

序章 13：00～13：20

- ・自己紹介
- ・時代の流れと、研究環境および研究内容や研究スタイルの変化

第一章 13：20～13：40

化学データサイエンスおよび人工知能の適用イメージ

- ・化学データサイエンスおよび人工知能による研究を、自動車による移動に例えると最も大事なものは正しく／正確／確実に目的地まで移動する技術の獲得と実施である
- ・化学データサイエンスおよび人工知能の適用分野は、化学研究の広範囲に及ぶ
アイディア次第で、自分の研究を精度高く、最新の研究に導ける
- ・WET研究にDRY研究の良さを加え、最新の融合研究に道を付ける

第二章 13：50～15：00

化学データサイエンスおよび人工知能の適用事例：様々なアプローチが可能

第三章 15：10～15：50

化学データサイエンスおよび人工知能の適用手順やパターン

第四章 16：00～16：50

化学データサイエンスおよび人工知能の適用上での留意事項

- ①化学分野の留意点
- ②データサイエンス実施上での留意点

まとめと提案 16：50～17：00

- ①化学データサイエンスおよび人工知能のまとめ
- ②今後の展開についての「**オートノマス創薬**」の提案
 - ・KY法の展開（クラス分類と重回帰型）
- ③自由討論

今後の化学データサイエンスやAIの展開

- ・化学データサイエンスやAIの**発展は今後も続く**。
- ・これらの技術が現在の**自動車からTVに変化**してゆくことも避けられない事実である。
- ・自動車からTVに変化することで、従来は出来たことが出来なくなる等が発生する。
- ・自動車で出来ることは極力自動車を有効に使うことが必要である。
- ・研究環境、特に**IT環境の変化は自動車からTV化への変革を促す**。
- ・**TV化することで新たに可能となる研究や作業の形**も出てくる。
例えば、リアルタイム創薬等の「**オートノマス創薬**」は評価や予測がTV化されることで可能となる。自動車ではこれは困難である。
- ・自動車でないと出来ないこともある。**ケースバイケースで採用**すべき。

今回学んだ内容について (基本原理よりは、適用事例や適用ルール等)

□時代の変化に耐える、対応できる手法や技術を学べば研究スタイルの変化や、分野の変更にも強い

今後、5年、10年先まで同じ研究手法を継続できる保証はない。研究の大きな流れは、研究分野の異なる研究者との分野間連携・融合や学際研究となっている。このような将来の研究発展や変化に寄与できる可能性を有するデータサイエンスや人工知能を武器と出来れば心強い。さらには、自分の研究に付加価値を付けることも可能となる。

□同じ手法や技術を学ぶにしても時代を背景とした変化に対応が必要

時代は常に変化している。特にIT関連技術やその周辺環境は急速に変化し続けている。同じことを学ぶにも、昔の古いままのスタイルや内容を学んでも意味が無い。常に時代の先端を行く形式で学ぶことが必要である。例えば、同じコンピュータ関連技術であっても、昔と現在では学ぶ内容は大きく変化している。

□化学データサイエンスやAIを取り巻く環境の変化

化学データサイエンスやAI自体も大きく変化しているが、さらに大きく変化しているのはコンピュータ関連技術である。これらの技術がデータサイエンスやAI自体の実施環境や学ぶ環境及び内容も大きく変化させた。特にAPIや解析ソフト及び関連ツールの充実によりプログラミングにかかわる作業は大きく変化した。この結果、現在はデータサイエンスやAIそのものを学ぶというよりは、その適用に関しての適用分野特有の環境に対応する技術の重要性が増大した。実用的観点では、データサイエンスやAIの基本原理よりも、適用に関する技術を学ぶことが重要となっている。

今後の化学データサイエンスやAIの変化と適用

□ 今後の化学データサイエンスやAIを取り巻く環境の変化

- ・ コンピュータ関連の様々な技術の向上と展開
- ・ データサイエンスやAI自体の発展
- ・ 多種多様な関連データベースの構築やビッグデータ
- ・ ケモメトリクス研究支援システムの発達と開発
- ・ 研究パターンの変化
- ・ 適用分野の拡大と融合研究の増加
- ・ その他

三段階での化合物評価／予測システムの展開と実現化

自動車からTVへの進化

□時代の変化と化学データサイエンスやAIの実施における実施環境の変化

時代の変化に従って、実施できるものと実施が不要になるものが出てくる。

いつまでも古い時代と同じ形式での研究業務が必要なわけではない。むしろ、新規の事項が出てきたり、従来は無視できたものが重要な問題になる等の変化が出る。

現在の化学データサイエンスやAIの実施環境はひと昔前の環境とは大きく異なっている。この観点で、今回のチュートリアルは現在の時代にふさわしい内容を理解するべく内容を組み立てた。

◇ひと昔前はコンピュータ関連の作業環境が貧弱だった。このため、データ解析ソフト自体の構築や、関連機能に関するプログラミングが主体で、解析自体の実行条件は緩くい、貧弱であった。

ひと昔前の主要な作業→コンピュータ周りの総ての環境整備が中心

現在はコンピュータ関連の環境が急速に改善され、昔は考えられないような潤滑な資源や高速性、ネットワークやデータベース等が利用できるようになった。

◇現在はデータ解析ソフトや化学に関連するコンピュータ環境が急速に充実している。

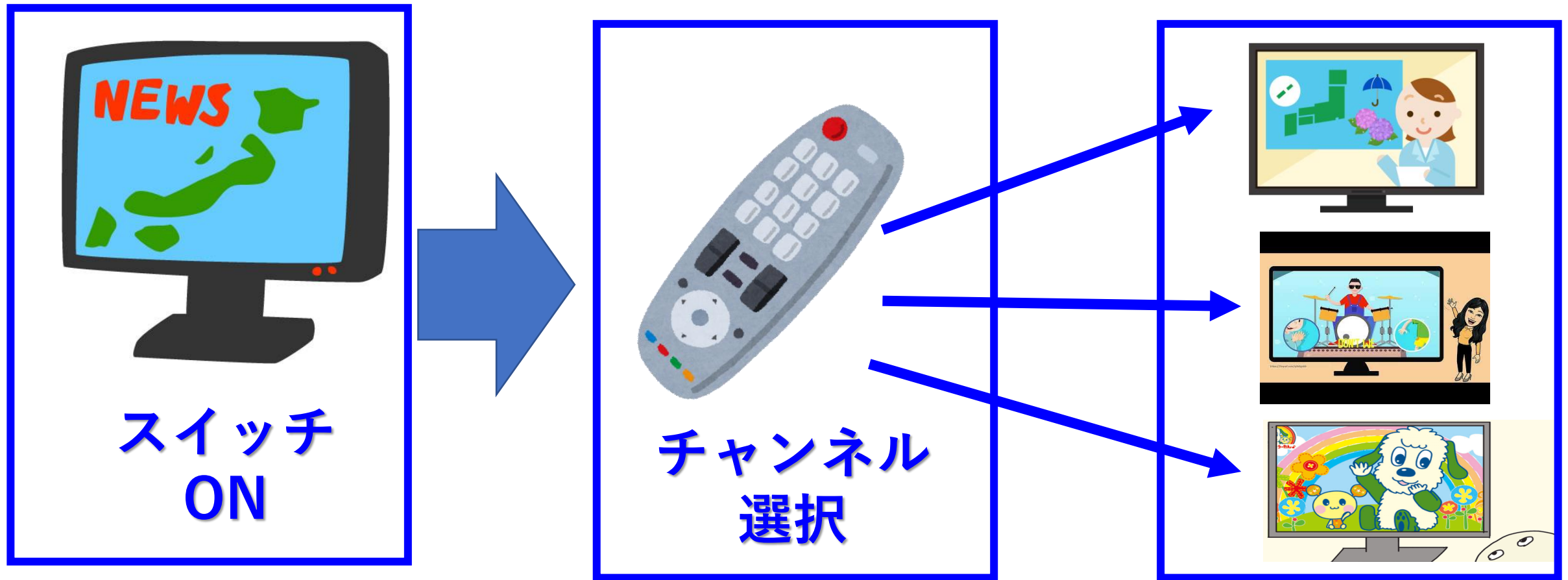
従って、プログラムに関する技術や、データサイエンス手法に関する基本知識等は不要になりつつある。変わって重要となっているのが、データサイエンス手法やAIの運用技術である。昔は無視できた内容が、データ量の急増等により無視不可能となっている。

現在の主要な作業→データサイエンスの運用関係への留意が重要

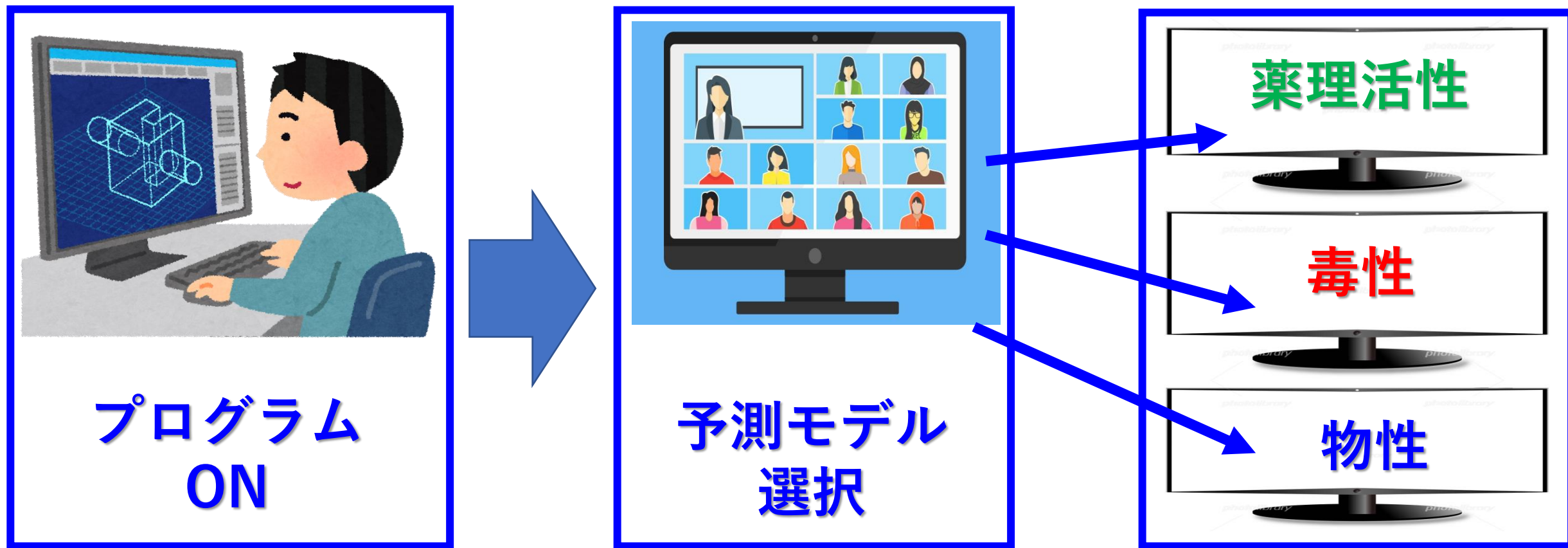
本チュートリアルは現在の化学データサイエンスやAIを実施するにあたって必要かつ留意すべき点をまとめて示した。

自動車の利用からTVの利用に変化した時の実施手順 三段階操作となる

今後の時代的な進歩による予測／評価研究実施イメージ

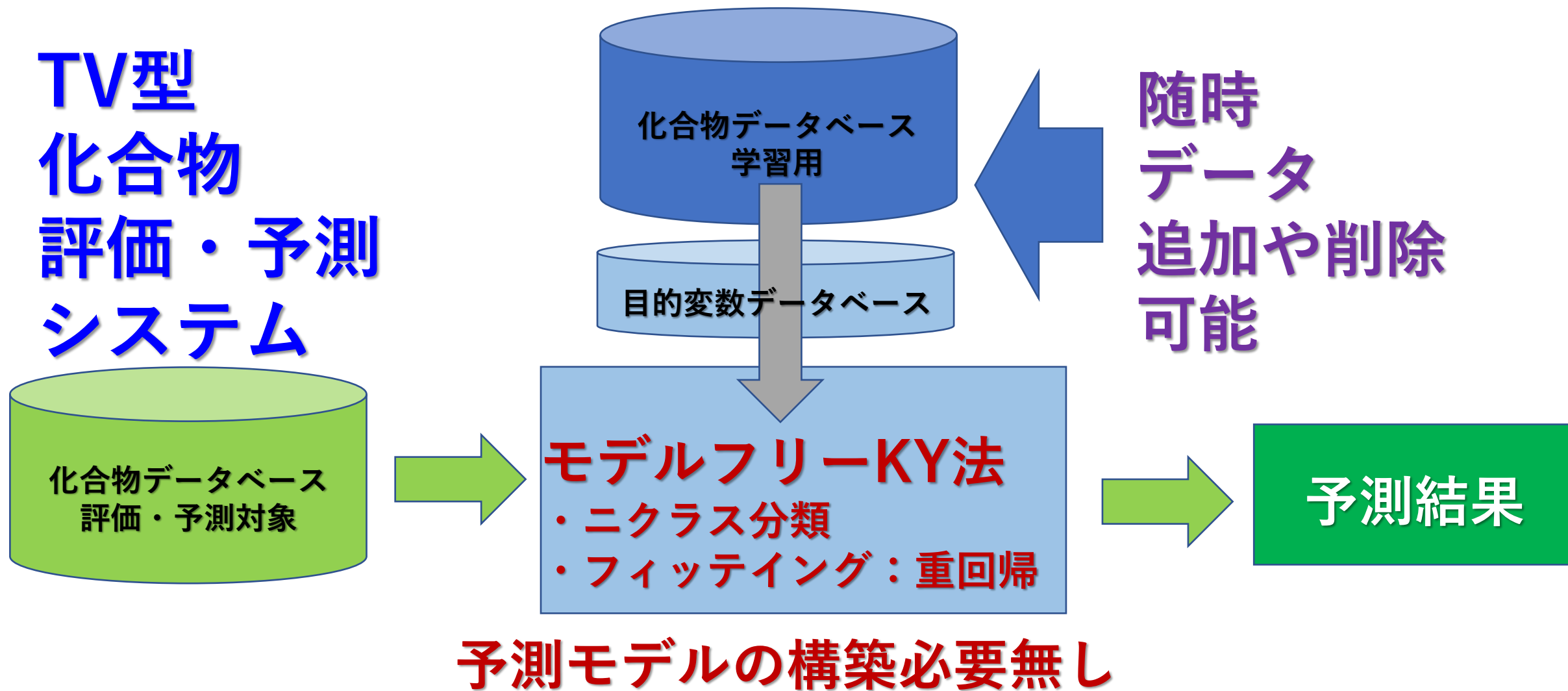


自動車の利用からTVの利用に変化した時の実施手順 三段階操作で完了→「オートノマス創薬」への第一歩



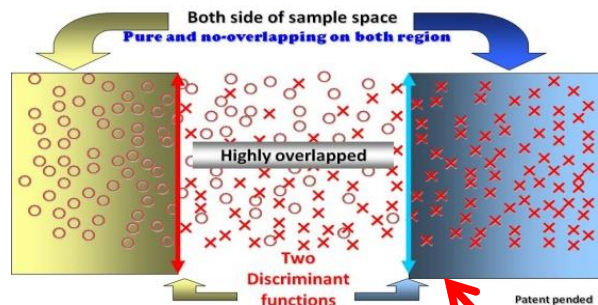
予測や評価プロセスの完全自動化による、
目的化合物の種々特性を三段階の操作で取得

化学データサイエンスやAIのTV化時代の創薬システム 「オートノマス創薬」



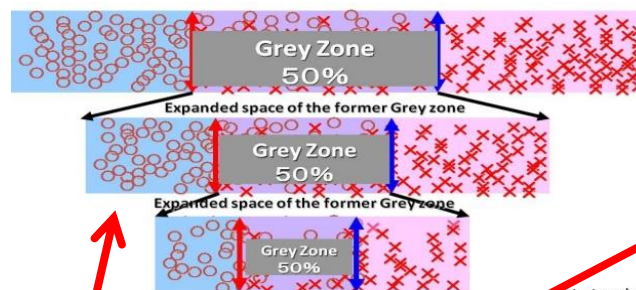
First basic concept of KY method

Spatial region on sample space



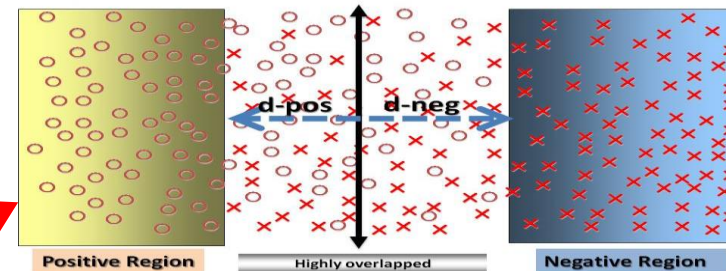
Second basic concept of KY method

Multi-steps for 100% classification

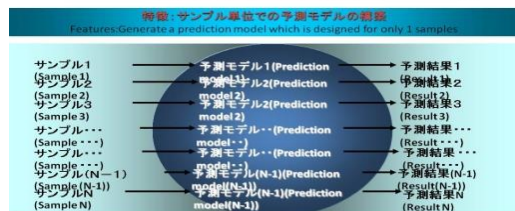


New approach to the "KY method"

by one discriminant function

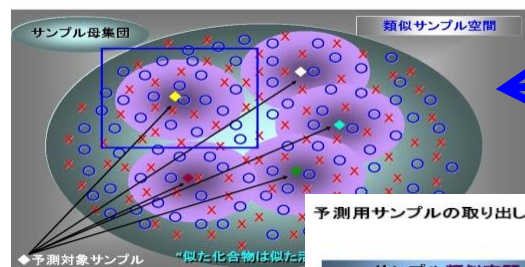


「テーラーメード・モデリング」による予測アプローチ
(Prediction approach by "Tailor-Made Modeling")

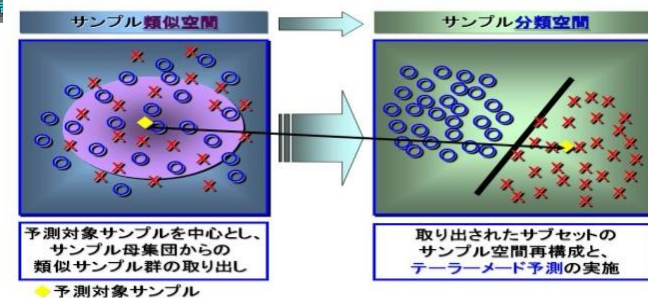


利点 (Merit) : 予測率が大幅に向上する (High prediction ratio will be achieved)
難点 (Weakness) : 計算時間がかかる (Need large calculation time)

サンプル母集団からの予測用サンプルの取り出し



予測用サンプルの取り出しと、テーラーメード予測



A series of KY methods

Discriminant Analysis	Fitting
Two model KY	KY Fitting with DA
Single model KY	KY Fitting with no DA
Model free KY	Model free KY Fitting

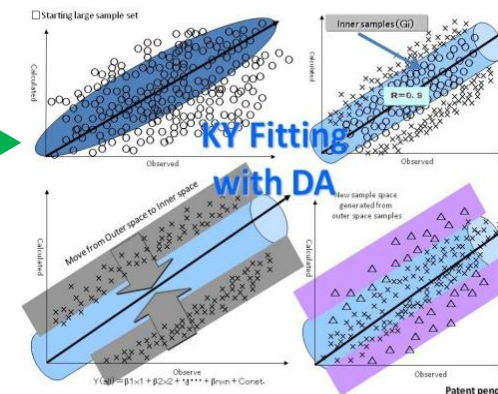
*Always carry perfect classification

*Always high coefficient of determination

Discriminant Analysis	Fitting
Tailor-made Modeling for DA	Tailor-made Modeling for Fitting

*Always carry high prediction ratio

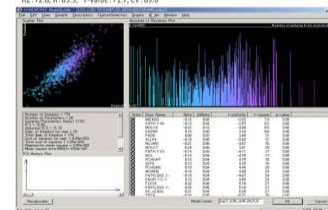
All methods were Patent Pending



◆KY method for fitting methods (Will be soon coming)

◆フィッティングKY法による解析(ステージ1)

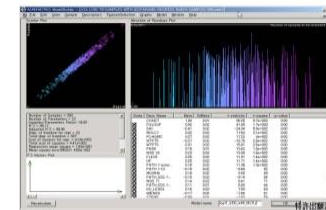
Step1: Inner sample set
Number of samples: 791, Number of used parameters: 28, Confidence ratio: 27.8
R2: 72.8, R: 85.3, F-value: 71.7, CV: 69.6



◆フィッティングKY法実証実験(2)

◆フィッティングKY法による解析(ステージ1)

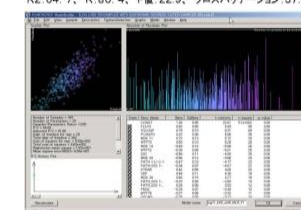
ステップ1: インナーサンプル
サンプル数: 398, パラメータ数: 22, 信頼性指標: 18.1
R2: 98.2, R: 98.1, F値: 428, クロスバリデーション: 94.4



◆フィッティングKY法実証実験(3)

◆フィッティングKY法による解析(ステージ1)

ステージ1: アウターサンプル
サンプル数: 393, パラメータ数: 29, 信頼性指標: 13.6
R2: 64.7, R: 80.4, F値: 22.9, クロスバリデーション: 57.5



TV型化合物評価・予測システムの構築提案

未来型創薬「オートノマス創薬」

予測手法としては「モデルフリーKY法」を適用する

モデルフリーKY法は、ニクラス分類とフィッティングの二種類存在
モデルフリーKY法の特徴は多数存在するが、以下のようなになる。

- ①信頼性や予測性が極めて高い
- ②同じ動作の繰り返しがメインなので、プログラミングが容易
- ③自動化が容易
- ④メンテナンス性が高い
- ⑤予測モデルの構築が不要なので、常に新規データを交えて解析可能
- ⑥サンプル数の変化への対応が強い。少数のサンプルデータから、ビッグデータまで対応可能

モデルフリーKY法の様々な特徴は、化合物の評価や予測の
自動車からTVへの進化や変換による「オートノマス創薬」実現

今後の研究形体：研究目的に応じた学習内容の選択

1. データサイエンスやAIのオールマイティなプロを養成

データサイエンスやAIを基本から知り、プログラミング等しながら解析等実施することは理想である。時間と、学習環境があり、細かな研究内容に答えるにはこのオールマイティな人材の養成が必要である。人力車から作成し、道路の敷設や舗装等の実施も可能でも一人の力では、最先端の研究レベルの研究実施はほとんど不可能な時代となっている。

さらに、化学データサイエンスやAIの研究レベルでの実施には適用分野の専門知識が必要なこととも明確である。両方を満たすことは昔と異なり現在は極めて困難である。

2. 自動車で例えられるような化学データサイエンスやAIの取得

現在の化学データサイエンスやAIを取り巻く環境はひと昔前とは劇的に変化した。コンピュータやプログラムに関する作業は殆ど必要なく、サンプルデータの収集等の環境も整っている。

解析手法の基本や原理、プログラムの内容等の知識は殆ど必要が無くなった。むしろ、解析を実行するための化学データサイエンスやAIの運用知識や、適用分野の専門知識による評価や洞察等が重要となってきた。ソフトの操作ではなく、解析の信頼性を高めるノウハウが重要となっている。

3. 化合物評価や予測のTV化への変化と展開への対応

研究スタイルは昔の**馬車レベル**から**自動車へと変化**した。今後はさらなるIT関連技術の発展に伴い、化学データサイエンスやAIの研究はますます自動化が進み、**TVのような三段階レベルの操作**で信頼の高い評価・予測結果を入手できる時代となる。このような環境で新たに出来ることが生じてくる。リアルタイム創薬、合成も含めた総合型創薬、インテグレートド概念に基づく創薬等、内容が高度でグレードの高いオートノマス創薬研究の実施が可能となる。今後の展開が楽しみである。

ご清聴ありがとうございました。

TS-02 日時：2021 年 10 月 25 日 13:00-17:00

半日で知る、化学分野のデータサイエンスおよび人工知能概要：

「FS-08:化学データサイエンスおよび人工知能討論、勉強会」立ち上げ会協賛

2021 年 10 月 28 日 13:00-14:30

A Half-day Overview of Data Science and Artificial Intelligence in the Field of Chemistry:

"FS-08: Chemical Data Science and Artificial Intelligence Discussion, Work Shop"
Kick-off Meeting Sponsorship

モデレーター：湯田 浩太郎 Kohtaro Yuta
株式会社インシリコデータ (In Silico Data, Ltd.)