

文脈と意味の対応密度最大化による 教師なし語義曖昧性解消

谷垣 宏一^{1,2,a)} 撫中 達司³ 匂坂 芳典⁴

受付日 2015年6月9日, 採録日 2015年12月7日

概要: 本論文では, 大量入手が容易なラベルなしコーパスを利用して, 対象テキストの全単語の語義曖昧性解消 (all-words WSD) を実現する教師なし学習方式を提案する. all-words WSD は膨大な種類の語義を扱うことから, 高コストな語義ラベル付きコーパスの構築を必要としない教師なし方式の実現が期待される. 提案法は, コーパスにおける語の出現文脈と語義の対応のまとまりの良さに着目し, 両者の直積の密度を最大化するように確率的な対応を求めることで教師なし all-words WSD を実現する. SemEval データセットを用いた評価実験結果より, 類似文脈に出現する異なり語が十分得られれば, ラベルなしコーパスだけで信頼性の高い曖昧性解消が可能であることを示す.

キーワード: all-words, 語義曖昧性解消, 多義性解消, 教師なし学習

Density Maximization of Context-to-sense Mapping for Unsupervised Word Sense Disambiguation

KOICHI TANIGAKI^{1,2,a)} TATSUJI MUNAKA³ YOSHINORI SAGISAKA⁴

Received: June 9, 2015, Accepted: December 7, 2015

Abstract: This paper proposes a novel unsupervised method employing large amount of unlabeled text corpora for all-words word sense disambiguation (WSD), which requires to discriminate huge variety of senses, thus unsupervised methods are desired to avoid constructing costly sense-labeled corpora. Given unlabeled corpora and a dictionary, the proposed method bases on the coherent correspondences between word contexts and word senses, and finds the all-words' senses that maximize mapping density in context-to-sense product metric space. Experimental results confirmed the efficacy of our unsupervised method by showing the reliability of disambiguation if sufficient variations of word-types are provided in similar context.

Keywords: all-words, word sense disambiguation, unsupervised learning

1. はじめに

語義曖昧性解消 (word sense disambiguation: WSD) は,

テキスト中の語が, 辞書で規定されたいずれの語義で用いられているかを文脈に基づいて識別するタスクである. WSD タスクの中でも all-words タスクは, 曖昧性解消の対象を特定の語に限定せず, 与えられたテキスト中に出現するすべての語を対象に語義を識別するタスクであり^{*1}, 辞書知識を広く言語処理に活用するための基礎技術として実用化が期待される. all-words WSD はタスクの性質上, 辞書に含まれるすべての語義を潜在的に識別対象とし, 膨大

¹ 三菱電機株式会社情報技術総合研究所
Information Technology R&D Center, Mitsubishi Electric Corporation, Kamakura, Kanagawa 247-8501, Japan

² 早稲田大学国際情報通信研究センター
Global Information and Telecommunication Institute, Waseda University, Shinjuku, Tokyo 169-8050, Japan

³ 東海大学情報通信学部
School of Information and Telecommunication, Tokai University, Minato, Tokyo 108-8619, Japan

⁴ 早稲田大学基幹理工学部
School of Fundamental Science and Engineering, Waseda University, Shinjuku, Tokyo 169-8555, Japan

a) Tanigaki.Koichi@ap.MitsubishiElectric.co.jp

^{*1} たとえば評価型ワークショップ Senseval/SemEval の英語 all-words タスクでは, 2~3 の新聞記事, 延べ 5,000 語程度が与えられ, 記事中の内容語 (辞書のエントリ) が対象語 (対象トークン) として指定されることが多い. 記事中の語は互いに意味的に関係しており, 曖昧性解消結果の手がかりとして利用される [1].

なクラスを扱う*2。さらに語義の分布は品詞と比べてドメインに強く依存することが知られている [3], [4]。こうした理由から、高コストな語義ラベル付きコーパスの構築を前提とする教師あり学習では実用を見込むことが難しく、辞書知識を利用した教師なし方式がさかんに研究されてきた。

辞書知識を利用した教師なし WSD の典型的な方法は、テキスト中で対象語から一定の範囲に出現している語を文脈語とし、文脈語と対象語の語義候補の間で、語積文中の語の重複率や、辞書階層中の語義の近さなどに基づく意味的類似度を計算して、最大スコアを与える語義を見つけるというものである [1], [3], [5]。また、語義を対象語ごとに独立推定する代わりに、テキストの一定範囲に出現する語群を対象に、ページランクや最適化の手法を適用して、各語の語義を同時推定する研究もある [6], [7]。こうした教師なし WSD の先行研究は、典型的にはテキスト中で対象語から一定範囲に出現する語を文脈語として利用する。そのような直接的な文脈語が曖昧性解消の手がかりとして有効であることは明らかであるが、一方で、手がかりを直接的な文脈語に限定することは WSD の限界を狭めると見ることもできる。

辞書知識を利用する先行研究の中には、語が出現する文脈の統計的傾向を大規模コーパスから抽出して、辞書知識と併用する研究がある。McCarthy らは語の分布類似度を文脈語の重みとし、語義の意味的類似度と併用することで支配的語義 (dominant word sense) を推定する方式を提案している [8]。Agirre らは対象語近傍に出現する実際の文脈語を用いる代わりに、大規模コーパスの分布類似度による k 最近傍語を用いて語義を推定し、その結果、ドメイン依存のデータセットでは実際の文脈語を用いるよりも高い性能が得られたことを報告している [9]。これらの先行研究は対象語の出現文脈を扱うものではなく、したがって WSD の方式として提案されたものではないものの、ラベルなしコーパスから得られる統計量が辞書知識ベースの WSD にも有効であることを示唆している。

本論文では、大量入手が容易なラベルなしコーパスを利用する新しい教師なし all-words WSD 方式を提案する。提案法は、コーパスにおける語の出現文脈と語義の対応のまとまりの良さに着目し、両者の直積の密度を最大化するように確率的な対応を求めることで語義の曖昧性を解消する。提案法について、本論文は以下の構成で述べる。まず 2 章では、提案法が着目する語義の性質について述べる。次に 3 章では、上記の性質をモデルとして定式化する。4 章では、与えられたコーパスよりモデルパラメータを最適化する方法を述べる。5 章では、文脈および語義に特定の類似度を用いた提案モデルの実装について述べる。本実装を SemEval-2 英語 all-words WSD データセットに適用した評価実験結果について 6 章で述べる。7 章では、提案法の

誤り傾向を分析し、教師なし学習の信頼性と、ラベルなしコーパス大量利用時の性能改善可能性を考察する。8 章は本論文のまとめである*3。

2. 語義の性質

本章では、教師なし all-words WSD の手がかりとして本論文が着目する語義の性質について述べる。本論文は、ラベルなしコーパスにおいて、種々の語の、出現文脈から語義への対応の相対的な関係 (対応のまとまり) に注目し、単一ドメインのコーパスでは良いまとまりが得られること、すなわち類似した文脈に出現する語群が互いに類似した意味を持つことを仮定する。このように相対的な関係に注目することで、曖昧性解消に利用する手がかり語を、対象語近傍の直接的な文脈語に限定することなく、ラベルなしコーパスから大量に抽出して利用することが可能となる。

たとえば、“*Exotic plants ...*”, “*Exotic trees ...*” などの文を含むコーパスが与えられたとする (図 1)。文中の語 “*plants*” には “工場”, “植物”, “桜客” の意味があり、この文においていずれの意味で用いられているかは曖昧である。同様に、別の文中の語 “*trees*” には “樹木” と “樹状図” の意味があり、曖昧性がある。ここで、2 つの語がともに “*Exotic*” (外来種の) の修飾を受けており出現文脈が似ていることを考慮すると、上述の仮定より 2 つの語は語義においても互いに類似することが期待される。語義の意味的非類似度が図中の 2 点間距離のように与えられるとすると、語 “*plants*”, “*trees*” の語義候補のうち、互いに類似した語義である “植物” と “樹木” がそれぞれもっともらしいと解釈できる。このように、語と語義の対応が曖昧で、個々の文脈からはいずれの語義が正しいかを識別できなくても、コーパス全体で種々の語と語義の対応を比較することにより、まとまりの良い対応を発見できる。

上述した対応のまとまりの良さに関する仮定は、分布仮説 [11] と呼ばれる、語の出現文脈の分布傾向に関する性質

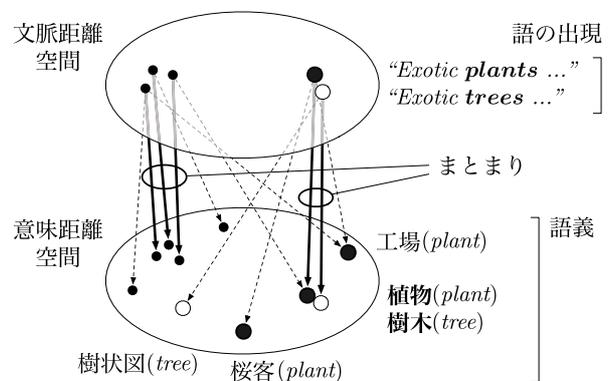


図 1 語義の曖昧性と距離空間の対応

Fig. 1 Word sense ambiguity and a mapping of metric spaces.

*2 たとえば SemEval の WSD タスクで用いられてきた WordNet [2] の英語 3.1 版では 11 万種類の概念 (synset) で語義を表す。

*3 本論文は、途中経過報告である文献 [10] の内容を発展・充実させたものである。

を、個々の語の出現（インスタンス）に対して仮定したものである。分布仮説はあくまでも分布の傾向であるため、単独で言い換えタスクなどに適用した場合には意味を識別する精度が必ずしも十分得られないことが知られているが [12]、意味の候補が辞書で規定される WSD においては、候補を識別する相対的な重みを得ることが可能と考える。

3. モデル

前章で述べた文脈と語義の対応のまとまりをモデルとして定式化する。定式化に用いる記号の一覧をあらかじめ表 1 に示しておく。いま、曖昧性解消の対象語 N 語からなるラベルなしデータセットを順序集合 $X = \{x_i\}_{i=1}^N$ とする。 x_i は i 番目の曖昧性解消対象語であり、文脈を持った個々の語の出現である。それぞれの x_i について辞書で規定される語義候補の集合を $S_i = \{s_{ij}\}_{j=1}^{M_i}$ とし、データセット N 語の S_i の順序集合を $\mathfrak{S} = \{S_i\}_{i=1}^N$ とする。ここで、各語に対し何らかの方法で語義を割り当てたとし、この割当てを決定するモデルパラメータの組を Θ とする (Θ の内容は後述する)。 Θ が、2 章で仮定した性質「類似した文脈に出現する語群は類似した意味を持つ」にどの程度従うかは、ラベルなしデータセット X と語義候補 \mathfrak{S} が与えられたときの対数尤度 \mathcal{L} を次式で定義して評価する。

$$\begin{aligned} \mathcal{L}(\Theta; X, \mathfrak{S}) & \quad (1) \\ & \equiv \ln \prod_i \sum_j P(s_{ij}|x_i, S_i, \Theta) p(x_i, s_{ij}|X_{-i}, \mathfrak{S}_{-i}, \Theta) \end{aligned}$$

ここで、 $P(s_{ij}|x_i, S_i, \Theta)$ は、語 x_i と語義の候補 S_i 、パラメータ Θ が与えられたとき、その語義、すなわち x_i の像が s_{ij} となる確率である。また、いま、ある文脈中の語 x_i によってある語義 s_{ij} が表出されるとする語義仮説を、二つ組 $\langle x_i, s_{ij} \rangle$ で表す。このとき式 (1) の $p(x_i, s_{ij}|X_{-i}, \mathfrak{S}_{-i}, \Theta)$ は、 x_i を除くデータセット X_{-i} とその語義候補 \mathfrak{S}_{-i} およ

表 1 本論文で用いる主な記号

Table 1 Major symbols used in this paper.

与えられるデータ X , \mathfrak{S} に関する記号	
X	語の出現（インスタンス）の順序集合。 $X = \{x_i\}_{i=1}^N$.
\mathfrak{S}	語義候補集合の順序集合。 $\mathfrak{S} = \{S_i\}_{i=1}^N$, $S_i = \{s_{ij}\}_{j=1}^{M_i}$.
N	X の要素数（延べ語数）.
M_i	S_i の要素数（ x_i の語義候補数）.
d_x	2 つの語の文脈距離を与える関数.
d_s	2 つの語義の意味距離を与える関数.
求めたいモデルパラメータ Θ に関する記号	
Θ	全モデルパラメータ $\Theta = \langle \pi_1, \dots, \pi_N, \mathbf{h} \rangle$.
π_i	語 x_i の語義が s_{ij} である確率 π_{ij} からなるベクトル.
\mathbf{h}	カーネルの文脈バンド幅と語義バンド幅 $\mathbf{h} = (h_x, h_s)$.
定義関数	
\mathcal{K}	文脈と意味の対応のカーネル関数.
\mathcal{R}	カーネルの外挿負担率.
\mathcal{L}	尤度関数.

び Θ に基づいて $\langle x_i, s_{ij} \rangle$ を予測する確率密度である。すなわち本方式は、文脈付きの語群から意味への対応の密度期待値を尤度として、これを最大化することにより WSD を実現する。

この尤度は leave-one-out 交差検証で定義する尤度であり、各 x_i の確率的なターゲット s_{ij} に対して、 x_i 以外の学習データ X_{-i} から確率密度を推定し、これをターゲット s_{ij} について周辺化して定義したものである。ラベルなしデータにおいてこのような交差検証が有効な理由は、all-words WSD では語義の候補が語によって異なるからであり、このため互いに類似した文脈に出現する語の間では、ある語の語義候補を、別の語の緩い教師情報として利用できるからである。

本論文では $P(s_{ij}|x_i, S_i, \Theta)$ の確率を直接、モデルパラメータ Θ の要素として扱い、パラメータ π_{ij} とおく。

$$P(s_{ij}|x_i, S_i, \Theta) \equiv \pi_{ij} \quad (2)$$

また、 x_i の各語義候補の確率をまとめてベクトル $\pi_i = (\pi_{i1}, \dots, \pi_{iM_i})$ と表す。 π_{ij} は確率としての制約条件 $\forall i \sum_j \pi_{ij} = 1$, および、 $\forall i, j 0 \leq \pi_{ij} \leq 1$ を満たす。この π_i を与えられたデータから推定するのが本論文の目的である。 π_i が定まれば、その中から確率最大の $\hat{j} = \arg \max_j \pi_{ij}$ を採択することで語義の曖昧性を解消できる。

ところで、本論文では文脈と意味の直積に連続性を仮定し、確率密度 $p(x_i, s_{ij}|X_{-i}, \mathfrak{S}_{-i}, \Theta)$ を推定する。これまで WSD の先行研究では、特徴空間で文脈を扱う際に文脈の連続性をしばしば仮定してきたが、語義は離散的な識別対象として扱うのが一般的であった [1]。本論文は、辞書に規定された膨大な語義が、本来連続な概念を量子化したものであり、それゆえ異なる語で名付けられた概念も、近傍ではスムーズな性質を持つと仮定する。このような仮定をおくことにより、ある語の語義仮説 $\langle x_i, s_{ij} \rangle$ は、類似した文脈に出現する別の語の語義仮説 $\langle x_{i'}, s_{i'j'} \rangle$ を平滑化して自然に外挿することが可能となる。外挿による確率密度は次式で得る。

$$\begin{aligned} p(x_i, s_{ij}|X_{-i}, \mathfrak{S}_{-i}, \Theta) & \\ & \equiv \frac{1}{N_{-i}} \sum_{i' \neq i} \sum_{j'} P(s_{i'j'}|x_{i'}, S_{i'}, \Theta) \\ & \quad \mathcal{K}(x_i, s_{ij}, x_{i'}, s_{i'j'}|\mathbf{h}) \end{aligned} \quad (3)$$

$$= \frac{1}{N_{-i}} \sum_{i' \neq i} \sum_{j'} \pi_{i'j'} \mathcal{K}(x_i, s_{ij}, x_{i'}, s_{i'j'}|\mathbf{h}) \quad (4)$$

ここで、 N_{-i} は X_{-i} の要素数（語の数）を表す。 \mathcal{K} は次式で定義するガウスカーネル

$$\begin{aligned} \mathcal{K}(x_i, s_{ij}, x_{i'}, s_{i'j'}|\mathbf{h}) & \\ & \equiv \frac{1}{2\pi h_x h_s} \exp \left[-\frac{d_x^2(x_i, x_{i'})}{2h_x^2} - \frac{d_s^2(s_{ij}, s_{i'j'})}{2h_s^2} \right] \end{aligned} \quad (5)$$

であり、 \mathbf{h} はこのカーネルのパラメータ h_x と h_s を便宜的にまとめて表したものである。なお、この \mathbf{h} を含めてモデルパラメータは $\Theta = \langle \pi_1, \dots, \pi_N, \mathbf{h} \rangle$ で構成する。式 (5) において、 d_x , d_s はそれぞれ文脈および語義について非類似度を与える任意の距離関数である*4。また、 h_x , h_s は文脈、語義それぞれの距離に対して平滑化の強さを制御する正の実数パラメータであり、以降ではバンド幅と呼ぶ。

本方式はカーネル密度推定 [13] の拡張である。一般的なカーネル密度推定が観測データを平滑化する方法であるのに対し、本方式は曖昧性のある仮説を扱うことを目的とする。このため、競合する仮説に確率を与え、周辺化して密度を推定する点が特徴である。語義仮説 $\langle x_i, s_{ij} \rangle$ に推定する密度は、他の語の高確率な語義仮説 $\langle x_{i'}, s_{i'j'} \rangle$ が、近傍に数多くあるほど高くなる。

本モデルによる外挿を模式的に図 2 に示す。概念的な分かりやすさを重視した表現上の制約で、文脈・語義とも 1 次元の軸として図示するが、実際には 2 点間の距離が距離関数 $d_x(\cdot, \cdot)$, $d_s(\cdot, \cdot)$ でそれぞれ定まる距離空間である。これら 2 つの距離空間が作る直積空間では、ある語の出現と、 M 個の語義候補との対応は、ある文脈を通して語義側に広がる超平面上の M 個の点であり、図中に中抜ききの円で示す 2 つ点はそうした 1 組の競合する対応を表している。対応を表す各点から上へ伸びる矢印の長さは確率の大きさを表し、矢印の始点から周囲に広がる陰影の濃さと面の高さは、確率で重み付けされたカーネルが外挿する確率密度の大きさを表している。

この外挿を用いて、他の語と語義の対応を予測する。図 2 に示した確率密度は、“*Exotic tree ...*” の文脈で出現する語 “*trees*” とその語義候補である “*樹状図*” および “*樹木*” との対応に基づく外挿を表している。この外挿は、“*trees*” と類似した文脈に出現する別の語 “*plants*” と語義の対応

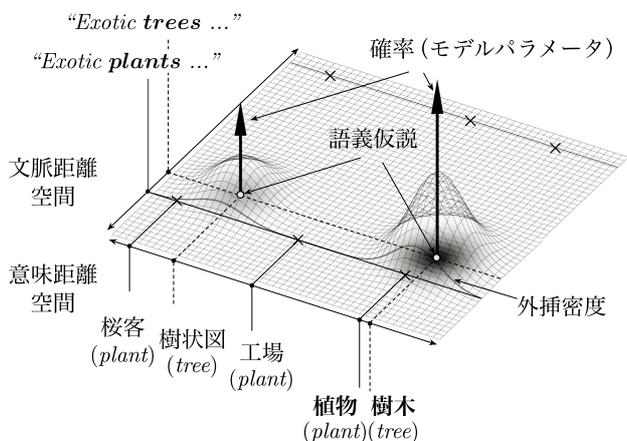


図 2 対応の外挿 (模式図)

Fig. 2 Extrapolation of mappings (schematic diagram).

*4 (非) 類似度は文脈・語義とも様々な方式が提案されており [1], [3], それら先行研究に本論文も従う。特定の非類似度を用いた実装については 5 章で述べる。

のうち、高確率な “*樹木*” に近い “*植物*” との対応を最も良く予測する。この外挿が及ぶ範囲はカーネルのバンド幅に依存する。バンド幅を適切に推定することで、近傍にある “*plants*” の語義を予測しながら、一方で紙面奥側の × 印の対応のように、文脈や語義が大きく異なる語の曖昧性解消には、“*trees*” のカーネルが影響を与えないようにすることができる*5。このように、本モデルによる曖昧性解消は、各対応に与える確率とカーネルのバンド幅の双方に依存するため、データセットに対して同時最適化する。

4. 最適化

与えられたデータセットに対しモデルパラメータを最適化することで、データセットの全語の語義曖昧性を解消する。最適なパラメータ Θ^* は、式 (1) の尤度 \mathcal{L} を、 x_i ごとの制約条件 $\forall i \sum_j \pi_{ij} = 1$ の下で最大化することによって得られる*6。この制約付き最大化の解は、ラグランジュの未定乗数法を用いると、各 x_i の制約条件に対応するラグランジュ乗数を λ_i として、次の制約なし最大化の解として求められる。

$$\Theta^* = \arg \max_{\Theta} \tilde{\mathcal{L}} \quad (6)$$

$$\tilde{\mathcal{L}} \equiv \mathcal{L} + \sum_i \lambda_i \left(\sum_j \pi_{ij} - 1 \right) \quad (7)$$

式 (7) を微分して $\nabla \tilde{\mathcal{L}} = \mathbf{0}$ とおき、 λ_i を消去して整理すると、パラメータ Θ の各要素 π_{ij} , h_x , h_s について以下の形式で最適解が得られる。

$$\pi_{ij} = \frac{\sum_{i',j'} \mathcal{R}_{ij}^{i'j'} + \sum_{i',j'} \mathcal{R}_{i'j'}^{ij}}{1 + \sum_j \sum_{i',j'} \mathcal{R}_{i'j'}^{ij}} \quad (8)$$

$$h_x^2 = \frac{1}{N} \sum_{\substack{i,i',j,j' \\ i' \neq i}} \mathcal{R}_{ij}^{i'j'} d_x^2(x_i, x_{i'}) \quad (9)$$

$$h_s^2 = \frac{1}{N} \sum_{\substack{i,i',j,j' \\ i' \neq i}} \mathcal{R}_{ij}^{i'j'} d_s^2(s_{ij}, s_{i'j'}) \quad (10)$$

ただし、 $\mathcal{R}_{ij}^{i'j'}$ は語義仮説 $\langle x_i, s_{ij} \rangle$ に対する $\langle x_{i'}, s_{i'j'} \rangle$ の外挿の負担率 (responsibility) であり、仮説の確率 π_{ij} , $\pi_{i'j'}$ を考慮して以下で定義する。

$$\mathcal{R}_{ij}^{i'j'} \equiv \frac{q(i, j, i', j')}{\sum_j \sum_{i' \neq i} \sum_{j'} q(i, j, i', j')} \quad (11)$$

$$q(i, j, i', j') \equiv \pi_{ij} \pi_{i'j'} \mathcal{K}(x_i, s_{ij}, x_{i'}, s_{i'j'} | \mathbf{h}) \quad (12)$$

ここで、式 (12) の右辺には π_{ij} と \mathbf{h} が含まれる。この

*5 紙面奥側の語の曖昧性は、図示していない別の近傍語のカーネルによって解消されることを期待している (図 3 参照)。

*6 もう 1 つの制約条件 $\forall i, j, 0 \leq \pi_{ij} \leq 1$ は明示的には扱わないが、後述する EM 法の適用においてこの制約条件が充足されることは、初期値と式 (8), (11) から明らかである。

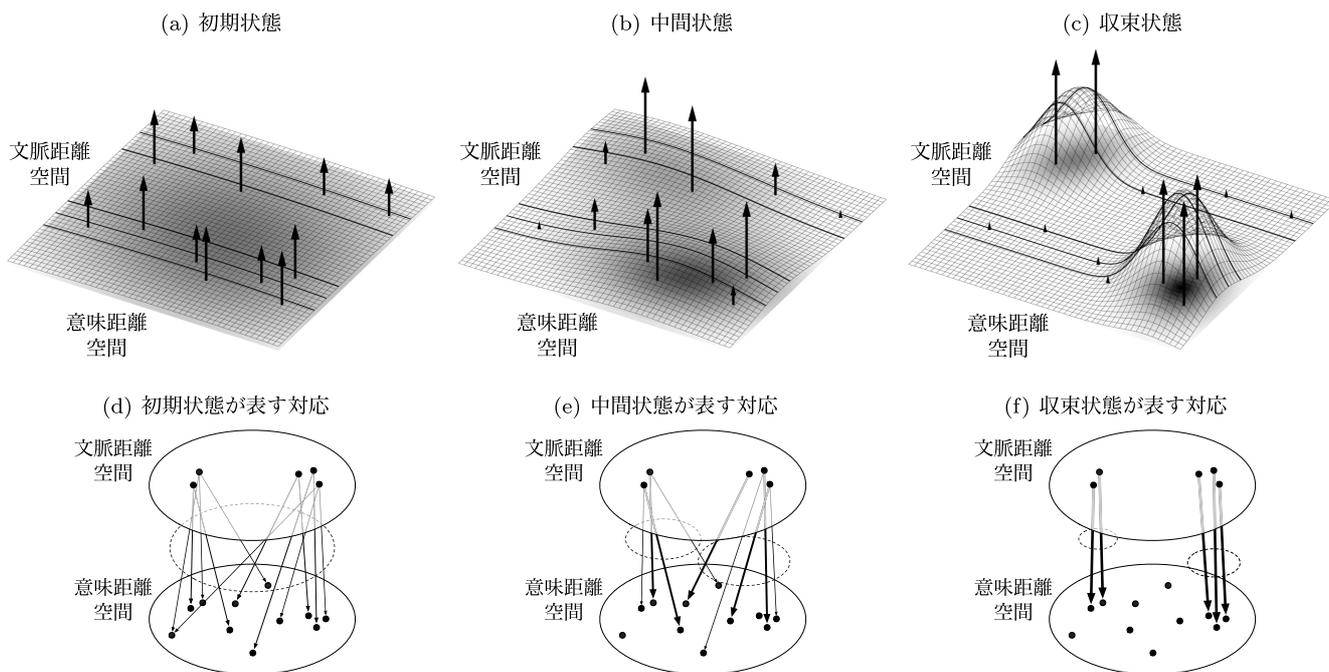


図 3 対応密度最大化による語義曖昧性解消 (模式図)

Fig. 3 Word sense disambiguation by maximizing mapping density (Schematic diagrams from pseudo 2D data simulation).

ため、式 (8)~(12) は循環定義になっており、解析的に解を求めることができない。そこで EM 法 (expectation maximization algorithm) [14] を適用して最適解を求める。EM 法を適用すると、以下のステップからなるパラメータ更新処理が得られる。

- **ステップ 1 (初期化)**：パラメータの初期値を、すべての i, j における π_{ij} と h_x, h_s に設定する。 π_{ij} の初期値は語義候補の数に応じた一様確率とし、 $\pi_{ij} \leftarrow 1/M_i$ とする。文脈と語義のバンド幅の初期値は、それぞれの距離空間におけるデータセットの平均 2 乗距離とし、文脈のバンド幅は $h_x^2 \leftarrow N^{-1} \sum_{i, i'} d_x^2(x_i, x_{i'})$ 、語義のバンド幅は $h_s^2 \leftarrow (\sum_i M_i)^{-1} \sum_{i, i', j, j'} d_s^2(s_{ij}, s_{i'j'})$ として初期化する。
- **ステップ 2 (E ステップ)**：パラメータ π_{ij}, h_x, h_s の現在の値を用いて、すべての i, j について負担率 R_{ij}^{ij} を式 (11) より求める。
- **ステップ 3 (M ステップ)**：現在の負担率 R_{ij}^{ij} を用いて、式 (8)~(10) よりすべての i, j に関するパラメータ π_{ij} と、 h_x, h_s を更新する。
- **ステップ 4 (収束判定)**：尤度を計算する。もし前回反復時と比べて尤度の増分が十分小さくなっているか、あるいは規定の反復回数に達したらパラメータ更新処理を停止する。そうでなければステップ 2 に戻って処理を継続する。

直感的には、ステップ 3 (M ステップ) において各パラメータは以下のように更新される。語義確率 π_{ij} は式 (8) より、語義仮説 $\langle x_i, s_{ij} \rangle$ がその近傍にある多数の仮説から

高い負担率で外挿されているほど、また、近傍にある多数の仮説を高い負担率で外挿しているほど、与えられたデータをもっともらしく説明する仮説として高い確率が与えられるように更新される。一方、文脈および語義のバンド幅 h_x, h_s は式 (9), (10) より、距離 d_x, d_s それぞれの分布において、実際に負担し合っている仮説間の平均 2 乗距離となるように更新される。

上述したパラメータ更新処理による語義曖昧性解消の動作を可視化するため、2次元ユークリッド空間の疑似データに本計算を適用した。この様子を図 3 上段の (a)~(c) に示す。また、図 3 下段の (d)~(f) には、上段の各状態が表す確率的な対応を模式的に示した (距離関係は厳密には上段の図と一致しない)。上段の図の紙面奥側/手前側が、下段の図では文脈距離空間の左側/右側に相当する。入力データは $N = 5$ 語のデータセットを模擬したものである。上段の図において、それぞれの語の出現文脈は、左上から右下方向へ伸びる 5 本の直線で示した。語義仮説の位置は、12 本ある上向きの矢印の始点に一致する。

EM 法の反復により、すべての語義仮説の確率とカーネルのバンド幅が同時推定される。図 3(a) の初期状態では、語義仮説はそれぞれ一様な確率を与えられている。バンド幅の初期値は平均 2 乗距離に設定され、過剰に平滑なカーネルとなる。確率密度は薄く直積空間全体を覆い、文脈と意味の対応は曖昧模糊としている。図 3(b) は 7 回目のパラメータ更新が終わった中間状態を示している。バンド幅は特に文脈側において縮退し、その結果、文脈側に 2 つのクラスタが見つかった。図 3(c) はパラメータ収束時

の状態 (25 回目の更新が終了した状態) を示している。すべての語義確率は、近傍の語義仮説間で揃って 1 または 0 に偏った解が得られており、2 章で仮定した語義の性質「類似した文脈に出現する語群が類似した意味を持つ」に従うように 5 語の曖昧性がすべて解消される。

このように、提案法は複数の語の語義仮説を相互作用させることで曖昧性を解消する。なお、データ間の相互作用はグラフ構造で扱うことも可能であり、WSD の先行研究においてもグラフベースの方式が複数提案されている。そうした先行研究では、グラフの節点を入力あるいは分類クラスのインスタンスに対応させるものが多い。たとえば、辞書知識ベースの方式では、WordNet の階層・リンク構造がしばしば利用される [6], [15]。これらは語義の概念 (synset) を節点とするグラフベース WSD である。コーパスベースの先行研究では、単語を節点としたグラフを用いる方式 [16] や語の出現文脈を節点とするグラフを用いる方式 [17] が提案されている。これらの先行研究と比べてとき、提案法は、グラフの節点に入力とクラス (語義) の間の仮説を対応させている点が特徴である。仮説を節点とする方式としては Mihalcea の先行研究 [18] があるが、これと比べて、提案法は枝を連続な空間における距離の概念の下で扱うことにより、節点が相互作用する範囲を最適化する点が特徴である。

5. 実装

本章では、後述の評価実験で実装した距離関数について述べる。先行研究において文脈の類似度と語義の類似度を同時に扱い、all-words WSD タスクに適用した例としては、McCarthy らの研究がある [8]。そこで、これと同じ類似度を本論文でも使い、距離に変換して距離関数 d_x , d_s を実装した。まず、文脈の類似度としては、構文的依存関係における共起で類似度を規定する、Lin の分布類似度 [19] を用いる。この分布類似度は語の出現ごとの類似度ではなく、語の異なり間での類似度であるため、厳密には本実装では出現ごとの曖昧性解消を目的とする WSD とはならないが、コーパスから安定した類似度を得やすいという利点がある。一方、語義の意味的類似度としては、McCarthy らの評価 [8] で最も良い性能が報告されている Jiang らの類似度 [20] と Lesk の類似度 [21] を用いる。前者の類似度は、概念辞書における分類階層の近さに基づくものであり、後者は語釈文に出現する語のオーバーラップに基づくものである。これらの意味的類似度の計算には Pedersen らの WordNet 類似度パッケージ [22] を利用した。

分布類似度および意味的類似度を距離として用いるため、 $d(\cdot, \cdot) \equiv -\alpha \ln((f(\cdot, \cdot) - f_{\min}) / (f_{\max} - f_{\min}))$ で変換して距離関数を得る。ただし、 d は d_x , d_s のいずれかの距離関数であり、 f , f_{\min} , f_{\max} はそれぞれ元の類似度関数と、その最小値、最大値を表す。また、 α はデータセット

(テストセット) において平均 2 乗距離が 1 となるよう定める正規化係数である。この正規化によりバンド幅 h_x , h_s の初期値はつねに 1 となる。

分布類似度を用いて文脈距離を構成したことにとともに、語の出現ごとの leave-one-out 交差検証で定義した尤度 \mathcal{L} を、語の異なりを単位とした交差検証に変更して計算する。これは、同種の語の出現間でつねに $d_x = 0$ となってバンド幅 h_x の推定にバイアスがかかるのを避けるためである。また、式 (4) による密度の計算には、文脈距離の k 最近傍にあるカーネルのみを用いる。これは計算を効率化するためであり、またノイズが影響しやすい分布類似度の低域を棄却することで性能の改善も期待できるからである。これらの変更を反映した実際の計算は以下となる。まず、式 (4), (8)~(11) における総和 $\sum_{i' \neq i}$ は k 最近傍の異なり語に限定した総和 $\sum_{i' \text{ s.t. } \mathcal{C}}$ で計算する。ここで、 \mathcal{C} は制約条件 $w_{i'} \neq w_i \wedge \{w_{i'}, w_i\} \in \mathfrak{N}^{(k)}$ を表し、 $\mathfrak{N}^{(k)}$ は w_i , $w_{i'}$ の少なくともどちらか一方が他方の k 最近傍であるような単語ペアの集合を表す。また、総和をとる範囲の変更により、式 (8) の分母中の 1 と、式 (9), (10) の分母 N は、実際の負担率の総和に置き換わり、それぞれ $\sum_{i', j', j' \text{ s.t. } \mathcal{C}} \mathcal{R}_{ij'}^{ij}$, $\sum_{i, i', j, j' \text{ s.t. } \mathcal{C}} \mathcal{R}_{ij}^{i'j'}$ となる。

6. 評価実験

提案法による all-words WSD の性能を評価し、教師なし辞書知識ベースの先行研究と性能を比較する。

6.1 実験条件

評価型国際ワークショップ Senseval/SemEval が公開する WSD データセットのうち、本論文執筆時点で最新の英語 all-words タスク用データセットは SemEval-2 のものであり [23]、このデータセットを本論文では用いた。本データセットは、テストデータおよびそれと同一ドメイン*7の背景データからなる。テストデータは 5,342 語 (語数は以下いづれも延べ) のテキストデータであり、うち曖昧性解消の対象語は 1,398 語 (名詞 1,032, 動詞 366) である。背景データは 270 万語のテキストであり、分布類似度の計算のみに用いた。テストデータ、背景データともプレーン・テキストとして配布されるため、辞書引きに必要な語の標準形と品詞の情報は、RASP パーザ [24] による構文解析結果より得た。

WSD 性能の評価は同タスクの規定に従い、タスクで配布される評価ツール scorer2 を用いて WordNet 語義 (synset) の厳密な一致 (fine-grained measure) による再現率と適合率を評価する。ツールが受け付ける回答形式は 2 種類あり、対象語の語義を決定的に回答する形式と、スコア付きで回答する形式のいずれかを選ぶことができる。提案法は確率

*7 SemEval-2 では特定ドメインを対象とした WSD の評価が行われ、環境ドメインのデータが用いられた [23]。

モデルであるため、特に断らない限り以下ではスコア付きの回答形式により、対象語のすべての候補語義に対し、語義確率 π_{ij} の値をスコアとして付与して評価した。本タスクにおける再現率・適合率の定義は、情報検索などにおける一般的な定義とは異なり、個々のターゲットについて出力の総和が1となるように正規化した確率カウントが用いられており、再現率は候補を複数出力しても有利にならない。そこで、以下では基本的に適合率の表記を省略し、タスクで正式なランキング基準になっている再現率のみを示す。なお、適合率と再現率の差は、前処理である品詞推定や複合語処理の誤りにより、語義の候補を WordNet から取得できなかったことによるものである。

実際の曖昧性解消の計算は、対象語を名詞と動詞の2つのデータに分けて別々に実行した。これは、今回の実装では文脈距離・意味距離とも異なる品詞間ではつねに0となって相互作用が発生せず、同時に計算する意味がないためである*8。得られた名詞・動詞の曖昧性解消結果を事後的に結合して scorer2 への入力とした。名詞・動詞とも EM 法の反復計算を 500 回行った状態を取束状態として評価した。文脈の距離は、5 章で述べたように分布類似度 [19] の k 最近傍語の関係に限定して構成する。 k の値は 10, 20, 30, 50, 100, ∞ を評価した ($k = \infty$ では最近傍語に限定せず、類似度が得られるすべての対象語との関係を用いる)。意味的類似度としては、Jiang らによる類似度 [20] と Lesk による類似度 [21] を用い、それぞれの結果を “JCN”、“Lesk” と表記する。

評価実験の目的は、語義の候補と類似度が与えられた条件で語義をスコアリングするスキーム (提案法) の効果を評価することである。この目的では、総合的なシステム性能を評価したタスクベンチマークとの比較は、必ずしも適当ではない。そこで主たる評価では、スコアリングスキームの異なる2つの先行研究を著者が再実装し、提案法と性能を比較する。比較対象の1つは、McCarthy らの方法 [8] であり、対象語との分布類似度が k 最近傍の語の語義候補を用いて、分布類似度と意味的類似度の積の総和で対象語の語義候補をスコアリングする。これを再実装して評価結果を “Mc2004” として示す。この方法は語義を対象語ごとに独立推定するのに対し、提案法はすべての対象語に同時

推定 (依存推定) する点が最大の違いである。もう1つの比較対象は、Patwardhan らの方法である [5]。これは、対象語の前後に出現している k 個の語を文脈語とし、それら文脈語の語義候補との意味的類似度の総和をスコアとして対象語の語義を推定する。ただし文脈語として使う語は、他の2手法に合わせてテストセットで対象語になっている語から選んだ。この評価結果を “Pat2007” として示す。この方法も Mc2004 同様、語義を対象語ごとに独立推定する。Pat2007 と比べた提案法の利点は、全対象語の語義を同時推定する点と、文脈の類似度を利用する点である。

6.2 実験結果

提案法と Mc2004, Pat2007 の再現率を図 4 に示す。図中、各点を中心とする縦の線分は 95% 信頼区間をブートストラップ・リサンプリング法で算出して示したものである。図中の MFS (most frequent sense : 最頻語義) は、テストセットとは別のコーパスにおける最頻語義を適用したときのベースライン性能を表し、タスクの公式報告 [23] より転記したものである。

図 4 (a), (b) から明らかなように、3 手法の性能はいずれも、利用する文脈 k 最近傍語数や意味的類似度の種別 (JCN/Lesk) に依存する。特に文脈 k 最近傍語数の条件に対しては、おおむね $k > 10$ で3手法とも右下がりの傾向が見られた。これは、分布類似度の低域では、低頻度素性の一致によるノイズが発生しやすいためと考えられ、低域を棄却して利用することで曖昧性解消性能が改善される。

3 手法の性能の相対的な関係に注目すると、利用する類似度の条件によらず、本評価実験を通じて提案法により Mc2004, Pat2007 を上回る性能が得られた。3 手法の性能差は JCN, Lesk の条件で異なり、後者で差が小さくなった。この理由としては、2つの語釈文のオーバーラップを計算するこの相対的な類似度が、まったく異なる単語ペアどうしの絶対的な近さを比較するには適さなかったことが考えら

*8 別の距離を実装する場合であれば、品詞ごとに計算を分ける必要はない。たとえば、注目する語の前後一定範囲に出現する語で構成する1次元ベクトルや、これを大規模コーパスにおける語の共起で展開して得られる2次元ベクトル [25]、あるいは、各語の分散表現を結合して構成されるパラグラフベクトル [26] などの類似度は、品詞が異なる語の間でも計算可能であり、本論文の文脈距離として利用することもできる。また、意味的類似度の導出に利用する概念辞書が品詞ごとの体系になっているため Jiang らの類似度では不可能だが、語釈文の類似度で計算する Lesk 類似度であれば、品詞が異なる語の語義間でも計算できる。これらの文脈距離と意味距離を用いることで、本実験においても名詞と動詞の語の曖昧性を同時に解消することが可能であり、少なくとも一定の名詞・動詞の関係を効果的に利用できる可能性がある。

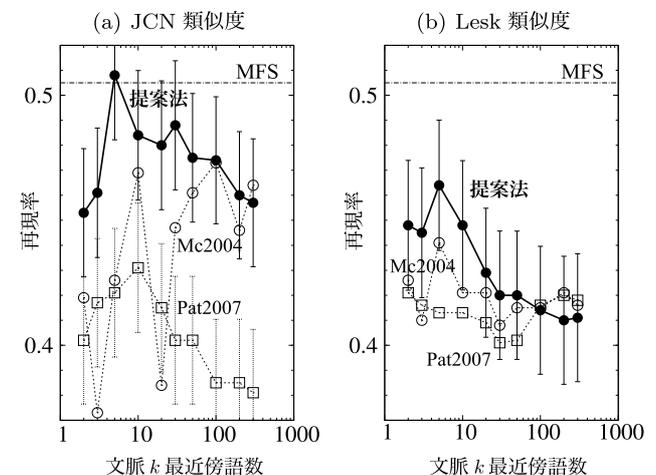


図 4 先行研究のスコアリングスキームとの性能比較
Fig. 4 Comparison with previous studies for scoring schemes.

れる。したがって、類似度に適切な正規化を加えることにより、Lesk の条件で性能が改善される可能性がある。これに対し JCN の条件では、提案法により Pat2007 と比べて有意に高い再現率が得られた。提案法による最良の再現率は $JCN/k = 5$ の条件で得られ、再現率は 0.508 であった。

6.3 SemEval-2 参加システムとの比較

参考のため、提案法による最良性能と最悪性能を SemEval-2 の参加システム [23] と比較した。提案法の性能は与えられる距離の定義に依存し (6.2 節, 図 4 参照), 最良性能は JCN ベースの意味距離と $k = 5$ 最近傍でフィルタした文脈距離との組合せで、最悪性能は $Lesk/k = 200$ との組合せで得られる。比較結果を表 2 に示す。表にはラベル付きコーパスを利用しないシステム (教師なし/知識ベースシステム) 全 20 エントリの一部を示した。“R”, “P” は全テストセットの再現率と適合率を表しており, “Rn”, “Rv” は名詞, 動詞それぞれの再現率である。順位は元の報告 [23] に倣い “R” の順としてある。この表に示すように, 提案法の最良性能は, SemEval-2 の教師なし/知識ベースシステム, および, MFS ベースラインのいずれをも上回った。一方, 最悪性能では 14 位のシステムを下回った。

なお, SemEval-2 の all-words WSD はドメイン適応をテーマに実施されており, 上位にランキングされたシステムは, それぞれに特徴的なドメイン適応処理を, 主たる曖昧性解消方式^{*9}に組み合わせている。たとえば, 1 位

表 2 SemEval-2 参加システム (教師なしのみ) との性能比較
Table 2 Performance comparison with the top five unsupervised systems in SemEval-2.

順位	参加システム	R	P	Rn	Rv
–	提案法 (ベスト)	.508	.510	.525	.462
–	MFS Baseline	.505	.505	.527	.443
1	CFILT-3 [27]	.495	.512	.516	.434
2	TreeMatch [28]	.493	.506	.516	.426
3	TreeMatch-2 [28]	.491	.504	.515	.425
4	kyoto-2 [29]	.481	.481	.487	.462
5	TreeMatch-3 [28]	.479	.492	.494	.434
...
14	IIITH2-d.r.l.ppv [30]	.422	.424	.456	.325
–	提案法 (ワースト)	.410	.412	.396	.450
15	UCF-WS-domain [31]	.392	.437	.377	.434
...
–	Random Baseline	.232	.232	.253	.172

^{*9} これら先行研究の曖昧性解消方式は, 文内など, 一定の文脈の関係を仮定できる語群の間で, 語義候補の意味的な関係だけを使って曖昧性を解消する方式が多い。これに対し提案法は, 文脈距離, 語義距離がともに保存される直積距離空間で曖昧性を解消するという点で特徴的である。提案法は計算対象の語群に一定の文脈の関係を仮定しないため, 文内に限らず, ラベルなしコーパスから手がかりとなる語を大量に抽出して利用することができる (2 章参照)。なお, 本実験結果はテストセットに閉じて計算したものであり, 大量コーパス利用時の性能評価は今後の課題である。

の CFILT-3 [27] は, グラフベースの曖昧性解消アルゴリズムを適用する際, あらかじめ辞書 (WordNet) にプルーニング処理を適用し, 対象ドメインのコーパスだけで観測された語をよく含む部分グラフに語義を限定して, 曖昧性解消の計算を行っている。2, 3, 5 位の TreeMatch [28] は, Lesk 類似度ベースの曖昧性解消アルゴリズムを適用する際, Web 検索で 3 TB の大規模ドメイン文書データを取得し, 統計量を素性への重み付けに利用している。4 位の kyoto-2 [29] は WordNet のグラフ構造に Personalized PageRank アルゴリズム [32] を適用して, 文脈語群の語義候補から互いに関連の強い語義を見つける。その際, テストセットに出現する実際の文脈語に代えて, より大規模な対象ドメインコーパスから分布類似度の k 最近傍語を抽出し, 文脈語として用いている。こうしたドメイン適応方式との組合せはシステム性能の改善に有効であり, 提案法への適用も可能と考える。

7. 考察

SemEval-2 データセットにおける動的特性と誤り傾向を分析し, 提案法による教師なし学習の信頼性と, ラベルなしコーパス大量利用時の性能改善可能性を考察する。

7.1 all-words の同時曖昧性解消効果

本節では, 4 章で模式的に示したモデルの動作 (図 3) が実際の WSD データでも同様に観測され, また正しい語義を発見するために機能したか検証する。まず, 前者について動作を確認するため, 6 章の評価実験結果より EM 法の反復計算でパラメータが変化の様子を調べた。語義確率が変化の様子を図 5 上段に示す。グラフ左端の初期状態では, 語義確率は語義候補数に応じて $1/2, 1/3, \dots$ の位

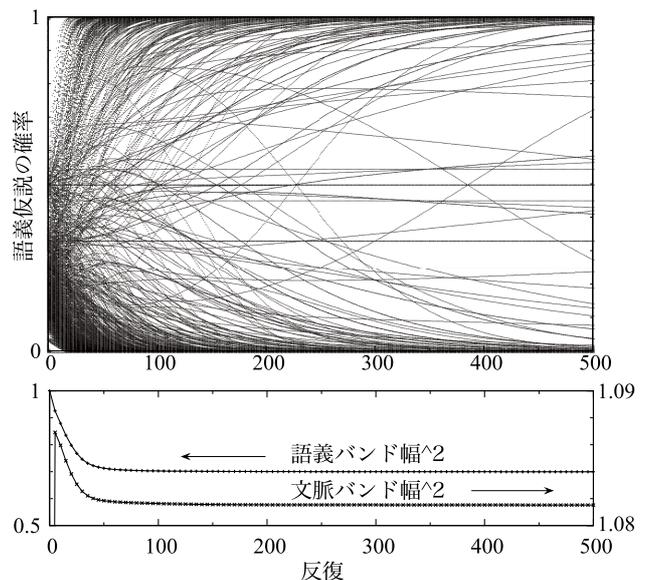


図 5 EM 法の反復によるパラメータ変化 (名詞のみ)
Fig. 5 Parameter updates by EM-algorithm (for nouns).

置にあるが、反復が進むに従い、1 または 0 のいずれかに寄っていく。グラフ右端の反復回数 500 では語義確率がほぼ完全に 1 または 0 に分離しており、曖昧性が解消されたことが分かる。また、図 5 下段のグラフは、バンド幅が変化する様子を示したものである。左側の縦軸は、文脈バンド幅 h_x^2 用の目盛を、右側の縦軸は、語義バンド幅 h_s^2 用の目盛を示している。このグラフから、反復が進むに従いバンド幅が徐々に狭まり、局所的なまとまりが発見されたことが読み取れる。これらの変化は図 3 で模式的に示した動作と符合するものであり、期待する曖昧性解消動作が実際の WSD データセットでも確認できた。

次に、上述のパラメータの反復更新が、実際に正しい語義を見つけるうえで有効に機能したかを確認する。非線形なカーネルの確率と幅を EM 法で反復更新する本方式では、初期状態と比べ、よりもっともらしい仮説間に限定して互いを外挿するようにまとまりが修正されていく。この過程で WSD の正解率も漸進的に改善されると期待される。そこで、各語の確率最大の語義候補をスコアなしで出力する「決定的評価」を反復ごとに行い、語義候補の中で順位に逆転が生じているかを調べた。本分析と次節の分析では、 k が小さく、利用する統計量が比較的少ない第 1 ピーク (JCN/ $k = 5$ および Lesk/ $k = 5$) の代わりに、 k が大きく、より典型的な特性の観測を期待できる第 2 ピーク (JCN/ $k = 30$ および Lesk/ $k = 10$) を分析した。分析結果を図 6 に示す。グラフは、反復回数に対する再現率の変化を、確率的評価・決定的評価で示したものである。なお、確率的評価とは前章までと同様に、確率をスコアとして全語義候補を出力した評価である。この図から、確率的評価だけでなく決定的評価でも、再現率は反復を通じて変化し、改善されていった様子を読み取れる。すなわち、提案法による反復計算が本実験では正しい語義を見つけるために有効に機能したといえる。

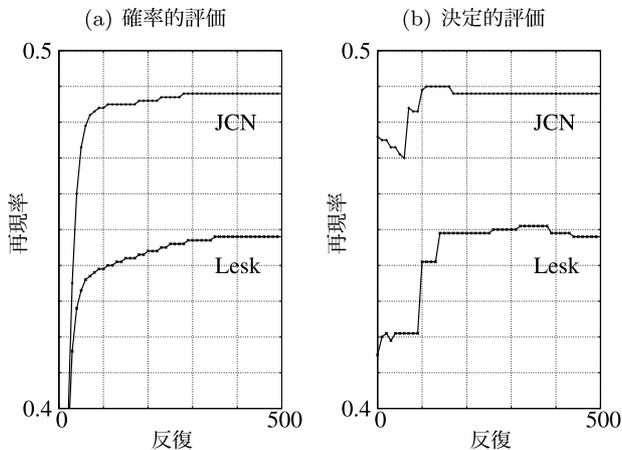


図 6 EM 法の反復にともなう再現率の変化 (全品詞)

Fig. 6 Performance improvement through iteration (all POSs).

7.2 教師なし学習の信頼性

本節では、提案法による教師なし学習の誤り傾向を分析し、パラメータ更新の信頼性について議論する。提案法は、語の出現 (文脈距離) と語義 (意味距離) の直積の分布をモデル化する。これは 2 章で述べた直積の性質「似た文脈に出現する語は似た意味を持つ傾向がある」に基づいているが、このような統計的傾向をデータから抽出するには十分な量のデータが必要である。特に直積の分布をモデル化するには、語義候補の十分なサンプルを必要とし、そのためには少なくとも、異なる語義候補を持つ異なり語を十分サンプルする必要がある。もし、語の異なりを十分得ることができず、ある語の近傍に出現する語群が特定の語に偏った場合は、語義の分布にバイアスがかかるため、学習の信頼性が低下すると予想される。

近傍語の異なりのランダムさは、単語パープレキシティによって計算できる。いま、注目する異なり語を $w \in V$ (V は語彙) とするとき、単語パープレキシティは $2^{H|w}$ で定義される。ただし、 $H|w$ は w 近傍の異なり語 $w' \in V$ に関するエントロピー $H|w \equiv -\sum_{w' \in V \setminus \{w\}} p(w'|w) \log_2 p(w'|w)$ である。ここで条件付き確率 $p(w'|w)$ は、 w の語義が近傍語 w' の影響によって決定される確率であり、確率密度比を用いて $p(w'|w) \propto \sum_{i: w_i=w} \sum_{i': w_{i'}=w'} \sum_{j, j'} q(i, j, i', j')$ で与えられる。

単語パープレキシティと学習の信頼性に相関があるかを 6 章の実験結果から分析した。単語パープレキシティと正解語義の語義確率の変化の関係を図 7 に示す (名詞のみ, JCN/ $k = 30$)。図 7 上段のグラフは反復 1~100 回目の変化量の総和を示しており、下段のグラフは 101~500 回目の変化量の総和を示している。反復 100 回の前後でグラフを分けているのは、図 5 の学習曲線より、おおむね 100 回程度でバンド幅が収束し、相互作用する範囲が落ち着くからである。図 7 において、上方向に伸びる棒グラフは

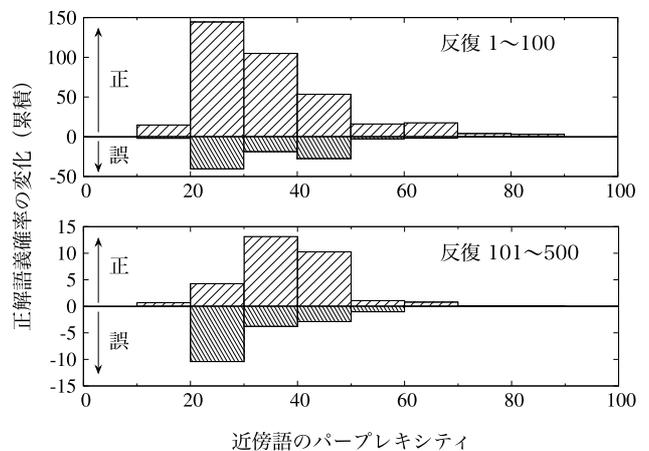


図 7 単語パープレキシティと正解語義確率の変化 (名詞のみ)

Fig. 7 Correlation between learning reliability and word-type perplexity with SemEval-2 nouns.

正解語義の確率が上がる変化（正しい変化）分の総和であり、下方向に伸びる棒グラフは正解語義の確率が下がる変化（誤った変化）分の総和である。解析の結果、図 7 右側領域にある、単語パープレキシティが 30 以上と比較的大きい語では、変化の大半にあたる 79%が正しい方向で生じた。したがって、近傍語の異なりが十分ランダムに得られれば、提案法をラベルなしデータに適用して信頼性の高い学習が可能といえる。

一方、グラフの左側領域にある、単語パープレキシティが 30 未満と小さい語では、特に下段の、反復も 101~500 と進んでバンド幅が絞られている状態において、正しい変化は全体で 32%と小さく、学習の信頼性が低い結果となった。なお、このように単語パープレキシティが低い語については、別のラベルなしコーパスから類似した文脈に出現する語を抽出してデータセットに追加することが可能であり、それによりグラフ右側領域と同等の学習の信頼性が得られ、WSD 性能が改善されると期待できる。

8. おわりに

大量入手が容易なラベルなしコーパスを利用して、対象テキスト中の全単語の語義曖昧性を解消する (all-words WSD) 方式を提案した。提案法は、コーパスにおける語の出現文脈と語義の対応のまとまりの良さに着目し、両者の直積の密度を最大化するように確率的な対応を求めることで教師なしの all-words WSD を実現する。

SemEval データセットを用いた評価実験の結果から、提案法により、文脈および意味的類似度を利用する先行研究と比べて性能が有意に改善されることを示した。また重要な性質として、類似文脈に出現する異なり語が十分得られれば、ラベルなしコーパスだけで信頼性の高い曖昧性解消が可能であることを実験的に示し、このため対象データセット中に異なり語が不足する場合でも、外部コーパスからラベルなしデータを追加することで性能の向上が期待されることを述べ、教師なし方式としての有用性を示した。

本論文では、文脈距離として比較的安定した統計量を利用できる分布類似度を実装したが、語の出現ごとの文脈距離 [25], [33], [34] を扱うことも可能であり、評価と改良は今後の課題として取り組みたい。

参考文献

- [1] Navigli, R.: Word sense disambiguation: A survey, *ACM Computing Surveys (CSUR)*, Vol.41, No.2, p.10 (2009).
- [2] Miller, G.A.: WordNet: A lexical database for English, *Comm. ACM*, Vol.38, No.11, pp.39–41 (1995).
- [3] Agirre, E. and Edmonds, P.: *Word sense disambiguation: Algorithms and applications*, Vol.33, Springer Science + Business Media (2006).
- [4] Resnik, P. and Yarowsky, D.: A perspective on word sense disambiguation methods and their evaluation, *Proc. ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How*, pp.79–86 (1997).
- [5] Patwardhan, S., Banerjee, S. and Pedersen, T.: UMND1: Unsupervised word sense disambiguation using contextual semantic relatedness, *Proc. 4th International Workshop on Semantic Evaluations*, pp.390–393 (2007).
- [6] Agirre, E. and Soroa, A.: Personalizing pagerank for word sense disambiguation, *Proc. 12th Conference of the European Chapter of the Association for Computational Linguistics*, pp.33–41 (2009).
- [7] Reddy, S. and Inumella, A.: WSD as a distributed constraint optimization problem, *Proc. ACL 2010 Student Research Workshop*, pp.13–18, Association for Computational Linguistics (2010).
- [8] McCarthy, D., Koeling, R., Weeds, J. and Carroll, J.: Unsupervised acquisition of predominant word senses, *Computational Linguistics*, Vol.33, No.4, pp.553–590 (2007).
- [9] Agirre, E., de Lacalle, O.L., Soroa, A. and Fakultatea, I.: Knowledge-based WSD on specific domains: Performing better than generic supervised WSD, *Proc. 21st International Joint Conference on Artificial Intelligence*, pp.1501–1506 (2009).
- [10] Tanigaki, K., Shiba, M., Munaka, T. and Sagisaka, Y.: Density Maximization in Context-Sense Metric Space for All-words WSD, *Proc. 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, Association for Computational Linguistics, Vol.1, pp.884–893 (2013).
- [11] Harris, Z.S.: Distributional structure, *Word* (1954).
- [12] 乾健太郎: 自然言語処理と言い換え, *日本語学*, Vol.26, No.13, pp.50–59 (2007).
- [13] Parzen, E.: On estimation of a probability density function and mode, *The Annals of Mathematical Statistics*, Vol.33, No.3, pp.1065–1076 (1962).
- [14] Dempster, A.P., Laird, N.M. and Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B (Methodological)*, pp.1–38 (1977).
- [15] Navigli, R. and Lapata, M.: Graph connectivity measures for unsupervised word sense disambiguation, *Proc. 20th International Joint Conference on Artificial Intelligence*, pp.1683–1688 (2007).
- [16] Véronis, J.: HyperLex: Lexical cartography for information retrieval, *Computer Speech & Language*, Vol.18, No.3, pp.223–252 (2004).
- [17] Niu, Z.-Y., Ji, D.-H. and Tan, C.L.: Word sense disambiguation using label propagation based semi-supervised learning, *Proc. 43rd Annual Meeting on Association for Computational Linguistics*, pp.395–402 (2005).
- [18] Mihalcea, R.: Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling, *Proc. Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp.411–418 (2005).
- [19] Lin, D.: Automatic retrieval and clustering of similar words, *Proc. 17th International Conference on Computational Linguistics*, Vol.2, pp.768–774 (1998).
- [20] Jiang, J.J. and Conrath, D.W.: Semantic similarity based on corpus statistics and lexical taxonomy, *arXiv preprint cmp-lg/9709008* (1997).
- [21] Lesk, M.: Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone, *Proc. 5th Annual International Conference on Systems Documentation*, pp.24–26 (1986).

- [22] Pedersen, T., Patwardhan, S. and Michelizzi, J.: WordNet::Similarity: Measuring the relatedness of concepts, *Demonstration Papers at HLT-NAACL 2004*, pp.38–41 (2004).
- [23] Agirre, E., de Lacalle, O.L., Fellbaum, C., Hsieh, S.-K., Tesconi, M., Monachini, M., Vossen, P. and Segers, R.: SemEval-2010 task 17: All-words word sense disambiguation on a specific domain, *Proc. 5th International Workshop on Semantic Evaluation*, pp.75–80 (2010).
- [24] Briscoe, T., Carroll, J. and Watson, R.: The second release of the RASP system, *Proc. COLING/ACL on Interactive Presentation Sessions*, pp.77–80 (2006).
- [25] Purandare, A. and Pedersen, T.: Word sense discrimination by clustering contexts in vector and similarity spaces, *Proc. Conference on Computational Natural Language Learning*, Boston, pp.41–48 (2004).
- [26] Le, Q. and Mikolov, T.: Distributed Representations of Sentences and Documents, *Proc. 31st International Conference on Machine Learning (ICML-14)*, pp.1188–1196 (2014).
- [27] Kulkarni, A., Khapra, M.M., Sohoney, S. and Bhattacharyya, P.: CFILT: Resource Conscious Approaches for All-Words Domain Specific, *Proc. 5th International Workshop on Semantic Evaluation*, pp.421–426 (2010).
- [28] Tran, A., Bowes, C., Brown, D., Chen, P., Choly, M. and Ding, W.: TreeMatch: A Fully Unsupervised WSD System Using Dependency Knowledge on a Specific Domain, *Proc. 5th International Workshop on Semantic Evaluation*, pp.396–401 (2010).
- [29] Soroa, A., Agirre, E., de Lacalle, O.L., Monachini, M., Lo, J., Hsieh, S.-K., Bosma, W. and Vossen, P.: Kyoto: An integrated system for specific domain WSD, *Proc. 5th International Workshop on Semantic Evaluation*, pp.417–420 (2010).
- [30] Reddy, S., Inumella, A., McCarthy, D. and Stevenson, M.: IIITH: Domain specific word sense disambiguation, *Proc. 5th International Workshop on Semantic Evaluation*, pp.387–391, Association for Computational Linguistics (2010).
- [31] Schwartz, H.A. and Gomez, F.: UCF-WS: Domain word sense disambiguation using web selectors, *Proc. 5th International Workshop on Semantic Evaluation*, pp.392–395, Association for Computational Linguistics (2010).
- [32] Haveliwala, T.H.: Topic-sensitive pagerank, *Proc. 11th International Conference on World Wide Web*, pp.517–526, ACM (2002).
- [33] Thater, S., Fürstenaу, H. and Pinkal, M.: Word Meaning in Context: A Simple and Effective Vector Model, *Proc. 5th International Joint Conference on Natural Language Processing (IJCNLP)*, pp.1134–1143 (2011).
- [34] Schütze, H.: Automatic word sense discrimination, *Computational Linguistics*, Vol.24, No.1, pp.97–123 (1998).



谷垣 宏一 (正会員)

1995年東北大学大学院情報工学科博士前期課程修了。同年三菱電機株式会社入社。1997年～2000年ATR音声翻訳通信研究所。現在、三菱電機情報技術総合研究所。2012年より早稲田大学国際情報通信研究センター招聘研究員。自然言語処理、音声言語処理の研究に従事。博士(国際情報通信学)。言語処理学会会員。



撫中 達司 (正会員)

1986年三菱電機入社。オペレーティングシステム(OS)、ネットワークシステム、ネットワークセキュリティの研究開発に従事。2015年より東海大学情報通信学部組込みソフトウェア工学科教授。製造業等の産業用システムへのIoT活用に関する研究開発に従事。第21回電気通信普及財団テレコムシステム技術賞受賞(2006年)、第62回電機工業技術功労表彰奨励賞受賞(2012年)。博士(工学)。IEEE Senior member, 電子情報通信学会会員, 本会シニア会員。



匂坂 芳典

1973年早稲田大学理工学部物理学科卒業。1975年同大学大学院修士課程修了。同年日本電信電話公社(現, NTT)武蔵野電気通信研究所入社。1986年より国際電気通信基礎技術研究所(ATR)に出向。2001年より早稲田大学教授, 現在の所属は早稲田大学基幹理工学部応用数理学科。工学博士。音声合成・音声認識を中心とした音声情報処理, 言語情報処理の研究に従事。日本音響学会, 電子情報通信学会, IEEE, 米国音響学会各会員。