

感情状態に基づく因果関係推論の一般化

井之上 直也 岡崎 直観 乾 健太郎

東北大学 大学院情報科学研究科

{naoya-i,okazaki,inui}@ecei.tohoku.ac.jp

1 はじめに

計算機の文章理解において重要な常識的知識として、事象間の因果関係の知識 (e.g., “怒り” が “攻撃” を引き起こす) がある。事象間の因果関係は、人手により書き尽くすことは困難と考えられており、近年では、大規模な Web コーパスを利用して確率統計的に因果関係の自動獲得を行う研究が盛んである [2, 13, 8, etc.]。

因果関係の知識の構築における重要な問題の一つに、事象の表現方法の問題がある。先行研究では (以降 “*John got an icecream for Mary.*” を例にとって説明する)、述語 (*get*) [2]、述語と項の組み合わせ (*get icecream*) [13, 16, 10]、分散表現 (*get, icecream* の単語ベクトルから構造的に計算) [8, 19] など、さまざまな表現方法が提案されている。これらの手法で獲得された因果関係の知識は、実際に言語解析への応用も始まっているが [12, 16, 10, etc.]、その効果は限定的であるという報告もある [18]¹。

そこで本稿では、登場人物の感情状態 (*affect state*) の変化に基づいて物語理解を行う Lehnert のプロット理論 [5] にヒントを得て、因果関係を表す事象の一般化の粒度として感情状態を採用し、その有用性を検証する。感情状態とは、人の感情に関する状態の総称であり、心理学・認知科学・言語学を中心として、その具体的な定義についてさまざまな議論がなされてきた [5, 7, 11, etc.]。自然言語処理の分野で最もよく用いられているのは、ポジティブ・ネガティブの 2 値の感情状態であるが、Rahman ら [12] は、この 2 値の感情状態に基づいて因果関係の認識を行い (i.e., 「ポジティブな (ネガティブな) 性質を帯びた名詞は、ポジティブな (ネガティブな) イベントの動作主になりやすい」)、共参照解析タスクにおいてその有効性を示した。例えば、例 (1) において、*he* は “angry” な状態 (=ネガティブな状態) であるため、“slap” イベント (=ネガティブなイベント) の動作主になりやすい、という因果関係が認識され、*he_(i)* の参照先が *John_(i)* に同定される:

(1) [*John_(i)*] slapped [*Bob_(j)*] because *he_(i)* was angry.

このようなアプローチは、多様な因果関係を説明できるように思われるが、2 値の感情状態に基づく一般化では、「感情的」「道徳的」といった観点の区別がないため、過度な一般化が起きてしまうという問題がある。例えば、

¹杉浦ら [18] は、Web コーパス約 7 億文から獲得した因果関係の知識を共参照解析に適用し、エラー分析を行っているが、知識の質・量の面で多くの改善の余地があると報告されている。

例 (2) においては、例 (1) と同様の理屈で *he_(j)* の参照先が *John_(i)* と同定されるが、これは誤りである:

(2) [*John_(i)*] slapped [*Bob_(j)*] because *he_(j)* hit his friends.

そこで本稿では、Plutchik らにより提案された感情の輪 (Wheel of Emotions) [11] をもとに、因果関係の一般化に有用と思われる 8 種類の感情状態を定義し、動詞・形容詞-感情状態間の対応辞書の作成、および感情状態の因果関係オントロジの記述により、上述の問題点の解消を試みる。より具体的には、次の二つの仮説のもとで因果関係認識の枠組みを設計し、その有用性を照応解析タスクにおいて検証する: (1) 因果関係を構成する事象は、感情状態をその項に投射 (projection) する機能をもつ、(2) 事象間の因果関係は、そこに登場するモノの感情状態の変化により表現できる。

本研究の狙いは、「傷づいた人は、誰かから癒やしを受けやすい」というように、感情状態という粗い粒度で事象を一般化することにより、(1) 文章には直接書かれないような因果関係を捉える、及び (2) 因果関係の知識を人手により精緻に設計することを可能にする、ということである。本稿の取り組みは実験段階であるものの、(1) 因果関係の表現に有効と思われる感情状態と因果関係オントロジを構築するための指針を示したこと、(2) fine-grained な感情状態に基づく因果関係認識の有効性を定量的に示したことが、本稿の貢献である。

本稿では、まず因果関係推論と評価極性の関連研究を概観し、問題点を述べる (2 節)。次に、上述の問題点を解消するために、Plutchik らの感情状態に基づく因果関係認識の枠組みを提案する (3 節)。最後に、代名詞照応解析のタスクにおいて、提案手法により精度よく因果関係を捉えられたことを報告する (4 節)。

2 関連研究

感情状態に基づく因果関係表現のアイデアは、AI 黎明期である 80 年代に提案された Lehnert のプロット理論 [5] に遡る。プロット理論は、物語の概要を登場人物の感情状態の変化で表現するというものであり、感情状態としては M (意思表示), + (ポジティブ), - (ネガティブ) の 3 値が用いられる。例えば、“*John wanted a car. Paul bought a car for John.*” という物語を、*John* の感情状態が M (*car* を *want* している) から + (*car* を得た = 目的を達成した) に変化した、と表現する。当初、プロットを自動生成する方法論は確立されていなかった

が、Goyal ら [4] は、自動獲得した辞書²に基づいて登場人物の感情状態を自動推定し、推定した感情状態間のリンクを自動同定する手法を示した。Goyal らの研究は、あくまでプロットの自動生成の研究であるため、共参照解析結果が事前に得られることを前提としており、感情状態に基づく因果関係の一般化が代名詞照応解析にどれほど有効かは評価されていない。

共参照解析の文脈においては、1節で述べたように、Rahman ら [12] の研究が最も関連深い。Rahman らは、感情分析 (sentiment analysis) の分野で構築された Subjectivity Lexicon [15] を利用し、2値の感情状態 (ポジティブ・ネガティブの2値) に基づいて、照応詞が参加している事象と先行詞候補が参加している事象の因果関係を捉え、これを機械学習器の素性の一部として用いている³。我々の研究は、Plutchik [11] の研究に基づく8種の感情状態を用いることで、2値ベースの手法の問題点を解消しようとするものである。

事象間の因果関係を、単語と抽象的な述語のマッピングと、抽象的な述語間の関係記述の組み合わせで表現するアプローチは、FrameNet [1] や佐野ら [17] の Phased Predicate Template Taxonomy (PPTT) に近いアイデアである。FrameNet は、自然言語のあらゆる単語を意味フレームに抽象化し、意味フレーム間の意味的關係を記述しようとするものである。PPTT は、万物に共通する、生成から消滅までのイベントを抽象的に定義した述語 (「モノが存在しない状態から存在する状態へと変化するイベント」クラスなど) の分類法であり、これらの述語の間の因果関係を記述することにより、高精度な因果関係知識の獲得が可能になると報告されている。これらの研究に対して、本研究は、感情状態の変化に着目して述語を抽象化している点で設計思想が異なっており、これらの研究とは相補的な関係にあると考えている。

感情分析の分野では、Subjectivity Lexicon [15], Effect WordNet [3] 等の2値の感情状態に基づく評価極性辞書や、WordNet-Affect [14]⁴, NRC-Emotion Lexicon [9]⁵ 等、多値の感情状態が付与された評価極性辞書が構築されている。しかし、因果関係の表現の観点から整備された感情状態辞書及びオントロジは存在しない。

3 感情状態に基づく因果関係推論の一般化

まず、本研究の概要を図1を用いて説明する。本研究では、(1)感情状態想起述語レキシコン (Affect State-Evoking Predicate Lexicon; AEPL)、(2)感情状態に基づく因果関係オントロジという2つの資源を構築する。解析時には、まず AEPL を用いることにより、入力文に出現した動詞・形容詞を手がかりとして、登場人物の感情状態を推定する。さらに、因果関係オントロジによ

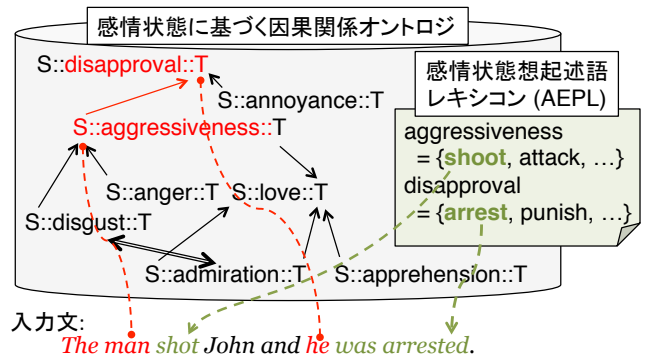


図1: 事象間関係オントロジに基づく因果関係認識の概要。S :: A :: T は、発信者 S、対象者 T であるような感情状態 A を表す。また、各感情状態の変項 S, T を結ぶ実線の矢印 \rightarrow は、因果関係を表す。感情状態間を結ぶ矢印 \leftrightarrow は、矛盾関係を表す。

り、感情状態レベルで因果関係の推論を行い、入力文内の因果関係の認識、および共参照関係の認定を行う。

例えば図1に示すように、AEPLを利用して、*the man* と *he* の感情状態について、それぞれの述語である *shoot*, *arrest* をもとに、それぞれ「aggressiveness (攻撃性)の発信者 (S::aggressiveness)」「disapproval (不承認性)の対象者 (disapproval::T)」という感情状態であることを推定する。さらに、因果関係オントロジにおいて、S::aggressiveness の状態の人は disapproval::T になりやすい (攻撃的な人は、不承認状態の対象になりやすい)、という因果関係が記されていることから、*the man* が disapproval::T 状態に遷移することを推定し、*the man* と *he* の共参照関係を認識する。

3.1 AEPL の構築

感情状態の分類法については、心理学・認知科学・言語学を中心としてさまざまな議論が行われてきたが [5, 7, 11, etc.], 研究者間でコンセンサスの取れた分類法は存在しない。自然言語処理では、評価極性 (sentiment polarity) と呼ばれる2値の感情状態 (ポジティブ・ネガティブ) が広く用いられてきたが、1節で述べたように、因果関係の表現には一般性が高すぎると考えられる。

そこで本研究では、さまざまな研究分野で幅広く用いられている Plutchik の Wheel of Emotions [11] をベースとした。Plutchik の研究では、感情状態を基本の8種類 (joy, trust, fear など) に分類し、それぞれを組み合わせることで、さらに8種類の感情状態 (love, submission, awe など) を定義している。本稿では、因果関係の記述に有用と思われる感情状態を、Winograd Schema Challenge (WSC) [6] と呼ばれる代名詞照応解析の問題集に基づいて選定した。WSC は、その問題の多くが、因果関係を正しく認識しないと解けない代名詞照応解析問題となっており、感情状態の選定に適した問題集であるといえる。この結果、選定された感情状態は、aggressiveness, disapproval, love, admiration, disgust, anger, submission, apprehension, annoyance の8種類

²動詞が目的語に与える影響を2値の評価極性レベルで記述したものである (e.g., punish: -, reward: +)。

³改良版として、Peng ら [10] の手法があり、同様に2値の感情状態に基づく手法を提案している。

⁴<http://wdomains.fbk.edu/wnaffect.html>

⁵<http://www.saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>

表 1: 感情状態想起述語レキシコン (AEPL) の例

投射する基本感情状態	動詞・形容詞の例
aggressiveness	shoot, throw at, laugh at
disapproval	arrest, scold, penalize
admiration	like, imitate, learn from
disgust	hate, dislike

となり、以降これらを基本感情状態と呼ぶ⁶。

また、WSC の訓練事例をもとに、動詞・形容詞がその項に投射する基本感情状態をデータベース化した（これが感情状態想起述語レキシコンである）。辞書の構築にあたっては、シソーラスを用いた同義語展開を行い、辞書のカバレッジの向上に務めた。作業の結果、496 エントリからなる辞書が得られた。辞書の一部を表 1 に示す。例えば、aggressiveness に関連付けられる動詞には、shoot などの物理的な攻撃を示す動詞のほか、laugh at などの精神的な攻撃を示す動詞が含まれる。また、disapproval については、arrest, penalize などの「懲罰」を表す動詞が含まれる。将来的には、クラウドソーシングなどを利用して辞書を大規模化するとともに、感情状態への関連付けタスクの妥当性を検証していく予定である。

3.2 AEPL を用いた感情状態推定

まず、本稿における感情状態を形式的に定義する。本稿では、モノの感情状態を、(1) モノが持つ基本感情状態 a と、(2) モノと a の関係（役割と呼ぶ） $r \in \{src, target\}$ (src : a を発信している、 $target$: 他者から a を受けている)、の組み合わせ $\langle a, r \rangle$ で表現する。例えば、図 1 では、*the man* は aggressiveness 状態を他者に発信しているため、その感情状態は $\langle aggressiveness, src \rangle$ である。

AEPL を用いて登場人物の感情状態を推定する際は、まず、構文解析結果より、感情状態の推定対象 x を govern する述語 p を同定する。辞書より、述語 p が投射する基本感情状態 a を同定し、 x が p の動作主である場合には、 $\langle a, src \rangle$ 、それ以外の場合には $\langle a, target \rangle$ を感情状態とする⁸。例えば、図 1 における *the man* に対しては、*the man* を govern する述語 *shoot* をクエリとして辞書を検索する。ここで *the man* は動作主であるため、推定結果は $\langle aggressiveness, src \rangle$ となる。

3.3 感情状態間の因果関係の記述

次に、3.2 節で定義した 8 種類の基本感情状態に基づく因果関係オントロジを構築する。因果関係オントロジには、感情状態間の因果関係のほか、事象間の反期待関係（ある状態または事象の発生が、期待に反する状態または事象の発生を引き起こしたという関係）の認識を行うために、基本感情状態間の矛盾関係を付与する。例えば、図 1 では、 $\langle anger, src \rangle$ 状態と $\langle aggressiveness, src \rangle$

状態が因果関係であり（ \rightarrow で示される）、admiration と disgust が矛盾関係である（ \Leftrightarrow で示される）。

構築したオントロジを用いて 2 つの事象間の因果関係認識を行う際は、まず事象間の接続関係（順接・逆接）を認識する⁹。順接関係である場合には、3.2 節の手法により推定された 2 つの感情状態が、因果関係で結ばれているかをチェックする。逆接関係である場合には、2 つの感情状態が矛盾関係で結ばれているか、かつ 2 つの感情状態の役割が同一かをチェックすることで、因果関係を認識する。例えば、admiration と disgust の矛盾関係を用いて、例文 “Although John hates Bob, John imitated Bob.” において、「John は Bob に対して admiration の感情を投射しており、通常は John は Bob に対して disgust の感情を投射しないが、例外的にそれが起こった」という反期待関係の認識を行う。

因果関係オントロジの作成にあたっては、(1) 3.2 節で言及した WSC の訓練事例を刺激とした人手による因果関係の同定、および (2) 感情状態間の関係の人手による精査、の二つを行った。例えば、WSC の訓練事例 “[John]_(i) kicked [Bill]_(j) because he_(i) was angry.” より、angry の状態にある人は kick しやすい、といった因果関係が考えられるため、これを AEPL により一般化し、 $\langle anger, src \rangle$ 状態にあるモノは、 $\langle aggressiveness, src \rangle$ 状態になりやすい、という因果関係を追加した。

結果として、32 個の関係が付与された。本稿では、人手により感情状態間の関係の有無を記述したが、今後の課題として、大規模 Web コーパス等から感情状態間の因果関係の強度を自動推定することなどが考えられる。

4 評価実験

3 節で提案した因果関係認識の枠組みを代名詞照応解析タスクの上で評価した結果を報告する。

4.1 実験設定

提案手法の評価タスクとして、Winograd Schema Challenge (WSC) [6] を用いた。WSC は、照応詞が参加している事象と先行詞候補が参加している事象の因果関係を正しく認識することが重要な代名詞照応解析の問題集であり、本研究の貢献を測るのに適したタスクだと考えられる。データセットについては、Rahman ら [12] が作成したデータ¹⁰のうち、テスト事例 565 問を評価に用いた。3 節で述べたように、AEPL 及び因果関係オントロジの構築の際は、テスト事例とは完全に独立の訓練事例のみを利用している。

共参照関係の認定にあたっては、代名詞の参加している事象と先行詞候補の参加している事象を入力として 3 節の因果関係認識手法を適用し、因果関係が認められた先行詞候補をシステムの出力とした。評価指標には、Precision (P), Recall (R), F 値を用いた。システムは、二つの先行詞候補について因果関係を認識した場合、出力を行わないものとした。

⁹本稿では、接続詞から判断した。

¹⁰<http://www.hlt.utdallas.edu/~vince/data/emnlp12/>

⁶データセットとしては、Rahman ら [12] が作成した WSC の問題集 7 の訓練事例 1,322 問を用いた。

⁸述語 p が catenative verb である場合 (to 不定詞を取っている try, want など) には、to 不定詞の後ろに来る述語をもとに基本感情状態を推定した。述語 p が否定されている場合には、3.3 節で説明する基本感情状態間の矛盾関係の情報を利用し、 a と矛盾関係にある基本感情状態を付与した。

表 2: WSC 解析性能の比較結果

モデル	訓練事例			テスト事例		
	P	R	F	P	R	F
RN12+W05	73.6	5.9	10.9	65.5	6.7	12.2
RN12+W05,G13	68.4	13.1	22.0	64.5	13.8	22.7
RN12+W05,C14	62.6	18.8	28.9	56.1	15.6	24.4
G13	63.2	4.5	8.5	64.6	5.5	10.1
G13+C14	50.3	6.3	11.2	54.8	7.1	12.5
Proposed+W05	73.2	8.2	14.8	70.0	8.7	15.4
Proposed+W05,G13	67.4	14.5	23.9	65.9	15.0	24.5
Proposed+W05,C14	62.5	20.2	30.5	58.4	17.2	26.5

本研究の比較対象のモデルとして、2 値の感情状態に基づく二つの因果関係認識モデルを用意した。一つ目は、Rahman ら [12]¹¹ の感情状態推定手法を用いたモデル (RN12) であり、二つ目は、Goyal ら [4] の推定手法を用いたモデル (G13) である。また、感情状態推定のための辞書として、Subjectivity Lexicon (W05) [15]¹², Patient Polarity Verbs Lexicon (G13) [4]¹³, +/- EffectWordNet (C14) [3]¹⁰ をそれぞれ使い、個別に評価した。提案手法については、提案モデルで先行詞が決定されない場合、Rahman らのベースラインモデルに back-off し、その出力を採用するようにした (Proposed)。

4.2 実験結果

実験結果を表 2 に示す。表 2 より、提案手法の解析性能が訓練事例とテスト事例の両方において上回っていることから、本研究で提案した感情状態に基づいて因果関係認識を行うことの有効性が確認された。一方で、適合率は比較的高いものの、再現率が低いことが確認された。

次に、提案モデルの振る舞いを明らかにするために、人手により解析結果を分析した。その結果、提案手法により正しく解析できた問題は 17 問であり、そのうち 10 問は Rahman らの手法 [12] では先行詞を捉えられない事例であることがわかった。また、この 17 問の中には、多値の感情状態を定義したことにより適切に先行詞を識別できたと考えられる事例が存在した。例えば、“[The man_(i)] killed [the boss_(j)] because he_(j) fired him.” と “[Someone_(i)] hit [Billy_(j)] in the face because he_(i) was angry.” というそれぞれの事例について、提案モデルは正しく先行詞を同定できていた。これらの事例は、1 節で述べたように、2 値の感情状態ではどちらも主語側が先行詞として認定されてしまう。

一方で、提案手法が解析を誤った問題は 3 問であった。例えば、“[The boy_(i)] escaped [the cop_(j)] but he_(i) continued to flee.” という事例では、まず the cop_(j) の感情状態が (submission, target) (屈服させた状態) と推定されたが、従属節が逆接関係にあるため反期待関係の推論が起り、従属節では the cop_(j) の感情状態が (submission, src) (屈服した状態) になると推定された (i.e., 屈服させた人が、予想外にも屈服した)。これに

¹¹Rahman ら [12] の改良版として Peng ら [10] の手法があるが、ほとんど解析性能が変わらなかったため、結果の表示を省いた。

¹²<http://mpqa.cs.pitt.edu/lexicons/>

¹³<http://www.umiacs.umd.edu/~amit/AESOP.html>

より、he_(i) の参照先が誤って the cop_(j) と同定されてしまった。しかし、ここでは、「行為 A を終えたが、予想外にもまた行為 A をした」というメタな逆接関係を認識する必要がある。現状、本事例における continued は cantenative verb であるために感情状態の推定対象となっていないため⁸、一段メタなレベルで逆接関係の認識を行う機構を入れる等の対策が考えられる。

5 おわりに

本稿では、感情状態に基づいて因果関係を一般化し、因果関係を認識する手法を提案した。先行研究では、2 値の感情状態に基づく一般化が行われていたのに対して、本研究では 8 値の感情状態を用いることにより、より精度よく因果関係を認識できることがわかった。今後の課題として、感情状態想起述語レキシコン (AEPL) の大規模化や、因果関係オントロジ内の因果関係の自動強度推定などが挙げられる。また、今回解けるようになった事例と、他の解析手がかり (述語の選択選好性、大規模に獲得した事象間関係知識 [18] など) により既に解けている事例との重なりを確認するために、本手法を既存の共参照解析器に統合し、評価を行っていく予定である。

謝辞 本研究は JSPS 科研費 15H01702, 15H05318 の助成を受けたものである。また、本研究は、JST, CREST の支援を受けたものである。

参考文献

- [1] C. F. Baker, C. J. Fillmore, and J. B. Lowe. The berkeley framenet project. In *Proc. of COLING*, pp. 86–90, 1998.
- [2] N. Chambers and D. Jurafsky. Unsupervised Learning of Narrative Event Chains. In *ACL*, pp. 789–797, 2008.
- [3] Y. Choi and J. Wiebe. +/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference. In *Proc. of EMNLP*, pp. 1181–1191, 2014.
- [4] A. Goyal, E. Riloff, et al. A computational model for plot units. *Computational Intelligence*, Vol. 29, No. 3, pp. 466–488, 2013.
- [5] W. G. Lehnert. Plot Units and Narrative Summarization. *Cognitive Science*, Vol. 5, No. 4, pp. 293–331, 1981.
- [6] H. J. Levesque. The Winograd Schema Challenge. In *AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning*, pp. 63–68, 2011.
- [7] J. R. Martin. Beyond exchange: Appraisal systems in english. *Evaluation in text: Authorial stance and the construction of discourse*, Vol. 175, , 2000.
- [8] A. Modi and I. Titov. Inducing neural models of script knowledge. In *CoNLL-2014*, p. 49, 2014.
- [9] S. M. Mohammad and P. D. Turney. Crowdsourcing a Word-Emotion Association Lexicon. Vol. 29, No. 3, pp. 436–465, 2013.
- [10] H. Peng, D. Khashabi, and D. Roth. Solving Hard Coreference Problems. In *Proc. of NAACL*, pp. 809–819, 2015.
- [11] R. Plutchik. The Nature of Emotions. *American Scientist*, Vol. 89, No. 4, pp. 344–350, 2001.
- [12] A. Rahman and V. Ng. Resolving Complex Cases of Definite Pronouns: The Winograd Schema Challenge. In *Proceedings of EMNLP-CoNLL*, pp. 777–789, 2012.
- [13] T. Shibata and S. Kurohashi. Acquiring strongly-related events using predicate-argument co-occurring statistics and case frames. In *IJCNLP*, pp. 1028–1036, 2011.
- [14] C. Strapparava and A. Valitutti. Wordnet affect: an affective extension of wordnet. In *Proc. of LREC*, Vol. 4, pp. 1083–1086, 2004.
- [15] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *Proc. of EMNLP*, pp. 347–354, 2005.
- [16] 井之上直也, 杉浦純, 乾健太郎. 共参照解析のための事象間関係知識の文脈化. 言語処理学会第 20 回年次大会論文集, pp. 717–720, 2014.
- [17] 佐野大樹, 鳥澤健太郎, Julien Kloetzer, Istvan Varga, 橋本力, 呉鍾勲. Phased Predicate Template Taxonomy (PPTT) を用いた意味的關係・スキーマの知識の大規模獲得. 言語処理学会第 20 回年次大会論文集, pp. 971–974, 2014.
- [18] 杉浦純, 井之上直也, 乾健太郎. 共参照解析における事象間関係知識の適用. 言語処理学会第 20 回年次大会論文集, pp. 713–716, 2014.
- [19] 大野雅之, 井之上直也, 松林優一郎, 岡崎直樹, 乾健太郎. 分散表現に基づく選択選好モデルの文脈化. 情報処理学会第 225 回自然言語処理研究会, p. To appear, 2016.