# 文字列に含まれる連の個数について

篠原　歩，草野一彦
東北大学大学院情報科学研究科
システム情報科学専攻　知能システム科学講座

本講演では，長さが $n$ の文字列に含まれる連の個数の最大数 $\rho(n)$にまつわる話題を我々の研究成果を中心に紹介する．

連(run)とは，文字列に含まれる 2 回以上の繰り返しで，それ以上，左にも右にも延長できないものいう．例えば，文字列 **abaababaaab** には，**aa**, **aaa**, **ababa**, **abaaba** という４つの連が含まれているが，これと同じ長さ 11 の文字列 **aaaaaaaaaaa** には 1 つしか連が含まれていない．一方，**aabaabbaabb** の中には **aa, aa, bb, aa, bb, aabaab, aabbaabb** という 7 つの連が含まれており，長さ 11 についてはこれが最多である．すなわち，$\rho(11)=7$ である．

Kolpakov と Kucherov は 1999 年に，入力として与えられた文字列に含まれるすべての連を数えあげる線形時間アルゴリズムを解析する中で，$\rho(n) = O(n)$ であることを示した．Rytter は 2005 年に，この定数項を解析し，$\rho(n) \leqq 5n$ を証明した．これを契機に，この定数をより厳密に求める研究が活発になり，3.44, 1.6, 1.52 と徐々にその上界を下げる証明が示され，現時点で最良の上界は 2009 年に Crochemore が示した 1.029 である．

一方，$\rho(n)$の下界については，2003 年に Franek らが示した $0.927n$ が，しばらくの間，最良であったが，我々は，2008 年に，新たな下界を与える文字列を見つけた．ここで用いたアプローチは，多くの連を含む文字列同士を連接することで，より多くの連を含む文字列を見つけ出そうという，遺伝的アルゴリズムにヒントを得たものであった．この方法を洗練させることで，この定数を 0.944565，さらには 0.9445756 にあげることができた．現在知られている最良の下界は，我々と Simpson が独立に見つけた 0.9445757 であるが，興味深いことに，この２つは異なる文字列の系列であるにもかかわらず，その定数（方程式の解として与えられる無理数）が厳密に一致している．

現在のところ，$\rho(n) < n$，すなわち，この定数は 1.0 未満であると予想されている．

また，$n$ が小さなところでは，計算機を使って文字列を網羅的に調べることで $\rho(n)$ は正確に求められる．この厳密な値を，アルゴリズムと実装の工夫でなるべく大きな $n$ に対して求めようという努力もなされている．ビット演算処理や GPU の活用，また探索空間の枝刈りなどを駆使することで，現時点では $n = 63$ まで求まっている（下の表）．

なお，長さ $n$ の文字列に含まれる連の平均数については，一般項が求まっている．

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho(n)$ | 0 | 1 | 1 | 2 | 2 | 3 | 4 | 5 | 5 | 6 | 7 | 8 | 8 | 10 | 10 | 11 | 12 | 13 | 14 | 15 | 15 |

| $n$ | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho(n)$ | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 27 | 28 | 29 | 30 | 30 | 31 | 32 | 33 | 35 |

| $n$ | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho(n)$ | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 |