称号及び氏名	博士（工学） Md. Rabiul Islam
学位授与の日付	2022 年 9 月 23 日
論　文　名	In-The-Wild Reading Activity Evaluation Using Self-Supervision
論文審査委員	主査　黄瀬　浩一 副査　本多　克宏 副査　能島　裕介

論文要旨

Reading is a cognitive human activity that decodes a series of written symbols to get meaning from them. In our daily life, we perform different reading activities, including the common task of reading plain texts for different purposes such as to learn information or facts, to entertain ourselves, or to understand other cultures or groups. The reading activity evaluation quantifies the read amount of each such activity. Another reading activity is solving problems and producing task outputs performed via reading. The evaluation of these activities assess the task outputs via estimating, for example, the number of confident or correct outputs. Therefore, by reading activity evaluation, we do not simply mean evaluating the activity of reading plain texts but also evaluating task outputs produced via reading. All these activities, including similar others, create a reading activity spectrum.

Reading activity evaluation gives quantitative and qualitative information on reading. For instance, reading materials are everywhere and reading detection by identifying subordinate categories of reading scripts, layouts, and document types allow obtaining reading habits. For example, reading Japanese vertical documents often indicates that the user reads novels or newspapers. Another instance is estimating confidence in the task outputs produced via reading. A correct but unconfident answer needs review to ensure the knowledge can be used in the future and an incorrect and confident answer requires more attention to revise the incorrect knowledge. Therefore, this allows us to produce better strategies for personalized review. Another instance is estimating the correctness of the task outputs produced via reading. The confidence and correctness are highly correlated. We, therefore, can also estimate the correctness of task outputs. This allows us to assess the quality of task outputs collected from the crowd (crowdsourcing) performed via reading by estimating the correctness.

The reading activities are reflected by reading behaviors, such as a reader having difficulty in understanding the contents of a document is characterized by low reading speed and frequent rereading. Therefore, a fundamental way of reading activity evaluation is analyzing reading behaviors.

Reading activity evaluation by analyzing data describing reading behaviors can be conducted in multiple ways. Classical machine learning algorithms, with handcrafted features employing data from laboratory settings, suffer from multiple issues, including poor performance. We need to let the user be free from the laboratory constraints by recording data outside the laboratory (in-the-wild), to capture natural reading behaviors and more realistic evaluation of the developed systems. Such noisy in-the-wild data obfuscates effective feature extraction that is needed with classical algorithms. Deep learning (DL) shows the path to handling such data but requires large labeled data. Obtaining large in-the-wild labeled reading behavior datasets is problematic because the annotation cost and time, diversity of employed devices, variations in specifications regarding sampling rates, and different deployment environments make dataset construction a challenging task. Moreover, a user may not be comfortable recording reading behaviors for an extended period because of privacy issues. Finally, considering all issues, constructing large in-the-wild labeled datasets for reading behavior analysis employing the fully supervised DL algorithms is always intricate.

A DL technique called self-supervision presents a potential solution in various domains, including physical human activity recognition to distinguish body movements that characterize activities. This technique can be applied as noncontrastive self-supervised learning (SSL) and contrastive self-supervised learning (contrastive learning). The self-supervision techniques pre-train the model using data synthesized from unlabeled data employing an automated process, followed by the target task (task of interest) training using small labeled data in a supervised manner. The data for pre-training are prepared without manual labeling, and the model, therefore, can be trained with a larger data size that improves model performances for the target task by enhancing generality and data efficiency.

However, this is unknown if similar approaches are effective for tackling the lack of labeled data issues in recognizing cognitive human activities such as reading, where the task is cognitively intensive with fewer bodily movements, and special sensors are needed to capture physiological signals such as eye movements that describe reading behaviors. This thesis aims to investigate self-supervision techniques in tacking the lack of large labeled data for reading activity evaluation by employing in-the-wild datasets that have been carried out in the following way.

We propose self-supervision methods and evaluate them for three different but related reading activity evaluation tasks placed at two extreme points on the reading activity spectrum. The first one is reading detection, a low-level reading activity relevant to the quantity of reading. The second one is confidence estimation, a high-level reading activity relevant to the quality of reading. The third one is correctness estimation which is also a high-level reading activity relevant to reading quality. In this case, the task is formatted for the quality assessment of crowdwork. The trials on both quantity and quality of reading activities allow obtaining a full picture of the effectiveness of the self-supervision technique across the reading activity spectrum.

As for the detailed content, the thesis is organized as follows:

In **Chapter 1**, we describe the background, problem statements, contributions, and outline of the thesis chapters in *Sections 1.1, 1.2, 1.3,* and *1.4*, respectively.

In **Chapter 2**, we analyze the literature to find directions to answer the research questions. In *Section 2.1*, we give an overview of reading behaviors and existing sensor technologies for capturing them. In *Section 2.2*, we present the existing self-supervision techniques that have solved various issues in many domains. In *Sections 2.3, 2.4,* and *2.5*, we give an overview of existing technologies for reading detection, confidence estimation, and quality assessment of crowdwork, respectively, showing that the existing reading analysis methods are mainly based on classical machine learning algorithms employing laboratory settings. Finally, *Section 2.6* conclude the chapter.

In **Chapter 3**, we present an in-the-wild data preparation workflow for reading detection. In *Section 3.1*, we give an overview of the necessity of in-the-wild studies and the motivation for developing an in-the-wild data preparation workflow. In *Section 3.2*, we present a prelude study to show the effect of laboratory restrictions and conditions on the user's naturalistic reading behaviors and labeled data. In *Section 3.3*, we present the proposed in-the-wild data preparation workflow describing the study design, sensor devices, and tools used by focusing on the directions obtained from the prelude study. In *Section 3.4*, we discuss related approaches, implications, and limitations of the proposed in-the-wild data preparation workflow. Finally, we conclude the chapter in *Section 3.5*.

In **Chapter 4**, we present the reading detection employing data recorded using an in-the-wild data preparation workflow. In *Section 4.1*, we describe the proposed SSL method. In contrast to the conventional reading versus not reading detection, we propose the SSL method for solving a fine-grained reading detection task that includes four subordinate classes. Reading detection is a low-level reading activity evaluation task reflected by eye movements and slight head and body movements. In such a fine-grained reading detection task, the large labeled data is always scarce, and this section presents a method to handle such issues using the SSL employing electrooculography data, describes eye movements and motion sensors data that describe head and body movements. Besides, in *Section 4.2*, we present a study to explore the above-mentioned best sensor's combination for fine-grained reading detection that includes background study, methods, evaluation, and results with experimental conditions and discussion to give a direction for computation cost-effective systems development. In *Section 4.3*, we present the reading detection by contrastive learning, a parallel counterpart of the proposed SSL method, by describing the background, employed contrastive learning method (SimCLR) with the required modifications for adaptation to the fine-grained reading detection, and results with experimental conditions and discussion. Finally, *Section 4.4* concludes the chapter by comparing proposed SSL and contrastive learning methods.

After showing that the proposed SSL method can handle the lack of large labeled data issues for the low-label reading activity evaluation task of reading detection, we extend the proposed SSL method for a high-level reading activity evaluation task in **Chapter 5**. We handled the task of confidence estimation in answering MCQs by analyzing eye gaze data that describes reading behaviors. We give an overview of the importance and issues arising from the lack of labeled data in solving this task in *Section 5.1*. In *Section 5.2*, we present the proposed SSL method for confidence estimation in answering MCQs. In *Section 5.3*, we present the data preparation by describing in-the-wild data recording and pre-processing for handling the length variability of eye gaze data for each MCQ. In *Section 5.4*, we present the evaluation and results by describing a novel approach to handling imbalanced data. Finally, we close the chapter by concluding that eye gaze tells us the confidence in the answer of MCQs in *Section 5.5*.

After having shown that the eye gaze carries a piece of information about confidence in the answer of a task, we introduce the first step toward a technique in quality assessment of crowdwork by predicting the correct answer of a task employing eye gaze data in **Chapter 6**. This is also a high-level reading activity evaluation task related to the quality of reading, and the lack of labeled data is a serious issue. In *Section 6.1*, we first introduce the quality assessment of crowdwork, including the drawbacks of the existing outcome-based quality assessment techniques and the possible solution by employing the worker behaviors. In *Section 6.2*, we propose two methods for quality assessment of crowdwork by analyzing the eye gaze data by presenting the task formation, feature extraction using the proposed SSL method, which is an extension of the proposed SSL method for confidence estimation with the necessary modifications, and the target task model generation. In *Section 6.3*, we present in-the-wild data preparation for evaluating the proposed methods. In *Section 6.4*, we present experimental conditions, results, and discussion. Finally, *Section 6.5* closes the chapter by showing that the eye gaze data show a richer fingerprint for the quality assessment of crowdwork.

Finally, in **Chapter 7**, we conclude the thesis with a summary of findings in *Section 7.1* and limitations and future work in *Section 7.2* with some lines of thoughts for further improvements.

審査結果の要旨

本論文は，実環境における人の読書行動の評価の目的のため，深層学習の一手法である自己教師あり学習の適用に関する研究についてまとめたものであり，以下の成果を得ている．

(1)　従来，人の読書行動の評価には，実験室環境で様々な制限を伴いつつ取得されたデータが用いられていた．一方で，技術の実用を考えたとき，データは制限のない実環境(In-the-wild)で取得されるべきである．その際に大きな問題となるのは，人のプライバシに配慮しつつ，いかに効率的・効果的にデータの取得とラベル付けを行うかという点である．本研究では，ウェアラブルカメラを用いた画像取得と，効率的な画像ブラウジングを用いた被験者自身によるラベル付けという新しい解法を提案し，有効性を確認した．

(2)　読書行動評価の量的な側面として読書検出を取り上げ，眼電位，加速度，角速度を計測可能な眼鏡型センサから得たデータに対して動作する手法を提案した．その際，ラベル付きデータの不足に対処するため，非対照学習（non-contrastive learning）と対照学習(contrastive learning)という 2 種類の自己教師あり学習を導入し，学習に用いるラベル付きデータの数を変化させながら精度を評価した．その結果，いずれの提案手法についても，ラベル付きデータの数にかかわらず，従来の古典的機械学習や単純な深層学習に基づく手法に比べて精度が高いこと，特にラベル付きデータの数が少ない場合に非対照学習が有効であることを確認した．

(3)　読書行動評価の質的な側面として多肢選択問題解答時の確信度推定を取り上げ，アイトラッカを用いて得た視線データに対して動作する手法を提案した．その際，上記と同様に，ラベル付きデータの不足に対処するため，非対照学習を導入し，学習に用いるラベル付きデータの数を変化させながら精度を評価した．その結果，ラベル付きデータの数にかかわらず，従来の古典的機械学習や単純な深層学習に基づく手法に比べて提案手法の精度が高いこと，特にラベル付きデータの数が少ない場合に，提案手法が有効であること確認した

(4)　読書行動は学習のみならず文書を扱う仕事においても避けられないことに鑑み，クラウドソーシングの文脈において，クラウドワーカが行う仕事の質評価に取り組んだ．具体的には，ワーカの仕事時の振る舞いに基づいて，仕事の結果の正誤を推定する手法，ならびに，多くのワーカが同じ結果を得るかどうかを推定する手法の 2 つを提案した．同様に，古典的機械学習に基づく手法と比較して，提案手法の精度が高いことを確認した．

以上の研究成果は，実環境における人の読書行動の評価に関する重要な知見を与えるとともに，本分野の学術的・産業的な発展に寄与するところが大である．また，申請者が自立して研究活動を行うに必要な能力と学識を有することを証したものである．

審査委員会は 2022 年 8 月 5 日，委員全員の出席のもとに，申請者に論文内容の説明を行わせ，関連する諸問題について試問を行った結果，合格と判定した．