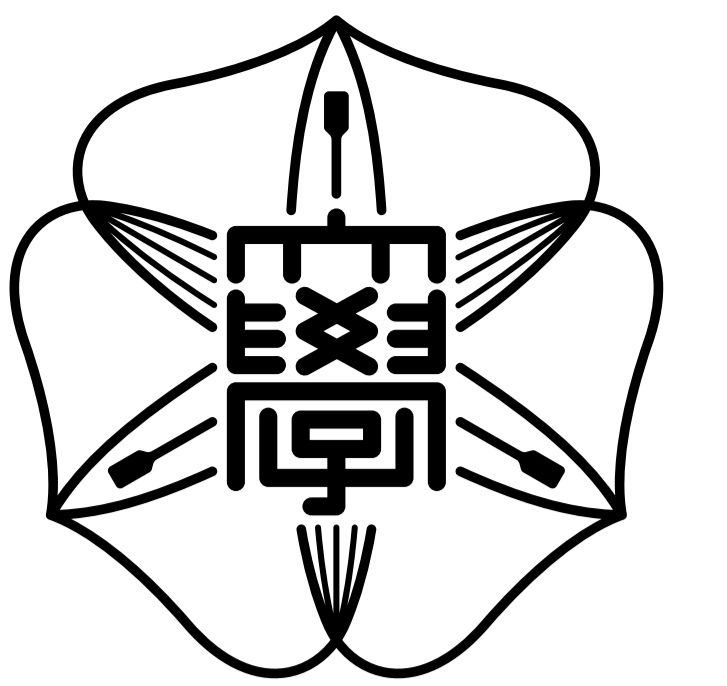




一階述語論理のプロパティ検査

Charles Jordan

JST ERATO 湊離散構造処理系プロジェクト



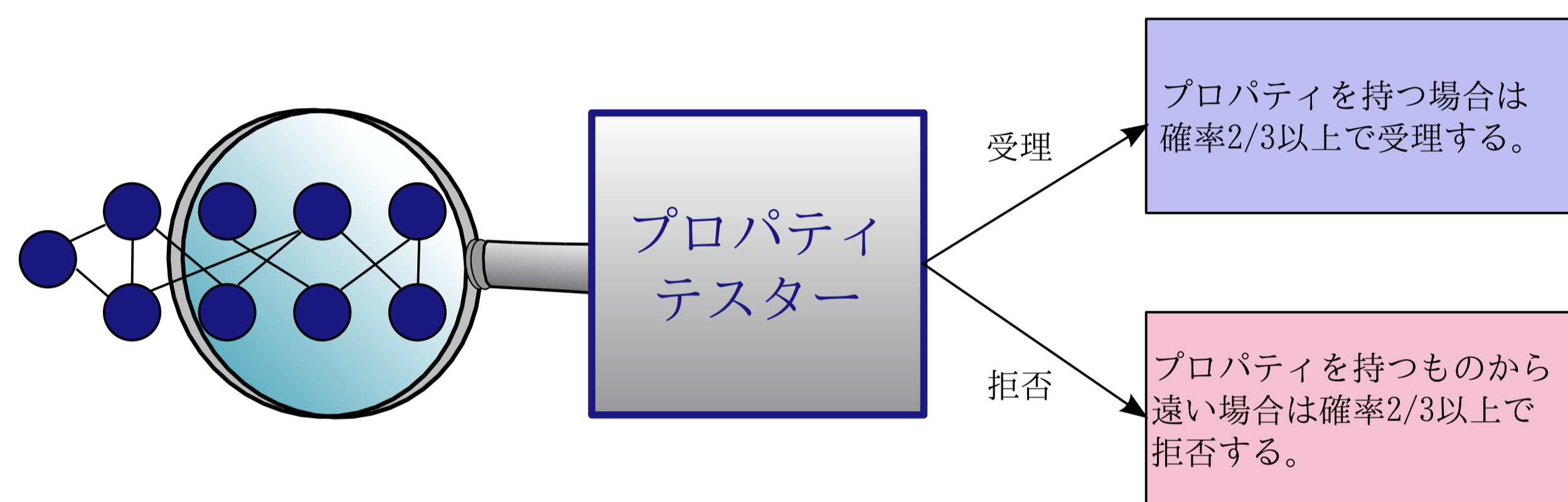
概要

近年は大規模なデータセットが増え、一種のデータセットでは正確なサイズが不明であり、全てのデータを一瞥することすら困難である。入力データ長に対して線形時間がなければほとんど何も計算できないと考えられがちだが、定数時間だけでも計算できるものが多数ある。その一つがプロパティ検査である。プロパティ検査とは、大規模なグラフやデータベースからランダムサンプリングを行い、そのごく僅かなサンプルを元に、確率的な近似アルゴリズムを用いて高速に結論を出力するタイプの検査手法である。

関係データベース等の大規模なデータセットに対して形式言語でクエリを行うことが自然であり、その言語が形式論理と関係が強い。形式検証等でも、論理式の確率的な近似が非常に自然な課題である。本研究では、一階述語論理のプロパティ検査について研究する。特に、一階述語論理で定義できるプロパティを検査可能と検査不可能なものに分類する。この分類問題はAlonら (Alon, *et al.*, *Combinatorica* 20(4):451-476, 2000) が初めて考えた課題である。

論理式の検査とは

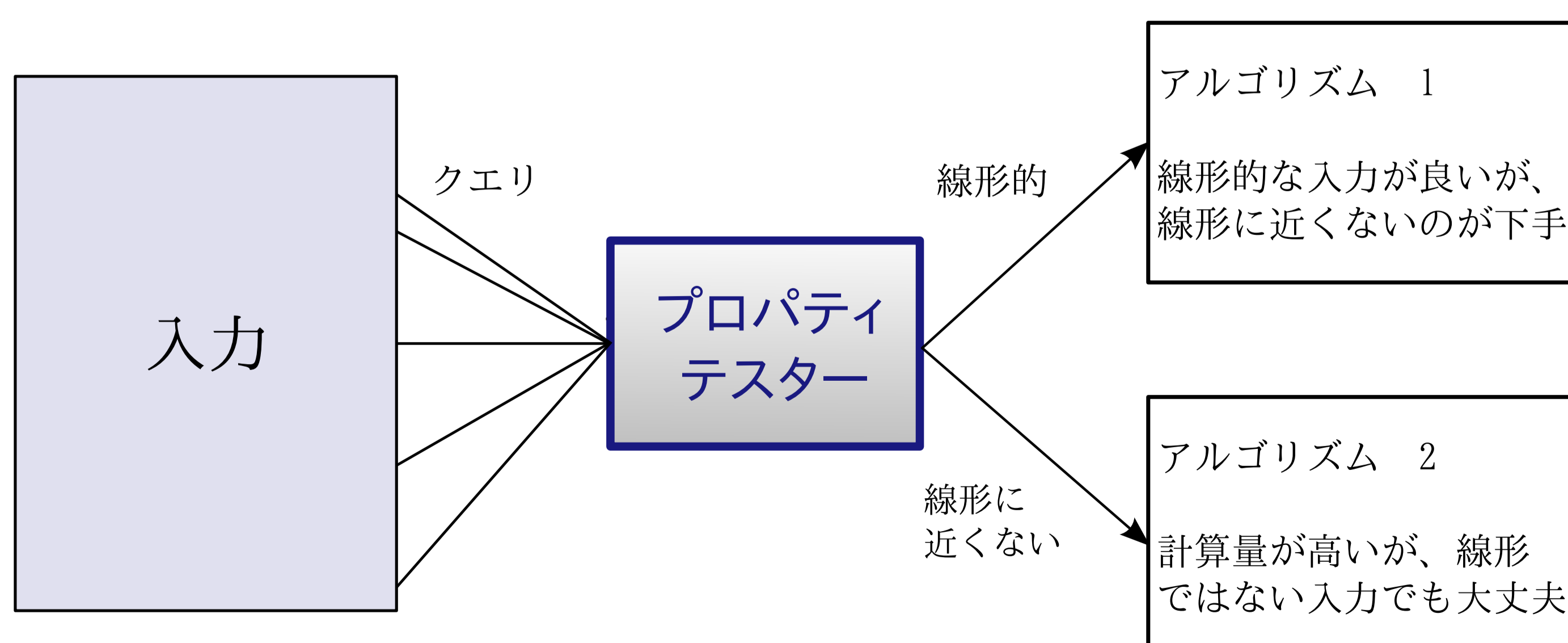
検査可能とは大規模なグラフやデータベース等が入力として与えた時、定数サイズのランダムサンプルだけを見、上の図のように目的の論理式を満たす場合高い確率で受理、目的の論理式を満たすものに近くない場合は高い確率で拒否できることである。



検査不可能なプロパティも一階述語論理で定義できる。一階述語論理式を検査するシステムを作る時、論理式の文法から検査アルゴリズムを自動的に作る必要がある。論理式の文法から検査可能と不可能なサブクラスに分け、検査可能なクラスに対して式の文法から検査アルゴリズムを作ることが本研究の目的である。

関係データベース等での応用

大規模な関係データベースがあるプロパティを満たすかどうかを確認したい場合は定数個だけのタプルを見、確率的に近似する検査が実用できると考えられる。SQL等の形式言語で書かれるクエリに対し、文法からテスターを自動的に作ることが望ましい。



データベース等でデータマイニングを考えると、データベースが何かのプロパティを満たす場合一つのアルゴリズム、そのプロパティから遠く離れている場合は違うアルゴリズムが使いたいと想像できる。その場合は検査アルゴリズムがフィルターとして考えられる。形式検証等でも論理式の検査が課題になる。

プロパティ検査での分類問題

我々の最近の研究成果を見ると一階述語論理で検査不可能なものを定義するため、最小十分な存在記号の数が1、全称記号の数が2と量子子の数が3だと示している。しかし、分類問題は量子子の数に限るわけではない。なので、量子子のパターンや述語記号のアリティと数についても考えると次の分類になる。

| 検査可能なクラス | 証明者 |
|----------------------------------------|-------------------------------------------------------------------------------------------|
| $[\exists^* \forall \exists^*, all] =$ | Jordan and Zeugmann (2009) |
| $[\exists^* \forall^*, all] =$ | Alon <i>et al.</i> (2000) |
| 単項 | Jordan and Zeugmann (2010b) Alon <i>et al.</i> (2001), McNaughton and Papert (1971) |
| 検査不可能なクラス | 証明者 |
| $[\forall^3 \exists, (0, 1)] =$ | Jordan and Zeugmann (2010a) Alon <i>et al.</i> (2000) |
| $[\forall \exists \forall, (0, 1)] =$ | Jordan and Zeugmann (2011) |
| $[\forall \exists \forall, (0, 1)]$ | Jordan (2012) |

Figure 1: プロパティ検査における現在の分類

BDDとプロパティ検査

プロパティ検査は論理式に限ることなく、BDDのプロパティ検査も課題になっている。Bollig (Bollig, *J. UCS* 12(6):710-724, 2006) はサーベイをしている。

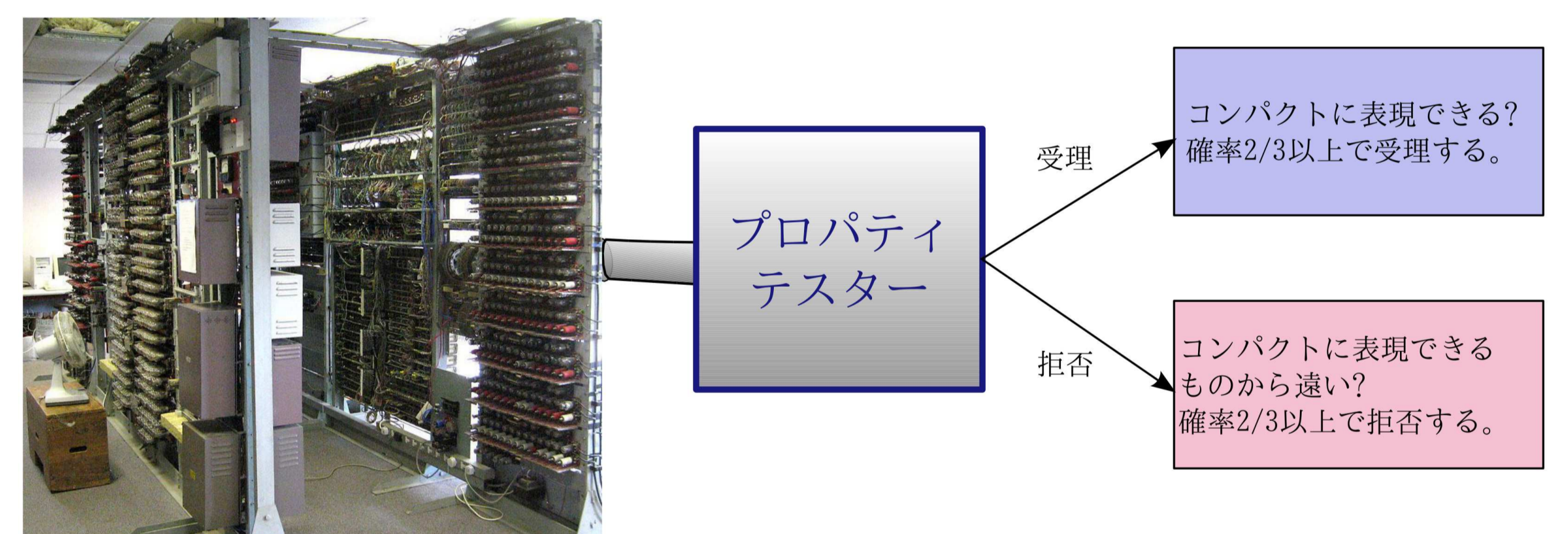


Figure 2: 写真はWikipediaのMaltaGCによる

例えば、何かの関数を計算する複雑なblack boxを与えられたときは、コンパクトに表現できるかどうかを知りたいとする。Black boxの中が見えないが、全ての入力($\Theta(2^n)$)を計算しないで答えたい。

特に、Brodyら (Brody, Matulef, Wu, *TAMC* 2011: 320-331) は入力のブール関数が定数幅のOBDDで表現できるかどうかというプロパティの検査について研究している。幅2で変数順序が決まった場合は $\tilde{O}(\log n)$ クエリでできるが、ほとんどの場合は $\Omega(n)$ クエリが必要になる。また、read-onceの定数幅のBDDで表現できるプロパティは検査可能であるが、read-twiceや線形幅等は検査できない。

これからの課題

1. 上の分類を完全に。特に、 $[\forall^2 \exists, (0, 1)] =$ の検査。
2. 一階述語論理は記述能力が低いので、線形代数的な演算子、順序やLFP等を加えた言語の検査。
3. Szemerédi regularity lemmaを有限関係ストラクチャーに拡張。
4. 検査不可能なものに対して、正確なクエリ計算量。
5. 量子的なDDや量子テスターの検査。