連想としての意味

持橋大地, 松本裕治
奈良先端科学技術大学院大学情報科学研究科

daiti-m@is.aist-nara.ac.jp
matsu@is.aist-nara.ac.jp

概要:

本論文では, 単語の意味を単語間の連想関係を表す確率分布として表現し, その定式化と連想確率の獲得について述べる. 単語の意味的な重みを表す指標として単語の共起確率分布の情報量から計算される連想情報量を提案し, 共起確率との組み合わせにより連想確率を計算する.

連想は Markov 過程の上で行われ, その状態確率分布として意味が定義される. 状態遷移として連想を行うことによって, 直接共起しない語の意味的な関係が表現できる.

また, 確率ベクトルとして捉えた意味のスケール変換として文脈を捉え, 先行単語集合の数を仮定しない非線型な更新式を提案し, これにより文脈の強化と順序への依存が表現できることを示す.

現実のテキストから意味を獲得し, 文脈をモデル化することで, 意味的類似度や文脈解析だけでなく, 情報検索などにおいて様々な実際的な意味処理が可能になる.

Meanings as association

Daichi Mochihashi and Yuji Matsumoto
Graduate School of Information Science, NAIST

daiti-m@is.aist-nara.ac.jp
matsu@is.aist-nara.ac.jp

Abstract

This paper describes meanings of a word by stochastic association.

First, we propose a new indicator of semantic informativeness of a word by its co-occurrence distributions. Second, we define the association probability by a combination of co-occurrence probability and the indicator.

Then, regarding context as a vector of scaling factors against semantic vector, we propose a nonlinear formula of context succession to show its validity in modeling reinforcement and order depencency of context.

Stochastic treatment of meaning and its aquisition from texts is useful in real semantic processing.

0　はじめに

意味論と構文論は言語学の二つの柱である. Chomsky 以後, 構文の解析から意味が構成できるとする「意味の構成性原理」(Frege,1890) が広まったが, 充分な結果を挙げたとは言えない. 意味とは大きく i) 構文的 ii) 語彙的 の二つに分けられるが [1], それは並列, あるいは構文上位に考えられるものではなく, イメージの抽象化としての静的な語彙的意味を動的に構成する方法として構文的意味が考えられる. syntagme は paradigm の上に存在するものなのである.

本論文ではそのような静的意味を連想の束として確率的に定式化することを試みる. この表現では, 直接の共起のない関係についても連想の連鎖により確率を与えることができる.

1 意味と意味ネットワーク

Saussure[1] の言うように, 語の意味とは畢竟他の語との関連で定義されるものであり, 語彙を単語／概念の意味ネットワークによって記述しようとする試みは Quillian(1968) に始まり, Shapiro(1971), Conceptual Graph(Sowa84) などによって提唱されてきた. この他, Generative Lexicon(Pustejovsky95) の Qualia Structure や, HPSG (Pollard&Sag85) の語彙記述における構造共有もそのリンクを基にする記述法から意味ネットワークの一種と考えることができる.

しかし, 長い研究の経過にもかかわらず, これらは大きな成果を上げたとはいえない. その理由は, 大きく分けて二つに考えることができる.

- 先験的なラベルの仮定
- 離散性.

これらの定式化はいずれも単語あるいは概念をノードとし, その間にラベル付きリンクを張ることによって構成されるものであるが, そのラベルがどの定義域の中から, どのように選ばれるかについてこれらの理論は根拠を与えていないため, 語の意味の定義は全く恣意的なものに終わらざるを得ない. また, 各リンクはある／なしの二元論によって離散的に表現されているため, あるリンクの存在／非存在が推論に決定的な影響を及ぼす他, A でありながら幾分 B の意味も含むというような意味の微妙な局面を表現することができない. 実際には我々は, 様々な意味解釈の間の重みづけに基づいて推論を行っているのであり, このような連続性をモデル化する必要がある.

それでは, どのようなモデルを考えれば良いのだろうか. ここで, 非常に単純な状況に立ち戻って考えてみたい. すなわち, 子供が全く知らない未知の単語 ℵ に出会った状況である. もし暗闇の中で「ℵ！」と単独に叫ばれたとすれば, 彼はこの語について全く何も知りようがない. しかし, 名差しを伴って「これが ℵ である」と言われたり, 「ℵ は α のことである」あるいは「$\beta_1\aleph\beta_2$」という文脈の中で言われるとき, そこに名差されたものと語, あるいは語の間の束縛 (binding) が生じ, ℵ から差されたイメージ, あるいは単語 α, β_1, β_2 などを<u>連想</u>して解釈ができるようになる. もちろんこれは一段階に留まるものではなく, α からさらに α_2, α_3 が, β_1, β_2 から β_3, β_4 が, $\cdots$ というように連想が広がっていき, それにつれて連想の重みが逓減していく (活性拡散 [2]). そこで, このような重みつき連想の束を意味と考えることができる. ここに従来のようなラベルは存在しないが, ラベル自体単語なのであるから, 適切な操作により同等の単語集合を連想させることが可能だと考えられる.

2 関連研究

語の意味を確率的な意味ネットワークによって表現しようとする試みはこれまでもいくつか行われている. 古くは [3] があるが, これはあらかじめ固定された重みの下で CFG の構文木を選択するものである.

小嶋 [4] では英語辞書からの確率計算を行い, 興味深い結果を得ているが, 辞書からの学習のため動的な環境に対応できない点や機能語の扱いなどの問題点がある. 廣 [5] はこれをコーパスからの学習に拡大しているが, 相互情報量を用いているため高頻度語のみを扱っており大きな問題を残している. 他には Dictionary Parsing Project[6] や MindNet[7] などの試みが行われているが, いずれも辞書からの獲得を主眼としたものであり, コーパスからの直接の計算は行われていない.

3 確率的表現

上の議論から, 語の意味は他の語との連想関係によって定義され, その手がかりは言語的な束縛＝共起であることが明らかになった. 共起関係にある二語はその共起頻度が高いほど関連性は基本的には高いと考えられるが, is, the のような機能語は共起頻度が大きくとも連想関係は弱く, また機能語とは言えないが連想関係の比較的弱い語 (do, some など) もある.

そこで, 共起空間における関係 c_{ij} から意味空間における関係 s_{ij} を構成することを考え, 以下のように定式化する.

<u>定義</u>: (共起空間) 自然数 $i, j \in \{1..n\}$ に対し, i 番目の単語と j 番目の単語との共起を根元事象 $\omega = c_{ij}$ とする. $\Omega_C = \{c_{ij}\}$ および Ω_C の部分集合を要素とする集合体 $\Xi_C = \wp(\Omega_C)$ に対し, N を事象から生起回数への関数として,

$$P_C(\omega) = \frac{N(c_{ij})}{\sum_{i,j} N(c_{ij})}, P_C(Z) = \sum_{\omega \in Z} P_C(\omega) \quad (1)$$

によって定義される確率空間 $\{\Omega_C, \Xi_C, P_C\}$ を共起空間と呼ぶ.

<u>定義</u>: (意味空間) i 番目の単語と j 番目の単語の共起空間における関係を $\omega = s_{ij}$, その確率を $P_S(\omega)$ とする.
$\Omega_S = \{s_{ij}\}$ および $\Xi_S = \wp(\Omega_S)$ に対し,

$$P_S(\omega) = a_{ij}, P_S(Z) = \sum_{\omega \in Z} P_S(\omega) \quad (2)$$

によって定義される確率空間 $\{\Omega_S, \Xi_S, P_S\}$ を意味空間と呼ぶ.

このとき, 意味を以下のように定義する.

定義: (意味) 意味空間上での集合 $w_i = \{s_{ij} | j \in \{1..n\}\}$ を意味空間における語と呼び, 各語 w_i を要素とする集合 $L = \{w_i\}$ を語彙と呼ぶ. L に値を取る確率過程 $c_t(\omega)$ $(\omega \in L)$ において,

$$\begin{aligned} a(w_j|w_i) &\equiv P(c_t(\omega) = w_j | c_{t-1}(\omega) = w_i) \\ &= \frac{P_S(w_i \cap w_j)}{P_S(w_i)} = \frac{a_{ij}}{\sum_j a_{ij}} \end{aligned} \quad (3)$$

と定めるとき, $a(x|y)$ を連想確率という. このとき, $c_t(\omega)$ は単純 Markov 過程となり, 意味 σ とは状態確率分布

$$\sigma = \{P(c_t(\omega) = w_i)\} \ (i \in \{1..n\}) \quad (4)$$

と定義する. とくに, 単語 w_i の意味 $\sigma(w_i)$ とは, $c_0(w_i)$ から始まって t 時間後の確率分布

$$\sigma_t(w_i) = \{P(c_t(\omega) = w_j | c_0(\omega) = w_i)\} \ (j \in \{1..n\}) \quad (5)$$

である.

3.1 連想確率の計算

そこで次に, 連想確率 $a(w_j|w_i) = \frac{a_{ij}}{\sum_j a_{ij}}$ を求めることが問題となる.

a_{ij} は w_i と w_j の意味的重なりを表し, 方向に依存しないから, これを相互情報量 $I(w_i; w_j) = \log \frac{p(w_i, w_j)}{p(w_i)p(w_j)}$ で表すことがまず考えられる [8][9][10]. しかし, この定義の問題点として, $p(w)$ の小さい低頻度語について過大に評価してしまい, $p(w_i, w_j)$ が充分に大きい高頻度語以外では有効な測度が得られないという問題が挙げられている. しかしながら, 我々の日常的な意味把握にとって低頻度語こそ重要な意味を持っていることが多く, それを切り捨てることはできない.

[11] に述べられているように, この原因は情報量の定義と我々の直感が相反していることから生じている. 情報理論では出現確率の低い記号を情報量が大きいとみなすが, 我々は逆によく目にする高頻度語を重要だとみなすからである.

Pearl[12] が述べているように, 我々はある記号の出現に際し, その記号の確率モデル全体の中での出現確率 $\sum_j P_C(c_{ij})$ を計算して情報量を判断しているわけではなく, その語の周囲のもっと局所的な情報を基にして判断していると思われる. そこで, a_{ij} それ自体ではなく, $\frac{a_{ij}}{\sum_j a_{ij}} = a(w_j|w_i)$ を求めればいいことに注意して, 共起空間において同様に Markov 過程 $p(w_j|w_i) = \frac{P_C(w_i \cap w_j)}{P_C(w_i)}$ を考える.

このとき, 連想確率 $a(w_j|w_i)$ は基本的には共起確率 $p(w_j|w_i)$ に比例すると考えられるが,

$$a(w_j|w_i) = p(w_j|w_i) \quad (6)$$

としたのでは適当でない. 上で述べたように, 連想語によって意味的な重みが違い, 意味の「薄い」語には連想関係が小さいと考えられるからである.

ここで, 機能語のように意味の「薄い」語に共通する特徴を考えてみると,

様々な語と平均的に共起する.

という特徴を挙げることができる. 共起関係から意味的関係が構成されるとするならば, 多くの単語と平均的に共起する語は一連想語当たりの連想確率が小さく, はっきりしたイメージを喚起させないはずである. そこで, 以下で共起確率分布の情報量を用いて連想情報量 (平均共起確率) $ap(x)$ を求め, これを用いた重み付けが単純共起確率 $p(w_j|w_i)$ の拡張になっていることを述べる.

3.2 連想情報量

単語 $x \in L$ に対して, その共起確率分布 $\{p(w|x)\}$ を考える. このとき, その情報量

$$H(x) = -\sum_{w \in L} p(w|x) \log p(w|x) \quad (7)$$

は情報量の意味より単語 x との共起に関する平均分岐数の対数であるから,

$$e^{H(x)} = \exp(-\sum_{w \in L} p(w|x) \log p(w|x)) \quad (8)$$

は x の共起に関する平均分岐数 (perplexity)[13] を表す. そこで, この逆数

$$\begin{aligned} e^{-H(x)} &= \exp(\sum_{w \in L} p(w|x) \log p(w|x)) \quad (9) \\ &= \langle p(w|x) \rangle_w \equiv ap(x) \quad (10) \end{aligned}$$

を一語当たりの平均共起確率と考えることができ, これを連想情報量 $ap(x)$ と呼ぶことにする.

この $ap(x)$ は単語の情報量を自然な形で反映する. 表1に, コーパスから計算した連想情報量の結果を示す. 定義より連想情報量は従来のように頻度に依存しないことから, 高頻度の語であっても持つ情報量は比較的高い場合もあることに注意されたい.

表 1: 連想情報量

x	$ap(x)$	(頻度)
の	0.000821556	(17425)
を	0.000907809	(9763)
が	0.00111244	(7980)
に	0.00113162	(9219)
は	0.00153034	(9152)
や	0.00168644	(955)
で	0.00190469	(1425)
だ	0.00212288	(4768)
も	0.00216023	(3017)
と	0.00232572	(5981)
など	0.00256123	(885)
:		
風土	0.0418808	(11)
契約	0.0418861	(15)
宮沢	0.0418896	(141)
豊富だ	0.0720602	(7)
北方	0.0720915	(34)
色彩	0.114388	(5)
自公民	0.114401	(29)
北海道	0.11455	(18)
質量	0.288675	(3)
重ね	0.288675	(3)
名刺	1	(2)
名望	1	(1)
模範	1	(1)

朝日新聞社説 1992/1-1992/6, 約 17000 行から計算. 5 単語の窓によって共起を定義した.

なお, (1) において共起事象から頻度への関数である N は様々に定義することができるが[1], 単純に生起回数を与えるものとした場合, (14) における共起確率分布のエントロピー $H(x)$ は共起が生じる毎に次式によって更新することができる. [2]

$$\begin{cases} H'(x) = & \frac{s}{s+1}H(x) - \log(s+1) + \frac{s}{s+1}\log s \\ & (c_k = 0\text{のとき}) \\ H'(x) = & \frac{s}{s+1}H(x) - \log(s+1) + \frac{1}{s+1}\{(c_k+1) \\ & \log(c_k+1) - c_k\log c_k + s\log s\} \\ & (c_k \neq 0\text{のとき}) \end{cases}$$

ただし, $c_k = N(c_{xk}), s = \sum_k c_k$.

$ap(x) = e^{H(x)}$ であるから, これにより, 共起が生じる毎に連想情報量 $ap(x)$ を $O(1)$ の計算量で更新することができる.

[1] 中心語からの距離に反比例させる [14], 強調されている場合には大きくするなどが考えられる.

[2] 証明は付録を参照.

3.3 重み付けと共起

本節では簡単のため, 以下 $w_i, w_j, \ldots$ を $i, j, \ldots$ と略記する.

i から j への連想確率 $a(j|i)$ は共起確率 $p(j|i)$ に依存するが, それだけでは適切でないことを 3.1で述べた. そこで, $p(j|i)$ を共起語 j の連想情報量 $ap(j)$ で重み付けした量

$$p(j|i)ap(j) = p(j|i)\langle p(w|j)\rangle_w \tag{11}$$

を考えると, Markov 性により

$$p(i,j,w) = p(i)p(j|i)p(w|j) \tag{12}$$

$$\therefore \frac{p(i,j,w)}{p(i)} = p(j|i)p(w|j) \simeq p(j|i)\langle p(w|j)\rangle_w \tag{13}$$

となり, 単純共起確率 (6) の拡張になっていることがわかる.

(11) を正規化した

$$a(j|i) = \frac{p(j|i)ap(j)}{\sum_j p(j|i)ap(j)} \tag{14}$$

によってコーパスから計算した連想確率の例を表2に示す. ここでは, 機能語の特徴を考慮しているため従来と異なり stop list を必要としない (完全自動学習が可能である) ことに注意されたい.

(11) はまた, 共起空間において Markov 過程を Bayesian Network[12] とみなした時, 連想を広げる際に共起語 j からさらに連想の広がる重みを evidence $\pi(j)$ とみなし, j への共起確率を posterior probability $\lambda(j)$ とみなした時の Bayesian Belief $\pi(j)\lambda(j)$ と考えることもできる.

4 確率的連想

(14) によって連想確率 $a(j|i)$ を計算すると, (??) よりそれを意味空間における Markov 過程の状態遷移確率とみなして遷移確率行列 A の要素 A_{ji} に対応させることができる.

このとき, (4) により単語 w_i の意味 $\sigma(w_i)$ を, 次のように t レベルの遷移により

$$\begin{aligned} \sigma_t(w_i) &= \frac{1}{t}(A\vec{w}_i + A^2\vec{w}_i + \cdots + A^t\vec{w}_i) \\ &= \frac{1}{t}\sum_{i=1}^{t} A^i\vec{w}_i \end{aligned} \tag{15}$$

として計算する. ただし, $\vec{w}_i$ は第 i 要素が 1, 他が 0 である長さ n の縦ベクトル ${}^t\underbrace{(0\ldots\overset{i}{1}\ldots0)}_{n}$ である.

表 2: 連想確率

水域:

カイリ	0.133754	(7)	
沿海	0.109562	(1)	
禁漁	0.062607	(1)	
操業	0.037416	(4)	
作れる	0.033712	(1)	
漁場	0.032824	(2)	
過密だ	0.032573	(1)	
漁獲	0.031461	(4)	
:			
へ	0.001043	(1)	
でも	0.001023	(1)	
も	0.000721	(1)	
	1		

影響:

有機	0.041175	(6)	
及ぼす	0.029010	(11)	
守備	0.026850	(1)	
出始める	0.024586	(3)	
同派	0.023275	(1)	
洋風だ	0.021365	(1)	
はかばかしい	0.020330	(2)	
水銀	0.019239	(6)	
与える	0.018079	(15)	
:			
くる	0.000344	(1)	
よる	0.000335	(1)	
から	0.000176	(1)	
	1		

これは [4] と同様活性拡散をモデル化したものだが，そこでの規則

$$v(T+1) = \phi(R(T) + R'(T) + e(T))$$

[R, R' は活性度 × リンクの重みの $\sum$，e は外部から与えられる活性，ϕ は値を $[0,1]$ に制限する関数.]

に比べ数学的意味が明確であるという特徴がある.

実際，Markov 確率行列の性質により A の固有値 $\lambda = 1$ に対する固有ベクトルを σ^* とおけば，$\lim_{n\to\infty} A^n \vec{w} = \sigma^*$ であるから[3]，(15) において $t \to \infty$ とすれば $\sigma(w) = \sigma^*$ となり，意味空間がエルゴード的であればすべての意味は t を充分大きくとれば σ^* に収束する[4]. この σ^* はそれ以上連想をくり返しても意味の変わらないベクトルであるから，これを (その意味空間における) 意味的なトートロジーと考えることもできる.

(5) のように意味を確率分布 $\sigma_n(w)$ で表すと，単語 w_i と w_j との類似度を $s(w_i, w_j) = \dfrac{\sigma(w_i)\cdot\sigma(w_j)}{|\sigma(w_i)||\sigma(w_j)|}$ のように測ることができる [8][9][10].

表 3に，σ_t を用いて計算した意味的類似度を t の値によって分けて示す. 当初関連の薄かった単語も，連想を広げることによって類似性が表れてくる場合があり，意味の遷移は $t = 4$ から $t = 10$ 程度で近似的に平衡に達することがわかる.

表 3: 連想による単語の意味的類似度

t	s(自然, 毒)	s(太陽, 水棲)
1	0.022863	0.003148
2	0.097671	0.012428
4	0.155306	0.028985
8	0.162944	0.031596
10	0.163072	0.031629
20	0.163095	0.031634

レイチェル・カーソン『沈黙の春』, 9 語の窓.

t	s(女性, 男性)	s(話す, 持つ)
1	0.116395	0.003745
2	0.212254	0.007996
4	0.384780	0.010564
8	0.393068	0.011252
10	0.393067	0.011277
20	0.393064	0.011284

朝日新聞天声人語:1992/1 月 –12 月, 13 語の窓.

5 文脈のモデル

上で得た意味ベクトルに重みづけを行うことによって，文脈に適合した意味に変換することを考える. [16] では先行する n 単語の意味ベクトルの分散が最小になるように変換を行い，よい結果を得ているが，この方法では同じ方向を持つ意味ベクトルが連続して入力されてもその文脈への強化が行われないこと，n の選び方が任意であり，先行単語集合として採用する／しないの二値に支配されることには問題が残る.

また，文脈は先行単語の意味の平均ではなく，それらの順序にも影響されるはずである.

そこで，文脈 $\mathbf{c}$ を意味ベクトルの各次元に対する拡大率として初期値 1 を与えて

$$\mathbf{c}_0 = (1, 1, \ldots, 1) \tag{16}$$

[3]証明はたとえば [15].

[4]厳密にはエルゴード的であったとしても，実際には意味過程は概略複数の連結成分から構成されると考えられるため，Markov 過程の状態遷移に関する情報量の凸増大性を考え合わせると，理論的な収束は現実的にはあまり問題とならない.

とし, 文脈 $\mathbf{c}$ における w の意味 $\sigma(w)|\mathbf{c}$ を

$$\sigma(w)|\mathbf{c} \equiv \alpha(w_0c_0, w_1c_1, \ldots, w_nc_n) \qquad (17)$$
$$(w_i \in \sigma(w), c_i \in \mathbf{c})$$

と定義する. ただし, α は正規化定数で, $\alpha = (\sum_i w_ic_i)^{-1}$ である.

ここで, $\mathbf{c}$ の絶対値は問題にならないから, $\mathbf{c}_0 = 1/n(1, \ldots, 1)$ としてよい. こうすることで, $\mathbf{c}$ を単語と同様に確率ベクトルとして扱うことができる.

このとき, 入力単語列 $w_0 \cdots w_n w_{n+1} \cdots$ に対し, 文脈 $\mathbf{c}$ を

$$\mathbf{c}_{n+1} = \lambda \cdot \sigma(w_n)|\mathbf{c}_n + (1-\lambda)\mathbf{c}_n \qquad (18)$$

によって更新する. λ は文脈感応性を表す定数であり, $0 \leq \lambda \leq 1$ である. $\lambda = 0$ のとき文脈は入力によって全く変化せず, $\lambda = 1$ のとき文脈は直前の入力単語の意味のみに追従する.

(18) は非線型な更新を行うことに注意されたい. 一般に入力単語列 $w_0 \cdots w_n w_{n+1} \cdots$ の順序によって文脈は異なったものになる.

例として, この文脈の定義を用いて宮澤賢治の短編小説「よだかの星」を解析した意味ネットワークにおいて, 単語列「大きな, 飛ばす, 殺す, 襲う」を与えたときの文脈を計算してみると表 4のようになり, 連想されるもの (鷹) に高い重みを持った文脈が得られることがわかる.

これに対し, 単純に意味ベクトルの平均をとる方法では関連する単語が文脈として表れるものの, それらに対して強い重みは得られない.

6 考察と展望

本研究では頻度による従来の単語の情報量基準を見直し, 共起確率分布から連想確率を定義したが, 完全に確率分布のみによってモデル化を行ったために逆に落ちている情報も存在する. すべての語に対して絶対頻度に関わりなく共起分布に同じ重みを与えているために, 表 2の'影響' の連想語にみられるように, 全体として低頻度の語が高頻度の語を抜いて強く連想される傾向にある.

この解決として, 高頻度の語に対して何らかの優先的な重み付けを与えることが考えられる. この重み付けは確率モデルとは独立であり, 文脈ベクトルに対する予備的操作として扱うことができるのではないかと考えている.

文脈モデルについては, 自然な効果である文脈の強化とその際の文脈感応性は (18) 式で扱えるものの, 文脈の変化が急峻に過ぎるという問題がある. 実

表 4: 文脈の強化

非線型(提案手法)

鷹:	0.175211
に:	0.077549
れる:	0.061118
は:	0.060922
風:	0.040431
を:	0.025111
とき:	0.024908
が:	0.021462
殺す:	0.013769
の:	0.010758
熊:	0.009982
横:	0.008415

線型(平均)

熊:	0.038979
鷲:	0.029454
れる:	0.027468
風:	0.027223
横:	0.022578
鷹:	0.019827
襲う:	0.019800
とき:	0.018622
一杯:	0.018163
血:	0.017837
もうすぐ:	0.015955
ぞ:	0.015315
殺す:	0.015148

際には文脈は入力語の意味ベクトルとの積で表されるような微分的動作と同様に，先行文脈の和の平均によって表されるような積分的動作によっても更新されていると考えられるため，今後は両者を統合するようなモデル化が必要となると思われる．

単語や文脈を確率ベクトルで表すことによって，様々な応用が可能である．本論文で述べた意味的類似性の計算の他にも，文脈を用いた照応関係の同定や選択制限 (selectional preference) の理解が離散的なルールを用いずに記述できる可能性がある．また，意味が連想によって表されることから，これを自然にメタファーのモデルとして用いることも可能であろう．また，実用面では実テキストに対して意味を学習することができることから，Query expansion など情報検索に関する意味処理として様々に有用となると考えている．

7 おわりに

本論文では，意味を Markov ネットワーク上の状態確率分布で表し，連想の定義と定式化を行った．共起空間から意味空間への変換に当たって，相互情報量を用いる従来研究とは異なり，人間の感じる情報量の基準に基づき共起確率の拡張として連想確率を定義し，関連して頻度によらない単語の情報量の基準として共起関係に基づく連想情報量を提案した．

また，確率分布のスケール変換として文脈を捉え，離散的な定数や素性を含まず文脈感応性のみに依存する文脈の更新法を提案した．

意味，および文脈の双方を確率ベクトルとして表現し，事前知識に依存せずにコーパスから意味ネットワークを構成し意味を計算することで，意味的類似性や文脈解析，照応解析などへの様々な応用が期待できる．

謝辞

本研究を進めるにあたり，当初より貴重な助言をいただいた奈良先端大 知識工学講座の石井信助教授に感謝します．

References

[1] Ferdinand de Saussure. 一般言語学講義. 岩波書店, 1972.

[2] A.M. Collins and E. F. Loftus. A spreading activation theory of semantic processing. *Psychological Review*, No. 82, pp. 407–428, 1975.

[3] David L. Waltz and Jordan B. Pollack. Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science*, No. 9, pp. 51–74, 1985.

[4] 小嶋秀樹, 古郡延治. 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.

[5] 廣恵太, 伊藤毅志, 古郡延治. コーパスから抽出した単語間類似度に基づく意味ネットワーク. 情報処理学会第 51 回論文集, Vol. 3, pp. 13–14, 1995.

[6] Univ. of Southern California ISI. Dictionary parsing project, 1998. http://www.isi.edu/natural-language/dpp/.

[7] Stephen D. Richardson, William B. Dolan, and Lucy Vanderwende. Mindnet: acquiring and structuring semantic information from text. Technical report, Microsoft SR-TR-98-23, May 1998. http://research.microsoft.com/scripts/pubDB/pubsasp.asp?RecordID=150.

[8] Kenneth Ward Church and Patrick Hanks. Word association norms, mutual information, and lexicography. In *Proc. of COLING 27*, pp. 76–83, 26-29 Jun 1989.

[9] Donald Hindle. Noun classification from predicate-argument structures. In *28th Proc. of COLING*, pp. 268–275, 1990.

[10] Fernando Pereira, Naftali Tishby, and Lillian Lee. Distributional clustering of english words. In *Proc. of 31th ACL*, pp. 183–190, 1993.

[11] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[12] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. The Morgan Kaufmann Series in Representation and Reasoning. Morgan Kaufmann, 1988.

[13] 北研二, 中村哲, 永田昌明. 音声言語処理. 森北出版, 1996.

[14] Kevin Lund, C.Burgess, and R.A.Atchley. Semantic and associative priming in high-dimensional semantic space. In *Proc. of the 17th Annual conferences of the Cognitive Science Society*, pp. 660–665, 1995.

[15] 有本卓. 確率・情報・エントロピー. 森北出版, 1980.

[16] 小嶋秀樹, 伊藤昭. 文脈依存的に単語間の意味距離を計算する一手法. 情報処理学会論文誌,

Vol. 38, No. 3, pp. 481–489, 1997.

[17] M. R. Quillian. *Semantic Information Processing*, pp. 216–270. MIT Press: Cambridge, MA., 1968.

[18] S. C. Shapiro. A net structure for semantic information strace, deduction and retrieval. In *Proc. of 2nd IJCAI*, pp. 512–523. Morgan Kaufmann: Los Altos, CA.

[19] J. F. Sowa. *Conceptual Structures: Information Processing in Mind and Machine.* Addison-Wesley, 1984.

[20] James Pustejovsky. *The Generative Lexicon.* The MIT Press, 1995.

[21] Carl Pollard and Ivan A. Sag. *Head-Driven Phrase Structure Grammar.* The University of Chicago Press, 1994.

付録: 連想情報量の計算

$$H(x) = \sum_{w \in L} p(w|x) \log p(w|x).$$

ここで, c_k を w_k の x との共起回数 $c_k = N(c_{xk})$, s をその総数 $s = \sum_{i=1}^{n} c_k$ とおけば, $p(w|x) = \frac{c_w}{s}$ であるから,

i) $c_k = 0$ のとき,

$$\begin{aligned} H' &= \sum_{i=1}^{n} \frac{c_i}{s+1} \log \frac{c_i}{s+1} + \frac{1}{s+1} \log \frac{1}{s+1} \\ &= \frac{1}{s+1} \sum_{i=1}^{n} [c_i \log c_i - c_i \{\log s + \log(s+1) - \log s\}] \\ &\quad + \frac{1}{s+1} \log \frac{1}{s+1} \\ &= \frac{s}{s+1} \sum_{i=1}^{n} \frac{c_i}{s} \log \frac{c_i}{s} + \frac{s}{s+1} \log \frac{s}{s+1} \\ &\quad + \frac{1}{s+1} \log \frac{1}{s+1} \\ &= \frac{s}{s+1} H + \frac{s}{s+1} \log s - \log(s+1). \end{aligned}$$

$$\therefore H' = \frac{s}{s+1} H + \frac{s}{s+1} \log s - \log(s+1). \quad \blacksquare$$

ii) $c_k \neq 0$ のとき,

$$c'_i = \begin{cases} c_i & (i \neq k) \\ c_i + 1 & (i = k) \end{cases}$$

とすれば,

$$\begin{aligned} H' - H &= \sum_{i=1}^{n} \frac{c'_i}{s+1} \log \frac{c'_i}{s+1} - \sum_{i=1}^{n} \frac{c_i}{s} \log \frac{c_i}{s} \\ &= \sum_{i=1}^{n} \{ \frac{c'_i}{s+1} \log \frac{c'_i}{s+1} - \frac{c_i}{s} \log \frac{c_i}{s} \} \\ &= \sum_{i \neq k}^{n} [\frac{c_i}{s+1} \{\log c_i - \log(s+1)\} - \frac{c_i}{s} (\log c_i \\ &\quad - \log s)] + \frac{c_k+1}{s+1} \log \frac{c_k+1}{s+1} - \frac{c_k}{s} \log \frac{c_k}{s} \\ &= \sum_{i \neq k}^{n} [-\frac{c_i}{s(s+1)} \log c_i + \frac{c_i}{s(s+1)} \log s + \frac{c_i}{s+1} \\ &\quad \log \frac{s}{s+1}] + \frac{c_k+1}{s+1} \log \frac{c_k+1}{s+1} - \frac{c_k}{s} \log \frac{c_k}{s} \\ &= -\frac{1}{s+1} \sum_{i \neq k}^{n} \frac{s_i}{s} \log \frac{c_i}{s} + \sum_{i \neq k}^{n} \frac{c_i}{s+1} \log \frac{s}{s+1} \\ &\quad + \frac{c_k+1}{s+1} \log \frac{c_k+1}{s+1} - \frac{c_k}{s} \log \frac{c_k}{s} \\ &= -\frac{1}{s+1} H + \frac{1}{s+1} \{-c_k \log c_k + (c_k+1) \\ &\quad \log(c_k+1) + s \log s\} - \log(s+1) \\ \therefore H' &= \frac{s}{s+1} H + \frac{1}{s+1} \{(c_k+1) \log(c_k+1) \\ &\quad - c_k \log c_k + s \log s\} - \log(s+1). \quad \blacksquare \end{aligned}$$