



CBI学会2019年大会、2019年10月21日(月)、13:00-17:00

〈チュートリアル〉
計算毒性学と
化学データサイエンスの基本

株式会社 インシリコデータ
湯田 浩太郎

■本日のフォーカストセッションの構成内容方針

1. 化学分野でのデータサイエンス実施上での留意点の討論
 2. 化学とコンピューターの融合に関する討論
 3. データサイエンスやAI等の適用に関する留意事項の討論
- 正しく且つ精度高く討論できることを目指す
 - 技術的な基本を広範囲に討論するものではない

例えば

車のメカニクではなく、運転して目的地に確実に着くことを目指す

ターボ、4駆、インバーター、他

道路交通法、地図、安全運転、他

本日のプログラム:

1. 13:00-13:05: (5分) 挨拶:株式会社 インシリコデータ 湯田浩太郎

2. 13:05-13:20(15分) ◆導入 計算毒性学と「化学データサイエンス」

計算毒性学でのコンピューター導入原理、二大毒性評価関連技術(化学多変量解析/パターン認識アプローチ、人工知能アプローチ)、データサイエンスから「化学データサイエンス」へ

3. 13:20-13:50(30分) ◇第一部 計算機化学(Computer Chemistry)関連

化合物保存形式、化合物命名法、化合物検索(完全一致、部分構造、2・3次元構造検索、他)手法、一元一項対応串刺し検索、化合物の扱い(縮合多環、互変異性、立体/幾何異性)、化合物表記(ケトエノール、ニトロニトロソ、他)

4. 13:50-15:20(90分) ◇第二部 化学多変量解析/パターン認識(ケモトリックス(Chemometrics))関連

化学パラメーター、2/3次元パラメーター、種々データ解析手法、過剰適合、偶然相関、線形/非線形性、特徴抽出、最少サンプル数、最少パラメーター数、クラスポピュレーション、次元変換/圧縮/縮小、分類率/予測率、要因解析、オートスケーリング、アウトライヤー/インライヤー、解析信頼性指標(サンプル数/パラメーター数)、KY(K-step Yard sampling)法、パーセプトロン、バックプロパゲーション、遺伝的アルゴリズム、ファジー理論、内挿/外挿問題、他

<15:20-15:40 休憩 20分>

5. 15:40-16:20(40分) ◇第三部 人工知能(Artificial Intelligence)関連

人工知能の歴史、ルールベース型人工知能、ニューラルネットワーク型人工知能、深層学習、サンプル数問題、要因説明問題、ルールのコンピューターへの組み込み、ネットワーク構造、LISP、FORTRAN、PYTHON、

6. 16:50-17:00(30分) ◇第四部 計算機科学(Computer Science)関連

データベース理論、プログラミング言語、クラスター、クラウド、スーパーコンピューター、ネットワーク、WEB、他

7. 16:50-17:00(10分) ◇討論および名刺交換会

挨拶: 株式会社 インシリコデータ 湯田浩太郎

基礎学習 研究期間

■経歴

1979年3月 東北大学薬学部大学院薬学研究科 薬品合成化学 博士過程終了
博士論文「抗腫瘍性多環式アルカロイドの合成研究」

1979年7月～1981年8月

米国ペンシルバニア州立大学化学科 P.C.Jurs研究室 リサーチアソシエイト
研究テーマ「パターン認識による発癌性予測」

1981年9月 豊橋技術科学大学 第5工学系

研究テーマ: 「NMRデータベースの構築」

1983年12月～2009年6月

富士通株式会社 化学・製薬関連企業研究所の研究システム構築とコンサルタント
2009年7月～2010年3月

(独) 国立環境研究所勤務 安全性予測システム関連業務に従事

2010年5月28日

株式会社 インシリコデータ 設立

取締役社長に就任

適用期間

適用分野

■専門および得意な分野・研究

- ・ 医薬化学 (Medicinal Chemistry)
- ・ 構造-活性相関 (Structure-Activity Relationships)
- ・ 化学多変量解析//パターン認識 (Chemical Multivariate & Pattern Recognition)
- ・ 計算機化学 (Computer Chemistry)
- ・ 計算毒性学 (Computational Toxicology)

富士通勤務時代の業務(1)

1983年から2009年(26年間)

◇コンピューターケミストリー関連業務

■コンピューターケミストリー関連分野での主たる業務

1. 化学関連システムの開発/設計/販売/教育/メンテナンス実施および支援業務
 - ・ CADD (Computer Aided Drug Design) 関連システム構築、教育支援
 - ・ ケモメトリクスおよび構造-活性相関支援システム構築、教育支援
 - ・ 化合物/反応等のデータベースシステム(MACCS/REACCS: 現ISIS)の汎用機への移植
2. データ解析支援業務
 - ・ ADAPT, TSAR, DIWA, SARD, **ADMEWORKS**システムを用いた構造-活性相関解析支援
 - ・ ADAPT, TSAR, **MODELBUILDER**システムを用いたケモメトリクス解析支援
 - ・ TOPKAT, **PREDICTOR**システムを用いた化合物のADME/毒性予測支援
3. コンサルタント関連業務
 - ・ 計算機化学関連一般教育
コンピューターケミストリー全般(データベース、化合物表記法、他)
 - ・ コンピュータによる構造-活性相関教育/解析実施
Hansch-Fujita法、パターン認識による構造-活性相関、毒性/分解性/代謝予測、
コンビナトリアルケミストリー/HTS全般、インシリコドラグデザイン支援
 - ・ ケモメトリクス教育/解析実施
多変量解析/パターン認識による化学およびバイオ関連データ解析
4. 情報収集業務
 - ・ コンピューターケミストリー、化学関連データ解析の全般的情報収集
 - ・ 富士通製品の競合製品調査

◇化学関連システムの設計・開発関連業務

■化学関連支援システムの設計・開発および支援実績

1. パターン認識による構造-活性相関・ケモトリクス支援システム；
ADAPT (米国ペンシルバニア州立大学 (基本部分構築)、富士通 (GUI 部分構築))
2. 人工知能によるバイオアナログウス化合物創出システムの設計支援と立ち上げ；
EMIL (京都大学農学部との共同開発)
3. 生体内代謝予測支援システム構築 (北里大学薬学部との共同開発)
4. 合成ルート創出支援システム構築；CASINO (通産省物質研究所との共同開発)
5. 化学スプレッドシートシステム設計支援；KAMELO (田辺製薬様との共同開発)
6. 化学ツールキット構築；CHECK (富士通開発)
7. 3次元構造-活性相関支援システム構築；SARD (富士通開発)
8. J01Sオンラインシステムの化合物データ移行設計支援

◇バイオ／蛋白インフォマティクス関連業務

■バイオ／蛋白インフォマティクス関連の業務内容

- ★本業務は1999年10月より始めたものであり、多くの実績はない。
- ★現在行っている業務中心にまとめる。

1. バイオ（特に解析系）関連研究の全般的調査
2. ポストゲノムにおけるバイオ解析システムの検討／評価

- バイオ関連支援システムの設計・開発および支援実績

1. モチーフ探索システム設計（国立遺伝学研究所）；初期の参加のみ（1999年）
モチーフ探索へのパターン認識適用の可能性の検討
2. SNP¹解析関連業務
 - ・徳島大学医学部ゲノム解析センター 板倉光夫教授による重点領域研究「ヒト遺伝子多型系の体系的解析による疾患遺伝子の解析」に参加
 - ・鎌谷直之教授（東京女子医科大学膠原病リウマチ痛風センター センター長）との遺伝子多型解析システム開発関連作業に従事
3. 発現プロファイル解析関連業務
- ・NEDOプロジェクト
“遺伝子発現頻度解析（通称 高原プロジェクト）”への参加（2000年度）
発現プロファイルデータベースシステムと連携する薬理活性データベース構築

- 蛋白関連支援システムの設計・開発および支援実績

1. 蛋白シミュレーション関連業務
 - ・NEDOプロジェクト（2003より参加、2007年終了）
“生体高分子構造情報利用技術開発（通称 中村プロジェクト）”への参加
2. 創薬バリューチェーン（大阪大学坂田教授主催）への参加（2005より参加）
 - ・産官学の3共同体制で行う新規医薬品開発を目指したコンソーシアム

◇環境関連業務

■環境関連の業務内容

1. 生体毒性（発ガン性）関連研究；J u r s 研究室での研究
 - ・ ADAPTシステムを用いた多変量解析／パターン認識による発ガン性予測
2. 環境ホルモン特性に関する研究
 - ・ ADAPTおよびT s a rシステムを用いた環境ホルモン特性に関する研究
 - ・ 環境関連支援システムの設計・開発および支援実績；
 1. 財団法人化学物質評価研究機構のWEB上での毒性予測システムの設計および開発支援
 2. 財団法人製品評価研究機構および通産省（現；文部通産省）からの調査依頼
「諸外国における化学物質規制に関する構造－毒性相関の適用」に関する調査
 3. 環境研究所（筑波）との化学物質毒性予測に関するシステム開発共同研究開始
 4. 国立医薬品食品衛生研究所との共同研究
 - ・ ADMEWORKSで用いる化審法および業審法関連化合物規制予測モデルの開発
 - ① Ames 試験予測モデルの構築（TA98およびTA100）：2004年度
 - ② 染色体異常試験（小核試験）予測モデル構築：2005年度
 5. 大阪府立環境研様との、皮膚感作性予測に関する共同研究実施：2006年度
 6. 国立環境研究所様、環境毒性予測に関する共同研究及びシステム構築
 - ・ 2007年度；甲殻類遊泳阻害試験48時間EC50、藻類生長阻害試験72時間EC50
 - ・ 2008年度；LogP 推算式の構築支援
 7. （財）日本自動車研究所様への研究業務支援
 - ・ 廃棄ガスと発がん性の相関解析支援
 8. 厚生労働省、感作性物質分類小委員会での皮膚／気道感作性予測モデルの構築

◇医療関連業務

■医療関連の業務内容

1. 日本医科大学様との医療を目指したNMRメタボノミクス関連研究支援
2. 皮膚および気道感作性予測に関する支援作業：感作性物質分類小委員会（福井大学様）

本日のプログラム:

1. 13:00-13:05: (5分) 挨拶:株式会社 インシリコデータ 湯田浩太郎

2. 13:05-13:20(15分) ◆導入 計算毒性学と「化学データサイエンス」

計算毒性学でのコンピューター導入原理、二大毒性評価関連技術(化学多変量解析/パターン認識アプローチ、人工知能アプローチ)、データサイエンスから「化学データサイエンス」へ

3. 13:20-13:50(30分) ◇第一部 計算機化学(Computer Chemistry)関連

化合物保存形式、化合物命名法、化合物検索(完全一致、部分構造、2・3次元構造検索、他)手法、一元一項対応串刺し検索、化合物の扱い(縮合多環、互変異性、立体/幾何異性)、化合物表記(ケトエノール、ニトロニトロソ、他)

4. 13:50-15:20(90分) ◇第二部 化学多変量解析/パターン認識(ケモトリックス(Chemometrics))関連

化学パラメーター、2/3次元パラメーター、種々データ解析手法、過剰適合、偶然相関、線形/非線形性、特徴抽出、最少サンプル数、最少パラメーター数、クラスポピュレーション、次元変換/圧縮/縮小、分類率/予測率、要因解析、オートスケーリング、アウトライヤー/インライヤー、解析信頼性指標(サンプル数/パラメーター数)、KY(K-step Yard sampling)法、パーセプトロン、バックプロパゲーション、遺伝的アルゴリズム、ファジー理論、内挿/外挿問題、他

<15:20-15:40 休憩 20分>

5. 15:40-16:20(40分) ◇第三部 人工知能(Artificial Intelligence)関連

人工知能の歴史、ルールベース型人工知能、ニューラルネットワーク型人工知能、深層学習、サンプル数問題、要因説明問題、ルールのコンピューターへの組み込み、ネットワーク構造、LISP、FORTRAN、PYTHON、

6. 16:50-17:00(30分) ◇第四部 計算機科学(Computer Science)関連

データベース理論、プログラミング言語、クラスター、クラウド、スーパーコンピューター、ネットワーク、WEB、他

7. 16:50-17:00(10分) ◇討論および名刺交換会

□計算毒性学と化学データサイエンス

1. 計算毒性学でのコンピューター導入基本技術

二大毒性評価関連技術

- ①化学多変量解析／パターン認識アプローチ
- ②人工知能アプローチ

2. データサイエンスから「化学データサイエンス」へ

データサイエンスはデータ処理の汎用的な呼び方である。

適用分野により様々な関連技術が展開される。

例えば、化学分野での多変量解析／パターン認識は

「ケモメトリックス」と呼ばれている。

化学分野でのデータサイエンスは

「化学データサイエンス」

1. 計算毒性学でのコンピューター導入原理

□化合物毒性の特徴とコンピューター導入の困難性

①メカニズム不明／複雑：メカニズム中心の議論困難

発現メカニズム等複雑で、毒性の種類も多い

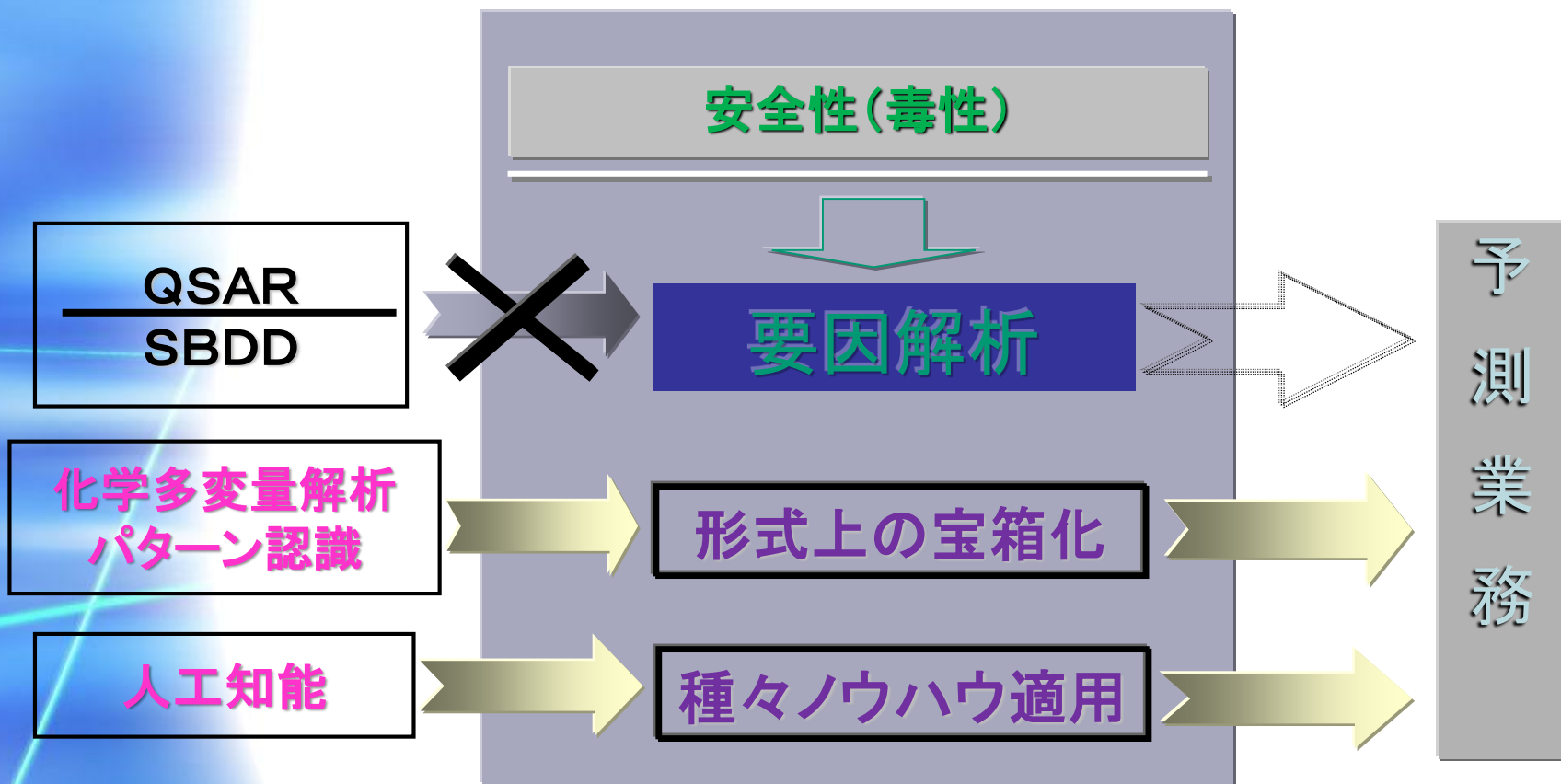
②化合物構造変化性極大：構造類似性による情報適用困難

メタンレベルからマクロライドまで、全化合物が評価対象

③高い予測率が要求される：予測結果重視の傾向

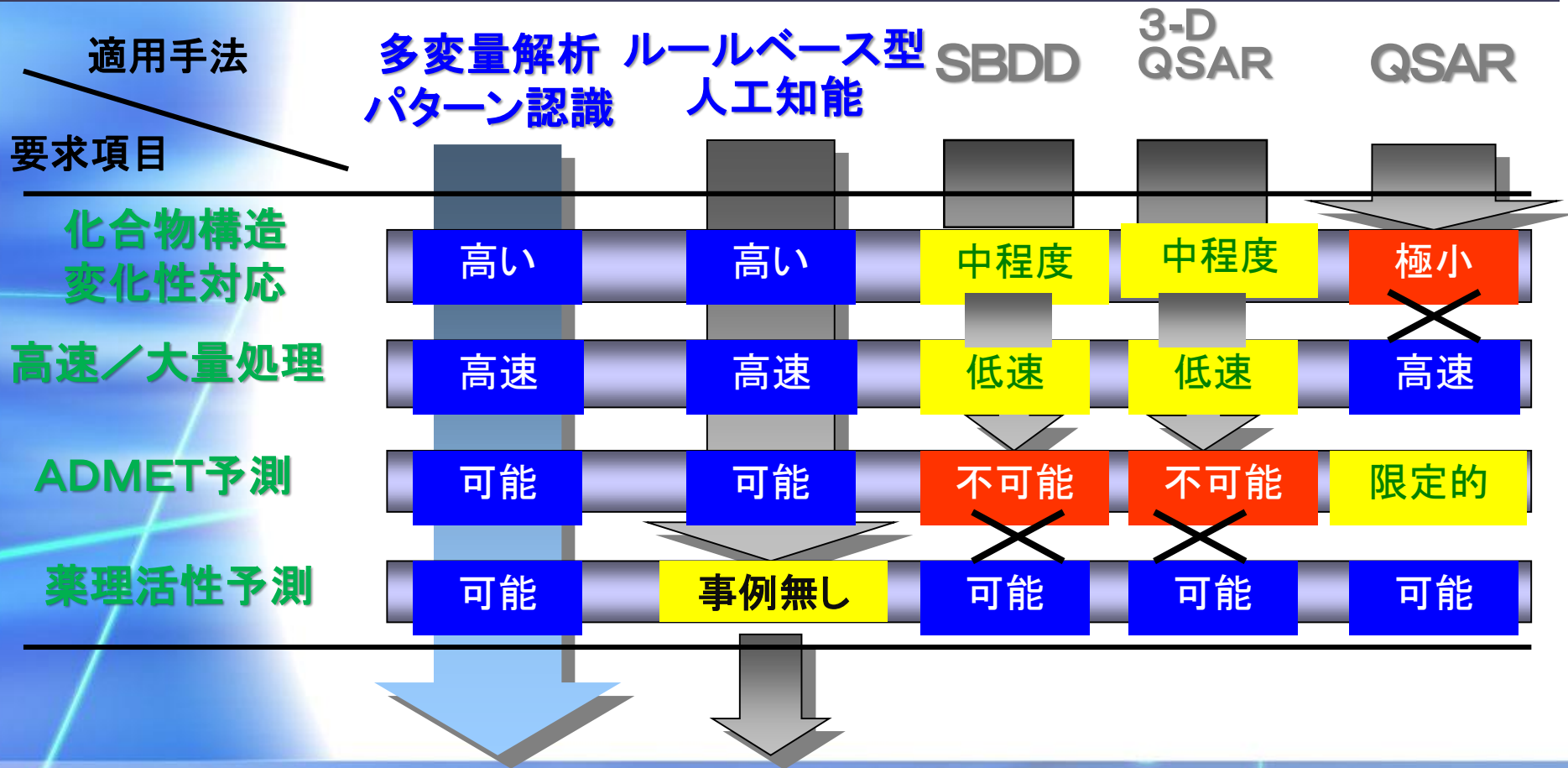
毒性予測は失敗が許されないなので、要求レベルが高い

コンピュータによる安全性予測へのアプローチ



□ 要求項目と適用手法の関係

安全性(毒性)、薬理活性、ADME予測上での 解決すべき要求項目と適用可能アプローチ



□ 二大毒性評価関連技術

毒性の特性より、毒性評価に適用可能なアプローチは適用内容や基本原理より以下の二種類となる

- ① 化学多変量解析／パターン認識アプローチ、
- ② 人工知能アプローチ

□ ICH-M7では、医薬品に含まれる不純物の遺伝毒性（変異原性）安全性評価に互いに相補的な

- ① データ解析アプローチと ② 人工知能アプローチの両方を採用するように求めている。

* 結果が矛盾する場合は、専門家によるレビューを実施

□データサイエンスから「化学データサイエンス」へ

「データサイエンス」はデータ処理の汎用的な呼び方である。
適用分野により様々な関連技術が展開される。
例) 化学分野での多変量解析／パターン認識は
「ケモメトリックス」と称される。

- ・化学分野でのデータサイエンスは「化学データサイエンス」
- ・創薬分野でのデータサイエンスは「**創薬**データサイエンス」
- ・毒性分野でのデータサイエンスは「**毒性**データサイエンス」
- ・物性分野でのデータサイエンスは「**物性**(マテリアルズ)データサイエンス」
- ・その他

□適用分野単位のデータサイエンス

基本技術

適用分野

名称

データサイエンス

多変量解析
パターン認識
AI

化学
創薬
毒性
物性
バイオ
蛋白
医療
環境
その他

化学
創薬
毒性
物性
バイオ
蛋白
医療
環境
その他

データサイエンス



ご清聴ありがとうございました

株式会社 インシリコデータ
湯田 浩太郎