

有限オートマトンと正規表現の等価性

大石巧 @noan6251

岡山大学 理学部数学科 3年生

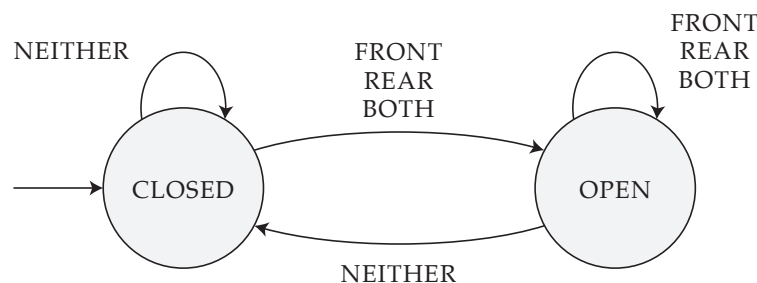
2015/12/13

0 前提知識

集合や組 (tuple) の定義や記法は既知とする。それ以外の前提知識は特に必要無い。

1 Introduction

有限オートマトンは「有限個の状態と遷移の組」と大雑把に定義することができる。そのため有限状態機械とも言われる。ここで言う状態は一般に使用される辞書的な意味の状態と認識して良く、遷移とはある状態から(別とは限らない)ある状態へ移行することである。遷移には条件があり、それを満たすことによってある状態から次の状態へ、そしてまた次の状態へ……と遷移していく。有限オートマトンは一つの開始状態と一つ以上の終了状態があり、終了状態で手続きが終わった時受理されたと言う。以下の図は自動ドアの動作を状態遷移図で記述したものである。終了状態がないため有限オートマトンではないが非常に近い構造をしている。



有限オートマトンは工学の様々な分野で応用されており、例えばネットワーク、人工知能、自然言語のモデル化などである。「有限オートマトンで受理できる形式言語(記号の列の意)」を正規言語という。「文字列の集合を一つの文字列で表現する方法」($\{x|P(x)\}$ は複数の元を一つの記法で表現する方法とも解釈できる)を正規表現といい、ある形式言語が正規言語であることと正規表現で表わされることは同値である。正規表現は、文字列のマッチングへ応用されており、非常に強力な文字列検索ツールとして利用される。

2 講演内容

本講演ではまず技術的な正規表現について10分程度の解説を行う。正規表現はプログラミング言語によって実装が異なるので、一般的なLinuxには標準で入っているであろうPerl、その正規表現について解説する。有限オートマトンの例を示した後に正確な定義を行う。幾つかの補題や定理を証明しながら、有限オートマトンを拡張し、定理として「ある形式言語 L に対して、 L が有限オートマトンで受理可能 $\Leftrightarrow L$ は正規表現で表わされる」を証明する。但し時間の都合上一部の証明は略証を示すに留める。

3 正規言語

n を自然数とする。ここに自然数は0を含む。ある文字 a に対して a^n は a を n 文字並べたものであると定義する。 ϵ は空文字列と言い、とりあえずは空集合の文字列版と思ってもらえれば良い。

例えば $\{a^n\}$ や $\{a^n b^m | n, m \in \mathbb{N}\} = \{\epsilon, a, b, aa, ab, aaa, \dots\}$ は正規言語である。

参考文献

- [1] 計算理論の基礎 [原著第2版] 1. オートマトンと言語, Michael Sipser, 共立出版 (2008)
- [2] 詳説正規表現第3版, Jeffrey E.F. Friedl, O'Reilly Japan (2008)