Lyrics Inducer using Bidirectional Long Short-Term Memory Networks

Jayashree Domala[1], Manmohan Dogra[1], and Anuradha Srinivasaraghavan[1]

Department of Computer Engineering, St. Francis Institute of Technology, Borivali-west, Mumbai

Abstract. Songs are melodic manifestations that are performed by individuals. These tunes are made altogether by both a lyricist who composes the verses and the artist who sings it. Verse writing in itself is an exceptionally selective and characterized issue. With the ever-expanding utilization of innovation and the way that they are effectively accessible to us, makes human lives comfortable. A lyricist can often have a mind block while considering verses or may even find it difficult to get an idea. The principal reason for this exploration is to enable the lyricist to get a motivation that can assist him in making better verses. To accomplish this, a profound learning method is utilized alongside the idea of natural language processing. Specifically, bidirectional LSTM (Long short term memory) networks are used for lyric generation. The proposed framework can exceptionally create versus relying upon the information seed and the scope of words.

Keywords: Long-short term memory network, text generator, natural language processing, deep learning

1 Introduction

The verses that are composed by a musician assume a significant function in the arrangement and formation of the song. It gives life to the music and makes the song whole and complete. But the work behind forming the lyrics is often tiresome and time-consuming. Additionally, if there is a writer's block or no inspiration, then the lyric writing process is delayed further. Machine learning has already done wonders in backing up humans by automating tasks completely or partially. In this case, also machine learning can be used as a tool to help the generation of lyrics.

The proposed system aims at generating lyrics according to the seed input which is given by the user. In this way, the lyrics will be generated following the idea that the writer wants. Backed by deep learning and Natural Language Processing (NLP) concepts, this system is trained on a unique dataset of 'romantic' English songs. Long Short Term Memory networks (LSTM) [1], an uncommon sort of recurrent neural network (RNN) are utilized for the learning process. These LSTM networks are equipped for recalling data for a more drawn out period because of the presence of cell states.The intuition is that the generation of music relies upon the arrangement of input words we feed and subsequently the next word generated must relate to it, therefore the utilization of the LSTM model is favored over the ordinary RNN model. Before the data can be given to the LSTM networks for learning, it is pre-processed by using NLP techniques. The insight is that since the lyrics consist of the words and strings, we may need to preprocess them using NLP to increase the performance. Specifically the bidirectional LSTM networks are used since the efficiency is better than the other models it's compared with, namely LSTM and gated recurrent unit (GRU) networks.

2 Related Work

Numerous researchers have worked using a multitude and varied algorithms on the lyrics generation techniques. The authors S. Pudaruth, et al., proposed a framework for automated generation of song lyrics using CFG's. Prior to the execution, an inside and out investigation was done for understanding the necessities of good verses. They completed their assessment by looking over the produced and existing verses present on the web. The outcomes were agreeable and produced verses were appraised as being existing verses [2]. Another research named Tra-la-Lyrics 2.0 was done by the author Gonçalo Oliveira, H. The exploration referenced a framework for the automatic generation of song lyrics on a semantic area. It is the improvement of the original Tra-la-Lyrics wherein the content is created which is characterized by at least one seed word. To gauge the advancement, the mood, the rhymes, and the semantic intelligence in verses delivered by the first Tra-la-Lyrics were investigated and contrasted with the verses created by the new launch of this framework, named Tra-la-Lyrics 2.0. The assessment demonstrated that, in the sections by the new system, words have higher semantic relationship among them and with the given seeds, while the rhythm is so far composed and rhymes are accessible. A study was led to affirm the aftereffects of the improvement of the verses by Tra-la-Lyrics 2.0 [3]. Furthermore, the author's Eric Malmi, et al., proposed DopeLearning. It explains the rap lyrics generation using a computational approach. They build up an expectation model to recognize the following line of the current verses from a lot of applicant next lines. The model is based on two artificial intelligence procedures. They are the RankSVM algorithm and a deep neural network model with a novel structure. Results show that the forecast model can recognize the valid next line among 299 erratically picked lines with a precision of 17%, i.e., more than 50 times more likely than by discretionary. They additionally utilize the forecast model to join lines from existing tunes, creating verses with rhyme and significance. An assessment of the delivered verses shows that as far as quantitative rhyme thickness, the technique beats the best human rappers by 21% [4]. In another paper, Kento Watanabe, et al., present novel generation models that catch the point changes between units exceptional to the verses, for example, stanza/theme and line. The research centers around the basic relations in Japanese verses. These changes are displayed by a Hidden Markov Model

(HMM) for representing themes and subject advances. As per the outcomes the language model is definitely more powerful than HMM-based models. However the HMM-based methodology effectively catches the between stanza/ensemble and between line relations. For confirmation, the models are assessed utilizing a log likelihood of lyrics generation and fill-in-the-spaces type test[5]. The authors Ramakrishnan A, et al., introduced a model for the programmed generation of Tamil lyrics. A corpus consisting of 10 melodies was utilized to prepare the framework to comprehend the syllable examples. Utilizing the prepared model, the syllabic example is speculated for another tune to create an ideal succession of syllables. The obtained sequence is presented to the Sentence Generation module. This module uses the Dijkstra's shortest path algorithm to think of an important expression coordinating the syllabic example [6]. Moreover, the author's Potash P., et al., exhibited the adequacy of a long short term memory language model for automatic rap lyrics generation. The model produces sections that are near in style to that of a given rapper. The model characterizes its rhyme plot, line length, and verse length. The examinations show that a Long Short-Term Memory language model delivers better "ghostwritten" verses than a standard model [7]. In the next paper, the authors Sung-Hwan Son, et al., have used deep learning for Korean song lyrics generation. They switched the K-pop verses information and utilized them as learning information. The setting between verses is considered in the proposed song lyrics generation method. Each time the model produces the verse, the model experiences upper randomization. It was affirmed that the verses produced utilizing the opposite information have a more characteristic setting than the verses created utilizing the forward information [8]. The authors Pablo Samuel Castro, et al., built up a framework which uses joined scholarly expressive structures and jargon for a verse age. They joined two separately trained language models into a framework that can deliver yield regarding the ideal tune structure while giving a wealth and decent variety of jargon that renders it all the more imaginatively engaging [9].Additionally, in another research the author's Fan H., et al., proposed a hierarchical attention based sequence to sequence model for Chinese lyrics generation. This model advances the subject of significance and consistency of generation by utilizing the encodings of word-level and sentence-level relevant data. For model training, a large Chinese lyrics corpus is likewise utilized. In the end, aftereffects of automatic and human evaluations demonstrate that the model can form total Chinese verses with one joined subject constraint [10]. The authors Manjavacas E., et al., used hierarchical modeling and conditional templates for the generation of hip hop lyrics. The model created is a straightforward system to separate and apply contingent formats from text pieces. The methodology that is proposed empowers start to finish preparing, focusing on formal properties of text, for example, musicality and rhyme, which are focal attributes of rap messages. A crossover type of hierarchical model is used. This intends to coordinate Language Modeling at two levels: word and character-level scales [11]. Lastly, in the paper proposed by the authors Naman Jain, et al., an automatics lyrics generator is executed for romanized Hindi. The proposed model uses simple techniques to catch rhyming examples previously and during the model training process in the Hindi language [12]. There has been quite a varied research and implementation in the domain of lyrics generation. The research paper proposes a different method by using the concept of deep learning LSTM networks specifically for romantic English songs. A succinct summary of the related papers is demonstrated in Fig.1.

Sr. No	Paper Title	Authors	Algorithm Used	Results
1	Automated generation of song lyrics using CFGs	S. Pudaruth, S. Amourdon and J. Anseline	Automated generation of song lyrics using CFG's	A majority of 52% of respondents thought that lyrics generated was from a human songwriter
2	Tra-la-Lyrics 2.0: Automatic Generation of Song Lyrics on a Semantic Domain	Gonçalo Oliveira	It is the improvement of the original Tra-la-Lyrics on a semantic area	Tra-la-Lyrics 2.0 got the highest scores, not only on the meaning but also sound, but there is still much room for improvement
3	DopeLearning: A Computational Approachto Rap Lyrics Generation	Eric Malmi, Pyry Takala, Hannu Toivonen, Tapani Raiko, and Aristides Gionis	Rap lyrics generation using a computational approach and expectation model	An accuracy of 17%, i.e., over 50 times more likely than by random, and it was ranked in the top 30 with 53% accuracy
4	Modeling Structural Topic Transitions for Automatic Lyrics Generation	Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui, and Masataka Goto	Hidden Markov Model (HMM) based system	This result indicates the superior effectiveness of line generation by the language model than by the contents models
5	Automatic generation of Tamil lyrics for melodies	Ramakrishnan A, A., Kuppan, S.	Sentence Generation module using the Dijkstra's shortest path algorithm	Generate a syllable pattern that closely matches the input tune.The identification of strong beats in the melody not considered though.
6	GhostWriter: Using an LSTM for automatic rap lyric generation	Potash, P., Romanov, A., & Rumshisky, A	Long short term memory language model	The correlation between rhyme density and max similarity for the n-gram model is 0.47.
7	Korean Song-lyrics Generation by Deep Learning	Sung-Hwan Son, Hyun-Young Lee, Gyu-Hyeon Nam,and Seung-Shik Kang	Deep learning based model	Confirmed that the lyrics generated using the reverse data, have a more natural context than the lyrics generated using forward data
8	Combining learned lyrical structures and vocabulary for improved lyric generation	Pablo Samuel Castro, Maria Attarian	Framework which uses joined scholarly expressive structures and jargon	Resulted desired song structure, while providing a richness and diversity of vocabulary.
9	Hierarchical Attention Based Seq2Seq Model for Chinese Lyrics Generation	Fan H., Wang J., Zhuang B., Wang S., Xiao J	A hierarchical attention based sequence to sequence model	Model is able to compose complete Chinese lyrics with one united topic constraint with BLEU score 0.288
10	Generation of hip-hop lyrics with hierarchical modeling and conditional templates	Manjavacas, E., Kestemont, M., & Karsdorp, F	Hierarchical modeling and conditional templates	Despite advantages of hierarchical modeling, effects of conditional templates did not compound and result is discouraging
11	Bollyrics: Automatic lyrics generator for romanised Hindi	Naman Jain, Ankush Chauhan, Atharva Chewale, Ojas Mithbavkar, Ujjaval Shah, Mayank Singh	Simple techniques to catch rhyming examples previously	54.5% of the paragraphs in N-Gram outputs and 64.5% of the paragraphs in LSTM model outputs were labeled as "Makes Sense"

Fig. 1. Comparison of related work

3 Implementation

The implementation is demonstrated in steps and shown in Fig.2:

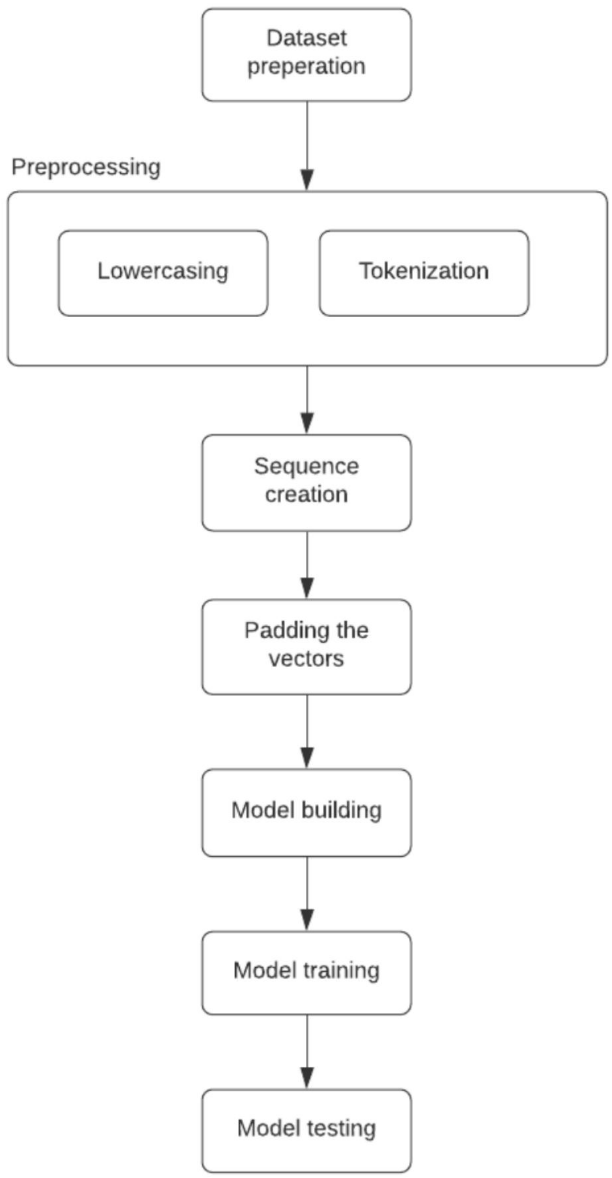


Fig. 2. Flowchart of the implementation

3.1 Dataset preparation

The dataset is unique and customized according to the use case. It consists of lyrics of English songs scraped from the web. The total number of words in the dataset is approximately 13k.

3.2 Preprocessing

The dataset is preprocessed to maintain consistency in data which will help further during the training of the model. The pre-processing techniques used are as follows:

Lowercasing The dataset contains both capital and small letters. Since the model will take the same word with a difference in capitalization as distinctive, it will create confusion. For instance, ‘The’ and ‘the’ is unique for the model to manage and may not anticipate the similarity on encountering it. Hence, making all the letters lowercase would wipe out this problem and improve the results.

Tokenization The data to be used in the model should be of numeric structure for further analysis and classification. This is accomplished by the process of ‘tokenization’. This helps by making the numeric tokens of each word present in the corpus.

3.3 Sequence creation

The next step is to turn the sentences into lists of values based on the tokens generated by the tokenizer. This is significant as the model will anticipate the next word on feeding the set of input words, we need to train it on a set of words with the next word as output to our model.

3.4 Padding the vectors

In padding, the list of sentences has been padded out into a matrix of the same length. This is achieved by putting the appropriate number of zeros before the sentence list. The matrix width is kept the same as the width of the longest sentence to make each vector of the same length.

3.5 Model building

Now, since all the required preprocessing is done, the next step is to build the model for training. The layers that are used for model building are as follows:

Embedding layer It is the first hidden layer of the neural network. The intuition is that words and related words are bunched as vectors in a multi-dimensional space. The input and output dimension along with the length is specified. For this layer, the yield is a 2D vector with one embedding for each word in the input sequence of words.

LSTM layer LSTM layer and bidirectional LSTM layer is used. The bidirectional layer helps in providing additional context to the network and results in a faster learning process. The bidirectional LSTM trains two LSTMs on the input sequence. The first one is the input sequence the way it is and the second on a switched duplicate of the input sequence.

Dropout layer This layer is introduced to reduce overfitting. At each update of the training phase, it works by arbitrarily setting the outgoing edges of hidden units to 0.

Dense layer This layer is fully connected. All the neurons in a layer are associated with those in the following layer. It is followed by the activation function which in this case is "softmax".

3.6 Model training

The model is compiled using the loss function "categorical cross-entropy" and optimized using "adam's" optimizer [13]. Then it is trained on 100 epochs by fitting the parameters.

3.7 Model testing

The model is tested with a set of input words to begin the lyrics with and word limit. The testing is to be carried out after preprocessing of the input and then parsing it through a function that predicts the next possible word for that phrase. This prediction runs in a loop to give out the chain of words to form lyrics.

The model used has 25 input variables and it is formed by firstly 'Embedding layers' with 160 neurons, the second layer is 'Bidirectional LSTM' with 400 neurons followed by a dropout layer of 0.2. Third is a LSTM layer with 100 neurons, followed by two dense layers with activation function 'relu' and 'softmax' respectively. The final dense layer of 23000 class is then followed by a compilation layer. A total set of trainable parameters (1,298) was compiled using 'adam' optimizer and 'categorical crossentropy' as the loss function. The layers of the CNN architecture can be seen in Fig.3.

4 Results

The dataset is trained on three models. These models are LSTM network, bidirectional network and GRU network model. The best results were obtained from the bidirectional LSTM model. This model is trained on approximately 18 lakh parameters for 100 epochs. The accuracy achieved is 82.6% and the loss is 0.40. The accuracy and the loss results for the models are shown in Fig.4. Since the

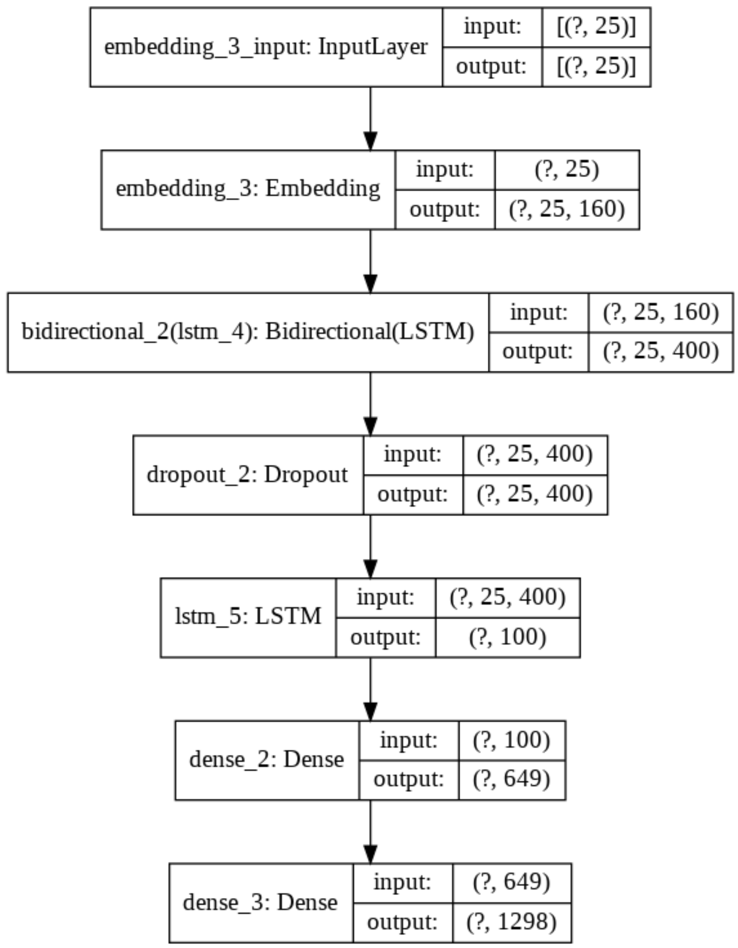


Fig. 3. Layers of the model

accuracy of all the models have very less difference, for more validity the output for different models are shown in Fig.5 for a sample input. The predicted lyrics for bidirectional LSTM model made more sense and had better sentence formation.

Conclusion

The lyrics inducer model was successfully built using natural language processing and deep learning. Bidirectional LSTM network has proved to efficiently remember the connection between the words which helps in predicting better song lyrics. The Training accuracy of 82% achieved can produce decently meaningful results according to the given seed input.

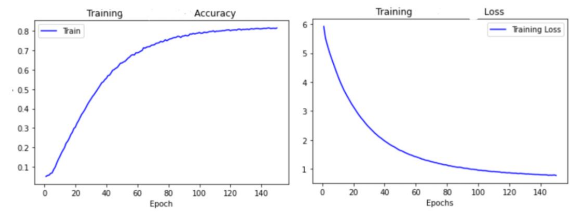
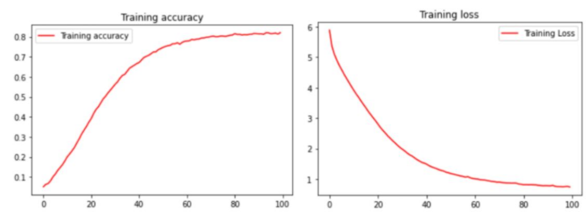
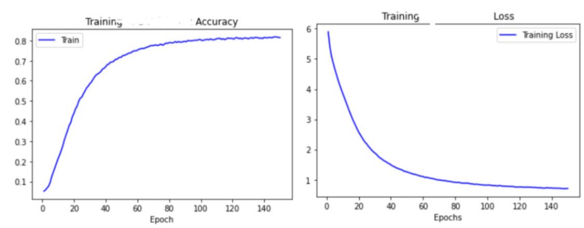
Sr. No	Algorithm	Training accuracy	Training loss	Plots
1	LSTM	81.50%	0.76	
2	Bidirectional LSTM	82.60%	0.43	
3	GRU	81.14%	0.71	

Fig. 4. The training accuracy and loss of LSTM,Bidirectional LSTM, and GRU

Future scope of the model could be to add more data to the corpus and train the model on more epochs. Furthermore, models can be built for separate genres for more genre-specific lyrics.

Input seed	Word limit	Prediction		
		LSTM	Bidirectional LSTM	GRU
Singing alone	50	Singing alone we make of what we met you fall my future in your temper is a promise me and now a bore is already gone gone gone gone gone gone already gone this gone love gone already indeed now already gone already gone already gone how to gone already gone already	Singing alone just about to get you crazy never give you everything so heavy come on the touch floor right there off doheny yeah every mood need you brr brr brr mind you need that lead eyes maybes is key for seventy minutes came like yeah every sin alone hey hey	Singing alone in my sleepless solitude tonight me laugh like i hates but i see my mind up yet we'll mend your heart and so big ground friend out missing wild diamond free free just water baby still it's no to i'm to give you after all these years no i'm not
Her hair	50	Her hair her hair falls perfectly without her trying to believe that now you let me go never let me go never let me gone beneath gone this high when you're already me gone love gone gone gone gone gone gone gone gone gone gone gone gone gone gone you already gone	Her hair curves i'm ghost just go unnecessary came south in ink and bleach like the '80's came shoot note we been trappin' like the '80's came shoot forever like me like through breaking but fm superhuman at the lot on the sea big jo ek jamati pourin' don't give you all	Her hair her hair falls perfectly without her trying at us to do us like i kissed you home me yet it's gone the pain truth just a little so used your eyes crazy crazy for you baby if you see me how to bend without the world caving in your faith
Toosie Slide	40	Toosie slide baby you got me like ah woo ah ow out to feeling what i'm feeling time with you baby i'm in love with you already gone this gone love to let it gone gone gone what you want to gone	Toosie slide you talking 'bout me i don't see a shade mm like to get me up in my bag to let me up and throw a tantrum all you just like to let me take you down i'm dwayne carter no	Toosie slide the whole world stops and stares for a while thing i can escape the stars they see the words of me i'll be your voice put a lips just kids girl lot just kids bit so dreaming wild love beautiful
The box	30	The box world is bright all right here all along in love babe again and again and again and again and again and again gone gone gone what gone this full gone	The box that tight they comfort in a man and bleach like monopoly baby say style I can carry dont do it is shine for me i don't know you care bout	The box whole world stops and stares for a while thing i can escape the stars they see the words of me i'll be your voice put a lips just kids girl

Fig. 5. Input and output results of LSTM,Bidirectional LSTM, and GRU

References

1. (n.d.). Institute of Bioinformaticspage. https://www.bioinf.jku.at/publications/older/2604.pdf
2. S. Pudaruth, S. Amourdon and J. Anseline, "Automated generation of song lyrics using CFGs," 2014 Seventh International Conference on Contemporary Computing (IC3), Noida, 2014, pp. 613-616, doi: 10.1109/IC3.2014.6897243.
3. Gonçalo Oliveira, H. (2015). Tra-la-Lyrics 2.0: Automatic Generation of Song Lyrics on a Semantic Domain, Journal of Artificial General Intelligence, 6(1), 87-110. doi: https://doi.org/10.1515/jagi-2015-0005
4. Eric Malmi, Pyry Takala, Hannu Toivonen, Tapani Raiko, and Aristides Gionis. 2016. DopeLearning: A Computational Approach to Rap Lyrics Generation. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 195–204. DOI:https://doi.org/10.1145/2939672.2939679
5. (n.d.). ACL Member Portal — The Association for Computational Linguistics Member Portal. https://www.aclweb.org/anthology/Y14-1049.pdf
6. Ramakrishnan A, A., Kuppan, S., & Devi, S. L. (2009). Automatic generation of Tamil lyrics for melodies. Proceedings of the Workshop on Computational Approaches to Linguistic Creativity - CALC '09. https://doi.org/10.3115/1642011.1642017

7. Potash, P., Romanov, A., Rumshisky, A. (2015). GhostWriter: Using an LSTM for automatic rap lyric generation. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. https://doi.org/10.18653/v1/d15-1221
8. Sung-Hwan Son, Hyun-Young Lee, Gyu-Hyeon Nam, and Seung-Shik Kang. 2019. Korean Song-lyrics Generation by Deep Learning. In Proceedings of the 2019 4th International Conference on Intelligent Information Technology (ICIIT '19). Association for Computing Machinery, New York, NY, USA, 96–100. DOI:https://doi.org/10.1145/3321454.3321470
9. Combining learned lyrical structures and vocabulary for improved lyric generation. (n.d.). arXiv.org. https://arxiv.org/abs/1811.04651
10. Fan H., Wang J., Zhuang B., Wang S., Xiao J. (2019) A Hierarchical Attention Based Seq2Seq Model for Chinese Lyrics Generation. In: Nayak A., Sharma A. (eds) PRICAI 2019: Trends in Artificial Intelligence. PRICAI 2019. Lecture Notes in Computer Science, vol 11672. Springer, Cham. https://doi.org/10.1007/978-3-030-29894-4_23
11. Manjavacas, E., Kestemont, M., & Karsdorp, F. (2019). Generation of hip-hop lyrics with hierarchical modeling and conditional templates. Proceedings of the 12th International Conference on Natural Language Generation. https://doi.org/10.18653/v1/w19-8638
12. Bollyrics: Automatic lyrics generator for romanised Hindi. (n.d.). arXiv.org. https://arxiv.org/abs/2007.12916
13. Diederik P. Kingma and Jimmy Lei Ba. Adam : A method for stochastic optimization. 2014. arXiv:1412.6980v9