

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	計算システムズ生物学 (金谷 重彦 (教授))		
学籍番号	2111105	提出日	令和 4年 7月 28日
学生氏名	小出 ゆり		
論文題目	機械学習に基づくterpenoids、phenylpropanoids、polyketide化合物の代謝経路による代謝物の分類		
要旨			
<p>Terpenoids、phenylpropanoids、polyketidesは炭素、水素、酸素を含む二次代謝産物の大部分を占めている。本研究では、KNApSAcK Core DBに蓄積された19,769種類の代謝物を、学術文献に基づき3大グループ(Terpenoids、phenylpropanoids、polyketides)の73サブグループに分類した。代謝物を化学的性質を含む分子Fingerprintとして表現し、その記述子を用いてRandom Forestモデルによる分類を行ったところ、トレーニング代謝物は94.06%、テスト代謝物は94.23%の精度でサブグループに分類できた。代謝経路に基づく代謝物の分類は煩雑な作業であるが、分子Fingerprintを用いた機械学習により、分類を達成できることが明らかになった。本研究は、二次代謝経路に基づく分岐した二次代謝産物の系統的・進化的理解に繋がると考えられる。データサイエンスは、統計学、数学、コンピュータサイエンス、情報科学などの技術や理論を用いた学際的な応用分野である。これらのリソースを組み合わせることで、二次代謝産物に関する有意義で実用的な知見を抽出することが可能になる。</p>			