

論文審査の結果の要旨

氏名 森 遼太

本論文は、生命情報学で現れる、高次元空間上の確率分布に関し、様々な特徴量の値の確率分布を効率的に計算する手法を提案している。一般に、配列アライメント問題や RNA 二次構造予測問題においては、入力配列サイズに指数的に依存する数の解候補を網羅的に探索することが必要となる。これらの問題では、動的計画法を用いて多項式時間で最尤解を計算できることが知られている。しかし、解空間の高次元性により最尤解の確率重みは通常非常に小さく、確率分布をあまりよく代表していない。このため、確率分布全体の構造を調べることができる計算手法が必要とされており、申請者はこの問題について研究した。

申請者は、まず、高次元空間における離散的な特徴量について、そのとりうる値の確率分布を厳密に計算する手法を開発した。離散的な特徴量としては、例えば二次構造予測問題では、与えられたリファレンス二次構造と各二次構造との塩基対的な距離などがある。申請者は、離散フーリエ変換と動的計画法のアルゴリズムを組み合わせることにより、このような離散的特徴量の高次元空間上の確率分布を多項式時間で計算する手法の開発に成功した。この手法を用いて RNA 二次構造の確率分布が予測構造の近傍にどの程度集中しているかを解析することができた。例えば、tRNA の二次構造予測がうまくいかない場合であっても、正解二次構造の周辺にはある程度の確率が集中していることや、リボスイッチの構造遷移経路に沿って、比較的高い確率重みが存在することを明らかにした。また、予測二次構造の周辺にどの程度確率が集中しているかを指標として、予測構造の信頼度を計算することができることも示した。

次に申請者は、アライメントスコアや二次構造自由エネルギーなど、連続値の特徴量の確率分布を計算するアルゴリズムの研究を行った。まず、連続値の値域を有限幅のビンに区切り、その連続値がその範囲内に収まる高次元空間の点の確率重みの総和を計算するための適切な窓関数について考察を行った。動

的計画法により、窓関数のフーリエ係数が正確に求めらねばならないという要請から、適切な窓関数を決定した。この関数を使うことにより、連続値の場合であっても多項式時間で特徴量の確率分布を必要なだけ正確に求めることができることを示した。申請者はこの手法を RNA 二次構造のエネルギー分布の計算に応用し、RNA 種による分布の特異性を調べた。

本論文の手法は宇宙の全原子数よりも多い解空間上の確率構造を詳細に調べることを可能にするもので極めて価値の高いものである。また、適用対象となる確率分布や特徴量は生命情報学に頻出するもので今後幅広い応用が期待される。この研究課題は申請者の創意に基づくもので、全編に渡って申請者が主体的に研究を進めたと判断される。

これらのことから、博士（科学）の学位を授与できると認める。

以上 1 1 8 5 字