Synonym Retrieval Using Word Vectors from Text Data

Kaname Kasahara
NTT Corporation
Kyoto, 619-0237, Japan
kaname@cslab.kecl.ntt.co.jp

Tsuneaki Kato
The University of Tokyo
Tokyo, 153-8902, Japan
kato@boz.c.u-tokyo.ac.jp

Christopher Manning
Stanford University
Stanford CA 94305-9040, USA
manning@cs.stanford.edu

Abstract

This paper proposes a method of computing a simulation of retrieving synonyms for a word. In this method, word vectors in a multi-dimensional space are derived from dictionary definitions and used to calculate degrees of semantic similarity between the words. Based on human subject experiments, standard synonyms for 200 Japanese words were collected for evaluating the simulation. These synonyms were associated by 200 subjects and judged to be appropriate by half of other 76 subjects. The average precision of the retrievals based on the proposed method using a Japanese dictionary was higher than that based on a method using co-occurrence vectors acquired from a one-year volume of Japanese newspaper articles. The method was refined from the viewpoints of how to reduce the dimensions of the space and how to select the vocabulary to be retrieved.

1 Introduction

The idea of using word vectors in a multi-dimensional space has been applied to measuring the degree of semantic similarity between words and between textual data. Recently, several methods of automatic acquisition of word vectors from textual data have been proposed (Hindle, 1990; Schutze, 1992; Niwa and Nitta, 1994) and applied to several kinds of textual processing such as solving word sense disambiguation (Niwa and Nitta, 1994), information retrieval (Schutze and Pedersen, 1995), and knowledge management (Yukawa et al., 2001).

The other application of word vectors built automatically is computer simulation of human judgment based on semantic memory of words. In psychological studies, word vectors whose features were acquired from human-subject experiments were used to guess the mental model of humans (Osgood, 1952; Deese, 1965). Therefore, there is a possibility that text-based word vectors can also be used to simulate human judgment of the semantic similarity between words. One of the basic tasks related to word similarity is the retrieval of synonym, since the simulation may be simply realized by calculating degrees of similarity between the vector of a given word and all other vectors and then by selecting words whose degrees of similarity are considerably high. Some studies of building text-based word vectors showed examples of finding similar words by using word vectors. However, there has been no quantitative evaluation of how close these simulations were to human judgment in synonym retrieval. This is mainly because that no appropriate database had been developed on human judgment of synonyms, especially in Japanese.

In this paper, we examine computer simulation of human synonym retrieval using word vectors made from text data. The paper also explains that how to build the standard database of synonyms derived from human-subject experiments and how close simulation is to the database.

2 Automatic Acquisition of Word Vectors

Several methods of building word vectors from text data and using them for different purposes and with different kind of source text data have been pro-

posed. However, they all share the following basic processes.

1. Getting "rough" word vectors whose elements are weights of features for the words from text data.

2. Transforming the rough word vectors into word vectors whose features are independent of each other.

3. Calculating a degree of similarity between the words using their word vectors.

In the following, we explain these procedures and the features of the existing methods.

2.1 Rough Word Vectors from Source Text Data

Features of words in text data are extracted from the data, and rough word vectors whose elements are "weights" of the features are made. Here, n (≥ 1) words of vocabulary, which are included in the text data and whose vectors are acquired, are defined as $N(= \{\mathrm{w}_1, \cdots, \mathrm{w}_n\})$. When the concepts of all the words in N can be represented by m $(geq1)$ features, a rough word vector of word w_i $(i = 1, \ldots, n)$ can be described as

$$w_i = (v_{i1}, \cdots, v_{im}). \tag{1}$$

v_{ij} $(j = 1, \cdots, m)$ is a numerical value of how strong the j th feature is related to word w_i and is denoted as "weight." The column vector whose elements are the rough word vectors becomes a "rough matrix," denoted as G_0.

Mainly, a text corpus or a dictionary was used as source data for getting a rough matrix. When a text corpus was used, the features of words were verbs for predicate and subject noun (Hindle, 1990), adjectives for modified nouns (Hatzivassiloglou and McKeown, 1993), words for neighboring word (Schutze, 1992), and so on. The IDs of documents included in a corpus were used as features of a word (Deerwester et al., 1990). When a dictionary was used as a source, the features of an entry word were the words in its definition (Kozima and Furugori, 1993; Niwa and Nitta, 1994; Kasahara et al., 1996).

Many definitions in a dictionary are described too briefly to get enough features for a rough word vector for the entry word. Therefore, these methods searched relations between words in the dictionaries and added related words as additional features. As an example, Kasahara's method is introduced. First, words in the definition of the entry word become features and the frequency of them become weights of rough word vectors. Next, G_0, the rough matrix of the vectors, are expanded.

$$G_{0D} = \alpha G_0 + \beta G_0^2 + \gamma G_0^T. \tag{2}$$

α,β, and γ are combination coefficients that are decided experimentally. G_0^2 means features that exist in each definition of a entry word. G_0^T are the features of words in definitions. Evaluation of the method made it clear that G_{0D} was better for judgment of semantic similarity between words than the original rough matrix or each element of G_{0D}.

2.2 Transformation into a Word Vector

When word vectors are used to calculate similarity between words, only elements of the same features in the two word vectors are usually compared. Therefore, features in the rough matrix should be changed into ones that are independent of each other. If a rough matrix is supposed to be linearly transformed into word matrix G, whose k $(\leq m)$ features are independent of each other, the function of transformation can be described as m by k matrix, K.

$$G = G_0 K. \tag{3}$$

Each transformed word vector $\acute{w}_i$ can also be represented by using the transformation matrix K.

$$\acute{w}_i = w_i K = (\acute{v}_{i1}, \cdots, \acute{v}_{ik}). \tag{4}$$

There are two types of transformation: a statistical method and a thesaurus-based one. In the statistical method, a rough matrix is regarded as a correlation matrix, and principal component analysis or singular value composition is applied to make a word matrix (Schutze, 1992). In the thesaurus-based method, each feature in the rough matrix is generalized into a category of the thesaurus (Kasahara et al., 1996).

2.3 Calculation of Degree of Similarity

The degree of similarity between word w_p and word w_q, in the vocabulary N, $sim(w_p, w_q)$, is calculated by using their vectors, $\acute{w}_p$ and $\acute{w}_q$, in word matrix G. The representative measure of similarity is cosine of the angle between word vectors.

$$sim(w_p, w_q) = \frac{\acute{w}_p \acute{w}_q}{|\acute{w}_p||\acute{w}_q|}. \quad (5)$$

3 Computer Simulation of Synonym Retrieval

In this section, we propose a method of computer simulation of retrieval using automatically acquired word vectors. First, the task of retrieval is defined as follows:

synonym retrieval task
is a task to output words in vocabulary N which are highly similar to the stimulus word w_i ($\in N$).

Using the word vectors, the basic procedures of simulating the task are: (1) to select text data for word vectors; (2) to build word vectors from the data; (3) to output highly similar words to the stimulus word. In the following subsections, each process for the proposed method is described.

3.1 Source Data

As mentioned in the previous section, text corpora or dictionaries were used to build word vectors. It remains unclear which data are better to use as the source for the task of synonym retrieval. In this paper, two series of word vectors based on a text corpus and on a dictionary are compared for the task. The number of vocabulary in the text data should also be considered. For example, in Japanese the number of words that more than half of the adult Japanese native speakers know is estimated to be about 66,000 from human-subject experiments (Amano and Kondo, 1998). In order to compare computer simulation with human judgment, synonyms should be retrieved from words of more than the number.

3.2 Word Vectors

There are several methods for carrying out linear transformation of a rough word matrix directly made from source text data into a word matrix of word vectors. This paper proposes a combination method using a SVD-based method, one of statistical method, and a thesaurus-based method.

3.2.1 SVD-based Method (Schutze, 1992)

Singular value decomposition, SVD, was adopted to transform a rough word matrix made from word occurrences into the word matrix.

It is known that any matrix G_0, can be decomposed as

$$G_0 = U\Sigma V^T. \quad (6)$$

When the rank of G_0 is r, U and V are orthogonal matrices whose dimensions are n by r and r by n ($U^TU = V^TV = I$). Σ is an r by r diagonal matrix whose elements are singular values of G_0. The matrix G_{0k} is the product of U_k, which consists of upper k columns of U, and V_k^T, which consists of upper k rows of V^T. This matrix is known to be the best approximation of G_0 of matrices whose ranks are less than k.

$$\begin{aligned} G_{0k} &= U_k \Sigma_k V_k^T \quad (7)\\ ||G_0 - G_{0k}||_F &= \min_{\text{rank(B)} \le \text{k}} ||G_0 - B||_F \\ &= \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}. \end{aligned}$$

Deerwester adopted this analysis in the information retrieval of a word by document matrix (Deerwester et al., 1990). Schutze adopted a cooccurrence matrix derived from a text corpus and regarded U_k as a word matrix. Equation (7) can be changed to

$$U_k = G_{0k} V_k \Sigma_k^{-1} \quad (U_k U = I). \quad (8)$$

When comparing this equation with equation (3), U_k can be regarded as word matrix G and $V_k\Sigma_k^{-1}$ can be regarded as transformation matrix K.

3.2.2 Thesaurus-based Method (Kasahara et al., 1996)

When features in a rough word matrix are described in words, a thesaurus can be used to generalize the features into its categories. Categories in a thesaurus are decided so that they are independent

of each other. If the categories can be regarded as the basis of word vectors, a rough word matrix can be transformed by using the thesaurus.

When N words are categorized into k classes in the thesaurus whose classes are represented as $\{th_1, \cdots, th_k\}$, the jth row vector, k_j in the thesaurus-based transformation matrix K in equation (3), K_T, can be represented as

$$k_j = \left(\frac{c(j,1)}{C(j)}, \cdots, \frac{c(j,i)}{C(j)}, \cdots, \frac{c(j,k)}{C(j)} \right)$$
$$C(j) = \Sigma_{i=1}^{k} c(j,i).$$

$c(j,i)$ is the function whose value becomes one if the jth feature in the rough word matrix is categorized in class th_i. In the other situations, $c(i,j)$ always becomes zero. The vector k_j distributes weights of the jth feature, v_{ij} ($i = 1, \cdots, n$), in G_0. When the jth feature in a rough vector is categorized into $C(j)$ classes, the ith feature in the vector k_j becomes $1/C(j)$ if the feature is categorized into class th_i.

Using this transformation matrix K_T, rough word matrix G_0 is changed into a word matrix whose features are independent of each other:

$$G_T = G_0 K_T. \tag{9}$$

A merit of the thesaurus-based model is that the transformation matrix does not depend on a rough word matrix. Therefore, when a new rough word vector, w_{n+1}, is added to the matrix, the other word vectors are not re-calculated to get word vector $\acute{w}_{n+1}$.

$$\acute{w}_{n+1} = w_{n+1} K_T. \tag{10}$$

3.2.3 Combination Method

The two existing methods differ in the kinds of source data that are used to judge relations between features in a rough word matrix. In the SVD-based method, only relations that can be derived from a rough word matrix itself are considered. Therefore, a word matrix can not be properly represented, if the scale of the source text data for the rough word matrix is small. In the thesaurus-based method, only relations described in the thesaurus are used. However, relations between classes in the rough word matrix are not considered.

In this paper, a new method of transforming a rough word into a word matrix is proposed in order to overcome the problems of the existing two methods. In the combination method, a rough word matrix is transformed by using the thesaurus method and then and re-transformed by using the SVD-based method.

$$G_T = G_0 K_T$$
$$G_{TM} = G_{Tk} V_k \Sigma_k^{-1}. \tag{11}$$

The first process of categorizing feature words into classes of a thesaurus has the effect of feature-smoothing, which can not be done by using only the rough word matrix. In the second process of the SVD analysis of the transformed matrix, whose features are the classes, has the effect of converting the classes related to each other into independent features.

3.3 Output

After word vectors are acquired, degrees of similarity between the vector of the stimulus word and all of the other word vectors are calculated and highly similar words are output as the synonym for the stimulus word. In order to gain precision of the simulation compared with human judgment, several words should always be excluded from the words to be retrieved. Among the words of a text corpus, such as a newspaper articles or entry words in a dictionary, there are rarely used words or difficult ones. Unfamiliar words will not be selected as a synonym of a stimulus word by people even when such words are highly similar to the stimulus.

In order to consider such a word attribute for the simulation, we used a database of word familiarity (Amano and Kondo, 1999) to restrict the words to be retrieved. The familiarity database was developed for about 80,000 Japanese words whose familiarity scores in the database were measured by 32 Japanese adults using a seven-point scale. Term frequency in a text corpus is another word attribute and it has high correlation with word familiarity. However, it was also reported that there were several familiar words such as “tamanegi” (onion) that appeared fewer than ten times in a fourteen-year sample of Japanese newspaper articles. Therefore, we

adopted word familiarity for the condition of words to be retrieved in the simulation.

4 Experimental Results

4.1 Standard data of synonym

A thesaurus, which is a database of words categorized into several classes, seems to be a standard database for simulating synonym retrieval. However, it is made by a few lexicologists, and it is not clear whether their judgments agree with those ones of ordinary people. Therefore, we collected standard synonyms through human-subject experiments.

First, we selected 200 daily-used Japanese words as stimulus words for the retrieval. Stimulus words should be selected from commonly known words because all of the subjects participating in the experiments should be familiar with the stimuli. For the same reason that we restricted the words to be retrieved as synonyms in the proposed simulations method, the database of word familiarity (Amano and Kondo, 1999) was used to select commonly known Japanese words as stimuli. The database contained about 80,000 words with scores of their word familiarity. Two hundred stimulus words were randomly selected from 28,764 words whose familiarity scores are more than five. The words were estimated to be commonly known by more than 90% of Japanese adults from human-subjects experiments (Amano and Kondo, 1998). The words below are examples of the 200 Japanese stimulus words and their translations.

Table 1: Examples of stimuli

"kuma" (bear), "kenchiku" (construction), "shouhi" (consumption), "hikohki" (airplane), "sakusen" (operation), "odoru" (dance), "youshoku" (western food), "uma" (horse), "hamusutah" (hamster)

Next, we explain the first experiment for collecting synonyms of the stimuli. One hundred subjects of university students as subjects were asked to write down associated synonyms for each of the 200 stimulus words on questionnaire sheets. They were required to write down as many words as they could in ten seconds for one stimulus word. As a result, about 40 associated synonyms were acquired for each stimulus on average.

For each stimulus word and its each associated synonym, 76 different university students were asked to judge whether the synonym was appropriate for the stimulus word. The associated synonyms that more than half of the subjects in the second experiment judged to be appropriate were regarded as standard human-judged synonyms. As a result, on average, about 6.4 standard synonyms were acquired for each of the 200 stimulus words.

We compared synonyms made with those in a one of the representative Japanese thesaurus (Susumu Ono, 1990), which categorizes about 60,000 words into 1,000 classes. For each stimulus word of 155 appearing in the thesaurus, the words that were included in the class of the stimulus word were assumed to be synonyms. It was found that the thesaurus-based synonym retrieval resulted in about only 35% of recall, which suggests that thesauri do not include enough synonyms that are judged to be appropriate by ordinary people. Moreover, the precision of the thesaurus-based retrieval was about 1.9%. Thesauri may be effect for several natural language processing tasks, but it was found that they can not fully simulate synonym judgment by average people. Therefore, we used the synonyms for 200 stimuli as the standard data.

For each stimulus word, a simulation was done and a synonym word list was compared with synonyms in the standard data. Averaged precision for eleven points from zero to one was calculated by using standard evaluation program for information retrieval (trec eval (Buckley, 1992)). When N, the vocabulary of a word matrix including in 200 L stimulus words, the averaged value of the precision of synonym retrievals was regarded as the evaluation value of the simulation of synonym retrieval, $Eval_1$.

$$Eval_1 = \frac{1}{11L} \sum_{i=1}^{L} \sum_{\substack{j=0 \\ step=0.1}}^{1} prec(st_i, recall \equiv j). \quad (12)$$

st_i is a stimulus of L words and $prec(st_i, recall = j)$ is a score of precision when the query is st_i and the recall score is j $(= 0, 0.1, \cdots, 1)$. When some standard synonyms are not included in the vocabulary N, they were discarded from the retrieval answers so that the evaluation value would not depend on the scale of the source text data. However, it is also important that the vocabulary N covers as many

standard vocabulary synonyms as possible. Therefore, the ratio of the number of standard synonyms that appeared in N to the number of all the standard synonyms, R_{max}, was also calculated to evaluate the simulation.

4.2 Source Text Data for Word Vectors

First, two kinds of text data were compared as the source of a word matrix.

> **DIC** 20 MBytes of Japanese dictionary Gakken Kokugo Jiten, whose number of entry words is 88,633
>
> **CPS** 122 MBytes of Japanese newspaper (Mainichi Shinbun) articles issued in 2,000

When the kinds of source data are different, corresponding methods of making a rough word matrix for the data are required. For DIC, Kasahara's method was adopted to build a rough matrix (Kasahara et al., 1996). For each of 88,633 entry words, the frequencies of appearance of all words in the definition of the entry word were counted, and an 88,633 by 88,633 rough word matrix was built and denoted as G_0. Equation (2) was adopted to build the expanded rough word matrix G_{0D} from G_0. We used linear combination efficiencies in equation (2) as 1.0(=α), 0.2 (=β), and 0.2(=γ), which were proposed in the original study (Kasahara et al., 1996).

For CPS, Schutze's method was adopted to build a rough word matrix from the corpus (Schutze, 1992). For all of the sentences of the corpus, 84,772 entry words appeared in the CPS more than three times, and the their 84,772 word rough vectors were made. Frequencies of cooccurrence in a sentence between each entry word and each of 5,000 highly appearing word of the entry words were counted. An 84,772 by 5,000 rough word matrix, denoted as G_{0C}, whose elements are frequencies of occurrence in each sentence of the corpus, was made. Sentences that consist of definitions in a dictionary can also be regarded as a text corpus. Therefore, Schutze's method was adopted in the corpus of definitions in DIC, and the 20,000 by 1,000 rough word matrix, G_{0DC}, was build.

For each of G_{0D}, G_{0C}, and G_{0DC} rough matrices, transformation to a word matrix whose number of dimensions was 136 was done by using the singular value composition

Results of evaluating the matrix from the viewpoint of synonym retrieval is shown in table 2. The baseline in the table is the result of simulation where a large Japanese thesaurus, Nihongo GoiTaikei (Ikehara et al., 1997), was used to judge whether words in the same category of the stimulus word were synonyms. The thesaurus categorizes about 300,000 words into about 3,000 classes. In the table, $Eval_1$, averaged precision of retrieval based on G_{0D}, were better than the baseline, G_{0C} and G_{0CD}. The rough matrices G_{0D} and G_{0DC} were built from the same dictionary. The results show that Kasahara's method, which is optimized for dictionary, is better than Schutze's method for the text corpus. We used G_{0D} in the following evaluations of the proposed method.

Table 2: Effect of source text data

rough word matrix	source	EVAL1	R_{max}
G_{0D}	DIC	0.051	0.622
G_{0DC}	DIC	0.028	0.325
G_{0C}	CPS	0.014	0.653
(baseline)	-	0.034	0.258

4.3 Representation of Word Vectors

Transformations of a rough word matrix into a word matrix were examined. Three methods of transformation, a thesaurus-method, an SVD-based method, and a combination of them, were adopted to the 88,633 by 88,633 rough matrix from the dictionary as described above, and word matrices were made. In the thesaurus-based method and the proposed method, Nihongo GoiTaikei was used. It categorizes 300,000 Japanese words into 2,715 classes, and the classes are related in a top-down tree structure whose maximum steps from the top class are twelve. When a class in it is substituted for its upper class, the number of classes in the thesaurus can be changed. In this experiment, 9, 30, 136, 392, and 2715 classes of thesauri were used for the transformation.

Figure 1 shows the results of evaluating the matrices in the task of similar word retrieval in several dimensions. Evaluation values of similar word re-

trieval based on matrices made from the three methods with the same dimensions of word matrices are almost the same when the dimensions are less than about forty. However, $Eval_1$ based on the word matrix of the combination method was much higher than the values based on the thesaurus-based method and the SVD-method with the same numbers of dimensions within 136.

There was a small difference between $Eval_1$ of the combination method with 200 and 250 dimensions. This tendency is similar to the results of the SVD-method. However, in the thesaurus-based method, $Eval_1$ monotonously gained in higher dimensions. When all of the 2,715 categories were used to represent the basis of the word matrix, $Eval_1$ of the matrix made by using the thesaurus-based method was 0.112, higher than the values in the figure. The thesaurus-based method also has the property of keeping the sparseness of the rough word matrix. Each rough word vector made from the dictionary whose number of dimensions was 88,633 has about 30 non-zero weights on average. In word matrices made from the SVD-based method and the combination method, almost all of the weights in the word vector became non-zero values in any selected number of dimensions. However, when the thesaurus-based method was adopted, numbers of non-zero values in a word vector was only about 46, even when the number of dimensions of the word matrix was 2,715. Therefore, we adopted the word matrix of 2715 dimensions derived using the thesaurus-based method from the viewpoints of precision of synonym retrieval and data size of the word matrix.

4.4 Selection of words for retrieval

Finally, vocabularies for synonym retrieval were evaluated. Table 3 shows a comparison of two vocabularies. The first line is the result when all of the entry words, 88,633, were used for the object to be retrieved. The second line is the result of using selected 26,371 words whose degrees of word familiarity were more than five and that are assumed to be known by about 94% of Japanese adult native speakers. The value of $Eval_1$ retrieved from the 26,371 words was twice as high as that retrieved from all of the words. On the contrary, only about 14% of synonyms in the standard database derived from the human subjects were lost. This selection of vocabulary resulted in a considerable improvement in the precision of synonym retrieval with small losses in recall.

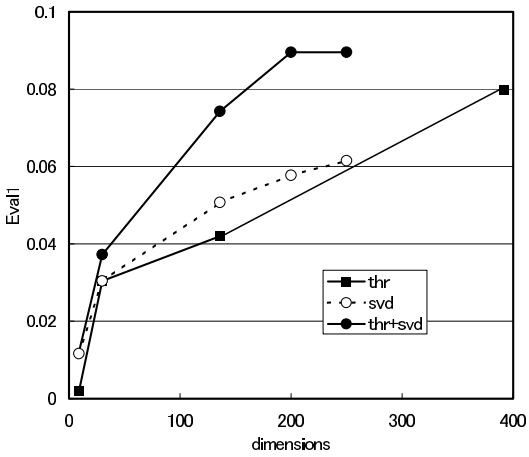

Figure 1: Evaluation of transformation

Table 3: Effect of selecting a vocabulary

vocabulary	$Eval_1$	R_{max}
88,633	0.112	0.622
26,371	0.213	0.486

Finally, Table 4 shows retrieved synonyms for the word “machi” (town) by using word vectors made by the proposed method. For this stimulus, standard synonyms were “toshi” (city), “machi” (town), “shigai” (town), “tokai” (city), and “taun” (town). The three standard synonyms of these five synonyms appeared in the list. Table 5 is the result of retrieval by using word vectors made from a large text corpus, CPS, by using Schutze’s method. No standard words in the list appears in the list. From the examples and the results of evaluating the simulations, the proposed method of simulating synonym retrieval was found to be fairly close to the judgment synonyms by humans.

5 Conclusion

In this paper, we investigated simulation of synonym word retrieval where word vectors were derived from text data, was studied. After selecting standard synonyms based on human-subject experiments, simulations were evaluated. First, it was found that dictionaries provide fairly promis-

Table 4: Retrieved synonyms for “machi” (town) based on proposed method

rank	retrieved synonyms	similarity
1	“shigai” (town)	0.66
2	“toshi” (city)	0.63
3	“beddotaun” (suburbs)	0.59
4	“taun” (town)	0.59
5	“denentoshi” (garden town)	0.57
6	“kinkou” (suburb)	0.55
7	“oodouri” (avenue)	0.55
8	“toukaidou” (toukaidou)	0.52
9	“shi” (city)	0.52
10	“hankagai” (downtown)	0.52

Table 5: Retrieved synonyms for “machi” (town) based on Schutze (1992)

rank	retrieved synonyms	similarity
1	“nigiwai” (bustle)	0.60
2	“machinami” (cityscape)	0.60
3	“sunahama” (sand beach)	0.56
4	“utsukushii” (beautiful)	0.56
5	“kawabe” (riverside)	0.56
6	“gareki” (rubble)	0.56
7	“aruku” (walk)	0.55
8	“fune” (ship)	0.55
9	“kuukann” (space)	0.55
10	“mizuumi” (lake)	0.55

ing source data for building a word matrix. Next, a transformation method using a thesaurus-based one and a SVD-based one at once was found to be the best for building word matrix whose features are independent of each other with the same number of dimensions. However, a word matrix of large number of dimensions, about 3000, was built using the thesaurus method and the resulted in the highest precision while keeping the sparseness of the rough word matrix. Finally, selecting vocabulary based on the degree of word familiarity was shown to be an effective way to improve precision of synonym retrieval.

It is not clear whether the proposed method of building a word matrix is effective for the other applications such as word sense disambiguation or information retrieval. In future work, the method should be evaluated for these applications.

Acknowledgment

This research was supported in part by the Research Collaboration between NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation and CSLI, Stanford University(, research project on Concept Bases for Lexical Acquisition and Intelligently Reasoning with Meanings.)

References

Shigeaki Amano and Tadahisa Kondo. 1998. Estimation of mental lexicon size with word familiarity database. In *Proc. of Intl. Conf. on Spoken Language Processing*, volume 5, pages 2119 – 2122.

Shigenari Amano and Tadahisa Kondo. 1999. *Goi-Tokusei (Lexical properties of Japanese) Vol. 1*. Sanseido.

Katsuji Bessho. 2001. Text segmentation using word conceptual vectors (in japanese). *Trans. of IPSJ*, 42(11):2650 – 2662.

Chris Buckley. 1992. Smart version 11.0. ftp://ftp.cs.cornell.edu/pub/smart.

S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41:391 – 407.

Jamese Earle Deese. 1965. *The Structure of Associations in Language and Thought*. The Johns Hopkins Press.

V. Hatzivassiloglou and K.R. McKeown. 1993. Towards the automatic identification of adjectival according to meaning. In *3lst Annual Meeting of the Association for Computational Linguistics*, pages 172–182.

D. Hindle. 1990. Noun classification from predicate-argument structures. In *In Proceedings of ACL*, pages 268–275.

Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentaro Ogura, Yoshifumi Oyama, and Yoshihiko Hayashi, editors. 1997. *The Semantic System, volume 1 of Goi-Taikei – A Japanese Lexicon*. Iwanami Shoten.

Kaname Kasahara, Kazumitsu Matsuzawa, and Tsutomu Ishikawa. 1996. Refinement method for a large–scale knowledge base of words. In *Working Papers of the Third Symposium on Logical Formalizations of Commonsense Reasoning*, pages 73–82.

Hideki Kozima and Teiji Furugori. 1993. Simiarity between words computed by spreading activation on an

english dictionary. In *In Proceedings of EACL-93*, pages 232–239.

Yoshiki Niwa and Yoshihiko Nitta. 1994. Co-occurrence vectors from corpora vs. distance vectors from dictionaries. In *In Proceedings of the International Conference on Computational Linguistics(COLING-94)*, pages 304–309.

C.E. Osgood. 1952. The nature and measurement of meaning. *Psychological Bulletin*, 49:197–237.

H. Schutze and J.O. Pedersen. 1995. Information retrieval based on word senses. In *Fourth Annual Symp. on Document Analysis and Information Retrieval*, pages 161–175.

H. Schutze. 1992. Dimensions of meaning. In *Proceedings of Supercomputing 92*, pages 787–796.

Masando Hamanishi Susumu Ono. 1990. *Ruigo Kokugo Jiten*. Kadokawa Shoten.

Takashi Yukawa, Kaname Kasahara, Tsuneaki Kato, and Toshiro Kita. 2001. An expert recommendation system using concept-based relevance discernment. In *Intl. Conf. on Tools with Artificial Intelligence*, volume 13, pages 257 – 264.