

# OSSに寄付をすべきか？

## 統計的因果推論による寄付バッジの効果分析

中才 恵太郎<sup>1,a)</sup> 畑 秀明<sup>1,b)</sup> 角田 雅照<sup>2,c)</sup> 松本 健一<sup>1,d)</sup>

**概要:** [背景] OSS の Eclipse は寄付を募集しており、2016 年 9 月より、バグレポートにおいて寄付者には寄付バッジをつけるという特典を提供している。寄付バッジを持つことは何らかのステータスになると思われるが、どのような効果があるかは明らかではない。[目的] 寄付バッジの効果を定量的に明らかにする。[手法] 寄付バッジの有無で、バグレポート投稿から初めてのコメントがつくまでの時間にどれほどの違いがあるのかを、統計的因果推論のアプローチで分析する。[結果] 寄付バッジを持つバグレポート投稿者のバグレポートはそうでないバグレポートと比べて 0.88 日早くコメントが得られていた。ただし、分析数が少ないため、統計的に有意な差ではない。[結論] さらにデータ数を増やして分析を進める必要はあるが、寄付バッジは対応を早める効果がある可能性がある。この効果が有用であればこの寄付バッジを得るために寄付をすることは合理的と判断できる。

### 1. はじめに

オープンソースソフトウェア (OSS) は無償で利用できるため、企業や官公庁、個人でも多くの利用がある。しかし、無償で利用できる OSS においてもサーバ費やソフトウェアメンテナンスのためのフルタイム開発者の賃金などが必要であるため、資金を集める必要がある。そこで、多くの OSS 運営団体では寄付を募っており、重要な資金源としている。

OSS の統合開発環境 (IDE) として有名な Eclipse プロジェクトでは寄付を募っており、寄付者に対し、特典を付与している。本稿では Eclipse の寄付の特典の一つである、バグ報告システム上の寄付バッジの効果について検証する。

Eclipse のバグレポートに関わった人のうち 1%ほどが寄付バッジが付与されている [1]。寄付バッジがついているバグレポートには、早期に対応しようという心理が開発コミュニティに働く可能性がある。しかし、Eclipse のバグ報告システム上の寄付バッジの効果を検証した研究は存在しない。Steinmacher らは新規開発者の OSS プロジェクト

への参入を妨げる要因について分析しており、その中で返事が返ってこないことを報告している [4]。寄付バッジに早期対応の効果があるとすれば、新規に OSS プロジェクトへ貢献したいと考えている人にこそ寄付は合理的であると考えられる。寄付バッジの効果进行分析に当たって、寄付バッジがついているバグレポートと寄付バッジがついていないバグレポートのバグレポート作成時から第三者によるコメントがつくまでの時間を比較することで寄付バッジの効果を分析する。

### 2. アプローチ

統計的因果推論の手法の一つである傾向スコアマッチング [2] と差分の差分分析を用いて分析する。効果を分析したい事象が介入する群を介入群、介入しない群を対照群とする。

傾向スコアマッチングは、分析したい事象に関するメトリクス以外のメトリクスが、介入群と類似した条件にある対照群を選出し、比較可能とする手法である。差分の差分分析は、介入群と対照群において、分析したい事象が起こる前と後のデータを使用する。対照群の分析したい事象が起こる前と後の変化量により介入群が分析したい事象に介入されない状況を仮定し、仮定した結果 (反実仮想) と実際の結果の差分から、分析したい事象の因果効果を推論する手法である。分析したい事象が起きる前と後の変化量から、分析したい事象の効果を推論する。

<sup>1</sup> 奈良先端科学技術大学院大学  
Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan

<sup>2</sup> 近畿大学  
Kindai University, Higashiosaka, Osaka 577-8502, Japan

a) nakasai.keitaro.nc8@is.naist.jp

b) hata@is.naist.jp

c) tsunoda@info.kindai.ac.jp

d) matumoto@is.naist.jp

### 3. 分析と結果

#### 3.1 データ

分析対象データの概要を表 1 に示す。過去でのバグレポートの寄付バッジの有無の特定には、寄付者リストを参照した。各バグレポートの経験年数を考慮するため、すべてのバグレポートからバグレポートの経験年数を計測した。

#### 3.2 介入群と対照群の抽出

##### コンポーネント群の選択

バグレポートの対象となるコンポーネントによってコメントがつくまでの時間に偏りがある。2016年3月1日から2017年3月1日までに寄せられたバグレポートそれぞれについて、コメントがつくまでの時間を ScottKnott クラスタリング [3] により、4つに分類した。そのうちもっともバグレポートが多く分類されたコンポーネント群のバグレポートのみを使用した。

##### 傾向スコアマッチング

寄付バッジがついたバグレポートを介入群、寄付バッジがついていないバグレポートを対照群とし、それぞれの投稿したバグレポートを分析する。寄付以外の条件が近いバグレポートを介入群と対照群として抽出するため、傾向スコアマッチングを用いる。傾向スコアを用いると、対象とする説明変数（バッジの有無）の共変量を調整できる。傾向スコアマッチングの結果を表 2 に示す。介入群、対照群は各 15 人となった。傾向スコアマッチングにより、介入群も対照群も経験年数が類似するグループとなった。

#### 3.3 差分の差分分析

Eclipse では 2016 年 9 月 1 日にバグ報告システムにおいて、寄付バッジを導入した。これを介入が起こったとみなし分析を行う。介入前を 2016 年 3 月 1 日から 2016 年 8 月 31 日まで、介入後を 2016 年 9 月 1 日から 2017 年 3 月 1 日とした。その両方にバグレポートを投稿したバグレポートを対象に分析した。

差分の差分分析の結果を図 1 に示す。縦軸がコメントがつくまでの日数である。バグレポートごとに各レポートのコメントがつくまでの時間の中央値をとり、対照群、介入群の中央値から因果効果を推定した。反実仮想が、寄付バッジありの群が仮に寄付バッジなしの状況を仮定したものである。反実仮想の値から介入後の値の差が因果効果である。寄付バッジがついていることにより 0.88 日早くコメントがつくことが推定された。

しかしながら、マン・ホイットニーの U 検定により、因果効果に有意差があるか検定し、p 値は 0.41 となった。ゆえに統計的に有意ではない。分析対象が少なかったからではないかと考えられる。

表 1 分析対象データの概要

Table 1 Analysis data

|        | 分析対象期間              | データ数    |
|--------|---------------------|---------|
| 寄付リスト  | 2007/12/30~2017/3/1 | 32,428  |
| バグレポート | 2001/10/19~2017/3/1 | 512,938 |

表 2 傾向スコアマッチングの結果

Table 2 Matching data

|              | 介入群の平均値 | 対照群の平均値 |
|--------------|---------|---------|
| バッジ導入前のレポート数 | 18.87   | 12.80   |
| バッジ導入後のレポート数 | 21.53   | 18.53   |
| 経験年数         | 6.87    | 6.67    |

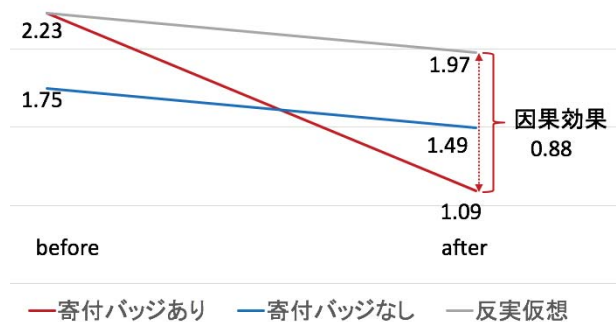


図 1 差分の差分分析の結果 (日)

Fig. 1 Result of Difference in Difference

### 4. おわりに

Eclipse の寄付バッジの効果を統計的因果推論の手法を用いて分析した。しかし今回は分析数が少なかったため統計的に有意な差は得られなかった。分析数を増やし、分析を進める必要はあるが、寄付バッジは対応を早める可能性があり、効果が有効であるならば、寄付バッジを得るために寄付をすることは合理的であると判断できる。

謝辞 本研究は JSPS 科研費 16H05857 の助成を受けた。

#### 参考文献

- [1] Nakasai, K., Hata, H., Onoue, S. and Matsumoto, K.: Analysis of Donations in the Eclipse Project, *2017 8th International Workshop on Empirical Software Engineering in Practice (IWESEP)*, pp. 18–22 (2017).
- [2] Rosenbaum, P. R. and Rubin, D. B.: The central role of the propensity score in observational studies for causal effects, *Biometrika*, Vol. 70, No. 1, pp. 41–55 (1983).
- [3] Scott, A. J. and Knott, M.: A cluster analysis method for grouping means in the analysis of variance, *Biometrics*, pp. 507–512 (1974).
- [4] Steinmacher, I., Conte, T., Gerosa, M. A. and Redmiles, D.: Social Barriers Faced by Newcomers Placing Their First Contribution in Open Source Software Projects, *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW 15*, New York, NY, USA, ACM, pp. 1379–1392 (2015).