

集約的手法による Error Prone モジュール予測における 最適モデル集合の選択

高橋利英[†] 小形 真平[†] 海谷 治彦[†] 海尻 賢二[†]

信州大学大学院理工学系研究科情報工学専攻[†]

1. はじめに

Error-prone モジュールとは未検出の誤りの起こる可能性の高いモジュールの事である。テストデータによるテストだけでは誤りの発見には十分ではないので人手による inspection 等が行われるが、網羅的に行う事は効率的ではない。そこで Error-prone モジュールと予測されたリスクの高いモジュールを重点的に調べる事で、効率的に保守作業を行う事ができる。このような考えから Error-Prone モジュール予測には多くの手法が提案されているが、最適な手法というものは見つかっていない。

予測器の決定には種々の要因（選択肢）があり、それによって予測の精度も異なってくる。また汎用的に最適な予測モデルは存在しないことが (Lessmann[1]) でいわれている。そこで、いくつかのアルゴリズムおよび訓練データに基づいて個々に作成した予測器の評価を集約することで、相互の欠点を補った平均的に優れた評価を行う予測器が作れるのではないかとという仮説に基づき様々な集約的手法が考えられている (Amasaki [2], Kamei [3])。集約的に予測器を使う手法はアルゴリズムおよび訓練データによる特異性があるという事実に基づいている。

我々は、これまでに予測器、予測モデルの再利用という立場から、訓練データおよびアルゴリズムの集約による効果の検証を行った[4]。アルゴリズムの集約手法は、複数の異なるアルゴリズムの予測結果を集約する事で、予測モデルのアルゴリズムが原因で予測精度が低くなる様な事態を防ぐことができるのではないかとこの考えに基づく。訓練データの集約は、複数の異なるプロジェクトを訓練データとした予測結果を集約する事で、訓練データの選択が原因で予測精度が低くなるというような事態を防げるのではないかとこの考えに基づく。

2. 研究の進捗状況

アルゴリズムの集約において、これまでの実験ではアルゴリズム集合を WEKA で利用可能なアル

ゴリズムの種類毎に代表を選出して構成した。これは、アルゴリズムの種類を網羅することで、複数の予測器が各々に特異性を持つのではないかと考えた為である。このようにアルゴリズム集合を構成する方法では、アルゴリズム集約の有効性ある程度は確認することができたが、本質的な改善にはならなかった。この原因は、一部の予測器には特異性がなく、アルゴリズムを組み合わせることで互いに補い合うような集約ができなかった為と考えている。そこで、アルゴリズム集合を互いに補い合うように構成する事で、アルゴリズムの集約手法をより有用な手法にできないかと考えた。

本稿では、より適切なアルゴリズム集合を構成するために、アルゴリズム対の評価方法による考察を行った。

3. アルゴリズム対の評価方法

あるアルゴリズムに基づく予測器に対して、別のアルゴリズムに基づく予測器をどのように選ぶべきかという観点でアルゴリズム対の評価方法を考察する。ベースアルゴリズムに基づく予測器の結果に対して、追加アルゴリズムに基づく予測器の結果がどのように作用するかを分析するために、図1のような定義を行った。

本研究では、複数の予測結果から最終的な分類を行うための集約手法として Voting を採用している。Voting とは、予測値の投票によって、複数の予測結果から最終的な分類を行う集約手法であり、予測時の目的によってユーザ指定の閾値は異なってくる。なお、本研究で採用している Voting の閾値は予測値 TRUE の数としている。

アルゴリズム対では Voting の閾値は、TRUE=1(OR)か TRUE=2(AND)を指定できる。OR は、より多くのバグありモジュールを見つけることを目的としており、Recall は大きくなるが、Precision は小さくなる傾向にある。AND は、より効率の良い保守作業を目的としており、Recall は小さくなるが、Precision は大きくなる傾向にある。

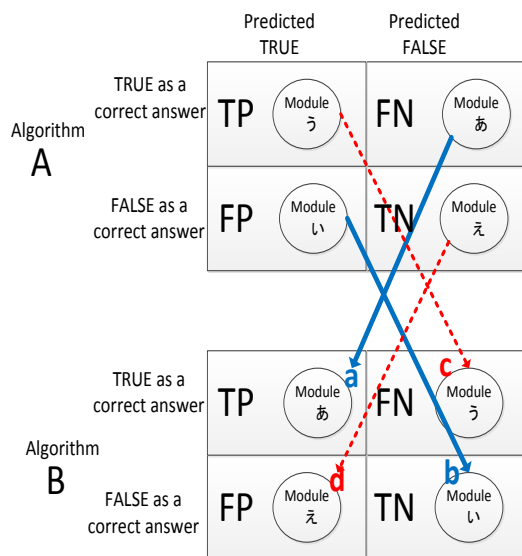
閾値が1の場合はベースアルゴリズムに対して、アルゴリズム対の評価方法の a と d のみ最

Selection of the best model in the aggregation based error prone prediction approach

[†] Toshihide Takahashi, Shinpei Ogata, Haruhiko Kaiya, Kenji Kaijiri (Shinshu University)

最終的な分類に寄与しており、a だけ TP 増加、d だけ FP 増加となるので a が大きく、d が小さい追加アルゴリズムがより優れているといえる。

閾値が 2 の場合はベースアルゴリズムに対して、アルゴリズム対の評価方法の b と c のみ最終的な分類に寄与しており、b だけ FP 減少、c だけ TP 減少となるので、b が大きく、c が小さい追加アルゴリズムがより優れているといえる。



- a: ベースアルゴリズムでFN、同モジュールが追加アルゴリズムでTPとなった数
- b: ベースアルゴリズムではFP、同モジュールが追加アルゴリズムでTNとなった数
- c: ベースアルゴリズムではTP、同モジュールが追加アルゴリズムでFNとなった数
- d: ベースアルゴリズムではTN、同モジュールが追加アルゴリズムでFPとなった数

図 1. アルゴリズム対を評価するための定義

4. アルゴリズム対の考察

独自に収集した OSS の 17 プロジェクトに対して、WEKA で適用可能な全アルゴリズムを対象としたアルゴリズム対の評価方法によってアルゴリズム間の補い合いの傾向を調べる実験を行った。

アルゴリズムの種類が同じであっても互いに補い合う組み合わせとなるペアが存在し、逆にアルゴリズムの種類が異なっても全く同じ分類となるケースを確認した。この結果は、アルゴリズムの種類が異なっても、各予測器が特異性を持つとは限らないことを示し、アルゴリズムの種類に基づくだけでは互いに補い合うような組み合わせが構成できないことを示している。

また、予測の目的に応じた予測精度向上に寄与する可能性の高いアルゴリズムは、ほとんどのプロジェクトで同じような傾向があることを確認できた。この結果から、予測の目的に応じてアルゴリズム集合に含ませるべきアルゴリズム

にはある程度の一般性があるのではないかと考えられる。

さらに、評価値 Recall の値が大きいアルゴリズムに対して、さらに精度を向上できる追加アルゴリズムが存在しないケースがあった。この結果は、アルゴリズムの組み合わせだけでは Recall 値の向上に限界があり、見つけることができないバグありモジュールが存在することを示した。

5. おわりに

本稿では、アルゴリズムの集約において、適切なアルゴリズム集合を構成するために、アルゴリズム対の評価方法によって、互いに補い合うペアが存在するのかを考察した。アルゴリズム対の評価によって、補い合いの度合いや分類特性の違い、類似性や優劣を分析することができた。また、Voting の閾値によって複数の予測結果がどのように寄与するかを考えた。

アルゴリズム対の評価は、あくまでも傾向を探る実験であり、実データに対して行うプロセスではない。アルゴリズム対の評価によって得られた、アルゴリズム対の情報に基づいて、3 つ以上のアルゴリズム集合へ適用する手法の考察および適用実験を行う。

今後は、3 つ以上のアルゴリズム集合の実験において、プロジェクトによらない一般に最適なアルゴリズム集合が存在するのか調査をし、その結果に基づいて実データに対して適切なアルゴリズム集合の選択方法を考えていく。

参考文献

[1] Stefan Lessmann, et al: Benchmarking Classification Models for Software Defect Prediction: A Proposed Framework and Novel Findings, IEEE Tr. on S.E., 34, 4 2008 pp. 485-406.
 [2] 天寄 聡介: fault-prone モジュール判別手法の予測結果の統合方法について、情報処理学会論文誌 52, 2 2011.
 [3] 亀井 靖高、森崎 修司、門田 暁人、松本 健一: 相関ルール分析とロジスティック回帰分析を組み合わせた fault-prone モジュール判別手法、情報処理学会論文誌 49, 12 2008 pp. 3954-3966.
 [4] 高橋 利英、大西 達也、小形 真平、海谷 治彦、海尻 賢二: 多数決手法による Error Prone 予測手法の評価、電子情報通信学会 KBSE 研究会 KBSE2012-7 2012