

意味空間のスケール変換による動的シソーラスの実現

小嶋 秀樹 ・ 伊藤 昭

郵政省 通信総合研究所 関西先端研究センター

〒 651-24 神戸市西区岩岡町岩岡 588-2

{xkozima, ai}@crl.go.jp

あらまし 本論文では、与えられた単語集合 C から文脈 (連想方向) を抽出し、それにもとづいて英単語 w, w' 間の距離 $d(w, w')$ を計算する手法を提案する。本手法では、「意味空間のスケール変換」によって、文脈の抽出とそれにもとづく距離計算を行なう。意味空間の中では、語彙 V の各単語 w が Q ベクトルとよばれる m 次元ベクトルで意味表現される。この Q ベクトルは、英語辞書にもとづいて生成した 2,851 次元の P ベクトルを主成分分析することによって得られる。本手法によって、与えられた文脈 C に動的に適応した意味空間が得られ、文脈依存的な単語間の距離を計算することができた。先行テキストを文脈とした後続単語の予測をととして本手法を評価し、与えられた文脈をよく捉えていることを確認した。

キーワード 意味距離, 単語類似度, 連想, シソーラス, 文脈処理.

Adaptive Scaling of a Semantic Space

Hideki Kozima and Akira Ito

Kansai Advanced Research Center
Communications Research Laboratory

Iwaoka 588-2, Iwaoka-cho, Nishi-ku,

Kobe 651-24, JAPAN

{xkozima, ai}@crl.go.jp

Abstract This paper proposes a computationally feasible method for measuring context-sensitive semantic distance between words. The distance is computed by adaptive scaling of the semantic space where each word in the vocabulary V is represented by a multi-dimensional Q -vector. The vectors are obtained through a principal component analysis on the P -vectors generated from an English dictionary. Given a word set C which specified a context, each dimension of the semantic space is scaled up or down according to the distribution of C on the dimension. In the scaled space, word distance becomes sensitive to the context which comes out from C . An evaluation through word prediction shows that the proposed distance works well as an context-sensitive word distance.

key words semantic distance, word similarity, association, thesauri, context processing.

1 はじめに

単語間の意味的な距離 (または類似度) は、自然言語処理や情報検索などの多くの分野で必要とされている。テキストを理解するには、それぞれの単語や文を解釈するための文脈を捉えることが必要であり、そのためには、全体的なテキスト構造 (text structure) [Grosz and Sidner, 1986] を捉えることが必要となる。このテキスト構造を捉えるためには、単語間の意味的な距離がよい手がかり (ボトムアップ情報) となる [Kozima and Furugori, 1994]。

いまでも多くの研究が、単語間の意味的な距離の計算を試みてきた。その先駆的研究である Osgood [1952] の意味微分 (semantic differential) は、心理実験にもとづいて単語の意味をベクトル化するものであった。最近の計算言語学における研究では、より再現性のある手法が提案されている。たとえば、Morris and Hirst [1991] は、Roget のシソーラスにもとづいて単語間の意味関係を検出している。Brown, *et al.* [1992] は、コーパスにおける共起性にもとづいて語彙をクラスタリングしている。Kozima and Furugori [1993] は、英語辞書から構成された意味ネットワーク上の活性伝播によって単語間の類似度を計算している。

これら従来の研究では、単語間の距離を文脈独立で静的なものとして扱っている。しかし、単語間の距離は、文脈依存で動的なものとして扱うべきである。なぜなら、ある単語から関連単語を連想する場合、どのような文脈を想定しているかによってその結果が異なるからである。たとえば、car に関連した単語はつぎの2つの文脈に沿って連想されるだろう。

car → bus, taxi, railway, ...

car → engine, tire, seat, ...

前者は「乗り物」、後者は「車の部品」という文脈 (連想方向) をもっている。たとえ文脈なしの自由連想でも、人間は何らかの文脈を想定した上で関連単語を連想していると思われる。

本論文では、このような連想の文脈依存性を英単語間の距離に反映させ、与えられた文脈に応じて動的に英単語間の距離を計算する手法を提案する。文脈は、複数の単語からなる集合 C によって与えることができる。上にあげた例は、与えられた単語集合 C から抽出した文脈にしたがって、その C に関連する単語を連想することとして、つぎの

ように書き直すことができる。

$C = \{\text{car, bus}\}$

→ taxi, railway, airplane, ...

$C = \{\text{car, engine}\}$

→ tire, seat, headlight, ...

一般に、 C を変化させることによって単語間の距離が変化する、その結果として連想される単語が変わる。そこで、本論文で扱う問題をつぎのように定式化する。

「手がかりとして与えられた単語集合 C から文脈 (連想方向) を抽出し、それにもとづいて任意の2単語 w, w' 間の距離 $d(w, w')$ を計算する。」

つまり、手がかり C を最もよく説明する文脈を見つけ、それを想定した上で単語間の距離 $d(w, w')$ を計算するというものである。

文脈の抽出とそれにもとづく距離計算を実現するために、我々は「意味空間のスケール変換」という手法を考えた。以下、つぎのような順序で本手法を説明していく。第2節では、意味空間をどのように構成するのかを説明する。この意味空間は、単語の意味が多次元ベクトルとして表現されるベクトル空間である。第3節では、意味空間のスケール変換について説明する。与えられた単語集合 C から文脈を抽出し、それにもとづいて意味空間のスケールを拡大・縮小するというものである。第4節では、本手法による距離計算の結果をいくつか例示する。第5節では、本手法の客観的評価を試みる。ここでは、先行するテキストを読み、後続するテキストに出現する単語を予測するという評価方法をとる。第6節では、ここで提案した手法をまとめ、今後の課題・展望などを述べる。

2 意味空間の構成

語彙 V の各単語は、意味空間における多次元の意味ベクトル「Qベクトル」として表現される。Qベクトルを生成するために、まず語彙 V の各単語について、2,851次元の「Pベクトル」を生成する。Pベクトルは、英語辞書から構成された意味ネットワーク上の活性伝搬 [Kozima and Furugori, 1993] によって生成される。このPベクトルについて主成分分析を行なうことによって、より少ない次元数をもつQベクトルが得られる。

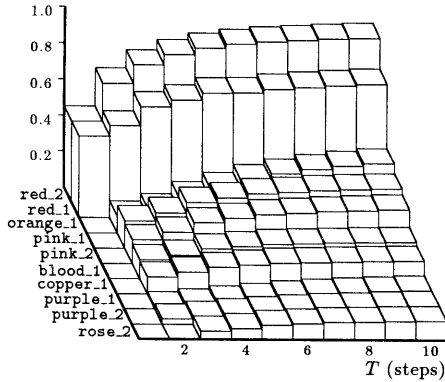


図1 意味ネットワークによる意味ベクトルの生成例
(見出し語 $w = \text{red}$ を活性化させたとき、活性化度上位10節点の活性化度分布を記録したもの。 $T = 10$ の活性化度分布を P ベクトル $P(w)$ とする。)

2.1 意味ネットワークによる P ベクトルの生成

語彙 V の各単語から P ベクトルへの変換には、英語辞書 LDOCE (*Longman Dictionary of Contemporary English*) から規則的に構成された意味ネットワークを利用する。また、ここで扱う語彙 V は、LDOCE の定義用語彙 LDV (*Longman Defining Vocabulary*) の 2,851 語とする。この意味ネットワークは、 $V (= \text{LDV})$ の各単語に対応する 2,851 個の節点と、それらの LDOCE における語義定義から得られる 295,914 本のリンクからなる。各節点は活性化をもつことができ、その活性はリンクをととして隣接する節点に伝播していく。

単語 w の P ベクトル $P(w)$ は、意味ネットワーク上の節点 w から活性を伝播させることによって計算される。節点 w を一定時間だけ活性化させることによって、その活性がネットワーク上に伝播し、各節点が w との意味的な連想度に応じた活性化度をもつようになる (図1 参照)。活性化が終了した時点 (実験では 10 ステップ目) での各節点 w_i ($i = 1, \dots, 2851$) の活性化度を $a(w_i)$ とすれば、目的とする P ベクトルは、第 i 成分を $a(w_i)$ とする 2,851 次元ベクトルである。

ここでは語彙 $V = \text{LDV}$ (2,851 語) としたが、実際には辞書 LDOCE のすべての見出し語 (約 56,000 語) について、2,851 次元の P ベクトルを計算することができる。なぜなら、LDOCE のすべての見出し語は LDV だけを使った語義定義を与えられているからである。LDV に含まれ

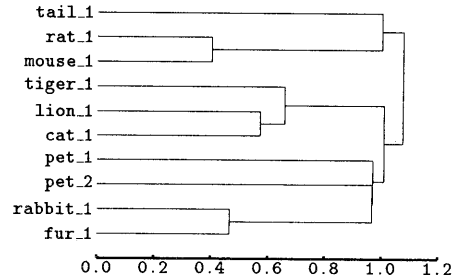


図2 P ベクトル間の階層クラスタリングの結果。
(クラスタ内の P ベクトルの重心間のユークリッド距離によるデンドログラムの一部分。)

ない単語の P ベクトルは、その語義定義に現われる各単語 ($\in \text{LDV}$) を同時に活性化させることによって計算すればよい。意味ネットワークと意味ベクトルの計算についての詳細は [Kozima and Furugori, 1993] を参照されたい。

P ベクトルの特徴をつかむために、 $\langle w, w' \rangle \in V^2$ について $P(w), P(w')$ 間のユークリッド距離を計算し、階層クラスタリング (重心法) を行なった結果を図2にあげる。得られたデンドログラムには、人間の直感に合った類似性 (rat/mouse や tiger/lion/cat など) が見られる。しかしながら、 P ベクトル間のユークリッド距離は文脈独立で静的なものである。これを文脈依存で動的なものにすることが、本研究の目標である。

2.2 主成分分析による Q ベクトルへの変換

主成分分析をととして、 P ベクトル $P(w)$ をより少ない次元数をもつ Q ベクトル $Q(w)$ に変換する。まず、 P ベクトルに新しい直交座標軸を与える。新しい座標軸は、つぎの3条件を満たすベクトル $X_1, X_2, \dots, X_{2851}$ によって与えられる。

1. $|X_i| = 1$.
2. X_i, X_j が直交 ($j \neq i$).
3. P ベクトル $P(w_1), P(w_1), \dots, P(w_{2851})$ を X_i に射影したときの分散を v_i とするとき、
$$v_1 \geq v_2 \geq \dots \geq v_{2851}.$$

言い換えれば、 P ベクトルの射影分散が最大となる軸が X_1 となり、それに直交して射影分散が最大となる軸が $X_2, \dots$ となる。

つぎに、得られた座標軸から $X_1, \dots, X_m$ ($m \ll 2851$) を取り出し、 m 次元の新しい意味空間を構成する。ただ

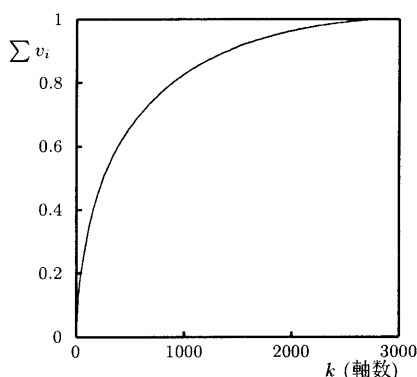


図3 新しい座標軸上のPベクトルの累積分散
(主成分軸 X_i 上にPベクトルを射影したときの分散 v_i の累積 $\sum_{i=1,k} v_i$ をグラフにしたもの.)

し、新しい座標系の原点は、Pベクトル $P(w_1), P(w_2), \dots, P(w_{2851})$ の平均 (重心) とする。座標軸 X_i におけるPベクトルの射影分散 v_i は、 X_i によって表現されている情報量 (つまり第 i 主成分軸の寄与率) を表わしている。図3に示すように、 $X_1, \dots, X_{100}$ の100次元だけで全情報の31.65%を表現でき、500次元で66.21%、1000次元で82.80%を表現できる。

最後に、各Pベクトル $P(w_i)$ を m 次元の新しい意味空間内のQベクトル $Q(w_i)$ に写像する。 $Q(w_i)$ の第 j 成分は、 $P(w_i)$ を座標軸 X_j に射影した値である。このようにして、各単語 w_i の意味は m 次元のQベクトル $Q(w_i)$ によって表現されることになる。我々は、このQベクトルによって表現される情報量のなかでノイズが占める割合を最小化することによって、 $m = 281$ という最適値を得た。ただし、ここでいうノイズとは、すべてのQベクトルによって表現される情報量のなかで、210語ある機能語 ($\in V$) が占める部分であるとした。第4節と第5節で報告する実験結果は、 $m = 281$ としたときのものである。

3 意味空間のスケール変換

Qベクトルの意味空間について「スケール変換」を行なうことによって、単語間の距離を文脈依存的かつ動的に計算することができる。Qベクトル間の単純なユークリッド距離は、Pベクトル間のユークリッド距離 (図2参照) と大きな違いはなく、ともに文脈独立で静的なものである。Qベクトルの意味空間を、与えられた文脈に適応するように

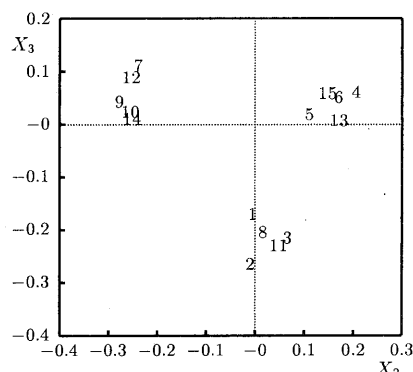


図4 部分意味空間での単語の分布
(主成分軸 X_2, X_3 によって張られる平面上に、本文中にあげた15単語 {1. after, 2. ago, ..., 15. vehicle} を散布したもの.)

スケール変換することによって、Qベクトル間の距離を文脈依存で動的なものにすることができる。

Qベクトルの意味空間からいくつかの次元を適切に選んで部分空間を構成することによって、意味的に関連した単語をクラスタにまとめることができる。たとえば、つぎにあげる15個の単語について考える。

1. after, 2. ago, 3. before, 4. bicycle,
5. bus, 6. car, 7. enjoy, 8. former, 9. glad,
10. good, 11. late, 12. pleasant, 13. railway,
14. satisfaction, 15. vehicle.

部分意味空間 (平面) $X_2 \times X_3$ のなかで、これらの単語は3つのクラスタ (「過去」・「乗り物」・「楽しさ」) に分離される (図4参照)。「乗り物」と「楽しさ」のクラスタは、 X_2 だけでは分離されないこともわかる。ここでは2次元の部分意味空間を例として取り上げたが、関連単語をクラスタ化するには一般により多くの次元が必要となる。

与えられた単語集合 C をクラスタ化するための部分意味空間の次元は、どのように選べばよいだろうか。ひとつの方法は、 C を射影したときの分散 (または標準偏差) が小さくなる次元を選択することである [小嶋・伊藤, 1995]。ここではそれを改良し、次元を選択する/しないという1/0の選択ではなく、各次元を距離計算に反映させる強さを「スケール変換」によって調節する。すなわち、

「単語集合 C が各次元 (座標軸) X_i 上でクラスタ化するように、 X_i のスケールを拡大または縮小する。」

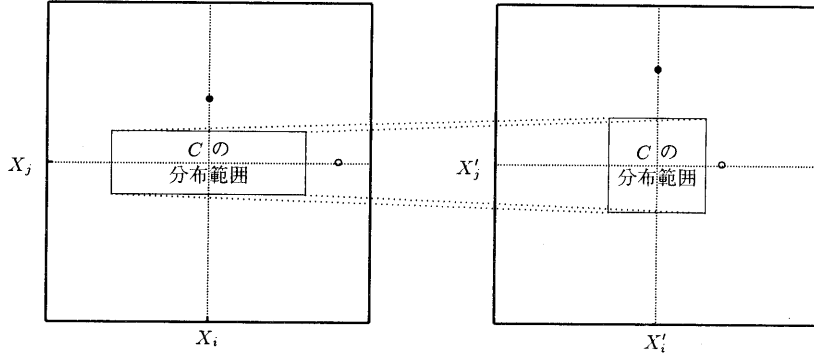


図5 意味空間のスケール変換の原理

もとの意味空間 $X_i \times X_j$ で偏平な分布範囲をもっていた C は、新しい意味空間 $X'_i \times X'_j$ で正方形や球に似た分布範囲をとるようになる。その結果、 C と単語 $\bullet/\circ$ との距離関係が変化する。）

こうしてスケール変換された意味空間のなかで、 C はクラスタをなす。また、任意の単語間の距離も、この意味空間のなかで計算できる。たとえば、図5に図示するように、 C が意味空間内で偏平な分布範囲をもつとする。このとき C が見かけ上のクラスタを形成するように各次元のスケールを調節するわけである。

意味空間のスケール変換にもとづく単語 w, w' 間の距離 $d(w, w')$ は、 $Q(w) = \{q_1, \dots, q_m\}$, $Q(w') = \{q'_1, \dots, q'_m\}$ とするとき、つぎのように計算される。

$$d(w, w') = \sqrt{\sum_{i=1, m} (f_i \cdot q_i - f_i \cdot q'_i)^2},$$

ここで、 f_i は第 i 次元（座標軸 X_i ）のスケール変換率であり、つぎのように計算される。

$$f_i = \begin{cases} 1 - r_i & (r_i \leq 1) \\ 0 & (r_i > 1), \end{cases}$$

$$r_i = SD_i(C)/SD_i(V),$$

ただし、 $SD_i(C)$ は単語集合 C を座標軸 X_i 上に射影したときの標準偏差、 $SD_i(V)$ は語彙 V についての標準偏差とする。もし、 C が X_i 上でまとまりのよいクラスタをなす（つまり $r_i \approx 0$ ）ならば、距離計算における X_i の重みは大きくなる（ $f_i \approx 1$ ）。逆に、 C が X_i 上でクラスタをなさない（つまり $r_i \gg 0$ ）ならば、距離計算における X_i の重みは小さくなる（ $f_i \approx 0$ ）。

この「意味空間のスケール変換」の計算コストは小さい。計算処理の全体は、おおまかに言って、初期処理・適

応処理・距離計算の3つの部分に分けられる。初期処理は、Pベクトルの生成（意味ネットワーク上の活性伝播）とQベクトルへの変換（主成分分析など）からなり、計算コストは大きい。しかしながら、ひとたび初期処理を済ませてしまえば、適応処理（与えられた単語集合 C にもとづく意味空間のスケール変換）は、 $f_1, \dots, f_m$ を計算するだけである。また、距離計算はふつうのユークリッド距離の計算とほとんど変わらない。

4 単語間の距離の計算例

意味空間のスケール変換による単語間の文脈依存的な距離計算をいくつか例示する。ここでは計算結果を直感的に捉えられるように、つぎのような問題を扱う。

「与えられた単語集合 C にもとづいて意味空間のスケール変換を行ない、語彙 V の各単語 w について C の各単語との平均距離

$$d_C(w) = \frac{1}{|C|} \sum_{w' \in C} d(w, w')$$

を計算する。」

距離 $d_C(w)$ が最も小さくなるものから k 個の単語を取り出し、単語集合 $C^+(k)$ とする。この $C^+(k)$ は、与えられた文脈 C に関連する単語からなるはずである。

文脈として $C = \{\text{bus, car, railway}\}$ を与えた場合、つぎのような関連単語 $C^+(12)$ が得られた。（各単語 w の距離 $d_C(w)$ も併せて示す。）

w	$d_C(w)$
car_1	0.10361
bus_1	0.10361
motor_1	0.17944
carriage_1	0.19403
motor_2	0.23213
passenger_1	0.26658
vehicle_1	0.26806
garage_1	0.31450
road_1	0.33824
inside_3	0.36867
wheel_1	0.37124
engine_1	0.38923

たとえば carriage や vehicle といった単語がみられるように、「乗り物」という文脈をよく捉えている。(表中の単語つけられた添字は、同形語 (homograph) を区別するためのものであり、辞書 LDOCE の見出し語につけられた添字と同じものである。)

一方、同じ bus を含む文脈 $C = \{\text{bus, ticket, tour}\}$ を与えた場合、つぎのような関連単語 $C^+(12)$ が得られた。

w	$d_C(w)$
bus_1	0.10053
ticket_1	0.10547
tour_2	0.11608
tour_1	0.13137
half_1	0.15715
abroad_1	0.16222
passenger_1	0.16306
make_2	0.16704
timetable_1	0.17118
garage_1	0.17145
part_1	0.17227
tourist_1	0.17557

この場合、文字どおり「バス旅行」に関する単語が連想されているのがわかる。

また、 $C = \{\text{read, book, school}\}$ とした場合、つぎのような関連単語 $C^+(10)$ が得られた。(距離 $d_C(w)$ の昇順にソートしてある。)

book_1, read_1, school_1, student_1,
university_1, education_1, someone_1,
something_1, print_2, that_1.

関連単語として student や education などがみられる。一方、上と同じように read を含む $C = \{\text{read, memory, machine}\}$ を与えた場合、つぎのような関連単語 $C^+(10)$ が得られた。

machine_1, memory_1, read_1, computer_1,
remember_1, someone_1, have_2, that_1,
instrument_1 feeling_2.

この場合は、computer といった単語がみられる。

5 後続単語の予測による評価

我々がテキストを読むとき、それまでに読んだ先行テキストの文脈にもとづいて、つぎにどんなテキストがつづくのかを予測する。この予測によって、記憶の探索空間をせばめることができ、照応・省略・曖昧性といった問題を解決する手がかりが得られる。これを裏返せば、後続単語をうまく予測できるならば、先行する文脈をよく捉えていると考えられる。

ここでは、提案した単語間の距離を評価するために、先行するテキストにもとづいて、後続するテキストの予測を試みる。すなわち、

「テキストの各文に現われる単語を、それに先行する n 文から予測する」

という課題を扱う。つまり、テキストを文の列 $S_1, \dots, S_N$ とするとき、ターゲット文 S_i ($i = n+1, \dots, N$) について、 $C = \{S_{i-n} \dots S_{i-1}\}$ という単語集合 (実際は重複をゆるす集合) を用意し、この C を文脈として意味空間のスケール変換を行なう。ただし、ここで扱う語彙 V' は、全語彙 V から機能語集合 F を除いたもの (つまり内容語) であるとする。

先行する n 文によって後続文に現われる単語をどのくらい予測できているのかを、つぎのような予測誤差として評価する。すなわち、ターゲット文 S_i の各単語 w が語彙 V' のなかで $r(w)$ 番目に小さな $d_C(w)$ をとるとする。このとき、

$$E_i = \frac{1}{|S_i|} \sum_{w \in S_i} \frac{r(w)}{|V'|}$$

をターゲット文 S_i の予測誤差として、テキストの各文 $S_1, \dots, S_N$ の予測誤差 E_i を計算する。 S_i がうまく予測されていれば、 S_i の各単語 w について C との距離 $d_C(w)$ が小さくなり、 E_i は 0 に近づく。

短編小説 (O. Henry: *Springtime à la Carte* [Thornly, 1960]) について、予測誤差 E_i の評価を行なった。このテキストは 1,620 語 / 110 文からなる (平均 14.7 語 / 文)。先

行テキストの文数 $n = 1, \dots, 8$ について, E_i の平均 $\bar{E}$ はつぎのようになった。

n	$\bar{E}$
1	0.324792
2	0.183826
3	0.162266
4	0.160213
5	0.163533
6	0.169595
7	0.174895
8	0.180140

予測を全く行なわなかった場合 ($n = 0$) の $\bar{E}$ の期待値は 0.5 である。また、単語が出現する先見確率を West [1953] による単語頻度調査をもとに計算し、それにもとづいて予測を行なった場合, $\bar{E} = 0.22364$ であった。上の評価実験では、この先見確率を利用していないにもかかわらず, $n = 1$ の場合をのぞいて $\bar{E}$ はより小さくなっている。

先行テキストが大量にあるほど (つまり文数 n が大きくなるほど) 後続単語の平均予測誤差 $\bar{E}$ が小さくなると考えられるが、実際には $\bar{E}$ を最小にする n が存在する。これは、テキストのもつ文脈がテキストの各位置で均一でないためであると考えられる。つまり、テキストの場面 (意味的な段落) が変わると、文脈 C によるターゲット文 S_i の予測誤差が大きくなるためである。16 人の被験者による場面分割を観察した結果 [Kozima, 1993; Kozima and Furugori, 1994] によれば、この評価実験に用いたテキストは 21 個の場面からなり、1 場面あたり平均 5.24 文からなる。この値は、評価実験での n の最適値 4 と何らかの関係があると思われる。

6 おわりに

本論文では、意味空間のスケール変換によって単語間の距離を文脈依存的かつ動的に計算する手法を提案した。語彙 V の各単語 w の意味は、この意味空間における Q ベクトル (実験では 281 次元) によって表現される。 Q ベクトルは、 P ベクトルについての主成分分析をとおして得られる。 P ベクトルは、英語辞書 (LDOCE) から規則的に構成された意味ネットワーク上の活性伝播によって得られる 2,851 次元ベクトルである。

意味空間のスケール変換は、文脈として与えられた単語集合 C に意味空間を適応させる手法である。すなわち、 C がクラスタを形成するように意味空間の各次元のスケールを C の分布特徴に合わせて拡大 / 縮小させる。スケール

変換後の意味空間では、任意の 2 単語 w, w' 間の距離 $d(w, w')$ が、 C の分布特徴 (つまり文脈) を反映したものとなる。先行テキストによる後続単語の予測を行なった結果、この単語間の距離が先行テキストの文脈をよく捉えていることを確かめた。

本手法では、意味ベクトルを英語辞書における単語間の参照関係から生成しているが、コーパスから抽出した単語の共起関係から生成することもできる。たとえば、相互情報量 [Church and Hanks, 1990] や N-gram 統計から (スパース性の問題があるが) 意味ベクトルを生成することができる。しかし、Niwa and Nitta [1994] が指摘するように、辞書から生成した意味ベクトルとコーパスをもとに生成した意味ベクトルは、互いに性質の異なった情報をもつと考えられる。これら 2 つの知識源を相補的に利用することによって、よりよく文脈を捉えることができると予想している。

人間のとくに優れた能力として、「情報の重要な部分に注目し、ほかの部分には注意を払わない」という能力や、「刻々と変化する環境に応じて、注意の方向を変えていく」という能力がある。このような「文脈」を捉える能力によって、人間は少ない認知コストで準最適解を求めることができ、いわゆる「フレーム問題」を回避することができる。ここで提案した意味空間のスケール変換は、このような能力の一部分を自然言語処理の立場から定式化したものといえる。たとえば、文脈として与えられる単語集合 C を短期記憶 (7 ± 2 チャンク) の内容とみなすことによって、環境の変化に自発的に適応していく能力をモデル化できるかもしれない。

今後の課題として、まず (1) 最適な次元数 m の計算やスケール変換率 f_i の計算を、より妥当性のあるものにする。ことと、(2) 後続単語の予測に、単語が出現する先見確率を導入すること、(3) コーパスにおける単語の共起関係から意味ベクトルを生成し、本手法を適用することなどがあげられる。将来的には、本手法を人間の「注意」のモデルに一般化し、変化する環境に適応する能力や、つぎに何が起こるのかを予測する能力などを説明できるようにしたい。

参考文献

[Church and Hanks, 1990] K. W. Church and P. Hanks : Word association norms, mutual information, and lexi-

cography, *Computational Linguistics*, Vol.16, pp.22-29.

[Brown *et al.*, 1992] P. F. Brown, V. J. Della Pietra, P. V. deSouza, J. C. Lai, and R. L. Mercer : Class-based n -gram models of natural language, *Computational Linguistics*, Vol.18, pp.467-479.

[Grosz and Sidner, 1986] B. J. Grosz and C. L. Sidner : Attention, intentions, and the structure of discourse, *Computational Linguistics*, Vol.12, pp.175-204.

[Kozima and Furugori, 1993] H. Kozima and T. Furugori : Similarity between words computed by spreading activation on an English dictionary, in *Proceedings of 6th Conference of the European Chapter of the Association for Computational Linguistics (EACL-93, Utrecht)*, pp.232-239.

[Kozima, 1993] H. Kozima : Text segmentation based on similarity between words, in *Proceedings of 31st Annual Meeting of the Association for Computational Linguistics (ACL-93, Ohio)*, pp.286-288.

[Kozima and Furugori, 1994] H. Kozima and T. Furugori : Segmenting narrative text into coherent scenes, *Literary and Linguistic Computing*, vol.9, No.1, pp.13-19.

[小嶋・伊藤, 1995] 小嶋 秀樹, 伊藤 昭 : 辞書にもとづいて語彙をクラスタリングする試み, 言語処理学会第1回年次大会, pp.205-208.

[Morris and Hirst, 1991] J. Morris and G. Hirst : Lexical cohesion computed by thesaural relations as an indicator of the structure of text, *Computational Linguistics*, Vol.17, pp.21-48.

[Niwa and Nitta, 1994] Y. Niwa and Y. Nitta : Co-occurrence vectors from corpora vs. distance vectors from dictionaries, in *Proceedings of 15th International Conference on Computational Linguistics (COLING-94, Kyoto)*, pp.304-309.

[Osgood, 1952] C. E. Osgood : The nature and measurement of meaning, *Psychological Bulletin*, Vol.49, pp.197-237.