

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	2011163	提出日	令和 4年 1月 20日
学生氏名	田口 雄大		
論文題目	非局所的な部分構造の共起を用いた天然物由来の部分構造の分散表現の学習		
要旨			
<p>植物や微生物あるいは海洋生物などの生物は生合成経路を経て天然物を生合成している。これらの天然物はその化学構造によって様々な生物学的機能を有している。このような天然物の化学構造と生合成経路の関連性を調べることは、天然物の生物学的機能の理解に繋がる可能性がある。Mol2vec は、分子の部分構造のベクトル空間への埋め込みを最適化する手法であり、生合成経路上での分子間の距離を埋め込み空間における類似性として評価することができる。しかし、データベースに登録されている化合物は、研究者の興味に偏って収集されているため、埋め込み類似度にバイアスがかかることがある。</p> <p>そこで本研究では、非局所的な部分構造の共起を考慮してMol2vecの埋め込みを改善することを提案する。その結果、生合成経路に含まれる分子の生物学的特性の情報を利用することで、生合成経路上での近さを埋め込み空間における距離として反映した埋め込みが可能であることが分かった。また、5-近傍探索により、生物学的特性の観点から関連する部分構造を得ることができることを示した。提案手法を用いることで、生合成経路において関連性のある部分構造の埋め込みを得ることができ、部分構造の埋め込み表現の類似性から、天然物の部分構造と生合成経路の関係を理解する手がかりを得ることが可能だと考えられる。</p>			