

# ICT、IoTやビッグデータ時代の ケモメトリックス／人工知能を知って 新たなチャレンジを…

株式会社 インシリコデータ  
湯田 浩太郎

<http://www.insilicodata.com>

# 時代の新しい三大潮流

## **ICT** : Information and Communication Technology

(情報通信技術)

情報技術に通信コミュニケーションの重要性を加味した言葉

## **IoT** : Internet of Things (物のインターネット)

従来は主にパソコンやサーバー、プリンタ等のIT関連機器が接続されていたインターネットにそれ以外の様々な"モノ"を接続することを意味

## **BD** : Big Data (ビッグデータ)

従来のデータ処理アプリケーションで処理することが困難なほど巨大で複雑なデータ集合の集積物を表す

常に時代は変わる ⇒ しかも、後戻りはしない

新しい時代の環境変化を先取りする事が必要

**ICT, IoTおよびビッグデータは避けられない**  
これらがもたらす様々な変化

データ内容、品質、リアルタイム性、量、他

データベース、データ処理法、適用分野拡大／変換

上記変化に応じられる分析機器の展開  
機能／精度／処理量／時間／ネットワーク／他

## Chemometrics : ケモトリックス (化学計量学)

数理科学、統計学、機械学習、パターン認識、データマイニングなどの手法により、(広義の) 化学分野における諸問題を解決しようとする分野

---

## AI : Artificial Intelligence (人工知能)

コンピュータを使って、学習・推論・判断など人間の知能の働きを人工的に実現するもの

# 世界最初の・・・

## □ケモメトリックス研究支援システム

**ADAPT** (Automated Data Analysis by Pattern recognition Techniques)

化合物と薬理活性／毒性／物性等との  
相関や要因解析研究支援システム

---

## □本格的推論システム(人工知能)

**Dendral** (Dendritic Algorithm)

未知の有機化合物を質量分析法で  
分析し、有機化学の知識を使って特定

## □最近の人工知能システム

### ハードウェアの進歩が支えている面が大きい

- ・手法的にはニューラルネットワーク(N.N)
- ・機械学習を中心に展開されている
- ・多変量解析/パターン認識と関連する事も多い
- ・適用分野が広範囲にわたっている
- ・ICT, IoT, ビッグデータ等との連携や活用を目指す

## □以前の人工知能システム

### 高度な知識をルール化して問題解決を目指す

- ・推論型のアプローチを取る(エキスパートシステム)
- ・研究者が有する既存のノウハウ活用が主たる目的

# 化学分野で現在展開されている人工知能システム

□歴史的に化学関連分野への人工知能適用の歴史は長い  
化学分野では数式に乗らない事項が多く、経験則が重要となることが多い⇒人工知能が活躍する地盤がある

□適用事例は多い

- ・機器スペクトルデータの解析支援システム
- ・有機合成支援システム
- ・毒性予測システム
- ・構造-活性相関支援システム
- ・創薬化学者支援システム
- ・その他

---

従来より展開されてきた化学分野の人工知能システムは、その展開上化学的なノウハウや考え方等のアナログ的な内容を、デジタルに変換する事が必要

## □最近の人工知能は機械学習がメインである

### 利点:

- ・大量のデータを扱える
- ・従来は人工知能で展開出来なかった内容を展開できる
- ・ノウハウ(ルール)等を必要としない:データがあれば良い  
ノウハウがない分野での展開が可能となる
- 新たな知見を発見出来る可能性がある

### 欠点:問題点

- ・化学的な知見をシステムに理解させられるか?
- ・結果のフィードバックが手法的に困難
- 新たな知見を人間が解釈できるレベルへの具象化が困難

## □過去に開発した人工知能システム

インシリコデータの湯田は富士通時代に以下のシステムの開発に関与し、推論型人工知能システムを開発した

- **EMIL (京都大学); 創薬支援システム**  
創薬研究者の化合物変換ノウハウをルールとして用いる
- **CASINO (産総研); 有機合成設計支援システム**  
スーパーコンピュータを用いて反応の組み合わせを計算
- **LogP推算システム (北里大学); 化学者のノウハウを適用**  
LogP計算するパラメータにノウハウを展開
- **毒性予測システム (CELI); WEB上での毒性予測**  
化学者のノウハウと多変量解析のハイブリッドシステム

## □過去に開発したケモメトリックス支援システム

インシリコデータの湯田は過去に以下のシステム開発に関与

### \* **ADAPT** : ペンシルバニア州立大学のJurs教授開発

(**A**utomated **D**ata **A**nalysis by **P**attern recognition **T**echniques)

世界初のケモメトリックス研究支援システム

米国ミニコン上で稼働⇒国内汎用機に移植

### \* **Model Builder** : (株)富士通九州システムズ開発

- ・上記ADAPTシステムのGUIを強化開発し、

- ・PC上に移植した

- ・新たに開発した最新の開発手法であるKY法等が組み込まれている

# インシリコ創薬への人工知能の部分適用事例

## マルチカテゴリースクリーニング実施事例

### 5種類の薬理活性に関する同時スクリーニング

1. CCK-A (Cholecystinin-A) 受容体拮抗剤; Malonamide誘導体、急性膵炎治療薬
2.  $Ca^{2+}$  遊離抑制作用; 1,5-benzothiazepine誘導体、降圧薬、抗狭心症薬
3. 抗潰瘍作用; 4-Bis[(aminoalkyl)amino]-9,10-anthracenediones誘導体
4. 抗菌作用; Quinolone誘導体
5. 抗アレルギー作用; Imidazo誘導体

### 薬理活性スクリーニング対象化合物群 (化合物ライブラリー)

人工知能システムEMILを用いて創出された、バイオアナログウス化合物群

### 薬理活性スクリーニング (多変量解析/パターン認識手法)

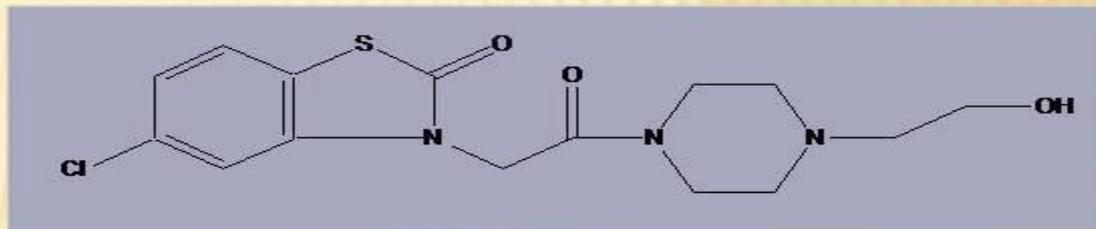
データ解析支援システムADAPTを用いて5種類の活性スクリーニングを行った

# 人工知能システムEMILを利用した バーチャルスクリーニング

## EMIL実行上での初期化合物と創出化合物群

### ■ 出発化合物構造式

Tiaramide  
(Antiasthmatic, Anti-inflammatory)



### ■ 適用した構造式変換ルール

医薬品関連変換ルール

### ■ 変換結果

第一次変換出力構造式数      78化合物

第二次変換出力構造式数      76化合物

(第一次変換化合物中の1化合物を出発化合物とした)

# 化学分野特有の問題

## 化合物構造式に始まり、化合物構造式に終わる

化学研究者の思考過程は化合物構造式で考え、  
相互コミュニケーションし、化合物構造式で答える。



人工知能システムが利用者である研究者と、  
化合物構造式で対話できることが必要

例：創薬研究者

薬理活性を強くするには、化合物構造式のどの部分を  
どのように変化させればいいのか？⇒研究者との対話必要

チェス、将棋、碁のように、盤上の座標を指定するようにはゆかない  
勝つだけで良いというわけでもない

## □化合物構造式に始まり、化合物構造式に終わる

・研究者の思考過程は総て化合物構造式で終始する

### ・化合物の表現の問題:

化合物名、分子式、二次元構造式、3次元構造式、等々  
同じ化合物が表現系により様々な形式を取り、それぞれの  
表現系が持つ情報の内容や情報量も異なる。



### ・入力の問題:

Journal や一般の化学文献が膨大な量あっても、単に  
文字情報を読み込ませただけで、化学情報を正確に人工  
知能システムに理解させることは困難

### ・結果の問題:

結果が出ても化学情報への変換が困難

## □人工知能による付加価値を付けたスペクトル機器の開発

スペクトル機器の機能や精度向上のみならず、  
人工知能との連携による新たな機能とは何であるかを追求

\* 従来型の競争市場から新市場への展開

## □ICT/IoT/ビッグデータ時代対応のスペクトル機器の開発

時代の変化により生じる新たな市場に適合する、  
新時代対応のスペクトル機器のあり方を追求

\* 新たな外部機器との連携機能の付加

\* 新分野にスペクトル機器を最適化する

## □従来および最新技術の俯瞰と適用および融合

多変量解析/パターン認識および人工知能の従来技術と  
新技術との差異を理解し、個別、融合等臨機応変に対応

## □化学分野特有の問題を認識しつつ新技術を適用

化合物構造式中心のアナログ文化を理解し、デジタルと融合する

**Thank you for your attention**

株式会社 インシリコデータ  
湯田 浩太郎

<http://www.insilicodata.com>