

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	ユビキタスコンピューティングシステム (安本 慶一 (教授))		
学籍番号	2211255	提出日	令和 6年 1月 18日
学生氏名	福光 嘉伸		
論文題目	クラウドソーシングベースのアノテーションタスクにおける回答の質低下の動的検知		
要旨			
<p>アノテーション作業をクラウドソーシングで行うことにより、低コストで機械学習のための学習データを収集できる。しかし、得られるデータの品質に大きなばらつきがあり、対価として報酬を付与すると可能な限り速く回答を行おうとする行動によって不良回答が発生する問題がある。誤ったラベルが含まれる学習データを用いると、機械学習モデルの精度が低下する恐れがあるため、不良回答の発生を検知・防止することが重要となる。先行研究では、評価尺度を取り入れずに努力の最小限化の傾向を検出することを目的として、スマートフォンの画面操作記録を特徴量とする機械学習による検出手法を提案しているが、この手法ではアンケートの回答が全て終わってからしか検出を行うことができない。そのため、データの母集団が限られている場合、不良回答の影響によってデータが不足する恐れがある。そこで、本研究ではアノテーションタスク、特に固有表現 (Named Entity) アノテーションを対象として不良回答をリアルタイムで検出することを目的とし、アノテーション作業中のクリックやマウスマウスカーソルの移動等の画面操作から得られる特徴量を用いた検出手法を提案する。オープンソースソフトウェアのアノテーションツールであるLabelStudioを用いて画面操作を記録するプラグインを開発し、学内学生およびクラウドワーカーを対象として画面操作データを収集して機械学習による不良回答検出を行った。LightGBM (LGBM) を分類アルゴリズムとして用い、Leave-One-Participant-Out交差検証による評価を行った結果、学内学生を対象とした実験では0.738のAccuracyが得られた。また、不良回答ではラベル付与毎の平均カーソル移動量・クリック回数・文字選択回数が多くなり、分類において重要であることが分かった。</p>			