

令和元年5月20日現在

機関番号：10101

研究種目：基盤研究(B)（一般）

研究期間：2015～2017

課題番号：15H02782

研究課題名（和文）メタ認知タスクを用いたヒューマンコンピューテーションの品質制御

研究課題名（英文）Quality Control of Human Computation Using Meta-Cognitive Tasks

研究代表者

小山 聡 (Oyama, Satoshi)

北海道大学・情報科学研究科・准教授

研究者番号：30346100

交付決定額（研究期間全体）：（直接経費） 9,640,000円

研究成果の概要（和文）：本研究では、計算機による情報処理の中で人間の能力を積極的に活用するというヒューマンコンピューテーションの考え方をさらに推し進めて、ヒューマンコンピューテーションの品質制御自体に人間の能力を活用する方式を追求した。確信度判断や多面的な評価といった人間の高度な認知能力と、計算機の高速かつ高精度な計算能力を利用した品質制御モデルを開発した。画像認識や推薦システム、教育といった様々な分野において、実サービスのデータや、実際の作業者を用いた評価実験によりその有効性を検証した。

研究成果の学術的意義や社会的意義

現在、機械学習などの人工知能技術による自動化が様々な分野において推し進められているが、依然として人間の判断や評価が必要な問題は多く存在しており、推薦システムや教育など、本質的に人間の介在が不可欠な問題も存在する。本研究はそのような人間を含むシステムにおいて人間の高度な認知能力を活用して品質制御を行う方法を提案しており、発展を続ける人工知能技術と組み合わせることで、様々な分野において製品やサービスの品質向上への貢献が期待できる。

研究成果の概要（英文）：In this research, we advanced the concept of human computation, which affirmatively uses human's ability in information processing by computer, to utilize human's ability in the quality control of human computation itself. We developed quality control models using human's high cognitive ability such as confidence judgment and multicriteria evaluation as well as high speed and high accuracy calculation ability of computer. In various fields such as image recognition, recommendation systems and education, the effectiveness of the approach was verified by data from actual services and experimental evaluation using real workers.

研究分野：人工知能

キーワード：人工知能 認知科学 メカニズム設計 推薦システム 教育応用

1. 研究開始当初の背景

ヒューマンコンピューテーションとは、計算機だけでは十分な性能を得ることが難しく、人間の認識や判断等を必要とするような課題において、人間を計算資源の一部として捉えることによって、計算機と人間の両者の力を組み合わせてこれを解決するという考え方である(Law and von Ahn, 2011). Amazon Mechanical Turk に代表されるクラウドソーシング市場によりインターネットを通して不特定多数の人的資源へのアクセスが可能となったこともあり、ヒューマンコンピューテーションは新たなサービスを実現するための手段として注目を集めている。例えば、これまで教師付き学習においてはラベル付きの訓練データを準備することがボトルネックになっていたが、クラウドソーシングを用いることで、比較的安価に大量のラベル付きデータを得ることが可能になり、機械学習を様々なサービスに適用することが容易になった。一方、不特定多数の作業員から得られた作業結果は、従来の専門家から得られた作業結果と異なり、能力不足や意図的な手抜きなどにより、誤りが多く含まれる可能性がある。そのため、品質制御(Quality Control)の問題は、ヒューマンコンピューテーションにおける重要な研究課題となっている(小山, 2014)。

2. 研究の目的

本研究では、計算機による情報処理の中で人間の能力を積極的に活用するというヒューマンコンピューテーションの考え方をさらに推し進めて、ヒューマンコンピューテーションの品質制御自体に人間の能力を活用することを目的とする。確信度判断や多面的な評価といった人間の高度な認知能力と、計算機の高速かつ高精度な計算能力を活用することで、情報処理の品質の向上を目指す。画像認識や推薦システム、教育といった様々な分野において、実サービスのデータや、実際の作業員を用いた評価実験によりその有効性を検証する。

3. 研究の方法

本研究の基本的なアプローチは、ヒューマンコンピューテーションの品質制御問題を、潜在変数を含んだ確率モデルで記述し、実データを用いてパラメータを推定するというものである。参加者への報酬を決定する問題に対しては、ゲーム理論に基づくインセンティブ設計の考えを用いる。商用のクラウドソーシングプラットフォームである Lancers (<https://www.lancers.jp/>) や Amazon Mechanical Turk (<https://www.mturk.com/>) 上でタスクのユーザインタフェースを設計し、作業を依頼して収集したデータとともに、楽天データセット (https://rit.rakuten.co.jp/data_release_ja/) などで公開されている実サービスの大規模データも用いて評価実験を行う。

4. 研究成果

(1) 人間による商品やコンテンツ、サービス(アイテムとよぶ)に対する主観的な評価は、推薦システムなどを構築する際に基本となるデータである。評価者間のばらつきの影響を抑えるために、複数の評価者の評価値を平均などで統合して統合評価値を算出し利用者に提示することが行われる。しかし、評価対象であるアイテムに対して評価者の数が少ない場合、評価の信頼性が確保できないため、実際の推薦システムなどでも一部のデータにしか統合評価値が提示されない場合が多い。そこで我々は、統合評価値を計算する問題を、仮想的な「平均的ユーザ」に対する推薦問題として定式化した。一方、たとえばユーザのホテルに対する全体的な評価(総合評価)だけでなく、立地や食事といった各観点での評価(多基準評価)を利用できる場合がある(図1)。多基準評価は、ユーザのアイテムに対する評価のより詳細な情報を含んでおり、これらを利用することで、単一評価のみを用いた場合よりも評価値統合や推薦の精度が向上することが期待できる。楽天データセットにおけるホテルに対する評価データとゴルフ場に対する評価データを対象に評価実験を行った結果、多基準評価データを利用することで予測精度が改善し、さらに多基準データを用いる方法の中では、我々が提案した多基準データのためのベイズ確率的テンソル分解モデルが最も優れていることを確認した。

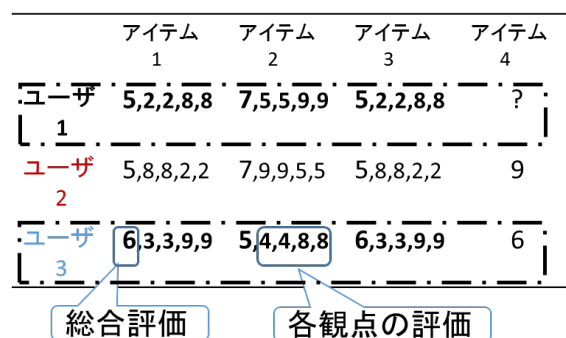


図 1 多基準評価データ

(2) クラウドソーシングの普及により、機械学習に用いるラベルなしデータを比較的容易に大量に得ることができるようになったが、依然としてデータの大規模化には限界がある。一方、画像データなどは、インターネットからほぼ無尽蔵に取得することができる。専門家が作成し

ラベルありデータとラベルなしデータを用いる学習は半教師付学習と呼ばれ従来から存在していたが、クラウドソーシングで獲得したラベルありデータを用いれば、機械学習の適用範囲を広げることができるとともに、予測精度も改善できることが期待できる。ただし、クラウドソーシングで獲得したラベルを用いる際には、作業員ごとのラベルの品質の違いを考慮して学習を行う必要がある。そこで我々は、深層生成モデルを用いて複雑な分布のデータに対しても半教師付学習を可能にしたモデル (Kingma et al., 2014) に複数の作業員によるラベル付けの過程を導入したモデルを提案した。画像分類および映画に対する評価のデータセットにおいて、人間の付与したラベルの数が十分ではない場合でも、半教師付き学習を実現することで、予測精度を向上できることを確認した。また、仮想的にランダムな結果を返す作業員 (スパマー) を導入した実験を行い、提案手法はそのような作業員の影響 (モデルにおける重み) を低減できることを確認した (図 2)。

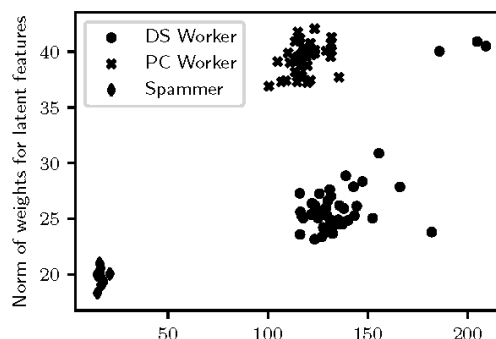


図 2 スパマーの影響の低減

人間の付与したラベルの数が十分ではない場合でも、半教師付き学習を実現することで、予測精度を向上できることを確認した。また、仮想的にランダムな結果を返す作業員 (スパマー) を導入した実験を行い、提案手法はそのような作業員の影響 (モデルにおける重み) を低減できることを確認した (図 2)。

(3) 多くの人々に将来起こりうるイベントについて予測してもらおう際、どのイベントが起こる / 起こらないという離散的な予測ではなく、それぞれのイベントが起こる主観的な確率を申告させることで、より正確な予測が行える可能性がある。ヒューマンコンピューテーションの品質制御においては、前述の方式のように作業を行った後で結果の統合などで品質を確保することに加え、最初から品質の高い作業結果が得られるような方策を考える必要がある。ヒューマンコンピューテーションは人間が関与するため、作業員が品質のよい作業を行うような動機付け (インセンティブ) を与える必要がある。たとえば、作業を行っても報酬がもらえない可能性が高ければ、タスク自体が実行されない可能性がある。我々は確率的な予測を行うタスクにおいて、実際に起こったイベントだけでなく、起こらなかったイベントに対してもイベントの類似度に応じて報酬を与えることで、参加者の報酬の分散を低減させるメカニズムを提案した。

ヒューマンコンピューテーションにおけるインセンティブとして金銭的なものを用いる場合、依頼者は自身の予算の制約の範囲で、望ましい結果が得られるように作業員に報酬を配分するためのメカニズムの設計を行う必要がある。しかし多くの依頼者にとって、このようなメカニズムを設計することは容易ではない。そこで本研究では、ヒューマンコンピューテーションのメカニズム自体を、クラウドソーシングで作業員に提案してもらおうことを試みた。画像へのタグ付けのタスクで提案されたメカニズムの例を図 3 に示す。さらに、提案されたメカニズムを評価するタスクも実行することで、メカニズムの妥当性や作業員に好まれる性質について分析を行った。

写真を4分割し、それぞれのパートに5人ずつ、合計20人に作業を依頼します。基本の報酬は45円とします (基本報酬の合計900円)。テキストにキーワードを改行して入力してもらいます。4つのパートごとに、一番多くキーワードをあげた人に成果報酬として25円をしいらいます (成果報酬の合計100円)。

図 3 提案されたメカニズムの例

(4) ヒューマンコンピューテーションの応用が期待される領域として教育がある。とくに、MOOC (Massive Open Online Courses) などのオンライン教育においては、課題の作成や受講者の評価に労力が掛かるが、それをヒューマンコンピューテーションで解決することが期待される。

そこで本研究では、数学教育への適用例として魔法陣パズルを取り上げ、ユーザのレベルに合った難易度の問題を提示することを目標とした。問題の難易度は、複数の受験者の解答から

項目反応理論 (Item Response Theory, IRT) を用いて推定することが可能である (豊田, 2005)。そこで、クラウドソーシングで問題に対する解答を募集し、難易度の推定を行った。さらに、解答事体に加えて、解答に対する自信 (確信度)、主観的な難易度、解答に掛かった時間を尋ねる補助的なタスクも実行し、項目反応理論によって推定される難易度とこれらの指標の間の相関を分析した (図 4)。項目反応理論で推定した難易度と、確信度や主観的な難易度は強い相関があり、お互いに代替可能であることが

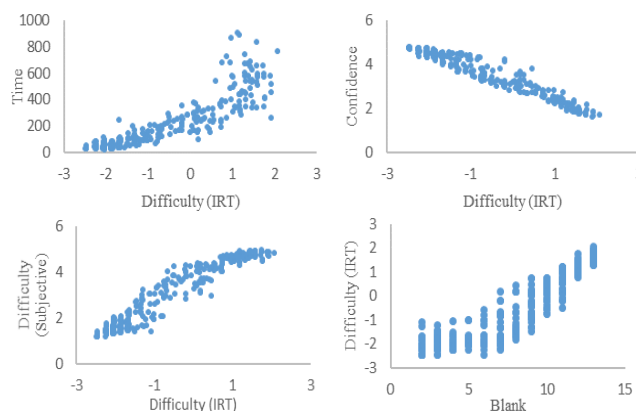


図 4 各指標と項目反応理論による難易度との相関

分かる。一方、解答にかかった時間は難易度の大きな問題で分散が大きく、空欄の数といった単純な指標も、難易度の分散が大きいことが分かる。

機械学習においては、特徴選択やモデルのハイパーパラメータの設定などで、人間の知識の導入や試行錯誤が必要となる。多くの参加者がこの作業を並行して行うデータ解析コンペティションは、ヒューマンコンピューションの一例と考えることができる。機械学習のコンペティションでは、現在の成績上位者を示すリーダーボードが参加者の意欲向上のために提示されることが多いが、これは一部のテストデータを用いた中間スコアにより計算されるため、何度も投稿を繰り返すと過学習を起こす危険があり、これを避けるために一日あたりの投稿回数を制限することが行われている。このような過学習を避けるために提案された Ladder (Blum and M. Hardt, 2015) と呼ばれる、参加者の従来の投稿に比べて有意に精度が向上している場合のみランキングを更新するアルゴリズムを教育用データ解析コンペティションにおいて実装し、実際のデータ解析の授業に用いて評価を行った。その結果、一日あたりの投稿回数の制限を撤廃することで、参加者の投稿回数が増え、上位入賞者の最終スコアも改善することが確認できた。

(5) 現在人工知能の様々な分野において、大規模なニューラルネットワークを用いた深層学習のアプローチが大きな成功を収め、これまで人間が行ってきた様々な作業を人工知能が代替していくと予想されている。人間の能力を積極的に活用しようというヒューマンコンピューションのアプローチは、一見これと対立するように見えるが、実は相補的な関係にあると我々は考えている。対象に対する主観的な評価などは人間からしか得られないものであり、(2)で示したようなアプローチでヒューマンコンピューションと深層学習を組み合わせることで、一層の活用が可能となる。今後の研究の展開としては、深層学習とヒューマンコンピューションを統合した品質制御やメカニズム設計の研究を推進する計画である。

< 引用文献 >

- E. Law and L. von Ahn: Human Computation, Morgan & Claypool, 2011.
- 小山 聡: ヒューマンコンピューションの品質管理, 人工知能, 29(1):27-33, 2014.
- D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling: Semi-supervised Learning with Deep Generative Models, NIPS, 2014.
- 豊田 秀樹: 項目反応理論・理論編 テストの数理, 朝倉書店, 2005.
- A. Blum and M. Hardt: The Ladder: A Reliable Leaderboard for Machine Learning Competitions, ICML, 2015.

5. 主な発表論文等

[雑誌論文] (計 2 件)

- Hiroki Morise, Satoshi Oyama, and Masahito Kurihara: Bayesian Probabilistic Tensor Factorization for Recommendation and Rating Aggregation with Multicriteria Evaluation Data. Expert Systems with Applications, Vol. 131, pp. 1-8, Elsevier, 2019, 査読有 (Impact factor: 3.768). DOI: 10.1016/j.eswa.2019.04.044
- 馬場 雪乃, 高瀬 朝海, 新 恭兵, 小山 聡, 鹿島 久嗣: 教育用データ解析コンペティション基盤の設計と実践. デジタルプラクティス, Vol. 9, No. 4, pp. 859-873, 2018, 査読無.

[学会発表] (計 29 件)

- Ryota Sekiya, Satoshi Oyama, and Masahito Kurihara: User-Adaptive Preparation of Mathematical Puzzles Using Item Response Theory and Deep Learning. 32nd International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE 2019), 2019, 査読有.
- Kyohei Atarashi, Satoshi Oyama, and Masahito Kurihara: Semi-supervised Learning from Crowds Using Deep Generative Models. 32nd AAAI Conference on Artificial Intelligence (AAAI 2018), 2018, 査読有 (Acceptance rate: 25%).
- Yukino Baba, Tomoumi Takase, Kyohei Atarashi, Satoshi Oyama, and Hisashi Kashima: Data Analysis Competition Platform for Educational Purposes: Lessons Learned and Future Challenges. Eighth AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI 2018), 2018, 査読有.
- Yuko Sakurai, Jun Kawahara, and Satoshi Oyama: Aggregating Crowd Opinions Using Shapley Value Regression. 12th Multi-Disciplinary International Conference on Artificial Intelligence (MIWAI 2018), 2018, 査読有 (Acceptance rate for full papers: 29%). DOI: 10.1007/978-3-030-03014-8_13

Hiroki Morise, Satoshi Oyama, and Masahito Kurihara: Collaborative Filtering and Rating Aggregation Based on Multicriteria Rating. First IEEE Workshop on Human-Machine Collaboration in Big Data (IEEE HMDData 2017), 2017, 査読有 (Acceptance rate for full research papers: 50%) .

DOI: 10.1109/BigData.2017.8258477

Yuko Sakurai, Masafumi Matsuda, Masato Shinoda, and Satoshi Oyama: Crowdsourcing Mechanism Design. 20th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA 2017), 2017, 査読有 (Acceptance rate: 41%) .

DOI: 10.1007/978-3-319-69131-2_32

Yuko Sakurai, Masafumi Matsuda, and Satoshi Oyama: Requesters' Personal Values, Just-world beliefs, and Their Choice of Incentive Mechanisms. Fifth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2017), 2017, 査読有.

Satoshi Oyama: Quality Control for Crowdsourced Multi-Label Annotation. 11th Korea-Japan Database Workshop (KJDB 2016), 2016, 招待講演.

Yuko Sakurai, Masato Shinoda, Satoshi Oyama, and Makoto Yokoo: Flexible Reward Plans for Crowdsourced Tasks. 18th International Conference on Principles and Practice of MultiAgent Systems (PRIMA 2015), 2015, 査読有 (Acceptance rate: 31%) .

DOI: 10.1007/978-3-319-25524-8_25

Hisashi Kashima, Satoshi Oyama, and Yukino Baba: Crowdsourcing for Big Data Analytics, 19th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2015), 2015, チュートリアル講演.

他 19 件

〔図書〕(計 1 件)

鹿島 久嗣, 小山 聡, 馬場 雪乃: ヒューマンコンピューテーションとクラウドソーシング, 講談社, 117 ページ, 2016.

6 . 研究組織

(1)研究分担者

研究分担者氏名：櫻井 祐子

ローマ字氏名：SAKURAI, Yuko

所属研究機関名：国立研究開発法人産業技術総合研究所

部局名：情報・人間工学領域

職名：主任研究員

研究者番号 (8 桁): 10396137

(2)研究協力者

研究協力者氏名：高瀬 朝海

ローマ字氏名：TAKASE, Tomoumi

研究協力者氏名：新 恭兵

ローマ字氏名：ATARASHI, Kyohei

研究協力者氏名：森瀬 寛己

ローマ字氏名：MORISE, Hiroki

研究協力者氏名：関屋 亮太

ローマ字氏名：SEKIYA, Ryota

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。