RHYME ANALYZER: AN ANALYSIS TOOL FOR RAP LYRICS

Hussein Hirjee **Daniel G. Brown**
University of Waterloo
Cheriton School of Computer Science
{hahirjee, browndg}@uwaterloo.ca

1. METHOD

We present a lyrical analysis and classification tool (`http://www.bioinformatics.uwaterloo.ca/~hahirjee/rhymes/`) focused specifically on rhyming style in rap lyrics. Five functions are available for processing lyric input to the Rhyme Analyzer: phonetic transcription, assigning similarity scores to lines, displaying detected rhymes, calculating rhyme features, and classifying the rapper.

1.1 Displaying Rhymes

For each syllable in the input lyrics, we find the longest and nearest rhyming matches from the current and preceding lines, after our previously described method [1]. For all participating words in each rhyme pair in the text, we apply one of five modified formatting styles: bold face, italic, red colour, underline, and strike-through. These formatting styles are not mutually exclusive, allowing words to be displayed as part of multiple rhymes (see Figure 1).

1.2 Analyzing Rhyme Style

We calculate a variety of statistical features about the detected rhymes in the input lyrics, providing a quantitative characterization of rhyming style. These statistics include rhymes per line and syllable, distribution of rhymes by syllable-length, proportion of internal and end rhymes, average rhyme scores, proportion of perfect rhymes, and complex internal rhyme features described by Alim [2].

1.3 Classifying Rappers

We use rhyme features calculated for songs from 53 popular albums by 25 famous rappers as the training set of instances for a Simple Logistic Regression from Weka Data Mining Software [3]. This model serves as the classifier used to evaluate the instance generated from the rhyme features of the input text. We return the most similar MC (from the set of 25) as the guessed writer of the lyrics.

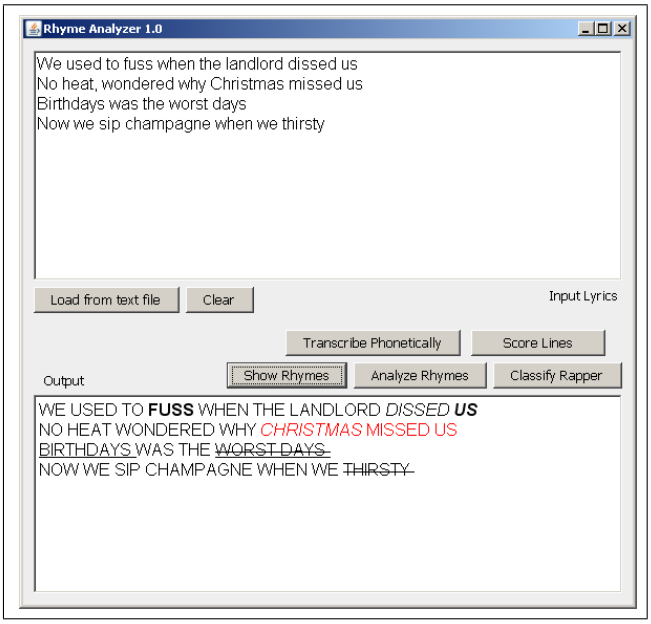

Figure 1. A visualization of detected rhymes from The Notorious B.I.G.'s "Juicy."

2. RESULTS

From the 25 rappers analyzed in depth, we found that calculated features appeared to be characteristic of artistic style, allowing for rhyme-based lyrical artist classification. We evaluated this classification using 10-fold cross-validation, in which a model was trained on 90% of the instances and used to classify the remaining 10%. The results were surprisingly good, with 314 out of 603 instances (52%) being classified correctly and a weighted F-measure of 0.516. All rappers were most often classified as themselves, with the exception of KRS-One who was most often classified as Rakim.

3. REFERENCES

[1] H. Hirjee and D.G. Brown: "Automatic Detection of Internal and Imperfect Rhymes in Rap Lyrics," *Proceedings ISMIR 2009*, 2009.

[2] H. Samy Alim: "On Some Serious Next Millennium Rap Ishhh" *Journal of English Linguistics*, Vol. 31, No. 1, pp. 60–84, 2003.

[3] M. Hall et al.: "The WEKA data mining software: An update" *ACM SIGKDD Explorations Newsletter*, Vol. 11, No. 1, pp. 10–18, 2009.