

## 論文内容の要旨

博士論文題目 Data-Intensive Science of Relationships Among Species, Volatile Organic Compounds and Biological Activities

(揮発性有機化合物と生物種および生物活性間を関係づける データ駆動型サイエンス)

氏名 Azian Azamimi binti Abdullah

(論文内容の要旨)

ビッグ・データ・サイエンスにおける最も重要なテーマは、ヘルスケアとエコロジーにおけるデータ駆動型サイエンスである。そのうちのひとつであるエコロジー研究の一環として、生物種、生物が生合成し、環境中に放出する揮発有機化学物質(VOC)、VOCの他の生物へ与える活性の三組の情報から法則性をマイニングすることを検討した。現在までに、科学論文で報告されているVOCを、上記の三組の情報を収集し、データベースを構築した。本データベースは、KNApSAcK Ecology DBとして整理し公開した(<http://kanaya.naist.jp/MetaboliteEcology/top.jsp>)。本データベースには8285エントリーのVOCと生物活性の関係情報が含まれている。KNApSAcK Ecology DBを活用し、教師なし機械学習の一方法である、次元クラスタリング、教師ありの機械学習としては、多層構造のニューラルネットワーク(DNN)、Gradient Boosting Machine (GBM)、ランダムフォレスト法(RF)、一般線形モデル(GLM)を活用し、VOCの化学構造と生物活性の関係を把握することを試みた。これらの結果を総合的に解釈し、VOCの化学構造の特徴により生物活性を説明することに成功した。また、化学構造からVOCの生物活性の予測についての可能性についても示唆された。

氏名	Azian Azamimi binti Abdullah
----	------------------------------

(論文審査結果の要旨)

平成 21 年 12 月 22 日に開催した公聴会の結果を参考に、平成 29 年 1 月 11 日に本博士論文の審査を実施した。以下に述べる通り、本博士論文は、本学位申請者が、独立した研究者として研究開発活動を続けていくために必要な素養を備えていることを示すものである。

Azian Azamimi binti Abdullah は、ビッグ・データ・サイエンスにおける最も重要なテーマであるヘルスケアとエコロジーにおけるデータ駆動型サイエンスのうち、エコロジー研究の一環として、生物種、生物が生合成し、環境中に放出する揮発有機化学物質 (VOC)、VOC の他の生物へ与える活性の三組の情報から法則性をマイニングすることを検討した。現在までに、科学論文で報告されている VOC を、上記の三組の情報を収集し、データベースを構築した。本データベースは、KNApSAcK Ecology DB として整理し公開した (<http://kanaya.naist.jp/MetaboliteEcology/top.jsp>)。本データベースには 8285 エントリーの VOC と生物活性の関係情報が含まれている。KNApSAcK Ecology DB を活用し、教師なし機械学習の一方法である、二次元クラスタリング、教師ありの機械学習としては、多層構造のニューラルネットワーク (DNN)、Gradient Boosting Machine (GBM)、ランダムフォレスト法 (RF)、一般線形モデル (GLM) を活用し、VOC の化学構造と生物活性の関係を把握することを試みた。これらの結果を総合的に解釈し、VOC の化学構造の特徴により生物活性を説明することに成功した。また、化学構造から VOC の生物活性の予測についての可能性についても示唆された。生物が放出する揮発物質と生物活性に焦点をあて、バイオインフォマティクス要素技術により解決した論文であり、情報科学とバイオサイエンスの境界領域の発展に貢献するものである。よって、本論文は、博士(工学)の学位論文としての価値があるものと認める。