

# 属性分類の為の逆問題解決型関連性計量システムの実現

政策・メディア研究科 修士課程 1年 引地志織

## 研究概要

本研究の目的は、専門知識がなければ理解が難しいデータ群（腸内細菌叢データ等の医療データ）を対象として、クラスタリングとデータマイニングを適用し、“データの属性分類に関連する指標の組み合わせ”を文脈として設定することにより、数値データから意味・知識の発見を行うシステムを構築することである。本研究では、“データ属性と指標の関連性”という既存手法では表現しにくい特徴を、先行研究により属性との関連性が知られていない場合でもデータの分布や連続性を捉えることによって、対象データの意味抽出（属性を分類するのに有意な部分空間抽出）を行うシステムを実現する。これにより、従来のクラスタリング手法とは異なり、腸内細菌叢データをヒト属性（国籍、性別、年齢等）ごとに分類するための主要な“要素”と“軸”を逆問題的に発見し、利用者の背景知識に関わらず、属性分類のための特徴量を視覚的に理解することが可能となる。

## 研究の社会的意義

本研究は、腸内細菌叢データやメタボロームデータという生命科学分野の専門的なデータへの適用により、新規治療法を提案できる応用可能性も持ち合わせており、情報分野と生命科学分野の両分野において新たな方向性を示すものとして期待される。情報分野では、結果として分類されている集合について、それらを分類する軸・因子を逆問題として求める方式として新規性があり、現在はデータマイニング手法の一種として利用されているクラスタリングにおける新しい知識発見の方法論としてインパクトを与えられる可能性がある。医療分野では、属性分類のために特徴量として、重要な指標、および、重要な指標の種類数も検討することが可能となり、データの利用の効率化、および、異なる分野の専門家の知識の相互利用が可能により、薬剤開発プロセスとの連携や新たな治療法の開発に貢献出来る可能性がある。

## 主な研究成果

2017年に11th IEEE International Conference on Semantic Computingにて、腸内細菌叢データをヒト属性ごとに分類する“意味的な軸”を発見するための「意味的逆問題解析方式 (Semantic Inverse Analysis)」とそのアルゴリズムとして、口頭発表を行った[1]。口頭発表後、そのシステムをもとに、応用可能性の高いシステム設計・実現に向けて、2軸までの一部の逆問題を解決する方式だけではなく、3軸以降の拡張や他データへの検証実験を行った。

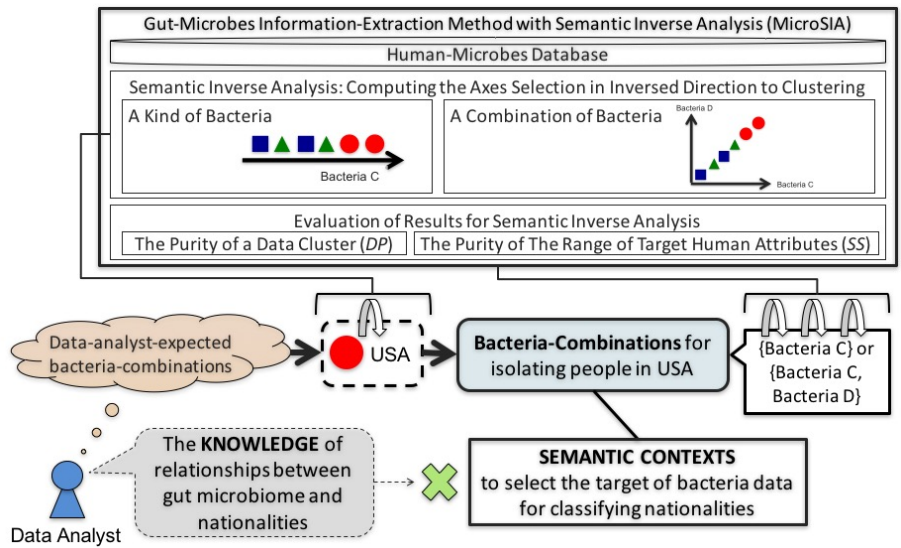


図1. 本研究が目的とする属性分類のための特徴量を発見・認知する逆問題解決型関連性計量環境

[1] "MicroSIA: A Gut-Microbes Information-Extraction Method with Semantic Inverse Analysis for Discovering Unique Bacteria-Combinations in Nationality", Hikichi, S. and Kiyoki, Y., 11th IEEE International Conference on Semantic Computing, 2017. San Diego, California, USA.