

生活保護業務支援のための質問応答システムの開発と評価

Development and Evaluation of a Question Answering System to Support Public Assistance Work

堂坂 浩二[†] 金子 和樹[†] 木村 幸司[‡] 伊東 嗣功[†] 石井 雅樹[†]
 Kohji Dohsaka Kazuki Kaneko Koji Kimura Hidekatsu Ito Masaki Ishii

1. はじめに

生活保護とは、資産や能力のすべてを活用しても生活に困窮する人に対し、国が経済的な援助を行う制度である。地域の安心した暮らしの実現のためには、被保護者の生活を再建し、労働市場への再挑戦を可能とする生活保護制度の安定運用が重要な役割を果たす。

保護申請者に対応し業務を行うのが生活保護ケースワーカーである。現在ケースワーカーの4分の1が1年未満、3分の2が3年未満の経験しか持たず、法律によりケースワーカーに従事するために必要とされている社会福祉主事資格の取得率が82%程度と、経験や能力が不足した人材が多いと言える[1]。これに加え、生活保護を受給している被保護人員数は200万人を超える高い値を維持しているが、生活保護世帯数あたりに定められたケースワーカーの標準配置人数における配置充足率は90%と足りていない[2]。このため、ケースワーカー一人当たりの生活保護業務の業務負担が非常に高く、ICT技術によるケースワーカーの業務支援の必要性が高まっている。特に、生活保護に関する手引き書である生活保護手帳[3]ならびに生活保護別冊問答集[4] (以下、簡単のため、まとめて生活保護手帳と呼ぶ) に記載された膨大な法令・通知から個別の業務に関連した部分を検索する機能の要望が高い[5]。

そこで、我々は、生活保護ケースワーカーが業務に関する質問を日本語で入力したときに、質問の回答の根拠となる生活保護手帳の該当箇所を出力する質問応答システムの開発に取り組んでいる。こうした質問応答システムを実現することにより、これまで熟練者の指導や膨大な紙の業務文書を頼りに業務を遂行してきたケースワーカーの業務効率の向上及び負担軽減につながることを期待できる。

2節で詳しく説明するように、質問応答システムは、質問文と応答文の対から成る質問応答データベースと、システムへの入力質問文に対して応答文を提示する質問応答アルゴリズムから構成される。本稿においては、このシステム構成において、質問応答データベースをいかに構築するか、質問応答データベース中の質問文やシステムへの入力質問文をいかに表現するかという2つの研究課題に焦点をあて、我々が開発している質問応答システムの構成法とシステム性能を評価した結果について報告する。

以下において、2節で質問応答システムの構成法と、データベースの構築法について述べる。3節において、質問文の表現方法として、自然言語処理における、word2vec [6]、fastText [7]、BERT [8]、Sentence BERT [9] という4種類の分散表現モデルを使った表現方法について説明する。4節では、4種類の分散表現モデルのそれぞれを使った場合の質問応答システムの性能を比較する。さらに、データ

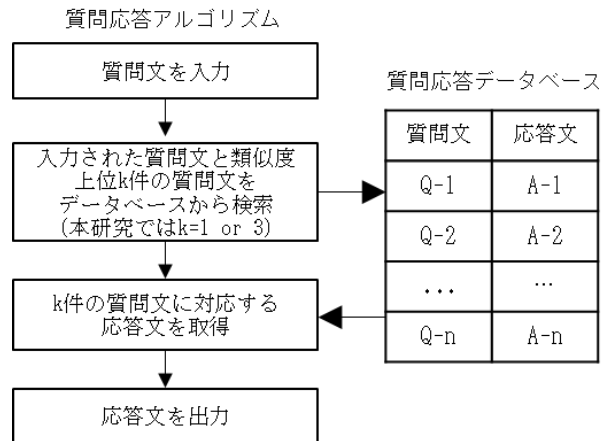


図 1 質問応答システムの構成

ベースの拡張がシステムの性能に及ぼす影響について論じる。最後に5節で本稿での報告をまとめる。

2. 質問応答システム

生活保護手帳は、生活保護の実施に関する法令・通知を集めたものであり、ケースワーカーは、保護の決定、実施の根拠を調べるため、知識源として生活保護手帳を参照する。質問応答システムは、生活保護手帳に記述された知識を効率的に引き出す仕組みと考えられる。生活保護手帳の中には、知識が問答の形式で表現されている部分があり、本研究ではその部分を問答と呼び、問答以外の部分を法令と呼ぶ。

図1に質問応答システムの構成を示す。質問応答システムは質問応答データベース(以降データベース)と質問応答アルゴリズムから構成される。データベースは質問文と応答文の対の集合である。質問文は生活保護業務に関する質問であり、応答文はその質問の回答の根拠となる生活保護手帳の問答あるいは法令である。図1では、質問文 $Q-i$ に応答文 $A-i$ が紐づいているが、実際には一つの応答文に複数の質問文が紐づけられる場合がある。

質問応答アルゴリズムは、システムに入力された質問文と類似した上位 k 件の質問文をデータベース中から検索し、それに紐づく応答文を出力する。本研究では k を1あるいは3としてシステムの性能を評価した。システムに入力される質問文、データベース中の質問文は、分散表現モデルにより数値ベクトルとして表現し、入力質問文とデータベース中の質問文との類似度はベクトルの \cos 類似度として計算する。

[†] 秋田県立大学 Akita Prefectural University

[‡] 北日本コンピューターサービス株式会社

KITANIHON COMPUTER SERVICE CO.,LTD

表 1: 質問応答データベースを構成する質問文の内訳

(A) 生活保護に関する問答の問の部分	(B) 専門家作成の質問文	(C) 専門家による質問文の言い換え	合計
750	1160	3850	5760

<p>質問文:</p> <p>「6 か月以上の入院を要する」場合又は「長期間にわたり入院を要する」場合については、どのような方法で判断したらよいか。</p>
<p>応答文:</p> <p>[問] 「6 か月以上の入院を要する」場合又は「長期間にわたり入院を要する」場合については、どのような方法で判断したらよいか。</p> <p>[答] かならずしも医療扶助の入院要否意見書によることとしなくてもよいが、少なくとも医師の診断証明により取り扱うこととされたい。なお、「長期間」とは、1年を超える期間をいうものである。</p>

図 2: データベース中の質問文と応答文の例 (生活保護手帳の問答から生成した例)

<p>質問文:</p> <p>生活保護を利用しながら大学に進学することは可能か？</p>
<p>応答文:</p> <p>次のいずれかに該当する場合は世帯分離して差しかえない。</p> <p>(1) 保護開始時において、現に大学で就学している者が、その課程を修了するまでの間であって、その就学が特に世帯の自立助長に効果的であると認められる場合。…</p>

図 3: データベース中の質問文と応答文の例 (専門家が生活保護手帳の法令文から質問を生成した例)

データベース中の質問文としては次の 3 種類がある。

- (A) 生活保護に関する問答の問の部分；この場合、質問文の回答根拠は問答となる
- (B) 法律の専門家により作成された、生活保護に関する問答・法令が回答根拠となるような質問文
- (C) 法律の専門家が(A)、(B)の質問文を言い換えた質問文

本研究においては、データベースや質問文の分散表現モデルを構築するために、秋田県健康福祉部において独自に

厚労省の開示資料などから集めたデータを利用した。我々は、この 3 種類の質問文を 2020 年 4 月から 2021 年 12 月にかけて作成した。表 1 に 3 種類の質問文の内訳を示す。(A) に関しては上記のデータを解析することにより抽出した。(B)、(C)に関しては、法律の専門家 2 名が、ケースワーカー経験者のコンサルティングを受けながら作成した。

図 2 は、データベース中の(A)のタイプの質問文、応答文の例である。生活保護に関する問答が応答文となっており、その問の部分が質問文となっている。図 3 は、データベース中の(B)のタイプの質問文、応答文の例であり、生活保護に関連する法令をデータベースの応答文とし、そこから法律の専門家が質問文を生成したものである。

3. 質問文のベクトル化手法

データベース中の質問文、システムへの入力質問文は、分散表現モデルにより数値ベクトル化される。自然言語の単語や文を数値ベクトルで表す方法としては、one-hot 表現があるが、語彙次元の疎なベクトルで表現されることとなり、意味が似ていても異なる単語同士の類似度は 0 になってしまうため、単語や文の意味の類似性を捉えることが難しいという課題があった。

これに対し、最近、分散表現モデルを使って単語や文の意味をニューラルネットワークで表現する研究が盛んである。分散表現モデルは、比較的次元の数値ベクトルを使って単語や文の意味を捉えることができ、様々な自然言語処理のタスクで高い性能を発揮してきた。本研究では、word2vec [6]、fastText [7]、BERT [8]、Sentence BERT [9] という 4 種類の分散表現モデルによる数値ベクトル化について比較した。それぞれの分散表現モデルについて簡単に説明する。

word2vec、fastText は「同じ文脈に現れる単語は類似した意味を持つ」という分布仮説に基づく、単語の分散表現モデルである[6, 7]。one-hot 表現とは異なり、単語の意味を捉えることに成功したが、文脈に応じて意味が変わる多義語の意味をうまく捉えることができないという課題がある。また、単語のベクトル表現であるので、文のベクトルを求める際には、文に含まれる単語の平均ベクトルとして計算するなどの工夫が必要となる。

BERT は、双方向 Transformer をベースとしたモデルであり、自己注意機構を用いることにより、文脈を捉えて単語や文の意味を表現することが可能である[8]。大規模コーパスによる事前学習モデルを個別タスクに応じてファインチューニングすることにより、自然言語処理の様々な個別タスクで高い性能を発揮し、汎用言語モデルと呼ばれる。しかし、BERT の事前学習モデルによって生成される文ベクトルは、必ずしも類似文検索には向いていない[9]。

Sentence BERT (以降、SBERT) は、BERT モデルをファインチューニングすることにより、類似文検索のための良質な文ベクトルを生成する手法である[9]。その考え方は、「ある文とそれに類似する文のペアを学習データとし、似た文から生成される文ベクトルが似たベクトルになるように BERT をファインチューニングする」というものである。

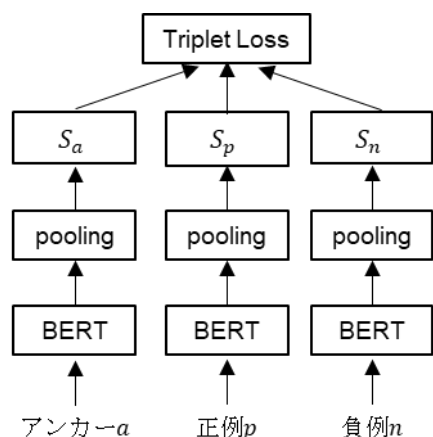


図 4: Triplet Loss を使った場合の Sentence BERT モデルのファインチューニング

本研究ではファインチューニングの際の損失関数として Triplet Loss を用いる。図 4 に Triplet Loss を用いた場合の SBERT モデルのファインチューニングの概要を示す。アンカーと呼ばれる基準となる文 a を与える。アンカーに対し、似ている (positive) 正例の文 p と似ていない (negative) 負例の文 n をそれぞれ指定する。このとき、Triplet Loss と呼ばれる損失関数を最小化するように、アンカー、正例、負例それぞれの分散表現 S_a, S_p, S_n が学習される。

Triplet Loss は次の式で表される[9] :

$$\max(\|S_a - S_p\| - \|S_a - S_n\| + \epsilon, 0)$$

$\|\cdot\|$ は距離尺度である。第 1 項目の $\|S_a - S_p\|$ を小さくすることにより S_a と S_p は似たようなベクトルに、第 2 項目の $\|S_a - S_n\|$ を大きくすることで S_a と S_n は離れたベクトルにできるように学習が進む。 ϵ はマージンと呼ばれ、1 項目と 2 項目の差が ϵ 以上となるようにする効果がある。

4. 質問文の分散表現による質問応答システムの性能評価

4.1 質問文の分散表現モデルの学習

3 節で述べた 4 種類の分散表現モデルを学習し、各モデルを使った場合の質問応答システムの性能を評価した。分散表現モデルの構築、システム評価は、次のライブラリを用いて行った。

- Python 3.9.4
- MeCab 1.0.4
 - 辞書 mecab-ipadic-NEologd v0.0.7
- fugashi 1.1.1
- transformers 4.16.0
- PyTorch 1.10.0+cu111
- SentenceBERT 2.1.0
- fastText 0.9.2
- genism 4.1.2

各分散表現モデルの構築方法について説明する。

(1) word2vec, fastText

word2vec, fastText のモデルは、生活保護関連の文章 1 万 2744 文章からを学習した。モデルの次元は 100 次元で

表 2: 質問文ベクトル化手法ごとの正答率

ベクトル化手法	1 位正答率	3 位内正答率
word2vec	7%	17%
fastText	42%	60%
BERT	15%	21%
SBERT	52%	74%

ある。単語の形態素解析には Mecab [10]、辞書には mecab-ipadic-NEologd [11] を用いた。

(2) BERT

BERT の事前学習モデルとして、東北大学乾研究室作成の日本語 Wikipedia による事前学習済みモデル[12]を使用した。768 次元の BERT-base を用いた。

(3) SBERT

本研究では、SBERT モデルのファインチューニングのため、まず、表 1 で示したデータのうち、(A)生活保護に関する問答の問の部分と(B)専門家作成の質問文を合わせて、1910 件の質問文をアンカーとして用いた。各アンカーの質問文に対して、アンカーと同じ応答文と紐づけられる質問文を正例とした。アンカーと同じ応答文と紐づけられた別の質問文がない場合、アンカー自体を正例として用いた。負例は、アンカーと同じ応答文とは紐づけられていない質問文から無作為に抽出した。結果として、2036 件のファインチューニング用の学習データが得られた。損失関数は Triplet Loss を用い、距離尺度はユークリッド距離、マージン ϵ は 1.0 とした。ファインチューニングの際のバッチサイズ 16、エポック数 2 とした。その他のパラメータそれ以外のパラメータは論文[9]と同様である。

4.2 質問応答システムの評価

4.1 節で述べた 4 種類の分散表現モデルを使って、データベース中の質問文と入力質問文の数値ベクトル化を行い、システムの性能評価を行った。

データベースには、表 1 に示した(A)生活保護に関する問答の問の部分と(B)専門家作成の質問文を合わせて、1910 件の質問文と、それに紐づく応答文を含めた。(C)の質問文の言い換えのデータはデータベースに含めていない。

テストデータは、表 1 に示した(C)の質問文の言い換えのデータのうち 256 件を用いた。テストデータとして用いた質問文は、データベースや分散表現モデルの学習データには含まれていない。テストデータの質問文に紐づく応答文が出力された場合を正答とし、システムの正答率を評価した。

表 2 にベクトル化手法ごとの 1 位正答率、3 位内正答率を示す。最も正答率が高いのは、SBERT モデルであったことが分かる。BERT に比べて飛躍的に正答率が向上しており、生活保護ドメインの文章を使って類似文検索タスクのファインチューニングを行う効果が大きいことが分かる。

Wikipedia から事前学習した BERT モデルは、生活保護に関する文章から学習した fastText モデルよりも正答率が低い。BERT 事前学習モデルは、様々な個別タスクでファインチューニングすれば高い性能を示すことが知られているが、BERT 事前学習モデルから生成されるベクトルを直

接的に類似文検索に用いても、必ずしも性能が発揮できないことが示唆される。

4.3 データベース拡張の効果

4.2 節で述べた評価実験では、テストデータに関しては、表 1 の(C)の質問文の言い換えのうち、256 件のみをテストデータとして使用した。また、(C)の質問文の言い換えのデータは、データベースや SBERT のファインチューニング用データには使用していない。

そこで、4.2 節の評価で最も高い性能を示した SBERT に焦点をあて、表 1 の(C)の質問文言い換え全件によって、データベースならび SBERT のファインチューニング用学習データを拡張した場合の効果調べた。

テストデータとしては、表 1 の (C) に含まれない質問文の言い換え 1140 件を用意した。4.2 節と比較して、さらに大きなサイズのテストデータを使って評価を行うことになる。データベースに含める質問文と SBERT のファインチューニング用データとして、次の 2 つの場合を比較した。結果を表 3 に示す。

(ア) 4.2 節と同様に、(A)の問答の問の部分、(B)の専門家作成の質問文を使った場合

(イ) (ア)に加えて、(C)の質問文の言い換え全件を使った場合

表 2 の SBERT の場合と表 3 の(ア)の場合は、テストデータの言い換え質問が 256 件から 1140 件に増えたことのみが異なる。表 2 の SBERT と比べて、表 3 の(ア)の正答率が向上している。より大きなサイズのテストデータでの評価であるので、より信頼性の高い結果とも言えるが、データベース中の質問文とテストデータの質問文の類似度が高くなったためとも考えられるため、今後の精査が必要である。

表 3 において、(ア)の場合と比べて、(イ)の場合の正答率が向上している。このことは、言い換え質問によってデータベースならびに SBERT のファインチューニング用学習データを拡張することは、正答率を向上させる効果があることを示唆している。しかし、この結果を見る限り、正答率向上の効果は限定的である。今後、コストをかけて、法律の専門家による言い換えによって質問文をさらに拡張しても、システムの性能の向上に顕著に貢献できるとは限らない。今後は、生活保護業務の担当者からの質問文を収集し、システムを評価することも必要と考えられる。

5. おわりに

本稿では、生活保護業務支援のための質問応答システムの構成法と質問応答データベース構築法について述べ、質問文のベクトル表現方式とデータベース拡張が市捨て者性能に及ぼす効果について調べた。データベース中の質問文や入力質問文のベクトル化手法は、word2vec、fastText、BERT に比べて、SBERT モデルを使った場合が高い正答率を示した。このことは、SBERT を使って、生活保護ドメインの文章で類似文検索タスクのファインチューニングを行う効果が大きいことを示している。また、法律の専門家による質問文の言い換えを使ってデータベースと SBERT のファインチューニング学習データを拡張することにより、システムの正答率が向上することが示されたが、質問文の

表 3 : SBERT で質問文をベクトル化した場合において、質問文言い換えによりデータベースと SBERT のファインチューニング用データを拡張した場合の正答率

ファインチューニング用学習データ	1 位正答率	3 位内正答率
(ア) 問答、 専門家作成質問	80.0%	92.2%
(イ) 問答、 専門家作成質問、 言い換え質問	82.5%	94.3%

言い換えによるデータ拡張の効果に関しては、さらに精査が必要である。

今後の課題としては、生活保護業務の担当者が実際に質問応答システムを使った場合の質問文を収集し分析することで、システムの改善を行うことがある。また、生活保護手帳の問答・法令の文章から質問文を自動生成できるか検討したい。

参考文献

- [1] 厚生労働省, 平成 28 年 福祉事務所人員体制調査について, <https://www.mhlw.go.jp/toukei/list/dl/125-1-01.pdf>, (最終閲覧日: 2021 年 1 月 28 日)
- [2] 厚生労働省, 生活保護の被保護者調査 (令和 3 年 1 月分概数) <https://www.mhlw.go.jp/toukei/saikin/hw/hihogosya/m2021/dl/01-01.pdf>, (最終閲覧日: 2021 年 1 月 28 日)
- [3] 生活保護手帳 2019 年度版, 中央法規出版, 2021
- [4] 生活保護手帳問答集 2019 年度版, 中央法規出版, 2019
- [5] 特定非営利活動法人 Child First Lab, 生活保護世帯に対するケースワーカー業務負担軽減のための効果的な人工知能 (AI) アプリの活用に関する調査研究事業報告書, 厚生労働省平成 30 年度社会福祉推進事業実施事業, <https://www.mhlw.go.jp/content/12200000/000525131.pdf>, (最終閲覧日: 2021 年 1 月 28 日)
- [6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean: Distributed Representations of Words and Phrases and their Compositionality, Proceedings of the 26th International Conference on Neural Information Processing Systems, Vol. 2, pp.3111-3119, 2013
- [7] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov: Enriching Word Vectors with Subword Information, Transactions of the Association for Computational Linguistics, Vol. 5, pp.135-146, 2017
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proc. 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, 2019
- [9] Nils Reimers and Iryna Gurevych: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, Proc. the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019
- [10] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237, 2004
- [11] Mecab-ipadic-NEologd v0.07, <https://github.com/neologd>
- [12] 東北大学 乾研究室, Pretrained Japanese BERT models, <https://github.com/cl-tohoku/bert-japanese>, 2021