

「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」

H25 年度
実績報告

平成 25 年度採択研究代表者

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

- 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
 - ・ 製造プラントにおける大規模運転データベースの管理手法の開発
 - ・ 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 (京都大学薬学研究科、教授)
- ② 研究項目
 - ・ バイオ関連情報を保有する化合物データの網羅的収集

(3)「泰地」グループ

- ① 主たる共同研究者: 泰地 真弘人 ((独)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

§ 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。我々はソフトセンサー構築に用いられるデータベースを管理するための指標 database monitoring index (DMI) および DMI を用いたデータベース管理手法を開発している。今年度は DMI におけるパラメータを最初のデータベースから自動的に決定する手法を考案した。本手法を用いてシミュレーションデータ解析および実プラントデータ解析を行った結果、DMI により適切なデータベース管理が可能であり、ソフトセンサーモデルの予測性能が向上することを確認した。

プロセス産業では製品の品質改善や生産性の向上が求められており、そのためには設定値変更や外乱に対して迅速かつ安定に制御を行わなければならない。我々は日常の運転データから得られる情報を利用してより効率的なプロセス制御を行うことを目的として、ソフトセンサーモデルとその逆解析を利用した制御手法である inverse soft sensor-based feed forward 制御法 (ISFF) を提案した。continuous stirred tank reactor (CSTR) のシミュレータを用いたケーススタディを行い、目的変数の設定値変更に対して提案手法が有効であることを示した。

(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。今年度は、開発項目「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」における「バイオ関連情報を保有する化合物データの網羅的収集」を行った。

具体的には、世界中で公表されている 14 種のデータベースからバイオ関連情報を保有する化合物データの網羅的収集を行い、本研究全体におけるデータソースの準備を行った。これらのデータの一部は、化合物情報、標的タンパク質情報、その他のバイオ情報の 3 種に分けて、<http://cgs.pharm.kyoto-u.ac.jp/ddida/>より公開している。

さらに、ビッグデータ規模のケミカル情報とバイオ情報の高速検索に対応できるデータ構造の検討を行い、多種多様なバイオ情報に対応するために、NoSQL 型のデータ構造を採用し、データベース構築を行った。

また、スーパーコンピュータ「京」を用いて、EBI/ChEMBL に登録されている化合物 195,522 に対して、ビッグデータ規模の相互作用予測(タンパク質 892 との組合せ総数 174,405,624)を行った。予測結果は、論文誌 Molecular Informatics に発表し、計算結果は <http://pharminfo.pharm.kyoto-u.ac.jp/services/cross-family/>より公開している。

(3)「泰地」グループ

大規模仮想化合物ライブラリの高度化を目指して次の三つの項目、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、について研究を実施している。今年度は特に次の二つの項目について研究を推進

した。

仮想大規模ライブラリの拡充にあたり、現在当グループにおいて化学構造情報産出の中心となっている東京大学船津研究室にて開発された化学構造創出エンジンに更なる高度化に向け、化学構造創出エンジンプログラムの内容解析と、大規模創出時の改善点についての検討を行った。

超大規模ライブラリからの有用情報検索技術の開発については、大量の情報から合理的な時間で有用な情報を利用者へ提示する手法を開発する必要があるため、本年度は基本データベース作成と部分構造検索機能のプロトタイプ構築を行った。また、今後の開発に向けて高速ストレージ装置の選定・調達を行った。また、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についても、基礎的な検討を開始した。

§ 3. 成果発表等

(3-1) 原著論文発表

論文詳細情報(国際)

1 J.B. Brown, Satoshi Niijima and Yasushi Okuno, “Compound-protein interaction prediction within chemogenomics: theoretical concepts, practical usage, and future directions” *Molecular Informatics*, 32, 906-921, 2013. (DOI: 10.1002/minf.201300101)