Predicting Sensorimotor Perception Strengths

Tobias Nusser
University of Konstanz
tobias.nusser@uni-konstanz.de

Abstract

The human perception is a complex concept and measuring on how humans perceive information is an intensively studied topic. Most often this is done manually through surveys and manual annotation by humans. The Lancaster Sensorimotor Norms data set measured the sensorimotor strength of 39,707 words across six perceptual modalities such as touch, hearing, smell, taste, vision and interoception. We use this data set as gold standard and train machine learning algorithms using word embeddings to predict perceptual strength values automatically. We achieve an aggregated mean absolute error of 0.345 across all modalities which suggests that word embeddings are able to model sensorimotor information of words and machine learning algorithms are then able to decently predict its perceived strength. Github Repository

1 Introduction

Sensorimotor information is central to how we experience the world and therefore plays a fundamental role in cognition (Lynott et al., 2019). As we interact with people, objects or the environment we perceive it through multiple channels. Most dominantly, humans perceive through vision, hearing, touching, smelling, tasting and interoception (Lynott and Connell, 2013). Using this information we form a conceptual view on the objects or people we interact with, or the feelings we experience and situations we encounter. Simultaneously we often link this newly gained information to one or more of the sensorimotor modalities in varying strengths. To quantify these modality-specific measures of sensory experience and test these theories of cognition researchers developed empirical tests and surveys (Lynott and Connell, 2009), which suggested these modalities to be the grounding mechanism for certain concepts (Connell et al., 2018). Most recently, Lynott and Connell et al. developed *The Lancaster Sensorimotor Norms: multidimensional measures of perceptual strength for 40,000 English words* (Lynott et al., 2019). A data set consisting of 39,954 English lemmas that are judged by humans based on their perceptual strength across the six modalities on a scale from 0 (meaning not experienced at all with that sense) to 5 (meaning experienced greatly with that sense). Using this data set as a gold standard it would be interesting to know if word embeddings and state-of-the-art machine learning algorithms are able to predict similar values as the human judges. This would indicate once more the significance and great performance of word embeddings in language tasks and furthermore suggest that word embedding models combined with machine learning models are able to mimic to a certain extent our understanding of language and semantic.

RQ: How good can we predict the sensory strength of experience of humans using word embeddings and machine learning algorithms?

2 Methodology

We used two different word embedding models - ConceptNet Numberbatch (Speer et al., 2016) with a 300 dimensional vector and BERT (Devlin et al., 2018) with a 768 dimensional vector. Whereas BERT handles arbitrary textual input, ConceptNets dictionary does not include a total of 835 words which is why we excluded them from the analysis for both approaches for better comparability. Table 1 shows the rating statistics for the remaining 38,872 words from which we then obtained word embedding vectors.

Since we want to produce numerical values, this

	Auditory	Gustatory	Haptic	Interoceptive	Olfactory	Visual
mean	1.518	0.325	1.074	1.034	0.391	2.900
std	0.992	0.699	0.932	0.881	0.621	0.902
25%	0.733	0.000	0.368	0.375	0.052	2.266
50%	1.388	0.117	0.769	0.785	0.187	2.941
75%	2.117	0.304	1.562	1.450	0.437	3.588

Table 1: A total of 38,872 words produced the following rating statistics for each of the six modalities.

	Auditory	Gustatory	Haptic	Interoceptive	Olfactory	Visual
ConceptNet + Linear Regression MAE	0.495	0.272	0.453	0.413	0.289	0.512
ConceptNet + Lasso Regression MAE	0.495	0.272	0.453	0.413	0.289	0.512
ConceptNet + Gradient Boosting MAE	0.494	0.226	0.448	0.409	0.257	0.512
ConceptNet + Neural Network MAE	**0.427**	**0.196**	**0.385**	**0.363**	**0.222**	**0.474**
BERT + Linear Regression MAE	0.601	0.330	0.538	0.490	0.323	0.574
BERT + Lasso Regression MAE	0.601	0.330	0.538	0.490	0.323	0.574
BERT + Gradient Boosting MAE	0.620	0.350	0.5528	0.499	0.342	0.580
BERT + Neural Network MAE	0.551	0.256	0.497	0.475	0.277	0.617

Table 2: Embedding + machine learning model and their mean absolute errors per modality

is a multi-output regression task. However we split up the whole task into six sub-tasks, creating six individual machine learning models that predict the strength values. We also split our data set into 31,097 training samples and 7,775 test samples (20%). We used four different regression models: **Linear Regression**, **Lasso Regression** (for additional variable selection and regularization) and **Gradient Boosting** with 350 estimators. Lastly, we used a **Neural Network** with one hidden layer. To counter overfitting we also included a dropout layers before and after the hidden layer. Furthermore, we used mean squared error (MSE) as loss function and Adam as optimizer. We trained the network until the validation error started to not decrease anymore for three consecutive epochs. The network has a single output node which represents the correct value that should have been predicted.

To evaluate the models, we used their fully trained state and predicted values for words out of the test set. We then compared the predicted value with the true value and calculated the error using the MSE and MAE.

3 Results

The best performing model is the neural network using the ConceptNet word embeddings with an aggregated MAE of 0.345 and MSE of 0.470 and therefore its prediction error is around 0.06 better compared to the other models. Additionally, the neural network outperforms the other models in every perceptual modality dimension. Worst performing dimension is the visual perception with a MAE of 0.474.

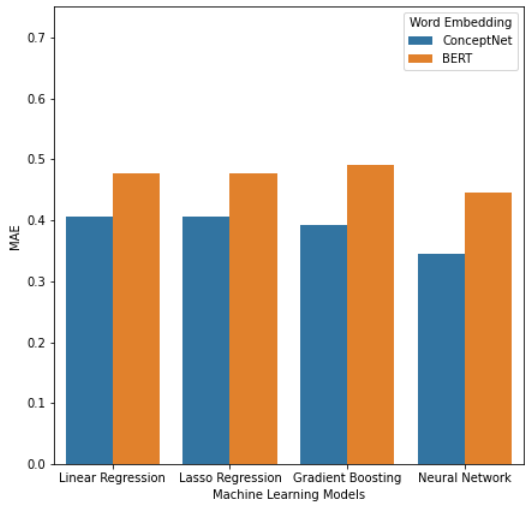

Figure 1: Aggregated MAE values for ConceptNet and BERT predicted using multiple ML models

4 Discussion

We used a small BERT model and did not fine-tune it on this task which is probably why ConceptNet outperformed it. Furthermore, BERT generally performs better on whole sentences, rather than on single words without contextual information. The overall performance of the predictions are quite decent, however the variance and standard deviation of the individual modalities is fairly low. Still one could argue, that the used word embeddings are a good enough representation of a semantic space that, combined with a learning algorithm, is able to decently predict sensorimotor perception strengths.

References

Louise Connell, Dermot Lynott, and Briony Banks. 2018. Interoception: the forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752):20170143.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dermot Lynott and Louise Connell. 2009. Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41(2):558–564.

Dermot Lynott and Louise Connell. 2013. Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior research methods*, 45(2):516–526.

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*, pages 1–21.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2016. Conceptnet 5.5: An open multilingual graph of general knowledge. *arXiv preprint arXiv:1612.03975*.