

熱化学計算問題の段階的解釈に基づく熱化学方程式の立式

経種 直之 佐藤 理史 小川 浩平 宮田 玲

名古屋大学大学院工学研究科

idane.naoyuki@c.mbox.nagoya-u.ac.jp

1 はじめに

大学入試では、言語運用能力に加え、多岐に渡る知識や能力が問われる。2011年から6年間に渡って実施された「ロボットは東大に入れるか」プロジェクトでは、主要5教科(国語、英語、数学、社会、理科)の大学入試問題の自動解法が研究された[1]。これらのうち、社会(歴史)、数学、英語の問題は比較的良好に解けたが、理科の問題の自動解法は非常に難しいことが判明した。

一般に、理科(物理・化学)の計算問題を解くためには、問題文を専門知識に基づいて解釈し、解釈結果を用いてある種の計算(推論)を行う必要がある。それゆえ、歴史の正誤問題の自動解法で通用したような、知識源(教科書)に対する表層的なテキスト照合では歯が立たない。

我々は、専門知識に基づく解釈と推論が必要な問題の具体例として、化学の熱化学計算問題を取り上げ、その自動解法に取り組んできた[2, 3]。熱化学計算問題は、熱化学反応を表す方程式を立式して解く計算問題である。必要な計算能力は、初等代数の計算スキルであり、この機械的実現は容易である。つまり、自動解法の中核は、いかにして問題文から熱化学方程式を立式するかという、問題文の解釈の機械的実現にある。前回の挑戦[3]では、開発用の問題の約半分しか解くことができなかった。今回は、熱化学方程式の立式を最終ゴールとして、そこに徐々に近づいていく複数のステップで問題文を解釈する方法を採用し、より多くの問題を解くことを目指す。

2 人間は問題をどう解くか

ここでは、図1に示す問題例を用いて、熱化学計算問題を解く標準的な方法を示す。

この問題が要求しているのは「プロパンの生成

水素、炭素(黒鉛)、プロパン C_3H_8 の燃焼熱は、それぞれ 286 kJ/mol、394 kJ/mol、2119 kJ/mol である。

プロパンの生成熱を求めよ。

(出典:「化学」数件出版, p114, 一部変更)

図1 熱化学計算の問題例

熱」である。生成熱の一般式は、以下で与えられる。

$$\text{構成元素のリスト} = \text{対象物質} + Q \text{ [kJ]} \quad (1)$$

プロパンの分子式は C_3H_8 であるので、その構成元素である炭素 C と水素 H の単体の分子式を代入すると、

$$a \text{ C(黒鉛)} + b \text{ H}_2 = C_3H_8 + Q \text{ [kJ]} \quad (2)$$

という式が得られる。ここで、 a と b は係数である。熱化学方程式では、両辺の元素の数が一致しなければならない。この制約より $a = 3$, $b = 4$ が得られ、最終的に

$$3 \text{ C(黒鉛)} + 4 \text{ H}_2 = C_3H_8 + Q \text{ [kJ]} \quad (3)$$

というプロパンの生成熱を表す熱化学方程式が得られる。

この式の Q の値を求めるために、問題文で示されている燃焼熱の情報を使用する。燃焼熱を表す熱化学方程式の一般式は、以下で与えられる。

$$\text{対象物質} + a \text{ O}_2 = b \text{ CO}_2 + c \text{ H}_2\text{O} + Q \text{ [kJ]} \quad (4)$$

水素、炭素(黒鉛)、プロパンのそれぞれに対して、単体の分子式を代入し、先と同じように係数を決定すると、以下の3つの熱化学方程式が得られる。

$$\text{H}_2 + \frac{1}{2} \text{ O}_2 = \text{H}_2\text{O} + 286 \text{ [kJ]} \quad (5)$$

$$\text{C(黒鉛)} + \text{O}_2 = \text{CO}_2 + 394 \text{ [kJ]} \quad (6)$$

$$C_3H_8 + 5 \text{ O}_2 = 3 \text{ CO}_2 + 4 \text{ H}_2\text{O} + 2119 \text{ [kJ]} \quad (7)$$

式(5)-(7)を用いて式(3)の Q を求めると、

$$Q = 107 \text{ [kJ/mol]} \quad (8)$$

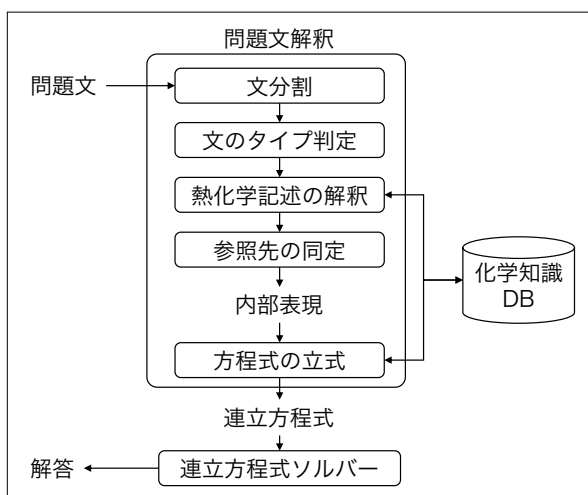


図2 自動解答システムの構成

という解が得られる。以上が、人間が熱化学計算問題を解く場合の標準的な方法である。

上記の方法を機械化することを前提に、問題を解くためにどのような知識・スキルが必要かを整理しよう。

物質(元素・化合物)に関する知識 「水素」、「炭素(黒鉛)」、「プロパン」が物質であるということに加え、分子式が与えられていない物質に対しては、単体の分子式が必要である。さらに化合物に対しては構成元素の情報が必要であり、中和反応などでは、液性(酸・アルカリ)の情報、イオン価の情報が必要となる。

熱化学反応に関する知識 「燃焼(熱)」、「生成(熱)」が熱化学反応であり、それらが一般に、どのような熱化学方程式で表現されるかという知識が必要である。

初等代数の計算スキル 反応式の係数を決定したり、複数の式から変数の値を決定するために、連立方程式を解くスキルが必要である。

日本語の理解スキル 当然のことながら、問題文を読み、どんな情報が与えられ、何を求めることが要求されているかを理解する必要がある。問題文の理解のためには、初等代数の計算スキルは不要であるが、物質(元素・化合物)に関する知識、および、化学反応に関する知識は必須である。なぜならば、これらの知識がなければ、問題が何を要求しているのか、理解不能である。

3 自動解答システム

作成した自動解答システムの構成を、図2に示す。システムは、問題文解釈と連立方程式ソルバー

表1 文のタイプ判定

物質記述	[物質名 = メタンとプロパンの混合気体] が [質量 = 標準状態において 33.6L] ある。
反応記述	[反応 = 水酸化ナトリウムの固体 4.0g を 0.10mol/L の塩酸 1L に加える] と [結果 = 10.1kJ の熱が発生した]
仮定記述	[記述物 = メタン CH ₄ の燃焼熱] は [値 = 890kJ/mol] である。
問い記述	[対象 = ブタン 1.0mol が完全燃焼するとき発生する熱量] を、熱化学方程式 (1)–(3) を用いて求めると何 [単位 = kJ] になるか。
背景記述	(それ以外 — 以降の処理の対象としない)

という2つのモジュールと、化学知識データベースから構成される。化学知識データベース(リレーショナルデータベース)は、元素、単体、分子、イオン、の4つのテーブルから構成され、化学式(分子式)、液性(酸・アルカリ)、状態(気体・液体・固体)、イオン価の情報などが記述されている。

システムの中核をなす問題文解析モジュールは、5段階のステップで問題文(テキスト)から連立方程式(熱化学方程式)を立式する。問題文は、句点であらかじめ分割し、1文1行の形式とする。問題文中に含まれる熱化学方程式は独立した行として扱い、先頭にタブを挿入して通常の文と区別するものとする。なお、問題文中に含まれる熱化学方程式は、式として解釈し、出力する。

3.1 文分割

後段の解析を容易にするために、ヒューリスティックを用いて文を分割する。たとえば、文(1)は文(2a), (2b), (2c)のように3文に分割する。

- (1) 二酸化炭素の生成熱を 394kJ/mol、水(液体)の生成熱を 286kJ/mol とすると、メタノール(液体)の生成熱は何 kJ/mol か。
- (2) a. 二酸化炭素の生成熱を 394kJ/mol とする。
b. 水(液体)の生成熱を 286kJ/mol とする。
c. メタノール(液体)の生成熱は何 kJ/mol か。

3.2 文のタイプ判定

文型パターンを用いて、文を表1に示す5つのタイプに分類する。背景記述以外の4タイプに対しては、その文の重要情報(物質名、質量、反応など)が書かれている範囲を同定する。

3.3 熱化学記述の解釈

同定された重要部分を解析し、熱化学問題の記述として解釈する。具体的には、重要部分に含まれて

```

{: type=>"反応記述",
 : args=>
  {: type=>"反応",
   : 種類=>"中和",
   : 反応前=>[{: type=>"物質",
                : 物質名=>"水酸化ナトリウム",
                : 状態=>"固体",
                : 質量=>"4.0g"},
               {: type=>"物質",
                : 物質名=>"塩酸",
                : 濃度=>"0.10mol/L",
                : 液体積=>"1L"}],
   : 反応後=>[{: type=>"熱量",
                : 値=>"10.1kJ"}]}

{: type=>"問い記述",
 : args=>
  {: type=>"問い",
   : 対象=>{: type=>"反応",
            : 種類=>"燃焼",
            : 反応前=>[{: type=>"物質",
                        : 物質名=>"ブタン",
                        : 質量=>"1.0mol"}],
            : 反応後=>[]},
   : 求めるもの=>"熱量",
   : 単位=>"kJ"}]}

```

図3 熱化学記述の解釈後の内部表現

いる用語が何(化学反応、物質名、質量、濃度など)を表しているのかを正規表現パターンにより決定し、必要な情報を抽出する。このステップにより、表1の反応記述と問い記述は、図3に示す内部表現に変換される。

3.4 参照先の同定

以下のような参照表現の参照先を同定する。

- (3) a. この混合気体を...
 b. このとき発生する熱量を...

前者の場合は直前の混合気体を指すものと、後者の場合は直前の化学反応を指すものと解釈し、内部表現を組み替える。

3.5 方程式の立式

以上のステップで文章解析は終了し、得られた内部表現に基づいて、熱化学方程式を立式する。熱化学方程式の一般式は、反応の種別によって定まる(表2)。立式の過程で、必要に応じて化学知識データベースを参照し、以下のことを行う。

1. 物質の分子式、および、構成元素の決定

表2 各反応の一般式(係数は省略した)

生成	構成元素のリスト = 対象物質 + Q [kJ]
燃焼	対象物質 + O ₂ = CO ₂ + H ₂ O + Q [kJ]
蒸発	物質(液体) = 物質(気体) - Q [kJ]
溶解	物質(固体) = 物質 aq + Q [kJ]
中和	酸性物質 + 塩基性物質 = 中和生成物の水溶液 + H ₂ O + Q [kJ]
結合エネルギー	物質の構成元素 1 つの気体リスト = 物質(気体) + 結合数 × 結合エネルギー

問題文

液体のメタノール CH₃OH を 16g とり、完全に燃焼させたところ、二酸化炭素と液体の水が生成して 363kJ の発熱があった。二酸化炭素の生成熱を 394kJ/mol、水(液体)の生成熱を 286kJ/mol とすると、メタノール(液体)の生成熱は何 kJ/mol か。

連立方程式(求める変数は Q₁、単位は kJ)

$$\begin{aligned}
 C(\text{黒鉛}) + 2 * H_2(\text{気}) + 1/2 * O_2(\text{気}) &= CH_3OH(\text{液}) + Q_1 \text{ [kJ]} \\
 1/2 * CH_3OH(\text{液}) + 3/4 * O_2(\text{気}) &= 1/2 * CO_2(\text{気}) + H_2O(\text{液}) + 363 \text{ [kJ]} \\
 C(\text{黒鉛}) + O_2(\text{気}) &= CO_2(\text{気}) + 394 \text{ [kJ]} \\
 H_2(\text{気}) + 1/2 * O_2(\text{気}) &= H_2O(\text{液}) + 286 \text{ [kJ]}
 \end{aligned}$$

図4 方程式の立式の具体例

2. 物質の標準状態での三体(液体、気体、固体)の決定
3. 物質の分子量に基づく、質量の mol への標準化
4. 酸・塩基の決定
5. 価数の決定

熱化学方程式に物質の分子式等を代入した後、係数が不定の場合は、両辺の元素数比較により係数を決定する。なお、このステップで立式される式には、熱化学方程式以外の式(たとえば、水の温度上昇を表す式)も含まれる。

以上の5ステップにより、問題文を連立方程式と求めるべき変数に変換する。具体例を図4に示す。

4 解答システムの性能

実装したシステムが、センター試験レベルの熱化学計算問題をどの程度解けるかを調べた。以下に示す出典から、熱化学計算問題を抜き出し、枝間はそれぞれ独立した問題として扱った。

開発用セット — システムの開発に使用

マーク式総合問題集 化学 I (河合塾) [2007, 2009, 2011, 2013, 2015, 2017, 2019]、大学入試センター試験実戦問題集 化学 I (大学入試模試センター) [2008, 2010, 2012, 2014, 2019]

評価用セット — システムの開発に使用せず

センター試験(本・追試験) [1997, 1999-2002, 2004, 2006, 2008, 2009, 2011-2016]、大学入試センター試験

表3 システムの性能

問題種別	開発用		評価用	
	正解	計	正解	計
1 与式のみで解答可能	7	7	6	7
2 定義上の反応熱	11	11	7	9
3 特定質量の反応	8	11	0	10
4 結合エネルギー	4	4	1	7
5 複数の反応	6	6	0	6
6 混合気体	2	3	0	13
7 その他	0	5	0	5
合計	38	47	14	57

表4 エラー分析

原因	開発	評価
o 本枠組では解けない	2	4
u 未実装	5	19
f 立式の不備	1	2
i 熱化学記述の解釈の失敗	1	15
s 文分割の失敗	0	3

実戦問題集 化学 I (代々木ゼミナール) [2006, 2008, 2010, 2012, 2014, 2016, 2017, 2019]

本システムの解答状況を表3に示す。この表に示すように、開発用セットは47問中38問(81%)を正しく解くことができたが、評価用セットは57問中14問(25%)しか解くことができなかった。

この表で、「与式のみで解答可能(1)」は、問題文中の熱化学方程式のみから解ける(問題文の解釈がほぼ不要な)問題であり、「定義上の反応熱(2)」は、立式に質量が絡まない問題である。これらの問題は、ほぼ解けるレベルに達している。

一方、「特定質量の反応(3)」は立式に質量が絡む問題、「結合エネルギー(4)」は結合エネルギーが関与する問題、「複数の反応(5)」は反応熱で水を温めるような複数の反応を内在する問題、「混合気体(6)」は混合気体を対象とする問題である。これらの問題は、開発用セットでは比較的解けていたが、評価用セットでは1問しか解けなかった。

表4に、解けなかった原因の分析結果を示す。質量や質量の比を求める問題は、開発用セットには3問しかなかったため、これらの解法を実装しなかったが、評価用セットには混合気体の問題が多く含まれており、その結果、質量や比を求める問題が全体で12問存在した。これらの問題は、原理的には、現在の連立方程式を立式して解くという枠組みで解くことができるが、熱化学記述の解釈と方程式の立式の両ステップを補強する必要がある。

評価用セットで「特定質量の反応(3)」の問題が解けなかった理由の大半は、「熱化学記述の解釈の

失敗(i)」である。質量が絡む問題では、問題記述の表現のバリエーションが多くなるが、今回使用した開発用セットでは、このバリエーションが十分にカバーできなかった。未知の問題に対する頑健性を向上させるためには、より広いバリエーションをカバーする必要がある。

表現のバリエーションには、分野固有のバリエーションと、日本語一般のバリエーションの2種類がある。開発に使用する問題例を増やせば、このバリエーションをある程度カバーできるとは考えられるが、両方のバリエーションを実際の問題例のみに求める方法論には限界がある。理想的には、後者のバリエーションをカバーする日本語の汎用文解析に分野固有知識を追加すれば、それらが連携して動くようなアーキテクチャが望ましい。今後は、そのような方向を模索していく必要がある。

5 おわりに

本稿では、熱化学計算問題の自動解法を示した。その中核は、日本語で書かれた問題文を熱化学方程式に変換する処理であり、そのためには、化学知識の援用が必要となる。開発用セットに対しては、前回の挑戦より多くの問題を解くことができたが、評価用セットに対する正解率は、前回と同程度に留まった。

本研究は、大学入試問題を対象とした自動解法の研究であるが、その本質は、テキスト理解にある。我々は、『『テキストを読む』ことは、読み手が持っている知識とテキストとの相互作用である』と考える。そのため、

- (4) 水素、炭素(黒鉛)、プロパン C₃H₈ の燃焼熱は、それぞれ 286kJ/mol, 394kJ/mol, 2219kJ/mol である。

というテキストに対して、書かれている情報を直接問う「プロパンの燃焼熱は？」のような問題ではなく、明示的に書かれていない「プロパンの生成熱は？」のような問題を扱う。このような問題を解く方法が十分に明らかになったならば、テキストの自動理解に近づいたことになる。

参考文献

- [1] 新井紀子, 東中竜一郎(編). 人工知能プロジェクト「ロボットは東大に入れるか」. 東京大学出版会, 2018.
 [2] 吉田達平, 松崎拓也, 佐藤理史. 大学入試化学の計算問題の自動解答システム. 人工知能学会第29回全国大会論

文集, 1K2-1, 2015.

- [3]加藤汰一, 松崎拓也, 宮田玲, 佐藤理史. センター試験「化学」計算問題の自動解答システムの作成. 言語処理学会第 25 回年次大会発表論文集, pp. 894–897, 2019.