

状態空間モデルを用いた Wikipedia からの 雑談対話システム用発話文の獲得

Acquiring Utterance Candidates for Chat-oriented Dialogue Systems from Wikipedia Based on a State Space Model

松本 知己^{*1} 酒井 正人^{*1} 中野 幹生^{*2}
Tomoki MATSUMOTO Masato SAKAI Mikio NAKANO

^{*1}株式会社 Nextremer ^{*2}株式会社 ホンダ・リサーチ・インスティテュート・ジャパン
Nextremer Co., Ltd. Honda Research Institute Japan Co., Ltd.

This paper presents a method for acquiring utterance candidates for chat-oriented dialogue systems in specific domains from Wikipedia articles. A previously proposed method extracts sentences from Wikipedia articles related to the domain, converts them to utterances suitable for chat, and excludes sentences inappropriate as system utterances using a classifier. The classifier uses position information of the sentence in the article, but it is based on a simple logistic regression. On the contrary, our proposed method uses a state space model in a Bayesian approach. Experimental results show the proposed method outperforms the previous method.

1. はじめに

我々は、特定のドメインの雑談を行う対話システムの構築を目指している。雑談対話システムの発話生成として、seq2seq のようなニューラルネットワークにより発話文を生成する方法 [Li 16] と、あらかじめ用意された発話文集合から文脈に合うものを選択する手法 [稲葉 17] がある。後者は自然な文が生成されるというメリットがあるが、あらかじめ人手で大量の文を用意するのは多大なコストがかかる。

そこで、自動的に対話システムの発話候補文を獲得する手法として、Wikipedia^{*1} を利用した方法が提案されている [太田 09, 杉本 17]。Wikipedia の記事は誰でも編集をすることが可能であるため情報の信頼性が心配されるが、一定の信頼性があることが分かっている [日下 12]。また Wikipedia の各記事は、記事ごとに差はあるが 1 つの記事にあるトピックについての情報が網羅されており、記事の数も膨大であるため、対話システムに用いる対話候補文の獲得源として注目されている。本稿では、Wikipedia を対話候補文の獲得源として注目した [杉本 17] の研究を発展させ、雑談対話システムのための新たな対話候補文の獲得方法を提案する。

本稿の構成は以下の通りである。第 2 章では、本研究に関連する研究を紹介する。第 3 章では、[杉本 17] の方法を概説した後、[杉本 17] のモデルを拡張したモデルについて説明し、第 4 章で提案したモデルの評価をする。第 5 章では考察を述べ、第 6 章では本稿のまとめと、今後の課題を述べる。

2. 関連研究

[太田 09] は、対話が盛り上がるような発話を行う雑談対話システムへの応用を目的に、Wikipedia から発話候補文の抽出を行う方法を提案している。[太田 09] は珍しい単語や単語の組み合わせで構成される文は人間にとって意外性があり、対話を盛り上げるような内容であると仮定し、TF-IDF、単語の共起性、文の長さを用いて計算されるスコアで文をランク付けする方法を提案し、そのスコアの値が高い文ほど、対話を盛り上

げる内容の文であるとしている。しかし、この方法では一部の特徴的な文だけが抽出されるため、これだけでは雑談対話システムの実現は難しい。

また、[杉本 17] は雑談対話システムのために Wikipedia の記事から自動的に文を抽出し、抽出した文を対話システムの出力として相応しいかという点でフィルタリング、及び口語体へ変換する方法を提案している。そして、こういった変換が施された文が、実際に対話システムの文として使用可能かどうか判別するモデルとして、文を抽出する際に付随する記事上における文の位置情報を特徴量とした、ロジスティック回帰判別モデルを提案している。しかし、[杉本 17] の判別モデルは、文の位置情報を十分に活用できていなかった。

本稿では、これら文の位置情報をベイジアンモデリングの枠組みに組み込むことで、[杉本 17] が提案したロジスティック回帰判別モデルを拡張する。

3. 提案手法

本章では、[杉本 17] が提案した、Wikipedia から発話候補文を獲得する方法について概説した後、獲得した文に付随する情報を用いた、新たな文使用可否判別モデルを提案する。

3.1 候補文の獲得

[杉本 17] が提案した、発話候補文を獲得する方法について概説する。まず、Wikipedia 記事の HTML の<p>タグ内に属する文章を抽出し、句点 (。) 区切りで文章を文に分割する。

次に、抽出した各文に対して、フィルタリングを行い、さらに口語体変換を行う。フィルタリングは、一文中に含まれる文字数や、一文で意味を理解できない文 (文脈依存の文)、対話システムへの使用が不適切と判断される表現を含む文などを排除することである。そして、フィルタリングによって排除されなかった文は口語体に変換される。この時点で、例えば、Wikipedia の「お好み焼き」トピックからは

お好み焼きは、小麦粉とキャベツなどを使用する鉄板焼きの一種で、日本の庶民的な料理なんですよ^{*2}

連絡先: 松本 知己, 株式会社 Nextremer, 東京都板橋区成増 1-30-13 トーセイ三井生命ビル 10F, 03-6915-6447, tomoki.matsumoto@nextremer.com

*1 日本語版 Wikipedia: <http://ja.wikipedia.org/>

*2 本稿では、クリエイティブ・コモンズ・表示・継承ライセンス 3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>) のもとで公表されたウィキペディア (<https://ja.wikipedia.org/>) の「お好み焼き」及び「ウイスキー」の項目を二次利用した。(参照日 2018 年)

といった文が抽出される。

こうして抽出された文が対話システムに使用可能かどうかを判定する。ここで、対話システムに使用可能かどうかの基準は、(1) 日本語として自然で、(2) その文だけで内容が理解でき、(3) その記事のトピックに関して述べている、である。この判別には、記事から文を抽出した際に得られる、以下の特徴量を用いたロジスティック回帰判別を用いる。^{*3*4}

- f_1 : header info (<h1>, <h2>, <h3>, <h4>)
- f_2 : header id (上から何番目のヘッダに所属するか)
- f_3 : paragraph id (ヘッダ中の何番目の段落に登場するか)
- f_4 : sentence id (所属する段落の何文目に登場するか)

3.2 モデル

本節では、[杉本 17] のロジスティック回帰を状態空間モデルへ拡張した、階層文状態空間モデルを提案する。

[杉本 17] は、ロジスティック回帰モデルの特徴量として記事における文の位置情報を使用したが、単に説明変数として使用しただけであったため、十分に位置情報をモデルに組み込めてはいなかった。そこで [杉本 17] のモデルを、以下を考慮したベイズモデルへ拡張する：

- それぞれの記事は、固有の特徴がある。
- それぞれの記事は、ヘッダーによる階層構造を持っている。
- 段落内において、ある文が使用可である確率が、その次の文の使用可否に影響を与える。
- 段落内において、文の使用可否の状態は連続的に変化する。

仮定 (i) は、それぞれの記事は現時点で手に入っているデータからは説明することができない、固有の特徴を持っていることを仮定している。これは、生態学において生物や植物の個体を解析する際に個体差と呼ばれているもので、記事差と呼ぶこととする。仮定 (ii) は、それぞれの記事はヘッダーごとに異なる傾向を示すであろうということを仮定している。例えば、ヘッダー 1 内の文は使用可となりやすく、ヘッダー 4 内の文は使用不可となりやすい、といった異なる傾向を持つことを仮定していることになる。仮定 (iii) は、例えば、ある文が使用可になる確率が高ければ、その次に続く文も使用可になりやすいといった傾向が存在することを仮定している。今回は、ある文に影響を与える文は、その文の 1 つ前の文、2 つ前の文の 2 文とした。仮定 (iv) は、段落内における文の先頭からの順番を時系列のインデックスとみなしたときに、段落内で文はある種の“状態”を持っており、その状態は前の文から次の文へ伝搬していくということを仮定している。このことを表現するために、今回はある文の状態とその 1 つ前の文の状態は似ているという 1 階差分の制約を課した。

以上をまとめるとモデルは以下となる。以下の式で、 $a = 1, \dots, A$ は記事番号、 $h = 1, \dots, 4$ はヘッダー番号、 $b = 1, \dots, B_a$ はパラグラフ番号、 $s = 1, \dots, S_{a,b}$ は文番号、 $t = 1, 2$ はそれぞれ 1 つ前、2 つ前の文を表す遅延番号、 $i = 2, 3$ は特徴量番号で、それぞれ 3.1 節で定義した特徴量 f_2, f_3 を表すインデックスとする。インデックス h には特徴量 f_1 を、インデックス s には特徴量 f_4 を用いることができる。また、 $y_{a,h,b,s}$ は

[杉本 17] の方法によって抽出された文のアノテーション結果であり、教師データとなる。つまり、抽出された文が使用可であれば $y_{a,h,b,s} = 1$ 、使用不可であれば $y_{a,h,b,s} = 0$ となる。

$$y_{a,h,b,s} \sim \text{Bernoulli}(p_{a,h,b,s}) \quad (1)$$

$$p_{a,h,b,1} = \text{logit}^{-1}(\gamma_a + \alpha_1 + \beta_{h,0} + \sum_{i=2}^3 \beta_{h,i} f_{a,h,b,1,i}) \quad (2)$$

$$p_{a,h,b,2} = \text{logit}^{-1}(\gamma_a + \alpha_2 + \eta_{h,1} p_{a,h,b,1} + \beta_{h,0} + \sum_{i=2}^3 \beta_{h,i} f_{a,h,b,2,i}) \quad (3)$$

$$p_{a,h,b,s} = \text{logit}^{-1}(\gamma_a + \alpha_s + \sum_{t=1}^2 \eta_{h,t} p_{a,h,b,s-t} + \beta_{h,0} + \sum_{i=2}^3 \beta_{h,i} f_{a,h,b,s,i}) \quad (4)$$

$$\gamma_a \sim \text{Normal}(\theta_\gamma, \sigma_\gamma^2) \quad (5)$$

$$\alpha_s \sim \text{Normal}(\alpha_{s-1}, \sigma_\alpha^2), \quad s = 2, \dots, S_{a,b} \quad (6)$$

$$\beta_{h,0} \sim \text{Normal}(\theta_{\beta_0}, \sigma_{\beta_0}^2) \quad (7)$$

$$\beta_{h,i} \sim \text{Normal}(\theta_{\beta_i}, \sigma_{\beta_i}^2) \quad (8)$$

$$\eta_{h,t} \sim \text{Normal}(\theta_{\eta_t}, \sigma_{\eta_t}^2) \quad (9)$$

$$\theta_\gamma, \theta_{\beta_0}, \theta_{\beta_i}, \theta_{\eta_t} \sim \text{Normal}(\mu, \rho^2) \quad (10)$$

$$\alpha_1, \mu \sim \text{Uniform}(-\infty, \infty) \quad (11)$$

$$\sigma_\gamma, \sigma_\alpha, \sigma_{\beta_0}, \sigma_{\beta_i}, \sigma_{\eta_t}, \rho \sim \text{Uniform}(0, \infty) \quad (12)$$

logit^{-1} はロジスティック関数、*Bernoulli* はベルヌーイ分布の確率質量関数、*Normal* は正規分布の確率密度関数、*Uniform* は一様分布の確率密度関数を表している。*Uniform*($-\infty, \infty$) 及び *Uniform*($0, \infty$) は無情報事前分布を表しており、事前情報を仮定しないことを意味している。

式 (5) は仮定 (i) を表している。それぞれの記事 a についての固有の特徴 γ_a の値は、平均値としては記事全体に共通の θ_γ であり、その平均の周りを共通の標準偏差 σ_γ で正規分布した結果、記事ごとに異なる特徴 γ_a が現れるというモデリングである。したがって、標準偏差 σ_γ は記事差がどれほど存在するか指標となり、小さな値を取れば、それぞれの記事の固有の特徴は平均周りに集中的に分布することになるため、あまり記事差が無いことを示唆し、逆に大きな値を取れば、それぞれの記事の特徴は平均から離れた値を取りやすくなるため、記事差があることを示唆する。

式 (7), (8), (9) は仮定 (ii) を表している。ヘッダー h ごとのロジスティック回帰における切片項 $\beta_{h,0}$ 、特徴量 f_i ($i = 2, 3$) に対する係数 $\beta_{h,i}$ と、 $t (= 1, 2)$ つ前の文が使用可である確率に対する係数 $\eta_{h,t}$ は、それぞれが各ヘッダーに共通の平均 $\theta_{\beta_0}, \theta_{\beta_i}, \theta_{\eta_t}$ の周りを標準偏差 $\sigma_{\beta_0}, \sigma_{\beta_i}, \sigma_{\eta_t}$ で正規分布した結果、ヘッダーごとに異なった値として現れるというモデリングである。したがって記事差の場合と同様に、これら標準偏差はヘッダーごとに異なる傾向を示すことの指標となる。

式 (3), (4) は仮定 (iii) を表現している。ロジスティック回帰モデルにおける確率を計算するロジスティック関数 logit^{-1} の中に、前の文が使用可になる確率とその係数が掛け合わされた形で入っている。

式 (6) は仮定 (iv) を表現している。ある文の状態 α_s とその前の文の状態 α_{s-1} は“似ている”という、以下の 1 階差分の

03 月 08 日)

*3 フィルタリングや口語体変換、文が対話システムで使用可能かどうかの判別の訓練データのアノテーションの方法の詳細は [杉本 17] を参照。

*4 [杉本 17] の検証では、 f_1, f_2, f_4 を説明変数として使用したロジスティック回帰判別モデルの精度が最も高かった。

	(f_2)	(f_3)	(f_2, f_3)	[杉本 17]
WAIC	0.572	0.576	0.571	-
精度	-	-	0.682	0.644

表 1: 提案モデルの WAIC と精度, 及び [杉本 17] のモデルの精度. (f_2) , (f_3) , (f_2, f_3) はそれぞれ提案モデル説明変数として f_2 だけを用いた場合, f_3 だけを用いた場合, f_2 と f_3 をともに用いた場合を表している.

制約をそれぞれの文の状態に課すことで, 文の使用可否の状態が連続的に変化することを表現している.

$$\alpha_s - \alpha_{s-1} \approx 0 \iff \alpha_s \approx \alpha_{s-1}$$

このことを, ある文の状態 α_s はその前の文の状態 α_{s-1} を中心に標準偏差 σ_α で正規分布するという形で表現したものが式 (6) であり, これは一般的に 1 階差分ガウス状態空間モデルと呼ばれている.

最後に, θ と下付き文字で表されるすべての正規分布の平均パラメータはすべて共通の平均 μ , 標準偏差 ρ の正規分布に従うとし, 段落内における先頭の文の状態 α_1 , μ , σ と下付き文字で表されるすべての正規分布の標準偏差パラメータ, そして ρ はすべて無情報事前分布に従うとする.

4. 評価

本章では, 提案手法のモデル選択と精度評価を行う. 提案モデルには, 説明変数として f_2 だけを用いる場合, f_3 だけを用いる場合, f_2 と f_3 をともに用いる場合の 3 パターンの自由度があるため, 変数選択の指標として WAIC (Widely Applicable Bayesian Information Criterion) を用いる [Watanabe 13]. WAIC は事後分布が正規分布で近似できない場合にも使用できる情報量基準であり, AIC (赤池情報量基準) を一般化したものである.

精度評価には 5-fold の交差検証を実施した. また, Wikipedia から抽出されたデータは [杉本 17] で使用されたものを使用した. [杉本 17] は料理分野に関する 60 記事から候補文を抽出していたため, その 60 記事を 5 分割することで訓練データ 48 記事, 検証データ 12 記事のペアが 5 つできる. これらを用いて交差検証を行い, 5 回の検証で得られた精度の平均値を評価指標とする. モデルの推定には, Stan を R 言語上で実行することができる, RStan を使用した [Stan Development Team 17]. Stan は, 事後分布に従うパラメータのサンプルを得ることができるサンプラーである.

検証結果は表 1 のようになった. WAIC によるモデル選択の結果, WAIC の値が最も小さい, 説明変数として f_2 と f_3 をともに用いるモデルが選択された. また, 選択されたモデルと [杉本 17] のモデルの精度を比較すると, 我々が今回提案するモデルの方が [杉本 17] のモデルよりも高精度に文の使用可否を判別できていることがわかる.

5. 考察

3.2 節で説明したように, 提案したモデルには 4 つの仮定を置いていた. 本章では, それぞれの仮定についての妥当性を確認する.

表 2 は, 3.2 節で提案したモデルの各正規分布の標準偏差パラメータの推定結果である. σ_γ は, 各記事は特有の効果を持つ

	mean	sd	2.5%	25%	50%	75%	97.5%
σ_γ	0.69	0.10	0.51	0.62	0.68	0.75	0.90
σ_0	2.87	3.94	0.77	1.35	1.97	3.11	10.2
σ_{η_1}	0.83	0.81	0.06	0.40	0.64	1.03	2.72
σ_{η_2}	0.49	0.73	0.02	0.16	0.34	0.62	1.86

表 2: モデルの各正規分布の標準偏差パラメータの推定結果. mean は平均値, sd は標準偏差を表す.

	mean	sd	2.5%	25%	50%	75%	97.5%
$\eta_{1,1}$	-0.02	0.53	-1.13	-0.34	0.00	0.31	0.98
$\eta_{2,1}$	0.54	0.27	0.03	0.35	0.54	0.72	1.07
$\eta_{3,1}$	0.39	0.31	-0.18	0.18	0.38	0.60	1.03
$\eta_{4,1}$	-0.35	0.57	-1.56	-0.72	-0.30	0.05	0.64
$\eta_{1,2}$	0.04	0.43	-0.82	-0.19	0.01	0.24	1.00
$\eta_{2,2}$	0.12	0.28	-0.40	-0.07	0.10	0.29	0.74
$\eta_{3,2}$	-0.13	0.33	-0.88	-0.31	-0.10	0.07	0.48
$\eta_{4,2}$	0.10	0.47	-0.76	-0.15	0.05	0.32	1.25

表 3: 1 つ前, 2 つ前の文が使用可となる確率に対する, ヘッダーごとの係数の推定結果. mean は平均値, sd は標準偏差を表す.

ているという仮定 (i) についての妥当性を表していると考えられる. 各記事の特徴は, 平均パラメータ θ_γ の値が定まると, その θ_γ を中心に平均的に標準偏差 0.69 でばらつく. 例えば, 他の変数の値が同じで記事差だけが 0.69 異なる記事同士を比較したとすると, 使用可の確率は $\exp(0.69) = 1.993$ と, 約 2 倍異なってくる. これは記事差が存在することを表していると言える.

σ_0 は, それぞれの記事はヘッダーによる階層構造を持っているという仮定 (ii) の妥当性を表していると考えられる. σ_γ の場合と同様の議論で, ヘッダーごとに異なる特徴を持つことができる. σ_{η_1} , σ_{η_2} はそれぞれ 1 つ前, 2 つ前の文が使用可となる確率に対する, ヘッダーごとの係数 $\eta_{h,1}$, $\eta_{h,2}$ の標準偏差である. 1 つ前, 2 つ前の文が使用可となる確率が当該文に与える影響がヘッダーごとに異なることを表しており, このことからヘッダーごとに異なる特徴を持っていると言える.

表 3 は, 1 つ前, 2 つ前の文が使用可となる確率に対する, ヘッダーごとの係数 $\eta_{h,1}$, $\eta_{h,2}$ の推定結果である. 特に 1 つ前の文が使用可となる確率に対する係数 $\eta_{h,1}$ について, ヘッダー 1 は 1 つ前の文が使用可となる確率に影響をあまり受けず, 逆にヘッダー 2 と 3 に属する文は正の影響を, ヘッダー 4 に属する文は負の影響を受ける傾向があることが分かる. このことから仮定 (ii) は妥当なものであると考えられ, また前の文が使用可である確率が次の文が使用可になる確率に影響を与えるという仮定 (iii) は妥当なものであると考えられる.

図 1 は, 推定された段落内における文の状態の推定結果を, 横軸が段落内における先頭からの文の順番, 縦軸が推定された文の状態として図示したものである. 段落の先頭の文が最も使用可になりやすく, 段落の深い箇所にある文であるほど使用可になりにくいという構造が推定されている. このことから, 段落内において, 文の使用可否の状態は連続的に変化するという仮定 (iv) は妥当なものであると考えられる.

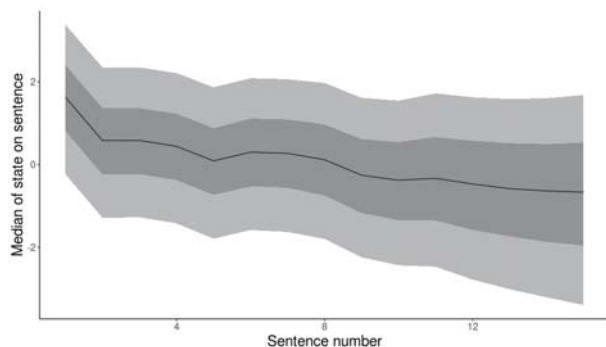


図 1: 文の状態推移. 横軸が段落内における先頭からの文の順番, 縦軸が推定された文の状態. 実線は 50%点, 濃い区間は 50%信用区間, 薄い区間は 80%信用区間を表している.

また, 以下は使用可とアノテーションされた文のうち, [杉本 17] のモデルでは使用不可と判別されていたが, 今回我々が提案したモデルでは正確に使用可と判別できるようになった文の例である.

英国スコットランドで造られるウイスキーをスコッチ・ウイスキーまたは単にスコッチと呼ぶんですよ

この文は「ウイスキー」トピックの記事において, 上から 18 番目のヘッダー番号 4 のヘッダーに属していた. ヘッダー番号 4 の文は使用不可になりやすく, また記事の下の方にあるヘッダーに属する文も使用不可になりやすい傾向にあるため, [杉本 17] のモデルはこの文を使用不可と判断したと考えられる. しかし, この文は段落内における先頭の文であったため, 先頭の文はより使用可になりやすいという, 我々が提案したモデルにおける文の状態の推定結果によって, この文が正しく使用可であると判別できるようになったと考えられる.

以上から, 3.2 節で提案した仮定は妥当なものであると考えられ, こういった文の構造のモデル化の妥当性が, 第 4 章における文の使用可否の判別性能の向上に繋がったと考えられる.

6. おわりに

本稿では, [杉本 17] が提案した雑談対話システムのために Wikipedia から発話候補文を獲得する方法によって得た候補文が使用可能であるかを判別する, 階層文状態空間モデルを提案した. また, [杉本 17] が提案したロジスティック回帰モデルと精度の比較検証を行い, 提案モデルの方がより高精度に文の使用可否判別が行えることを確認した.

今後の課題として, 今回は単に説明変数として用いたヘッダーの順番 f_2 と段落の順番 f_3 のさらなる活用がある. 今回のモデルではヘッダー情報を表す f_1 と段落内での文の順番を表す f_4 を, それぞれ階層化の対象と状態変化の対象としてモデルに組み込むことができた. しかし, ヘッダーの順番 f_2 と段落の順番 f_3 は単に説明変数として用いただけであったため, これらを適切にモデルの中に組み込む方法を考える必要がある. また, 現在は文の位置情報だけをモデルに用いており, 仮に同じ構造をもつ記事が複数あった場合, 記事の内容や文の意味に関わらず全く同じ使用可否判別を行うことになる. より高精度な判別のためには, 文の意味を組み込んだモデルが必要となる.

参考文献

- [稲葉 17] 稲葉 通将, 高橋 健一: 非タスク指向型対話システムのためのリカレントニューラルネットワークを用いた発話候補ランキング, 電子情報通信学会論文誌 D, Vol.J100-D, No.6, pp. 661-671, 2017.
- [稲葉 12] 稲葉 通将, 平井 尚樹, 島海 不二夫, 石井 健一郎: 非タスク指向型対話エージェントのための統計的応答手法, 電子情報通信学会論文誌, vol.95, pp. 1390-1400. 2012.
- [太田 09] 太田 知宏, 島海 不二夫, 石井 健一郎: 発話生成を目的とした Wikipedia からの文抽出, 人工知能学会全国大会講演論文集, 2009.
- [日下 12] 日下 九八: ウィキペディア: その信頼性と社会的役割, 情報管理, vol.55, pp. 2-12. 2012.
- [杉本 17] 杉本 俊, 植木 拓, 林 宏幸, ニコルズ エリック, 中野 幹生: 特定ドメイン雑談対話システムのための Wikipedia を用いた発話文の生成, 人工知能学会 言語・音声理解と対話処理研究会, vol.81, pp. 98-99. 2017.
- [東中 14] 東中 竜一郎: 雑談対話システムに向けた取り組み, 人工知能学会研究会資料 SIG-SLUD-70, pp. 65-70. 2014.
- [樋口 11] 樋口 知之: 予測にいかす統計モデリングの基本—ベイズ統計入門から応用まで, 講談社. 2011.
- [Gelman 13] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, Donald B. Rubin: Bayesian Data Analysis, Third Edition, CRC Press. 2013.
- [Li 16] Jiwei Li, Michel Galley, Chris Brockett, Georgios PSpithourakis, Jianfeng Gao, Bill Dolan: A persona-based neural conversation model, In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL-16), pp. 994-1003. 2016.
- [Stan Development Team 17] Stan Development Team: RStan: the R interface to Stan. R package version 2.16.2. 2017.
- [Watanabe 13] Sumio Watanabe: A Widely Applicable Bayesian Information Criterion, Journal of Machine Learning Research, vol.14, pp. 867-897. 2013.