The need for optimal models of language processing is universal. Industries have long examined natural language processing to gain insights in meeting the needs of customers. Models of language classification also contribute to the design of assistive technologies and teaching methods for individuals that may experience difficulties in reading, writing, processing audio information, or in speech. Semantic properties of words are an essential aspect of linguistic models. The purpose of this project is to explore evolutionary computing in testing and potentially improving word modeling. Due to the time constraints, this will be a partial novel research and partial algorithm comparison project. Explore how to optimize/utilize textual data with genetic algorithms.

The set used is from Buchanan et al. (2019) consisting of over 16,000 concept-feature responses. In a concept-feature task, respondents are provided with a cue consisting of a single word to which they are asked to provide a typed response. These responses are analyzed and nominally classified. Words were classified as part of speech, including adjectives, nouns, and verbs. These were further classified as either concrete or abstract terms. Concrete terms are more easily visualized and sensed, while abstract terms require more higher-order thinking. The process identifies stop words, generates lemmas, groups of words with properties in common. Relationship encoding is based on distance. Baayen et. al suggests word composition and semantic meaning is secondary to linear discriminative learning, emphasizing phonemes, particularly trigrams, over morpheme based theories.

Research Objectives include

Optimize lemmatization

Replication using predictive modeling toward alternative dataset

Development of proper semantic matrices

Predictive methods from DISC model applied toward reduced word forms present in conversational speech

Expanding hidden layers present in DISC model and testing results.

References

Baayen, R. H., Chuang, Y. Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning. Complexity, 2019.

Buchanan, E. M., De Deyne, S., & Montefinese, M. (2019, July 24). A practical primer on processing semantic property norm data. https://doi.org/10.31219/osf.io/qb8en