

# クラウドソーシングを用いた仮説入手・検証の自動化

米良 俊輝<sup>†</sup> 平木 理恵<sup>††</sup> 若宮 翔子<sup>†††</sup> 森嶋 厚行<sup>††††</sup> 荒牧 英治<sup>†††</sup>

<sup>†</sup> 筑波大学 情報学群情報メディア創成学類 〒305-8550 茨城県つくば市春日 1-2

<sup>††</sup> 筑波大学 図書館情報メディア研究科 〒305-8550 茨城県つくば市春日 1-2

<sup>†††</sup> 奈良先端科学技術大学院大学 〒630-0192 奈良県生駒市高山町 8916-5

<sup>††††</sup> 筑波大学 知的コミュニティ基盤研究センター 〒305-8550 茨城県つくば市春日 1-2

E-mail: <sup>†</sup>toshiki.mera.2015b@mlab.info, <sup>††</sup>s1421602@u.tsukuba.ac.jp, <sup>†††</sup>{wakamiya,aramaki}@is.naist.jp, <sup>††††</sup>mori@slis.tsukuba.ac.jp

あらまし 計算機のみでは解決困難な問題の1つとして、仮説の形成・検証がある。本研究では、仮説のもっともらしきや内容の重複を判断する指標を定義して用いることで、専門家による介入なしに、計算機処理とクラウドソーシングを組み合わせて一連の作業を自動で行う手法を提案する。しかし、自動化する上で単純な手法を用いたのではタスクサイズやタスクの計算量が大きくなる問題があるため、タスクサイズの固定化やキャッシュを用いた効率化の手法も合わせて提案する。また、実際のクラウドソーシングサービスを用いた実験によって本手法を評価するとともに、実験の結果に基づくシミュレーションを行うことで効率化についての評価を行う。

キーワード クラウドソーシング, ヒューマンコンピューテーション, 処理効率化

## 1. はじめに

近年、問題解決のために必要な作業(タスク)を、ネットワーク上の群衆(ワーカ)に依頼するクラウドソーシングシステム [1] が注目を集めている。クラウドソーシングシステムは Human Computation [2] を大規模かつ効率的に実現するために活用されることがある。Human Computation とは、人間を明確な計算資源とみなしてシステムに組み込み作業を行わせることである。これにより計算機には解決困難な問題を人間の力を利用して解決する研究が行われている。

計算機にとっては困難であり人間には比較的容易な作業の一つとして、仮説の形成がある。例えば、ある疾患が最近急増している原因を知りたいという事例を考える。これについて、計算機では既知の解を検索することは可能でも、新たな仮説をゼロベースで形成して答えを導くことは困難である。しかし人間であれば、自身の経験などからもっともらしい仮説を考えついで原因を導くことができる可能性がある。

そこで、このような仮説をクラウドソーシングによって入手することを考える。このときワーカから回答として得られる仮説の内容は玉石混淆であり、その中からもっともらしい仮説を選別することは容易ではない。

クラウドソーシングで仮説を入手する研究はこれまでも行われている。医学研究の分野では、成人の肥満に関する仮説をクラウドソーシングで収集し、新しい有用な仮説が得られることなどが示されている [3]。加えて、疾病の原因に関する仮説の入手をクラウドソーシングで行い、入手した仮説を再びクラウドソーシングによって検証する手法の有用性を示す研究も行われている [4]。しかしこれらの研究では、クラウドソーシングによって入手した多数の仮説の中から、もっともらしい仮説を選別する作業を全て専門家の手作業に委ねており、大きな手間と

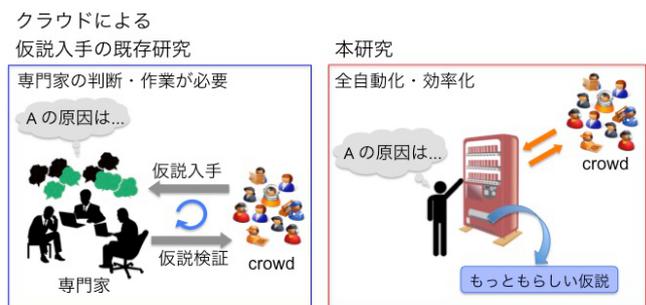


図1 本研究の目的

コストがかかっていた。

そこで本論文では、ある問題に関するもっともらしい仮説を入手・検証する流れを、クラウドソーシングと計算機処理を組み合わせることで専門家の力を借りずに自動化する手法を提案する。提案手法では、図1のように「Aの原因を知りたい」という入力を依頼者から受け取り、最終的に「Aの原因となるもっともらしい仮説」を出力する。これらの処理はクラウドを介して全自動で行われるが、後述するように、単純な手法ではタスクサイズやタスクの生成時間が非常に大きくなって非効率的である。したがって、本論文では(1)常に一定数の仮説を選択して検証することでタスクサイズを固定し、さらに、(2)仮説の最ももらしきの偏りに着目した中間結果のキャッシュによる効率化手法についても合わせて提案する。そしてこれらの手法の評価のために実際のクラウドソーシングサービスを利用した実験についても報告する。

本論文の貢献は次の通りである。

(1) 完全自動の仮説入手・検証。我々の知る限り、本論文はクラウドソーシングを利用した仮説・検証の全自動化について議

論する初の論文である。専門家による仮説の判断・作業を必要とせず、統計的な手法を利用してもっともらしい仮説を選択する。これまでにクラウドソーシングによって仮説を入手・検証する試みは存在しているが、人手を挟まず完全に自動で行う研究は存在しておらず、この点は大きな貢献と考える。

(2) シミュレーションと実験による検証。完全自動化に関しては、実際にクラウドソーシングにより実験をおこなった結果を報告する。また、中間結果のキャッシュによる効率化手法に関しても、実際の実験データを用いたシミュレーションを行った結果を報告する。

本論文の構成は次の通りである。まず、2. 節では本研究の関連研究について述べる。次に3. 節では仮説の入手・検証の自動化をどのように実現するかについて述べる。4. 節では、単純な手法を用いたのでは膨大になるタスクサイズを一定に保つための手法について説明する。5. 節では、4. 節における仮説の選択を行う際に、仮説ごとの選択される確率に注目し、中間結果を保持することで処理を効率化する手法について説明する。6. 節では提案手法の有効性を確かめるための実験について述べる。7. 節はまとめと今後の課題である。

## 2. 関連研究

これまで、クラウドソーシング処理の効率化に関しては様々なものが議論されてきた [5] [6] [7] [8] [9] が、仮説の入手・検証の処理効率化については我々の知る限り存在しない。

また、主に医学研究の分野において、仮説をクラウドソーシングで入手し、もっともらしいものを選別する研究はいくつか存在する。例えば、成人の肥満に関して幼少期に見られる兆候をクラウドソーシングで収集し、まだ十分に立証されていない新しい有用な仮説を入手する研究がある [3]。しかし、これは仮説の入手のみをクラウドソーシングによって行うものであり、仮説を入手した後の選別や評価はすべて専門家が行う。また、アレルギーの原因に関する仮説をクラウドソーシングで入手し、新たなリスクを発見し、既知のリスクは再発見できることを示した研究 [4] も存在する。これは、仮説の入手以外に、入手した仮説の検証もクラウドソーシングで行うものであったが、入手した仮説の中から検証するべき仮説を選ぶ作業は専門家が手作業で行う。一方、本研究ではこれらの専門家による作業をなくし、全自動化することを目的としている。具体的には、仮説を評価する指標を定義して確率・統計的な処理を行うことで、既存研究では仮説入手・専門家による選別・仮説検証というステップを繰り返して行っていたものを、専門家の作業なしで仮説入手と検証を同時に実施できるようにすることで実現することを考える。

## 3. 仮説入手・検証の自動化手法

本節では、実際に仮説を入手しそれらを検証するために用いるタスクの内容、及びそのタスクへの回答をもとに仮説のもっともらしさをどのようにして定義するかについて述べる。

### 3.1 概要

提案手法の概要を図2に示す。また、提案手法における入力

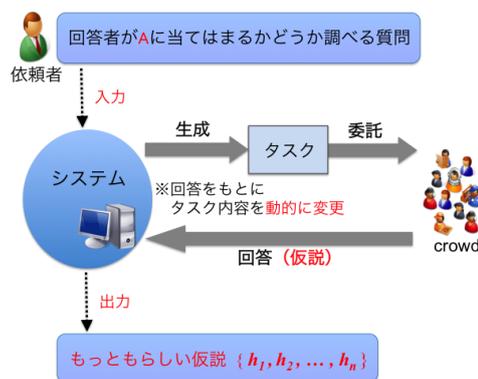


図2 提案手法の概要

図3 タスク内の質問文の一例

と出力は次の通りである。

入力：

回答者の状態が  $A$  であるか  $\neg A$  であるかを判定する質問

出力：

$A$  の原因として考えられるもっともらしい仮説  $\{h_1, h_2, \dots, h_n\}$

提案手法では、最初に入力を受け取るとそれをもとにタスクを生成してクラウドソーシングプラットフォームへ投入する。そして、回答として得られた仮説に当てはまるかを尋ねる質問を生成して検証のための質問とし、新しいタスク内に動的に追加する。それを再びクラウドソーシングプラットフォームへ投入するという流れを繰り返す。このように仮説の入手とその検証を並行して行うという形式をとる。

### 3.2 提案手法で用いるタスク

図3のようにタスクはアンケート形式となっており、次に示すような三つの質問群に大別される。

#### 質問 A アウトカム質問

回答者の状態が  $A$  か  $\neg A$  かを判定するための質問。

#### 質問 B 仮説プール投入質問

回答者から仮説  $h_i$  を得るための質問。文末を指定した自由入力形式とする。また、回答として得られた仮説の集合を仮説プールと呼ぶ。

#### 質問 C 仮説検証質問

上記の質問 B で得られた仮説から、その仮説に当てはまるか

表 1 集計表の例

T	仮説 $h_i$	
	YES ( $h_i$ )	NO ( $\neg h_i$ )
状態: A	a (人)	b (人)
状態: $\neg A$	c (人)	d (人)

どうかを尋ねる質問として生成する。それを以降で生成されるタスク内に掲載することで、その仮説のもっともらしさを検証するための質問。例えば図 3 の中で、質問 B に対して「虫歯になる (ことが原因だと思う)」という入力を得たならば、「虫歯になる ことがある □ はい □ いいえ」のような質問を生成して、以降のタスクに質問 C として掲載する。

はじめに依頼者が入力したアンケートの質問 A および質問 B を含むタスクを生成し、以降は入手したタスクの回答をもとに質問 C の内容を追加した新たなタスクをシステムが動的に生成する。

### 3.3 仮説のもっともらしさ

前述したタスクの回答をもとに、仮説のもっともらしさを数値化するための指標について定義する。ある仮説のもっともらしさを表す値を  $P(h_i)$  とする。この値は、表 1 のような、質問 A において判定した回答者の状態ごとに質問 C への回答数を集計した表 ( $T$  とする) の値から計算して求めるものとする。すなわち、 $P(h_i)$  の値は、集計表  $T$  の値を引数とする関数  $f$  として計算できる値、 $P(h_i) \equiv f(a, b, c, d)$  と定義する。

関数  $f$  の計算に用いる具体的な指標は、この集計結果の値を用いて算出できる範囲で、扱う問題に応じて適当なものを利用するものとする。例えば、相関係数やオッズ比などである。そして求めた  $P(h_i)$  の値が大きいほどもっともらしい仮説であるとみなす。

## 4. 仮説の選択によるタスクサイズの固定

前節の質問 C において、これまで入手したすべての仮説を質問としてタスクに掲載すると、実施タスク数に比例してタスク内の質問数が線形に増え続けタスクサイズが膨大になる問題がある。そこで本節では、入手したすべての仮説をタスクに掲載するのではなく、ある基準に基づいて一定数の仮説を選択して掲載することでタスクサイズを常に一定に保つ手法について考える。

### 4.1 仮説選択の目標

提案手法において、仮説を選択するための目標は次の通りである。

**目標 1.** 最終的にもっともらしい仮説として出力されそうな仮説を選択する。理想的には、最終的に選択されない仮説は不要であるため、できるだけ最終的に仮説として出力される可能性が高いものを選択したい。

**目標 2.** 内容が本質的に同じ仮説は選択しない。重複した内容の仮説ばかりを選択して一つのタスクに掲載すると、内容が本質的に異なる他の仮説が選択される機会を奪うことになる。そのため、内容が本質的に同じ仮説は除外して選択されること

## Algorithm 1 仮説選択のアルゴリズム

**Input:**  $m, H_t$

**Output:**  $H$

```

1:  $H \leftarrow \phi$ 
2: while ( $|H| < m$ )  $\wedge$  ( $H_t \neq \phi$ ) do
3:    $h_i \leftarrow \text{getByWeightedRandom}(H_t)$ 
4:   delete  $h_i$  from  $H_t$ 
5:   if  $\text{checkCorrelate}(h_i, H)$  is true then
6:     add  $h_i$  to  $H$ 
7:   end if
8: end while

```

が望ましい。例えば、3.2 節の質問 B で「虫歯になる」と「虫歯がたくさんある」という二つの仮説が得られた場合、これらは本質的には内容が同じであると考えられるため、一方の仮説は除外して選択を行う。

### 4.2 仮説選択のアルゴリズム

本節では 4.1 節で述べた 2 つを目標に、各タスクに掲載する有限個の仮説を選択する手法を提案する。具体的には、ワーカーからタスクを要求され、発行する際に、Algorithm 1 を呼び出す。これは、その時点での仮説プールから  $m$  ( $\in \mathbb{N}$ ) 個の仮説を選択して出力するアルゴリズムである。ここでは、ある時点  $t$  で入手済みである全ての仮説の集合を  $H_t$ 、次のタスクで質問 C として掲載するために選択済みの仮説の集合を  $H$  と定義する。

Algorithm 1 では、目標 1,2 を実現するためにそれぞれ下記に説明するヒューリスティックスペースの手法を用いる。

**目標 1 を実現するための手法:** これに関しては、その時点での  $P(h_i)$  の値に基づく重み付きランダムによって仮説の選択を行う。この重み付きランダムにおいて、各  $P(h_i)$  が選択される確率  $Q(h_i)$  は次で定義される。

$$Q(h_i) = \frac{P(h_i)}{\sum_{h_j \in H_t} P(h_j)}$$

この考え方はある回答数の時点でもっともらしい仮説は最終的にもっともらしい仮説である可能性が高い、というヒューリスティクスに基づくものである。Algorithm 1 中の  $\text{getByWeightedRandom}$  関数は、 $H_t$  を引数として受け取り、要素を一つ返すものであり、この手法を実装している。

**目標 2 を実現するための手法:**  $H_t$  から  $H$  に追加する候補として  $h_i$  が選択されたとき、もし、その時点での  $H$  に、ある  $h_j$  が存在し、 $h_i$  と  $h_j$  に相関がある場合は、 $h_i$  は  $H$  には追加しない。具体的には、相関の計算時にこれまでの質問 C の結果から表 2 のようなクロス集計表を作成する。これは、内容が重複する仮説は、回答に強い相関が出るというヒューリスティクスに基づくものである。Algorithm 1 中の  $\text{checkCorrelate}$  関数は、 $h_i$  と  $H$  を引数として受け取り、 $h_i$  と全ての  $h_j \in H$  との間で相関を計算する。相関のある組み合わせが 1 つでも存在すれば  $\text{false}$  を返す。逆に全ての組み合わせにおいて相関がない場合には  $\text{true}$  を返す。

Algorithm 1 全体の動作は次の通りである。まず入力として、

表 2 相関計算に用いるクロス集計表の例

		仮説 $h_i$	
		YES ( $h_i$ )	NO ( $\neg h_i$ )
仮説 $h_j$	YES ( $h_j$ )	$a$ (人)	$b$ (人)
	NO ( $\neg h_j$ )	$c$ (人)	$d$ (人)

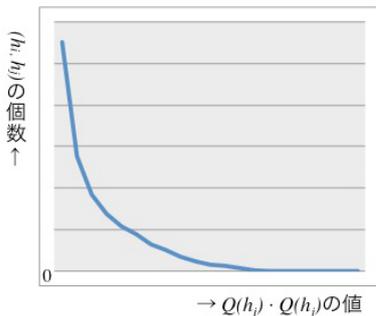


図 4 予想される  $Q(h_i) \cdot Q(h_j)$  の値ごとの組み合わせの数の偏り

$m$  と  $H_{all}$  をとる。1 行目では、 $H$  を用意する。2 行目以降は、 $H$  の要素数が  $m$  と等しくなるか  $H_i$  の要素数が 0 以下になるまで繰り返される。3 行目では、getByWeightedRandom 関数によって  $H_i$  から要素  $h_i$  を一つ得る。4 行目では、 $h_i$  を  $H_i$  から除外する。5 - 7 行目では checkCorrelate 関数によって  $h_i$  を  $H$  に追加するかどうかを判定し、true の場合は  $h_i$  を  $H$  に追加する。

## 5. 中間結果キャッシュの利用による処理効率化

提案手法では、以前のタスク結果に基づいてタスクを毎回動的に生成しているが、その計算量は必ずしも小さくない。特に、4.2 節の checkCorrelate 関数における相関の計算では仮説同士の組み合わせを考える必要があるが、これは仮説の総数が  $n$  個の時に全体で約  $n^2$  通り存在する。しかし、クラウドソーシングによるタスク処理を想定しているため、タスク生成にかかる時間はできるだけ短い事が望ましい。

### 5.1 中間結果キャッシュ

効率化の一つのアプローチは、記憶領域を利用して、空間計算量を増やす代わりに時間計算量を減らすことである。本提案手法では、仮説はもっともらしさ  $P(h_i)$  の値に応じた確率  $Q(h_i)$  によって選択されるが、 $Q(h_i)$  においては確率が低いものが占める割合が大きく、図 4 のような偏りが存在すると予想される。したがって、選択される確率が高い仮説の組み合わせに関するクロス集計表を固定長のキャッシュで保持し、キャッシュに無ければ再計算を行うという手法が効果的であると予想される。

本論文では、このクロス集計表を  $d$  個格納可能なキャッシュを用意し、サイズを超えた場合には  $Q(h_i) \cdot Q(h_j)$  が小さい値を持つクロス集計表から捨てていく手法を提案する。キャッシュに存在しないクロス集計表を必要とする場合には、これまでの計算結果から再計算してクロス集計表を求めるものとする。

## 6. 評価実験

本節では、これまでに示した自動化および効率化の手法についての評価を行う。はじめに 6.1 節で評価実験の概要について説明し、6.2 節でタスク内容や用いる手法についての詳細を述べる。6.3 節では実験により得られた仮説内容の評価について説明する。6.4 節で実験結果を利用したシミュレーションについて説明し、6.5 節では実験とシミュレーションの結果について述べる。6.6 節は結果についての考察である。

### 6.1 実験概要

評価実験では、Yahoo!クラウドソーシング [10] と Crowd4U [11] を利用して提案手法についての評価を行った。実験では 2818 タスクを実施した。実験で扱う問題は「うつ病となる原因は何か」とし、入力として回答者がうつの状態か否かを判定するための質問を受け取ってタスクを生成し、Yahoo!クラウドソーシングを通してワーカーに委託した。うつ病を選択した理由は、すでに専門家によって研究が進められているものであり、入手された仮説の妥当性の判断が可能であるからである。

5.1 節で述べたキャッシュを利用した効率化に関しては、入手した回答データをもとに提案手法を用いたシミュレーションを行い、提案手法と他の手法との結果を比較した。

### 6.2 実験の詳細

本節では実際のタスクの内容のほかに、実際に仮説選択に用いる指標や実験の流れについて説明する。

#### 6.2.1 タスク内容

本実験で用いたタスクの内容について説明する。

##### • 質問 A

図 5, 図 6 に示すような、回答者がうつの状態であるか否かを判定するための質問を行う。具体的には、ベック式抑うつ評価尺度 [12] を用いて回答者の状態を判定する。これには回答が 4 択である 21 問の質問を用いる。質問の各回答には点数が設定されており、その合計が一定値以上となった場合に臨床的な意味でのうつ状態みなす。

##### • 質問 B

図 7 に示すような、「うつになるのは、どうしてだと思いますか」という質問を行う。回答フォームで、「ことが原因だと思う」に接続するような文字列を入力させるものとする。これによってうつになる原因に関する仮説を入手する。

##### • 質問 C

図 8 に示すような、質問 B で入手した仮説の文章の末尾に「ことがありますか?」を付加した、その仮説を検証するための質問を行う。 $m = 10$  とし、仮説プールから提案手法によって 10 個の仮説を選択して掲載する。

#### 6.2.2 仮説選択に用いる指標

本実験における仮説選択に用いた指標やパラメータについて説明する。

##### (1) 仮説のもっともらしさ

本実験では、その仮説のもっともらしさを示す値  $P(h_i)$  にはオッズ比を用いる。これは医学分野の関連研究 [3] においても採用されている指標である。 $P(h_i)$  は 3.3 節の表 1 の値を用い

- ◇該当するものを選んでください。
- 1)  私は落ち込んでいない  
 私は落ち込んでいる  
 私は落ち込んでいるから急に元気にはなれない  
 私はとてつもないほど落ち込んでいて不幸だ
  - 2)  私の将来について特に失望していない  
 私の将来について失望している  
 私の将来に期待するものはない  
 私の将来には希望がもてないし、物事は良くならないと思う
  - 3)  私は自分が失敗するとは思わない  
 私は他人よりは失敗してきたと思う  
 今までのことを考えると失敗を繰り返してきたと思う  
 私は人間としてまったくだめだと思う
  - 4)  日常生活では大変満足している  
 日常生活の出来事を楽しんでいない  
 私は何にも本当に満足できない  
 私はどんなことにも満足できないし退屈だ
  - 5)  私は特に罪悪感をもっていない  
 私は時々罪悪感を感じている  
 私は多くの時間罪悪感を感じている  
 私はいつも罪悪感を感じている
  - 6)  私は罰を受けている（いわば罰が当たっている）とは思わない  
 私は罰せられているかもしれないと思う  
 私は罰せられるだろうと思う  
 私は罰せられていると思う
  - 7)  私は自分自身に失望していない  
 私は自分自身に失望している  
 私は自分自身にいやがさがしている  
 私は自分自身が嫌いだ
  - 8)  私は自分は他人よりは良くないとは思わない  
 私は自分の弱さや失敗について自分自身を責めている  
 私は自分の欠点をいつも自分のせいにする  
 嫌なことが起こるとそれを自分自身のせいにする
  - 9)  自殺について考えたことはない  
 自殺について考えたことはあるが実行したことはない  
 自殺したいと思う  
 もし機会があったら自殺するだろう
  - 10)  私はいつもより悲しい気持ちではない  
 私はいつもより悲しい気持ちでいる  
 私はいつも悲しい気持ちでいる  
 私はいつも泣いていたが、今では泣きたいと思っても泣けない

図 5 質問 A の内容 (1)

- 11)  私はこれまでのようにいららしない  
 私は今までより簡単に悩んでしまうし、いららする  
 私はいつもいららを感じる  
 私はいららすらも感じなくなった
- 12)  私は他人に対する興味を失っていない  
 私は以前より他人に興味を持たなくなった  
 私は他人への興味を大部分失った  
 私は他人への興味を失った
- 13)  私は自分なりの判断力がある  
 私は今までのような判断力に乏しい  
 私は以前よりも物事の判断に困難を感じる  
 私は何も判断することができない
- 14)  私は自分の現実よりも悪くは見えない  
 私は年をとり魅力が失って見えるのではないかな気になる  
 私はだんだん魅力がなくなったように思う  
 私は自分の見かけが見苦しくなってきたと信じている
- 15)  私は以前と同様に仕事ができる  
 何かをしよとすると昔よりも余分な努力をしなければならぬ  
 何かを始めるときずっと頭張らなくてはならぬ  
 私は何もしたくない
- 16)  私はいつものように良く眠れる  
 私はいつものように良く眠れない  
 私は以前よりも 1～2 時間早く目が覚めるしそれからはなかなか眠れない  
 私は以前より数時間早く目が覚めるし再び眠れない
- 17)  私はいつもより特に疲れたりほしくない  
 私は以前よりも簡単に疲れる  
 私は何をやっても疲れる  
 私は余り疲れるので何もできない
- 18)  食欲は普通だ  
 食欲は以前よりは良くない  
 今は食欲がない  
 食欲が全くない
- 19)  最近大きな体重の減少はない  
 最近 2 kg 以上体重が減った  
 最近 5 kg 以上体重が減った  
 最近 7 kg 以上体重が減った
- 20)  私は健康について特に気にしない  
 私は体の問題について気にしている  
 私は体のことが大変気になるので他のことを考えるゆりがない  
 体の問題について大変悩んでいるので他のことは何も考えられない
- 21)  性についての興味は特に変わっていない  
 以前より性に対する興味が増した  
 今では性に対する興味が大変減少した  
 性に対する興味は全くなくなった

図 6 質問 A の内容 (2)

◇うつになるのは、どうしてだと思いますか？

ことが原因だと思う。  
 (例) [自分のことを過剰に意識する]ことが原因だとおもう。

図 7 質問 B の内容

考えすぎることがありますか？	<input type="radio"/> はい <input type="radio"/> いいえ
自分を好きになれないことがありますか？	<input type="radio"/> はい <input type="radio"/> いいえ
まわりを気にしすぎる ことがありますか？	<input type="radio"/> はい <input type="radio"/> いいえ
考えすぎることがありますか？	<input type="radio"/> はい <input type="radio"/> いいえ
人間関係の悪化による ことがありますか？	<input type="radio"/> はい <input type="radio"/> いいえ
ストレス ことがありますか？	<input type="radio"/> はい <input type="radio"/> いいえ
後ろ向きになる ことがありますか？	<input type="radio"/> はい <input type="radio"/> いいえ
完璧な自分を求める ことがありますか？	<input type="radio"/> はい <input type="radio"/> いいえ
考え過ぎることがありますか？	<input type="radio"/> はい <input type="radio"/> いいえ
友達が少ない ことがありますか？	<input type="radio"/> はい <input type="radio"/> いいえ

図 8 質問 C の内容の一例

て次の式で表される。

$$P(h_i) = (a \times d) / (c \times b)$$

## (2) 仮説間の相関

仮説同士の内容の相関を調べるために相関係数を求める。仮説を検証するための質問のうち、ある二つに対して同時に回答した人数というのは、タスク実施数と比較して値の増加が非常に緩やかであるため、独立性の検定を行うのに十分な標本数を得ることは困難である。本研究の提案手法では、タスクを実施しつつ、ある時点での回答同士の相関を使って仮説の内容の相関を予想するというヒューリスティックスペースの手法を用いている。したがって検定によって厳密に相関を調べるのではなく、相関係数が設定した閾値を超えた場合に強い相関があると判定することとした。具体的には、クラメールの連関係数が 0.5 を超えた場合に、相関があると判定する。

## (3) 新規入手仮説の優先

仮説の選択は基本的に前述した二つの指標によって行う。しかし、入手した直後の仮説は全て、本実験における  $P(h_i)$  としたオッズ比の値が 0 となる。仮説はこの値を重みとしてランダム選択されるため、入手した直後の仮説はそのままでは選択さ

れない。したがって新規に入手した仮説に関しては、一定の回答数を得るまでの間は例外的に固定値の  $P(h_i)$  を割り当てることで優先して選択されるようにする。本実験においては、新規に入手した仮説に対して回答数が 10 に達するまでは  $P(h_i)$  に固定値として 5 を割り当てて実験を行った。

### 6.2.3 実験の流れ

はじめに入力として受け取った質問 A および質問 B を掲載したタスクを、並列処理のために 100 件生成する。以降は 1 タ

スク実施されるたびに1タスク生成するため、101タスク目以降からは、動的に生成された質問Cが掲載されたタスクが終了まで続く。

### 6.3 入手した仮説内容の評価

本手法の有用性を示すために、共同研究グループの医学研究者に、医学的見地から入手した仮説内容の評価を依頼した。これまでに知られている仮説においては、脳で生じる生物学的問題のような自分では認識できないような専門的な仮説については再発見できなかったが、本人が認識できる状態に関する仮説については250タスク実施後の上位10個の仮説にて網羅的に再発見することができているとの報告を受けた。

したがって、既にある程度一般に認知されている問題に対して、回答者本人が認識できる状態に関する仮説を入手することにおいては一定の有用性がある場合があることがわかった。

### 6.4 実験結果に基づいたシミュレーション

実験によって得たタスクへの回答データを用いて、中間計算の中間結果を保持するキャッシュの有効性を確認するためのシミュレーションを実施した。シミュレーションは、キャッシュサイズ  $d$  およびキャッシュの削除手法を変えて実施し、それぞれでキャッシュ上に存在しない場合に再計算を行った回数を比較した。本シミュレーションにおいては、5.1節で提案した  $Q(h_i) \cdot Q(h_j)$  が高確率のものを保持、つまり低確率順に消していく手法の他に、入手が古い順、ランダム(平均)の2つの手法を用いた。

### 6.5 実験結果およびシミュレーション結果

本実験で実施したタスクの総数は2818であったが、途中の処理結果を解析した結果、中盤以降はオッズ比による  $P(h_i)$  の値が5を大きく超える仮説が多数出現していることが判明した。6.2節で述べた通り、この実験では一つのタスクで検証のために掲載する仮説の数を10としており、新規に入手した仮説には固定の  $P(h_i)$  の値として5を割り当てることで選択されやすくしている。したがって、仮説プールにおける全ての仮説のうち上位10個以上のオッズ比の値が5を超えている状態でのタスク結果は、新規に入手した仮説が選択されやすくなっているとは言えず、少なくとも適切な結果ではないと考えられる。そのため、上位10個の仮説のオッズ比の最小値が5以下となった1~250タスク目を有効な結果としてシミュレーションに用いることとした。

シミュレーションの結果を図9に示す。最終的に発生した組み合わせの数は3432、キャッシュを参照した回数は6474回であった。結果的には三つの手法において、いずれのキャッシュサイズの時でもキャッシュの効果が確認できた。しかし、それぞれの手法における再計算回数については、想定ほど大きな差は見られなかった。

### 6.6 考察

シミュレーションにおいて、提案手法とその他の手法の間に想定していたほどの差が見られなかった理由としては、図10のように、 $Q(h_i) \cdot Q(h_j)$  のばらつきが予想よりも極端であったことが挙げられる。具体的には、 $Q(h_i) \cdot Q(h_j)$  の値が非常に小さい仮説の組み合わせが全体に占める割合が極端に高く、

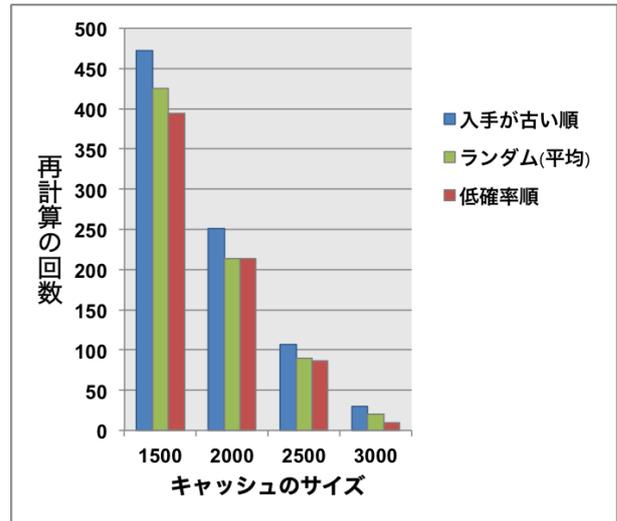


図9 各手法におけるキャッシュサイズごとの再計算回数

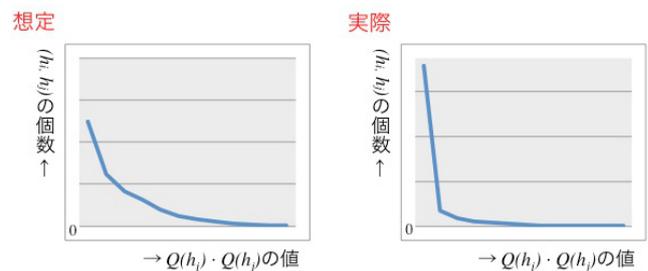


図10  $Q(h_i) \cdot Q(h_j)$  の値ごとの組み合わせの数のばらつき

$Q(h_i) \cdot Q(h_j)$  の値が比較的大きい仮説の割合が低かったため、結果的に他の手法でも  $Q(h_i) \cdot Q(h_j)$  の値が小さいものを多く削除することになり、想定ほどの差は生じなかったと考えられる。 $Q(h_i) \cdot Q(h_j)$  の値に十分なばらつきが生じなかった理由として、今回の実験では結果的にシミュレーションに利用できるタスク数が少なくなったことが考えられる。そのため、今後は十分な回答数のサンプルを得るために、実施タスク数を増やして実験するとともに、新規に入手した仮説に割り当てる  $P(h_i)$  の値を固定値ではなく、その時点での  $P(h_i)$  の最高値とすることなどを検討している。

## 7. まとめと今後の課題

本研究では、クラウドソーシングを用いた仮説の入手と検証を同時に行い、専門家による介入無しで、もっとも面白い仮説を入手するための手法を提案した。また、タスクサイズを一定に保つ方法、および、中間結果のキャッシュを用いた効率化手法についても提案した。実験及びシミュレーションを行い、少なくとも今回の結果に関しては、ある程度もっとも面白い仮説が入手できていること、および、提案したキャッシュを用いた効率化が有効であることを示した。

今後の課題としては、提案手法が、まだ十分に研究されていない問題に対して適用して、未知の仮説を得ることが可能かどうかを評価することが挙げられる。また、より大規模な実験を

行って効率化処理の有効性を再評価することや、因果関係の検証のために、単なるアンケートではなく実際にワーカに作業を依頼する事も考えられる。

## 謝 辞

本論文の一部は科研費基盤研究 (#25240012) の支援による。

## 文 献

- [1] A. Doan, R. Ramakrishnan, A. Y. Halevy, “Crowdsourcing systems on the world-wide web”, *Communications of the ACM* 54(4): 86-96, 2012.
- [2] Edith Law, Luis Von Ahn. “Input-agreement: a new mechanism for collecting data using human computation games.” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009.
- [3] Kirsten E. Bevelander, Kirsikka Kaipainen, Robert Swain, Simone Dohle, Josh C. Bongard, Paul D. Hines, Brian Wansink, ” “Crowdsourcing novel childhood predictors of adult obesity”, *PLoS One* 9: e87756, 2014.
- [4] 荒牧英治, 四方朱子, 渡部恵理子, 宮部真衣, 白田泰如, 綾屋紗月, 熊谷晋一郎, ”クラウドソーシングによるアレルギー・リスク推定 ?仮説形成から実験までの研究を半自動で行う試み? ” , *NLC2014-43*, pp.133-138, 2014.
- [5] Adam Marcus, Eugene Wu, David Karger, Samuel Madden, Robert Miller, “Human-powered Sorts and Joins”, *PVLDB* 2011, vol. 5, no. 1, pp. 13-24, 2011.
- [6] Marcus Adam, Wu Eugene, Madden Samuel, Miller Robert C, “Crowdsourced Databases: Query Processing with People”, *CIDR*, pp. 211-214, 2011.
- [7] Franklin Michael J., Kossmann Donald, Kraska Tim, Ramesh Sukriti, Xin Reynold, “CrowdDB: answering queries with crowdsourcing”, *SIGMOD Conference*, pp.61-72, 2011.
- [8] Parameswaran Aditya, Polyzotis Neoklis, “Answering Queries using Humans, Algorithms and Databases”, *CIDR*, pp. 160-166, 2011.
- [9] 三津石智巳, 森嶋 厚行, 品川 徳秀, 青木 秀人, “Crowdsourced Join Pre-filter による Human-powered Join 処理効率化の評価”, *DEIM Forum* 2013, 2013.
- [10] Yahoo!クラウドソーシング, <http://crowdsourcing.yahoo.co.jp/>
- [11] Crowd4U, <http://crowd4u.org/>
- [12] Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J, “An inventory for measuring depression.”, *Arch Gen Psychiatry* 4: 561-571, 1961.