Implementation of instance-based distributional semantic model in predicting dominancy of homonyms

Soo Hyun Ryu
University at Buffalo
Department of Linguistics
soohyunr@buffalo.edu

Brendan T. Johns
University at Buffalo
Department of Communicative Disorders and Sciences
btjohns@buffalo.edu

Abstract

The purpose of the present paper is to investigate the functionality of linguistic instance-based Distributional Semantic Model, referred to as the Instance Theory of Semantics (ITS). Specifically, this paper examines whether ITS could preserve sophisticated semantic information inherent to homonyms so that it becomes able to discern the dominancy of multiple meanings attached to homonyms. We suspect that the accuracy of ITS in determining dominancy of homonyms would increase if the model is trained with more linguistic instances since instance-based DSM has the advantage of storing diverse contexts attached to homonyms. For this reason, we examine whether the frequencies of homonyms during the model training would affect the performance in determining dominant meanings of homonyms. Our result shows that as more linguistic instances ITS experience in the encoding process, the better it becomes in determining the dominancy of meanings of homonyms. This provides a support that ITS model has a capacity to accumulate diverse senses pertaining to homonyms while other DSMs merge multiple senses into a single prototypical vector.

1 Introduction

Distributional Semantic Models (DSMs) learn the meaning of words from statistical redundancies contained in text. The Instance Theory of Semantics (ITS) is a new type of DSM which uses individual instances of language as the basis for semantic processing (Jamieson et al., 2018). While other DSMs are constructed to represent prototypes of words, ITS enables words to be represented on the basis of multiple language instance traces using the combination of BEAGLE model (Jones and Mewhort, 2007) and MINERVA2 framework (Hintzman, 1984,1986,1988). Based on the ability to preserve abundant linguistic experiences pertaining to each word in the model, ITS successfully simulates human cognitive behavior in language processing, such as word sense disambiguation (Jamieson et al., 2018). This result implies that ITS could store more delicate semantics of words on the basis of episodic memories of linguistic experience, showing potential to complement mental lexicon models. To support the functionality of ITS for the richer semantic representation, this paper aims to show that as the model is exposed to more linguistic instances, it can develop more sophisticated understanding of word senses. For this purpose, we investigate how the number of occurrences of homonyms in the encoding process affects the performance of ITS model in determining which meaning of the homonym is dominant. We hypothesize that the more instances the model is exposed to, the more accurately the model could determine the dominancy of homonyms. The result implies that as the frequency of homonyms increases in the encoding process, ITS can better determine dominant meanings of homonyms. This suggests that ITS model, which stores individual linguistic instances, could successfully preserve diverse meanings inherent to homonyms and discern the extent to which each meaning of word is dominant when it is trained with enough number of linguistic instances.

2 ITS Model

ITS model(Jamieson et al., 2018) enables words to be represented with expectation vectors on the basis of multiple language instance traces, using the combination of BEAGLE model (Jones and Mewhort, 2007) and MINERVA 2 framework (Hintzman, 1984,1986,1988), the former of which is an encoding scheme and the latter of which involves retrieval operations. To say, ITS model consists of two main parts: encoding of linguistic instances and retrieving semantic information from the memory.

2.1 Encoding - BEAGLE Model

In the process of encoding, each linguistic instance, a single episodic memory, is represented with a single vector, being accumulated in the memory. ITS model implements BEAGLE model (Jones and Mewhort, 2007) as the encoding process. In this model, every word is represented with a n dimensional vector *(environmental vector)*, which consists of random values taken from a normal distribution with a mean of zero and a variance of *1/n*. The memory of linguistic instances is also represented with a single n dimensional vector with summing the word vectors appearing in the context by the following equation:

$$c_i = \sum_{j=1}^{j=h} w_{ij} \tag{1}$$

where c_i is the i^{th} linguistic instance in memory, h is the number of words in the instance, and w_{ij} corresponds to the j^{th} word in the i^{th} context. After the encoding process, memory matrix ends up with having the size of *m x n*, where m is the number of linguistic instances and n is the number of dimension of vectors . While in the original BEAGLE model introduces the notion of *'forgetting'* to assume data loss by deleting some features in the memory, the present study does not suppose any data loss.

2.2 Retrieval - MINERVA 2

2.2.1 Retrieval from single words

From the memory constructed on the basis of linguistic instances, ITS extracts expected semantic information when a word is provided as a probe. Specifically, the provided word activates each instance trace in the memory by calculating the similarities between the word and the trace, which is referred to as *'activation'*, as below:

$$a_i = \left(\frac{\sum_{j=1}^{j=n} p_j \cdot M_{ij}}{\sqrt{\sum_{j=1}^{j=n} p_j^2}\sqrt{\sum_{j=1}^{j=n} M_{ij}^2}} \right)^3 \tag{2}$$

where a_i is the activation for i^{th} instance in the memory given a probe p, p_j is j^{th} value of the environmental vector of the probe, and M_{ij} is the j^{th} value in the i^{th} instance in the memory.

The values of activation are then multiplied by the corresponding memory traces, the sum of which turns into be the *'expectation vector'* of the probe. The equation for expectation vector is provided below:

$$E(W) = \sum_{i=1}^{n} (a_i > 0) a_i \times M_{\mathrm{i}} \tag{3}$$

In order to retrieve instances whose semantics is similar to the probe, we only use traces that have positive similarity to the probe in the process of getting expectation vectors (Johns and Jones, 2015).

2.2.2 Retrieval from a set of multiple words

To extract semantic information from a set of multiple words, *'Joint probe'* needs to be implemented (Jamieson et al., 2018). The joint probe of the word set calculates the activation by taking the product of the activations of all the words in the set by the following equation:

$$a_i = \prod_{k=1}^{k=h} \left(\frac{\sum_{j=1}^{j=n} p_j \cdot M_{ij}}{\sqrt{\sum_{j=1}^{j=n} p_j^2}\sqrt{\sum_{j=1}^{j=n} M_{ij}^2}} \right)^3 \tag{4}$$

where h is the number of words in the set.

The equation to obtain the expectation vector for joint probes is same as (3).

3 Determining dominant meanings of homonyms

3.1 Data

To construct memory, Touchstone Applied Science Associates (TASA) corpus (Landauer and Dumais, 1997) is used. Each of 37,636 paragraphs in the corpus was considered as a unit of linguistic instance.

In order to test the performance of the trained model in determining dominant meanings of homonyms, norms from Armstrong et al. (2012) are used as target data. In Armstrong et al. (2012) 's data, the human-rated relative frequencies for meanings of homonyms, definitions of which are derived from wordsmyth.net (Parks et al., 1998), are provided. For example, a homonym *'bank'* in the norms has two meanings, the first of which is related to *'river'* and the second of which is related to *'money'*, having relative frequencies of 22 and 78, respectively. Out of 554 homonyms provided in Armstrong et al. (2012)' norms, we excluded homonyms with more than two meanings for the sake of simplicity. We also excluded homonyms if either of their definitions consist of zero or one content words since it is implausible that the definitions represent their semantics with zero or one content word without bias. As a result, 387 homonyms from Armstrong et al. (2012) norms were included to test the performance of ITS model in determining dominancy of homonyms.

3.2 Dominant Meaning Prediction

3.2.1 Procedure

We supposed that when a meaning is dominant over the other in a homonymous word, the expectation vector obtained from the joint probe of the definition of dominant meaning would show bigger similarity to the expectation vector of the target word.

To predict which meaning of homonyms is dominant using ITS model, we firstly extracted expectation vectors of target words and two definitions pertaining to them. Then we calculated the similarities between the expectation vector of the target word and the ones of definitions. Lastly, we compared the similarities to determine which definition would be dominant.

3.2.2 Analysis and Results

With all homonyms taken into account regardless of their frequencies, ITS shows chance level accuracy in determining dominant meanings of homonyms. Out of 387 homonyms with two discrete definitions, ITS succeeds in selecting dominant meaning for 211 homonyms, which is accuracy of 54%. As the frequencies of homonyms increase, however, the ability of ITS model in predicting the dominant meaning significantly improves. While the words with frequency less than 300 and the ones with frequency greater than 300 and less than 600 both obtain 53% of accuracy, the homonyms with frequency greater than 600 and less than 900 shows the accuracy of 59%. Even more, when homonyms with frequency greater than 900 are considered, the accuracy is raised to 71%.

Frequency	Num of words	Accuracy
$0 \sim 300$	296	53%
$300 \sim 600$	45	53%
$600 \sim 900$	23	59%
$900 \sim$	24	71%

Table 1: Accuracy in determining dominant meaning depending on the frequencies of homonyms

3.3 Difference in dominancy between human-rated data and ITS prediction

3.3.1 Procedure

To examine the performance of ITS in determining dominant meanings of homonyms more closely, we also investigated how error terms of ITS prediction correlate to the frequencies of homonyms. We calculated error terms by taking squared differences between the dominancy calculated from human rated-frequencies and the ones calculated based on ITS model.
As Armstrong et al. (2012) calculated the dominancy as *(biggest relative frequency - second biggetst relative frequency) / (biggest relative frequency)*, the dominance of ITS prediction was calculated by *(biggest similarity - second biggest similarity) / (biggest similarity)*. In the cases where ITS fails to predict dominant meaning correctly, dominance from ITS model was converted into negative value.

3.3.2 Analysis and Results

refer to the previous paper Results of the Pearson correlation indicate that there is a significant negative association between error terms of ITS prediction and frequencies of homonyms in the training corpus, ($r(385)$=-.11, $p < 0.05$).

4 Discussion

In this paper, we provide a corroborating evidence for the functionality of Instance Theory of Semantics(ITS), which is an instance-based distributional semantic model, in preserving diverse meanings associated with a single word. This study suggests that the more numbers of linguistic experiences ITS has, the better performance it shows in predicting dominant meaning of homonyms. This result is noteworthy in two respects.

First, the result suggests that the ITS which preserves diverse linguistic instances related to each word shows a possibility of building a more sophisticated models of mental lexicon. Since the superiority of ITS is based on the accumulated linguistic experiences, more experiences it is exposed to would lead to richer and more delicate representation of each homonyms. With accumulating every single instance rather than putting accumulated experiences together as a prototypical vector, ITS model would outperform previous DSMs in representing human mental lexicon, which is constructed based on diverse instances (Jamieson et al., 2018). Moreover, the result of this study showing that more linguistic instances can lead to increasing improvement of ITS in predicting dominancy of homonyms suggests a possibility that ITS could describe how human beings develop their understanding of word senses as their linguistic experience accumulate increasingly.

Secondly, the functionality to determine the dominant meaning of homonyms and to predict the degree of dominancy in homonyms would have a applicability in a variety of Natural Language Processing (NLP) tasks. Without doubt, Word Sense Disambiguation (WSD) has been one of the long-standing challenges in the field of NLP. In order to tackle the WSD issue, the dominancy of ambiguous words has been considered as an important factor for the lexical disambiguation. Given the result of this study suggesting that ITS can improve its ability to predict dominancy of homonyms with enough number of linguistic experiences encoded, ITS trained with a larger number of linguistic instances would contribute to enhance performance in WSD tasks.

References

Blair C Armstrong, Natasha Tokowicz, and David C Plaut. 2012. edom: Norming software and relative meaning frequencies for 544 english homonyms. *Behavior research methods*, 44(4):1015–1027.

Douglas L Hintzman. 1984. Minerva 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16(2):96–101.

Douglas L Hintzman. 1986. “schema abstraction” in a multiple-trace memory model. *Psychological review*, 93(4):411.

Douglas L Hintzman. 1988. Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological review*, 95(4):528.

Randall K Jamieson, Johnathan E Avery, Brendan T Johns, and Michael N Jones. 2018. An instance theory of semantic memory. *Computational Brain & Behavior*, 1(2):119–136.

Brendan T Johns and Michael N Jones. 2015. Generating structure from experience: A retrieval-based model of language processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 69(3):233.

Michael N Jones and Douglas JK Mewhort. 2007. Representing word meaning and order information in a composite holographic lexicon. *Psychological review*, 114(1):1.

Thomas K Landauer and Susan T Dumais. 1997. A solution to plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211.

R Parks, J Ray, and S Bland. 1998. Wordsmyth english dictionary–thesaurus. university of chicago.