

레벨 교차 트리를 이용한 연관 서비스 탐사

황정희*

요약

사용자는 시간적, 공간적 상황에 따라 다양한 정보를 요구한다. 상황변화에 맞는 서비스 정보를 제공하는 것이 중요하다. 그러므로 사용자의 행동 및 서비스 이력의 최신정보를 기반으로 마이닝하여 최적의 서비스를 사용자에게 제공해야 한다. 본 논문에서는 시공간 정보 및 서비스 정보 온톨로지를 기반으로 사용자의 서비스 사용 이력을 이용하여 연관 있는 서비스 규칙을 탐색하기 위한 마이닝 방법을 제안한다. 이를 위해 서비스 온톨로지 계층에 대한 레벨 교차 기반의 연관 서비스 규칙을 발견한다. 제안된 마이닝 방법은 일정한 시간과 공간에 대한 시기별, 위치별, 연령별에 대한 연관 서비스 패턴을 발견할 수 있으므로 사용자의 상황변화에 양질의 서비스를 제공할 수 있는 기반이 된다.

키워드 : 서비스 규칙, 연관규칙, 온톨로지, 데이터 마이닝

Association Service Mining using Level Cross Tree

Jeong Hee Hwang*

Abstract

The various services are required to user in time and space. It is important to provide suitable service to user according to user's circumstance. Therefore it is need to provide services to user through mining by latest information of user activity and service history. In this paper we propose a mining method to search association rule using service history based on spatiotemporal information and service ontology. In this method, we find the associative service pattern using level-cross tree on service ontology. The proposed method is to be a basic research to find the service pattern to provide high quality service to user according to season, location and age under the same context.

Keywords : Service rule, Association rule, Ontology, Data mining

1. 서론

상황 인식 컴퓨팅이란 사용자의 입력을 최소화하여 상황에 따라 언제 어디서나 적절한 서비스를 받을 수 있는 환경을 제공하는 것이다. 상황 인식 컴퓨팅에 대한 요구가 높아지고 있으며

사용자의 위치와 상황에 맞는 서비스를 제공하기 위한 애플리케이션이 개발되고 있다[1, 2]. 그리고 사용자에게 적절한 서비스를 제공하기 위해서는 유용한 지식을 추출하는 효율적인 데이터 마이닝 기법이 필요하게 됨에 따라 기존 서비스 제공 환경의 한계를 극복한 새로운 개념의 서비스 환경을 구축하기 위한 연구들이 있다[3, 4]. 유용한 서비스를 제공하기 위해 함께 고려되어야 하는 점은 다양한 사용자들의 서비스 요청에 각 사용자의 특성에 맞게 적절한 서비스를 어떻게 제공할 것이냐는 점이다. 사용자의 시간과 위치와 같은 상황정보에 동적으로 반응하여 서비스 발견, 조합하는 사용자 중심의 서비스 제공 메커니즘이 필요하다[5].

※ 교신저자(Corresponding Author): JeongHee Hwang
접수일:2014년 06월 01일, 수정일:2014년 07월 10일
완료일:2014년 08월 04일

* 남서울대학교 컴퓨터학과

Tel: +82-41-581-2108, Fax: +82-41-581-2100

email: jhhwang@nsu.ac.kr

■ 이 논문은 2014년도 남서울대학교 학술연구비 지원에 의해 연구되었음

기존의 데이터 마이닝은 개념적인 계층을 고려하지 않고 단일개념에서 각 항목들간의 유용한 패턴과 규칙을 추출한다. 그러나 관심있는 도메인에서의 다계층 구조를 고려한 데이터 마이닝은 상세하고 구체적인 패턴과 규칙을 발견할 수 있는 특징이 있다. 다계층을 고려한 데이터 마이닝은 각 계층에서 최소 지지도를 만족하는 최대 빈발항목 집합에 대해서만 항목 크기를 증가시키면서 후보항목집합을 생성하고, 빈발한 항목들에 대한 하위계층의 빈발 항목집합을 계속 탐색해간다[6]. 그러나 이 방법의 단점은 계층간의 연관성있는 의미를 도출하기는 어렵다. 즉, 같은 레벨에서의 연관 규칙은 발견할 수 있지만 레벨이 다른 항목간의 연관 규칙을 발견하기 어렵다는 문제점이 있다. 이에 대한 해결 방법으로 계층간의 연관빈발항목을 발견할 수 있다는 레벨 교차 알고리즘[7, 8]이 제안되었다. 그러나 이러한 알고리즘은 레벨단위로 최소 지지도를 만족하지 않는 데이터 항목을 필터링하고 후보항목을 생성하여 빈발 항목을 탐색하기 때문에 수행시간이 많이 소요되고, 레벨 단위별로 필터링을 하기 때문에 발견 가능한 레벨간의 연관 항목을 발견하지 못하게 되는 문제가 있다. 따라서 본 논문에서는 온톨로지에 대한 계층정보를 고려한 연관 서비스 항목을 추출하기 위해 후보항목을 생성하지 않고 레벨 교차 정보를 트리로 생성하여 빈발 규칙을 발견하는 방법을 제안한다.

또한 기존의 마이닝 방법은 사용자의 빈발 행동패턴만을 추출하여 각 사용자의 상황과는 무관한 서비스 제공이 이루어질 수 있는 문제점이 있었다. 사용자의 프로필 즉, 성별, 나이 등의 개인정보 그리고 사용자의 시공간 상황과 연관된 서비스 패턴을 마이닝하면 사용자 컨텍스트에 따른 서비스를 제공할 수 있게 된다[9, 10, 11]. 그러므로 본 논문에서는 시공간 정보 및 서비스의 계층구조를 갖는 온톨로지를 기반으로 사용자와 연관된 도메인에서 관심있는 규칙을 발견하기 위해 온톨로지의 상위계층에서 하위계층으로의 연관탐색을 통해 의미있는 빈발 서비스 패턴과 규칙을 발견하여 사용자에게 제공하기 위한 시스템 구조 및 마이닝 방법을 제안한다.

데이터 마이닝에 의한 연령대별 서비스 패턴

을 탐사하고 이를 이용하면 사용자의 유형별 행동 패턴으로 사용자의 선호할만한 서비스를 예측할 수 있어 잠재적 고객을 유도할 수 있고 사용자의 만족을 극대화시킬 수 있다. 그러므로 시간과 공간에 따른 연령대별 서비스 패턴 발견을 통해 선호하는 유용한 서비스 정보를 적절하게 제공 및 추천할 수 있는 기반이 된다.

2. 관련연구

상황정보(context)란 사용자가 상호 작용을 하는 시점에 가용한 모든 정보이다. 이는 일반적으로 사람, 그룹, 객체의 위치, 식별, 활동, 상태 등을 포함한다. 상황정보는 사용자의 현재 활동과 같이 개인적인 것일 수도 있으며, 현재 사용 중인 기기와 같이 기술적인 것일 수도 있고, 온도, 위치 또는 시간과 같이 환경적인 것일 수도 있다. 상황인지 서비스 개인화는 이러한 상황정보의 수집 및 교환을 통해 인식하고 해석 및 추론과 같은 처리 과정을 거쳐 사용자에게 상황에 적절한 맞춤형 서비스를 제공한다[11].

상황인지 컴퓨팅은 사용자의 요청에 의해서나 사용자의 요청이 없는 경우에도 필요한 시점에 상황에 맞는 정보와 서비스를 지능적이고 능동적으로 제공할 수 있어야 한다. 이를 위해서는 주변의 다양한 상황정보가 잘 모델링되고, 수집되고, 분배되어야 함은 당연하다. 상황인지 서비스를 제공하기 위해서 수집된 상황정보는 사용자 정보와 결합되어 생활의 한 단위로 나타내어질 수 있는 고차원 정보로 분석 및 추론이 되어야 한다. 즉, 적절한 정보를 저장하고 체계적이고 효율적으로 공유하는 방법이 상황인지 시스템의 성공여부를 판단하는 중요한 기술이다[11].

온톨로지는 특정 도메인 내에 존재하는 단어를 정형화하고 명확한 개념으로 기술함으로써 의미적인 요소를 부여할 수 있다[9, 10]. 사용자의 환경과 배경을 개념적인 도메인으로 온톨로지를 구성하는 컨텍스트 온톨로지는 시간과 공간 그리고 사용자의 행동에 대한 정보를 모두 포함하므로 정확한 서비스를 제공할 수 있는 기초가 된다. 즉, 미리 정의된 컨텍스트 온톨로지 스키마를 사용하여 개인화된 서비스를 상황에 따라 언제 어디서나 제공할 수 있다. 유비쿼터스 컴퓨팅 환경의 프레임워크로서 기존의 [9]는 상

황인지 시스템을 위한 온톨로지를 제공한다. 특히 웹 언어들을 하나로 통합하여 정보를 공유하고 이들의 사실적인 정보들을 의미로 바꾸어 제공해 주는 서비스를 구현하였다. 그리고 [11]은 유비쿼터스 상황인지 시스템을 컨텍스트 온톨로지 및 추론 엔진, 컨텍스트 레벨조절 모듈, 컨텍스트 관리 모듈 등 3가지 구성요소로 나누어 컨텍스트 정보를 컴퓨터가 식별할 수 있는 수준의 정보로 가공하는 과정으로 구분하였다.

데이터 마이닝은 다양한 데이터에 대한 논리적인 구성이 가능하고 유용할 것으로 생각되는 숨겨진 패턴을 발견하기 위한 과정이다. 즉, 다양한 데이터에서 유용한 패턴을 찾아 의사결정에 도움을 주고 나아가 이러한 결정에 대한 결과를 예측할 수 있도록 한다. 특히 연관규칙은 데이터베이스의 트랜잭션에서 항목간의 연관성에 기초한 신뢰도와 지지도를 기반으로 하여 유용한 정보를 탐색한다. 데이터 마이닝 기법에서 상세하고 구체적인 패턴과 규칙을 발견할 수 있는 다계층 연관규칙(multiple level association rules)[6]은 Apriori 알고리즘을 기초로 대용량 데이터베이스로부터 효율적인 다계층 연관규칙 추출을 위해 Top-down 방식의 탐색방식을 이용한다. 다계층 구조를 가지고 있는 데이터베이스에서 Apriori 알고리즘을 이용하여 상위계층에서 하위계층으로 깊이 탐색하여 의미있는 패턴과 규칙을 발견한다. 즉, food 카테고리의 빈발항목인 우유와 빵에 대한 계층구조에서 2%우유와 white 빵이 빈발하다면 빈발하지 않은 같은 계층의 다른 항목은 제거되고 빈발한 항목의 하위계층에 있는 항목만을 추가 탐색하여 최종적인 빈발항목, Foremost의 2% milk와 OldMils의 white bread를 연관규칙으로 발견한다.

일반화된 연관규칙에 관한 기존의 접근방법[6]은 level-cross association rule로써 후보항목집합의 각 항목에 대한 개념 계층상의 모든 조상들을 트랜잭션에 추가하여 확장된 트랜잭션에 대해 지지도를 계산하는 방법이다. 이와 같은 연관규칙의 단점은 계산량이 많아지고 또한 계층간의 연관항목은 발견하지 못한다는 점이다. 즉, 각 항목들의 발생빈도를 계산하기 위한 데이터베이스 스캔을 레벨에 따라 하게 되어 항목 증가로 인해 계산량은 기하급수적으로 늘어나게 되고, 같은 계층에 대한 연관 항목만 발견한다.

[7]은 레벨간의 빈발 항목을 발견하기 위해 Apriori 알고리즘에 기반한 후보항목을 생성한다.

본 논문에서는 서비스의 빈발 패턴을 마이닝하기 위하여 후보항목 발생이 없는 일반화된 연관규칙 알고리즘을 기반으로 레벨 교차 트리를 이용하여 연관규칙을 발견한다. 이것은 시공간 정보를 포함하는 온톨로지를 기반으로, 새롭게 도출되는 서비스 패턴 규칙을 이용하여 변화하는 사용자의 상황에 맞는 최신의 서비스 정보를 제공할 수 있는 기반이 된다.

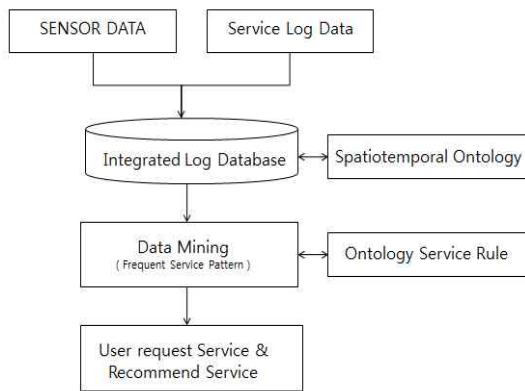
3. 서비스 상황 온톨로지를 이용한 시스템

시공간 정보를 기반으로 사용자의 서비스 이용 패턴을 발견하기 위한 시스템 구조는 어플리케이션 계층(application layer), 서비스 프로바이더 계층(service layer), 물리 계층(physical layer)으로 구성된다.

어플리케이션 계층은 사용자로부터 서비스 내용에 대한 질의 및 제공받은 최적의 서비스를 사용자에게 제공해 주는 응용계층이다. 센싱 계층은 상황 정보를 센서네트워크를 통해 센서 정보를 수집하는 가장 하위계층이다. 그리고 서비스 제공을 위해 가장 중요한 과정을 포함하고 있는 서비스 계층은 상황 정보의 기초를 구성하고 있는 시공간 및 서비스 온톨로지와 사용자 행동 패턴과 서비스 규칙을 발견하기 위한 데이터 마이닝 엔진, 그리고 새로운 서비스 생성 및 기존의 서비스 내용을 저장하고 있는 데이터베이스 등의 세부 구조를 포함한다. 온톨로지 엔진은 사용자에게 제공되는 기존 서비스 이력을 저장하고 마이닝 결과로부터 얻어진 새로운 규칙은 지속적으로 추가되어 저장된다. 온톨로지는 마이닝 수행에 필요한 시공간 온톨로지 정보를 제공하여 사용자의 상황에 적합한 서비스를 추출하는 기반이 된다. 즉, 온톨로지 데이터베이스에는 기존에 제공된 서비스 규칙의 저장 및 마이닝에 의해 생성된 새로운 규칙을 저장한다. 그리고 새로운 규칙이 발견되면 기존의 규칙보다 더 높은 우선순위를 부여하여 최근에 발견된 새로운 규칙을 우선적으로 적용한다.

(그림 1)은 세부적인 데이터 마이닝의 처리 구조를 보여준다. 사용자 요청 서비스 및 추천 서비스를 추출하는 과정은 세 단계로 이루어진다. 첫째, 데이터 통합 단계로 시공간 정보에 따른 사용자의 물리적인 센서 정보와 기존의 서비스 사용 이력 정보를 통합하는 과정이다. 이는 사용자의 서비스 요구 및 추천을 위한 사용자의 물리적 시공간 정보와 관련된 기존의 서비스 정보를 필터링하여 통합한다. 두 번째 단계는 데이터 마이닝 단계에서 연관규칙을 사용하여 빈발 서비스 패턴을 추출한다. 서비스가 요구된 시공간 정보와 연관된 통합 서비스 데이터로부터 온톨로지에 저장된 서비스를 고려하여 사용자에게 적합한 서비스 제공을 위해 유용한 정보를 추출한다. 마지막 단계는 사용자의 서비스 요구 및 추천에 필요한 추출 서비스 정보를 제공하는 것이다.

(그림 1) 마이닝 처리구조



(Figure 1) Mining Process

시공간 관계는 시공간 객체와 관련된 사건들 간의 인과 관계(casual relationship)를 탐사하는데 매우 중요한 의미를 가지므로 시공간 특성을 함께 고려하는 온톨로지 설계가 필요하다. 시간과 공간을 함께 고려하는 온톨로지를 통해 사용자의 행동에 대한 적절한 서비스 제공이 가능하다.

사용자의 공간과 시간을 식별하는 시공간 정보는 사용자의 상황정보를 나타내고 상황에 따른 서비스 이력을 구분할 수 있게 하는 기준이 된다. 시간은 일정한 시간간격으로 구분한 정보

이고, 공간정보는 공간상의 좌표 (x_i, y_i) 에 해당하는 일반화된 위치 정보이다. 위치 일반화는 공간의 좌표 값에 대해 일정한 구역(zone)으로 일반화하여 사용자 및 객체의 위치를 식별한다.

사용자들은 자신들의 위치나 주변 환경에 상관없이 최적의 서비스 사용을 원한다. 서비스는 제공 가능한 기기, 위치, 실행 조건 등과 같은 기능에 대한 정보를 추상화하여 기술하고 계층적으로 구조화 하는 단위이다. 기존에 저장되어 있는 서비스 규칙집합에서 사용자의 요구에 맞는 서비스가 존재하지 않는다면 새로운 서비스를 생성하거나 기존의 서비스를 조합하여 제공해 주는 것이 필요하다. 서비스 온톨로지는 사용자의 행동과 밀접한 관련이 있는 기기나 서비스를 표현한다. 본 논문에서는 서비스 온톨로지를 일반적인 정보 서비스 중심의 클래스 계층으로 표현한다.

시공간 기반의 서비스 정보를 나타내는 온톨로지는 시간(time), 공간(location), 서비스(service)의 계층 구조로 구성된다. 즉, 사용자를 위한 서비스 정보는 시간과 공간 정보가 연관되어 있다. 시간 정보는 일상적인 생활의 행동패턴으로 나누어 구분할 수 있는 서비스 제공의 기준이 되는 주중, 주말 정보로 구분하고 이에 대한 하위레벨로 오전, 오후, 야간으로 나누어 일반화하였다. 위치 정보는 사용자들의 일상생활에서 가장 많은 시간을 보내게 되는 장소인 홈, 직장, 다운타운으로 구분하여 일반화하고 해당 장소에서 많은 사람들이 이용하는 공간으로 다시 세분화한다. 홈 공간은 방, 부엌, 거실로, 직장은 사무공간, 로비, 카페테리아 등으로 세분화한다. 서비스는 가이드, 추천, 예약으로 구분되며, 가이드는 다시 하위 레벨의 교통, 날씨로, 추천서비스는 전시회 및 상가 추천으로, 예약서비스는 호텔 및 식당 예약으로 세분화하였다.

4. 마이닝 알고리즘

사용자에게 제공하였던 서비스 이력정보들을 트랜잭션의 항목으로 고려하여 시간, 공간 정보와 연관된 서비스 연관 규칙을 발견하고, 발견된 서비스 규칙에 의해 사용자에게 최적의 서비스 제공을 적용할 수 있다. 즉, 일정한 기준이 되는

시공간에서 자주 사용되는 연관 서비스 패턴을 발견한다. 발견된 패턴은 언제, 어디서, 어떤 서비스를 자주 이용하는지 알 수 있고, 연관성 있는 서비스 패턴의 발견으로 서비스 조합을 확장하여 더 다양한 서비스 및 추천이 가능하다.

온톨로지에 저장된 서비스 규칙은 기존에 사용자에게 제공되었던 서비스 정보와 임의 사용자 또는 시공간과 연관되어 제공하는 규칙이 고정적으로 저장되어 있다. 그러나 최근에 사용자가 이용하거나 제공되기를 바라는 서비스는 시간의 흐름에 따라 변할 수 있다. 이러한 변화의 흐름을 반영할 수 있는 방법으로 최근에 사용된 사용자의 서비스 이력 데이터에 대해 연관 규칙을 발견하고, 온톨로지에 저장된 서비스에서 더 많은 서비스 조합을 통해 다양한 서비스를 제공할 수 있는 기반을 마련하는 것이 중요하다.

<표 1>은 온톨로지에 저장된 서비스 규칙의 예를 보여준다. 사용자 ID, 성별, 나이, 시간과 위치정보에 따라 이용 가능한 서비스 규칙의 예이다. 여기서 시공간 및 서비스 정보는 온톨로지 구조에 따라 레벨 정보를 고려하여 부여된 상세한 서비스 정보를 의미한다.

<표 1> 온톨로지의 서비스 규칙

user	sex	age	time	space	service
U1	F	28	Td1	Lh1	Aa3, Bb1
U2	M	37	Te2	Ld3	Aa3, Cc3
U3	M	23	Td1	Lw2	Cb2, Ac2
U4	F	52	Te3	Lh3	Bc3, Aa1

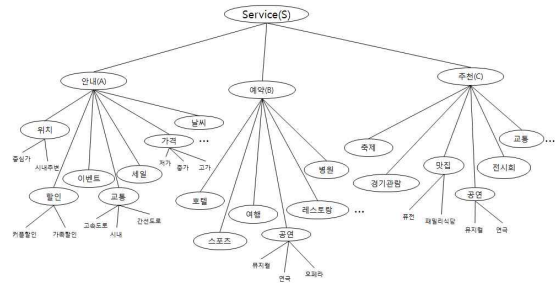
<Table 1> Service rule in ontology

서비스 빈발 패턴의 의미는 일정한 시간과 위치에서 같은 서비스가 자주 이용된다는 것을 말한다. 예를 들어, 사용자별 특정 시간과 장소에서의 빈발 서비스 패턴 (Aa3, Bb1, Cc2)은 자주 사용되는 연관 서비스를 의미하며, (Aa3, Bb1), (Aa3, Cc2), (Bb1, Cc2)도 빈발한 서비스 패턴이 된다. 이러한 서비스 규칙은 기준이 되는 특정 시간정보 또는 위치정보에 대한 서비스 계층내에서 시간별, 위치별, 연령별과 같은 관심 있는 서비스 연관규칙도 발견할 수 있다.

본 논문에서는 마이닝 과정을 설명하기 위하여, 20대 남녀를 대상으로 서비스 사용 트랜잭션에 대한 정보를 추출하여 마이닝을 수행한다고

가정한다. 최소 지지도를 3으로 하여, 서비스 온톨로지에 대한 계층정보를 사용하여 마이닝을 수행하는 과정을 설명한다.

(그림 2) 서비스 온톨로지



(Figure 2) Service ontology

시간, 공간, 서비스 클래스로 구분되는 상황 온톨로지에서 서비스 온톨로지를 구체화하여 표현한 것이 (그림 2)이다. (그림 2)의 서비스 온톨로지 계층 정보에 대한 단순한 표현을 위해 첫 번째 레벨은 대문자로, 두 번째 레벨은 소문자로, 세 번째 레벨은 숫자로 표시하여 트랜잭션의 항목을 나타낸다. 서비스 정보에 대한 알파벳 표기방법은 같은 레벨의 서비스에 대해 순차적으로 알파벳을 부여한다. 즉, 1-레벨은 대문자 알파벳 표기로써 안내(A)-예약(B)-추천(C)으로 표기하고, 같은 부모노드를 갖는 2-레벨의 서비스 내역은 소문자로 부여한다. 안내 서비스에는 위치(a)-할인(b)-이벤트(c)-교통(d)-세일(e)-가격(f)-날씨(g), 예약 서비스는 호텔(a)-스포츠(b)-여행(c)-공연(d)-레스토랑(e)-병원(f), 추천 서비스는 축제(a)-경기관람(b)-맛집(c)-공연(d)-전시회(e)-교통(f)으로 부여한다. 또한 이들의 하위계층은 같은 부모를 갖는 자식노드에 대해 숫자 1부터 차례로 부여한다. 예를 들면, Aa2는 안내서비스-위치-시내주변의 서비스 정보를 의미한다. 다음의 <표 2>는 서비스 이용에 대한 트랜잭션 데이터의 예이다.

<표 2>의 트랜잭션 데이터에서 1-레벨의 A, B, C 기준으로 1-레벨 중심의 1차적인 빈발 항목을 필터링한다. 즉, A**, B**, C**간의 연관 빈발항목을 추출한다. <표 2>의 트랜잭션에서는 1-레벨 항목기준의 빈발도는 A** :12, B** :9, C** :7으로 최소 지지도 3을 모두 만족한다. 그러므로 1-level 빈발항목의 하위레벨인 2-레벨

기준으로 빈발도를 만족하는 항목을 추출한다. 트랜잭션에 대한 항목별 빈발도는 Aa*:3, Ad*:3, Ab*:3, Af*:2, Ae*:1, Bc*:1, Bf*:1, Bd*:4, Ba*:1, Bb*:1, Cc*:5, Cf*:1이다. 각 항목별에 대해 주어진 최소 지지도 3을 만족하는 항목은 Cc*:5, Bd*:4, Aa*:3, Ad*:3, Ab*:3 이므로 필터링된 트랜잭션 데이터의 결과는 <표 3>이다.

<표 2> 트랜잭션 데이터

TID	Items
100	Aa2, Bc1, Cc1, Bf2, Cd1, Af1
200	Ad3, Bd2, Cc1, Ba1, Ab1
300	Cf2, Aa1, Af2, Ae1, Cc2, Bd1
400	Cc2, Bd2, Ad2, Bb1, Ab1, Be1
500	Cc1, Ab1, Ad1, Aa2, Bd2

<Table 2> Transaction data

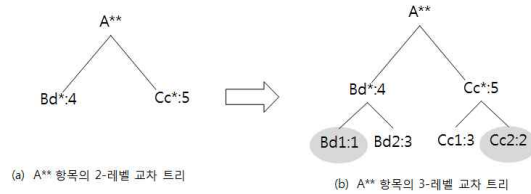
<표 3> 필터링 데이터

TID	Items
100	Aa2, Cc1
200	Ab1, Ad3, Bd2, Cc1
300	Aa1, Bd1, Cc2
400	Ab1, Ad2, Bd2, Cc2
500	Aa2, Ab1, Ad1, Bd2, Cc1

<Table 3> Filtered Data

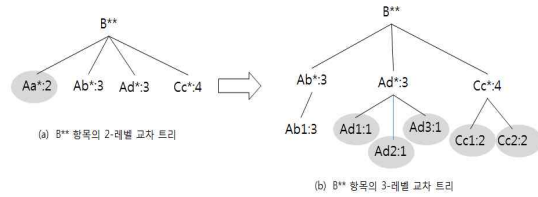
다음은 <표 3>의 트랜잭션 항목간의 연관성을 구성하는 레벨 교차 트리를 생성한다. 트리를 구성할 때 해당항목을 기준으로 연관항목의 하위레벨간의 빈발도를 계산하여 최소 지지도를 만족하면 해당 항목 하위레벨의 연관 빈발도를 체크하여 트리를 생성한다. 연관 항목간의 트리를 구성할 때 기준이 되는 항목과 같은 항목의 하위 레벨에 대해서는 트리를 구성하지 않는다. (그림 3-5)은 빈발 항목 A**, B**, C**에 대한 레벨간 교차 트리를 생성하는 과정을 보여준다. (그림 4)의 (a)에서 최소 지지도를 만족하지 못하는 Aa*항목은 더 이상 연관 빈발도를 계산하지 않는다.

(그림 3) 항목 A**의 레벨 교차 트리



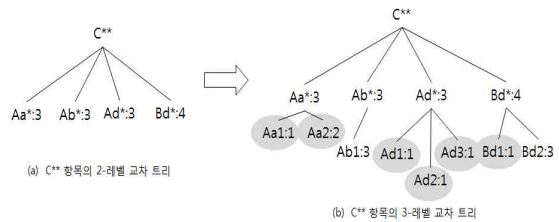
(Figure 3) Item A** level-cross tree

(그림 4) 항목 B**의 레벨 교차 트리



(Figure 4) Item B** level-cross tree

(그림 5) 항목 C**의 레벨 교차 트리



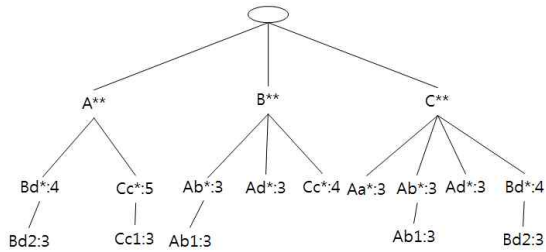
(Figure 5) Item C** level-cross tree

각 항목의 1-레벨 기준으로 레벨 교차를 이용하여 최소 지지도를 만족하지 못하는 항목을 제거한 결과의 빈발 연관항목 트리는 (그림 6)과 같다. (그림 6)에서 발견되는 연관 레벨 교차 항목은 A**~Bd2~Cc1, B**~Ad*~Ab1~Cc*, C**~Aa*~Ab1~Ad*~Bd2이다. 여기서 가장 하위 레벨과의 연관항목인 B**~Ab1의 경우, 이것은 B**~Ab1, B**~Ab*, B**~A** 간의 연관 관계도 포함되어 있음을 의미한다.

(그림 6)과 같은 빈발 레벨교차 트리로부터의 최대 빈발항목은 개별항목 기준의 공통된 빈발 항목을 의미한다고 할 때 최대 빈발항목은 Ab1~Ad~Bd2~Cc가 된다. 원시 데이터의 정보를 결부시켜 연관규칙의 항목을 설명하면 20대를 대상으로 하는 최대 빈발패턴 서비스는

Ab1-Ad-Bd2-Cc(커플할인안내-교통안내-연극 공연예약-맛집추천)가 추출된다. 즉, 추출된 결과는 같은 레벨간의 연관규칙뿐만 아니라 다른 레벨간의 연관규칙도 함께 발견할 수 있으므로 다양한 서비스 제공이 가능하다.

(그림 6) 빈발 연관 레벨교차 트리



(Figure 6) Frequent associative level-cross tree

이와 같은 빈발 레벨교차 트리를 구성하는 과정을 단계별로 설명하면 다음과 같다.

1단계: 트랜잭션에 대한 정보를 기준으로 온톨로지에 의한 1-레벨 빈발항목을 기준으로 최소 지지도를 만족하는 후보항목을 선별한다. <표 2>의 트랜잭션 데이터에서 1-레벨의 항목은 A**, B**, C**이며 모두 빈발하므로 빠른 실행을 위해 하위레벨의 2-레벨 기준으로 선별된 트랜잭션의 항목 즉, Aa*:3, Ab*:3, Ad*:3, Bd*:4, Cc*:4 으로 필터링하여 <표 3>을 만든다. 만약 1-레벨의 항목으로 충분한 필터링이 가능하면 1-레벨 기준으로 수행한다.

2단계 : 빈발한 1-레벨 빈발항목을 기준으로 다른 항목의 하위계층인 2-레벨 항목들과의 지지도를 계산한다. 예를 들어 1-레벨 항목 A**항목과 같이 발생하는 항목들, B**, C**항목의 2-레벨에 대한 빈발 지지도를 계산한다. 빈발항목을 발견하기 위하여 (그림 3)과 같은 레벨 교차 트리를 이용하여 해당 항목과 함께 발생하는 항목에 대한 발생빈도에 따라 지지도를 증가시킨다. 최소 지지도를 만족하는 2-레벨의 Bd*, Cc*에 대해서는 이들의 하위항목인 3-레벨의 빈발도를 계속 탐색한다. 그리고 3-레벨에서의 지지도는 Bd1:1, Bd2:3, Cc1:3, Cc2:2이므로 최소지지도를 만족하지 못하는 Bd1:1, Cc2:2는 제거한다. 이와 같은 방법으로 1-레벨의 최소 지지도를 만족하는 다른 항목들인 B**, C**에 대해서도 같

은 방식으로 빈발항목을 계산하여 온톨로지를 구성하는 다계층간의 빈발 레벨교차트리를 생성한다.

3단계: 생성된 레벨교차 트리를 이용하여 각 항목별 빈발한 연관항목과 공통된 최대 빈발항목집합을 추출한다. 1-레벨의 A**에 대한 연관 빈발항목은 같은 레벨의 B**, C**, 그리고 2레벨의 Bd*, Cc*, 3-레벨의 Bd2, Cc1이다. 즉, A**-B**-C**-Bd*-Cc*-Bd2-Cc1이다. 여기서 빈발항목 관계에서 상위레벨과 하위레벨의 관계가 중복되는 경우 더 상세한 정보를 포함하는 하위레벨의 항목을 고려한다. 즉, B**와 Bd* 보다는 하위레벨인 Bd2가 연관 항목으로 추출된다. 이것은 Bd2가 빈발하면 상위항목인 Bd*와 B**는 당연히 빈발하기 때문이다. 따라서 A**에 대한 빈발 항목집합은 {A**-Bd2-Cc1}이 된다. 이와 같은 방법으로 빈발 항목, B**, C**에 대한 빈발 항목 집합을 생성한다. 그리고 이들 연관 항목에 대한 빈발 공통항목을 최대 빈발 항목으로 추출한다.

마이닝 과정에 대한 알고리즘은 다음과 같다.

(그림 7) 마이닝 알고리즘

```

Input: transaction table T[1], min_sup
Output: frequent service items

if i_level =1 {
    find frequent_service_1_level_item,
    F[i_level, 1_item] ≥ min_sup
    descending by frequency of F[i_level, 1_item]
    make filtered_table, T[2]
}
for (i_level=2; i_level ≤ max_level; i_level++) {
    for (j=1; j ≤ transaction_cnt; i++) {
        if frequent item on T[2],
            F[i_level, 1_item ] ≥ min_sup;
            make frequent_itemset;
    }
    Make cross_level_frequent_itemset;
}
Find maximum frequency item from
cross_level_frequent_itemset;
    
```

(Figure 7) Mining algorithm

제안된 마이닝 알고리즘은 동일한 시간 및 공

간에서의 연령별 서비스 제공이력 데이터를 수집하여 발견한 규칙을 통해 같은 연령대 사용자에게 서비스 제공 및 추천에 유용하게 활용될 수 있다.

5. 실험 및 평가

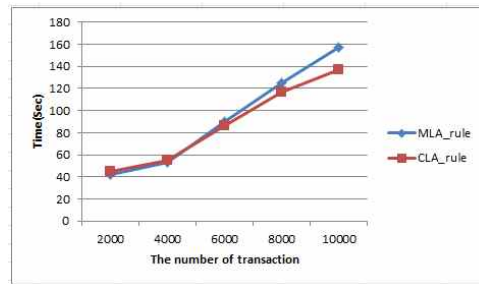
본 논문에서 제안한 트리 구조 및 알고리즘의 성능을 분석하기 위하여 온톨로지 규칙 베이스에는 특정 위치에 따라 제공 가능한 서비스 조합의 규칙이 저장되어 있다고 가정한다. 실험은 Intel Quad Core 2.5GHz, 2GB 메모리에서 JAVA로 구현하였다.

사용자들의 서비스 사용 로그 데이터베이스에는 사용자의 연령, 성별 등에 대한 기본정보 및 시공간 정보를 포함한 서비스 사용 내역을 포함하고 있다. 이러한 서비스 사용 로그 데이터베이스에서 동일한 위치기준의 같은 연령대의 사용자에 대한 서비스 사용 이력 정보를 이용하여 빈발 서비스 패턴을 추출한다. 실험에는 10,000개의 트랜잭션을 대상으로 한다. 각 트랜잭션에는 랜덤하게 2~3개의 기본 서비스 정보를 포함하도록 하고, 본 논문에서 적용한 레벨 교차 마이닝 알고리즘을 이용하여 연관된 서비스 빈발 패턴을 발견하였다. [7]의 다계층 연관규칙 알고리즘(MLA_rule)과 레벨 교차 연관규칙 알고리즘(CLA_rule)의 수행속도를 비교한 결과를 (그림 8)에서 보여준다. 실험결과에서 각 레벨단위로 최소 지지도를 만족하지 않는 항목을 제거하여 필터링을 수행하고 후보항목을 생성하는 MLA_rule은 후보항목을 생성하지 않는 CLA_rule보다 더 많은 실행시간이 소요됨을 알 수 있다. 마이닝 초기에는 근사한 수행시간이 소요되지만 트랜잭션의 수가 많아질수록 MLA_rule의 수행시간과 차이가 많이나는 것을 알 수 있다. 이것은 트랜잭션 수의 증가로 탐사되는 항목이 많을수록 MLA_rule에 비해 CLA_rule이 더 빠른 처리 속도를 보이는 것을 알 수 있다.

다음 실험에서는 트랜잭션의 증가에 따라 MLA_rule 알고리즘과 CLA_rule 알고리즘에서 발견되는 규칙의 수를 비교하였다. MLA_rule에서는 각 레벨에서의 최대 빈발항목의 합계를 계산하였고, CLA_rule에서는 레벨 교차를 통해 생

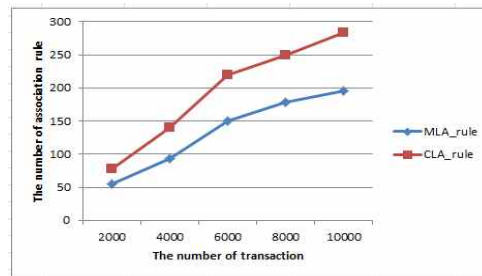
성된 빈발항목 수를 계산하였다. (그림 9)는 마이닝 알고리즘으로 발견된 연관규칙의 수를 비교한 결과이다. 결과에서 보는 것처럼 CLA는 MLA_rule 알고리즘에 의한 결과보다 평균 1.45배의 규칙을 생성하는 것을 알 수 있다. MLA_rule은 레벨단위로 필터링을 통해 후보항목을 생성하는 과정에서 최소 지지도를 만족하지 않는 해당 레벨의 항목이 제거되어 발견 가능한 레벨간의 연관 빈발 항목을 탐색하지 못하게 되는 결과를 초래한다. 이에 반해 CLA_rule은 1레벨 또는 2레벨에서만 필터링 과정을 거쳐 하위레벨의 항목간의 연관 정보를 탐색하므로 2레벨보다 낮은 하위 레벨로 마이닝을 수행하면서 더 많은 연관 항목을 발견하는 결과를 볼 수 있었다.

(그림 8) 수행시간



(Figure 8) Running time

(그림 9) 연관규칙의 수



(Figure 9) The number of association rules

데이터의 의미 관계를 정의하는 개념 계층은 정해진 표준이 존재하는 것이 아니기 때문에 서로 다른 레벨에 있는 항목일지라도 연관성 있는 데이터가 될 수 있다. 따라서 본 논문에서 적용하는 레벨 교차 비교를 통한 알고리즘 적용이 실제적으로 사용자에게 유용한 서비스를 제공하

기 위한 정보 탐색을 위해서 활용될 가치가 높다고 할 수 있다.

6. 결론

본 논문에서는 일정한 시공간 정보에서의 특정 사용자 그룹에 대해 연관된 서비스 패턴을 추출하는 마이닝 기법을 제안하였다. 추출된 서비스 빈발패턴은 성별, 나이 등의 개인 기본정보와 시공간 정보를 기반으로 서비스 온톨로지 계층정보를 고려하여 연관 패턴을 추출하므로 사용자들에게 더 다양한 서비스를 제공할 수 있고 유용한 정보를 추천할 수 있다. 실험을 통해 레벨 교차 트리 개념을 적용하여 더 많은 연관규칙을 발견할 수 있다는 결과를 볼 수 있었다. 중복된 규칙이 많이 생성될 수 있는 문제 그리고 생성되는 트리의 사이즈를 줄일 수 있는 방안에 대한 추가 연구가 필요하다.

References

- [1] M. Oh, J. Lee, B. Chang, J. Ahn, K. Doh, "A Programming Environment for Ubiquitous Computing Environment," ACM SIGPLAN Notices Volume 42(4), pp. 14-22, 2007.
- [2] S.Wang, J. Min, B.Yi, "Location Based Service for Mobiles: Technologies and Standards," IEEE International Conference on Communication (ICC), 2008.
- [3] L. Brisson, M. Collard, "An Ontology Driven Data Mining Process," The Tenth International Conference on Enterprise Information Systems, 2008.
- [4] A. Bellandi, B. Furlletti, V. Grossi, A. Romei, "Ontology-driven Association Rules Extraction: a Case of Study, The International Workshop on Contexts and Ontologies: Representation and Reasoning, 2007.
- [5] M. Takenmoto, et al., "A Service-Composition and Service-Emergence Framework for Ubiquitous-Computing Environments," Proc. of the 24th International Symposium on Applications and the Internet Workshops, 2004.

- [6] J. Han, Y. Fu, "Discovery of Multiple-Level Association Rules from Large Databases," IEEE Transactions on Knowledge and Data Engineering, 11(5), 1999.
- [7] R. S. Thakur, R.C. Jain and K. R. Pardasani, "Mining Level-Crossing Association Rules from Large Databases," Journal of Computer Science 2(1), pp. 76-81, 2006.
- [8] V. Ramana, M. Rathnamma, A. Reddy, "Methods for Mining Cross Level Association Rule In Taxonomy Data Structures," International Journal of Computer Applications, Vol. 7, No. 3, 2010.
- [9] C. Harry, F. Tim, "An Ontology for Context-aware Pervasive Computing Environments," Workshop on Ontologies and Distributed Systems, IJCAI Press, 2003.
- [10] Q. Ni, "Service Composition in Ontology enabled Service Oriented Architecture for Pervasive Computing," Workshop on Ubiquitous Computing and e-Research, 2005.
- [11] M. Khedr, A. Karmouch, "Negotiating Context Information in Context-aware Systems," IEEE Intelligent Systems, 2004.



황정희

2001년 :충북대학교 전자계산학과 (이학석사)
2005년 :충북대학교 전자계산학과 (이학박사)

2001년~2006년 :경우씨시스템(주) 연구소장
2006년~현재 :남서울대학교 컴퓨터학과 조교수
관심분야 : XML 및 웹 데이터베이스, 유비쿼터스 컴퓨팅, 데이터 마이닝