

MAGICIAN 養成講座

Materials Genome/Informatics and Chemo-Informatics Activate Networks

第3b回ポリマーの配合処方設計

2018.8.25 横浜国大非常勤講師 山本博志

第3a回では、フッ素系ゴムを例題に、Materials Informatics と Materials Infrastructure を中心に解説を行った。最初は Materials Integration も一緒に書いてしまおうか思ったのだが、あまりに長くなるので分割することにした。材料の複合化と言った場合、従来の処方設計と、最新のコンピュータを用いた複合化設計と2種類ある。当然、Materials Integration という時には、後者の複合化設計をさす。しかし、従来の処方設計に対しても、Materials Genome は非常に有効だ。

フッ素ゴム系のパッキンの配合設計

先回述べた方法で設計されたポリマーは生ゴムと呼ばれる。この生ゴムに様々な副資材を配合して、混練し、架橋させることによってゴムが得られる。主な添加剤として受酸剤、充填剤、加硫剤、可塑剤と加工助剤が挙げられる。受酸剤は加硫の時に生成する酸性物質を中和するために加えられ、主なものに、MgO, PbO, ZnO, CaO, Ca(OH)₂がある。充填剤としてはカーボンブラック、無機物のフィラーがある。加硫剤としてはポリアミン、ポリオール、パーオキサイドなどの過酸化合物がある。可塑剤としてはDOP, PET、低分子量FKM、加工助剤としてはカルバナワックス、ポリエチレンなどがある。

多くの場合、生ゴムの製造メーカーが最終製品のパッキンまで製造することはなく、加工メーカーが生ゴムも含め資材を調達し、パッキンを作成する。こうした原料は様々な資材メーカーが販売している商品なので、原料の詳しい諸元はわからない。そうした商品名しか分からない資材を使って素材を設計して行くのにも Materials Genome は有効に使える。

まず、フッ素樹脂ハンドブックからデータを集め、テーブルを作成する。最後の4列が出来上がったゴムの物性値になる。FKV 種類からトリアジンまではゴムのシステム要件なる。

FKV種類			カーボンブラック						受酸剤					パーオキサイド					トリアジン	100%引張応力	破断強度	伸度	硬度	
G751	G701	G901	MT	FT	SRF	APF	FEF	HAF	MgO	MgO1	PbO	ZnO/P	CaO	CaOH2	3M	V	D	F	PH2.5B	TAIN	Kg/cm2	Kg/cm2	%	JIS A
100	0	0	20	0	0	0	0	0	3	0	0	0	0	2	0	0	0	0	0	0	36	137	200	71
100	0	0	20	0	0	0	0	0	3	0	0	0	0	4	0	0	0	0	0	0	43	160	200	71
100	0	0	20	0	0	0	0	0	3	0	0	0	0	6	0	0	0	0	0	0	50	158	200	71
100	0	0	20	0	0	0	0	0	3	0	0	0	0	8	0	0	0	0	0	0	56	164	200	73
100	0	0	20	0	0	0	0	0	3	0	0	0	0	10	0	0	0	0	0	0	58	161	190	74
0	100	0	20	0	0	0	0	0	3	0	0	0	0	6	0	0	0	0	0	0	49	142	210	70
0	100	0	20	0	0	0	0	0	0	3	0	0	0	6	0	0	0	0	0	0	52	134	210	71
0	100	0	20	0	0	0	0	0	0	0	3	0	0	6	0	0	0	0	0	0	49	115	220	72
0	100	0	20	0	0	0	0	0	0	0	0	3	0	6	0	0	0	0	0	0	63	119	170	73
0	100	0	20	0	0	0	0	0	0	0	0	0	3	6	0	0	0	0	0	0	44	99	220	71
0	100	0	20	0	0	0	0	0	0	0	0	0	3	6	0	0	0	0	0	0	42	98	220	68
0	100	0	20	0	0	0	0	0	2	0	0	0	0	6	0	0	0	0	0	0	49	129	210	71
0	100	0	20	0	0	0	0	0	4	0	0	0	0	6	0	0	0	0	0	0	53	152	210	71
0	100	0	20	0	0	0	0	0	6	0	0	0	0	6	0	0	0	0	0	0	53	158	210	70
0	100	0	20	0	0	0	0	0	8	0	0	0	0	6	0	0	0	0	0	0	58	171	210	71
0	100	0	20	0	0	0	0	0	10	0	0	0	0	6	0	0	0	0	0	0	60	169	220	72
0	100	0	20	0	0	0	0	0	15	0	0	0	0	6	0	0	0	0	0	0	59	176	230	74
0	100	0	20	0	0	0	0	0	20	0	0	0	0	6	0	0	0	0	0	0	56	148	310	81
0	100	0	20	0	0	0	0	0	3	0	0	0	0	6	0	0	0	0	0	0	50	149	210	70
0	100	0	20	0	0	20	0	0	3	0	0	0	0	6	0	0	0	0	0	0	94	185	190	77
0	100	0	20	0	0	20	0	0	3	0	0	0	0	6	0	0	0	0	0	0	121	209	170	81
0	100	0	20	0	0	20	0	0	3	0	0	0	0	6	0	0	0	0	0	0	131	207	160	81
0	100	0	20	0	0	20	0	0	3	0	0	0	0	6	0	0	0	0	0	0	133	207	170	85
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.5	2	25	178	420	67
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.5	4	29	207	380	72
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.5	6	32	247	370	73
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	1.57	0	0	0	0	0	25	212	360	70
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	1.73	0	0	0	0	26	223	340	69
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	0	2.78	0	0	0	28	215	350	72
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.74	0	0	28	220	370	72
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.5	0	26	217	370	70
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	4	29	204	380	70
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	26	196	350	71
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	4	29	210	380	73
0	0	100	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	4	29	208	380	73

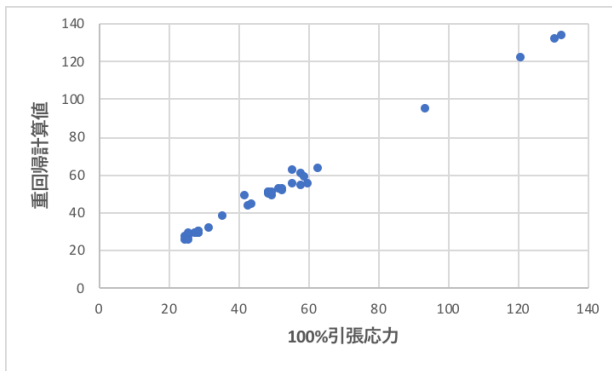
これは、目的変数が4つと説明変数が21、データ数が35ある場合の解析になる。受酸剤、トリアジン以外は商品名

で、その説明変数は加えた量だけがわかっている。それが目的変数にどのような影響を与えているかがわかれば、4

つの目的変数のうち3つを固定し、残りの1つの値だけをあげたいときにどのような処方を取ればいいのかを計算で出すことができる。

それでは、4つの目的変数に対して、順番に重回帰法で予測式を作成しよう。ソフトウェアは授業で配布したYSBを用いても、Excelに搭載されている重回帰分析でもどちらでも良い。

まず、100%引張応力を計算してみよう。



上図に示すように、ほぼ綺麗に100%引張応力が計算できていることがわかる。ここで作成された式の意味を確認しておこう。

100%引張応力=

$$\begin{aligned} &0.7462 * G751 + 0.7629 * G701 + 0.6917 * G901 \quad (\text{生ゴム}) \\ &- 2.2666 * MT - 2.2584 * FT - \\ &0.0592 * SRF + 1.2938 * APF + 1.7934 * FEF + 1.8920 * HAF \quad (\text{カーボンブラック}) \\ &+ 0.6610 * Mg0h + 1.2573 * Mg0l + 0.2573 * Pb0 + 1.4772 * Zn0 / P - \\ &0.8456 * Ca0 + 2.8500 * CaOH2 \quad (\text{受酸剤}) \\ &+ 0.7285 * 3M + 1.2392 * V + 1.4906 * D + 2.3815 * F + 0.4526 * PH2.5B \\ &\quad (\text{パーオキサイド}) \\ &+ 1.0176 * TAIN \quad (\text{トリアジン}) + 0.0156 \quad (\text{定数}) \end{aligned}$$

という式で表されていることが理解できるだろう。

生ゴムの種類としては、100%引張応力を一番高くするのは、G701、一番小さくするのはG901であることがわかる。

カーボンブラックの種類としては、MT, FT, SRFは100%引張応力を小さくし、APF, FEF, HAFはこの順に大きくする。

受酸剤はCa0のみ100%引張応力を小さくする。

パーオキサイドは全て100%引張応力を大きくするが、パーオキサイド-Fがその効果が最も高い。

従って、100%引張応力を一番大きくしたければ

G701/HAF/Mg0l-CaOH2/F/TAIN系が良いことになる。

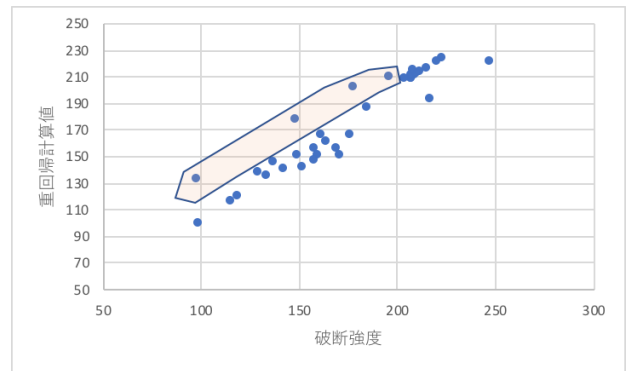
逆に一番小さくしたければ、

G901/MT/Ca0-CaOH2/系が良いことになる。

ただし、この式はあくまで、100%引張応力に関するもので、

処方を変えれば、破断強度、伸度、硬度の値は変化し、また添加物の与える効果は、100%引張応力の時の効果とは当然異なる。また、処方する時の量も考えに入れて置く必要がある。生ゴムの係数は大きくは変わらないが、使用量は100部なので、影響は最も大きくなる。

次に破断強度に関して、同じように重回帰分析をしてみよう。(伸度、硬度については自分でやってみよう)



先ほどと異なり、少し精度が低くなる。重回帰分析では各説明変数の係数は固定の値になる。ところが実際には、ある量を超えたら急に値が大きくなる(非線形性)、ある組み合わせの時に特異的に値が小さくなる(相互作用)ことがある。そのような現象に対しては重回帰計算では精度が出なくなる。それでは、どうしたらいいか? 非線形回帰のニューラルネットワーク法(NN法)を用いるのも手ではあるが、NN法は使いこなすのが非常に難しい技術である。何故だろうか? このぐらいの非線形性であれば簡単に精度がアップされ、しかもそれが何故だかわからないからだ。従ってNN法で予測されるものを作ってみても予想された性能が出ない。NN法で学習させ直して、を繰り返すことになる。そして、ついに諦める。つまり、この程度のデータ数で、予測性能が十分期待できるニューラルネットワークを構築するのは、それなりの知識と経験が必要だからだ。慣れてくると比較的簡単にできるようになる。

どうしたら慣れることができるのだろうか?

データをよく見ることに尽きる。

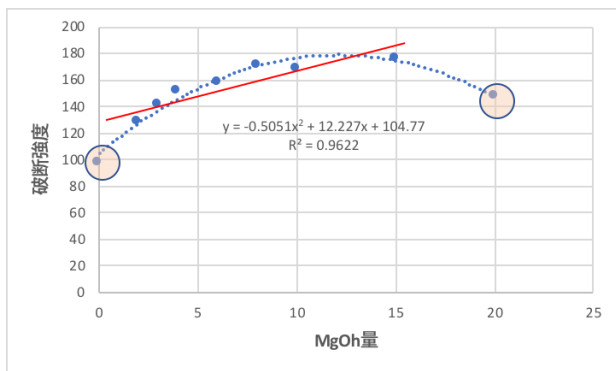
合わないデータは実験誤差だろうと消去するようでは、化学系のデータサイエンティストは目指さないほうが良い。

それでは、先ほどの図で特定した4つのデータを、テーブルで特定しよう。RUN-11, 18, 24, 33が相当する。

RUN-11, 18はMg0hが関与する系、RUN-24, 33はPH2.5BとTAINが関与する系であることがわかる。

さらに詳しく見て行こう。

受酸剤を用いる系は必ず CaOH₂ とそれ以外の受酸剤のペアで用いている。そこで、テーブルの中で MgOh の量だけをふったRUNを抽出してサブセットのテーブルを作成する。



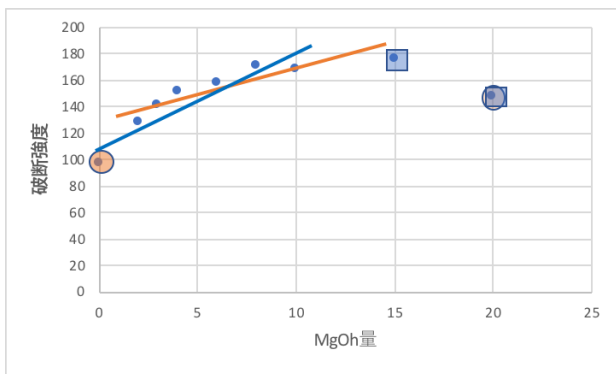
グラフ化して見ると明らかであるが、RUN-11 は左端、RUN-18 は右端の処方になる。重回帰解析では、結果は線形になる。そこで他の受酸剤の系も含めると、図中の赤い線が、平均的に一番誤差が小さくなると解釈される。しかし、実際は上に凸の2次関数になることがカーブフィットの結果からわかる。従って非線形性を導入しなければ予測式の精度は上がらない。しかし、このデータは MgOh の量以外は全て一定の時に成立しているので、単純に上に凸の2次関数を予測式に組み込むのは難しい。

取り敢えず破断強度予測式の、精度を上げるのに RUN-11, 18 を除外して精度を上げるか、他のものを除外するか選択が別れる。そして最終的に出来上がる予測式のパフォーマンスも異なる。

その際に一番重要なのは、「実験を行っている研究者とよく相談すること」であろう。

その際に、予測式を合理的に説明するのは、データサイエンティストの責務になる。

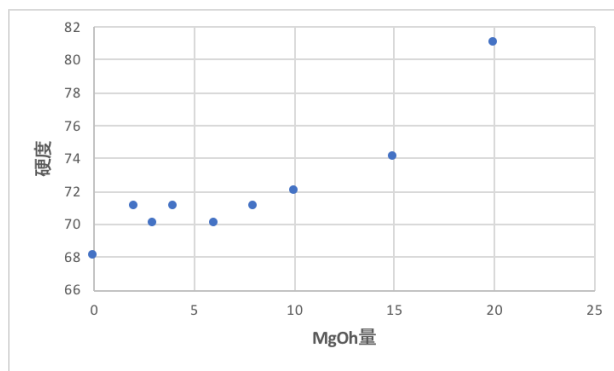
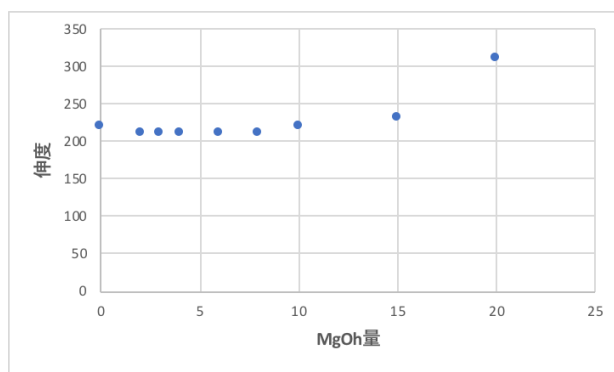
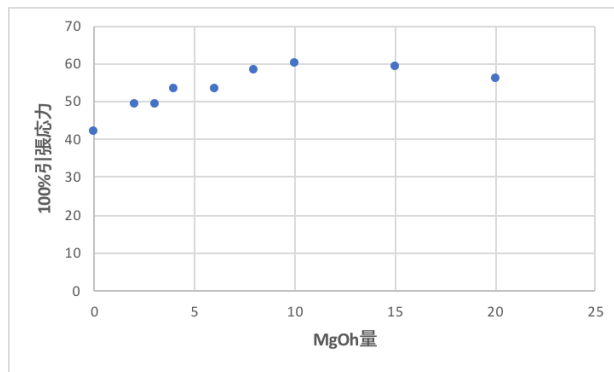
一つの解釈を示そう。



一番左と一番右を外した場合は、重回帰分析の結果になり、平均的には良くなる(赤い線)。しかし、この場合は、一番

右2つを外して考えるべきである(青い線)。

MgOh の量を変化した時の、他の物性をプロットすると次のグラフになる。特に伸度について、一番右の点は、急に伸度が高くなっている。つまり、MgOh の量が 15 を超えると、伸度が大きくなり、破断強度が低下する。ゴムとしてだらしのないゴムになる。



また、硬度や100%引張強度についても右の2つは異常値をとる。そのような通常のゴムの特性を外れた領域を探索したいなら、右の二つは外せないが、通常のゴム領域で探索したいのであれば、右の二つは外すべきである。ただし、通常のゴム領域を外れたものを探索したいなら、明らかにデータ不足なので、受酸剤量を大きくしたデータを取得すべきである。右の2つを外すのであれば、予測式としては受酸剤量が多いものに対しては予測できないことを納得して頂きたい。

こうした事を、実験系の研究者と打ち合わせるのである。

RUN-24, 33 の PH2. 5B と TAIN が関与する系については、自分で考えてみよう。

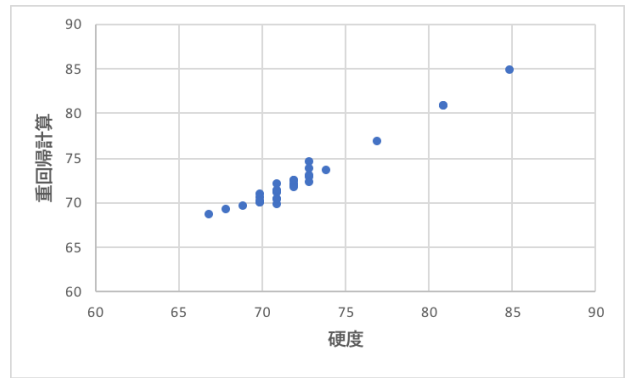
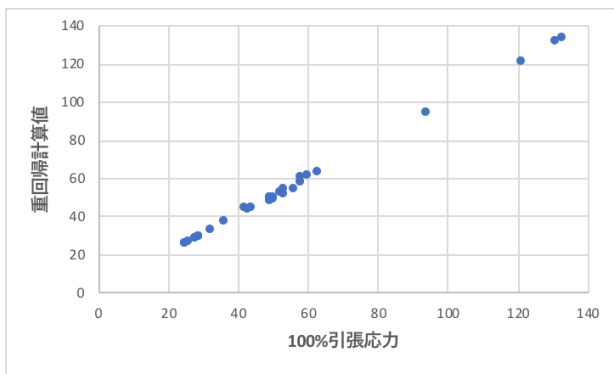
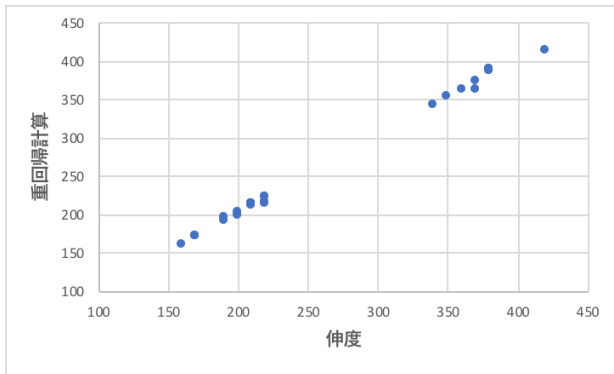
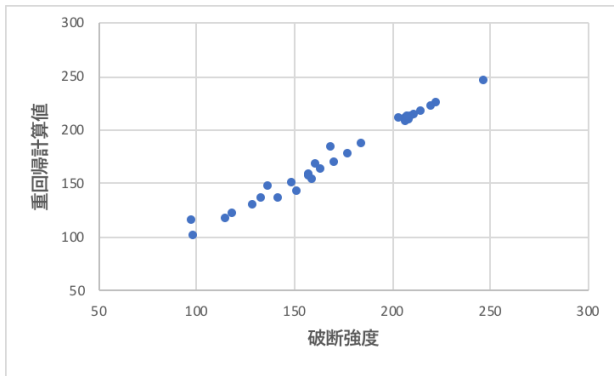
RUN-24, 25, 26, 31 を比較して見解を導いてみよう。

RUN-33, 25, 32, 34, 35 を比較して見解を導いてみよう
ヒントは PH2. 5B と TAIN は両者をペアで使うもので、ペアでないものの取り扱いをどうするか？

もう一つのヒントは実験誤差の可能性だろう。

実験誤差を非線形性や相互作用に落とし込んで予測式を作成すると、多くの場合性能低下につながる。非線形性を考慮するのは、データのクレンジング（第5回で説明）を行い、それでも必要であった解に留めたほうが良いだろう。

筆者の理解で4つほどデータを除外して重回帰解析した結果を以下に示そう。



受酸剤の多い系には使えないが、非線形性を導入しなくても十分な精度で実験結果を再現できていることがわかるだろう。逆に言えば、重回帰分析で精度の出る系をニューラルネットワーク法のような非線形回帰解析する事の無意味さも理解できるだろう。

今回の場合は、生ゴムを2種類以上混合した系、カーボンブラックを混合した系、パーオキシドを混合した系は無い。(そうした系が増え、重回帰解析でどうしても精度が出なくなったら、その時にNN法を利用すれば良い。)

ここまで出来上がると、Excel上に計算機が構築できる。処方を入力すると、物性値を計算する。

B	C	D	E	F	G	H
生ゴム	G751	G701	G901			
	100	0	0			
カーボンブラック	MT	FT	SRF	APF	FEF	HAF
	20	0	0	0	0	0
受酸剤	MgOh	MgO1	PbO	ZnO/P	CaO	CaOH2
	3	0	0	0	0	2
パーオキシド	3M	V	D	F		
	0	0	0	0		
パーオキシド・ペア	PH2.5B	TAIN				
	0	0				
予測値	100%引張	伸度	破断強度	硬度		
	37.0	202.3	146.5	70.4		
制限	生ゴム、カーボンブラック、パーオキシドは1種類のみ選択。					
	受酸剤は、入れる際には、CaOH2とのペアで。					
	パーオキシド・ペアはペアで値を入れ、他のパーオキシドと併用しない					

海外のメーカー品と比べ、自社の製品は（例えば）破断強度が足りない。他の物性はそのままだ、破断強度を10%だけ向上させるのに、処方をどういじるか？ エクセル上で処方をいろいろ検討して、一番よくなりそうなもので実際に実験を行う。結果が予測値に合えば良いが、合わなければ予測式を再構築する。この過程を素早く回せるかが一番重要になる。

各項目の重回帰係数が与えられているので、例えばカーボンブラックの種類を変えたら、各物性値がどう変わるかは理解しやすいだろう。それにしても4つの目的変数を全て満足する処方を設計するのは大変かもしれない。その際には、遺伝的アルゴリズムを用いて、自動探索してしまうのは一つの方法であろう。そこまでやれば、胸を張って最新

の Materials Integration と呼べるだろう。制限事項をプログラムに反映させるのは少し手間かもしれないがチャレンジしてほしい。

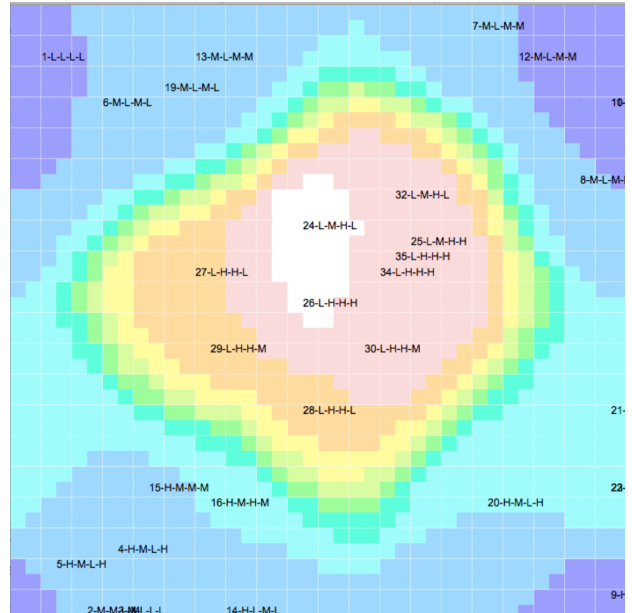
自己組織化ニューラルネットワーク法

SOM(Self Organization Map)という解析方法がある。これは定性的解析法として優れているので、授業でもよく取り上げている。この方法を用いてフッ素ゴムの4つの物性を検討してみよう。SOMのプログラムは所定のURLにアクセスして用いる。まずは、下図のテーブルを用意する。100%引張強度、破断強度、伸度、硬度を順にソートをかけて、データにL,M,Hのラベル付けをする。そして、RUN-no も含め、24-L-M-H-L の様に表現する。

RUN	Label	100%引張応力	破断強度	伸度	硬度
24	24-L-M-H-L	25	178	420	67
11	11-M-L-M-L	42	98	220	68
28	28-L-H-H-L	26	223	340	69
6	6-M-L-M-L	49	142	210	70
19	19-M-L-M-L	50	149	210	70
14	14-H-L-M-L	53	158	210	70
27	27-L-H-H-L	25	212	360	70
32	32-L-M-H-L	29	204	380	70
1	1-L-L-L-L	36	137	200	71
3	3-M-L-L-L	50	158	200	71
2	2-M-M-L-M	43	160	200	71
12	12-M-L-M-M	49	129	210	71
7	7-M-L-M-M	52	134	210	71
13	13-M-L-M-M	53	152	210	71
15	15-H-M-M-M	58	171	210	71
10	10-M-L-M-M	44	99	220	71
8	8-M-L-M-M	49	115	220	72
16	16-H-M-H-M	60	169	220	72
29	29-L-H-H-M	28	215	350	72
30	30-L-H-H-M	28	220	370	72
25	25-L-M-H-H	29	207	380	72
9	9-H-L-L-H	63	119	170	73
4	4-H-M-L-H	56	164	200	73
26	26-L-H-H-H	32	247	370	73
35	35-L-H-H-H	29	208	380	73
34	34-L-H-H-H	29	210	380	73
5	5-H-M-L-H	58	161	190	74
20	20-H-M-L-H	94	185	190	77
22	22-H-M-L-H	131	207	160	81
21	21-H-H-L-H	121	209	170	81
23	23-H-H-L-H	133	207	170	85

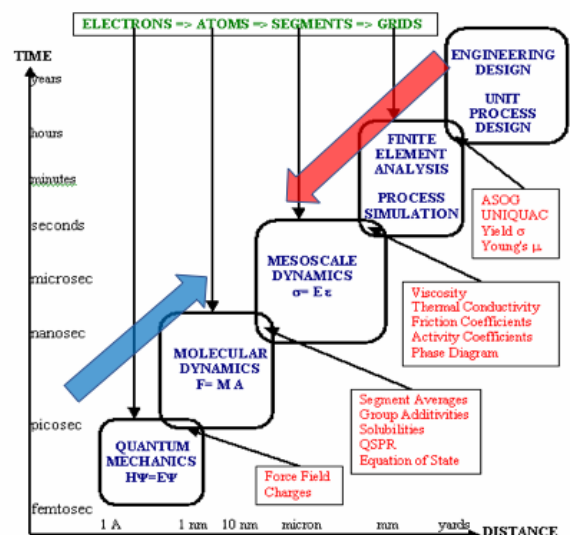
このテーブルをタイトル、RUN-No まで全てコピーし、SOMのプログラムにペーストし、自己組織化させる。つまり、100%引張強度、破断強度、伸度、硬度を4次元のベクトルとした時に、似たベクトルは似た2次元上の位置にマッピングしてみる。多次元ベクトルを2次元マッピングした結果は下図の様に表現される。24-L-M-H-Lは図中のちょうど真ん中にある。このL-M-H-Lのパターンが近いものは、近い位置に配置される。4つが同じパターンではほぼ重なり、3つが同じものは近い位置に来る。例えばRun-24の破断強度をMからHに変えた、L-H-H-Lの処

方を検討しようとした時には、RUN-27,28を参考にすれば良いことがわかる。RUN24とRUN-27,28ではパーオキサイドの種類が違う。中々そうした系統の異なる実験はしないものなので、情報学のこうした提案は新鮮な提案になるはずだ。



これまでのMIの主流は、第1回で紹介した Multiscale Computational Hierarchy を基礎にしている。原子、分子、分子集合体、物質と、図中左下から右上に向かって研究が進行している。

Multiscale Computational Hierarchy for Materials Simulations.



しかし、この方法で Material Integration するには、生ゴムと、例えばカーボンブラックの相互作用を見積もり、架橋が時間とともにかかり粘度が変化する過程を解き明かさなくてはならない。しかも、カーボンブラックの諸元の違いなどを知ることなしに。そこで、この方法で左下から

右上に登っていこうとすると、材料の複合化は非常に難しい、乗り越えるのが大変な壁があることがわかる。しかし方向を逆に考えて、すでにあるマクロの物性値を構造に落とし込んでいくと、比較的簡単に複合化を達成できることを今回示した。いわゆる計算機化学というと、反応の様なミクロな、非常に短い時間の問題を解くには有利である。化学工学は実時間、実サイズを扱う。化学情報学は、そうした化合物の持つ情報を、全て受け入れてしまう懐の深い学問である。コンピュータを使うので計算機化学の一部と思われがちだが、それは大間違いである。

雑感

筆者は授業の際に、毎年同じアンケートを行う。年々文章は短くなり、語彙は減って行っている。三行を超える長い文章の意味を捉える能力が低くなっている。長い解説も同じであろうから、分割することにした。SNSの時代、しようがないことではあろうが、ここまで本を読まない大学(院)生は世界に稀であろう。デジタル・ネイティブなどと言われている様であるが、全く逆だと思う。学生はマテリアル・ゲノムやインダストリー5などという言葉すら知らない。知らないものは検索もできないので、無いのと同じだ。筆者のHPなども学生にとっては存在しないのと同じなのであろう。それでも、今の時代、筆者の書いたものは自動的に翻訳され世界中の学生が読む。自動翻訳させるなら、複雑な係り受けはしない。難しい単語は使わない。文章は短めに。あれ？日本の大学生が読んで理解できる様

に書けば自動翻訳させやすいのか。。。そんな状況で売り手市場とか言って浮かれている。もう4-5年してAI-ロボットが本格的に実用化された時に“そんなの聞いていないよ”とならない様に祈るしかない。

ビッグデータと自然言語解釈、画像認識の技術は、いまは華やかであるが、化学や最先端素材開発と親和性が高いわけでは無い。この分野にはビッグデータなどそもそも存在しないし、MOやMDでデータを水増しさせようにも、商品名しかわからない様な材料は計算のしようが無い。しかし、現実問題としては今回紹介した様な方法で、研究を加速しているところは実在している。

Pirika [マテリアル・ゲノム](#) のページ

以下PDF

[第1回 イントロダクション](#) 2018.8.23

[第2回 データ収集と昔ながらのやり方](#) 2018.8.24

[第0回 物性推算と逆設計と呼んでいた時の話](#) 2000.8.28

なんと18年前！

[第3a回 ポリマー設計と3つのMI \(その1\)](#) 2018.9.3

[第3b回 ポリマー設計と3つのMI \(その2\)](#) 2018.9.3

[第4a回 MIに適した簡単なデータベースの利用法](#)
2018.9.4

[第4b回 複雑なポリマーのデータベース化](#) 2018.9.7

プレゼン用：[MIを使う時のデータベース構築法](#)
2018.9.11

プレゼン用：[複雑なポリマーの設計とDB](#) 2018.9.15

[第5回 データのクレンジング](#) 2018.8.28

[第6a回 ニューラルネットワーク法の初歩](#) 2018.9.25

[第6b回 ニューラルネットワーク法を使ったDrug Design](#)
2018.9.22

第7回 遺伝的アルゴリズム(GA)を理解しよう