

実験計画書

名称

ワーカーの振る舞いによる低寝室ワーカーの検出

背景

クラウドソーシングでデータを集める際に多くの場合考えなくてはいけないこととして、ワーカーの品質が保証できないこと、それによってどのように低品質ワーカーを検出し省くかを考えなくてはいけない。検出方法について既存のものであれば、多数決やgold setなど回答した結果から検出する方法があるが、回答から検出しようとするそれが適当にやったのか、本当に間違えただけかを考慮できない。そこで回答ではなく、回答中のワーカーのふるまい方に注目して低品質ワーカーの検出ができるかどうかに興味を持ったため今研究を行う

仮説

PC限定だが、回答する際、適当に答える人はマウスがほとんど動かない、もしくはほぼ一定の動きをするのではないかと、逆に真剣に悩んでいればある程度ランダムにマウスが動くのではないかと仮定し、低品質ワーカー、高品質ワーカーのふるまいの違いが分かればフィルターをかけることができる

目的

今実験の目的は2つ

- 高品質ワーカー、低品質ワーカーのマウスの動きに差があるか検証
- 実際にマウスの動きで検出が行えるか実証する

方法

- 対象：不特定多数の人々
- 実施場所：AMT(不可能ならクラウドソーシング?)
- 手続き：まず、低品質ワーカーのマウスの動きのベースとなるデータを作成、それを何らかのタスクで実践する。
 - ここは現在悩み中、マウスの動きを取得する流れは同じだが、低品質ワーカーのモデルの取得について、ベースとなるデータのフィルタをgold setや多数決など回答そのものから検出するか、それともランダムフォレストを用いるか。データに差があれば引き続き実験、時間など他の要因と絡めて考えるもあり
- タスク：検討中
- 用いるデータ：検討中 (タスクもデータも既存のものを使用する予定)

得られる結果

低品質ワーカーの検出の新たな提案(ただしまだまだ提案段階)

評価方法

検討中、ランダムフォレストとかの検出率と比べてみる？ 細かく出せなくてもいいのでどれくらい検出できたか見せるだけでもよし。

人権配慮

とくになし(制限なし)

データ保管の方法

ハードディスク、10年間保管