二段階 fine-tuning を用いた条件付きなぜ型質問応答技術

二宮 大空 †‡　呉 鍾勲 ‡　Julien Kloetzer‡　飯田 龍 †‡　鳥澤 健太郎 †‡
† 奈良先端科学技術大学院大学 先端科学技術研究科
‡ 国立研究開発法人 情報通信研究機構
ninomiya.hirotaka.ng8@is.naist.jp
{rovellia, julien, ryu.iida, torisawa}@nict.go.jp

1 はじめに

近年，SQuAD[1] のような事実を名詞で回答するファクトイド質問応答タスクについては，汎用的な事前学習言語モデル BERT（Bidirectional Encoder Representations from Transformers）[2] 等の発展によりその性能が人間の性能を超えたことが報告されている．一方，説明的な回答が求められるノンファクトイド質問応答は，未だ発展途上であり，ノンファクトイド質問応答の一つであるなぜ型質問応答などにおいて盛んに研究が進められている [3, 4, 5, 6]．

本研究では，より高度ななぜ型質問応答を実現するために，「なぜ地球温暖化が進むと海水面が上昇する」のような，なぜ型質問に条件（「地球温暖化が進む」）が含まれるより複雑な質問に対して回答する質問応答を対象とする．本稿では，このような質問応答を条件付きなぜ型質問応答と呼ぶ．この条件付きなぜ型質問に回答するためには，質問「なぜ（事象 A）と（事象 B）？」に対し，事象 A の因果的帰結であり，かつ，事象 B の原因であるものを回答する必要があるため，単純ななぜ型質問（「なぜ（事象 B）？」）と比べてより難しい問題となっている．一方で，「なぜ地球温暖化が進むと海水温が上昇する？」のような条件付きなぜ型質問は，「地球温暖化が進むとどうなる？」といったどうなる型質問応答 [7, 8, 9] の結果，得られた回答が「海水温が上昇する」であった場合にその理由を問うための自然な質問の形式だと考えられるため，この条件付きなぜ型質問応答の技術を発展させることで，単純な質問応答に付随して発生する，さらなる深掘りを行う質問に適切に回答することが可能になると考えられる．

ただし，我々が知る限り，条件付きなぜ型質問応答に特化した学習・評価用データは存在しない．そこで，本研究では，まず表 1 に示すような条件付きなぜ型質問応答のデータを作成した．このデータ作成では，日本語 Web4 億文書を対象に検索して得られたパッセージ[1)]（以降，回答候補パッセージ）に対して，人手でアノテーションを行った．さらに，このデータを用い，BERT を fine-tuning することで回答候補パッセージのランキングを行う分類器を開発した[2)]．その際，なぜ型質問応答に関連する 4 つのタスク（因果関係認識 [10]，因果関係連鎖判定，なぜ型質問応答 [3][4]，ファクトイド質問応答 [11]）の学習データも用い，最終的な条件付きなぜ型質問応答の fine-tuning の前に実施する fine-tuning，もしくは，条件付きなぜ型質問応答との Multi-task learning を行うことで，条件付きなぜ型質問応答の性能が向上するかの調査も行った．この結果，組み合わせるタスクと組み合わせ方によってはベースラインである単純な BERT の fine-tuning と比較して有意に性能が向上することを明らかにした（P@1 で約 1.2%の向上）．

表 1　条件付きなぜ型質問応答の具体例

条件付きなぜ型質問	なぜ歯周病になると心臓病を引き起こす
回答候補パッセージ	歯周病に罹患した患者の多くが心臓病を引き起こしていることが学会で発表されました．... 発表された論文によると歯周病菌が心臓の血管を詰まらせるからだそうです．炎症によって歯肉内に歯周病菌が侵入する可能性があります．
ラベル	1（回答候補パッセージが答えを含む）

2 関連研究

Oh ら [3] の日本語を対象としたなぜ型質問応答手法では，因果関係を認識する問題を系列ラベリング

1） パッセージは連続する 5 文からなる．
2） 一般的な質問応答では回答は一つとは限らないため，適切な回答を不適切な回答より上位にランキングするランキング問題として扱われる．

問題と捉え，Conditional Random Fields を用いて文書中の原因部と結果部を特定した．それにより一文内の因果関係と文をまたがる因果関係を含む多様な形式で表される因果関係を自動抽出した後，抽出した因果関係をなぜ型質問応答の回答リランキング時に利用し，一文内の因果関係と文をまたがる因果関係が共に回答リランキングの性能向上に寄与していることを示した．さらに，Oh ら [4] はなぜ型質問応答において敵対的学習を用いたモデルを提案し，回答精度を大幅に向上させた．

また，英語を対象にしたノンファクトイド質問応答に関して，Fan ら [12] は回答が長文となる質問応答データセット ELI5 を作成している．ELI5 のデータ中の約 45%が why から始まるなぜ型質問であるが，ELI5 の全データを対象にした評価では Seq2Seq をベースにした手法の性能は人間の性能を大幅に下回っていることが報告されている．また，Hashemi ら [13] は Web から抽出したノンファクトイド質問を含む質問応答データセットを作成した．

3 条件付きなぜ型質問応答のデータ作成

1 節で述べたように，条件付きなぜ型質問は典型的には「なぜ（事象 A）と（事象 B）？」という形式で記述されるため，事象 A と事象 B に埋まる表現の対が必要となる．この 2 つの事象を表す表現は一般的には「なぜ地球温暖化が進むと海水温が上昇する？」の「地球温暖化が進む」と「海水温が上昇する」の対のように因果関係が成り立つ対であると考えられるため，そのような対を自動で収集して質問を作成する．一方，そのような人工的な条件付きなぜ型質問だけでは人間が記述する多様な条件付きなぜ型質問をカバーできない可能性があるため，自然な条件付きなぜ型質問についてもデータを作成し，評価を行う．

データ作成の手順は下記の通りである．

手順 1：条件付きなぜ型質問の作成

Train，Dev1，Dev2 の質問の作成方法として，まずは Hashimoto ら [7] の因果関係認識器を Web40 億文書に適用して得られた因果関係を表す原因句と帰結句の対を利用し，「なぜ（原因句）と（帰結句）」とすることで人工的な条件付きなぜ型質問を作成する．一方，最終評価用のデータ（Test）の作成時には，質問は Web40 億文書から抽出した「なぜ」を含む単語列から抽出した条件付きなぜ型質問である．詳細を付録 A で述べる．質問の作成後，質問が曖昧でない自然なものであるかをアノテータ 3 名による多数決で判定した．

手順 2：回答候補パッセージの抽出とアノテーション

手順 1 で得られた各質問に対して，Murata らの検索手法 [14] を用いて，Web4 億文書から質問に対する回答が含まれている可能性が高い回答候補パッセージを最大 20 件抽出した．さらに，質問と回答候補パッセージを見て，回答候補パッセージ中に質問に対する答えが含まれているかをアノテータ 3 人で判定し，多数決でラベルを決定した．[3)]

本研究で作成したデータの[4)]の総事例数とそれぞれの正例の割合は表 2 に示す通りである．[5)]

表 2　データセットの統計

データ	件数	質問数	正例
Train	44,707	2,250	3,518 (7.9%)
Dev1	6,375	321	433 (6.8%)
Dev2	12,771	642	1,029 (8.1%)
Test	9,896	488	771 (7.8%)

4 比較手法

条件付きなぜ型質問応答とは同じ質問応答という観点で関連するファクトイド質問応答（以降，FQA）[11]，なぜ型質問応答（以降，WhyQA）[3, 4]，また，因果関係を対象とするという点で関連すると考えられる因果関係認識（以降，Causality）[10]，因果関連連鎖判定（以降，Chain）の学習データ（それぞれの詳細は表 3 参照）も利用することで条件付きなぜ型質問応答の性能が向上するという仮説に基づき，下記の Step1 と Step2 からなる二段階の fine-tuning 手法を実験する．

Step1　事前学習済みの BERT モデルを Causality，Chain，WhyQA，FQA のいずれか（もしくはその組み合わせ）のデータで fine-tuning する．

Step2　Step2 で得られた fine-tuning 済み BERT モデルを Causality，Chain，WhyQA，FQA のいずれか（もしくはその組み合わせ）のデータと条件付きなぜ

3）　Train，Dev1，Dev2 においては，答えが含まれていると 1 人以上判定したものは正例である可能性が高いと考え，それらに対して質問に対する答えを回答候補パッセージから抽出する作業をアノテータ 3 人で行った．その後，抽出可能であるかにより最終的なラベルを決定した．

4）　Train，Dev1，Dev2 で質問が重複しないようにランダムに分割する．

5）　Oh[4] が利用している一般的ななぜ型質問応答データの正例の割合は Train，Dev，Test でそれぞれ 12.5%，23.4%，24.1% であり，それらと比べて正例の割合が小さく，ランキングがより困難な問題となっている．

表 3　二段階 fine-tuning で用いるタスクの詳細

タスク	概要	具体例
条件付きなぜ型質問応答（CoWhyQA）	条件付きなぜ型質問に対する答えが回答候補パッセージ中に含まれているかを判定する二値分類タスクである．入力単語列は「[CLS]（質問）[SEP]（回答候補パッセージ）[SEP]」となる．	[CLS] なぜ動物を飼うと婚期が遅れる [SEP]... 愛くるしさゆえに，多くの時間を費やしてしまうそう．...[SEP]
因果関係認識（Causality）[10]	与えられた 2 つの句に因果関係が成立するかを判定する二値分類タスクである．句は必ず名詞，助詞，述語の順で構成されている．入力単語列は「[CLS]（原因句の候補）[SEP]（結果句の候補）[SEP]」となる．	[CLS] 地球温暖化が進む [SEP] 海水面が上昇する [SEP]
因果関係連鎖判定（Chain）	与えられた 4 つの句（句 A，句 B1，句 B2，句 C）に句 A と句 B1 の間，句 B2 と句 C の間に因果関係が成り立ち，因果関係の連鎖全体として正しいかを判定する二値分類タスクである．句は名詞，助詞，述語の順で構成されており，句 B1 と句 B2 は名詞が一致，かつ，助詞と述語に対する「活性/不活性」の意味的極性 [10] が一致している．入力単語列は「[CLS]（句 A）[SEP]（句 B1）[SEP]（句 B2）[SEP]（句 C）[SEP]」となる．	[CLS] 地球温暖化が進む [SEP] 海水温が高くなる [SEP] 海水温が上昇する [SEP] 腸炎ビブリオが増加する [SEP]
なぜ型質問応答（WhyQA）[4]	物事の理由や原因を尋ねるなぜ型質問と，それに対する回答候補パッセージが与えられ，質問に対する答えが回答候補パッセージ中に含まれるかを判定する二値分類タスクである．入力単語列は「[CLS]（質問）[SEP]（回答候補パッセージ）[SEP]」となる．	[CLS] なぜ寒天は体にいい？ [SEP] 寒天は水を含むと膨らみますので，... 満足感が得られ，ダイエットに効果的です．[SEP]
ファクトイド質問応答（FQA）[11]	ファクトイド質問と回答候補となる名詞と回答候補を抽出したパッセージが与えられ，回答候補が質問に対する答えとして適切かを判定する二値分類タスクである．入力単語列は「[CLS]（質問）[SEP]（回答候補）[SEP]（パッセージ）[SEP]」となる．	[CLS] 地球温暖化は何を引き起こす [SEP] 海水面上昇 [SEP] 地球温暖化は海水温上昇を引き起こすと言われています．[SEP]

条件付きなぜ型質問の作成に用いた Hashimoto ら [7] の因果関係認識器は因果関係認識のデータで学習されているため，条件付きなぜ型質問応答に関連する情報を含まないように，Train，Dev1，Dev2 中の条件付きなぜ型質問の作成に用いた句が含まれる事例を，因果関係認識と因果関係連鎖判定のデータからそれぞれ 3,675 件，985 件除外した．表 4 はこれらを除外した件数である．

表 4　関連するタスク 4 種類のデータセット

タスク	Causality	Chain	WhyQA	FQA
Train	103,777	9,106	17,000	174,765
Dev	11,389	909	2,000	10,881

型質問応答（以降，CoWhyQA）のデータを用いて Multi-task learning でさらに fine-tuning する．

5　実験

5.1　実験設定

実験で利用する BERT のモデルサイズは Devlin ら [2] の $\text{BERT}_{\text{LARGE}}$ に従う．BERT の事前学習には，Kadowaki ら [8] と同様に，自動検出された因果関係を含むテキスト約 22 億文のコーパス（データサイズは 353GB であり，T5[15] の事前学習で利用されているコーパスの約半分に相当する [11]）を用いる．このコーパスは，Oh ら [3] の因果関係認識器により検出された文とその前後の文で構成された 7 文のパッセージから得られた文集合からなる．事前学習時のバッチサイズは 4,096 であり，token 最大長 128 で 100 万ステップ，さらに追加で token 最大長 512 で 10 万ステップ事前学習を行っている[6)]．

6）事前学習と fine-tuning 時には Kadowaki ら [8] と同様に learning rate warmup (warmup rate=0.01) を用いた．

fine-tuning 時のバッチサイズは 32 であり，学習率は 8e-6，9e-6，1e-6 の 3 通り，epoch 数は 1，2，3 の 3 通りの組み合わせである 9 通りから Dev の P@1（最上位の回答の精度）が最大となるパラメタを手法ごとに選択した．Step1 のモデル選択に関しては WhyQA の場合は Dev1 の P@1 で，それ以外は Dev1 の平均精度でベストなモデルを選択した．最終的な条件付きなぜ型質問応答の評価では P@1 に加え，MAP（回答の平均精度の平均値）でも評価し，各手法の結果を比較する．ベースラインとして，前述の因果関係を含むテキスト約 22 億文のコーパスで事前学習した BERT モデルを条件付きなぜ型質問応答のデータだけで fine-tuning したものを利用する．

また，二段階 fine-tuning で用いるデータセットの事例数を表 4 に示す．二段階 fine-tuning の Step1，Step2 におけるパラメタ探索範囲や学習方法，その他の実験設定は共に fine-tuning 時の設定に従う．

5.2　実験結果

結果を表 5 に示す．表の (c)〜(f) が Step1 の fine-tuning の学習データを変更した場合の結果であるが，この内，WhyQA と FQA で fine-tuning した場合が Dev1，Dev2，Test で一貫して性能が向上しており，因果関係認識に関するデータ（Causality，Chain）

表 5　二段階 fine-tuning に関する実験結果

ID	Step1 で用いるタスク	Step2 で用いるタスク	Dev1	Dev2	Test		
			P@1	P@1	P@1	相対値	MAP
(a)	Oracle		35.83	42.37	40.81	100	40.81
(b)	BERT（ベースライン）		28.04	30.53	31.72	77.7	32.79
(c)	Causality	CoWhyQA	27.41 (-0.63)	32.09† **(+1.56)**	31.11 (-0.61)	76.2	33.02
(d)	Chain	CoWhyQA	27.41 (-0.63)	31.62† **(+1.09)**	30.91† (-0.81)	75.7	32.86
(e)	WhyQA	CoWhyQA	28.97 **(+0.93)**	31.78† **(+1.25)**	32.53† **(+0.81)**	79.7	33.75
(f)	FQA	CoWhyQA	28.97 **(+0.93)**	32.55† **(+2.02)**	32.73† **(+1.01)**	80.2	33.80
(g)	なし	CoWhyQA+Causality	28.97 **(+0.93)**	30.37 (-0.16)	32.12 **(+0.40)**	78.7	33.14
(h)	なし	CoWhyQA+Chain	28.35 **(+0.31)**	31.78 **(+1.25)**	31.11 (-0.61)	76.2	32.58
(i)	なし	CoWhyQA+WhyQA	28.97 **(+0.93)**	31.15 **(+0.62)**	**32.93**† **(+1.21)**	**80.7**	33.87
(j)	なし	CoWhyQA+FQA	29.28 **(+1.24)**	31.15† **(+0.62)**	32.32 **(+0.60)**	79.2	32.91
(k)	FQA	CoWhyQA+WhyQA	**29.60 (+1.56)**	**33.18**† **(+2.65)**	**32.93**† **(+1.21)**	**80.7**	**34.20**
(l)	Causality+FQA	CoWhyQA	**29.60 (+1.56)**	**33.18**† **(+2.65)**	32.32 **(+0.60)**	79.2	33.47

Oracle は全ての質問に対して適切に回答できた場合の理想的な精度を，相対値は「(手法ごとの P@1) / (Oracle の P@1) × 100」で求めた値である．P@1 における括弧内の値は BERT（ベースライン）の P@1 との性能差を表しており，Dev2 と Test においては McNemar 検定（有意水準 5%）で有意差が確認されたものを † で示す．

よりも質問応答に関するデータ（WhyQA，FQA）で Step1 の学習を行うことに効果があると考えられる．また，表 5 の (g)〜(j) が Step1 の fine-tuning を行わず，Step2 で Multi-task learning を行った結果であるが，この結果については特に WhyQA で Multi-task learning した場合の Test の P@1 が最も高くなっており，対象となる条件付きなぜ型質問と最もタスクが近いなぜ型質問応答との Multi-task learning に効果があったと考えられる．

また，表 5 の (c) と (d) に着目した場合，Dev2，つまり，質問の形式が「なぜ（事象 A）と（事象 B）」の場合には P@1 が向上しているのに対し，自然な質問の形式である Test においては P@1 がベースラインよりも低下するという結果となった．これは Causality や Chain の学習データの個々の事象を表す表現が必ず名詞+助詞+述語であるのに対し，Test の質問ではそれらの表現が「運動すればなぜ内臓脂肪が減るのでしょうか？」の「運動する」のように必ずしも名詞+助詞+述語の形式とはなっていないことが影響している可能性がある．

最後に，Step1 の選択肢と Step2 の選択肢を組み合わせることで性能が向上するかについても調査を行った．具体的には Step1 で効果があった Causality と FQA（表 5 の (c)〜(f) で Dev2 の P@1 の上位 2 つ）を利用する場合としない場合の組み合わせ 4 通り，また，Step2 で効果があった Chain，WhyQA，FQA（表 5 の (g)〜(j) で Dev2 の P@1 がベースラインよりも良かったもの）を利用する場合としない場合の組み合わせ 8 通り，合わせて 32 通りに対して実験を行い，Dev2 の P@1 が最大となる組み合わせを決定した．この結果，表 5 の (k)Step1=FQA，Step2=CoWhyQA+WhyQA，(l)Step1=Causality+FQA，Step2=CoWhyQA の 2 つが同じ P@1 だったため選択された．ただし，表に示すように，この 2 つで Test の P@1 が良かった (k) のその値は，(i) に示す Step1=なし，Step2=CoWhyQA+WhyQA と変わらないため，Step1 と Step2 の組み合わせは効果がなく，Step1 もしくは Step2 のいずれかで類似するタスクの学習を行うだけで十分であることがわかる．一方で，(k) と (l) の Dev2 の P@1 の値は他の (b)〜(j) のいずれをも上回っているため，質問の形式が学習データと同じ場合においては Step1 と Step2 で複数のタスクを組み合わせることに効果があることがわかる．このように，現状では質問の形式（人工的な質問 or 自然な質問）が学習データと異なることが，二段階の fine-tuning が有効であるかに影響していると考えられるため，今後は質問の形式が異なっていたとしても性能を劣化させない頑健な手法を開発していく必要がある．

6　おわりに

本研究では，条件が含まれるなぜ型質問に対する応答技術の開発を目指し，条件付きなぜ型質問応答のデータを作成した．さらに，BERT を用いた分類器を構築し，条件付きなぜ型質問応答に関連するタスクを用いて二段階で fine-tuning を行うことで，分類器の性能向上を図った．実験の結果，二段階 fine-tuning を行うことで，最上位の回答の精度 P@1 がベースラインである BERT を約 1.2%上回った．

参考文献

[1] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2383–2392, Austin, Texas, November 2016. Association for Computational Linguistics.

[2] 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.

[3] Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Motoki Sano, Stijn De Saeger, and Kiyonori Ohtake. Why-question answering using intra- and inter-sentential causal relations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1733–1743, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[4] Jong-Hoon Oh, Kazuma Kadowaki, Julien Kloetzer, Ryu Iida, and Kentaro Torisawa. Open-domain why-question answering with adversarial learning to encode answer texts. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4227–4237, Florence, Italy, July 2019. Association for Computational Linguistics.

[5] Fatima T AL-Khawaldeh. Answer extraction for why arabic questions answering systems: Ewaq. *arXiv preprint arXiv:1907.04149*, 2019.

[6] Andrei Dulceanu, Thang Le Dinh, Walter Chang, Trung Bui, Doo Soon Kim, Manh Chien Vu, and Seokhwan Kim. Photoshopquia: A corpus of non-factoid questions and answers for why-question answering. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[7] Chikara Hashimoto, Kentaro Torisawa, Julien Kloetzer, Motoki Sano, István Varga, Jong-Hoon Oh, and Yutaka Kidawara. Toward future scenario generation: Extracting event causality exploiting semantic relation, context, and association features. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 987–997, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[8] Kazuma Kadowaki, Ryu Iida, Kentaro Torisawa, Jong-Hoon Oh, and Julien Kloetzer. Event causality recognition exploiting multiple annotators' judgments and background knowledge. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5816–5822, Hong Kong, China, November 2019. Association for Computational Linguistics.

[9] Canasai Kruengkrai, Kentaro Torisawa, Chikara Hashimoto, Julien Kloetzer, Jong-Hoon Oh, and Masahiro Tanaka. Improving event causality recognition with multiple background knowledge sources using multi-column convolutional neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31, 2017.

[10] Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Jong-Hoon Oh, and Jun'ichi Kazama. Excitatory or inhibitory: A new semantic orientation extracts contradiction and causality from the web. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 619–630, Jeju Island, Korea, July 2012. Association for Computational Linguistics.

[11] 関直哉, 水野淳太, 門脇一真, 飯田龍, 鳥澤健太郎. ファクトイド質問応答における bert の pre-trained モデルの影響の分析. 言語処理学会第 26 回年次大会発表論文集, pp. 105–108, 3 2020.

[12] Angela Fan, Yacine Jernite, Ethan Perez, David Grangier, Jason Weston, and Michael Auli. ELI5: Long form question answering. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3558–3567, Florence, Italy, July 2019. Association for Computational Linguistics.

[13] Helia Hashemi, Mohammad Aliannejadi, Hamed Zamani, and W Bruce Croft. Antique: A non-factoid question answering benchmark. In *European Conference on Information Retrieval*, pp. 166–173. Springer, 2020.

[14] Masaki Murata, Sachiyo Tsukawaki, Toshiyuki Kanamaru, Qing Ma, and Hitoshi Isahara. A system for answering non-factoid japanese questions by using passage retrieval weighted based on type of answer. In *NTCIR*, 2007.

[15] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*, 2019.

付録

A 条件付きなぜ型質問の抽出

Web40 億文書から抽出した「なぜ」を含む単語列から条件付きなぜ型質問を抽出する際，質問中に条件を表す部分を含む必要があるため，以下を全て満たすものを対象とする．単語や品詞の連続は+で表す．

- 動詞，または形容詞+接尾辞を単語列中に 2 つ以上含む．
- 「と，ば，たら」で区切った時，前後半共に動詞，または形容詞+接尾辞が存在する．
- 以下のいずれかを満たす．
  - 動詞+「と」を含む．ただし，動詞は「いう，言う，よぶ，呼ぶ，思う，感じる，する」を除く．
  - 末尾が「ば，たら」の動詞を含む．
  - 動詞+末尾が「ば，たら」の接尾辞を含む．
  - 形容詞+接尾辞+「と」を含む．
  - 形容詞+末尾が「ば，たら」の接尾辞を含む．