

# 集合知メカニズムとしてのベイズ自白剤 (Bayesian Truth Serum) の有効性

SP12137 非線形物理学研究室 橋本真佑

## 1. はじめに

「企業のネットが星を被い電子や光が駆け巡る」情報化された現在、さまざまなヒトの提供する情報をうまく集約していく仕組みが求められている。そのとき問題になるのが情報のノイズである。ノイズには様々なものがあり、情報がお互いに影響しあうため、多数のユーザーの提供する情報の平均値を計算すればノイズがキャンセルして真実が浮かび上がるようにはなっていない。この情報の集約の難しさは、企業が商品開発のために多数の消費者(モニターなど)を集めて行うアンケート調査でも同じである。正直に回答した消費者には謝礼をはずむことが出来ればインセンティブとして作用し、正確な情報を集めることが出来るかもしれない。しかし、消費者が「正直に」回答したことをどうやって判定すればよいのだろうか? また、「正直に」回答したとしても、ユーザーの提供する情報は主観的なものである。そこから客観的な価値のある情報を抽出するにはどうすればよいのか? D.Prelec は、ベイズ自白剤 (bayesian truth serum) というアイデアを提案した [1]。それは、回答の際に、同時に他の回答者の回答についても予測も尋ね、人々の予想より多数のヒトの選択する選択肢を選択すると高くなるなる「情報スコア」IS (information score) と、人々の回答をどれぐらい正確に予測できたかを示す「予測スコア」PS (prediction score) の線形結合としてベイズ自白剤スコア BTS を計算し、BTS の高い回答は正直な回答と考えるものである。

本研究では、二択のクイズを用いて、BTS の有効性を検証する。二択のクイズに対して、(i) 正解を知っているか、(ii) 正解の選択肢、(iii) 正解を知っているの回答者の比率 (%), の3度の質問を行い、IS, PS, BTS を計算し、これらのスコアと (i)(ii) の回答との関係を調べる。

## 2. 実験の説明

実験は 2015 年 12 月 8 日から 14 日にかけて北里大学理学部学生を被験者としてのべ 104 名を対象に行った。被験者は 100 問の二択のクイズに (i)(ii)(iii) の3度回答する。実験者は理学部物理学の川崎と佐野、院生の吉田の3名。被験者の報酬は 1050 円とし、100 問回答終了後に支払った [2]。二択のクイズは 200 問用意し、初回は最初の 100 問、二回目は残りの 100 問回答するとして被験者は最大 2 回実験に参加した。クイズは、過去に実施した情報カスケード実験での正答率が 40% から 70% の難しい問題から選んでいる。最初の 100 問のみに回答した被験者は 74 名、全 200 問に回答した被験者は 30 名となった。以下の解析には、最初の 100 問に回答した 74 名分のデータを用いている。

クイズ番号  $i$  を  $i \in I = \{0, 1, \dots, |I| = 100\}$ 、被験者番号  $id$  を  $id \in ID = \{1, \dots, |ID| = 74\}$  で表す。また、3 つの質問に対する回答は、(i) 知っているかどうかは、知っているなら 1、知らないなら 0 として  $A_{id}^i \in \{0, 1\}$ , (ii) 選択した選択肢は、正解なら 1、不正解なら 0 として  $X_{id}^i \in \{0, 1\}$ , (iii)(i) の質問に知っているか答えた被験者の比率の推定値は  $p_{id}^i \in [0, 1]$  で表すものとする。

## 例

漫画「ワンピース」に登場する、バギー海賊団船長「道化のバギー」の懸賞金はいくら? (ルフィーと最初に会ったころ)

A. 1300万ベリー

B. 1500万ベリー

①あなたは正解を知っていますか? 自信がある場合に○、ない場合はXと書いてください

②正解を選んでください。

③理学部生100人中、何%が正解を知っていると思いますか?

%

Fig.1: 被験者募集のポスターで用いた実験説明。

## 3. ベイズ自白剤 BTS の計算方法

全 100 問 74 名の 7400 の回答中の正答率は 57.4% であった。一方、(i) で yes と回答した回答数は 1402 あり、その正答率は 70.0% であった。つまり、正解を知っていると回答しても正解であるとは限らない。しかし、(i) で yes、(ii) で間違いを選んだ回答者 A は嘘つきで、(i) で yes、(ii) で正解を選んだ回答者 B が正直者とは限らない。回答者 A は単に勘違いしているだけかもしれない、回答者 B は正解を知らないのにたまたま選んだ選択肢が正解だったのかもしれない。

そこで、BTS を計算する。まず、問題  $i \in I$  に対する (i) 「知っている、知らない」の回答の分布  $(a^i, 1 - a^i)$  を求める。

$$a^i = \frac{1}{|ID|} \sum_{id \in ID} A_{id}^i$$

また、(iii) の予想の平均値  $(p^i, 1 - p^i)$  を幾何平均を用いて評価する。

$$\ln p^i = \frac{1}{|ID|} \sum_{id \in ID} \ln p_{id}^i$$

情報スコア  $IS_{id}^i$  は、被験者集団の予想より被験者集団の回答が多数派  $a^i > p^i$  (surprisingly common) なら  $A_{id}^i = 1$  のスコアを正にし、逆に少数派 (surprisingly uncommon) なら  $A_{id}^i = 0$  のスコアを正にする。スコアは common, uncommon の度合いで定まる。

$$IS_{id}^i = A_{id}^i \cdot \ln a^i / p^i + (1 - A_{id}^i) \cdot \ln(1 - a^i) / (1 - p^i).$$

予測スコア  $PS_{id}^i$  は、主観的な確率分布  $(p_{id}^i, 1 - p_{id}^i)$  と被験者全体での回答の分布  $(a^i, 1 - a^i)$  の「距離」としてカルバックライブラーダイバージェンスで計算する。

$$PS_{id}^i = \ln p_{id}^i / a^i + \ln(1 - p_{id}^i) / (1 - a^i)$$

BTS は、 $\alpha \in (0, 1)$  をパラメータとして次の式で計算する。

$$BTS_{id}^i = \alpha IS_{id}^i + (1 - \alpha) PS_{id}^i.$$

## 参考文献

[1] D. Prelec, Science vol.306, 462(2004).

[2] 科研費「実験経済物理学による社会的学習と集団知の創発過程の解明」を使用。