

平成29年度 情報工学コース卒業研究報告要旨

関 研究室	氏 名	吉村玲勇
卒業研究題目	重み付きオートマトンの計数ツールおよび 重み付きレジスタオートマトンへの拡張	
<p>プログラムの脆弱性を定量化するための手法の一つに、与えられた制約式を満たす系列の個数を数えるという方法がある。例えば、ウェブアプリケーションにおけるクライアントコードの脆弱性検査では、コードが生成する可能性がある系列のうち攻撃者に悪用されるパターンをもつ系列を数える。系列の数え上げを行うための計数ツールがいくつか実装されている。</p> <p>形式級数 (formal series) は形式言語の自然な拡張であり、これはアルファベット上の各系列に対して半環の元を重み (または係数) として割り当てる。形式級数の部分クラスに認識可能級数があり、これは正規言語の自然な拡張である。認識可能級数 S は3つ組 (λ, μ, γ) で表現される。 λ は初期状態を表す行ベクトル、 μ は入力語から状態遷移行列への写像、 γ は終状態を表す列ベクトルである。入力語 w に対して認識可能級数 S における重みを $(S, w) = \lambda\mu w\gamma$ と定義する。 (λ, μ, γ) を重み付きオートマトン (Weighted Automata, WA) または S の線形表現という。すなわち WA は、有限オートマトンにおいて各状態に入力に従って重みを割りつけることにより拡張した計算モデルである。</p> <p>先行研究では WA に対する計数問題を定義し、それを解くアルゴリズムが提案されているが、実データに基づく有効性は確認されていない。計数問題は S を形式級数、 d を自然数とすると $CC(S, d) = \sum_{ w =d} (S, w)$ を求める問題と定義される。これは S における長さ d の系列の重みの和を求める問題である。提案されたアルゴリズムの時間計算量は1回の行列計算に要する時間を η とおくと $O(\eta \log d)$ である。</p> <p>本研究ではまず提案された WA の計数アルゴリズムを実装し、ベースラインとなる2つの計数アルゴリズムと先行研究で提案されたアルゴリズムの実行時間を比較することにより、その有効性を確認した。計数ツールは WA と系列長を入力とする。実験用に、文字数の数え上げ、Arctic な半環、マルコフモデルなどの異なる性質の複数の WA を用意した。各オートマトンを入力として計数ツールを実行し、系列長を変化させながら3つの計数アルゴリズムの実行時間を測定した。実験の結果、どの WA を入力とした場合も理論的な時間計算量に近い実行時間が得られた。すべての結果を比較して、いずれの場合も3つの計数アルゴリズムの実行時間の関係性は同じになり、提案された計数アルゴリズムの有効性が確認できた。</p> <p>有限オートマトンの別の拡張として、レジスタをもちデータ値に対する限定された演算を許すレジスタオートマトンがある。近年、レジスタオートマトンに重みを導入し拡張した重み付きレジスタオートマトン (Weighted Register Automata, WRA と略) が提案された。WRA はレジスタを用いてデータ値の記憶や比較を表現でき、重みで遷移や操作のコストを表すことができる。記号とデータ値の対の系列 $(a_1, d_1) \cdots (a_n, d_n)$ をデータ語とよぶ。WRA は、各データ語に重みを割り当てたデータ級数を表現する。本研究では WA の計数アルゴリズムを WRA へ拡張する。そのために WRA の射影級数を定義する。これは、WRA の表現するデータ級数において、各データ語のデータ値部分を消去して得られる通常の語に、もとのデータ級数における重みの総和を重みとして与えたものである。本研究では任意に与えられた WRA \mathcal{A} の射影級数を表現する WA を \mathcal{A} から構成可能であることを証明した。これにより、WRA の射影級数に対する計数問題を WA の計数アルゴリズムによって効率よく解くことができる。</p>		