

# Entropy-Based Hyper Question: 回答のバイアスを踏まえた 少数選抜の手法について

## Entropy-Based Hyper Question: Methodology to Identify Expertise Considering Answer Bias

田丸 陽稀<sup>†</sup>, 藤崎 樹<sup>‡</sup>, 馬場 雪乃<sup>†</sup>, 植田 一博<sup>†</sup>  
Haruki Tamaru, Itsuki Fujisaki, Yukino Baba, Kazuhiro Ueda

<sup>†</sup>東京大学, <sup>‡</sup>東北大学

The University of Tokyo, Tohoku University  
tamaru-haruki327@g.ecc.u-tokyo.ac.jp

### 概要

人々の意見を適切に集約した結果が時に高い精度を生む現象は集合知と呼ばれ, さまざまな手法が提案されている. 中でも, 集団内の成績優秀者を推定し, その意見のみを用いる少数選抜は高い精度を誇る. 本研究では, 少数選抜の一手法である **Hyper Question** に注目し, この手法が回答に偏りがある際に精度が落ちることを明らかにした. また, この限界に対し, エントロピーを利用することで回答の偏りを回避する応用手法を提案する.

キーワード: 集合知, 少数選抜, クラウドソーシング, エントロピー

### 1. はじめに

適切な方法で個々人の意見を集約することで, 意思決定の精度が向上する現象を集合知 ("The crowd of wisdom"; e.g. Surowiecki, 2004) と呼ぶ.

集合知を利用する手法として, 多数決のように, 集団全員の意見を集約する手法が知られている. しかし, 集団内の回答の偏りや外れ値に引きずられ, 結果が悪化する可能性がある (Simmons et al., 2011) (表 1). こうした問題を回避する方法として, 集団内でも特に意思決定の精度の高い回答者のみを抽出する, 少数選抜と呼ばれる手法の開発が近年進んでいる. 代表的な手法の一つに **Hyper Question** (Li et al., 2017) が挙げられる.

**Hyper Question** は, 精度の高い回答者 ("専門家"と呼ぶ) の回答が類似するという特徴を利用し, 専門家を選抜する. 専門家同士の回答は, 意思決定の精度が高くはない回答者 ("非専門家") 同士に比べて, 正答率が高い以上, 回答が類似する傾向が強い. そこで, まずいくつかの問題をまとめて問題のセットを作成, 比較し, 似た回答を持つ回答者のみを選抜する (表 2). 次に, それぞれの単一の問題に各回答を置き換え, 各問題で多数決などによってどの回答が多いのかを分析する. このような形で, 問題に詳しい回答者を抽出した後, 彼らの回答を集約することで, 専門家のみでの検討を可能にする. **Hyper Question** は, 従来手法の多数決などと比較

しても高い精度を誇る (Li et al., 2017).

一方で, 回答者の類似度による選抜という方法では, 集団内の回答の偏りを回避できない可能性がある. ある誤った選択肢に回答が集中している場合, 回答は当然類似するため, 結果的に誤答が選ばれてしまう危険性が指摘できる. このようにある誤った選択肢が多く選ばれる場合は, 人間にバイアスがある以上, ある程度現実に起こり得ると考えられる.

そこで, 本研究では, こうした回答に偏りがある場合でも有用な手法の提案を試みる.

表 1 回答の偏った回答データの例

	問題 1 (正解: A)	問題 2 (正解: A)	問題 3 (正解: A)
回答者 $\alpha$	A	A	A
回答者 $\beta$	A	A	A
回答者 $\gamma$	B	C	A
回答者 $\delta$	B	A	B
回答者 $\epsilon$	B	B	C

問題 1, 2, 3 に対し, 回答者  $\alpha, \beta, \gamma, \delta, \epsilon$  が A, B, C の三種類の選択肢から回答を行った場合の回答データセット. 各問題について, 最も多い回答を水色に塗っている. 多数決を利用した場合, 問題 1 を見ると, 正解 A に対して B が選ばれる結果となっている.

表 2 **Hyper Question** の回答データセットの例

	問題 1,2 (正解: A,A)	問題 1,3 (正解: A,A)	問題 2,3 (正解: A,A)
回答者 $\alpha$	A,A	A,A	A,A
回答者 $\beta$	A,A	A,A	A,A
回答者 $\gamma$	B,C	B,A	C,A
回答者 $\delta$	B,A	B,B	A,B
回答者 $\epsilon$	B,B	B,C	B,C

**Hyper Question** の例. **Hyper Question** では, 問題を一つずつ検討するのではなく, 複数の問題をセットにした問題 (超問題) について検討する. ここでは, 表 1 のデータセットの各問題に対する回答を, 2 問ずつ超問題としてまとめた. 複数の問題に対する回答を比較することで, 専門家同士の回答の類似度を比較することができ, 問題 1 に対しても正しい回答が選択されている.

## 2. 回答の歪みに対する Hyper Question の振る舞い

本研究ではまず、シミュレーションを用いて、Hyper Question が回答構造の偏りに対して弱いという仮説を検証した。シミュレーションでは、Li et al. (2017) で用いられた方法を踏襲しつつ、回答に偏りが発生しない場合と発生しない場合の2条件を作成し、結果を比較した。

具体的なシミュレーションについて説明する。まず両条件に共通した部分を説明し、次に異なる部分を説明する。

まず問題に回答するエージェント 20 人を作成した。エージェントは、与えられた問題に対して用意された選択肢の中から一つを出力する。エージェント集団はそれぞれの問題に対する正答率を設定されていた。エージェント集団は正答率が高い集団（専門家）と正答率が低い集団（非専門家）の2種類に分けられた。エージェントのうち、専門家の人数は{2,4,6}の中からランダムに決定され、正答率は{0.8,0.9,0.95,1.0}の中からランダムに決定された。残りのエージェントは正答率が低い非専門家であり、正答率は正規分布からランダムに決定されるとした。この正規分布の平均(0 から 1)と分散(0 から 0.2)を変数とすることで、「エージェント集団の能力」を変化させた。

次に、25 問の問題を作成した。問題の選択肢数は、2 択から 4 択まで変化させた。そして、各問題で正解となる選択肢をランダムで決定した。

以上のように作成した問題をエージェント 20 人に回答させた。なお、エージェントが正解を出力するか不正解を出力するかは正答率をもとに決定された。

以上のような 20 行 25 列の回答データセットを、変数の組合せごとに 100 個作成した。

2 条件で違いがあったのは、エージェントが不正解した時の行動であった。回答に偏りが無い条件の場合、エージェントは、不正解の選択肢から一つをランダムに選択し、出力した。このとき、どの不正解の選択肢も、選ばれる確率は等しかった。

回答に偏りがある条件においては、不正解の選択肢のうち、一つの選択肢の選択率を高くした。これにより、一つの回答に偏る状況を作成した。選択率は平均 0.6, 分散 0.05 の正規分布から、問題ごとにランダムに設定した。

結果を図 1 に示す。図 1 は横軸が非専門家の正答率

を決定する正規分布の平均、縦軸がそれぞれの手法の正答率であり、オレンジ線が Hyper Question, 青線が多数決である。回答に偏りが無いときに比べ、回答に偏りがあるときに Hyper Question の精度が低下している点が指摘できる。

以上のように Hyper Question は、回答の偏りがある場合に脆弱性があることが指摘できる。そこで、次に回答の偏りに対しても強い手法の提案を試みる。

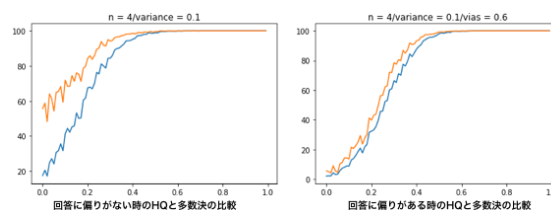


図 1 回答構造に偏りがある場合とない場合の Hyper Question と多数決の比較。

横軸が非専門家の正答率を決定する正規分布の平均、縦軸が各手法の正答率。オレンジ線が Hyper Question, 青線が多数決の精度を表す。問題の選択肢は 4 択。

## 3. Entropy-Based Hyper Question

一般的に、ある問題集合を観察した時、回答構造に偏りのある問題もあれば、比較的偏りの小さい問題もあると考えられる。回答構造に偏りが少なければ、Hyper Question はある程度精度を発揮する。したがって、回答構造の偏りを回答だけで判断できる指標があれば、ある程度精度の期待できる偏りのない問題から順番に検討できる。この際、回答者それぞれの能力を推定し、専門家を最初に選抜することができれば、その後回答構造に偏りのある問題についても正答を抽出できると期待される。

そこで、本研究ではエントロピーを利用して回答構造の偏りを測定する手法である Entropy-Based Hyper Question (EBHQ) を提案する。エントロピーはある事象のもつ情報量の期待値であり、ある選択肢が他の選択肢よりも大きく選ばれたときに小さい値を取る。逆に、どの選択肢も均等に選ばれている場合は大きい値を取る。このエントロピーに基づいて問題を降順にソートすることで、上述のように回答構造に偏りのない問題から検討を行うことができる(図 2)。提案手法では、各超問題に含まれる単一問題のエントロピーの総和を、超問題のエントロピーとした。

次に提案手法では、各問題の検討時に、選ばれた回答者に重みを与え、逐次的に重みを増やすことで徐々に専門家を推定することを試みた。ここで、望ましい重みの与え方として、初期の試行では与える票数を小さくし、試行回数が増えるにつれ一回で与える票数を徐々に増やしていく方法が考えられる。もし誤って初期の試行で非専門家に大きな重みを与えた場合、最終的に正しく専門家を同定できない可能性があるためである。

初期の試行では票を小さく与え、徐々に票の数を増やすという手法は、回答の偏りを使って表現できる。すなわち、各超問題の回答の偏りが小さいときには小さい票数を与え、回答の偏りが大きくなるにつれて与える票数も大きくすることを考える。回答の偏りは平均情報量を用いて表現できる。平均情報量は回答の偏りが小さい時に大きい値を取る。そのため、平均情報量のとりうる最大値 $\max(H)$ からの差、すなわち

$$\max(H) - H(X)$$

は、回答の偏りが小さい時に小さい値を取るようになる。この値を利用して票を与えることを考える。一方、平均情報量の値は問題の選択肢数によって変化する。そこで、問題の選択肢数に依存せず重み付けを行うために、平均情報量のとりうる最大値 $\max(H)$ で割って正規化を行った値、すなわち

$$\frac{\max(H) - H(X)}{\max(H)}$$

を票として与える。具体的には、例えば回答者 $\alpha$ の回答が問題 $X$ において選ばれた時 $\alpha$ の票数 $w_\alpha$ は、

$$w_\alpha = \frac{\max(H) - H(X)}{\max(H)}$$

と計算される。

上記の重みを用いて、Hyper Question を検討するごとに、選ばれた回答者に重み付けを行う。2章の「回答の偏りがある場合」と同条件のシミュレーションによって、提案手法を、Hyper Question および多数決と比較した。

その結果を図3に示す。提案手法 EBHQ (緑線)は、Hyper Question (オレンジ線)と多数決 (青線)を上回る精度を示しており、また精度が収束する速度も速いことが読み取れる。

こうした優位性が現実のデータに対しても発揮できるかどうかを調べるため、次に実データを用いて EBHQ の精度を他手法と比較した。利用したデータは、(Li et al., 2017) で用いられたデータのうち、特に利用

可能性が高いと思われる2種類の4択問題のデータ (薬学に関する問題, ITに関する問題) である。その結果を図4に示す。提案手法 EBHQ が他の2つの手法よりも高い精度を発揮していることが確認できる。

	問題 1,2	問題 1,3	問題 1,4	問題 2,3	問題 2,4	問題 3,4		問題 2,4	問題 3,4	問題 1,3	問題 2,3	問題 1,4	問題 1,2
回答者 $\alpha$	AA	AA	AA	AA	AA	AA		AA	AA	AA	AA	AA	AA
回答者 $\beta$	AA	AA	AA	AA	AA	AA		AA	AA	AA	AA	AA	AA
回答者 $\gamma$	A,C	AA	AB	C,A	C,B	A,B		C,B	A,B	AA	C,A	A,B	A,C
回答者 $\delta$	A,D	A,C	AB	D,C	D,B	C,B		D,B	C,B	A,C	D,C	A,B	A,D
エントロピー	0.0	1.5	1.0	1.3	2.5	1.8		2.5	1.8	1.5	1.3	1.0	0.0

図2 エントロピーを用いたソート法。

各超問題が持つ単一問題のエントロピーの総和を超問題のエントロピーとする。

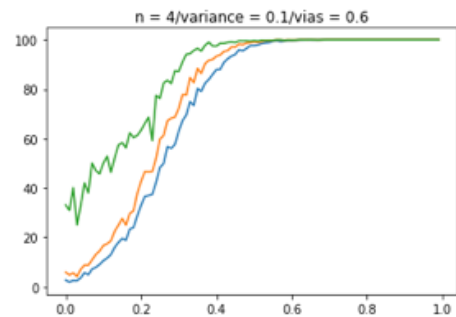


図3 EBHQ, Hyper Question, 多数決の精度

横軸が非専門家の正答率を決定する正規分布の平均、縦軸が各手法の正答率。オレンジ線がHyper Question, 青線が多数決の精度, 緑線がEBHQを表す。問題の選択肢は4択。

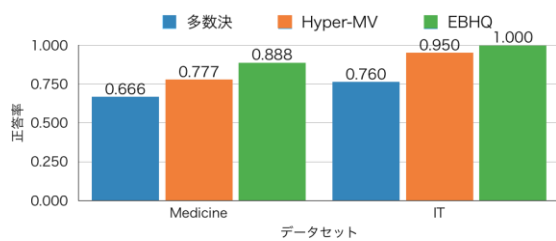


図4 実データに対するEBHQ, Hyper Question, 多数決の精度

それぞれのデータは薬学に関するデータ (Medicine), ITに関するデータ (IT) である。

## 4. 結論

本研究では Hyper Question が回答の偏りに脆弱性があるという問題点を指摘し、エントロピーを用いて最初に専門家を推定する手法を提案した。提案手法は、シミュレーションおよび実データによる検討において、

従来手法を上回る精度を発揮した。提案手法は、Hyper Question という優れた手法に認知科学的な観点から検討を加え、さらに情報科学的な概念であるエントロピーを援用することで、回答傾向だけから専門家を精度よく抽出することを可能にしており、集合知研究を現代的な観点から推し進めている点で意義があると考えられる。将来的に、株式市場などの、選択肢が数多くある問題で、一部の専門家を非専門家のバイアスから見つけ出すような場面での利用を検討したい。

## 文献

- [1] Li, J., Baba, Y., & Kashima, H. (2017). “Hyper questions: Unsupervised targeting of a few experts in crowdsourcing.” In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 1069-1078.
- [2] Simmons, J., Nelson, L. D., Galak, J., & Frederick, S. (2011) “Intuitive biases in choice vs. estimation: Implications for the wisdom of crowds”. *Journal of Consumer Research*. 38(1):1–15.
- [3] Surowiecki, J. (2004). “The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economics, Societies and Nations”, Doubleday.