

■吉田 悠一 情報学プリンシプル研究系 助教

【タイトル】

## ビッグデータ時代に不可欠な高速アルゴリズムを考案

【本文】

インターネットの Web やゲノムなど、大規模なデータを扱うことが一般的となった今日、膨大な情報量を実用的な速度で解析できる計算の手法（アルゴリズム）の開発が求められています。私は、データの大きさにかかわらず一定の時間で処理できる「定数時間アルゴリズム」の研究を進めています。

### すべてを見ずに問題を解く

「頂点」とそれを結ぶ「枝」からなる図形を「グラフ」とよびます。たとえば、グラフ中のどの枝も両端の頂点の色が異なるように、3色で塗り分けることが可能かどうかを調べるアルゴリズムをつくるとします。普通はグラフ全体を見て、可能なら Yes、不可能なら No と判定していくのですが、グラフが大きければそれだけ計算時間も長くなってしまいます。大規模なグラフの場合、いつまでたっても処理が終わらないということになりますから、この問題は計算機では解けないと数学の世界では考えられているのです。

しかし、データのすべてではなく一部分だけを見て判断できれば、計算のスピードを大幅に速めることができるはずです。そこで、Yes、No の判定を緩めて、可能なら Yes、可能からほど遠いものは No、それ以外のは見ない、という手法を考えました。これが「定数時間アルゴリズム」です。この手法を使えば、グラフの大きさに関係なく決まった時間で判定することができます。今後、実用に応用できれば、これまで解析がむずかしかった大規模データを簡単に扱えるようになるでしょう。

### 巨大グラフを解析する高速アルゴリズムの開発

「グラフ」というのは、実は世の中にたくさん見られます。ひとつは Web やソーシャルネットワークなどの複雑ネットワークとよばれるものです。Web の各ページや Facebook の個人を頂点、Web 間のリンクや Facebook の友だち関係を枝と見なすとグラフになります。もうひとつは、交通のネットワークを表した地図グラフで、交差点や駅を頂点、道路や線路を枝と見なします。これらのネットワークは近年急速に膨張していて、やがて今のアルゴリズムでは解析できない巨大なサイズになると予想されます。私は、そうした巨大なネットワークを巨大グラフととらえ、解析を可能にする高速アルゴリズムの開発を目指したプロジェクトにも携わっています。

グラフの性質をふまえてモデルをつくり、予想を立てて実験し、確認するという、これまでコンピューターサイエンスの分野にはなかった“研究のループ”をこのプロジェクトを通して確立し、より実用的なアルゴリズムの開発を目指したいと思っています。

(取材・構成 財部 恵子)