

平成 17 年度科学研究費補助金実績報告書 (研究実績報告書)

1. 機関番号 1 4 6 0 3 2. 研究機関名 奈良先端科学技術大学院大学
3. 研究種目名 特別研究員奨励費 4. 研究期間 平成 17 年度 ~ 平成 19 年度
5. 課題番号 1 7 ・ 2 8 3 0
6. 研究課題名 生物配列の高次構造記述向き形式文法とその構造予測への応用

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
	カトウ, ユウキ 加藤, 有己	情報科学研究科	特別研究員 (DC1)

8. 研究分担者(所属研究機関名については、研究代表者の所属研究機関と異なる場合のみ記入すること。)

研究者番号	研究分担者名	所属研究機関名・部局名	職名
	カトウ		
	カトウ		
	カトウ		
	カトウ		
	カトウ		

9. 研究実績の概要(国立情報学研究所でデータベース化するため、600字~800字で記入。図、グラフ等は記載しないこと。)

形式文法の構文解析技術を応用した、RNAの2次構造予測が注目を浴びている。特に、シュードノットと呼ばれる、1次構造上で塩基対が互いに交差して現れる依存関係を持つ部分構造をモデル化するための形式文法がいくつか提案されている。機能が未知のRNAは数多くあり、構造が類似していれば機能も類似するという生物学の経験則から、RNAの2次構造を知ること、その機能の解明につながるものと期待されている。本研究では、文脈自由文法の自然な拡張モデルであり、高い解析能力を持つ多重文脈自由文法(MCFG)を用いて、RNAの2次構造解析を行った。我々の先行研究では、シュードノットを含むRNA 2次構造記述のための生成能力が最小であると考えられるMCFGの部分クラスを同定した。本年度は、その部分クラスの規則に記号出力確率、状態遷移確率を付与した確率多重文脈自由文法(SMCFG)を導入した。次に、SMCFGに対して、入力塩基配列長の5乗のオーダーで確率最大の導出木を計算する構文解析アルゴリズム(CYKアルゴリズム)を設計した。また、文字列が出力される確率を計算するインサイドアルゴリズム及びアウトサイドアルゴリズムを設計し、両者を組み合わせて用いるEMアルゴリズムで、文法の確率パラメータを学習することが理論的に可能であることを示した。最後に、計算機実験を行うためにC言語を用いてCYKアルゴリズムを実装した。なお、プログラムへの入力にはSMCFGとRNA塩基配列である。ここで、Rfamと呼ばれるデータベースに登録されており、シュードノットを構成するRNAファミリーに属す複数の塩基配列に対して2次構造予測を行った。SMCFGによる予測2次構造の精度(感度)は、Rfamの2次構造アノテーションと比較した結果99%以上となり、高い精度を示した。従って、RNA 2次構造予測に対して提案手法は有効であると考えられる。

成果の公表を見合わせる必要がある場合は、その理由及び差し控え期間等を記入した調書(A4判縦長横書1枚)を添付すること。

10. キーワード

- (1) 確率多重文脈自由文法 (2) RNA 2次構造 (3) シュードノット
- (4) (5) (6)
- (7) (8) (裏面に続く)

11. 研究発表(平成17年度の研究成果)
〔雑誌論文〕 計(2)件

著者名	論文標題		
Y. Kato	Stochastic Multiple Context-Free Grammar for RNA Structure Analysis		
雑誌名	巻・号	発行年	ページ
情報処理学会研究報告	2005・128	2005	7-14

著者名	論文標題		
Y. Kato	Stochastic Multiple Context-Free Grammar for RNA Pseudoknot Modeling		
雑誌名	巻・号	発行年	ページ
NAIST Technical Report		2006	

著者名	論文標題		
雑誌名	巻・号	発行年	ページ

著者名	論文標題		
雑誌名	巻・号	発行年	ページ

著者名	論文標題		
雑誌名	巻・号	発行年	ページ

著者名	論文標題		
雑誌名	巻・号	発行年	ページ

〔図書〕 計(0)件

著者名	出版社		
書名	発行年	総ページ数	

12. 研究成果による工業所有権の出願・取得状況
計(0)件

工業所有権の名称	発明者	権利者	工業所有権の種類、番号	出願年月日	取得年月日