

令和 6 年 6 月 25 日現在

機関番号：13701

研究種目：基盤研究(B) (一般)

研究期間：2019～2022

課題番号：19H04218

研究課題名(和文)クラウドソーシングにおける協調動作による大規模創造的作業に関する研究

研究課題名(英文)Research on large-scale creative work by cooperation in crowdsourcing

研究代表者

鈴木 優 (Suzuki, Yu)

岐阜大学・工学部・准教授

研究者番号：40388111

交付決定額(研究期間全体)：(直接経費) 13,300,000円

研究成果の概要(和文)：クラウドソーシングと呼ばれるインターネット上で共同で分担して作業を行うプラットフォームを用いて、創造性の高い作業を行う方法の構築についての研究を行った。本研究では、作業者の品質測定および高品質な作業結果の二つの点に重点を置き研究を行った。作業者の品質に関する客観的および主観的な特徴を抽出し、それらの特徴を機械学習アルゴリズムに入力することによって、作業者の品質を測定した。また、作業成果の品質を向上させるためにマルチタスク学習を行い、そのための品質測定を自動的に行うことができるようにするため、Webアプリケーションの構築を行った。評価実験を行い、提案システムの有効性を確かめた。

研究成果の学術的意義や社会的意義

本研究における成果によって、高い品質のデータを大量に作成するための方法を構築することができた。また、マルチタスク学習に必要なデータセットの構築を機械学習と人手の組み合わせにより高品質で作成することができた。複数の作業者により一つのタスクを行う場合の合意形成過程について明らかとなったため、高い精度で機械学習による分類や予測を行うために必要な方法が明らかになった。実際に企業において提案Webシステムを用いたため、実用的な精度であることも明らかになった。

研究成果の概要(英文)：Using crowdsourcing, a working platform by unspecified workers over the Internet, we construct a method for developing creative sentences. In our research, we mainly focus on two research topics; 1) how to measure the quality of outputs, and 2) how to aggregate multiple outputs for improving quality of outputs. To solve 1), we extract features from outputs and construct machine learning models. To solve 2), we use multi-task learning method. We developed Web applications to accomplish these tasks. We evaluate our system to confirm the effectiveness of our proposed method.

研究分野：情報科学

キーワード：クラウドソーシング 機械学習 データ品質 マルチタスク学習

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

本研究ではクラウドソーシングにおいて創造的な作業を大規模に行う際に、高品質な作業結果を得るための協調的作業環境を構築することを目的とする。クラウドソーシングとは大量の作業員によってインターネット上で共同で分担して作業を行い、作業量に応じた報酬額を作業員に支払う仕組みである。人手により大量の作業を短期間に終了させる必要がある場合に用いられている。ところが、多くの作業員によって創造性の高い作業を行うことは、作業員相互の意思疎通が困難であることから行われていない。そこで、高い品質で創造性の高い作業、つまり人間の直感に合う作業を行う方法を構築する。

2. 研究の目的

本研究では、作業員の品質測定システムを構築することにより、協調的な作業環境において高品質な作業結果を得ることを目的とする。具体的には、Twitter や X, YouTube などに投稿されたメッセージに対して、何に対してどのような感情を持つのか、高い精度で予測する機械学習モデルを構築する。機械学習モデルの精度を向上させるためには、入力となる訓練データの品質が向上すること、機械学習モデルを工夫することの二つについて取り組んだ。

このときに課題となる点は、教師データの品質、および構築するために必要な時間、費用である。分析対象となるメッセージが大量となったときや、多くの点で分析を行う必要があるとき、現実的な時間で分析を行うためには全ての手順をクラウドソーシングにより人手で行うことは現実的ではない。たとえば、ある製品に関して必要な分析項目（デザインや色、重さなど）が 50 個あったとき、クラウドソーシングにより各メッセージに対してどの分析項目について書かれているかを選択することは極めて複雑であり、多くの時間と費用がかかるため、LLM により処理することが有効であるといえる。一方で各分析項目が判明しておりそれぞれの分析項目に対してメッセージ投稿者がどのような感情（ポジティブやネガティブなど）を持っているかを分析するときには、クラウドソーシングにより人手で処理することが有効である。そこで本研究ではクラウドソーシングと LLM を組み合わせ、複雑な処理を現実的な時間、費用で処理可能なシステムの構築を目指す。

3. 研究の方法

以下の方法で研究を行った。

1. 作業員の作業品質の測定

作業員は多ければ多いほど、不適切な作業員の特定は容易になる。ところが、作業員を増やすことはコスト増大につながり、現実的な費用での作業が困難となる。そこで、機械学習により作業員を模倣するモデルを構築することができないかどうか、それにより作業員の品質測定が可能かどうかを調査した。

まず作業員ごとに、作業結果を模倣する機械学習モデルを構築する方法についての研究を行った。実データを利用して実験を行った。この実験では、全ての作業員による作業結果を入力としたモデルを構築し、一人の作業員の作業結果によりファインチューニングを行う方法（方法 1）と、一人の作業員の作業結果だけをを用いる方法（方法 2）を比較した。この結果、予想に反して方法 2 の精度が方法 1 と比べ高いことが明らかになった。これは、作業員により特徴的な作業結果がファインチューニングによる方法では失われてしまうためである。また、ここで構築した機械学習モデルと実際の作業を併用し、作業員の品質を測定する方法を構築した。その結果、実際の作業だけを使った場合と比較して、低品質な作業員の特定に関して精度が向上することが確認できた。

2. マルチタスク学習

マルチタスク学習を用いることを前提としたデータセット構築を行い、それに必要なデータセットの構築を行った。既存のデータセットではメッセージそれぞれに評判に関するラベルだけが付与されており、マルチタスク学習に用いることができなかった。本研究では一つのメッセージに対して様々な質問を行い、その回答を収集しているため、マルチタスク学習による精度向上が期待できる。また、LLM は BERT などによる既存の分類器よりも質問によっては高い精度で分類できる場合があるため、効果的に組み合わせることによって精度を維持しつつ時間的および金銭的成本を削減することが可能となる。

3. Web サービスの構築

クラウドソーシングにより事前に設定したタスクを、作業者が入力となるメッセージデータに対して処理する。クラウドソーシング作業画面を図に示す。Web上で主にスマートフォンを用いて作業を行うことができるよう、クラウドソーシングによる作業環境を構築した。なぜなら、従来の評価実験においてクラウドソーシングを行ったとき、80%以上の作業者がスマートフォンもしくはタブレットにより作業を行っているためである。作業者は、画面中央にあるメッセージを閲覧し、画面下にあるボタンのうち対応するものを押すことにより、次のメッセージもしくは次の質問が現れる。

構築したシステムはフロントエンドとバックエンドの二つで構成されており、フロントエンドは作業者のブラウザ上で、バックエンドは作業依頼者が準備したサーバ上で動作する。フロントエンドとバックエンドは両方ともJavaScriptで実装されており、フロントエンドはReact.js、バックエンドはNode.jsがベースとなっている。バックエンドで用いるデータベースはPostgreSQLをORMとしてPrisma.jsを介して用いている。フロントエンドとバックエンドとの通信はGraphQLによってJSONをPOSTすることによって行っている。本システムは今回の実験だけではなく様々なタスクを処理することが可能なように設計されているため、今回のタスク以外でも利用することが可能である。



図：クラウドソーシング作業画面

4. 研究成果

ここで構築した機械学習モデルと実際の作業を併用し、作業者の品質を測定する方法を構築した。その結果、実際の作業だけを使った場合と比較して、低品質な作業者の特定に関して精度が向上することが確認できた。

さらに、機械学習モデルとしてマルチタスク学習を用いる方法についての研究を行った。これは、入力と出力が対となっている通常の機械学習と異なり、一つの入力から複数のタスクを解くモデルを構築することにより、精度向上を目指す方法である。この方法を利用するためには、様々なタスクに対応する学習データが必要となり、さらにコストが増大する。そこで本研究では、マルチタスク学習に対応する学習データ構築が可能であることを確認した。さらに、構築した学習データによる機械学習モデルが精度向上に寄与することを示した。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Kuwabara Yuki, Suzuki Yu	4. 巻 32
2. 論文標題 Automatic Stopword Generation Based on Attention for Document Classification Using Neural Networks	5. 発行年 2024年
3. 雑誌名 Journal of Information Processing	6. 最初と最後の頁 487 ~ 495
掲載論文のDOI (デジタルオブジェクト識別子) 10.2197/ipsjip.32.487	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Kimura Yusuke, Komamizu Takahiro, Hatano Kenji	4. 巻 47
2. 論文標題 An Automatic Labeling Method for Subword-Phrase Recognition in Effective Text Classification	5. 発行年 2023年
3. 雑誌名 Informatica	6. 最初と最後の頁 315-326
掲載論文のDOI (デジタルオブジェクト識別子) 10.31449/inf.v47i3.4742	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 吉丸 直希, 木村 優介, 楠 和馬, 波多野 賢治	4. 巻 22-J
2. 論文標題 部分構造を考慮した化合物分散表現の食材分類タスクにおける効果	5. 発行年 2024年
3. 雑誌名 日本データベース学会和文論文誌	6. 最初と最後の頁 online
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 古田 朋也, 鈴木 優	4. 巻 J105-D
2. 論文標題 Wikipedia校閲時の事実確認作業における誤り箇所の自動推定	5. 発行年 2022年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 288 ~ 296
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2021DEP0003	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

[学会発表] 計24件(うち招待講演 0件/うち国際学会 15件)

1. 発表者名 Yu Suzuki
2. 発表標題 Measuring Quality of Workers by Goodness-of-Fit of Machine Learning Model in Crowdsourcing
3. 学会等名 International Database Engineering & Applications Symposium(IDEAS2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Yu Suzuki
2. 発表標題 An Autonomous Crowdsourcing System
3. 学会等名 International Conference on Database and Expert Systems Applications (DEXA2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Yu Suzuki, Yoshiki Yoneda, Akiyo Nadamoto
2. 発表標題 Analysis of Behavioral Facilitation Tweets for Large-Scale Natural Disasters Dataset using Machine Learning
3. 学会等名 International Conference on Database and Expert Systems Applications (DEXA2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Tomoya Furuta, Yu Suzuki
2. 発表標題 A Fact-checking Assistant System for Textual Documents
3. 学会等名 IEEE International Conference on Multimedia Information Processing and Retrieval (IEEE MIPR 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Kenji Fukumoto, Rinji Suzuki, Hiroyuki Terada, Masafumi Bato and Akiyo Nadamoto
2. 発表標題 Comparison of Deep Learning Models for Automatic Generation of Product Description on E-commerce site
3. 学会等名 International Conference on Information Integration and Web Intelligence (iiWAS2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Futo Yamamoto, Yu Suzuki and Akiyo Nadamoto
2. 発表標題 Extraction and Analysis of Regionally Specific Behavioral Facilitation Information in the Event of a Large-scale Disaster
3. 学会等名 IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '21) (国際学会)
4. 発表年 2021年

1. 発表者名 Tomoya Furuta, Yu Suzuki
2. 発表標題 A Fact-checking Assistant System for Textual Documents
3. 学会等名 International Conference on Multimedia Information Processing and Retrieval (国際学会)
4. 発表年 2021年

1. 発表者名 Yusuke Kimura, Kazuma Kusu, Kenji Hatano, Tokiya Baba
2. 発表標題 Automatic Terminology Extraction using A Dependency-Graph in NLP
3. 学会等名 Proceedings of the 9th World Congress on Information and Communication Technologies, (国際学会)
4. 発表年 2020年

1. 発表者名 Taisho Sasada, Zhaoyu Liu, Tokiya Baba, Kenji Hatano, Yusuke Kimura
2. 発表標題 A Resampling Method for Imbalanced Datasets Considering Noise and Overlap
3. 学会等名 Procedia Computer Science (国際学会)
4. 発表年 2020年

1. 発表者名 Rinji Suzuki, Akiyo Nadamoto
2. 発表標題 Extracting Rhetorical Question from Twitter
3. 学会等名 Proceedings of The 22nd International Conference on Information Integration and Web-based Applications & Services (iiWAS2020) (国際学会)
4. 発表年 2020年

1. 発表者名 笹田 大翔, 馬場 睦也, 木村 優介, 波多野 賢治
2. 発表標題 データの複雑性を考慮した不均衡データに対するリサンプリング手法
3. 学会等名 情報処理学会第 82 回全国大会
4. 発表年 2020年

1. 発表者名 Yoshiki Yoneda, Yu Suzuki, and Akiyo Nadamoto
2. 発表標題 Behavioral Facilitation Tweets with Basis Extraction
3. 学会等名 IEEE International Conference on Data Mining Workshops (国際学会)
4. 発表年 2019年

1. 発表者名 Yoshiki Yoneda, Yu Suzuki, and Akiyo Nadamoto
2. 発表標題 Detection of Behavioral Facilitation information in Disaster Situation
3. 学会等名 International Conference on Information Integration and Web-based Applications & Services (国際学会)
4. 発表年 2019年

1. 発表者名 米田吉希, 見塚圭一, 鈴木優, 瀬本明代
2. 発表標題 機械学習を用いた災害時行動促進情報抽出手法の比較検討
3. 学会等名 Webとデータベースに関するフォーラム
4. 発表年 2019年

1. 発表者名 Kosuke Wakasugi, Futo Yamamoto, Yu Suzuki, Akiyo Nadamoto
2. 発表標題 Feature Analysis of Regional Behavioral Facilitation Information Based on Source Location and Target People in Disaster
3. 学会等名 International Conference on Big Data Analytics and Knowledge Discovery (DaWaK) (国際学会)
4. 発表年 2023年

1. 発表者名 Nana Ota, Yu Suzuki
2. 発表標題 A Label Aggregation Method Using Worker Quality in Crowdsourcing
3. 学会等名 International Conference on Information Integration and Web-based Applications & Services (iiWAS) (国際学会)
4. 発表年 2023年

1. 発表者名 Yuki Kuwabara, Yu Suzuki
2. 発表標題 Attention Based Stopword Generation for Neural Network Based Text Processing
3. 学会等名 International Conference on Information Integration and Web-based Applications & Services (iiWAS) (国際学会)
4. 発表年 2023年

1. 発表者名 鈴木 優, 城所 祥太, 北村 拓斗, 太田 奈那, 三島 惇也, 本田 文隆, 長徳 弘一, 中村 勇太, 佐々木 和彦, 笠丸 陽一
2. 発表標題 機械学習とクラウドソーシングを用いた顧客の声の自動分類
3. 学会等名 第16回データ工学と情報マネジメントに関するフォーラム (DEIM2024)
4. 発表年 2024年

1. 発表者名 エルゲン 瑛夏, 鈴木 優
2. 発表標題 List-wise手法における順位に基づいた多段階分割と学習データの上位遷移による能動学習
3. 学会等名 第16回データ工学と情報マネジメントに関するフォーラム (DEIM2024)
4. 発表年 2024年

1. 発表者名 川上 大凱, 鈴木 優
2. 発表標題 動画の属するコミュニティ情報を考慮した動画再生回数予測手法の提案
3. 学会等名 第16回データ工学と情報マネジメントに関するフォーラム (DEIM2024)
4. 発表年 2024年

1. 発表者名 城所 祥太, 鈴木 優
2. 発表標題 クラウドワークの作業結果と投票結果から生成される確率分布を利用した結果集約手法
3. 学会等名 第16回データ工学と情報マネジメントに関するフォーラム (DEIM2024)
4. 発表年 2024年

1. 発表者名 若杉 広介, 鈴木 優, 灘本 明代
2. 発表標題 ユーザの性格タイプの差異に基づく災害時行動促進情報理解の分析
3. 学会等名 第16回データ工学と情報マネジメントに関するフォーラム (DEIM2024)
4. 発表年 2024年

1. 発表者名 沢田 凌一, 鈴木 優
2. 発表標題 誹謗中傷検出タスクの学習を補助するためのサブタスクの提案
3. 学会等名 第16回データ工学と情報マネジメントに関するフォーラム (DEIM2024)
4. 発表年 2024年

1. 発表者名 城所 祥太, 鈴木 優
2. 発表標題 クラウドソーシングにおけるラベル付け難易度を考慮したタスク割当手法
3. 学会等名 電子情報通信学会データ工学研究会研究報告
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	灘本 明代 (Nadamoto Akiyo) (30359103)	甲南大学・知能情報学部・教授 (34506)	
研究 分担者	波多野 賢治 (Hatano Kenji) (80314532)	同志社大学・文化情報学部・教授 (34310)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------