

参照表現ベイジアンネットワークによる参照解決の評価

船越 孝太郎^{†1} 中野 幹生^{†1}
徳永 健伸^{†2} 飯田 龍^{†2}

我々は、対話システムのための、ベイジアンネットワークを用いた参照表現のモデルを提案した。提案モデルは、記述、照応、直示を統一的に扱い、参照表現の理解（参照解決）と生成に用いることができる。本稿では、REX-J 日本語マルチモーダル参照表現コーパスの一部を用いて、提案モデルによる参照解決の予備的な定量評価を行う。

1. はじめに

参照表現とは話し手が関心を持つ特定の事物を聞き手に対して指し示す言語表現である。参照表現には記述表現（「机の上のコップ」）、照応表現（先行文脈を伴っての「それ」）、直示表現（外界への指差しを伴っての「それ」）が含まれるが、これらは従来別々に研究されることが多かった。特に言語生成の分野では記述表現の生成^{(1),(2),(6)}が、テキスト解析の分野では照応表現の解析^{(5),(8),(9)}が重点的に研究されてきた。

対話システムにおいて、ある程度複雑な課題領域での対話を扱おうとすると、参照表現の動的な理解と生成の両方が必要になり、また記述も照応も扱う必要がでてくる^{*1}。特にマルチモーダルな対話システムにおいては、照応だけでなく直示も重要になる。一方、実際の対話では、記述、照応、直示を明確に区別することは難しい。そのため、理解と生成の両方に用いることができ、記述、照応、直示を区別せずに統一的に扱える枠組みの実現が望まれる。このような枠組みとして、我々は参照表現ベイジアンネットワークを提案した⁴⁾。

参照表現ベイジアンネットワークの枠組みでは、ベイジアンネットワークの形で、個々の参照表現の確率モデルを動的に（つまり対話進行中に）構築する。確率モデルは、ある表現

が特定の対象を参照していると思われる確からしさを確率の形で与えるので、それを元に参照解決（参照表現の理解）を実現できる。また、その確からしさを指標として、複数の候補のなかから最良と予想される参照表現を選択することで、参照表現の生成も実現できる^{*2}。

先稿⁴⁾では、枠組みの提案と定性的な考察にとどまり、定量的な評価を示していない。そこで本稿では、REX-J コーパス¹¹⁾の一部を用いて、参照表現ベイジアンネットワークによる参照解決の予備的な評価を行う。

まず、2節で提案アプローチの重要な要素である参照ドメイン^{3),(10)}について説明する。次に、3節で参照表現ベイジアンネットワークの概要を説明し、4節で本稿での評価に用いるREX-J コーパスを説明する。5節でREX-J コーパス上で参照表現ベイジアンネットワークを用いるために必要な各種のデータやモデルの実装を説明する。6節で実験結果を示し、7節で結論と今後の課題を述べる。

2. 参照表現と参照ドメイン

参照表現は話し手が関心を持つ特定の事物（以後これを指示対象とよぶ）を聞き手に指し示す。参照ドメインとは指示対象を含む集合であり、知覚的あるいは談話的な作用によって対話者に認識される心的表象である^{3),(10)} ^{*3}。

参照表現は参照ドメインを前提として作られる。つまり、指示対象を x 、その時前提となっている参照ドメインを d ($x \in d$) としたとき、 x に対する参照表現 e は、 x を d 中のその他の要素 $d \setminus \{x\}$ から弁別する表現であり、 e がどのような表現になるかは x だけでなく $d \setminus \{x\}$ によって決まる。対話のある時点においてどのような参照ドメインが前提とされているのかは、それまでの対話履歴（以後、**談話**とよぶ。談話には発語行為だけでなく、物体の操作といった物理的な行為も含まれる）とその時の世界の様子（以後、**状況**とよぶ）に依存するが、一般に暗黙的である。人間の認知的な特性や、文化的・言語的慣習、および使用された参照表現が含む情報（例えば「別の箱」という参照表現は、参照ドメインが複数の箱を含んでいるという情報を持っている）によって特定されていると考えられる。

参照ドメインの例を図1を用いて説明する。図1は、2人の人間が対話をしながら、コン

^{†1} (株)ホンダ・リサーチ・インスティテュート・ジャパン

^{†2} 東京工業大学大学院情報理工学研究科

^{*1} ごく単純な課題領域の内では、往々にして参照表現と指示対象が一对一に決まっているので、表現と対象の対応ルールを事前に与えれば事足りる。また照応はないものとする 것도多い。現在研究・実用化されている対話システムでもまだこのような領域内に留まるものが多いが、本研究はその先を念頭においている。

^{*2} ここでの「生成の実現」は、指示対象を特定するという参照表現の最低限の機能を満たすという意味に限られる。言語生成においては機能的な側面だけでなく質的な側面も重要であるが、生成の質を高めるには、よい表現候補を生成するアルゴリズムや首尾一貫性を評価するモデルなどと組み合わせる必要がある。

^{*3} 参照ドメインは単なる集合ではなく、焦点などの情報も付随している。また対話システムにおける課題領域（タスクドメイン）や、対話の場面（フォーマル/インフォーマルの区別など）といった概念とは異なる。

コンピュータ画面上でタングラムという7つのピースからなるパズルを解いている場面である(詳細は4節を参照)。7つのピースに振られている(1)-(7)の番号は判り易さのために後から記したもので、どのピースが何番のピースであるかはパズルを解く当事者には判らない。指示対象となりうる対象全てを含む参照ドメイン、つまり7つのピース全てを含む参照ドメインを以後、**基底ドメイン**とよぶ。

前述のように、参照ドメインは知覚的・談話的に認識される。この図1の状況において、人間は知覚的群化¹²⁾の作用により、{1,2,6}や{5,7}といったピースの集合を認識する。これらの集合は参照ドメインとなり得る。図1の場面では1番のピースを動かして2番のピースに寄せており、{1,2,6}に対して焦点が当たっている状況なので、次の瞬間に{1,2,6}が参照ドメインとして前提される可能性は高いと予想される。一方、もしここで基底ドメインを前提として「正方形と中くらいの三角形をくっつけてみよう」といった発話がなされれば、聞き手には{3,6}も集合として認識され、次の瞬間に参照ドメインとして前提される可能性が高くなるだろう。

前述のように、参照表現は参照ドメインを前提として作られる。特定の参照ドメインが共有されていない段階では基底ドメインが前提とされ、その中で指示対象を弁別可能にするため複数の属性情報が同時に用いられることになる。例えば、図1の状況で、1番のピースを特定するためには「右の大きな三角形」のような参照表現が用いられることになる。しかし、例えば{1,2,6}を参照ドメインとして前提できるような文脈(談話と状況をまとめて**文脈**とよぶ)であれば、「右の三角形」とだけ言って済ませることもできる。

以後、単なる集合と参照ドメインを区別するために、参照ドメインを要素を列挙して記述する場合は[1,2,6]のように鍵括弧で括る。参照ドメイン内の焦点については、焦点の当たっている要素を*記号で示す(例えば、焦点が1番のピースであれば[1*,2,6])。特定の参照ドメインをラベルで参照する場合は、@_mのように@記号と添字を用いて示す。基底ドメインは常に@₀で表す。

3. 参照表現ベイジアンネットワーク

本節では、参照表現ベイジアンネットワーク(以後、Referring Expression Bayesian Networkを略してREBNとよぶ)について説明する。説明は本稿で必要な分に限るので、詳しくは先稿⁴⁾を参照されたい。

まず最初に、参照表現の分類と、参照表現の表層構造を表現するために用いるREX-graphの定義を行う。次に、REX-graphを元にREBNを構築する際の基本構造となるWCXD

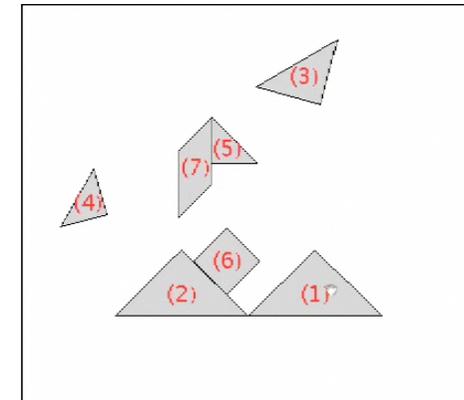


図1 タングラムパズル

構造について説明し、REBNを用いてどのように参照解決を行うのかを示す。

3.1 REX-graph

REX-graphは語で表現された概念間の依存構造のことである。ここでいう語とは、概念に対する単位的な言語表現を意味する。厳密な意味での語よりも広く、単語だけでなく複合語や句(特に固有表現などの場合)まで含めて考える。例えば、「もう一つの」という3単語からなる表現は、ANOTHERという概念を意味する1つの「語」と捉える。この語のことを概念表現ともよぶ。

個体、個体の部分、個体の集合(群)、ある特定の空間や時間、出来事など、人間が1つのまとまりとして認識するものへの参照表現を**単位参照表現**とよぶ。例えば、「赤いボール」、「あの夫婦」、「隣の部屋」、「昨日」といった参照表現は単位参照表現である。単位参照表現を2つ以上含む参照表現を**複合参照表現**とよぶ。例えば、「僕のボール」、「机の上」、「去年の12月」といった参照表現は複合参照表現である。REBN自体は複合参照表現も扱えるが、本稿では単位参照表現しか扱わないので、複合参照表現についてのこれ以上の説明は割愛する。

REX-graphは単位参照表現をノードとする有向グラフである。しかし、先ほど述べたように複合参照表現は扱わないので、本稿で取り扱うREX-graphは全てノード1つだけからなる一番単純な形のグラフとなる。ノードである単位参照表現は、空白で区切った語を丸括弧で括って表す。例えば「大きい 三角形」は

(大きい 三角形)

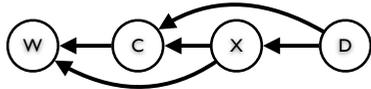


図2 WCXD 基本構造

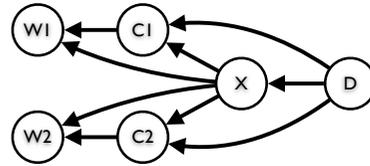


図3 2語からなる参照表現ベイジアンネットワーク

と表す。同様に、「もう1つの三角」は
(もう1つの 三角)
となる。

読みが曖昧な場合もあるので、言語表現から適切な REX-graph を抽出する過程は必ずしも単純ではない。例えば「その右の三角形」という表現には、「その右にある三角形」(代行指示, 複合参照表現)と「右側のその三角形」(指定指示, 単位参照表現)の2つの読み¹³⁾が可能である。しかし本稿では、適切な REX-graph が与えられると前提とし、REX-graph の抽出についてはこれ以上議論しない。

3.2 WCXD 基本構造

参照表現ベイジアンネットワークは、図2に示す WCXD 構造を基本とする。

WCXD 基本構造は参照表現の生成過程を表現するもので、 W, C, X, D を確率変数とするとき、 W が語の選択、 C が概念の選択、 X が指示対象の選択、 D が参照ドメインの選択を表現している。すなわち、参照表現の生成過程を以下のように考える。まず、参照ドメイン D (何について述べるか) が決定され、その中から指示対象 X (どれについて述べるか) が決定される。その指示対象 X を言語化するため、その対象に関する概念 C (どの特徴を提示するか) が決定され、最後に語 W (どう表現するか) が決定される。観測されるのは W だけである。

図2に示した WCXD 基本構造はそれ自体で、語1つだけからなる参照表現 (例えば「正方形」) に対する参照表現ベイジアンネットワークが持つネットワーク構造となっている。図2中の REBN を確率式で表現すると、ベイジアンネットワークが仮定するマルコフ性により以下ようになる。

$$P(W, C, X, D) = P(W|C, X)P(C|X, D)P(X|D)P(D)$$

この REBN を用いて、ある参照表現の指示対象が x である確率 (x の参照確率) を求めるには、 W が証拠として与えられた下で、既存の手法を用いて $P(W, C, X, D)$ を周辺化し $P(X = x|W)$ の値を調べればよい。

2語で1つの対象を参照する場合 (例えば「大きい 三角形」) は、図3に示す形の REBN を用いる。 W_1 が「大きい」、 W_2 が「三角形」に対応する。この場合 REBN の確率式は $P(W_1, W_2, C_1, C_2, X, D)$ となるが、図2の場合と同様、周辺化によって $P(X|W_1, W_2)$ を求めればよい。

REBN の構造は参照表現が与えられてから決まり、確率変数の定義域も動的に変わるので、条件付き確率分布を事前に収集したデータから一括で学習することはできない。そのかわりに、個々の確率分布 ($P(W|C, X), P(C|X, D), P(X|D), P(D)$) を関数の形で実装して与える (5節参照)。

REBN を用いた参照表現の理解、すなわち、参照表現によって指示されている対象 x の特定 (参照解決) は、以下の様に行う。

- (1) 言語表現からの REX-graph の抽出：
語で表現されている概念間の依存構造を抽出する。前述の通り、本稿では適切に抽出できるものと仮定する。
- (2) 参照表現ベイジアンネットワークへの変換：
REX-graph は意味的な構造とはいえ、まだ表層表現のレベルで概念間の関係性を表現したものに過ぎないので、これを REBN に変換する。
- (3) 参照表現ベイジアンネットワーク上での推論：
REBN が与えられれば、指示対象の特定は一般のベイジアンネットワークに関する推論手法を用いて行える。REBN を周辺化して $P(X|E)$ を得、最も確率が高い対象 x' を選ぶ。ここで E は REBN 中の語ノード (W_i) の集合である。すなわち $E = \{W_1, \dots, W_n\}$ 。数式で表現すれば、

$$x' = \operatorname{argmax}_{x \in \mathcal{D}(X)} P(x|e) \quad (1)$$
 となる。ここで、 e は E 中の変数への値割り当て (観測値)、 $\mathcal{D}(X)$ は確率変数 X の定義域である。

REBN では、参照ドメイン自体も参照可能な「群」と捉える。これによって「その2つの三角形」のような集合に対する参照表現を単体の物体に対する参照表現の場合と同様に扱うことを可能にする。従って、常に $\mathcal{D}(D) \subset \mathcal{D}(X)$ が成り立つ。

表 1 REX-J コーパス中の (a) 対話と (b) 参照表現に対するアノテーションの例

(a)				(b)							
始端時刻	終端時刻	話者	発話 (括弧内がピースへの参照表現)	対話 ID	話者 ID	役割	始端時刻	終端時刻	参照表現	指示対象	REX-graph
13.222	19.052	SV	あつえつとじゃあ、うーん (おっきい三角) あるじゃない?	D08_08_00	S08_08_00	SV	17.345	18.390	おっきい三角	012	(おっきい 三角)
15.472	15.823	OP	はい。	D08_08_00	S08_08_00	SV	20.758	21.368	それ	1	(それ)
19.001	19.337	OP	はい。	D08_08_00	S08_08_00	SV	23.394	24.720	右側のおっきい三角	1	(右側のおっきい 三角)
19.908	31.275	SV	えつとー (それ) を、は、(右側のおっきい三角)、あつ(これ) ほつといて、で (その) 隣に同じ感じで、おの、(このおっきい三角) をもつこ並べてみよ。	D08_08_00	S08_08_00	SV	25.084	25.277	これ	1	(これ)
				D08_08_00	S08_08_00	SV	26.512	26.671	その	1	(その)
				D08_08_00	S08_08_00	SV	28.871	29.747	このおっきい三角	2	(この おっきい 三角)
31.715	32.223	OP	はい。	D08_08_00	S08_08_01	OP	46.497	48.204	おっきな三角形	12	(おっきな 三角形)
39.380	41.701	OP	え、これつてこつちから質問をしていいんですか?	D08_08_00	S08_08_01	OP	51.958	52.228	両方	12	(両方)
46.271	53.638	OP	と、(おっきな三角形) を、何かおむすびが2つ並ぶ感じで、(両方) 隣に並べるのはいい。								

4. REX-J コーパス

本稿では、REBN を用いた参照解決の評価に REX-J コーパス¹¹⁾ を用いる。REX-J コーパスは、ピースを組み合わせて目標の形を作るパズルタスク (タングラムパズル) を設定し、指示者と操作者の間の対話を記録したコーパスで、言語情報だけでなく、パズルの操作やマウスカーソルの位置を記録したマルチモーダルコーパスである。パズルの操作は操作者が PC 上で行う (図 1 参照)。

タングラムパズルではピースを組み合わせて目標の形を作ることを目指すが、組み合わせ方は伏せられ、目標の形の外形 (輪郭) のみが示される。目標の形は指示者だけに与えられているので、タスクを達成するために指示者 (SV) と操作者 (OP) の協調的な対話が生まれる。一部の対話データでは指示者と操作者の視線情報も利用可能であるが、本稿では用いない。

REX-J コーパスでは、ピースを指す参照表現に対して、対話 ID、話者 ID、話者の役割、始末端の時刻 (対話開始時からの経過時間)、及び指示対象のピースの情報がアノテーションされている。今回の実験では、これらのアノテーションに加えて、各参照表現に対し REX-graph をアノテーションした。表 1 に、ある時間内の対話と、それに対するアノテーションの例を示す。

表 1(a) の最初の参照表現の指示対象が 012 となっているのは、図 1 中の 1 番と 2 番のピースのどちらを指しているか解釈しても間違いではないことを意味している (英語の不定名詞句に相当)。一方表 1(a) の最後の発話に現れる 2 つの参照表現がそれぞれ 12 となっているのは、図 1 中の 1 番と 2 番のピースからなる集合を指していることを意味している

表 2 概念辞書 (一部)

概念	概念表現
OBJ	もの やつ
FIG	図形 図形
TRI	三角形 三角 直角三角形
SQR	正方形 正方形

表 3 静的適合度表 (一部)

概念	(1)	(2)	(3)	(4)	(5)	(6)	(7)
OBJ	1	1	1	1	1	1	1
FIG	0.3	0.3	0.3	0.3	0.3	0.3	0.3
TRI	1	1	1	1	1	0	0
SQR	0	0	0	0	0	1	0.8

※ (1)-(7) はピースの番号

(群への参照)。以降では、前者を OR 表現、後者を AND 表現とよぶ。

5. 実 装

本節では、REX-J コーパスの上で REBN を用いるために必要となる、概念辞書、静的適合度表を説明する。そして REBN を構築する際に条件付確率分布表を生成する手順について説明する。単位参照表現に限れば、条件付確率分布表の作成方法さえ判れば、ネットワーク構造の構築手順に特別な注意が必要な部分はない。3.2 節に述べたようにノードを作って接続するだけである。従って、ネットワーク構造を構築する手順については特に説明しない。

5.1 概念辞書

表 1(b) のようにアノテーションした REX-graph 中に現れる概念表現を抜き出し、それらに対応する概念を意味するシンボル (OBJ など) と組みにして、表 2 のように定義する。

5.2 静的適合度表

概念辞書中に定義した概念のうち、静的なもの、つまり対話中に適合度が変化しないものについて、静的適合度表を表 3 のように与える。適合度は [0, 1] の実数値で表現する。静

的適合度の意味を確率的に記述すると、 $Pr(C_c = \text{true}|X = x) = r$ となる。これは対象 x が概念 c に適合する程度が r ということの意味する。つまり、 $r = 0$ は「まったく適合しない」、 $0 < r < 0.5$ は「あまり適合しない」、 $r = 0.5$ は「どちらともいえない」、 $0.5 < r < 1$ は「まあまあ適合する」、 $r = 1$ は「強く適合する」ことを意味する。静的適合度表の中の数値はこの意味に沿う形で先験的に決めた。

例えば、表 3 によれば、6 番のピースは三角形ではないが四角形であるという意味になる。また、全てのピースは FIG (図形) という概念にはあまり適合しないということになるが、これは REX-J コーパスにおいて「図形」という表現が参照するものは多くの場合複数のピースが集まった群であり、単独のピースを参照するケースは稀であることを表している。

5.3 $P(W|C, X)$

REX-J コーパスでは、 W の選定に C とは別に X の情報が必要になることはないので、 $P(W|C, X) = P(W|C)$ と見なせる。 W の値は観測値なので、既に判っている (w とする)。 $\mathcal{D}(C)$ は、この観測値から前述の概念辞書を引いて、可能性のあるものを列挙して定める。これに加え、次に説明する $P(C|X, D)$ のために $\mathcal{D}(C)$ には ELSE という特別な値を含める必要がある。従って、 w が「図形」の場合であれば、 $\mathcal{D}(C) = \{\text{FIG}, \text{ELSE}\}$ となる。

この ELSE に対応させて、 $\mathcal{D}(W)$ にも観測された語 w に加え、else という特別な値を含める。すなわち $\mathcal{D}(W) = \{w, \text{else}\}$ とする。

本来であれば $P(W = w|C = c)$ はコーパス内の分布を反映させるのが妥当だが、本稿では簡単のため、概念 c に対して可能な概念表現が N 個であった場合、一律に $P(W = w|C = c) = 1/N$ とし、 $P(W = \text{else}|C = c) = (N - 1)/N$ とする。一方、 $P(W = w|C = \text{ELSE})$ は事前に与える極小さな値 ϵ とする。従って、 $P(W = \text{else}|C = \text{ELSE}) = 1 - \epsilon$ となる。

5.4 $P(C|X, D)$

$P(C|X, D)$ は $\mathcal{D}(C)$ が決まらなければ与えることができない。そこで、 $\mathcal{D}(C)$ によらない $P(C_c = \text{true}|X = x, D = d)$ を元にして動的に構築する。 $P(C_c = \text{true}|X = x, D = d)$ は、 c が静的な場合は $P(C_c = \text{true}|X = x)$ とみなし、静的適合度表から得た値をそのまま用いる。 c が動的 (状況依存) な場合は、 x と d に応じて手続き的な計算によって推定する。

例えば「右の」(RIGHT) という概念を考える。図 1 の状況において、 d が基底ドメインである場合には、例えば 5 番のピースの RIGHT への適合度は高くない (0.5 程度あるいはそれ以下と考えるのが自然)。一方 d が [5, 7] ならば、5 番のピースの RIGHT への適合度は 1 になる。これは d の含むピースの水平方向の座標の中で一番左側のものを u_l 、一番右側のものを u_r とし、 x の水平方向座標を u_x とすれば、 $P(C_{\text{RIGHT}} = \text{true}|X = x, D = d) = (u_x - u_l)/(u_r - u_l)$

表 4 $P(C_c = \text{true}|X = x, D = d)$ から $P(C = c|X = x, D = d)$ を得るためのアルゴリズム

```

Input:  $\mathcal{D}(C)$  及び全ての  $c \in \mathcal{D}(C) \setminus \{\text{ELSE}\}$  についての  $P(C_c = \text{true}|X = x, D = d)$ 
Output:  $P(C|X = x, D = d)$ 
 $n \leftarrow 0, s \leftarrow 0, S = \mathcal{D}(C) \setminus \{\text{ELSE}\}$ 
for all  $c$  in  $S$  do
     $r[c] \leftarrow P(C_c = \text{true}|X = x, D = d)$  { 概念  $c$  の適合度 }
     $s \leftarrow s + r[c]$ 
     $n \leftarrow n + (1 - r[c])$  { 残差  $(1 - r[c])$  の和 }
end for
 $r[\text{ELSE}] \leftarrow n/|S|$  { 残差の平均を ELSE の適合度にする }
 $s \leftarrow s + r[\text{ELSE}]$ 
for all  $c$  in  $\mathcal{D}(C)$  do
     $P(C = c|X = x, D = d) \leftarrow r[c]/s$  { 適合度を正規化して確率値にする }
end for
    
```

と計算できる。

$P(C|X = x, D = d)$ は、表 4 に示すアルゴリズムによって、概念 c 毎に推定した $P(C_c = \text{true}|X = x, D = d)$ から得られる。ELSE は 5.3 節で説明した特殊な概念である。これによって、 x が $\mathcal{D}(C) \setminus \{\text{ELSE}\}$ 中のいずれの概念にも適合しない状態を表現できる。ELSE を含まないと、 $\sum_c P(C = c|X = x, D = d) = 1$ という制約によりほとんど適合しない概念に対しても $P(C = c|X = x, D = d)$ の値が大きくなってしまう。

5.5 $P(X|D)$

$P(X|D)$ の推定には、SVM^{rank*1}を用いた予測器 $\text{rank}(x, \mathcal{D}(X), \theta)$ を用いる⁷⁾。 $\text{rank}(x, \mathcal{D}(X), \theta)$ は、文脈情報 θ を元に、あるピース x が現時点で参照される度合いを $\mathcal{D}(X)$ に対するランキングの形で予測する。出力される値は実数値で、値が大きいくほど上位にランクされることを意味する。 θ には、現在の時刻と、過去の参照、操作、オンマウス (カーソルがピースの上に乗る状態) の時刻付きの履歴が含まれている。

この $\text{rank}(x, \theta)$ を用いて、 $P(X = x|D = d; \theta)$ を以下の式で求める。 \mathcal{N} は、 $\sum_{x \in \mathcal{D}} P(X = x|D = d; \theta) = 1$ となるための正規化係数である。

$$P(X = x|D = d; \theta) = \begin{cases} \mathcal{N}/(1 + \exp(-\text{rank}(x, \theta))) & : x \in \mathcal{D} \\ 0 & : x \notin \mathcal{D} \end{cases}$$

前述のように $\mathcal{D}(X)$ は全ての参照ドメインを含んでおり、 $\text{rank}(x, \mathcal{D}(X), \theta)$ は個々のピー

*1 http://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html

スも参照ドメインも同列にランキングする。しかし、REX-J コーパスが持っている操作、オンマウスの情報はすべて個々のピースに関する情報である。そこで、これらに関する各時点での参照ドメインの素性値は、それが含むピースの情報を元に決定する。例えば、ある参照ドメインが「過去に操作されたことがあるかどうか」という 1/0 の素性は、それが含むピースのどれか 1 つでも操作されたことがあれば素性値を 1 とする。

5.6 $P(D)$

$P(D = d)$ は、参照ドメイン d の顕現性を元に推定する。顕現性についての基本的な考え方は単純で、直前に参照されたり知覚的群化によって認識された参照ドメインほど、顕現性が高いと考える。参照ドメインにはこのような考えに基づいて数値（非負の実数）を付与し、その数値に従って降順にソートしたリストで管理する。以下、リストの更新方法、知覚的群化の近似的な実装、その結果に基づく $P(D)$ の推定方法について説明する。

5.6.1 参照ドメインのリストの更新

参照ドメインの新規導入と顕現性の更新は、参照解決の直前と直後の 2 つのタイミングで行う。参照ドメインのリストへの追加はいずれの場合も同じ操作で行う。

5.6.1.1 参照解決の直前：知覚的群化による参照ドメインの導入・更新

ある時刻 t における参照表現を受け取ると、まずその時刻の状況に対して、後述する知覚的群化手法を適用し、参照ドメインの候補を取得する。ここで得られた候補のうち、時刻 t の直前に操作されていたピースを含む候補 $@_m$ だけをリストに追加する。追加方法は後述する。また、各参照ドメインにはその要素の中のどれに焦点が当たっているのかが記憶されている（どれにも焦点が当たっていない場合もある）。 $@_m$ の焦点には、直前に操作されていたピースを設定する。

5.6.1.2 参照解決の直後：参照解決結果による導入・更新

本稿の実験では参照解決を対話の先頭から順に行うが、このときそれまでの正解は全て判っているものと仮定する。そこで、参照解決を行った参照表現の指示対象（正解）が参照ドメイン（群）であり、この時点のリストにそれが含まれていなかった場合は、正解の参照ドメインをリストに追加する。また、正解が単体のピースであった場合は、そのピースを含む最も顕現性の高い参照ドメイン $@_m$ がリストの先頭に来るようにそれまでの最大の顕現性より 1 だけ大きい顕現性を与える。 $@_m$ の焦点には、その正解のピースを設定する。

5.6.1.3 参照ドメイン $@_m$ のリストへの追加方法

もし既に $@_m$ と同じ要素からなる参照ドメイン $@_k$ がある場合は、 $@_m$ を新しく追加するのではなく、その既存の参照ドメイン $@_k$ がリストの先頭に来るようにそれまでの最大の

顕現性より 1 だけ大きい顕現性を与える。但し、参照ドメイン $@_k$ が既に先頭にある場合はその顕現性を変更しない。もしリストに $@_m$ と同一視できるものがない場合は、その参照ドメインの補集合を取り、これを $@_n$ とする。更に $@_m$ と $@_n$ の 2 つを要素に持つ参照ドメイン $@_o = [@_m*, @_n]$ を作る。 $@_o$ の焦点には $@_m$ を設定している。この 3 つの参照ドメイン $@_m, @_n, @_o$ をリストに加えるが、それぞれの顕現性はこれまでの最大の顕現性に +1, +0.5, +0.5 した値にする。 $@_n$ および $@_o$ は、「残りのピース」のような表現に対して参照解決を行うために必要となる。

5.6.2 知覚的群化

先行研究¹²⁾ で示されている知覚的群化のシミュレーション手法を REX-J の状況に適用すると、比較的ピースがばらけている対話の中盤までの状況では妥当な結果を返すが、ピースが密集してくる終盤では直感とあまり一致しない結果を出す傾向が見られた。

そこで、今回はより簡便な近似的方法で知覚的群化のシミュレーションを行う。具体的には、ある 2 つのピース間の最短距離が極短い l ピクセル以下である時にその 2 つのピースは接触していると見なし、接触してるピースの集合だけを群として認識する。この方法を図 1 の状況に適用すると、[1, 2, 6] と [5, 7] が認識される。タングラムパズルでは必ずピースを接触させて目標の形を作るので、このような方法でも有意な群だけを抽出できる。

5.6.3 $P(D)$ の推定モデル

$P(D = d)$ を推定するためのモデルとして以下の 3 種類を用意し、6 節の実験にて比較する。ここで、参照ドメイン d の顕現性を $\text{sal}(d)$ で表す。

5.6.3.1 一様モデル

ドメインの顕現性を考慮しない。このモデルは $P(D)$ の推定の影響力を見るために 1 種のベースラインとして用意する。

$$P(D = d) = 1/|\mathcal{D}(D)|$$

5.6.3.2 線形モデル

顕現性に比例して確率値を線形配分する。すなわち、

$$P(D = d) = \frac{\text{sal}(d)}{\sum_{d' \in \mathcal{D}(D)} \text{sal}(d')}$$

5.6.3.3 指数モデル

顕現性の値に指数関数を適用してから線形配分することで、顕現性の高い参照ドメインをより重視する。

$$P(D = d) = \frac{\exp(\text{sal}(d))}{\sum_{d' \in \mathcal{D}(D)} \exp(\text{sal}(d'))}$$

5.7 各種概念の適合度の実装

参照表現に現れた概念 c を適切に解釈するためには、適合度 $P(C_c = \text{true} | X = x, D = d)$ を求める関数の実装が要となる。既に「右の」(RIGHT)の実装方法については説明した。ここでは、REX-J コーパスに現れるいくつかの概念に対する適合度の実装を説明する。

5.7.1 「もう一つの」(ANOTHER)

参照ドメイン d の要素の中で焦点となっているものを $\text{focus}(d)$ で表す。このとき、

$$P(C_{\text{ANOTEHR}} = \text{true} | X = x, D = d) = \begin{cases} 1 & : \text{ if } \text{focus}(d) \neq x \text{ and } x \in d \\ 0 & : \text{ otherwise} \end{cases}$$

5.7.2 「残りの」(REST)

5.6.1.3 節で説明した $@_o = [@_{m*}, @_n]$ のように 2 つの群を要素に持つ参照ドメインの場合に限って、焦点の当たっていない方の群の適合度を 1 とする。それ以外の場合は 0 とする。すなわち、

$$P(C_{\text{REST}} = \text{true} | X = @_n, D = @_o) = \begin{cases} 1 & : \text{ if } @_o = [@_{m*}, @_n] \\ 0 & : \text{ otherwise} \end{cases}$$

5.7.3 「両方」(BOTH)

x が群でその要素が 2 つであるとき 1 とする。すなわち、

$$P(C_{\text{BOTH}} = \text{true} | X = x, D = d) = \begin{cases} 1 & : \text{ if } |x| = 2 \\ 0 & : \text{ otherwise} \end{cases}$$

5.7.4 「図形」(FIG)

先に述べたように、この表現は組み上がっているピースの群を参照する。そこで、 x が単一ピースの場合 ($\text{single}(x) = \text{true}$) は静的適合表から得た数値 r を、 x が群であり、現時点で互いに接続して形をなしている場合 ($\text{shape}(x) = \text{true}$) 1 とする、

$$P(C_{\text{FIG}} = \text{true} | X = x, D = d) = \begin{cases} r & : \text{ if } \text{single}(x) \\ 1 & : \text{ if not } \text{single}(x) \text{ and } \text{shape}(x) \\ 0 & : \text{ otherwise} \end{cases}$$

5.7.5 「全部」(ALL)

REX-J に出現する「全部」という表現は必ずしも基底ドメインを参照していないので、任意の群を参照しうる概念として捉える必要がある。REBN の枠組みの中で ALL を扱うためには、少し特殊な仕組みを用意する必要がある。ここでは全ての参照ドメインは自分自身への参照 (自己参照) をその要素に含むと考える。つまり、 $@_m = [e_1, e_2, \dots, e_n, @_m]$ 。その上で、

$$P(C_{\text{ALL}} = \text{true} | X = x, D = d) = \begin{cases} 1 & : \text{ if } x = d \\ 0 & : \text{ otherwise} \end{cases}$$

とする。ただし、上記の「両方」などの概念に関して適合度を計算する場合には、自己参照子は無視する。もし自己参照を持たせないとすれば $@_{m'} = \{@_m\}$ のような参照ドメインを考える必要があるが、そうすると参照ドメインの数が常に 2 倍になり、計算量の面で負荷が大きくなってしまう。

6. 評価実験

REX-J コーパスを用いた参照解決実験について説明する。ベイジアンネットワークを用いた推論には、BNJ^{*1}を使用した。

6.1 実験条件

実験は以下の条件設定で行う。

- 単位参照表現だけを扱う。
- SV/OP の発話を区別しない。
- 未来の情報は使用しない。
- 人間は、場合によっては参照表現が発話された後の話し手の視線の情報なども利用して参照解決を行うが、実験では参照表現が発話され始めた時刻までの情報しか利用しない。
- 構文解析等の前処理は完全である。
- コーパスにアノテーションした REX-graph を入力として利用する。
- 直前までの参照表現の正解は全て判っている。
- コーパスにアノテーションされている正解データを用いて参照解決を行う。
- OR 表現の場合は正解候補のいずれかを指示対象と判定すれば正解とする。

*1 <http://bnj.sourceforge.net/>

表 5 実験に用いた 6 対話

対話 ID	D00	D04	D08	D12	D16	D20	合計
単位参照表現の数 (S)	57	100	123	64	57	82	483
S 中の OR 表現の数	3	2	4	4	2	2	25
S 中の AND 表現の数	13	7	17	9	11	12	69
S 中の代名詞表現の数	20	46	59	25	18	28	196
複合参照表現の数 (参考)	1	4	0	0	0	0	5

表 6 参照解決結果 (正解率 (%))

$P(D)$ 推定方法 種別	単一ドメイン (@ ₀)			一様モデル			線形モデル			指数モデル		
	単体	集合	全体	単体	集合	全体	単体	集合	全体	単体	集合	全体
単複情報なし	76.1	4.4	65.8	66.7	21.7	60.3	68.8	15.9	61.3	69.3	26.1	63.2
単複情報あり	77.8	4.4	67.3	68.8	46.4	65.6	68.8	46.4	65.6	70.8	46.4	67.3

表 7 改訂手法による参照解決結果 (正解率 (%))

$P(D)$ 推定方法 種別	単一ドメイン (@ _{0'})			一様モデル			線形モデル			指数モデル		
	単体	集合	全体	単体	集合	全体	単体	集合	全体	単体	集合	全体
単複情報なし	73.4	27.5	66.9	76.3	27.5	69.4	77.3	26.1	70.0	71.3	30.4	65.4
単複情報あり	78.5	39.1	72.9	80.2	46.4	75.4	80.9	44.9	75.8	75.1	43.5	70.6

正解データが 012 なら、参照解決の結果が 1 か 2 なら正解となる。

前述のパラメータは以下の値に設定した: $\epsilon = 0.0001$ (5.3 節), $l = 2$ (5.6.2 節). $P(X|D)$ の推定に用いる SVM^{rank} のモデルには REX-J コーパスの 08 年度収集のデータ全てで訓練したものを使用する。

6.2 データ

実験には REX-graph のアノテーションが完了している 08 年度収集分の中の 6 対話のみを用いる。この 6 対話はすべて異なる話者による対話だが、同一の課題に取り組んでいる。表 5 に実験に用いた 6 対話に現れた参照表現の内訳を示す (対話 ID は D08_08_nm を Dnm と略記している)。群への参照 (AND 表現) は 69 個で、全体の 14.3% である。また、代名詞表現 (「これ」「それ」「あれ」) の数は 40.6% を占めている。本稿では複合参照表現は扱わないと述べたが、そもそも今回使用したデータには複合参照表現はほとんど現れない。

6.3 実験結果

表 6 に参照解決の結果 (正解率) を示す。「 $P(D)$ 推定方法」という見出しの行は、 $P(D)$ の推定方法を表している。このうち「一様」、「線形」、「指数」は 5.6.3 節で示した 3 つの推定モデルを使用した場合に対応する。「単一ドメイン」は、参照ドメインを考慮しない従来

手法的な条件であり、この場合、 D の値となり得るのは基底ドメイン @₀ だけとなる。つまり、常に $D(D) = \{@_0\}$ の状態で参照解決を行う。この条件をベースラインとして考える*1。5.6.1 節で説明した参照ドメインリストの更新も行われず。この場合、群への参照は「全部」という表現で基底ドメインを参照する場合しか扱うことができなくなる。

「種別」は正解データを元にした参照表現の種別を表す。「単体」は単体ピース (1~7 のどれか) を参照する参照表現 (OR 表現も含む。総数 414 個)、「集合」は群 (1~7 の組み合わせ) を参照する参照表現 (すなわち AND 表現、総数 69 個)、「全体」は前記 2 つを併せた全ての参照表現を指す (総数 483 個)。

「単複情報なし」と「単複情報あり」は、指示対象が単体か集合かを区別する情報を REBN に与えるかどうかを意味している。英語のような言語では、名詞や代名詞の単数形/複数形を使い分けるので、指示対象が単体か集合かはかなりの場合で事前に知ることができる。日本語の場合では単数か複数かを表層から判断することは英語の場合より難しいが、機械学習などを用いればある程度の精度で判定することは可能であろう。従って、「単複情報あり」の場合の結果が、提案手法の現状の上限性能と見なせる。

表 6 を見ると、推定モデルに関しては、一様/線形/指数の中で、指数モデルが最も良い結果となっている。しかしながら差はそれほど大きくない。

「単体」の正解率を見ると、従来法に相当する単一ドメインの場合が最も正解率がよいという結果になっている。しかしながら、例えば D00 の対話において、LEFT (「左の」という概念を用いた単体ピースへの参照表現は 3 回出現し、「単一ドメイン」の場合いずれも不正解である一方、「線形モデル」の場合は全て正解できていた。このことから、参照ドメインの導入は、期待した効能を発揮している一方で、他の部分に悪影響を及ぼしてしまっていると予想される。データを概観したところ、特に不正解が目立つのは代名詞の場合であった。

そこで、参照表現が代名詞の場合には、各時点の $D(X)$ の要素を全て含むただ 1 つの参照ドメイン @_{0'} を $D(D)$ の要素とし、それ以外の場合は従前の通りに参照解決を行うようにして、再度実験を行った (これを改訂手法とよぶ)。また、単一ドメインの場合においても、談話・状況から参照ドメインを導入し*2、各時点での参照ドメインとして @_{0'} を対象とするようにして実験を行った。この結果を表 7 に示す。

*1 Iida ら⁷⁾ は同じ REX-J コーパスを用いて単一ドメイン条件に相当する設定で参照解決実験をしている。ただし彼らの実験では概念への適合度は考慮されていない (形などいくつかの属性についてピースと参照表現が矛盾していないかどうかだけ見ている) ので、本稿の実験結果と同列には比較できない。

*2 但し、「残りの」を理解するための補集合 @_m とそれを包含する @₀ は作っても役に立たないので作らない。

改訂手法では、全体に性能が向上する傾向が見られた。しかしながら、一様／線形／指数モデル間での性能差ははっきりしなかった。特に単複情報ありの場合には、一様モデルと線形モデルの間にほとんど違いはなかった（これらの間のパーセンテージの違いは、個数の上では1つか2つの違いしかない）。また、単複情報ありの場合、単体、集合の両方で指数モデルの結果が一様・線形モデルより悪くなっており、代名詞を区別して処理をしなかったときの傾向と変わってしまっている。今後データ数を増やして検証したい。

対話毎の結果を眺めると、指数モデルの性能がよい場合、線形モデルの性能がよい場合とがあり、話者毎に異なる適切なモデルが異なる可能性も伺えた。対話毎に分けると母数が小さいので本稿では詳細は述べないが、これについても今後データ数を増やして分析したい。

集合への参照は、最大の場合でも50%に届かなかった。データを眺めると、「それ」「これ」の代名詞で直示的にピース群を指している場合を拾えていないことが多いようであった^{*1}。このような時、人間はマウスカーソルでピース群上に円を描いて明示することが散見されるが、これは今回の実験では取り込めていない。今後、このようなジェスチャ的な情報も活用できるよう検討したい。また、操作に関する概念などで、まだ適合度の推定関数を実装できていないものが複数あり、これらについても実装を進める必要がある。

7. おわりに

本稿では、参照表現ベイジアンネットワーク⁴⁾の予備的な定量評価を行った。評価には、REX-J参照表現コーパスの中の6対話を用いた。参照表現ベイジアンネットワークは、記述、照応、直示を区別せずに統一的に扱うことができることが1つの特徴であり、また参照ドメイン（その時々において前提とされる指示対象の集合）を考慮することができる。

実験により、常に全てのピースを参照ドメインとして前提とする従来の方法では正しく参照解決できなかった表現も、参照表現ベイジアンネットワークによってその時々において適切な参照ドメインを考慮して正しく参照解決できることが確認された。

同時に、単純に全ての表現を同様に扱うだけでは、従来手法よりも正解率が悪化する種類の参照表現があることも確認された。この問題は、代名詞のときだけ参照ドメインを単一の基底ドメインに限定することで解決でき、最終的に全体の性能も従来手法より向上する傾向が確認された。今後は更にデータの数を増やして評価を続ける予定である。また今回 $P(X|D)$ の推定には単一のモデルを使用したがる、これについても、代名詞とそれ以外で別々

のモデルを使い分けることでより性能が高まる可能性がある⁷⁾ので今後試みる予定である。

参 考 文 献

- 1) Dale, R. and Reiter, E.: Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions, *Cognitive Science*, Vol.18, pp.233-263 (1995).
- 2) Dale, R. and Viethen, J.: Referring Expression Generation through Attribute-Based Heuristics, *Proc. the 12th European Workshop on Natural Language Generation (ENLG)*, pp.59-65 (2009).
- 3) Denis, A.: Generating Referring Expressions with Reference Domain Theory, *Proc. the 6th International Natural Language Generation Conference (INLG)* (2010).
- 4) 船越孝太郎, 中野幹生, 徳永健伸, 飯田 龍: ベイジアンネットワークを用いた参照表現モデルの提案, 情報処理学会研究報告, Vol.2011-NL-202, No.4 (2011).
- 5) Hobbs, J.R.: Resolving pronoun references, *Lingua*, Vol.44, pp.311-338 (1978).
- 6) Horacek, H.: Generating referential descriptions under conditions of uncertainty, *Proc. 10th European Workshop on Natural Language Generation (ENLG)* (2005).
- 7) Iida, R., Kobayashi, S. and Tokunaga, T.: Incorporating Extra-linguistic Information into Reference Resolution in Collaborative Task Dialogue, *Proc. 48th Annual Meeting of the Association for Computational Linguistics*, pp.1259-1267 (2010).
- 8) Mitkov, R.: *Anaphora Resolution*, Studies in Language and Linguistics, Pearson Education (2002).
- 9) Ng, V.: Supervised Noun Phrase Coreference Research: The First Fifteen Years, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp.1396-1411 (2010).
- 10) Salmon-Alt, S. and Romary, L.: Generating Referring Expressions in Multimodal Contexts, *Proc. the INLG 2000 workshop on Coherence in Generated Multimedia* (2000).
- 11) Spanger, P., Yasuhara, M., Iida, R., Tokunaga, T., Terai, A. and Kuriyama, N.: REX-J: Japanese Referring Expression Corpus of Situated Dialogs, *Language Resources and Evaluation* (2010). Online First, DOI: 10.1007/s10579-010-9134-8.
- 12) Thórisson, K.R.: Simulated Perceptual Grouping: An Application to Human-Computer Interaction, *Proc. the 16th Annual Conference of the Cognitive Science Society*, pp.876-881 (1994).
- 13) 庵 功雄: 「この」と「その」の文脈指示的用法再考, 一橋大学留学生センター紀要, Vol.5, pp.5-16 (2002).

*1 一方、その後同じ群に対して照応的に代名詞が続く場合は比較的正確にしていた。