

帰納論理プログラミングを用いた高分子の組成と物性との関係に関する考察 Discussion of Relationship between Compositions and Properties of Polymers using Inductive Logic Programming

力規晃† 越村三幸‡ 橋本司‡ 西田光生§ 阿部幸浩§ 藤田博‡ 長谷川隆三‡
Noriaki Chikara Miyuki Koshimura Osamu Hashimoto Mitsuo Nishida Yukihiko Abe Hiroshi Fujita
Ryuzo Hasegawa

1. はじめに

化学反応を利用して、原料から所望の性能を満たす化合物を効率よく生成するには、化合物を生成する原料の濃度、流量、温度、即ち反応条件に関連したパラメータを適切に設定する必要がある。

最適な反応条件は、実験を繰り返し行うことによって求めるのが普通であるが、実験には時間とコストがかかるため、最小の実験回数で最適な条件を求めることが肝要である。

そのため本研究では、実験初期の時点でそれまでに得られたデータを基に所望の性能を満足する実験条件を推定することを目的とし、帰納論理プログラミング手法の応用を試みた。すなわち、まだ目標を満たしていない実験データから差分を用いた簡単な方法で予測データを生成し、これを帰納論理プログラミングで学習することで、目標条件を満たす条件パラメータのルールの抽出を行った。

2. 取り扱うデータの概要

本研究で取り扱うデータは化学反応の条件パラメータ(説明変数)とその性能を示す値(目的変数)とがセットになっている。条件パラメータには組成項があり、組成の合計を一定値にするという制約がある。

与えられるデータ件数についてはデータマイニングの対象のデータとしては少ない。具体的には 10 数件~50 件程度である。これは、データを得るための化学実験に長期間要することと、コストが高いためである。そのため、少ないデータからルール抽出を行うというところに挑戦課題がある。

また、データ中の性能を表わす値は 1 つ以上あり、それぞれにクリアしたい目標値があるが、全ての目標値を満たすデータは既存データに含まれていない。これは、そのような実験データが得られれば、それで実験の目的は達せられ、実験をそれ以上続ける必要が無くなるからである。本研究の目的はすべての目標条件を満たす条件を推定することである。そのために、実験データの近傍において、条件パラメータと目標条件との間に線形性を仮定し、仮想的に実験データを合成する。こうして得られた仮想実験データには、目標条件を満たすものもあれば、満たさないものもある。

そして、仮想実験データと実際の実験データを帰納論理プログラミングによって解析し、目標条件を満たすような条件パラメータを導く。導かれた条件が、実験化学

者によって、妥当と判断されれば、化学実験によって本当に目標条件を満たすかどうか実証される。

3. 帰納論理プログラミングの適用

本研究では帰納論理プログラミングを用いて説明変数と目的変数間のルール抽出を行う。

3.1 帰納論理プログラミング

帰納論理プログラミング(Inductive Logic Programming: ILP)[1]とは一階述語論理に基づいた機械学習の手法である。ILP の一般的な枠組みは、正例 E^+ と負例 E^- 、背景知識 B が節集合で与えられ、

$$\begin{cases} B \models E^+ \\ B \cup E^- \not\models \square \end{cases} \quad (1)$$

であるとき、つまり、背景知識では正例が説明できず、しかも、背景知識と負例が矛盾しないとき、

$$\begin{cases} H \cup B \models E^+ \\ H \cup B \cup E^- \not\models \square \end{cases} \quad (2)$$

を満たす、つまり背景知識に仮説を加えると正例を説明でき、しかも、背景知識と仮説と負例は矛盾しない仮説 H を見つけることである。

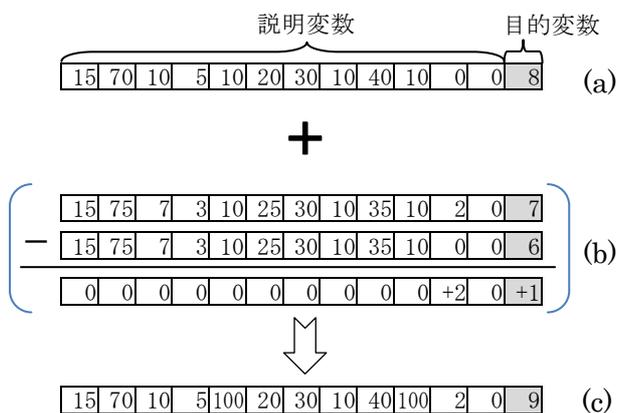


図 1 実験条件の推定方法

3.2 目標値を満足する実験条件の推定方法

本研究において、予め得られている実験データには目標性能を満足するデータは含まれていない。つまり正例はない。このため、各パラメータ特性はそのデータ点近傍では線形性を示すと仮定して、目標とする条件を推定した。

図 1 に条件推定の方法を示す。データの右端の値(目的変数)が性能を表わすパラメータ属性であり、この値によって目標値を満たすか否かが判定される。条件推定は既

† 徳山工業高等専門学校, Tokuyama College of Technology

‡ 九州大学, Kyushu University

§ 東洋紡績株式会社, TOYOBO CO. LTD.

存データに2つのデータの差分を加えて新たな条件を生成することにより行われている。

まず、(b)の部分のように任意の2組の実験値の各々の要素の差を取り、差分値を生成する。そしてある既存データ(a)に(b)で生成した差分値を加えて、新たなデータ(c)を生成する。これにより既存データでは存在しない目標値を満たす条件を生成できる。

3.3 正負例と背景知識

任意の既存データで差分を加えるデータ合成を行い、目標の条件と条件パラメータの制約を満たすものを正例とし、目標条件や制約を満たさないものを負例とする。また、既存データは目標条件を満たさないため負例とする。

また、本研究では差分を行った実数値を直接扱う。このとき ILP では値がわずかに違うが近い値が全く別の値として扱われる。このため ILP で既存データの条件と極めて近い条件をルールとして抽出することが予想される。これの対策として、既存データにわずかなノイズを加えたデータも負例として加えることにより、既存データに近い条件がそのままルールとして抽出されないようになる。

本研究の背景知識は、表1のような述語表現で、各正負例の条件パラメータを記述したものである。

表 1 本研究で用いる知識表現

知識表現	意味
p_aN(E, In)	事例 E の条件パラメータ aN は In である。
p_gN(E, In)	事例 E の条件パラメータ gN は In である。
p_aaN(E, In)	事例 E の条件パラメータ aaN は In である。
p_M(E, In)	事例 E の条件パラメータ M は In である。

但し、ここで N は 1 以上の整数であり、M は a~x の文字を示す。

4. 実験

本研究では帰納論理プログラミングシステムとして Aleph[2] Version5 を利用する。この Aleph は Prolog で記述されているため、Yap-Prolog 6.2[3] という Prolog を用いる。

また、実験に用いた PC は CPU が Corei7 2.7GHz でメモリが 8GHz、OS は Windows7(64bit) である。

4.1 実験(1)

データ数は 47 事例であり、条件パラメータは 24 属性ある。条件パラメータの合計に制約あり、5つの性能を表わすパラメータの目標条件を満たす条件パラメータのルールを抽出する。データ合成により得られた正例 78 事例、負例 9674 事例でルール抽出を行った。図 2 に得られたルールの例を示す。

図 2 において抽出ルールは 2 行目以降であり、「属性 G=0.4, 属性 K=0.0, 属性 N=1.0 の時に目標値になる」ということを意味している。また、図 2 の一行目の "Pos cover=3" からこのルールは正例を 3 つ満たしていることが分かる。

```
[Pos cover = 3 Neg cover = 0]
p_ex(A) :- p_g(A,g_0_4),
           p_k(A,k_0_0),p_n(A,n_1_0).
```

図 2 実験(1)の抽出ルールの例

4.2 実験(2)

データ数は 20 事例であり、条件パラメータは 2 つのグループとグループに属さないものがあり、一つのグループには 5 つの属性が属する。2 つのグループには 9 つの属性が属する。それぞれのグループの属性の合計に制約がある。また、どちらのグループにも属さないパラメータも 2 つある。5 つの性能を表わすパラメータの目標条件を満たす条件パラメータのルールを抽出する。データ合成により得られた正例 17 事例、負例 1103 事例でルール抽出を行った。図 3 に得られたルールの例を示す。

```
[Pos cover = 2 Neg cover = 0]
p_ex(A) :-
           p_a1(A,a1_55_0),p_g1(A,g1_6_0).
```

図 3 実験(2)の抽出ルールの例

図 3 において抽出ルールは 2 行目以降であり、「属性 A1=55.0, 属性 G1=6.0 の時に目標値になる」ということを意味している。また、図 3 の一行目の "Pos cover=2" からこのルールは正例を 2 つ満たしていることが分かる。

5. 考察

各パラメータがそのデータ点近傍では線形性の特性を持つという仮定のもとでデータを合成して、ルール抽出を行ったため、仮定や近似の妥当性を考えなければならない。本研究で得られたルールは、実験化学者が経験的に得ている感覚と非常に良く一致しており、今回対象とした化学反応のデータに関しては本研究の推論方法は妥当であると考えている。

6. おわりに

化学反応の入出力の関係で、望ましい出力を得る条件のルールを帰納論理プログラミングで抽出する実験をおこない、適切なルールを得ることができた。今後、今回扱ったデータ以外の多くのデータで本研究の手法を試すとともに、他の手法と定量的な比較が必要だろう。

謝辞

本研究は科研費(21300054)の助成を受けたものである。

参考文献

- [1] 古川康一, 尾崎 知伸, 植野 研, "帰納論理プログラミング", 共立出版 (2001)
- [2] A.Srinivasan, "The Aleph Manual", <http://www.comlab.ox.ac.uk/oucl/research/areas/machlearn/Aleph/> (1999)
- [3] "Yap Prolog", <http://www.dcc.fc.up.pt/~vsc/Yap/>