

3가원소 컨텍스트 기반의 클래스계층화분석 도구의 개발

강유경, 황석형, 김동순
선문대학교 컴퓨터공학부
e-mail:yukyung.kang@gmail.com

Development of a Hierarchical Classes Analysis Tool based on Triadic Context

Yu-Kyung Kang, Suk-Hyung Hwang, Dong-Soon Kim
Div. of Computer Engineering, Sunmoon University

요 약

매일 방대한 양의 다양하고 복잡한 데이터가 발생하는 오늘날 정보의 홍수 속에서 유용한 정보를 추출하기 위해서는 상당한 시간과 노력이 소요된다. 본 논문에서는 주어진 데이터로부터 공통속성을 갖는 객체단위의 정보를 추출, 분석하여 동치관계를 기반으로 클래스계층구조로 표현하는 클래스계층화 분석기법을 소개한다. 또한, 다양하고 복잡한 실세계 데이터들 속에 숨겨져 있는 개념들과 관계들을 파악하여 수월하게 클래스계층구조를 구축하기 위하여, 본 연구에서 개발 중인 3가원소 컨텍스트를 토대로 하는 클래스계층화분석 자동화지원도구를 제안한다.

1. 서론

오늘날 통신기술과 컴퓨터의 발달로 지식과 정보의 양은 기하급수적으로 증가하고 있으며, 매일 발생하는 방대한 양의 데이터들 속에서 손쉽게 유용한 정보를 추출하기 위한 다양한 데이터 분석기법들이 제안되고 있다[1].

최근 주목받고 있는 형식개념분석기법(FCA : Formal Concept Analysis)은 도메인 내의 다양한 데이터들로부터 개념들(Concepts)을 추출하고, 개념들 사이의 상-하위관계를 파악하여 개념계층구조(Conceptual hierarchy)를 구축하기 위한 정형화된 데이터분석기법이다[2,3]. 형식개념분석기법을 적용함으로써, 실세계의 데이터에 내포된 개념들에 대한 계층구조표현과 각종 질의정보처리 등이 가능하므로 데이터분석에 필요한 유용한 정보를 제공할 수 있다.

또한, 3가원소개념분석기법(TCA:Triadic Concept Analysis)은, 2가원소(객체, 속성)를 기반으로 개념추출 및 계층화분석을 수행하는 형식개념분석기법(FCA)을 확장하여 3가원소(객체, 속성, 조건)로 구성된 데이터에 대한 개념분석기법으로서 형식개념분석기법에서 다루는 데이터보다 다양하고 복잡한 실세계정보를 추출, 표현, 분석할 수 있다[4].

3가원소개념분석기법은 3가원소 컨텍스트(triadic context)를 토대로 3가원소개념(triadic concept)들을

추출하고 그들 사이의 관계를 파악하여 3가원소개념 격자(triadic lattice)를 구축함으로써, 데이터로부터 개념구조를 가시화할 수 있다. 그러나, 현재, 3가원소개념분석기법을 지원하는 자동화도구가 개발되지 않고 있는 상태이므로, 수작업에 의한 3가원소개념의 추출 및 계층구조의 가시화 등을 위해서는 매우 많은 시간과 노력이 필요하며, 구축된 3가원소개념 격자 또한 상당히 복잡하므로, 이를 토대로 하는 데이터분석 및 질의처리 등의 작업에는 상당한 어려움이 발생할 수 있다.

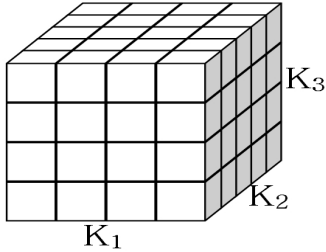
주어진 데이터들의 동치관계를 기반으로 클래스계층구조로 표현하는 클래스계층화분석기법(HCA: Hierarchical Classes Analysis)[5]은, 3가원소개념분석의 제반문제점을 해결하기 위한 유용한 도구로서 활용될 수 있다. 따라서 본 연구에서는 데이터로부터 정보를 추출/분석하여 동치클래스 계층구조로 표현하는 클래스계층화분석기법을 소개하고, 3가원소 컨텍스트를 입력으로 하여 클래스계층화분석을 수행하는 지원도구를 개발하고 있다.

본 논문은 다음과 같이 구성된다. 2장에서는 3가원소 컨텍스트와 클래스계층화분석기법의 기본적인 제반정의들을 소개하고, 클래스계층화분석방법을 설명한다. 그리고 3장에서는 본 연구에서 개발된 클래스계층화분석도구에 대해서 소개하고, 4장에서는 결

론과 향후 연구과제에 대해서 설명한다.

2. 3가원소 컨텍스트기반의 클래스계층화분석기법

3가원소 컨텍스트는, 어떤 객체가 특정한 상태/조건하에서 어떤 속성값을 갖는지에 대한 정보를 표현하고 있는, 일종의 데이터테이블이다(그림1 참조).



(그림 1) 3가원소 컨텍스트의 개념

[정의 1] 3가원소 컨텍스트 $K = (K_1, K_2, K_3, Y)$ 는 객체들(Objects)의 집합 K_1 , 속성들(Attributes)의 집합 K_2 와 상태들(Conditions)의 집합 K_3 그리고 그들 사이의 관계 $Y \subseteq K_1 \times K_2 \times K_3$ 로 구성된다. 즉, $(o, a, c) \in Y$ 는 어떤 객체 o 가 c 상태일 때 속성 a 를 갖는다는 것을 의미한다■.

표 1은 $K_1 = \{o_1, o_2, o_3, o_4, o_5\}$, $K_2 = \{a, b, c, d\}$, 그리고 $K_3 = \{I, II, III\}$ 로 구성된 3가원소 컨텍스트의 예이다.

<표 1> 3가원소 컨텍스트

| | I | | | | II | | | | III | | | |
|-------|---|---|---|---|----|---|---|---|-----|---|---|---|
| | a | b | c | d | a | b | c | d | a | b | c | d |
| o_1 | X | X | | | | | | | | | | |
| o_2 | X | X | | | | | | | | | | |
| o_3 | X | | | X | X | | | X | X | | | X |
| o_4 | X | X | | X | X | X | X | X | | | | X |
| o_5 | | | | | X | X | | | | | | |

클래스계층화분석기법은 주어진 데이터로부터 동치클래스(Equivalent class)를 추출하고 그들 사이의 순서관계(Order relation)를 파악하여 계층구조화된 형태로 가시적으로 표현한다. 클래스계층화분석에 관련된 제반정의들은 다음과 같다.

[정의 2] $i \in \{1, 2, 3\}$ 일 때, 주어진 두 요소 $a, b \in K_i$ 에 대해서, a 와 b 의 동치관계(Equivalence relation)는 다음과 같이 정의한다.

$$a \equiv b \Leftrightarrow (a)^* = (b)^*$$

단, $\{i, j, k\} = \{1, 2, 3\}$ 일 때 $j < k$ 고 $a_i \in K_i$ 에 대해서, $(a_i)^* = \{(a_j, a_k) \in K_j \times K_k \mid (a_i, a_j, a_k) \in Y\}$.■

예를 들어, 표1에서 객체 o_1 와 o_2 은 정의 2에 의해 동치이다.

$$o_1 \equiv o_2 \Leftrightarrow (o_1)^* = (o_2)^* = \{(a, I), (b, I)\}.$$

[정의 3] $i \in \{1, 2, 3\}$ 일 때, $a \in K_i$ 를 포함하는 동치클래스(Equivalent class)는 다음과 같이 정의된다.

$$[a] := \{b \in K_i \mid a \equiv b\} \\ := \{b \in K_i \mid (a)^* = (b)^*\}. \blacksquare$$

객체들의 집합(K_1), 속성들의 집합(K_2), 상태들의 집합(K_3) 각각에 대한 동치클래스는 다음과 같다.

$$K_1 = \{[o_1], [o_3], [o_4], [o_5]\}, \\ K_2 = \{[a], [b], [c], [d]\}, \\ K_3 = \{[I], [II], [III]\}.$$

이와 같은 객체/속성/상태동치클래스들을 각각 객체클래스, 속성클래스, 그리고 상태클래스라고 부른다.

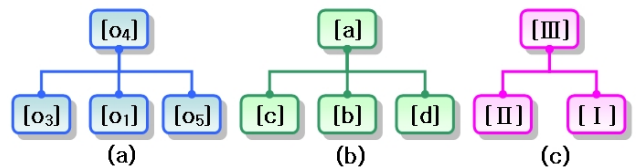
[정의 4] $i \in \{1, 2, 3\}$ 일 때, 주어진 두 클래스 $[a], [b] \in K_i$ 사이에 순서관계(Order relation)는 다음과 같이 정의된다.

$$[a] \succ [b] \Leftrightarrow (a)^* \subseteq (b)^*.$$

$i \in \{1, 2, 3\}$ 이고 $[a], [b] \in K_i$ 일 때, $[a]$ 는 $[b]$ 의 하위클래스(subclass)이며, 동시에 $[b]$ 는 $[a]$ 의 상위클래스(superclass)이다■.

$i = 1$ 일 때, K_1 의 두 클래스 $[o_4]$ 과 $[o_5]$ 에 대해서, $(o_4)^* = \{(a, I), (b, I), (d, I), (a, II), (c, II), (d, II), (a, III), (d, III)\}$ 이고, $(o_5)^* = \{(a, II), (c, II)\}$ 이다. $(o_5)^* \subseteq (o_4)^*$ 이므로, $[o_4]$ 는 $[o_5]$ 의 상위클래스이고, $[o_5]$ 는 $[o_4]$ 의 하위클래스이며, $[o_5] \succ [o_4]$ 나타낸다.

위의 정의2~4를 토대로 표1에 주어진 3가원소 컨텍스트의 객체/속성/상태집합 각각에 대한 클래스를 추출하여 계층화하면 그림2와 같다.

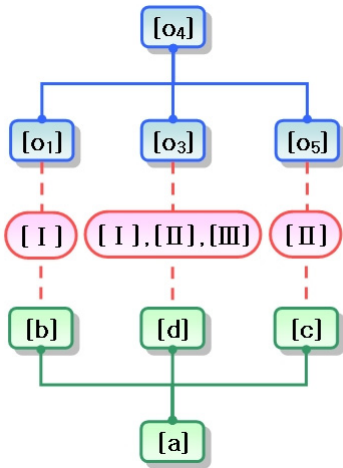


(그림 2) 객체/속성/상태클래스의 계층구조

객체/속성/상태집합 사이의 관계를 총괄적으로 파악하기 위하여 각각의 클래스계층구조를 통합하여 3가클래스계층구조(Triadic Class Hierarchy)를 구축할 수 있다. 그림 3은 표1에 대한 객체클래스계층구조(그림2의(a))의 클래스와 속성클래스계층구조(그림2의 (b))의 클래스를 상태클래스들을 통해서 연결시킨 3가클래스계층구조이다. 3가클래스계층구조에서 각 노드들은 객체 또는 속성 클래스이고, 각 노드들을 연결하는 선은 클래스들 사이에 순서관계를 나타내며, 타원형 노드는 상태클래스를 나타낸다. 객체클래스는 연관관계가 있는 속성 클래스와 상태클래스를 통해서 점선으로 연결되고, 연결된 속성클래스

의 모든 상위클래스들과도 관계가 있다. 속성클래스 또한 연결된 객체클래스의 모든 상위클래스들과 관계가 있다.

예를 들어, 객체클래스[o₃]은 상태클래스[I],[II],[III]를 경유하여 속성클래스[a],[d]와 연결되어 있으며, 객체클래스[o₃]은 [I],[II],[III]상태일 때, 속성클래스[a],[d]를 갖는다는 것을 의미한다. 한편, 속성클래스[c]는 상태클래스[II]를 통해서 객체클래스[o₅]와 연결되어 있고, [II]상태일 때, 속성클래스[c]를 갖는 객체클래스는 [o₅]와 [o₄]임을 나타낸다.



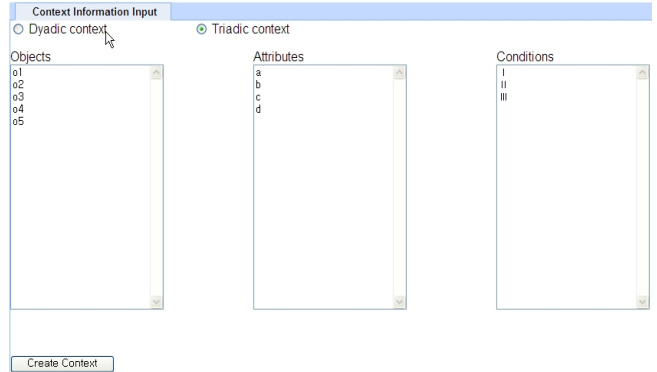
(그림 3) 표1에 대한 3가클래스계층구조

3. 클래스계층화분석도구(HCA Wizard)

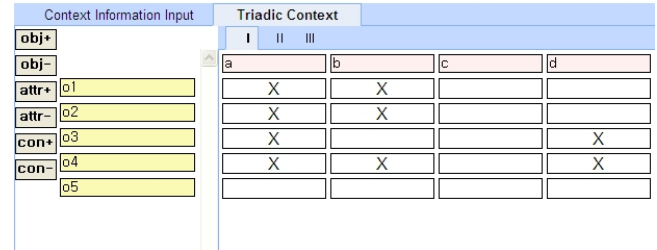
3.1. 기능

클래스계층화분석을 보다 수월하게 적용하기 위해 앞 절의 제반 정의들을 이용하여 2가원소 및 3가원소 컨텍스트를 지원하는 클래스계층화분석도구를 개발하였으며, 다음과 같은 기능들을 제공한다.

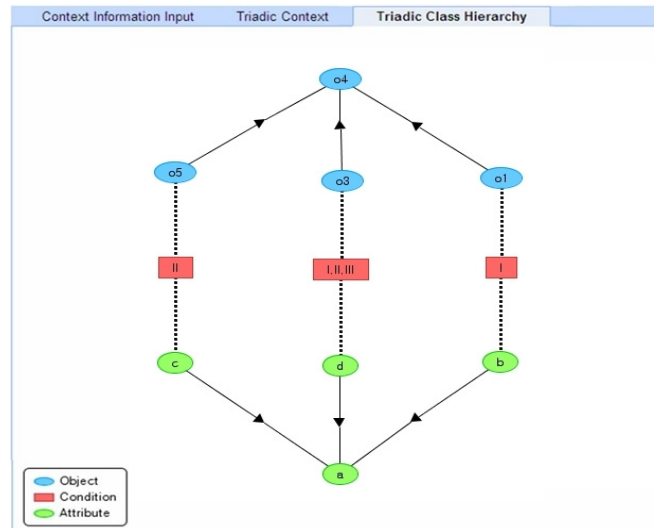
- **2가원소 및 3가원소 컨텍스트 생성 및 편집:**
그림4, 5와 같이, 클래스계층화분석도구는 사용자가 입력한 객체들, 속성들 그리고 상태들의 정보로부터 2가원소 컨텍스트 뿐만 아니라 3가원소 컨텍스트를 생성하고 편집할 수 있는 기능을 제공한다.
- **동치클래스 추출 및 클래스계층구조 가시화 :**
클래스계층화분석도구는 컨텍스트로부터 동치클래스들을 추출하여 그들 사이에 순서관계를 토대로 클래스 계층구조들을 구축하여 가시화한다. 그림6의 클래스계층구조에서 파란색 노드는 객체클래스를 나타내고 초록색 노드는 속성클래스이며, 빨간색 사각형 정점은 상태클래스이다. 노드들 사이의 화살표는 클래스들 사이에 상위-하위 관계를 나타내고, 점선들은 특정 상태클래스를



(그림 4) 객체/속성/상태 정보입력 화면



(그림 5) 3가원소컨텍스트 데이터 입력 화면



(그림 6) 3가클래스계층구조 가시화 화면

경유하는 객체클래스들과 속성클래스들 사이의 연결을 보여준다.

3.2. 실험

클래스계층화분석도구의 유용성을 검증하기 위해 본 논문의 참고문헌[2, 3, 4, 5]를 Bibsonomy(<http://www.bibsonomy.org>)에서 검색하여 사용자와 태그들을 추출하여 클래스계층화분석도구를 적용하여 본 연구에서 개발된 도구의 유용성과 적용범위, 그리고 문제점 등을 파악하기 위한 실험을 수행하였다. 이 실험에서 3가원소컨텍스트를 구성하는 객체/속성/상태집합은, 표2와 같이 각각, 4개의 참고문헌에 대해 Bibsonomy로부터 추출한 8명의 사용자, 7개 태그 그리고 4개의 참고문헌으로 규정하였다.

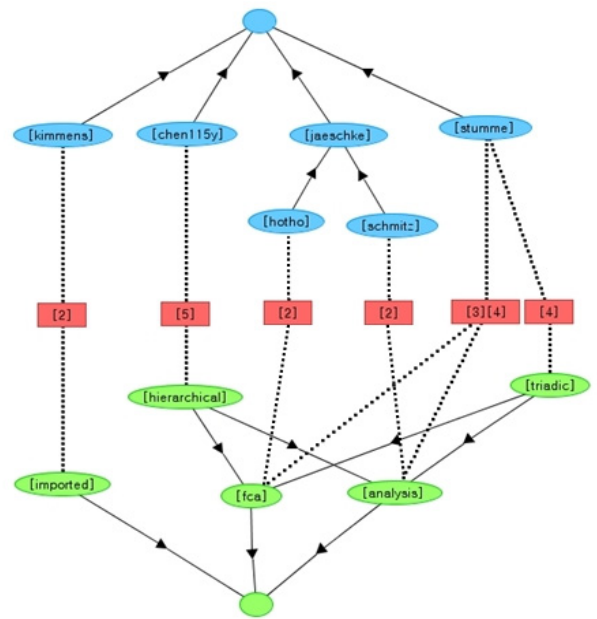
그림7에는 본 연구에서 개발된 클래스계층화분석

<표 2> Bibsonomy로부터 추출한 데이터에 대한 3가원소 컨텍스트

| | [2] | | | | | | [3] | | | | | | [4] | | | | | | [5] | | | | | | | | |
|----------|-----|---|---|-----|---|---|-----|---|---|---|-----|---|-----|---|---|---|---|-----|-----|---|---|---|---|---|-----|---|---|
| | i | a | c | fca | f | t | h | i | a | c | fca | f | t | h | i | a | c | fca | f | t | h | i | a | c | fca | f | t |
| kimmens | X | | | | | | | | | | | | | | | | | | | | | | | | | | |
| evol | X | | | | | | | | | | | | | | | | | | | | | | | | | | |
| tommens | X | | | | | | | | | | | | | | | | | | | | | | | | | | |
| jaeschke | | X | X | X | X | | | | | | | | | | | | | | | | | | | | | | |
| schmitz | | | | X | | | | | | | | | | | | | | | | | | | | | | | |
| hotho | | X | X | | X | | | | | | | | | | | | | | | | | | | | | | |
| stumme | | | | | | | | X | X | X | X | | | | X | X | X | X | X | | | | | | | | |
| chen15y | | | | | | | | | | | | | | | | | | | | | | X | X | X | X | | X |

i = imported, a = analysis, c = concept, f = formal, t = triadic, h = hierarchical

도구에 의해 수행된 실험의 결과 일부분을 나타낸 3가클래스계층구조이다. 3가클래스계층구조는 다음과 같이 해석된다. 사용자 chen15y는 태그"analysis" "fca", "concept", "formal", "hierarchical"를 참고문헌 [5]에 태깅하였으며, 사용자 stumme는 참고문헌[3]과 [4]에 "fca", "analysis", "concept", "formal"이라는 같은 태그를 사용하였고, 참고문헌[4]에는 "triadic"이라는 태그도 사용하였다. 특히, kimmens, evol과 tommens는 참고문헌[2]에 "imported"이라는 같은 태그를 사용하였다. 이와 같이, 같은 논문에 같은 태그를 할당한 사용자들을 동치클래스[kimmens]로 표현하였고, 사용자 jaeschke, schmitz, hotho, stumme, chen15y에 의해 모든 참고문헌([2],[3],[4],[5])에 함께 사용된 "analysis", "concept", "formal" 태그들을 동치클래스[analysis]로 표현하였다. 이 실험결과를 통해서, 본 연구에서 개발된 클래스계층화 분석도구는 3가원소컨텍스트로 구성된 데이터로부터 공통요소 및 관련관계, 그리고 다양한 부가정보들을 수월하게 추출/분석하여 클래스계층구조로 가시화할 수 있음을 알 수 있다.



(그림 7) Bibsonomy데이터를 대상으로 실시한 실험결과

로 확장하기 위한 제반 연구와 대량의 데이터를 의미적으로 분석하기 위해 데이터베이스를 연동하는 방법에 대해 연구할 필요가 있다.

4. 결론 및 향후 연구 과제

본 연구에서는 3가원소 컨텍스트를 토대로 클래스계층화분석기법을 이용하여 보다 수월하게 도메인에서 제공된 데이터들 속에 숨겨져 있는 개념들과 관계들을 파악하여 클래스계층구조를 구축하고 관리하기 위한 자동화도구를 개발하고 있다.

본 논문에서는 현재개발중인 클래스계층화분석도구의 핵심기능을 소개하였다. 본 연구결과물을 사용하여 기존의 형식개념분석기법에서 적용하기 어려웠던 3가원소형 데이터를 3가원소 컨텍스트로 표현하고, 이를 토대로 클래스계층화분석을 수행함으로써 3가클래스계층구조를 추출할 수 있다. 향후 연구과제로서는, 다양한 도메인에서 활용할 수 있는 도구

참고문헌

[1] Pang-Ning Tan and Michael Steinbach and Vipin Kumar, Introduction to Data Mining, Pearson, 2006
 [2] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer, 1999.
 [3] C. Carpineto, G. Romano, Concept Data Analysis: Theory and Applications, Wiley, September, 2004.
 [4] R. Wille, "The Basic Theorem of Triadic Concept Analysis", Order12, pp.149-158, 1995.
 [5] Chen, Y.H. and Yao, Y.Y. "Formal concept analysis based on hierarchical class analysis", Proceedings of The 4th IEEE International Conference on Cognitive Informatics (ICCI 2005), pp.285-292, 2005.