ニューラルネットワークを用いた日本語学習者の文章における不自然箇所検知

鈴井　克徳†　　若林　　啓††

† 筑波大学 情報学群 知識情報・図書館学類　〒 305–8550 茨城県つくば市春日 1-2
†† 筑波大学 図書館情報メディア系　〒 305–8550 茨城県つくば市春日 1-2
E-mail: †kts-sz@klis.tsukuba.ac.jp, ††kwakaba@slis.tsukuba.ac.jp

あらまし　近年，日本語学習に対し，日本語教師が不足している．文章における修正すべき不自然な箇所をコンピュータによって検知する手法は，このような状況を改善するのに有用であると考えられる．既に提案されている誤用検知の手法には，文法的な誤用の種類を定め，分類問題として誤用を指摘する手法がある．しかし，この手法では検知できる誤りは設定した誤りの種類に限定され，また文法的には正しくても不自然である表現には対応できない．本研究では，自然な日本語文章のコーパスを大量に用いて日本語の特徴を抽出し，この特徴を用いて，文章中のそれぞれの箇所の自然さを推定する手法を提案する．自然さを推定するモデルとして，ある箇所の尤度を，その前後の文脈から推定するニューラルネットワークを用いる．実験では，日本語学習者による文章の各箇所を，人手により自然か不自然かを判断し，これとモデルによって推定された尤度との対応を評価することで手法の有効性を確認した．

キーワード　ニューラルネットワーク言語モデル，不自然箇所検知

1. はじめに

国際交流基金の 2015 年の調査結果 [1] によれば，多くの国・地域では教育機関で日本語を学ぶ学習者数が増加傾向にあり，これらの国・地域では日本語教育に対する需要が確認されている．一方で，教育機関あたりの教師数は 2012 年の調査結果に比べ，僅かに減少している．同調査結果では，独学で日本語を学習している学習者等についても，インターネット環境の普及に伴い増加しているとの報告が一部の国から上がっていると報告している．したがって，Web サービスを活用した外国語学習支援が重要性を増していると考えられる．

Lang-8 (注1) は，外国語学習を支援する Web サービスのひとつである．Lang-8 では，ユーザーが，学習中の言語で書いた作文を投稿する．それに対し，その言語を母語とする別のユーザが添削を行う．投稿者は，あるときには添削者でもあり．これは言語交換というコミュニケーションの形態である．母語でない言語で文章を書くということは，手間や時間のかかる作業であるが，Lang-8 において，ユーザは余暇の時間を使っていることが想定される．このため，多くのユーザが継続的な利用に困難を感じて離れてしまう．また，学習者が独力で書ききっても，あまりに母語話者にとって意味の取れない文章は，添削されないまま忘れ去られてしまう．このことから，コンピュータシステムによって母語でない文章を書く際の負担を減らすことが，言語交換のようなコミュニケーションを活発化させるためにも有用であると考えられる．

本研究では，日本語学習者による誤用を含む文章に対して，誤用と考えられる箇所を自動で検知する手法を提案する．例えば，学習者の文として，「今日は天気が晴れます」という入力が与えられたとき，これに対して文章が不自然であることを推定する手法の提案を目指す．この文章は文法的には正しいものの，日本語の母語話者はほとんど使わず，このような場合には「今日は天気がいい」「今日は晴れます」という文章を用いている．本研究では，日本語の母語話者が書いた自然な文章のコーパスを大量に用いて，各単語が当該の文脈においてどの程度よく用いられるのかを推定するモデルを学習することで，このような不自然さの検出を行うことを目指す．この目的を達成するため，ニューラルネットワークを用いて，母語話者による自然な文章における形態素の並び方を学習し，学習者の文章を与えたときに，ニューラルネットワークの出力から不自然な並びと判断できる部分を誤用として検知するモデルを提案する．

2. 関連研究

2.1 規則に基づく誤用検知

textlint (注2) のような，規則に基づいて校正支援を行うツールがある．これらのツールでは，人手でルールを管理することを前提としており，想定しうるあらゆる誤りに対して複雑なルールを作る必要があるが，起こりうるあらゆる誤りを人手で作ったルールで網羅することは困難である．更に，自然言語はルールに基づいて運用されるのではなく，実際に用いられている自然言語を観察した結果としてルールが存在しているため，それらのルールが常に成り立つと考えるのは適当でない．

2.2 分類に基づく誤用検知

文章中の表現が誤用であるか誤用でないかを分類するという，分類問題として定式化するアプローチが研究されている．

(注1)：http://lang-8.com

(注2)：https://github.com/textlint/textlint

Rozovskaya ら [2] は，英語の文章について，冠詞，前置詞，動詞の形，名詞の単複といった英文法上の誤りの種類を設定し，分類機を用いて誤りを分類している．Sun ら [3] は，Convolutional Neural Network を用い，英語の冠詞の誤りを訂正する手法を提案している．誤用を分類するアプローチの場合，このように，文法上想定される各種の誤用の定義を明確にした上で，設定した誤りの種類に対してしか分類ができない．

このことは,「黒板を洗う」「天気が晴れる」というような，文法上間違いとは言えないが不自然であるような例で特に問題になる．このような文法上間違いとはいえないが不自然であるような間違いを訂正することの重要性は Sakaguchi ら [4] によって指摘されている．

2.3 機械翻訳

水本ら [5] は，Lang-8 の添削ログから，学習者の書いた文書を原言語，添削者によって直された文章を目的言語とみなしたパラレルコーパスを得て，これを基に，未知の学習者の文章を機械翻訳によって訂正するという手法を提案している．Neubig ら [6] は，話し言葉の忠実な書き起こしを，議事録に掲載する文章に翻訳する手法を提案している．機械翻訳によるアプローチには，巨大なパラレルコーパスが必要となり，これを用意するのが難しいという問題がある．

3. 提案手法

本研究では，ニューラルネットワークを用いて，自然な日本語で書かれた文章の特徴を学習したモデルを，不自然な箇所の検知に利用する．

3.1 ニューラルネットワーク言語モデル

ニューラルネットワークを用いて文章の特徴を学習した言語モデルを，ニューラルネットワーク言語モデル (Neural Network Language Model; NNLM) という．NNLM では，$(n-1)$ 個の単語列が与えられた時，その次に単語 w が出現する確率を推定する．これは，n-gram モデルによる単語の出現確率の推定を，ニューラルネットワークを用いて行うことに対応する．図 1 に，Bengio らによるニューラルネットワーク言語モデル [7] の構造を示す．本稿では，これを Bengio モデルと呼ぶ．

Bengio モデルでは，まず各単語を，one-hot 表現と呼ばれる，その単語に割り当てられた索引の要素のみが 1，他を 0 で表現したベクトルで表現する．one-hot 表現により，文章を単語列 $\mathbf{S} = w_1, w_2, ..., w_{n_S}$ と表現する．このとき， w_i は N 次元のベクトルである．$\mathbf{S}$ の任意の j 番目の単語 w_j の確率を予測するとき，w_j の直前から $n-1$ 個前までの連続する単語 $\mathbf{w_{j-n+1}}, \mathbf{w_{j-n+2}}, ..., \mathbf{w_{j-1}}$ をモデルの入力とする．これらを $N \times P$ の射影行列 C により $\mathbf{c} = (\mathbf{Cw_{j-n+1}}, \mathbf{Cw_{j-n+2}}, ..., \mathbf{Cw_{j-1}})$ と連結した 1 つのベクトルに射影する．隠れ層として，$P \times H$ の重み行列 W_h と H 次元のバイアスベクトル b_h により $\mathbf{d}' = \mathbf{W_h c} + \mathbf{b_h}$ へとアフィン変換を行う．$\mathbf{d}'$ を tanh 関数を活性化関数として非線型変換を行い，$\mathbf{d}$ を得る．$\mathbf{d}'$ の k 番目の要素を d'_k ，$\mathbf{d}$ の k 番目の要素を d_k とすると， $d_k = \tanh d'_k$ と表される．$\mathbf{c}$ を $(n-1)P \times N$ の重み行列 W_o と N 次元のバイアスベクトル $\mathbf{b_o}$ によりアフィン変換したベクトルと，$H \times N$ の重み行列 U によりアフィン変換したベクトルを足し合わせ，$\mathbf{o}' = \mathbf{W_o c} + \mathbf{b_o} + \mathbf{Ud}$ を得る．Softmax 関数を $\mathbf{o}'$ を正規化し，モデルの出力 $\mathbf{o}$ を得る．$\mathbf{o}'$ の k 番目の要素を o'_k ，$\mathbf{o}$ の k 番目の要素を o_k とすると $o_k = e^{o'_k} / \sum_{l=1}^{N} e^{o'_l}$ と表される．この出力を o_k は $\mathbf{w_j}, \mathbf{w_{j+1}}, ..., \mathbf{w_{j+n-2}}$ の次に続く単語のベクトル $\mathbf{w_{j+n-1}}$ の k 番目の要素が 1 である．

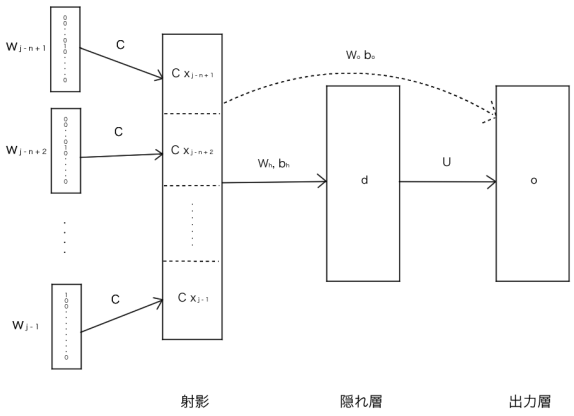

図 1　Bengio モデル構造図

射影行列 C，重み行列 W_h, W_o, U，バイアスベクトル $\mathbf{b_h}$, $\mathbf{b_o}$ が学習すべきパラメータであり，交差エントロピー $E = \ln \mathbf{o^T w_{j+n-1}}$ (注3) を損失関数として．誤差逆伝播法により学習を行う．

本研究で提案するモデルでは，前述のニューラルネットワーク言語モデルとは異なり，周辺の単語列から単語の出現確率を推定する．図 2 に，本研究で提案するモデルの構造を示す．ネットワークの構成は Bengio らによる言語モデルと同じであるが，入力には $\mathbf{w_{j-2}}, \mathbf{w_{j-1}}, \mathbf{w_{j+1}}, \mathbf{w_{j+2}}$ を用い，出力は $\mathbf{w_j}$ の尤度を推定するように学習を行う．これにより，後ろに続く単語に対して不自然であると考えられる単語を見つけ出すことができる．

4. 実　　験

4.1 実験方法

前章において示した提案モデルを Python 3.6.3(注4) と TensorFlow 1.3.0(注5) により実装した．射影行列 C，重み行列 W_h, W_o, U は Xavier の初期値により初期化し，バイアスベクトル $\mathbf{b_h}, \mathbf{b_o}$ は零ベクトルで初期化した．

学習に用いる日本語の自然言語コーパスには，Yahoo! 知恵袋 (注6) における 2004 年 4 月 1 日から 2009 年 4 月 7 日までの質問のうちの 3,000,000 文章，および 2016 年 9 月 7 日取得の日本語版 Wikipedia ダンプデータから本文を取り出したものの

(注3)：Bengio らのモデルにおいては，これにフロベニウスノルムを加えているが，ここでは用いなかった．
(注4)：https://www.python.org/
(注5)：https://www.tensorflow.org/
(注6)：https://chiebukuro.yahoo.co.jp/

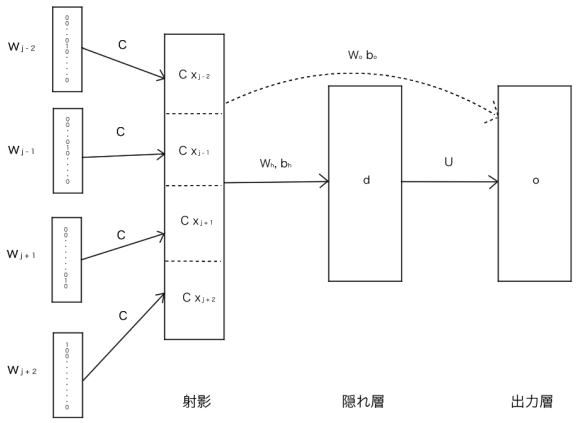

図 2　提案モデル構造図

2 種類を用い，それぞれのコーパスによって 2 つのモデルを学習した．

まず，これらのコーパスを文章に分割し，文章ごとに形態素解析エンジン MeCab (注7) を用いて単語に分割し，それぞれのコーパスについて出現頻度が高い 30,000 語の索引を作成する．これは，モデルの学習の際に，計算資源の都合により，語彙数を制限する必要があるためである．MeCab の辞書には IPAdic 2．7．0 (注8) を用いた．

モデルの入力に用いる単語の one-hot 表現には，出現頻度が高い 30,000 語を表す要素の他に，出現頻度が低いため索引されなかった単語を表す要素，文頭より前または文末より後ろを表す要素の 2 要素を追加し，1 単語あたり大きさ 30,002 のベクトルとする．出力も同様の大きさ 30,002 のベクトルとする．この他のハイパーパラメータは，C による射影の次元数を $P = 100$，隠れ層の次元数を $H = 600$ とした．学習には勾配降下法を用い，この際の学習率は 0.01 とした．Yahoo! 知恵袋については 3,000,000 文章を 1 エポック，Wikipedia については全件を 1 エポックとし，10 エポックの学習を行った．

4.2　評価方法

寺村誤用例集 (注9) からひらがな・カタカナ・漢字のみから為り，漢字を 1 文字以上含む文章を無作為に 100 個選んだ．これらを学習と同様に MeCab により単語に分割し，文章中に用いられている単語について，モデルによって推定された尤度を評価する．

表現が「自然である」かどうかという判断は，複数人が人手により判断を行い，「自然である」と判断した人数により評価できると考えられる．したがって，クラウドソーシングサービス Lancers (注10) において，図 3 に示した指示をし，文章の箇所について，図 4 の形式の質問への回答を得ることで，「自然である」か「不自然である」か，箇所ごとに得た複数人の回答を集計した．各箇所に対する回答者は 5 人としたが，質問作成の都合により，3 箇所についてのみ 10 人の回答を得ている．

(注7)：http://taku910.github.io/mecab/
(注8)：https://ja.osdn.net/projects/ipadic/
(注9)：http://teramuradb.ninjal.ac.jp/
(注10)：https://www.lancers.jp/

文章中の指定された箇所が不自然かどうかお答えください。

例1) 大声で話す人は [ほど] んといません。
→「ほとんど」の間違いと考えられるので「不自然である」とお答えください。

例2) 今日は天気 [が] いい。
→ 特に不自然な文章ではないので「自然である」とお答えください。

例3) 大声で [話す] 人はほどんといません。
→「話す」の部分は特に間違いと考えられる部分ではないので「自然である」とお答えください。

図 3　Lancers での依頼内容

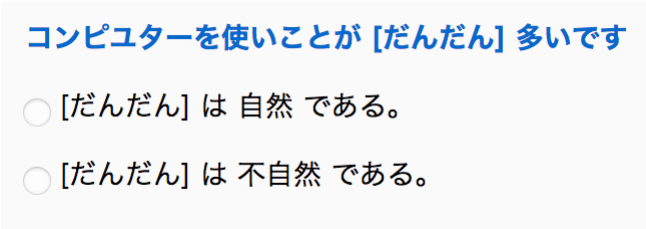

図 4　Lancers での質問例

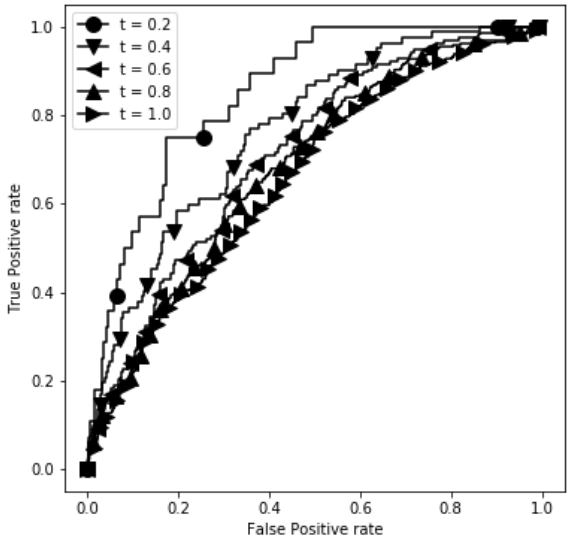

図 5　Yahoo! 知恵袋コーパスで学習したモデルによる ROC 曲線

各箇所について，「『自然である』と答えたのが回答者のうちの $100t$ %未満である $(0 < t \leqq 1)$」場合に「不自然である」とする．このクラウドソーシングの結果を正解と考え，モデルの推定した尤度の ROC 曲線を求め，これに基づく AUC スコアをモデルの評価値とする．

4.3　実験結果

図 5，6 に，Yahoo! 知恵袋コーパスと Wikipedia コーパスを用いて学習を行ったそれぞれのモデルについての ROC 曲線を示す．t の値ごとの検出すべき不自然箇所数、および、2 つのモデルによる ROC 曲線の AUC スコアは表 1 の通りである．

いずれのモデルについても，「5 人中 5 人が『不自然である』

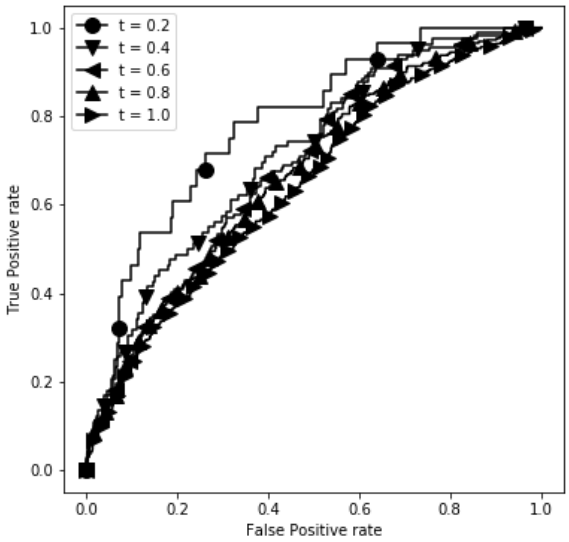

図 6　Wikipediak コーパスで学習したモデルによる ROC 曲線

表 1　t の値ごとの検出すべき不自然箇所数および各モデルによる ROC AUC スコア

t の値	検出すべき箇所の数	知恵袋スコア	Wikipedia スコア
$t = 0.2$	28	0.85040	0.78385
$t = 0.4$	82	0.76603	0.71609
$t = 0.6$	154	0.71353	0.68888
$t = 0.8$	287	0.68450	0.67274
$t = 1.0$	551	0.65751	0.64357
判定対象箇所	計 1,634 箇所		

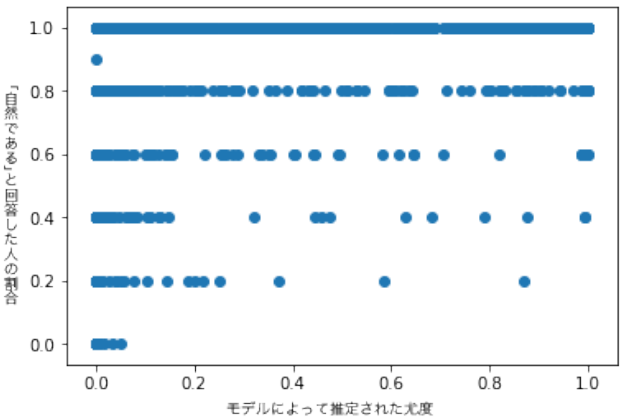

図 7　Yahoo! 知恵袋コーパスで学習したモデルによる散布図

と回答した」箇所のみを「不自然である」と考えた場合にはスコアが 0.8 程度となり，モデルによって推定される尤度は，不自然な箇所を判断するのに有効であることが確認された．

モデル間を比較すると，Wikipedia コーパスによるモデルは Yahoo! 知恵袋によるモデルよりもやや低いスコアとなった．原因として，日本語学習者による文章には「です」「ます」調の文章が多く含まれるが，Wikipedia にはこのような文章が現れないことが考えられる．

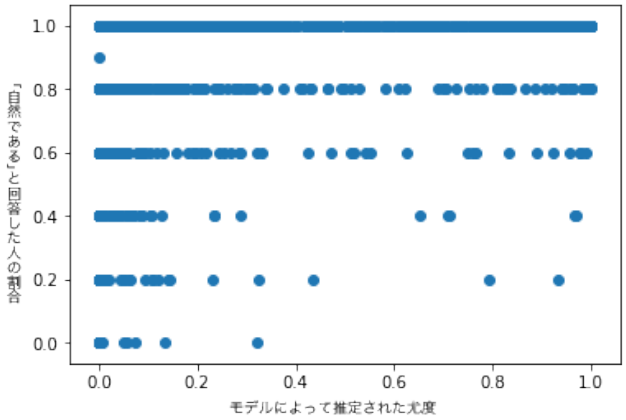

図 8　Wikipediak コーパスで学習したモデルによる散布図

一方，それぞれのモデルについて「自然である」と回答した人数とモデルによって推定される尤度との相関係数を求めると，Yahoo! 知恵袋コーパスによるモデルでは 0.24，Wikipedia コーパスによるモデルでは 0.22 となり，有意な相関は確認されなかった．散布図を図 7，8 に示す．これは、多くの人が「自然である」と回答する箇所に対して推定される尤度が低くなる傾向があるためと考えられる．

5. 結　　論

本研究では，日本語学習者による誤用を含む文章中の，誤用と考えられる箇所を自動で検知する手法を提案した．Yahoo! 知恵袋コーパス，および Wikipedia コーパスを用いてそれぞれモデルを構築し，クラウドソーシングによって得られた回答と比較する実験を行ったところ，「自然である」と回答したのが 5 人中 0 人 (または 10 人中 1 人以下) の場合のみを不自然な箇所と考え，ROC AUC スコアを計算した場合には，両モデルにおいて 0.8 程度となり，モデルの有効性が確認された．

これにより，前後のたかだか 2 単語ずつのみを入力として，自然な文章コーパスから言語の特徴を学習し，日本語学習者による文章を同様に入力として与え，適当な閾値を設定し，これを下回る箇所を「不自然である」とするシステムの有効性が示唆された．

しかし，本研究では，実際に日本語学習者によって評価を行っていないので，実際に日本語学習者が作文を行う際に，「役に立つ指摘をしている」かどうかを評価してもらうような実験が必要であると考えられる．

また，本研究では，たしかに前後 2 単語のみの入力でも有効性は確かめられたが，もっと離れた呼応関係の間違いが文章中に含まれることも考えられる．そのためには，係り受け構造を入力として自然な文章の特徴を学習するモデルの検討が必要であると考えられる．

文　　献

[1] 国際交流基金．2015 年度「海外日本語教育機関調査」結果（速報）．http://www.jpf.go.jp/j/about/press/2016/dl/2016-057-1.pdf, 2015. (参照 2016-12-10).

[2] Alla Rozovskaya and Dan Roth. Building a state-of-the-art grammatical error correction system. *Transactions of the Association for Computational Linguistics*, 2, 2014.

[3] Chengjie Sun, Xiaoqiang Jin, Lei Lin, Yuming Zhao, and Xiaolong Wang. Convolutional neural networks for correcting english article errors. In *Proceedings of the 4th CCF Conference on Natural Language Processing and Chinese Computing - Volume 9362*, NLPCC 2015, pages 102–110, New York, NY, USA, 2015. Springer-Verlag New York, Inc.

[4] Keisuke Sakaguchi, Courtney Napoles, Matt Post, and Joel Tetreault. Reassessing the goals of grammatical error correction: Fluency instead of grammaticality. *Transactions of the Association for Computational Linguistics*, 4:169–182, 2016.

[5] 水本 智也, 小町 守, 永田 昌明, and 松本 裕治. 日本語学習者の作文自動誤り訂正のための語学学習 sns の添削ログからの知識獲得. **人工知能学会論文誌**, 28(5), 2013.

[6] Graham NEUBIG, 森 信介, and 河原 達也. 重み付き有限状態トランスデューサーと対数線形モデルを用いた話し言葉の整形. **情報処理学会研究報告**, (2), 2009.

[7] Y Bengio, R Ducharme, P. Vincent, , and C. Jauvin. A neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137–1155, 2003.