

ビッグデータ解析者が欲しいサイズのデータグループが得られる新ツール

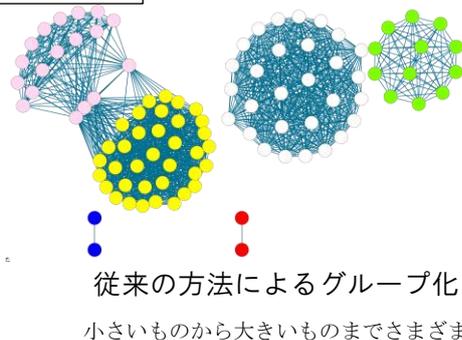
大規模な多変量データを実用的にグループ化できる 新ツールを開発・公開！

大阪府立大学（学長：辻 洋）生命環境科学域 応用生命科学類の尾形善之准教授と公益財団法人かずさDNA研究所の鈴木秀幸博士の研究グループは、ビッグデータの解析時に用いるデータの小さなグループ分け時に効果的な新手法として、大規模な多変量データを構成する要素を効率的かつ実用的にグループ化できるツール「ConfeitoGUI」を開発・公開しました。このツールは、関連ネットワーク解析という手法を応用した特許技術（特許第 6318334 号）に基づき開発しております。またこの研究成果は 2018 年 10 月 24 日 AM 4:00（日本時間）に、科学と医学分野の科学雑誌「PLOS ONE」にてオンライン公開されました。

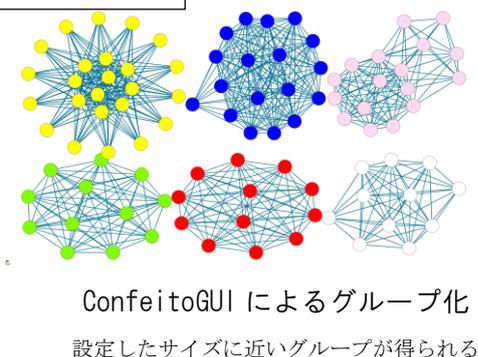
■本取り組みのポイント■

- ・従来のツールでは、グループの大きさを調整することが難しかったが、今回の「ConfeitoGUI」ではユーザーが期待するサイズのグループを多く得られるようになり、解析者の負担軽減に寄与した。
- ・すでに遺伝子発現データなどのバイオサイエンスのデータに応用され、実験科学的な検証をするために相応しい数の要素を含むグループが得られている。
- ・ツール名「confeito」は金平糖の語源から取り、実際に金平糖の製造工程も本ツールのアルゴリズムの参考とした。
- ・この研究成果は、2018年10月24日に PLOS ONE 誌にてオンライン公開された。
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0206075>
- ・「ConfeitoGUI」ツールは、以下の URL で公開している。
<http://www.plant.osakafu-u.ac.jp/~kagiana/confeito/>

これまで



ConfeitoGUI



【研究内容について】

私たちは多変量データの要素をその変数の値に基づいてグループするアルゴリズムを構築してきました。従来の方法は統計学的には優れたアプローチではありますが、得られるグループのサイズを指定することができないため、注目したグループについて実験科学的な検証をしたい研究者にとっては、何度も不慣れな再解析をすることが求められていました。

私たちはそうした研究者の声に応えるべく、ユーザーが欲しいサイズのグループを得られるアルゴリズムを考案し、特許技術を取得しました（特許第 6318334 号）。そして、この特許技術に基づいて、ユーザーフレンドリーな「ConfeitoGUI」ツールを開発しました。「confeito」というのはポルトガル語で金平糖の語源です。従来の方法で得られたグループのネットワークは大きさがバラバラになりやすいのですが、ConfeitoGUI で得られたグループは大きさが揃っていることが多く、まるでカラフルな金平糖を眺めている様子だと思い名付けました。実際に、金平糖の製造工程もこのアルゴリズムの参考にしています。



お菓子の金平糖

本研究成果は、2018年10月24日に PLOS ONE 誌にてオンライン公開されました。

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0206075>

また、ConfeitoGUI ツールは、以下の URL からダウンロードできます。

<http://www.plant.osakafu-u.ac.jp/~kagiana/confeito/>

1. 背景

最近になって、分析機器や観測機器のハイスループット化に伴い、世間ではビッグデータが注目されるようになってきました。こうしたビッグデータを有効活用するためには、ビッグデータのままで見るのではなく、多変量解析という手法を使って効率的に小さいグループに分けて見ていくことが望ましいと考えられます。こうしたグループ分けをするツールはいろいろとありますが、統計学的な判断に従って分けるために、分けられたグループのサイズが大きいものや小さいものがさまざまになります。使う側としては、見つかったグループの中を詳しく見ていくために、ある程度の大きさのグループが欲しい場合が多いのではないかと考えました。

2. 研究成果

そこで、私たちは多変量解析のひとつである相関ネットワーク解析という手法を基に、ビッグデータを効率的に分ける ConfeitoGUI ツールを開発しました。ConfeitoGUI ツールでは、ユーザーが欲しいグループの大きさをあらかじめ決めておくことで、その大きさのグループを効率的に取り出すことができるようにしました。大きさを調整したとはいえ、統計学的に有意なグループを得ることができます。例を挙げると、遺伝子を研究する分野であれば、ConfeitoGUI で得られた遺伝子のグループをさらに実験的な手法で詳細に調べるために利用できます。この場合、グループのサイズが小さすぎたり大きすぎたりすると、正しい結果につながらない可能性が出てきます。つまり、小さすぎると本来目的としている遺伝子が含まれていなかったり、逆に大きすぎるとすべて実験をするのが大変だったりします。

3. 今後の展開

現在、ConfeitoGUI はビッグデータをさらに効率的に分ける方法を取り入れようとしています。これまでの方法だと、グループに入れられない孤立したものが出てくることがありました。そこで、そうした孤立したものをさらにグループに入れるように調整しています。この調整が完了することで、ビッグデータをすべてグループに分けることができるようになります。また、ビッグデータになると、同じ種類のデータばかりでなく、数値データと文字データのように異なるタイプのデータが混在してきます。そうした異種のデータ解析にも活用できるように、開発を進めています。