

LiNGAM を用いた因果関係同定による工数データセットの分析

近藤 将成[†] 水野 修^{††}

[†] 京都工芸繊維大学 工芸科学部

^{††} 京都工芸繊維大学 情報工学・人間科学系

E-mail: [†]m-kondo@se.is.kit.ac.jp, ^{††}o-mizuno@kit.ac.jp

あらまし ソフトウェアの開発において、工数予測は重要な課題である。工数予測の手法としては、回帰分析などを用いてソフトウェアメトリクスからモデルを作成する研究が行われている。ここで、サンプル数の不足などから企業横断的 (Cross-Company) データセットが、モデル作成に用いられることがあるが、このデータセットが工数予測に対して良いデータセットなのか悪いデータセットなのかを判断することは難しい。本報告では、因果関係推定手法を用いた新しい工数データセットの評価指標を提案した。オープンデータを用いた実験より、我々の評価指標は正しく工数データセットの良さを評価できていることを確認した。

キーワード 因果関係, Linear Non-Gaussian Acyclic Model (LiNGAM), Bayes LiNGAM, 工数予測

Analysis on Effort Datasets by Causal-Effect Relationship Using LiNGAM

Masanari KONDO[†] and Osamu MIZUNO^{††}

[†] School of Science and Technology, Kyoto Institute of Technology

^{††} Faculty of Information and Human Sciences, Kyoto Institute of Technology

E-mail: [†]m-kondo@se.is.kit.ac.jp, ^{††}o-mizuno@kit.ac.jp

Abstract The effort estimation is an important task in the software development. Previous research works proposed models using machine learning techniques such as regression analysis and so on. The previous works sometimes used open cross-company datasets as training data because of the lack of sufficient number of samples. In such cases, the goodness of dataset affects the result of models. We thus have to know the goodness of datasets for effort estimation model. In this paper, we propose a new evaluation measure using causal-effect relationships. The empirical experiment shows that our the new evaluation measure could evaluate an accuracy of effort datasets.

Key words Causal-Effect Relationship, Linear Non-Gaussian Acyclic Model (LiNGAM), Bayes LiNGAM, Effort Estimation

1. はじめに

ソフトウェアの開発において、工数予測は重要な課題である。なぜなら、期日までにソフトウェアをリリースするために、設計、コーディング、およびデバッグ作業などのそれぞれの作業の日程を適切に定める必要があるからである。もし、工数を適切に見積もることができたならば、ソフトウェアのリリースが期日より遅れることを無くすることができる。また、適切な人員配置が可能になり、コスト削減や、作業員の負担軽減につながる [1]。

実際の工数予測では、多くのソフトウェアプロジェクトが、予定を超過しているという報告が寄せられている。例えば、カオスレポートでは、89%のプロジェクトがコストの面で、予定を超過していると報告されている [2]。また、他の文献でも概ね

30%から 40%程度の予定超過が見られるとの報告がある [3]。

この問題を解決するために、工数予測モデルが研究されている [4]。工数予測モデルでは、予測したいプロジェクトの工数メトリクスに対してモデルを適用し、工数を予測する。こうしたモデルには、教師有り学習の機械学習手法が用いられることがある。教師有り学習の適用のためには、学習データを集める必要がある。その場合、予測したい対象のプロジェクトを持つ企業の他のプロジェクトのデータを用いる方法が一番考えやすい。これを、企業内 (Within-Company) データセット (WC) と呼ぶことにする。工数予測モデルにおいては、1つの企業のデータを用いるべきであるという報告もある [5]。しかし、WCモデルには実用上、以下の3つの問題点があることが指摘されている [6], [7]。

(1) 1つの企業から十分な過去のデータの情報を集めるた

めには、非常に大きな時間がかかる可能性がある。

(2) 使用するに耐えうる十分なデータを集める間に、企業が用いる技術は変化し、また、古いプロジェクトが現在のプロジェクトを代表できる時間は長くはない可能性がある。

(3) 一貫したルールによってデータを集めることが必要であるため、ケアが必要である。

WCの問題を解決するために考えられたのが、企業横断的 (Cross-Company) データセット (CC) である。これは、複数の企業からそれぞれのプロジェクトの工数マトリクスを集めて工数データセットを作成する方法であり、WCと比較して多くのデータを集めることが可能になる。これにより、WCで問題であった、十分なデータを集められないことや、集められたとしても多くの時間が必要となるなどの問題を解決することができる。しかし、CCも問題点を持っている。例えば、データセットの不均一性 (マトリクスの測定手法に差がある) や、収集されるマトリクスの相違などである [8]。

WCおよびCCは、どちらがより良い工数データセットであるかについて、多くの研究がなされている [6], [7], [9]。しかし、これらの研究では、結論がまとまらず、どちらの工数データセットがより良いのかについては未だに答えが出ていない [10]。

本報告では、工数マトリクスを持った工数データセットにおいて、Linear Non-Gaussian Acyclic Model (LiNGAM) [11] を基礎となるモデルとして、さらにこれを発展させた Bayes LiNGAM [12] によるファンクションポイントマトリクス間における因果関係同定を用いた手法によって、複数の工数データセットから、どの工数データセットがより工数予測モデルに適用するのに適しているのかを判別する手法を提案する。この手法は、各工数データセットに対して得点を与えるため、WCおよびCCの比較に限らず、工数予測マトリクス (具体的には5つのファンクションポイント) をを持った工数データセットに適用可能である。本報告においてはオープンなデータセットを用いた実証実験を行い、手法の有効性を確かめた。

また、本報告における研究設問 (Research Question: RQ) として、以下を設定した。

(RQ) 工数予測モデルの精度を、工数データセットのマトリクス間の因果関係を用いて評価できるか?

本報告の構成は以下のようになっている。節2.において、統計的因果推論の概説および、研究の理論であるLiNGAMおよびBayes LiNGAMについて説明する。節3.では、実験についての説明と、その結果、およびRQの回答を与える。節5.で、本報告をまとめる。

2. 統計的因果推論

因果関係は、工学において非常に重要な関係である。なぜなら、説明変数と目的変数の間に因果関係があるため、種々の状況に対応できる一般的なモデルの構築が可能になるためである。

統計的因果推論では、因果関係をデータから推測する方法論

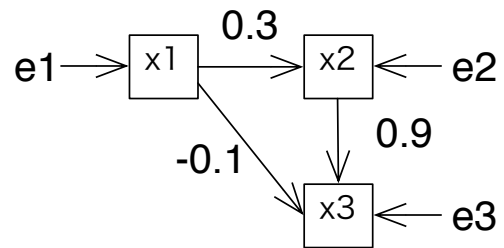


図1 DAGの例

のことである。フレームワークには、反実仮想モデルや構造法的式モデルなどがある [13]。

今回用いた手法は、構造法的式モデルに分類されるものである。図1に示すような有向非巡回グラフである DAG (Directed Acyclic Graph) を用いて、要因 (マトリクス) 間の関係を記述する。この時、その関係を線型性およびガウス性から、線型性および非ガウス性に変更することにより、因果関係を同定できるようにした [11]。

2.1 LiNGAM

LiNGAMはいくつかの仮定の元に成り立つ因果モデルである。仮定とは以下に示す3つである。

1. DAG G が、観測変数 $x_i (i = 1, 2, \dots, m)$ によって、表現可能である。
2. それぞれの観測変数 x_i は、DAGの上流側にある既に値が決まっている観測変数 x_j 、その係数 b_{ij} 、外的要因 e_i および定数項 c_i による、線形関数によって値が決まる。この線形関数は、以下の式で表される。

$$x_i = \sum_{k(j) < k(i)} b_{ij} x_j + e_i + c_i \quad (1)$$

ここで、 $k(\cdot)$ は、DAGにおける因果的順序を示す。つまり、 $k(j) < k(i)$ が示すのは、観測変数 x_i の親ノードをすべて用いることである。

3. 外的要因 e_i は、確率変数である。この確率変数は、分散が非ゼロの非ガウス分布に従う。また、 e_i は互いに独立である。したがって、 $p(e_1, \dots, e_n) = \prod_i p_i(e_i)$ を満たす。

この3つの仮定を満たす時に、我々は一意に因果関係を推定することが可能になる。よって、このモデルが、線形 (Linear)、非ガウス (Non-Gaussian)、非巡回 (Acyclic) を含んでいることから、Linear Non-Gaussian Acyclic Model (LiNGAM) と名付けられた [11]。

また、これらの仮定は、観測変数が因果関係を全て充足していることを示しており [14]、したがって、非観測データが存在しないことがわかる [15]。

2.2 Bayes LiNGAM

本報告では、LiNGAMモデルをベイズ推定によって拡張したモデルを、因果関係の同定に用いる。このモデルを Bayes LiNGAM と呼ぶ [12]。Bayes LiNGAMでは、可能性のあるすべての DAG に対して事後確率を計算し、観測データのみから因果関係を推定しようという手法である。この時に計算される事後確率は、ベイズの定理より、以下の式で計算される。

$$P(G_m|D) = \frac{p(D|G_m)P(G_m)}{p(D)} \quad (2)$$

ここで、 G_m ($m = 1, \dots, N$) は、存在する可能性のある N 個の DAG をそれぞれ表す。 N は存在する可能性のある DAG の数である。 S を観測変数の数とすると、 $D = \{x^1, \dots, x^S\}$ は、観測変数のデータセットを示している。 $P(D)$ は、正規化定数である。 $P(G_m)$ は、すべての DAG の事前分布である。 ここには、あらかじめ我々が持っている観測変数に関する情報を付加することができる。 もし、何も事前情報がなければ、一様な値を用いる。 最後に、尤度 $p(D|G_m)$ は、次の式によって更新される。

$$p(D|G_m) = \int p(D|\theta, G_m)p(\theta|G_m)d\theta \quad (3)$$

ここで、 θ は、すべてのパラメータを含む変数である。 つまり、係数 b_{ij} 、定数 c_i 、および外的要因 $p_i(e_i)$ を持つ変数である。 ここで、式 3 の近似を計算するために、積分、 $p(D|\theta, G_m)$ および $p(\theta|G_m)$ を計算する必要がある。 $p(D|\theta, G_m)$ は、LiNGAM の定義により次のように変換される。

$$p(D|\theta, G_m) = \prod_i p_i(x_i - \sum_{k(j)<k(i)} b_{ij}x_j - c_i). \quad (4)$$

$p_i(e_i)$ の密度のモデルを仮定した時に、式 4 は計算される。 $p_i(e_i)$ は、非ガウスの確率分布を取ることができる。 ここでは、元の論文で採用されていた 2 つの非ガウスの確率分布を紹介する。 ラプラス分布とガウス分布を組み合わせた分布と、混合ガウス分布である。 ラプラス分布とガウス分布を組み合わせた確率分布より、混合ガウス分布の方が、因果関係の推定において精度が良いことが示されている [12]。 よって、混合ガウス分布を採用した。 混合ガウス分布は以下の式で定義される。

$$p_i(e_i) = \sum_j^m \pi_j N(\mu_j, \sigma_j^2) \quad (5)$$

ここで、 m は、定められた混合されるガウス分布の数を示す。 また、 μ_j および σ_j^2 は、 j 番目の混合されたガウス分布の平均と分散を示す。 π_j は、 $\sum_j \pi_j = 1$ を満たすように定められる。 $p(\theta|G_m)$ は、 b_{ij} が平均 0、分散 1 のガウス分布にしたがっており、 c_i が 0 と仮定し計算される。 積分は、ラプラス近似を用いて近似される。

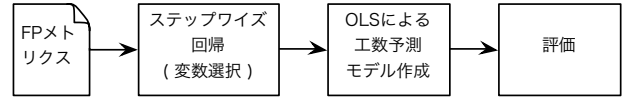
3. ケーススタディ

3.1 工数データセット

表 1 に示す、China データセット [16] および、Finnish データセット [17] の 2 つの工数データセットを実験に用いた。 工数データセットの選定基準は以下の 2 つである。

1. サンプル数が 400 以上あること。
2. IFPUG のファンクションポイント法で示されている 5 つのメトリクスを持っている。

1. は、Bayes LiNGAM で精度が良い分析を行うためには、サンプル数が 100 個以上あることが望ましい。 そのため、10-分割交差検証法を行うことも含めて、サンプル数が 400 以上とした。 2. は、工数との因果関係があると考えられているメトリクスであり、さらに頻繁に用いられているものを選択した。



FP: ファンクションポイント
OLS: 最小2乗回帰

図 2 事前実験における工数データセットの評価手順

3.2 事前実験

まず、今回用意した 2 つの工数データセットに対して、どちらの工数データセットがより優れているかを分析する。 実験手順を図 2 に示している。

- 1) 工数データセットからファンクションポイントメトリクスを取り出す。
- 2) 5 つのファンクションポイントから、予測に適するものをステップワイズ重回帰分析によって選択する。
- 3) 最小 2 乗回帰によってモデルを作成する。
- 4) MMRE などの各種評価指標を計算する。

実験では 10 回 10-分割交差検証法を行う。 そのため、手順 1. の工数データセットとは、学習データのことを指す。 また、手順 4. の評価時に用いられるデータはテストデータである。 工数予測モデルを作成するために用いたメトリクスは、提案手法と同じファンクションポイントである。 これについては、提案手法の節で説明する。

用いた評価指標は、絶対誤差平均値 (Mean Magnitude Related Error: MMRE)、絶対誤差中央値 (Median Magnitude Related Error: MdMRE)、Pred(25)、および標準偏差である。 Magnitude Related Error (MRE) とは、式 6 で表される指標である。

$$MRE = \frac{|E^a - E^e|}{E^a} \quad (6)$$

ここで、 E^a は実工数 (actual effort)、 E^e は見積もり工数 (estimated effort) を示す。 これにより、予測された工数と実際の工数の差から、工数予測モデルの性能を評価することができる。 MRE が大きい場合、予測された工数と実際の工数との間に隔たりがあることを示していることになる。 そのため、この予測モデルは良くないことがわかる。 小さい場合、予測された工数と実際の工数との間の差が小さいことを示すことになる。 そのため、この予測モデルは良いことがわかる。 MMRE とは、MRE の平均を取ったものであり、式 7 で表される。

$$MMRE = \frac{1}{N} \sum_{i=1}^N \frac{|E^a - E_i^e|}{E^a} \quad (7)$$

MdMRE とは、MMRE のように平均を取る代わりに、中央値を用いる方法である。 Pred(x) とは、MRE の値が x 以下のサンプルの割合を示す指標である。 MRE では、小さな値がより優れているため、Pred(x) では、値が大きい方が望ましいことがわかる。 これらの指標は、モデルの良さを計測するための指標である。 しかし、我々は、これらの指標を工数データセットの良さを計測するために用いた。 同じモデルを用いた場合、工数データセットの差が計測可能だからである。

最後に、サンプルごとの MRE の標準偏差を計算した。 計算

表1 分析対象とした工数データセット

データセット名	サンプル数	マトリクス数	使用されたマトリクス名
China	499	19	Interface, Output, Enquiry, Input, File, Effort
Finnish	407	46	IntFP, OutFP, InqFP, InpFP, EntFP, Worksup

表2 ステップワイズ重回帰分析による工数データセットの比較

	China	Finnish
MMRE	1.101	0.926
MdMRE	0.589	0.545
Pred(25)	0.205	0.251
標準偏差	2.033	1.402

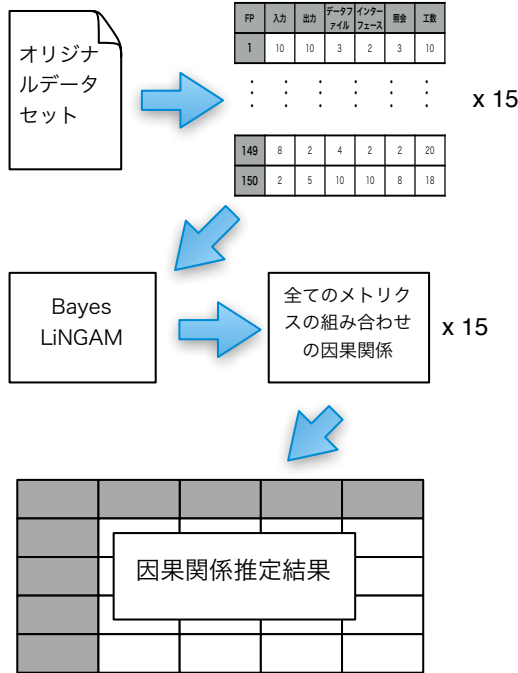


図3 提案手法の因果抽出手順

式は式8で表される。

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^N (E_i^e - \bar{E}^e)^2} \quad (8)$$

ここで、 \bar{E}^e は、予測工数の平均を表す。今回は、標準偏差を用いて、工数データセットのサンプル間のばらつき具合を計測した。サンプル間のばらつきが大きければ大きいほど、予測が難しい工数データセットであることがわかる。よって、工数データセットの良さを判断できる。

分析の結果を表2に示す。全ての指標において、Finnish データセットの方が優れていた。よって、工数予測において、China データセットよりも Finnish データセットの方が優れている工数データセットであることがわかった。

3.3 提案手法

提案する評価指標の計算方法を以下に示す。

1) それぞれの工数データセット (China および Finnish) から 150 個のサンプルを無作為抽出することを 15 回行い、それぞれの工数データセットにつき 15 個のデータセットを作る。

2) それぞれの工数データセットにおいて、15 個のデータセットそれぞれで、全ての組み合わせのマトリクス間の因果関係の事後確率が Bayes LiNGAM により計算される。

3) ここまでで、それぞれのマトリクスペア (m_a, m_b とする) に存在する可能性のある 3 通りの因果関係 ($m_a \rightarrow m_b, m_a \leftarrow m_b, \text{none}$) の事後確率がそれぞれ 15 個計算されている。それぞれのマトリクスペアで抽出された因果関係の数を $r_{\leftarrow}^{a,b}, r_{\rightarrow}^{a,b}, r_{\text{none}}^{a,b}$ とする。これらのうち、最大のものを $R_1^{a,b}$ 、二番目に大きいものを $R_2^{a,b}$ とする。

4) $R_1^{a,b}$ および $R_2^{a,b}$ 間の差を計算する。具体的には、式9によって計算される。

$$D^{a,b} = R_1^{a,b} - R_2^{a,b} \quad (9)$$

この値をそれぞれの工数データセットの評価に用いる新しい評価指標とする。

それぞれの工数データセットにおけるマトリクス間の因果関係の抽出までの手順を図3に示す。最終的に得られる因果関係推定結果の例を表4, 表5に示す。表4について詳しく述べる。この表は $R_1^{a,b}$ と $R_2^{a,b}$ の因果関係を示している。各セルの左側が $R_1^{a,b}$ 、右側が $R_2^{a,b}$ を示している。また、右矢印は、列のマトリクス (m_a) から行のマトリクス (m_b) への因果関係 ($m_a \rightarrow m_b$) を示しており、左矢印はその逆 ($m_a \leftarrow m_b$) である。None は因果関係が無い場合である。丸括弧の数字は、作成した 15 個のデータセットの内、幾つのデータセットでその因果が最大の事後確率を示したかを表している。例えば、IntFP および OutFP 間の因果関係では、因果関係が存在しないという事後確率が最大であったデータセットの数が 15 個中 9 個あり ($r_{\text{none}}^{\text{IntFP,OutFP}} = 9$)、これが $R_1^{\text{IntFP,OutFP}}$ となる。また、IntFP から OutFP への因果関係を示したデータセットの数が 15 個中 3 個あり ($r_{\rightarrow}^{\text{IntFP,OutFP}} = 3$)、これが $R_2^{\text{IntFP,OutFP}}$ となる。最後に、OutFP から IntFP への因果関係を示したデータセットの数が 15 個中 2 個あり ($r_{\leftarrow}^{\text{IntFP,OutFP}} = 2$)、ここには掲載されていない。残りの 1 つのデータセットでは、因果関係の抽出に失敗して計算ができていない。このように因果関係を抽出し、 $R_1^{\text{IntFP,OutFP}}$ と $R_2^{\text{IntFP,OutFP}}$ を求め、その差 $D^{\text{IntFP,OutFP}}$ を計算する。

もし、 D^{*} が大きければ、オリジナルの工数データセットから無作為にサンプルを取り出し構成された 15 個のデータセットそれぞれで同じような因果関係が抽出できたことを示す。つまり、その工数データセットにおけるマトリクス間の因果関係がより強固であると考えられる。逆に、 D^{*} が小さければ、因果関係を同定することが難しいため、マトリクス間の因果関係が弱いと考えることができる。因果関係が強いマトリクス間では、予測などがより容易にできる可能性があるため、本報告では、この D^{*} を工数データセットの評価に用いた。

分析するメトリクスには、IFPUG (International Function Point Users Group) のファンクションポイント法で示されている5つのメトリクスおよび工数を用いた、IFPUGで示されている5つのファンクションポイントとは、入力 (Inputs), 出力 (Outputs), データファイル (Data Files), インターフェース (Interfaces), および照会 (Inquiries) である。

3.4 RQ: 工数予測モデルの精度を、工数データセットのメトリクス間の因果関係を用いて評価できるか?

3.4.1 動機

因果関係を考慮した指標は、介入効果を期待できるため、予測という点でメリットがあると考えられる。そのため、因果関係を用いた指標で工数データセットを評価できれば、工数データセットの選択において優れた指標となるのではないかと考えた。以上より、提案手法で定義した因果関係を考慮した difference, D^{**} という指標を用いて、工数データセットの良さを評価し、 D^{**} の有効性を確かめた。介入効果とは、因果関係において、原因を変化させるとそれに伴って結果が変化するという効果である [13]。この時、結果を変化させても、原因が変化しない可能性がある。しかし、通常の相関ではどちらが原因で、どちらが結果であるかわからないため、説明変数に結果のメトリクスを使用してしまう可能性がある。一方で、因果関係においては、原因と結果がはっきりしているため、このような心配がいらぬ。

3.4.2 アプローチ

実験では、それぞれの工数データセットで全メトリクスペアの D^{**} を計算し、検定を行うことで、有意差があるかないかを確かめた。その後、 D^{**} の平均値と中央値を確かめることで、どちらが因果関係という観点から有意に優れているのかを確かめた。正規性を仮定できなかったため、検定に用いたのは、ノンパラメトリックなウィルコクソンの符号順位和検定と順位和検定である。符号順位和検定で、個別のメトリクスペアに焦点を当てて差を検定し、また、順位和検定で工数データセット全体に焦点を当てて検定した。実験の際は、Bayes LiNGAMのパラメータである混合数を変化させながら行なった。混合数とは、Bayes LiNGAMの外的要因 e に採用した混合ガウス分布の混合数を指す。式5における m に当たる。混合数を増やすと、より複雑な確率分布を表現することができる。よって、より個別の工数データセットに合わせた因果関係の抽出が可能となる。

3.4.3 結果

表3に混合数を変化させた時の、それぞれの工数データセットにおけるウィルコクソンの符号順位和検定および順位和検定の p 値が示されている。有意水準は0.05とした。混合数が5の時に、それぞれの検定で有意差が得られた。よって、混合数が5の時に、ChinaデータセットおよびFinnishデータセットの間および、それぞれのメトリクスペア間には因果関係を考慮した我々の指標の視点から有意差が存在することが確かめられた。

表4と表5は、それぞれのデータセットにおいて混合数が5の時に得られた因果関係の推定結果から上位2つずつを表したものである。ここから difference, D^{**} を計算した結果を表6に示す。ここでは、それぞれの工数データセットにおける D^{**} の

表3 ウィルコクソンの符号順位和検定および順位和検定の p 値

混合数	符号順位和検定	順位和検定
2	0.889	0.892
3	0.0910	0.0523
4	0.0615	0.0490*
5	0.0281*	0.00576*

*は、有意水準 $\alpha = 0.05$ を満たすことを意味する。

値の平均値および中央値を示している。これより、混合数が4および5の時には、Finnishデータセットの方が D^{**} の値が大きいことがわかる。表4と表5を比較してみても、Finnishデータセットでは明確に因果関係が推定できていることが確認できる。我々の指標 D^{**} は、大きい方がより良い工数データセットであると定義した。この結果より、FinnishデータセットはChinaデータセットよりも因果関係の点から優れていると結論づけられる。

また、この結果は、実際に工数予測モデルを作成し、その予測精度を確かめた時の結果と一致する。これより、我々の指標 D^{**} は、工数データセットの良さを測定することができることがわかった。

4. 妥当性への脅威

この節では、本報告の妥当性への脅威について、構成概念妥当性、内的妥当性、外的妥当性、および信頼性という観点から分析する。

4.1 構成概念妥当性

今回は、工数メトリクスとして、ファンクションポイントを用いた。ファンクションポイントは、その測定基準が定められており、また、工数を予測する上で多くの場合に用いられる指標である。また、多くの工数データセットが採用しているメトリクスである。そのため、尺度として採用したことは適切であったと考えている。

4.2 内的妥当性

本報告では、ファンクションポイントおよび工数間の因果関係を仮定し、その因果関係を分析することを行なった。工数は様々な要因を持つ結果であり、他のメトリクスからの因果関係が無いとは言い難い。そのため、他の要因からの因果関係を完全に排除して実験を行えているとは言い難い。この妥当性は、今後の研究で検討していく必要が非常に高いと考えている。

4.3 外的妥当性

今回の実験は、2つの工数データセットに応用されている。これは、非常に小さなエンピリカルな実験であり、今後さらに多くの事例に対して応用していく必要があると考えている。

4.4 信頼性

Bayes LiNGAMは、誰でも簡単に使用できるように一般公開されている。また、実験に用いたデータセットも自由に利用可能なものである。そのため、本報告に示した手順で実験を進めることで、実験を再現することは容易であると考えている。

表4 Finnish データセット (混合数: 5) における因果関係推定結果 (上位 2 つ)

	OutFP	InqFP	InpFP	EntFP	Worksup
IntFP	None(9), →(3)	None(10), →(2)	→(10), ←(2)	None(9), →(3)	→(10), None(2)
OutFP		→(12), ←(2)	→(15),	→(13), ←(2)	→(9), ←(4)
InqFP			→(9), ←(6)	→(12), ←(3)	←(9), →(6)
InpFP				→(9), ←(6)	→(11), ←(4)
EntFP					→(12), ←(3)

表5 China データセット (混合数: 5) における因果関係推定結果 (上位 2 つ)

	Output	Enquily	Input	File	Effort
Interface	←(3), None(3)	None(6), ←(2)	None(4), ←(3)	→(3), ←(3)	→(3), ←(3)
Output		None(10), ←(3)	←(8), →(6)	←(7), →(5)	→(9), ←(4)
Enquily			→(13), ←(1)	→(11), None(2)	→(9), ←(4)
Input				←(9), →(5)	←(8), →(6)
File					→(9), ←(5)

表6 R_1^{**} および R_2^{**} 間の差 D^{**} の平均値と中央値

混合数	China		Finnish		中央値
	平均値	中央値	平均値	中央値	
4	4.93	3	7.80	8	8
5	4.07	4	7.40	8	8

5. 結 論

本報告では、因果関係を考慮した新しい工数データセットの評価指標を提案した。エンピリカルな実験より、我々の評価指標は正しく工数データセットの良さを測定できることがわかった。

本報告のソフトウェア工学に対する寄与は、以下に示す点である。

- **因果関係をソフトウェア工学の工数予測に応用する 1 つの方法を示した。** 因果関係を用いた工数予測は、今まで行われてきてはなかった。本報告では、LiNGAM を応用した Bayes LiNGAM を用いて工数メトリクスおよび工数間の因果関係を同定し、それを評価する方法を初めて示した。

6. 謝 辞

本研究を遂行し本報告を執筆するにあたって貴重な助言を頂きました九州大学の亀井靖高先生、Queen's University の Prof. Ahmed Hassan, Dr. Cor-Paul Bezemer に深く感謝します。また、本研究の一部は日本学術振興会科研費 15K00096 の補助を受けています。

文 献

[1] 野中誠, 鷺崎弘宜. 演習で学ぶソフトウェアメトリクスの基礎, 2009.

[2] Standish Group. The chaos report (1994), 1994.

[3] Kjetil Molokken and Magen Jorgensen. A review of software surveys on software effort estimation. In *Empirical Software Engineering, 2003. ISESE 2003. Proceedings. 2003 International Symposium on*, pp. 223–230. IEEE, 2003.

[4] Jianfeng Wen, Shixian Li, Zhiyong Lin, Yong Hu, and Changqin

Huang. Systematic literature review of machine learning based software development effort estimation models. *Information and Software Technology*, Vol. 54, No. 1, pp. 41–59, 2012.

- [5] PAM Kok, BA Kitchenham, and J Kirawkowski. The mermaid approach to software cost estimation. *ESPRIT' 90*, pp. 296–314. Springer, 1990.
- [6] Lionel C. Briand, Khaled El Emam, Dagmar Surmann, Isabella Wiecek, and Katrina D. Maxwell. An assessment and comparison of common software cost estimation modeling techniques. In *Proceedings of the 21st International Conference on Software Engineering, ICSE '99*, pp. 313–322, New York, NY, USA, 1999. ACM.
- [7] E. Mendes, C. Lokan, R. Harrison, and C. Triggs. A replicated comparison of cross-company and within-company effort estimation models using the isbgs database. In *11th IEEE International Software Metrics Symposium (METRICS'05)*, pp. 10 pp.–36, Sept 2005.
- [8] Jia Chen, Ye Yang, Wen Zhang, and Gregory Gay. Measuring the heterogeneity of cross-company dataset. In *Proceedings of the 11th International Conference on Product Focused Software*, pp. 55–58. ACM, 2010.
- [9] R Jeffery, M Ruhe, and I Wiecek. A comparative study of two software development cost modeling techniques using multi-organizational and company-specific data. *Information and Software Technology*, Vol. 42, No. 14, pp. 1009 – 1016, 2000.
- [10] B. A. Kitchenham, E. Mendes, and G. H. Travassos. Cross versus within-company cost estimation studies: A systematic review. *IEEE Transactions on Software Engineering*, Vol. 33, No. 5, pp. 316–329, May 2007.
- [11] Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, and Antti Kermiinen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, Vol. 7, No. Oct, pp. 2003–2030, 2006.
- [12] Patrik O. Hoyer and Antti Hyttinen. Bayesian discovery of linear acyclic causal models. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI '09*, pp. 240–248, Arlington, Virginia, United States, 2009. AUAI Press.
- [13] 宮川雅巳. 統計的因果推論 — 回帰分析の新しい枠組み —. 朝倉書店, 2004.
- [14] Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.
- [15] Judea Pearl. Causality: models, reasoning, and inference. *IIE Transactions*, Vol. 34, No. 6, pp. 583–589, 2002.
- [16] The promise repository of empirical software engineering data, 2015.
- [17] Boyce Sigweni, Martin Shepperd, and Pekka Forselius. Finnish Software Effort Dataset. 3 2015.