

# Relationship between Geographical Location and Evaluation of Developer Contributions in GitHub

---

Ayushi Rastogi, Nachiappan Nagappan,  
Georgios Gousios, Andre van der Hoek

(紹介担当:阿萬@愛媛大)

# 概要

---

## 【目的】

GitHub における**開発者の地理的な位置**と**貢献の評価**の関係に関する定量的な分析

## 【方法】

最も活発な**17カ国**の参加者からの**7万件を超えるプルリクエスト**を分析し、開発者の地理的な位置とプルリクエストの採択の関係をモデル化

## 【結果】

**国によって採択率に違い**あり: スイスと日本が高く、中国とドイツは低い

プルリクエストが**同じ国の開発者に評価されたとき**は採択率が高まる傾向にある

# 背景: 地理的なバイアス

---

- 人の**地理的な位置** (国や地域) という属性が**評価に影響**するということは長く知られてきた

(例) オリンピック等での審判<sup>[22]</sup>, 論文の引用<sup>[18]</sup>

- 近年, オープンソース開発では, **貢献者の性別**が評価のバイアスになっている (**女性の方が採択されにくい傾向**) という報告<sup>[25]</sup>  
→ **地理的なものでも違いが出るかも?**

# 背景：プルリクエスト

---

- 今日，GitHub 上では約半数のプロジェクトがプルリクエストベースの開発を行っている
- **プルリクエストの採択**に**影響するかもしれないファクター**が研究されている
  - 開発者の技術的・社会的な評判<sup>[7][14][19][26]</sup>
  - プロジェクトの規範への準拠<sup>[14][16][26]</sup>
  - プルリクエストでの変更の規模，品質，テーマや目的<sup>[14][26][30][24]</sup>
- **どの国に住んでいるのかも影響するかも？**

# 収集データ

---

- プルリクエストの**投稿者(submitter)**と**担当者(integrator)**それぞれの**居住国**
  - GitHub プロファイル等を解析
- これ以外は**先行研究で議論されていたもの**を使用
  - ただし、性別データは今回はうまく使えていない
- 表1参照(次頁)

**Table 1: Factors influencing pull request acceptance**

Characteristic	Measure			
	Tsay et. al [26]	Gousios et. al [14]	Our study	Variable name
<b>Project characteristics</b>				
Maturity	#months in existence	-	#months in existence	proj_months_existence
Team size	#contributors	#active core team members	#active core team members	-
Popularity	#watchers	#watchers	#watchers	proj_watchers
Size of code	-	#non-comment LOC	#non-comment LOC	proj_nloc
Openness to external contributions	-	%external contributions	%external contributions	proj_external_contribs
Code quality	-	#test LOC per 1,000 LOC	#test LOC per 100,000 LOC	proj_test_loc_per_llcc
	-	#test cases per 1,000 LOC	#test cases per 100,000 LOC	-
	-	#asserts per 1,000 LOC	#asserts per 100,000 LOC	-
<b>Developer characteristics</b>				
Status in community	#followers	#followers	#followers	dev_followers
Status in project	direct commit access (yes=1; no=0)	-	direct commit access (yes=1; no=0)	dev_commit_access
Social closeness	followed integrator prior (yes=1; no=0)	-	followed integrator prior (yes=1; no=0)	dev_followed_integrator
	watched project prior (yes=1; no=0)	-	watched project prior (yes=1; no=0)	dev_watched_project
Experience	-	#previous pull requests on project	#previous pull requests on project	dev_prev_pull_requests
	-	previous pull request success rate	previous pull request success rate	dev_success_rate
	-	-	#months of project participation	dev_months_participation
<b>Pull request characteristics</b>				
Uncertainty of pull request	#comments	#comments	#comments	pr_comments
Size of change	#changed LOC	#changed LOC	#changed LOC	pr_changed_loc
Quality	#files changed	#files changed	#files changed	pr_changed_files
	inclusion of tests (yes=1; no=0)	#changed LOC in tests	inclusion of tests (yes=1; no=0)	pr_test_inclusion
<b>Geographical location</b>				
Country of submitter			country	geo_country
Same country submitter/integrator			same country (yes=1; no=1)	geo_same_country

# 対象データ

---

- GitHub においてプルリクエストの多い上位1% に相当する **1,069 個のプロジェクト**: 以下の条件に合う **70,740 個のプルリクエスト**
  - マージされた(採択)プルリクエストとされなかった(不採択)プルリクエストのみを対象; **インテグレータ本人によるプルリクエストは除外** (なんと37%もあった)
  - 1,000件以上のプルリクエストのあった国に限定 (**17カ国**: アメリカ, イギリス, ドイツ, フランス, カナダ, 日本, ブラジル, オーストラリア, ロシア, オランダ, 中国, スペイン, イタリア, スイス, スウェーデン, イタリア, ベルギー) ※これらで83%を占めていた

# 統計的手法

---

## ■モデル

ロジスティック回帰モデル

## ■目的変数

プルリクエストの採否

## ■説明変数

【モデル1】 従来のメトリクス + **投稿者の国**

【モデル2】 (同上) + 投稿者と管理者が**同国か**

# 結果：投稿者の国

---

(アメリカを100として)

## ■採択されやすい

- スイス(146), オランダ(130), 日本(128), イギリス(114), カナダ(113)

## ■採択されにくい

- スウェーデン(81), ドイツ(78), ブラジル(76), イタリア(73), 中国(68)

## ■どちらともいえない

- ベルギー(109), スペイン(108), オーストラリア(105), インド(102), フランス(102), ロシア(94)

# 結果：投稿者と担当者が同じ国の人かどうか

---

■ 17ヶ国中16ヶ国（インド以外）では、**同じ国の人からの投稿の方が採択されやすい**という傾向にあった

- インド： -7.4%

- アメリカ：+2.4%

.....

- 中国： +13.4%

# まとめ

---

- 新たな視点として、**開発者の地理的な位置（国，地域）**に着目
- プルリクエストの投稿者の**国によって採択率に違い**が見られた
- 投稿者（submitter）と管理者（integrator）が**同じ国の人間だと、採択率が高まる**傾向が見られた
- 理由・原因の解明はこれから