

Word Vectors and Two Kinds of Similarity

Akira Utsumi and Daisuke Suzuki

Department of Systems Engineering
The University of Electro-Communications
1-5-1 Chofugaoka, Chofushi, Tokyo 182-8585, Japan
utsumi@se.uec.ac.jp, dajie@utm.se.uec.ac.jp

Abstract

This paper examines what kind of similarity between words can be represented by what kind of word vectors in the vector space model. Through two experiments, three methods for constructing word vectors, i.e., LSA-based, cooccurrence-based and dictionary-based methods, were compared in terms of the ability to represent two kinds of similarity, i.e., taxonomic similarity and associative similarity. The result of the comparison was that the dictionary-based word vectors better reflect taxonomic similarity, while the LSA-based and the cooccurrence-based word vectors better reflect associative similarity.

1 Introduction

Recently, geometric models have been used to represent words and their meanings, and proven to be highly useful both for many NLP applications associated with semantic processing (Widdows, 2004) and for human modeling in cognitive science (Gärdenfors, 2000; Landauer and Dumais, 1997). There are also good reasons for studying geometric models in the field of computational linguistics. First, geometric models are cost-effective in that it takes much less time and less effort to construct large-scale geometric representation of word meanings than it would take to construct dictionaries or thesauri. Second, they can represent the implicit knowledge of word meanings that dictionaries and thesauri cannot do. Finally, geometric representation is easy to revise and extend.

A vector space model is the most commonly used geometric model for the meanings of words. The basic idea of a vector space model is that words are represented by high-dimensional vectors, i.e., *word vectors*, and the degree of semantic similarity between any two words can be easily

computed as a cosine of the angle formed by their vectors.

A number of methods have been proposed for constructing word vectors. Latent semantic analysis (LSA) is the most well-known method that uses the frequency of words in a fraction of documents to assess the coordinates of word vectors and singular value decomposition (SVD) to reduce the dimension. LSA was originally put forward as a document indexing technique for automatic information retrieval (Deerwester et al., 1990), but several studies (Landauer and Dumais, 1997) have shown that LSA successfully mimics many human behaviors associated with semantic processing. Other methods use a variety of other information: cooccurrence of two words (Burgess, 1998; Schütze, 1998), occurrence of a word in the sense definitions of a dictionary (Kasahara et al., 1997; Niwa and Nitta, 1994) or word association norms (Steyvers et al., 2004).

However, despite the fact that there are different kinds of similarity between words, or different relations underlying word similarity such as a synonymous relation and an associative relation, no studies have ever examined the relationship between methods for constructing word vectors and the type of similarity involved in word vectors in a systematic way. Some studies on word vectors have compared the performance among different methods on some specific tasks such as semantic disambiguation (Niwa and Nitta, 1994) and cued/free recall (Steyvers et al., 2004), but it is not at all clear whether there are essential differences in the quality of similarity among word vectors constructed by different methods, and if so, what kind of similarity is involved in what kind of word vectors. Even in the field of cognitive psychology, although geometric models of similarity such as multidimensional scaling have long been studied and debated (Nosofsky, 1992), the possibility that different methods for word vectors may cap-

ture different kinds of word similarity has never been addressed.

This study, therefore, aims to examine the relationship between the methods for constructing word vectors and the type of similarity in a systematic way. Especially this study addresses three methods, *LSA-based*, *cooccurrence-based*, and *dictionary-based methods*, and two kinds of similarity, *taxonomic similarity* and *associative similarity*. Word vectors constructed by these methods are compared in the performance of two tasks, i.e., multiple-choice synonym test and word association, which measure the degree to which they reflect these two kinds of similarity.

2 Two Kinds of Similarity

In this study, we divide word similarity into two categories: taxonomic similarity and associative similarity. Taxonomic similarity, or categorical similarity, is a kind of semantic similarity between words in the same level of categories or clusters of the thesaurus, in particular synonyms, antonyms, and other coordinates. Associative similarity, on the other hand, is a similarity between words that are associated with each other by virtue of semantic relations other than taxonomic one such as a collocational relation and a proximity relation. For example, the word *writer* and the word *author* are taxonomically similar because they are synonyms, while the word *writer* and the word *book* are associatively similar because they are associated by virtue of an agent-subject relation.

This dichotomy of similarity is practically important. Some tasks such as automatic thesaurus updating and paraphrasing need assessing taxonomic similarity, while some other tasks such as affective Web search and semantic disambiguation require assessing associative similarity rather than taxonomic similarity. This dichotomy is also psychologically motivated. Many empirical studies on word searches and speech disorders have revealed that words in the mind (i.e., mental lexicon) are organized by these two kinds of similarity (Aitchison, 2003). This dichotomy is also essential to some cognitive processes. For example, metaphors are perceived as being more apt when their constituent words are associatively more similar but categorically dissimilar (Utsumi et al., 1998). These psychological findings suggest that people distinguish between these two kinds of similarity in certain cognitive processes.

3 Constructing Word Vectors

3.1 Overview

In this study, word vectors (or word spaces) are constructed in the following way. First, all content words t_i in a corpus are represented as m -dimensional feature vectors w_i .

$$w_i = (w_{i1}, w_{i2}, \dots, w_{im}) \quad (1)$$

Each element w_{ij} is determined by statistical analysis of the corpus, whose methods will be described in Section 3.3. A matrix M is then constructed using n feature vectors as rows.

$$M = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \quad (2)$$

Finally, the dimension of row vectors w_i is reduced from m to k by means of a SVD technique. As a result, any words are represented as k -dimensional vectors.

3.2 Corpus

In this study, we employ three kinds of Japanese corpora: newspaper articles, novels and a dictionary. As a newspaper corpus, we use 4 months' worth of Mainichi newspaper articles published in 1999. They consist of 500,182 sentences in 251,287 paragraphs, and words vectors are constructed for 53,512 words that occur three times or more in these articles. Concerning a corpus of novels, we use a collection of 100 Japanese novels "Shincho Bunko No 100 Satsu" consisting of 475,782 sentences and 230,392 paragraphs. Word vectors are constructed for 46,666 words that occur at least three times. As a Japanese dictionary, we use "Super Nihongo Daijiten" published by Gakken, from which 89,007 words are extracted for word vectors.

3.3 Methods for Computing Vector Elements LSA-based method (LSA)

In the LSA-based method, a vector element w_{ij} is assessed as a tf-idf score of a word t_i in a piece s_j of document.

$$w_{ij} = tf_{ij} \times \left(\log \frac{m}{df_i} + 1 \right) \quad (3)$$

In this formula, tf_{ij} denotes the number of times the word t_i occurs in a piece of text s_j , and df_i denotes the number of pieces in which the word t_i occurs. As a unit of text piece s_j , we consider

a sentence and a paragraph. Hence, for example, when a sentence is used as a unit, the dimension of feature vectors w_i is equal to the number of sentences in a corpus. We also use two corpora, i.e., newspapers and novels, and thus we obtain four different word spaces by the LSA-based method.

Cooccurrence-based method (COO)

In the cooccurrence-based method, a vector element w_{ij} is assessed as the number of times words t_i and t_j occur in the same piece of text, and thus M is an $n \times n$ symmetric matrix. As in the case of the LSA-based method, we use two units of text piece (i.e., a sentence or a paragraph) and two corpora (i.e., newspapers or novels), thus resulting in four different word spaces.

Note that this method is similar to Schütze’s (1998) method for constructing a semantic space in that both are based on the word cooccurrence, not on the word frequency. However they are different in that Schütze’s method uses the cooccurrence with frequent content words chosen as indicators of primitive meanings. Burgess’s (1998) “Hyperspace Analogue to Language (HAL)” is also based on the word cooccurrence but does not use any technique of dimensionality reduction.

Dictionary-based method (DIC)

In the dictionary-based method, a vector element w_{ij} is assessed by the following formula:

$$w_{ij} = \left(f_{ij} + \alpha \sqrt{\sum_k f_{ik} f_{kj}} + \beta f_{ji} \right) \times \log \frac{n}{df_j} \quad (4)$$

where f_{ij} denotes the number of times the word t_j occurs in the sense definitions of the word t_i , and df_j denotes the number of words whose sense definitions contain the word t_j . The second term in parentheses in Equation (4) means the square root of the number of times the word t_j occurs in a collection of sense definitions for any words that are included in the sense definitions of the word t_i , while the third term means the number of times t_i occurs in the sense definitions of t_j . The parameters α and β are positive real constants expressing the weights for these information. (Following Kasahara et al. (1997), these parameters are set to 0.2 in this paper.)

Equation (4) was originally put forward by Kasahara et al. (1997), but our dictionary-based method differs from their method in terms of how the dimensions are reduced. Their method groups

together the dimensions for words in the same category of a thesaurus, but our method uses SVD as we will described next.

3.4 Reducing Dimensions

Using a SVD technique, a matrix M is factorized as the product of three matrices $U\Sigma V^T$, where the diagonal matrix Σ consists of r singular values that are arranged in nonincreasing order such that r is the rank of M . When we use a $k \times k$ matrix Σ_k consisting of the largest k singular values, the matrix M is approximated by $U_k \Sigma_k V_k^T$, where the i -th row of U_k corresponds to a k -dimensional “reduced word vector” for the word t_i .

4 Experiment 1: Synonym Judgment

4.1 Method

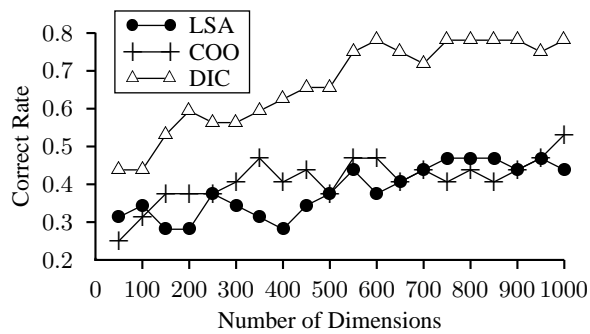
In order to compare different word vectors in terms of the ability to judge taxonomic similarity between words, we conducted a synonym judgment experiment using a standard multiple-choice synonym test. Each item of a synonym test consisted of a stem word and five alternative words from which the test-taker was asked to choose one with the most similar meaning to the stem word.

In the experiment, we used 32 items from the synonym portions of Synthetic Personality Inventory (SPI) test, which has been widely used for employment selection in Japanese companies. These items were selected so that all the vector spaces could contain the stem word and at least four of the five alternative words. For comparison purpose, we also used 38 antonym test items chosen from the same SPI test. Furthermore, in order to obtain a more reliable, unbiased result, we automatically constructed 200 test items in such a way that we chose the stem word randomly, one correct alternative word randomly from words in the same deepest category of a Japanese thesaurus as the stem word, and other four alternatives from words in other categories. As a Japanese thesaurus, we used “Goi-Taikai” (Ikehara et al., 1999).

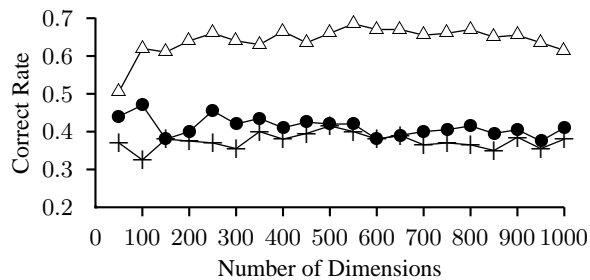
In the computer simulation, the computer’s choices were determined by computing cosine similarity between the stem word and each of the five alternative words using the vector spaces and choosing the word with the highest similarity.

4.2 Results and Discussion

For each of the nine vector spaces, the synonym judgment simulation described above was con-



(a) SPI test items



(b) Computer-generated test items

Figure 1: Correct rates of synonym tests

ducted and the percentage of correct choices was calculated. This process was repeated using 20 numbers of dimensions, i.e., every 50 dimensions between 50 and 1000.

Figure 1 shows the percentage of correct choices for the three methods of matrix construction. Concerning the LSA-based method (denoted by LSA) and the cooccurrence-based method (denoted by COO), Figure 1 plots the correct rates for the word vectors derived from the paragraphs of the newspaper corpus. (Such combination of corpus and text unit was optimal among all combinations, which will be justified later in this section.) The most important result shown in Figure 1 is that, regardless of the number of dimensions, the dictionary-based word vectors outperformed the other kinds of vectors on both SPI and computer-generated test items. This result thus suggests that *the dictionary-based vector space reflects taxonomic similarity between words better than the LSA-based and the correlation-based spaces.*

Another interesting finding is that there was no clear peak in the graphs of Figure 1. For SPI test items, correct rates of the three methods increased linearly as the number of dimensions increased, $r = .86$ for the LSA-based method, $r = .72$ for the correlation-based method and $r = .93$ for the dictionary-based method (all $ps < .0001$), while correct rates for computer-generated test items

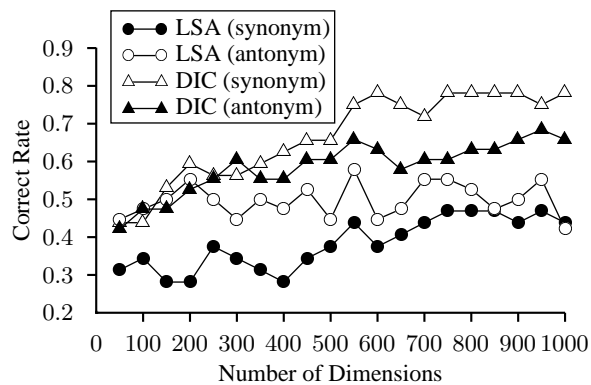


Figure 2: Synonym versus antonym judgment

were steady. Our finding of the absence of any obvious optimal dimensions is in a sharp contrast to Landauer and Dumais's (1997) finding that the LSA word vectors with 300 dimensions achieved the maximum performance of 53% correct rate in a similar multiple-choice synonym test. Note that their maximum performance was a little better than that of our LSA vectors, but still worse than that of our dictionary-based vectors.

Figure 2 shows the performance of the LSA-based and the dictionary-based methods in antonym judgment, together with the result of synonym judgment. (Since the performance of the cooccurrence-based method did not differ from that of the LSA-based method, the correct rates of the cooccurrence-based method are not plotted in this figure.) The dictionary-based method also outperformed the LSA-based method in antonym judgment but their difference was much smaller than that of synonym judgment; at 200 or lower dimensions LSA-based method was better than the dictionary-based method. Interestingly, *the dictionary-based word vectors yielded better performance in synonym judgment than in antonym judgment, while the LSA-based vectors showed better performance in antonym judgment.* These contrasting results may be attributed to the difference of corpus characteristics. Dictionary's definitions for antonymous words are likely to involve different words so that the differences between their meanings can be made clear. On the other hand, in newspaper articles (or literary texts), context words with which antonymous words occur are likely to overlap because their meanings are about the same domain.

Finally, we show the results of comparison among four combinations of corpora and text units for the LSA-based and the cooccurrence-based

Table 1: Comparison of mean correct rate among the combinations of two corpora and two text units

Method	Newspaper		Novel	
	Para	Sent	Para	Sent
SPI test				
LSA	0.383	0.366	0.238	0.369
COO	0.413	0.369	0.255	0.280
Computer-generated test				
LSA	0.410	0.377	0.346	0.379
COO	0.375	0.363	0.311	0.310

Note. Para = Paragraph; Sent = Sentence.

methods. Table 1 lists mean correct rates of SPI test and computer-generated test averaged over all the numbers of dimensions. Regardless of construction methods and test items, the word vectors constructed using newspaper paragraphs achieved the best performance, which are denoted by bold-faces. Concerning an effect of corpus difference, the newspaper corpus was superior to the literary corpus. The difference of text units did not have a clear influence on the performance of word spaces.

5 Experiment 2: Word Association

5.1 Method

In order to compare the ability of the word spaces to judge associative similarity, we conducted a word association experiment using a Japanese word association norm “Renso Kijun-hyo” (Umemoto, 1969). This free-association database was developed based on the responses of 1,000 students to 210 stimulus words. For example, when given the word *writer* as a stimulus, students listed the words shown in Table 2. (Table 2 also shows the original words in Japanese.)

For the simulation experiment, we selected 176 stimulus words that all the three corpora contained. These stimuli had 27 associate words on average. We then removed any associate words that were synonymous with the stimulus word (e.g., *author* in Table 2), since the purpose of this experiment was to examine the ability to assess associative similarity between words. Whether or not each associate is synonymous with the stimulus was determined according to whether they belong to the same deepest category of a Japanese thesaurus “Goi-Taikei” (Ikehara et al., 1999).

In the computer simulation, cosine similarity

Table 2: Associates for the stimulus word *writer*

Stimulus: writer (作家)				
Associates:				
novel	pen	literary work	painter	
小説	ペン	著作	画家	
book	author	best-seller	money	
本	作者	ベストセラー	金	
write	literature	play	art	work
書く	文学	劇	芸術	作品
popular	human	book	paper	pencil
流行	人	書物	紙	鉛筆
lucrative	writing	mystery	music	
儲かる	文章	推理	音楽	

between the stimulus word and each of all the other words included in the vector space was computed, and all the words were sorted in descending order of similarity. The top i words were then chosen as associates.

The ability of word spaces to mimic human word association was evaluated on mean precision. Precision is the ratio of the number of human-produced associates chosen by computer to the number i of computer-chosen associates. A precision score was calculated every time a new human-produced associate was found in the top i words when i was incremented by 1, and after that mean precision was calculated as the average of all these precision scores. It must be noted here that, in order to eliminate the bias caused by the difference in the number of contained words among word spaces, we conducted the simulation using 46,000 words that we randomly chose for each corpus so that they could include all the human-produced associates.

Although this computational method of producing associates is sufficient for the present purpose, it may be inadequate to model the psychological process of free association. Some empirical studies of word association (Nelson et al., 1998) revealed that frequent or familiar words were highly likely to be produced as associates, but our methods for constructing word vectors may not directly address such frequency effect on word association. Hence, we conducted an additional experiment in which only familiar words were used for computing similarity to a given stimulus word, i.e., less familiar words were not used as candidates of as-

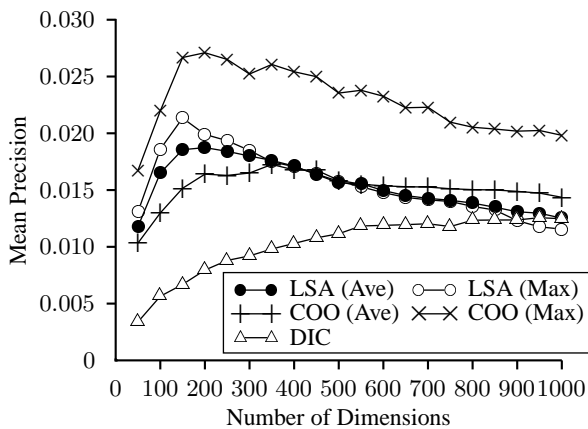


Figure 3: Mean precision of word association judgment

sociates. For a measure of word familiarity, we used word familiarity scores (ranging from 1 to 7) provided by “Nihongo no Goitaikei” (Amano and Kondo, 2003). Using this data, we selected the words whose familiarity score is 5 or higher as familiar ones.

5.2 Results and Discussion

For each of the nine vector spaces, the association judgment simulation was conducted and the mean precision was calculated. As in the synonym judgment experiment, this process was repeated by every 50 dimensions between 50 and 1000.

Figure 3 shows the result of word association experiment. For the LSA-based and the cooccurrence-based methods, two kinds of mean precision were plotted: the average of mean precision scores for the four word vectors and the maximum score among them. (As we will show in Table 3, the LSA-based method achieved the maximum precision when sentences of the newspaper corpus were used, while the performance of the cooccurrence-based method was maximal when paragraphs of the newspaper corpus were used.) The overall result was that the dictionary-based word vectors yielded the worst performance, as opposed to the result of synonym judgment. There was no big difference in performance between the LSA-based method and the cooccurrence-based method, but the maximal cooccurrence-based vectors (constructed from newspaper paragraphs) considerably outperformed the other kinds of word vectors.¹ These results clearly show that *the LSA-based and*

¹These results were replicated even when all the human-produced associates including synonymous ones were used for assessing the precision scores.

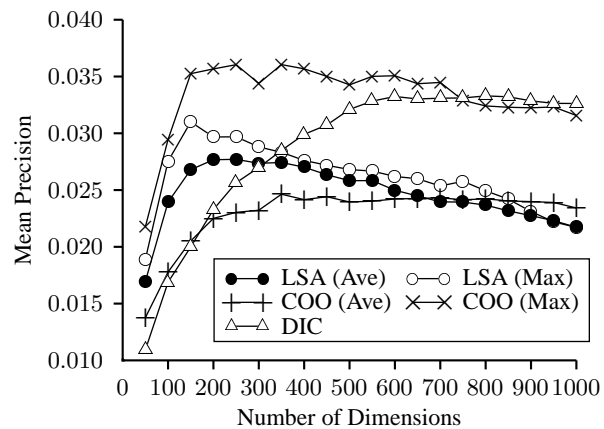


Figure 4: Average precision of word association judgment for familiar words

the cooccurrence-based vector spaces reflect associative similarity between words more than the dictionary-based space.

The relation between the number of dimensions and the performance in association judgment was quite different from the relation observed in the synonym judgment experiment. Although the score of the dictionary-based vectors increased as the dimension of the vectors increased as in the case of synonym judgment, the scores of both LSA-based and cooccurrence-based vectors had a peak around 200 dimensions, as Landauer and Dumais (1997) demonstrated. This finding seems to suggest that some hundred dimensions may be enough to represent the knowledge of associative similarity.

Figure 4 shows the result of the additional experiment in which familiarity effects were taken into account. As compared to the result without familiarity filtering, there was a remarkable improvement of the performance of the dictionary-based method; the dictionary-based method outperformed the LSA-based method at 350 or higher dimensions and the cooccurrence-based method at 800 or higher dimensions. This may be because word occurrence in the sense definitions of a dictionary does not reflect the actual frequency or familiarity of words, and thus the dictionary-based method may possibly overestimate the similarity of infrequent or unfamiliar words. On the other hand, since the corpus of newspaper articles or novels is likely to reflect actual word frequency, the vector spaces derived from these corpora represent the similarity of infrequent words as appropriately as that of familiar words.²

The result that the cooccurrence-based word

Table 3: Comparison of mean precision among the combinations of two corpora and two text units

Method	Newspaper		Novel	
	Para	Sent	Para	Sent
All associates				
LSA	0.016	0.017	0.015	0.015
COO	0.023	0.018	0.012	0.008
Familiar associates				
LSA	0.0261	0.0258	0.024	0.023
COO	0.033	0.027	0.018	0.014

Note. Para = Paragraph; Sent = Sentence.

vectors constructed from newspaper paragraphs achieved the best performance was again obtained in the additional experiment. This consistent result indicates that *the cooccurrence-based method is particularly useful for representing the knowledge of associative similarity between words*. The relation between the number of dimensions and mean precision was unchanged even if a familiarity effect was considered.

Finally, Table 3 shows the comparison result among four kinds of word vectors constructed from different corpora and text units in the experiment with and without familiarity filtering. The listed values are mean precisions averaged over all the 20 numbers of dimensions. As in the case of synonym judgment experiment, word vectors constructed from newspaper paragraphs achieved the best performance, although only the LSA-based vectors had the highest precision when they were derived from sentences of newspaper articles. As in the case of synonym judgment, the newspaper corpus showed better performance than the novel corpus, and especially the cooccurrence-based method showed a fairly large difference in performance between two corpora. This finding seems to suggest that word cooccurrence in a newspaper corpus is more likely to reflect associative similarity.

6 Semantic Network and Similarity

As related work, Steyvers and Tenenbaum (2005) examined the properties of semantic network, an-

²Indeed, the dictionary-based vector spaces contained a larger number of unfamiliar words than the other spaces; 63% of words in the dictionary were judged as unfamiliar, while only 42% and 50% of words in the newspapers and the novels were judged as unfamiliar.

other important geometric model for word meanings. They found that three kinds of semantic networks — WordNet, Roget’s thesaurus, and word associations — had a small-world structure and a scale-free pattern of connectivity, but semantic networks constructed from the LSA-based vector spaces did not have these properties. They interpreted this finding as indicating a limitation of the vector space model such as LSA to model human knowledge of word meanings.

However, we can interpret their finding in a different way by considering a possibility that different semantic networks may capture different kinds of word similarity. Scale-free networks have a common characteristic that a small number of nodes are connected to a very large number of other nodes (Barabási and Albert, 1999). In the semantic networks, such “hub” nodes correspond to basic and highly polysemous words such as *make* and *money*, and these words are likely to be taxonomically similar to many other words. Hence if semantic networks reflect in large part taxonomic similarity between words, they are likely to have a scale-free structure. On the other hand, since it is less likely to assume that only a few words are associatively similar to a large number of other words, semantic networks reflecting associative similarity may not have a scale-free structure. Taken together, Steyvers and Tenenbaum’s (2005) finding can be reinterpreted as suggesting that WordNet and Roget’s thesaurus better reflect taxonomic similarity, while the LSA-based word vectors better reflect associative similarity, which is consistent with our finding.

7 Conclusion

Through two simulation experiments, we obtained the following findings:

- The dictionary-based word vectors better reflect the knowledge of taxonomic similarity, while the LSA-based and the cooccurrence-based word vectors better reflect the knowledge of associative similarity. In particular, the cooccurrence-based vectors are useful for representing associative similarity.
- The dictionary-based vectors yielded better performance in synonym judgment, but the LSA-based vectors showed better performance in antonym judgment.
- These kinds of word vectors showed the distinctive patterns of the relationship between

the number of dimensions of word vectors and their performance.

We are now extending this work to examine in more detail the relationship between various kinds of word vectors and the quality of word similarity involved in these vectors. It would be interesting for further work to develop a method for extracting the knowledge of a specific similarity from the word space, e.g., extracting the knowledge of taxonomic similarity from the dictionary-based word space. Vector negation (Widdows, 2003) may be a useful technique for this purpose. At the same time we are also interested in a method for combining different word spaces into one space, e.g., combining the dictionary-based and the LSA-based spaces into one coherent word space. Additionally we are trying to simulate cognitive processes such as metaphor comprehension (Utsumi, 2006).

Acknowledgment

This research was supported in part by Grant-in-Aid for Scientific Research(C) (No.17500171) from Japan Society for the Promotion of Science.

References

- Jean Aitchison. 2003. *Words in the Mind: An Introduction to the Mental Lexicon, 3rd Edition*. Oxford, Basil Blackwell.
- Shigeaki Amano and Kimihisa Kondo, editors. 2003. *Nihongo-no Goitokusei CD-ROM (Lexical properties of Japanese)*. Sanseido, Tokyo.
- Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *Science*, 286:509–512.
- Curt Burgess. 1998. From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model. *Behavior Research Methods, Instruments, & Computers*, 30(2):188–198.
- Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas L. Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society For Information Science*, 41(6):391–407.
- Peter Gärdenfors. 2000. *Conceptual Spaces: The Geometry of Thought*. MIT Press.
- Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentaro Ogura, Yoshifumi Ooyama, and Yoshihiko Hayashi. 1999. *Goi-Taikei: A Japanese Lexicon CDROM*. Iwanami Shoten, Tokyo.
- Kaname Kasahara, Kazumitsu Matsuzawa, and Tsutomu Ishikawa. 1997. A method for judgment of semantic similarity between daily-used words by using machine readable dictionaries. *Transactions of Information Processing Society of Japan*, 38(7):1272–1283. in Japanese.
- Thomas K. Landauer and Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240.
- Douglas L. Nelson, Cathy L. McEvoy, and Thomas A. Schreiber. 1998. The university of south florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>.
- Yoshiki Niwa and Yoshihiko Nitta. 1994. Co-occurrence vectors from corpora vs. distance vectors from dictionaries. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING94)*, pages 304–309.
- Robert M. Nosofsky. 1992. Similarity scaling and cognitive process models. *Annual Review of Psychology*, 43:25–53.
- Hinrich Schütze. 1998. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.
- Mark Steyvers and Joshua B. Tenenbaum. 2005. The large-scale structure of semantic network: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78.
- Mark Steyvers, Richard M. Shiffrin, and Douglas L. Nelson. 2004. Word association spaces for predicting semantic similarity effects in episodic memory. In Alice F. Healy, editor, *Experimental Cognitive Psychology and Its Applications*. American Psychological Association, 2004.
- Takao Umemoto. 1969. *Renso Kijunhyo (Free Association Norm)*. Tokyo Daigaku Shuppankai, Tokyo.
- Akira Utsumi, Koichi Hori, and Setsuo Ohsuga. 1998. An affective-similarity-based method for comprehending attributional metaphors. *Journal of Natural Language Processing*, 5(3):3–32.
- Akira Utsumi. 2006. Computational exploration of metaphor comprehension processes. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society (CogSci 2006)*.
- Dominic Widdows. 2003. Orthogonal negation in vector spaces for modelling word-meanings and document retrieval. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 136–143.
- Dominic Widdows. 2004. *Geometry and Meaning*. CSLI Publications.