

# コードレビュー分析におけるデータクレンジングの影響調査

戸田 航史<sup>1</sup> 亀井 靖高<sup>2</sup> 吉田 則裕<sup>3</sup>

**概要:** 本論文ではコードレビュー分析に対してデータクレンジングが与える影響を調査する。調査では、オープンソースソフトウェア開発プロジェクトである Android, Chromium, OpenStack の 3 プロジェクトを対象とした。クレンジングはレビューとレビュー開始・終了日時の 2 つに対して実施した。レビューへのクレンジングとして、ビルドやテストの自動化を行う bot の除去を、レビュー開始・終了日時へのクレンジングとして、実際のレビューの状況をふまえた補正を行った。3 プロジェクトから取得したデータを用いて、レビュー、レビュー開始・終了日時について、クレンジングを行ったデータと行わなかったデータをそれぞれ作成、比較した。比較の結果、bot によるレビューは OpenStack では全体の 19.4% を占めていること、レビュー開始・終了日時へのクレンジングの有無により、レビュー期間に有意な差が出ることが分かった。この結果から、データクレンジングはコードレビュー分析に影響を与えうることが分かった。さらに具体的なコードレビュー分析として、レビュー経験とレビュー期間の関係(相関)の分析を対象に、レビュー、レビュー開始・終了日時のデータクレンジングを実施したデータと実施しなかったデータを分析し、その結果を比較した。比較の結果、クレンジングを行わなければ相関係数に影響を与えることが分かった。

本招待論文は、情報処理学会論文誌に掲載されました  
「コードレビュー分析におけるデータクレンジングの影響調査」 [1] についてご紹介いただくものです。

## 参考文献

- [1] 戸田航史, 亀井靖高, 吉田則裕: コードレビュー分析におけるデータクレンジングの影響調査, 情報処理学会論文誌, Vol. 58, No. 4, pp. 845-854 (2017).

---

<sup>1</sup> 福岡工業大学  
<sup>2</sup> 九州大学大学院  
<sup>3</sup> 名古屋大学