図書館情報メディア研究科修士論文

マイクロタスク設計支援のための
ユーザフィードバックの収集・選択手法

2017年3月

201521639

林亮太

マイクロタスク設計支援のためのユーザフィードバックの収集・選択手法
A Method to Collect and Select User Feedbacks for Improving Microtask Designs

学籍番号: 201521639
氏名: 林 亮太
Ryota HAYASHI

近年，群衆の知や力を利用して，計算機だけでは解決が困難な問題に取り組むクラウドソーシングが注目を集めている．クラウドソーシングとは，リクエスタと呼ばれる問題解決を望む人が，ワーカと呼ばれる不特定多数の人々にタスクの処理を委託することである．クラウドソーシングのうち，委託するものが短時間で処理可能なタスク（マイクロタスク）であるものをマイクロタスク型クラウドソーシングと呼ぶ．

マイクロタスク型クラウドソーシングにおける重要な課題として，どのようにしてマイクロタスク内の質問文の曖昧性を除去するかということが挙げられる．マイクロタスク内の質問文が曖昧であると，質問文が同じタスクでもそれぞれのワーカが様々な解釈をしてしまい，タスク結果の品質が下がってしまう．

そこで本研究では，タスクへの回答とその回答理由の組をクラウドソーシングによって収集し，リクエスタに提示することを提案する．これは，タスクを処理してもらう際に，ワーカには質問文に対する回答だけでなく，なぜそのような回答に至ったかという理由も入力してもらうことによって実現する．例えば，ワーカに赤ちゃんが写っている写真を見せ，「この写真は暴力的な内容を含みますか？」と問うタスクがあったとすれば，ワーカには「いいえ，暴力的ではありません」と回答してもらい，さらに理由として「なぜなら赤ちゃんは可愛く，赤ちゃんを嫌いな人などいないからです」などと答えてもらうようにする．この回答と回答理由の組を本論文では Viewpoint と呼ぶ．Viewpoint をリクエスタに提示することで，リクエスタはワーカの質問文に対する解釈を知ることができ，質問文改善の材料とすることができる．

しかし，収集する Viewpoint の数が多くなると，そのすべてをリクエスタが確認するのは困難となるため，質問文改善に有用な Viewpoint を選択することが重要となる．本論文では，Viewpoint に記述された文の論理構造に着目し，実験を通して，その論理構造が有用な Viewpoint を選択するための手がかりでとなることを明らかにした．さらに，コーパスを作成し，論理構造を自動的に判定する分類器を構築した．

さらに，本論文では，論理構造を用いた手法と，エントロピーを用いた手法の比較実験をおこなった．エントロピーを用いた手法とは，複数のワーカからの回答が割れているタスクで得られた Viewpoint を有用と判定するものである．エントロピーを用いた手法では，回答の割れ具合を見るために，1 つのタスクに対して複数の回答が必要である．実験の結果，論理構造を用いた手法が，有用な Viewpoint を選択において，エントロピーを用いた手法と同程度の性能を持つことを示した．

研究指導教員: 森嶋厚行
副研究指導教員: 関洋平

マイクロタスク設計支援のための
ユーザフィードバックの収集・選択手法

筑波大学

図書館情報メディア研究科

2017年3月

林亮太

目次

表目次

図目次

1 はじめに

近年，群衆の知や力を利用して，計算機だけでは解決が困難な問題に取り組むクラウドソーシングが注目を集めている．クラウドソーシングとは，リクエスタと呼ばれる問題解決を望む人が，ワーカと呼ばれる不特定多数の人々にタスクの処理を委託することである．クラウドソーシングのうち，委託するものが短時間で処理可能なタスク（マイクロタスク）であるものをマイクロタスク型クラウドソーシングと呼ぶ．

マイクロタスク型クラウドソーシングにおける重要な課題として，どのようにしてマイクロタスク内の質問文の曖昧性を除去するかということが挙げられる．マイクロタスク内の質問文が曖昧であると，質問文が同じタスクでもそれぞれのワーカが様々な解釈をしてしまい，タスク結果の品質が下がってしまう．実際に，第 3.3.1 項で説明したマイクロタスクの一つでは，簡単な質問文であるにも関わらず曖昧性を含んでいたため，多様な解釈の存在により正答率は 75.5% しかなかった．このような曖昧性を除去するには，リクエスタが多くのワーカたちがとりうる解釈を把握していなければならない．しかしながら，熟練のリクエスタでさえもそのような多様な解釈を把握することは困難である．したがって，多様な解釈をフィードバックとして提供し，リクエスタに気付きを与えることは質問文改善に有用であると考えられる．

そこで本論文では，タスクへの回答とその回答理由の組を収集し，その中から質問文改善に有用なものを自動的に選択する手法を提案する．この回答と回答理由の組を本論文では Viewpoint と呼ぶ．Viewpoint をリクエスタに提示することで，リクエスタはワーカの質問文に対する解釈を知ることができ，質問文改善の材料とすることができる．本手法は直接高品質なタスクの結果を選択する手法とは異なり，そのような手法と組み合わせることも可能である．

本論文の貢献は以下の 3 つである．

(1)Viewpoint の収集方法の提案．マイクロタスクを処理してもらう際に，ワーカには質問文に対する回答だけでなく，なぜそのような回答に至ったかという理由も入力してもらう．これは，マイクロタスクの最後に回答理由を入力するためのフォームを作成することによって実現する．例えば，ワーカに赤ちゃんが写っている写真を見せ，「この写真は暴力的な内容を含みますか？」と問うタスクがあったとすれば，ワーカには「いいえ，暴力的ではありません」と回答してもらい，さらに理由として「なぜなら赤ちゃんは可愛く，赤ちゃんを嫌いな人などいないからです」などと答えてもらうようにする．本論文では，この回答と回答理由の組のことを Viewpoint と呼ぶ．先ほどの例では，組（「いいえ，暴力的ではありません」,「なぜなら赤ちゃんは可愛く，赤ちゃんを嫌いな人などいないからです」）が Viewpoint に該当する．なお，本手法においては回答理由の入力はワーカの任意とし，回答理由の入力による追加の報酬は与えないものとする．

(2)Viewpoint へのアノテーションガイドラインと有用な Viewpoint の選択手法の提案．収集した Viewpoint の数が多くなると，その中からより有用な Viewpoint を抽出することは重要な課題となる．本手法では，有用な Viewpoint かどうかを判断する基準として汎用性を用いる．汎用性とは，その Viewpoint が質問文に回答する際に, より多くのデータに対して役立つという性質である．例えば，「この写真には不快な内容が含まれますか？」という質問に対しての汎用性の低い Viewpoint の例として以下のようなものが挙げられる．

（「不快な内容を含みません」,「野球のバッドは不快ではないから」）
（「不快な内容を含みません」,「日没の写真はたいてい不快でないから」）
（「不快な内容を含みません」,「これは果物であり，不快ではないから」）

これらの Viewpoint は，「不快な内容を含まない」ものの名前を単に挙げているだけである．一般に「不快な内容を含まない」ものは非常に多く存在する．その中から，このように「不快な内容を含まない」ものを

ひとつ取り上げたとしても，取り上げたもの以外のものに関してはワーカがどのように考えているかわからない．ゆえにリクエスタは，このような Viewpoint を見ても，ワーカが質問文に対してどのように解釈したのかを把握することができない．これに対して，汎用性の高い Viewpoint の例としては（「不快な内容を含みません」,「ヌードや暴力的な表現が含まれていないから」）などが挙げられる．これはワーカの判断基準をはっきりと述べているため，ワーカが質問文に対してどのように解釈したのかを把握するには十分であると考えられ，質問文改善に有用であるといえる．

クラウドソーシングを用いた主観評価実験により，我々は Viewpoint 内の論理構造がこの汎用性を決める重要な要因であることを発見した．そこで，論理構造を表す型を Viewpoint に付与するためのアノテーションガイドラインを提案し，それに則って型を付与した Viewpoint のコーパスを作成した．

(3) 評価実験の実施．本論文では様々な評価実験の結果を示している．まず，クラウドソーシングを用いて Viewpoint の汎用性を評価する主観評価実験が挙げられる．この実験では，収集した Viewpoint についてワーカにその Viewpoint の生成元となったデータとは異なるデータについて判断する際に役立つかを問う．例えば，野球のバットが映った写真（写真 A とする）を見せて「これは不快な内容を含みますか？」と問うタスクから，(「不快ではありません」,「野球のバットは不快ではないから」）という Viewpoint が得られたとする．このとき主観評価実験では，まず，次の 2 つの条件を満たす写真 B をランダムに選ぶ．

(1) 写真 A とは違うものが写っている
(2) 質問（ここでは，「これは不快な内容を含みますか？」）に対する正しい回答が，Viewpoint の回答 (ここでは，「不快ではありません」) と同じ

この写真 B について，得られた Viewpoint を用いて「『いいえ，不快ではありません．なぜなら，野球のバットは不快ではないから』という意見が寄せられました．この意見は写真 B が「不快でない」と判断するときに役立つと思いますか？」とワーカに問うタスクを作成し，委託する．このとき，「役立つ」と答えられれば，その Viewpoint は汎用性が高いといえる．この実験の結果は，提案したアノテーションガイドラインがより汎用的な Viewpoint を選択するのに役立つことを示した．さらに，主観評価実験の有効性を検証するための実験として，実際に質問文の書き換えを行ってもらい，その際に役に立った Viewpoint を選択してもらう実験も行った．

また，有用な Viewpoint を選択するのに有効と考えられる手法として，エントロピーを用いた手法が挙げられる．これは，回答の分布が広い，すなわち，「はい」「いいえ」などの選択肢に投票する人数がより割れているデータに対する Viewpoint を有用とする手法である．このエントロピーを用いた手法と，提案した論理構造を用いる手法を比較を行った．その結果，提案手法はエントロピーを用いた手法と同程度の品質で有用な Viewpoint を選択できることを示した．提案手法は論理構造の判定のためのテキスト解析しか行わず，1 つのタスクに対して複数のワーカにタスクを処理してもらう必要がない．

2 関連研究

我々の知る限りでは，マイクロタスク型クラウドソーシングにおける，有用な Viewpoint の選択手法に関する研究は存在しない．クラウドソーシングの分野では，最も関連性の高い研究は [1] で行われている．この研究では，「明らかでない属性は必ずしも簡単に命名できない．それにもかかわらず，人々はそれぞれの画像への解釈のもと組織的な役割を果たす．ここで，関連し合う明らかでない属性の集合を “解釈の次元” と呼ぶ．“解釈の次元” は人々が画像を比較するように頼まれたときに出現し，関連し合うと考えられる社会に関するイメージへの重要な洞察をもたらす」と述べられている．また，クラウドソーシングを用いて，明らかでない属性を発見する手法も提案している．[1] では画像の明確でない属性の発見に焦点を当てているのに対し，我々の研究ではタスクの記述の明確でない属性の発見に焦点を当てている．

自然言語処理の分野では，我々の考えに最も近い技術は前提条件検出である．前提条件検出での課題は，質問文の中の前提条件を見つけることである．前提条件とは，何らかの情報を提供する命題である．例えば，「あなたはいつ結婚しましたか？」という質問文には，結婚しているという前提条件が含まれる．結婚していない人は何と回答すればよいか分からないため，回答はランダムとなる．作成した質問文に意図しない前提条件が含まれることを避けるため，前提条件検出の手法は開発された [2, 3]．前提条件検出は原文推論 [4] の一部である．原文推論には，ある文章から他の文章を推論する様々なタスクがある．原文推論では書かれたテキストの意味のみについて焦点を当てているのに対し，本研究では書き手（Viewpoint を提供するワーカ）のコメントを分類するための読み手（リクエスタ）の要求に焦点を当てている．リクエスタへのワーカのコメントの有用性は，テキストで記述された明確な事実や関係よりも，抽象的な概念である．

最後に，データの品質はクラウドソーシングにおいて重要な問題である．質問文改善の他にも，タスク結果の品質を改善する方法は存在する．クラウドソーシングを行う前には，タスク設計を改善したり，良いタスク結果をもたらすワーカを選択する機会がある [5, 6]．この段階において，元のタスク結果が芳しくないと思われるとき，タスクの質問文を訂正するのに用いることができる．タスクを訂正した後，データの品質を改善するために，適切でないタスク結果やワーカをフィルタしたり，結果を統合する機会がある [7, 8]．本論文ではこの段階については焦点を当てないが，この段階において Viewpoint を用いることは今後の課題である．例えば，回答理由が極端に短い結果をフィルタすることが挙げられる．回答理由があまりにも短いと，真面目に答えていないと考えられるからである．

3 提案手法

本章では，まず 3.1 節で Viewpoint について説明する．次に，3.2 節で Viewpoint を用いたタスクの質問文の改善を支援する手法の概要を説明する．

Viewpoint の数が膨大となると，タスクの質問文を改善するリクエスタがそのすべてを確認するのは困難となる．したがって，収集した Viewpoint の中から質問文改善に有用であるものを選択することは重要である．そこで 3.3 節では，まず実際に Viewpoint を収集し，それらを分類するために Viewpoint の型を定義する．さらに，その定義に基づいて，収集した Viewpoint に手動で型のアノテーションを行い，自動的な Viewpoint の型のアノテーションのためのコーパスを作成する．

4 章では作成したコーパスを用いて分類器を構築した．また，4 章において示す実験結果は，特定の型を持つ Viewpoint が有用でないものが多い傾向にあることを示し，本章で提案するアノテーションスキーマが有用な Viewpoint を選択する手がかりであることを発見した．

3.1 Viewpoint とは

本論文では，Viewpoint とは質問に対する回答とその回答理由の組である．例えば，「これはコーヒーですか？」に対する Viewpoint として（「コーヒーである」，「カップに入っている黒い液体だから」）が挙げられる．

Viewpoint はタスクの質問文改善に役立つと考えられる．なぜなら，リクエスタの意図しないワーカの判断基準を知ることができるからである．例えば，お湯に溶かして飲むスティックタイプのインスタントコーヒーの画像を見せて，「インスタント食品ですか？」と問うタスクがあったとする．リクエスタの意図としては，インスタントコーヒーはお湯に溶かすだけで飲むことができるので「はい」と回答してほしい．しかし，ワーカの中に「コーヒーは飲料であるため，食品ではない」と考える人がいたとする．この場合，リクエスタはそのように考える人がいることに気付けないため「いいえ」と回答する人が出てしまい，タスク結果の品質が下がる．

ここで，「コーヒーは飲料であるため，食品ではない」と考える人の Viewpoint（「インスタント食品ではありません」，「飲料であり食品ではないから」）をリクエスタに提示すれば，そのように考える人の存在に気付くことができる．したがってリクエスタは，質問文改善についての手がかりを得ることができる．このように，Viewpoint は質問文改善を支援するのに役立つと言える．

3.2 提案手法概要

提案手法は次の 2 つのステップからなる．

ステップ 1. 質問に対する回答だけでなく，その回答理由も述べてもらうことにより，回答と回答理由の組，すなわち Viewpoint を収集する．具体的には，元となるマイクロタスクに，ワーカがその回答を行った理由を問う質問文とその入力フォームを追加し，図 3.1 のようなマイクロタスクを作成する．このマイクロタスクをワーカに委託することによって Viewpoint を収集する．

ステップ 2. 収集した Viewpoint の中から，有用と思われるものを選択し，リクエスタに提示する．本研究では Viewpoint の論理構造に着目し，論理構造に基づいて Viewpoint の型を定義する．さらに 4 章において，特定の型を持つ Viewpoint が有用でないものが多い傾向にあることを示し，それらをフィルタしたとしても質問文の改善支援に影響がないことを示す．また，型を判定する分類器を構築することによって，

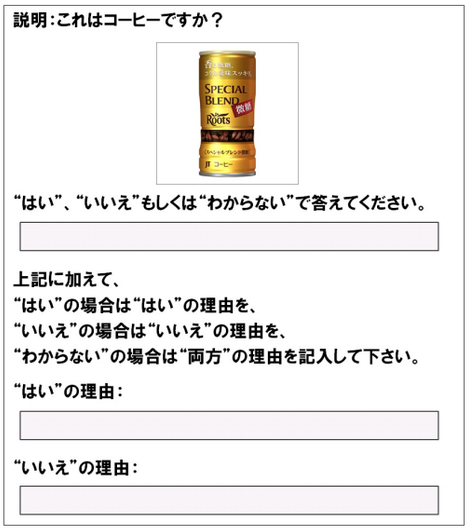

図 3.1: Viewpoint 収集用タスク

自動的に有用と思われるものを選択可能にする．

3.3 Viewpoint へのアノテーション

3.2 節のステップ 2. を実現するため，Viewpoint の収集とそれらへの型のアノテーションを手動で行うことによってコーパスを作成する．このコーパスは，他のマイクロタスクから得られた Viewpoint への自動的なアノテーションのために用いる．

3.3.1 Viewpoint の収集

■タスクの作成 まず，Viewpoint を収集するためのマイクロタスクを作成する．このマイクロタスクのデータには，日本の電子商取引サイトである ASKUL[*1]で販売されている商品の画像データを用い，それぞれのマイクロタスクでは画像データに映る商品が特定のカテゴリに属すかをワーカに問う．ASKUL では，多くの商品が階層構造を持つカテゴリに分類されているため，ASKUL で取り扱われている全ての商品画像データを用いるとマイクロタスクの数が膨大になってしまう．そこで，次の条件を満たす「飲料／食料」カテゴリのサブカテゴリに属する商品画像データを用いる．

(1) 多様な商品を含むようにするため，少なくとも 3 つのサブカテゴリを持つ．
(2) カテゴリ名を簡潔にするため，カテゴリ名が他のカテゴリ名を含むもの（例：「コーヒー／お茶」など）でない
(3) カテゴリに属するデータ数が膨大になるのを防ぐため，広い概念を指すもの（例：「贈りもの」など）でない

その結果，「飲料／食料」カテゴリの 7 つのサブカテゴリ「コーヒー」，「炭酸飲料」，「緑茶」，「紅茶」，「お茶」，「調味料」，「インスタント食品」が選出された．

ここから，それらに属する 28 個の商品画像データと，カテゴリに属するかどうかを問う 7 つの質問文 (図 3.2) を作成した．この 7 つの質問を，28 個の商品画像データに対して行うこととし，元となるマイクロタスクを $7 \times 28 = 196$ 個作成した．

[*1] http://www.askul.co.jp/

次に，元となるマイクロタスクに，ワーカがその回答を行った理由を問う質問文とその入力フォームを追加し，図 3.1 のようなマイクロタスクを作成する．ワーカは「これはお茶ですか？」といった 1 つ目の質問に対して「はい」「いいえ」「わからない」のどれかで回答し，なぜその回答を選んだのかの理由を 2 つ目の質問で答える．

これはコーヒーですか？
これは炭酸飲料ですか？
これは緑茶ですか？
これは紅茶ですか？
これはお茶ですか？
これは調味料ですか？
これはインスタント食品ですか？

図 3.2: コーパス作成に用いる元の質問

■タスクの委託　Yahoo!クラウドソーシングを通して，作成されたマイクロタスク 1 つにつき，20 人のワーカに問い合わせ，20 個の結果を得る．ワーカの必要人数の削減のため，重複のない 5 つのマイクロタスクを 1 セットとし，それぞれのワーカに 1 セット単位でタスクの委託を行った．このセットを作成するためにダミーのマイクロタスクを 4 個作成し，マイクロタスク数が 5 で割り切れる数（$196 + 4 = 200$）に設定した．また，ワーカに対する 1 セットあたりの報酬は 2 円に設定し，ワーカは多くとも 1 セットのタスクしか処理できないよう設定した．得られる最大の Viewpoint の数は $200 \times 20 = 4000$ であるが，理由を問う 2 つ目の質問に対する回答が任意であるため，結果として得られた Viewpoint の数は 1413 であった．また，このタスクを処理したワーカの人数は，387 人であった．

3.3.2　本論文で用いる表記法

本論文では，Viewpoint の収集を行うタスク結果の表記法として，次を用いる．W をワーカの集合，Q を質問文の集合，A を取り得る判断の集合（$A = \{$ はい, いいえ $\}$），C を取り得る選択肢の集合（$C = A \cup \{\text{Unsure}\}$）とする．このとき，タスク結果の各インスタンスを (w, q, d, a, r) という組で表す．ここで，r は，ワーカ $w \in W$ が提示された画像データ d を対象とする質問文 $q \in Q$ に対して判断 $a \in A$ を行った理由である．ワーカが q に対して「わからない」と回答した場合，そのワーカには「はい」「いいえ」両方の判断の理由を入力してもらう．その場合，我々は 2 つのインスタンス $(w, q, d, \text{はい}, r_1), (w, q, d, \text{いいえ}, r_2)$ を手に入れる．

また，Viewpoint は理由 r と結論 c からなると考えられ，これを論理式 $r \to c$ で置き換える．例えば，ある画像を見せて「これはコーヒーですか？」という問うタスクに対して Viewpoint（「コーヒーである」，「これはカップに入った黒い液体であるから」）があったときに，r は「これはカップに入った黒い液体」，c は「これはコーヒーである」を表す．

ここで，Viewpoint $r \to c$ はタスク結果のインスタンス (w, q, d, a, r) から入手することができる．なぜなら，もし a が「はい」（もしくは「いいえ」）であれば，c は肯定形（あるいは否定形）の記述に定まるからである．例えば，先ほどコーヒーの例における c は，「これはコーヒーですか？」という質問 q に対する「はい」という判断 a によって，「これはコーヒーである」という結論 c が導き出せる．したがって本論文では，必要であれば $r \to (q, a)$ という記法で Viewpoint を表すこととする．

3.3.3　Viewpoint の型

論理構造の観点から，Viewpoint の型を定義する．Viewpoint を表す論理式 $r \to c$ において，回答理由 r あるいは c が肯定的な意味を持つ場合それを “P”，否定的な意味を持つ場合はそれを “N” で表すこととする．まず，N と P の組み合わせによって，次の単純な 4 つの型を定義する．

PP 型 $(r \to c)$: （例）ラベルにコーヒーと書いてあるので，これはコーヒーである．
NP 型 $(\neg r \to c)$: （例）コーヒーでない証拠がないので，これはコーヒーである．
PN 型 $(r \to \neg c)$: （例）これはカレーである．したがって，コーヒーではない．
NN 型 $(\neg r \to \neg c)$: （例）コーヒー豆から作られていないので，コーヒーではない．

これらに加えて，次の 2 つの複合型を定義する．ここで複合型とは，回答理由に肯定形と否定形の両方が含まれているものであり，回答理由 $r_1, \neg r_2, \ldots$ 間の論理演算子（「かつ」「または」など）は区別しない．

PNP 型 $(r_1, \neg r_2 \to c)$: （例）ラベルにコーヒーと書いてあり，かつコーヒーでないという証拠がないため，コーヒーである．
PNN 型 $(r_1, \neg r_2 \to \neg c)$: （例）液体の色が黒でなく，透明であるので，コーヒーでない．

また，「コーヒーだからコーヒーである」というような Viewpoint をトートロジーと呼ぶ．

定義 1（トートロジー） PP 型 $(r \to c)$ あるいは NN 型 $(\neg r \to \neg c)$ のうち，$r = c$ となるものを, トートロジーと呼ぶ．

トートロジーかそうでないかを区別するために，トートロジーとなっている Viewpoint を表す 2 つの型を定義する．

P 型 $(c \to c)$: （例）コーヒーであるから，コーヒーである．
N 型 $(\neg c \to \neg c)$: （例）これはコーヒーでない．したがって，コーヒーではない．

トートロジーにおいては，前件 r が実用的な内容を提供しないことから，前件 r を表す “P”, “N” を取り除くことによってトートロジーを表現する．また，PNN 型において，その一部分を構成する NN 型がトートロジーとなっている場合，PN 型に分類する．

3.3.4 アノテーションガイドライン

Viewpoint の収集を行った後，各 Viewpoint に対して，型を表すラベルを手動で付与する．ここで，Viewpoint $r \to (q, a)$ は，タスク結果のインスタンス (w, q, d, a, r) から得られる．a が「はい」であったときには，c が肯定形（P）となるため，PP 型, NP 型, PNP 型のどれかを型を表すラベルとして付与する．また，a が「いいえ」であったときには，c が否定形（N）となるため，PN 型, NN 型, PNN 型のどれかを型を表すラベルとして付与することとなる．ただし，例外として次のようなケースには，wrong 型とする．

- 回答 a が「はい」（あるいは「いいえ」）であるのにも関わらず，「いいえ」（あるいは「はい」）と回答した理由を入力するためのフォームにその回答理由を記述している．
- フォームに記述された内容が，理由を表す文章となっていない（「これはカレーですか？」，「あああああ」など）
- フォームに記述された内容が，文脈に沿ったものでない（質問「これはお茶ですか？」に対して「いいえ」と回答しているにも関わらず，入力された理由が「これは茶葉を使用して作られたものなので，お茶です」というように，「はい」の回答理由を記述している）

表 3.1 に，アノテーション例を示す．各 Viewpoint について，アノテーションの根拠を次から示す．
PP1. 「粉末緑茶という単語がパッケージに書かれている」が回答理由 r であり，「お茶である」が結論 c であるため，PP 型となる．

型	ID	質問文	回答	理由を入力するためのフォームに記述された文章
PP 型	PP1	これはお茶ですか？	はい	粉末緑茶という単語がパッケージに書かれているので，お茶である．
	PP2	これは炭酸飲料ですか？	はい	サイダーと記載してあるため．
NP 型	NP1	これはインスタント食品ですか？	はい	茶葉をポットに入れて作る，といった本格的な作り方の食品ではないので，はい，となりうる．
PN 型	PN1	これはコーヒーですか？	いいえ	ドレッシングはコーヒーではないので
	PN2	これはお茶ですか？	いいえ	コーヒー
NN 型	NN1	これは緑茶ですか？	いいえ	ミルクティーに緑茶の成分は入っていないから
	NN2	これは調味料ですか？	いいえ	調味料は，料理に味をつけるもの．
PNN 型	PNN1	これはコーヒーですか？	いいえ	これはハーブ茶であり，コーヒーは原料に含まれないので，コーヒーではない．
	PNN2	これはインスタント食品ですか？	いいえ	インスタント食品とは手軽に簡単に利用できるものだが，これは豆を煎ってから使用するものなので，いいえ，になりうる．

表 3.1: アノテーションの例

PP2. 「サイダーと記載してある」が回答理由 r であり，回答 a が「はい」であるため，結論 c が「炭酸飲料である」となり，PP 型となる．ここで，結論 c は判断 a によって決まるため，理由を入力するためのフォームに記述された文章には，結論 c が書かれていなくともよい．

NP1. 回答理由 r が「茶葉をポットに入れて作る，といった本格的な作り方の食品ではない」という否定形であり，回答 a が「はい」であるため，NP 型となる．

PN1. 一見すると NN に見える．しかし，文章中の「コーヒーでない」は回答理由 r を述べたものではなく，結論 c を述べたものである．回答理由 r を述べている部分は「(これは) ドレッシング」であるので，PN となる．このようなパターンの Viewpoint は非常に多かった．

PN2. 回答 a が「いいえ」であり，理由を入力するためのフォームに記述された文章は「これはコーヒーである．したがって，お茶ではない」と解釈できるため，PN 型となる．

NN1. 回答理由 r が「ミルクティーに緑茶の成分は入っていない」で，否定形であり，回答 a が「いいえ」である．したがって，NN 型となる．

NN2. これは複雑なケースである．理由を入力するためのフォームに記述された文章は「調味料は，料理に味をつけるもの」であり，否定形の表現は含まれない．しかし，以下の理由で NN 型となる．まず，回答 a が「いいえ」であるので，結論 c が「調味料でない」となり，候補は NN 型か PN 型，PNN 型に絞られる．つぎに，入力された文章「調味料は，料理に味をつけるもの」は「調味料であれば，料理に味をつけるものである」ということを意味する．これの対偶をとると「料理に味をつけるものでないならば，調味料ではない」という文章になる．回答 a から導かれた結論 c が「調味料ではない」であるので，この文章が厳密に Viewpoint を表す文章となっていると考えられる．したがって，回答理由 r は「料理に味をつけるものでない」，結論 c が「調味料ではない」であるので，NN 型となる．一般に，その Viewpoint に関して (1) 質問が「これは X ですか？」である．(2) 記述された文章に，X であるための必要条件が含まれる．(3) 回答 a が「いいえ」である．の 3 つを全て満たすとき，その Viewpoint を NN 型となる．

PNN1. 記述された文章には，「これはハーブ茶であり」「コーヒーは原料に含まれない」という 2 つの回答理由 $r_1, \neg r_2$ と，「コーヒーではない」という結論 c が含まれる．r_1 が肯定形，$\neg r_2$ が否定形であるので，PNN 型とする．このように，回答理由に肯定形と否定形の両方を含み，結論が否定形であるものが PNN 型の典型的なパターンであった．

PP 型	(P)NP 型	PN 型	NN 型	PNN 型	P 型	N 型	W 型	Total
247	2	824	110	95	23	4	108	1,413
17.5%	0.1%	58.3%	7.8%	6.7%	1.6%	0.3%	7.6%	100%

表 3.2: 各 Viewpoint の型の数と割合

PNN2. 記述された文章は，「インスタント食品とは手軽に簡単に利用できるもの」「これは豆を煎ってから使用するもの」という理由を表す部分と，「いいえ，になりうる.」という結論 c を表す部分からなる．理由を表す部分の 1 つ目に関しては，NN2 で示したパターンと同じで，(1) 質問が「これはインスタント食品ですか？」である．(2) 文章内に「インスタント食品とは手軽に利用できるもの」というインスタント食品であるための必要条件が含まれている．(3) 回答 a が「いいえ」である．という 3 つ条件が成り立つため，NN 型となる．また，理由を表す部分の 2 つ目については，回答理由 r が肯定形，回答 a が否定形である典型的な PN 型である．したがって，全体としてみると PNN 型となる．ここで説明したように，文章内に 2 つ以上の回答理由を含む場合，まずそれぞれの理由に関してどの型に当てはまるかを考えた上で，それらを結合することによって，（複合型の）Viewpoint の型を決定する．

3.3.5 インターアノテータアグリーメントとコーパスの統計

本項では，インターアノテータアグリーメントと，アノテーションしたコーパスの統計について報告する．ここで，インターアノテータアグリーメントとは，2 人以上のアノテータ間でどのくらい合意がとれるかを表す尺度である．アノテーションガイドラインに従い，収集した Viewpoint に対して 2 人がアノテーションを行った．その結果，2 人のアノテータ間で異なるアノテーションがされたのは 1,413 個中わずか 28 個であった．2 人のアノテータ間のアグリーメントを表す一般的な尺度である κ 統計量 [9] は，0.968 という非常に高い値となり，アノテーションガイドラインが明確であることが示された．

また，表 3.2 に，各 Viewpoint の型の数と割合を示す．W 型は wrong 型の略記である．この統計から，PN 型の数が多い一方，PNP 型,NP 型がコーパス内に 1 つしか存在しないことがわかる．PNP 型，NP 型が少ないのは，次のような理由によるものと考えられる．NP 型である Viewpoint $\neg r \rightarrow c$ の対偶をとると，$\neg c \rightarrow r$ となる．ここで，多くのケースでは，c を満たさないものは非常に多く存在する．そして，回答理由 r となりうるような，それらに共通する属性を表すシンプルなフレーズを提供することは難しいと考えられる．例えば，結論 c が「コーヒーである」であったとき，コーヒーでないものの共通の属性を言い表すことは簡単ではない．したがって，NP 型，PNP 型においては否定形の回答理由が作りにくく，その結果数が少なかったのだと考えられる．

4 実験

本章では実験を通して，論理構造を用いたアノテーションスキーマが，有用な Viewpoint を選択するのに役立つことを示す．ここで，本研究では，有用な Viewpoint を選択する基準として，汎用性を用いる．汎用性とは 1 章で述べたように，その Viewpoint が質問文に回答する際により多くのデータに対して役立つという性質である．汎用性の高い Viewpoint は，ワーカの質問文に対する解釈をはっきりと表している点で，質問文改善に有用であると言える．そこで実験では，クラウドソーシングを用いて Viewpoint の論理構造と汎用性の関係を調査する．本論文では，この実験を主観評価実験と呼ぶ．

さらに，実際に Viewpoint を用いて質問文の書き換えを行い，書き換え後のタスク結果が良くなったことを示すとともに，汎用性の低いものが多い特定の型は提示せずとも書き換えには影響がないことを示す．

また，有用な Viewpoint を選択する他の手法として考えられる，エントロピーを用いた手法を取り上げ，論理構造を用いた提案手法との比較を行う．

4.1 主観評価実験

■主観評価タスクの作成 3.3.1 項で述べたように，Viewpoint $r \to (q, a)$ を得るために，ワーカにある画像データ d を対象とした質問 q に対して回答 a を下した理由 r を尋ねた．主観評価タスクは，このようにして得られた各 Viewpoint が，他の画像データ $d'(\neq d)$ を対象とした質問 q に対して回答するときに役立つかを尋ねるものである．主観評価タスクを作成するため，収集した Viewpoint と，質問に対する正しい解答を含むデータ（ゴールドスタンダードデータ）を用いる．具体的には，タスクは以下のようにして作成する．

ステップ 1. (w, q, d, a, r) から得られた Viewpoint $r \to (q, a)$ それぞれに対して，正しい回答が a であるデータ $d'(\neq d)$ をランダムに選ぶ．これは，質問 q に対する回答を，d と d' で同じにするためである．ここで，型が wrong である Viewpoint に対しては，タスクを作成しない．その理由は，wrong 型の Viewpoint はそもそも意味をなさないためである．

ステップ 2. ステップ 1 によって選ばれた d' それぞれに対して，図 4.1 に示すような主観評価タスクを作成する．主観評価タスクには 2 つの質問が含まれる．1 つ目の質問は，その Viewpoint を収集したときに行った質問 q と同じものである．この質問は，提示する d' に対して，ワーカが正しい回答 a ができるかを確認するために行う．2 つ目の質問は，d' を対象とする質問 q に回答するときに，提示する Viewpoint が役立つかどうか問うものである．

このタスクについて複数人に問い合わせたときに，2 つ目の質問に対して「役立つ」と答えた人が多ければ，その Viewpoint はより汎用性が高いものであるといえる．言い換えれば，「役立たない」と答えた人が多ければ，その Viewpoint は特定のデータ d にしか当てはまらない，汎用性の低い Viewpoint ということになる．

■ワーカ 多数決をとるため，生成した各主観評価タスクに対して，重複のない 9 人のワーカに問い合わせた．その結果，主観評価タスクを行ったワーカの人数は，1058 人であった．

■実験結果 表 4.1 に，主観評価タスクの結果を示す．各行は，Viewpoint の型ごとに結果を集計したものとなっており，各行 2 列目は提示する d' に対して，少なくとも 1 人のワーカが正しい回答 a を行った Viewpoint の数である．また 3 列目と 4 列目は，2 列目で示した人数のうち，過半数が d' に対して「役立つ」と回答した Viewpoint について，それぞれ数と割合で示している．

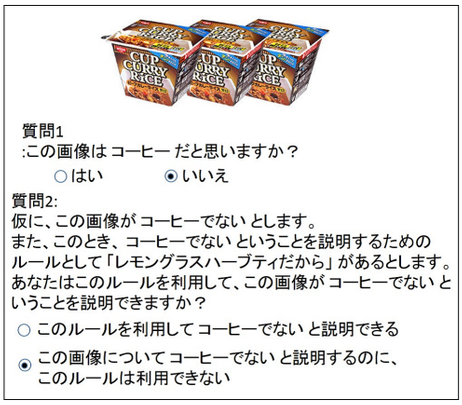

図 4.1: 主観評価タスク

type	1 人以上が正しい判断をした数	過半数が「役立つ」と判断した数	割合
PP 型	247	96	38.9%
NP 型	1	0	0.0%
PNP 型	1	1	100.0%
PN 型	824	115	14.0%
NN 型	110	48	43.6%
PNN 型	95	22	23.1%
P 型	23	22	95.7%
N 型	4	3	75.0%

表 4.1: クラウドが「役立つ」と判断した Viewpoint の数と割合

この結果から，他のデータに対して「役立つ」と答えた人の割合が，各型で明らかに異なることがわかる．重要なのは以下の点である．

- P 型と N 型は，他のデータに「役立つ」とされる Viewpoint の割合が明らかに高い．これは，P 型と N 型が「インスタント食品であるから，インスタント食品である」というようなトートロジーであり (3.3.3 項参照)，どのデータに対しても当てはまるものであったからであると考えられる．
- PN 型は，他のデータに「役立つ」とされる Viewpoint の割合が，PP 型と NN 型に比べて低い．これは，PN 型は「これはブラックコーヒーであるので，炭酸飲料ではない」など特定のデータ（ここでは，「ブラックコーヒー」）に対する判断にしか役立たないものが多かったからであると考えられる．言い換えれば，PN 型の Viewpoint $r \to c$ において，r が表現するデータの集合が小さいものが多いということであり，多くのケースで d' がその集合の中に含まれなかったということである．一方で，PP 型や NN 型における r は，「これは飲み物ではないから（炭酸飲料ではない)」など，特定のデータだけに依拠したものではなく，比較的汎用性が高いものが多いことが示されている．
- PNN 型については，若干 PN 型よりも，他のデータに対して「役立つ」と回答されている割合が高い．これは，PNN 型の Viewpoint は「これはお茶であり，食品ではない．したがって，インスタント食品ではない.」というような，汎用性の低い PN 型と比較的汎用性の高い NN 型の複合型であり，何人かのワーカが汎用性の低い PN 型の部分（ここでは，「これはお茶であり」）を無視したからであると考えられる．

「役立つ」と回答される割合は高ければ，汎用性が高い，すなわち r の表す集合が大きいということである．ここで，Viewpoint $r \to c$ において，r の表す集合は大きくなれば大きくなるほど c に近づいていく．したがって，「役立つ」と回答される割合は高ければ r と c が表す集合が似ているということが言え，r がそれぞれのワーカの c に対する解釈を表しているといえる．実験結果から，「役立つ」と回答される割合が比較的高かった PP 型，NN 型，PNN 型は c の特徴をうまく表現していると考えられ，リクエスタがワー

カの解釈を把握しやすいといえる．3.3.5 項で説明したように，NP 型と PNP 型の数は少なくなるため，これらを取り除いたとしても大きく Viewpoint の数を減らすことにはならない．一方で，PN の数はかなり多くなるため，これらを取り除けばかなりの数の Viewpoint を減らすことができる．また，P 型と N 型の Viewpoint はトートロジーであるため，割合が高くとも提案手法には役立たないことに注意されたい．

これらから，論理構造の観点から提案したアノテーションガイドラインが，収集した多くの Viewpoint の中から，有益なものを抽出する際の重要な手がかりとなることを示した．

4.2 主観評価実験結果の検証実験

主観評価実験の結果は，PN 型が汎用的でないことから有用でないことが示唆された．そこで，PN 型の Viewpoint が質問文の書き換えに影響を与えないことを検証するため，質問文の書き換えを以下の 2 つの設定で行ってもらい，それらをクラウドソーシングした結果の品質を比較した．

- 設定 A: 質問文を書き換える人に全ての Viewpoint を見せる
- 設定 B: 質問文を書き換える人に PN 型を除いた全ての Viewpoint を見せる．

■実験方法　まず，作業者に質問文の書き換えを行ってもらう．本実験には， 3.3.1 章でのタスク結果の品質が低かった 3 つの質問文「これはインスタント食品ですか？」「これは調味料ですか？」「これはお茶ですか？」を用い，各質問文に対して設定 A，設定 B それぞれで書き換えを行ってもらった．ここで，それぞれの作業者には 3 つの異なる質問文を書き換えてもらうが，その際 3 つすべての設定が同じにならないようにする．例えば，「これはインスタント食品ですか？」を設定 A で，「これは調味料ですか？」を設定 B で，「これはお茶ですか？」を設定 A で書き換えてもらうようにする．このような組み合わせは全部で 6 通りあるため，6 人の作業者を書き換えのために雇った．3 つの質問文の組み合わせを 6 人に書き換えてもらった結果，$3 \times 6 = 18$ 個の書き換えられた質問文を得た．

これらの 18 個の質問文それぞれと 28 個の画像データを組み合わせて，合計 $18 \times 28 = 504$ 個のタスクを生成し，それらをワーカに委託した．6 つのタスクを 1 セットとし，それぞれのワーカは 1 セット単位でにタスクを処理してもらった．報酬は 1 セットにつき 3 円に設定し，1 人のワーカにつき 3 セットまで処理できるようにした．

これら 504 個のマイクロタスクを Yahoo!クラウドソーシングを通して，作成されたタスク 1 つにつき，20 人のワーカに問い合わせ，20 個の結果を得た．その結果，全部で $504 \times 20 = 10,080$ の結果を得た．その際のワーカの数は 620 人であった．

■結果　表 4.2 に，書き換え前と書き換え後の質問文を用いたタスクの正答率を示す．設定 A と設定 B の両方の設定において，書き換え前よりタスクの正答率が上がっている．このことから，もとの質問文が曖昧であったとしても，Viewpoint を用いて質問を書き換えることによって高品質のタスク結果を得られることがわかる．ここで，「わからない」と回答した人の割合がわずかに増えている．これは質問文がより説明的になったことにより，ワーカが質問に回答することに慎重になったからであると考えられる．また，表 3.2 に示した通り，PN 型の Viewpoint が占める割合は 58.3% に上るが，それらを取り除いたとしても質問文の書き換えには大きな影響がないことがわかる．この結果は，提案した論理構造を用いたスキーマがより有用な Viewpoint を選択するのに有用であることを示している．

もとの質問文	もとの質問文による正答率	設定 A での正答率	設定 B での正答率
これはインスタント食品ですか？	75.5% (4.3%)	88.8% (4.5%)	90.1% (1.1%)
これは調味料ですか？	88.9% (1.4%)	96.6% (4.0%)	94.6% (1.5%)
これはお茶ですか？	91.2% (1.7%)	94.4% (5.3%)	92.5% (2.6%)

表 4.2: 書き換え前と書き換え後の質問文を用いたタスクの正答率．括弧内の割合は「わからない」と回答した人の割合を表す．

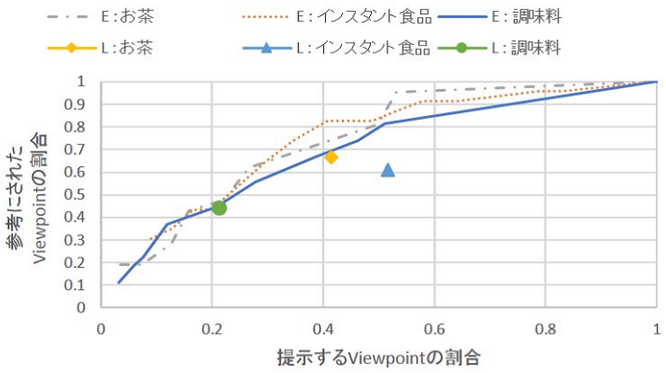

図 4.2: 比較実験の結果． L は論理構造を用いた手法，E はエントロピーを用いた手法を表す．

4.3 エントロピーを用いた手法との比較

有用な Viewpoint を選択するのに有効であると思われるもう一つの方法としてエントロピーを用いた手法が挙げられる．これは，回答の分布が広い，すなわち，「はい」「いいえ」などの選択肢に投票する人数がより割れているデータに対する Viewpoint を有用とする手法である．このエントロピーを用いた手法と提案手法である論理構造を用いた手法を実験によって比較した．その結果，同じタスクに対して複数の回答が必要ない提案手法が，エントロピーを用いた手法と同程度の効果が期待できることがわかった．

■実験方法 4.1 節において，6 人の作業者に質問文と Viewpoint を見せてタスクの書き換えを行ってもらった．その際，書き換えの参考にした Viewpoint を尋ねた．さらに，提案手法，エントロピーを用いた手法それぞれにおいて，有用であるとされた Viewpoint のうちに，作業者が書き換えの参考にした Viewpoint がいくつ含まれているのかを調べた．ここで，3.3.1 項において，それぞれのマイクロタスクにつき 20 人に問い合わせて複数（20 個）の回答を得ているため，エントロピーを計算することが可能である．

■実験結果 図 4.2 のグラフに結果を示す．x 軸が書き換えの際に見せた Viewpoint の数であり，y 軸は提示した Viewpoint に含まれる有用な Viewpoint の割合である．有用とされた Viewpoint は全体の平均で 9.3 個であった．エントロピーを用いた手法では，第 3.3.1 章で得られた回答から計算されたエントロピーが高い順に，Viewpoint を順位付けする．いま，t をマイクロタスク，a を t に対する回答とする．ここで，a の発生確率を与える関数 $P_t(a)$ が与えられたとき，エントロピー $H(t)$ は $H(t) = -\Sigma_a P_t(a) \log P_t(a)$ [10] として計算される．Viewpoint を提示する際は，上位の Viewpoint を提示するが，上位何位までの Viewpoint を提示するかによって，提示する Viewpoint の数が変わる．したがってグラフには複数の点がプロットされることになるが，提案手法との比較がしやすいように，図 4.2 ではそれらを線でつなぎ，点を表示しないようにしている．一方で，論理構造を用いた手法は，提示する Viewpoint の数は一定（PN 型以外の

Viewpoint の数）であるので，一点のみがプロットされる．

グラフでは，提案手法を表す点が，ほぼエントロピーを用いた手法を表す線上に存在するのが確認できる．したがって，提案手法は，エントロピーを用いた手法と同程度の品質で有用な Viewpoint を抽出することができるといえる．ここで，論理構造を用いた提案手法は，タスクに対して 1 人の回答しか必要とせず，1 タスクにつき必ず複数人の回答が必要になるエントロピーを用いた手法よりも少ないタスク数で用いることができる．

ただし，インスタント食品かどうかを質問したタスクでは，論理構造を用いた提案手法がエントロピーを用いた手法よりも，有用な Viewpoint の割合が 24.7% 低い．この理由は，PN 型の中にもいくつか汎用性が低くないものが存在し，それらは書き換えの役に立つとされていたからであった．このような Viewpoint を，論理構造を用いた提案手法ではフィルタしてしまうため，さらなる改良が必要であると考えられる．その一例として，PN 型でも，Viewpoint の回答理由が長い文章となっている場合はフィルタしない，といったものが考えられる．なぜなら，汎用性が低くない Viewpoint は説明的になり，回答理由が長くなると考えられるからである．このような提案手法の改良は今後の課題である．

4.4 分類実験

Viewpoint の数が増えるにつれ，それを分析するリクエスタの負担も増えてしまう．したがって，分類器による Viewpoint のフィルタリングがリクエスタにとって必要となる．ここで，PN 型の Viewpoint に汎用性が低いものが多く，PP 型，NN 型，PNN 型の Viewpoint は比較的汎用性が高いものが多いことは 4.1 節で既に確認した．本項では，PP 型，NN 型，PNN 型を有用な Viewpoint，それ以外を有用でない Viewpoint として，自動的に分類する分類器の作成方法と，その性能について述べる．

■分類器の作成 Viewpoint の型を推定するため，単純な線形分類器を用いる．ライブラリとして libliner [11] を用い，L2-regularized logstic regression を指定したうえで，その他のパラメータはデフォルトのものを使用した．文書分類と同じように，Viewpoint i を入力ベクトル $x_i \in R^m$ で表現し，出力ラベル y_i が Viewpoint の型を表すとする．入力ベクトルは，以下の 4 ステップで作成する．

(1) 各 Viewpoint $r \rightarrow (q, a)$ の q を分析し，質問対象物を特定する．例えば，q =”これはインスタント食品ですか？” の場合，「インスタント食品」が質問対象物となる．その後，Viewpoint の回答理由 r の文字列の中から，質問対象物を特別なシンボル T で置き換える．
(2) (1) で生成した文字列の中から，名詞を特定し，すべて特別なシンボル O で置き換える．
(3) (2) で各 Viewpoint に対して生成された文字列の終端に特殊なシンボルを付与した後，それらを鎖状につなぐ．このようにしてできた一つの文字列の中から，極大部分文字列 [12, 13] を抽出する．ここで，極大部分文字列は，本質的に分類器の重み学習に必要な部分文字列そのものである [12]．
(4) 各極大部分文字列の出現数を，入力ベクトルの特徴量として用いる．すなわち，入力ベクトル x_i の j 次元目の値 n を，対応する極大部分文字列がその Viewpoint i 内に出現した回数であるとする．
(5) 入力ベクトルの次元に値がバイナリとなる 5 つの次元を追加する．そのうち 3 つの次元は，「はい」「いいえ」「わからない」がそれぞれ選択されたかどうかを表すものである．残り 2 つの次元は「はい」の理由を入力するフォーム，「いいえ」の理由を入力するフォームに入力したかどうかを表すものである．この次元の追加は，wrong 型の Viewpoint をうまく検出するために行う．

訓練データとして，次の飲料に関する質問文から得られた Viewpoint の集合を用いた．

1. これはコーヒーですか？

2. これは炭酸飲料ですか？
3. これはお茶ですか？
4. これは紅茶ですか？
5. これは緑茶ですか？

検証データとして最後の質問文である「これは緑茶ですか？」から得られた Viewpoint の集合を用いた．また，テストデータとして，以下の 2 つの質問文から得られた Viewpoint の集合を用いた．

1. これは調味料ですか？
2. これはインスタント食品ですか？

このように，訓練データとテストデータで使用されている語彙が異なるように意図的に設定した．その理由は，学習された分類器が特定のドメインに依存していないことを示すためである．

■分類器の評価　評価に用いる尺度は，精度 (正しくラベル付けされた Viewpoint の割合)，適合率 P（分類器が「有用である（PP 型, NN 型, PNN 型のどれかである）」と判断した Viewpoint のうち，実際に有用であった Viewpoint の割合），再現率 R(有用な Viewpoint のうち，分類器が「有用である」と判断した Viewpoint の割合)，そして P と R の調和をとる F 値 F_1（$F_1 = 2PR/(P + R)$)）を用いる．テストの結果，精度は 85.3% であった，また，F 値について，最も良かった結果は適合率 $P = 85.8\%$ ，再現率 $R = 65.5\%$ となったときの $F_1 = 74.3\%$ であった．この結果は，現在のコーパスの大きさでは自動的なアノテーションは難しいことを示している．インターアノテータアグリーメントの値が高いことから，さらなる改善の余地があると考えられる．

5 おわりに

本論文では，質問文の改善を支援するため，回答と回答理由の組（Viewpoint）を収集し，リクエスタに提示することを提案した．また，収集した Viewpoint の数が増えると，リクエスタがそれら全てを分析することは難しい．そこで，玉石混淆な Viewpoint の中から，有用なものを選択する手法の提案も行った．有用なものを選択する基準として汎用性を用い，汎用性が Viewpoint の論理構造に関係していることを発見した．さらに，有用な Vewpoint を選択する他の手法としてエントロピーを用いる手法を取り上げ，論理構造を用いる提案手法との比較を行った．提案手法は，1 つのタスクにつき複数の回答を必要とせず，エントロピーを用いる手法と同程度の品質で，有用な Viewpoint を選択することができた．最後に，作成したコーパスを用いて，自動的に有用な Viewpoint を選択するための分類器を構築した．

今後の課題は次の 2 つである．(1) 汎用性が低く，有用でないと見なした論理構造（PN 型）の Viewpoint の中にも，汎用性が高いものが存在した．このような Viewpoint をとりこぼさないようにするには，PN 型であっても回答理由が長いものはフィルタしないなどの工夫が必要である．(2) 分類器の性能が十分であるとはいえない．これは，コーパスのサイズが小さすぎるからであると考えられる．したがって，クラウドソーシングを用いてさらに Viewpoint を収集し，それらに手動でアノテーションを行うことによって，コーパスのサイズを拡大する必要がある．

謝辞

本修士論文の執筆にあたって，たくさんの方々にお世話になりました．

主指導教員の森嶋厚行教授からは，研究のことだけにとどまらず，ものの考え方，上手なプレゼンの仕方，日々の生活において心がけるべきことなど多岐にわたりご指導頂きました．深く感謝致します．

また，自分の研究に際して，ヤフー株式会社の清水伸幸様には毎週の打ち合わせをはじめ熱心なご指導を頂いたの同時に，Yahoo!クラウドソーシングを用いた実験でもご協力を頂きました．深く感謝いたします．

関洋平教授には，副指導教員を引き受けて頂き，研究に対して鋭いご指摘を頂きました．深く感謝致します．

さらに，森嶋研究室の方々には，ゼミでのご意見や論文・プレゼンの添削等で大変お世話になりました．先輩の福角さん，丹治さん，権守さん，櫻井さん，平木さん，同期の根本君，太田さん，後輩の熊井君，中村君，米良君，鈴木君，佐々木さん，林君，橋本君，水澤君，小林君，岩本君に感謝いたします．

加えて，合同ゼミでたくさんのご指導を頂いた杉本重雄教授，阪口哲夫准教授，永森光晴講師，および合同ゼミメンバの皆様に感謝いたします．

参考文献

[1] Melenhorst, Mark and Menéndez Blanco, María and Larson, Martha. A Crowdsourcing Procedure for the Discovery of Non-Obvious Attributes of Social Images. Proceedings of the 2014 International ACM Workshop on Crowdsourcing for Multimedia. International ACM Workshop on Crowdsourcing for Multime-dia (CrowdMM '14). ACM, New York, NY, USA, 45-48. 2014.

[2] Katja Wiemer-Hastings and Peter Wiemer-Hastings. DP: A Detector for Presuppositions in Survey Questions. In Proceedings of the Sixth Conference on Applied Natu- ral Language Processing (ANLC '00). Association for Computational Linguistics, Stroudsburg, PA, USA, 90-96. 2000

[3] Tremper, Galina. Weakly Supervised Learning of Presupposition Relations Between Verbs. Proceedings of the ACL 2010 Student Research Workshop. 97-102, 2010

[4] Weisman, Hila and Berant, Jonathan and Szpektor, Idan and Dagan, Ido. Learning Verb Inference Rules from Linguistically-motivated Evidence. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 194-204, 2012

[5] Djellel Eddine Difallah and Gianluca Demartini and Philippe Cudré-Mauroux. Pick-a-crowd: tell me what you like, and i'll tell youwhat to do. 22nd International World Wide Web Conference, WWW'13, Rio de Janeiro, Brazil, May 13-17, 2013. 367-374. 2013

[6] Ailbhe Finnerty andPavel Kucherbaev andStefano Tranquillini andGregorio Convertino. Keep it simple: reward and task design in crowdsourcing. Biannual conference of the Italian chapter of SIGCHI, CHItaly '13,Trento, Italy - September 16 - 20, 2013. 14:1-14:4. 2013.

[7] Nguyen Quoc Viet Hung and Nguyen Thanh Tam and Ngoc Tran Lam and Karl Aberer. An Evaluation of Aggregation Techniques in Crowdsourcing. Web Information Systems Engineering - WISE 2013 -14th International Conference, Nanjing, China, October 13-15, 2013,Proceedings, Part II. 1-15. 2013.

[8] Srikanth Jagabathula andLakshminarayanan Subramanian andAshwin Venkataraman. Reputation-based Worker Filtering in Crowdsourcing. Advances in Neural Information Processing Systems 27:Annual Conferenceon Neural Information Processing Systems 2014, December 8-13 2014,Montreal, Quebec, Canada. 2492-2500. 2014.

[9] Artstein, Ron and Poesio, Massimo. Inter-coder Agreement for Computational Linguistics. Comput. Linguist. 555-596. 2008.

[10] Christopher D. Manning and Prabhakar Raghavan and Hinrich Schütze. Introduction to information retrieval. Cambridge University Press. 2008.

[11] Fan, Rong-En and Chang, Kai-Wei and Hsieh, Cho-Jui and Wang, Xiang-Rui and Lin, Chih-Jen. LIBLINEAR: A Library for Large Linear Classification. J. Mach. Learn. Res. 1871-1874. 2008

[12] Okanohara, D. and Tsujii, J. Text Categorization with All Substring Features. SIAM International Conference on Data Mining (SDM). 838-846. 2009.

[13] Gallé, Matthias. The Bag-of-repeats Representation of Documents. Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. 1053-1056. 2013.

発表論文一覧

1. 林亮太, 清水伸幸, 森嶋厚行. “クラウドソーシング用マイクロタスク設計支援のためのユーザフィードバックの収集手法”. 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016), 7pages, 福岡, 2015-3.