Volume 6, Issue 4, April 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Evaluation of Semantic Similarity Measures for Analysis of fMRI Data

Accamma I.V*
Research Scholar
VTU, Belgaum, India

Suma H.N.
Dept of Medical Electronics
BMSCE, Bangalore,

Dakshayini M.
Dept of Information Science
BMSCE, Bangalore, India

Abstract— How language is represented in the brain has been a topic of interest to many researchers. The advent of functional neuroimaging techniques has enabled us to see the brain at work. This paper compares the behavior of semantic similarity measures against the neural representation of words obtained by capturing the functional Magnetic Resonance imaging (fMRI) activations. It is an attempt to see which measure comes close to the representation used by the brain. The distance between words as given by the semantic similarity measures is compared against distance between neural representations of words. It can be seen that the Leacock measure consistently comes very close to the brain's representation of words.

Keywords— fMRI, Semantic similarity, procrustes distance,

I. INTRODUCTION

Two decades after the invention of fMRI; it has become the modality of choice for understanding the neural code underlying cognition [1]. Extensive research is underway to use fMRI activation profiles in decoding studies [2, 3]. The aim of these studies is to be able to predict the underlying mental state given the fMRI image. Many cognition based studies involve a task for the subject under the scanner that is based on language constructs. Decoding models for such tasks are based on semantic information. The datasets for such studies are collected by very controlled experiments. Some of these datasets are shared in the public domain. Though datasets are available, the number of observations vis-a-vis the number voxels is very low. This results in the curse of dimensionality problem. To improve the predictability of the algorithms it would help to have more datasets. Since each experiment deals with a unique scenario, it is difficult to combine the datasets. Using semantic similarity as a measure of commonality between the stimulus is an approach to combine the datasets.

In this paper we evaluate and analyze the various semantic similarity measures. The distance between words in the language is measured using semantic similarity. This distance is compared against the distance between neural representations of words. The neural representation is taken as the procrustes distance between voxel point clouds.

II. RELATED WORK

Neuroimaging and behavioural studies demonstrate that when humans understand a word, they activate the sensory-motor neurons with respect to perceiving and acting on the real-world objects [4]. Therefore a neural signature for the word is formed based on activated neurons. In studies involving predictions based on fMRI data, semantic information of stimulus words have been used for classification of the fMRI data. From using machine learning techniques to predict the semantic category of a viewed word [5] research has evolved to predicting the exact word [6]. This was possible by creating the semantic representation of nouns using the co-occurrence statistics of that noun with 25 verbs in the massive Google corpus. This representation coupled with fMRI activation profiles of the nouns were used by a machine learning algorithm to predict activations of untrained nouns. The fMRI activations of the same set of nouns were predicted based on features from definitional text [7] as well as WorldNet [8]. This framework has also been used to train classifiers and predict nouns across languages [9]. Factor analysis also has been used to arrive at the underlying factors common to multiple nouns [10]. Not just nouns, adjective-noun phrases were also predicted by creating a multiplicative model using the co-occurrence of adjectives and nouns with 5 sensory verbs [11].

III. SEMANTIC SIMILARITY

Semantic similarity measures the similarity between two concepts in the taxonomy of a language. Language semantics are mostly given by the nouns in the language. Therefore, semantic relations between nouns are considered in this paper. Nouns can be related in four different ways. The is-a (hyponym/hypernym) relation between nouns is the most common type of relation. The other types are part meronym/part holonym (part-of), member meronym/member holonym (member-of) and substance meronym/substance holonym (substance-of).

There are two types of measures for is-a relation. Path based measures use the length of the path connecting the nouns on the taxonomy. The edges are considered as a parameter. Information content based measures use the amount of information shared by the nouns. [12], [13] give a good review of WorldNet based semantic similarities.

This work aims to find a semantic similarity measure that mimics the similarity of neural signatures between words. 8 different measures are considered which are described below.

A. Based on path length

Path measure uses the is-a taxonomy to calculate the similarity of two words as the shortest path that connects them. The **Wu & Palmer (WuP)** measure calculates similarity by considering the depth of the two words in the taxonomy and that of their Least Common Subsumer (LCS)[14]. **Leacock** and Chodorow's measure takes the maximum depth of taxonomy and the shortest path into account for calculating the similarity [15].

B. Based on information content

Resnik uses the information content (IC) of the LCS[16]. It assumes that for two given concepts, similarity is depends on the IC that subsumes them in the taxonomy. **Lin** [17] and **Jiang** [18] use the sum of IC of individual concepts in addition to the IC of the LCS. The Lin measure uses the sum to scale the IC of LCS while Jiang takes the difference of the sum and the IC of the LCS.

C. Based on other types of semantic relatedness

The **Hirst** approach [19] suggests that two words are close if their lexical chain is a path that is not too long and does not change direction too often. **Lesk** [20] measures the extent of overlaps in the definition of the two words.

IV. DATA AND PREPROCESSING

The fMRI dataset considered is the well studied dataset of concrete noun from Carnegie Mellon University. This dataset consists of fMRI activation profiles obtained when 9 subjects were shown pictures of 60 different nouns. Each stimulus was presented 6 times. This data is available in a pre-processed format. The available data has been corrected for slice timing, motion, and linear trend. The data is spatially normalized into MNI space, resampled into 3x3x6 mm3 voxels and converted to Percent Signal Change format. A single fMRI mean image was created for each of the stimulus presentations by taking the mean of the images collected at 4s, 5s, 6s, and 7s from the time the stimulus was shown.

V. FEATURE SELECTION

After preprocessing, the size of the dataset is further brought down by removing the non-informative voxels. Not all voxels carry information that is pertaining to the stimulus. Most voxels are not active for all stimuli. The feature selection step selects and retains only those voxels that are pertinent.

The feature selection is carried out in the neural representational space. The stimulus is considered against a neutral stimulus (*celery* in case of noun dataset) and the voxels that help to discriminate best between these two stimuli are selected. This is accomplished by the use of feature selection algorithm based on mutual information. Therefore, the voxels that have maximum mutual information with the class label but minimum mutual information between them are chosen [21]. This step can reduce the voxels by a considerable number.

Figure 1 shows how the mutual information between the selected voxels and class labels changes with each new voxel selection. The first few voxels that are selected have high mutual information with the class label. It decreases as the voxel numbers increase and becomes almost constant at 100. This behaviour is seen for all subjects. Only 3 subjects are shown here to have clarity.

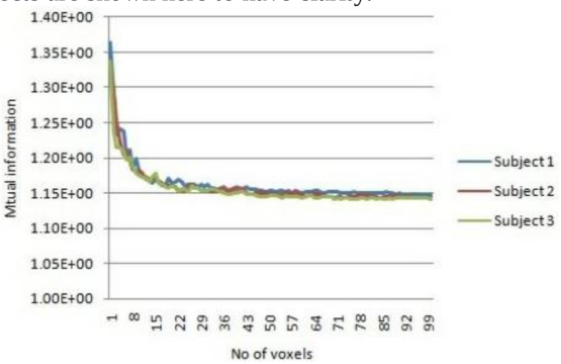

Fig. 1. The mutual information between the selected voxels and the class label

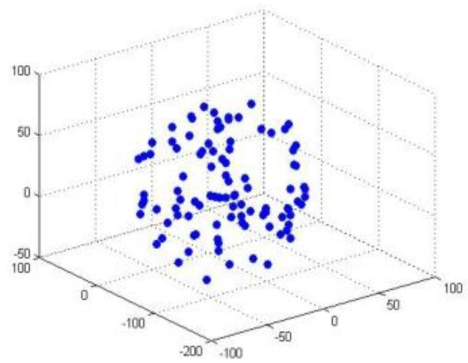

Fig. 2. The Point Cloud of neural representation of *Bottle* in Subject1

The 100 voxels that are chosen are then considered in their geometrical space. The x,y,z co-ordinates that characterize the position of the voxel will make up the cloud representing the stimulus. Figure 2 shows an example of a voxel cloud of neural representation of the stimulus *Bottle* in Subject 1.

VI. RESULTS AND DISCUSSION

12 words were chosen from the dataset each from the categories body parts, kitchen items and buildings. These categories were chosen because they are well studied [22], [23] in fMRI analysis. The distance between the words was

analysed. Each of the semantic similarity measure discussed above was applied using the Wordnet Similarity for Java (WS4J) [24].

The point clouds of voxels that are selected using feature selection are used as neural representations of the words. The procrustes distance between the point clouds is used as a distance measure.

Figure 3 shows the semantic distances and neural distances from a body part (*eye*) to other words.

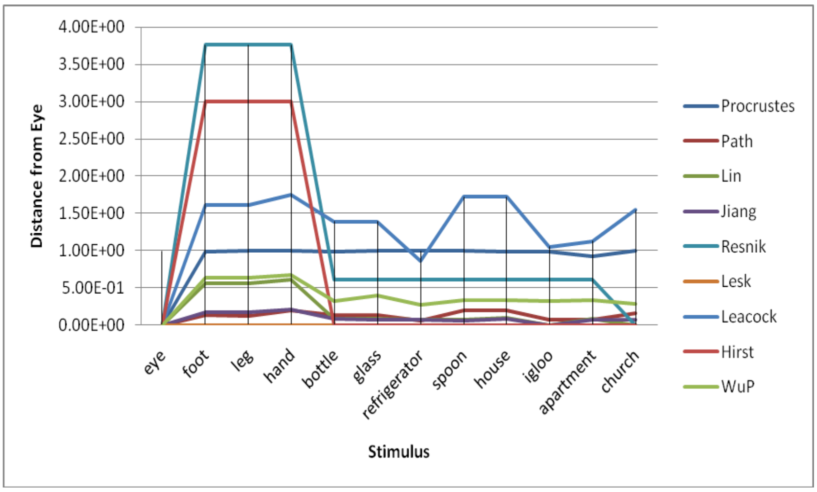

Fig. 3: The distance from eye to other words for Subject 1

Table 1 gives semantic similarity measures that are significantly correlated (P<0.05) with procrustes distance. The Path based measures WuP, Path and Leacock show high correlation with Leacock correlated highest by 0.77. Though the table shows data only for first three subjects, it was found that all the subjects exhibited similar behavior.

Table I Correlated Measures (P¡0.05) To Body Parts

Subject	Stimulus	Path	Lin	Jiang	Resnik	Lesk	Leacock	Hirst	WuP
1	Eye	✓					✓		✓
	Foot						✓		
	Leg	✓					✓		
	Hand						✓		
2	Eye	✓					✓		✓
	Foot						✓		
	Leg	✓					✓		
	Hand						✓		
3	Eye	✓					✓		✓
	Foot						✓		
	Leg	✓					✓		
	Hand						✓		

Figure 4 shows the semantic distances and procrustes distances from a kitchen item (*bottle)* to other words.

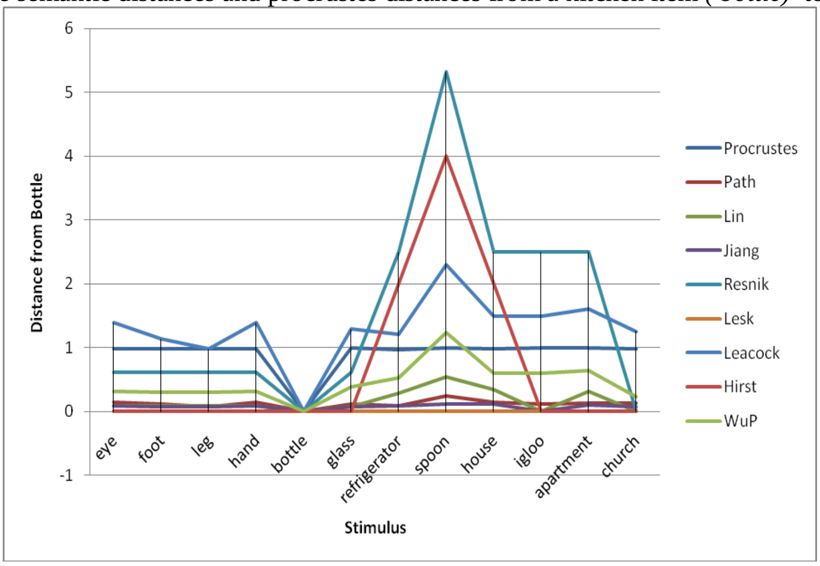

Fig. 4. Subject1: The distance from *bottle* to other words for Subject 1

Table 2 gives measures that are significantly correlated (P<0.05) with procrustes distance. Path, Jiang, Lesk, Leacock and WuP are correlated measures with Lesk having the highest correlation of .99. Though the table shows data for three subjects, it was found that all the subjects exhibited similar behaviour.

Table 2 Correlated Measures (P<0.05) To Kitchen Items

Subject	Stimulus	Path	Lin	Jiang	Resnik	Lesk	Leacock	Hirst	WuP
4	**Bottle**	✓		✓		✓	✓		
	Glass	✓		✓		✓	✓		✓
	Refrigerator	✓		✓		✓	✓		✓
	Spoon	✓		✓		✓	✓		
5	**Bottle**	✓		✓		✓	✓		
	Glass	✓		✓		✓	✓		✓
	Refrigerator	✓		✓		✓	✓		✓
	Spoon	✓		✓		✓	✓		
6	**Bottle**	✓		✓		✓	✓		
	Glass	✓		✓		✓	✓		✓
	Refrigerator	✓		✓		✓	✓		✓
	Spoon	✓		✓		✓	✓		

Figure 5 shows the semantic distances and procrustes distance from a building (house) to other words.

Table 3 gives measures that show significantly correlation (P<0.05) with procrustes distance for building stimulus. Again Path based measures are most correlated. Lin and Jiang also show correlation for the stimulus Igloo. Again all subjects exhibited similar behavior.

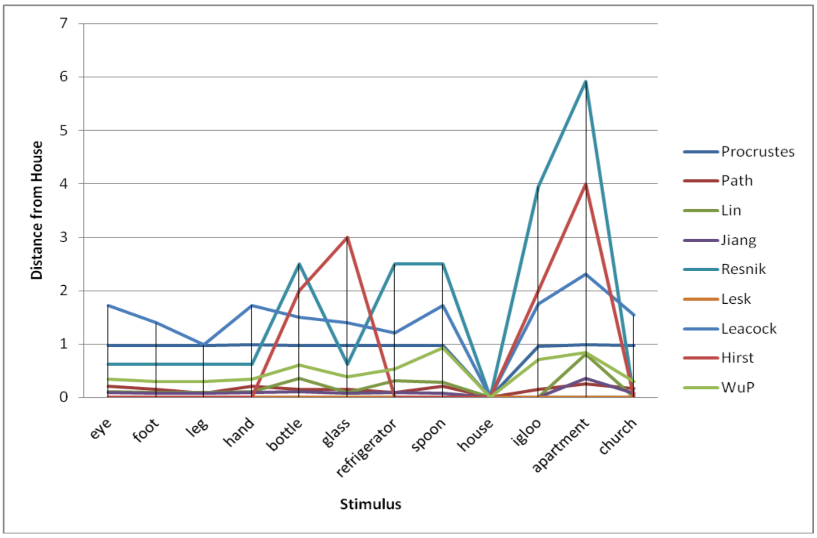

Fig. 5. The distance from *house* to other words in Subject1

Table 3 Correlated Measures (P<0.05) To Buildings

Subject	Stimulus	Path	Lin	Jiang	Resnik	Lesk	Leacock	Hirst	WuP
7	**House**	✓					✓		✓
	Igloo	✓	✓	✓			✓		
	Apartment	✓					✓		✓
	Church	✓					✓		✓
8	**House**	✓					✓		✓
	Igloo	✓	✓	✓			✓		
	Apartment	✓					✓		✓
	Church	✓					✓		✓
9	**House**	✓					✓		✓
	Igloo	✓	✓	✓			✓		
	Apartment	✓					✓		✓
	Church	✓					✓		✓

For exemplars from all the three categories it was seen that semantic distances between words calculated based on path lengths come close to the distances between voxel clouds representing those words in the brain.

VII. CONCLUSION

This paper compared the semantic distances between nouns to the neural distances between the fMRI profiles of those nouns. 8 different semantic similarity measures where compared against neural distances measured using procrustes distance. It was observed that in general Path based measures correlate highly with procrustes distance. Leacock distance consistently showed high correlation for all the stimuli in all the subjects. Therefore, these distances could be used in further studies involving semantic relations.

ACKNOWLEDGMENT

The work reported in this paper is supported by the college through the TECHNICAL EDUCATION QUALITY IMPROVEMENT PROGRAMME [TEQIP-II] of the MHRD, Government of India

REFERENCES

[1] Cabeza, R. and Nyberg, L., 2000. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of cognitive neuroscience*, *12*(1), pp.1-47.

[2] Norman, K.A., Polyn, S.M., Detre, G.J. and Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in cognitive sciences*, *10*(9), pp.424-430.

[3] Haynes, J.D. and Rees, G., 2006. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), pp.523-534.

[4] Pecher, D. and Zwaan, R.A. eds., 2005. *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge University Press.

[5] Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M. and Newman, S., 2004. Learning to decode cognitive states from brain images. *Machine Learning*, *57*(1-2), pp.145-175..

[6] Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A. and Just, M.A., 2008. Predicting human brain activity associated with the meanings of nouns. *science*, *320*(5880), pp.1191-1195.

[7] Pereira, F., Botvinick, M. and Detre, G., 2010, June. Learning semantic features for fMRI data from definitional text. In *Proceedings of the NAACL HLT 2010 First Workshop on Computational Neurolinguistics* (pp. 1-9). Association for Computational Linguistics.

[8] Jelodar, A.B., Alizadeh, M. and Khadivi, S., 2010, June. WordNet Based Features for Predicting Brain Activity associated with meanings of nouns. In*Proceedings of the NAACL HLT 2010 First Workshop on Computational Neurolinguistics* (pp. 18-26). Association for Computational Linguistics.

[9] Buchweitz, A., Shinkareva, S.V., Mason, R.A., Mitchell, T.M. and Just, M.A., 2012. Identifying bilingual semantic neural representations across languages. *Brain and language*, *120*(3), pp.282-289..

[10] Just, M.A., Cherkassky, V.L., Aryal, S. and Mitchell, T.M., 2010. A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PloS one*, *5*(1), p.e8622.

[11] Chang, K.M.K., Cherkassky, V.L., Mitchell, T.M. and Just, M.A., 2009, August. Quantitative modeling of the neural representation of adjective-noun phrases to account for fMRI activation. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2* (pp. 638-646). Association for Computational Linguistics.

[12] Meng, L., Huang, R. and Gu, J., 2013. A review of semantic similarity measures in wordnet. *International Journal of Hybrid Information Technology*,*6*(1), pp.1-12.

[13] Gomaa, W.H. and Fahmy, A.A., 2013. A survey of text similarity approaches. *International Journal of Computer Applications*, *68*(13).

[14] Wu, Z. and Palmer, M., 1994, June. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics* (pp. 133-138). Association for Computational Linguistics.

[15] Leacock, C. and Chodorow, M., 1998. Combining local context and WordNet similarity for word sense identification. *WordNet: An electronic lexical database*, *49*(2), pp.265-283.

[16] Resnik, P., 1995. Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007*.

[17] Lin, D., 1998, July. An information-theoretic definition of similarity. In *ICML*(Vol. 98, pp. 296-304).

[18] Jiang, J.J. and Conrath, D.W., 1997. Semantic similarity based on corpus statistics and lexical taxonomy. *arXiv preprint cmp-lg/9709008*.

[19] Hirst, G. and St-Onge, D., 1998. Lexical chains as representations of context for the detection and correction of malapropisms. *WordNet: An electronic lexical database*, *305*, pp.305-332.

[20] Banerjee, S. and Pedersen, T., 2002. An adapted Lesk algorithm for word sense disambiguation using WordNet. In *Computational linguistics and intelligent text processing* (pp. 136-145). Springer Berlin Heidelberg.

[21] Peng, H., Long, F. and Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy.*Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *27*(8), pp.1226-1238.

[22] Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L. and Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), pp.2425-2430.

[23] Shinkareva, S.V., Mason, R.A., Malave, V.L., Wang, W., Mitchell, T.M. and Just, M.A., 2008. Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One*, *3*(1), p.e1394.

[24] Shima, H., 2014. WS4J WordNet Similarity for Java.

