

MAGICIAN 養成講座

Materials Genome/Informatics and Chemo-Informatics Activate Networks

第2回 データ収集、データ処理

2018.8.25 横浜国大非常勤講師 山本博志

第1回でも触れたように、実験結果を統計解析して、その予測式をもとに材料の研究開発を進める事は、3-40年前から普通に行われていた事である。その際には今と違ってビッグデータが必要などとは言わなかった。データが1点では何もわからない。2点あればだいたいの傾向がわかる。3点あれば誤差も含めて傾向がはっきりしてくる。大事なものはある現象に対してどれだけ、「知識と経験に基づいた合理的な式が立てられたか」だろう。それを知識も経験もないAIにやらせようとしたら、良いもの、悪いものを大量に教えるしかない。良いものだけを教えたら、当然、悪くなることを理解できないAIになる。実験のうち99%は役に立たないデータしか取れず、特許や論文に残っているものはせいぜい1%の成功例だろう。AIにはそれをバランスよく教えないといけない。最先端の話ではなく、2-30年前には普通に行われていた温故知新の技術では、データをどう集めたか、その集めたデータをどう処理したかを再確認しておこう。

データマイニング

マイニングというのは、探鉱(有意義な鉱石を土の中から発掘する)のことだ。そこでデータマイニングというのは、膨大なテキスト、画像の中から、有意義なデータを探し出すことだ。

昔何かで読んだのだが、面白い実験があった。例えばある文章がある。その中からランダムに文字を虫食いにする。何%虫食いになったら意味が通じなくなるかを作者ごとに調べていた。確か筒井康隆と小林秀雄を比較していたと記憶しているが、かたや30%虫食いでも意味が通じるのに対しもう一方は5%で意味が通じなくなる。デジタル化時代どちらが有用な情報源かといえば当然、30%虫食いでも意味が通じる方であろう。OCR(光学文字認識)と言っても完全では無い。文字の大きさがマチマチ、縦書き、横書き、アルファベットとギリシャ文字、数字が混在の化学系の論文、特許をAI搭載のOCRソフトに読ませたところで、認識率は95%を越えるのも大変であろう。そうした不完全なデータからAIが知識を獲得するのは非常に困難となるだろう。

それに対して人間の脳は非常に曖昧にデータマイニングする機能を持っている。下の文章をさーっと流し読みしてみよう。

「みささん、まだまださむいですが、おんげきですか。かぜなどひいてないいいですか。」

ほとんど違和感なく意味がつかめると思う。

Typoglycemiaとは、単語を構成する文字を並べ替えても、

最初と最後の文字が合っていれば読めてしまう現象のことである。脳のこの機能の為に、文章はいくら推敲しても直せないものは直せないことになる。(最近のWordなどは自動で赤い波線を引いてくれるが、修正候補は出てこない)当然検索しても探せない文章になってしまうが、人間が読めば意味の通じる文章になる。

こうしたことから、これから先は、e-実験ノートなどを導入して失敗例も含めて情報を蓄積していこうという一つの流れがある。それはそれで正しいと思う。

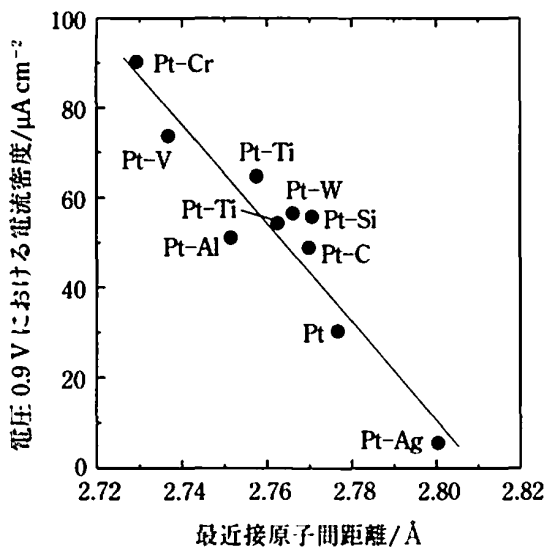
そして、過去については手書きの文章は無理でも、印刷されたものは(精度は低いかもしれないが)デジタル化してテキストデータマイニングしよう。過去のグラフなどをデジタル化して画像認識しよう。などと言う議論に繋がっていくのだろう。その際に、現在のAIが成功しつつある、自然言語解析、画像認識がビッグデータ解析によって行われているので、材料開発もビッグデータが必要と誤解しているように思える。いくら材料系論文、特許解釈用のAIを開発したところで、人間の脳には叶わない。それなら人間の手間のかかる部分だけをアシストするソフトを開発して、解釈は人間が行うしかないと思う。

テキストデータマイニング

今更、筆者が言うまでも無いことだが、Googleの検索はテキストデータマイニングの最高峰に位置していると思う。例えば、「燃料電池 触媒」と入れて検索すれば、「fuel cell」と英語表記のページも検索に引っかかってくる。Typoglycemiaでわざと「feul cell」と入れて検索しても、自動的に「fuel cell」の結果を表示する。それなら

ば、わざわざテキストデータマイニングのソフトを開発しなくても、Google の検索を社内用のサーバーにも適用すれば良いだけになる。法人向け Google Apps を利用するのが一番簡単であろう。PDF のファイルを OCR する機能も、Google ドライブには搭載されている。筆者のような個人ユーザーでは使えるような値段では無い。幸い、筆者は Mac ユーザーなので、昔から (Apple Sherlock: 1998 年) ファイル内の単語まで検索できた。現在 700 冊の専門書、PDF、特許 (36044 件、23GB) が Mac のハードディスクに溜め込まれている。

ファイルは OCR されているので、”燃料電池 触媒” と検索すると、1054 件ヒットする。その中で、科学マスター講座 触媒化学 丸善出版 (2011) に、白金と卑金属との合金でさらなる高活性化が図られているとある。



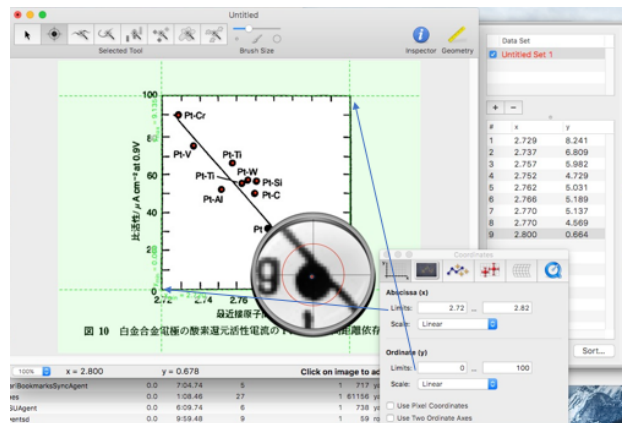
元文献：

V. Jalan et al., J. Electrochem. Soc., 130, 2299 (1983)
 書籍にはグラフは記載されているがテーブルは無い。そこでグラフをデジタル化する必要がある。

画像データマイニング

例えば上図のようなグラフがあった時にデジタイザーを使って数値を拾うことは、昔はよくやられていた。様々な計測装置はペンでアナログなチャートを残すだけだった。民生用のハードウェア・デジタイザーは 1983 年 Apple II 用が初めてだそう。その後、マッキントッシュになってからは、すぐにソフトウェア・デジタイザーがあったと思う。そこで、筆者はもう 30 年近くこうしたグラフから数値を起こして利用してきたことになる。マック用で筆者が愛用しているのは、GraphClick というフリーウェアだ。様々なタイプの画像を読み込み、フレームとス

ケールを設定した後、拡大鏡を使いながら読み取り点をクリックして行くとデータがデジタルで得られる。非常に優れたソフトであるが、ついに開発は止まってしまった。授業の際には、学生はほとんど Windows なので、Windows 用のデジタイザー・ソフトを探してもらっている。毎年ちゃんと見つけれられているようなので、Windows 版にも優れたフリーウェアがあるのだろう。



それでは、こうしたソフトの、AI 自動化はどうであろうか？ 論文とか特許からグラフを探してきてグラフをテーブル化できるだろうか？ このくらいシンプルなグラフならどうにかなるかもしれないが、Pt-Ti に横棒が入っているのをちゃんと理解できるか？ 混み合った領域でどの点がどれだかちゃんと認識できるか？ 複数のスレッドが存在する場合、マーカーが重なっているのが理解できるか？ フレームの一部を省略する 2 重波線などを理解できるか？ などなど問題は多いように思える。

また、このくらい鮮明な画像であれば誤認識は少ないだろうが、Ti と TI などは OCR での誤認識が多い。C5H11 (シクロペンチル基) など、正式に下付きの数字にした、C₅H₁₁ を使っていると、C₅H₁₁ と、C5 が Cs (セシウム) になったりする。

実は、このグラフには Pt-Ti は 2 回出て来る。

TI の誤植かも知れない。

同じように作ってもこのぐらいの誤差が出る。

Ti には α と β フォームがあるからだろう。

人間の実験者であれば、最低そのぐらいは考えるだろうが、AI に自動的に行わせた場合、どこまでを期待するか？ は難しい問題になる。

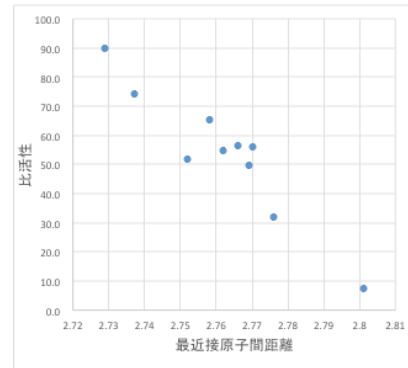
結局ある程度まではソフトウェアが行い、詰めのところでは人間が対話型で処理しなければならないのでは無いだろうか。その対話を通じて成長できるような AI であれば、教えれば教える程、人間は楽になって行く。しかし、AI がその特徴の高速性を発揮して、対話が何時までも律速

になるのならビッグデータ収集という目的からは外れてしまう。

おそらく、コンピュータや AI に適したデータはあくまでもテーブル・データなのだろう。人間の脳と言う低速で曖昧なプロセッサでも直感的に理解できるようにグラフ化された段階で、情報は縮退されてしまっている。そこから情報を引き出すためには、修士のレベルの専門性を理解

Alloy	最近接原子間距離	比活性
Cr	2.729	89.9
V	2.737	74.1
Ti	2.758	65.4
Al	2.752	51.8
Ti	2.762	55.0
W	2.766	56.7
Si	2.77	56.0
C	2.769	49.8
Pt	2.776	31.8
Ag	2.801	7.5

し、直感を持った AI が必要になる。当分の間は人間が関与しなければならないのであれば、テキストデータマイニングの段階で、重要な論文、特許を優先していくという判別が重要となる。ここでは取りあえず、人力でデジタル化したグラフデータを解析する事を考えてみよう。もとの図と比べて、ほぼ問題のない精度でデジタル化できていると言えるだろう。



識別子(Descriptor)の収集

Pt 合金の比活性は最近接原子間距離でほぼ決まる事がグラフから読み取れる。それでは、このグラフに無い原子との合金の比活性はどのように予測できるだろうか？片端から合金を作成し、X線などを用いて原子間距離を測定し、最適合金の種類と配合比を決めていけば良い。そうした実験手法を HTP(High Through Put)法といい、それはそれで合理的な実験手法である。

それに対して、原子の持つ化学的な情報から、絞り込んで行くのが Chemo-Informatics(化学情報学)的なやり方である。マテリアル・ジノムや MI(material informatics)はこちらの手法をとる。それを行うには、原子の持つ化学的な

情報、識別子(Descriptor)を収集する。

例えば、フリーウェアの Periodic Table of the Elements という Javascript のプログラムは周期律表の原子をクリックすると様々な原子情報を与えてくれる。化学便覧や化学データハンドブック | 無機・分析編にも重要なデータが記載されている。こうした情報を比活性のテーブルに加えて行く。

この部分の自動化についてはどうであろうか？単純な原子については 100 種類程度しか存在しないので、ある程度は自動化も可能であると思われる。

Periodic Table of the Elements

Click on an element to view its properties

Symbol Color

- Black: Solid
- Blue: Liquid
- Red: Gas
- Grey: Synthetically Created

Background Color

- Green: Nonmetals
- Light Blue: Alkali Metals
- Light Purple: Alkali Earth Metals
- Light Green: Transition Metals
- Light Cyan: Other Metals
- Pink: Metalloids
- Yellow: Halogens
- Light Orange: Noble Gases
- Light Grey: Rare Earth Metals

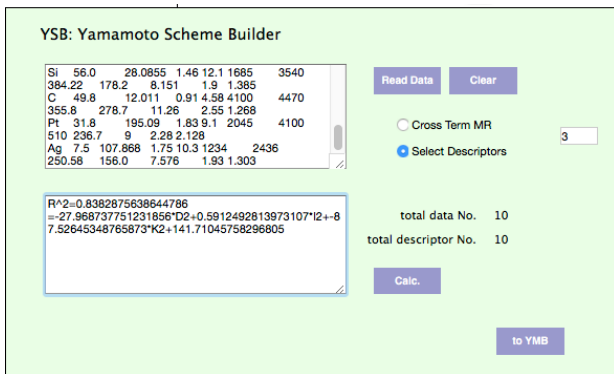
Close Element Window

Density: 21.4 g/cm³
 Oxidation States: 4; 2
 Electron Configuration: [Xe]4f14-5d9-6s1
 State: Solid
 Melting Point: 2045 K
 Boiling Point: 4100 K
 DH_{vaporization}: 510 kJ/mol
 DH_{fusion}: 19.6 kJ/mol
 Specific Heat: 0.13 J/(g·K)
 Group: 10 (VIII)
 Family: Transition Metal
 Crystal: Fcc
 Atomic Volume: 9.1 Å³
 Atomic Radius: 1.83 Å
 Covalent Radius: 1.3 Å
 Thermal Conductivity: 0.716 1/(μW·cm)
 Electrical Conductivity: 0.0966 1/(μW·cm)
 Ionization Potential: 9 eV
 Electronegativity: 2.28

Alloy	比活性	Atomic Mass	Atomic Radius	Atomic Volume	Melting Point	Boiling Point	$\Delta H_{\text{vaporization}}$	SP	Ionization Potential	Electronegativity	ElectronAffinity
Cr	89.9	51.996	1.85	7.23	2130	2945	344.3	218.2	6.766	1.66	0.66
V	74.1	50.9415	1.92	8.78	2175	3682	447.02	225.6	6.74	1.63	0.5
Ti	65.4	47.9	2	10.64	1943	3562	421	198.9	6.82	1.54	0.2
Al	51.8	26.98154	1.82	10	933.25	2740	293.4	171.3	5.986	1.61	0.46
Ti	55.0	47.9	2	10.64	1943	3562	421	198.9	6.82	1.54	0.2
W	56.7	183.85	2.02	9.53	3680	5828	824	294.0	7.98	2.36	0.6
Si	56.0	28.0855	1.46	12.1	1685	3540	384.22	178.2	8.151	1.9	1.385
C	49.8	12.011	0.91	4.58	4100	4470	355.8	278.7	11.26	2.55	1.268
Pt	31.8	195.09	1.83	9.1	2045	4100	510	236.7	9	2.28	2.128
Ag	7.5	107.868	1.75	10.3	1234	2436	250.58	156.0	7.576	1.93	1.303

そして、比活性と相関の高い識別子を絞り込んでいく。テーブル中のSPというのはSolubility Parameter (溶解度パラメータ)という物性値で、蒸発潜熱を体積で割った値、ルートを取ったものになる。物質の相溶性などの指標として重要な値である。

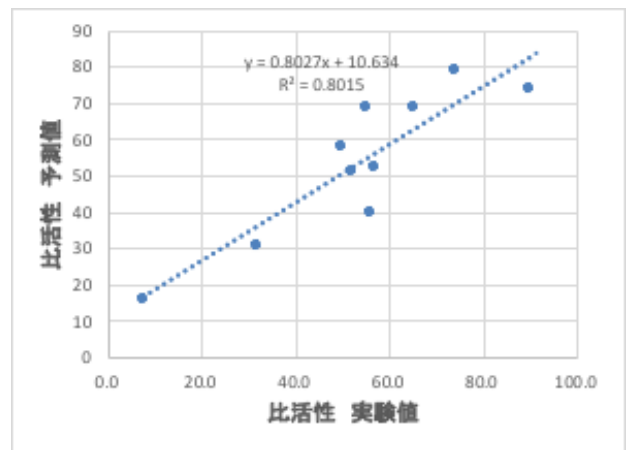
多くの場合、一つの識別子とだけで比活性と高い相関(単相関)が得られる事は無い。特に炭素(C)は金属では無いので識別子の値が異常値を取る。そうした時には複数の識別子を使って重相関を検討する。複数の識別子から有用な識別子を選択して重相関式を作るには、変数選択重回帰法を使う。ソフトとしては、大学の授業の初回に配布しているYMSBを利用する。使い方は授業の際に徹底的にやっているのだからわかるだろう。(初めての方は、"Pirika MOOC"で検索して使い方を理解しておいてほしい。)



先ほど作った比活性のテーブルをコピーして、データを読み込み、識別子を3つ選択し、比活性を予測する式を作成する。

$$\text{比活性} = -27.969 \cdot \text{Atomic Radius} + 0.5912 \cdot \text{SP} - 87.5264 \cdot \text{electron negativity} + 141.7104$$

となった。結果をグラフで表示すると下図のようになる。

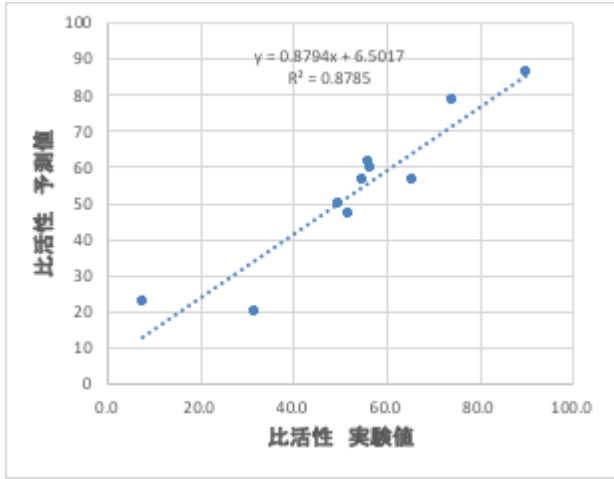


YSB プログラムは、あくまでも”相関係数が高くなるように”3つの変数を選択する。YSBが化学的に意味のない変数を選択していたら、いくら相関係数が高くても予測性能は出ない。そこで化学者の勘と経験で、その列を取り除いて、変数選択を繰り返す。こうした解析を繰り返しても相関係数が高くないことがある。その時には、次のことを考える。

1. 入力ミスは無いかな?
2. 識別子として足りないものは無いかな?
3. 現象として、非線形現象では無いかな?

多くの場合、1.が原因で精度がでない。Typoglycemiaの効力で、何度見直しても入力ミスは見つからないので苦労する。2.の識別子として足りないものを考える時には、プラスに一番大きく外れるものと、マイナスに一番大きく外れるものをマークして、その違いを説明する因子を考える。YSBが作った式で、プラスに外れるものは、Ti、マイナスに外れる物はCrだった。TiはElectron Affinityがこの中で一番小さい。また、Crについては遷移金属の中ではAtomic Radiusが一番小さい。そうした識別子を必ず入れるとしたら他の識別子はどれを選べば良いかな? このテーブルには無い、他の識別子は無いかななどを考える。3.の非線形性を考えるのは最後の段階だ。この程度のデータ数ではニューラルネットワーク法を用いるのは意味がない。各項目のlogを取ったり、ルートを取ったり、クオースタムを考えたり徐々に非線形性を上げていく。

最終的にはどんな式になったろうか？
 自分考えた比活性の予測式を書き出してみよう。
 比活性=



筆者の作った比活性の予測値を実験値に対してプロットすると上図のようになる。YSB が自動的に選択した識別子と比べ、特に高活性の領域での精度が高くなっていることが分かると思う。ここは非常に大事な点で、例えば、推算精度（相関係数）が劣っていても、自分が検討したい領域（今回は高活性な領域）で精度が出る式が、目的に対して、よりベターな予測式であることを忘れてはいけません。

なんでも、かんでもニューラルネットワークに放り込めば、全領域でぴったり合う予測式が作成され、新しい合金触媒が設計される事を期待するのであれば、マテリアル・ジオノムに手を出すのはお勧めしない。意外と地道な作業の積み重ねになる。

YSB の選んだ識別子は非常に参考になる。変数が3つ以上になると線形で考えても訳がわからなくなる。また項目間の単純な相互作用を見つけるクロスタム重回帰法も搭載されているので、多くの場合シンプルでかつ強力な解析結果を手に入れることができる。こうした YSB によるアシストを受けながら、材料設計を行なった場合、出来上がった予測式は、作成した人間に強く依存してしまう。偉そうな事を言っても、触媒設計など一度もやったことのない身としては、結果がどのくらい妥当なのかは判断できない。

予測結果の検証

まず、自分で作成した予測式を使って、周期律表の原子全てに対して、比活性を予測してみよう。ここから先は、学生によって答えは、まちまちになる。筆者が作成した予測式で説明するが、答えが異なっていて

も一向に構わない。

Alloy	ctivit	Radius	Ionization Potential	ElectronAffinity	予測値
H		0.37	11.473	0.754	40.1
Li		1.34	5.392	0.62	19.1
Be		1.25	9.322	-2	162.7
B		0.9	8.298	0.28	155.0
C		0.77	11.26	1.268	49.7
N		0.75	14.534	-0.07	-22.3
O		0.73	13.618	1.462	-44.6
F		0.71	17.422	3.399	-258.6
Na		1.54	5.139	0.546	-73.0
Mg		1.45	7.646	-1.8	96.9
Al		1.3	5.986	0.46	47.0
Si		1.18	8.151	1.385	61.5
P		1.1	10.486	0.743	82.9
S		1.02	10.36	2.077	43.3
Cl		0.99	12.967	3.615	-47.7

そして、比活性の高い順にソートをかけて、高活性になる原子を特定しよう。

Alloy	比活性予測値
Be	162.7
B	155.0
Mn	126.6
Hg	112.3
Fe	101.3
Mg	96.9
Zn	95.2
Co	89.0
Cr	86.5
P	82.9
V	78.6
Re	78.3
As	76.6
Ni	74.8

さて、比活性が高くなる原子種は特定されたが、如何せん非常勤講師の筆者では実験による検証は行えない。そこで書籍、ネット上にある情報から検証を行ってみる。このグラフを取り出した書籍には次のように記載されている。

「白金を卑金属と合金化して高活性化することは、リン酸形燃料電池(PAFC)の実用化過程で活発に研究された。チタン、クロム、鉄、コバルト、ニッケルなどと合金化することにより、最大で3倍程度の活性増大が得られている。」比活性の高いと予測した、Cr, V はもともとグラフに記載されていた。Fe, Co, Ni は書籍に記載されていた。Hg や As は毒性の観点から使いたく無いので除外する。

他の原子について、“燃料電池 Pt-B 触媒”のように Google で検索をかけてしまう。例えばホウ素の場合、トヨタ特許(WO2014034357A1)

酸化耐性と触媒活性とを両立する燃料電池用電極触媒 炭化ホウ素に担持された貴金属

山口大学論文、Pt を MgO に担持

埼玉大学論文、Pt を ZnO に担持

などのように、B, Mg, Zn に関してはPtの担持体として優れていることがすでに知られていた。

また、Mn に関しては、田中貴金属(Spring-8 ワークショッ

ブ) Pt-Co-X 3元系合金触媒でMnが最高の性能という資料が見つかったので、Mnに関しても良い性能になることがすでに知られているようだ。

そうすると、今回設計した、比活性(ただし筆者の予測式)の高い原子のほとんどは、既に知られているものであることが検証される。

Alloy	比活性予測値	
Be	162.7	
B	155.0	トヨタ特許(担持)
Mn	126.6	田中貴金属(合金)
Hg	112.3	毒性?
Fe	101.3	書籍に記載
Mg	96.9	山口大学論文(担持)
Zn	95.2	埼玉大学論文(担持)
Co	89.0	書籍に記載
Cr	86.5	グラフに記載
P	82.9	リン酸型燃料電池
V	78.6	グラフに記載
Re	78.3	
As	76.6	毒性?
Ni	74.8	書籍に記載

残るものは、BeとReでこれらはGoogleの検索には引っかかってこない。原子の周期律表の位置から考えて、Beは酸化ベリリウムとしてPt担持用、Reは合金触媒用として使うのだろう。この二つに関して、コンピュータの計算だけで特許を書いてしまうのがマテリアル・ゲノムによる材料設計の面白さであり、怖さでもある。論文と違って特許は、正しくなくても良い。Pt-BeO担持、Pt-Re合金は性能が高い(可能性がある)。そんな特許が認められる時代になってしまった事が、マテリアル・ゲノム時代の一番の問題で、(その善し悪しはおいておいて)どう対応すれば良いのか? を問われている。

ここで、使った触媒設計の遺伝子は、
原子半径
イオン化エネルギー
電子親和力
の3つという事になる。

また、田中貴金属のPt-Co-Xという3元系の触媒のデータから、3元系の触媒に関して最適なものをマテリアル・ゲノムで探してしまう事も可能だろう。

学生さんが作成した予測式で、これ以外の原子は出てきたでしょうか? そうした触媒系はネットで見つかったでしょうか?

大学での研究はともかく、企業の研究であれば、こうしたマテリアル・ゲノムの開発手法を取り入れない理由は全く無いように思える。スパコンが必要なわけでもなく、学

生が個人で持っているような性能のPCで十分計算できる。高価なソフトを導入しなければならないわけでもない。PDCAを非常に高速に回すことができるようになる。新たな実験結果、参照論文、競合他社から出た特許などから新たな情報が得られたら、即座に予測式を立て直すフットワークの軽さが成功を左右する。

例えば、BeやReを試して、結果が悪かった場合、すぐに予測式を作り直す。もしくは、諦めずに実験のやり方を変えるなど様々な対策が取れるようになるはずだ。

日本でのこうした研究が進まなくても、海外では当たり前のように使われ出している。待つてはくれない。日本の論文、特許がマテリアル・ゲノムの検証に使われる、3元系への拡張に使われる。「そんな時代への対応が十分か」が問われ出している。

大学での教育

マテリアル・ゲノムが有効だと言っても、今回示したように、化学者としての地力の高さも必要な事も指摘した。それでは企業はともかく、大学ではどういう教育によって学生の地力を高めたら良いのだろうか? 実は、ここで取り上げたグラフと同じものが、2001年に発行の、化学総説49、新型電池の材料化学という書籍中でも使われている。そこでは、分子軌道法を使った反応機構が説明されている。

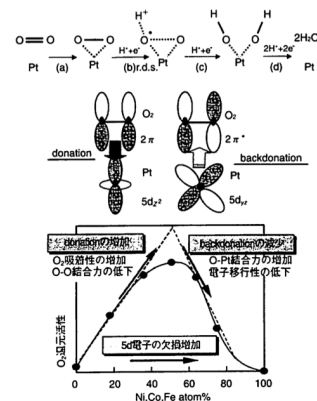


図 11 Pt-Fe合金などにおける表面白金層の電子構造修飾による酸素還元活性の増大機構¹⁰⁾。

これを見たときに、酸素のπ軌道と白金のd軌道の反応である事を理解して、塗りつぶされた原子軌道の違いを学生が理解できるかどうかは重要な教育上の問題として残る。それでは、どんな原子ペアが良いのかを全て分子軌道法で遷移状態を求めて評価する事を学生に求めるのかは、意見は分かれるだろう。現在のところ分子軌道計算はそこまで精度は出ないように思える。最低限、電子配置とイオン化エネルギー、電気陰性度の値から、
Pt: [Xe]4f14 5d9 6s1

Ionization Potential 9eV

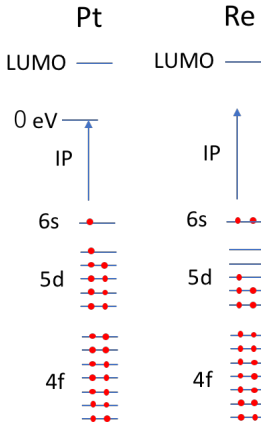
Electronegativity 2.28

Re: [Xe]4f14 5d5 6s2

Ionization Potential 7.88eV

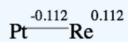
Electronegativity 1.9

である事を理解し、



電子配置の図を頭に思い浮かべられる必要はあると思う。また、電荷平衡法による、原子上の電荷が次のように求められる事を知っていることは重要だろう。

$$\left(\frac{\partial E}{\partial Q}\right)_{A0} = \frac{1}{2}(IP + EA) = \chi_A^2$$



今回の触媒設計の遺伝子は、原子半径、イオン化エネルギー、電子親和力であることが示唆されたが、この3つがわかれば、電荷平衡法で原子上の電荷が計算で出すことができる。電荷支配の反応なら、この3つの情報が必要十分条件である可能性が高い。

そうしたことを理解できるように学生を育成することは、教育機関としての大学の役目であるように思える。

その上で、実験結果の無い系について予測して可能性を示唆できるなら、高度な地力を持った学生として、AIアシストを受けながら将来にわたって生き延びられる人材になることができると言えるので無いかと思う。

雑感

自分が住んでいる次元の下の次元は理解できるけど、同次元やそれ以上の次元は脳には理解できないのかもしれない。人間の住んでいる次元が、3次元なのか、時間軸が4次元目の軸なのかで、2次元グラフしか理解できない、3次元グラフまで理解できるかの差はあるかもしれない。程度の差はあれ4次元以上のグラフは人間にはお手上げだろう。かと言って一旦縮退させてしまった情報は復元で

きない。

どのようにしてAIにデータマイニングさせたらいいかは、難しい問題として残る。しかも解析までAIに任せるなら、失敗データをどう収集させたらいいのかわからない。そのあたりの解決は数学者や哲学者に任せるしか無いように思える。

大学でこうした授業を取り入れてしまうと、実は評価の際に困ったことが起きる。どんな予測式であれ、それを評価の高低には結び付けられないからだ。精度が低くても、たまたま性能の良い合金触媒になっているかもしれないし、その逆もあるかもしれない。運も実力のうちなので、授業にちゃんと参加していれば皆高い点数を差し上げている。大学の入試までは、結局は記憶力の高いものが勝つ。社会に出てからは答えのないところで競わなくてはならない。答えの無いところで競うなら、地力をつけて、AIアシストをうまく利用して、目的に対して高い確率の道を選ぶことは学生のうちに身につけておいて欲しいと思う。

今回紹介した事例は、2-30年前のやり方で、マテリアル・ゲノムと呼べるような代物では無い。分析結果も計算結果も非常に少なくとも何とか方向性を見出して材料設計を行ってきた古き良き時代の研究手法だろう。でも技術の進歩によって、そうした研究も、とても楽に結果が出せるようになってきたと思う。

データ収集に関しては、インターネットの進歩によって、非常に楽になった面と、マテリアル・インフォマティクスの進歩によって非常に不便になった面がある。楽になった面はすぐにわかるだろう。情報はPDFなどの形で容易に入手できるようになった。マテリアル・インフォマティクスによって不便になったというのはどういう意味であろう？ 最近、特許などで顕著なのだが、MIやマテリアル・ゲノムをやっているような先端企業は、他社がデータを逆解析できないようにデータ記載を工夫するようになってしまった。案外1970-90年代の論文、特許の方がMI用のデータとして優れたものが記載されているように思える。

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