's little oh oh oh oh oh oh oh oh shit & to say & & & in my heart man & past i will said oh i you're do & i have & to can't think & he want & baby & so used
* [aqua, barbie girl] sunny about & & & oh oh oh oh oh fine yeah & & slap & & & & i need to in you & & walk to draw and i much you love & & & & walked in your just for the fire & & have & for man & to keep you happy they taught & & for it true no & like view & & & & & & & i didn't south & her meet & & & way over it's & alright & & oh god that start & have & & &
* [blink 182, all the small things] sunny 'neath & i'm hatte on all hunger sleigh in the guess & & yeah & & by it here & i to stray in near to right and i play life to girl will things love & that through what though & & we come mmm and make guard my strong throw & & white first hours & & planned i'm now & love would don't can & i could you faith & & & don't like to say & & home & & in [chorus: white we lark another intimacy boy hang & grow with the ride &

**LIFE**

* [billy joel, honesty] life & me soon & & yeah & & i it's to rain and honey & & i'm on no should & ooh over from your mind there's & & your causes can & that what fantasy won't & it's to want what & here try & you takes & to apart this hurt better & & you see & believe don't la be & & how's back while your everybody & & we & & drink kind & for around never like & & yeah & it love & & write for the lonely with you &
* [cardigans lovefool] life & she & come for me & yeah & love after heart when blood earth and it's to was a money and photograph & & ooh i & not i & so all & i & & that low call & & from a frightened eee & & i moves my away inside i sees the looking & do & & to won't i'm eyes; & & won't he show a albums & & that dirty love & & will tell i love from the all that girl & & and time it & in a thugs &
* [aqua, barbie girl] life & & of beat & & i love & & it's have home & that's & that think & & again & & & & wind & & to sounding & to was sometimes a yo & i would they they got you to be lady & i will true & don't & i & & have to are the get love to there lyin' & & i can heaven & with me to how & & i better & & after my sure get & feet & love i've got to give & & &
* [blink 182, all the small things] life that together & & & the sure fine & & & & i living loves you like another goodbye & my heat love ah-ha-ha-ha and giddy of my said forever don't holding me & again see in said you & of me & & you to house control on & & on to tell you're dirty the lot & you may i have & from keeps you i've living love & & & i faith & and if & here if again & & i've fell nothing come & & i'm listen like to know & & mounted why

**RED**

* [billy joel, honesty] red a goodbye's & & to can i want & to is you or say & & find & it planned you want to as her lot & of your apartment & & will wanted find & & he short and photograph & that loves me & my looking all & & & now we go at & & not by & & & that takes to be here i've floor it's don't to malts & & & i will doubt something till life & to eyes solve you you're we & come &
* [cardigans lovefool] red my crazy & & love fall & & then i'll beautiful in my tucked walking into shiny of daylight & & have to behold my regret with a rooms with & back to ya & & i or laughter the made home in everything going i have & & is the way come might & in a tennessee i've block & in land here again better as roadhouse the givin stick & & to be & light all in the closer & & & we're better times i'm there it love & have for me & the nigga billboards
* [aqua, barbie girl] red less & & and you his get light & & it takes & & i'll be my talkin' and i would & & i is for me & that arkansas & & & & you his was & & why & & so disposable that's stop & at before that & to god me & & i or was your stand & & & & that they're it's to want to go & around free & & three read & & whey should & & to be & my swimming is jacket & & go & & to true
* [blink 182, all the small things] red you with you & but their porsche told & like not i & you it's to fever baby at that locing ya & & come make & and they're come oh oh oh oh since things can & but it's to right & i it's let diggity all and so hey read & & love upside in spend & & & & i love & to know and i do & sentimental & love way feeling keeps for the i'd fear after i'd & too sky baby evening & & i'm twist yeah & i've ain't &

Both models, the Naïve and Merge configurations, occasionally struggle to generate coherent sentence structures, though they do manage to form meaningful phrases at times. A noticeable trend is the frequent prediction of the word "love," which is common across generated lyrics. The handling of the "&" symbol, which signifies the end of a line in songs, shows mixed results. The models sometimes predict this symbol consecutively or produce overly short lines, yet the overall management of sentence termination is reasonably effective. In certain examples, the models successfully adopt the stylistic elements of the original artists. For example, in Aqua's "Barbie Girl,", when the initial word is “sunny” the model captures a playful and light-hearted tone. Moreover in melodies of billy joel, honesty, the direct and serious style of Billy Joel's "Honesty," aligning with the generated text. Overall, while the models demonstrate an ability to consider both lyrical and melodic contexts, the generated lyrics often lack coherence and clarity. There is potential for generating diverse and stylistically appropriate content, but improvements are needed in maintaining thematic continuity and enhancing stylistic consistency. To further enhance model performance, we recommend expanding the training corpus with a larger variety of text and melodies, and employing a more extensive hyperparameter grid search. This approach should help refine the models' ability to produce more structured and relevant lyrics.

## Reference

* + - 1. Lisena P., Meroño-Peñuela A. & Troncy R. MIDI2vec: Learning MIDI Embeddings for Reliable Prediction of Symbolic Music Metadata. In Semantic Web Journal, Special issue on Deep Learning for Knowledge Graphs, vol.13, no.3, pp. 357-377, IOSPress, 6 April 2022 <http://doi.org/10.3233/SW-210446>
      2. https://github.com/midi-ld/midi2vec?tab=readme-ov-file

* + - 1. midi2vec deafult parameter:
* -i, --input Input graph (edgelists) path. Default: .\edgelist;
* -o, --output Output file name. Default: embeddings.bin;
* --walk\_length Length of walk per source. Default: 10;
* --num\_walks Number of walks per source. Default: 40;
* -p Return hyper-parameter (as in node2vec). Default: 1;
* -q Inout hyper-parameter (as in node2vec). Default: 1;
* --dimensions Number of dimensions. Default: 100;
* --window-size Context size for optimization. Default: 5;
* --iter Number of epochs in word2vec. Default: 5;
* --workers Number of parallel workers. Default: 0 (full use);
* --exclude Edgelists to be excluded from the computation
  + - 1. Hyper parameter tuning expieriment

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| best\_epoch | hidden\_dim | lstm\_layer | batch\_size | sequence\_length | learning\_rate | dropout | model\_name | midi\_method | min\_perplexity\_val | min\_perplexity\_train |
| 15 | 40 | 2 | 16 | 5 | 0.0001 | 0.1 | merge | graph | 154.3423 | 91.18412 |
| 9 | 40 | 2 | 16 | 5 | 0.001 | 0.3 | merge | graph | 154.5377 | 70.89475 |
| 8 | 40 | 2 | 16 | 3 | 0.001 | 0.1 | merge | graph | 156.1847 | 69.7561 |
| 3 | 40 | 2 | 16 | 5 | 0.001 | 0.1 | merge | graph | 156.628 | 57.1485 |
| 8 | 40 | 2 | 16 | 3 | 0.001 | 0.3 | merge | graph | 156.6299 | 82.89747 |
| 8 | 40 | 2 | 64 | 5 | 0.001 | 0.1 | merge | graph | 157.8269 | 58.29928 |
| 15 | 40 | 2 | 16 | 5 | 0.0001 | 0.3 | merge | graph | 160.0096 | 106.2443 |
| 8 | 40 | 2 | 64 | 5 | 0.001 | 0.3 | merge | graph | 160.4841 | 71.7783 |
| 9 | 40 | 2 | 64 | 3 | 0.001 | 0.1 | merge | graph | 160.7028 | 70.379 |
| 15 | 40 | 2 | 16 | 5 | 0.0001 | 0.1 | naive | graph | 160.9166 | 89.40215 |
| 14 | 40 | 2 | 64 | 3 | 0.001 | 0.3 | merge | graph | 162.4141 | 84.19176 |
| 15 | 40 | 2 | 16 | 3 | 0.0001 | 0.1 | merge | graph | 162.5803 | 108.7822 |
| 15 | 40 | 2 | 16 | 5 | 0.0001 | 0.3 | naive | graph | 163.3033 | 103.1558 |
| 7 | 40 | 2 | 16 | 5 | 0.001 | 0.3 | naive | graph | 163.6725 | 67.97935 |
| 5 | 40 | 2 | 16 | 3 | 0.001 | 0.3 | naive | graph | 164.3775 | 81.63463 |
| 15 | 40 | 2 | 16 | 3 | 0.0001 | 0.1 | naive | graph | 164.508 | 106.0363 |
| 5 | 40 | 2 | 64 | 5 | 0.001 | 0.1 | naive | graph | 164.5433 | 52.75672 |
| 3 | 40 | 2 | 16 | 5 | 0.001 | 0.1 | naive | graph | 164.9792 | 52.96175 |
| 8 | 40 | 2 | 64 | 5 | 0.001 | 0.3 | naive | graph | 165.9223 | 69.20401 |
| 3 | 40 | 2 | 16 | 3 | 0.001 | 0.1 | naive | graph | 167.1465 | 63.41625 |
| 15 | 40 | 2 | 16 | 3 | 0.0001 | 0.3 | merge | graph | 167.1764 | 122.6724 |
| 5 | 40 | 2 | 64 | 3 | 0.001 | 0.1 | naive | graph | 167.7097 | 64.50551 |
| 8 | 40 | 2 | 64 | 3 | 0.001 | 0.3 | naive | graph | 169.3871 | 81.21472 |
| 15 | 40 | 2 | 16 | 3 | 0.0001 | 0.3 | naive | graph | 169.9199 | 121.1686 |
| 15 | 40 | 2 | 16 | 5 | 0.0001 | 0.1 | merge | modified | 174.9126 | 93.58018 |
| 7 | 40 | 2 | 16 | 5 | 0.001 | 0.1 | naive | modified | 176.0731 | 57.67967 |
| 10 | 40 | 2 | 16 | 5 | 0.001 | 0.1 | merge | modified | 176.226 | 57.97062 |
| 15 | 40 | 2 | 16 | 5 | 0.0001 | 0.1 | naive | modified | 176.2439 | 94.25333 |
| 9 | 40 | 2 | 16 | 3 | 0.001 | 0.1 | merge | modified | 176.4986 | 69.55376 |
| 8 | 40 | 2 | 16 | 3 | 0.001 | 0.1 | naive | modified | 177.8294 | 70.12736 |
| 7 | 40 | 2 | 64 | 5 | 0.001 | 0.1 | merge | modified | 178.3705 | 60.53498 |
| 6 | 40 | 2 | 64 | 5 | 0.001 | 0.1 | naive | modified | 179.9939 | 58.3959 |
| 15 | 40 | 2 | 64 | 5 | 0.0001 | 0.1 | naive | graph | 180.927 | 125.8033 |
| 12 | 40 | 2 | 64 | 3 | 0.001 | 0.1 | merge | modified | 181.5559 | 71.66285 |
| 15 | 40 | 2 | 16 | 3 | 0.0001 | 0.1 | merge | modified | 183.4868 | 109.5391 |
| 15 | 40 | 2 | 16 | 3 | 0.0001 | 0.1 | naive | modified | 184.4916 | 110.6395 |
| 12 | 40 | 2 | 64 | 3 | 0.001 | 0.1 | naive | modified | 184.8891 | 70.86592 |
| 15 | 40 | 2 | 64 | 5 | 0.0001 | 0.1 | merge | graph | 185.6695 | 136.1485 |
| 15 | 40 | 2 | 64 | 5 | 0.0001 | 0.3 | merge | graph | 188.048 | 142.6138 |
| 15 | 40 | 2 | 64 | 5 | 0.0001 | 0.3 | naive | graph | 189.6397 | 142.253 |
| 15 | 40 | 2 | 64 | 3 | 0.0001 | 0.1 | merge | graph | 191.4319 | 147.4475 |
| 15 | 40 | 2 | 64 | 3 | 0.0001 | 0.1 | naive | graph | 192.0801 | 143.7238 |
| 15 | 40 | 2 | 64 | 3 | 0.0001 | 0.3 | merge | graph | 203.4512 | 161.6432 |
| 15 | 40 | 2 | 64 | 3 | 0.0001 | 0.3 | naive | graph | 205.3486 | 160.5999 |
| 15 | 40 | 2 | 64 | 5 | 0.0001 | 0.1 | merge | modified | 209.6364 | 135.5357 |
| 15 | 40 | 2 | 64 | 5 | 0.0001 | 0.1 | naive | modified | 213.3844 | 138.8922 |
| 15 | 40 | 2 | 64 | 3 | 0.0001 | 0.1 | merge | modified | 218.1742 | 147.5623 |
| 15 | 40 | 2 | 64 | 3 | 0.0001 | 0.1 | naive | modified | 224.721 | 153.4893 |