Learning free word associations from texts

Manfred Wettler (wettler@psycho.upb.de) & Petra Seidensticker (karas@psycho.upb.de)
University Paderborn; 33095 Paderborn Germany

Introduction

On the basis of the statistical distribution of words in texts Burgess (1998), and others have computed semantic representations producing humanlike performance in tasks that are considered to presuppose language understanding. These representations are calculated in three steps: First the co-occurrences of words are counted. A co-occurrence is defined as the joint appearance of two words or within a maximum distance from each other. In the second step for each word a context vector is computed whose elements are the normalized co-occurrences with all other words of the vocabulary. In a third step similarities between words are computed as the dot products of their context vectors. These models make no assertions about how the postulated semantic representations are learned. We have developed and implemented an unsupervised learning algorithm which is based on the stimulus sampling theory of Estes (1950). It computes incrementally, while reading, a net of associations between words that correspond to the normalized co-occurrence values. These associations agree with the results observed in the free word association experiment.

Method

According to stimulus sampling theory, the associative strength between two events i and j increases by a constant fraction of the maximally possible increment whenever these two events co-occur. If associative strengths range between 0 and 1 we have:

$$a_{i,j}(t+1) = (1-\Theta) \cdot a_{i,j}(t) + \Theta \qquad (1)$$

where $a_{i,j}(t)$ denotes the associative strength before and $a_{i,j(t+1)}$ after time t. Θ is the learning rate. If i does not occur in the context of j, then the strength of the association between i and j diminishes by a constant fraction of the existing associative strength.

$$a_{i,j}(t+1) = (1-\Theta) \cdot a_{i,j}(t) \qquad (2)$$

Learning of word associations has been simulated with the so-called window technique. In this technique a window with a constant number of words is moved from left to right across a text. At each step the word at the left side of the window is covered and the next word at the right side is uncovered. The associations between all pairs of words which are inside the window are then strengthened by application of equation 1 and all associations between the words in the window and words which are not in the window are weakened by application of equation 2. We have used texts of a German newspaper with 34 million words as learning material. All inflected words had been replaced by their root forms. The associative matrix describes the relations between 814 words. These are the 100 stimulus words and all responses which were given by at lest 5 of the 331 participants in the German association norms from Russel (1970). The following results have been obtained with a Θ of $2.1 \cdot 10^{-5}$ and a window size of 20.

Results

We compared the computed associations with the results of the free association study by Russel (1970). For 29 of the 100 stimulus words the word with the strongest computed association to the stimulus is equal to the primary response, i.e. the response most frequently given by the participants in the Russel study. In comparison, the participants in the study of Russel (1970) produced on average 22.5 primary responses. For 10 stimulus words the observed primary response is equal to the second-strongest computed association to the stimulus word. For 55 of the 100 stimulus words the primary response belongs to the five strongest computed associations to the stimulus.

Our results show that contiguity learning might be an important factor for the build-up of semantic representations.

Acknowledgments

The work was supported by the Deutsche Forschungsgemeinschaft and by the Heinz-Nixdorf-Institute.

References

Burgess, C. (1998). From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model. *Behavioral Research Methods, Instruments, and Computers, 30,*(pp. 188-198).

Estes, W.K. (1950). Toward a statistical theory of Learning. *Pychological Review ,57,* (pp. 94-107)

Russel, W.A. (1971). The complete German language norms for responses to 100 words from the Kent-Rosanoff word association test. In L. Postman & G. Keppel (Eds.),*Norms of word association.* New York: Academic Press.