

ベイズの定理を利用した日本語意味役割付与モデルの構築 Building a Japanese Semantic Role Labeling Using Bayesian Inference

岸本 廉†
Ren Kishimoto

竹内 孔一†
Koichi Takeuchi

1. はじめに

本論文では日本語の入力文に対して意味役割付与を行うシステムをベイズの定理を利用しモデル化する研究を行う。意味役割付与とは 1 文内の各係り受け関係に対して意味的な分類を与え、述語に対して概念フレームを付与する解析である。モデルの構築に意味役割付与データを利用するが低頻度の事例が多い。よって本研究では事前分布を仮定したベイズの定理に基づく意味役割付与モデルを提案する。事前分布としてディリクレ分布を利用して意味役割の同定に必要な特徴量を組み合わせた関数を定義し、ベイズ推定により学習データからパラメータの値を得るための定式化を行う。模擬的な学習データに対して得られた確率分布が妥当であったことを報告する。

1.1 研究背景

自然言語処理の分野において文を理解する手法の一つに、意味役割付与という手法がある。意味役割付与とは、文章中の述語の係り受け関係にある項との意味関係を付与する手法である。図 1 は、「客がタクシーを乗り逃げする」という文に意味役割付与をした結果である。この例では、「乗り逃げする」という動詞に対して、「逃亡」という概念フレームを付与し、「客が」に「動作主」、「タクシーを」に「対象」という意味役割を付与している。



図 1 意味役割付与の例

この概念フレームを意味役割は竹内研究室で開発している述語項構造ソーラスに基づく*1。BCCWJ-PT による意味役割付与の先行研究がニューラルネットワークで行われているが^[1]、本研究ではベイズによるモデル化を行う。ベイズの定理による学習には事前分布が用いられているため、低頻度でも極端な分布にならない。この性質を利用できるためベイズの定理による学習は有効な手法と考えられる。

2. ベイズの定理を利用した意味役割付与モデル

本研究の意味役割付与タスクでは、述語項構造ソーラスによる構造の枠組みを利用して、意味役割を付与するモデルを考える。

2.1 変数定義

本節では、文章に対して述語の概念フレームと対応する項の意味役割が付与されているデータを仮定した際の意味役割付与モデルの変数を定義する。文書の総数を D および項の総数を P とインデックス付けされている。また、項の表層、品詞、格助詞、態および動詞の表層の総数を A , Po , Re , Vo および V とする。これらより、観測変数を次のように定義する。

- $arg_surf_{d,p} \in \{0,1\}^A (\sum_{a=1}^A arg_surf_{d,p,a} = 1)$: 文書 d の p 番目の項の表層
- $pos_{d,p} \in \{0,1\}^{Po} (\sum_{po=1}^{Po} pos_{d,p,po} = 1)$: 文書 d の p 番目の項の品詞
- $rel_{d,p} \in \{0,1\}^{Re} (\sum_{re=1}^{Re} rel_{d,p,re} = 1)$: 文書 d の p 番目の項の格助詞
- $voice_d \in \{0,1\}^{Vo} (\sum_{vo=1}^{Vo} voice_{d,vo} = 1)$: 文書 d 番目の態
- $v_surf_d \in \{0,1\}^V (\sum_{v=1}^V v_surf_{d,v} = 1)$: 文書 d 番目の動詞の表層

概念フレームの総数を K 、意味役割の総数を R と固定した場合、次のように出現比率が得られる。

- $\theta \in (0,1)^K (\sum_{k=1}^K \theta_k = 1)$: 概念フレームの出現比率
- $\phi_k \in (0,1)^R (\sum_{r=1}^R \phi_{k,r} = 1)$: k 番目の概念フレームに対応する各意味役割の出現比率
- $\theta'_k \in (0,1)^{Vo} (\sum_{vo=1}^{Vo} \theta'_{k,vo} = 1)$: k 番目の概念フレームに対応する各態の出現比率
- $\theta''_k \in (0,1)^V (\sum_{v=1}^V \theta''_{k,v} = 1)$: k 番目の概念フレームに対応する各動詞の表層の出現比率
- $\phi'_r \in (0,1)^A (\sum_{a=1}^A \phi_{r,a} = 1)$: r 番目の意味役割に対応する各項の表層の出現比率
- $\phi''_r \in (0,1)^{Po} (\sum_{po=1}^{Po} \phi_{r,po} = 1)$: r 番目の意味役割に対応する各品詞の出現比率
- $\phi'''_r \in (0,1)^{Re} (\sum_{re=1}^{Re} \phi_{r,re} = 1)$: r 番目の意味役割に対応する各表層格の出現比率

また、潜在変数を以下に示す。

- $sem_d \in \{0,1\}^K (\sum_{k=1}^K sem_{d,k} = 1)$: 文書 d が K 個の概念フレームのうちどれから生成されたかを示す潜在変数
- $role_{d,p} \in \{0,1\}^R (\sum_{r=1}^R role_{d,p,r} = 1)$: 文書 d の p 番目の項が R 個の意味役割のうちどれから生成されたかを示す潜在変数

潜在変数および観測変数の集合表現は一文字目を大文字にすることで示す。それぞれの変数と事例を表 1 に示す。

† 岡山大学大学院自然科学研究科

*1 <http://pth.cl.cs.okayama-u.ac.jp/>

2.2 生成過程

前節までで定義した変数を利用して、意味役割ラベルの生成過程を定義する。各潜在変数および各観測変数は各パラメータからカテゴリ分布に従い生成されると仮定する。また各パラメータは各ハイパーパラメータからディリクレ分布に従い生成されると仮定する。以上の生成過程および各ハイパーパラメータを図2に示す。

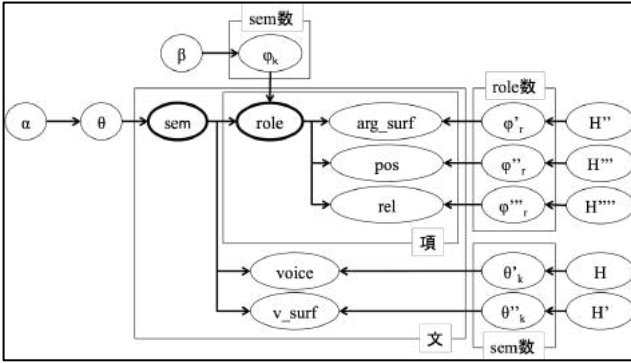


図2 意味役割付与モデル

2.3 同時確率分布

ラベルに関する同時確率分布を以下に示す。

$$\begin{aligned}
 & p(\text{Sem}, \text{Voice}, \text{V_surf}, \text{Role}, \text{Arg_surf}, \text{Pos}, \text{Rel}, \\
 & \theta, \theta', \Phi, \Phi', \Phi'', \Phi''') = \\
 & p(\text{Sem}|\theta)p(\text{Voice}|\text{Sem}, \theta')p(\text{V_surf}|\text{Sem}, \theta'') \\
 & p(\text{Role}|\text{Sem}, \Phi)p(\text{Arg_surf}|\text{Role}, \Phi') \\
 & p(\text{Pos}|\text{Role}, \Phi'')p(\text{Rel}|\text{Role}, \Phi''') \\
 & p(\theta)p(\theta')p(\theta'') \\
 & p(\Phi)p(\Phi')p(\Phi'')p(\Phi''')
 \end{aligned}$$

3. 変分推定

以上で定義した意味役割付与モデル内の各パラメータをデータから推定する手法として、本研究では変分推定を利用する。

3.1 事後分布の近似

図2で示したモデルはカテゴリ分布とディリクレ分布の混合モデルである。したがって、直接計算すると莫大な計算量になり実行することが困難である。そこで、以下のような平均場近似を利用する。

$$\begin{aligned}
 & p(\text{Sem}, \text{Role}, \theta, \theta', \Phi, \Phi', \Phi'', \Phi''') \\
 & \approx p(\text{Sem})p(\text{Role})p(\theta)p(\theta')p(\Phi)p(\Phi')p(\Phi'')p(\Phi''')
 \end{aligned}$$

3.2 近似事後分布

変分ベイズ推定より導出した Sem の近似事後分布を以下に示す。

$$q(\text{sem}_d) = \text{Cat}(\text{sem}_d|\eta_d)$$

ただし、

$$\begin{aligned}
 & \eta_{d,k} \propto \exp\{\langle \ln \theta_k \rangle_{q(\theta_k)}\} \\
 & + \sum_{v_0=1}^{V_0} \text{voice}_{d,v_0} \langle \ln \theta'_{k,v_0} \rangle_{q(\theta'_{k,v_0})} \\
 & + \sum_{v=1}^V v_{\text{surf}_{d,v}} \langle \ln \theta'_{k,v} \rangle_{q(\theta'_{k,v})} \\
 & + \sum_{p=1}^P \sum_{r=1}^R \langle \text{role}_{d,p,r} \rangle_{q(\text{role}_{d,p,r})} \langle \ln \phi_{k,r} \rangle_{q(\ln \phi_{k,r})}
 \end{aligned}$$

ここで、Cat はカテゴリ分布を示している。

4. 小規模な模擬データによる評価実験

本実験で用いる模擬データを説明する。観測した変数を1としそれ以外を0とした疎行列を定義する。偏ったデータを作成するために、sem と role は0番目と対応する変数、各観測変数は1回ずつ順番に観測するデータを作成する。実際のデータで推定する場合は項の数を観測する必要があるが、本実験では3で固定する。K, R, Vo, V, A, Ps, Reをそれぞれ10, 5, 2, 20, 50, 5, 10とする。

4.1 実験結果および考察

提案したモデルで推定した sem と role の分布を図3、図4に示す。sem は10カテゴリ中、出現が多いクラス0の確率が高い。また role も出現が多いクラス0だけ確率が高い。よって、意図通り計算できていることがわかる。

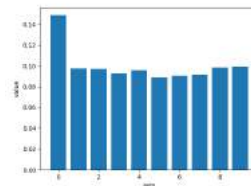


図3 sem の分布

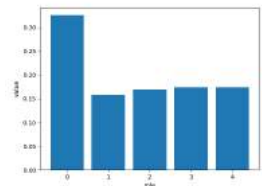


図4 role の分布

5. おわりに

本研究では、日本語の係り受け関係に意味的な関係と述語の概念フレームを付与する日本語意味役割付与をベイズの定理を利用してモデル化した。模擬データによる評価実験を行った結果、模擬データの分布を反映した確率分布が得られたことを実験的に示した。

謝辞

本研究は国立国語研究所機関拠点型基幹研究プロジェクト「統語・意味解析コーパスの開発と言語研究」および科研費(課題番号15H03210, 19H00552)の助成を受けた。

参考文献

- [1] 岡本 拓哉, 竹内 孔一, 石原 靖弘 “ニューラルネットワークを利用した日本語意味役割付与モデルの構築”, 情報処理学会論文誌, Vol.60, No.11 pp.2063-2074 (2019) .