

フィルタリングのための 隠語の有害語意検出機能の 意味解析システム SAGE への組み込み

橋本 広美[†] 木下 嵩基^{††}
原田 実^{††}

概要：原田研究室では、文章中の単語の語意の決定および係り受け関係にある2文節間の深層格の決定を行う意味解析システム SAGE の開発を行ってきた。しかし、従来の EDR の共起辞書を用いた語意決定方法では、隠語において、隠された語意と一般の語意とを区別するための共起レコードの登録が膨大になりその実現が難しかった。本研究では、隠語の有害語意と文脈に登場する他の語（周辺語）の語意との共起頻度を辞書化しこれを基に隠語の語意を決定することで、有害語意検出精度を98%に向上できた。

The function that detect harmful word sense from slang built into the semantic analysis system SAGE for filtering

Hiromi Hashimoto[†], Takanori Kinoshita^{††}
and Minoru Harada^{††}

Abstract: We have developed semantic analysis system SAGE which decides the meaning of a word in sentences and the deep case between two clauses having dependency relation. However, deciding the meaning of a word in SAGE requires the collocation dictionary of EDR, and the registration of the co-occurrence record to distinguish a harmful meaning and a general meaning of slang becomes huge, so the achievement was difficult. In this research, the harmful meaning detection accuracy was improved to 98% by making a dictionary containing the co-occurrence frequency of harmful meaning of slang with the meaning of other words appeared in the surroundings and deciding the meaning of slang based on this.

[†]青山学院大学大学院 理工学研究科理工学専攻 知能情報コース
^{††}青山学院大学 理工学部 情報テクノロジー学科

1. 背景・研究目的

近年の IT の急激な発展によって、大量の文書データからの知識の発掘（テキストマイニング）などの分野で、文章の意味解析への期待が高まっている。

原田研究室では、EDR 電子化辞書[1]に記載された情報を元に、文章中の単語の語意の決定および係り受け関係にある2文節間（主辞同士）の深層格の決定を行う意味解析システム SAGE[2][3][4][5][6]の研究開発を行ってきた。

一方で、誰もが自由に web ページを閲覧すること、作成することができるようになった結果、有害情報を含む記事を未成年者に見せないようにする必要性に迫られている。

しかし従来の EDR の共起辞書を用いた語意決定方法では、隠語において、隠された語意と一般の語意とを区別するための共起レコードの登録が膨大になりその実現が難しかった。また、係り受け木で葉に当たる語においては単語辞書における語意の出現頻度のみから決定するので有害語意を正しく検出できなかった。本研究では、この問題を解決するために、語の語意は、周りに出現する語（周辺語）によって決定されるという判断を元に改良を行う。すなわち、当該語と直接係り受け関係にないが、その周辺に(実際は一定範囲の前に)出現して文脈でのテーマを表しているような語との共起頻度を語意決定に用いることで、隠語における語意決定の精度向上を行う。

2. 従来の SAGE

2.1 基本的考え方

SAGE は日本語を意味解析し、結果を文節や形態素ごとにそれらの意味や品詞や深層格（他の文節との役割的關係）などを保持したリストの集合として表現する。これは、文節を頂点、係り受け関係にある文節間の深層格を辺と考えると、図 1 のような意味グラフとして表現される。

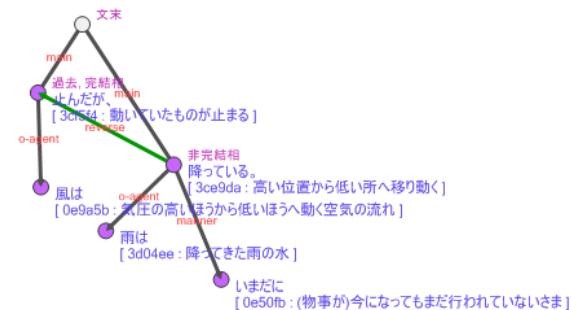


図 1: SAGE の意味解析結果を示す意味グラフ

図 1 において、紫色の丸は文節を、白色の丸は文末を表し、文節間の矢線は係り受け関係および深層格を表示している。格の向きは、係り先→係り元とし、黒い辺は係り受け関係にある文節間の深層格関係、緑の辺は並列の深層格を表す。表示される語意は各文節の主辞（主要となる形態素）が表す概念を EDR 辞書の概念 ID と語意見出しで表示した。なお、文中の最後の文末節は主述語の文節への main 格を付与している。

2.2 システム概要

SAGE の解析手順を、図 2 を用いて説明する。SAGE の処理に移る前に、まず JUMAN と KNP[7] によって日本語文章の形態素解析および係り受け解析を行う。その後、各文節・形態素ごとにそれ自身の主辞及びその他の構成形態素の語意決定をすると共にこの文節に係る文節との間の深層格を決定する処理を行う。これらの決定においては EDR の共起辞書を参照し、原文と共起事例との類似性を元に統計的に語意と格を決定する。具体的には、受け側文節主辞 u 、係り側文節主辞 k 、係り側文節の共起関係子 j を引数 (u, j, k) にして共起辞書を検索し、該当レコードが語意 μu と $m k$ と深層格 $d j$ を持つ頻度を計算し、これを元に 3 次元の語意-格総合評価値配列 $P(\mu u, d j, m k)$ を作成する。語意の決定においては、葉の方から計算し、語意確率 $P(\mu u, d j, m k)$ を最大にする語意 $m k$ を採用し、葉以外では、係り側の決定された語意 $m k_0$ を固定して、 $P(\mu u) = \sum_{d j} P(\mu u, d j, m k_0)$ として計算し、これを最大にする語意 μu を採用する。

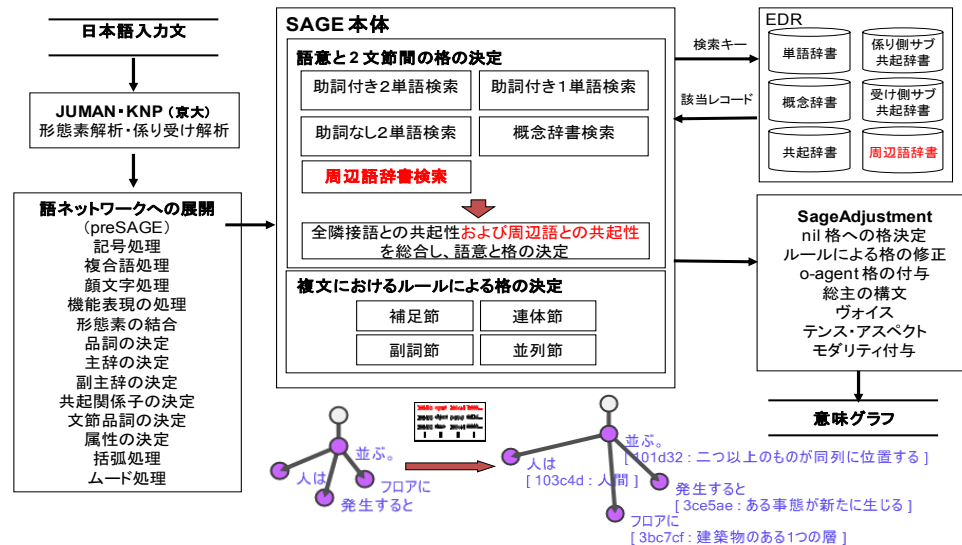


図 2: SAGE における処理の流れ

出力結果は図 3 のように表示される。文節や形態素の概念 ID、品詞 ID、深層格 ID 等必要な情報を出力している。図 1 はこれを視覚化ツール vivi で表示したものである。

```
[sg_v200]
f: 1, 風は、は, ME, 2, [, [, [, ], [, ]
s: 2, 風, カゼ, , 0e9a5b, FTM, JN1, ,
s: 3, は, ハ, , 3ca448, FJJ, JJO, ,
f: 4, 止んだか、, , DO, 5, [, [oa], [, [, ], [, ], [断定, 現実, 判定, 非完結相]
s: 5, 止んだ, ヤンダ, 止む, 3cf5f4, DOS, JVE, 子音動詞マ行, タ形
s: 6, か, カ, , 3ca448, SEJ, JJO, ,
s: 7, , , , 10010b, TOT, JSY, ,
f: 8, いまだに、, φ, HU, 9, [, [, [, ], [, ], [, ]
s: 9, いまだに, イマダニ, , 0e50fb, FUK, JD1, ,
f: 10, 雨は、は, ME, 11, [, [, [, ], [, ]
s: 11, 雨, アメ, , 3d04ee, FTM, JN1, ,
s: 12, は, ハ, , 3ca448, FJJ, JJO, ,
f: 13, 降っている。、, , DO, 14, 15, [ma8, oa10], [rv4], [, ], [断定, 現実, 判定, 非完結相]
s: 14, 降って, フツテ, 降る, 3ce9da, DOS, JVE, 子音動詞ラ行, タ系連用テ形
s: 15, いる, イル, , 0e52f0, DOB, JAX, 母音動詞, 基本形
s: 16, , , , 0ee33b, KUT, JSY, ,
e: 17, null, null, [mn4, mn13]
```

図 3: SAGE による意味グラフ出力結果例

3. 周辺語の定義

周辺語とは、当該語よりも前に出てくる一定範囲内の語で、文脈のテーマを表すような語を示す。

例えば、「都内の指定駅にて S を手渡しします。S と草のユーザーのみ、取引実績のある顧客のみです。」「牧場に遊びに行った。ヤギがおいしそうに草を食べていた。」という 2 つ例文では、「草」という語の語意が異なっている。この 2 つの「草」という語は、共に意味グラフにおける葉になるため、従来の語意決定においては、共起辞書から、係り側、受け側の表記と共起関係子をもとに検索した場合のみから語意確率を求め、「植物としての草」という語意に決定されていた。

しかし、前者の例文においては、「草」は「覚醒剤という薬剤」という意味をもち、周辺に「都内」、「指定駅」、「S」、「手渡し」が出現している。後者においては、「草」は「植物としての草本」という意味をもち、周辺に「牧場」、「ヤギ」が出現している。これらの語の語意を周辺語として登録し、意味解析に利用する。


```

f: 1, Sと,と,ME,2,,[],[],[],S,[]
   s: 2, S,,,0aa858,KIG,JSY,,
   s: 2, S,,,Of5370,KIG,JSY,,
   s: 3, と,と,,3ca448,KKJ,JO,,
f: 4, 草の,の,ME,5,,[],[an1],[],[]
   s: 5, 草,クサ,,Oedddb,FTM,JN1,,
   s: 5, 草,クサ,,Of5370,FTM,JN1,,
   s: 6, の,ノ,,3ca448,SEI,JO,,
f: 7, ユーザーのみ,のみ,ME,8,,[mo1,mo4],[],[],[]
   s: 8, ユーザー,,,3be933,FTM,JN1,,
   s: 9, のみ,ノミ,,3ca448,FJJ,JO,,
   s: 10, ,,,,10010b,TOT,JSY,,
f: 11, 取引実績の,の,ME,13,12,[],[],[],[]
   s: 12, 取引,トリヒキ,,100c7a,SAM,JN1,,
   s: 13, 実績,ジツセキ,,Of73bf,FTM,JN1,,
   s: 14, の,ノ,,3ca448,KKJ,JO,,
f: 15, ある,φ,ME,16,,[oa11],[],[],[]
   s: 16, ある,アル,,1fa2f7,FTM,JN1,
f: 17, 顧客のみです,.,DA,18,,[ma7,ao15],[],[],[断]
   s: 18, 顧客,コキヤク,,3cf103,FTM,JN1,,
   s: 19, のみ,ノミ,,3ca448,FJJ,JO,,
   s: 20, です,デス,だ,2621ba,HAN,JD,判
   s: 21, ,,,,0ee33b,KUT,JSY,,
e: 22, null,null,[mn17]
    
```

図 6: 隠語を含む掲示板に対する SAGE 解析結果の例

5. 周辺語辞書

周辺語データベースを語意決定に利用するため、頻度から語意確率を計算し、登録した周辺語辞書を作成する。語意確率として、語意を決定したい語 x があり、その前にある周辺語 y の語意の第 5 階層上位概念が s である時、語 x の語意が m である確率 $P(m,s)$ を以下のように定義する。

$N(m,s)$: 周辺語 s と、語意 m が同時に現れる頻度

M : 全コーパス内に出現するすべての名詞の語意

$$p(m, s) = \frac{N(m, s)}{\sum_{x \in M} N(x, s)}$$

周辺語辞書レコードの形式は、表 2 のようにする。

表 2: 周辺語辞書レコードの形式

m	s	P(m, s)
---	---	---------

確率計算の例を表 3 に示す。ここでは、「S」という語があり、その前にある周辺語の第 5 階層上位概念が「10a479」（店）である時、「S」の語意が「0f5370」（覚せい剤という薬剤）である確率を求めている。

表 3: 周辺語辞書の確率計算の例

レコード番号	m	s	N(m,s)	
SAW00004729	0ff865	天然	10a479	1
SAW00011750	10082a	都内	10a479	1
SAW00013610	101aa8	夏	10a479	1
SAW00016970	10082a	都内	10a479	4
SAW00024524	3cf180	状態	10a479	5
SAW00030053	3cf2a9	一部	10a479	1
SAW00001602	10a479	店	10a479	7
SAW00011859	0f644a	お客	10a479	4
SAW00019190	0c2918	氷	10a479	1
SAW00026629	103c4d	人	10a479	3
SAW00028522	0f512f	ビル	10a479	1
SAW00023797	3cec66	質問	10a479	4
SAW00027266	0f8ebd	炭	10a479	1
SAW00031279	3ce7e6	地方	10a479	1
SAW00005847	1f35bb	九州	10a479	1
SAW00013312	3be1d9	買い物	10a479	4
SAW00016062	0f5370	S	10a479	4
SAW00016298	1faf20	紹介	10a479	1
SAW00019768	0f697d	新規	10a479	4
SAW00028866	1faea4	女性	10a479	4
SAW00030599	1f2518	信州	10a479	1
SAW00030830	10a479	店舗	10a479	3
SAW00033266	1fae9b	住宅	10a479	1

表 3 の例を計算すると以下ようになる。

$$N("0f5370", "10a479") = 4$$

$$P(0f5370, 10a479) = \frac{N("0f5370", "10a479")}{\sum_{x \in M} N(x, "10a479")} = \frac{4}{58} = 0.069 \dots$$

6. 周辺語を用いた語意決定方法

6.1 周辺共起確率ベクトル

周辺共起確率ベクトル R は、語意を決定したい語 x における語意候補 $M\{m_1, m_2 \dots m_i\}$ ごとに周辺語の第 5 階層上位概念リスト $S\{s_1, s_2 \dots s_j\}$ を用いて次式のように求める。

$$R(m) = \sum_{y \in S} P(m, y)$$

周辺語を取得する範囲については、解析する文章によって適切な範囲を設定する。新聞記事などの場合は、文脈がはっきりしているため、少なめで良い。しかし、掲示板など、様々な人が書き込む文章においては、文脈がはっきりしないことが多いため、文章全体の話題を得るために、周辺語の取得範囲を多くとる。

6.2 意味グラフにおける葉の語意決定

従来の SAGE では 2.2 節で論じたように、意味グラフの各節の語 k の語意 mk は、共起辞書から、係り側、受け側の表記と共起関係子をもとに検索し、求めた語意確率ベクトル P を用いて決定していた。しかし、この方法では隠語などの正しい語意をとることはできない。

そこで、従来の語意決定に用いられていた語意確率ベクトル P の値に、周辺語辞書より求めた周辺共起確率ベクトル R を加算した和ベクトルを用い、この和ベクトル内で最大の確率の語意 mk を語 k の語意と決定する。これにより、辞書中の頻度だけでなく、文脈を考慮した語意決定が可能になる。

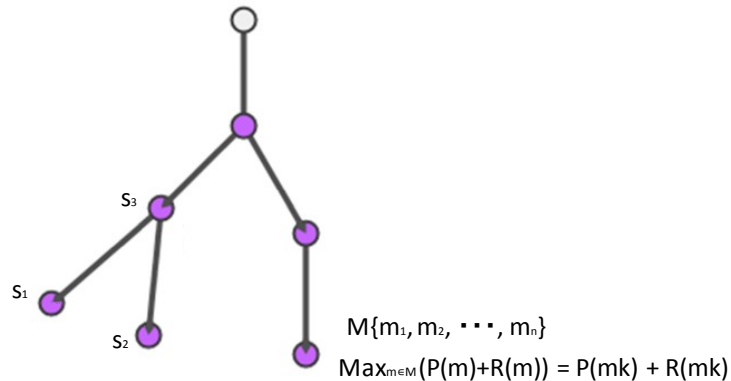


図 7: 葉の語意決定

6.3 意味グラフにおける葉以外の語意決定

意味グラフの葉以外の語では、係り側の語意がすでに決定されているため、共起辞書から求めた 3 次元の語意-格総合評価値配列 $P(\mu, dj, mk)$ の係り側の決定された語意 mk_0 を固定して、 $P(\mu) = \sum_{dj} P(\mu, dj, mk_0)$ として計算しこれを最大にする語意 μ を採用していた。葉の語意決定に比べると、精度は良いが、係り受け関係にある語だけでは、隠語の判定は難しい。そこで葉の場合と同様に、従来の確率ベクトル P に周辺共起確率 R を加えた和ベクトルを利用することで、より広範囲の語を手掛かりとして語意を決定することが可能になる。

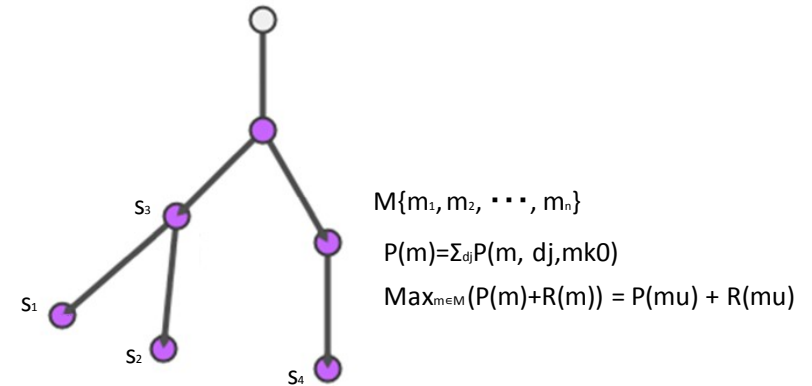


図 8: 葉以外の語意決定

6.4 未知語の語意決定

未知語とは、共起辞書において語意決定したい係り側、受け側、共起関係子の組に該当するレコードがなかった場合や、係り側の語意が決定されたことで該当レコードがなくなった語を示す。このような場合は従来、単語辞書の語意別頻度から求めた語意確率ベクトル P のみで語意決定を行っていた。この方法では、文脈などは全く考慮されないことになる。これを解決するために、語意確率ベクトル P と周辺共起確率ベクトル R の和を利用し、共起関係にはない語からも情報を得られるようにした。

7. 実験及び評価

本研究では有害文における隠語の語意決定精度向上を目指した。よって、評価実験は、一般的な語意と表記の同じ隠語についての正誤判定を行った。実験に用いる例文には、隠語が多く含まれる web 上の掲示板を用いた。総語意数 13871 のうち、一般的な語意と表記の同じ隠語 198 語を評価対象とする。

周辺語辞書の追加による隠語判別精度を調べるため、周辺語を利用する場合、利用しない場合それぞれについて、解析を行い、解析結果より、隠語の語意を確認する。このとき、有害語であることが明らかな表記のものは集計しない。

表 4：評価実験結果

	周辺語なし	周辺語あり
有害語検出数	77/198	195/198
有害語検出率	38.9%	98.5%

集計の結果は、表 4 のようになった。

周辺語を利用した場合は、利用しなかった場合に比べて有害語の判定率が、38.9% から 98.5% に向上した。周辺語辞書の追加によって、隣接する語以外の語からの情報を得ることにより、有害であることを判断しにくい隠語についても、正しい語意を割り当てることが可能になったといえる。

有害語意検出率が向上した要因としては、隠語の周りには、特徴的な単語が多く現れることが多い、ということが考えられる。

現段階では、語意を確認できた隠語についてのみ登録を行ったが、実際のフィルタリングへの応用のためには、さらに多くの隠語を確認し、登録する必要がある。隠語は、すぐに新しいものが生まれ、語意自体も変化してゆくことが多いため、最新のものを追加してゆく作業には手間がかかるが、周辺語辞書の構築は半自動的に進めるため、隠語と係受け関係にある語を多数用意してそれらの語意を決定して共起辞書を更新することに比べると、比較的容易に行える。

今後の課題としては、周辺語収集方法の充実により、多くの隠語に対応し、実際のフィルタリングへの応用を目指す。

参考文献

- 1) (株)日本語電子辞書研究所: EDR 電子化辞書仕様説明書(第2版), (株)日本語電子辞書研究所(2002).
- 2) 原田実, 尾見孝一郎, 岩田隆志, 水野高宏: 日本語文章からの意味フレーム自動生成システム SAGE(Semantic frame Automatic GEnerator)の開発研究, 人工知能学会第13回全国大会論文集, pp. 213-216 (1999).
- 3) 原田実, 水野高宏: EDR を用いた日本語意味解析システム SAGE, 人工知能学会論文誌, Vol.16, No.1, pp.85-93 (2001.1).
- 4) 原田実, 田淵和幸, 大野博之: 日本語意味解析システム SAGE の高速化・高精度化とコーパスによる精度評価, 情報処理学会論文誌, Vol.43, No.9, pp.2894-2902(2002.9).
- 5) 川口純一, 青木洋, 松田源立, 原田実: 意味解析システム SAGE の精度向上, 情報処理学会第69回全国大会論文集, 1C-04, 第2分冊, pp. 77-78 (2007.3).
- 6) 梅澤俊之, 加藤大知, 松田源立, 原田実: “意味解析システム SAGE の精度向上—モダリティと副詞節について—”, 情報処理学会研究報告, Vol.2009-NL-191 No.4, pp. 1-8, (2009.5).
- 7) 京都大学情報学研究科知能情報学専攻能メディア講座言語メディア研究室(黒橋研究室), <http://nlp.kuee.kyoto-u.ac.jp/>
- 8) 独立行政法人 情報通信研究機構: 平成21年度 新規委託研究「インターネット上の違法・有害情報の検出技術の研究開発」 研究計画書 http://www2.nict.go.jp/q/q265/s802/info/20090422koubo/theme_b001_koubo.pdf