ABSTRACT

Poetry, as a special form of literature, is crucial for computational linguistics. It has a high density of emotions, figures of speech, vividness, creativity, and ambiguity. Poetry poses a much greater challenge for the application of Natural Language Processing algorithms than any other literary genre.

Our system establishes a computational model that classifies poems based on similarity features like rhyme, diction, and metaphor. Metaphor is a part of diction indeed, due to the complexity involved in its detection, we dedicate an entire chapter to it.

For rhyme analysis, we investigate the methods used to classify poems based on rhyme patterns. First, the overview of different types of rhymes is given along with the detailed description of detecting rhyme type and sub-types by the application of a pronunciation dictionary on our poetry dataset. We achieve an accuracy of 96.51% in identifying rhymes in poetry by applying a phonetic similarity model. Then we achieve a rhyme quantification metric *RhymeScore* based on the matching phonetic transcription of each poem. We also develop an application for the visualization of this quantified *RhymeScore* as a scatter plot in 2 or 3 dimensions.

For diction analysis, we investigate the methods used to classify poems based on diction. First the linguistic quantitative and semantic features that constitute diction are enumerated. Then we investigate the methodology used to compute these features from our poetry dataset. We also build a word embeddings model on our poetry dataset with 1.5 million words in 100 dimensions and did a comparative analysis with GloVe embeddings.

Previous work on metaphor detection relies on either rule-based or statistical models, none of them applied to poetry. Our method focuses on metaphor detection in a poetry corpus, but we test on non-poetry data as well. It combines rule-based and statistical models (word embeddings) to develop a new classification system. Our system has achieved a precision of 0.759 and a recall of 0.804 in identifying one type of metaphor in poetry by building a machine learning model. We also build a deep learning model for metaphor detection that achieves a precision of 0.831 and a recall of 0.836 in identifying one type of metaphor in poetry. We also develop an application for generic metaphor detection in any type of natural text.