

Group Contribution Method의 작용기 분석을 위한 전문가 시스템 개발

이건홍 · 양기주* · 정준영 · 이인범

포항공과대학 화학공학과 공정산업의 지능자동화연구센터/산업과학기술연구소 이공부문 화공연구부

*한국통신 소프트웨어 연구소

(1993년 5월 10일 접수, 1993년 7월 5일 채택)

Development of an Expert System for Functional Group Analysis in Group Contribution Method

Kun-Hong Lee, Gijoo Yang*, June Young Jung and In Beum Lee

Automation Research Center, Dept. of Chem. Eng., Pohang University of Science and Technology/
Chemical Engineering Division, Research Institute of Industrial Science and Technology

*Software Research Laboratories, Korea Telecom

(Received 10 May 1993; accepted 5 July 1993)

요 약

주어진 분자식으로부터 작용기의 조합을 얻어내는 컴퓨터 프로그램을 개발하였다. 이 프로그램은 rule-based 전문가 시스템의 형태를 가지고 있으며, common LISP를 사용하였다. 추론기관은 전방향 추론을 채택하여 직접 개발하였고, 지식베이스는 meta rule 80개, select rule 34개 및 combine rule 7개로 구성되었다. Joback Group Contribution Method에 따라서 작용기를 설정하였으며, 8개 이하의 그룹으로 구성된 화합물들에 대하여 오류없이 작용기의 조합을 구할 수 있었다.

Abstract—A computer program which gave a functional group set from a given chemical formula was developed. This program has the form of a rule-based expert system, using the common LISP. The inference engine was developed using the forward chaining technique, and the knowledge base was composed of 80 meta rules, 34 select rules and 7 combine rules. The functional groups were chosen following the Joback group contribution method, and the functional group sets were successfully obtained for the substances of less than 8 groups.

1. 서 론

화학공정의 설계에는 그 대상이 되는 물질들의 물성치 데이터가 필수적이다. 물성치는 실험을 통하여 얻는 것이 가장 정확하나, 때로는 경제적인 이유로 인하여 추정된 물성치를 사용하게 된다. 특히, 새로운 물질을 합성하거나 많은 성분이 섞여 있는 혼합물을 다루게

되는 경우에는 추정된 물성치에 주로 의존하게 된다. 물성치를 추정하는 방법은 크게 2가지로 나누어 진다. 첫째는 많은 실험데이터의 회귀분석(regression analysis)을 통하여 얻어진 특정한 함수형태를 가진 상관식들이며, 거의 대부분의 물성치 예측법이 이에 속한다. 이러한 상관식들은 상당히 정확한 예측치를 제공하나, 그 적용범위가 매우 제한적이라는 단점을 가진다. 둘

책은 Group Contribution Method(이하 GCM)를 이용하는 방법이며 제시된 작용기(functional group)의 조합으로 이루어지는 모든 물질에 적용이 가능하나, 예측치의 정확도가 물질에 따라서 불규칙하게 변화할 뿐 더러 경우에 따라서는 매우 부정확한 예측치를 제공한 다.

GCM은 한 분자를 이루는 작용기들의 특성들을 알면 그 작용기들의 조합으로 이루어지는 분자의 특성을 예측할 수 있다는 가설에 바탕을 두고 있다. 그러나, 이러한 가설이 정확하게 성립되는 경우는 구성 원자의 원자량으로부터 분자량을 계산하는 경우가 유일하며, 다른 모든 경우에는 다소간의 오차를 수반하게 된다. 특히 분자의 크기가 커질수록, 분자가 복잡해질수록 오차는 증가하게 되어, 결과적으로 그 사용범위가 제한되게 된다. 이를 역으로 생각해 보면, GCM의 정확도를 개선할 수 있는 방법이 고안되기만 한다면, 광범한 적용범위로 인하여 GCM은 물성치 추정에 있어서 가장 이상적인 방법으로 부각될 가능성이 있다.

GCM이 정확한 예측을 하지 못하는 원인으로는 몇 가지를 생각할 수 있다[1]. 첫째는 전체의 특성은 부분의 특성을 통하여 예측할 수 있다는 GCM의 기본 가정 자체가 옳지 않기 때문이다. 둘째는, 특정한 GCM이 가진 결함이다. 같은 물질에 대해서도 GCM에 따라서 정확도에 많은 차이가 있으며, 이는 GCM을 개발할 때 어떤 작용기를 기본 그룹으로 선택하였으며, 그 선택의 근거는 무엇인가, 또 선택된 작용기의 특성치를 어떠한 방법을 사용하여 결정하였는가에 따라서 달라지게 된다. 따라서 GCM의 사용자 입장에서는 특정한 GCM을 선택할 때 이러한 점을 반드시 고려하여야 한다. 셋째는, 주어진 GCM을 사용할 때 발생하는 사용자에 의한 오류이다. GCM의 사용자는 작용기라는 조각을 조립하여 주어진 물질의 구조를 구성하는 퍼즐작업을 수행하여야 하며, 이 과정에서 오류를 범할 수 있다. 이는 Ambrose법과 같은 복잡한 GCM에서 발생하기 쉬우며, 다른 GCM의 사용에서도 흔히 발생되고 있다.

3번째의 오류는 몇 가지의 물질을 대상으로 하는 사용자에게는 별로 심각한 문제가 아니지만, GCM의 정확도를 개선하기 위해서 가능한 한 많은 수의 물질들을 다루어야 하는 연구자의 입장에서는 매우 중요한 문제로 부각될 수 있다. 본 연구에서는 이상의 3가지 오류중 3번째의 오류를 방지하기 위하여, 주어진 분자를 작용기들로 쪼갤 때 사용자의 수작업을 대신할 수 있는 컴퓨터 프로그램을 개발하였다.

2. 연구의 방향 설정

분자의 물성치를 작용기들로부터 예측하려는 시도는

Table 1. Various group contribution methods

Physical property	Developer
Critical temperature	Ambrose, Joback
Critical pressure	Ambrose, Joback
Critical volume	Ambrose, Joback
Boiling temperature	Joback
Freezing temperature	Joback
Liquid heat capacity	Chueh-Swanson, Misernard
Ideal gas heat capacity	Joback, Yoneda, Thinh, Benson
Heat of formation	Joback, Yoneda, Thinh, Benson, Cardozo
Gibbs free energy of formation	Joback, Thinh
Entropy	Yoneda, Thinh, Benson
Activity coefficient	Wilson(ASOG), Fredenslund(UNIFAC)
Gas viscosity	Reichenberg
Liquid viscosity	Orrick-Erbar, Grunberg-Nissan
Thermal conductivity	Roy-Thodos

매우 오랜 역사를 가지고 있으며[2], 그 적용 범위로 탄화수소, 유리[3], 고분자[4] 및 생체분자 등 다방면에 걸쳐 있다. 화학공정의 설계에 필요한 열역학적 물성치를 예측하기 위한 GCM으로는 Ambrose 법[5, 6], Joback 법[7], Fedors 법[8], Thinh 법[9, 10] 등 1차 물성치를 예측하기 위한 방법들과 ASOG 법[11] 및 UNIFAC 법[12] 등의 2차물성치(여기서는 활동도계수를 의미함)를 얻기 위한 방법들이 개발되어 있으나, 정확도의 문제로 인하여 추정결과를 전적으로 신뢰받지는 못하고 있다. Table 1은 현재까지 개발된 여러 가지 GCM들로 추정 가능한 물성치들을 나타내고 있다[13]. 본 연구에서는 위에 열거된 GCM중 가장 넓은 적용범위를 가지고 있는 Joback의 GCM을 대상으로 선택하였다. Joback법은 Ambrose법에 비하여 훨씬 간단하면서도 비슷한 정도의 정확성을 보이는 것으로 알려져 있다[13].

Joback법에서는 40개의 작용기들을 조합하여 모든 분자들의 구조를 재현하여야 한다. 따라서, 어떤 분자가 N개의 작용기로 구성되어 있을 때 40^N 가지의 그룹집합을 대상으로 타당성을 조사하여야 한다. 예를 들면, 어떤 분자가 3개의 작용기의 조합으로 이루어진다고 할 때 검색의 대상이 되는 그룹집합의 수는 64000개나 되며, 이 숫자는 분자의 크기가 증가함에 따라서 지수적으로 증가하게 된다. 따라서, 모든 가능한 그룹집합을 대상으로 한 검색은 현실적으로 불가능하며, 검색의 대상이 되는 그룹집합의 갯수를 줄일 수 있는 합리적인 방법을 모색하여야 한다.

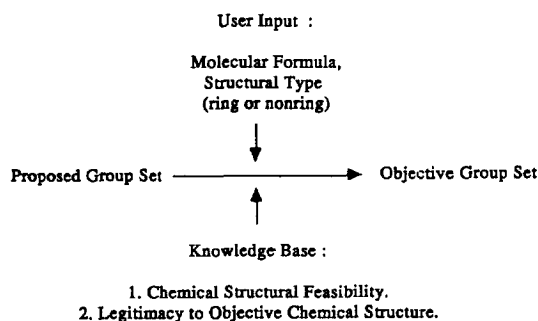


Fig. 1. Functional scheme of the expert system.

주어진 그룹집합으로부터 조립된 분자의 화학구조적 타당성의 조사는 검색 대상을 줄일 수 있는 가장 합리적인 방법이다. 예를 들면, (-CH₃, >CH₂)와 같은 그룹집합은 조립후 화학 결합이 하나 남게 되므로 완성된 분자를 구성하는 그룹집합의 검색 대상에서는 제외될 수 있을 것이다. 그러나, 대상분자가 복잡해지면 어떤 그룹집합을 제외할 것인지가 간단하지 않게 되며, 100%의 확실성을 보장하는 방법을 고안하는 것도 사실상 불가능하게 된다.

본 연구에서는 이 문제를 rule-based 전문가 시스템의 도입으로 해결하고자 하였다. Rule-based 전문가 시스템의 장점은 단편적이고 제한적인 법칙(rule)들을 사용하여 개략적인 해답을 얻어낼 수 있다는데 있으며, 프로그램상에서는 논리연산 부분과 법칙(rule)부분이 분리되어 있어서 추후에 법칙들을 추가함으로써 해답의 정확도를 향상시킬 수 있다는데 있다. 이는 예상치 못한 구조의 분자를 새로 접하게 되었을 때 그에 해당하는 법칙을 추가함으로써 문제를 해결할 수 있다는 장점이 있다.

작용기의 조합을 사용해서 분자의 구조를 조립하는 전문가 시스템의 시초로는 DENDRAL이 있으며, 고가이기는 하지만 상업화된 프로그램의 구입도 가능하다 [14]. 그러나, DENDRAL은 화학조성식, mass spectrum 및 NMR spectrum을 입력데이터로 요구하므로 본 연구에서 보다는 훨씬 적은 숫자의 대상을 검색하게 된다. 특정한 GCM에 맞는 전문가 시스템을 만들기 위해서는 DENDRAL을 수정하는 방법도 있으나, DENDRAL의 프로그램 코드를 입수하기는 어려우므로 본 연구에서는 GCM을 대상으로 하는 전문가 시스템을 자체적으로 개발하였다 [15].

3. 전문가 시스템의 개요

Fig. 1은 위의 문제를 해결하기 위한 전문가 시스템의

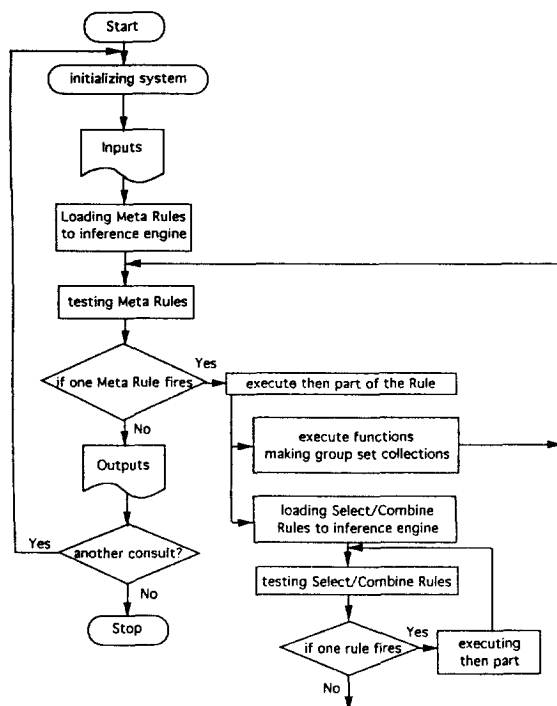


Fig. 2. Flow diagram of the inference engine.

기능을 설명하는 개념도이다. 이 전문가 시스템에서는 분자식과 ring 구조의 유무를 입력받은 후, 지식베이스에 내장되어 있는 규칙들을 이용하여 제한된 그룹집합의 화학구조적 타당성 및 목적 화합물에 대한 적합성을 검토하게 된다. 이러한 작업을 가능한 모든 그룹집합에 대하여 반복함으로써 최종적인 그룹집합을 최적의 경로를 통해서 찾도록 하였다. 전문가 시스템을 개발하는 경우, 상용화된 전문가 시스템 툴(tool)을 이용하는 경우가 대부분이나, 본 연구에서는 common LISP를 사용하여 전방향 추론(forward chaining)을 하는 추론기관(inference engine)을 직접 개발, 사용하였다. Fig. 2는 본 연구에서 사용된 추론기관의 알고리즘을 개략적으로 나타낸 것이다.

3-1. 알고리즘

본 연구에서 개발된 전문가 시스템의 알고리즘은 다음과 같이 5단계로 나누어지며, 입력된 분자식으로부터 시작하여 각 단계를 거쳐서 최종적으로 그 분자를 표현하는 그룹집합을 찾게 된다.

3-1-1. 1단계(전처리 단계)

분자식을 입력받아서 화합물을 구성 그룹으로 조립하는데 필요한 정보를 재구성한다. 그룹들은 크게 세 부류로 나누어진다. 이들은 탄소를 포함하는 그룹들,

산소, 질소, 황을 포함하는 그룹들, 그리고 할로젠 원소를 포함하는 그룹들이다. 따라서 분자식에 포함된 원소에 따라서 그룹 표현에 사용될 각 부류의 그룹의 갯수를 결정하는 것이 이 단계에서 이루어진다. 예를 들면, 에틸 알콜(ethyl alcohol)의 분자식(C_2H_6O)가 입력된 경우에, 이 분자는 탄소를 포함하는 그룹 2개, 산소를 포함하는 그룹 한 개로 표현된다고 할 수 있다. 한편, 탄소와 산소, 탄소와 질소를 동시에 포함하는 $>C=O(\text{nonring})$, $>C=O(\text{ring})$, $-COOH$, $-COO-$, $O=CH-$, $-CN$ 등의 그룹들은 탄소를 바탕으로한 그룹들에 포함된다.

3-1-2. 2단계(초기 그룹 형성 단계)

전처리 단계에서 결정된 세 부류의 그룹들을 ring 구조의 존재 유무에 따라서 탄소를 기본으로 하는 그룹들로 이루어진 초기 주 그룹집합(initial main group set)과 산소, 질소, 황을 포함하는 그룹들로 이루어진 초기 부 그룹집합(initial sub group set), 그리고 할로젠 원소를 갖는 그룹들로 이루어진 초기 할로젠 그룹집합(initial halogen group set)들로 나누어서 각각 독립적으로 형성한다. 이 때 초기 주 그룹집합과 초기 부 그룹집합에는 동일한 그룹이 각각 하나씩 포함되어 있으나 초기 할로젠 그룹집합에는 입력된 화합물을 표현하는데 필요한 각 할로젠 원소만큼 그 원소에 해당하는 그룹을 포함함으로써 완성된 형태로 구성되어진다.

3-1-3. 3단계(group set collection의 형성)

이 단계에서는 초기 주 그룹집합으로부터 1단계에서 결정된 갯수만큼의 그룹들로 구성된 그룹 집합들로 이루어진 group set collection(x_{carbon})을 형성한다. 우선은 점진적인 결합 방법의 기초를 마련하기 위해 초기 주 그룹집합을 스스로 결합시켜서 두 개의 그룹들로 구성된 그룹집합들로 이루어진 group set collection(2)를 만든다. 여기서 점진적인 방법이란 이미 만들어진 group set collection(a)와 group set collection(b)로부터 group set collection(a+b)를 만드는 방법을 말한다. 즉 화합물을 표현하는데 4개의 그룹을 필요로 하는 경우에는 앞 단계에서 만들어진 group set collection(2)를 자신끼리 결합하여 4개의 그룹을 갖는 그룹집합들의 모임인 group set collection(4)를 만든다. 또 6개의 그룹을 필요로 하는 분자의 경우에 group set collection(6)은 초기 그룹집합과 group set collection(2)를 결합하여 group set collection(3)을 만들고 나서 이 group set collection(3)을 자신끼리 결합하여 만들어진다. 점진적 방법은 초기 그룹집합에서 시작하여 하나 하나 그룹 첨가하는 방법에 비해서 그룹집합의 수가 증가할수록 효율적으로 탐색의 영역을 줄이는 효과를 가져온다. 이 단계에서는 점진적 방법을 통해서 최소한의 경로를 거쳐서 입력된 분자식을 표현하는데 필요한 만큼의 탄소를 포함하는 그룹들로

이루어진 group set collection(x_{carbon})을 만든다. 이 단계에서는

$$|\text{group set collection}(a)| + |\text{group set collection}(b)| \\ \rightarrow |\text{group set collection set}(a+b)|$$

가 성립되도록 지식베이스를 통해 결합이 효과적으로 제한받게 된다. 여기서 $|\text{group set collection}(a)|$ 는 group set collection(a)에 포함된 그룹집합의 갯수를 말한다.

3-1-4. 4단계(최종 결합 단계)

이 단계에서는 초기 주 그룹집합으로부터 시작하여 3단계에서 완성된 group set collection(x_{carbon})과 초기 부 그룹집합, 초기 할로젠 그룹집합을 결합하여 최종적인 final group set collection을 형성한다. 이 단계에서도 앞의 3단계에서와 마찬가지로 화학 구조적 타당성 및 목적 화합물의 적합성이 고려되어 결합이 제한받는다.

3-1-5. 5단계(group set 선택단계)

분자를 표현하는데 필요한 갯수만큼의 그룹을 포함하는 그룹집합들의 모임인 final group set collection 으로부터 입력된 화합물에 해당하는 그룹집합을 선택한다. 이를 위해서 2개의 규칙이 고려되는데, 그 첫 번째는 최종단계의 그룹집합으로서의 타당성을 검토한다. 대상 화합물의 구조가 비환구조(nonring)일 경우 최종단계의 그룹집합은 결합되지 않은 자유전자를 갖지 않는다. 즉 $(-CH_2- -CH_3 -CH_3)$ 는 최종단계의 그룹집합으로서 타당하지만 $(-CH_2- -CH_2- -CH_3)$ 는 하나의 자유전자가 남게 되므로 완성된 화합물을 표현할 수 없다. 반면에 환구조(ring) 화합물일 경우는 최종적인 그룹집합에 같은 타입의 결합 두 개를 포함해야 한다. 즉 $(-CH_2- -CH_2- -CH_2-)$ 는 최종단계의 그룹집합으로서 타당하지만 $(-CH_3 -CH_2- -CH_3)$ 는 최종단계의 그룹집합이 될 수 없다. 다음은 화합물의 각 원소의 갯수와 그룹집합의 원소의 갯수가 같은 그룹집합을 찾음으로써 최종 목표에 도달하게 된다.

3-2. 지식베이스의 개발

본 연구에서는 상용 전문가 시스템 툴에서 가장 보편적으로 이용되고 있는 생산규칙(production rule)을 이용하여 지식베이스를 구축하였다. 생산규칙은 if로 시작하는 조건부와 then으로 시작하는 실행부로 나누어지는데, 조건부에서는 실행부의 내용이 실행되는데 필요한 조건들이 열거되어 있으며, 추론기구는 각 규칙의 모든 조건이 만족되었을 때 실행부를 실행하도록 한다.

한편 지식베이스는 규칙들을 저장하는 장소로서, 본 연구에서는 Fig. 3에서 볼 수 있는 바와 같이 추론의 속도 및 기억장치 비용을 고려하여 구조화된 지식베이

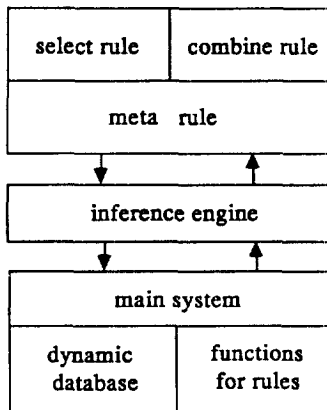


Fig. 3. Structure of the expert systems.

Meta Rules

```

RULE chem1
if (process step 1)
  (others_o is unknown)
  ($ include oxygen 1)
then (others oxygen 1)
    (others_o is known)

RULE chem2
if (process step 1)
  (others_o is unknown)
  ($ include oxygen 2)
then (others oxygen 2)
    (others_o is known)

RULE chem3
if (process step 1)
  (others_o is unknown)
  ($ not-match goal form-letters o)
then (others_o is known)

RULE chem3
if (process step 1)
  (others_n is unknown)
  ($ include nitrogen 1)
then (others nitrogen 1)
    (others_n is known)

RULE chem4
if (process step 1)
  (others_n is unknown)
  ($ not-match goal form-letters n)
then (others_n is known)

RULE chem5
if (process step 1)
  (others_s is unknown)
  ($ include sulfur 1)
then (others sulfur 1)
    (others_s is known)

RULE chem6
if (process step 1)
  (others_s is unknown)
  ($ not-match goal form-letters s)
then (others_s is known)

```

Select Rule

```

RULE select1
if (process step 2)
  ($ eq_max h_num h)
  (context bond -)
then (select is done)

RULE select2
if (process step 2)
  (select is undone)
  ($ eq_max h_num h)
  (context bond -)
then (select is done)

RULE select 3
if (process step 2)
  (select is undone)
  ($ less 4 goal class)
  ($ more_length context bond 4)
then (select is done)

```

Combine Rule

```

RULE comb1
if ($ check_bond_combine context data)
)
then (combine test done)

RULE comb2
if (process step 3)
  ($ combine-test-of-step-3)
then (combine test done)

RULE comb3
if (combine test undone)
  (context bond -)
  (data bond -)
  ($ rst_larg_used context data)
  ($ rst_larg_used data context)
  ($ check_bond_length context data)
-)
then ($ combine -)

```

Fig. 4. Sample rules in the knowledge base.

스를 구축하였다. 즉 규칙들을 각 기능상 특성에 따라서 분류하여 meta rule, select rule, combine rule로 나누었다. 전체적으로 추론기구는 meta rule을 추론하면서 필요에 따라서 select rule과 combine rule을 추론한다. Meta rule은 모두 80개의 규칙으로 구성되어 있으며, 초기 그룹집합의 형성 및 초기 추론에 필요한 환경을 마련하고, 전체 과정의 진행상황을 관리하여 다음 단계의 작업을 실행시키는데 필요한 지식들과 점진적인 방법을 통해서 입력된 분자식을 표현하는 최종적인 그룹집합을 포함하고 있는 final group set collection을 형성하는 전략들로 구성되어 있다. Select rule은 34개의 규칙들로 이루어졌으며, 추론의 중간 단계에서 이미 만들어진 group set collection에 포함된 그룹집합들을 검색하여 이 group set collection에 포함된 그룹집합들을

걸르는(filtering)에 필요한 지식들로 이루어진다. Combine rule은 모두 7개의 규칙들을 가지며, 두 개의 그룹 집합을 결합하는데 있어서 화학구조적 타당성과 대상 화합물을 표현하는데의 적합성을 고려하여 두 그룹 집합의 결합을 제한한다. Fig. 4에 이들 3가지 규칙들의 예를 몇 가지 제시하였다.

3-2-1. Meta rule

이 부분에 저장되는 규칙들은 앞의 방법론에서 설명한 1단계에서 5단계를 마칠 때까지 각 단계에서 필요한 함수들을 실행시키고, 작업의 완결을 확인한 후 다음 단계의 과정을 처리하도록 하는 총괄적인 역할을 수행한다. 이 부분의 규칙들이 추론되는 과정에서 select rule과 combine rule들이 부분적으로 필요할 때마다 추론되어진다. 또 이 규칙들은 점진적으로 다음 group set collection을 형성해가는데 필요한 규칙들을 포함하고 있다. 예를 들면 group set collection(8)의 경우 group set collection(2)로부터 group set collection(4)를 만들고 두 개의 group set collection(4)로부터 group set collection(8)을 만든다. 마지막으로 group set collection(x_{carbon})의 완성을 확인하고, 이를 탄소와 수소 이외의 원소를 모은 초기 부 그룹집합 및 초기 할로젠 그룹집합과 결합하여 final group set collection을 만든다. 이로부터 분자식과 각 그룹집합을 비교하여 완전히 일치되는 그룹집합을 선택하고 전체 추론이 종결된다. 여기서 final group set collection은 입력된 분자식으로부터 얻은 정보에 의해 그룹집합의 가짓수를 최소화하여 얻어진 그룹집합들의 집합이므로, final group set collection에는 분자식을 정확히 표현하는 그룹집합 이외에도 같은 그룹 갯수를 갖는 그룹집합들이 포함되어 있다. 그러므로 각 원소의 갯수를 비교하여 정확히 입력 분자식을 표현하는 그룹집합을 결정하게 된다.

3-2-2. Select rule

각 group set collection이 가능한 최소의 그룹집합을 포함하도록 그룹집합을 여과하는 역할을 한다. 이 부분의 규칙들은 meta rule이 2단계의 과정을 처리하면서부터 이용된다. 2단계에서는 입력된 분자의 각 원소의 갯수와 구조형태를 이용하여 3 부류의 초기 그룹집합이 각기 최소한의 그룹으로 구성되도록 한다. 첫번째로 분자식에 포함되어 있는 원소 이외의 원소를 갖는 그룹은 초기 그룹집합의 형성에서 제외한다. 다음 단계로, Joback GCM에서는 그룹들이 환구조(ring system)와 비환구조(nonring system)로 나누어져 있으므로 입력된 데이터(ring구조의 유무를 의미함)에 따라서 선별적으로 그룹들을 선택한다. 그리고 분자식으로부터 그 분자가 가질 수 있는 단일결합, 이중결합, 삼중결합의 갯수를 예측할 수 있으므로 그 분자에 포함되지 않을 결합형태를 갖는

그룹은 초기 그룹집합에서 제외시킨다. 이 예측은 다음과 같은 방법으로 할 수 있다.

$$\begin{aligned} N_{ai} &= \sum (A_i - 2) * N_i + 2 \\ M_h &= N_{ai} - N_{hai} \\ M_{db} &= (N_{ai} - N_h - N_{hai}) / 2 \\ M_{tb} &= \text{truncate}[(N_{ai} - N_h - N_{hai}) / 4] \\ N_{db} &= M_{db} - 2 * N_{ib} \end{aligned}$$

여기서 N_h 와 N_{hai} 는 각기 입력된 분자식의 수소와 할로젠 원소의 갯수이다. 또, N_{ib} 는 입력 분자가 갖는 삼중결합의 갯수이며 N_{db} 는 입력된 분자가 갖는 이중결합의 갯수이다. N_{ai} 는 분자내에 하나의 결합을 갖고서 결합되는 그룹의 갯수를 나타내며, N_i 는 분자내의 주요 성분(탄소, 산소, 질소, 황)의 갯수이다. A_i 는 이들 주요 성분이 가질 수 있는 결합의 갯수를 나타낸다. 즉 탄소는 4, 산소는 2, 질소는 3, 그리고 황은 4개의 결합을 갖는다. 또 M_h , M_{db} , M_{tb} 는 각기 그 분자내에 포함될 수 있는 수소, 이중결합, 그리고 삼중결합의 이론적 최대수를 나타낸다. 여기서 N_h 와 M_h 가 같으면 이중결합이나 삼중결합은 포함될 수 없으므로 이들 결합을 포함하는 그룹은 초기 그룹집합의 형성에서 제외된다. 또 3단계와 4단계에서는 이론적 최대 이중결합의 갯수와 삼중결합 갯수를 이용하여 이들 보다 많은 수의 이중결합이나 삼중결합을 갖는 그룹집합은 여과시킨다. 마지막으로 5단계에서는 final group set collection에서 분자를 나타내는 최종 단계의 그룹집합으로서 타당한 것만을 선택한다. 여기에 관련된 규칙은 방법론에서 설명되었다.

3-2-3. Combine rule

두 개의 group set collection(x_a)와 group set collection(x_b)를 통해서 group set collection(x_c), ($x_c = x_a + x_b$)를 만들 때 각 group set collection으로부터 추출된 2개의 그룹집합의 결합 형태를 검토하여 같은 형태의 결합을 갖는 그룹집합끼리 결합을 이루도록 하거나, 분자를 표현하기 위해서 필요한 나머지 원소의 갯수를 결합 상대방 그룹집합의 각 원소의 갯수와 비교하여 상대방 그룹집합의 각 원소의 갯수가 나머지 원소의 갯수보다 작은 경우에 서로 결합시킨다. 그리하여 대상화합물을 만드는데 적합한 구조의 그룹집합들이 group set collection(x_c)에 포함되도록 한다. 이 규칙들은 방법론에 나타난 3단계에서 이용된다. 이와 같이 세부분의 지식베이스는 meta rule이 순환적으로 실행되면서 각 단계별로 필요한 select rule과 combine rule을 불러들여 실행하게 된다.

4. 결과 및 고찰

Fig. 5는 cyclopropane에 대한 추론 과정을 보여주는

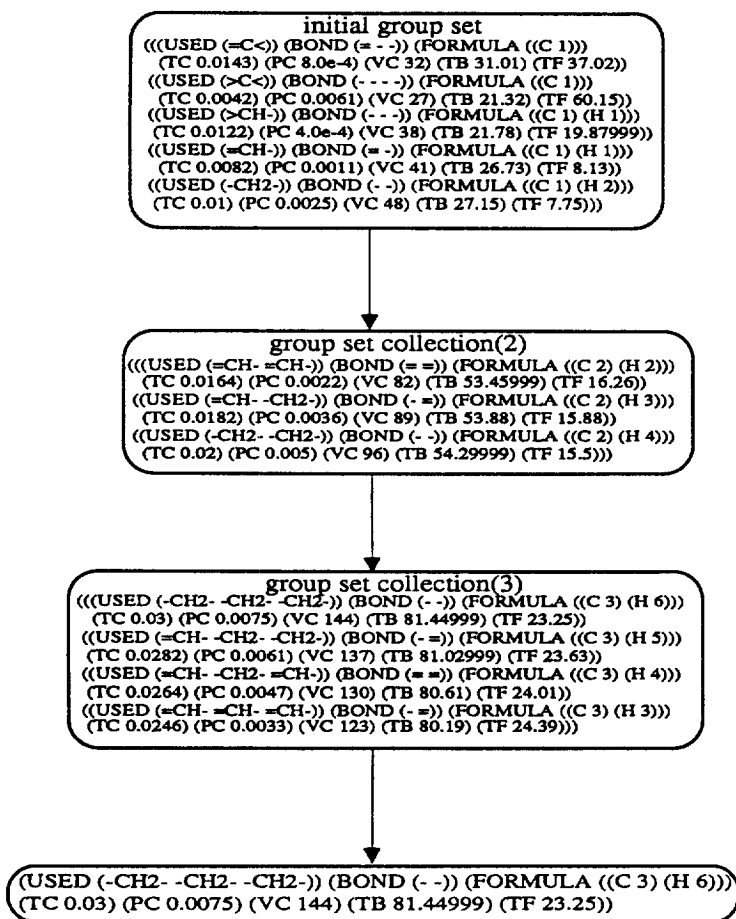


Fig. 5. Example of operations in the expert system.

예이다. 우선 주함수(main function)를 실행하면 사용자에게 분자식, ring구조의 존재 유무, 그리고 화학명을 입력하도록 차례로 묻는다. 이에 대해 (C 3 H 6), ring, cyclopropane을 입력하면 프로그램 내부적으로 Fig. 5에서와 같이 초기 그룹집합, group set collection(2), group set collection(3)을 만들고 최후에는 (-CH2- -CH2- -CH2-)를 입력 분자식에 해당하는 그룹집합으로 선택한다. 그 과정을 상세히 살펴보면 다음과 같다.

추론은 meta rule로부터 시작하여 입력데이터로부터 group set collection(x_{carbon})에서 x_{carbon} 을 3으로 정한다. 우선 초기 그룹결합을 만들기 위해서 select rule을 추론한다. 여기서는 분자의 원소가 탄소와 수소로만 이루어져 있으므로 이외의 원소를 갖는 그룹의 모임인 초기 부 그룹집합과 초기 할로젠 그룹집합은 형성되지 않는다. 그리고 구조는 환형체인 것만을 선택한다. 이로부터 Fig. 5에서 볼 수 있는 바와 같이 40개의 그룹들

로부터 5개의 group들로 압축된 초기 그룹집합을 형성한다. 다음은 meta rule로부터 $x_{carbon}=3$ 인 group set collection을 만들기 위해서는 group set collection(2)를 만들고 이것과 초기 그룹집합을 이용해서 group set collection(3)을 만든다는 전략에 따라서 먼저 group set collection(2)를 만든다. 이 때 combine rule은 초기 그룹집합에서 두 개씩 그룹을 선택하여 화학구조적 타당성 및 대상 화합물에 대한 적합성을 검토하여 타당한 경우에만 group set collection(2)에 첨가한다. 이로써 group set collection(2)는 3개의 화학적으로 타당한 그룹집합을 포함하게 된다. 다음은 초기 그룹집합과 group set collection(2)에서 각각 하나씩의 그룹집합을 선택하여 combine rule에 의해서 구조적으로 타당하다고 판정된 그룹집합들을 group set collection(3)에 첨가한다. Group set collection(3)에는 결과적으로 4개의 그룹집합이 포함된다. 마지막으로 select rule은 환구조를 이

```

What is the structure of chemical?
1. ring
2. nonring
Ans : 1
Enter the data in the following form: (C 6 H 6)
(C 6 H 6)
Enter The Name of The Chemical
BENZENE
The results of the analysis are:

(USED (=CH- =CH- =CH- =CH- =CH- =CH-))

(BOND (= -))

(FORMULA ((C 6) (H 6)))

(TC 0.0492)

(PC 0.0066)

(VC 246)

(TB 160.38)

(TF 48.78)

boiling_temperature      : 358.38      (K) (predicted)
critical_temperature     : 569.7095    (K)
critical_temperature     : 561.475     (K) (from data-base)
critical_pressure        : 47.69389    (bar)
critical_volume          : 263.5       (cm3)

*****ERRORS*****
Error of boiling_temperature      : 1.466589 (%)
Error of critical_temperature_from_calculated_Tb: 1.335749 (%)
Error of critical_temperature_from_dbase_Tb   : 0.128953 (%)
Error of critical_pressure        : 2.466471 (%)
Error of critical_volume          : 1.737452 (%)

```

Fig. 6. Example of input and output screen(The values for TC, PC, VC, TB and TF denote the Δ values in Joback GCM).

루는 그룹집합으로서의 타당성을 검사하고 이를 통해 걸러진 group set collection(3)의 각 그룹집합의 원소의 갯수와 분자식을 비교하여 분자식과 일치하는 그룹집합 (-CH₂- -CH₂- -CH₂-)을 최종적으로 선택하게 된다. Fig. 5에서 각 그룹집합의 bond는 결합에 쓰이고 난 나머지 여분의 결합을 나타내고 formula는 그룹집합의 각 원소의 갯수를 나타낸다.

Fig. 6은 전문가 시스템에 의한 입력과 출력의 화면을 나타낸 것이다. 분자의 구조, 분자식, 분자의 이름이 입력에 해당하는데 분자의 이름은 단순히 실험데이터를 검색함으로써 예측치와 실험치간의 오차를 계산하기 위해 필요한 것이다. 출력의 첫째줄에 전문가 시스템에 의하여 얻어진 작용기들이 나타나 있다.

Table 2에는 본 연구에서 개발된 전문가 시스템을 사용하여 작용기들의 조합을 구해 본 결과를 요약하였다. 이 Table의 첫째열은 참고문헌 13의 부록으로 수록된

물성치 데이터베이스에서 각 물질에 부여한 고유번호이며, 2-41열의 숫자는 대상 분자를 Joback GCM에서 사용되는 40개의 작용기로 조합할 때(Table 3 참조), 각 작용기가 몇 개씩 들어있는지를 나타내고 있다. 본 연구에서는 C1부터 C8까지 임의로 67종의 분자를 선택하였으며, 모든 경우에서 정확한 작용기들의 조합을 구할 수 있었다.

본 연구에서 개발된 전문가 시스템은 몇 가지 제약과 가지고 있다. 첫째, 적용대상이 제한적이다. 존재가능한 화합물들의 구조는 매우 다양한 형태를 가질 수 있으나, 본 전문가 시스템의 지식베이스는 Joback GCM의 작용기들로부터 일반적인 화학구조의 규칙성을 통하여 유추할 수 있는 화합물들에 국한되고 있다. 비환구조(nonring system)의 화합물에서는 -O-, -N=, -S= 그룹을 제외한 모든 그룹들을 포함한 물질들의 구조의 추론이 가능하다. 여기서 한 가지 제약점은 초기 부 그룹집합에 대한 결

[illegible]

리고 추가의 규칙들이 필요하다.

본 연구에서 개발된 전문가 시스템을 활용하면 많은 수의 분자들을 작용기들의 집합으로 자동적으로 쪼갤 수 있으므로, GCM의 활용이나 개발 등에 유용하게 사용될 수 있다.

감 사

본 연구를 지원하여 주신 산업과학기술연구소 및 공정산업의 지능자동화연구센터에 감사드립니다.

사용기호

- A_i : number of attachments in the main elements (C, O, N, S)
 M_{db} : maximum number of double bonds in a molecule(calculated from the number of main elements)
 M_h : maximum number of hydrogen in a molecule
 M_{tb} : maximum number of triple bonds in a molecule
 N_{db} : number of double bonds in a molecule
 N_h : number of hydrogens in a molecule
 N_{hal} : number of halogen elements in a molecule
 N_i : number of main elements in a molecule
 N_{tb} : number of triple bonds in a molecule

참고문헌

- 이진홍, 임재곤 : 산업과학기술연구소 보고서, 9173

- F-22, 포항(1989).
- Langmuir, I.: "Third Colloid Symposium Monograph", Chemical Catalog Co.(1925).
- Bondi, A.: "Physical Properties of Molecular Crystals, Liquids, and Glasses", John Wiley & Sons, New York(1968).
- Van Krevelen, D.W.: "Properties of Polymers", 2nd ed., Elsevier, Amsterdam(1976).
- Ambrose, D.: *NPL Rep. Chem.*, **92** (1978) corrected(1980).
- Ambrose, D.: *NPL Rep. Chem.*, **98** (1979).
- Joback, K. G. and Reid, R. C.: *Chem. Eng. Commun.*, **57**, 233(1987).
- Fedors, R. F.: *Chem. Eng. Commun.*, **16**, 149(1982).
- Thinh, T.-P., Duran, J.-L. and Ramalho, R. S.: *Ind. Eng. Chem. Process Design Develop.*, **10**, 576(1971).
- Thinh, T.-P. and Trong, T. K.: *Can. J. Chem. Eng.*, **54**, 344(1976).
- Kojima, K. and Tochigi, K.: "Prediction of Vapor-Liquid Equilibria by the ASOG Method", Kodansha, Tokyo(1979).
- Fredenslund, A., Gmehling, J. and Rasmussen, P.: "Vapor-Liquid Equilibria Using UNIFAC", Elsevier, Amsterdam(1977).
- Reid, R. C., Prausnitz, J. M. and Poling, B. E.: "The Properties of Gases & Liquids", 4th ed., McGraw-Hill, New York(1988).
- Firebaugh, M. W.: "Artificial Intelligence", Boyd & Fraser, Boston(1988).
- 정준영 : 석사논문, 포항공과대학, 포항(1993).