

◆特集1◆

AI, 生成 AI の R & D への導入・活用の最前線
～研究 DX のためのデータプラットフォームの構築～

データ探査及び分析を意識した 研究データの蓄積と構造化

上島 豊 (株) キャトルアイ・サイエンス 代表取締役



《PROFILE》

略歴：

1992 年 3 月	大阪大学工学部 原子力工学科 卒業
1997 年 3 月	大阪大学大学院工学研究科 電磁エネルギー工学専攻 博士課程修了
1997 年 4 月	日本原子力研究所 博士研究員
2000 年 4 月	日本原子力研究所 研究職員
2006 年 3 月	日本原子力研究開発機構 (旧日本原子力研究所) 退職
2006 年 4 月	キャトルアイ・サイエンス設立 代表取締役 就任

主な参加国家プロジェクト：

文部科学省 e-Japan プロジェクト「ITBL プロジェクト」, 「バイオグリッドプロジェクト」
総務省 JGN プロジェクト「JGN を使った遠隔分散環境構築」
文部科学省リーディングプロジェクト「生体細胞機能シミュレーション」

主な受賞歴：

1999 年 6 月	日本原子力研究所 有功賞 「高並列計算機を用いたガガ粒子シミュレーションコードの開発」
2003 年 4 月	第 7 回サイエンス展示・実験ショーアイデアコンテスト文部科学大臣賞「光速の世界へご招待」
2004 年 12 月	第 1 回理研ベンチマークコンテスト 無差別部門 優勝

主な著作：

培風館「PSE book シミュレーション科学における問題解決のための環境 (基礎編)」ISBN : 456301558X
培風館「PSE book シミュレーション科学における問題解決のための環境 (応用編)」ISBN : 4563015598
培風館『ベタフロップス コンピューティング』ISBN978-4-563-01571-8
臨川書店『視覚とマンガ表現』ISBN978-4-653-04012-5

1 はじめに

現在の R & D 領域では、データ分析や管理は、極めて属人的な扱いである。客観的なデータ生成、分析が要求される理学、工学領域で、この属人性は大きな問題を孕んでいる。研究というものとは創造的な活動であり、個人の才能、発想に起因する「なぜ、そう考えたか？」の部分に属人性が必要なことは当然だ。しかし、どのようにデータを生成し、どのように分析し、結論を導いたかは、属人的では問題で、客観的かつトレーサブルであるべきである。実際、センサーや計算機の能力向上により、データの生産性が向上し、扱うべきデータが膨大になり、詳細記録の欠如、偶発的データ取り違い、主観的データ操作が発生する余地が増大し、データ生成、分析プロセスの信頼性が大きく揺らいでいる。

実際、私は弊社を設立する前は研究機関にてコンピュータ、ネットワークの最先端技術を駆使し、自然科学、工学研究を約 10 年間行っていた。その中で R & D データが属人的に処理され、その管理状態がデータの信頼性及び有効活用性を大きく阻害し、共有化及びインフォマティクス分析、AI 化が進まないことを経験した。

本記事では、私自身の 10 年の R & D 経験と弊社の 19 年の R & D 支援実績から得た「データ探査及び分析を意識した研究データの蓄積と構造化」に関して、簡単に解説する。

2 R & D 部門における研究・実験データ蓄積の実情

まず、「何のためにデータを蓄積するのか？」を考えてみよう。それは、過去行った実験や解析のデータを共有し、利活用するためであろう。そういう観点を加味して、R & D 部門における研究・実験データ蓄積の状況を見てみると、研究・実験データが蓄積はされているが、共有し、利活用することはほとんどできていないというのが実情である。そのようになってしまっている原因は、いくつかあり、そのすべてをクリアしない限り、共有し、利活用できるデータ蓄積には至らないのである。以下、その原因を解決すべき順番に列挙しておく。

- 1) データを記録したものが個人フォルダの中にしかない。
- 2) データを記録したものは参照できるが、どこに何があるのかが属人管理である。

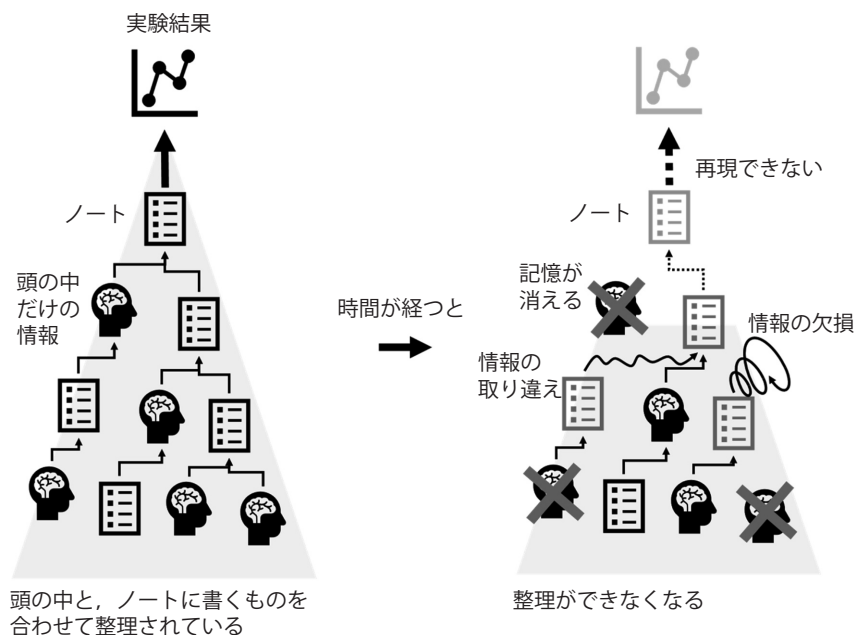


図1 属人的なデータ管理とその特性

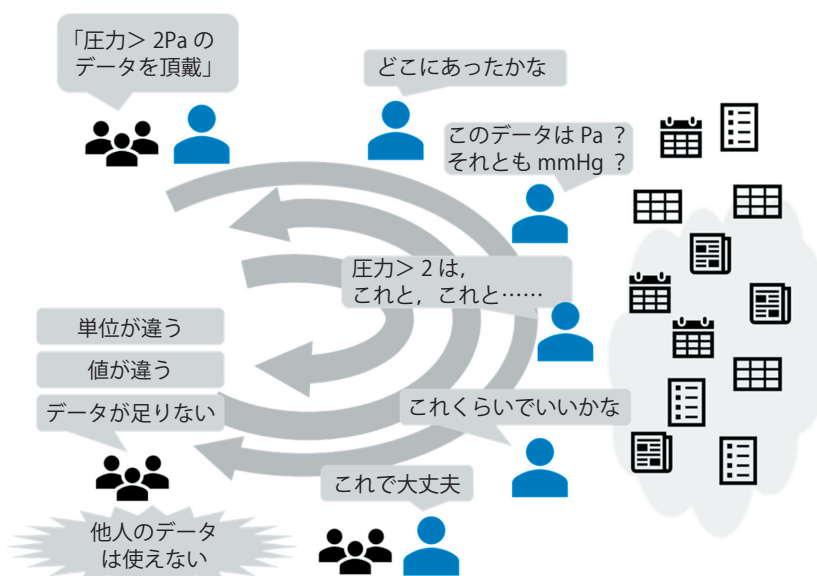


図2 属人的に管理されたデータが引き起こす問題

3) データ記録に抜け（必須な実験パラメータ値が記録されていない等）がある。

4) データ記録の項目名が記録者、時期によってバラバラで、統一されていない。

これらの問題の解決方法は、研究開発リーダーのバックナンバー及び書籍「ケムインフォマティックスにおけ

るデータ収集の最適化と解析手法」の私の執筆部分に詳しく記載しているので、そちらを参照してほしい。次章では、「4) データ記録の項目名が記録者、時期によってバラバラで、統一されていない」に関して、さらに一歩踏み込んで、どのように統一すべきかを具体的な例とともに説明していく。

データ記録では単純構造で項目間に 3 論理的関係性を持たない項目名を 使うべき

前章では、「R & D 部門における研究・実験データ蓄積の実情」に関して、説明を行った。本章では、「データ記録では単純構造で項目間に論理的関係性を持たない項目名を使うべき」と題し、記録するデータの項目名はどのように決めるべきかを論じる。

項目名決定において重要なことは、データ分析を前提にし、データ抽出／検索、分析を行いやすい項目名にしておかないといけないということである。データ抽出／検索、分析を行いにくい項目名だと、結局、データを探したり、分析したりするのに大きな手間がかかるため、データ共有を大きく阻害するのである。R & D 部門以外の他の業務系の場合は、項目名決定時にデータ分析のことを重視せず、どのようにデータを入力、登録するかを考え、項目名を決定する。つまり、データ入力のしやすさを前提にして、項目名を決定しているのである。R & D 部門以外の他の業務系の場合は、データ抽出／検索、分析が固定的で、追加、変化が発生しないので、データ入力のしやすさを前提にして項目名を決定しても、その後、分析しやすい項目に内部で変換してデータを登録したり、数学的基盤に則った変数を項目名にしたりしなくとも、大きな問題にはならないのである。一方、R & D 部門では、項目が多く、項目の追加や変更も頻繁である中で、様々なパターンのデータ抽出／検索、分析が必要となるので、データ抽出／検索、分析を行いやすい項目名にしておかないと、データ抽出／検索、分析し易さを維持するのが大変になり、使われなくなってしまうのである。以下、R & D 部門でデータを記録していくための項目名を決めるときに注意すべき点を列挙しておく。

3.1 すべてのデータは、項目名－項目値という 単純構造に落とし込む

これは、「データを分析するときには、2次元の表形式データになっている必要がある」ことからくる条件である。Excelなどで、グラフを描いてデータ分析をするときには、必ず項目名－項目値の形になっていることは理解できると思う。しかし、これは結構守られていないのである。例えば、実験ノートや実験データを記録するExcelでは、表1のようにになっていることが多いと思う。

表1 項目名－項目値という単純構造になっていない項目のデータ記録

乾燥工程				焼結工程	
1 段目		2 段目			
温度	風速	温度	風速	温度	風速

実は、この状態は項目名－項目値の単純構造にはなっていないのである。「純粋な項目名は温度と風速だけで、乾燥工程、焼結工程や1段目、2段目などは、分類的なもの」と捉えて、受け入れてしまってはだめなのである。もし、そう考えてしまうと温度＝120という項目名－項目値を見たときに、それが乾燥工程の温度なのか？焼結工程の温度なのか区別がつかないことになる。項目名は重複がなくユニークな命名でないと、データ分析時にそれが何の値なのか厳密に分からず、困ってしまうのである。つまり、乾燥工程、焼結工程や1段目、2段目などの分類的情報を分類情報と考えるのではなく、それら分類情報を含めて項目名にしなければならないのである。Excelでセル結合を使っているデータ記録は、この段階で失格ということになる。結局、先ほどのデータを項目名－項目値の単純構造にすると、表2のようになる。

表2 項目名－項目値という単純構造になっている項目のデータ記録

1 段乾燥温度	1 段乾燥風速	2 段乾燥温度	2 段乾燥風速	焼結温度	焼結風速

つまり、Excel で言うと 1 行目の 1 セルずつに項目名が列挙され、2 行目以降は項目値だけが並び、それ以外の分類情報などは一切無い形が、項目名－項目値の単純構造の項目ということである。「Excel でセル結合を使っている、ピボットテーブルやマクロを使えば、データ分析はできるのでは？」という方もいらっしゃると思う。もちろん、それは間違っていない。ただし、R & D 部門で取り扱う様々な実験すべてに対応するようにできるかという点で非常に悩ましいはずである。ピボットテーブルやマクロは、ある一定の決まりきったことを何度か行う場合は、便利なのだが、そもそもそれを作るのが大変なのである。経理や営業などの業務系で、いつも同じ項目データで同じ処理をするのであれば、ピボットテーブルやマクロは便利なのだが、扱う材料や処理プロセスがどんどん変わっていく R & D 部門では、ピボットテーブルやマクロを作成し、メンテナンスすることの負担が大きく、現実的には運用できなくなってしまうのである。実際、使われなくなったマクロや GUI が施された Excel が乱立し、收拾がつかなくなってしまう経験のある人も多いのではないと思う。R & D 部門のデータ記録は、単純で「それでいいのか？」と思ってしまうが、項目名－項目値の単純構造でなければならないのである。「Simple is best.」である。

3.2 項目名－項目値の項目間に論理的関係があってはいけない

「項目名－項目値の項目間に論理的関係があってはいけない」が、何を意味しているのかピンとこない人も多いと思う。以下で、例を挙げて、説明する。

表 3 の原材料名 1 と原材料濃度 1、原材料名 2 と原材料濃度 2 は、2 つの項目がペアになっている。これが「項目間に論理的関係がある」ということである。また、上記は 1、2 を入れ替えても同じ実験になるという対称性もこの項目の論理的関係性として埋め込まれている。ペアや対称性といった項目自体に内在させられた関係性は、物理（自然科学）とは関係のない人為的な、項目名命名による関係性である。そして、データを絞り込んだり、分析する場合には、この項目の論理的関係からくる論理制約（ペアや対称性）を排除する必要があり、非常に面倒な作業が必要になってくる。したがって、項目名－項目値の項目間に論理的関係がある項目名でデータ記録をしてはいけないのである。ちなみに、表 3 の項目名から項目名－項目値の項目間に論理的関係を排除すると表 4 のような項目名になる。

さらに詳しい説明は、研究開発リーダーのバックナンバー及び書籍「ケモインフォマティクスにおけるデータ収集の最適化と解析手法」の私の執筆部分に詳しく記載しているので、そちらを参照してほしい。

表 3 項目名－項目値の項目間に論理的関係があるデータ記録

実験 ID	原材料名 1	原材料濃度 1	原材料名 2	原材料濃度 2	引張強度
EXP1	エチレン	80	プロピレン	20	10
EXP2	ブタン	75	エチレン	25	12
EXP3	プロピレン	60	ブタン	40	8

表 4 項目名－項目値の項目間に論理的関係がないデータ記録

実験 ID	エチレン濃度	プロピレン濃度	ブタン濃度	引張強度
EXP1	80	20	0	10
EXP2	25	0	75	12
EXP3	0	60	40	8

4 記録するデータの項目は、データを絞り込み、分析で使う項目名にすべき

前章では、「データ記録では単純構造で項目間に論理的関係性を持たない項目名を使うべき」と題し、項目間に論理的関係のある項目を作ってはいけない理由と回避方法を説明した。本章では、「記録するデータの項目は、データを絞り込み、分析で使う項目名にすべき」と題し、前章よりさらに深く、記録するデータの項目名に関して、どうあるべきかを論じる。

例えば、エチレン、プロピレン、ブタンを一定の圧力で混ぜて合成し、その合成物の粘度を測定するだけの単純な工程を考える。項目間に論理的関係性を持たない一番単純な項目名は、次のようなものになるだろう。

a) エチレン重量 (g), プロピレン重量 (g), ブタン重量 (g), 合成圧力 (Pa), 粘度 ($\text{Pa} \cdot \text{s}$)

この項目名で、望みのデータを絞り込むことができるだろうか？粘度が $10 (\text{Pa} \cdot \text{s})$ 以上の実験結果を集めたい場合は、問題なさそうだが、材料（エチレン、プロピレン、ブタン）の配合を条件にすることを考えるとどうだろうか？「エチレンが 10g 以上含んだ」というような条件で絞り込みたいだろうか？多分そうではないはずである。「エチレンが 10 重量%以上含んだ」のように、比率で指定したいはずである。これは「合成反応は、材料比率が同じであれば、近い特性を有するものになる」という暗黙の前提があり、エチレン 10g , プロピレン 20g , ブタン 70g の実験で生成される合成物とエチレン 15g , プロピレン 30g , ブタン 105g の実験で生成される合成物の特性（粘度）は、ほぼ同じであると考えられているからである。つまり、同じ材料比率の実験データを集め、合成圧力が変わると特性がどう変わるかなどを確認したいはずであり、その場合、重量単位はデータ絞り込みの条件としては使えないのである。

実は、合成物の粘度の合成圧力依存性を調べる場合も、エチレン重量 (g), プロピレン重量 (g), ブタン重量 (g), 合成圧力 (Pa) のような項目では困る。X 軸に合成圧力 (Pa), Y 軸に粘度 ($\text{Pa} \cdot \text{s}$) を設定して、X-Y プロットを描けばいいだけなのでは？と思うかもしれない。次章でより詳しく説明するが、実は、データ分析をするためにはそれだけではダメで、X 軸以外の実験パラメータ、つまり、エチレン重量 (g), プロピレン重量 (g), ブタン重量 (g) のそれぞれが同じ値のデータごとにデータ

を分類して、X-Y プロットを別グラフとして描く必要がある。そうしないと、X-Y プロットに合成圧力 (Pa) 以外の依存性が紛れ込んでしまい、正しい合成圧力依存性を見ることができないのである。

それなら、エチレン重量 (g), プロピレン重量 (g), ブタン重量 (g) のそれぞれが同じ値のデータを探して、それでプロットすればいいのではないか？と思うかもしれない。しかし、それでは「エチレン 10g , プロピレン 20g , ブタン 70g 」のデータを抽出すると、「エチレン 15g , プロピレン 30g , ブタン 105g 」のデータは漏れてしまうことになる。しかし、これらはエチレン、プロピレン、ブタンの比率は同じなので、量効果（容器壁面、表面効果など）が無視できると仮定するなら、粘度など生成物の物理特性は非常によく似たものになるはずであり、同じグラフにマーカー種を変えたぐらいで実験結果をプロットしたいはずで、データを絞り込むときには双方抽出してほしいはずである。つまり、容器壁面、表面効果などの量効果はほとんど効かないことが多く、量効果を分離した分析をすることが多い場合は、総重量を分母にした重量百分率の項目が便利だということである。

b) エチレン濃度 (wt%), プロピレン濃度 (wt%), ブタン濃度 (wt%), 総重量 (g), 合成圧力 (Pa), 粘度 ($\text{Pa} \cdot \text{s}$)

ここで注意が必要なのは、総重量 (g) という変数が必要だということである。a) で 5 変数だったものが、b) では 6 変数になり、変数が増えているではないか？と思うかもしれない。しかし、エチレン濃度 (wt%) + プロピレン濃度 (wt%) + ブタン濃度 (wt%) = 100 という、変数間の縛りがあり、独立な変数は a) と同じなので、これで問題ないのである。逆に、総重量 (g) という変数を忘れると独立な変数が減ってしまうことになるので、忘れないように注意が必要である。

エチレン濃度 (wt%), プロピレン濃度 (wt%), ブタン濃度 (wt%) のそれぞれが同じ値のデータごとにグラフを描くとすると「エチレン 10g , プロピレン 20g , ブタン 70g 」と「エチレン 15g , プロピレン 30g , ブタン 105g 」は、同一グラフにプロットすることができ、a) での問題は解消されるのである。また、このような項目にすると、X 軸に総重量 (g), Y 軸に粘度 ($\text{Pa} \cdot \text{s}$) を設定し、エチレン濃度 (wt%), プロピレン濃度 (wt%), ブタン濃度 (wt%), 合成圧力 (Pa) がそれぞれ一定値のデータで、グラフを描くと総重量依

存性がほとんどないことも簡単に確認できる。もし、総重量依存性が大きい場合は、何か実験で失敗をしているか？、特異な量効果（容器壁面、表面効果など）が効いているか？等が原因である可能性があり、追試実験を行う必要があることもこの変数系ではわかるのである。

本章で伝えたかったことを纏めると、記録蓄積するデータの項目は、実験時の記録の取り易さではなく、実験の独立変数を意識し、データをどのように絞り込み、どのように分析するかを具体的にイメージし、それに沿った項目になっていなければならないということである。

5 教科書などであまり触れられていない多変量データ分析の重要な注意点

前章では、「データ記録では単純構造で項目間に論理的関係性を持たない項目名を使うべき」と題し、データを記録する場合における項目名の注意点を説明した。本章では、「教科書などであまり触れられていない多変量データ分析の重要な注意点」と題し、蓄積、共有した実験データを分析する場合に非常に重要にもかかわらず、データ分析の教科書にはあまり触れられていない注意点に関して、説明する。

世の中では、MI, AI が叫ばれているが、MI, AI はアイデアを提供してくれるという点では心強いが、「それを良いと思った理由」は提示してくれないので、データ分析、

理解という観点では無力である。実際、MI, AI を活用するにしても、MI, AI が提示した案について、実データを分析し、その背景理由を探ることを辞めてしまっはいけない。そもそも、それを辞めるなら研究者はいらないということになる。背景理由を探るための最も初歩的なデータ分析方法が義務教育時代から親しんだ X-Y プロットである。たかが、X-Y プロットであるが、されど X-Y プロットである。結局、様々な学术论文でも賑やかなの 3D 可視化や画像などがあつたとしても、論文で一番重要な部分は X-Y プロットになっているのは、偶然ではない。

X-Y プロットは、「Y 軸項目値の X 軸項目値依存性」を表すもので、データ分析、理解というものの基本的な部分は、まさしくこれを使う必要がある。一般的に X 軸項目に実験パラメータを割り付け、Y 軸項目に実験結果項目を割り当てる。実は、注目している実験パラメータを X 軸項目に指定し、依存性を確認したい実験結果項目を Y 軸項目に指定しただけでは駄目なのである。まず、どんな実験でも実験パラメータが一つということはあるはずで、注目している実験パラメータ以外にも実験パラメータがたくさんあるはずである。その実験パラメータを全く無視して、X-Y プロットを描いてはいけない。実際、そのようなことをすると「Y 軸項目値の X 軸項目値依存性」を見ているようで、表には出てきていない X 軸項目以外の実験パラメータ依存性を X 軸項目値依存性と誤認してしまう恐れがある。

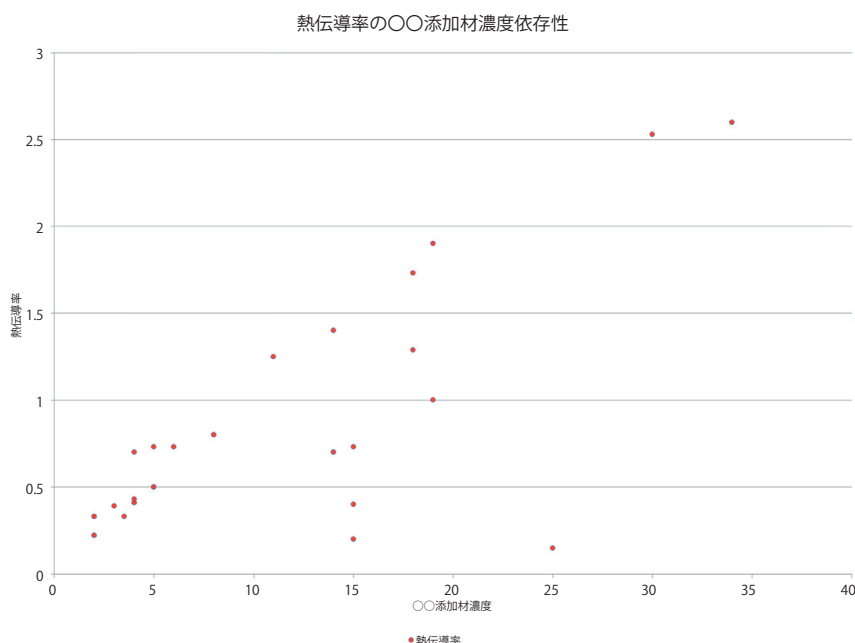


図3 X 軸以外の実験パラメータが一定であるデータ毎にデータ分類されていない X-Y プロット

上記 X-Y プロットは、様々な実験結果に対して、注目している〇〇添加剤濃度を X 軸項目に指定し、依存性を確認したい熱伝導率を Y 軸項目に指定したもので、添加剤濃度を上げていけば、添加剤濃度 1%あたり、熱伝導度は 0.07 ~ 0.08 ずつ比例して上がっていく傾向にあることが読み取れてしまう。

実は、上記 X-Y プロットは問題なのである。「注目している実験パラメータを X 軸項目に指定し、依存性を確認したい実験結果項目を Y 軸項目に指定」する前に X 軸項目以外の実験パラメータが一定であるデータ毎にデータを分類しておく必要があるのである。多工程の実験では、実験パラメータは数百から千程度になるので、「X 軸項目以外の実験パラメータが一定であるデータ毎にデータを分類しておく」作業は、結構大変な作業になる。しかし、この作業を省略して、「Y 軸項目値の X 軸項目値依存性」をグラフ化してはいけない。検討したい X 軸項目以外の実験パラメータが確定している場合は、「X 軸項目以外の実験パラメータが一定であるデータだ

けにデータを絞り込む」ことは、ひたすら実験パラメータの数だけ Excel の filter をかけていけばいいだけなので、大変な作業にはなるができてしまう。しかし、そもそもその一定にしたい値がよくわかっていない場合は、何らかの方法で「X 軸項目以外の実験パラメータが一定であるデータ」毎に分類しておき、分析をしたい実験パラメータが一定であるデータ塊を選ぶ必要がある。

Excel で数百から数千の filter をかけることも大変だが、「X 軸項目以外の実験パラメータが一定であるデータ」毎に分類するのは、手動処理では永遠に終わらないほどの作業量になってしまう。例えば、Excel で 1000 実験パラメータ項目を「実験パラメータが一定であるデータ毎に分類」しようとする、1 個目のパラメータ項目の値を filter で 1 個指定し、次のパラメータ項目も値を filter で 1 個指定するという作業を 1000 項目すべてに行い、やっと実験パラメータが一定であるデータが 1 塊だけ分類できる。その後、最初のパラメータ項目の値を別の項目値に filter を変え、同じことを全項目の全

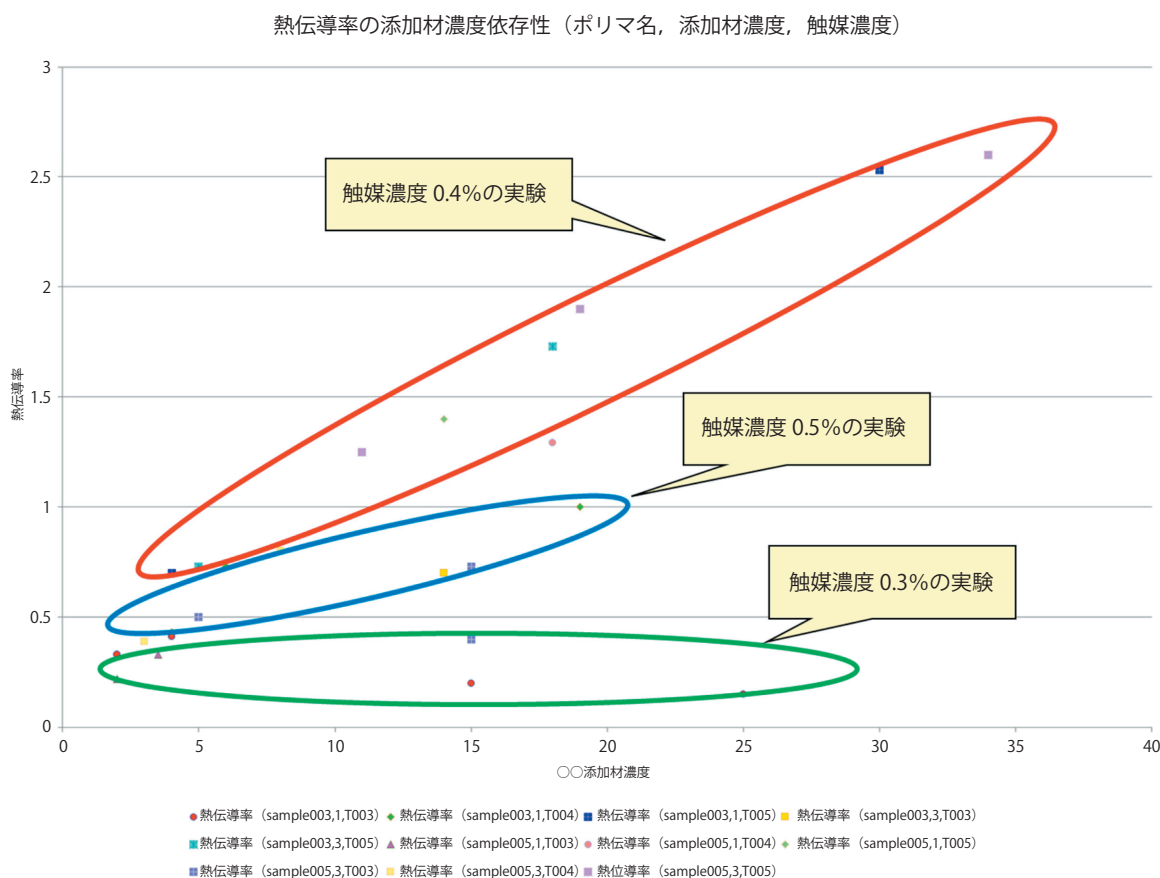


図4 X 軸以外の実験パラメータが一定であるデータ毎にデータ分類された X-Y プロット

項目値のパターンを抜けなく filter をかけていく必要がある。もし、項目値が2個ずつだとして、filter を1秒で行なえたとしても、全パターンの2の1000乗パターンを確認するには、 2^{1000} 秒=2の後に0が300個ぐらい付いた数(秒)=宇宙の寿命よりはるかに長くかかってしまい、現実的ではない。実際には、何列か filter をかけると結果が0件(0行)になることがほとんどなので、ここまで時間はかからないが、たぶんやりきる人はいないはずである。したがって、多工程の実験を行う R & D 部門では、こういうことが簡単に実施できるツールを整備しておくことが必須ということになる。

先ほどの図3では、X軸項目以外の実験パラメータ値で分類をしないで、X-Yプロットを描いたが、図4では追加しているポリマー種類や触媒濃度などのX軸以外の実験パラメータの値が同じデータ毎にデータを分類(マーカー種を分け)し、データをプロットしたものである。そうすると「添加剤濃度1%あたり、熱伝導度は0.07~0.08ずつ比例して上がっていく傾向がある」のは、触媒濃度を0.4%にした時だけで、0.3%の場合は、そもそも添加剤濃度依存性がないことがわかる。X軸以外の実験パラメータの値が同じデータ毎にデータを分類しないと、結論をミスリードしてしまいかねないのである。

実は、データ蓄積、共有を行うまでは、この作業は発生しないか、したとしても大きな負担ではない。データ蓄積、共有を行うまでは、自分が依存性を調べたいと思う実験パラメータ数種だけを変動させて、それ以外の実験パラメータは固定した実験を行うはずであり、また、実験直後にデータ分析を行うはずである。その場合、ほとんどの実験パラメータは固定されており、また、自身も何を固定したという記憶が残っているため、ここで議論している「X軸の項目以外の実験パラメータが一定であるデータ」を分類することは簡単である。X軸の項目以外の実験パラメータを一定にして実験をするので、そもそも分類する必要がないことがほとんどである。つまり、こんなグラフを描きたいと思ってから、そのグラフが簡単に描けるような実験をしているということになる。

しかし、データ蓄積、共有をし、遥か過去の自分のデータや他研究者のデータを含めて分析をする場合、「どの実験パラメータが固定されているのか?」は、実データを確認するまで分からなく、また、殆ど場合は、

実験パラメータの大多数が変動してしまっているはずである。実は、データ蓄積、共有をしない場合は、「実験前にグラフが描きやすいように実験パラメータを固定しているか」、少し前の実験であれば自身の記憶を活用することで「どの実験パラメータが固定されているのか」が分かるので、比較的楽にデータ分析ができるだけなのである。データ蓄積、共有を始めると自身の記憶を使わずにデータ分析をすることが強いられ、その分析の大きな変化を研究者が認識できておらず、当惑してしまうのである。データ蓄積、共有をする前に、自身の記憶を活用せず、実データをしっかり確認しながらデータ分析をする習慣をつけていくことが重要なのである。

データ分析を行うときに、X軸に実験パラメータではなく、特性値を設定することもあるはずである。例えば、試作物の粘度と熱伝導度の関係性を知りたい場合は、X軸項目に粘度を設定し、Y軸項目に熱伝導度を設定して、X-Yプロットを描くはずである。ただ、この時にも幾つか気を付けなければならないことがある。特性値は実験パラメータのように項目値が実験パラメータに対して一対一の関係のある項目でない為、「粘度が〇〇の時に熱伝導度が××」というプロットがあっても、あらゆる実験でそれが成り立つかと言えば違う。実験パラメータが違って「粘度が〇〇」になることは、十二分にあり得て、その時、「熱伝導度が××」にならないことも十二分にあり得るのである。だからと言って、「粘度が〇〇」になるあらゆる実験パラメータを実験で確かめ、「粘度が〇〇」になる実験パラメータのすべてで、「熱伝導度が××」になることを調べるのは、現実的でない。もちろん、こういうことは、皆は承知で、いくつかの異なる実験パラメータの結果で、この手のグラフを描いているはずだが、どんな実験パラメータでもその傾向が同じかどうかはしっかりと確かめられていないことが多いはずである。つまり、たまたま、偏ったパラメータの実験での限定的な依存性(偽依存性)を一般的な依存性と勘違いしてしまう可能性が高いのである。したがって、この手のグラフを描く時には、事前に実験パラメータと当該X軸、Y軸に設定する予定の特性値の相関を事前に分析しておき、どのパラメータ領域での粘度と熱伝導度の関係性を評価しようとしているかを明確化しておく必要があるのである。

連載へ続く お楽しみに

参考文献

- 1) 川田重夫, 田子精男, 梅谷征雄, 南多善, 上島豊, 他
PSE book シミュレーション科学における問題解決のための環境
(応用編), 川田重夫, 田子精男, 梅谷征雄, 南多善 共編, 培風館,
(2005), p69-82
- 2) 谷啓二, 奥田洋司, 福井義成, 上島豊
ペタフロップスコンピューティング,
矢川元基 監修, 培風館, (2007), p183-202
- 3) 牧野圭一, 上島豊, 視覚とマンガ表現, 臨川書店, (2007),
p1-5, 221-229
- 4) 上島豊, 月刊「研究開発リーダー」6月号, 技術情報協会, (2023),
p63-68
- 5) 上島豊, 月刊「研究開発リーダー」7月号, 技術情報協会, (2023),
p86-91
- 6) 上島豊, 月刊「研究開発リーダー」8月号, 技術情報協会, (2023),
p78-82
- 7) 上島豊, 月刊「研究開発リーダー」7月号, 技術情報協会, (2024),
p77-84
- 8) 上島豊, 月刊「研究開発リーダー」9月号, 技術情報協会, (2024),
p73-81
- 9) 上島豊, 月刊「研究開発リーダー」11月号, 技術情報協会, (2024),
p85-96
- 10) 上島豊, 月刊「研究開発リーダー」1月号, 技術情報協会, (2025),
p74-82
- 11) 上島豊, 月刊「研究開発リーダー」3月号, 技術情報協会, (2025),
p68-73
- 12) 上島豊, 他
研究開発部門への DX 導入による R & D の効率化, 実験の短縮化,
技術情報協会, (2022), p195-221
- 13) 上島豊, 他
ケムインフォマティクスにおけるデータ収集の最適化と解析手法,
技術情報協会, (2023), p39-74
- 14) 上島豊, 他
実験の自動化・自律化による R & D の効率化と運用方法,
技術情報協会, (2023), p159-199
- 15) 上島豊, 他
少ないデータによる AI・機械学習の進め方と精度向上, 説明可能
な AI 開発, 技術情報協会, (2024), p112-127