

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

船越 孝太郎^{†1} 中野 幹 生^{†1}
徳 永 健 伸^{†2} 飯田 龍^{†2}

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する。提案モデルは、記述、照応、直示を統一的に扱い、参照表現の理解と生成に用いることができる。本稿では、提案モデルの概要とその利用方法を説明し、様々な参照表現をどのようにベイジアンネットワークとして表現するのかを示す。そして音声対話システムへの統合に向けた議論を行い、関連研究について述べる。

On a Bayesian Network-based Model for Referring Expressions

KOTARO FUNAKOSHI,^{†1} MIKIO NAKANO,^{†1}
TAKENOBU TOKUNAGA^{†2} and RYU IIDA^{†2}

A Bayesian network-based model available both for resolution and generation of referring expressions in dialogue systems is presented. The model handles descriptions, anaphoras, and deixes in a unified way. This paper describes the overview and use of the model, and shows in what manner various expressions are handled as Bayesian networks. Finally integration of the model to spoken dialogues systems and related work are discussed.

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、記述表現、照応表現、直示表現を含む。参照表現はほとんどの対話において現れる要素である

ので、問題領域によって単純・複雑の違いはあっても、対話システムは参照表現を理解（解決）、生成する機能を備えておく必要がある。

本稿では、対話システムのための^{*1}確率に基づく参照表現のモデルを提案する。確率に基づく参照表現のモデルを提案するのは以下の理由による。前述の通り、対話システムは参照表現の理解と生成を共に行う必要がある。しかし、それぞれを別々に実装・保守するのはコストがかかる。また、理解と生成の間に容易に齟齬が生じるため、自分の生成した表現を自分で正しく理解できないようなシステムを作ってしまう危険が大きい^{*2}。そこで、理解と生成を同一の参照表現モデルに基づいて実現することが望まれる。確率に基づく参照表現の理解モデルを用意すれば、後述するように、参照表現の生成もそのモデルを用いて実現できる^{*3}。また、参照表現はしばしば曖昧であるが、その曖昧さの程度を確率の形で表現できることは、理解においても生成においても利点を持つ。理解においては、2つの参照対象の間で曖昧な表現であっても、どちらか一方に対する参照確率が他方に対するそれよりも十分に大きければ、確認発話を省略して対話を先に進める判断ができる。生成においては、ある簡潔な表現が指示対象に対して十分に高い参照確率を持てば、それが厳密には曖昧な表現であっても（聞き手が正しく推定することを期待して）選択することができる。

本稿の提案モデルは確率モデルの枠組みとしてベイジアンネットワークを用いる。これにより、ベイジアンネットワークのために開発された手法やツールを利用できる。また、視覚的なネットワークで表現することで、確率式を直接扱うよりも、様々な言語表現に対応する定式化や考察を行い易くなる。提案モデルを用いる際には、後述する WCXD 構造を基本として、参照表現毎に1つのベイジアンネットワーク（以後、これを**参照表現ベイジアンネットワーク**とよぶ）を構成する。この WCXD 構造上での確率に基づく推論が、記述、照応、直示の全てを統一的に扱うことを可能にする。

まず2節で提案モデルとその利用方法を説明し、次に3節にて様々な参照表現をどのようにベイジアンネットワークとして表現するのかを示す。具体的には、照応表現（one-anaphora及び指示詞）に加え、場所と位置、時間、事象、群、補集合、部分への参照について議論する。そして4節にて音声対話システムへの統合に向けた議論を行う。最後に5節で関連研

^{†1} (株) ホンダ・リサーチ・インスティテュート・ジャパン
Honda Research Institute Japan Co., Ltd.
^{†2} 東京工業大学
Tokyo Institute of Technology

^{*1} 主に課題指向の対話システムを想定している。個々のシステムは、ホテル予約のような限られた課題ドメインを扱うことが前提となる。

^{*2} システムが使った参照表現はユーザによって模倣され（interactive alignment⁶⁾）システム自身に戻ってくる。

^{*3} 理解と生成では最適化を行う際に重視する側面などが異なるため全く同一の処理とはならないが、本稿ではそのような差異は提案モデルの外側で扱うと想定する。2.3節と2.4節を比較されたい。

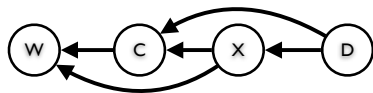


図1 WCXD 基本構造

究について述べた後、6節でまとめと本文中で触れなかったその他の課題を述べる。

2. 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とその利用方法（理解と生成）について説明する。簡単のため、以降では参照表現ベイジアンネットワークを REBN (Referring Expression Bayesian Network) とよぶ。また、確率変数 V の定義域を $\mathcal{D}(V)$ のように表す。

2.1 ベイジアンネットワーク

ベイジアンネットワークは情報の間の依存関係を非循環有向グラフと条件付確率表の形で表現したモデルで、様々な自然言語処理のタスクに応用されており²⁰⁾、参照表現に関する先行研究もいくつかある。Burger と Connolly²⁾ は照応解決に特化したベイジアンネットワークを提案している。Weissenbacher¹⁹⁾ は、非照応的な用法の it の特定にベイジアンネットワークを用いている。これらに対し、本稿では、照応に限らず、記述、直示も含む参照表現の参照対象の特定（参照解決）を統一的に扱うモデルを提案する。

ベイジアンネットワークで表現する情報間の依存関係として因果関係が考えられることが多いが、ベイジアンネットワークの定義自体は因果関係を求めるものではなく⁹⁾、実際に多様な関係が条件付確率の形で数値化され用いられる²⁰⁾。本稿の提案モデルにおいても、因果関係（生成過程）だけでなく、位置関係、包含関係（部分-全体関係）なども表現する。

2.2 WCXD 基本構造

1節で述べたように、REBN は WCXD 構造を基本とする。図1に WCXD 構造を示す。

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、 W が語の選択、 C が概念の選択、 X が参照対象（対話者が置かれた状況で認識されている物理的な事物や、談話中に言語的に導入された事物）の選択、 D が参照ドメイン^{5),15)} の選択を表現している。参照ドメインとは参照対象を含む集合 ($\mathcal{D}(X)$ の部分集合) であり、談話的あるいは知覚的な作用によって対話者に認識される心的表象である^{*1}。対話

*1 心的表象としての参照ドメインには、集合の要素について、顕現性や焦点などの情報も付随している。

システムにおける課題ドメインや、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^{*2}。参照ドメインは、別の参照ドメインをその要素として包含することで、階層構造をなす^{*3}。

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、その中から参照対象 X （どれについて述べるか）が決定される。その参照対象 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^{*4}。観測されるのは W だけである。概念 C の選択は、対象 X だけでなく X を包含する参照ドメイン D にも依存し、語 W の選択は、概念 C だけでなく対象 X にも依存する。この2つの依存関係の必要性については後述する。

図1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「机」）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1中の REBN を確率式で表現すると以下ようになる。

$$P(W, C, X, D) = P(W|C, X)P(C|X, D)P(X|D)P(D) \quad (1)$$

この REBN を用いて、ある参照表現の参照対象が 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「赤い机」）は、図2に示す形の REBN を用いる。 W_1 が「赤い」、 W_2 が「机」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1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。

2語で2つの対象を参照する場合（例えば「僕の机」）は、図3に示す形の REBN を用いる（煩雑なので、 X から W へのリンクと D から C へのリンクは省略してある）。 W_1 が「僕」、 W_2 が「机」に対応する。この場合は、 $P(W_1, W_2, C_1, C_2, X_1, X_2, D_1, D_2)$ から $P(X_2|W_1, W_2)$ を求める。

参照表現ベイジアンネットワークの構造は参照表現が与えられてから決まり、確

*2 これらは、対話システム構築時に決まる背景パラメータとして、モデルの外側で固定されていると考える。

*3 例えば、ある家具店の中の「家具の集合」、その中の「椅子の集合」、「机の集合」など。

*4 ここでいう語とは、概念 C 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）まで含めて考えている。例えば、「もう一つの」という3単語からなる表現は、another という概念の1つの「語」と捉える。

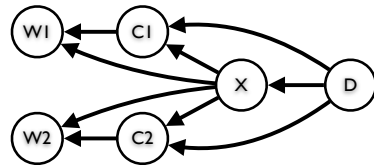


図 2 2 語からなる参照表現ベイジアンネットワークの例 1

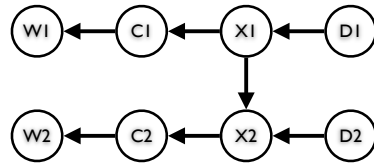


図 3 2 語からなる参照表現ベイジアンネットワークの例 2
(W-X 間, C-D 間のリンクは省略)

率変数の定義域も動的に変わるので、ネットワークの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 ($P(W|C, X)$, $P(C|X, D)$, $P(X|D)$, $P(D)$ など) を関数の形で実装して与える。勿論、これらの個々の関数をコーパスから得た統計量に基づく条件付き確率表の形で実装することは可能である。

確率分布をどのように与えるのかは後回しにして、REBN を用いた参照表現の理解と生成について先に説明する。

2.3 理 解

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) 言語表現からの構造抽出：

語で表現されている概念間の依存構造を抽出する。ここでいう依存構造とは意味的な構造であり、係り受け解析などの構文解析の結果とは異なる。ここで抽出される構造を **REX-graph** とよぶ。REX-graph については、下で簡単に説明する。

(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を表現

したものに過ぎないので、これを REBN に変換する必要がある。この変換を実現するアルゴリズムについては本稿では議論しない。

(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、参照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、

$$x' = \operatorname{argmax}_{x \in \mathcal{D}(X)} P(x|e) \quad (2)$$

となる。 e は E 中の変数への値割り当て (観測値, 証拠) を表す。もし第 2 位のものとの差が小さければ、対話システムの場合なら確認対話を駆動してもよい。

2.3.1 REX-graph

前述のように、REX-graph は語で表現された概念間の依存構造のことである。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** とよぶことにする。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** とよぶことにする。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。

REX-graph を、単位参照表現をノードとする有向グラフとして定義する。ノードである単位参照表現は、ある参照対象の属性表現 (語) の集合として、空白で区切った属性表現を丸括弧で括って表す。例えば「赤いボール」は

A:(赤い ボール)

と表す。A はノードのラベルである。このラベルを使って単位参照表現 A,B 間の依存関係 (ノード間のリンク) を、 $A \rightarrow B$ と表す。 $A \rightarrow B$ は、B の参照対象は A の参照対象によって決まることを表す。図 2 中の REBN は、この REX-graph から構築される。

これにより「机の上」は

A:(机) B:(上) $A \rightarrow B$

と表す。図 3 中の REBN は、この REX-graph から構築される。複合参照表現において、主辞 (「上」) の方の参照を **主参照**、それ以外の参照 (主参照を特定するための基準点への参照) を **補参照** とよぶことにする。

「左の机の引き出しの中」なら、

A:(左) B:(机) C:(引き出し) D:(中) $A \rightarrow B$ $B \rightarrow C$ $C \rightarrow D$

となる。

「椅子と机の間のボール」のような三項関係は

A:(椅子) B:(机) C:(間) D:(ボール) $A \rightarrow C \ B \rightarrow C \ C \rightarrow D$

のように表せる。

言語表現から適切な REX-graph を構築する過程は必ずしも簡単ではないのでこれ自体も研究の課題となるが、本稿ではこれ以上議論しない。また、読みが曖昧な表現であれば、その読みに応じて複数の REX-graph が作られることになる。すなわち、参照表現の理解は

$$(i', x') = \operatorname{argmax}_{i \in \mathcal{I}, x \in \mathcal{D}(X)} P^i(x|e) \quad (3)$$

と再定式化できる^{*1}。ここで、 \mathcal{I} は可能な読みの集合で、 $P^i(X|E)$ は i 番目の読みに対して作られた REBN から得られた参照確率の分布である。

2.4 生成

REBN を用いた参照表現生成は、generate-and-test 方式で行える^{*2}。すなわち、なんらかの方法で参照表現の候補を生成し、それを REBN を用いて評価（候補毎に 1 つの REBN を作る）し、最も評価値の高いものを選択する。表現候補の生成方法、および評価と選択の方策にはさまざまな工夫が可能であるが、本稿では特定のものを詳細に議論することはしないで概略を示す。評価の方策には、(1) 参照確率だけを考慮する、(2) 表現の生成確率を考慮する、(3) ディストラクタ（目的の参照対象以外の存在）との確率値の差の大きさを考慮する、といったものが考えられる。選択の方策には、(A) 生成し得るものを全て作ってから評価値最大のものを選ぶ、(B) 短いものから逐次生成・評価し閾値以上の評価値を持つものを見つけた時点で終了する、(C) 人間の特性を考慮した方策に沿って逐次生成・評価し閾値以上の評価値を持つものを見つけた時点で終了する^{*3}、といったものが考えられる。

上記の (A) の方策の下で、定式化の例を示す。まず、参照表現候補の集合 \mathcal{E} を生成する。次に、参照表現毎に REBN を構築し、参照対象を表す変数について周辺化した $P^e(X|E=e)$ を得る。参照対象 x に対して高い参照確率を持つ表現を選択するだけであれば、これは

*1 あくまで一例であり、これ以外にも様々な定式化が考えられる。例えば、「読み」の自然さを確率の形で考慮することも可能である。その上で i に関して周辺化してしまうことも考えられる。どのような定式化がよいかは定量的な評価で決定すればよい。

*2 REBN は（確率モデル論的な意味で）生成的なモデルであるので、もっと積極的にモデルを言語生成に利用できる可能性もある。例えば REBN の X, D などを観測値として与えることで、妥当そうな少数の表現候補だけを留意して、探索範囲を小さくすることも可能と思われる。しかしこの可能性についてはこれ以上議論しない。

*3 参照表現生成で広く知られている incremental algorithm⁴⁾ はこれに相当する。

$$e' = \operatorname{argmax}_{e \in \mathcal{E}} P^e(X=x|E=e) \quad (4)$$

となる。表現自体の自然さ（生成確率）を表現した生成モデル $P(\hat{E})$ を考慮して、

$$e' = \operatorname{argmax}_{e \in \mathcal{E}} P^e(X=x|E=e)P(\hat{E}=e) \quad (5)$$

とすることもできる^{*4}。

2.5 確率分布の与え方

REBN が必要とする各種の（条件付き）確率分布の与え方の基本的な指針を説明する。

2.5.1 $P(D)$

参照ドメイン d ($d \in \mathcal{D}(D)$) は 1 つ以上の参照対象を含む集合であり、 $P(D)$ は今何について話しをしているのかについての確率分布である。

例えば、ある家具店の中で、客が家具について店員に話しかけたとする（例えば「すみません、ちょっといいですか」）。このとき店の中の全ての家具の集合である参照ドメイン d_0 が最も確率が高い。その後、客がイスについてまず質問を始めると（例えば「イスを探しているのですが」）、店の中の全てのイスを含む参照ドメイン d_1 の確率が最も高くなる。さらに、客が特定のイスをピックアップして検討を始めれば（例えば「このイスかあのイスがいいと思っているのですが」）、その特定のイス 2 つを含む集合が参照ドメイン d_2 として認識され、かつ確率が最も高くなる。このように $\mathcal{D}(D)$ は対話の進行に沿って動的に更新される。実際のシステムでどのように $\mathcal{D}(D)$ を更新するかはここでは触れない。

Denis⁵⁾ は、これらの参照ドメインを、度数で表現した顕現性の順に並ぶリストで管理している。最も単純な $P(D)$ の設定方法の 1 つは、アドホックであるが、リストの順に確率値を線形配分する方法である。この方法で一般的な問題に対しては十分なのか、さらに改善が必要なのかは今後具体的な評価を通して検討する必要がある。

2.5.2 $P(X|D)$

$P(X=x|D=d)$ はある参照ドメイン d が与えられたときに、その中に含まれる参照対象 x が参照される確率を表現する。従って、 $\mathcal{D}(X)$ は $\mathcal{D}(D)$ に含まれる全ての要素であるが、実質的には $\mathcal{D}(X) = d$ である ($x \in d$ でなければ $P(x|d) = 0$ となる)。 $P(X|D)$ は、過去の文脈（直前の発話で参照されたか）や現在の状況（話し手の視線など）をもとに、ある対象が次に参照される確率を動的に与える。例えば小林ら²³⁾ が提案するような手法で順

*4 $\mathcal{D}(\hat{E}) = \mathcal{E}$ であるが、 $\mathcal{D}(E) \neq \mathcal{E}$ なので、式 5 中の確率式は $P(X, \hat{E})$ とは形式上は等価でないことに注意されたい（実質的にはほぼ等価と予想される）。つまり 1 つの REBN から直接的に $P(\hat{E}|X)$ を求めることはできない。

位付け器を構成し、その出力値を確率値に変換することで実現できる。物理世界への直示的な参照表現は主にこの情報に基づいて理解される。

また、 X には単体の対象に限らず対象の集合、すなわち参照ドメイン自体もその値として取ることも許す。これにより集合への参照（例えば「あそこの2人」）も扱うことができる。当然ながらこのためには、 $D(D)$ の要素として、参照ドメインをその要素として含む参照ドメインも含まれている必要がある。つまり参照ドメインは入れ子状になる。

2.5.3 $P(C|X, D)$

$P(C = c|X = x, D = d)$ はある対象 x を参照するために概念 c が選択される確率を表現している。一義的にはその概念 c の対象 x への適合度 (relevancy) によって決まる（例えば c が「赤色」という概念であれば、 x がどれくらい赤っぽいかによる）が、「ちゅうくらい」などの相対的な概念は比較対象が決まらなければ適合度そのものを測ることができない。そのため、前件として D を含んでいる。暗黙的に $x \in d$ である。

純粹に生成的に考えれば $D(C)$ は問題領域に現れる全ての概念の集合であるが、REBN 構築時は参照表現（正確には REX-graph）が与えられていることが前提なので、通常 $D(C)$ は観測された W の値からただか数個の概念の集合に限定できる。例えば「イス」という言語表現が観測されているとき、それに対応する C の値としては chair も sofa もあり得るが、table という概念を考える必要はない*1。

2.5.4 $P(W|C, X)$

$P(W = w|C = c, X = x)$ はある対象 x を参照するために概念 c が選択されたとき、それが実際にどのような言語表現として現れるかを表している。一般的には $P(W|C, X) \approx P(W|C)$ と考えて支障ないが、一部の表現は X に依存すると考えると都合が良い。固有名は $P(W = \text{田中さん}|C = \text{name}, X = \text{human1})$ のように、 X に依存するものとして表現できる。もしモデルを単純化するため $P(W|C, X)$ から X を排除しようとすれば、「田中」という人物に対して、対象（インスタンス、 $X = \text{human1}$ ）だけでなく、その人物だけを表す概念（クラス、 $C = \text{tanaka}$ ）も用意する必要が生じる。また、「あれ」や「これ」などの指示代名詞は、話し手及び聞き手と参照対象 x の間の物理的・心理的な距離に依存する。 X を排除するために、これらの代名詞は個々の代名詞に対応する特別な認知的概念から生じると考えることもできる。つまり「あれ」に対して far-object のような概念を想定

し、 $P(W = \text{あれ}|C = \text{far-object})P(C = \text{far-object}|X = x, D = d)$ のように考え、対象 x が「あれ」とよばれる確率はある時点での対象 x の概念 far-object に対する適合度に依存すると考える。しかしこの考え方では、後述する参照の二重性をうまく扱えなくなるという問題が生じる。

$P(W|C, X)$ の確率分布は事前にコーパスを収集することで頻度分布として統計的に推定できる*2が、実際には対話の進行に沿って動的に更新されるべきである。例えば「山田さんの社員番号を教えてください」という発話の後に「田中さんの番号も教えてください」という発話がなされれば、2番目の発話の中の「番号」は電話番号ではなく社員番号のことであると理解するのが普通である。これは2番目の発話の時点では $P(W = \text{番号}|C = \text{tel\#}, X)$ より $P(W = \text{番号}|C = \text{staff\#}, X)$ の方が高くなることで説明できる*3。また、対話者が使用する言語表現を動的に共通化して行く interactive alignment⁶⁾、特に lexical alignment⁸⁾ も同じような仕組みで対処できると推察される。この $P(W|C, X)$ の動的な更新をどのように行うかは今後の検討課題である。

2.5.5 $P(X'|X)$

$P(X' = x'|X = x)$ は対象 x と対象 x' の間の関係性の有無を表現する。どのような関係性かは x と x' によって決まる。図3中の X_1 から X_2 へのリンクがこれに相当する。例えば「僕の本」という表現を考える。ある本 x' に関して、話し手 x （「僕」）に所有されている、あるいは、話し手 x によって書かれたという事実があれば、 $P(X' = x'|X = x)$ の値は1に近くなる。そうでないことがはっきりしていれば $P(X' = x'|X = x)$ の値は0に近づき、関係性が不明であれば0.5である。

図3のREBNを動かすためには、 $P(X_2|X_1, D_2)$ を与える必要がある。REBNにおいては $P(X_2|X_1, D_2) \approx P(X_2|D_2)P(X_2|X_1)$ と近似できると考えている*4。

3. 各種参照表現の取り扱い

本節では、さまざまな参照表現をREBNでどのように「表現」するか、事例別に説明する。

3.1 one-anaphora

口語対話においては「やつ」のようないわゆる one-anaphora が頻出する。例えば、「山

*1 これは、ある具体物としての机（例えば小さな作業台）が chair として認識されて「イス」と表現されることを排除するものでない。

*2 この推定は $P(W|C, X) \approx P(W|C)$ として行うことになる。

*3 $P(W|C, X)$ が X を考慮するかどうかは、 C の値によって決まると考えている。この例の場合は、前段の例とは変わって X は無視すればよい。

*4 一般的にはこの近似は成り立たない。この近似の妥当性は今後検証する必要がある。

田さんの社員番号を教えてください」に続く「田中さんのやつも」*1のような発話である。英語の場合なら、例えば“that blue one”。これは、2.5.4節で既におおよそ説明しているように、 $P(W|C, X)$ を文脈に合わせて動的に更新することで対処できると考えられる。例えば、先の2.5.1節での家具店の例の場合、 $P(W = \text{やつ} | C, X)$ の $D(C)$ は、イス、机、タンスなど様々な概念を含むが、「イス」に付いて話している間は $P(W = \text{やつ} | C = \text{chair}, X)$ の確率がその他の概念よりも格段に高くなればよい*2。

参照表現、特に照応表現を理解する場合には、その表現が現れる格や「も」のような格助詞の持つ情報も重要である。このような情報は過去の文脈と併せてパラメータ θ として与え、 $P(W|C, X)$ の算出に利用する。すなわち $P(W|C, X; \theta)$ 。

2.5.4節で触れた参照の二重性について説明する。参照表現においては、概念への参照と存在への参照が二重になっていることがある。例えば、「山田さんの靴をとって」という発話の後の「田中さんのやつも」という表現は、概念的には同じ「靴」を参照し、同時に存在としては別の「靴」を参照している。このとき、田中さんの靴自体を知らずに参照対象を特定できなかったとしても聞き手には「やつ」が何か判るので、「田中さんの靴はどれ？」のように問い返すことができる。REBNにおいて推論対象の変数である X と観測変数である W の間に C を挟んでいる理由は、この二重性を扱うためである。

3.2 指示詞

指示代名詞と指示形容詞については、互いに少し異なった扱い方をする。

まず指示形容詞については、指示形容詞に相当する特別な概念を仮定する。例えば「そのボール」の場合、図2中の構造のREBNが使用され、で「その」に対応する C_1 には、demonstrativeという概念が入ると考える。demonstrative概念が表層で「その、あの、この、それらの、これらの、あれらの」のうちのどの形を取るかは X の値によって決まる。英語などの言語における冠詞も指示形容詞と同様の扱い方で処理できると考えている。一方、「それ」「あれ」のような指示代名詞の場合は3.1節の場合と同様、 C には「靴」のような具体的な概念が入ると考える。これにより「それ」が指すものが「どれであるか」だけでなく「何であるか」も同時に推定できる。

指示形容詞に戻る。「この」や「その」には、指定指示と代行指示の2つの用法がある^{21);22)}*3。

*1 「田中さんのもの」のように形式名詞（準体助詞）の「の」が使われる場合も、「の」を「やつ」と同等に捉えることで対処できる。

*2 この場合も X は無視できる。

*3 金水²²⁾は「その」だけが代行指示可能としている。

先の「そのボール」は指定指示の例である。それに対し「リングを食べたがその味はひどかった」というときの「その」は「その」という意味であり「リング」を指す。これは代行指示の例である。代行指示の場合は図3の構造を与える必要がある。このためには、表層表現からREX-graphを構成する段階で、指定指示か代行指示かを区別する必要がある。

3.3 場所と位置

場所（空間）は明瞭な境界を持たないことが多いが、参照対象となる1つの領域と捉える。従って、「机の右のボール」という表現に対しては図4中のREBNを与える。「右」が W_2 に対応する。「右」と参照しうる空間は大量にあるので、全ての可能性を探索するのは非効率的である。従って実装においては、 W_1 （「机」）の参照対象である X_1 の候補を絞り込んだ上で、それらに対する「右」の候補だけを X_2 の定義域として設定する必要があると思われる。このとき、 D_2 は不要かもしれない。

$P(X_2 = x_2 | X_1 = x_1)$ によって、ある空間 x_2 の「 x_1 の右」の解釈としての尤もらしさを表現できる。例えば、「〇〇の右」というとき、実際にどの空間を意図しているのかは参照枠に依存する。参照枠は明確に言語化されないことがほとんどだが、話し手や聞き手と対象（「〇〇」）の位置関係や対象の性質（前後左右の方向性を内在するかどうか）などによって、一定の傾向を示す。このような傾向を $P(X_2 | X_1)$ の分布として表現することで、参照枠について曖昧な表現の解釈を順序づけられる。

$P(X_3 = x_3 | X_2 = x_2)$ によって、空間 x_2 中でのある対象 x_3 の位置の典型性（適合性）を表現できる。例えば「正門の真正面の建物」といえば、多少遠くても門と直行する線上にある建物のことであろうし、単に「正門の前の建物」といえば、多少正面からはずれていても比較的近くの建物のことと受け取るのが普通であろう。このような分布はポテンシャル関数¹⁸⁾を用いて表現できる。

参照対象としての場所とは別に、ある存在の動的・相対的な属性として位置を考える。例えば「真ん中のボール」というような表現の「真ん中」は2.5.3節で取り上げた「ちゅうくらの」と同様の相対属性と捉え、図2中の構造のREBNを与える。この「真ん中」を位置ではなく場所として捉える場合は、参照の基準点（「机の右」の場合の「机」）が省略されていると考え、これに対処する特別な措置を講じる必要がある。属性としての位置という考え方を採用すれば、そのような仕組みは不要となる。その代わりに、言語表現からREX-graphを構成する際に場所か位置かで構造を作り分ける仕組みが必要になるが、現時点ではなるべくREBNを単純にするほうがよいと考えている。また、日本語には「右」や「真ん中」に相当する形容詞はないが、英語にはrightやmiddleといった形容詞があり、属

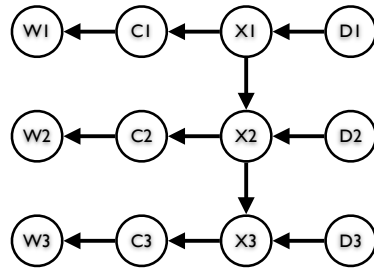


図4 3語・3単位参照表現に対応する参照表現ベイジアンネットワーク
(W-X間, C-D間のリンクは省略)

性的に捉えたほうが自然ともいえる。

「机の右のボール」のような複合参照表現を、REBNで参照解決することには大きな利点がある*1。「机の右のボール」という表現は「ボール」に対して「机の右にある」という制約を与えているだけでなく、「机」に対しても「ボールが右にある」という制約を与えている。このことを考慮せずに、補参照から主参照に向けて逐次的に参照解決をするアルゴリズムを作っても、「机」の参照解決ができないために停止してしまう。また、このような参照表現に対して対話的に「机」の曖昧性を解消しようとする、「ボールが右にある机」であることを考えれば話し手には「机」がどれかはあまりに自明なため、話し手に非人間的で不快な印象を与える。しかし、逐次的な処理を基本としたままこれを回避しようとする、複雑なアルゴリズムを用意しなければならなくなる。REBNを用いる場合は、補参照と主参照を同時的に推定できるので、このような問題を回避できる。

3.4 時間

時間は「場所」と同様に考えることができる。時間に関する表現には、「今日」のように比較的確な境界を表現も、「今」のように境界のはっきりしない表現もあるが、これは場所についても「東京都」のように境界のはっきりしたものもあれば、「前」のように境界のはっきりしないものがあるのと同じである。「キリストが生まれた日」のような表現はやは

*1 「机の上のボール」の方が例として判り易いかもしれない。以下のような状況を想像して欲しい。部屋の中にくつろぎながら机があり、床にボールがいくつも転がっている。このとき1つだけ机の上に乗っているボールがあるとする。このとき、「机」について共同注意を事前に築いていなくても、「机の上のボール」と言うだけで1つのボールを特定できるはずである。

り場所と同様参照対象を持つ表現と考えるのがよく、そうでなければ「その日は何曜日？」のような質問をうまく扱えなくなる。

これに対し、「さつき」のような表現は、場所に対する位置のように、属性表現として考えた方が扱い易い。「さつき」を場所的に考える利点も特に思い当たらない。一方、「昨日」や「今日」のような時間的な直示表現は、場所的に考えるべきか位置的に考えるべきかの判断が難しい。場所のように考えた方が一貫性があるように思えるが、「昨日」や「今日」が対話中に先行してもそれを「その日」と参照することは稀であれると思われる。アプリケーションによってはこれらを属性表現と考えてしまった方が実装の手間が少ないと思われる。これは、システムを構築する時にアプリケーション毎に判断すべきことと思われる。

3.5 事象

事象（出来事）も参照対象として考える。例えば、「昨日買ったボール」という表現に対しては、図4中のREBNを与える*2。

場所、時間、事象、存在（もの）はそれぞれ独立した参照ドメイン空間で管理し、「昨日買ったボール」の場合であれば D_1, D_2, D_3 の定義域は全て別個に考えるのが良いと考える。このためには、REBN構築時には、それぞれの参照対象が場所、時間、事象、存在のいずれであるのかを明らかにできる必要がある。

3.6 群

図5に「右の3つのボール」に対応するREBNの構造を示す。前述のように、 X の値として3要素からなる参照ドメインが入ると考えることで群に対する参照を扱う。数については、quantityという数量に対応する概念があると考えられる。つまり「3つ」に対応する C_2 の値がquantityとなる。

「赤と白の2つのボール」のように、群と群の中の要素について同時に参照することがある。この参照表現に対するREBNの構造を図6に示す。 W_1 と W_2 がそれぞれ「赤」と「白」に対応し、 W_3, W_4 が「2つのボール」に対応する。ここでの $P(X_1|X_3)$ および $P(X_2|X_3)$ は包含関係にあるかどうかという制約を0/1で表現する。

「赤と白の2つのボール」には、「2つの赤白まだらのボール」という読みも可能である。その場合に作られる構造は、図6の構造から X_1 と X_2 を除き、 X_3 から C_1, C_2 それぞれにリンクを張ったものになる。

「夫婦」のような集合概念は、特に数量の指定がなくても群を参照対象とする。数の制

*2 「昨日」も参照対象を持つと考える場合。

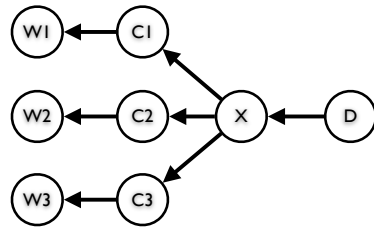


図5 3語・1単位参照表現に対応する参照表現ベイジアンネットワーク
(W-X間, C-D間のリンクは省略)

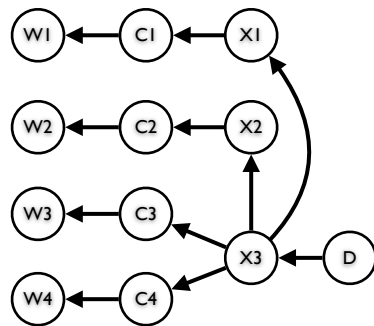


図6 包含関係を含む参照表現ベイジアンネットワークの例
(W-X間, C-D間のリンクは省略)

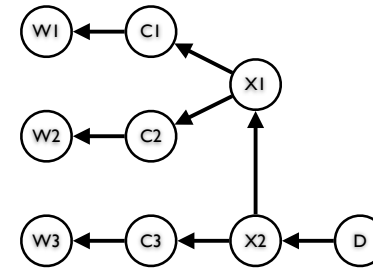


図7 部分に言及する参照表現ベイジアンネットワークの例
(W-X間, C-D間のリンクは省略)

約は、概念に関するオントロジ的知識により $P(C|X, D)$ の中に暗黙的に組み入れることもできる。すなわち $C = \text{couple}$ なら、 X の値となる参照ドメインの要素数が2のときだけ、 $P(C|X, D)$ が1に近い値を持つようにする。しかし、もっと直接的に $D(X)$ から要素数の一致しないものを除いてしまう方が計算の効率はよいと思われる。

3.7 補集合

「もう一つの机」のような補集合表現も、動的な相対属性として扱える。「もう一つの」という表現が *another* という概念に対応すると仮定すると、「もう一つの机」に対する REBN 構造は図2中のものと同じになる。参照ドメイン d が2つの机 x_1, x_2 を含み、 x_1 が直近の過去に参照されていた (d 中では x_1 に焦点が当たっていた) とする。その場合、 $P(C = \text{another}|X = x_2, D = d) > P(C = \text{another}|X = x_1, D = d)$ となることで、「もう一つの机」が x_2 を参照することが判る。

3.8 部分への言及

「角の赤い机」という表現は机の一部(「角」)に言及することで「机」を特定している。図7にこの表現に対応する REBN の構造を示す。

部分に言及する表現において、厳密には曖昧でも、実際にはほとんど曖昧性のない参照表現がある。例えば、図8に示す状況²⁴⁾では、論理的には A,B 両方のテーブルは同じ程度に「角の赤い机」である(寧ろ面積的には A の方が角が赤い)が、人間はこの表現が B をさすものであると理解する。これは、B の「角」の視覚的顕現性の高さによると考え、それを $P(X_1|X_2)$ の値の大きさに反映することで対処できる。 $D(X_1)$ は $D(X_2)$ の要素となる物体の部分の集合になる。

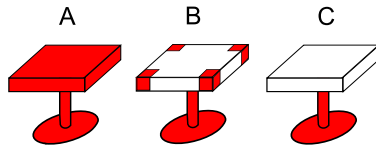


図8 視覚的顕現性によって曖昧性が消失する例

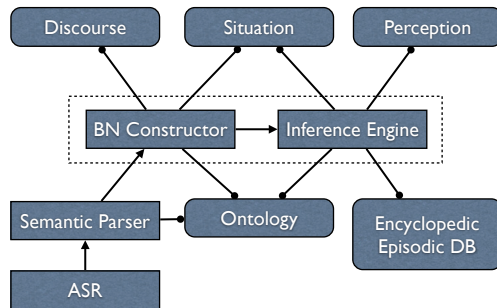


図9 音声対話ロボットにおける参照表現理解のためのモジュールアーキテクチャ

4. 対話システムとの統合に向けて

2節で提案したモデルを実装した参照表現理解装置を、多様な対話システムで用いることのできる汎用的なモジュール（以後、参照表現エンジンとよぶ）とするためには、外部モジュールとのインターフェースやプロトコルを十分に明らかにしておく必要がある。事前に参照対象 ($D(X)$) や参照ドメイン ($D(D)$) についての情報が全て判っている閉じた世界を対象にするのであれば、参照表現エンジンは、入力として (1) REX-graph, (2) それまでの談話文脈, (3) 現在の世界の状況（参照対象, 参照ドメイン, 話し手の視線情報などを含む）を受け取り、参照対象の参照確率分布を出力する単純なフィルタ型のモジュールとして機能すればよい。しかし、実際は参照対象や参照ドメインが未確定である開いた世界を対象とすることが多いので、汎用性を求めるなら参照表現エンジンもそれを前提とした設計を必要とする。

開いた世界を対象とし、物理空間も非物理空間（各種データベースなどの情報空間）も対

話の対象とするシステムの例として、モノの運搬や情報アクセスに応える音声対話ロボットを考える。図9に、音声対話ロボットにおける参照表現理解のためのモジュールアーキテクチャを示す。図中の四隅が直角の枠は入力に対して計算処理をするモジュールを表し、四隅の丸い枠は各種の情報を管理するモジュールである。破線で囲まれた部分（BN-ConstructorとInference Engine）が参照表現エンジンに相当する。ASRによって音声認識された発話はSemantic ParserによってREX-graphに変換され、BN-Constructorに渡される。BN-Constructorは、談話情報（Discourse）、状況情報（Situation）、概念知識（Ontology）を参照してREBNを構築する。最後にInference Engineが与えられたREBNに基づいて参照解決を行う。

ロボットが認識していない物理世界の対象をユーザが参照したとき、まずInference Engineはそれが $D(X)$ の中に存在しない対象であることを何らかの方法によって判断する必要がある。これに対する1つのアプローチは、 $D(X)$ の中に常に「その他」の存在elseを含め、適当な参照対象がない場合に、elseの確率が高まるようにする方法である（どのようにこれを実現するかも今後の課題である）。次に、Inference EngineはPerceptionモジュールに働きかけ、参照表現の中に与えられた言語情報を元に物体の探索を駆動する。候補となる物体が見つかった場合には、その情報をSituationに反映し、参照解決を行う。それでも曖昧性のために一定の確信を持って参照解決できない場合や、適当な参照対象が見つからない（依然としてelseの確率が高い）場合には、対話行動管理モジュール（図9外）に必要な情報を与えて、適切な対話行動や探索行動を駆動する必要がある。 $D(D)$ に関しても同様で、ユーザとロボットの知覚的群化¹⁷⁾に齟齬があれば、Perceptionモジュールや対話行動管理モジュールに働きかけて齟齬を解消する必要がある。

先に例としてあげた電話番号や社員番号を検索する対話のような、情報アクセス対話においても同様である。大きなデータベースの内容を全て $D(X)$ として扱えば、ペイジアンネットワークが直ぐに計算不可能な大きくなり、リアルタイムでの応答が行えなくなるので、Situationモジュールが管理する $D(X)$ のサイズはできるだけ小さくすることが望ましい。そのためには先の物体探索の場合と同様、 $D(X)$ に存在しない対象は、動的にデータベースモジュール（Encyclopedic Episodic DB）から取得し、Situationに反映する必要がある。

今後、以上のようなモジュール間のインタラクションを可能にするインターフェースとプロトコルの整理を行う必要がある。

5. 関連研究

Cho と Maida³⁾ は記述表現（「赤い机」のように属性情報で対象を特定する表現）の参照解決のための独自の確率的枠組みを提案している。彼らのモデルは、本稿の提案モデルから C と X の部分だけを取り出し、 C を観測変数と考えたようなものになっている。照応は考慮されていないので、3.1 節で述べた参照の二重性は扱えない。複合参照表現もそのままでは扱えない。参照ドメインに類似する考えとして、primary, secondary, tertiary という 3 つの焦点領域を設定しているが、本稿の提案モデルほど柔軟性はない。

Roy¹⁴⁾ は確率モデルを用いた参照表現生成の枠組みを提案している。Roy らの枠組みでは、記号接合から表層生成の問題まで幅広く扱え、それら全てをコーパスから統計的に学習できる。しかしながら、扱える場面と表現は非常に限られている。Cho と Maida 同様、記述表現が対象であり、照応や直示は対象としていない。

船越ら²⁴⁾ も、参照表現の理解と生成のための確率モデルを提案している。船越らのモデルは、特に物体の部分への言及（3.8 節参照）を扱うための拡張が特徴的であるが、従来個別に研究・対処されてきた記述、直示、照応を統一的に扱うことの重要性を指摘するとともに、提案モデルによってそれが実現できる可能性を議論している。しかしながら、具体的にどのように直示や照応を扱うかは示していない。また、提案モデルそのままでは単位参照表現しか扱えず、群への参照も扱えない。参照対象の集合 ($D(X)$) も予め判っていることが前提となっている。本稿の提案モデルはこれらの問題を解決したモデルになっている。

Salmon-Alt ら¹⁵⁾ や Denis⁵⁾ は、参照対象の集合（参照ドメイン）を動的に構造化して言語生成に用いることの必要性を指摘し、そのための各種のアルゴリズムを提案している。本稿の提案モデルはこの参照ドメインの考え方を取り入れ、確率的に扱えるようにした。

Lison ら¹¹⁾ は、Markov Logic Network (MLN)¹³⁾ を使った参照表現理解の枠組みを提案している。MLN では論理式の形で事前知識を制約として与えることができるが、数量的な知識は制約毎の重みの形でしか与えられず、ベイジアンネットワークに比べて多様な関係性を数値的に表現し難い。また、重みは制約間の複雑な依存関係によって決まるもので、人手で予め適切に与えることは難しく、どうしてもデータからの学習が前提となる。そのため、本稿で述べたようなモデルを MLN で表現することは容易ではない。また MLN は実装技術的にも発達途上であり、少しドメインを大きくするとすぐに計算量が増大して、処理できなくなってしまう問題もある。

6. おわりに

本稿では、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。提案モデルによって、記述、直示、照応を含む多様な参照表現を統一的な枠組みで扱えると期待できるが、今後定量的な評価を行う必要がある。最初の定量的評価は、REX-J コーパス¹⁶⁾ を用いて行う予定である。本稿ではモデルの大枠を示しただけであり、条件付き確率表の動的な更新方法などについては見通ししか示していない。定量的評価に先立って、まずこれらを明らかにする必要がある。また、カテゴリー転換や分配的解釈²²⁾などをどう扱うや、McShane¹²⁾ が示す参照表現にまつわる様々な問題にどのように対処するのかも考えて行く必要がある。本稿では談話構造については議論しなかったが、意図構造⁷⁾ や談話表示構造^{1),10)} の形で議論されてきた意味解釈上の制約を、REBN の中にどのように取り込むのかも検討を要する課題である。

参考文献

- 1) Asher, N. and Lascarides, A.: *Logics of Conversation*, Cambridge (2003).
- 2) Burger, J.D. and Connolly, D.: Probabilistic Resolution of Anaphoric Reference, *Proc. AAAI Fall Symposium on Intelligent Probabilistic Approaches to Natural Language* (1992).
- 3) Cho, S. and Maida, A.: Using a Bayesian Framework to Identify the Referent of Definite Descriptions, *Proc. AAAI Fall Symposium on Intelligent Probabilistic Approaches to Natural Language* (1992).
- 4) 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).
- 5) Denis, A.: Generating Referring Expressions with Reference Domain Theory, *Proc. the 6th International Natural Language Generation Conference (INLG)* (2010).
- 6) Garrod, S. and Pickering, M.J.: Joint Action, Interactive Alignment, and Dialog, *Topics in Cognitive Science*, Vol.1, No.2, pp.292–304 (2009).
- 7) Grosz, B. and Sidner, C.: Attention, Intentions and the Structure of Discourse, *Computational Linguistics*, Vol.12, pp.175–204 (1986).
- 8) Janarthnam, S. and Lemon, O.: Learning Lexical Alignment Policies for Generating Referring Expressions for Spoken Dialogue Systems, *Proc. the 12th European Workshop on Natural Language Generation (ENLG)*, pp.74–81 (2009).
- 9) Jensen, F.V. and Nielsen, T.D.: *Bayesian Networks and Decision Graphs*, Springer, second edition (2007).

- 10) Kamp, H. and Reyle, U.: *From Discourse to Logic*, Kluwer Academic Publishers (1993).
- 11) Lison, P., Ehrler, C. and Kruijff, G.-J.M.: Belief Modelling for Situation Awareness in Human-Robot Interaction, *Proc. the 19th International Symposium on Robot and Human Interactive Communication (RO-MAN)* (2010).
- 12) McShane, M.: Reference Resolution Challenges for Intelligent Agents: The Need for Knowledge, *IEEE Intelligent Systems*, Vol.24, No.4, pp.1541–1672 (2009).
- 13) Richardson, M. and Domingos, P.: Markov Logic Networks (2006).
- 14) Roy, D.: Learning Visually-Grounded Words and Syntax for a Scene Description Task, *Computer Speech and Language*, Vol.16, No.3 (2002).
- 15) Salmon-Alt, S. and Romary, L.: Generating Referring Expressions in Multimodal Contexts, *Proc. the INLG 2000 workshop on Coherence in Generated Multimedia* (2000).
- 16) 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.
- 17) 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).
- 18) Tokunaga, T., Koyama, T. and Saito, S.: Meaning of Japanese spatial nouns, *Proc. the Second ACL-SIGSEM Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, pp.93 – 100 (2005).
- 19) Weissenbacher, D.: A Bayesian Network for the Resolution of Non-anaphoric Pronoun it, *Proc. NIPS 2005 Workshop on Bayesian Methods for Natural Language Processing* (2005).
- 20) 秋葉友良：自然言語処理におけるベイジアンネット，人工知能学会誌， Vol.17, No.5, pp.553–558 (2002).
- 21) 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).
- 22) 金水 敏：日本語の指示詞における直示用法と非直示用法の関係について，自然言語処理， Vol.6, No.4, pp.67–91 (1999).
- 23) 小林俊平， 飯田龍， 徳永健伸， 船越孝太郎， 中野幹生：非言語情報を用いた協調作業対話における照応解析，言語処理学会第 16 回年次大会， pp.800 – 803 (2010).
- 24) 船越孝太郎， スバンガーフィリップ， 中野幹生， 徳永健伸：不均一な物体を考慮した参照表現の確率的モデル，人工知能学会第 23 回全国大会 (2009).