Automated Rhyme Detection in Text

Nguyen Thanh Vinh
VNU University of Engineering and Technology
21020673@vnu.edu.vn

Abstract

Rhyme detection play a crucial role in various fields such as poetry analysis, lyrics composition, and speech recognition. This paper presents an automated approach for detecting and visualizing rhymes in a given text. We propose a step-by-step methodology that leverages natural language processing techniques to identify rhyming patterns and create visually appealing representations of the detected rhymes. Our approach combines linguistic analysis, pattern matching, and data visualization to provide a comprehensive and intuitive understanding of rhyming structures within a text.

1 Introduction

Rhyming words is a simple task for humans, but an involved one for machines. For humans, identifying rhymes often relies on an innate understanding of language, sound patterns, and cultural exposure. We intuitively recognize the rhyming patterns within a poem or a song without a conscious effort. However, for machines, this task presents several challenges. Machines lack the inherent knowledge and context that humans possess, making it necessary to develop computational methods to analyze texts and identify rhymes systematically.

2 What is a rhyme?

A rhyme is a phonetic similarity between words or phrases, particularly in their ending sound. It is a poetic device that involves the repetition of similar or identical sounds at the end of lines or within lines of a poem or song. Rhymes are often used to create a sense of rhythm, musicality, and aesthetic appeal in written and spoken language.

In a rhyme, the similarity in sound can occur at various levels. The most common type of rhyme is an exact or perfect rhyme, where the sounds of the final stressed syllables of two or more words are identical, such as "cat" and "hat". In constrast is the imperfect rhyme or slant rhyme, which occurs when the sounds are similar but not identical. There's also multisyllabic rhymes, which involve the rhyming of two or more syllables, for example "elephant" and "relevant".

Rhymes can also be classified based on their position within a line. End rhymes are the most common type and involve words or phrases that rhyme at the end of lines. Internal rhymes occur within a line of poetry, where words within the same line rhyme with each other.

3 Word Pronounciation

Unlike Vietnamese and many other languages where pronunciation of a word can be inferred from its spelling, English words are often far from people's expectations. For example, the words "tough", "though", "through" and "thought" are all pronounced differently, despite having similar spellings. To determine if two words rhyme, we need to know how they are pronounced.

A word is made from a sequence of syllables. Syllables are the basic units of sound in spoken language. They are formed by combining one or more phonemes. Phonemes are the smallest distinct units of sound in a language that distinguish one word from another. They are the individual speech sounds that make up words.

Phonemes can be further categorized into vowels and consonants. Vowels are speech sound made with your mouth fairly open, while consonants are speech sounds made with your mouth fairly closed. In English, the vowels include the sounds represented by the letters a, e, i, o, u, and sometimes y, while the consonants include all the other letters of the alphabet.

A syllable is made up of an optional onset, nucleus, and an optional coda. The onset is the initial consonant sound or sounds that occur before the vowel sound in a syllable. The nucleus is the central vowel sound and typically the most prominent

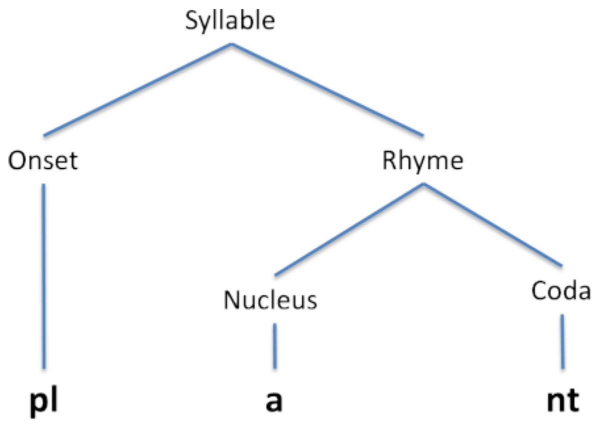

Figure 1: Composition of a syllable

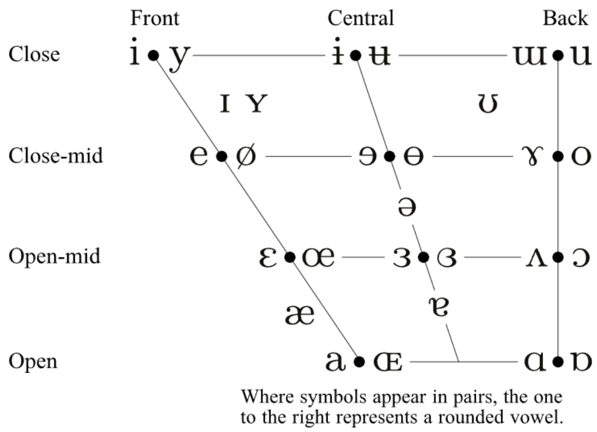

Figure 2: IPA Vowels Chart with features of height (vertical axis), frontness (horizontal axis) and rounding

part of a syllable. The coda is the final consonant sound or sounds that occur after the vowel sound in a syllable. Both the onset and/or coda may be empty. Example in figure 1.

4 Phonetic Features

To determine if 2 words rhyme, we need to determine the phonetic similarity of them. To do this we group vowels and consonants into classes based on their phonetic features, similar to previous works (Bay et al., 2019). Vowels are classified based on phonetic features as follows and shown in figure 2:

- Height: Height refers to the position of the tongue in the oral cavity during the production of a vowel. It describes the vertical distance between the highest point of the tongue and the roof of the mouth. Vowels can be classified as high, mid, or low.
- Backness: Backness describes the position of the tongue during the production of a vowel relative to the back of the mouth. Vowels can be classified as front, central, or back.
- Rounding: Rounding refers to the shape of the lips during the production of a vowel. Vowels can be classified as rounded or unrounded.
- Stress: Stress is a phonetic feature that refers to the emphasis or prominence given to certain syllables within a word or phrase. It involves the variation in loudness, pitch, and duration of a syllable to highlight its importance in the overall rhythmic structure of a language. Vowels can be classified as stressed or unstressed.

Consonants are classified based on phonetic features as follows and shown in figure 3:

- Place of Articulation: This feature describes where in the vocal tract the constriction occurs during the production of a speech sound.
- Manner of Articulation: This feature refers to how airflow is obstructed or modified during the production of a speech sound.
- Voicing: Voicing refers to whether the vocal cords vibrate or remain silent during the production of a speech sound. Sounds produced with vocal cord vibration are voiced, while sounds without vocal cord vibration are voiceless.

From the phonetic features of vowels and consonants, we put them in groups of similar sounding vowels and consonants, which will be used to calculate a vowel score and a consonant score.

5 Rhyme Detector Pipeline

5.1 Preprocessing

The text is preprocessed to remove punctuation, convert to lowercase, and tokenize into words.

5.2 Grapheme-to-Phoneme

Grapheme-to-Phoneme (G2P) is the task of converting a word from its written form to its spoken form. For this task we used the CMU Pronouncing Dictionary, which contains over 134,000 words with their pronunciations and stress patterns encoded using a modified form of the ARPABET system. We found that the ARPABET system is much more readable than the IPA system, and it is also easier to implement. For the words that are not found in the CMU Pronouncing Dictionary, we use a neural network model (Park and Kim, 2019) trained on

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

Figure 3: IPA Consonants Chart with features of manner of articulation (vertical axis), place of articulation (horizontal axis), and voicing (voiceless on left, voiced on right)

the CMU Pronouncing Dictionary to predict the pronunciation

5.3 Syllabification

After obtaining the pronunciation of words, the next step is to separate them into syllables. Syllabification is the process of dividing words into their constituent syllables, which helps in identifying the onset, nucleus, and coda of each syllable.

Syllabification is a crucial step because rhymes often occur at the syllable level. By separating the words into syllables, we can check for multisyllabic rhymes and analyze the rhyming patterns more effectively and accurately.

To syllabify the pronunciation, a syllabify module is employed. This module applies specific rules to determine syllable boundaries based on the phonetic characteristics of the language. These rules consider factors such as vowel sounds and consonant clusters to accurately divide the pronunciation into syllables.

5.4 Scoring rhymes

For rhyming, only the nucleus and coda matter, so at this stage, we remove the onset of each syllable.

We define 0 as the lowest possible rhyme score and 1 as the highest rhyme score, reserved for perfect rhymes. The rhyme score is calculated from the vowel score and the consonant score, with the vowel score weighted more heavily than the consonant score to reflect the importance of vowel sounds in rhyming. In the case where there are no consonants in both syllables, the rhyme score is equal to the vowel score.

We calculate the vowel score by first comparing the vowel sounds of the two syllables. If the vowel sounds are identical, a vowel score of 1 is assigned, indicating a high degree of similarity. If the vowel sounds belong to a predefined group of similar vowels, a vowel score of 0.75 is assigned. This captures a moderate level of similarity. In cases where the vowel sounds do not fall into any of the predefined groups, a vowel score of 0.5 is assigned, representing a lower degree of similarity.

To further refine the vowel score, we compare the vowel stress of the syllables. If the vowel stress is the same in both syllables, the vowel score is unchanged, indicating no penalty for differing stress patterns. However, if the vowel stress differs between the syllables, the vowel score is multiplied by 0.85, reflecting a slight reduction in similarity due to stress mismatch.

Moving on to the consonant score, we line up the consonant sounds of the two syllables and compare them one by one. If the number of consonant sounds differs between the syllables, the shorter list is padded with empty strings to ensure equal comparison. For each corresponding pair of consonant sounds, like vowel sounds, if the consonant sounds are identical, the consonant score is multiplied by 1. If the consonant sounds belong to a predefined group of similar consonants, the consonant score is multiplied by 0.75. If the consonant sounds do not fall into any of the predefined groups, the consonant score is multiplied by 0.5.

5.5 Creating Rhyme Groups

To group syllables into rhyme groups, we iterates through the lines of input text and compares the syllables of different word to identify rhyming patterns. The process compares the end of the current line and the next three lines to account for end rhymes, as well as comparing word within the same line and the next line to account for internal rhymes. If two syllables have a rhyme score past our rhyme score threshold, the code assigns the syllables to the same group. It also handles cases where syllables already belong to separate groups by merging them into a single group.

5.6 Visualization

The detected rhyme group are visualized using a graphical representation where each group is assigned a unique color. These colors are then used to highlight the rhyming syllables in the input text. The visualization provides a clear and intuitive way to identify rhyming patterns within a text. It is also a popular technique used in a lot of videos on Youtube to highlight rhymes in rap songs, for example in figure 5. In our implementation, we can only syllabify the pronunciation of words instead of the actual words, so we also display the original text in the visualization to make it easier to read.

6 Results

The output of the rhyme detector is shown in figure 4, and for comparison, the same text with rhyme scheme highlighted by hand is shownn in figure 5. The results of the rhyme detector is shown in table 1. In the input text which includes 309 rhyming syllables creating 51 rhyme groups and 42 non rhymes, the rhyme detector correctly identified 279 rhymes, 28 non-rhymes, with 30 false negatives and 21 false positives. The rhyme detector has an accuracy of 0.83, a recall of 0.93, and a precision of 0.90. When taken into consideration word pronunciations incorrectly predicted by the G2P model, the rhyme detector has an accuracy of 0.87, a recall of 0.94, and a precision of 0.93.

	With G2P	Without G2P
True positive	279	279
True negative	28	28
False Positive	21	18
False Negative	30	21
In Wrong Group	10	8
Accuracy	0.83423913	0.867232
Recall	0.93	0.939394
Precision	0.90291262	0.93

Table 1: Results of the rhyme detector. False positives include syllables detected in wrong group.

7 Discussion

As rhyming is a subjective matter, our method is not perfect. The speaker can choose to pronounce a word in a way that is not expected, and the rhyme detector will not be able to detect the rhyme.

There are a lot of room for improvement for the rhyme detector still. The G2P model used is not state-of-the-art, leading to inaccuracies. The rhyme detector also doesn't actually learn from the position of the rhymes and employs a greedy algorithm to detect all rhyme pair within 2 lines even when not intentional by the author of the text. We also didn't actually classify the vowels and consonants into feature groups, and in future work we plan to do use a machine learning algorithm to optimize the weight of each feature group for better accuracy of the rhyme scoring function.

Code for our implementation of this paper can be found on GitHub[2].

References

Benjamin Bay, Paul Bodily, and Dan Ventura. 2019. Dynamically scoring rhymes with phonetic features and sequence alignment. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 1581–1585.

Kyubyong Park and Jongseok Kim. 2019. g2pe. `https://github.com/Kyubyong/g2p`.

[1] `https://www.youtube.com/watch?v=tlhealMskXo`

[2] `https://github.com/21020673/rhyme-detector`

Tripping off the beat kinda, dripping off the meat grinder

TRIH1-PIH0NG AO1F DHAH0 BIY1T KIH1N-DAH0 DRIH1-PIH0NG AO1F DHAH0 MIY1T GRAY1N-DER0

Heat niner, pimping, stripping, soft sweet minor

HHIY1T NAY1-NER0 PIH1M-PIH0NG STRIH1-PIH0NG SAA1FT SWIY1T MAY1-NER0

China was a neat signer, trouble with the script digits

CHAY1-NAH0 WAA1Z AH0 NIY1T SAY1-NER0 TRAH1-BAH0L WIH1DH DHAH0 SKRIH1PT DIH1-JHAH0TS

Double dipped, bubble lipped, subtle lisp midget

DAH1-BAH0L DIH1PT BAH1-BAH0L LIH1PT SAH1-TAH0L LIH1SP MIH1-JHAH0T

Borderline schizo, sort of fine tits though

BAO1R-DER0-LAY2N SHIY1-TOW0-SOW0 SAO1RT AH1V FAY1N TIH1TS DHOW1

Pour the wine, whore to grind, quarter to nine, let's go

PAO1R DHAH0 WAY1N HHAO1R TUW1 GRAY1ND KWAO1R-TER0 TUW1 NAY1N LEH1TS GOW1

Ever since ten eleven, glad she made a brethren

EH1-VER0 SIH1NS TEH1N IH0-LEH1-VAH0N GLAE1D SHIY1 MEY1D AH0 BREH1DH-RAH0N

Then it's last down, seven alligator seven

DHEH1N IH1TS LAE1ST DAW1N SEH1-VAH0N AE1-LAH0-GEY2-TER0 SEH1-VAH0N

At the gates of heaven, knocking - no answer

AE1T DHAH0 GEY1TS AH1V HHEH1-VAH0N NAA1-KIH0NG NOW1 AE1N-SER0

Slow dancer, hopeless romancer, dopest flow stanzas

SLOW1 DAE1N-SER0 HHOW1-PLAH0S ROW0-MAE1N-SER0 DOW1-PEH2ST FLOW1 STAE1N-ZAH0Z

Figure 4: Output from the rhyme detector. In the syllables, 0 denotes unstressed, 1 denotes primary stress, and 2 denotes secondary stress.

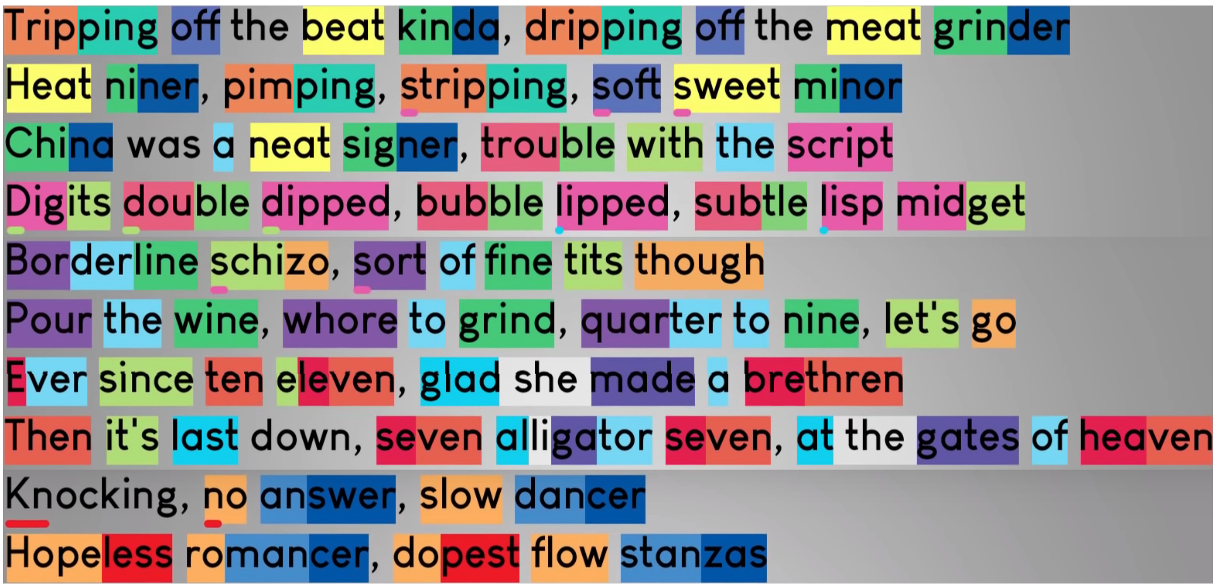

Figure 5: The same text with rhyme scheme highlighted by hand, taken from the video "MF DOOM - Meat Grinder | Rhymes Highlighted" [1] by Youtube user "Highlighted"