

タスク推薦システムを用いたクラウドソーシングプラットフォーム

新山 光太郎[†] 遠山 元道[‡]

慶應義塾大学理工学部情報工学科^{†‡}

1. はじめに

コンピュータだけでは解くことが難しく、人間の判断が必要とされる問題を解決する方法の一つに、クラウドソーシングが挙げられる。不特定多数の人にタスク処理を依頼するクラウドソーシングのサービスを扱う市場には Amazon Mechanical Turk (AMT) や oDesk 等が存在する。

クラウドソーシングプラットフォームにおけるユーザインタフェースは非常に重要であり、それぞれのサービスによって異なる。

これらのサービスではただタスクを無造作に提供するのではなく、回答者がキーワードやタグ等から自分でタスクを検索するプルモード、プラットフォームが回答者の適正に合ったタスクを推薦するプッシュモードを扱っている。本論文では後者を扱った研究の提案をする。

2. タスク推薦システム

プッシュモードを扱う際に、タスクや回答者に幾つかのパラメータを割り当てる。それらのパラメータから回答者に適したタスクを提供するシステムとしてタスク推薦システムを提案する。以下でそのパラメータを紹介する。

2.1 タスク難易度

タスクの難しさを扱う指標としてタスク難易度を用いる。タスクの提供者は難しさに応じて自身の判断で 0.1~1.0 の値を決定する (1.0 が最も難しいものとする)。また、タスクには例えば「Java のコーディング」、「英語から日本語への翻訳」などのタグ付けをプラットフォーム内で行う。

2.2 回答者レベル

オンラインでの不特定多数の回答者が全て同じ技術や知識を持っていないことは明らかである。したがって、それぞれの回答者にその能力に応じたレベルを適当に割り当てる必要がある。

このレベルは回答者一人に対して単一なものではなく、タスクのタグごとにパラメータを同じく 0.1~1.0 の間で決定する。(図 1 に概念図を示す)

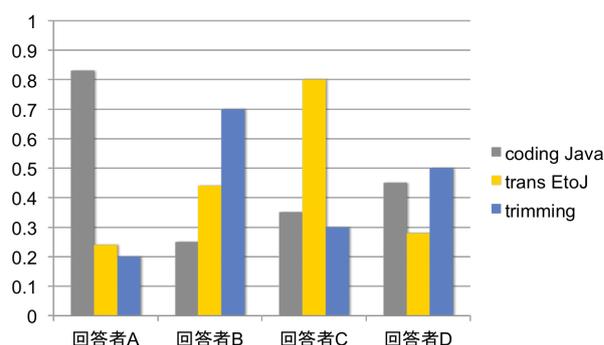


図 1 回答者レベルの例

このパラメータの値を決定する方法としてゴールドクエスチョンを用いる。ゴールドクエスチョンとはあらかじめプラットフォームが用意しており、答えが明確な問題である。この正答度によってパラメータの値を決定する。

タグ x のタスクに関する回答者レベルは

$$L_x = 1 - \frac{N_m}{N_a}$$

で与えられる^[1]。このとき、 N_a はゴールドクエスチョンの問題数、 N_m は正答できなかった問題数である。

また、タスク提供者が決定したタスク難易度 t を閾値とし、それ以下のレベルの回答者にはそのタスクを提供しないこととする^[1]。また、著しく低い回答者レベルを持つ回答者がいた場合、適切な回答をしていないと考えられるので、任意の閾値 t_r を下回るような回答者は除外する。このように、ゴールドクエスチョンは協力的でないような回答者を発見する際にも有用である。

両パラメータのマッチングはそれぞれの回答者レベルに一番近い値のタスク難易度を持つタスクを提供する。値が同じようなタスクに関してはその中から無造作に提供する。

Crowdsourcing Platform Using Task Recommendation System

[†] Koutarou Niiyama

[‡] Motomichi Toyama

^{†‡} Dept. of Information and Computer

Science, Keio University Yokohama, Japan

2.3 デバイス推薦

以上で紹介した以外にも、タスクの種類には様々なものがある。例えば 1 枚の写真の中から池だけをトリミングしたり、流星の軌跡になぞって線を引くタスク等が実在する。

これらのタスクを処理する場合、PC 端末で範囲を選択したり、曲線を引いたりするよりはスマートフォンで指でなぞる方がより容易であると考えられる。しかし一方で、先に挙げたコーディングや翻訳などは PC 端末の方が早く回答できると予想できる。

このようにデバイスごとに回答の精度や速度が変わることが考えられるので、タスク提供者が推薦するデバイスを決められるようにする。

3. 実験目的と方法

タスク難易度と回答者レベルのマッチングをすることで得られる正答度とタスク処理速度の向上、またデバイス推薦をすることで更なるタスク処理時間の短縮が期待できる。これらの性能を測るために以下のような実験方法を提案する。

3.1 回答者レベルの値の決定

自身が所属する遠山研究室の学生達を回答者とする。回答者はゴールドクエスチョンに回答する。その正答度からタスク推薦システムに沿ってパラメータを割り当てる。

ゴールドクエスチョンには「Java プログラミング」、「英語から日本語への翻訳」、「トリミング」のタグを持つタスクを扱う。それぞれの難易度は 0.1~1.0 まで 0.1 刻みずつ 10 問、計 30 問用意する。また、スパム除外の閾値である t_r は同研究室の学生が行うため考慮しないこととする。

3.2 タスク難易度とデバイス推薦の決定

タスクは AMT や oDesk 等、その他実在するクラウドソーシングプラットフォームのタスクから無造作に取り出す。タスク難易度と推薦するデバイスは著者らが任意に決定する。

3.3 タスク処理

回答者が実際にタスクを処理する。その際、以下の手順に従う。
方法 1: タスク推薦システムを用いずに、無造作に回答者にタスクを提供する。
方法 2: タスク推薦システムのうち、タスク難易

度と回答者レベルのみを用いて回答者に適当なタスクを提供する。

方法 3: 方法 2 に加えてデバイス推薦も行い、回答者にタスクを適当な提供する。

4. 評価方法

方法 1 と方法 2 において、タスクごとの正答度を比較する。その正答度は著者らが任意に 0.1~1.0 の値で決定する。また、全てのタスクの処理時間を計測し比較する。

次に、方法 2 と方法 3 において、全てのタスクの処理時間を計測し比較する。

5. 今後の課題

ゴールドクエスチョンの数と難易度は多すぎる、もしくは難しすぎると回答者に負荷をかけてしまう恐れがある。しかし一方で、タスクのタグごとに決定される回答者のパラメータは実世界で扱うとなるとかなりの数がある。したがって、適切にゴールドクエスチョンを設ける必要がある。

また、実験精度を改良するためにタスクの回答による正答度の決定はなるべく明確な基準を設定しなければならないと考えられる。

ゴールドクエスチョンの他にも、タスク処理の正答度を動的にフィードバックして回答者レベルを決定することで、更なる正答度の向上が得られると Bragg らは述べている^[2]。

さらに、タスクの数は統計的に評価が信用できる件数を調査し、用意しなければならない。

6. まとめ

本論文ではクラウドソーシングプラットフォームにおいてタスク処理の正答度の向上と処理時間の短縮が期待できるタスク推薦システムを提案した。

文献

- [1] Oleson, D., Sorokin, A., Laughlin, G. P., Hester, V., Le, J., & Biewald, L. Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing. *Human computation*, 11(11). (2011).
[2] Bragg, J., Kolobov, A., Mausam, M., & Weld, D. S. Parallel Task Routing for Crowdsourcing. In *Second AAAI Conference on Human Computation and Crowdsourcing*. (2014, May).