



Title	メタボロミクスのための高性能ピークピックアップと キーワードレコメンド法の開発
Author(s)	金澤, 慎司
Citation	大阪大学, 2021, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/82297
rights	©2021. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論文内容の要旨

氏 名 （ 金 澤 慎 司 ）	
論文題名	メタボロミックスのための高性能ピークピッキングとキーワードレコメンド法の開発
論文内容の要旨 メタボロミックスは、生物試料中に含まれる多数の代謝中間体含量を測定する技術である。生体中の代謝中間体含量には、遺伝的な要因に加え、外部環境が反映するため、医学分野では、疾患群と健常者群の比較からがん早期診断のためのバイオマーカー探索などに用いられている。食品分野でも品質の客観的な評価を可能とするなど、メタボロミックスは様々な分野で重要な役割を担っている。 メタボロミックスは、主に4つの技術 [1. データ取得、2. 生データ処理、3. データ解析、4. データ解釈] から構成される。このうち、液体クロマトグラフィー質量分析装置 (LC-MS) で得られた信号データからピーク面積を決定する (ピークピッキング) 作業では、自動ピークピッキング法の開発が進められてきた。しかし、精度が熟練作業レベルに達していないため、確認・手動修正に数時間から半日程度の時間が費やされているという課題がある。データ解釈では、含量が増減した代謝物のリストから、関連するキーワードを連想する作業が行われるが、「解釈担当者」の知識に大きく依存し、文献調査に日から週スケールの作業が必要であることも深刻な問題となっている。そこで、本研究は 1) 自動ピークピッキング作業の正確さを熟練作業並みとし、修正作業を不要にすること。2) キーワード連想作業を自動化するキーワードレコメンド手法を開発して、キーワード連想作業の属人性を解消し、作業時間を短縮することを目標とした。 学位論文は、第1章から第4章より構成される。第1章では、本研究の背景と目的、論文の構成について記述した。第2章では、正確にラベル付けされた学習データを生成する疑似クロマトグラム生成器を開発し、生成した疑似クロマトグラムを用いてピークピッキングニューラルネットワークの深層学習を行った。その結果、ピークピッキングニューラルネットワークは初めて熟練者と同等の性能を示し、生データ処理にかかる時間は実質ゼロになった。 第3章では、第2章で得られた「代謝物リスト」から、関連度の高いキーワードを提案する手法を開発した。アンシエーション分析の手法を活用することで、代謝物リストと「解釈担当者の既知のキーワード」から、統計的に有意な「解釈担当者の知識外のキーワード」をレコメンドする新たな手法を開発した。これにより、日から週スケールの調査時間が約1.5時間になった。 第4章では、本研究の結果を総括し、今後の展望について述べ、結言とした。	

論文審査の結果の要旨及び担当者

氏 名 (金澤 慎司)		
	(職)	氏 名
論文審査担当者	主 査	教授 松田 史生
	副 査	教授 清水 浩
	副 査	准教授 戸谷 吉博
	副 査	准教授 瀬尾 茂人

論文審査の結果の要旨

本論文は、生物試料中に含まれる多数の代謝中間体含量を測定するメタボロミクスの高性能化、省力化を目指したものである。その実現のために、ピークピッキングニューラルネットワークの開発に加え、メタボロミクス分析結果からの仮説生成を支援する技術の開発を志向している。メタボロミクスは、主に4つの技術 [1. データ取得、2. 生データ処理、3. データ解析、4. データ解釈] から構成される。これまでに、1. データ取得および3. データ解析部分は高性能化、省力化が行われてきた。本論文では、2. 生データ処理について、自動ピークピッキング作業の正確さを熟練作業員並みとし、修正作業を不要にすること、4. データ解釈について、キーワード連想作業を自動化するキーワードレコメンド手法を開発して、作業の属人性を解消、省力化することを目標としている。

第1章は序論である。本研究の背景と目的について述べている。

第2章では、液体クロマトグラフィー質量分析法で得られた信号データから、ピーク面積を決定する(ピークピッキング)作業を担う自動ピークピッキング法の性能が低く、確認・手動修正が必要とされている課題の解決を行っている。まず、正確にラベル付けされた学習データを生成する、疑似クロマトグラム生成器を新規に開発している。ついで、開発した生成器で生成した疑似クロマトグラムを用いてピークピッキングニューラルネットワークの深層学習を行い、学習後のピークピッキングニューラルネットワークが熟練者と同等の性能を持つまで高性能化したことを示している。また、これにより、生データ処理にかかる時間を実質なしにするまで省力化したことを示している。

第3章では、データ解釈段階において行われる、含量が増減した代謝物のリストから、関連するキーワードを連想する作業が、「解釈担当者」の知識に大きく依存し、文献調査に労力が必要とされている課題の解決を行っている。まず、「代謝物リスト」から、関連度の高いキーワードを提案する手法を開発している。アソシエーション分析の手法を活用することで、代謝物リストと「解釈担当者の既知のキーワード」から、統計的に有意な「解釈担当者の知識外のキーワード」をレコメンドする新たな手法を開発している。これにより、日から週スケールの調査時間を大きく省力化しうることが示している。

第4章では、本研究で得られた知見をまとめ、展望について述べている。

このように、本論文ではメタボロミクスを構成する4つの技術のうち、ピークピッキングニューラルネットワークの高性能化(第2章)、データ解釈作業におけるキーワード連想の省力化(第3章)に初めて成功した。本研究をもとに、メタボロミクス技術の高性能化、省力化が可能になれば、医学分野でのバイオマーカー探索、食品分野での客観的な品質評価等の実現につながることが期待されることから、情報科学の発展に寄与することが期待される。したがって、本論文は博士(情報科学)の学位論文として価値のあるものと認める。