

Suraj Janakiraman (sxj170022)
Dmitrii Obideiko (dxo200006)
CS 4395 Human Language Technologies
Due Date: 11-13-22

Portfolio: Reading ACL Papers: DeepRapper: Neural Rap Generation

This summary is focused on the analysis of the paper “DeepRapper: Neural Rap Generation.”

The author list is as follows: (1) Lanqing Xue, the first author is a student at The Hong Kong University of Science and Technology. (2) Kaitao Song graduated with a B.S. in computer science and engineering from the Nanjing University of Science and Technology. (3) Duocai Wu is pursuing research at the Software School of Fudan University. He is also a member of ACM (Association of Computing Machinery). (4) Xu Tan is a Principal Research Manager within the Machine Learning Group at Microsoft Research Asia. (5) Nevin L. Zhang received a PhD in Applied Math at Beijing Normal University in 1989. He also received a PhD in Computer Science at the University of British Columbia in 1994. He has been part of the Artificial Intelligence (AIJ), the International Journal of Approximate Reasoning (IJAR), and Journal of Artificial Intelligence Research (JAIR). (6) Tao Qin is a Senior Principal Researcher/Manager at Microsoft Research AI4Science Asia. He is also a senior member of ACM and IEEE, and a professor (PhD advisor) at the University of Science and Technology of China. (7) Tie-Yan Liu works at Microsoft Research AI4Science Asia as a scientist and an Assistant Managing Director. (8) Wei-Qiang Zhang has a Ph.D from Tsinghua University. He is currently an Associate Professor at the Department of Electronic Engineering at Tsinghua University.

The paper “DeepRapper: Neural Rap Generation,” calls into discussion in regards to a DeepRapper which is defined in the paper as “ a Transformer-based rap generation system that

can model both rhymes and rhythms” (69, Xue et. al). In essence, the authors Lingquin Xue and the others discuss and describe the invention of the DeepRapper that has the potential to act as an artificial replacement of notorious rappers such as Kanye West, DJ Khaled, Drake, etc. However, to perform any machine learning, deep learning, natural language processing, or artificial intelligence operations, data is a mandatory prerequisite. Before this project began, the authors realized that “there was no available dataset with rhythmic beats.” (69, Xue et. al.). The solution to this problem was to collect, process, and combine large amounts of data into a dataset. To implement the solution in the previous sentence, the authors implemented a technique known as data mining. The process of data mining would collect and process “ a large number of rap songs with aligned lyrics and rhythmic beats”(69,, Xue et. al) to formulate a potential dataset that can be used for performing Deep Learning algorithms used to create the DeepRapper.

Previous works of the rap generation were primarily focused on lyric generation. More particularly they were interested in developing strategies for rhyme modeling. For example, Nikolov developed a strategy where at first the program generated rap lyrics and then it added words that rhyme to the lyrics. Unfortunately, his strategy wasn’t consistently successful as the program couldn't guarantee a rhyme pattern for every single line. None of the prior projects that were focused on developing rap music produced any promising results due the nature of a complex structure in lyrics of any rap song. One of the major reasons why it’s so difficult to develop a program that generates rap lyrics is because not enough rap datasets are currently made available in order to teach a computer to make lyrics for rap music.

One of the major contributions that the authors have offered with respect to this paper is the data mining for the rap dataset. In the previous works, authors such as Nikolov and Potash

focused on lyric generation for rap music, leaving out a model that determines rhyme patterns and beats in rap music (69, Xue et.al.). The data mining for the rap set that was created by these authors made sure that rhythmic beats would act in alignment with the rhythmic lyrics (71, Xue et.al.). The data mining process consisted of 5 important stages: data crawling, vocal and accompaniment separation, vocal and lyric alignment, beat detection, and lyric and beat alignment (71, Xue et.al.). Another major contribution introduced in this paper is the autoregressive language model. As opposed to a standard autoregressive language model, a reverseorder language model can be used “to generate sentences from right to left.” (72,Xue et.al.). The overall autoregressive language model is necessary to the modeling of rhymes that complement rap lyrics. The DeepRapper also has this amazing capability of attempting to create N-gram rhymes. An N-gram rhyme occurs when “the current sentence and the previous sentence keep the same rhyme for the last N consecutive tokens.” (73, Xue et.al.). However the N-gram rhyme scheme is weak without two key aspects: rhyme modeling and rhyme constraint. (76, Xue et.al.). Nonetheless, the N-gram rhyme scheme is an excellent mechanism for counting rhymes in a given rap song. Another important contribution is the usage of the beat frequency, where a current beat interval has been defined strictly “as the number of words between current [BEAT] and the next [BEAT]” (76, Xue et.al). The beat frequency is a good mechanism in this case for providing and modeling a beat that can be used to complement rap lyrics. All these contributions are what make the DeepRapper, a potential Deep Learning version of the standard rapper in today’s society. However, despite all of its impressive features, it still has plenty of improvements.

On an objective level, the authors have chosen five metrics to evaluate the DeepRapper

project: Perplexity, Rhyme Accuracy, Rhyme Density, Combo-N, and Beat Accuracy. In the experimental results, the authors show on Table 2 (page 75), that with respect to Perplexity, Rhyme Accuracy, and Rhyme Density the DeepRapper performs better as opposed to the Baseline (standard autoregressive language model without the rhyme constraint and rhyme modeling), and Baseline + PT. Notice that the Perplexity for the DeepRapper is much lower as opposed to the Baseline and Baseline +PT models indicating that a lower perplexity level is a better perplexity. With respect to Combo-N, the Table 4 in page 76 showcases Combo-N levels for 1-gram (unigram), 2-gram (bigram) and 3-gram (trigram) rhymes. The DeepRapper model performed best as it contained both the rhyme constraint and the rhyme modeling. The Beat accuracy is measured through a mechanism known as the beat frequency control. Table 5 in page 76, mentions a First Order distribution and a Second Order distribution. The Beat frequency added on to the DeepRapper has increased the beat accuracy for the DeepRapper. On a subjective level, the DeepRapper performs better with respect to the following aspects: (1) the presentation of the theme of the rap and rap lyrics, (2) how fluent and refined the rap lyrics are in the rap song, (3) “the quality of the rhyme,” and (4) “the diversity of the rhyme,” (Xue et.al. 75) as opposed to the Baseline and Baseline + PT models in page 75.

Lanquin Xie has received 47 citations on google scholar in total since 2017. Kaitao Song received a total of 2,275 citations on google scholar. There is no page on Google Scholar discussing the number of citations regarding Duocai Wu. Xu Tan has received approximately 4,883 citations in total on google scholar. Nevin L. Zhang has 6,010 citations in total on google scholar. Tao Qin has received 16,335 citations on google scholar. Wei-Qiang Zhang has received 1,167 citations on google scholar. Tie-Yan Liu has received 40.729 citations which is the most

citations out of all the authors. Even though some of the impressive credentials of authors like Kaito Song, Xu Tan, Nevin L. Zhang, Tao Qin, Wei-Qiang Zhang, and Tie-Yan Liu, it is also clear that these authors have shown great interest in machine learning and natural language processing. Natural Language Processing at its basic level has been almost certainly about understanding human conversation between two people. These authors took NLP to another level by focusing on how natural language processing can be used to emulate certain musical talents such as rapping.

References

Section 1. Author Affiliations:

1. **Lanqing Xue:** <https://scholar.google.com/citations?user=6zAraEIAAAAJ&hl=en>
2. **Kaitao Song:** <https://ieeexplore.ieee.org/author/37086531230>
<https://scholar.google.com/citations?user=LLk9dR8AAAAJ&hl=en>
3. **Duocai Wu:** There is no google scholar article indicating the number of citations this person received.
<https://dl.acm.org/profile/99659259928>,
<https://www.researchgate.net/profile/Duocai-Wu-2>
4. **Xu Tan:** <https://www.microsoft.com/en-us/research/people/xuta/>,
<https://scholar.google.com/citations?user=tob-U1oAAAAJ&hl=zh-CN>.
- 5.
6. **Nevin L. Zhang:** <https://www.cse.ust.hk/faculty/lzhang/>,
https://scholar.google.com/citations?user=18_xlPUAAAAJ&hl=en
7. **Tao Qin:**
<https://www.microsoft.com/en-us/research/people/taoqin/>,
<https://scholar.google.com/citations?user=BI4SRU0AAAAJ&hl=en>
8. **Wei-Qiang Zhang:** <https://sites.google.com/site/weiqzhang/>,
<https://scholar.google.com/citations?user=OKupdYQAAAAJ&hl=en>
9. **Tie- Yan-Liu:**
<https://www.microsoft.com/en-us/research/people/tyliu/>,

<https://scholar.google.com/citations?user=Nh832fgAAAAJ&hl=en>.

Section 2. Sources for Summary:

Xue, Lanqing, et al. "DeepRapper: Neural Rap Generation with Rhyme and Rhythm Modeling." *ACL Anthology*, <https://aclanthology.org/2021.acl-long.6/>.