

「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」

H28 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

- 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
  - ・モデルの自動的メンテナンス手法の開発
  - ・運転監視・プロセス制御のための知識抽出

### (2)「奥野」グループ

- ① 主たる共同研究者：奥野 恭史（京都大学大学院医学研究科、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

- ① 主たる共同研究者：泰地 真弘人（国立研究開発法人理化学研究所 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

(4)「堀」グループ

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

- ・ QMLB(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
- ・ QMLB 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

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 研究題目

- ・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

産業プラントにおいては測定困難なプロセス変数をリアルタイムに推定する手法としてソフトセンサーが広く使用されている。前年度に開発した手法により適切なモデルが自動的に構築されたとしても、モデルはプロセスの状態変化によって予測性能が低下してしまう。この性能低下を防ぐために様々な種類の適応型モデルが開発されている。

本研究では様々なプロセス状態およびプロセス状態の変化に対応するため、複数の非線形回帰モデルを組み合わせて予測する開発した。各モデルはそれぞれ特定の時間のプロセス状態に特化したモデルであり、予測したい時刻の状態に応じてモデルの重みを決定することで様々な状態において高精度な予測が可能となった。

これまで開発したソフトセンサーを活用したプロセス制御手法は、単入力単出力システムに適用範囲が限られていた。本研究では多入力多出力システムにおいて出力変数の値を入力すると入力変数をどのように変化させれば効率的に制御可能か計算するシステムを開発した。

### (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 28 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、2)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、独自のデータベースを構築した。また、化合物-タンパク質-フェノタイプ(表現型)の薬理活性の発現メカニズムを意識した相互作用解析・予測のための数理モデルの構築を実現した。さらには、Deep learning による新規化合物に対する標的タンパク質予測や活性予測のための化合物-タンパク質相互作用の予測器を完成させ、新規化合物の予測を実施した。

### (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。そのために必要となる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を把握するための可視化システムといった関連技術の開発を行った。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構築を開始した。

### (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(QMLB)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

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し、昨年度開発を行った、船津グループの体験型ポータルを研究の発展に応じて拡張し、維持管理を行った。新たに、吉田グループの体験型ポータルの第1版を開発し、三好グループの体験型ポータルの第1版 $\alpha$ 版を開発し、各々公開可能にした。