

文脈依存的に単語間の意味距離を計算する一手法

小嶋 秀樹[†] 伊藤 昭[†]

本論文では、単語間の意味距離を文脈依存的に計算する手法を提案する。各単語は、英語辞書から抽出された多次元ベクトルとして、意味空間と呼ばれるベクトル空間における点に写像される。文脈から独立した意味距離は、このベクトル間の距離として計算すればよい。文脈に依存した意味距離は「意味空間のスケール変換」によって計算する。文脈の手がかりとして単語（キーワードなど）の集合が与えられると、この単語集合が均整のとれた分布を持つように、意味空間の各次元のスケールを拡大・縮小する。このスケール変換によって、意味空間における任意の2単語間の距離は与えられた単語集合の意味的な分布に依存した値となる。先行テキストに基づく後続単語の予測によって本手法を評価した結果、本手法が先行テキストの文脈をよくとらえていることを確かめた。

A Context-sensitive Measurement of Semantic Word Distance

HIDEKI KOZIMA[†] and AKIRA ITO[†]

This paper proposes a computationally feasible method for measuring context-sensitive semantic distance between words. The distance is computed by adaptive scaling of a semantic space. In the semantic space, each word in the vocabulary is represented by a multi-dimensional vector which is extracted from an English dictionary. Given a word set C which specifies a context, each dimension of the semantic space is scaled up or down according to the distribution of C in the semantic space. In the semantic space thus transformed, distance between words becomes dependent on the semantic distribution of C . An evaluation through a word prediction task shows that the proposed measurement successfully extracts the context of a text.

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ためにはテキスト構造¹⁾をとらえることが必要となるが、単語間の意味距離はそのためのよい手がかり（ボトムアップ情報）となる。また、情報検索ではテキスト間の類似度をとらえることが必要となるが、単語間の意味距離はその1次近似を与える。

いままでも数多くの研究が単語間の意味距離の計算を試みてきた。その先駆的研究である Osgood²⁾の意味微分 (semantic differential) は、多数の被験者に対する心理実験を通して単語の意味を計量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、より客観的で再現性のある手法が提案されている。たとえば、Morris³⁾は Roget のシソーラスに基づいて単語間

の意味関係の有無を判定している。Brown⁴⁾はコーパスにおける共起性に基づいて語彙をクラスタリングしている。また、Kozima⁵⁾は英語辞書から構成し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、単語間の意味距離を文脈から独立した静的なものとして扱っている。しかし実際には、単語間の意味距離は文脈に依存して動的に変化する。たとえば単語 **car** からは、次のように、異なる文脈ごとに異なる関連語が連想される。

- **car** → **bus, taxi, railway, ...**
- **car** → **engine, tire, seat, ...**

前者は「乗物」としての機能、後者は「機械」としての構造を文脈としている。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うえで関連語を連想していると考えられる。

本論文では、単語間の意味距離を文脈依存的に計算することを試みる。ここでは、文脈の手がかりとして単語（キーワードなど）の集合を与える。たとえば「乗物」という文脈ならば、単語集合 {**bus, taxi**} を与え

[†] 郵政省通信総合研究所関西先端研究センター
Kansai Advanced Research Center, Communications
Research Laboratory, MPT

る。すると、上にあげた例は、次のように書き換えられる。

- car/{bus, taxi}
→ railway, bicycle, airplane, ...
- car/{engine, tire}
→ seat, wheel, headlight, ...

一般に、文脈が変われば単語間の意味距離も変わり、その結果として連想される関連語が変わる。そこで本論文で扱う問題を次のように定式化する。

与えられた単語集合 C から文脈を抽出し、それに基づいて任意の 2 単語 w, w' 間の意味距離 $d(w, w' | C)$ を計算する。

文脈の抽出とそれに基づく意味距離の計算のために、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、次のような順序で本手法を説明していく。2章では「意味空間」をどのように構成するのかを説明する。意味空間とは、単語間の意味距離を計量化するためのベクトル空間である。Salton ら⁶⁾は、情報検索の研究の中で、テキストを語句の出現特徴 (tf-idf など) を表すベクトルに写像し、テキスト間の意味距離を計量化している。本手法でも同じようにベクトル空間を利用して単語間の意味距離を計量化する。ただし、各単語を英語辞書から抽出した多次元ベクトルとして意味空間内に写像し、単語間の意味距離をこのベクトル間の距離として計算する。3章では、意味空間の「スケール変換」について説明する。文脈の手がかりとして単語集合 C が与えられると、意味空間における C の分布特徴として文脈を抽出する。この分布特徴に基づいて意味空間の各次元を拡大・縮小し、 C が意味空間内で均整のとれた分布をとるようにする。この結果、意味空間における任意の 2 単語間の距離は C の意味的な分布特徴を反映した値となる。4章では、本手法による意味距離の計算結果をいくつか例示する。5章では、先行テキストに基づいて後続単語を予測することを通して、本手法の客観的評価を試みる。6章では、本手法をまとめ、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、図 1 に示す操作によって、多次元ベクトル「Q ベクトル」に変換される。まず、英語辞書を知識源として、各単語を「P ベクトル」という 2851 次元の実数ベクトルに変換する。次に、この P ベクトルの集合について主成分分析を行い、各 P ベクトルをより次元数の小さな Q ベクトルに変換する。この Q ベクトルの空間が、単語間の意味距離を計算

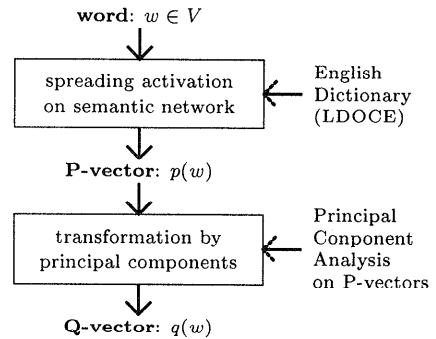


図 1 単語から Q ベクトルへの変換
Fig.1 Mapping words onto Q-vectors.

するための意味空間となる。

2.1 単語から P ベクトルへの変換

語彙 V の各単語 w を、意味ネットワーク上の活性伝播によって、P ベクトル $p(w)$ に変換する。ここではその概要を述べるが、意味ネットワークと活性伝播の詳細については、文献 5) および 7) を参照されたい。

この変換に利用する意味ネットワークは、英語辞書 LDOCE (*Longman Dictionary of Contemporary English*) から規則的に構成され、2851 語の定義用語彙 LDV (*Longman Defining Vocabulary*) の各単語に対応する 2851 個の節点と、それらの LDOCE における語義定義から得られる 295914 本のリンクからなる。また、各リンクには単語の使用頻度から計算された重みがつけられている。LDOCE のすべての語義定義は LDV だけを用いて記述されているため、この意味ネットワークは LDV の各単語を LDV で説明する閉じた体系となっている。

意味ネットワークの各節点は活性度を持つことができ、その活性は重み付きのリンクを通して隣接する節点に伝播していく。ある節点 n を一定時間にわたって活性化させることによって、その活性が意味ネットワーク上に伝播し、各節点は n との語義定義上の連想度に応じた活性度を持つようになる。図 2 は、単語 red に対応する節点を 10 ステップの時間にわたって活性化させたとき、最も強く活性化した 10 節点*の活性度の時間的変化を記録したグラフである。

単語 w の P ベクトル $p(w)$ は、 w に対応する節点を活性化させることによって得られる意味ネットワーク上の活性度分布である。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 n_i ($i = 1, \dots, 2851$) の活性度を a_i ($0 \leq a_i \leq 1$) とすれば、 $p(w)$ は a_i を

* 節点名の末尾につけた数字によって品詞の違いを区別している。

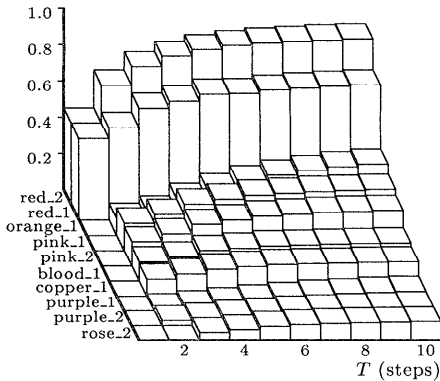


図2 活性伝播によるPベクトルの生成例

Fig. 2 Generating a P-vector by spreading activation.

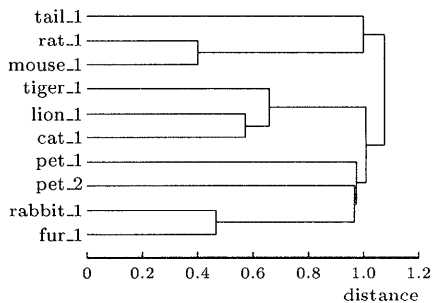


図3 Pベクトルの階層クラスタリング

Fig. 3 Hierarchical clustering of P-vectors.

第 i 成分とする 2851 次元の実数ベクトルとなる。

本論文では、対象とする語彙 V を 2851 語の LDV とした。LDV は比較的小さな語彙ではあるが、LOB コーパス (Lancaster-Oslo/Bergen Corpus) の 83.07% をカバーする*。また、 V を LDOCE のすべての見出し語 (約 56000 語) に拡張することも容易である。なぜなら、LDV に含まれない見出し語についても、その語義定義に現れる各単語 (LDV の要素) を同時に活性化させることによって、P ベクトルの生成が可能だからである。

意味的に類似した単語は、類似した P ベクトルを持つ。P ベクトル $p(w)$ は、 w からみた LDV の各単語の語義定義上の連想度を表す。単語 w, w' が意味的に類似していれば、それぞれから連想される単語は類似したものになり、P ベクトル $p(w), p(w')$ 間の距離は小さくなる。図 3 は、すべての単語対 $\{w, w'\} \in V^2$ 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) を行った結果の一部分である。このデンドログラムから、P ベクトル間

の距離が、*rat/mouse* や *tiger/lion/cat* など、単語間の意味的な距離をとらえていることが分かる。しかしながら、この距離は文脈から独立した静的なものである。この距離を文脈に依存した動的なものにすることが、本研究の目的である。

2.2 P ベクトルから Q ベクトルへの変換

主成分分析を通して、2851 次元の P ベクトルをより次元数の小さな Q ベクトルに変換する。与えられた単語集合 C の文脈は、意味空間の各次元ごとに、 C の分布特徴としてとらえられる。しかし、各節の活性化度を表す P ベクトルは強いスパース性**を持っているため、P ベクトルの各次元ごとに C の分布特徴をとらえることは難しい。そこで、各次元が C の分布特徴を効率良く表現するように、主成分分析を通して 2851 次元の P ベクトルをより次元数の小さな Q ベクトルに変換する。まず、各単語 $w \in V$ の P ベクトル $p(w)$ からなる集合 P について主成分分析を行い、すべての主成分を求める。次に、上位 m 個の主成分によって各 P ベクトルを m 次元で近似する。この m 次元ベクトルが Q ベクトルとなる。以下、この変換と次元数 m の選定について説明する。

まず、 P について主成分分析を行い、すべての主成分を求める。次の条件を満たす 2851 次元ベクトルの列 $X_1, X_2, \dots, X_{2851}$ を求めれば、 X_i は第 i 主成分の方向を表すベクトルとなる。

- $|X_i| = 1$ ($i = 1, \dots, 2851$).
- X_1 は、すべての方向の中で P を射影したときの分散 v_1 が最大になる方向をとる。
- X_i ($i = 2, \dots, 2851$) は、 X_1, \dots, X_{i-1} のいずれとも直交する方向の中で P を射影したときの分散 v_i が最大になる方向をとる。

X_i は符号に関して自由であるが、本手法では本質的でないため符号の違いを無視する。

次に、上位 m 個の主成分によって各 P ベクトルを m 次元の Q ベクトルで近似する。 X_1, X_2, \dots, X_m を基底とする直交座標系 (原点は P の重心) を用意し、もとの P ベクトルをこの新しい座標系で表現する。この結果得られる m 次元ベクトルが Q ベクトルであり、その第 i 成分はもとの P ベクトルを第 i 主成分方向の座標軸に射影した値となる。この座標変換によって、 P は Q ベクトルの集合 Q に変換される。

Q ベクトルは少ない次元で P ベクトルの持つ情報を効率良く表現できる。図 4 は、第 i 主成分上の P の分散 (つまり固有値) v_i の累積割合をグラフにした

* 形態素解析によって、出現単語と基本形 (V の要素) との可能な対応を調べている。

** P ベクトルの長さ $|p(w)|$ の平均は 0.8278 であった。

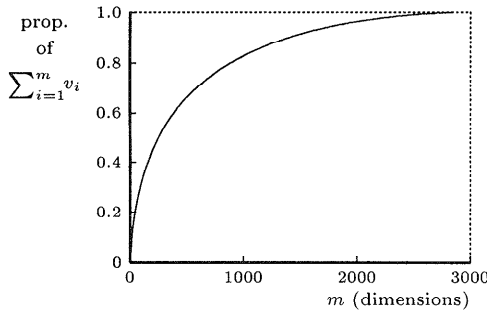


図4 分散 v_i (第 i 主成分の固有値) の累積割合
Fig. 4 Cumulative proportion of eigen values.

ものであり、 Q ベクトルの次元数 m と Q によって表現される情報量の関係を表している。 $m = 200$ のとき、 Q は P の持つ情報量の 45.11% を表現でき、500 次元では 66.21%，1000 次元では 82.80% を表現できる。このように次元数 m を増やせば情報量も増えるが、同時に P の持つノイズ（つまり語義定義の持つノイズ）を多く含むようになる。

我々は Q ベクトルの次元数 m を 281 とした。この場合、 Q は P の持つ情報量の 52.66% を表現できる。この次元数は、 Q の全情報量のうち機能語*の Q ベクトルが占める割合を最小化することによって得られる。つまり、機能語の持つ情報量をノイズと見なし、 Q の S/N 比を最大化しているわけである。4 章と 5 章で述べる実験結果は $m = 281$ のときのものである。

3. 意味空間のスケール変換

Q ベクトルの空間をスケール変換することによって単語間の意味距離に文脈依存性を与える。 Q ベクトル間の距離は、 P ベクトル間の距離と大きな違いはなく、どちらも文脈から独立した静的なものである。そこで図 5 に示すように、文脈の手がかりとして与えられた単語集合 C に適応するように、 Q ベクトルの空間をスケール変換する。スケール変換後の意味空間では、 Q ベクトル間の距離が C に依存した値となる。

以下、とくに断わらないかぎり、「ベクトル」は Q ベクトルを指し、「意味空間」は Q ベクトルの空間を指すものとする。

3.1 部分意味空間と関連語クラスタ

意味空間からいくつかの主成分（すなわち次元）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、第 2 主成分と第 3 主成分から構成される部分意味空間に、次の 15 個の単語をプロットしてみる。

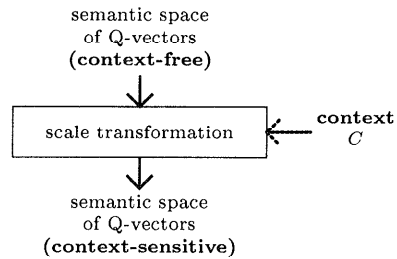


図5 意味空間のスケール変換
Fig. 5 Scale transformation of the semantic space.

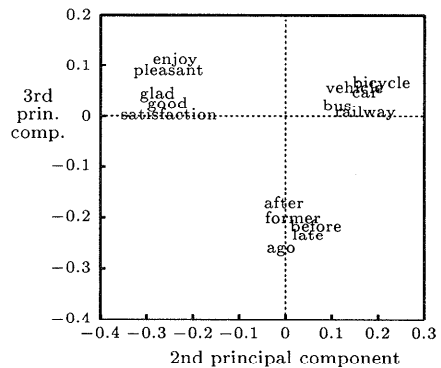


図6 部分意味空間における関連語クラスタ
Fig. 6 Word clusters in a semantic subspace.

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.

図 6 に示すように、これらの単語は 3 つのクラスタ（「楽しみ」「乗物」「過去」）を形成する。また、第 3 主成分だけでは「乗物」と「楽しみ」とを区別できないことも分かる。ここでは 2 次元の部分意味空間を例示したが、与えられた関連語集合をクラスタ化するためには一般により多くの主成分が必要となる。

関連語集合 C が与えられたとき、これをクラスタ化するために必要十分な主成分を選択するには、どうすればよいだろうか。上の例で見たように、適切に選ばれた主成分上では C が偏った分布をとる。したがって、 C を射影したときのバラツキが語彙 V のそれにくらべて十分小さくなるような主成分を選択すればよい。小嶋ら⁸⁾は、この考えに基づいて、語彙をクラスタリングすることを試みている。以下では、この手法を改良し、主成分を選択する・しないという 2 値的な方法ではなく、各主成分に連続的な重みを与えて新しい意味空間を構成することを試みる。

3.2 スケール変換による文脈への適応

意味空間のスケール変換は、与えられた単語集合 C

* 限定詞・冠詞・前置詞・代名詞・接続詞など 210 語。

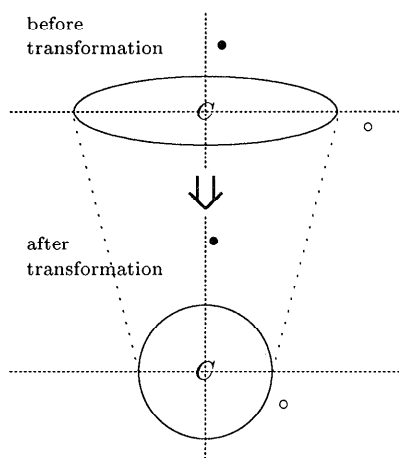


図7 スケール変換による文脈 C への適応
Fig.7 Adaptive scaling for a given context C .

に適応するように意味空間を再構成する手続きであり、次のように記述できる。

C を射影したときのバラツキが、語彙 V のそれにくらべて十分小さくなるように、各主成分のスケールを拡大または縮小する。

つまり、 C のバラツキが大きな主成分は、そのスケールを縮小することによって、距離計算への影響を小さくする。逆に、 C のバラツキが小さな主成分は、そのスケールを拡大することによって、距離計算への影響を大きくする。

このスケール変換の結果、 C は意味空間内で均整のとれたクラスタを形成するようになり、意味空間は C の意味的なまとまりを最もよく反映したものとなる。たとえば図7に示すように、意味空間内で C が扁平な（超楕円状の）分布範囲を持つとき、各主成分のスケールを調節して、 C が均整のとれた（超球状の）分布範囲を持つようにするわけである。この座標変換によって、ベクトル間の距離関係が変化する。図7に示した2つのベクトル ●, ○ は、スケール変換の前後で C との距離関係を大きく変化させている。

例として、図7の横軸を「本の厚さ」、縦軸を「本の背の高さ」として、スケール変換による文脈 C への適応を説明しよう。●は「普通の厚さで背の高い本」、○は「非常に厚くてやや背の低い本」となる。ある本棚に並べられた本の集合 C が「厚さはまちまちだが背の高さは普通」であるとき、変換前は ● のほうが C の重心に近いが、変換後は ○ のほうが近くなる。

このスケール変換によって、単語集合 C の持つ文脈に依存した単語 w, w' 間の意味距離 $d(w, w'|C)$ は、対応するベクトルを $q(w) = (q_1, \dots, q_m)$, $q(w') =$

(q'_1, \dots, q'_m) とするとき、次のように計算される。

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

ここで f_i は第 i 主成分の拡大率であり、次のように計算される*。

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

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

$SD_i(C)$ は第 i 主成分上の C の標準偏差、 $SD_i(V)$ は語彙 V の標準偏差とする。もし C が第 i 主成分上でクラスタ化している（つまり $r_i \approx 0$ ）ならば f_i は大きくなる。逆に、 C が第 i 主成分上でクラスタ化していない（つまり $r_i \gg 0$ ）ならば f_i は小さくなる。

以上の計算手続きによって、単語間の文脈依存的な意味距離を小さな計算コストで求めることができる。手続きの全体は、おおまかに初期処理・拡大率計算・距離計算の3つの部分に分けられる。初期処理は、前節で説明した 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）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れば、単語集合 C が与えられるごとに各主成分の拡大率 f_i を再計算するだけでよい。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. 意味距離の計算例

意味空間のスケール変換による単語間の意味距離の計算例をいくつかあげる。ここでは計算結果を直観的にとらえられるように、次のような問題を扱う。

与えられた単語集合 C について、語彙 V を距離 $\bar{d}(w, C)$ の昇順にソートし、その先頭から k 個の単語を取り出した集合 C^{+k} を求める。

ここで $\bar{d}(w, C)$ を次のように定義する。

$$\bar{d}(w, C) = d(w, \bar{w}|C).$$

ただし、 \bar{w} の Q ベクトル $q(\bar{w})$ を、 C の各単語 w_i の Q ベクトル $q(w_i)$ の平均とする。つまり、 $\bar{d}(w, C)$ は w から C の重心までの距離となる。 C^{+k} は C と意味的に関連する単語の集合となるはずである。

$C = \{\text{bus, car, railway}\}$ を与えた場合、表1に示す関連語集合 C^{+12} が得られた。たとえば carriage

* $|C|$ が十分に大きければ $f_i = 1/r_i$ とする方が自然だが、 $|C|$ が実際の範囲（数十語程度）では本文中の定義の方が良い結果をもたらす。

表1 単語集合 {bus, car, railway} からの連想
Table 1 Association from {bus, car, railway}.

$w \in C^{+12}$	$\bar{d}(w, C)$
car_1	0.1039
railway_1	0.1131
bus_1	0.1141
carriage_1	0.1439
motor_1	0.1649
motor_2	0.1949
track_2	0.1995
track_1	0.2024
road_1	0.2038
passenger_1	0.2185
vehicle_1	0.2274
engine_1	0.2469

表2 単語集合 {bus, ticket, tour} からの連想
Table 2 Association from {bus, ticket, tour}.

$w \in C^{+12}$	$\bar{d}(w, C)$
bus_1	0.1005
ticket_1	0.1055
tour_2	0.1161
tour_1	0.1314
half_1	0.1572
abroad_1	0.1622
passenger_1	0.1631
make_2	0.1670
timetable_1	0.1712
garage_1	0.1715
part_1	0.1723
tourist_1	0.1756

や vehicle といった単語がみられるように、「乗物」という文脈をよくとらえていることが分かる。一方、同じ bus を含む $C = \{\text{bus, ticket, tour}\}$ を与えた場合、表 2 に示す関連語集合 C^{+12} が得られた。この場合は「旅行」に関する単語が連想されている。

また、 $C = \{\text{read, magazine, paper}\}$ とした場合、次のような関連語集合 C^{+10} が得られた。

{paper_1, read_1, magazine_1, newspaper_1, print_2, book_1, print_1, wall_1, something_1, article_1}.

「読物」に関する単語がみられる。一方、 $C = \{\text{read, memory, machine}\}$ とした場合、次のような関連語集合 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「計算機」に関する単語がみられる。どちらの場合も単語 read が C に含まれているが、得られた関連語集合は大きく異なっている。

5. テキスト予測による評価

先行テキストに基づいて後続単語を予測することを通して、本手法によって計算された意味距離を評価する。ここでは、次のような課題を通して、本手法の評

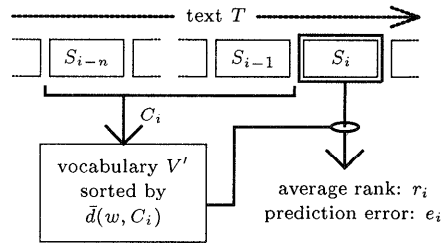


図8 先行テキストに基づく後続文の予測
Fig.8 Prediction of the succeeding sentence.

価を行う。

テキストの各文に現れる単語を、それに先行する n 文から予測する。

予測のターゲットとなる文とそれに先行する n 文は、多くの場合、同じ文脈の中にあると考えられる。したがって、先行テキストに基づいて後続文をうまく予測できるならば、先行テキストの文脈をうまくとらえていることがいえる。

具体的には図 8 に示すような課題を扱う。テキスト T を文列 S_1, \dots, S_N (S_i は先頭から i 番目の文) とする。ターゲット文を S_i ($i > n$) とするとき、それに先行する n 文 S_{i-n}, \dots, S_{i-1} に現れる単語からなる集合を C_i とする*。このとき、次の手続きによって T の予測誤差 \bar{e} を計算する。

- 語彙 V' の各単語 w を距離 $\bar{d}(w, C_i)$ の昇順にソートする。ただし、 V' は語彙 V から機能語 (210 語) を除いたものとする。
- ソートされた V' における S_i の各単語 (V' の要素のみ) の平均順位 r_i を求め、 S_i の予測誤差 e_i を $r_i/|V'|$ とする。
- T 全体を通しての予測誤差 \bar{e} を、 e_{n+1}, \dots, e_N の平均とする。

後続単語をうまく予測できた場合、 $\bar{e} \approx 0$ となり、予測がランダムな場合、 \bar{e} の期待値は 0.5 となる。

O. Henry の短編小説 “Springtime à la Carte”¹⁰⁾ について、予測誤差 \bar{e} の評価を行った。このテキストは 110 文 (1620 語) からなる。先行文数 $n = 1, \dots, 8$ について、このテキストの \bar{e} を本手法によって計算した結果を表 3 の「スケール変換あり」の欄にあげる。どの n についても、予測がランダムな場合にくらべて e が小さくなった。また、意味距離を使わずに、West⁹⁾ の 500 万語のコーパスにおける出現頻度順に語彙 V' をソートした場合、 $\bar{e} = 0.2236$ となった。本

* 正確には、要素の重複を許す集合であり、その度数をスケール変換に反映させている。

表3 O. Henry の短編小説における予測誤差 \bar{e}
 Table 3 Prediction error \bar{e} of O. Henry's short story
 (with and without scale transformation).

n	スケール変換あり	スケール変換なし
1	0.3248	0.3410
2	0.1838	0.2778
3	0.1623	0.2837
4	0.1602	0.2905
5	0.1635	0.2960
6	0.1696	0.3007
7	0.1749	0.3038
8	0.1801	0.3067

手法による結果は、このような出現頻度を利用していないにもかかわらず、 $n=1$ の場合を除いて \bar{e} はより小さくなっている。

スケール変換の有効性を確かめるために、スケール変換を行わない単純なユークリッド距離を用いて後続単語の予測を行った。この結果を表3の「スケール変換なし」の欄にあげる。どの先行文数 n についても、スケール変換を行わなかった場合は、行った場合に比べて予測誤差 \bar{e} が大きくなっており、スケール変換の有効性が示された。

先行テキストが長くなるほど後続テキストの予測が正確になると考えられるが、実際には予測誤差 \bar{e} を最小にする先行文数 n が存在した。スケール変換を行った場合は $n=4$ 、行わなかった場合は $n=2$ のとき、 \bar{e} が最小となっている。この理由として、文脈がテキスト全体にわたって一様ではないこと¹⁾が考えられる。テキストの場面(意味段落)境界^{11),12)}の付近では、先行テキストと後続テキストが異なる場面に含まれるため、 e_i が大きくなる。場面ごとに変化する文脈をとらえるには n が小さいほうが有利であるが、小さすぎると文脈そのものをとらえきれない。このことが n に最適値を与えていると考えられる。

6. おわりに

本論文では、単語間の意味距離を文脈に依存して動的に計算する一手法として「意味空間のスケール変換」を提案した。

意味空間とは、単語間の意味距離を計量化するためのベクトル空間である。まず、英語辞書に基づいて構成された意味ネットワーク上の活性伝播によって、各単語 w を 2851 次元の P ベクトル $p(w)$ に変換する。次に、P ベクトルの集合について主成分分析を行い、得られた主成分を利用して $p(w)$ を 281 次元の Q ベクトル $q(w)$ に変換する。単語間の意味距離は、この Q ベクトル間の距離として計算される。

意味空間のスケール変換は、文脈の手がかりとして

与えられた単語集合 C に意味空間を適応させる手続きである。 C が与えられると、意味空間の各次元を拡大・縮小することによって C をクラスタにまとめる。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 $d(w, w'|C)$ (すなわち $q(w), q(w')$ 間のユークリッド距離) が C の意味的な分布特徴に依存した値となる。先行テキストに基づく後続単語の予測を通して、本手法による意味距離がテキストの文脈をよくとらえていることが確かめられた。

本手法では英語辞書を知識源としているが、コーパスにおける単語の共起関係(相互情報量¹³⁾や n -gram 統計⁴⁾など)から Q ベクトルを生成することも可能である。しかしながら、辞書から抽出したベクトルとコーパスをもとに生成したベクトルは、互いに異なった情報を持つと考えられる¹⁴⁾。これら 2 つの知識源を相補的に利用することが必要であろう。

今後の課題として、まず、Q ベクトルの次元数 m の最適値を、5 章で述べたテキスト予測による評価を通して、最尤推定することがあげられる。つまり、テキストの文脈を最もよくとらえられる次元数を推定するわけである。また、単語が出現する先見確率を単語間の意味距離に反映させることもあげられる。4 章で例示した関連語の連想では、**that/have/something** といった単語が出力された。これらは出現頻度が高く意味の弱い機能語であり、本来は連想されにくいはずである。より人間の直感に合った連想を実現するには、各単語の意味の強さを考慮すべきであろう。

本手法は、人間の持つ「文脈を動的にとらえる能力」を自然言語処理の立場からモデル化したものである。この能力によって、人間は小さな認知コストで最適な解を求めることができ、フレーム問題を回避することができる。将来的には本モデルを一般化し、人間の持つ「情報の重要な部分に注目する能力」や「刻々と変化する環境に適応する能力」をモデル化していきたい。

参考文献

- 1) Grosz, B.J. and Sidner, C.L.: Attention, Intentions, and the Structure of Discourse, *Computational Linguistics*, Vol.12, No.3, pp.175-204 (1986).
- 2) Osgood, C.E.: The Nature and Measurement of Meaning, *Psychological Bulletin*, Vol.49, No.3, pp.197-237 (1952).
- 3) Morris, J. and Hirst, G.: Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text, *Computational*

- Linguistics*, Vol.17, No.1, pp.21-48 (1991).
- 4) Brown, P.F., Della Pietra, V.J., deSouza, P.V., Lai, J.C. and Mercer, R.L.: Class-based *n*-gram Models of Natural Language, *Computational Linguistics*, Vol.18, No.4, pp.467-479 (1992).
 - 5) Kozima, H. and Furugori, T.: Similarity between Words Computed by Spreading Activation on an English Dictionary, *Proc. 6th Conference of the European Chapter of the Association for Computational Linguistics (EACL-93, Utrecht)*, pp.232-239 (1993).
 - 6) Salton, G. and McGill, M.J.: *Modern Information Retrieval*, McGraw-Hill, New York (1983).
 - 7) 小嶋秀樹, 古郡廷治: 単語の意味的な類似度の計算, 電子情報通信学会技術研究報告, AI92-100, pp.81-88 (1993).
 - 8) 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 (1995).
 - 9) West, M.: *A General Service List of English Words*, Longman, Harlow (1953).
 - 10) Thornley, G.C.: *British and American Short Stories*, Longman, Harlow (1960).
 - 11) Kozima, H.: Text Segmentation based on Similarity between Words, *Proc. 31st Annual Meeting of the Association for Computational Linguistics (ACL-93, Ohio)*, pp.286-288 (1993).
 - 12) 小嶋秀樹, 古郡廷治: 単語の結束性にもとづいてテキストを場面に分割する試み, 情報処理学会研究会報告, 93-NL-95-7, pp.49-56 (1993).
 - 13) Church K.W. and Hanks, P.: Word Association Norms, Mutual Information, and Lexicography, *Computational Linguistics*, Vol.16,

No.1, pp.22-29 (1990).

- 14) Niwa, Y. and Nitta, Y.: Co-occurrence Vectors from Corpora vs. Distance Vectors from Dictionaries, *Proc. 15th International Conference on Computational Linguistics (COLING-94, Kyoto)*, pp.304-309 (1994).

(平成8年4月2日受付)

(平成8年12月5日採録)



小嶋 秀樹 (正会員)

昭和63年電気通信大学電気通信学部情報数理工学科卒業。平成2年同大学大学院電気通信学研究科情報工学専攻博士前期課程修了。平成6年同専攻博士後期課程修了。工学博士。同年郵政省通信総合研究所入所。現在、同所関西先端研究センター知識処理研究室研究官。認知科学、計算言語学の研究に従事。人工知能学会、言語処理学会、認知科学会、ACL各会員。



伊藤 昭 (正会員)

昭和47年京都大学理学部物理学科卒業。昭和54年同大学大学院理学研究科博士課程修了。理学博士。同年郵政省電波研究所(現通信総合研究所)入所。現在、同所関西先端研究センター研究調整官、知識処理研究室長兼務。ヒューマンインタフェース、自律エージェント、人工生命などの研究に従事。人工知能学会、認知科学会、電子情報通信学会、ACL、ACM各会員。