1 はじめに

最小無矛盾決定性有限オートマトン (DFA) 問題は、計算学習理論において重要な問題の一つであり、理論的な研究は古くから行われている。最小無矛盾 DFA 問題とは、入力である文字列に矛盾しない、出来る限り状態数の小さい DFA を出力する問題である。1978 年には、Gold[1] と Angluin[2] によって $NP-$ 完全であることが示されている。また、$P \neq NP$ の仮定のもとで、Li と Vazirani[3], Simon[4] は多項式時間では定数倍近似ができないことを示し、さらに、Pitt と Warmuth[5] は近似率が opt^k(ただし opt は最適解のサイズ、k は定数) または $n^{1/14}$(ただし n は入力サイズ) の多項式時間近似アルゴリズムが存在しないことを示している。一方、Trakhtenbrot と Barzdin[6] は、入力として与えられる文字列集合が完全であれば、線形時間で解けることを示した。完全であるとは、長さ $k > 0$ 以下の文字列全てが正または負の例として、入力で与えられるということである。また、Angluin[2] は $g(x)-$ 不完全な入力についても定義し、$g(x) = d[\log x]$(ただし d は正定数) のときクラス P に属し、$g(x) = x^{\epsilon}$ (ただし ϵ は正定数) のとき $NP-$ 完全であることを示した。$g(x)-$ 不完全であるとは、長さ k に対し完全な文字列集合について、m 個 (ただし $m < g(2^{(k+1)})$) の例が欠けている文字列集合である。

状態数に関する

・個の最大

このへんどれもよくわからんぞ…

接頭辞集合とは、正負の例を一本の文字列の先頭から途中 (あるいは末尾まで) の文字列に限る例の集合のことである。この接頭辞集合に制限した最小無矛盾 DFA 問題でも、計算量的に困難であることが上埜ら [7] によってわかっている。ただし、すべての接頭辞が正または負の例である完全接頭辞集合については、効率よく解ける可能性は残されている未解決問題である。たとえば、文字の種類が 1 であるときは、効率よく解ける問題であることが最近わかった。

(接尾辞)

本研究では、完全接頭辞集合に無矛盾ななるべく小さなオートマトンを出力する効率のよいアルゴリズムを提案し、この問題について考察する。まず最小無矛盾 DFA 問題について詳しく述べる。次に提案する近似アルゴリズムの説明する。最後に具体的な事例をアルゴリズムに当てはめ、問題について考察を行~~った~~う。

入れる

後で行う (文章中で) ので未来形

引用する.

2 最小無矛盾 DFA 問題について

2.1 決定性有限オートマトン

決定性有限オートマトン (DFA) は、状態と入力によって次に遷移すべき状態が定まる有限オートマトンである。DFA は (S, Σ, T, s, A) の 5 要素から構成され、それぞれは以下の~~性質をもつ~~。ように定義される。

- 状態の有限集合 (S)
- 有限のアルファベット (Σ) ~~1978 年には、Gold~~
- (部分) 遷移関数 $(T : S \times \Sigma \rightarrow S)$
- 初期状態 $(s \in S)$
- 受理状態の集合 $(A \subseteq S)$

i 1.
ii 2.
iii 3.
4.
iv 5

\cdot

$x_0 \cdot x_1 \cdot \cdots \cdot x_{n-1}$

\cdots

の は 区切り

ここで M は、$M = (S, \Sigma, T, s, A)$ という DFA ~~であり~~、$X = x_0, x_1, ..., x_{n-1}$ は Σ に含まれる文字から構成される文字列とする。遷移関数 T を $S \times \Sigma^* \rightarrow Q$ に自然拡張し、文字列 $x \in \Sigma^*$ について、$T(s, x) \in A$ ならば M は文字列 x を受理するという。また、文字列集合 $L(M) = x \in \Sigma^* | T(s, x) \in A$ を M が受理する言語という。M のサイズ $|M|$ は M の状態数 S とする。

\{ \}

DFA

2.2 最小無矛盾 DFA 問題

\{ \}

Σ は有限アルファベットを表す。以降、特にことわらない限り、$\Sigma = a, b$ である。$\Sigma *$ は長さ 0 の空文字列 λ を含む Σ 上の文字列すべての集合を表す。長さ $n > 0$ の文字列 $s = s[1].s[2]....s[n]$ について、正整数 $i \leqq n$ で定められる部分列 $s[1]...s[i]$ を s の接頭辞 (プレフィクス) であるといい、$s[1, i]$ あるいは $s_{(i)}$ と書くことにする。

Σ^*

最小無矛盾 DFA 問題とは、与えられた文字列の集合の組 (P,Q) に対し、P に含まれる文字列はすべて受理し、Q に含まれる文字列はすべて受理しない DFA で状態数が最小のものを求めるものである。すなわち、オートマトン A で定義される言語 $L(A)$ が $P \subseteq L(A)$, かつ $Q \cap L(A) =$ で、状態数 $|A|$ が最小となる、そのような A を見つける問題である。

$s[1] \cdot s[2] \cdot \dots$ $s[n]$

このとき、P を正の例の集合, Q を負の例の集合とよび、組 (P,Q) を標本と呼ぶ。標本 (P,Q) を文字列の集合 S とラベル (関数):S→+,-の組 (S,) で表すこともできる。

内容: 定義域 (集合) → 値域 (集合) なので, たとえば

$l : S \rightarrow \{+, -\}$

数式で統一

…を…といい (および) …は…であるといい

両接名

文字列 s と 3 ラベル : $1, ..., |s| \to +, -, *$ の組で定義される標本 $P = s(i)|1 \leqq i \leqq |s| and(i) = `+', Q = s(i)|1 \leqq i \leqq |s| and(i) = `-'$ を接頭辞集合という。接頭辞~~それぞれを~~例~~にした~~集合~~(サンプル)~~のことである。~~集合というのは、正例と負例の任意の部分集合のことである。~~一本の文字列に対して、その接頭辞にクラスラベル (~~今回は~~正か負のラベル) を貼ったもので、正のクラスラベルが貼られた文字列 (~~受理する文字列~~) が正例、負のクラスラベルが貼られた文字列 (~~却下する文字列~~) が負例である。接頭辞集合が~~ラベル : $1, ..., |s| \to +, -$ で定義できるとき~~、つまり、すべての接頭辞が正または負の例である列の集合であるとき、完全であるという。

とする

すべてを

ラベルが + または − の

すべての接頭辞を含むとき

?

シングルクオートとプライム（ダッシュ）はちがう。

`−'