**Interpretable Language Model with Metaheuristic Optimization Using Word2Vec Embeddings**

**Abstract**

Large Language Models (LLMs) are typically deep neural networks that compute complex non-linear functions to map input vectors, representing preceding text, to output vectors, representing subsequent text. However, the opaque nature of these models raises significant concerns regarding their use in fields such as education, law, finance, politics, and medicine. This research aims to introduce a novel approach to language modeling that enhances interpretability and cost-efficiency. We propose employing metaheuristic optimization to simulate the non-linear mappings of large language models. Initially, words will be vectorized using contextual embeddings like word2vec. Each observed word embedding will then be represented as a function of its preceding word embedding(s). Finally, metaheuristic optimization will be utilized to optimize the cosine similarity and Euclidean distance between the predicted and observed text embeddings.

**Research Goals**

Enhancing cooperation between humans and intelligent systems relies on increasing the perceived trustworthiness of the latter. The high-dimensional hyperparameter space in neural networks fundamentally hinders user trust. By reducing or eliminating this complexity, we aim to achieve:

* Lower computational costs
* Improved interpretability
* Enhanced robustness in terms of technical security and soundness
* Greater fairness

**Model Description: The** process begins by segmenting sentences into sets of n sequential words, or n-grams, which are then converted into vectors using pre-trained word2vec embeddings. For each n-gram, a function of the first n-1 word vectors will predict the nth word vector across all sentences. These predictions will be based on multiple polynomial (potentially recursive) non-linear functions. Metaheuristic optimization will then be applied to optimize the cosine similarity and Euclidean distance between the predicted and observed word embeddings. The parameters for the proposed functions will be determined using metaheuristic optimization. Initial experiments will be conducted using the Wall Street Journal version of the Penn Treebank Dataset. Both cosine similarity and Euclidean distance measures will be utilized to compute the objective function, as both vector phase and magnitude are critical for vector analysis.