

「科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進
のための次世代アプリケーション技術の創出・高度化」

H29 年度
実績報告書

平成 25 年度採択研究代表者

船津 公人

東京大学大学院工学系研究科
教授

医薬品創薬から製造までのビッグデータからの知識創出基盤の確立

§ 1. 研究実施体制

(1)「船津」グループ(東京大学)

① 研究代表者:船津 公人(東京大学大学院工学系研究科 教授)

② 研究項目

・運転監視・プロセス制御のための知識抽出

* Iterative Optimization Technology (IOT) による calibration free による製品
品質予測手法の開発

* 製品製造時における主製品、副製品の製造制御を目的とした、多入力多出力シ
ステムの精度向上

(2)「奥野」グループ(京都大学)

① 主たる共同研究者:奥野 恭史 (京都大学医学研究科 教授)

② 研究項目

・ケミカル情報とバイオ情報の統合化と高速処理を可能にするデータ構造とアルゴリズムの開
発

・ケミカル情報とバイオ情報の相互作用ビッグデータ解析を可能にする数理的モデルの開発

(3)「泰地」グループ(理化学研究所)

① 主たる共同研究者:泰地 真弘人 (理化学研究所生命システム研究センター コア長)

② 研究項目

・ 仮想大規模ライブラリの拡充

・ 超大規模ライブラリからの有用情報計算技術の開発

- ・ 超大規模仮想ライブラリのコンテンツ可視化技術の開発
- ・ グループ間連携実証プラットフォームの開発(→H29 年度計画書「研究実施項目」に記載)

(4)「堀」グループ(山口大学)

①主たる共同研究者:堀 憲次 (山口大学大学院創成科学研究科 教授)

② 研究項目

- ・QMRDB(化学反応に関わる量子化学計算結果のライブラリ)のデータ収集
- ・QMRDB データを用いた TSDB (遷移状態データベース) の開発
- ・TSDB に関するプログラム開発

(5)「杉本」グループ(北海道大学)

①主たる共同研究者:杉本 雅則 (北海道大学大学院情報科学研究科 教授)

②研究項目

- ・ビッグデータ応用領域研究成果体験型ポータルの開発

§ 2. 研究実施の概要

(1)「船津」グループ

通常、ソフトセンサーによる生産プロセスの状態監視は、calibration model を構築することで実現するが、モデル構築のためのデータ取得が難しい場合やコストの問題から、Calibration-free のソフトセンサーの実現が求められている。今年度はこのための手法として Iterative Optimization Technology (IOT) による手法を開発を実施した。この技術により、例えば医薬品錠剤製造プロセスにおいてスペクトル情報のみからの有効成分含量などの高精度推定が可能となった。この成果は、S. Shibayama, H. Kaneko, K. Funatsu, Formulation of excess absorption in infrared spectra by numerical decomposition to allow effective process monitoring, *Computers & Chemical Engineering*, 113, 86-97, 2018.において公表されている。

また、ソフトセンサーを活用することで効率的なプロセス制御が可能となるが、製品製造時における主製品、副製品の製造制御を目的に、複数の複数の制御変数を複数の操作変数により制御するシクムの開発に取り組んだ。

(2)「奥野」グループ

奥野グループは、「大量のタンパク質 対 化合物情報からの創薬指針の抽出」を担当している。平成 29 年度では、開発項目 1)ケミカル情報とバイオ情報の統合化と高速処理を可能にするデータ構造とアルゴリズムの開発と、3)ケミカル情報とバイオ情報の相互作用ビッグデータからの創薬指針の抽出を行った。具体的には、バイオ関連情報を保有する化合物関連データベースを引き続き、網羅的に収集した。また、Deep learning による化合物-タンパク質相互作用の予測器を用いた新規化合物に対する標的タンパク質予測や活性予測および、創薬指針となる創薬標的タンパク質の同定と医薬品候補化学構造の抽出を行なった。

•Kei Terayama, Hiroaki Iwata, Mitsugu Araki, Yasushi Okuno, Koji Tsuda. “Machine Learning Accelerates MD-based binding pose prediction between ligands and proteins” *Bioinformatics*. 34(5), pp770-778, 2018

•Nakaoku, T., Kohno T., Araki M., Niho, S., Chauhan, R., Knowles, P.P., Tsuchihara, K., Matsumoto, S., Shimada, Y., Mimaki, S., Ishii, G., Ichikawa, H., Nagatoishi, S., Tsumoto, K., Okuno, Y., Yoh, K., McDonald, N.Q., Goto, K. “A secondary RET mutation in the activation loop conferring resistance to vandetanib” *Nature Communications* 9(1) : 625, 2018.

(3)「泰地」グループ

創薬過程の初期段階において重要な役割を担う大規模仮想化合物ライブラリの開発と高度化を目的とする。数十億規模の化合物を含む仮想ライブラリを構築し、同時にこれら全ての化合物についての合成経路情報を保持することで、より広大な化合物空間からの有用化合物の探索や、その化合物の合成計画立案時の基礎情報提供を可能とするシステムを構築する。これまでに開発した化学構造創出エンジン、大規模化合物データからの効率良い情報検索技術、および視覚的にラ

イブラリ内容物を把握するための可視化システム等の関連技術の改良を進めた。また、他グループとの連携を行いチーム全体での知識創出基盤としてのプラットフォームを開発した。

(4)「堀」グループ

本プロジェクトは、(1)薬物候補化合物ライブラリの構築、(2)候補化合物のディープラーニング手法を用いたスクリーニング、(3)候補化合物の製造プロセスのコントロール、に関する3つのサブプロジェクトで構成されている。我々のグループの役割は、(2)と(3)の間をつなぐ、提案された合成法で薬物候補化合物が合成可能かどうかを、理論計算により検証することを目的としている。

化学反応の進行の可能性は、(a)遷移状態(TS)の有無、(b)反応の活性化自由エネルギーの高さ、等から理論計算により評価することができる。そのため、薬物候補化合物の合成経路は、最終的には合成実験が必要となるものの、コンピュータのみを用いた理論計算により判断できる。我々は、合成経路の妥当性、副反応評価などについて定量的かつ迅速に判断を行うための手法(*in silico* スクリーニング)の開発を行ってきた。本グループ研究では、この手法の適用性を拡大する目的で、化学反応に関する量子化学計算結果を集めたライブラリ(QMRDB)の充実と、それを活用した遷移状態データベース(TSDB)の構築と、それらにかかわるプログラム開発を行った。

(5)「杉本」グループ

体験型ポータルシステムのアーキテクチャの検討と本採択課題に対応したポータル開発のための開発課題事前調査、ポータルの試作開発を前年度より継続して行った。

体験型ポータルとは開発された個々の分析手法やそれらを組み合わせたデータ分析過程が再利用可能な形式でライブラリに登録されていて、ユーザは、個々のプロジェクトが公開した分析過程を体験できるだけでなく、その一部の分析手法を別の手法に取り替えたり、別の分析過程と組み合わせたりして、個々の手法や分析のユースケースが知識発見にどのように寄与するかを体験できることを実現するものである。他のプロジェクトや、ビッグデータ基盤技術の研究領域のプロジェクトで開発された手法やユースケースの再利用ができるようにすることができるとなると嬉しい。これにより、個々のプロジェクトにおいても、新しいデータ分析過程の創成が促進されると期待できる。

ビッグデータ応用研究領域の成果が、個別のプロジェクトの成果の単なる羅列に終わることなく、共通の応用技術の研究開発の促進に貢献できるように、本領域の成果を、サンプル・データセットも含めて、誰もが個々の分野における知識発見のためのデータ分析過程を体験できるような体験型ポータルとして公開することを目指す。今年度は船津プロジェクト、吉田プロジェクトのポータルの維持管理と、三好プロジェクトのポータルの継続開発を行ったことに加え、新しく2プロジェクトのポータルの開発を目標として、越村プロジェクトと西浦プロジェクトの体験型ポータルの開発課題事前調査と、ポータルの試作開発を行った。