Capturing the Essence of a Phrase: Extracting Physical and Sensory Information from Text

Abhinav Gupta and **Jesse Thomason**
University of Southern California, Los Angeles, CA

Abstract

We developed a model that extracts sensory and physical data across 11 modalities from word embeddings, revealing intriguing insights. Notably, names of places were often associated with region-specific foods. We're leveraging these sensorimotor associations to enhance tools like recommender systems for recipe substitutions and emoji prediction—suggesting emojis based on a sentence's sensory context. This study opens new avenues for integrating sensory information into natural language processing tasks.

1 Introduction

Human cognition relies on the sensory and motor inputs from various organs, and people associate different sensorimotor modalities with different objects. For example, the concept of an 'apple' is closely tied to gustatory perception, while 'music' aligns with auditory perception. The Lancaster Sensorimotor Norms, a dataset from Lancaster University (Lynott et al., 2019), captures these associations by providing sensorimotor strengths for around 40,000 phrases across 11 modalities. These norms offer qualitative insights into how words are linked to sensory experiences.

Traditional lexical word embeddings such as Word2Vec and GloVe, while effective in capturing semantic and syntactic relationships between words based on their co-occurrence patterns in large text corpora, are not explicitly learned to store any qualitative data or sensorimotor attributes about the words they encode. However, by analyzing what words occur in similar contexts, these embeddings might be able to implicitly encode some sensorimotor attributes. The primary focus of this project is to study the relationship between lexical word embeddings and modern contextual word embeddings from large language models (such as BERT and GPT) to the Lancaster Sensorimotor Norms.

We hypothesize that patterns of word usage in large corpora may reflect their underlying sensorimotor associations with the physical world. Support for this hypothesis opens possibilities for physical and sensory inference using text alone, a possibility hotly debated in modern language grounding discourse. Rejecting this hypothesis opens an orthogonal set of opportunities to inject learned representations with sensorimotor norm information, potentially enabling that same physical and sensory inference. The ability to infer physical and sensory information from text could also significantly enhance several downstream applications, opening up new possibilities for tasks like sarcasm detection, emoji prediction, and recommender systems.

2 Background

The Lancaster Sensorimotor Norms database, retrieved from the Lancaster university's website (Lynott et al., 2019) consists of norms of sensorimotor strength across six perceptual modalities (touch, hearing, smell, taste, vision, and interoception) and five action effectors (mouth/throat, hand/arm, foot/leg, head excluding mouth/throat, and torso). The data was gathered by surveying 3,500 participants, who rated the degree to which each of the 39,707 words in the collection was associated with specific perceptual and action modalities, using a scale from 0 to 5.

3 Methods

3.1 Dataset Preparation

In preparing the datasets, we faced challenges aligning phrases from the Lancaster Sensorimotor Norms with the chosen word embeddings, Word2Vec and GloVe. Some phrases did not have a direct match in the embeddings, especially multiword phrases. For these cases, we averaged the vectors of each individual word in the phrase to represent it. If a phrase had no corresponding vec-

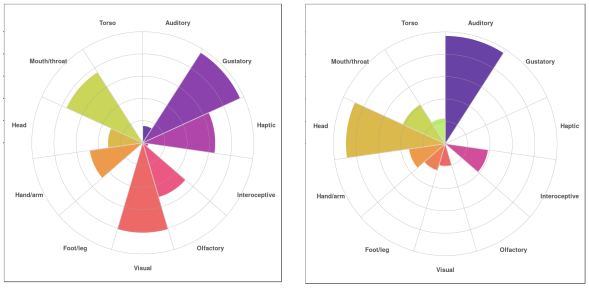

Figure 1: Lancaster Norms plots for the word 'apple' on the left and the word 'music' on the right, illustrating the difference in sensorimotor

tors in the embeddings, it was excluded from the final dataset.

3.1.1 Word2Vec - Lancaster.

The dataset comprised the Lancaster Sensorimotor Norms and Word2Vec embeddings imported from the open source Gensim library (Řehůřek and Sojka, 2010). A parallel corpus was created by attempting to match each Lancaster sensorimotor vector with its corresponding Word2Vec embedding. A parallel corpus was created by aligning each Lancaster sensorimotor vector with its corresponding Word2Vec embedding and after handling unmatched phrases as described, the final dataset consisted of 33,798 parallel words.

3.1.2 Glove - Lancaster.

Similar to Word2Vec - Lancaster, this dataset was created by matching each Lancaster sensorimotor vector with its corresponding Glove embedding (Pennington et al., 2014). This dataset had a size of 34,346 parallel words.

3.1.3 BERT CLS - Lancaster.

For each phrase in the Lancaster dataset, we used the BERT base model (Devlin et al., 2018) to extract the CLS token embeddings. This allowed us to create a parallel corpus by aligning each BERT CLS token embedding with its corresponding Lancaster sensorimotor vector.

3.1.4 Intersection Data.

"Intersection words" were selected as the words or phrases that appear in all three datasets—Word2Vec, GloVe, Lancaster Norms—and the BERT vocabulary. For each of the original datasets, we created a new Intersection version, limiting the content to these common words. This ensured that all datasets contained the same words, allowing for a fair comparison of model performance across different embeddings. The results presented in this paper are based on this Intersection dataset.

3.2 Modeling Experiment

3.2.1 Baseline Model.

The Baseline Model calculates the mean of the sensorimotor vectors from the training set and uses this average to make predictions for all test inputs. It serves as a benchmark for evaluating more complex models.

3.2.2 K-Nearest Neighbors (KNN).

The KNN model predicts sensorimotor attributes by finding the five most similar examples in the training set based on cosine similarity. These similarities are used to compute a weighted average of the sensorimotor vectors, which forms the final prediction. The similarity scores are scaled between 0 and 1 to ensure that closer matches contribute more significantly to the prediction.

3.2.3 Neural Network.

The Neural Network was implemented using PyTorch, featuring one hidden layer with dimensions of 64 and 128 neurons. The input size varies based on the embedding type: 300 for Word2Vec, 100 for GloVe, and 768 for BERT CLS. ReLU activation functions were applied between layers, and the network's output is an 11-dimensional vector corresponding to the Lancaster Sensorimotor Norms. The network is trained using the Mean Squared Error Loss function and the Adam optimizer, with a learning rate of 0.001. Training spans 10 epochs with a batch size of 128, utilizing a DataLoader for efficient data handling.

3.3 Word and Emoji Clustering

After training a model to extract sensorimotor data from word embeddings, we analyzed the BERT vocabulary. Most words clustered around the average sensorimotor vector from the Lancaster Dataset, so we sorted BERT CLS tokens by their distance from this average and found the five nearest neighbors for each for the outliers. Additionally, we tested if emojis could be used for sensorimotor prediction by tokenizing them with the BERT tokenizer, extracting CLS embeddings, and identifying the closest words using a similar method.

Embedding	Baseline	Neural Net	KNN
Word2Vec	0.028	0.014	0.0138
GloVe	0.027	0.017	0.017
BERT CLS	0.027	0.016	0.021

Table 1: Average Test losses accross various models predicting sensorimotor information from different word embeddings

4 Results and Discussion

4.1 Performance by Embedding

We evaluated Word2Vec, GloVe, and BERT CLS embeddings using a Baseline model, Neural Network, and K-Nearest Neighbors (KNN). Table 1 summarizes the average test losses for each combination.

4.1.1 Word2Vec Results

The Baseline achieved a loss of 0.028, while the Neural Network reduced it to 0.014, and KNN slightly outperformed it with a loss of 0.0138.

4.1.2 GloVe Results

The Baseline had a loss of 0.027, with both Neural Network and KNN showing similar performance at 0.017.

4.1.3 BERT CLS Results

The Baseline and Neural Network both performed similarly, with the Neural Network slightly better at 0.016. KNN had a higher loss of 0.021, suggesting that BERT CLS is less effective at capturing sensorimotor attributes.

4.2 Modality Analysis

All word embeddings excelled in predicting 'Gustatory' attributes, achieving a low loss of 0.0056. This strong performance suggests potential for developing food-related applications, such as food recommendation systems or recipe suggestions, leveraging the accurate prediction of taste-related concepts.

4.3 Place Names and Food Neighbors

Table 2 reveals that place names like "napoli" and "roma" have food items as their closest sensorimotor neighbors, such as 'LEMONADE' and 'CAKE.' This suggests a cultural or contextual link between these locations and specific foods. Such associations could be utilized in food recommendation systems tied to regional cuisines or in context-aware NLP tasks, enhancing applications like sentiment analysis and recommendation engines.

Word	Sensorimotor Neighbors
napoli	['LEMONADE', 'ORANGE JUICE', 'EGGNOG', 'LISTERINE', 'EDIBILITY']
roma	['CAKE', 'CREME BRULEE', 'CROISSANT', 'MANGO', 'SWEET PEPPER']
padua	['EGGNOG', 'MARMALADE', 'POTPIE', 'CHEDDAR', 'MILK CHOCOLATE']
laval	['MARMALADE', 'ESCARGOT', 'WASABI', 'LENTIL', 'CHARDONNAY']
mmm	['DELECTABLY', 'SWEETLY', 'MALTY', 'ORALLY', 'MUNCH']
lagos	['HEADCHEESE', 'GUMBO', 'TRUFFLE', 'SALAD OIL', 'PARSLEY']
podcast	['STORY', 'LYRIC', 'BIGMOUTH', 'POEM', 'CONVERSATIONAL']
luisa	['GUMBO', 'ESCARGOT', 'MESQUITE', 'UNEATABLE', 'CHARDONNAY']

Table 2: Word Neighbors in Sensorimotor Lancaster Space

Emoji	Sensorimotor Neighbor
🍎	['PINEAPPLE', 'APRICOT', 'RED CABBAGE', 'BOYSENBERRY', 'BELL PEPPER']
😋	['INDULGE', 'CONSUMING', 'NUTRITIOUS', 'NUTRIENT', 'NICOTINE']
🤦	['MISCOMMUNICATION', 'CRITICIZE', 'INARTICULATELY', 'ACCUSE', 'SNARKY']
🍣	['CASH REGISTER', 'XYLOPHONE', 'FIDDLE', 'BLOW DRYER', 'CD PLAYER']
🗣️	['COMPLAININGLY', 'ENUNCIATOR', 'PERSUADE', 'QUIETING', 'YELP']

Figure 2: Table showing certain emojis and their sensorimotor neighbors

4.4 Emoji Analysis

Figure 2 shows sensorimotor neighbors for selected emojis. For example, the emoji showing an orange is associated with food-related terms like 'PINEAPPLE' and 'APRICOT,' while the last emoji with a man talking links to communication actions such as 'PERSUADE' and 'YELP.' These findings suggest that sensorimotor embeddings capture nuanced emoji meanings instead of just describing the emoji with contextual neighbors of the emoji's description. This could enhance tasks like emoji prediction and sentiment analysis by providing richer context.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Dermot Lynott, Louise Connell, Marc Brysbaert, James Brand, and James Carney. 2019. The lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 english words. *Behavior Research Methods*, 52(3):1271–1291.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks, pages 45–50, Valetta, MT. University of Malta.