DEIM Forum 2022 E31-2

コンテキストデータセットを用いたコンテキスト検索・ノウハウ読解

白　書霆† 李　廷軒† 朱　福主† 宇津呂武仁†

† 筑波大学大学院　システム情報工学研究群　知能機能システム学位プログラム
〒 305-8573 茨城県つくば市天王台 1-1-1

あらまし　本論文では，非事実型機械読解においてノウハウを対象とした機械読解に注目する．インターネット上のノウハウサイトを情報源として用いたコンテキストデータセットの構築方法を提案し，提案手法を用いることで数千規模のコンテキスト集合を含むコンテキストデータセットを容易に作成できることを示す．さらに，構築したコンテキストデータセットに対して，質問の回答が含まれるコンテキストを検索する手法，および，機械読解を行う手法を適用する．具体的には，TF-IDF 法，および，BERT モデルに基づく 3 種類の検索モデルと，BERT 機械読解モデルを併用したコンテキスト検索・機械読解のアプローチをコンテキストデータセットに適用し，評価実験を行った結果について述べる．

キーワード　機械読解，ノウハウ，TF-IDF，BERT

1 はじめに

自然言語処理分野における機械読解タスクは，自然言語で記述された質問文とコンテキストに対して，コンテキスト中から質問の回答を抽出するタスクとして定式化される．機械読解タスクは，回答対象の種類によって，固有名詞や数量などの事実を回答対象とした事実型機械読解，および，物事のやり方や理由などの非事実を回答対象とした非事実型機械読解の二種類に大別される．事実型機械読解においては，英語版 Wikipedia の記事から作成した機械読解タスク用データセットである SQuAD [10] に対して，最新の深層学習による機械読解モデル (例えば，BERT [3]) は，人間による読解の性能を上回ることが知られている [1]．一方，非事実型機械読解においても，いくつかの研究事例が知られているが，ノウハウを対象とした機械読解の研究事例 [2] においては，事実型機械読解モデルにおいて用いられているモデルを用いることにより，ノウハウ機械読解においても一定の性能が達成できることが報告されている．この事例では，インターネット上でノウハウが掲載されているウェブサイト (本論文ではノウハウサイト [7] と呼ぶ) を選定し，その中のコラムページを情報源として作成した訓練・評価事例を用いることにより，一定の性能を持つノウハウ機械読解モデルが訓練できることを示している．

ノウハウ読解モデルの枠組みを図 1 に示す．この図に示すように，ノウハウ読解モデルに限らず，一般に読解モデルの枠組みは，質問文とコンテキストに対して，コンテキスト中から質問の回答を抽出するタスクとして定式化される．そのため，任意の質問に対して回答を与えるためには，読解モデルとは別の枠組みによって，回答を含む可能性があるコンテキストの候補集合を収集し，それらのコンテキスト候補集合に対して読解モデルを適用することによって回答候補を得る，という過程が必要となる．そこで，文献 [1] によって，情報検索と機械読解の二つのタスクを同時に扱う大規模機械読解タスクが提案された．文献 [1] の大規模機械読解タスクの定式化においては，情報検索部分のモデル化においては TF-IDF 法によってコンテキスト候補を収集し，収集されたコンテキスト候補に対して機械読解モデルを適用するアプローチが採用された．また，文献 [5] では，大規模機械読解タスクにおける検索部分を対象として，BERT [3] を用いた検索モデルが提案され，複数の事実型機械読解データセットにおいて BM25 法より高い検索精度を達成できた．そして，提案された検索モデルと BM25 スコアを併用することにより，検索精度がさらに改善されたことが報告された．

これらの背景をふまえて，本論文では，任意のノウハウに関する質問に対して回答を与えるための枠組みを構築することを目的とする．そのための手段として，図 2 に示す枠組みのもとで，TF-IDF 法によるコンテキスト検索 [1]，BERT 検索モデル，TF-IDF 法と BERT 検索モデルの併用の 3 種類の検索モデル，および，ノウハウ機械読解モデル [2], [8] を併用するモデル化によって，ノウハウ機械読解タスクをコンテキスト検索・機械読解タスク化する．文献 [2], [8] においては，インターネット上のノウハウサイトから収集したコラムページを情報源としてノウハウ機械読解モデル [2], [8] の訓練・評価事例を作成している．これをふまえて，本論文では，図 3 の例に示すように，文献 [2], [8] においてインターネット上のノウハウサイトから収集したコラムページ中の段落のうち，ノウハウ機械読解モデルの訓練・評価事例作成には用いられなかった段落を全て収集し，コンテキスト検索・機械読解タスクにおいて検索されるコンテキスト C' (検索されるだけのコンテキストであり，評価用の質問回答事例として，質問に対する回答を含むコンテキストとし

1：https://rajpurkar.github.io/SQuAD-explorer/

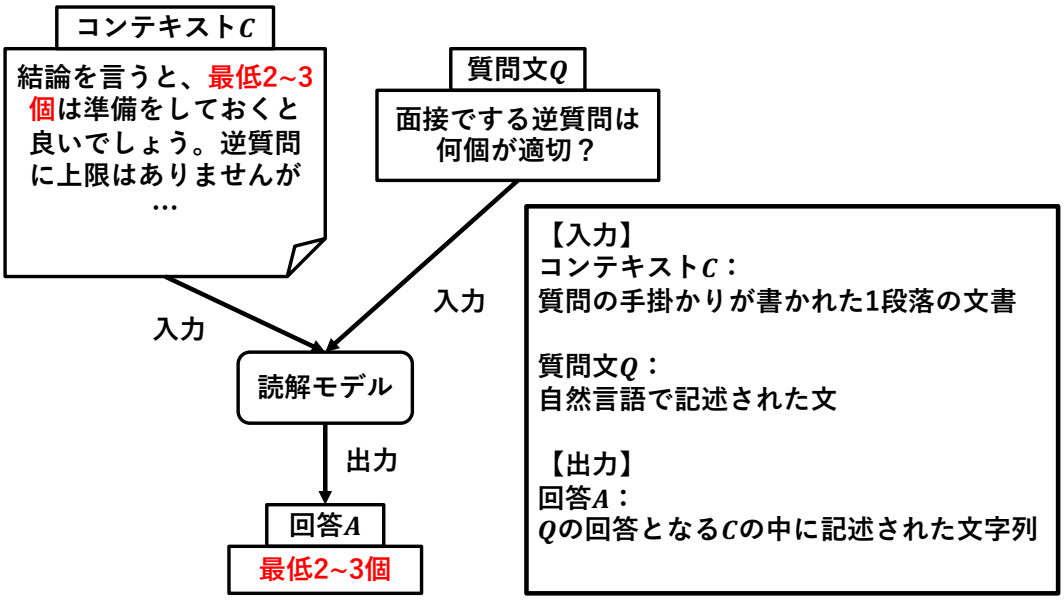


図 1　ノウハウ読解モデルの枠組み

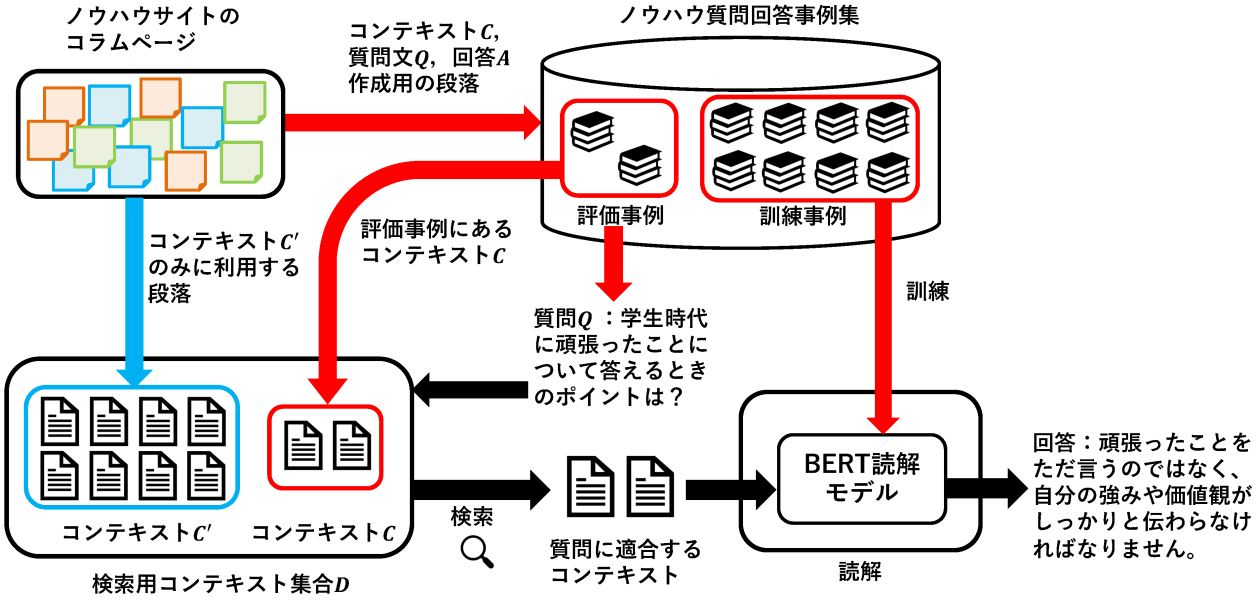


図 2　ウェブ上のコラムページを情報源とするノウハウ読解用コンテキストデータセットの作成

て事前に設定することはしない [2]) の情報源として用いる．本論文では，以上の手順を経て，検索用コンテキスト C' として数千規模のコンテキスト集合を含むノウハウ機械読解用コンテキストデータセットを作成した結果について述べる．さらに，図 2 に示すように，ノウハウ機械読解モデルの評価事例，および，検索用コンテキスト集合によって構成されるコンテキストデータセットの和集合を対象としてコンテキスト検索・ノウハウ機械読解の評価実験を行った結果について述べる．

2　ノウハウ機械読解用コンテキストデータセット

本節では，検索用コンテキスト C' の収集，および，コンテキスト検索・ノウハウ機械読解における検索用コンテキストデータセットの作成方法について述べる．

文献 [2], [8] では，「就職活動」，「結婚」，「マンション」，「花粉症」，「虫歯」，「食中毒」の六つのクエリ・フォーカスを対象として，ノウハウサイトを選定し，その中のコラムページを収集した．そして，収集した各ページから最大五段落を選定して回答可能，および，回答不可能なノウハウ質問回答事例作成用のコンテキストとして使用し，回答可能なノウハウ質問回答事例 (コンテキスト C，質問文 Q，回答 A の組)，および，回答不可能なノウハウ質問回答事例 (コンテキスト C，質問文 Q，回答 A' =「」(空白) の組) を作成した．そこで，本節では，文献 [2], [8] において使用した各ウェブページ中で，文献 [2], [8] で選定された最大五段落に含まれず，かつ，以下の基準をみたす段落を検索用コンテキスト C' として収集する．

2：ただし，評価実験の結果において，評価用質問回答事例に対する回答として適切な回答が偶然出力されることは起こり得る．

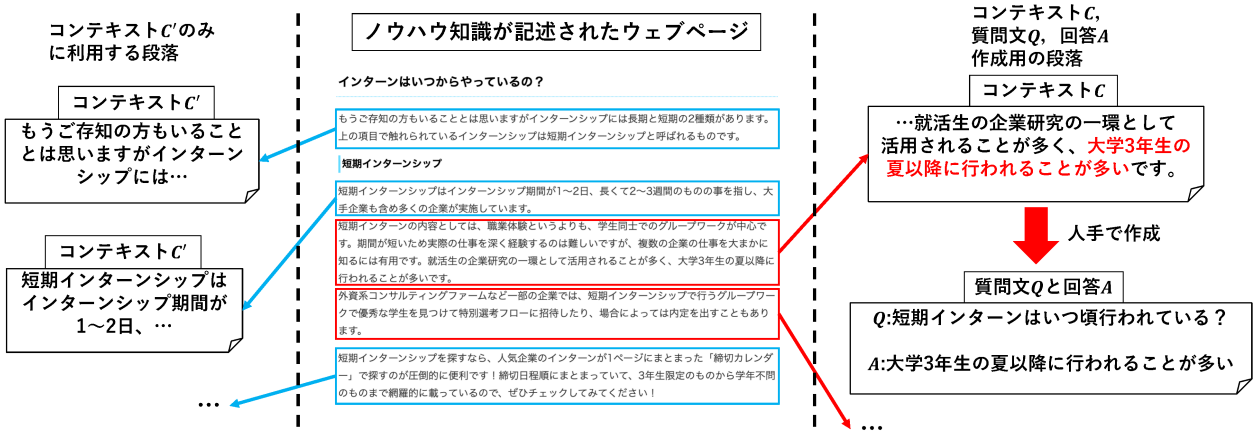


図 3　コラムページを情報源とするコンテキストの収集およびノウハウ質問回答事例の作成（出典：「就活はいつから始まる？選考解禁の時期と 21 卒・22 卒学生の動き方」（https://internshipguide.jp/columns/view/shukatsu_sched_1））

(i)　BERT [3] による機械読解モデルを適用する際の制約をふまえて，段落内の形態素数の上限を 290 とする [3]．

(ii)　段落内の文字数の下限を 30 文字とする．

(iii)　段落内に URL が含まれない．

(iv)　段落内にメールアドレスが含まれない．

(ii) では，文献 [2], [8] において作成されたノウハウ質問回答事例のコンテキスト C の平均文字数が 106.2，標準偏差が 77.6 であることをふまえて，コラムページにおいて文字数の少ない段落は，文章の流れを作るための段落であることが多く，ノウハウに関する知識が含まれることが少ないことを確認した結果として，段落内の文字数の下限を設定した．(iii) および (iv) では，URL，あるいは，メールアドレスが含まれている段落は，その URL 先のページへの誘導や連絡先の情報を提示するための段落である場合が多いため，除外対象とする．

「就職活動」,「結婚」,「マンション」,「花粉症」,「虫歯」,「食中毒」の六つのクエリ・フォーカスの各々について，使用したウェブページ数，および，ノウハウ質問応答回答事例作成用コンテキスト数・検索のみに利用するコンテキスト数を表 1 に示す．一つのコラムページに対するコンテキスト収集の例を図 3 に示す．以上の手順を経て，図 2 に示すように，検索用コンテキスト C'，および，ノウハウ質問回答事例の評価事例中のコンテキスト C を混合し，検索用コンテキスト集合 D とする．

3　BERT 検索モデル

本節では，文献 [5] にもとづき構築した BERT 検索モデルの構造，訓練の方法，および，検索手順について述べる．

この BERT 検索モデルでは，質問のエンコーダ E_q，および，コンテキストのエンコーダ E_c として独立した 2 つの BERT [3][4] モデルを利用する．入力された質問 Q とコンテキスト C それぞれに BERT [3] モデルを適用し，出力された CLS トークンの分散表現を質問 Q とコンテキスト C の分散表現とする．エンコードされた Q と C の類似度尺度としては，次式のコサイン類似度を用いる．

$$sim(Q,C) = \frac{E_q(Q) \cdot E_c(C)}{\| E_q(Q) \| \, \| E_c(C) \|} \tag{1}$$

モデルの訓練においては，1 個の質問 Q，正解が含まれる 1 個のコンテキスト C^+，および，正解が含まれない n 個のコンテキスト C^- の組み合わせを 1 組の訓練データ

$$(Q_i, C_i^+, C_{i,1}^-, \cdots, C_{i,n}^-) \tag{2}$$

として用意し，同様の組み合わせを m 組集めたものを訓練データの集合 T とする．

$$T = \{(Q_i, C_i^+, C_{i,1}^-, \cdots, C_{i,n}^-) \mid i = 1, \ldots, m\} \tag{3}$$

訓練時における 1 組の訓練データに対する損失関数としては次式を用いる．

$$L(Q_i, C_i^+, C_{i,1}^-, \cdots, C_{i,n}^-) = -log \frac{e^{sim(Q_i, C_i^+)}}{e^{sim(Q_i, C_i^+)} + \sum_{j=1}^{n} (e^{sim(Q_i, C_{i,j}^-)})} \tag{4}$$

また，質問 Q と正解が含まれるコンテキスト C の組から上述の訓練データの組を作成するために，「in-batch negatives」の仕様を導入する．この仕様においては，バッチサイズが B であるミニバッチ内の 1 個の質問 Q_i に対して，同一のミニバッチにある他の $B-1$ 個の QC ペアの中のコンテキストを，正解が含まれないコンテキスト C_i^- とする．その結果，B 組の訓練データが作成され，各組の訓練データ中には，1 個の質問 Q_i，正解が含まれる 1 個のコンテキスト C_i^+，および，正解が含まれない $B-1$ 個のコンテキスト C_i^- が含まれる．

3：日本語形態素解析においては MeCab (https://taku910.github.io/mecab/)，および，mecab-ipadic-NEologd (https://github.com/neologd/mecab-ipadic-neologd) を用いた．

4：BERT 検索モデルの実装としては，HuggingFace 版 (https://github.com/huggingface/transformers) の BERT モデルを利用した．事前学習モデルとしては，多言語モデル (Multilingual Cased Model) を採用した．

検索時には，訓練した BERT モデルを用いて，事前に質問と検索対象となるすべてのコンテキストをエンコードする．1 つの質問に対して，(1) 式の類似度尺度のもとで上位 n 個のコンテキストを検索結果として出力する．検索部分の実装においては，近傍探索ライブラリ Faiss [4] を利用した．

4 評　　価

4.1 評 価 手 順

評価実験では，以下 3 種類の検索モデルを使用する．

(i) TF-IDF モデル

(ii) BERT 検索モデル

(iii) (i) と (ii) のスコアを併用した「TF-IDF+BERT」モデル

文献 [1] の TF-IDF モデル [5] に日本語ストップワードリスト SlothLib[6] のストップワードを追加し，「就職活動」，「結婚」，「マンション」，および，「花粉症・虫歯・食中毒」の各々の検索用コンテキスト集合に対して，個別に TF-IDF モデルを構築する．

また，文献 [2], [8] のノウハウ質問回答事例の中で，「就職活動」，および，「結婚」についての回答可能な訓練事例から，質問 Q と正解が含まれるコンテキスト C を抽出し，QC ペアの訓練データを作成する．使用したノウハウ質問回答事例の数を表 3(b) に示す．作成した訓練データを用いて，3 節で述べた BERT 検索モデルの訓練を行う．訓練時のハイパーパラメータの設定として，エポックを 20，バッチサイズを 32 に設定する．初期学習率を 10^{-5} に設定し，最後のエポックの終了時に 0 になるように線形的に減衰させる．

「TF-IDF+BERT」モデルにおいて，質問 Q とコンテキスト C の TF-IDF 特徴ベクトルの内積を TF-IDF モデルのスコア $S_T(Q,C)$ とし，BERT 検索モデルでエンコードした Q と C のコサイン類似度を BERT 検索モデルのスコア $S_B(Q,C)$ とする．1 個の質問 Q_i に対して，n 個のすべての検索候補コンテキストとのスコア $S_T(Q_i,C_j)(j=1,\cdots,n)$ を求め，最小値が 0，最大値が 1 になるように Min-Max 法で正規化する．1 個の検索候補コンテキスト C_j との「TF-IDF+BERT」モデルのスコア $S_{T+B}(Q_i,C_j)$ は次式となる．

$$S_{T+B}(Q_i,C_j) = S_T(Q_i,C_j)_{normalized} + S_B(Q_i,C_j) \qquad (5)$$

S_{T+B} に基づき，検索候補コンテキストのランキングを行い，上位 n 個のコンテキストを検索結果として出力する．

文献 [2], [8] のノウハウ質問回答事例，および，事実を回答対象とした機械読解データセット [11] を利用し，以下の 2 種類のデータを用いて BERT [3][7] の fine-tuning を行い，ノウハウ機械読解モデルを作成する [8]．

(i) 「就職活動」，および，「結婚」についてのノウハウに関する質問回答事例

(ii) 事実に関する質問回答事例，および，(i) の両方を用いる事実・ノウハウ混合質問回答事例

ノウハウ質問回答事例における質問の数，および，ノウハウ・事実に関する質問回答事例の数を，表 2，および，表 3 に示す．図 2 における検索用コンテキスト集合 D の検索の枠組みとしては，2 節で述べた MeCab，および，mecab-ipadic-NEologd によって分かち書きされたコンテキスト集合に対して，上述の 3 種類の検索モデルを適用する．検索モデルの評価方法として，検索結果上位 n 個 (本論文では，$n = 1,\ldots,10,20,50$ として，1 ～ n 位までを評価する) のコンテキストの中で，正解が含まれるコンテキストが存在する場合，検索成功と判定する．評価事例にあるすべての質問のうち，「検索成功した質問の割合」を算出する．そして，検索結果上位のコンテキストの各々に対してノウハウ機械読解モデルを適用し，モデルが出力した確信度の最も高い回答を選ぶ．最後に，参照用回答の形態素列に対して F1 スコアを算出する．人手評価においては，モデルの回答と参照用回答を対照し，「完全一致（EM），部分一致（PM），不一致」の 3 段階の評価基準に加え，「その他の回答（another answer, AA）」の評価基準も取り入れ，人手評価を行う．「その他の回答」の判定基準として，「参照用回答とは異なるが，質問の回答になり得るほどの情報量が含まれている」のような回答が該当する．そして，完全一致，部分一致，および，その他の回答の事例数の割合を算出する（EM+PM+AA）．

4.2 評 価 結 果

検索モデルの評価結果を図 4 に示す．クエリ・フォーカス「就職活動」においては，BERT 検索モデルの性能は TF-IDF モデルとの間で大きな差はないが，「就職活動」以外においては，BERT 検索モデルの性能が他の 2 種類のモデルより大幅に低いことが分かった．また，「マンション」以外のクエリ・フォーカスにおいては，「TF-IDF+BERT」モデルの性能が一番高く，「マンション」においても TF-IDF モデルと同等の性能を示した．BERT 検索モデルの性能が他の 2 種類のモデルより低いことの原因として，ノウハウ質問回答事例の作成において，先にノウハウサイトのコラムページからの段落をコンテキストとして人手で選定し，選定したコンテキストに対して質問を作成するという手順をとっていたため，コンテキスト中のキーワードを多く含む質問が作成される傾向にあることが挙げられる [9]．そのため，質問とコンテキストの意味の特徴に注目する BERT 検索モデルよりも，キーワードの特徴に注目する TF-IDF モデルの方が全体的に高い性能を達成した．そして，TF-IDF モデルと BERT 検索モデルを併用することにより，コンテキスト中

5：https://github.com/facebookresearch/DrQA

6：日本語ストップワードリスト（http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt）

7：読解モデルにおいては, TensorFlow 版（https://github.com/google-research/bert）の BERT モデルを利用した．事前学習モデルとしては, 多言語モデル（Multilingual Cased Model）を採用した．

8：ハイパーパラメータの設定として，エポックを 20，バッチサイズを 8，学習率を 0.00003 に設定した．

9：質問文に含まれるストップワードと助詞以外の形態素のうち，回答が含まれるコンテキスト中に出現する形態素の割合の平均値は，「就職活動」が 0.62，「結婚」が 0.72，「マンション」が 0.74，「花粉症・虫歯・食中毒」が 0.64 であった．

表 1　使用したウェブページ数および収集したコンテキスト数

クエリ・フォーカス	使用したウェブページ数	ノウハウ質問回答事例作成用コンテキスト数		検索のみに利用するコンテキスト数
		訓練事例	評価事例	
就職活動	293	1,478	98	4,675
結婚	182	1,386	98	2,868
マンション	50	—	100	491
花粉症・虫歯・食中毒	51	—	100	962

表 2　ノウハウに関する質問数

クエリ・フォーカス	訓練事例作成用	評価事例作成用
就職活動	795	50
結婚	799	49
マンション	—	50
花粉症・虫歯・食中毒	—	49

表 3　質問回答事例数

(a) 事実に関する質問回答事例

訓練・評価	コンテキスト，質問文，回答の組数 (回答可能/回答不可能)
訓練	27,427/28,742
評価	50/50

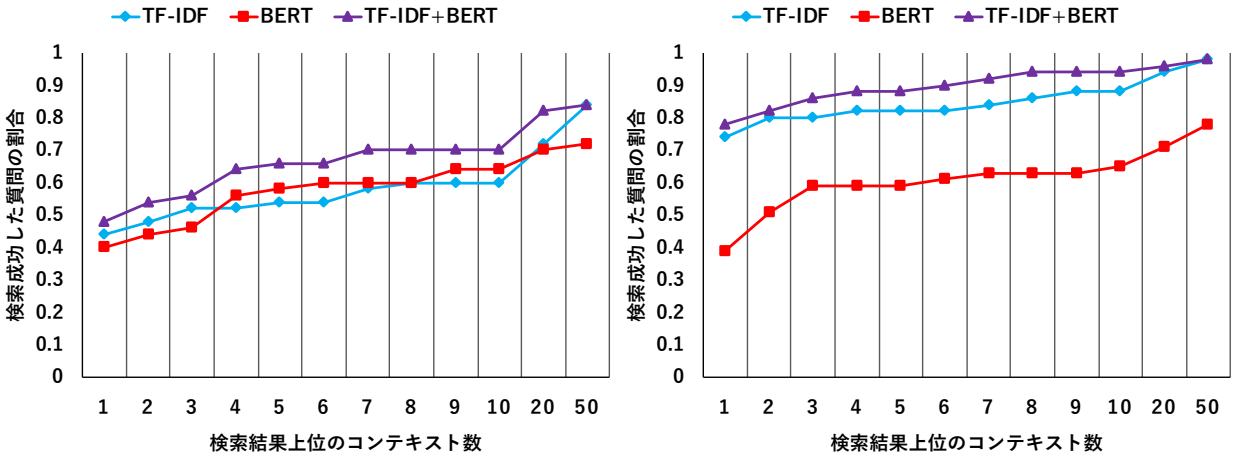
(b) ノウハウに関する質問回答事例

クエリ・フォーカス	コンテキスト，質問文，回答の組数 (回答可能/回答不可能)	
	訓練事例	評価事例
就職活動	807/807	50/50
結婚	807/807	50/50
マンション	—	50/50
花粉症・虫歯・食中毒	—	50/50

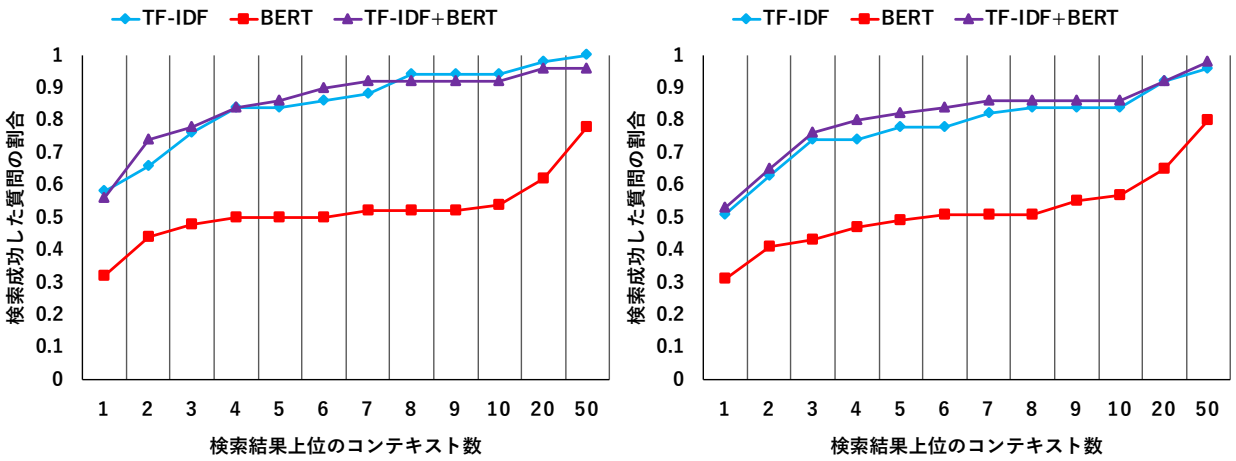
のキーワードを多く含まない質問に対しても検索性能が向上し，結果的に「TF-IDF+BERT」モデルの全体的な性能が最も高くなった．「就職活動」において BERT 検索モデルの性能が良かった原因としては，「就職活動」の質問においては，コンテキスト中のキーワードが含まれる割合が他のクエリ・フォーカスよりも少ないことが挙げられる．

3 種類の検索モデルによる検索結果に対して，ノウハウ質問回答事例，および，事実・ノウハウ混合質問回答事例で訓練した読解モデルを用いた場合の読解の自動評価結果を図 5，および，図 6 に示す．検索性能の影響を受けて，「就職活動」以外のクエリ・フォーカスにおいては，BERT 検索モデルを用いた場合の性能が一番悪かった．また，TF-IDF モデル，および，「TF-IDF+BERT」モデルにおいては，クエリ・フォーカスごとのデータセットにも依存するが，検索結果上位のコンテキスト数の増加に伴い，性能が下がる傾向にあることが分かる．検索結果上位のコンテキスト数の増加に伴い再現率は上がるが，最良解を選択する際に検索順位を考慮していないため，適合率は下がり，両者が相殺し，適合率が下がる要因の方が勝ることが原因である．また，文献 [2], [8] における評価結果と同様に，読解モデルの訓練事例として用いられていないクエリ・フォーカスである「マンション」，および，「花粉症・虫歯・食中毒」においても，「就職活動」，および，「結婚」と比較して一定以上の性能が得られていることから，コンテキスト検索・ノウハウ機械読解においても，異なる話題の間で読解モデルの横断的適用がある程度可能であることが分かった．

ノウハウ質問回答事例で訓練した読解モデルを用いた場合の読解の人手評価結果を図 7 に示す．この結果においては，「就職活動」，および，「結婚」の両方から総合的に判断して，「TF-IDF+BERT」モデルの検索性能が相対的に最も高いことが分かる．



(a) 就職活動　(b) 結婚



(c) マンション　(d) 花粉症・虫歯・食中毒

図 4　3 種類の検索モデルの評価結果

5　関 連 研 究

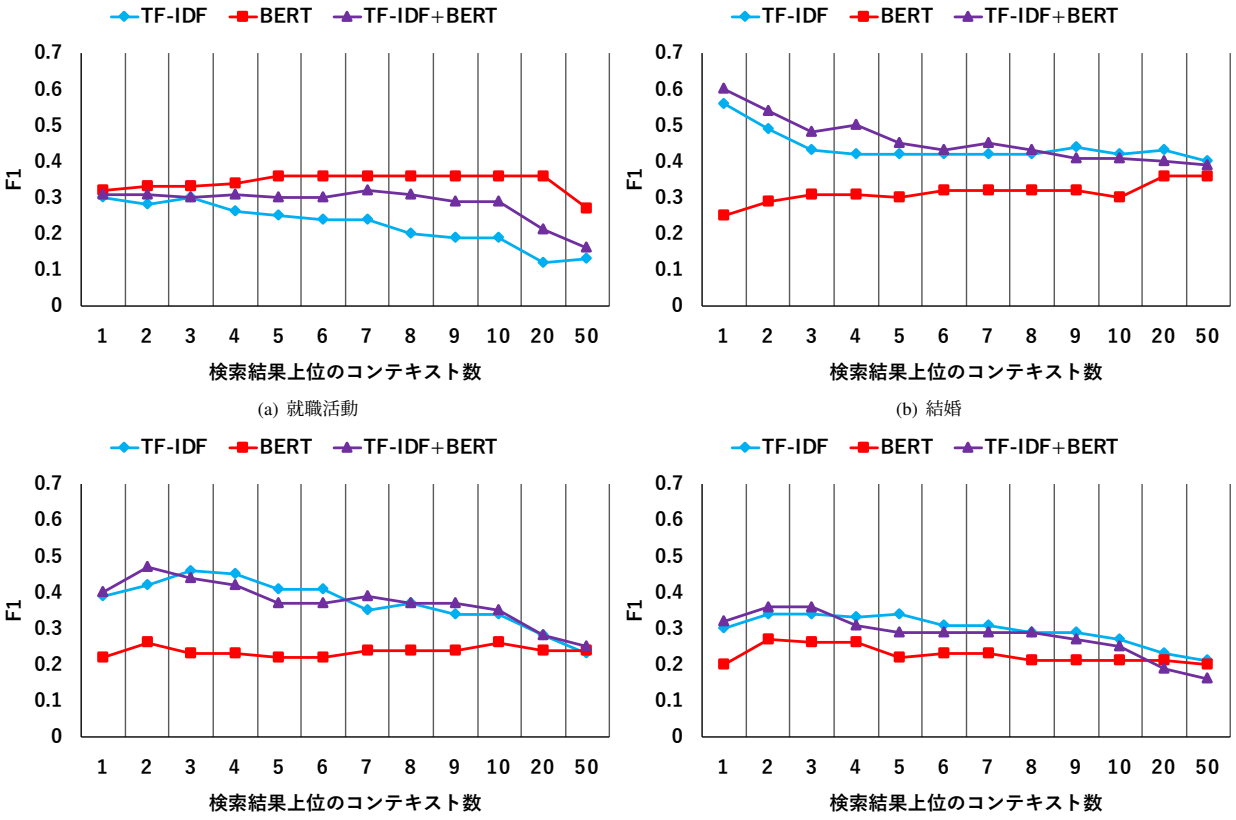
本論文に関連して，大規模機械読解の関連研究としては，事実型機械読解を対象とする研究事例 [1], [5], [6], [9] が挙げられる．文献 [1] においては，TF-IDF 法によるコンテキスト検索の後，RNN を用いたニューラル読解モデルを適用することにより大規模機械読解を実現した．それに対して，文献 [5] においては，BERT [3] を用いた検索モデルを適用した後，ニューラル読解モデルを適用することにより大規模機械読解システムを構築した．また，文献 [9] においては，情報検索と機械読解のマルチタスク学習により大規模機械読解を実現している．その他，検索モデルおよび読解モデルを一つのモデルとして訓練する大規模機械読解モデル [6] も提案されている．これに対して，本論文では，ノウハウ機械読解のタスクにおいて，文献 [1], [5] に基づき，TF-IDF 法によるコンテキスト検索，BERT 検索モデル，TF-IDF 法と BERT 検索モデルの併用の 3 種類の検索モデルを適用した後，BERT [3] を用いたニューラル読解モデルを適用することにより，コンテキスト検索・機械読解を実現した．また，文献 [5] に比べて，本研究では評価実験を通じて検索結果上位のコンテキスト数の増加による機械読解の性能の変化を明らかにした．

6　お わ り に

本論文では，文献 [2], [8] においてインターネット上のノウハウサイトから収集したコラムページ中の段落のうち，ノウハウ機械読解モデルの訓練・評価事例作成には用いられなかった段落を全て収集し，コンテキスト検索・機械読解タスクにおいて検索されるコンテキストの情報源として用いる方式を提案した．そして，数千規模のコンテキスト集合を含むノウハウ機械読解用コンテキストデータセットを作成した結果について述べた．また，ノウハウ機械読解モデルの訓練・評価事例，および，検索用コンテキスト集合によって構成されるコンテキストデータセットの和集合を対象としてコンテキスト検索・ノウハウ機械読解の評価実験を行った結果について述べた．今後の課題としては，訓練事例の追加，構造の改良などによる BERT 検索モデルの性能向上，他のスコア計算式の導入による「TF-IDF+BERT」モデルの性能向上，および，複数の回答候補の中から，検索順位を考慮して最良解を選択する手法を導入することが挙げられる．

謝辞

本研究は科研費 19H04417 の助成を受けたものである．



(a) 就職活動　(b) 結婚

(c) マンション　(d) 花粉症・虫歯・食中毒

図 5　「3 種類の検索モデル＋ノウハウ質問回答事例で訓練した読解モデル」の読解の自動評価結果

文　献

[1] D. Chen, A. Fisch, J. Weston, and A. Bordes. Reading Wikipedia to answer open-domain questions. In *Proc. 55th ACL*, pp. 1870–1879, 2017.

[2] 陳騰揚, 前田竜治, 李宏宇, 銭澤長, 宇津呂武仁, 河田容英. ウェブ上のコラムページを情報源とする回答不可能なノウハウ質問応答事例の作成. 言語処理学会第 26 回年次大会論文集, pp. 315–318, 2020.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional Transformers for language understanding. In *Proc. NAACL-HLT*, pp. 4171–4186, 2019.

[4] J. Johnson, M. Douze, and H. Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, Vol. 7, No. 3, pp. 535–547, 2021.

[5] V. Karpukhin, B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W. Yih. Dense passage retrieval for open-domain question answering. In *Proc. EMNLP*, pp. 6769–6781, 2020.

[6] K. Lee, M.-W. Chang, and K. Toutanova. Latent retrieval for weakly supervised open domain question answering. In *Proc. 57th ACL*, pp. 6086–6096, 2019.

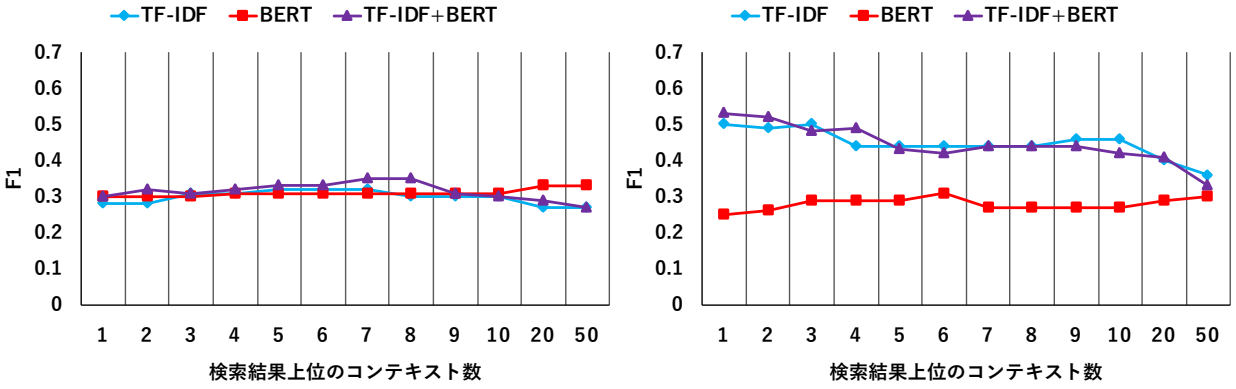
[7] 李佳奇, 趙辰, 林友超, 丁易, 川畑修人, 宇津呂武仁, 河田容英. トピックモデルおよび分類器学習を用いたノウハウサイトの同定. 第 10 回 DEIM フォーラム論文集, 2018.

[8] 李廷軒, 白書霆, 鈴木勢至, 宇津呂武仁, 河田容英. コミュニティ QA サイト上の質問回答事例に対するノウハウ読解. 第 35 回人工知能学会全国大会論文集, 2021.

[9] 西田京介, 斉藤いつみ, 大塚淳史, 浅野久子, 富田準二. 情報検索とのマルチタスク学習による大規模機械読解. 言語処理学会第 24 回年次大会論文集, pp. 963–966, 2018.

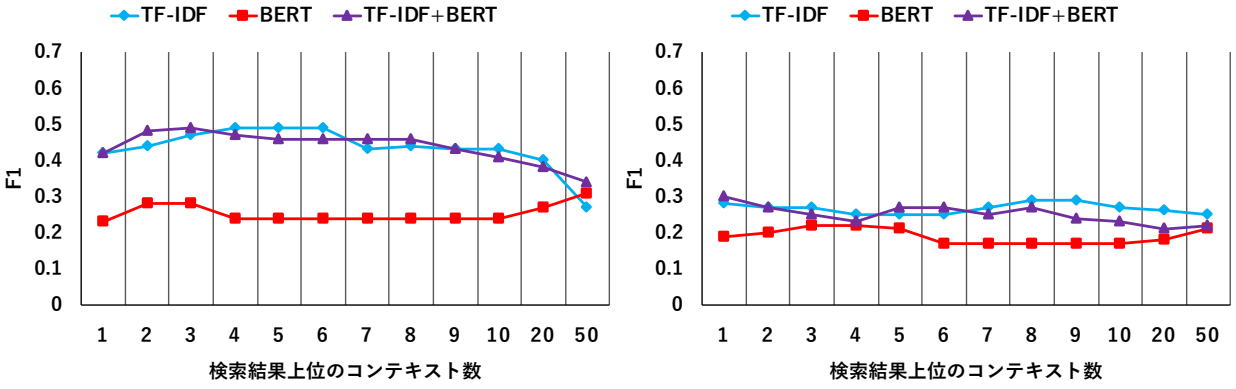
[10] R. Pranav, Z. Jian, L. Konstantin, and L. Percy. SQuAD: 100,000+ questions for machine comprehension of text. In *Proc. EMNLP*, pp. 2383–2392, 2016.

[11] 鈴木正敏, 松田耕史, 岡崎直観, 乾健太郎. 読解による解答可能性を付与した質問応答データセットの構築. 言語処理学会第 24 回年次大会論文集, pp. 702–705, 2018.



(a) 就職活動

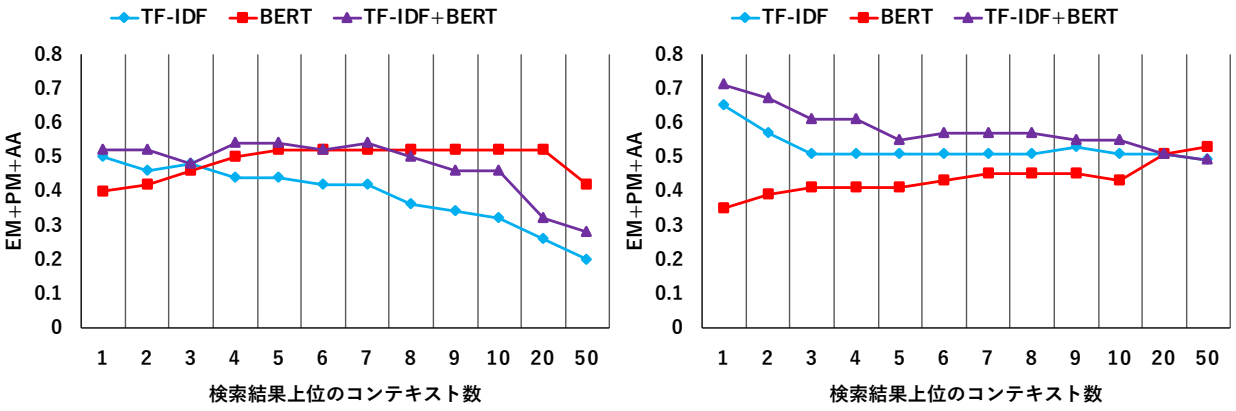
(b) 結婚



(c) マンション

(d) 花粉症・虫歯・食中毒

図 6　「3 種類の検索モデル＋事実・ノウハウ混合質問回答事例で訓練した読解モデル」の読解の自動評価結果



(a) 就職活動

(b) 結婚

図 7　「3 種類の検索モデル＋ノウハウ質問回答事例で訓練した読解モデル」の読解の人手評価結果