ITESM Campus Monterrey

Proof of Concept: Natural Language Processing: Exploring Word Embeddings

Luis Alberto Portilla López
September 20th, 2024

Research Stay - Going beyond Artificial Intelligence: Artificial Emotions

TC3073 | Group 573

Introduction

Word embeddings have become fundamental in Natural Language Processing (NLP), offering numerical vector representations of words that capture semantic relationships based on context. These representations enable a wide range of applications, from sentiment analysis to machine translation. However, the effectiveness of embeddings depends on their quality, which varies significantly depending on the evaluation method employed. This PoC aims to compare different word embeddings using both intrinsic and extrinsic evaluation methods to determine their strengths and limitations.

Business Problem

Evaluating the quality of word embeddings is critical for their effective application in various NLP tasks. However, existing evaluation methods face several challenges:

- Variability Across Tasks: Embeddings that perform well in one context (e.g., sentiment analysis) may perform poorly in another (e.g., syntax parsing), highlighting the need for context-specific evaluations.

- Subjectivity in Evaluation: Intrinsic evaluations, which rely on human judgments, can be subjective and inconsistent, affecting the reliability of results.

- Bias and Fairness: Many embeddings reflect biases present in their training data, impacting their performance and fairness in applications like sentiment analysis or named entity recognition.

These challenges necessitate a comprehensive evaluation approach that considers both intrinsic linguistic properties and extrinsic task performance.

Proposed Solution

To address these challenges, this PoC proposes a comparative evaluation of different word embeddings using intrinsic and extrinsic methods. The approach involves:

- Intrinsic Evaluation:
  - Relatedness Task: Measuring the similarity between word pairs using datasets like WordSim-353.
  - Coherence Task: Implementing an intrusion task to test the consistency of semantic clusters within the embeddings.
- Extrinsic Evaluation:
  - Sentiment Analysis: Using word embeddings in a sentiment classification model to assess their impact on task performance.
  - Noun Phrase Chunking: Evaluating the embeddings' effectiveness in syntactic tasks by measuring accuracy in chunking noun phrases.
- Feature Engineering and Preprocessing
  - Preprocessing: Standardize text data by lowercasing, tokenizing, and removing stopwords to ensure consistent input for embedding evaluations.
  - Embedding Preparation: Train or source embeddings (e.g., GloVe, Word2Vec, FastText) using similar corpora to ensure a fair comparison.
- Model Optimization and Training
  - Intrinsic Task Implementation:
    - Use cosine similarity and correlation metrics to compare human judgments with model-generated similarity scores.
    - Perform the coherence intrusion task using human evaluators on crowdsourcing platforms like Amazon Mechanical Turk.
  - Extrinsic Task Implementation:
    - Integrate embeddings into sentiment analysis and chunking models using frameworks like NLTK, SpaCy, or Hugging Face.
    - Evaluate task performance using accuracy, precision, recall, and F1-score metrics.

Expected Outcomes

The evaluation of word embeddings through this PoC aims to achieve the following:

- Improved Understanding of Embedding Performance: By comparing embeddings across intrinsic and extrinsic evaluationswe can highlight which embeddings are best suited for specific tasks.

- Insights into Task-Specific Strengths: Results will help identify the strengths and weaknesses of embeddings in different contexts, guiding their application in real-world NLP tasks.

- Bias and Fairness Assessment: Evaluating how embeddings handle biased or sensitive contexts, informing future approaches to training fairer embeddings.

Conclusion

This PoC outlines a structured approach to evaluating word embeddings using intrinsic and extrinsic methods, offering insights into their contextual performance. By addressing current evaluation challenges, this approach will guide the selection and adaptation of embeddings for specific NLP tasks, enhancing the overall effectiveness of these fundamental language models.