

ノンパラメトリックベイズによる意味役割推定

松尾 健太 石原 靖弘 竹内 孔一

岡山大学大学院自然科学研究科

{kmatsuo,ishihara,koichi}@cl.cs.okayama-u.ac.jp

1 はじめに

近年、インターネット上でさまざまな情報が交換されており情報交換の手段は画像、音声、言語などさまざまである。言語による情報交換では記述の仕方は情報発信者によってさまざまで同じ意味を表すにも同一の文となるとは限らない。それらの文を構造化して扱うことができれば、自動翻訳、文章要約などさまざまな言語処理に応用することができる。構造化の方法として文の構造を述語語義と意味役割として捉えるというものがある [1]。これによりそれぞれの文を抽象化して扱うことができ、それらを自動で判定する研究が盛んに行われている。述語語義と意味役割の推定にはルールによるもの [2] やグラフ分割によるもの [3] など、さまざまであるが本研究ではノンパラメトリックベイズによる述語語義と意味役割の推定を扱う。従来、英語文やドイツ語文におけるノンパラメトリックベイズによる意味役割推定は行われているが [4][5]、調査した限りでは日本語文に対してノンパラメトリックベイズによる推定は行われていない。そこで本研究では日本語文を対象としたノンパラメトリックベイズによる意味役割推定について述べる。ノンパラメトリックベイズではデータに応じて扱うモデル自体の複雑さを学習することができ、いくつかのパラメータで記述される分布ではなく柔軟な分布を考えることができる。英語では文内での主語との位置関係が意味役割を決める有力な情報となるが、日本語では項が格をとり主語や述語に対して自由な位置を取ることから、主語との位置関係を意味役割推定に用いることができない。このような日本語の特徴を踏まえたうえで述語との係り受け関係を依存木として表したモデルを構築し、意味役割推定を行った。さらに半教師あり学習も行った。以下の章ではまず 2 章で手法と依存木によるモデルについて述べる。3 章で実験の結果を示し、4 章で結果の考察、5 章でまとめを述べる。

2 提案手法

本節ではまず、依存木について述べ、例文を用いて依存木の各要素について説明する。次に、依存木のモデルとサンプリングする変数について述べる。最後にサンプリング時に学習データを用いる方法を説明する。

2.1 依存木

項の意味役割は係り先の述語の態や項の品詞、格などにより決定される。このことから述語との係り関係を表す依存木をモデルとした。依存木は本来、項や述語の表層により表現されるが、本研究では項の表層部分を項の意味役割、述語の表層部分を述語の語義としたモデルを用いる。依存木の性質は [6] を参考にし、文の中の依存関係が満たす性質 (公理) として挙げられている 4 つの性質のうち 1 と 2 を満たす依存木を構築した。

1. 1 つの要素のみが独立している
2. 他の要素はすべて何らかの要素に直接依存する
3. どの要素も 2 つ以上の要素に直接依存しない
4. もし A が B に直接依存し、要素 C が線形順序をなす連鎖において両者の間に位置する場合、C は A または B または A と B の間に位置する他の要素に直接依存する

性質 1 の独立した要素とは述語の語義であり、他の要素は述語と係り関係にある項の意味役割や項の意味役割を決定する要素となる品詞、格情報などとなる。以下に「私は本を買った」を例文とした依存木の図を載せる。この文では「購入」が述語「買った」の語義、「動作主」と「対象」は「購入」と係り関係にある項の意味役割である。「動作主」と「対象」は「購入」に依存している。品詞「名詞」と格「ハ格」が「動作主」に、品詞「名詞」と格「ヲ格」が「対象」に依存している。また、述語の態「past」と品詞「動詞」は述語の語義「購入」に依存している。このようなモデルをノンパラメトリックベイズの言語モデルとして用いる。

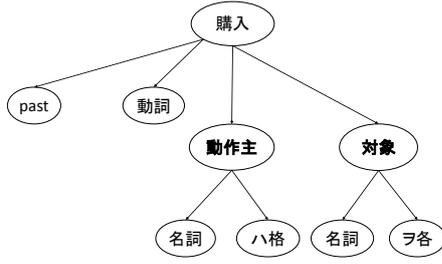


図1 依存木のモデル

2.2 ノンパラメトリックベイズ

ここでは依存木をノンパラメトリックベイズのモデルとした場合、どのようにモデル化されるかについて述べる。図2を見てもらいたい。依存木はノード T_t, T'_t

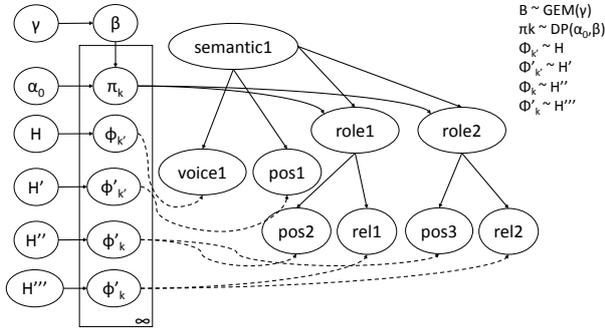


図2 ノンパラメトリックベイズの依存木モデル

を持ち、 t と t' はそれぞれノードのクラス semantic(語義) とクラス role(意味役割) を表す。semantic はデータ pos_t (品詞), $voice_t$ (態) を role はデータ pos'_t (品詞), rel'_t (格) を持つ。ここで T_t, T'_t の確率 $p(T_t)$ と $p(T'_t)$ は次のように定義される。

$$p(T_t) = p(pos_t | semantic_t) p(voice_t | semantic_t) p(semantic_t)$$

$$p(T'_t) = p(pos'_t | role'_t) p(rel'_t | role'_t) p(semantic_t | role'_t) p(role'_t)$$

ここで、role は semantic に依存して確率分布が表されるので HDP[8] を用いる。HDP は確率基底測度 G_0 を共有した複数の DP であらわされ、 G_0 も DP によって導出される。 $G_k \sim DP(\alpha_0, G_0)$ また、 G_0 は $G_0 \sim DP(\gamma, H)$ で導出される。ここで stick-breaking construction[7] から HDP は次のように解釈される。

$$G_0 = \sum_{k'=1}^{\infty} \beta_{k'} \delta_{\phi_{k'}}$$

$$G_k = \sum_{k'=1}^{\infty} \pi_{kk'} \delta_{\phi_{k'}}$$

$$\beta \sim GEM(\gamma)$$

$$\pi_k \sim DP(\alpha_0, \beta)$$

$$\phi_{k'} \sim H$$

本手法では role は確率基底測度を共有した複数の DP から生成されるとし、semantic の各 G'_k を $\pi_{k'}, \phi_{k'}, \phi'_{k'}$ で、role の各 G_k を π_k, ϕ_k, ϕ'_k で表す。 $\pi_{k'}$ は semantic のクラス k' の確率分布、 $\phi_{k'}$ と $\phi'_{k'}$ は各データの確率分布のパラメータを表す。 π_k はクラス k の semantic に依存する role の確率分布、 ϕ_k と ϕ'_k は role に対するデータの確率分布のパラメータを表す。従ってこのモデルは以下のように定義される。

$$\beta | \gamma \sim GEM(\gamma),$$

$$\pi_k | \alpha_0, \beta \sim DP(\alpha_0, \beta),$$

$$\phi_{k'} \sim H, \phi'_{k'} \sim H', \phi_k \sim H'', \phi'_k \sim H''',$$

$$role_{t'} | semantic_t \sim Multinomial(\pi_{semantic_t}),$$

$$word_t | semantic_t \sim F(\phi_{semantic_t}),$$

$$pos_t | semantic_t \sim F'(\phi'_{semantic_t}),$$

$$pos_t | role_t \sim F''(\phi_{role_t}), rel_t | role_t \sim F'''(\phi'_{role_t}).$$

これはパラメータ ($\phi_{semantic_t}, \phi'_{semantic_t}, \phi_{role_t}, \phi'_{role_t}$) に従う分布があり、各データは対応するパラメータによって表される分布から条件付き独立で生成されることを表している。semantic のクラスを表す $semantic_t$ は他のクラスに依存しないのでパラメータ α_0 、基底測度 γ に従うディリクレ過程に従って生成される。

2.3 推定

推定では、データが与えられたときの事後確率を最大にするような semantic, role のクラスの集合を求める。しかし、クラス数が無限にあるのですべてのクラスを考慮して事後確率を求めることができない。そこでギブスサンプリング[8]を用いて各データに対して順次サンプリングを行い近似的に事後確率を求める。サンプリングでは z, m, β をサンプリングする。詳細は[8]を見てもらいたい。

Sampling z :

z_t, z'_t はデータ t に対しての semantic, role それぞれのクラスを表す。これらは次のように動的に求める。

$$p(semantic_t | voice_t, pos_t) \sim$$

$$p(voice_t | semantic_t) p(pos_t | semantic_t) p(semantic_t)$$

$$p(role_t | pos_t, rel_t, semantic'_t) \sim$$

$$p(pos_t | role_t) p(rel_t | role_t) p(role_t)$$

ここでデータの事後確率は $p(x_t | z_t) = \frac{n_{x_t k} + \alpha_0}{n_{x_t} + N \alpha_0}$ となる。 x_t は t と対応する word, pos, rel を、 z_t は t と対応する role または semantic を表す。 $n_{x_t k}$ はクラス k のデータ x の総数であり、 N はデータ x の総数である。各データごとのサンプリングでは現在出現しているクラスの K 個と未知のクラスを表すクラスについて確率

分布を算出し、データのクラスをサンプルする。未知のクラスを表すクラスはまだ出現していないクラスすべてを表している。サンプルされたものが未知のクラスであった場合、新しいクラスとし K に加え、新しく未知のクラスを定義したのち後述する β を計算し直す。

Sampling m :

ここでは変数 \mathbf{m} を導入する。 \mathbf{m} は各 semantic に依存している role のクラスごとの数を表し、 m_{ij} は semantic のクラス i に依存している role のクラス j の数を表す。 \mathbf{m} の条件付き確率は $p(m_{ij} = m | z, \beta, \alpha_0) \propto S(n_{ij}, m)(\alpha_0 \beta_j)^m$ と表せる。ここで、 $S(n, m)$ は第一種スターリング数である*1。

Sampling β :

パラメータ β は role のクラスそれぞれの確率を表す。 β は \mathbf{m} を用いて表すことができ、次に示すディリクレ分布を用いてサンプリングできる。 $(\beta_1, \dots, \beta_K, \sum_{k'=K+1}^{\infty} \beta_{k'}) \sim Dirichlet(m_{.1}, \dots, m_{.K}, \gamma)$, $m_{.k} = \sum_{k'=1}^K m_{k'.k}$ 。ここで、 $m_{.k}$ はすべての semantic のクラスにおいての role のクラス k の数を表す。

2.4 半教師あり学習

ノンパラメトリックベイズではサンプリング時は各クラスがどの意味役割を表しているかわからない。このため学習データを用いるには学習データ及びサンプリングの一部に工夫が必要になる。まず、あらかじめ学習データに意味役割ごとにクラスを割り当てておく。例えば「対象」は 1, 「目的」は 2, というように同じ意味役割には同じクラスを割り当てる。これらのデータを学習データとし、クラスが割り当てられていないデータ(テストデータ)とともにサンプリングを行う。クラスがあらかじめ割り当ててあるデータはクラスのサンプリングを行わず、テストデータにのみサンプリングを行う。学習データはサンプリングの際 n_{xk} , n_{ij} , などに影響を与える。このように学習データを利用し、各サンプリングに用いる。

3 実験

実験では 2662 文に対して係り受け解析まで自動で行い、人手で意味役割と語義を付与したデータを用いた*2。比較は CRF と行った。CRF では項の連続列に

*1 $S(0, 0) = S(1, 1) = 1$, $s(n, 0) = 0$ for $n > 0$, $s(n, m) = 0$ for $m > n$, and $S(n + 1, m) = s(n, m - 1) + nS(n, m)$ for others

*2 koichi et al. Annotating Semantic Role Information to Japanese Balanced Corpus. (to appear)

対して項の素性や項と前後の項との素性を特徴に用いて学習し意味役割を付与しており、本手法とは条件が異なるが目安として載せている。

3.1 実験方法

半教師あり学習 (sIDT:semi-supervised learning for Infinite Dependent Tree) では、3.4 で述べた方法を用い、学習データ 1331 文はサンプリングせずテストデータ 1331 文にサンプリングを行う。教師なし学習 (uIDT:unsupervised learning for Infinite dependent model) ではテストデータにサンプリングを行う。サンプリング時のパラメータ α_0 , α'_0 , γ , γ' は同じでそれぞれ 1.0, 1.0, 0.6, 0.6 と設定し、サンプリングは 500 回行った。サンプリング後の各クラスにどの意味役割を付与するかはサンプリングのみでは決めることができないので、各クラスごとに正解データと比較し、クラス内にもっとも多かった意味役割をそのクラスの意味役割とした。CRF では意味役割項の表層、読み、基本形、品詞と述語の表層、述語の基本形、助詞とそれらの前後の組み合わせを特徴としている。

3.2 実験結果

以下の表に各手法の意味役割の precision, recall, f 値を示す。

表1 各手法の precision, recall, f 値

手法	precision(%)	recall(%)	f 値 (%)
uIDT	34.61	34.61	34.61
sIDT	42.76	42.76	42.76
CRF	48.77	37.26	42.25

uIDT および sIDT の precision, recall はクラスを割り当てるため、f 値と同じ値になっている。sIDT がわずかに CRF よりも高い精度を出している。

4 考察

ここではサンプリングされたクラス、クラスの割り当て方、サンプリング回数について述べる。まずサンプリングされたクラスについて述べる。各クラスの数「対象」が最も多く 64, ついで「動作主」が 25, 「原因」が 9, 「連語副詞(仮)」が 8 と続いている。「対象」のクラスには「対象(身体部分)」や「対象(生成物)」などが含まれており、「身体部分」や「生成物」などの副分類ではなければ識別できていると考えた。また、「動作主」, 「経験者」は同じクラスに分類されることが多く、格が「は」で品詞が「名詞-固有名詞-人名」である項がクラスを占めている。特徴的なクラスもいくつか存在

し、例えば「原因」のクラスの一つでは含まれているのはすべて「原因」の項で格が「ので」の項のみであった。他には「起点」のクラスでは格が「から」で品詞が「名詞-普通名詞-一般」など、決定的な特徴により分類が行われていた。上で示したのはクラスの数であり意味役割自体の数ではない。「連語副詞(仮)」は8クラスあるが、そのうちのいくつかのクラスは複数種類の意味役割項が含まれており、たまたま「連語副詞(仮)」の数だけ他の意味役割より1多いことで「連語副詞(仮)」と割り当てられたというクラスがある。このようなクラスは本研究で用いた特徴では分けることができない曖昧な項が集まったクラスを表しており、識別することができていない項のクラスであると言える。次にクラスの割り当て方について述べる。提案手法ではサンプリング結果と正解データを照らし合わせて意味役割に割り当てている。しかし、実際には正解がわからないため、どのようにクラスに意味役割を割り当てるかが問題になる。半教師あり学習であれば、クラスを固定した学習データを元に、意味役割を割り当てることができる。また、半教師あり学習にも言えるが、教師なし学習でも未知のデータの中に正解データを加えてサンプリングを行うことで、正解データから各クラスに意味役割を割り当てることができる。ただし、サンプリング後のクラスに学習データが含まれていなければこの方法でも割り当てることができないので、実際に未知のデータに対して意味役割を推定する場合、この問題を考える必要がある。次にサンプリング回数について述べる。ギブスサンプリングでは無限回のサンプリングを行うことを仮定している。しかし実際には無限回のサンプリングはできないので、ある程度の回数を設定する。なのでサンプリングがどのくらい必要かが問題になる。テストデータの初期値に正解クラスを設定して、sIDTと同じ条件でサンプリングした結果、精度は49.09%であった。このことから、提案手法のモデルと学習データに対して実験で設定したサンプリング回数は少ないと考えた。ただし、これは精度が49.09%となるクラスの状態が精度42.25%となるクラスの状態よりも安定している必要がある。さらにこの二つの状態の間に精度が落ち込む部分があった場合、これをサンプリングによって越える必要がある。

5 おわりに

本研究では依存木によって述語と意味役割項の係り関係をモデル化し、ノンパラメトリックベイズによる

意味役割推定を行った。実験では2662文を用いて学習を行い、CRFと同等の精度で意味役割の推定を行うことができた。クラスへの意味役割の割り当て方などが今後の課題である。

謝辞

本研究の遂行にあたってJSPS科研費(26370485)の助成を受けた。

参考文献

- [1] Jane B. Grimshaw. Argument Structure (Linguistic Inquiry Monographs). The Mit Press. 1990.
- [2] 池田 吉優, 竹内 孔一. 意味役割と述語の概念を付与するシステムの構築. 信学技報. Vol. 114, NLC2014-39, pp. 55-60, 2014.
- [3] Joel Lang and Mirella Lapata. Similarity-Driven Semantic Role Induction via Fraph Partitioning. Association for Computational Linguistics. Vol. 40, pp. 633-669, 2014.
- [4] Titov, Ivan and Klementiev, Alexandre. Crosslingual Induction of Semantic Roles. Association for Computational Linguistics. pp. 647-656, 2012.
- [5] Ivan Titov and et al. Semi-Supervised Semantic Role Labeling. Approaching from an Unsupervised Perspective. 2009.
- [6] 麻生 英樹. 句構造と依存構造について. The 26th Annual Conference of the Japanese Society for Artificial Intelligence. 2012.
- [7] Jayaram Sethuraman. A CONSTRUCTIVE DEFINITION OF DIRICHLET PRIORS. Statistica Sinica. pp. 639-650, 1994.
- [8] Yee Whye Teh, Michael I Jordan, Mathew J Beal and David M Blei. Hierarchical Dirichlet Processes. American Statistical Association. Vol. 101, pp. 1566-1581, 2012.
- [9] Tamura, Akihiro and Watanabe, Taro and Sumita, Eiichiro and Takamura, Hiroya and Okumura, Manabu. Part-of-Speedch Induction in Dependency Trees for Staticistical Machine Translation. Association for Computational Linguistics. pp. 841-851, 2012.