

生命の言語の理解をめざして

加藤 有己
関 浩之

京都大学化学研究所
奈良先端科学技術大学院大学情報科学研究科



[受賞論文]

- RNA Pseudoknotted Structure Prediction Using Stochastic Multiple Context-Free Grammar
- 加藤有己(奈良先端科学技術大学院大学情報科学研究科, 現在, 京都大学化学研究所), 関浩之(奈良先端科学技術大学院大学情報科学研究科), 嵩忠雄(奈良先端科学技術大学院大学情報科学研究科, 2007年3月逝去)
- 情報処理学会論文誌, Vol.47, No.SIG17 (TBIO1), pp.12-21 (2006)

このたび、標記の論文で本会論文賞をいただくことになり、大変光栄に存じます。査読者ならびに論文を高く評価して下さった皆様に厚く御礼申し上げます。

近年、情報科学で培われた技術を駆使して、分子生物学における問題の解決を目指すバイオインフォマティクスと呼ばれる分野の研究が盛んに行われています。遺伝情報を担う DNA からの転写産物である RNA にはタンパク質に翻訳されない noncoding RNA (ncRNA) と呼ばれるものが存在し、高等生物ではゲノム中で ncRNA に転写される割合が高いことが知られています。そのため、生体内での ncRNA の機能に注目が集まっています。生物学の経験則から、構造と機能の間には相関があると言われており、機能推定のためには構造を知ることが重要になります。しかしながら、RNA の立体構造を実験的に決定することは容易ではありません。そのため、情報科学的アプローチに基づいて、塩基の結合情報のみを表す 2 次構造を配列データから予測するさまざまな手法が提案されています。文脈自由文法等の形式文法を用いた手法は RNA2 次構造のモデル化における成功例の 1 つです (図-1 参照)。ここでは、2 次構造予測を文法の構文解析に帰着させます。

我々が RNA 構造解析の研究を始めたきっかけは、標記論文の共著者である故嵩名誉教授の勧めからでした。嵩教授は形式言語理論において今日 CYK アルゴリズムとして知られる文脈自由文法の構文解析アルゴリズムを提案され、また符号理論においても多大な功績をあげられました。2002 年に米国 Howard Hughes 医学研究所の Eddy 教授から前述のアルゴリズムの照会を受けられ、CYK アルゴリズムの拡張版が RNA2 次構造予測に用いられていることをお聞きになり、我々にできることはないか考えるきっかけを与えてくださいました。RNA2 次構造には、塩基配列上で塩基対を表す弧が交差するシ

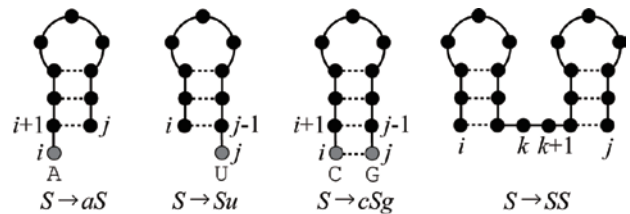


図-1 RNA2 次構造と文脈自由文法

ュードノットと呼ばれる部分構造が存在することもあり、文脈自由文法では記述できないことが知られていました。そこで我々は、1988 年に嵩らが提案した多重文脈自由文法を用いるという着想に至りました。この文法は文脈自由文法の自然な拡張であり、交差する依存関係を表現するのに適しています。標記論文では、新たに提案した確率的拡張モデルに対する構文解析アルゴリズムを用いてシールドノットを考慮した 2 次構造予測を行い、良い予測精度を得ることができました。

DNA, RNA 塩基配列, タンパク質アミノ酸配列等を言語とみなし、それらを生成する文法を構築することによって生命の言語を理解する試みは、今後ますます期待されると信じてやみません。

(平成 20 年 4 月 28 日受付)

平成 19 年度論文賞の受賞論文紹介

加藤 有己(正会員) ykato@kuicr.kyoto-u.ac.jp

2007 年奈良先端科学技術大学院大学博士後期課程修了。博士(工学)。同年、日本学術振興会特別研究員 PD。現在、京都大学化学研究所特定研究員。バイオインフォマティクスに関する研究に従事。

関 浩之(正会員) seki@is.naist.jp

1987 年大阪大学博士後期課程修了。工学博士。同年、大阪大学助手。同講師、助教授等を経て、1996 年奈良先端科学技術大学院大学教授。現在に至る。形式言語理論、ソフトウェア基礎理論に関する研究に従事。