An Associative Concept Dictionary for Natural Language Processing: Text Summarization and Word Sense Disambiguation

Jun Okamoto\textsuperscript{1} and Shun Ishizaki\textsuperscript{2}

\textit{Graduate School of Media and Governance, Keio University}

\textsuperscript{1} juno@sfc.keio.ac.jp

\textsuperscript{2} ishizaki@sfc.keio.ac.jp

We constructed an Associative Concept Dictionary based on the results from large-scale association experiments with participants in order to develop simulation models and systems of understanding natural language. In this paper, we first briefly explain the construction of this dictionary and describe its unique features; we then show how to apply it to text summarization and word sense disambiguation. The text summarization method uses this dictionary for calculating the importance scores of sentences. A Contextual Semantic Network that includes the semantic relations and quantitative distance information among words is constructed as a model of human contextual understanding using this dictionary. We compare the quality of the summarization with that of human participants and that using conventional methods such as term frequencies. Our method shows that the quality of summarization is better than that of conventional methods. The word sense disambiguation method uses a Dynamic Contextual Network Model constructed using this dictionary. An interactive activation method on the network is used in this system as a model for the human dynamic contextual understanding to identify a word’s meaning. The results show that such dynamic features are effective.

Key words: Associative Concept Dictionary, Dynamic Contextual Network Model, spreading method, important sentence extraction, word sense disambiguation

\textit{Journal of Cognitive Science} 12: 259-276, 2011

\textcopyright 2011 Institute for Cognitive Science, Seoul National University
1. Introduction

1.1 Associative Concept Dictionary
Large scale concept dictionaries such as WordNet (Miller et al., 1993; Fellbaum, 1998), EuroWordNet (Vossen, 1998), FrameNet (Baker & Fillmore, 1998), and EDR (EDR, 1996) have existed since the 1990s. The first three were constructed by psychologists or linguists. WordNet is an English electronic concept dictionary that was created at Princeton University. It organizes English words and phrases into synonym sets (i.e., synsets) representing the underlying lexical concepts. The synsets are linked with each other via their relationships, such as super-sub and antonym. EuroWordNet, which was created in the same way as WordNet, is a multilingual database of several European languages. FrameNet is based on the Frame Semantics by C. Fillmore at Berkeley, which used the Frame semantics (case frames) of verbs. The last one, EDR, is a Japanese concept dictionary and was created by computer engineers for machine translations between Japanese and English.

Our proposed Japanese concept dictionary, which is based on the results of large-scale human association experiments, is a model of the hierarchical structure of human concepts and the semantic relations among concepts. In comparison with WordNet and EDR it was found to have reasonable structures. Our concept dictionary contains new associative relations features due to the quantitative distances among concepts calculated using the linear programming method. This dictionary has good potential for use in semantic analysis in natural language processing (NLP) tasks. The conventional concept dictionaries calculate distances using the number of links between concepts along its hierarchical structure. However, since the concepts of daily life have fine sub-categorization, the number of links between them increases and, as a result, the distances become too long. Conventional dictionaries have few applications for systems of natural language understanding because the distances among the concepts in the dictionaries are difficult to measure. It is, therefore, difficult to show their usability in high level applications. In this paper, our dictionary is applied to text summarization and word sense disambiguation systems in the ways described in the
1.2 Text summarization
Text summarization methods generally require deep semantic processing and background knowledge in order to approach the level of human results. Much of the previous work has used superficial clues (Watanabe, 1996) and ad hoc heuristics. In summarizing texts, the word frequency approach or connectionist approach (Hashida et al., 1987) have often been used for calculating the importance scores of words; the importance scores of sentences are then calculated as the sum of those of the words in the sentences (Mochizuki & Okumura, 2000) (Zechner, 1996). Summarization based on a concept dictionary such as WordNet uses a semantic graph-based representation for the documents (Plaza, 2009).

The Contextual Semantic Network is used in our research to calculate the importance scores for sentences given in the input document. The results are compared with those from human summarization experiments and those from conventional methods using word frequencies. The comparison shows that the system does well in summarization tasks.

1.3 Word sense disambiguation
Word sense disambiguation is one of the most difficult problems in the field of NLP because it needs contextual meanings. A lot of previous work on such disambiguation used the co-occurrence of words in context. Several machine learning algorithms, such as the Naive Bayes method or the Support Vector Machine (Murata et al., 2003), are based on such co-occurrence. The effectiveness of neural network approaches to word sense disambiguation has been suggested (Waltz & Pollack, 1985). Not only neural network architecture but also large-scale machine-readable dictionaries for word sense disambiguation have been exploited (Veronis & Ide, 1990). For knowledge-driven advanced language processing, the WordNet glosses are disambiguated by increasing the connectivity of WordNet concepts (Moldovan, 2004).

Human-beings can assign an appropriate word sense to an ambiguous word in a sentence based on the words that follow it. We propose a Dynamic Contextual Network Model for word sense disambiguation, where
the Contextual Semantic Network has a structure that dynamically changes depending on each input word sequentially in an input sentence. In this model, the network architecture is based on the proposed concept dictionary, which includes the semantic relations among concepts/words, and the relations can be represented using the quantitative distances between them. In the word sense disambiguation process, an interactive activation model is used to identify a word’s meaning on the Contextual Semantic Network.

2. Associative Concept Dictionary

Background knowledge is crucial for computers to better understand contextual information as well as the syntactic or shallow semantic information from texts. The Associative Concept Dictionary (hereafter ACD) was created based on the results of large-scale online association experiments in which many participants simultaneously used the campus network at Shonan Fujisawa Campus of Keio University. The details of the ACD were described in the Journal of NLP (Okamoto & Ishizaki, 2001), which was written in Japanese, and the essential parts of the paper are described below.

In the experiments, the stimulus words that were used were fundamental nouns used in Japanese elementary school textbooks in which homonyms were excluded. Fifty participants focused on each stimulus word. In the experiment, a participant was requested to associate from ten stimulus words with a given set of semantic relations, such as hypernym, hyponym, part/material, attribute, synonym, action, and situation, and to input the associated words using a Japanese input system.

All of the associated words in the ACD have distances to the stimulus words calculated by using the following linear programming method. The distance $D(x, y)$ between concepts, $x$ and $y$, is shown by means of the following equation (Okamoto & Ishizaki, 2001):

\[
D(x, y) = 0.81F(x, y) + 0.27S(x, y),
\]

where $F(x, y) = N_x/(n_{xy} + \delta)$, $\delta = N_x/10 - 1$, $(N_x \geq 10)$ and $s(x, y) = 1/n_{xy} \sum_{i=1}^{n_{xy}} s_{xyi}$. $N_x$ denotes the number of participants who received the stimulus word $x$, and $n_{xy}$ denotes the number of participants who inputted
the associated word $y$ for a stimulus word $x$ with a semantic relation. Furthermore, $\delta$ denotes a factor introduced to limit the maximum value of $F(x, y)$ to 10, and $s_{xyi}$ denotes an order of the associated word $y$ by a participant $i$ for a given stimulus word $x$.

The ACD is built using quantified distances and is organized in a hierarchical structure in terms of hypernyms and hyponyms. The attribute information is used to explain the features of a given word. It also includes the action and situation concepts related to the stimulus words. It contains 1100 stimulus words and the total number of the associated words is about 280,000. It has about 64,000 different associated words when the overlapping words are not counted. The experiments have been going on for more than ten years. The total number of participants for the experiments has been increasing every year and is currently more than 5000 people. Figure 1 shows a description for the stimulus word “魚:fish” in the ACD. The second column represents the semantic relations given to the participants. The third column shows the associated words for “fish”. The numerals in <1> represent the frequencies of participants who gave the same associated word. The numerals in <2> are the averages of the participants’ association.

|         | <1> | <2> | <3> |
|---------|-----|-----|-----|
| 魚:fish | 0.60| 1.23| 1.52|
| 魚:fish hypernym 生物:living-thing | 0.16| 2.13| 3.95|
| 魚:fish hypernym 食べ物:food | 0.20| 3.00| 8.91|
| 魚:fish hypernym 物:object | 0.46| 3.10| 2.33|
| 魚:fish hyponym 鯖:sea-bream | 0.50| 3.52| 2.35|
| 魚:fish hyponym 魚:fish | 0.70| 2.54| 1.73|
| 魚:fish part/material 魚:fish | 0.44| 3.00| 2.37|
| 魚:fish part/material 目:eyes | 0.40| 4.15| 2.81|
| 魚:fish part/material 骨:bone | 0.46| 1.57| 1.92|
| 魚:fish attribute ヌルヌル:slimy | 0.18| 1.89| 3.63|
| 魚:fish attribute おいしい:delicious | 0.70| 1.14| 1.35|
| 魚:fish synonym 魚類:fishes | 0.96| 1.25| 1.12|
| 魚:fish action 泳ぐ:swim | 0.56| 2.75| 2.01|
| 魚:fish action 食べる:eat | 0.86| 1.23| 1.19|
| 魚:fish situation 海:sea | 0.58| 2.76| 1.97|

**Figure 1.** Concept dictionary descriptions for the stimulus word “fish.” Part of the associated concepts are presented here. They are grouped by the seven semantic relations and are sorted in the order of conceptual distance in each group.
order, where 1.0 is the first place of the association. The numerals in <3> expressed the calculated conceptual distances using equation (1). We have distributed at no cost a CD-ROM containing the ACD to more than fifty research organizations who sent email requests to Prof. Ishizaki, one of the authors and agreed to a simple contract.

We compared the associative concept dictionary with EDR and WordNet using the distance information described in our dictionary (Okamoto & Ishizaki, 2001). Several familiar words in daily human life were chosen to compare the three dictionaries. These words, for example, tree, car, train, and fruit, have a relatively large number of associative words. We calculated the distances between those words and their hypernyms in these three dictionaries. The results of the principal component analysis show that the conceptual structure of ACD is more similar to that of WordNet than that of EDR.

3. Summarization method using the Associative Concept Dictionary

Text summarization has conventionally been accomplished by extracting important sentences from a document based on various superficial cues. For example, in such conventional methods, the frequency of the occurrence of a given word in a document has often been used in calculating the importance scores of sentences. In this research, the Contextual Semantic Network (hereafter CSN) is developed using the ACD. A spreading activation model is used to calculate a word’s score on the CSN, where the activation values on the network are calculated using quantitative distances among the concepts.

3.1 Extraction of important sentences based on word scores

3.1.1 Important sentence extraction using CSN

In the proposed summarization method, the CSN is used to calculate the importance scores of the sentences in the input document. One cannot use only the information obtained from the word co-occurrence in its context, but must also use that from a comparatively rich network with quantitative distances and contextual information for extracting important sentences.

The following steps are the procedural details for the CNS construction
according to an input document for summarization tasks (Figure 2). These steps are used again for word sense disambiguation in Section 4.1.1.

1. The part of speech information for words (nouns, adjectives, adverbs, verbs, and so on) and the dependency structure of each sentence in an input document are obtained using Cabocha, which is a Japanese dependency structure analysis software program (Kudo & Matsumoto, 2002).

2. A CSN is constructed using the semantic relations among the words extracted from the ACD, where the information is obtained by conducting a dependency structure analysis.

   A) When an input word $w_i$ is included in the ACD as a stimulus word, a network is constructed around the $w_i$ in the CSN using the information in the ACD. A new network is constructed from the stimulus word by tracing the semantic relation paths until the accumulated distance reaches a certain numerical level.

   B) Links are added between an input word $w_i$ and the other words with dependency relations to it in the sentence.

For example, let the input sentences be “ガラパゴスにはゾウガメがいる．その亀は島の中を歩いている：There is a giant tortoise in Galapagos. This turtle is walking around the island.” In this text, “ゾウガメ: A giant tortoise” is a hyponym of “亀: a turtle”. The importance score of “a turtle” is calculated using the distance between “turtle” and “giant tortoise” in the
Jun Okamoto and Shun Ishizaki

CSN. "ガラパゴス: Galapagos" is an island and has a situation relation with "ゾウガメ: a giant tortoise". "歩く: Walk" is a verb concept of "a giant tortoise" and also that of "a turtle". We can construct an intra-document network since all the words are included in the ACD. In addition, some hypernym words are added in the CSN such as "生物: living-thing", which is a hypernym of "a turtle."

The activation value of each node is calculated using a spreading activation method on the CSN. The initial values \( V_a(0) \) of the words in sentence \( k \) of the input document are calculated using equation (2) below. Next, the activation value of each node \( N_a \) is calculated by means of equation (3).

\[
\begin{align*}
(2) & \quad V_a(0) = 1.0 \times S(j, N_a), \\
(3) & \quad V_a(t + 1) = V_a(t) - \theta V_a(t) + \sum_{b=1} \left( V_b(t)/\alpha D_{ab} \right),
\end{align*}
\]

where the calculation is repeated until \( \sum_{a=1} (V_a(t) - V_a(t + 1))^2 \) reaches a certain small value. The decay parameter \( \theta \) is assumed to be 0.1 based on our test experiments. \( V_a(t) \) is the activation value of node \( N_a \) at time \( t \). \( S(j, N_a) \) is the number of nodes (word) \( N_a \) appearing in document \( j \). \( D_{ab} \) is the distance between two concepts. \( V_b(t) \) is the activation value of the node connected to node \( N_a \). \( \alpha \) is a normalized value where the total number of links is divided by the maximum distance. \( P_{jkl} \) is the score of words \( (N_a) \) in sentence \( k \) in document \( j \). For each important sentence score \( (T_{jk}) \), the sum of the keywords-weights is divided by the number of words \( (L_{jk}) \), as shown in the following equation:

\[
(4) \quad T_{jk} = \left( \sum_{l=1}^{L_{jk}} P_{jkl} \right) / L_{jk},
\]

\( j = 1, 2, \ldots N, \quad (N: \text{No. of documents}), \)
\( k = 1, 2, \ldots M_j, \quad (M_j: \text{No. of sentences in document } j), \)
\( l = 1, 2, \ldots L_{jk}, \quad (L_{jk}: \text{No. of words in } k\text{th sentence in document } j). \)

3.1.2 Extracting important sentences based on word frequency

We use the following important sentence extraction methods to compare them with our proposed methods. The input texts are morphologically ana-
An Associative Concept Dictionary for NLP

The importance scores are calculated using the root morphemes of the nouns, adjectives, adverbs, and verbs. Pronouns and certain nouns (number, counter suffix, and so on) are not included in the calculations.

In this method, the word $w_{jkl}$ is defined as the $l$th word in the $k$th sentence in document $j$. The importance score $P_{jkl}$ of word $w_{jkl}$ is calculated using the following equation (5),

$$P_{jkl} = \sum_{k=1}^{L_j} \sum_{j=1}^{M_j} F_{jkl} \cdot \log N/n_{jkl},$$

where $F_{jkl}$ is the frequency of word $w_{jkl}$ in document $j$, $N$ is the total number of documents, and $n_{jkl}$ is the number of documents where word $w_{jkl}$ appears. The sentence scores ($T_{jk}$) are calculated using equation (4).

3.1.3 Important sentence extraction by human participants

We used eight documents extracted from Japanese elementary school textbooks because the ACD is constructed of the basic nouns in these textbooks. These documents contain about 17 sentences on average (range 10-23). All the documents focus on a single topic in natural science and consist of titles and documents.

We carried out an experiment with forty participants all of whom were native Japanese speakers and students from Keio University. We requested the participants to choose the five most important sentences from the document and to arrange them in the order of their importance in the document. The importance scores were from 5 to 1. An importance score of 5 is given to the most important sentence and a score of 1 to the fifth most important sentence. Next, the important sentences are sorted on the basis of the sum of the importance scores given to each sentence by all the participants.

Next, we calculated the degree of coincidence using Kendall’s coefficients of concordance ($W$) among the participants’ results of extracting important sentences. Kendall’s coefficients of concordance show that there was a degree of agreement among the sentences ordered by the participants. A high $W$ value means that the ordering results by the participants were consistent with each other. Table 1 shows Kendall’s coefficients of concordance ($W$) and the numbers of sentences in the eight documents.
Jun Okamoto and Shun Ishizaki

Kendall’s coefficients of concordance ($W$) of three of the documents (D3, D4, and D7) were relatively lower than the others. However, the chi-square analysis indicates that the difference is not statistically significant.

### 3.2 Evaluations and discussion

#### 3.2.1 Evaluation of our method and conventional methods

In order to show the effectiveness of the ranks of the sentences obtained by our system, we compared our results with those extracted by the participants and those automatically obtained using the conventional methods based on the word frequencies. First, importance scores from 10 to 6 were given respectively to the top five sentences that the participants chose. For a sentence chosen both by the participants and by one of the other methods, the value of correspondence ($C$) is calculated using the following equation (Okamoto & Ishizaki, 2003).

\[
C = \sum_{i=1}^{5} R_i (p, m) - \Delta r_i (p, m),
\]

where $R_i (p, m)$ is the importance score (range 10-6) of the top five sentences that the participants extracted, and $\Delta r_i (p, m)$ is a value obtained by subtracting the rank of a sentence extracted by the participants from the rank of the sentence extracted by the system. The above equation compares the ranks using the computational methods (our method and the word frequency method) with those by the participants. For example, if sentences (15, 10, 2, 6, 8) are the human extraction results and (15, 8, 10, 14, 5) are the computational ones, the value of correspondence is calculated using equation (6), such as $C = 10 + (9 - 1) + (6 - 3) = 21$. The correspondence of extracted sentences ($CS$) is the number of same sentence extractions between the human and computational method results. In the previous example, $CS$ was 3.

| Document No. | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 |
|--------------|----|----|----|----|----|----|----|----|
| No. of sentences | 15 | 18 | 22 | 21 | 9  | 9  | 19 | 18 |
| Kendall’s $W$ | 0.44 | 0.45 | 0.35 | 0.38 | 0.57 | 0.54 | 0.37 | 0.48 |

Table 1. Kendall’s coefficients of concordance ($W$) for sentence rankings

---

Kendall’s coefficients of concordance ($W$) of three of the documents (D3, D4, and D7) were relatively lower than the others. However, the chi-square analysis indicates that the difference is not statistically significant.
Table 2 lists a comparison of our method with the word frequency methods. The data in the “Our method” rows in Table 2 was obtained by comparing the top five sentences that our method chose using CSN with those chosen by the participants. The data in the “Word frequency method” rows in Table 2 was obtained by comparing the top five sentences using the word frequencies and those chosen by the participants. The CS in the table is the number of sentences from the top five sentences extracted using the automatic methods that correspond to the sentences extracted by the participants. In Table 2, the C for our method was high. The results show that this method, using the CSN, was more effective than the word frequency method. The Cs of some documents (D3, D4, and D7) are relatively lower than the ones from other documents. These results correspond with the results of Kendall’s coefficients of concordance (W), which were also relatively low (see Table 1).

4. Word sense disambiguation using the Associative Concept Dictionary

Numerous Japanese ideographs have several different meanings and pronunciations. For example, the ideograph “額,” which has two pronunciations, /hitai/ and /gaku/, in which the former represents “the forehead of the human body” and the latter “the amount of money” or “a frame”. In the association experiments, such ideographs were given as stimulus words followed by their pronunciations to avoid ambiguities.
4.1 Homographic ideogram understanding by a Dynamic Contextual Network Model

A Dynamic Contextual Network Model (hereafter DCNM) disambiguates word senses by means of a spreading activation model in the CSN. In addition, this network model is not static but dynamic, where the network structure changes depending on the context of the words in the sentences. In the proposed model, when the network is not yet rich enough to disambiguate word senses and when the model cannot decide the appropriate meaning of the homographic ideograms, it will get another input word from the input sentence sequentially and add it to the network to enhance it. Using the dynamic model we can assign appropriate word senses to ambiguous words by comparing the activation values and choosing the best one for the homographic ideogram meanings.

4.1.1 Construction of CSN for homographic ideograph

The CSN is constructed using the ACD for word sense disambiguation. The following steps (C and D) are added to steps A and B for the CSN construction using the content words in Section 3.1.1.

C) When a homographic ideograph has two meanings, we provide two nodes that correspond to each of the two meanings in the CNS. The two nodes are connected with an inhibitory link.

D) All the words associated from one of the two nodes are connected to the other node by inhibitory links. The same procedure is done to the other node, where all the associated words from it are connected to the first one using inhibitory links.

Figure 3 shows an example of a CSN. The square shape nodes are the input words in the sentence for the two meanings of the homographic ideographs and are connected with an inhibitory link. The oval shape nodes are added using the ACD and are connected with the excitatory links. The dotted lines are the inhibitory links. The associated word nodes of a homographic ideograph are connected with other homographic ideographs using the inhibitory links. “Museum” is a situational concept of “frame”. “Picture” and “face” do not exist in the sentence but are obtained from the ACD.
4.1.2 Activation value calculation in the Network

The activation value of each node is calculated based on the interactive activation model (McCelland & Rumelhart, 1981) on the CSN. We define the maximum activation level as 1.0. An initial value \( a_i(0) \) and each initial value \( a_i(t) \) for the sequential input words are calculated using the following equations.

\[
(7) \quad a_i(0) = \left(1.0 \times S_{ki}\right) / 2, \text{ and}
\]

\[
(8) \quad a_i(t) = \left(1.0 \times S_{ki} + a_i(t-1)\right) / 2,
\]

where \( 1.0 \) is the normalization value for \( S_{ki} \), which is the number of nodes \( N_i \) that appear in sentence \( k \). Next, the new activation value \( a_i(t+1) \) of each node \( N_i \) at time \( t+1 \) is calculated using the following equation.

\[
(9) \quad a_i(t+1) = a_i(t) - \theta \cdot a_i(t) + e_i(t),
\]

where the decay parameter \( \theta \) is assumed to be 0.1 and \( e_i(t) \) expresses the influence of its neighbors at time \( t \). When the neighbors of a node are active, they affect the activation value of the node by their excitatory or inhibitory connections, depending on the link between the two nodes. Those excitatory and inhibitory influences are combined using a simple equation (9) to yield a net input to the node. Thus, \( n_i(t) \) represents the net input to the node using
the following equation.

\[ n_i(t) = \sum (a_j(t) / \alpha D_{ij}), \]

where \( a_j(t) \) denotes the activation value of node \( N_j \) connected with node \( N_i \), \( \alpha \) is a constant weight, given by the total number of links of the CSN, and \( D_{ij} \) denotes the distance between two nodes, \( N_i \) and \( N_j \). In this article, the CSN is constructed by tracing the semantic relation paths while accumulating the distance until the value of 5.0 is exceeded. Therefore, the value of \( D_{ij} \) with an inhibitory link is assumed to be 5.0. When the net input is excitatory, \( n_i(t) > 0 \), the effect on the node, \( \varepsilon_i(t) \), is given by the following equation.

\[ \varepsilon_i(t) = n_i(t)[M - a_i(t)], \]

where \( M \) is the maximum activation level of the node and is set to 1.0. When the net input is inhibitory, \( n_i(t) \leq 0 \), the effect of the input on the node is given by the following equation.

\[ \varepsilon_i(t) = n_i(t)[a_i(t) - m], \]

where \( m \) is the minimum activation level of the node and is set to 0.

### 4.1.3 Enhancement of the Contextual Semantic Network

Human beings try to assign appropriate word senses to a given ambiguous word in a sentence using the contextual information located before the word; when they cannot disambiguate the word, they will use the words that follow the ambiguous word. In our model, if there is only a slight difference between the activation values of the two nodes for a given input word, it enhances the network by adding the words that follow it from the input sentence. Figure 4 shows an example of enhancing CSN depending on the words that follow it.

### 4.2 Simulation for WSD by the proposed method

Let us take as input the following Japanese sentence: “壁のピカソの絵の額”
The frame of Picasso’s picture dropped from the wall, struck my head, and my forehead bled.

In this sentence, “frame” and “forehead” are English expressions that correspond to the homographic ideograph “額” in Japanese. The input word order follows the Japanese order. At first, we construct the CSN based on the ACD for the first input word “wall”. Then, the nouns in the sentence, including the ambiguous words, are added sequentially to the network and all the nodes activate each other by the method outlined in Section 4.1.2.

Figure 5 shows the activation values of the homographic ideographs

Figure 4. Example of enhancing CSN. The activation values of the nodes correspond to the word senses of the homographic ideograms.
and the other input words in the simulation, where each input word spans twenty time cycles. The horizontal axis represents time, and the vertical axis represents the activation values. The words in the rectangles are the major words selected from the input sentence. For the homographic ideogram, “frame” has a bigger value than that of “forehead” (see broken-line circle in Figure. 5) in the first half of the sentence, so we can assign the appropriate pronunciation and meaning to “frame” as that from the Japanese ideograph “額.” However, since the homographic ideogram “forehead” has a bigger value than that of “frame” (see thick-line circle in Figure. 5) in the last half of the sentence, in this case we can assign “forehead” to the Japanese ideograph “額.”

5. Conclusion and Future Work

In this article we have proposed a simulation model for word sense disambiguation and an application system for document summarization. Both use contextual information and the quantitative distances among the concepts in the ACD. The method for summarization shows that it is better than the conventional ones and the method for disambiguation shows that it can dynamically identify the meaning of ambiguous words according to the input words.

The ACD is currently a small size dictionary compared with other concept dictionaries. We will extend it to a large-scale dictionary by automatically extracting concepts from the corpora. This extension will be useful for higher level contextual understanding systems.

The evaluation of the word sense disambiguation method using DCNM is a future project from the NLP point of view. As a preliminary evaluation, we used the multinomial Naive Bayes text classification method. We have obtained success rates of roughly 90% for both methods. We need to increase the test data to conduct a more precise evaluation.

Human word sense disambiguation processes have been observed using MEG (Magnetoencephalogram) (Ihara et al., 2008). We also have observed such brain activities using multi-channel Near Infrared Spectroscopy (NIRS). These studies will be useful for enhancing our DCNM as a model for human language understanding as well as computer systems for NLP.
NIRS is expected to be useful for the real-time analysis of human brain activity during language processing.

Acknowledgement

We wish to express our gratitude to Ms. Nao Tatsumi, a graduate student at Keio University, for making a number of helpful suggestions.

References

Baker, F. C., Fillmore, J. C., and Lowe, B. J. 1998. The Berkeley Framenet Project. Proc. of the 36th Annual Meeting of the Association for Computational Linguistics and International Conference on Computational Linguistics, 1, 86-90.

Chouekha, Y., and Lusifnan, S. 1985. Disambiguation by short contexts. Computers and the humanities, 19, 147-157.

Fellbaum, C. 1998. WordNet: An electronic lexical database. MIT Press.

EDR. 1996. EDR Electronic Dictionary Version 1.5 Technical Guide. EDR TR2-007.

Hashida, K., Ishizaki, S., and Isahara, H. 1987. Connectionist approach to the generation of abstracts, natural language generation. New Result in Artificial Intelligence Psychology and Linguistics, 149-156.

Ihara, A., Fujimaki, N., Wei, Q., Hayakawa, T., and Murata, T. 2008. MEG analyses on lexico-semantic process. Clinical Electroencephalography, 50, 9, 531-539.

Ishikawa, K., Ando, S., and Okumura, A. 2002. Evaluating text summarization using multiple correct answer summaries. Journal of NLP, 9, 4, 33-53.

Kudo, T., and Matsumoto, Y. 2002. Japanese Dependency Analysis using Cascaded Chunking. Proc. of CONLL 2002, 63-69.

McCelland, J.L., and Rumelhart, D.E. 1981. An interactive activation Model of context effects in letter perception: Part 1. An Account of Basic Findings, Psychological Rev., 88, 5, 375-407.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K., and Tengi, R. 1993. Five papers on WordNet, CSL Report 43, Cognitive Science Laboratory Princeton University.

Mochizuki, H., and Okumura, M. 2000. A comparison of summarization methods based on task-based evaluation. Proc. of LREC2000, 633-639.

Moldovan, D.I., and Novischi, A. 2004. Word sense disambiguation of WordNet glosses. Proc. of Computer Speech & Language, 301-317.

Murata, M., Utiyama, M., Uchimoto, K., Ma, Q., and Isahara, H. 2003. CRL at
Japanese dictionary based task of SENSAVAL-2 -- Comparison of various types of machine learning methods and features in Japanese word sense disambiguation --. *Journal of NLP*, 10, 3, 115-133.

Okamoto, J., and Ishizaki, S. 2001. Construction of associative concept dictionary with distance information, and comparison with electronic concept dictionary. *Journal of NLP*, 8, 4, 37-54.

Okamoto, J., and Ishizaki, S. 2003. Evaluating a method of extracting important sentences using distance between entries in an Associative Concept Dictionary. *Journal of NLP*, 10, 5, 139-151.

Okamoto, J., and Ishizaki, S. 2010. Homographic Ideogram Understanding Using Contextual Dynamic Network. *Proc. of LREC2010*, 1180-1186.

Plaza, L., Diaz, A., and Gervás, P. 2010. Automatic summarization of news using wordnet concept graphs. *IADIS International Journal on Computer Science and Information Systems*, 5, 1, 45-57.

Veronis, J., and Ide, N. M. 1990. Word sense disambiguation with very large neural networks extracted from machine readable dictionaries. *Proc. of Coling1990*, 389-394.

Vossen, P., editor. 1998. Euro WordNet. Kluwer.

Waltz, D. L., and Pollack, J. B. 1985. Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science*, 9, 51-74.

Watanabe, H. 1996. A method for abstracting newspaper articles by using surface clues. *Proc. of the 16th International Conference on Computational Linguistics*, 974-979.

Zechner, K. 1996. Fast generation of abstracts from general domain text corpora by extracting relevant sentences. *Proc. of the 16th International Conference on Computational Linguistics*, 986-989.