

ソフトウェア開発の定量的分析における傾向スコアの活用

角田 雅照¹ 天喜 聰介²

概要：ソフトウェア開発において収集されるデータでは、項目間に相互に関連が見られる場合があり、これにより誤った結論を導いてしまう場合がある。本稿では、その関連を考慮する方法として傾向スコアに着目する。傾向スコアにより、それらの項目の個別の影響を明らかにすることができます。さらに、傾向スコアによるインターネット調査の補正についても取り上げ、傾向スコアのソフトウェア工学への適用可能性について議論する。

Using Propensity Score for Empirical Analysis of Software Development

MASATERU TSUNODA¹ SOUSUKE AMASAKI²

Abstract: On a datasets collected in software development, there are relationships between attributes, and it sometimes causes spurious relationship. To consider the relationships, we focus on propensity score. Using the propensity score can identify influence of each attribute separately. Additionally, we pick up calibration of a web-based questionnaire using the propensity score. This paper discusses application of the propensity score to software engineering research.

1. はじめに

大規模なソフトウェアにおいて、品質の低下、納期の遅延、コストの超過などのプロジェクトの失敗を防ぐためには、定量的なデータに基づいてプロジェクトを管理することが重要となる。定量的なプロジェクト管理を支援するために、ソフトウェア開発プロジェクトにおいて収集されたデータを分析し、知見を得ることが行われる。例えば、生産性に影響する要因の分析[4]などが行われており、分析の結果は、プロジェクト計画やプロセス改善に役立てることができる。

ソフトウェア開発において収集されるデータでは、項目間に相互に関連が見られる場合があり、これにより誤った結論を導いてしまう場合がある。例えば、あるツールを導入することにより、生産性向上の効果があるかどうかを明らかにしたいとする。この場合、データを分析した結果、ツールの導入有無により生産性に差異があったとしても、必ずしもツールの導入に効果があったと結論付けることはできない。例えば、ツールを導入している企業のうち、多くの企業では技術者のスキルが高いとすると、生産性に差異を生じさせている主な要因は技術者のスキルである可能性がある。このような、原因と結果の両方の変数に影響している変数が存在することを交絡と呼ぶ。図 1 に交絡の例を示す。

ツールの導入と技術者のスキルに関連がなければ、ツー

ルの導入と技術者のスキルを説明変数、生産性を目的変数として重回帰分析を行い、ツールの導入効果を確かめることができる。しかし、上記の例のように、ツールが使われている企業では技術者のスキルが高いといった場合では、重回帰分析を行っても、ツール単独の効果を確かめることはできない。なお、生産性を予測することを目的とし、ツールの生産性に対する効果を明らかにする必要がないならば、重回帰分析などが適用可能である。

本稿では、ソフトウェア開発において収集される項目に相互に関連が見られる場合に、その関連を考慮する方法として、傾向スコア[6]に着目する。さらに、傾向スコアのその他の活用方法についても議論する。

2. 傾向スコア

2.1 概要

原因と結果の両方の変数に影響している変数が存在する場合、それを考慮して分析する方法として、傾向スコア(propensity score)を用いた分析方法が提案されている。傾向スコアを用いた分析では、前章で述べたようなデータの偏りが存在する場合でも、ある説明変数(一般的には二値をとるダミー変数)の目的変数に対する影響を、他の説明変数の影響を除外して明らかにすることができます。

例えば、喫煙と大腸がんの関係を分析する場合、飲酒量が喫煙と大腸がんの両方に関係していると、喫煙の有無でデータを層別しても、大腸がんに対する飲酒量の影響が除去されず、喫煙の大腸がんに対する影響を知ることができない[2]。傾向スコアはこのようなデータの分析に有効であり、モデルの誤設定に強いなど、従来用いられていた共分

^{†1}近畿大学
Kindai University
^{†2}岡山県立大学
Okayama Prefectural University

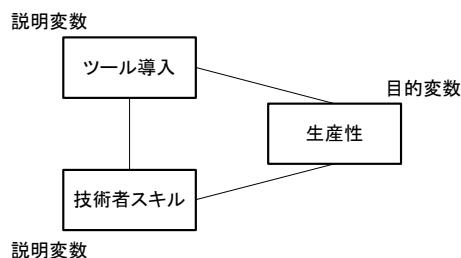


図 1 ソフトウェア開発データにおける交絡の例

Figure 1 An example of confounding on the software development dataset.

散分析などの手法よりも有利な点がいくつか指摘されている[2]。

傾向スコアを用いた分析は2つの手順からなる[2]。まず、他の説明変数を用いて、着目する説明変数の値を推定するモデルを作り、傾向スコアを算出する。例えば、ツールの利用有無が、生産性に与える影響を知りたいとする。この場合、ツールの利用有無を目的変数、技術者のスキル、開発ソフトウェアの複雑度などを説明変数とするモデルを作る。モデルの作成にはロジスティック回帰分析が用いられることが多い。傾向スコアとは、この例ではツール利用有となる確率となる。

次に、マッチングなどにより分析を行う。マッチングとは、傾向スコアが近く、かつ着目する説明変数の値（この例ではツールの利用有無）が異なるケース同士をペアにして分析することである。傾向スコアが類似している場合、スコアの算出に用いた説明変数、すなわち技術者のスキルや開発ソフトウェアの複雑度などが同程度であり、これらを無視することができる。ペア間の目的変数の差（この例では生産性）が、着目する説明変数（ツールの利用有無）の効果となる。

傾向スコアの計算対象となる変数は二値を取るダミー変数であるが、複数の値に対応する方法がいくつか提案されている。

2.2 定量的分析における活用

ソフトウェア工学のデータでは、説明変数間に関連があることが多い。例えば、金融業向けのソフトウェアはメインフレームで開発することが多い。プロジェクトの生産性に対する要因を明らかにしようとしても、従来法では金融業、メインフレームそれぞれの影響を個別に明らかにすることは難しい。傾向スコアを適用することにより、個別の影響を明らかにできることが期待される。ソフトウェア工学分野における傾向スコアの適用は、著者らの知る限り文献[5]などのごく一部であるが、データに偏りがあることが多いため、傾向スコアの適用が向いていると考えられる。傾向スコアを生産性分析[4]などに適用することにより、新たな知見が得られる可能性がある。

傾向スコアの一般的な使い方は、ある説明変数の効果を

分析することであるが、それ以外にインターネット調査のデータ補正[1]に用いられることがある。無作為標本抽出による調査と異なり、インターネット調査は偏りがある。その偏りを除去するために傾向スコアを利用している。まず、Web調査と無作為標本抽出による調査の両方からデータを収集する。このとき、一般的なアンケート項目とともに、性別や都市規模、学歴などを収集する。次に、「Web調査かそうでないか」という変数を作成し、性別や都市規模、学歴からその変数を予測するモデルを作成し、予測結果を傾向スコアとする。最後に、算出された傾向スコアを重みとして、インターネット調査のアンケート項目の解答の補正を行う。

我々の知る限り、ソフトウェア工学分野において、傾向スコアによるインターネット調査の補正を行った研究は存在しない。ソフトウェア工学においても、インターネット調査が活用されつつあるが[3]、同様のデータ補正を行うことにより、研究の信頼性がさらに高まる可能性がある。

3. おわりに

本稿では、ソフトウェア開発において収集されるデータの分析に適していると考えられる、傾向スコアについて説明し、適用可能な分析を示した。適用する際に問題となる点は、ソフトウェア開発に関するデータにおいて、傾向スコアを推定するモデルを適切に作ることができるかどうかであると考えられる。今後の課題として、傾向スコアによるデータ補正を応用し、工数見積もりモデルやバグモジュール予測モデルの精度向上に応用できないかを検討する。

謝辞 本研究の一部は、文部科学省科学研究補助費（基盤C：課題番号 25330090、萌芽研究：課題番号 26540029）による助成を受けた。

参考文献

- [1] 星野崇宏、前田忠彦：傾向スコアを用いた補正法の有意抽出による標本調査への応用と共変量の選択法の提案、統計数理、vol.54, no.1, pp.191-206 (2006).
- [2] 星野崇宏、岡田謙介：傾向スコアを用いた共変量調整による因果効果の推定と臨床医学・疫学・薬学・公衆衛生分野での応用について、保健医療科学、vol.55, no.3, pp.230-243 (2006).
- [3] 河村智行、高野研一、情報システム開発の成否に影響を与える組織文化の要因の研究：情報処理学会論文誌、vol.53, no.12, pp.2854-2864 (2012).
- [4] Maxwell, K., Wassenhove, L. and Dutta, S.: Software Development Productivity of European Space, Military, and Industrial Applications, *IEEE Transactions on Software Engineering*, vol. 22, no. 10, pp.706-718 (1996).
- [5] Ramasubbu, N. and Balan, R.: The impact of process choice in high maturity environments: An empirical analysis. In Proc. of International Conference on Software Engineering (ICSE), pp.529-539 (2009).
- [6] Rosenbaum, P. and Rubin, D.: The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, vol.70, no.1, pp.41-55 (1983).