

決定木を用いた基礎的有機化合物の発火点予測

林 亮子^{1,a)}

概要：発火点とは、可燃物を空気中で加熱すると自然発火する温度である。安全性に直結する発火点は古くから実験的に調べられているが、近年では新規の化合物が大量に生成されており、危険を伴う発火点調査を迅速に行えるよう支援が必要である。そこで、著者らは近年データマイニングを用いて発火点の予測を試みている。本稿では、古くから知られていて発火点などの諸量がよく調べられている炭化水素および類例分子を選び、データマイニングの手法の一つである決定木を用いて発火点の分類ルールを検討した結果を報告する。入力データに分子量、融点、沸点、炭素原子個数、酸素原子個数および特徴的な部分構造の件数を使用し、統計プログラミング言語 R の決定木パッケージ rpart を用いて決定木を作成した。その結果、今回用いた物質の発火点に対して最も影響力が大きい分岐ルールは、ベンゼン環の有無であることがわかった。さらに、化学的な部分構造の観点から入力データを再検討した結果、決定木の過学習状態が改善したので、本稿で報告する。

キーワード：データマイニング、決定木、発火点、分子、有機化合物

1. はじめに

近年のデータマイニング技術 [1][2] は非常な発展を遂げ、誰でも簡単にデータマイニングを利用できるようになった。また、データマイニングに代表されるデータ科学は物質科学関連分野においても、実験、理論、計算に次ぐ「第4の科学手法」と言われはじめのほど注目されている。本稿ではデータマイニングの物質科学関連分野への応用事例を報告する。

発火点 [3][4] は、可燃物を空気中で加熱すると自然発火する温度であり、工業的に重要な量である。そのため、古くから知られた多くの物質において発火点は調べられており、データが蓄積されている。一方、化合物 [5][6] はすでに数千万種類が知られており、さらに日々新しい物質が生成されている。そのため、データマイニングによって新規の物質の発火点予測ができると、安全対策や化合物の有効利用の観点から実用的にも役立つものと考えられる。

化学分野においては、計算機の黎明期から化学物質の性質を調べるために計算機の積極的な利用が試みられており、ケモインフォマティクス [7] と呼ばれる分野が形成されている。ケモインフォマティクスでは、分子の性質を表す量を記述子と呼ぶ。記述子設定の方法論についてはまだ定説

はなく、データ分析の目的によって必要な記述子を検討する必要がある。ケモインフォマティクス分野においても発火点の予測は試みられている [8][9]。しかし、分子の性質や構造と発火点の関係については、まだ不明な点が多い。

本稿では、主要なデータマイニング手法の一つである決定木を用いて発火点予測を試みる。決定木を使用すると発火点の分岐ルールが得られ、記述子の発火点における寄与度を議論することができるため、先行研究とは異なる観点から発火点を議論できる可能性がある。本稿の著者グループは、これまでに文献 [10][11] において発火点を調べたが、十分な予測精度が得られなかった。精度を改善するためにデータ内容を確認したところ、相当数の誤りがあることがわかった。そこでデータを全面的に再確認の上訂正し、記述子も再設定を行った結果を本稿で報告する。

本稿では、多様な化学物質の中でも炭素原子、水素原子、酸素原子のみで構成された分子を扱う。これらの分子は化学物質の中でも構造が簡単な基本的なものであり、古くから存在が知られていて、性質もよく調べられている。既知の知見と分岐ルールや予測結果を比較することで、データマイニングが適切に行われたかどうかを確認できる。

化学では「基」という概念があり、複数の分子で特徴的に見られる原子の組み合わせと結合形態をまとめたグループを指す。基は分子の性質と関係が深いことが知られている。原子の種類に従って多様な基が存在するので、扱う原子の種類を限定することによって、考慮すべき基の種類を

¹ 金沢工業大学
KIT, 3-1 Yatsukahoh, Hakusan, Ishikawa 924-0838, Japan
^{a)} ryoko@neptune.kanazawa-it.ac.jp

限定し、問題の複雑さを調整することができる。今回は扱う原子を3種類に限定したが、それでもなお分子中の部分構造として10種類以上の記述子が考えられる。

第2節以降の本論文の構成を述べる。第2節は類例研究を紹介する。第3節はデータの作成方法を説明し、記述子の設定内容を述べる。第4節では決定木を作成し、得られた発火点の分類ルールを議論する。第5節は本稿で得られた結果をまとめ、今後の課題を述べる。

2. 類例研究

まず、定量構造活性相関分野における先行研究を紹介する。Tsai, Chen, Liaw は820件のデータに4個の記述子とその線形結合式を用いて最大誤差 $89K$ 、平均誤差 $36K$ で有機化合物の発火点を予測した [8]。Tsai らは、計算化学により得られる大量の数値データから、記述子として有効なものを選択している。Shi, Chen, Chen は分子を部分的に分割した8~12種類のフラグメントを記述子とし、265件の有機化合物データを使用して、実際の発火点との誤差が $50K$ から $90K$ 程度で発火点を予測した [9]。

岡田と、本稿著者である林は炭化水素と類例分子21種類の分子量、融点および沸点をデータに用いて自己組織化マップを作成し、分子間の類似度を利用した発火点の予測を試みた [10]。その結果、炭素原子が単結合だけで直鎖型に結合したアルカンでは、ある程度の予測が可能であったが、芳香族の発火点予測はできなかった。そこで、林と中田は炭素、酸素、水素のみから構成されている炭素原子10個以内の有機化合物に注目して245種類の学習データと10種類の試験データを用意し、発火点を決定木によって調べた [11]。しかし、得られた決定木は多くの部分が過学習であり、改善の余地があった。

3. 使用データの概要

3.1 分子データの作成法

本章では、使用したデータの概要を述べる。本研究では発火点予測を目的とするため、実用上十分な精度で発火点が既に決定した物質に関する情報を収集する必要がある。発火点の情報をまとめて公開しているものに「国際化学物質安全性カード」 [3] がある。ICSC は、基本的に単独の分子だけを含む純物質についてデータを作成することを基本としているので、本研究の目的にも適合する。本稿でデータに含める物質は、国際化学物質安全性カードにおいて、分子の元素組成を示す組成式、沸点、融点、発火点が記載されているものを中心とする。なお国際化学物質安全性カードでは、発火点は「発火温度」と記載されている。

本稿では記述子として、特徴的な結合や基がその分子に含まれる個数を使用する。それらを元素の種類とその比率のみを示す組成式から決定するのは困難であるため、分子

構造を表示する構造式を目視し、さらに SMILES 記法を参照して決定した。SMILES 記法は次節で紹介する。構造式と SMILES 記法は、科学技術振興機構 (JST) が作成する「日本化学物質辞書 Web」 [5] を参照した。沸点、融点、構造式、SMILES 記法のいずれかがこれまで紹介したウェブサイトに掲載されていない物質が10件程度あったため、wikipedia、「職場のあんぜんサイト」 [4] および「Chemical Book」 [6] を補助的に利用した。

本稿のデータ作成で行った例外的な扱いを述べる。

昇華する物質 一部の物質は、固体を加熱すると固体から直接気体になる「昇華」を起こす。そのため、大気圧程度の状態では昇華点のみが知られている物質がある。その場合は、昇華を「液化と気化が同時に同じ温度で起こる」と解釈し、沸点と融点の両方に昇華点を用いた。この扱いにより、例えば身近な物質で炭素間の三重結合を持つアセチレンをデータに加えることができる。

混合物質 一部の物質は、異性体が混合した状態で発火点を決定していて、その場合は構造が一意に決まらない。構造が大きく異なる異性体は一般に物理的性質も大きく異なって分離精製が可能であり、多くの場合は異性体を分離して発火点を決定している。一方混合状態で発火点を決定する場合は、例えば鏡像異性体などのように、沸点や融点の差異が小さく、特徴的な部分構造が共通する物質が混合していて、実用的な分離が困難である。そのような場合は、異性体でも本研究で使用する基や結合の種類は共通するのでそのまま扱って問題はない。そのため、可能な範囲で構造式を調査し、基や結合の記述子を決定した。

沸点の欠損 発火点は決定していても、加熱すると分解してしまうために沸点が決定できない物質もある。そのような物質は今回使用しないものとした。

3.2 データの修正と修正後のデータ分布

本稿で行ったデータの修正内容を以下に述べる。国際化学物質安全性カードはウェブページとしての閲覧のみ可能であり、csv形式等での電子ファイル出力ができないために手入力で電子化している。本稿作成にあたって全データを再確認したところ、単純な入力ミスが10件程度存在した。次に、基などの特徴的な部分構造の記述子を構造式で確認すると、間違いが20件ほど存在した。化学物質は慣用的な名称と化学構造に由来する命名法によるものなど、一般に複数の名称を持つことが多い。データ確認に際して一意性を保証するため、米国化学会が構築管理を行うCASデータベースが付している識別番号CAS番号を使用したところ、同じ分子の重複使用が3件存在したため、それらを削除した。また、加熱すると分子が分解するために沸点

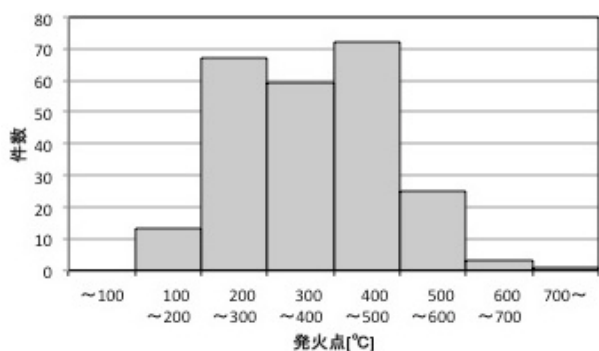


図 1 使用データ 240 件における発火点の分布

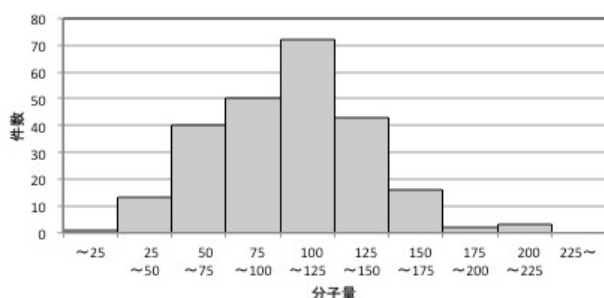


図 2 使用データ 240 件における分子量の分布

データが欠損する分子が 2 件あり、これらも削除した。そのため文献 [11] で用いた学習用データは、最終的に 240 件となった。

今回用いたデータの発火点と分子量の分布を示す。まず発火点の分布を図 1 に示す。図 1 によると、最も温度の低い発火点でも 100 度以上であり、最高の発火点は 700 度であった。ほとんどの物質は 200 度から 500 度の範囲にある。また、類例研究の発火点分布とおおむね同様である。次に、分子量の分布を図 2 に示す。本稿では炭素原子個数を 10 個以内に限定していることから、非常に大きい分子は除外される。図 2 によると、分子量 100 から 125 までの分子が最も多く、その階級から離れるにつれて、件数は減少する。

3.3 分子の構造と記述子

分子の構造から記述子を決めるために SMILES 記法を併用するので、その概要を紹介する。化合物の一つであるジケテンは組成式が $C_4H_4O_2$ であり、SMILES 記法で表すと C1(=O)OC(=C)C1 となる。SMILES 記法の詳細は文献 [7] などに譲るが、分子の化学構造を文字列で表現するもので、原子を元素記号で表し、二重結合を記号「=」、三重結合を記号「#」で表し、構造の主要部分からの分岐を () 内に入れて表す。環構造は結合部の原子に識別番号を付す。図 3 の例では、最初の炭素原子が環構造の起点で、その炭素原子に酸素が二重結合で枝分かれして結合しており、その炭素原子に酸素原子 1 個、次に炭素原子が 2 個単結合して最初の炭素原子に戻ることで環構造ができています。水

素原子は 4 個含まれるが、結合可能な箇所には結合しているものとして省略して扱い、SMILES 記法では明記しない。

SMILES 記法では識別番号を用いて環構造を明示するため、SMILES 記法を併用することで環構造の個数を確定することができる。また、芳香族構造をなす原子は小文字で書くこととなっているため、環構造と区別することができる。一方で、SMILES 記法は一つの化学構造に対して一意に決まらないことが欠点であり、近年では様々な改良版が提案されているが、歴史的な経緯もあり、多くのデータベースでは現在も SMILES 記法が記載されている。

次に、構造式と記述子の関係を説明する。構造式は、分子構造を簡略化して平面上に図示したものである。基本的には構造式は、情報科学で「グラフ」と呼ばれるデータ構造と考えることができ、節点が原子またはよく知られた部分構造に相当して、原子記号を用いて記載する。結合関係にある原子間を辺で結んだものである。辺が直線 1 本であるときは単結合であり、二重結合、三重結合の場合はそれぞれ二重線、三重線で節点間を結合する。慣例として構造式では、自明と考えられる水素原子の記載を省略し、また炭素原子も特別な場合を除き、辺の折れ曲がりて存在を表現する。構造式は原子位置の三次元情報を含まないため、立体構造を扱うことはできないが、本稿の範囲では大きな問題はない。

ジケテン $C_4H_4O_2$ を例として、構造式と特徴的な部分構造を図 3 に示す。図 3(a) のように、ジケテンは 3 個の炭素原子と 1 個の酸素原子が環構造をつくり、酸素が環構造の炭素原子 1 個と二重結合している。また、環構造のもう一つの炭素原子に CH_2 が二重結合している。

このジケテンが含む特徴的な部分構造を図 3(b) および図 3(c) に示す。図 3(b) に示すように、炭素原子に酸素原子が二重結合している部分をカルボニル基 (ケトン基と呼ばれることもある) という。また、炭素原子に CH_2 が二重結合している部分は炭素間二重結合として扱う。さらに、カルボニル基の炭素原子に酸素原子が結合し、その先に結合している炭素原子までも一つの特徴的な部分構造であり、エーテル結合と呼ばれている。エーテル結合からさらに原子が結合した先は、もとのエーテル結合のもう一方の原子に結合しているため、ジケテンは環状エーテルである。図 3(b) に示した部分構造は互いに独立している。

またジケテンでは、図 3(c) に示すように、エーテル結合の片方の炭素原子とカルボニル基が直接結合している構造を持つが、これらをまとめてエステル結合と呼ぶ。このように、分子の特徴的な部分構造は、入れ子の関係を持つ場合があり、記述子として部分構造を使用する際には、入れ子の関係を持つものの取扱いを決める必要がある。ジケテンでは、エステル結合から先がもとのエステル結合に戻るため、環状エステルでもある。

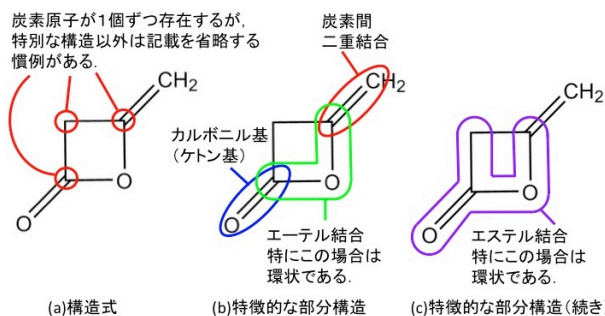


図 3 構造式と特徴的な部分構造の例(ジケテン $C_4H_4O_2$, SMILES 記法では C1(=O)OC(=C)C1)

本稿における分子の特徴的な部分構造の取扱いを説明する。本稿ではベンゼン環, 炭素間二重結合, 炭素間三重結合, 水酸基, アルデヒド基, カルボニル基, エーテル結合, 環状エーテル結合, カルボキシル基, エステル結合, 環状エステル結合, 環構造, 直鎖型構造の数を記述子とする。これらの記述子で, 炭素, 酸素, 水素からなる化合物がもつ特殊な構造の主要なものを網羅している。これらは構造式を目視して数を判定するが, SMILES 記法をあわせて確認することで誤りを減らすことができる。また, すでに述べたようにいくつかの構造は他の構造を包含しているが, 発火点に対して上位階層の構造が影響する場合と, 下位階層の構造が単独で影響する場合のいずれも可能性があるため, 重複して数えるものとする。

本稿および文献 [11] で採用した手法に基づいて作成した図 3 のジケテンのデータを表 1 に示す。表 1 では, 文献 [11] 中での記述子の扱いを「旧版」, 本稿での扱いを「新版」として両方を示した。表 1 に示すように, 本稿では分子の性質を表す連続値として分子量, 沸点, 融点, 発火点を用いる。そして, 分子の特徴的な原子個数として炭素原子個数, 酸素原子個数を用いる。さらに, 記述子として特徴的な部分構造の個数を用いる。構造が存在しない場合は 0 とする。環状構造, 枝分かれがない構造である直鎖構造およびアルデヒド基も発火点に影響する可能性があるため, 本稿では新規に加えた。本稿は「決定木を用いて発火点の決定ルールを調べる」という考え方であるため, 入力データには予測対象である発火点を含む。本稿で扱うデータでは, 以上の内容を欠損値なしで持つ。

ジケテンは図 3 に示したように環状エステル構造を 1 個持つ。旧版の記述子では包含関係にある記述子では上位階層の記述子だけを数える。そのため図 3 の旧版ではエステル結合だけを 1 としており, カルボニル基とエーテル結合は数えず 0 としている。一方, 新版の記述子では包含関係にある記述子を重複して数えるため, エステル結合, カルボニル基, 環状エーテル, エーテル結合, 環構造の全てが 1 である。

4. 発火点を分類する決定木

4.1 旧版記述子を用いた決定木

まず, 240 件の分子で旧版の記述子を用いたデータによる決定木を作成する。決定木を作成すると過学習が起きていることが多いので, 文献 [1] の手順に従い, 枝刈りを検討する。今回使用する `rpart` パッケージ中の関数 `rpart` は, 決定木作成時にデータセットをランダムに分割して交差確認も行い, デフォルトでは 10 分割交差確認を行う。そして, 関数 `plotcp` はブルーニングに必要な情報を図示するのに用いる。

240 件の分子で旧版の記述子によるデータを入力し, デフォルト設定で作成した決定木における `plotcp` の出力を図 4 に示す。図 4 は下側の第一横軸が複雑度 cp , 上側の第二横軸が決定木の葉ノード個数であり, 縦軸は相対交差判定誤差である。デフォルトでは $cp = 0.01$ まで分岐を行うため, 図 4 の第一横軸右端は $cp=0.01$ である。複雑度 cp は木が成長するとおおむね減少するが, 木の大きさに対して一様ではなく, 増加することもある。

決定木では, 一標準偏差ルールと呼ばれる経験則があり, (相対交差判定誤差の最小値) + (標準偏差) を枝刈りの基準値としてよく用いる。図 4 ではその基準値が約 0.8 であり, 図 4 中の水平線として表示されている。多くの場合, その水平線より下になった最小の葉ノード数までを採用し, それよりも先の分岐は過学習によるものとして採用しないことで枝刈りを行う。

図 4 によると, 葉ノード数 2 で相対交差判定誤差は一旦

表 1 使用データ例(環状エステル, ジケテン $C_4H_4O_2$, 旧版で定義していない記述子は「-」と記載する。)

| 記述子 | 旧版 | 新版 |
|------------------|------|------|
| 分子量 | 84.1 | 84.1 |
| 沸点 [澁] | 127 | 127 |
| 融点 [澁] | -7 | -7 |
| 炭素原子個数 | 4 | 4 |
| 酸素原子個数 | 2 | 2 |
| ベンゼン環 | 0 | 0 |
| 炭素間二重結合 $C=C$ | 1 | 1 |
| 炭素間三重結合 $C\#C$ | 0 | 0 |
| 水酸基 $-OH$ | 0 | 0 |
| アルデヒド基 $-CHO$ | - | 0 |
| カルボニル基 $-C(=O)-$ | 0 | 1 |
| エーテル結合 $-O-$ | 0 | 1 |
| 環状エーテル | - | 1 |
| カルボキシル基 $-COOH$ | 0 | 0 |
| エステル結合 $-COO-$ | 1 | 1 |
| 環状エステル | - | 1 |
| 環構造 | - | 1 |
| 直鎖構造 | - | 0 |
| 発火点 [澁] | 274 | 274 |

基準値よりも明確に小さくなり、その後基準値を超えることはないが、基準値付近に留まった状態である。図4において基準値を厳格に適用すると、葉ノード2、すなわち根のみの分岐で木の成長を止めることとなる。しかし、この基準値と複雑度は試行によりある程度の数値変動があり、また相対交差判定誤差は葉ノード数4までおおむね一定で、5以上になると再び下り始めるので、ここでは葉ノード数4までを採用した木を参考に示す。

240件の分子で旧版の記述子を用いて作成した、葉ノード数4までを採用した決定木を図5に示す。図5では3つの分岐があるが、根に近い分岐ほど重要なルールである。図5で各節点内の上側の数値はその節点に属する分子の平均発火点であり、下側はその節点に属する分子が全データに占める割合をパーセンテージ表示したものである。そのため、根ノードでは上側は全データの平均発火点、下側は100%となる。各節点の直上に分岐ルールが記載されており、そのルールを満たす分子は左子へ、満たさない分子は右子へ分類される。

図5は入力ファイルで用いた記述子名をそのまま用いた表示であるため、根ノードの分岐ルールから順番に説明を加えると、以下ようになる。

ルール1: ベンゼン環を持たない分子は全体の83%あり、平均発火点は343 焓、ベンゼン環を持つ分子は17%あって平均発火点は494 焓である。

ルール2: (ベンゼン環を持たない分子のうち) エーテル結合を2個以上持つ分子は全体の8%あって、平均発火点は246 焓である。エーテル結合を1個持つ、または持たない分子は全体の75%あって平均発火点は352 焓である。

ルール3: (ベンゼン環を持たない、エーテル結合が1個以下の分子のうち) 炭素原子個数が7個以上の分子は20%あり、平均発火点は306 焓、炭素原子個数が6個以下の分子は56%あって平均発火点は369 焓である。

以上のルールのうち、ベンゼン環の有無は文献[11]でも根ノードのルールとなっており、共通した結果となっている。2つ目のルールはエーテル結合個数を用いているが、これは文献[11]では上位になかったルールであり、データ作成上の錯誤を修正したため現れたルールと考えられる。3つ目のルールは炭素原子個数を用いているが、その本質は分子サイズと考えられ、数値的に差が出やすい記述子としてたまたま炭素原子個数を採用したものと考えられる。炭素原子個数の他に分子サイズに関連の深い記述子は、分子量、沸点、融点であることが知られている。文献[11]でも分子サイズを用いたルールが根の右子と左子の両方に現れており、分子サイズは重要度の高い記述子であることがわかる。

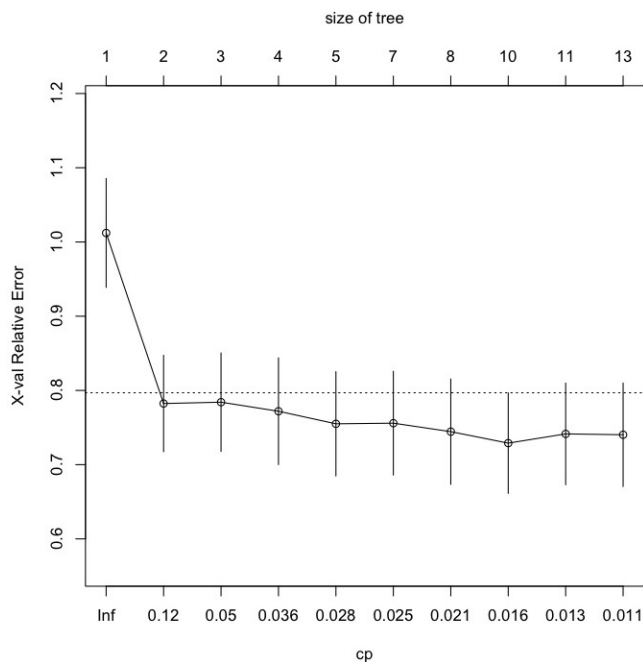


図4 修正後の学習データ240件を用いて作成した決定木における複雑度と交差判定誤差の関係

4.2 新版記述子を用いた決定木

次に、新版の記述子による240件の分子の学習データを用いて決定木を作成する。まずデフォルト設定で作成した決定木におけるplotcp関数の出力を図6に示す。図6によると、一標準偏差ルールに従っても4個の葉ノードが利用できるため、葉ノード4個で枝刈りを行う。新版の記述子では、一標準偏差ルールに従っているため、図5の決定木よりも過学習の影響が少ないことが期待できる。

新版の記述子による、枝刈り後の決定木を図7に示す。図7のルールは以下の通りである。

ルール1: ベンゼン環を持たない分子は全体の83%あり、平均発火点は343 焓、ベンゼン環を持つ分子は17%あって平均発火点は494 焓である。

ルール2: (ベンゼン環を持たない分子のうち) カルボニル基を持たない分子は全体の55%あって、平均発火点は314 焓である。カルボニル基を1個以上持つ分子は全体の28%あって平均発火点は400 焓である。

ルール3: (ベンゼン環を持たない、カルボニル基を持たない分子のうち) 分子量が67以上の分子は43%あり、平均発火点は290 焓、分子量が67以下の分子は12%あって平均発火点は401 焓である。

ルール1は旧版記述子で得られたものと全く同じであり、ベンゼン環の扱いは旧版と新版で全く同じであるために同じ結果になっている。2個目のルールは、部分構造に関する記述子の扱いが旧版と異なるために、新規に現れたルールと考えられる。3個目のルールは旧版記述子で得られたものと同様に、本質は分子サイズに関するルールであるものと考えられる。

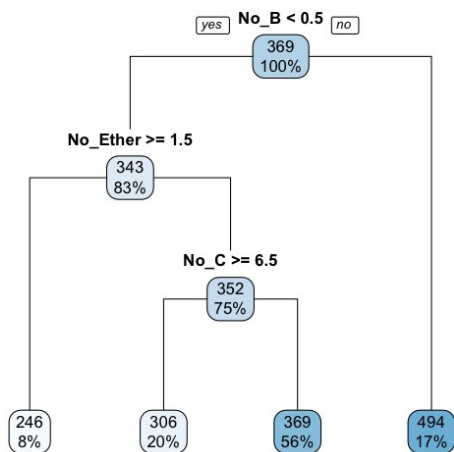


図 5 修正後の学習データ 240 件を用いて作成した枝刈り後の決定木

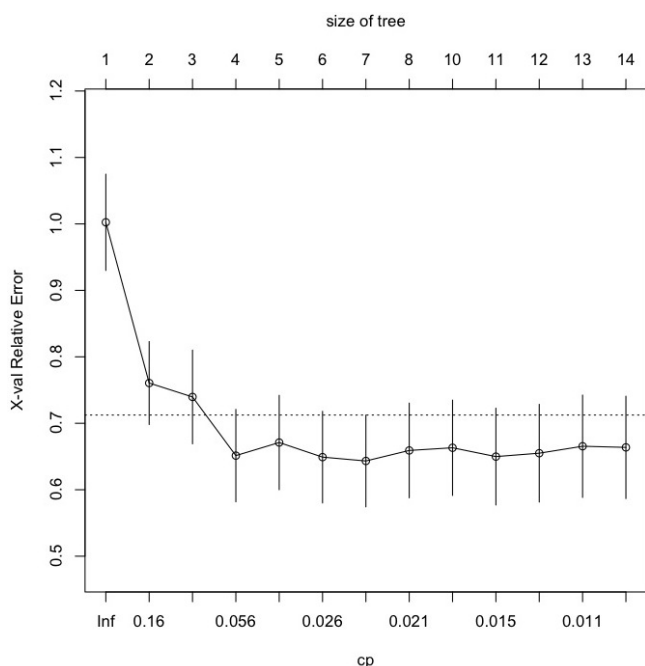


図 6 新版の記述子を用いた学習データ 240 件を用いて作成した決定木における複雑度と交差判定誤差の関係

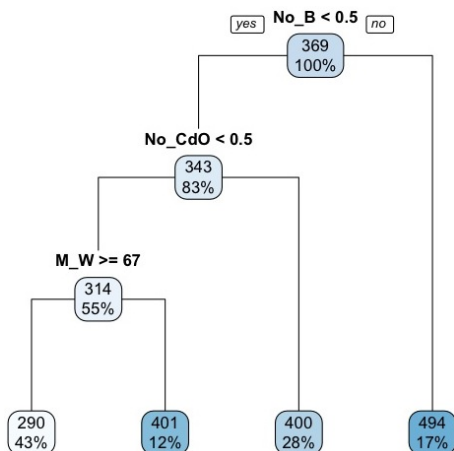


図 7 新版の記述子を用いた学習データ 240 件を用いて作成した枝刈り後の決定木

5. おわりに

本稿では、R の rpart パッケージを用いて炭化水素および類例分子の発火点を分類する決定木を作成した。各分子を特徴付ける記述子には、分子量、融点、沸点などの物理量に加えて特徴的な部分構造の個数を使用した。得られた決定木は過学習の影響が大きかったため、特徴的な部分構造の種類を増やし、包含関係の扱いを見直して記述子を再設定した。その結果、過学習の影響を減らした決定木を得ることができた。

「決定木を用いて発火点を予測する」ということは、「葉ノードのどれかに予測したい分子を割り当てる」ということであるため、葉ノードが 10 個程度以上あって、葉ノードが持つ平均発火点の差分が要求誤差程度になる必要がある。現段階では枝刈り後の葉ノードが 4 個しかないため、過学習を避けて発火点を予測することはまだ困難である。今後は、過学習の影響が少ない決定木を作成しても葉ノードが 10 個程度になる状態を目指したい。また、過学習に強いランダムフォレストなどの他の手法の利用も検討する。

謝辞

本研究の一部は科研費 16K13739 の助成を受けて行われた。関係各位に感謝する。

参考文献

- [1] 豊田秀樹：データマイニング入門，東京図書株式会社 (2008).
- [2] 平井 有三：はじめてのパターン認識，森北出版株式会社 (2012).
- [3] 国立医薬品食品衛生研究所 (NIHS)：国際化学物質安全性カード (ICSC) 日本語版，入手先 <http://www.nihs.go.jp/ICSC/> (2016.3.17).
- [4] 厚生労働省：職場のあんぜんサイト，入手先 <http://anzeninfo.mhlw.go.jp/> (2016.3.17).
- [5] 科学技術振興機構 (JST)：科学技術総合リンクセンター J-GLOBAL，入手先 <http://jglobal.jst.go.jp/> (2016.3.17).
- [6] Chemical Book，入手先 <http://www.chemicalbook.com/> (2016.3.17).
- [7] J.Gasteriger, T.Engel 編集，船津公人，佐藤寛子，増井秀行訳：ケモインフォマティクス 予測と設計のための化学情報学，丸善株式会社 (2005) .
- [8] Tsai, F.-Y., Chen, C.-C., Liaw, H.-J.: *A model for predicting the auto-ignition temperature using quantitative structure property relationship approach*, Procedia Engineering 45, 512-517, (2012).
- [9] Shi, J., Chen, L., Chen, W.: *Prediction on the auto-ignition temperature using substructural molecular fragments*, Procedia Engineering 84, 879-886, (2014).
- [10] 岡田 彩，林 亮子：競合学習を用いた炭化水素分子の類似度マップ，平成 26 年度電気関係学会北陸支部連合大会，講演論文集 F27 (2014).
- [11] 中田 侑江，林 亮子：炭化水素および類例分子の発火点決定木，情報処理学会研究報告，Vol. 2016-MPS-107, No. 14, (2016).