

## 平成 25 年度 情報工学コース卒業研究報告要旨

酒井 研究室	氏 名	佐 藤 太 貴
卒業研究題目	正規表現の同値性判定アルゴリズムにおける簡約の簡易化に基づく高速化	
<p>正規表現は Kleene によって考案された表記法で、文字列の集合を有限個の記号列で表現する。現在では各種プログラミング言語、コンパイラ、文字列の検索や置換を行うときなどに利用されている。</p> <p>異なる形の正規表現が同一の言語を表しているかを判定する手法として、正規表現を最小の決定性有限オートマトン (DFA) に変換し比較する手法がある。これに対し、Antimirov と Mosses により 2 つの正規表現を決定的線形正規表現と呼ばれる状態に書き換え、それを微分することで得られた正規表現に対し再度同値性判定を行うことで、異なる正規表現の同値性を判定する手法が提案された。この手法は Almeida, Moreira らによって効率化され、実験により DFA に変換する手法より多くの場合で高速に同値性を判定できることが示された。</p> <p>Antimirov と Mosses が提案したアルゴリズムでは、正規表現の同値性判定を行う上で正規表現を簡約形にすることが重要である。彼らの簡約化では、正規表現に対して簡約可能な限り ACI(正規表現の選言の結合律, 交換律, 冪等性) を法として繰り返し簡約する。これにより、扱う正規表現を簡約形に限定できるためアルゴリズムが効率化できる。さらに簡約形に制限することでアルゴリズムの停止性も保証される。しかし、彼らの簡約化はアルゴリズム中で何度も実行する必要がありコストがかかる。Almeida, Moreira らは簡約化のコストを削減するために、正規表現を常に簡約形が維持されるような高度なデータ構造で表すことで簡約化処理を行わないようにし、コストの削減を図った。しかし、このデータ構造が複雑であるため実装を行うことが容易ではない。</p> <p>そこで本研究では、実装が容易な正規表現本来の構造をデータ構造として使用する。そのため、簡約化処理を適宜適用しなければならない。そこで簡約化のコスト削減のために簡易簡約化を提案する。最初に簡約化規則について詳細に解析し、停止性を破壊する可能性があるのは選言の冪等規則 <math>\alpha + \alpha \rightarrow \alpha</math> のみであることを明らかにする。さらに、他の規則は最初に一度だけ適用すれば、アルゴリズム中で再び適用されることがないことも明らかにする。これらを踏まえて、簡約化のコスト削減のため簡約化を冪等規則とそれ以外に分ける。簡約化では ACI を法としていたが、簡易簡約化は冪等規則を AC を法として行い、正規表現に対して部分的に適用するだけで十分であることを明らかにする。すなわち、正規表現 <math>\alpha_1 + \dots + \alpha_n</math> に対して、<math>\alpha_i = \alpha_j</math> となるような <math>\alpha_j</math> に対して冪等規則を適用し、<math>\alpha_i</math> の内側は無視する。また、冪等規則以外の規則については規則を追加することで ACI を法とせずに行い、線形時間での簡約を可能とする。これらにより簡約化を簡易化することで、アルゴリズム全体の簡約化処理回数を減らすことができ、コストが削減される。</p> <p>Almeida, Moreira らが効率化したアルゴリズムを基に簡易簡約化を用いたアルゴリズムを構築する。本論文では簡易簡約化を用いたアルゴリズムを紹介し、このアルゴリズムと Almeida, Moreira らのアルゴリズムとで実行時間を比較し、簡易簡約化を用いたアルゴリズムが多くの場合で高速であることを実験により検証する。</p>		