

事前学習モデルに基づく日本語形態素解析器における辞書の利用

田村稔行 河原大輔
早稲田大学理工学術院

tam.fse3982@ruri.waseda.jp dkw@waseda.jp

概要

事前学習モデルに基づく日本語形態素解析器では、従来の形態素解析器で用いられていた辞書を用いないことが一般的である。しかし、単語の区切りや品詞が登録された辞書は形態素解析において重要な知識源となる。本論文では、事前学習モデルに基づく日本語形態素解析器の入力として、辞書引きによる単語区切り位置の情報を与える手法を提案する。提案手法は、複数のコーパスにおいて、辞書を用いない手法と比較して高い精度を示すことを確認した。

1 はじめに

日本語の文においては単語の間に区切りが明示されない。計算機で日本語を解析する場合、まず単語の区切りを認識する必要がある。日本語の形態素解析においては、品詞や活用の付与と同時に単語への分かち書きを行う。

従来の形態素解析器である JUMAN¹⁾、MeCab [1]、Juman++ [2] などは辞書に登録されている単語の情報に基づき、最も尤度の高い単語列を採用することによって分かち書きを行う。一方で、近年の KWJA [3] などの事前学習モデルベースの形態素解析においては、事前学習モデルが日本語解析に関する有用なパラメータを獲得していることから、辞書を分かち書きに用いていない。しかし、単語の区切りや品詞が定義されている辞書の情報は分かち書きや品詞付与において重要な知識源となる。

本研究では、事前学習モデルを用いた形態素解析において辞書の情報を利用する手法を提案する。具体的には、モデルへ入力文に加えて辞書から得た単語の分割位置の情報を与える。提案手法の有効性を検証するため、3種類のコーパスと2種類の辞書において分かち書きと品詞・活用付与の実験を行う。

2 関連研究

辞書を使用する形態素解析の手法として、ラティスを用いるものと点予測のものがある。ラティスは辞書に存在する単語と、辞書に存在しない単語(未知語)を処理する規則に基づいて生成される。生成されたラティスから最も尤度の高いパスを決定し、これが単語の分割となる。ラティスの作成は辞書に基づいて行い、機械学習により調整された重みを用いてスコア算出し、そのスコアによりパスを決定することが一般的である。JUMAN では人手で、MeCab では CRF [4] で重みを調整している。Juman++ も同様に辞書に基づいたラティスを生成するが、線形モデルと RNN の組み合わせによりスコアを算出する。

点予測は、周囲の判定情報を用いずに各文字や単語について分かち書きや品詞推定などを個別に行う手法である。KyTea [5] は点予測の形態素解析器である。文字 n-gram や文字種 n-gram、単語辞書に基づく特徴量を用いて分かち書きと読み・品詞推定を行う。

事前学習を用いた形態素解析器として、事前学習モデル RoBERTa [6] を用いた日本語統合解析器 KWJA がある。KWJA はパイプライン処理によって形態素解析を行う。まず文字モデルにて誤り文字を訂正し、次の文字モデルにて分かち書きと単語正規化を行う。そして単語(サブワード)モデルにより形態素解析を実行する。KWJA の形態素解析においては、単語分割が決まった後に、各単語の先頭トークンの隠れ状態から品詞と活用を予測し、辞書に基づく意味情報を付与する。辞書は分かち書きおよび品詞・活用付与には用いない。また、Tolmachev ら [7] は、双方向 LSTM [8] と Transformer [9] のそれぞれを用いた、辞書を用いない形態素解析器を提案した。この解析器は事前学習を用いないが、大量の自動解析コーパスを学習に用いる。モデルは入力文中の各文字について、分かち書き、品詞・活用付与をマルチタ

1) <https://nlp.ist.i.kyoto-u.ac.jp/?JUMAN>

スクで行う。本研究ではこの出力の構造を用いる。

事前学習モデルに辞書などの知識を統合している研究を紹介する。Sun ら [10] は、事前学習モデルは大量の生コーパスを用いて学習を行っているものの、そのような生コーパスは言語知識や一般常識に関する明示的な知識に欠けていると指摘している。モデルに知識を与えるため、多層の Transformer-XL [11] からなるモデルに知識グラフを埋め込み、単語や知識に関するタスクによる事前学習を提案している。

他の研究例として、BERT [12] の構造を流用したモデルが存在する。事前学習の際に Masked Language Modeling や Next Sentence Prediction と並行して語彙的な関係を分類するタスクを行うモデル [13] や知識に関するタスクを行うモデル [14] がある。また、ファインチューニングにおいて追加の層を用いて知識グラフの情報を統合するモデル [15] などが存在する。

本研究では、既存の事前学習モデルに辞書情報を統合することでの精度向上を目的とするため、ファインチューニングを対象とする。また、形態素解析というタスクに着目し、辞書の情報を、単語の区切り位置の情報としてモデルに与える。

3 事前学習モデルに基づく形態素解析における辞書の利用

本研究では、文字単位のタスクである分かち書きを行うことから、日本語コーパスで事前学習された文字ベースのモデルを用い、分かち書きと品詞・活用付与タスクでファインチューニングする。

モデルの入力としては、入力文の各文字の埋め込みに加え、辞書から得た単語の区切り位置の情報を与える。具体的には、文の各文字について、その文字を単語の先頭、中間、末尾とする単語が辞書に収録されていればそれぞれ B, I, E のフラグを立てる。B は単語の先頭の文字、E は単語の末尾の文字、I は 3 文字以上の単語の中間の文字を示す。例えば、図 1 において、「じかん」という単語が辞書に存在するとする。この場合、入力文中の「じ」に B、「か」に I、「ん」に E のフラグを立てる。入力文中の一つの文字が、辞書中のある単語において B となることもあれば、他の単語において I や E となることもある。例えば図中の「む」には B, E の 2 種類のフラグが立っている。このため、各文字に対して B, I, E の 3 種類の 2 値フラグを与える。以下、これを辞書フラグと呼ぶ。モデルは辞書フラグのそれぞれを埋め込みとして学習し、各文字の埋め込みと足し合わせる。

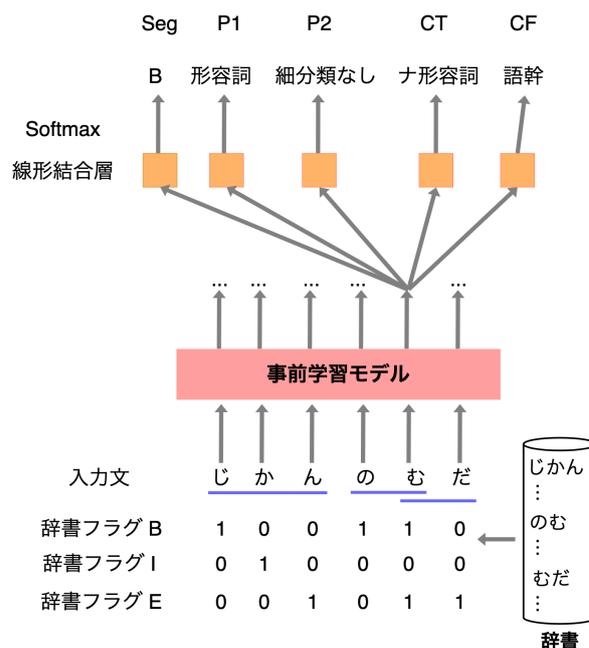


図 1 提案手法。入力文に対して事前に辞書引きをして辞書フラグを形成し、事前学習モデルに入力する。線形結合層が各タスクについて出力する。

モデルの出力としては、事前学習モデルが出力する隠れ状態を各タスク向けの線形結合層に入力し、入力文の各文字について分類を行う。分かち書き (Seg) については BMES の体系を用いて分類する。S は 1 文字からなる単語、B は単語の先頭の文字、E は単語の末尾の文字、M は 3 文字以上の単語の中間の文字を示す。品詞・活用付与については、品詞 (P1)、品詞細分類 (P2)、活用型 (CT)、活用形 (CF) を JUMAN 体系にもとづいて分類する。

4 実験

辞書情報を与えた事前学習モデルベースの形態素解析器の有効性を検証するため、複数のコーパスを評価対象として形態素解析を行い、精度を算出する。

4.1 実験設定

4.1.1 モデルと辞書

ファインチューニングする事前学習モデルとして、文字単位の RoBERTa-BASE²⁾ および RoBERTa-LARGE³⁾ を用いる。学習は異なる 3 つの seed 値を設定して行い、平均値を算出する。実験を行う際のハイパーパラメータを付録 A にて示す。

2) <https://huggingface.co/ku-nlp/roberta-base-japanese-char-wwm>
 3) <https://huggingface.co/ku-nlp/roberta-large-japanese-char-wwm>

辞書については、JUMAN 辞書⁴⁾と森羅⁵⁾を用い、Trie による辞書引きのライブラリである darts-clone⁶⁾を使用して辞書引きを行い、学習・評価データに辞書フラグを与える。森羅は Wikipedia のタイトルとその拡張固有表現 (ENE) における分類情報を持つ辞書である。辞書引きは、文の最初の文字を開始位置とし、その文字から n ($n = 2, 3, \dots$) 文字の各範囲について辞書引きした結果を集計する。この際、 n は辞書引きの範囲が文末の文字に達する場合が最大値である。この作業を開始位置が文末になるまで繰り返す。この際、1 文字からなる単語は有効な情報ではないと判断し、2 文字以上からなる単語の区切り位置についてのみから辞書フラグを形成する。

JUMAN 辞書については、辞書引きの結果を品詞などで区別せずに B, I, E の 3 種類の埋め込みを与える。森羅については、ENE [16] (Version: 7) の「1. 名前」中の 12 種類の大分類それぞれについて B, I, E の埋め込みを与える。つまり、合計 36 種類の埋め込みをモデルに入力する。

実験において、辞書を与えないモデル (辞書なしと表記)、JUMAN 辞書を与えるモデル (JUMAN と表記)、森羅を与えるモデル (ENE と表記)、JUMAN 辞書と森羅を与えるモデル (JUMAN+ENE と表記) の 4 つの条件を比較する。なお、JUMAN+ENE においては、JUMAN 辞書の B, I, E のフラグ 3 個と森羅のカテゴリごとの B, I, E フラグ 36 個を合わせた 39 個の埋め込みをモデルに入力する。

4.1.2 学習・評価データセット

ファインチューニングには、人手のデータセットと自動解析のデータセットの 2 種類を用いる。人手でアノテーションされたデータセットとして京都大学テキストコーパス [17] (以下 KC) と京都大学ウェブ文書リードコーパス [18] (以下 KWDLC) を用いる。自動解析したデータセットとして、CC-100 [19, 20] の日本語テキストから全角スペースと URL を含む文を除いて Juman++ で自動解析したものを用いる。評価用データセットとして、KC と KWDLC、そして不満調査データセットタグ付きコーパス⁷⁾ (以下 FKC) を用いる。なお、KC および KWDLC は GitHub 上の標準スプリットを用いて学習用と評価用に分割する。各データセットに収録されている文数を表 1 に示す。

4) <https://github.com/ku-nlp/JumanDIC>

5) <http://shinra-project.info/>

6) <https://github.com/s-yata/darts-clone>

7) <https://github.com/ku-nlp/AnnotatedFKCCorpus>

表 1 各コーパスの分量

コーパス	文の数	
	学習用	評価用
KC	36, 623	1, 783
KWDLC	13, 856	2, 195
CC-100	250, 988	-
FKC	-	1, 282

KC と KWDLC を学習に用いたモデル (KC+KW と表記) と、KC と KWDLC と C-100 を学習に用いたモデル (KC+KW+CC100 と表記) の両方で実験を行う。

4.1.3 評価

分かち書き及び品詞・活用付与の精度を評価する。分かち書きについては、各単語の開始位置と終了位置の組が合致しているかにもとづいて F1 を算出する。品詞・活用付与については、品詞、品詞細分類、活用型、活用形を評価し F1 を算出する。この際、単語の開始位置である文字について分かち書きと各タスクを両方正解した場合のみ正解したとして精度を算出する。つまり、品詞・活用付与の各精度は、分かち書きと各タスク両方を正解した精度 (F1) である。

また、JUMAN 辞書を用いた形態素解析器として Juman++ (Version: 2.0.0-rc3) を、事前学習モデルを用いた形態素解析器として KWJA (Version: 1.1.0) についても比較対象とする。なお、公開版の Juman++ は評価用データセットも学習に用いているため参考として掲載する。KWJA についても、本実験で評価のみに用いている FKC の一部を学習に用いている。さらに、KWJA については形態素解析の前段階で入力誤り訂正を行うため、入力文と出力文の文字数が異なる場合がある。本実験の評価では、文中で文字数が増減した箇所があると、その文の残りの部分について評価を正しく行うことができない。そのため、KWJA の評価についてはこの問題に対する処理を行わない場合の精度と、入力誤り訂正により文字数が変化した文を取り除いた場合の 2 つの精度を掲載する。

4.2 結果と考察

KC+KW より KC+KW+CC100 を学習に用いた場合に全体的に良い結果が得られたため、KW+KW+CC100 のみの結果を表 2 に示す。KC+KW を学習に用いた結果は付録 B に示す。Seg, P1, P2, CT, CF はそれぞれ分かち書き、品詞、品詞細分類、活用型、活用形を指す。以下では、各辞書の情報を与えたモデルについて、他のモデルとの比較を考察する。

表2 KC+KW+CC100を用いた学習による分かち書きと品詞・活用付与の精度(F1). KWJAのFKCでの精度とJuman++の精度以外について、表中で最も高い精度を示したモデルの値を太字で示した。なお、最高値と2番目に高い値の差が0.01以内の場合は2番目に高い値も太字で示した。

解析器	KC(新聞)					KWDLC(Web)					FKC(ユーザー不満)					
	Seg	P1	P2	CT	CF	Seg	P1	P2	CT	CF	Seg	P1	P2	CT	CF	
Juman++(2.0.0-rc3)	99.84	99.40	98.72	99.18	99.18	99.40	98.85	98.03	98.74	98.90	98.99	98.42	97.99	97.27	97.45	
KWJA	97.06	96.68	95.85	91.08	91.79	95.92	95.47	94.29	90.71	91.17	93.27	92.15	91.56	83.99	83.83	
KWJA(タイポ修正文除去)	98.94	98.90	98.11	95.93	96.66	98.34	98.17	97.01	96.36	96.87	99.26	98.88	98.36	96.66	96.64	
BASE	辞書なし	99.37	99.26	98.59	98.26	98.23	98.73	98.47	97.42	97.98	98.22	98.88	98.15	97.50	96.66	96.77
	JUMAN	99.50	99.34	98.69	98.33	98.36	98.82	98.52	97.49	97.96	98.22	99.00	98.32	97.74	96.87	96.92
	ENE	99.33	99.24	98.56	98.11	98.14	98.74	98.47	97.43	97.98	98.22	98.87	98.18	97.53	96.65	96.77
	JUMAN+ENE	99.52	99.34	98.66	98.26	98.32	98.81	98.50	97.49	97.98	98.25	99.01	98.30	97.65	96.79	96.93
LARGE	KWJA	97.42	97.33	96.31	94.51	94.94	98.06	97.99	97.15	94.13	95.05	95.70	95.06	94.67	90.15	90.43
	KWJA(タイポ修正文除去)	98.08	98.07	97.06	96.37	96.82	98.76	98.83	97.97	96.19	97.17	98.08	98.07	97.06	96.37	96.82
	辞書なし	99.44	99.33	98.74	98.39	98.35	98.83	98.53	97.52	98.04	98.26	98.87	98.17	97.54	96.70	96.87
	JUMAN	99.53	99.40	98.83	98.55	98.65	98.86	98.60	97.59	98.14	98.32	99.01	98.33	97.73	96.86	97.03
	ENE	99.41	99.31	98.72	98.27	98.23	98.80	98.53	97.50	98.09	98.28	98.92	98.21	97.62	96.81	96.94
	JUMAN+ENE	99.53	99.38	98.79	98.47	98.41	98.89	98.59	97.58	98.07	98.30	98.98	98.32	97.71	96.77	96.96

表3 分かち書きの改善例。“/”は形態素の区切りを示す。モデルはいずれもLARGE。

モデル	分かち書き文
辞書なし	ヒゲ/も/じゃで/、/伸び/放題/の/髪/。
JUMAN	ヒゲもじゃで/、/伸び/放題/の/髪/。
辞書なし	同じく/わさび/醤油/や/お寿司/(中略)/。
JUMAN	同じ/くわさび/醤油/や/お寿司/(中略)/。
辞書なし	まじ/時間/の/むだ/。
JUMAN	まじ/時間/の/むだ/。

JUMAN 辞書による影響 JUMAN 辞書を与えたモデルを、辞書を用いないモデルと比較すると、すべてのデータセット・タスクにおいて提案手法の精度が上回った。このことから、辞書による単語の区切り位置情報が、分かち書きのみならず品詞・活用付与の精度向上にも貢献しているといえる。提案手法ではJUMAN 辞書を品詞の区別なく与えたが、品詞ごとに辞書フラグを与えることで、品詞・活用付与の各タスクにおいてさらに精度が向上する可能性がある。

各評価用データセットの分かち書きについて、辞書情報を与えないモデルでは間違えていたがJUMAN 辞書を与えたモデルで正解した文の例を表3に示す。各例は、区切り位置により文意が異なる表現について、提案手法が正解した例である。1例目はKC, 2例目はKWDLC, 3例目はFKCの文である。1例目では「ヒゲもじゃ」、2例目では「わさび」、3例目では「むだ」が辞書引きされ改善した。

森羅の影響 森羅を与えたモデルは辞書なしのモデルと近いが若干低い精度を示し、JUMAN 辞書と森羅を与えたモデルはJUMAN 辞書を与えたモデルと近いが若干低い精度を示した。森羅のエントリの区切りが評価用データセットのJUMAN 体系と異なる

ため、精度の向上に寄与しなかったと考えられる。一方で、森羅に収録されている長い複合名詞の始まり・終わりは分かち書きの精度向上に寄与する可能性があるため、辞書をフィルタリングすることで精度が向上する可能性がある。また、森羅によるENEの分類など、詳細な分類や分野に特化した追加情報を形態素解析器が出力できれば下流タスクにとって有益であると思われる。

既存の形態素解析器との比較 Juman++と比較すると、提案手法はKCの品詞分類と品詞細分類、そしてFKCの分かち書きにおいて同等か上回る精度を示した。一方で、KWDLCにおいてはJuman++がすべての指標について提案手法の精度を上回った。ただし、本実験で用いたJuman++は公開版であり、評価用に用いたテストデータを学習している。KWJAと比較すると、KWDLCのP1, P2, FKCのSeg, P1, P2においてKWJAの精度が上回り、その他では提案手法の精度が上回った。ただし、KWJAは構文解析などの他のタスクとマルチタスク学習を行っており、形態素解析に最適化されていないため、若干精度が低い。また、KWJAは本実験で評価に用いたFKCの一部を学習に用いている。

5 おわりに

本研究では、日本語の事前学習モデルに入力文に加えて辞書の情報を与えて形態素解析を行い精度を評価した。結果として、辞書情報を与えることで分かち書きだけでなく品詞・活用付与の精度も向上することを確認した。今後は、どのような辞書であれば形態素解析の精度向上につながるのかを明らかにしたい。

謝辞

本研究は JSPS 科研費 JP21H04901 の助成を受けて実施した。

参考文献

- [1] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to Japanese morphological analysis. In **Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing**, 2004.
- [2] Daisuke Kawahara Hajime Morita and Sadao Kurohashi. Morphological analysis for unsegmented languages using recurrent neural network language model. In **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**, p. 2292–2297, 2015.
- [3] 植田暢大, 大村和正, 児玉貴志, 清丸寛一, 村脇有吾, 河原大輔, 黒橋禎夫. Kwja : 汎用言語モデルに基づく日本語解析器. 第 253 回自然言語処理研究会, 京都, 2022.
- [4] Andrew McCallum John Lafferty and Fernando Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In **Proceedings of the Eighteenth International Conference on Machine Learning**, Vol. 30, pp. 282–289, 2001.
- [5] Graham Neubig, Yosuke Nakata, and Shinsuke Mori. Pointwise prediction for robust, adaptable Japanese morphological analysis. In **Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies**, pp. 529–533, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.
- [6] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. arXiv, 2019. abs/1907.11692.
- [7] Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. Shrinking Japanese morphological analyzers with neural networks and semi-supervised learning. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 2744–2755, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [8] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. **Neural Computation**, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, L ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 30. Curran Associates, Inc., 2017.
- [10] Yu Sun, Shuohuan Wang, Shikun Feng, Siyu Ding, Chao Pang, Junyuan Shang, Jiayang Liu, Xuyi Chen, Yanbin Zhao, Yuxiang Lu, et al. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. arXiv, 2021. abs/2107.02137.
- [11] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc Le, and Ruslan Salakhutdinov. Transformer-XL: Attentive language models beyond a fixed-length context. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 2978–2988, Florence, Italy, July 2019. Association for Computational Linguistics.
- [12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [13] Anne Lauscher, Ivan Vulić, Edoardo Maria Ponti, Anna Korhonen, and Goran Glavaš. Specializing unsupervised pretraining models for word-level semantic similarity. In **Proceedings of the 28th International Conference on Computational Linguistics**, pp. 1371–1383, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.
- [14] Yoav Levine, Barak Lenz, Or Dagan, Ori Ram, Dan Padnos, Or Sharir, Shai Shalev-Shwartz, Amnon Shashua, and Yoav Shoham. SenseBERT: Driving some sense into BERT. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 4656–4667, Online, July 2020. Association for Computational Linguistics.
- [15] Zhe Zhao Zhiruo Wang Qi Ju Haotang Deng Ping Wang Weijie Liu, Peng Zhou. K-BERT: Enabling language representation with knowledge graph. In **Proceedings of AACL 2020**, 2020.
- [16] Satoshi Sekine. Extended named entity ontology with attribute information. **LREC**, 2008.
- [17] Sadao Kurohashi and Makoto Nagao. Building a Japanese parsed corpus. In **Treebanks: Building and Using Parsed Corpora, Text, Speech and Language Technology**, p. 249–260. Springer Netherlands, Dordrecht, 2003.
- [18] Masatsugu Hangyo, Daisuke Kawahara, and Sadao Kurohashi. Building a diverse document leads corpus annotated with semantic relations. In **Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation**, pp. 535–544, 2012.
- [19] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 8440–8451, Online, July 2020. Association for Computational Linguistics.
- [20] Guillaume Wenzek, Marie-Anne Lachaux, Alexis Conneau, Vishrav Chaudhary, Francisco Guzmán, Armand Joulin, and Edouard Grave. CCNet: Extracting high quality monolingual datasets from web crawl data. In **Proceedings of the Twelfth Language Resources and Evaluation Conference**, pp. 4003–4012, Marseille, France, May 2020. European Language Resources Association.

A 学習時のハイパーパラメータ

表 4 に, 学習を行った際のハイパーパラメータを掲載する.

表 4 学習時のハイパーパラメータ

Name	Value
learning rate (BASE)	5e-5
learning rate (LARGE)	3e-5
epoch	3
batch	16
seed	13, 35, 42

B 異なる学習データセットを用いた場合の精度

表 5 に, 本文中に掲載していない KC+KW で学習を行った際の精度を掲載する.

表 5 KC+KW を用いた分かち書きと品詞・活用付与の精度 (F1). KWJA の FKJ での精度と Juman++ の精度以外について, 表中で最も高い精度を示したモデルの値を太字で示した.

Analyzer	KC (新聞)					KWDLC (Web)					FKJ (ユーザー不満)					
	Seg	P1	P2	CT	CF	Seg	P1	P2	CT	CF	Seg	P1	P2	CT	CF	
Juman++ (2.0.0-rc3)	99.84	99.40	98.72	99.18	99.18	99.40	98.85	98.03	98.74	98.90	98.99	98.42	97.99	97.27	97.45	
KWJA	97.06	96.68	95.85	91.08	91.79	95.92	95.47	94.29	90.71	91.17	93.27	92.15	91.56	83.99	83.83	
KWJA (タイポ修正文除去)	98.94	98.90	98.11	95.93	96.66	98.34	98.17	97.01	96.36	96.87	99.26	98.88	98.36	96.66	96.64	
BASE	辞書なし	99.16	99.06	98.45	97.73	97.88	98.62	98.43	97.37	97.51	97.83	98.50	97.82	97.21	95.88	95.94
	JUMAN	99.36	99.24	98.58	98.06	98.23	98.85	98.55	97.46	97.62	97.93	98.83	98.03	97.38	96.14	96.38
	ENE	97.79	97.96	97.10	96.01	96.25	98.55	98.36	97.25	97.44	97.75	98.43	97.75	97.02	95.80	96.02
	JUMAN+ENE	97.43	97.63	96.63	95.60	95.80	98.79	98.49	97.33	97.45	97.73	98.78	97.98	97.33	96.06	96.35
	KWJA	97.42	97.33	96.31	94.51	94.94	98.06	97.99	97.15	94.13	95.05	95.70	95.06	94.67	90.15	90.43
	KWJA (タイポ修正文除去)	98.08	98.07	97.06	96.37	96.82	98.76	98.83	97.97	96.19	97.17	98.08	98.07	97.06	96.37	96.82
LARGE	辞書なし	99.21	99.16	98.60	97.93	98.01	98.73	98.52	97.56	97.57	97.86	98.72	98.03	97.43	96.07	96.46
	JUMAN	99.41	99.32	98.78	98.24	98.38	98.92	98.63	97.68	97.71	97.97	98.84	98.14	97.52	96.19	96.56
	ENE	98.76	98.76	98.08	97.36	97.58	98.72	98.51	97.46	97.56	97.83	98.70	98.01	97.37	96.14	96.38
	JUMAN+ENE	98.81	98.78	98.11	97.44	97.53	98.90	98.63	97.61	97.82	98.08	98.97	98.26	97.56	96.52	96.79