

마코브체인을 이용한 韓國 化學者의 公式 커뮤니케이션의 構造的 分析

A Structural Analysis of the Formal Communication
of Korean Chemists by Using Markov Chains

金 賢 姬*
(Kim, Hyun-Hee)

抄 錄

本 論文에서는 다음 두가지 가설들을 검증하기 위해서 1967~83 년 동안에 化學分野 學術誌들에 실린 3,815 편의 論文들을 分析하였다. 첫번째 가설은 마코브체인을 이용하여 특정한 주제 분야의 研究者들의 下位主題間 移動過程을 기술하고 豫測할 수 있다는 것이며, 두번째 假說은 마코브체인의 推移確率行列을 이용하여 해당 주제분야의 知的 構造를 多次元 空間上에 나타낼 수 있다는 것이다.

研究의 結果는 마코브체인이 韓國 化學者들의 7 개의 하위주제간의 이동과정을 연구하는데는 물론 化學分野의 知的構造를 2次元 지도상에 규명하는 데에도 應用될 수 있음을 보여주고 있다.

ABSTRACTS

The purpose of this study is to verify the following two hypotheses by using a test collection of 3,815 documents on the subject of chemistry. First hypothesis is that a Markov chain model can be used to describe and predict authors' movements among subareas of a discipline. Second hypothesis is that a transition matrix of the Markov chain can be applied to de-

* 明知大學校 圖書館學科.

scrib the intellectual structure of a discipline on the multidimensional space.

The results of this study have shown that the Markov chain is a good model to be used to study the movement of Korean chemists in 7 subtopics in chemistry and understand the intellectual structure of chemistry.

I. 序 論

1. 研究의 目的

個人이나 組織에 있어서 그것의 活動에 影響을 미치는 어떤 事件의 발생과 그 사건에 대비하여 준비하여야 하는 과정사이에 시간적 차이가 존재하는 경우가 많다. 이러한 事前的 準備로서 우리는 계획을 세워야 하고 계획을 세우기 위해서는 미래의 特定狀況에 대한 가정을 미리 설정해야 한다. 이렇게 미래에 대한 가정을 수립하는 작업이 바로 豫測活動이다.¹⁾ 도서관이나 온라인 檢索시스템 등 情報시스템에 있어서도 이러한 豫測이 전제되어야만 정보의 選擇·檢索·서비스 등의 諸般計劃이 가능하게 되므로 豫測의 正確性은 效率的인 情報시스템의 설계 및 운영에 매우 중요한 요소가 된다.

마코브체인은 自然科學에서는 물론 人文·社會科學에서도 널리 應用되는 豫測模型으로²⁾ 圖書館學·情報學分野에서는 학술 커뮤니케이션의 動的過程을 記述·豫測하는데 이용되어 情報의 選擇, 서지적 데이터베이스의 질적 여과 및 효율적인 情報서비스에 應用되어질 수 있다. 이 모형은 기호 논리학, 민속 음악학 등 특정한 주제분야의 하위주제간의 저자들의 이동과정을 분석해서 해당분야의 發展過程을 記述·豫測하는데 應用되었고,^{3) 4)} 또한 마코브체인은 經濟學, 社會學 등 13개의 주제분야로 구성된 社會科學分野의 文獻들의 引用文獻들을 13개의 주제

1) 尹錫喆, 「計量經營學」, (서울:經文社, 1989), pp.625~626.

2) 金基永·郭魯均, 「計量意思決定論-經營問題解決의 計量的 接近-」, (서울:法文社, 1989), p.477.

3) W. Goffman & K.S. Warren, *Scientific Information Systems and the Principle of Selectivity* (N.Y.: Praeger, 1980), pp.87~91.

4) M.L. Pao & L. McCreery, "Bibliometric Application of Markov Chains," *Information Processing & Management*, 22:1(1986):7~17.

별로 分析하여 社會科學分野의 發展過程과 科學·技術分野와 社會科學 分野의 文獻들의 參考文獻들을 두 분야별로 分析하여 科學의 발전과정을 記述·豫測하는데에도 應用되었다. 5)

이러한 몇가지 研究成果가 있기는 했어도 마코브체인이 學問發展過程을 기술·예측할 수 있다는 가설을 충분한 실험데이터를 가지고 검증한 연구가 거의 없었기 때문에 本稿에서는 이러한 가설을 충분한 데이터를 이용하여 검증해 보고, 아울러 마코브체인이 특정한 분야의 科學者들의 下位主題間 推移確率行列을 이용하여 同時引用法과 마찬가지로 學問의 知的構造를 多次元空間의 側面에서 分析할 수 있다는 가설을 검증하는데에 目的을 두고자 한다.

2. 研究의 方法

마코브체인이 學問發展過程을 기술·예측할 수 있다는 첫번째 가설을 검증하기 위한 실험 데이터는 學術院 發刊의 「學術總覽」(vol.33, 化學篇(IV))을 통해서 얻었다. 6) 이 抄錄誌에는 27개의 化學分野 學術誌에서 발췌한 1967~83년 동안 韓國 化學者 3,197名과 外國人 化學者 145명이 쓴 3,815種의 論文들이 7개의 下位主題別로 분류되어 있다. 먼저 韓國 化學者 3,197명 중에서 1967~73년 동안 2편 이상의 논문을 쓴 313명이 쓴 總論文數 968편을 分析하여 각 하위주제의 初期狀態確率, 推移確率, 固定確率을 구한 다음 고정확률에서 예측된 대로 化學者들이 각 하위주제에 분포되어 있는지를 확인하기 위해서 1967~73년 동안 2편 이상의 논문을 발표했으면서 1974~83년 동안 1편 이상의 논문을 발표한 290명의 저자가 1974~83년 동안에 쓴 總論文數 1,773편을 分析해 보았다. 7)

다음은 두번째 가설을 검증하기 위해서 1967~83년 동안 논문 4편 이상을 쓴 648명의 化學者들의 하위주제간의 推移過程을 分析해서 얻은 推移確率을 基礎資料로 해서 多次元蓄積技法(Multidimensional Scaling, MDS)을 이용하여 2

5) P. Zunde & V. Slamecka, "Predictive Model of Science Progress," *Information Storage & Retrieval*, 7(1971): 103~109.

6) 「學術總覽」, 第33輯, 化學篇(IV), (서울: 學術院, 1986).

7) 總論文數 968篇과 1,773篇은 한 論文에 共著者數가 2名 이상일 때는 같은 論文이라도 2篇 이상으로 取扱하여 集計한 統計이다.

次元의 공간에 하위주제를 나타내는 점들을 매핑하고 이들을 다시 群集分析(Cluster Analysis) 하였다. 多次元蓄積과 群集分析을 위해서 SYSTAT 패키지의 MDS 프로그램과 CLUSTER 프로그램을 이용하였다.

3. 假 說

本 研究에서 검증하고자 하는 假說은 다음과 같다.

- ① 마코브체인을 이용하여 특정한 主題分野 研究者들의 하위주제간의 移動過程을 기술하고 예측할 수 있다.
- ② 마코브체인의 특정한 主題分野 研究者들의 下位主題間의 推移確率行列을 이용하여 해당 주제분야의 知的構造를 多次元 空間上에 나타낼 수 있다.

II. 理論的 背景

1. 確 率 過 程

自然·社會現象이나 人爲的 실험을 통하여 결과(outcome)가 얻어지게 되는 과정은 두 가지 경우에 귀착된다. 첫째는 매회 관찰 또는 측정에서 일정한 결과만이 얻어지게 되는 決定論的 現象(deterministic phenomenon) 이고, 둘째는 주사위(die)를 던지는 게임에 있어서와 같이 매회 얻어지게 되는 결과들이 일정치 않을 경우이다. 이런 현상을 非決定論的 現象 또는 偶然現象(random phenomenon) 이라고 부른다.⁸⁾

前者의 경우를 해석하기 위한 理論模型으로서는 함수, 미분방정식 등이 있으며 後者의 경우에는 결과들이 얻어지게 될 경향성 내지는 法則性을 주는 理論模型인 確率, 確率過程 등이 있다.⁹⁾

確率過程이란 일반적으로 시간 $t(a \leq t \leq b)$ 를 보조변수로 갖는 確率變數 $x(t)$ 의 집합 $\{x(t), a \leq t \leq b\}$ 를 말하며 이산·시 확률과정과 연속·시 확률과정으로 구분된다. 確率過程 $\{x(t)\}$ 에서 補助變數 t 가 정수값 $t = 0, \pm 1,$

8) 구자홍, 「確率論」, (서울: 민음사, 1988), pp.11 ~ 12.

9) 上揭書.

士 2, …… 만을 취하며 변화할 때 이것을 이산·시 확률과정이라 하며, t 가 連續적으로 변할 때 이것을 연속·시 확률過程이라 한다.^{10) 11)}

2. 마코브체인模型

(1) 마코브체인의 特性

마코브체인은 $n + 1$ 번째단계 狀態確率에 n 번째단계 狀態確率의 영향만 받고 그 이전단계의 狀態確率의 영향을 받지 않는 이산·시 확률過程의 특별한 경우이다.¹²⁾ 마코브체인의 特性을 설명하기 위해서 研究機關間에 科學者들의 이동과정을 예로 들어 설명해 보기로 한다. 현재 세 개의 癌 研究機關 a, b, c 가 있고, 이 분야에 종사하는 네명의 科學者 A, B, C, D 가 있다고 가정해 보자. 또한 각 科學者는 반드시 세 研究機關中 한 기관에 고용되며, 年初마다 각 科學者는 자유롭게 현재 속해 있는 기관에서 다른 기관으로 옮겨가거나 혹은 그대로 같은 機關에 머무를 수 있다고 가정해 보자. 여기서 각 科學者가 選擇할 수 있는 기관을 상태라고 하고 상태의 집합 $\{ a, b, c \}$ 를 狀態空間이라고 할 수 있다.¹³⁾

예로 든 이 특정경우를 狀態空間이 유한집합인 유한마코브체인으로 기술하기 위해서는 다음과 같은 특성을 가지고 있어야 한다.

- ① 이산·시 확률過程이어야 한다.
- ② 狀態空間이 유한집합이어야 한다.
- ③ 각 상태의 예측된 확률은 바로 이전단계의 상태에만 의존해야 한다.

이 경우 科學者들의 機關(狀態)間的 이동이 제한된 간격, 즉 1년에 한번씩 발생하기 때문에 이산·시 확률過程이며 상태공간이 $\{ a, b, c \}$ 로 유한집합이며, 또한 癌 研究처럼 급격하게 발전하는 분야에서 科學者들의 현재의 성과가 研究機關에 僱傭되는데 유일한 고려의 대상으로 생각될 수 있기 때문에, 각 科學者의 다음 단계의 상태는 현 상태에 의해서 결정된다고 할 수 있다. 4년동안 科學者들의 研究機關交替의 推移를 조사한 다음 그 결과를 <表 1>로 표시하고 이를 다시 3次正方向行列로 표시하여 <表 2>와 같이 정리하였다.

10) 鄭英鎮, 「數理統計學: 確率과 統計」, (서울: 喜重堂, 1987), pp.323~340

11) 朴乙龍 등, 「콘사이스 數學辭典」, (서울: 創元出版社, 1986), pp.773~774.

12) 최희찬, 이대영, 신철재, 「理工系 學生을 위한 應用數學」, (서울: 韓新文化社, 1981), pp.509~515.

13) M.L.Pao & L.McCreery, *op.cit.*

< 表 1 >

4년동안의 科學者들의 移動過程

a	b	c
A ₁	C ₁	B ₁
C ₂	D ₁	D ₂
A ₂	B ₂	D ₃
A ₃	B ₃	D ₄
C ₃	B ₄	A ₄ *
C ₄		

註 : *는 表에서 研究者 A의 경우, 機關 a에서 3년동안 (A₁, A₂, A₃) 勤務한 後 4년째 (A₄)에 機關 c로 옮겼음을 알 수 있다.

< 表 2 >

科學者 研究機關交替의 推移

~에서 \ ~로	a	b	c	合 計
a	4	0	1	5
b	1	2	1	4
c	0	1	2	3
合 計	5	3	4	12

(2) 推移確率과 初期狀態確率

위의 예를 마코브체인으로 完全히 記述하기 위해서는 앞에서 논의한 세가지 特性을 갖는것 외에 먼저 각 狀態의 初期確率과 推移確率을 알아야 한다.¹⁴⁾

먼저 각 狀態의 初期確率은 첫 해에 각 研究機關에 의해 고용된 科學者數를 總 科學者數로 나눈 값으로 얻어지는데, < 表 1 >에서 보면 첫 해에는 機關 a에 科學者 A가, 機關 b에 科學者 C, D가, 그리고 機關 c에 科學者 B가 雇傭되어 있었기 때문에 a, b, c에 대한 初期狀態確率は < 表 3 >과 같다.

다음으로 推移條件確率は 일반적으로 P_{ij} 로 나타내는데 이는 현재 상태 i 에서 다음 상태 j 로 진행될 확률을 뜻한다. < 表 2 >를 推移確率行列로 변환시키면 < 表 4 >와 같다.

< 表 4 >에 표시된 推移行列의 제1행 P_{aa} , P_{ab} , P_{ac} 는 a라는 상태에서 각각 a, b, c라는 상태로 옮겨가는 確率을 보여주고 있다. 狀態 a에서 한번 이

14) W.Goffman & K.S. Warren, *op. cit.*

〈表 3〉 初期狀態確率

a	b	c
0.25	0.50	0.25

〈表 4〉 推移確率行列 P

~에서 ~로	a	b	c
	0.80	0	0.20
a	0.25	0.50	0.25
b	0	0.33	0.67
c			

동할 때 a, b, c 중 어느 한 상태로 옮겨가게 되므로 반드시 $P_{aa} + P_{ab} + P_{ac} = 1$ 이 성립된다.¹⁵⁾ 이 경우에 P_{aa} 의 條件確率は 狀態 a 에서 狀態 a 로 이동할 확률이기 때문에 機關 a 에 계속 2년동안 研究者가 머무는 경우를 계산함으로써 얻어질 수 있다. 이 경우는 $A_1 \rightarrow A_2, A_2 \rightarrow A_3, C_2 \rightarrow C_3, C_3 \rightarrow C_4$ 로 총 4번이며 機關 a 에서 b 로 옮겨간 사람은 아무도 없고 機關 a 에서 c 로 옮긴 경우는 $A_3 \rightarrow A_4$ 로 단 한번 있다. 따라서 상태 a 에서 a, b, c 로의 推移條件確率は $4/5 (= 0.80), 0, 1/5 (= 0.20)$ 이 된다(〈表 4〉의 제1행 참조).

3. 豫 測

마코브체인의 주된 目的은 시스템의 미래의 행위를 예측하는 것이며, 그 결과는 推移確률과 현 단계의 狀態確률이 결합되어 구해지는 다음 단계의 狀態確률과 推移確률을 n 乘(n 제곱乘)하여 구할 수 있는 시스템의 장기적인 安全狀態의 確률을 나타내고 있다.^{16) 17) 18)}

1) 未來의 狀態確率

위의 예에서 다음 기간의 각 研究機關에 雇傭될 科學者들의 比率 즉 狀態確률을 예측하고자 하는데 이 예측은 初期狀態確률과 推移確률을 곱하여 구할 수 있다.

15) 최희찬·이대영·신철재, 前揭書.

16) 尹錫喆, 前揭書, pp.511~530.

17) 金基永·郭魯均, 前揭書, pp.477~504.

18) 郭魯均·崔泰成, 計量經營學 OR/MS (서울: 茶山出版社, 1988), pp.605~628.

$$\begin{array}{ccc}
 & a & b & c \\
 \begin{array}{ccc}
 a & b & c \\
 [0.25 & 0.5 & 0.25]
 \end{array}
 \times
 \begin{array}{ccc}
 a & b & c \\
 \left[\begin{array}{ccc}
 0.80 & 0.00 & 0.20 \\
 0.25 & 0.50 & 0.25 \\
 0.00 & 0.33 & 0.67
 \end{array} \right]
 \end{array} \\
 \\
 \begin{array}{ccc}
 a & b & c \\
 = [0.33 & 0.33 & 0.34]
 \end{array}
 \end{array}
 \quad 19)$$

다음 기간의 狀態確率¹⁹⁾이 계산된 후 그 다음 기간의 狀態確率의 計算도 可能하다. 두 기간후의 狀態確率은 다음 기간의 狀態確率과 推移確率을 곱하면 구해진다.

$$\begin{array}{ccc}
 [0.33 & 0.33 & 0.34]
 \end{array}
 \times
 \begin{array}{ccc}
 \left[\begin{array}{ccc}
 0.80 & 0.00 & 0.20 \\
 0.25 & 0.50 & 0.25 \\
 0.00 & 0.33 & 0.67
 \end{array} \right] \\
 \\
 = [0.34 & 0.28 & 0.38]
 \end{array}$$

위의 計算方式은 2기간 후의 科學者 比率을 계산하는데 1기간후의 狀態確率을 사용하고 있는데, 이는 初期狀態確率을 사용하면서 推移確率을 自乘한 결과와 같다. 즉,

$$\begin{array}{ccc}
 [0.25 & 0.5 & 0.25]
 \end{array}
 \times
 \begin{array}{ccc}
 \left[\begin{array}{ccc}
 0.80 & 0.00 & 0.20 \\
 0.25 & 0.50 & 0.25 \\
 0.00 & 0.33 & 0.67
 \end{array} \right]^2 \\
 \\
 = [0.34 & 0.28 & 0.38]
 \end{array}$$

앞서와 같이 n 期間 後 마코브過程을 다음과 같이 나타낼 수 있다.

$$\begin{array}{ccc}
 [0.25 & 0.5 & 0.25]
 \end{array}
 \times
 \begin{array}{ccc}
 \left[\begin{array}{ccc}
 0.80 & 0.00 & 0.20 \\
 0.25 & 0.50 & 0.25 \\
 0.00 & 0.33 & 0.67
 \end{array} \right]^n
 \end{array}$$

그러나 實際的으로 3기간 후의 마코브過程은 豫測分析에 별로 사용되지 않는다. 왜냐하면 n 이 增加함에 따라서 推移確率의 停滯的 性格도 변화하여 새로운

19) $a = 0.80 \times 0.25 + 0.25 \times 0.5 + 0 \times 0.25 = 0.33$
 $b = 0 \times 0.25 + 0.5 \times 0.5 + 0.33 \times 0.25 = 0.33$
 $c = 1 - 0.33 - 0.33 = 0.34$

推移確率²⁰⁾이 필요해지기 때문이다.

2) 固定確率

마코브체인理論에 의하면 만약 推移確率 P 가 定規 마코브체인²¹⁾이면 여러 단계를 거친 후 각 행에 있는 確率값들의 분포가 똑같은 固定確率行列 L 을 얻게되는데, 일단 固定確率行列 L 에 이르면 行列의 確率값들은 시간에 따라 더 이상 변화하지 않는다.²²⁾ 이 경우는 2⁵ 단계를 거친 후 固定確率行列 L 에 이르게 된다. <表 5>와 <表 6>은 각각 固定確率 L 과 각 狀態의 再現期間 (recurrent period)을 나타내고 있다.

固定行列 L 은 몇 년 후에 科學者들의 33%, 27%, 40%가 기관 a, b, c에서 각각 발견될 수 있음을 예측하고 있다. 初期狀態確率과 固定確率을 비교하면 機關 c로 科學者들이 몰려올 것을 예측할 수 있는데, 이는 機關 c의 固定確率²³⁾이 40%로 初期狀態確率 25%보다 훨씬 높기 때문이다.

다음은 再現期間을 설명하면 科學者가 研究機關 a에서 다른 研究機關으로 이동한 후 원상태 a로 돌아오는 것을 a의 再現期間이라 하는데 b와 c의 再現期間도 같은 맥락에서 설명될 수 있다. 再現期間의 계산은 固定確率의 逆數이다.²³⁾ <表 6>을 설명하면 研究機關 a에 雇傭된 科學者가 研究機關을 바꿔서 研究機關 b와 c에 머물다가 다시 研究機關 a로 돌아올 때까지의 平均期間이 3.0만큼 소요된다는 것이다. b와 c의 再現期間 3.7과 2.5도 같은 方法으로 설명이 可能하다.

<表 5> 固定確率行列 L

a	b	c
0.33	0.27	0.40

<表 6> 再現期間

a	b	c
3.0	3.7	2.5

20) 金基永·郭魯均, 前掲書, p.483.

21) 定規 마코브체인에서는 각 狀態가 자기 다른 狀態로부터 接近되어질 수 있다. 따라서, 推移行列의 대각원소 P_{ii} 가 만약 1이라면 ($P_{ii}=1$) 自己狀態에서 이탈하지 못하는 경우로 定規 마코브체인이 아니라 하겠다.

22) 尹錫喆, 前掲書.

23) W. Goffman & K. S. Warren, op. cit.

Ⅲ. 實 驗

1. 第 1 實 驗

특정한 주제분야의 研究者들의 下位主題間의 移動過程을 마코브체인으로 기술·예측할 수 있다는 첫번째 假說을 검증하기 위한 實驗데이터는 學術院發刊의 「學術總覽」(第 33 輯)을 통해서 얻었다. 이 抄錄誌는 1967~83년 동안 韓國化學者 3,197명과 外國人 化學者 145명이 쓴 3,815種의 論文들이 7개의 下位主題別로 분류되고 있다.

먼저 이 抄錄誌에서 1967~73년 동안 2편 이상의 論文을 쓴 313명이 저술한 總論文數 968편을 分析對象으로 하였다. 968편의 논문들을 저자별로 분류하여 각 저자가 쓴 論文들을 데이터베이스 管理시스템인 dBASEⅢ Plus를 이용하여 출판연도를 1차 정렬키로, 下位主題코드를 2차 정렬키로 한 후 올림차순으로 정렬하였다. 정렬결과를 分析해 보니 같은 해에 相異한 下位主題코드를 갖는 論文을 2편 이상 발행한 저자들이 상당수 있었는데, 이런 경우에 마코브체인의 特性에 따라서 이들 論文의 發行時期가 각기 달라야 하는데 이를 위해서 論文이 게재된 學術誌의 發行月일을 기준으로 하여 論文의 出版時期를 결정하였으며 우연히도 學術誌의 發行月日마저 같은 경우는 出版時期를 임의로 통제해 주었다.

위의 方法에 따라서 313명의 각 化學者가 쓴 論文들을 出版時期에 따라서 정렬한 후 워드프로세서인 Word Perfect에 각 저자의 코드, 각 저자가 쓴 論文數, 각 저자가 쓴 올림차순으로 정렬된 論文들의 하위주제코드와 發行年度를 入力하여 313개의 레코드로 구성된 DOS 텍스트 화일로 構築한 후 GW-BASIC으로 작성한 프로그램에서 이 DOS 텍스트 화일을 入力파일(input file)로 하여 각 下位主題의 初期狀態確率, 推移確率, 固定確率을 구하였다.

각 하위주제의 初期狀態確率은 313명의 각 저자가 맨 처음으로 쓴 313편의 論文들을 7개의 下位主題別로 집계한 결과 論文數는 각 下位主題코드 A,B,C,D,E, F,G에 대하여 60, 27, 38, 37, 57, 21, 73이다. 각 下位主題의 初期狀態確率은 <表 7>에 나타나 있다.

다음은 1967~73년 동안 313명의 化學者들의 하위주제간의 推移의 數를 계산하여 <表 8>에 표시하고 이를 기초로하여 推移行列 P인 <表 9>를 만들

었다. 推移行列 P는 2⁷ 번째 단계에서 固定確率分布에 수렴한다. <表 10>은 固定確率行列을 나타낸 것이다.

<表 7> 初期狀態確率(1967~73)

A : 物理化學 = 0.1917	E : 生 化 學 = 0.1812
B : 無機化學 = 0.0863	F : 高分子化學 = 0.0671
C : 分析化學 = 0.1214	G : 工業化學 = 0.2332
D : 有機化學 = 0.1182	

<表 8> 化學者들의 下位主題間的 推移의 數 (1963~73)

~에서 / ~로	A	B	C	D	E	F	G	合 計
A	162	13	7	11	2	1	30	226
B	13	38	7	5	1	2	7	73
C	6	7	52	2	18	1	10	96
D	14	3	0	72	12	3	4	108
E	3	1	10	7	138	0	4	163
F	1	1	1	2	2	69	2	78
G	30	5	7	2	2	2	176	224
合 計	229	68	84	101	175	78	233	968

<表 9> 推移確率行列 P (1967~73)

~에서 / ~로	A	B	C	D	E	F	G	合 計
A	.7168	.0575	.0310	.0487	.0088	.0044	.1327	1.0000
B	.1781	.5205	.0959	.0685	.0137	.0274	.0959	1.0000
C	.0625	.0729	.5417	.0208	.1875	.0104	.1042	1.0000
D	.1296	.0278	.0000	.6667	.1111	.0278	.0370	1.0000
E	.0184	.0061	.0613	.0429	.8466	.0000	.0245	1.0000
F	.0128	.0128	.0128	.0256	.0256	.8846	.0256	1.0000
G	.1339	.0223	.0313	.0089	.0089	.0089	.7857	1.0000

< 表 10 >

固定確率行列 L (1967~73)

A	B	C	D	E	F	G
.2359	.0617	.0759	.0908	.2048	.0724	.2585

< 表 11 >

初期狀態確率과 固定確率의 比較 (1967~73)

下位主題	初期狀態確率	固定確率	變化率(%)
物理化學	0.1917	0.2359	+123
無機化學	0.0863	0.0617	-29
分析化學	0.1214	0.0759	-37
有機化學	0.1182	0.0908	-23
生化學	0.1812	0.2048	+113
高分子化學	0.0671	0.0724	+108
工業化學	0.2332	0.2585	+111

< 表 12 >

下位主題의 再現期間 (1967~73)

A	B	C	D	E	F	G
4.24	16.21	13.18	11.01	4.88	13.81	3.87

< 表 11 >에서 각 下位主題의 初期狀態確率과 固定確率을 比較해 본 결과 物理·生·高分子·工業化學 分野의 固定確率이 初期狀態確率보다 높게 나타났으며 無機·分析·有機化學 分野의 固定確率은 初期狀態確率보다 낮게 나타났는데 이는 韓國 化學者들의 物理·生·高分子·工業化學 分野에로의 관심이 점차 증가하고 있음을 보여주고 있다. 각 하위주제의 平均 再現期間이 < 表 12 >에 나와 있다. 再現期間이 가장 짧은 工業化學의 경우 工業化學에 관한 논문을 발표한 化學者가 일단 工業化學 分野에서 떠났다가 다시 工業化學 分野로 돌아올 때까지 平均 3.87 편의 논문을 발표해야 한다는 것을 의미한다. 이 데이터에서는 313 명의 化學者들이 출판한 平均 論文數가 3.09 이기 때문에 平均으로 化學者들은 일단 벗어난 하위주제로 다시 돌아가지 않음을 알 수 있다.

다음은 固定行列 L에 의해서 예측된 化學者들의 確率分布가 타당성이 있는지를 검증하기 위해서 1967~73년까지 2편 이상의 논문을 썼으면서 동시에 1974~83년까지 1편 이상의 논문을 쓴 290명의 저자들이 저술한 總論文數 1,773편

〈表 13〉 5년單位 觀測確率

下位主題	物理化學	無機化學	分析化學	有機化學	生 化 學	高分子化學	工業化學
觀測確率 (1974~78)	0.2450	0.0597	0.0721	0.0908	0.2674	0.0560	0.2090
觀測確率 (1979~83)	0.2012	0.0464	0.1115	0.1032	0.2074	0.1115	0.2188

〈表 14〉 10년單位 觀測確率

下位主題	物理化學	無機化學	分析化學	有機化學	生 化 學	高分子化學	工業化學
觀測確率 (1974~83)	0.2211	0.0525	0.0936	0.0976	0.2346	0.0863	0.2143

〈表 15〉 初期狀態確率, 固定確率, 觀測確率과의 比較

下位主題	物理化學	無機化學	分析化學	有機化學	生 化 學	高分子化學	工業化學
初期狀態 確 率	0.1917	0.0863	0.1214	0.1182	0.1821	0.0671	0.2332
豫測確率 (固定確率)	0.2359	0.0617	0.0759	0.0908	0.2048	0.0724	0.2585
觀測確率 (1975~83)	0.2211	0.0525	0.0936	0.0976	0.2346	0.0863	0.2143

을 7 개의 하위주제별로 5년 단위와 10년 단위로 각각 分析하였다. 〈表 13〉은 5년 단위 觀測確率이고 〈表 14〉는 10년 단위 觀測確率이며, 〈表 15〉는 각 하위주제의 初期狀態確率, 豫測確率(固定確率), 觀測確率을 비교한 것이다.

실제 데이터를 分析한 결과 예측에서 처럼 無機·分析·有機化學을 연구하는 化學者들의 비율이 감소하고 있고, 物理·生·高分子化學 分野의 化學者들의 비율이 증가하고 있으나 工業化學만은 研究者들의 비율이 조금 증가하리라는 예측과는 달리 약간 감소하고 있는 경향을 보이고 있다. 이와 같이 7년간의 化學者들의 하위주제간의 推移過程을 기초로하여 구한 각 下位主題의 固定確率은 그 이후 10년 동안의 化學者들의 각 下位主題에 대한 비율을 대체적으로 예측해 주고 있다고 할 수 있겠다.

2. 第 2 實 驗

두번째 實驗에서는 1983년 이후의 각 下位主題에 대한 化學者들의 비율을 예측하기 위해서 1967~83년 동안 논문 4편 이상을 쓴 648명의 저자가 쓴 總論文數 4,598편을 分析하여 각 下位主題의 初期狀態確率, 推移確率, 固定確率을 구하였다. <表 16>은 初期狀態確率과 固定確率과를 比較·分析해 놓은 데이터이다.

<表 16>과 1976~73년 동안의 初期狀態確率과 固定確率을 比較·분석한 <表 11>을 比較해 보면 <表 11>에서 가장 큰 비율로 증가하는 분야가 物理化學이고, 가장 큰 비율로 감소하는 분야가 分析化學으로 나타났으며, <表 16>에서는 각각 高分子化學, 無機化學이 여기에 해당된다. 이 두 表를 比較해 본 결과 가장 두드러진 차이는 <表 11>에서는 生化學의 固定確率이 初期狀態確率에 비해 +113%의 증가를 보이고 있는데 반해서 <表 16>에서는 生化學의 固定確率이 初期狀態確率과 比較해서 -1%의 減少現象을 보이고 있는 점이다. 따라서 1983년 이후에는 無機·分析·有機·生化學 分野에 대한 化學者들의 비율이 감소현상을 보일 것이며 化學者들의 관심이 高分子·物理·工業化學 分野로 쏠릴 것으로 예측할 수 있다.

<表 17>은 각 下位主題의 再現期間을 나타낸 것으로 1976~83년 동안 4편

<表 16> 初期狀態確率과 固定確率의 比較(1967~83)

下 位 主 題	初期狀態確率	固 定 確 率	變 化 率 (%)
物 理 化 學	0.1744	0.1925	+110
無 機 化 學	0.0802	0.0513	- 36
分 析 化 學	0.1034	0.0811	- 21
有 機 化 學	0.1019	0.0916	- 10
生 化 學	0.2562	0.2541	- 1
高 分 子 化 學	0.0679	0.1064	+157
工 業 化 學	0.2160	0.2229	+103

<表 17> 下位主題의 再現期間(1967~83)

下位主題	物理化學	無機化學	分析化學	有機化學	生化學	高分子化學	工業化學
再現期間	5.19	19.49	12.33	10.92	3.94	9.40	4.49

이상을 쓴 648 명의 化學者들의 平均 論文數가 7.1 편이기 때문에 재현기간이 平均 論文數보다 적은 物理·生·工業化學 分野에서는 平均적으로 일단 化學者들이 이 분야를 떠난 후 다시 해당분야로 돌아온다고 할 수 있고 그 외의 下位主題 分野들에서는 平均적으로 일단 벗어난 下位主題로 다시 돌아가지 않음을 알 수 있다.

3. 第 3 實 驗

마코브체인의 推移行列을 기초 데이터로 하여 특정한 主題分野의 知的構造를 多次元空間上에 나타낼 수 있다는 假說을 검증하기 위해서 먼저 1967~83년 동안 4 편 이상을 쓴 648 명의 化學者들이 쓴 總論文數 4,598 편을 분석하여 구한 <表 18>의 推移行列을 기초 데이터로 하여 多次元蓄積技法을 이용하여 2次元의 공간에 下位主題를 나타내는 점들을 매핑하였다.

多次元蓄積技法이란 두 사물간의 類似性 혹은 相異性에 따라 각각의 사물을 空間上에 점으로 나타냄으로써 데이터內에 숨겨진 구조를 드러내거나 데이터를 보다 이해하기 쉽게 해 주는 技術로, 두 사물간의 類似度가 낮을수록 혹은 相異度가 높을수록 두 점은 空間上에서 멀리 떨어져서 배열된다²⁴⁾ 여기서는 下位主題間의 類似度를 下位主題間의 研究者들의 推移確率을 이용하여 측정했는데, 즉 C 라는 下位主題에서 E 라는 下位主題로의 推移確率 P_{CE} 가 P_{CA} 보다 크다면 하위주제 C 와 E 와의 類似度가 C 와 A 와의 類似度보다 더 높다고 가정하였다. 推移行列 P <表 18>를 바로 多次元蓄積의 入力物로 이용하는데 몇가지 문제점이 있어서 이 행렬을 다음과 같이 수정하였다. 즉 推移確率行列 P 가 非對稱行列이기 때문에 P_{ij} 와 P_{ji} 의 원소를 합하여 $P_{ij}(=P_{ji})$ 의 원소로 처리하고 推移行列 P 의 대각원소(P_{ii})를 모두 1로 수정하여 推移行列 P' <表 19>를 만들었다. <圖 1>은 <表 19>인 推移行列 P' 를 入力物로 하여 SYSTAT 패키지의 多次元蓄積 프로그램(MDS)을 이용하여 出力物로 나온 하위주제 지도이다. 다음은 <圖 1>의 下位主題 지도에 표시된 7개의 下位主題들을 동질적인 집단으로 묶어주기 위해서 群集分析을 하였다. 群集分析이란 對象들이 지니고 있

24) H.D. White & B.C. Griffith, "Author Cocitation: A Literature Measure of Intellectual Structure," *JASIS*, 32:3(1981): 163~171.

〈表 18〉

推移確率行列 P (1967~83)

~에서 ~로	A	B	C	D	E	F	G	合計
A	.7152	.0507	.0378	.0485	.0313	.0162	.1003	1.0000
B	.1835	.5000	.0935	.0504	.0252	.0180	.1295	1.0000
C	.0934	.0455	.5530	.0202	.1995	.0101	.0783	1.0000
D	.0961	.0297	.0229	.6110	.1396	.0572	.0435	1.0000
E	.0285	.0042	.0603	.0502	.8192	.0092	.0285	1.0000
F	.0367	.0131	.0105	.0577	.0289	.7979	.0551	1.0000
G	.0803	.0315	.0254	.0142	.0295	.0407	.7785	1.0000

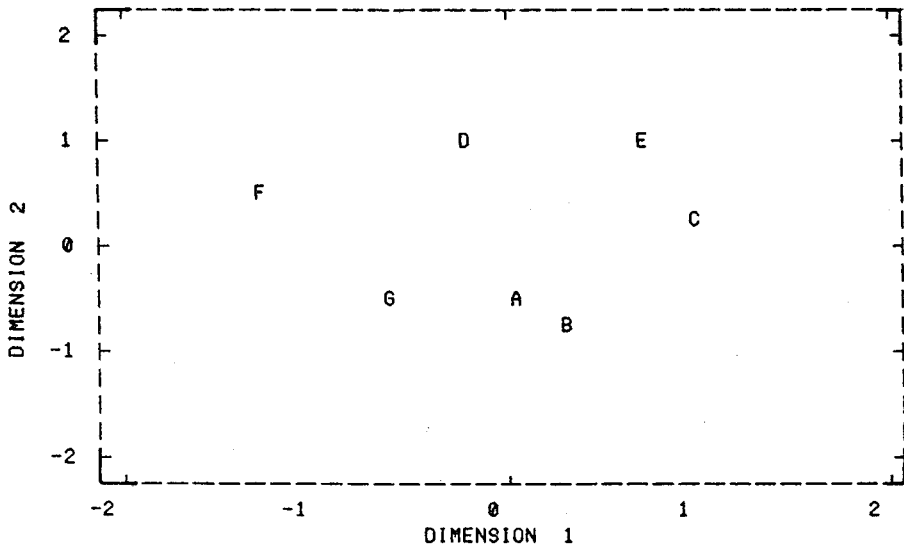
〈表 19〉

推移確率行列 P'

~에서 ~로	A	B	C	D	E	F	C
A	1.0000						
B	.2342	1.0000					
C	.1321	.1390	1.0000				
D	.1446	.0801	.0431	1.0000			
E	.0598	.0294	.2598	.1898	1.0000		
F	.0529	.0311	.0206	.1149	.0381	1.0000	
G	.1806	.1610	.1037	.0577	.0580	.0958	1.0000

〈圖 1〉

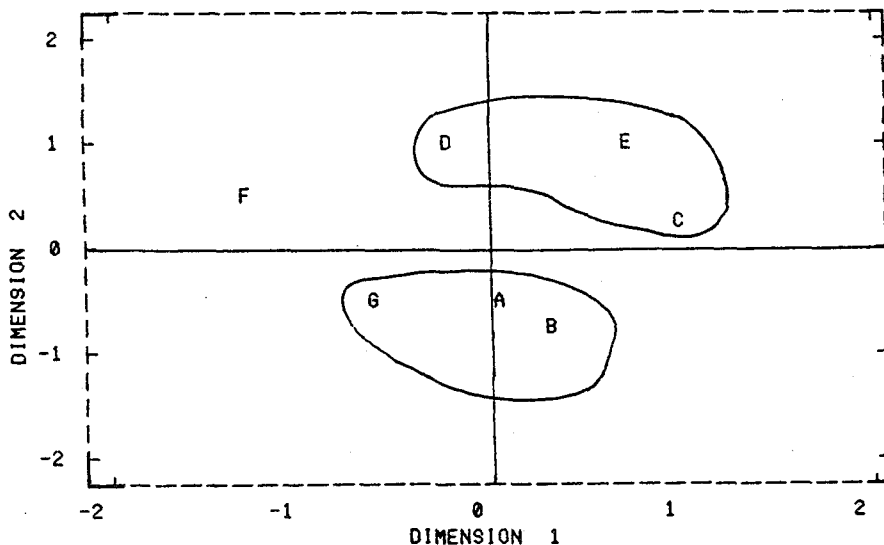
下位主題 地圖



는 다양한 특성의 類似性을 바탕으로 同質的인 집단으로 묶어주는 方法으로 同一集團內에 있는 대상들의 공통된 특성들을 조사하기 위한 目的으로 이용된다.²⁵⁾ <圖 2>는 <表 19>인 推移行列 P' 를 入力物로 하여 SYSTAT 패키지의 群集分析 프로그램 (CLUSTER)을 이용하여 出力物로 나온 群集分析에 의한 下位主題 地圖이다.

<圖 1>에 나타난 각 下位主題의 위치를 살펴보면 하위주제 지도 전체에 광범위하게 분포되어 있어서 각 하위주제가 대체로 獨立된 領域임을 나타내 보이고 있는데 그 중에서도 物理化學과 無機化學과의 관계가 비교적 密接함을 알 수 있다. 群集分析 결과 <圖 2>에서는 두 개의 下位主題 群集이 형성되었는데 有機化學, 生化學, 分析化學이 모여서 하나의 群集을 이루고 工業化學, 物理化學, 無機化學이 모여서 또 다른 군집을 이루었는데 高分子化學은 어느 군집에도 속하지 않고 따로 떨어져 있다. 각 群集內의 중심 하위주제는 生化學과 物理化學이다. 지도의 중심에 위치한 物理化學分野는 <表 19>에 의해서 알 수 있듯이 6개의 他下位主題들에서 物理化學에로의 推移條件確率의 합이 가장 큰 0.8042로서 他下位主題들과의 관계가 가장 密接한 하위주제이다. 그 어느 군집에도 속하지 않고 딱 떨어져 고립된 모습을 보이는 高分子化學分野는 6개의 他下位主

<圖 2> 群集分析에 의한 下位主題 地圖



25) K.W.McCain, "Longitudinal Author Cocitation Mapping: The Changing Structure of Macroeconomics," *JASIS*, 35; 6(1984): 351~359.

題들에서 高分子化學에로의 推移條件確率의 합이 가장 적은 0.3534로서 他下位主題들과의 관계가 가장 미미한 편이다.

IV. 結 論

本論文은 시스템의 미래의 행위를 예측하는 技法中 하나인 마코브체인 模型을 이용하여 學問의 發展過程을 기술·예측할 수 있다는 첫번째 假說과 또한 마코브체인의 推移確率行列을 이용하여 學問分野의 知的構造를 多次元空間上에 규명할 수 있다는 두번째 假說을 검증하기 위해서 化學分野 學術誌들에 실린 3,815 편의 論文을 이용하여 실험을 하였다.

第1實驗은 첫번째 假說을 검증하기 위해서 「學術總覽」(第33輯)에서 1967~73년 동안 2편 이상의 論文을 쓴 313명의 化學者들이 저술한 總論文數 968편을 분석하여 각 하위주제의 初期狀態確率, 推移確率 및 固定確率을 구하였다. 固定確率에 의해서 예측된 각 下位主題에 대한 研究者들의 비율이 타당성이 있는지를 검증하기 위해서 313명 중에서 1974~83년 동안 1편 이상의 논문을 쓴 290명을 선정하여 이들이 쓴 總論文數 1,773편을 7개의 下位主題別로 집계하여 분석한 결과인 觀測確率을 豫測確率(=固定確率)과 비교해 보니 예측된대로 無機·分析 有機化學에 관한 論文을 발표한 研究者들의 비율이 낮아졌으며 物理·生·高分子化學에서는 그들의 비율이 높아지고 있었다. 그러나 工業化學만은 研究者들의 비율이 조금 높아지리라는 예측과는 달리 다소 낮아지고 있었다. 결론적으로 7년동안 化學者들의 下位主題間의 推移過程을 分析하여 얻은 각 下位主題의 固定確率은 앞으로의 科學者들의 確率分布를 대체적으로 예측해 주고 있다고 말할 수 있겠다.

第2實驗은 1983년 이후의 각 下位主題에 대한 化學者들의 비율을 豫測하기 위해서 1967~83년 동안 648명의 저자가 쓴 總論文數 4,598편을 분석하여 각 下位主題의 初期狀態確率과 固定確率을 구하여 比較·分析하였다. 분석결과 1983년 이후에는 初期狀態確率과 비교해서 無機·分析·有機·生化學 分野의 化學者들의 관심이 高分子·物理·工業化學 分野로 쏠릴 것으로 예측된다.

第3實驗은 동시인용법을 이용하여 特定分野의 知的構造를 규명하듯이 마코브

체인의 推移行列을 이용하여 2次元의 공간에 下位主題를 나타내는 점들을 매핑해 보았는데 대체로 각 下位主題가 獨立된 영역임을 나타내 보이고 있었고 群集分析結果 有機化學, 生化學, 分析化學이 모여서 하나의 군집을 이루고 工業化學, 物理化學, 無機化學이 모여서 또다른 군집을 이루었는데 高分子化學은 어느 군집에도 속하지 않고 따로 떨어져 있다. 지도의 중심에 위치한 物理化學 分野는 6개의 他下位主題들에서 物理化學에로의 推移條件確率의 합이 가장 큰 他下位主題들과의 관계가 가장 密接한 하위주제이며 그 어느 군집에도 속하지 않고 꼭 떨어져 고립된 모습을 보이는 高分子化學은 6개의 他下位主題들에서 高分子化學에로의 推移確率의 합이 가장 작은 他下位主題들과의 관계가 가장 미미한 下位主題였다. 세번째 實驗結果를 同時引用頻度에 의한 유사성을 바탕으로 多次元 蓄積技法과 群集分析에 의해서 규명한 化學分野의 知的構造와 比較·分析해 보는 것도 의미가 있을 것이다.

〈 參 考 文 獻 〉

1. 郭魯均·崔泰成, 「計量經營學 OR/MS」, (서울: 茶山出版社, 198).
2. 구자홍, 「確率論」, (서울: 民音社, 1988).
3. 金基永·郭魯均, 「計量意思決定論-經營問題解決의 計量的 接近-」, (서울: 法文社, 1989).
4. 金石英, “國內文獻의 著者同時引用에 관한 研究”, 「情報管理研究」 19;1(1988):73~94.
5. 김영진, 「論文的 同時引用을 통한 知的構造의 규명에 관한 研究」, 延世大學校 大學院 碩士學位論文, 1986.
6. 金賢姬, “學術 커뮤니케이션의 數量學的 分析에 관한 研究”, 「圖書館學」 14(1987): 93~130.
7. 朴乙龍 등, 「콘사이스數學辭典」(서울: 創元出版社, 1986).
8. 尹錫喆, 「計量經營學」(서울: 經文社, 1986).
9. 鄭英鎮, 「數理統計學; 確率과 統計」(서울: 喜重堂, 1987)
10. 최희찬·이대영·신철재, 「理工系 學生을 위한 應用數學」(서울: 韓新文化社, 1981).
11. Cinlar, E., *Introduction to Stochastic Processes* (Englewood Cliffs, N. J.: Prentice-Hall, 1975).

12. Garfield, E: *Mapping the Structure of Science* (N.Y.: Wiley, 1979).
13. Goffman, W. & Warren, K.S., *Scientific Information Systems and the Principle of Selectivity* (N.Y.: Praeger, 1980).
14. Isaacson, D.L. & Madsen, R.W., *Markov Chains Theory and Applications* (N. Y. :Wiley, 1976).
15. McCain, K.W. "Longitudinal Author Cocitation Mapping: The Changing Structure of Macroeconomics," *JASIS* 35; 6(1984): 351~359.
16. McCreery, L.S. "Bibliometric Study of Ethnomusicology, A Humanities Subject", (ph.D Diss., Case Western Reserve Univ., 1985).
17. Pao, M.L. & McCreery, L.S. "Bibliometric Application of Markov Chains", *Information Processing & Management* 22;1(1986):7~17.
18. Seung- hee Sohn. "A Quality Filetring Method for Biomedical Literature", (ph. D Diss., Case Western Reserve Univ., 1985).
19. Small, H.G. "Co- citation Context Analysis and the Structure of Paradigms", *J. of Docu.* 36;3 (1980): 183 ~ 196.
20. White, H.D. & Griffith, B. C. "Author Cocitation: A Literature Measure of Intellectual Structure", *JASIS* 32;3 (1981) : 163~171.
21. Zunde, P. & Slamecka, V. "Predictive Model of Science Progress", *Information Storage & Retrieval* 7(1971) : 103 ~ 109.