

「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」

平成 25 年度採択研究代表者

H27 年度
実績報告書

船津 公人

東京大学工学系研究科
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 (東京大学工学系研究科、教授)
- ② 研究項目
 - ・プラント運転モニタリングのための自動的モデル構築システムの開発
 - ・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(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 (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、副センター長)
- 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検索技術の開発
 - ・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(山口大学)

① 主たる共同研究者:堀 憲次(山口大学理工学研究科、教授)

② 研究項目

- Laueらのテキストにある人名反応から30種類を選び、それらの反応解析を、2種類以上の化合物について計算を行い、結果を登録する。その際、基本構造について検討を行う。
- DB データ自動登録プログラム(aRegistration)を実装する。
- 類似反応を検索するプログラム(cStructure)を実装する。
- TS 初期構造自動生成プログラム(iStructure)のためのアルゴリズムを開発する。
- 前年度に構築した「合成経路スクリーニングクラウドシステム」の拡充・整備を行う。
- TS Technology(株)においてQMRDB 及びTSDBの準備ができ次第、有料で公開を開始する

§ 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。前年度に開発した自動的にソフトセンサーモデルを構築するシステムについて、インプットデータがノイズの強いデータの場合でもスムージング手法を駆使することで良好な予測を行うことに成功した。あるデータベースが与えられた際に自動的にノイズ処理を行った後に予測的なソフトセンサーモデルを構築する手法を開発した。

上述のシステムにより適切なソフトセンサーモデルが構築できたとしてもプラントの状態変化によって性能が劣化してしまう。この劣化に追従する適応型モデルの性能をモニタリングするため、適応型モデルの予測性能およびそのモデルを適応できるデータ範囲を推定する指標を開発した。

プラントの安全かつ安定的な運転のためにはプラントの運転状態を監視しなければならない。プラントでは膨大な測定データが蓄積されていることからデータに基づいたプロセス管理に着目し、本年度はプラントにおいて測定されたデータを入力することで異常の検出、異常の種類の特定、および異常に関与するプロセス変数の特定を行う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 27 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モデルの開発を行った。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網羅的に収集し、NoSQL 型の MongoDB 形式によって独自のデータベースを構築した。また、26 年度に引き続き化合物とタンパク質の相互作用データのビッグデータ化に対応できる新たな機械学習法として Deep Learning (深層学習) に基づく手法の開発を行い、本研究室が保有する 400 万件の化合物とタンパク質の相互作用データを用いた機械学習を実現した。さらには、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様々な数理モデルの検討・評価を実施した。

(3)「泰地」グループ

創薬過程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。ライブラリが含む化合物の数を数十億規模まで増加させ、同時にこれら全ての化合物について合成経路情報を付与することで、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。そのためには化学構造創出エンジンの改良、大規模化合物データからの効率良い情報検索技術の開発、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の開発を行うことが不可欠となるため、これら当該研究を実施した。

(4)「堀」グループ

- (1) Laue らのテキストにある人名反応から 30 種類を選び、それらの反応解析を、2 種類以上の化合物について計算を行い、結果の登録をおこなった。
- (2) DB データ自動登録プログラム(aRegistration)を実装した。
- (3) 類似反応を検索するプログラム(cStructure)を実装した。
- (4) TS 初期構造自動生成プログラム(iStructure)のためのアルゴリズムを開発するとともに、半自動的に入力を作成する iStructure ver.1 を完成させた。
- (5) 研究加速化経費により、「合成経路スクリーニングクラウドシステム」に高速演算サーバ群の設置を完了した。それに伴い、Gaussian09 プログラムに特化したスケジューラを実装した。
- (6) TS Technology(株)において QMRDB 及び TSDB の準備ができ次第、有料で公開を開始した。