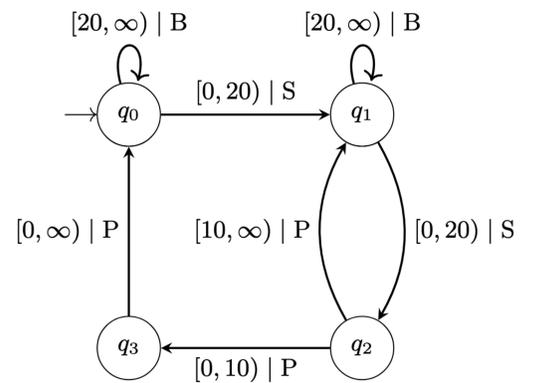


記号的ミーリオートマトンの能動的学習

記号的:無限個の入力文字
ミーリオートマトン:多値出力

入江 堅吾 和賀 正樹 末永 幸平
京都大学大学院情報学研究科



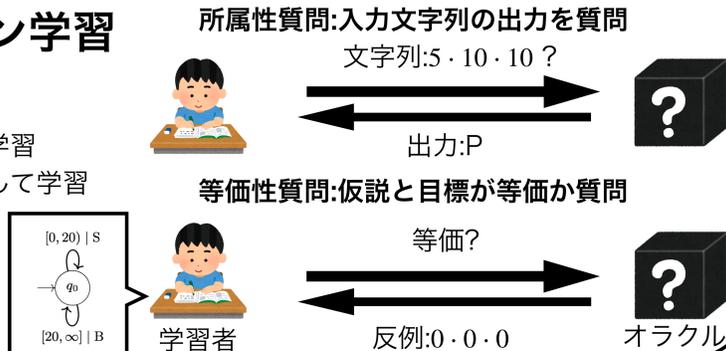
①:記号的ミーリオートマトン

入力アルファベット:無限集合 出力:多値
遷移に述語と出力
述語を決めると定義される

例
入力アルファベット:有理数全体
述語:有理数上の区間
入力:5 · 17 · 3 出力:P

②:能動的オートマトン学習 [Angluin '87]

学習者はオートマトンを正確に学習
オラクルに二つの種類の質問をして学習
・ 所属性質問
・ 等価性質問



③:能動的オートマトン学習

+ PAC (Probably Approximately Correct)
[Valiant '84][Angluin '87]

学習者はオートマトンを「高い確率」で「大体正しく」学習
所属性質問のみで学習
・ 十分な数のテストケースを生成し所属性質問で出力を比較し
等価性質問を代替

①+②:記号的ミーリオートマトンの能動的学習[本研究] アルゴリズム

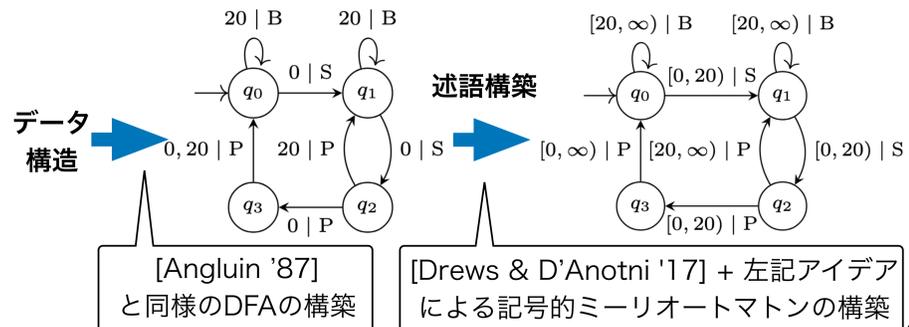
[Drews & D'Antoni '17]

[本研究]

○ 記号的オートマトンの学習

○ 記号的ミーリオートマトンの学習

× 状態数に関して2次の計算量の証明 ○ 状態数に関して1次の計算量の証明



難点:無限の入力文字に対する出力の学習

アイデア:重要な入力文字集合の明示的な保持

①+③:記号的ミーリオートマトンのPAC学習[本研究]

[Maler & Mens '17]

[本研究]

○ 記号的オートマトンのPAC学習

○ 記号的ミーリオートマトンのPAC学習

× 述語がR上の区間のみ

○ PAC学習可能な述語

× 確率1で停止/計算量の議論がない

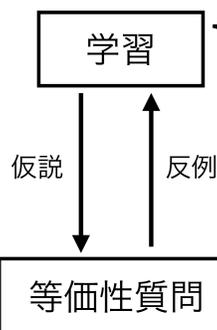
○ 常に停止/述語が多項式回学習可 ⇒ 全体も多項式回学習可

PAC:所属性質問による等価性質問の代替

関数クラスCがPAC学習可能:次の条件を満たす学習アルゴリズムと多項式関数polyが存在
1. 任意の $0 < \epsilon, \delta \leq 1$ 、任意の確率分布 μ 、任意の目標関数 $c_{tgt} \in C$ に対して、
2. μ に従って得られる訓練データのサイズが、 $poly(\epsilon, \delta, size(c_{tgt}))$ 以上あれば、
3. 学習アルゴリズムは確率 $1 - \delta$ 以上で、 $\sum_{x \in X} \chi(c_{tgt}(x) \neq c_{hyp}(x)) \cdot \mu(x) \leq \epsilon$ となる c_{hyp} を学習

PACの保証:

「高い確率(1- δ 以上の確率)」で「大体正しく(誤差が ϵ 以下)」
学習されている



反例を使ってデータ構造にデータを記録していき状態と述語を学習

PACの保証を付けるために
十分な数のテストケースを生成
所属性質問を使って
仮説と目標の出力を比較

アイデア[本研究]

I:述語の誤りに起因する反例を無視

述語をPAC学習可能と仮定
述語を学習するためのデータサイズが十分大きくなった後は...

述語は高い確率で
大体正しいはず!

	[Maler & Mens '17]	[本研究]
状態数の不足に起因	反例扱い🐼	反例扱い🐼
述語の誤りに起因	反例扱い🐼	反例扱いしない🐼

II:テスト回数を増加

述語の誤りに起因する反例を無視するため、誤差が増加
誤差が増加してもPACの保証が付くように
[Maler & Mens '17]よりも生成するテスト数を増加させる

ϵ_p : 述語のPAC学習の誤差の精度
 δ_p : 述語のPAC学習の信頼度の確率
 m : 反例の最大の長さ

	[Maler & Mens '17]	[本研究]
i回目の等価性質問の テスト数	$\frac{-\log \delta + (i+1)\log 2}{\epsilon}$	$\frac{-\log \delta + (i+1)\log 2}{\epsilon - m(\epsilon_p + \delta_p) + m^2\epsilon_p\delta_p}$

停止性・PAC学習可能性[本研究]

定理:本アルゴリズムは停止する

略証:状態数の不足に起因する反例は仮説のオートマトンの状態数を必ず増加させる
状態数の不足に起因する反例のみ反例扱いすることからオートマトンの状態の有限性から停止する

定理:述語が多項式回学習可能 ⇒ 本アルゴリズムも多項式回学習可能

略証:述語がPAC学習可能である仮定から述語の誤りに起因する反例が発生する確率は多項式で抑えられる
よってテスト数も多項式で抑えられる

今後の課題

PAC学習可能な述語のケーススタディ
・ PAC学習の既存研究を活用
実用的なベンチマークでの実験