

伝達内容を考慮して言語化するニューラル言語生成の検討

生田 和也^{1,a)} 品川 政太郎^{1,b)} 吉野 幸一郎^{1,c)} 鈴木 優^{1,d)} 中村 哲^{1,e)}

概要: 自然言語による情報提示は情報案内システムのユーザビリティを高めると期待されており、文生成はこの実現のための重要な課題である。これまでの文生成手法の多くは、伝えたい情報をあらかじめ用意した文のテンプレートに埋め込むことで行われてきたが、表現の多様性を確保するのが難しいという問題があった。この問題点に対し、近年注目されているニューラル言語モデルは、提示すべき情報を損なわずに生成文の自然性を改善することができるという報告がなされている。本研究では、伝達内容を 1-hot のコンテンツベクトルとして与えた上で、このコンテンツベクトルの入力に応じた文生成を可能とするシステムの構築を目指す。本稿では先行研究のモデルを参考に、提示すべき情報を損なわずに文生成が可能かどうかの検討を行った。

1. はじめに

情報基盤技術の発展により、実世界の様々な場所にセンサを取り付け、モノとインターネットをつなげる取り組みがなされている。このような取り組みの中で得られるデータの多くは、数値情報のような構造化された時系列情報が想定される。このような情報を分析しようとする際にはグラフのような形で可視化したものを用いる事が多い。しかし、系列が長期に渡ったり種類が多様である場合、視覚的に何が起きているのかを俯瞰的、直感的に理解することは難しい。そこで、計算機の持っている情報をいかにして理解しやすい形式で表現するかが重要となる。

我々が新しい情報を獲得したり情報交換を行う際に、日常的に用いるプロトコルの 1 つに自然言語がある。自然言語は大半のユーザが直感的に情報伝達を行う事ができる手段であり、計算機の扱う情報を言語として表現する事が出来れば、これらの情報の直感的な理解に繋がる可能性がある。また、言語による表現は視覚以外からも伝達可能となる。最近では、決済短信や市場変動情報等の膨大な情報を言語と対応付ける事で、直感的な情報伝達を行う研究が盛んに行われている [1][2]。

本研究では、様々な場所からセンサデータが収集されると想定し、取得した時系列情報を自然言語で提示するシス

テムの構築を目標とする。実現にあたり、センサデータから伝えたいコンテンツを決定する「What to say」と、コンテンツをどのように書き起こすか決める「How to say」を分離して構築することで、様々な入力に対応した文生成を可能とする枠組みを目指す。特に本稿では、「How to say」に主眼を置き、伝えたいコンテンツが既に定まっている時に、どのように文生成を行うかについて検討する。

2. 関連研究

自然言語による文生成は、対話システムの応答文生成 [3] や天気予報文生成 [4][5]、Web ページの要約 [6] 等、多岐のドメインに渡って研究されてきた。1 章において述べたように、文生成は、「What to say」と「How to say」という 2 つのタスクをまたいで行われる。このうち「How to say」については、文の骨組みに情報を埋め込むテンプレートによる文生成 [4][7][8] や、言語モデルを用いて文生成を行う研究 [2][9] などがある。

文献 [4] では、風速や波の高さなどの予測情報から天気予報の説明を行うシステムを構築している。このシステムでは「What to say」として、予測された時系列情報から発話内容を決定し、「How to say」で予め設定したテンプレートに内容を直接埋め込む事で文生成を行っている。文献 [8] では、発話内容を全体説明と詳細説明に分離した上でテンプレートを構築することで、天気予報やロボカップの実況など、複数ドメインにおける文生成を可能にしている。テンプレートによる文生成の問題点として「文の多様性がテンプレートの数に依存する」という点が挙げられる。

¹ 奈良先端科学技術大学院大学情報科学研究科
Nara Institute of Science and Technology

a) ikuta.kazuya.hy5@is.naist.jp

b) shinagawa.seitaro.si8@is.naist.jp

c) koichiro@is.naist.jp

d) ysuzuki@is.naist.jp

e) s-nakamura@is.naist.jp

文の多様性は生成文の自然性に直結しており、伝達する情報の多様性に応じて必要なテンプレートも膨大となるが、テンプレートの作成は人的コストがかかる場合が多い。文献 [7] では、テンプレートをコーパスから大量に自動構築し、意味構造の関係を示したフローグラフに対して適切なテンプレートを自動選択するというアプローチを用いている。これにより、人手によるコストをかけずに柔軟な文生成が実現可能ではあるが、依然として多様性の問題は残る。

言語モデルを用いて動的に文生成を行う研究として、株価の推移を時系列情報として動向を記述する研究 [2]、将棋の盤面から解説文を生成する研究 [9] などがある。これらではテンプレートの代わりにコーパスの文を元に N-gram 言語モデルを構築して文生成に利用している。近年では言語モデルを Neural Network を用いて構築する研究も行われており、文献 [3] では、対話システムの応答文の生成に Recurrent Neural Network (RNN) を用いることで、自然性や多様性などにおいて従来法より高い評価を得ている。

文生成において「良い文」とはどのようなものかを一概に決定付けるのは難しいが、文献 [10] では、生成文の評価指標として (1) 妥当性、(2) 流暢性、(3) 可読性、(4) 多様性が重要であるとしている。

テンプレートを用いた文生成について考えると、意味表現の埋め込みによる生成文は流暢性が問題となる可能性があり、またテンプレートの設計方法によっては生成文の多様性が犠牲となる場合が予見される。一方で、言語モデルを用いた生成手法は、生成すべきものを生成出来るかどうかの妥当性に課題が有ることが多い。よって本研究は RNN を用いた文生成を、妥当な内容が出力されるように制御する手法について検討する。

3. コンテンツベクトルを制約とした文生成

本研究では、言語モデルを用いた文生成を行う。言語モデルとは、文を単語列とみなし単語列のモデルに対する尤もらしさを測るものである。例えば文 S を単語列 $S = w_0w_1w_2\dots w_n$ として考えると、この文の生起確率 $P(S)$ は以下の式で一般化される。

$$P(S) = P(w_0) \prod_{i=1}^n P(w_i | w_0, w_1, \dots, w_{i-1}) \quad (1)$$

一般的によく用いられる N-gram 言語モデルは (1) に N 重マルコフ性を仮定し、右辺第二項を

$$P(w_i | w_0, w_1, \dots, w_{i-1}) \sim (w_i | w_{i-(N-1)}, \dots, w_{i-1}) \quad (2)$$

と近似して使用する。

この場合考慮される文脈は対象単語の直前の $N - 1$ 語に限られるため、N が小さい場合は長期系列における文の関係性を考慮することが難しく、逆に大きくした場合は対象

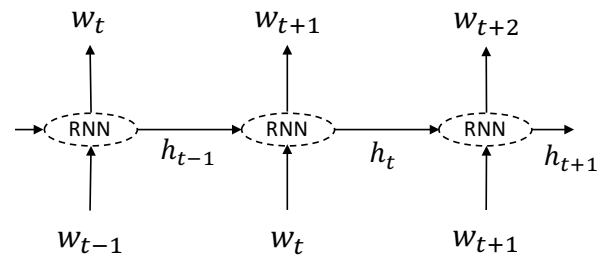


図 1 RNN 言語モデル

単語列の出現頻度が減少するため、ゼロ頻度問題が生じ、有限の学習データから正しくモデルを構築することが難しい。これに対し、近年 RNN を用いることで、必要に応じて履歴の内容を次の単語の予測に用いる言語モデル [11] が利用されており、本研究でもこれを用いる。

RNN は Neural Network の中でも内部に再帰結合を有するものを指し、音声や言語等の系列データの振る舞いを学習する際に有効であるとされている。

図 1 は RNN 言語モデルのネットワークである。以降では、便宜上これをニューラル言語モデルと呼ぶ。このネットワークでは、モデルを学習させるために、各タイムステップ t で単語 w_t を逐次的に与え、次の単語 w_{t+1} を予測するように学習を行う。ここで、ネットワークが出力する予測信号を y_t^w とすると、予測信号 y_t^w は以下の式で表される。

$$y_t^w = \text{softmax}(W_1 h_t) \quad (3)$$

$$h_t = \tanh(W_2 w_t + W_h h_{t-1}) \quad (4)$$

式 (4) を見ると、 h_t は h_{t-1} を用いることで定義されている事がわかる。これは、各タイムステップにおいて h が過去の情報を保持して更新されるということである。これにより、理論的に (1) 式を近似する事無くモデル化しているとみなす事ができる。

上記を用いて出力される y_t^w は単語 w_{t+1} を正解とする教師信号と比較され、それらを近づけるように W を最適化する事で文脈の学習を行う。

3.1 ニューラル言語モデルによる文生成

ニューラル言語モデルにおける文生成では、構築したモデルに従い次に出現するであろう単語の確率分布を出力する。この分布を用いて、単語列の長さが一定長に達するか終了記号 <eos> を出力するまで単語をサンプリングする事で文を生成する。

何も制約を加えずに単純に言語モデルからサンプリングを行う場合、生成される文は尤度の高い流暢なものとなってしまう、明示的な内容の制御を行う事が難しい。文献 [3] では、対話システムにおける発話内容をコンテンツベクトルとして与え、これを制約とする事で生成文を制御している。本稿ではこれを参考に、伝えたいコンテンツをベクトルとして与えることで文生成を行う。

```
<s> <num> 日の 沖縄 地方は、 高気圧に 覆われ
おおむね 晴れる でしょう。 <eos>
<s> 静岡県は、 おおむね 曇りで 雨の 降っている
所 があります。 <eos>
<s> <num> 日は、 晴れで 明け方 まで 曇り でしょう。 <eos>
```

図 2 コーパスの一例

3.2 ドメインとコーパス

文生成を行うドメインは、時間変化に応じて伝えたい内容が変化するものとして天気予報説明文とした。コーパスの収集には、Livedoor の WeatherHacks^{*1}の天気予報概況文を用いた。2016年10月から2017年1月の内、APIにより収集した16日分を用いてコーパスとし、前処理を経て得られた9421文のうち75%の7066文を学習対象とした。

前処理として、Mecabを用いた単語分割、数字の共通表現置き換え、低頻出単語の未知語置き換えを行った。前処理後のコーパスの一例を図2に示す。

3.3 コンテンツベクトルの設計

生成する内容を制御するために、伝えたい内容(コンテンツ)をベクトルとして管理するコンテンツベクトル d を用いる。このベクトルの各次元はコンテンツに対応しており、例えば天気予報文について「朝」、「昼」、「夜」という3つの状態を管理したい場合、これを管理するベクトル d の次元は $\dim(d) = 3$ となる。この d は 1-hot 形式で使用し、「1」か「0」の値のみを初期値として持つような形となる。このうち1つの状態のみを取り扱う際は、

$$\begin{aligned} d_{\text{morning}} &= [1, 0, 0] \\ d_{\text{noon}} &= [0, 1, 0] \\ d_{\text{night}} &= [0, 0, 1] \end{aligned}$$

というような形を指定する事で表現できる。

本稿では、コンテンツベクトルで管理する内容として「地域名」を用いる。これは、生成器に擬似的に与えるコンテンツベクトルを生成する際に曖昧性が少ないと考えられるためである。実際にコンテンツベクトルを設計するにあたり、固有表現の語彙が豊富である辞書 Neolog^{*2}を用い、地名の抽出を行った。その結果今回得られた地名は「関東」「東京」等を含む54種類となり、各地名を1つの次元に割り当て $\dim(d) = 54$ とした。図2の一文目の例では、対応する d は、「沖縄」に対応している値のみ「1」となり、他は「0」であるベクトルで表現される。このベクトル d を意味表現として用いて、文脈とコンテンツの出現を同時に最適化させることで、生成される内容の制御を行う。

*1 http://weather.livedoor.com/weather_hacks/webservice

*2 Neologism dictionary based on the language resources on the Web for Mecab
<https://github.com/neologd/mecab-ipadic-neologd>

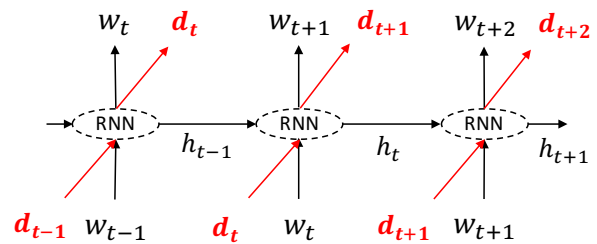


図 3 本研究で用いるネットワーク構造

3.4 学習と生成

本研究で用いたネットワーク構造を図3に示す。

図3の w は単語、 d はコンテンツベクトルである。本アーキテクチャでは、この2つを入力に用いて学習を行う。

学習の初めに開始記号 w_0 と文に対応した d_0 が与えられ、タイムステップ毎に逐次的に輸入される。また、タイムステップ t において w_{t+1} がコンテンツベクトル d に存在している場合、対応するコンテンツの値を「0」とする更新を行い、 d_{t+1} とする。

このコンテンツベクトル d の状態を学習するために、(4)式を以下のような形に変更して用いる。

$$h'_t = \tanh(W_2 w_t + W_h h_{t-1} + W_d d_t) \quad (5)$$

また、タイムステップ t における d_{t+1} を予測する新たな出力 y_t^d を以下のように定義して使用する。

$$y_t^d = \text{sigmoid}(W_3 h_t) \quad (6)$$

このようにして得られた出力 y_t^d と (3) 式による y_t^w を最適化する事で、文脈とコンテンツを同時に学習する。

文を生成する際は、 w_0 と、生成文に含めたい任意のコンテンツに対応するインデックスを「1」としたコンテンツベクトル d_0 を与えて行う。これにより作られる文は式(1)によって条件付けられた上、さらに式(5)(6)によるコンテンツベクトル d の制約を加える事で

$$P(w_i | w_0, w_1, \dots, w_{i-1}) = P(w_i | w_0, w_1, \dots, w_{i-1}, d) \quad (7)$$

によって生成され、 d の初期状態に応じて生成文が変更される。また、生成時に当該コンテンツが選択された時、理想的には対応するコンテンツベクトルの値は「0」へ更新される。それにより、 d の推移から生成文が妥当であるか判断することが可能となる。

ここまでで示した言語モデルを用いて、初期コンテンツベクトル d_0 を与えた時、確率分布に従い生成確率の高い単語を探索して単語を選択することで文生成を行う。生成は、文が規定長に達するか終了記号が選択されるまで行う。

3.5 ビームサーチ

生成候補の単語探索を行う際、全ての候補を探索対象とすることは計算量的に現実的ではない。そこで、計算量の

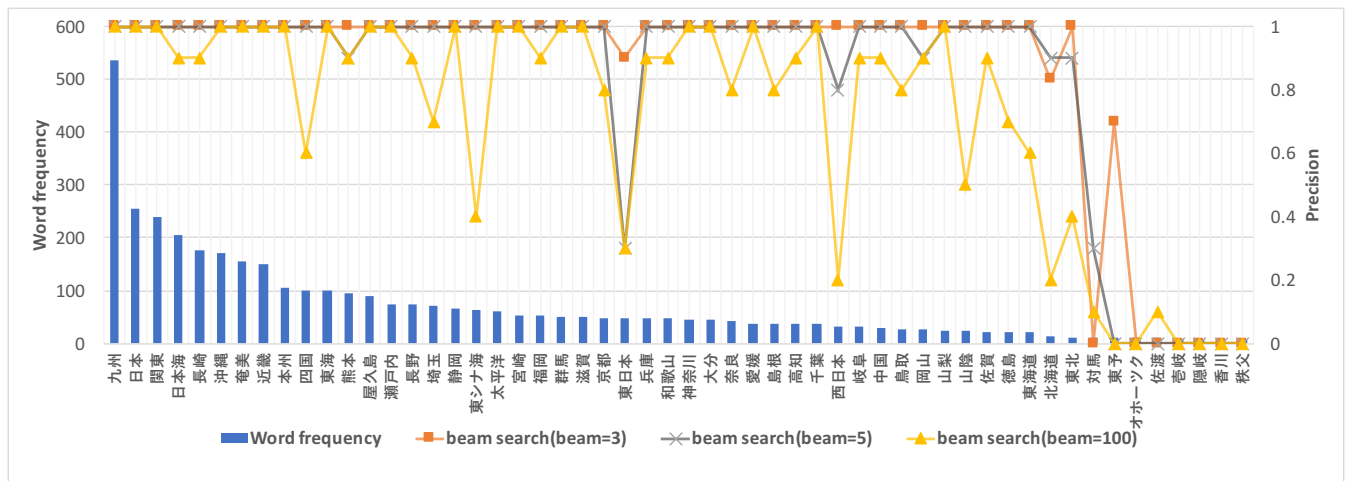


図 4 単語毎の Precision と語彙数

削減のためビームサーチを用いる。これは単語を全探索する代わりに、各ステップで生起確率の高い上位 k 個の系列のみを保持する探索手法である。この k をビーム幅と呼ぶ。探索の過程で終了シーケンスが出たものから順にスタッキングを行い、規定長に達した時点で生起確率の高い順にリランキングを行い、生成結果とする。

4. 実験と評価

4.1 実験条件

本稿では、コンテンツベクトルを用いたニューラル言語モデルによる生成文制御について検討した。評価にあたっては、コンテンツベクトルの初期状態 d_0 を反映した文生成が行われているかどうかを測るため、生成文に指定したコンテンツが含まれているかどうかで評価を行った。具体的には、コンテンツベクトル d の各要素を「1」とした 54 種類のベクトルを用意し、各ベクトルを初期値とした生成文のうちコンテンツを含んでいる割合を Precision として評価に用いた。

生成に際してはビームサーチを用い、評価には生成文の生起確率上位 10 文を用いた。ビーム幅は 2, 3, 4, 5, 10, 20, 30, 50, 100 とした場合についてそれぞれ評価を行い、生成における最大文長はコーパス内の文で最も単語数の多かった 60 単語を上限とした。

4.2 実験結果

実験結果を表 1 に示す。表 1 の Precision 内の数値がスコアであり、括弧内の値は、Precision の内訳 (コンテンツを包有する文の数/生成文の数) である。なお 60 語生成時点で終了記号を出力していない文については適切な文生成が行えていないものとし、評価対象外とした。この結果、beam size = 3 の時に最も Precision が高く、beam size を

表 1 beam 幅における Precision の変化

method	beam size	Precision
Beam search	2	0.861 (155/187)
	3	0.862 (343/398)
	4	0.809 (389/481)
	5	0.827 (407/492)
	10	0.798 (424/531)
	20	0.777 (414/533)
	30	0.761 (407/535)
	50	0.728 (393/540)
	100	0.719 (388/540)

広く取るほどスコアが下がる結果となった。

まず生成文の数 (表 1 括弧内の分母) であるが、ビーム幅が広いほど増えている。これは単語の探索幅が広がるため終了シーケンスに辿り着きやすく、逆にビーム幅が小さいほど探索範囲が狭いため局所的なループに陥りやすいためだと考えられる。このため、ビーム幅が大きいほど柔軟な文生成が行われていると推察できる。

次に生成文のうち対応コンテンツが含まれている文の数 (括弧内の分子) であるが、これはビーム幅が大きくなるにつれて減少の傾向にある。これは、探索幅が広がると生成される文が多様となり、その結果コンテンツを出力する前に終了シーケンスが生成される場合が多くなってしまったためと考えられる。

次に、各コンテンツごとの Precision を図 4 に示す。各コンテンツにおける Precision を見ると、学習に用いたコーパス内の出現頻度が低い語彙ほど小さくなるという結果となった。これは、学習に用いる文数が少なかったため文脈やコンテンツを適切に学習できなかったためだと考えられる。またこれらの結果から、学習コーパスに当該コンテンツを含む文が 20-30 文程度あれば、Precision は十分な値となることが分かった。

図 5 は、実際に生成された文の例と、生成中のコンテン

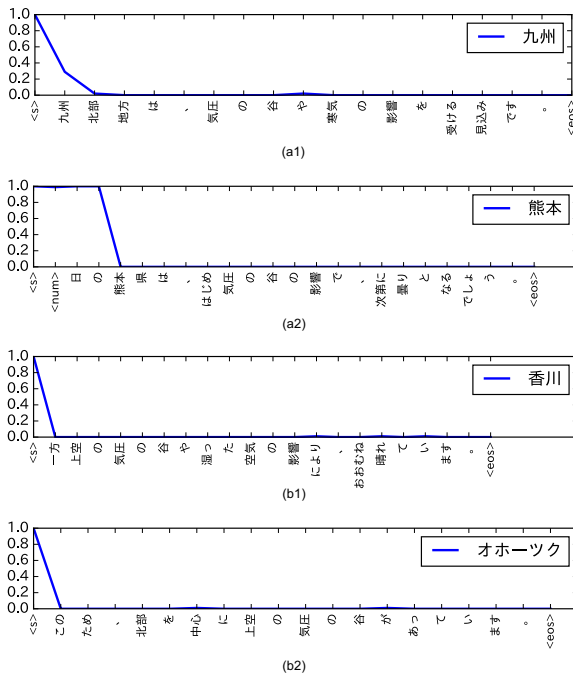


図 5 生成文と対応コンテンツ値の推移

ツベクトルの値である。図 5 (a1)(a2) は生成文にコンテンツが含まれたもの、図 5 (b1)(b2) はコンテンツが含まれなかったものである。この結果から、十分なサンプル数が学習コーパスに含まれなかった単語については、当該語をサンプリングしていないにもかかわらずコンテンツベクトルの値が「0」に遷移している。これは、当該語を生成する確率が低く、言語モデルが文脈を優先しているのではないかと考えられる。

この対策として、低頻度単語を「共通表現を表すタグ」に置き換え、クラス化するという対応が考えられる。例えば、「朝」「昼」「夜」などの時間帯を表す単語は、「時間帯」を表す 1 つの共通表現として統合が可能である。この場合、サンプリングされた「時間帯」といった共通表現に「朝」というような単語を埋め込む事で文を表層化する。しかしこの方法はテンプレート埋め込みによる文生成に近く、場合によっては文意に沿わない埋め込みを行いかねない。そのため、コンテンツをどのような形で中間表現として設計するかが重要となる。

5. おわりに

本稿では、既に伝えたい内容が定まっているという想定のもと、その情報を含んだ文を出力する事のできるニューラル言語モデルの検討を行った。手法として、伝えたい内容を表現したコンテンツベクトルを設計し、言語モデルに文脈とそのコンテンツベクトルを同時に学習させることで、生成文が制御可能であることを確認した。

今後の課題として、複数の内容が同時に与えられた場合

に生成文の制御が可能であるか検討する必要がある。伝えたい内容が多岐に渡るほどコンテンツベクトルは複雑になり、学習の難易度は増加する。この対策として、単語を内容毎に共通表現に置き換えて未知語やベクトルの次元を減らすなどの対応を取る等が考えられ、その場合は共通表現をどのような形でコンテンツベクトルとして設計するかが重要となる。また、実際に生成された文が自然であるかどうかの評価も必要となる。今回の評価尺度では自然性を考慮していないため、文意が通じるのかどうか定かではない。今後は主観評価による流暢性や多様性の検証も行い、柔軟な文生成が可能であるか検討していく予定である。

謝辞 本研究成果は、国立研究開発法人情報通信研究機構 (NICT) の委託研究「ソーシャル・ビッグデータ活用・基盤技術の研究開発」により得られたものです。

参考文献

- [1] 磯沼 大, 藤野 暢, 浮田 純平, 村上 遥, 浅谷 公威, 森 純一郎, 坂田 一郎: 業績変動を考慮した決算短信からの重要文抽出, 研究報告自然言語処理 (NL), pp.1-6 (2016)
- [2] Aoki, Kasumi, and Ichiro Kobayashi. : Linguistic summarization using a weighted N-gram language model based on the similarity of time-series data, Fuzzy Systems (FUZZ-IEEE), IEEE International Conference on. IEEE, pp.595-601 (2016)
- [3] Tsung-Hsien Wen, Milica Gasic , Nikola Mrksic , Pei-Hao Su, David Vandyke and Steve Young : Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp.1711-1721 (2015)
- [4] Sripada, S. G., Reiter, E. and Davy, I. : SUMTIME-MOUSAM: Configurable Marine Weather Forecast Generator, Expert Update, Vol. 6, No. 3, pp.4-10 (2003)
- [5] Kondadadi, R., Howald, B. and Schilder, F.: A Statistical NLG Framework for Aggregated Planning and Realization, Proc. 51st Annual Meeting of the Association for Computational Linguistics, pp.1406-1415 (2013)
- [6] 長谷川 隆明, 西川仁, 今村 賢治, 菊井 玄一郎, 奥村 学携帯端末のための Web ページからの概要文生成人工知能学会論文誌, Vol.25, No.1, pp.133-143 (2010)
- [7] 山崎 健史, 吉野 幸一郎, 前田 浩邦, 笹田 鉄郎, 橋本 敦史, 船富 卓哉, 山肩 洋子, 森 信介: フローグラフからの手順書の生成, 情報処理学会論文誌, Vol.57 No.3 pp.849-862 (2016)
- [8] Angeli, G., Liang, P. and Klein, D.: A simple domain independent probabilistic approach to generation, Proc. 2010 Conference on Empirical Methods in Natural Language Processing, pp.502-512 (2010)
- [9] 亀甲 博貴, 三輪 誠, 鶴岡 慶雅, 森 信介, 近山 隆: 対数線形言語モデルを用いた将棋解説文の自動生成, 情報処理学会論文誌, Vol.55, No.11, pp.1-10 (2014)
- [10] Amanda Stent, Matthew Marge, and Mohit Singhai : Evaluating evaluation methods for generation in the presence of variation, CICLing (2005)
- [11] Tomas Mikolov, Martin Karafit, Lukas Burget, Jan Cernocky , and Sanjeev Khudanpur. : Recurrent neural network based language model, InterSpeech, Vol2 (2010)