

最大エントロピー法による下位範疇化の確率モデル学習 および統語的曖昧性解消による評価

宇津呂 武仁 宮田 高志 松本 裕治
奈良先端科学技術大学院大学 情報科学研究科

本論文では、統語解析における解析結果の優先付けを目的として、動詞と格要素の共起データから、動詞の下位範疇化の確率モデルを学習する手法を提案する。特に、格の依存関係・格要素の名詞の汎化レベルの曖昧性を統合的に取り扱う。確率モデル学習の手法としては、最大エントロピー法を採用し、格の依存関係および格要素の汎化レベルという二つの異なる情報を素性によって表現し、確率モデルに組み込む。また、格の依存関係に応じて、いくつかの確率モデルのタイプを提案し、EDR 日本語共起辞書を用いた確率モデル学習の実験、および下位範疇化タスクにおける性能測定を通して、これらのモデルの性能を実験的に比較・検討する。

Maximum Entropy Model Learning of Subcategorization Preference and its Evaluation in Syntactic Disambiguation

Takehito UTSURO Takashi MIYATA Yuji MATSUMOTO
Graduate School of Information Science, Nara Institute of Science and Technology

This paper proposes a novel method for learning probabilistic models of subcategorization preference of verbs. Especially, we propose to consider the issues of *case dependencies* and *noun class generalization* in a uniform way. We adopt the maximum entropy model learning method and apply it to the task of model learning of subcategorization preference. Case dependencies and noun class generalization are represented as *features* in the maximum entropy approach. The feature selection facility of the maximum entropy model learning makes it possible to find optimal case dependencies and optimal noun class generalization levels. We describe the results of the experiment on learning probabilistic models of subcategorization preference from the EDR Japanese bracketed corpus. We also evaluate the performance of the selected features and their estimated parameters in the subcategorization preference task.

1 はじめに

一般に、高精度な統語的曖昧性解消を実現するためには、統語レベルだけの知識だけでは不十分で、語彙あるいは意味レベルの知識を効果的に用いる必要があり、コーパスからの言語知識抽出においても、統語レベルの知識とあわせて、語彙・意味レベルの知識を取り込んだ形で抽出しなければならない。ここで、語彙・意味レベルの知識をどのような形式で表現して抽出するかについては大きく分類すると以下の二つのアプローチがある。1) 文法規則・係り受け関係などの統語的な知識に語彙・意味レベルの知識を埋め込んだ形で記述し、この形式の上で知識を抽出するアプローチ [Black93, Magerman95, Collins96]、2) 文法規則・係り受け関係などの統語的な知識と語彙・意味レベルの知識を独立に記述し、知識の抽出も独立に行うアプローチ [李96a, 乾96]。これらのうち、2) のアプローチでは、1) のアプローチの問題として、語彙・意味レベルの知識が統語的な知識に依存しすぎるため、統語レベルの依存関係と意味レベルの依存関係が同型でない場合に問題が生じる点や、統語知識と意味知識の組み合わせでパラメータ数が増加する点を指摘し、統語知識と意味知識を独立に扱う手法の優位性を主張している。

本論文でも、基本的には、2) のアプローチにしたがって統語レベルの知識と語彙・意味レベルの知識が独立であるとみなし、特に、語彙・意味レベルの知識の抽出について議論する。さらに、語彙・意味レベルの知識の中でも、動

詞の共起知識に焦点をあて、統語的曖昧性解消を目的として動詞の共起知識をコーパスから抽出する手法を提案する。

一般に、コーパスから、共起知識あるいは格フレームなど動詞の語彙知識を学習する際には、

- 1) 格の依存関係
- 2) 格要素の汎化レベル

の二つの点を考慮して学習を行う必要がある。1) は、コーパス中の共起データにおける複数の格の共起を観測したときに、これを、格フレーム中の必須格の共起に相当する強い依存関係ととらえるのか、あるいは、独立事象に相当する任意格的な格が偶然共起したととらえるのかという問題に相当する。一方、2) は、格要素の名詞を観測したときに、これを概念階層のどのレベルの概念クラスの共起知識としてとらえるのかという問題に相当する。

これまで、コーパスからの動詞の語彙知識抽出に関する研究として、すでにいくつかの研究がなされているが、その多くはこれらの二つの問題を個別に扱っている。例えば、[Resnik93, Li95] では、動詞と一つの格要素の共起知識の記述において、シソーラス中の最適な汎化レベルを求める手法が提案されており、統語的曖昧性解消を対象として、抽出された共起知識の性能の実験的な評価も行っている。ただし、動詞と一つの格要素の間の共起知識にとどまっておらず、統語的曖昧性解消を直接の目的として、複数の格を含む格フレームを抽出し、その性能を実験的に評価するには至っていない。また、[Li96b] においては、[Li95] の手法

によりそれぞれの格の最適な汎化レベルを求めてから、複数の格要素の間の依存関係を学習する手法が提案されているが、格の依存関係と最適な汎化レベルが独立に学習されており、これら二つを同時には考慮していない。

これに対して本論文では、統語解析における解析結果の優先付けを直接の目的として、1) 格の依存関係、および、2) 格要素の汎化レベル、を同時に考慮して、動詞と複数の格の格要素の共起データから、動詞の下位範疇化の確率モデルを学習する手法を提案する。本論文では、まず、一つまたは複数の(部分的)下位範疇化フレームから、動詞・格要素の名詞の共起の用例を生成するモデルを提案し、これを確率モデルとしてとらえる。確率モデル学習の手法としては、最大エントロピー法 [Pietra95, Berger96] を採用し、これを動詞の下位範疇化の確率モデル学習に適用する。最大エントロピー法では、様々な情報を素性として表現して確率モデルに組み込む機構を備えており、動詞の下位範疇化の確率モデル学習においては、これが大きな利点の一つになる。具体的には、上で述べた格の依存関係および格要素の汎化レベルという二つの異なる情報を素性によって統合的に表現し、確率モデルに組み込むことができる。また、確率モデル学習の際には、上で述べたように、格の依存関係および格要素の汎化レベルの曖昧性を考慮して学習を行わなければならない。これに対し、最大エントロピー法においては、各素性が規定する事象集合が重複することが可能であり、これによって格の依存関係および格要素の汎化レベルの曖昧性を容易にモデルに組み込むことができる。

本論文では、格の依存関係に応じて、いくつかの確率モデルのタイプを提案し、これらのモデルの性能を実験的に比較・検討する。実験として、EDR 日本語電子化辞書 [EDR95] の日本語共起辞書を用いて、下位範疇化の確率モデル学習を行った結果について述べる。さらに、統語解析結果の優先付けにおいて、学習された下位範疇化確率モデルの性能を評価し、各タイプの確率モデルの比較を行う。

2 動詞・格要素の共起用例の生成モデル

本節では、まず、一つまたは複数の(部分的)下位範疇化フレームから、動詞・格要素の名詞の共起の用例を生成するモデルを提案する。

2.1 データ構造

2.1.1 動詞・格要素の共起の用例

ここでは、簡単のため、動詞・格要素の共起データ e を、動詞 v および、格 p と格要素の名詞の意味クラス c の組合わせとして以下のように表現する。

$$e = \begin{bmatrix} \text{pred} : v \\ p_1 : c_1 \\ \vdots \\ p_k : c_k \end{bmatrix} \quad (1)$$

ただし、ここでは、格要素の名詞の多義性は無視する。また、シソーラスを、意味クラスを節点とする木構造の型階層とみなし、各意味クラス c_1, \dots, c_k は、この型階層の葉クラスであるとする。さらに、型階層中の意味クラス間の包含関係を \subseteq で定義し、意味クラス c_1 が意味クラス c_2 に包含されるといふ関係を $c_1 \subseteq c_2$ と記述する。¹

¹本節の定義においては、名詞の意味クラスの曖昧性を無視したが、実際の実装においては、曖昧性のない上位クラスと曖昧な葉クラスとの間の包含関係を、以下のように定義することにより、名詞の意味クラスの多義性を考慮している。すなわち、ある意味クラス c が曖昧性のある葉

2.1.2 下位範疇化フレーム

下位範疇化フレーム s は、動詞 v および、格 p と格要素の名詞に対する意味制約を表すクラス c の組合わせとして以下のように表現する。

$$s = \begin{bmatrix} \text{pred} : v \\ p_1 : c_1 \\ \vdots \\ p_l : c_l \end{bmatrix} \quad (2)$$

ただし、各意味クラス c_1, \dots, c_l は、シソーラスの任意のクラスであるとする。また、下位範疇化フレーム s のうちで、動詞 v に関する部分を s_v 、格 p および格要素の名詞クラス c に関する部分を s_p とする。

$$s = s_v \wedge s_p = [\text{pred} : v] \wedge \begin{bmatrix} p_1 : c_1 \\ \vdots \\ p_l : c_l \end{bmatrix} \quad (3)$$

2.1.3 下位範疇化フレームと共起の用例の包含関係

動詞・格要素の共起の用例 e が下位範疇化フレーム s に包含されるといふ関係 \subseteq_{sf} を以下のように定義する。

$$e \subseteq_{sf} s \iff \begin{array}{l} s \text{ の全ての格 } p \text{ について, } p \text{ が必} \\ \text{ず } e \text{ にもあり, } s \text{ の持つ } p \text{ の意味} \\ \text{クラス } c_s \text{ が, } e \text{ の持つ } p \text{ の意味ク} \\ \text{ラス } c_e \text{ を包含する } (c_e \subseteq c_s). \end{array}$$

関係 \subseteq_{sf} は、二つの下位範疇化フレーム間の包含関係としても適用可能である。

2.2 下位範疇化フレームからの動詞・格要素の共起の用例の生成

次に、格の依存関係および格要素の汎化レベルの曖昧性を考慮して、一つまたは複数の(部分的)下位範疇化フレームから、動詞・格要素の名詞の共起の用例を生成するモデルの基本的な考え方を述べる。まず、動詞・格要素の共起用例 e が以下のように与えられたとする。

$$e = \begin{bmatrix} \text{pred} : v \\ p_1 : c_{e1} \\ \vdots \\ p_k : c_{ek} \end{bmatrix}$$

このとき、動詞・格要素の共起用例 e を生成し得る下位範疇化フレーム s の条件として、 s が e を包含するといふ条件を仮定する。

$$e \subseteq_{sf} s$$

そして、下位範疇化フレーム s からの動詞・格要素の共起用例 e の生成を以下のように記述する。

$$s \rightarrow e \quad (4)$$

以下では、このような簡単な生成モデルを用いて、格の依存関係の曖昧性の問題、および格要素の名詞クラスの汎化レベルの曖昧性の問題について説明する。

2.2.1 格の依存関係の曖昧性

格の依存関係の曖昧性の問題は、コーパス中の共起データにおける複数の格の共起を観測したときに、これを、格フレーム中の必須格の共起に相当する強い依存関係ととらえるのか、あるいは、独立事象に相当する任意格的な格が偶然共起したととらえるのかという問題に相当する。

例えば、

クラス C_1 を包含することは、意味クラス c が、曖昧性のある葉クラス C_1 の(曖昧性のない)葉クラスのうちの少くともどれか一つを包含することと同等である。

例1 子供が公園でジュースを飲む。

という文を動詞・格要素の共起の用例 e として型付き素性構造で表現すると, c_c, c_p, c_j をそれぞれ “子供”, “公園”, “ジュース” に相当する葉クラスとして, 以下のように記述できる。

$$e = \begin{bmatrix} \text{pred: 飲む} \\ \text{が: } c_c \\ \text{を: } c_j \\ \text{で: } c_p \end{bmatrix} \quad (5)$$

ここで “子供”, “公園”, “ジュース” をそれぞれ包含するクラスとして, “人間”, “場所”, “飲料” に相当する $c_{hum}, c_{plc}, c_{bev}$ を考え,

$$c_c \leq c_{hum}, c_p \leq c_{plc}, c_j \leq c_{bev}$$

これらのクラスを格要素の意味制約として持つような下位範疇化フレームを考える。このとき, 「が」「を」「で」のそれぞれの格の間の依存関係としては, 様々なもの考えることができる。例えば, 「が」「を」「で」の三つの格が互いに依存関係にある下位範疇化フレームから共起の用例 e が生成されたと考えると, 以下のように書ける。

$$\begin{bmatrix} \text{pred: 飲む} \\ \text{が: } c_{hum} \\ \text{を: } c_{bev} \\ \text{で: } c_{plc} \end{bmatrix} \rightarrow e \quad (6)$$

また, 「が」「を」の二つの格は強い依存関係にあるが「で」格は任意格的な独立な関係にあると考える場合は, 以下のような部分下位範疇化フレームの組合わせから共起の用例 e が生成されたと書くことができる。

$$\begin{bmatrix} \text{pred: 飲む} \\ \text{が: } c_{hum} \\ \text{を: } c_{bev} \end{bmatrix} \rightarrow e \quad (7)$$

$$\begin{bmatrix} \text{pred: 飲む} \\ \text{で: } c_{plc} \end{bmatrix} \rightarrow e \quad (8)$$

さらに, 「が」「を」「に」の三つの格が互いに独立の関係にあると考える場合は, 以下のような部分下位範疇化フレームの組合わせから共起の用例 e が生成されたと書くことができる。

$$\begin{bmatrix} \text{pred: 飲む} \\ \text{が: } c_{hum} \end{bmatrix} \rightarrow e \quad (9)$$

$$\begin{bmatrix} \text{pred: 飲む} \\ \text{を: } c_{bev} \end{bmatrix} \rightarrow e \quad (10)$$

$$\begin{bmatrix} \text{pred: 飲む} \\ \text{で: } c_{plc} \end{bmatrix} \rightarrow e \quad (11)$$

2.2.2 格要素の名詞クラスの汎化レベルの曖昧性

格要素の汎化レベルの曖昧性の問題は, 格要素の名詞を観測したときに, これを概念階層のどのレベルの概念クラスの共起知識としてとらえるのかという問題に相当する。

例えば, 上の例1について, “人間”, “飲料” それぞれの上位クラスとして, “動物”, “液体” に相当する c_{ani}, c_{liq} を考え,

$$c_{hum} \leq c_{ani}, c_{bev} \leq c_{liq}$$

これらのクラスを格要素の意味制約として持ち得るような下位範疇化フレームを考える。ここでは, (7)式による動詞・格要素の共起用例の生成に注目し, 共起の用例 e の “子供” および “ジュース” を表す葉クラス c_c, c_j の汎化レベルとして, c_{hum}, c_{ani} および c_{bev}, c_{liq} のそれぞれ二通りずつが可能であるとする。すると, 共起の用例 e を生成し得る下位範疇化フレームの可能性としては, (7) および (12)~(14) 式の4通りを考えることができる。

$$\begin{bmatrix} \text{pred: 飲む} \\ \text{が: } c_{ani} \\ \text{を: } c_{bev} \end{bmatrix} \rightarrow e \quad (12)$$

$$\begin{bmatrix} \text{pred: 飲む} \\ \text{が: } c_{hum} \\ \text{を: } c_{liq} \end{bmatrix} \rightarrow e \quad (13)$$

$$\begin{bmatrix} \text{pred: 飲む} \\ \text{が: } c_{ani} \\ \text{を: } c_{liq} \end{bmatrix} \rightarrow e \quad (14)$$

2.3 格の依存関係に応じた生成モデルの定義

2.2.1節で述べたように, 格の依存関係の可能性としては, 様々なもの考えることができる。また, ある動詞・格要素の名詞の共起用例が与えられた場合にも, 格の依存関係の違いによって, これを生成可能な下位範疇化フレームが異なってくる。本節では, 格の依存関係の可能性として, いくつかの異なったものを考え, それらに応じて, 動詞・格要素の名詞の共起用例の生成モデルを定義する。

2.3.1 部分フレームモデル

まず, 与えられた動詞・格要素の名詞の共起用例 e における格の依存関係として特に何も仮定せず, 前節で述べた生成モデルと同じように, 動詞・格要素の共起用例 e を生成し得る下位範疇化フレーム s の条件として, s が e を包含するという条件だけを仮定する。

$$e \leq_s s$$

この場合, 下位範疇化フレーム s は, 共起用例 e の全ての格を持つ必要はなく, そのうちの一部を持っていればよい。ここでは, この生成モデルを部分フレームモデル (partial-frame model) と呼ぶ。

前節と同様に, この生成モデルにおいて, 下位範疇化フレーム s から共起用例 e を生成することを以下のように記述する。

$$s \rightarrow e \quad (15)$$

2.2.1節の (6)~(14) 式の例は, 全て部分フレームモデルの条件を満たしており, 部分フレームモデルにおける共起用例の生成の例になっている。

2.3.2 1フレームモデル

次に, 与えられた動詞・格要素の名詞の共起用例 e における格の依存関係として, e 中の全ての格が依存していると仮定し, 動詞・格要素の共起用例 e を生成し得る下位範疇化フレーム s の条件として, s が e の持つ格と同じ格だけを持ち, しかも s が e を包含するという条件を仮定する。

$$e = \begin{bmatrix} \text{pred: } v \\ p_1: c_1 \\ \vdots \\ p_k: c_k \end{bmatrix}, s = \begin{bmatrix} \text{pred: } v \\ p_1: c'_1 \\ \vdots \\ p_k: c'_k \end{bmatrix}, e \leq_s s \quad (16)$$

ここでは, この生成モデルを1フレームモデル (one-frame model) と呼ぶ。例えば, 2.2.1節の (5) 式の共起用例 e が与えられたとすると, (6) 式の生成の例が1フレームモデルの条件を満たしており, 1フレームモデルにおける共起用例の生成の例になっている。

2.3.3 独立格モデル

逆に, 与えられた動詞・格要素の名詞の共起用例 e における格の依存関係として, e 中の全ての格が独立であると仮定し, 動詞・格要素の共起用例 e を生成し得る下位範疇化フレーム s の条件として, s は e の持つ格のうちの一つだけを持ち, しかも s が e を包含するという条件を仮定する。

$$s = \begin{bmatrix} \text{pred: } v \\ p_i: c'_i \end{bmatrix} (1 \leq i \leq k), e \leq_s s$$

ここでは, この生成モデルを独立格モデル (independent-frame model) と呼ぶ。例えば, 2.2.1節の (5) 式の共起用例 e が与えられたとすると, (9)~(11) 式の生成の例が1フレームモデルの条件を満たしており, 1フレームモデルにおける共起用例の生成の例になっている。

2.3.4 独立フレームモデル

次に、実際に訓練集合中の共起用例の格の依存関係を統計的手法により検出し、この依存関係を反映した共起用例の生成モデルを定義することを考える。具体的には、まず、与えられた動詞・格要素の共起用例 e に対して、部分的な下位範疇化フレームというものを導入し、次に、統計的手法により互いに独立であると認定された部分的な下位範疇化フレームの組から共起用例 s を生成するというモデルを定義する。ここでは、このモデルを独立フレームモデル (independent-frame model) と呼ぶ。

部分下位範疇化フレーム

まず、与えられた共起用例 e に対して、2.3.2節の1フレームモデルの条件を満たす下位範疇化フレーム s を考える。すなわち、(16)式にあるように、 e と同じ格だけを持ち、しかも e を包含する下位範疇化フレーム s を考える。

次に、 s と同じ動詞 v を持ち、 s の持つ格のうちいくつかを持つものを s_i とし、これを部分下位範疇化フレームと呼ぶ。そして、 s を互いに素な部分下位範疇化フレームの組 $\{s_1, \dots, s_n\}$ に分割することを考える。各部分下位範疇化フレーム s_1, \dots, s_n は互いに素であるので、どの二つをとっても共通の格を持たず、しかもこれらの単一化 $s_1 \wedge \dots \wedge s_n$ は、 s に等しくなる。

$$s = s_1 \wedge \dots \wedge s_n \quad (17)$$

$$s_i = \begin{bmatrix} \text{pred: } v \\ \vdots \\ \text{pi}_{ij}: \text{ci}_{ij} \\ \vdots \end{bmatrix}, \quad \forall j \forall j' \quad \text{pi}_{ij} \neq \text{pi}_{i'j'} \quad (i, i' = 1, \dots, n, \quad i \neq i') \quad (18)$$

部分下位範疇化フレームの組の独立性

まず、以下のように、全ての部分下位範疇化フレーム s_1, \dots, s_n に包含される共起用例 e ($e \subseteq_{sf} s_1, \dots, s_n$) の条件付生起確率 $p(e|v)$ の和によって、部分下位範疇化フレームの条件付結合確率 $p(s_1, \dots, s_n | v)$ を推定する。²

$$p(s_1, \dots, s_n | v) \approx \sum_{e \subseteq_{sf} s_1, \dots, s_n} p(e | v) \quad (19)$$

これらの推定値を用いて、部分下位範疇化フレームの組の独立性に対して数学的に厳密な定義を与えると、以下のようなになる。

部分下位範疇化フレームの組 s_1, \dots, s_n が独立であるための必要十分条件は、 s_1, \dots, s_n のどの j 個の部分集合 s_{i_1}, \dots, s_{i_j} ($j = 2, \dots, n$) についても、以下の式が成り立つことである。

$$p(s_{i_1}, \dots, s_{i_j} | v) = p(s_{i_1} | v) \cdots p(s_{i_j} | v) \quad (20)$$

しかし、厳密にこの条件を満たすことを要求するのは厳しすぎるため、本論文では、独立性判定の閾値を表すパラメータ α ($0 \leq \alpha \leq 1$) を導入して、(20)式のかわりに以下の式が成り立つことが独立性の必要十分条件としている。

$$\alpha \leq \frac{p(s_{i_1}, \dots, s_{i_j} | v)}{p(s_{i_1} | v) \cdots p(s_{i_j} | v)} \leq \frac{1}{\alpha} \quad (21)$$

この条件においては、独立性の判定パラメータ α が 0 に近付けば独立性の判定条件が緩くなり、逆に 1 に近付けば条件が厳しくなる。

² 共起用例 e の条件付生起確率 $p(e | v)$ の推定値は、コーパスからの最尤推定値 $\text{freq}(e)/\text{freq}(v)$ を用いる。

独立な部分下位範疇化フレームの組からの生成

下位範疇化フレーム s から共起用例 e を生成する場合と同様に、独立な部分下位範疇化フレームの組 $\{s_1, \dots, s_n\}$ から共起用例 e を生成することを以下のように表す。

$$\{s_1, \dots, s_n\} \rightarrow e \quad (22)$$

例

いま、2.2.1節の(5)式の共起用例 e が与えられたとする。ここで、(6)式における共起用例 e の生成の場合のように、 e の全ての格が依存しているとする、独立フレームモデルにおいては、共起用例 e の生成を以下のように記述する。

$$\left\langle \begin{bmatrix} \text{pred: } \text{nomu} \\ \text{ga: } \text{Chum} \\ \text{wo: } \text{Cbev} \\ \text{de: } \text{Cpic} \end{bmatrix} \right\rangle \rightarrow e$$

また、(7)および(8)式における共起用例 e の生成の場合のように、「が」「を」の二つの格は強い依存関係にあるが、「で」格は任意格的な独立な関係にあると考える場合は、独立フレームモデルにおいては、共起用例 e の生成を以下のように記述する。

$$\left\langle \begin{bmatrix} \text{pred: } \text{nomu} \\ \text{ga: } \text{Chum} \\ \text{wo: } \text{Cbev} \end{bmatrix}, \begin{bmatrix} \text{pred: } \text{nomu} \\ \text{de: } \text{Cpic} \end{bmatrix} \right\rangle \rightarrow e$$

3 最大エントロピー法による確率モデル学習

本節では、最大エントロピー法による確率モデル学習 [Pietra95, Berger96] の形式的説明を簡単に挙げる。

3.1 最大エントロピー原理

ここでは、確率モデル学習の目的を、与えられた訓練集合から、文脈 $x \in \mathcal{X}$ において出力値 $y \in \mathcal{Y}$ を出力するプロセスの確率的振舞い、すなわち条件付確率分布 $p(y | x)$ を推定することとする。

まず、訓練集合中の事象 (event) (x, y) の観測値を大量に集め、訓練集合中の経験的確率分布 (empirical probabilistic distribution) $\tilde{p}(x, y)$ を以下のように推定する。

$$\tilde{p}(x, y) \equiv \frac{\text{freq}(x, y)}{\sum_{x, y} \text{freq}(x, y)}$$

ただし、 $\text{freq}(x, y)$ は事象 (x, y) の訓練集合中での生起頻度である。

次に、訓練集合中のどのような現象に注目して確率分布を推定するのかを表す二値の関数 $f(x, y)$ を導入し、これを素性関数 (feature function) と呼ぶ。具体的には、各素性関数 f_i について、この関数が真となる事象 x および y の集合 V_{xi} および V_{yi} が規定されていると考え、この集合にしたがって素性関数 f_i が以下のように定義される。

$$f_i(x, y) = \begin{cases} 1 & (x \in V_{xi} \text{ かつ } y \in V_{yi} \text{ の場合}) \\ 0 & (\text{それ以外の場合}) \end{cases}$$

また、一般に確率モデル学習の際には、大量の素性からなる素性の候補集合 \mathcal{F} から、活性化された素性の部分集合 $S \subseteq \mathcal{F}$ が選択され、これらによって事象 (x, y) および確率分布 $p(y | x)$ が記述される。

次に、実際に確率モデル学習を行う際には、活性化された素性集合 S 中の各素性 f_i について、学習すべき確率分布 $p(y | x)$ による素性 f_i の期待値 (左辺) と経験的確率分布 $\tilde{p}(x, y)$ による素性 f_i の期待値 (右辺) が等しいとする以下の制約等式 (constraint equation) を課す。

$$\sum_{x, y} \tilde{p}(x) p(y | x) f_i(x, y) = \sum_{x, y} \tilde{p}(x, y) f_i(x, y) \quad \text{for } \forall f_i \in S$$

そして、これらの制約等式を満たす確率分布 $p(y|x)$ のうちで、以下の条件付エントロピー $H(p)$ を最大にする最も「一般的な」モデルが、求めるべきモデル p_* であるとする。

$$H(p) \equiv - \sum_{x,y} \tilde{p}(x)p(y|x) \log p(y|x)$$

$$p_* = \arg \max_{p \in \mathcal{C}(S)} H(p) \quad (23)$$

3.2 パラメータ推定

(23)式を満たす確率分布は必ず存在し、それは以下の確率分布 $p_\lambda(y|x)$ で記述される。

$$p_\lambda(y|x) = \frac{\exp\left(\sum_i \lambda_i f_i(x,y)\right)}{\sum_y \exp\left(\sum_i \lambda_i f_i(x,y)\right)} \quad (24)$$

ただし、 λ_i は各素性 f_i のパラメータ (parameter) である。また、実際にエントロピーを最大にする最適なパラメータ λ_i^* を推定するには、Improved Iterative Scaling (IIS) アルゴリズム [Pietra95, Berger96] と呼ばれるアルゴリズムが用いられる。

3.3 素性選択

実際の確率モデル学習は、素性の候補集合 \mathcal{F} から、活性化された素性集合 S を選択することによって行われる。[Pietra95, Berger96] では、訓練集合に対して以下の対数尤度

$$L_{\mathcal{F}}(p) \equiv \sum_{x,y} \tilde{p}(x,y) \log p(y|x) \quad (25)$$

を計算し、対数尤度の増加分が最も大きくなる素性を一つずつ活性化された素性集合 S に追加するというアルゴリズムが用いられている。

4 最大エントロピー法による下位範疇化の確率モデル学習

次に、最大エントロピー法を下位範疇化の確率モデル学習に適用する手法を述べる。

4.1 事象

事象 (x,y) は、動詞・格要素の名詞の共起用例 e によって記述される。まず、下位範疇化フレームの場合と同様に、共起用例 e を動詞 v に関する部分 e_v と格 p および格要素の名詞クラス c に関する部分 e_p に分割する。

$$e = e_v \wedge e_p = \{ \text{pred} : v \} \wedge \begin{bmatrix} p_1 : c_1 \\ \vdots \\ p_k : c_k \end{bmatrix}$$

そして、事象 (x,y) のうちの文脈 x および出力値 y を、それぞれ、動詞 v および格 p および格要素に関する部分 e_p とし定義し、訓練集合中の事象を (v, e_p) と記述する。

$$x \equiv v$$

$$y \equiv e_p$$

4.2 素性

素性 f は、各 (部分) 下位範疇化フレームに対応して定義される。2.3.1 ~ 2.3.3 節の部分フレーム/1 フレーム/独立格モデルの場合は、各下位範疇化フレーム s に対して、素性関数 $f_s(v, e_p)$ が定義される。また、2.3.4 節の独立フレームモデルの場合は、(22) 式の生成モデルにおける各部分下位範疇化フレーム s_i に対して素性関数 $f_{s_i}(v, e_p)$ が定義される。各素性は、独立なパラメータ λ を持つので、この定義

により、各 (部分) 下位範疇化フレームが独立なパラメータを持つことになる。また、確率モデル学習の際には、訓練集合中の共起用例に対して、格の依存関係・格要素の名詞クラスの汎化レベルを考慮して、あらゆる可能な (部分) 下位範疇化フレームを作り、素性の候補集合 \mathcal{F} を構成する。

以下では、2.3 節で定義された各モデルに対して、素性の形式的定義を与える。

4.2.1 部分フレームモデル

各下位範疇化フレーム s に対して、素性関数 $f_s(v, e_p)$ は、共起用例 (v, e_p) が下位範疇化フレーム s に包含される場合に限り 1 を返すように定義される。

$$f_s(v, e_p) = \begin{cases} 1 & (e = (\text{pred} : v) \wedge e_p) \preceq_{sf} s \text{ の場合} \\ 0 & \text{(それ以外の場合)} \end{cases}$$

4.2.2 1 フレームモデル

各下位範疇化フレーム s に対して、素性関数 $f_s(v, e_p)$ は、共起用例 (v, e_p) が下位範疇化フレーム s と同じ格だけを持ち、しかも s に包含される場合に限り 1 を返すように定義される。

$$e = \begin{bmatrix} \text{pred} : v \\ p_1 : c_1 \\ \vdots \\ p_k : c_k \end{bmatrix}, \quad s = \begin{bmatrix} \text{pred} : v \\ p_1 : c'_1 \\ \vdots \\ p_k : c'_k \end{bmatrix}$$

$$f_s(v, e_p) = \begin{cases} 1 & (e = (\text{pred} : v) \wedge e_p) \preceq_{sf} s \text{ の場合} \\ 0 & \text{(それ以外の場合)} \end{cases}$$

4.2.3 独立格モデル

各素性は、格を一つだけ持つ下位範疇化フレームに対応している。各下位範疇化フレーム s に対して、素性関数 $f_s(v, e_p)$ は、共起用例 (v, e_p) の持つ格のうちの一つが下位範疇化フレーム s に存在し、しかも共起用例 (v, e_p) が s に包含される場合に限り 1 を返すように定義される。

$$e = \begin{bmatrix} \text{pred} : v \\ p_1 : c_1 \\ \vdots \\ p_k : c_k \end{bmatrix}, \quad s = \begin{bmatrix} \text{pred} : v \\ p_i : c'_i \end{bmatrix} \quad (1 \leq i \leq k)$$

$$f_s(v, e_p) = \begin{cases} 1 & (e = (\text{pred} : v) \wedge e_p) \preceq_{sf} s \text{ の場合} \\ 0 & \text{(それ以外の場合)} \end{cases}$$

4.2.4 独立フレームモデル

まず、訓練集合中の共起用例 e から、独立な部分下位範疇化フレームの組のうち、 e を生成可能な組が集められ、集合 $SF(e)$ が構成される。^{3 4}

$$SF(e) = \left\{ (s_1, \dots, s_n) \mid (s_1, \dots, s_n) \rightarrow e \right\}$$

そして、各下位範疇化フレーム s に対して、素性関数 $f_s(v, e_p)$ は、共起用例 $e = (v, e_p)$ を生成可能な部分下位範疇化フレームの組の集合 $SF(e)$ に、 s を含む組 (s_1, \dots, s_n) が少くとも一つ存在する場合に限り 1 を返すように定義される。

$$f_s(v, e_p) = \begin{cases} 1 & (\exists (s_1, \dots, s_n) \in SF(e = (\text{pred} : v) \wedge e_p)) \text{ の場合} \\ 0 & \text{(それ以外の場合)} \end{cases} \quad (26)$$

³厳密には、ある組 (s_1, \dots, s_n) が集合 $SF(e)$ に含まれるためには、 (s_1, \dots, s_n) のうちのどの部分下位範疇化フレームを二つ以上に分割してより細かな組 $(s'_1, \dots, s'_n, \dots, s'_{i+1})$ を作っても、これが独立性の条件 (21) 式を満たさないことが必要である。

⁴実際に、学習された確率モデルを訓練集合に含まれないテスト事象 e^{*0} に適用する際には、訓練集合から推定された部分下位範疇化フレームの確率値を用いて、部分下位範疇化フレームの組の独立性が判定され、集合 $SF(e^{*0})$ が構成される。

4.3 素性選択およびパラメータ推定

訓練事象の集合を \mathcal{E} 、素性の候補集合を \mathcal{F} とすると、3.3 節で述べた方法により活性化された素性の集合 S が構成され、3.2 節で述べたように、IIS アルゴリズムにより各素性のパラメータが推定される。そして、事象の条件付確率分布 $p_S(e_p | v)$ が以下のように推定される。

$$p_S(e_p | v) = \frac{\exp\left(\sum_{f_i \in S} \lambda_i f_i(v, e_p)\right)}{\sum_{e_p} \exp\left(\sum_{f_i \in S} \lambda_i f_i(v, e_p)\right)} \quad (27)$$

4.4 下位範疇化による統語解析結果の優先付け次に、確率モデル学習の終了後、活性化された素性の集合 S および確率モデル $p_S(e_p | v)$ を用いて、実際の入力文を構文解析した結果の構文木を優先付ける方法を述べる。

4.4.1 基本モデル

入力文を w とし、その統語解析結果の集合を $T(w)$ 、統語解析結果の一つを $t \in T(w)$ 、 t に含まれる動詞・格要素の共起用例の集合を $E(t)$ とする。このとき、統語解析結果 t の下位範疇化尤度 $\phi(t)$ は、各動詞・格要素の共起用例 $e \in E(t)$ の条件付確率 $p_S(e_p | v)$ の積で定義される。

$$\phi(t) \equiv \prod_{(v, e_p) \in E(t)} p_S(e_p | v) \quad (28)$$

入力文 w の統語解析結果 t としては、集合 $T(w)$ の中から、この下位範疇化尤度 $\phi(t)$ が最大のものを選ぶ。

$$\hat{t} = \arg \max_{t \in T(w)} \phi(t)$$

4.4.2 格被覆に関するヒューリスティックス

上で述べた基本モデルと併せて、ここでは、活性化された素性の集合 S が、実際に共起用例 e_p の全ての格に適用されているかどうかの情報を利用するヒューリスティックスを導入し、その効果を次節の実験において評価する。

まず、共起用例 (v, e_p) のどの格についても、対応する素性が素性集合 S に存在することを判定する格被覆 (case-covering) 関係 \preceq_{cv} を以下のように定義する。

$$(v, e_p) \preceq_{cv} S \iff (v, e_p) \text{ のどの格 } p \text{ (および } p \text{ でマークされる格要素の名詞の葉クラス } c_l) \text{ についても、素性集合 } S \text{ 中のいずれかの素性に対応する下位範疇化フレーム } s \text{ で、同じ格 } p \text{ を持ち、しかもその格の名詞クラス } c_s \text{ が葉クラス } c_l \text{ を包含する } (c_l \preceq c_s) \text{ のものが存在する。}$$

そして、この格被覆関係を用いて、以下の条件を満たす場合に、共起用例 (v_1, e_{p1}) が共起用例 (v_2, e_{p2}) に優先されるとみなす。

$$(v_1, e_{p1}) \preceq_{cv} S, (v_2, e_{p2}) \not\preceq_{cv} S$$

また、このヒューリスティックスを用いて、以下のような手順により、入力文 w の統語解析結果の集合 $T(w)$ の要素の優先付けを行う。(i) 活性化された素性集合 S に対して、格被覆関係を満たす共起用例の数が多いものを優先する。(ii) (i) の条件で差がつかない場合は、活性化された素性集合 S に対して、格被覆関係を満たす共起用例の条件付確率の積の値が大きいものを優先する。(iii) (i), (ii) の条件でも差がつかない場合は、活性化された素性集合 S に対して、格被覆関係を満たさない共起用例の条件付確率の積の値が大きいものを優先する。

表 1: 動詞「買う」の場合の選択された素性の例 (独立フレームモデル, $\alpha = 0.9$)

最初の 10 回で選択された素性			
順位	素性	名詞クラス/名詞の例	事象数
1	を:1404	切符, 証券	22
2	を:1524	土地	16
3	を:1553	株	23
4	を:14	生産物	158
5	を:1196	通貨, 単位	32
6	を:1301	怒り	9
7	を:1151	反発	11
8	を:1462	電子機器	9
9	を:1451	入れ物, 容器	2
10	を:1302	反感	8

二つ以上の格に対応する素性のうち、最も早く選択されたもの(上位 5 個)

順位	素性	名詞クラス/名詞の例	事象数
30	が:1259, を:13	が:国, を:国債	2
53	に:1200, を:14	に:私, を:生産物	1
54	に:121, を:145	に:人間, を:製品	1
61	に:12, を:1404	に:人間, を:切符	1
62	に:1205, を:140	に:子供, を:みやげ	1

5 実験および評価

5.1 コーパスおよびシソーラス

EDR 日本電子化辞書 [EDR95] の日本語共起辞書から、出現頻度 50 回以上の動詞 835 個について動詞・格要素の共起を合計 153,014 個抽出し、これを用いて下位範疇化優先度学習の実験を行った。EDR 日本語共起辞書の動詞・格要素の共起 153,014 個には、格助詞・係助詞相当語句が約 270 語含まれている。また、シソーラスとしては、「分類語彙表」[国研 93] を用いた。分類語彙表は 6 層の階層構造から構成され、階層構造の葉の部分 (第 6 層目) に総数約 60,000 語の単語 (葉クラス) が分類されている。また、その体の類 (名詞シソーラス) の部分には、総数約 45,000 語の名詞 (葉クラス) が分類されている。

5.2 下位範疇化の確率モデル学習の実験

以下の条件のもとで下位範疇化の確率モデル学習の実験を行った。すなわち、i) 名詞クラスを「分類語彙表」の 4 レベル以上 (桁数 4 桁以下) に限定した。また、ii) 本論文の確率モデルにおいては、動詞間の独立性を仮定しているので、訓練集合を動詞ごとに構成し、動詞ごとに独立して確率モデル学習を行った。⁵

各動詞について、訓練集合の事象数は 200~500 であった。また、素性の候補集合のサイズは、格独立モデルで 200~400、1 フレームモデル/独立フレームモデル (独立性判定パラメータ $\alpha = 0.5/0.9$) で 500~1,300、部分フレームモデルで 650~1,550 であった。

選択された素性の例

動詞「買う」の場合について、独立フレームモデル (独立性判定パラメータ $\alpha = 0.9$) の確率モデル学習によって実際に選択された素性の例を表 1 に示す。表では、まず、素性選択の過程において最初の 10 回で選択された素性を示す。ここで、EDR 日本語共起辞書から構成した訓練集合

⁵[白井 97] では、名詞クラスを用いず、格の情報だけを用いて、最大エントロピー法により格の依存関係を学習し、統語解析における性能向上に利用している。ここでは、一つの確率モデルの中に全ての動詞の情報を記述し、全ての動詞の事例を含む一つの訓練集合を用いて確率モデル学習を行っている。

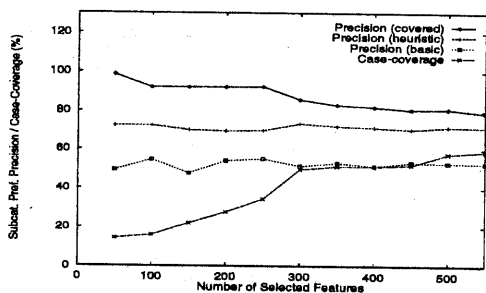


図1: 独立フレームモデル ($\alpha = 0.9$) における下位範疇化の正解率 r_b, r_h, r_c およびテスト事象の格被覆率の比較

においては、一つの格しか持たない共起用例が約75%も含まれていたため、上位10個の素性は、いずれも格一つの部分下位範疇化フレームに相当するものであった。そこで、二つ以上の格を持つ部分下位範疇化フレームに相当する素性のうちで、最も早く選択されたもの上位5個を、これとは別に示す。

表中では、各素性は、格および「分類語彙表」の名詞クラスコードのペアで表現されている。また、素性のほかに、名詞クラスの説明と名詞の例、および訓練集合中でその素性関数が真となる事象数を示す。

5.3 下位範疇化による優先付けの性能の評価

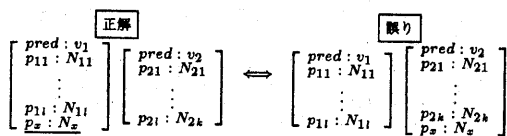
5.3.1 評価方法

学習されたパラメータが、統語解析における優先付けでどの程度の性能を示すかを評価するために、以下のような仮想的な実験を行った。

いま、 N, N_x が格要素の名詞、 p, p_x が格助詞、 v_1, v_2 が動詞であるとして、以下のような列からなる埋め込み文もしくは複文を考え、後置詞句 N_x-p_x が、 v_1 もしくは v_2 の格要素になり得るという曖昧性を想定する。

$$N_x-p_x \dots N-p-v_1-N-p \dots N-p-v_2$$

ここで、後置詞句 N_x-p_x が、 v_1 の格要素になる場合を正解とし、 v_2 の格要素になる場合を誤りとして、この二通りの場合の下位範疇化の確率値を計算し、どの程度の正解率が達成できるかを調べた。



実際の評価用の共起用例の収集は、 v_1 および v_2 のテスト集合から正解共起用例を一つずつ選んで、これから誤り共起用例を作成するという手順で行った。「買う」「飲む」「重ねる」の三つの動詞の共起用例をテスト集合から抽出し、 v_1, v_2 の組み合わせとして三通りの場合について評価用の共起用例を作成した。

また、下位範疇化の性能の評価は、(i) 4.4.1節の基本モデルによる正解率 r_b 、(ii) 4.4.2節の格被覆に関するヒューリスティックスを用いた場合の正解率 r_h 、(iii) 活性化された素性集合 S との間で格被覆関係を満たす共起用例を対象とした場合の正解率 r_c 、の3種類の正解率を測定して行った。

5.3.2 実験結果

まず、独立フレームモデル (独立性判定パラメータ $\alpha = 0.9$) の場合において、素性選択の過程における3種類の正解率 r_b, r_h, r_c の変化の様子を図1に示す。また、活性化された素性集合 S との間で格被覆関係を満たす共起用例がテスト事象集合に含まれている割合を測定し、この変化の様子が図中に示す。

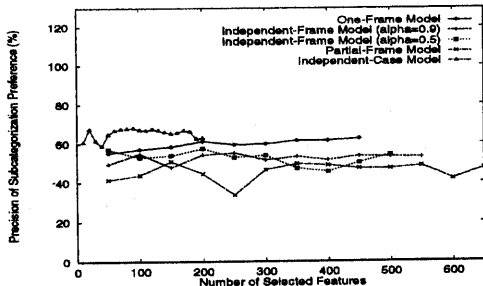
図から分かるように、下位範疇化の正解率は r_c, r_h, r_b の順に低くなっている。これは、活性化された素性集合 S に適当な素性が含まれていて、与えられた共起用例に対して格被覆関係を満たす場合には、下位範疇化タスクにおいて学習された確率モデルが十分高い性能を示すことを表している。しかし、与えられた共起用例が格被覆関係を満たさない場合で、4.4.2節で述べたヒューリスティックスを用いた場合には、下位範疇化タスクにおける性能が低下する。また、4.4.2節で述べたヒューリスティックスを用いず、4.4.1節の基本モデルにしたがう場合は、下位範疇化タスクにおける性能がさらに低下する。

次に、素性選択の過程における3種類の正解率 r_b, r_h, r_c およびテスト事象の格被覆率の変化の様子を、1フレーム/独立フレーム/部分フレーム/独立格モデルの間で比較したものを図2に示す。図中の (b)(d) のグラフから、格被覆に関するヒューリスティックスを用いた場合、および格被覆を満たす場合のいずれにおいても、他のモデルに比べて独立フレームモデルが高い性能を示すことが分かる。また、独立フレームモデルにおける独立性判定パラメータ α の違いに注目すると、独立性の条件が厳しい方が、 r_h, r_c の二つの正解率で測定した下位範疇化タスクの性能が高いことが分かる。ただし、独立フレーム/1フレームモデルのように、格の独立性に条件を課す、または格の独立性を認めないモデルにおいては、図中の (c) のグラフのように、テスト事象の格被覆率が低くなっている。これらの結果から、格の独立性の条件が厳しいほど、訓練集合に特化 (overfit) した学習が行われるため、テスト事象への適用率は悪くなるが、下位範疇化タスクにおいては、格の独立性を厳しく判定した方が性能が上がるというトレードオフの関係があることがわかる。

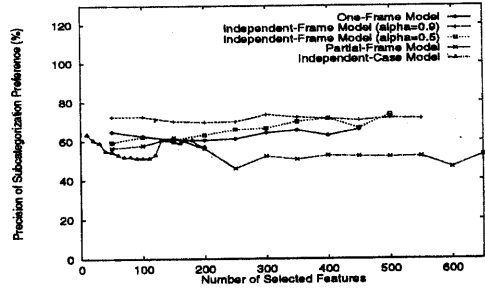
6 おわりに

最大エントロピー法により、下位範疇化の確率モデル学習を行う手法を提案した。実験の結果、下位範疇化タスクにおける性能を測定した結果、部分下位範疇化フレームの組の独立性を厳しく判定するモデルが最も性能が高いことが分かった。今後は、下位範疇化タスクの性能を落とすことなく、テスト事象への適用率を確率モデル学習をいかにして実現するかが大きな課題である。3.3節で述べた素性選択の方法 (すなわち、[Pietra95, Berger96] の方法) では、訓練集合の対数尤度の増加分が最大となる素性の一つずつ活性化させるという単純な手順をとっているが、下位範疇化の確率モデル学習では、このような方法は十分ではないといえる。今後は、選択された素性の適用率も考慮にいれたきめ細かな素性選択の手法を考える必要がある。⁶ また、今

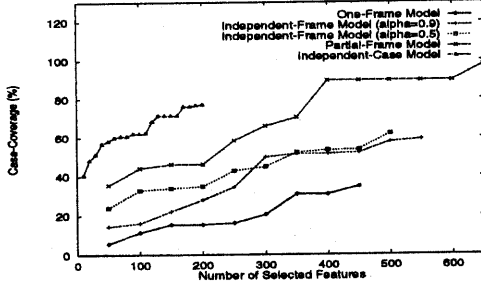
⁶我々は、確率モデルとして Bayesian Network を用いた下位範疇化の確率モデル学習 [宮田 97] についても研究を行っている。Bayesian Network を用いる手法においては、訓練集合中の全ての事象の全ての格の情報が必ず確率モデルに反映されるが、本論文の最大エントロピー法による手法においても、この条件を満たす素性選択アルゴリズムを適用することは可能である。その際には、MDL 原理 [Rissanen89] のような



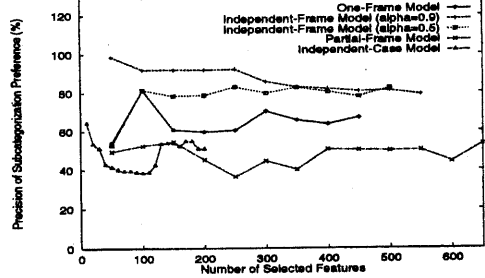
(a) 基本モデルの正解率



(b) 格被覆に関するヒューリスティクスを用いた場合の正解率



(c) テスト事象の格被覆率



(d) 格被覆関係を満たす事象の正解率

図 2: 素性選択の過程における下位範疇化の正解率およびテスト事象の格被覆率の変化

後、「分類語彙表」の葉クラスを含む詳細なクラス情報を用いると、素性の候補集合のサイズが大きくなること予想されるが、そのような場合でも、選択された素性の適用率も考慮にいれたきめ細かな素性選択の手法を適用すれば、下位範疇化の性能および素性の適用率の双方に関して最適な素性集合が選択できると考えられる。⁷

謝辞: 最大エントロピー法に関して、多くの貴重な情報を提供して頂いた東京工業大学の乾健太郎氏、白井清昭氏に感謝致します。

参考文献

- [Berger96] Berger, A. L., Pietra, S. A. D. and Pietra, V. J. D.: A Maximum Entropy Approach to Natural Language Processing, *Computational Linguistics*, Vol. 22, No. 1, pp. 39-71 (1996).
- [Black93] Black, E.: Towards History-based Grammars: Using Richer Models for Probabilistic Parsing, *Proceedings of the 31st Annual Meeting of ACL*, pp. 31-37 (1993).
- [Collins96] Collins, M.: A New Statistical Parser Based on Bigram Lexical Dependencies, *Proceedings of the 34th Annual Meeting of ACL*, pp. 184-191 (1996).
- [EDR95] 日本電子化辞書研究所, EDR 電子化辞書仕様説明書 (1995).
- [乾 96] 乾健太郎, 白井清昭, 徳永健伸, 田中穂積: 種々の制約を統合した統計的日本語文解析, 情報処理学会研究報告, Vol. 96, No. 114 (96-NL-116), pp. 35-42 (1996).
- [国研 93] 国立国語研究所: 分類語彙表, 秀英出版 (1964,1993).

情報量基準によってモデルの評価を行うことも可能であると考えられる。⁷我々は、確率モデル学習に基づかない下位範疇化優先度学習手法 [Utsuro97] において、「分類語彙表」の葉クラスを含む詳細なクラス情報を用いて学習を行い、本論文と同様の下位範疇化タスクに適用した結果、最高で 96% という高い性能を得ている。[Utsuro97] の手法と比較すると、本論文の最大エントロピー法に基づく手法の利点として、全く同じ枠組みの中で、格および格要素以外の様々な情報を素性として記述できることが挙げられる。そのような情報としては、例えば、動詞の態・アスペクト・慣用句などの統語的屬性がある。

[Li95] Li, H. and Abe, N.: Generalizing Case Frames Using a Thesaurus and the MDL Principle, *Proceedings of International Conference on Recent Advances in Natural Language Processing*, pp. 239-248 (1995).

[李 96a] 李航: 心理言語学原理に基づいた確率的曖昧性解消法, コンピュータソフトウェア, Vol. 13, No. 6, pp. 489-501 (1996).

[Li96b] Li, H. and Abe, N.: Learning Dependencies between Case Frame Slots, *Proceedings of the 16th COLING*, pp. 10-15 (1996).

[Magerman95] Magerman, D. M.: Statistical Decision-Tree Models for Parsing, *Proceedings of the 33rd Annual Meeting of ACL*, pp. 276-283 (1995).

[宮田 97] 宮田高志, 津津呂武仁, 松本裕治: Bayesian Network による下位範疇化の確率モデルおよびその学習, 情報処理学会研究報告, Vol. 97, No. (97-NL-119) (1997).

[Pietra95] Pietra, S. D., Pietra, V. D. and Lafferty, J.: Inducing Features of Random Fields, CMU Technical Report CMU-CS-95-144, School of Computer Science, Carnegie Mellon University (1995).

[Resnik93] Resnik, P.: Semantic Classes and Syntactic Ambiguity, *Proceedings of the Human Language Technology Workshop*, pp. 278-283 (1993).

[Rissanen89] Rissanen, J.: *Stochastic Complexity in Statistical Inquiry*, Vol. 15 of *Series in Computer Science*, World Scientific Publishing Company (1989).

[白井 97] 白井清昭, 乾健太郎, 徳永健伸, 田中穂積: 最大エントロピー法による格の従属関係の学習, 言語処理学会第 3 回年次大会論文集, pp. 337-340, 言語処理学会 (1997).

[Utsuro97] Utsuro, T. and Matsumoto, Y.: Learning Probabilistic Subcategorization Preference by Identifying Case Dependencies and Optimal Noun Class Generalization Level, *Proceedings of the 5th ANLP*, pp. 364-371 (1997).