

検証パターンに注目した機械学習に基づくモデル検査手法の評価

張 超群 岸 知二

早稲田大学

1. 研究の背景と目的

モデル検査 (Model Checking) は形式手法のひとつであり、プログラムや設計の正しさを数理論理学に基づき全数検査をすることによって検証するために用いられる技術である。モデル検査はプロパティが満たされない場合には反例が出力される。検証するシステムのプロパティを記述するためによく LTL [1] が用いられる。モデル検査の利用における課題のひとつに状態爆発問題がある。Zhu ら [2] は、機械学習とモデル検査を組み合わせた手法を提案している。複数の Kripke 構造と LTL 式で記述されたプロパティに対してモデル検査を行い、結果でラベル付けして学習モデルを作成し、未知の Kripke 構造とプロパティのモデル検査の結果を推定する。手法の全体像を図 1 に示す。機械学習の分類アルゴリズムを用いることで LTL モデル検証の結果を予測し、状態爆発の問題を回避しつつ一定精度の推定結果が得られることを示している。

Zhu らの手法はランダムに生成した Kripke 構造と LTL 式を用いている。現実のモデル検査で用いられるプロパティには一定の類型があると考えられる。そこで学習や推定に用いるプロパティをそうした類型にあてはまるものに限することで、予測精度がどのように変わるかを調べる。また、先行研究では用いられていないアルゴリズム、ニューラルネットワークも用いた。

2. パターンの影響の評価

2.1. 概要

本研究では、機械学習とモデル検査を組み合わせた手法に基づいて、システムモデルを構築データセットとすることにより、機械学習モデルを訓練し、Dwyer の時相仕様のパターン [3] から 3 つを選び、予測精度の比較を行う。

2.2. 実験設定

2.2.1. システムモデルの準備

本実験では異なる 24 個のシステムモデルを対象とする。システムモデルは Kripke 構造で、そ

Evaluation of machine learning based model checking methods focusing on validation patterns

†ZHANG CHAOQUN ·Waseda University

‡KISHI TOMOJI ·Waseda University

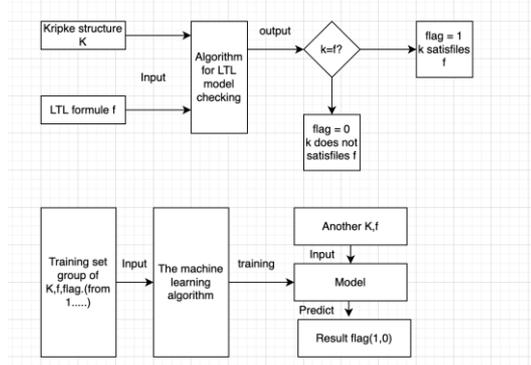


図 1.zhu らの手法の全体像[2]

れぞれ 3 つの状態を持ち、3 つの原子命題を含む。

2.2.2. プロパティの準備

本実験では、Dwyer [3] によって提案されたプロパティパターンから標準的な LTL 式のパターンである Absence, Universality, Existence の 3 つのパターンを選択した。それぞれ意味は区間の中、「ずっとない」、「ずっとある」、「ある」を表す。各パターンには 5 つの LTL 式が含まれているので、15 の LTL 式を実験に用いた。

2.2.3. モデル検査の実施

15 の LTL 式を使ってモデル検査を行い、結果を得た。なお各 LTL 式は 3 つの命題変数を持ち、システムモデルは 3 つの原始命題を持つため、当てはめ方は 6 通りあるが、全部を用いた。

2.2.4. データ構築

Kripke 構造を先行研究と同様の方法で数字列としてデータ化する。図 2 は本実験で用いた一つ Kripke 構造のひとつである。ここで状態中の数字は 3 つの原始命題の真偽を表している。データ化においては、各状態での原始命題の真偽を列挙し (s0:011, s1:110, s2:110) 次に遷移関係 (s0s1:01, s1s2:12, s2s0:20) を列挙する。従ってこの Kripke 構造は 011110110011220 とという数字列になる。プロパティのデータ化の方法とモデル検査の結果のデータ化も先行研究と同様であり、LTL 式に Kripke 構造中の原始命題をあてはめた文字列をデータとした。性質の成立・不成立を 1 と 0 で表した。以上より、2160 個のデータを作成した。

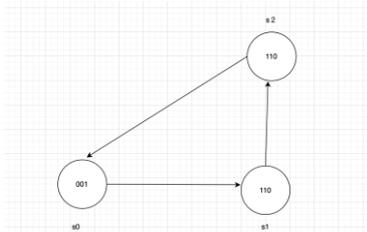


図 2.Kripke 構造の例

2.2.5.データセット構築

2.2.4 で作られたデータはプロパティパターン Absence, University, Existence それぞれに 720 個ずつとなる。7 種類のデータセットを作る。

Absence, University, Existence: 名に示すようにパターン。ALL: すべてのデータ。AbandUn, UnandEx, AbandEx: 頭文字に対応した二つのものだけからなるデータ。上記を用いて、各実験のためのデータセットを以下のように作る。

●実験 A (全体パターン): ALL をシャフルした後、720 個データを抽出して、その中で 80% を訓練セット、残り 20% をテストセットとする。

●実験 B (2 つのパターン学習と予測が異なる): 2 つのプロパティパターンに対応したシャフルした後、576 個データを抽出して訓練セットとし、別のプロパティパターンからランダムに 144 個データを抽出してテストセットとした。3 種類の訓練・テストセットのペアが作られる。

●実験 C (2 つのパターン学習と予測同じ): 2 つのプロパティパターンに対応したデータセットから 720 個データを抽出して、その中で 80% を訓練セット、残り 20% をテストセットとする。3 種類の訓練・テストセットのペアが作られる。

●実験 D (1 つのパターン): シャフルした後、720 個データを抽出して、その中で 80% を訓練セット、残り 20% をテストセットとする。

2.2.6.モデル作成と予測

訓練セットを用いてモデルを作成し、テストセットで予測を行う。Zhu らが用いたアルゴリズムを用いた。それ以外にもニューラルネットワーク (NN) を使っている。各実験は 5 回ずつ行い精度の平均値をとって、実験に用いたデータセットを表 3 に示す。

3.実験結果と考察

3.1 実験結果

実験結果を表 3 に示す。実験には 5 回の平均値をとって、精度は四捨五入して小数点以下 2 位までを示す。

4.評価と考察

4.1 評価

実験結果より、同じパターンを用いることで精

表 3. DT による実験結果

	データセット	DT 精度	LR 精度	NN 精度
実験 A	ALL	0.69	0.68	0.70
実験 B	AbandEx+	0.66	0.67	0.67
	Universality			
	AbandUn+	0.66	0.66	0.70
	Existence			
	ExandUn+	0.68	0.66	0.70
実験 C	AbandEx	0.69	0.69	0.73
	AbandUn	0.69	0.68	0.71
	ExandUn	0.67	0.70	0.74
実験 D	Absence	0.66	0.67	0.68
	Existence	0.65	0.66	0.70
	Universality	0.67	0.68	0.69

度が著しく向上することはなく。総合的には実験 C の方が実験 B よりも少なく良い結果となっており、これは同じ状態の同じモデルに基づくデータであれば、より精度が高いことを表している。実験 A は良好な精度を維持し、これは、ランダムサンプリングの偶然性と関係している可能性がある。実験 D について現時点では詳細な検討はしていない。複数のアルゴリズム結果を見ると、分類アルゴリズムより NN の精度が高い。これは、分類アルゴリズムがシステムモデル構造と LTL 式をうまく認識できていないことが原因だと思われる。分類アルゴリズムに特に大きな差はなく、DT アルゴリズムの全体的な精度は LR ほど安定しておらず、これは DT アルゴリズムが複数のフィーチャーを持つデータを分類した結果に偏っているためとも考えられる。

4.2 考察

今は初期的な評価として、実験の回数比較的に少なくなつて、実験過程で実験が不安定であり、得られた結果の信頼性が限られていることが今実験の欠点である。本手法の有用性には現時点では未知数であるが、今後より信頼性、正確性の高い評価を進めたい。

参考文献

[1]Pnueli A. The temporal logic of programs[C]//18th Annual Symposium on Foundations of Computer Science (sfcs 1977). IEEE, 1977: 46-57.
 [2]Zhu, Weijun, Huanmei Wu, and Miaolei Deng. "LTL model checking based on binary classification of machine learning." IEEE Access 7 (2019):135703-135719.
 [3]<https://matthewbdwyer.github.io/psp/patterns/ltl.html>