

사용자 웹 사이트 방문 시간을 고려한 연관 규칙

강형창* · 김철수* · 이동철**

*제주대학교 전산통계학과

**제주대학교 경영정보학과

Association Rule by Considering Users Web Site Visiting Time

Hyung-Chang Kang* · Chul-Soo Kim* · Dong-Cheol Lee**

*Dept. of Computer Science and Statistics, Cheju National University

**Dept. of Management Information System, Cheju National University

We can offer suitable information to users analyzing the pattern of users. An association rule is one of data mining techniques which can discover the pattern. We use an association rule which considers the web page visiting time and we should the pattern analyse of users. The offered method puts the weights in Web page visiting time of the user and produces an association rule. Weight is web page visiting time unit divide to total of web page visiting time. We offer rather meaningful result the association rule by Apriori algorithm. This method that proposes in the paper offers rather meaningful result Apriori algorithm

Keywords : association rule, pattern discovery, web page visiting time

1. 서 론

웹 사용자들이 웹 사이트 방문 이탈을 하지 않게 하기 위해서는 사용자들이 편리하게 정보를 얻을 수 있도록 서비스해야 한다. 대부분의 웹 사이트는 사용자의 단계적인 클릭행위를 통해 접근하는 방식으로 운영되고 있기 때문에 고객별로 차별화되지 않은 일률적 정보를 제공함으로써 사용자가 필요로 하는 정보 탐색에 있어 많은 시간과 노력이 소요 된다. 이러한 결과는 웹 사이트를 이용하는 사용자의 전반적인 만족도 하락을 초래한다.

웹 사이트에서 사용자에게 알맞은 정보를 제공하는 전략을 세우기 위해서는 사용자 개개인의 행동 패턴에 대한 정보가 필요하다. 이와 같은 정보를 기반으로 사용자 개개인의 특성에 맞는 동적인 웹 페이지 구성이나 링크정보를 제공할 수 있다. 차별화된 서비스를 제공하기 위해서는 사용자가 웹 사이트에 방문하는 접근경로 및 방문 시간 등의 정보가 필요하게 된다. 인터넷 사용

자가 웹 사이트를 방문하면 웹 서버에는 사용자가 요청한 서비스나 방문 페이지(page view) 및 방문 시간(visiting time) 등에 대한 정보를 파일 형태로 저장하게 되며 이를 웹 로그 파일(web log file)이라 한다. 웹 로그 파일을 이용하여 웹 사이트 사용자들의 행동 패턴을 분석하여 차별화된 서비스를 제공할 수 있다[4].

데이터 마이닝(data mining)은 데이터베이스 또는 대용량의 데이터에 내재되어 있는 의미 있는 정보를 찾는 데 쓰인다. 웹 로그 파일을 분석하기 위해 데이터 마이닝을 웹에 적용한 것이 웹 마이닝(web mining)이다[5, 8].

웹 마이닝은 웹 사이트 개선 및 광고 효과 측정 또는 상품 추천 등에 응용될 수 있다.

웹 마이닝에서 사용자의 패턴을 파악하고 분석하기 위해 주로 사용되는 기법으로 연관 규칙 탐색과 군집분석을 들 수 있다[1]. 특히 연관 규칙 탐색은 웹 페이지 간 빈발항목집합(frequent itemsets)을 찾아내고 이들로부터 연관 규칙을 생성한다. 기존 대부분의 연관 규칙 탐색기법들은 빈발하게 나타나는 데이터들만 탐색 대상으

로 하기 때문에 빈발하게 발생되지 않는 데이터들은 연관 규칙이 생성되지 않고, 빈발하게 나타나지 않는 데이터들을 탐색 대상으로 하는 알고리즘들도 있으나 이러한 알고리즘들은 데이터들 사이의 연관정도만 고려하였기 때문에 의미 없는 연관 규칙을 탐색하게 되는 문제가 있다[2].

인터넷 사용자에게 따라 웹 페이지를 방문하는 시간은 다르다. 따라서 사용자들이 웹 페이지를 방문하는 시간을 고려하여 연관 규칙을 탐색할 필요가 있다.

본 논문에서는 사용자들이 웹 사이트를 방문하였을 때 방문하는 웹 페이지별 방문 시간을 고려하여 연관 규칙을 탐색하는 경우 빈발항목집합을 구성하기 위한 방법을 제안한다.

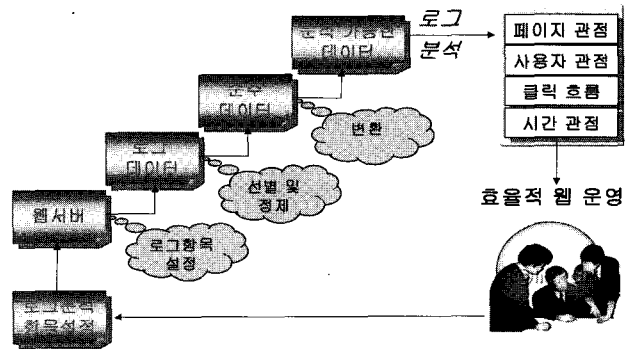
2. 관련 연구

2.1 웹 로그 파일과 웹 마이닝

웹 로그 파일은 사용자들이 웹 사이트에 접속할 때마다 요청 시간, 요청 페이지 등과 같은 정보를 웹 서버에 파일 형태로 저장되는 정보이다. 웹 로그는 웹 마이닝 분야에서 가장 많이 사용되는 정보로 사용자들의 행동 패턴을 분석하기 위한 정보들이 포함되어 있다. 하지만 실제 웹 로그 파일에는 이미지 정보나 스크립트 정보와 같은 불필요한 정보가 많이 포함되어 이를 제거하기 위한 전처리 과정이 필요하다. 전처리 과정을 통한 웹 로그 파일을 통해 특정 IP를 가진 사용자가 페이지끼리 이동한 경로를 추출하여 웹 사이트의 구조적인 정보를 유추할 수 있으며, 사용자의 행동 패턴 분석이 가능하다. 이러한 웹 로그 파일은 웹 로그를 수집하는 수준에 따라 클라이언트 수준, 프록시 수준, 서버 수준으로 구분할 수 있는데 분석을 위하여 사용되고 있는 대부분의 웹 로그 데이터는 웹 서버 수준에서 얻어지는 것이다.

웹 마이닝은 “웹 상에서 얻어지는 데이터를 마이닝” 하는 것으로 분석 대상의 유형에 따라 웹 구조 마이닝(web structure mining), 웹 내용 마이닝(web content mining), 웹 사용 마이닝(web usage mining)으로 구분할 수 있다. 웹 구조 마이닝은 웹 내용을 기술하는데 사용하는 구조화된 정보를 분석하는 과정으로서 하이퍼텍스트로 구성된 문서들의 구조에 대해 마이닝하는 것이다. 웹 내용 마이닝은 웹 사이트를 구성하는 페이지 내용 중에서 텍스트, 이미지, 오디오, 동영상 등과 같은 다양한 데이터로부터 얻어지는 내용에 대한 분석을 통하여 웹 사이트에 대한 유용한 정보를 찾아내는 과정이다. 마지막으

로 웹 사용 마이닝은 웹 서버에 저장된 웹 로그 파일을 이용하여 사용자들의 행동 패턴에 대한 정보를 분석하는 과정이다. 웹 사용 마이닝은 웹 사이트에서 사용자들의 웹 페이지 사용 패턴을 분석하고, 사용자가 웹 서핑하면서 발생하는 로그 데이터와 사용자가 직접 작성한 등록정보 등에 의해 얻어지는 데이터를 사용하여 수행한다. 웹 사용 마이닝은 접속 패턴을 찾는 작업과 개별 사용자의 사용 패턴을 분석하여 차별화된 서비스를 제공하기 위해 사용된다. 웹 마이닝은 대부분 웹 사용 마이닝을 의미한다. 웹 마이닝은 로그분석 항목 설정, 로그 항목 설정, 선별 및 정제·변환, 로그 분석, 웹 사이트에의 결과 반영의 일련의 과정으로 진행된다[4]. 다음 <그림 1>은 웹 마이닝 수행과정을 흐름도를 보여주고 있다.



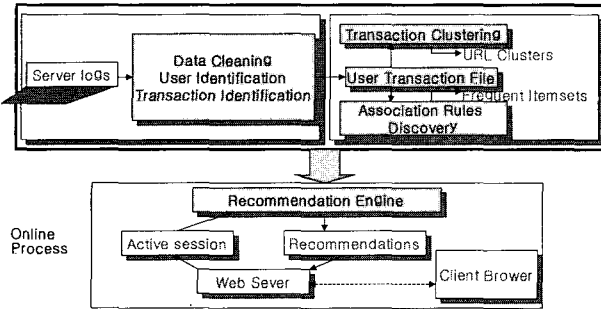
<그림 1> 웹 마이닝 분석 흐름도

2.2 웹 개인화

웹 개인화란 사용자의 요구에 가장 적절한 서비스를 개인별로 제공하는 것으로 사용자의 성향을 파악하고, 이에 맞추어 사이트를 적용하거나 혹은 변화시켜 서비스를 제공하여 사용자들이 해당 사이트로부터 좀더 쉽고, 빠르며, 효과적으로 적절한 정보를 제공할 수 있도록 해준다.

웹 개인화는 사용자의 요구에 맞는 자료를 제공해 주는 것부터 사용자의 요구를 예상하여 적절한 정보를 제공해 주는 것까지 범위가 다양하다. 기존 접근법들은 웹 개인화를 위해 사용자 프로파일(개인 신상 정보)에 많이 의존해 왔다. 그러나 프로파일이 가지고 있는 문제점(사용자들의 주관 개입, 서술적 정보의 부정확성, 프로파일의 정적인 성질 등) 때문에 정확한 프로파일 확보가 어렵다. 따라서 최근에는 자동적으로 사용자 프로파일을 발견하기 위해 웹 이용 마이닝과 같은 기법들을 이용하고 있다. 다음 <그림 2>는 웹 사이트를 개인화시키는데 있어 일반적으로 사용되어지는 아키텍처이다[4].

<그림 2>는 웹 로그 파일로부터 얻어지는 사용자 정보에 대해 정제하고, 사용자 정의 및 트랜잭션을 정의한 후에 패턴을 분석하기 위해 트랜잭션에 대해 군집화를 시행하거나 또는 연관 규칙을 생성하기 위한 아키텍처이다.



<그림 2> web personalization system 아키텍처

2.3 연관 규칙 탐색과 Apriori 알고리즘

웹 마이닝 분석 목적 중 하나는 “사용자들이 웹 사이트에서 어떤 패턴을 가지는가?”를 알아내는 것이기 때문에 ‘패턴 발견’은 매우 중요하다. ‘패턴 발견’을 위해 적용 가능한 마이닝 기법에는 연관 규칙(association rule), 순차 연관 규칙(sequential association rule), 군집화(clustering), 분류(classification) 등이 있다. 이러한 패턴발견의 여러 기법을 적용하여 웹 사이트에서 사용자 패턴을 분석함으로써 성향을 파악하고 행동을 예측할 수 있다. 특히 연관 규칙 탐색(association rule discovery)은 웹 페이지 관점에서 비목적성 기법으로 목표 변수 없이 규칙 관계를 설명할 수 있다는 장점이 있다.

연관 규칙 탐색은 최소 지지도(minimum support)와 최소 신뢰도(minimum confidence)를 만족하는 규칙이며 2 단계로 구성된다[7].

단계 1 : 빈발항목집합을 구성한다. 빈발항목집합은 미리 결정된 최소 지지도를 만족하는 트랜잭션(transaction) 지지도(support)를 가지는 항목들의 모든 집합(itemsets)을 의미한다.

단계 2 : 단계 1에서 구성된 빈발항목집합을 이용하여 연관 규칙을 생성한다. 모든 빈발항목집합에 대해 빈발항목집합의 모든 공집합이 아닌 부분집합들을 찾고, 각각의 부분집합에 대하여 최소 신뢰도를 만족하는 규칙을 생성한다.

연관 규칙 탐색 알고리즘의 성능은 단계 1에서 결정된다. 빈발항목집합을 찾기 위한 여러 방법들이 연구되

어졌다. 대부분의 연관 규칙 알고리즘은 빈발항목집합을 생성하기 위해 빈발가능성이 높은 후보항목집합(candidate itemsets)을 생성한다. 후보항목집합들 중에서 빈발항목집합을 생성하기 위해 데이터베이스를 읽어가면서 후보항목들에 대한 지지도를 계산하게 되며 후보항목집합이 많을수록 많은 시간과 상당량의 메모리가 요구된다.

연관 규칙 탐색 알고리즘 중에서 Apriori 알고리즘은 후보항목집합을 먼저 구성한 후 빈발항목집합을 구성하는 알고리즘으로써 사전 지식(priori knowledge)을 이용하여 빈발항목집합을 생성한다[6].

Apriori 알고리즘에서 빈발항목집합을 구성하기 위한 방법은 k번째 항목집합이 k+1번째 항목집합을 발견하기 위해 레벨단위로 진행하여 반복 접근한다. 빈발항목집합을 생성하기 위한 순서는 다음과 같다. 첫째 1-빈발항목집합을 찾는다. 이 집합을 L_1 으로 나타내면 L_1 은 2-빈발항목집합인 L_2 를 찾는데 사용되며 L_2 는 다시 3-빈발항목집합 L_3 를 찾는데 이용되는 식으로 계속되어 더 이상의 k-빈발항목집합이 없을 때까지 진행된다.

빈발항목집합을 레벨단위로 생성하는 것을 효과적으로 개선하기 위해서 Apriori 알고리즘 특성인 “모든 공집합이 아닌 빈발항목집합의 부분집합은 반드시 빈발하다는 특성”과 “빈발하지 않은 집합들의 상위 집합들(super sets)은 빈발하지 않다”는 이용하여 탐색 공간을 감소시키는데 사용할 수 있다[7].

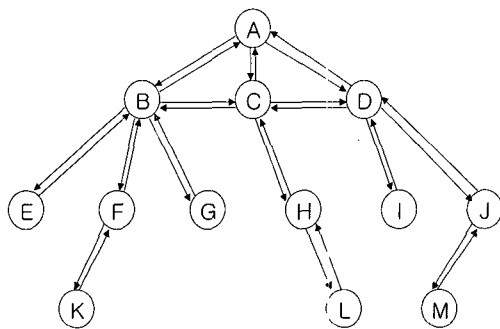
Apriori 알고리즘은 데이터베이스에 존재하는 모든 데이터들은 유사한 빈도수를 가지고 있는 것으로 가정하고 연관 규칙을 탐색한다. 그러나, 실제 많은 응용에서는 상대적으로 많은 빈도를 가지고 있는 데이터들이 존재하는가 하면 상대적으로 적은 빈도를 가지고 나타나는 데이터들도 있다. Apriori 알고리즘에서 상대적으로 적은 빈도를 갖는 데이터들에 대한 연관 규칙을 생성하는 경우, 최소 지지도는 낮게 설정되어야 한다. 그러나 최소 지지도를 너무 낮게 설정하면 빈발항목집합을 구성하기 위한 후보항목집합들이 상대적으로 많이 늘어나게 되고, 빈발항목집합을 구성하는 시간이 길어지게 된다[2].

인터넷 사용자들이 웹 사이트에 방문하는 페이지간 연관 규칙을 탐색하기 위해서는 사용자에 따른 웹 페이지 방문을 고려하게 되는데, 웹 페이지 빈발항목집합을 구성하기 위해 Apriori 알고리즘을 적용하게 되면 상대적으로 적게 방문한 웹 페이지는 빈발하지 않게 된다. 그러나 사용자에 따라 웹 페이지를 방문하는 것이 다르기 때문에 상대적으로 적게 방문하는 웹 페이지라 하더라도 방문 시간이 길면 빈발항목집합으로 간주하여야 한다[2, 3].

3. 웹 페이지 방문 시간을 고려한 연관 규칙 탐색

인터넷 사용자가 웹 사이트에 접속하여 웹 페이지를 방문하는 순서 및 방문 시간, 방문 횟수는 다르다. 다음 <그림 3>을 보면, 어떤 한 인터넷 사용자가 (A, E, I) 웹 페이지를 방문하는 것은 ¹(A, B, E, B, C, D, I), ²(A, B, E, B, A, D, I), ³(A, B, E, B, C, A, D, I) 등의 웹 페이지 이동 경로가 나타날 수 있다. (A, B, E, B, C, D, I) 페이지 이동 경로에서 A 웹 페이지에서 E 웹 페이지를 방문하기 위해서는 B 웹 페이지를 방문해야 한다. 그러나 B 웹 페이지는 E 웹 페이지를 방문하기 위한 경로에 불과하다. 이런 경우 방문하였다 하더라도 방문 시간이 상대적으로 적을 수밖에 없다. 이러한 웹 페이지의 경우 방문 횟수가 많아 빈발하더라도 빈발항목집합으로 볼 수 없다. 즉 사용자들이 웹 사이트에 접근한 경로 및 방문 순서, 방문 페이지, 방문 시간 등의 정보가 다르기 때문에 웹 로그 데이터에 대한 연관 규칙을 탐색 하는 경우 사용자에 따른 웹 페이지 방문 시간을 고려하여야 한다.

본 논문에서는 사용자들의 웹 페이지간 연관 규칙 탐색을 위해 페이지 방문 시간을 고려한 빈발항목집합을 다루고, Apriori 알고리즘에 의한 빈발항목집합과 비교한다.



<그림 3> 웹 페이지 구조

3.1 웹 로그 파일을 이용한 웹 페이지 연관 규칙 탐색

<표 1>은 사용자에 따라 웹 페이지를 방문한 이동 페이지 및 방문 페이지와 그 시간을 의미하고, 방문 시간이 0시간은 다른 페이지로 이동하기 위해 거쳐야하는 페이지를 의미한다. 다음 <표 2>는 웹 로그 파일을 정제된 데이터이다.

<표 2>는 사용자에 따라 웹 페이지를 방문했을 때 시간 단위를 주어 해당 웹 페이지의 방문 시간 단위가 최대한 값으로 할당된 것이다. 예를 들어 (A, E, I) 페이지

를 방문하는 경우 이동 경로는 (A, 3), (B, 1), (E, 5), (B, 0), (A, 0), (D, 1), (I, 2)이고, A 페이지는 3시간 단위와 0시간 단위 2개중 최대값인 3시간 단위 방문을 의미한다.

방문 시간이 10초 미만은 방문 시간 단위를 0으로 처리하고, 방문 시간이 10분 이상은 방문 시간 단위를 10으로 처리하였다. 그리고 연관 규칙을 생성하기 위한 각 웹 페이지 방문 시간 단위에 따른 가중값은 사용자가 웹 페이지를 방문한 총 시간 단위 합을 계산한 후 각 페이지 방문 시간 단위로 나누어 계산하였다.

<표 1> 전처리 된 웹 로그 데이터

User ID	(방문 페이지) : (이동 및 방문 페이지, 방문 시간 단위)
ID001	(A, E, I) : (A, 3), (B, 1), (E, 5), (B, 0), (A, 0), (D, 1), (I, 2)
ID002	(A, G, H, M) : (A, 0), (B, 1), (G, 3), (B, 0), (A, 0), (C, 1), (H, 2), (C, 0), (A, 0), (D, 1), (J, 1), (M, 9)
ID003	(A, K, L) : (A, 2), (B, 3), (F, 1), (K, 6), (F, 0), (B, 0), (A, 0), (C, 1), (H, 2), (L, 5)
ID004	(A, K, M) : (A, 1), (B, 1), (F, 2), (K, 4), (F, 0), (B, 0), (A, 0), (D, 2), (J, 2), (M, 10)
ID005	(A, I, L) : (A, 0), (D, 1), (I, 4), (D, 0), (A, 1), (C, 3), (H, 7), (L, 5)

<표 2> 정제된 웹 로그 데이터

User ID	(방문 페이지) : (이동 및 방문 페이지, 방문 시간 단위)
ID001	(A, E, I) : (A, 3), (B, 1), (D, 1), (E, 5), (I, 2)
ID002	(A, G, H, M) : (A, 0), (B, 1), (C, 1), (D, 1), (G, 3), (H, 2), (J, 1), (M, 9)
ID003	(A, K, L) : (A, 2), (B, 3), (C, 1), (F, 1), (H, 2), (K, 6), (L, 5)
ID004	(A, K, M) : (A, 1), (B, 1), (D, 2), (F, 2), (J, 2), (K, 4), (M, 10)
ID005	(A, I, L) : (A, 1), (C, 3), (D, 1), (H, 7), (I, 4), (L, 5)

웹 페이지 당 가중값 =

$$(\text{웹 페이지 방문 시간 단위} \div \text{사용자의 전체 웹 페이지 방문 시간 단위 합}) \dots\dots\dots (1)$$

웹 페이지 방문 단위당 시간 가중치는 다음 <표 3>과 같다.

<표 3> 방문 시간에 따른 방문 시간 단위 가중값

방문 시간 단위	방문 시간	방문 시간 단위	방문 시간
1	10초~1분	6	5분~6분
2	1분~2분	7	6분~7분
3	2분~3분	8	7분~8분
4	3분~4분	9	8분~9분
5	4분~5분	10	9분~10분

웹 페이지 연관 규칙 탐색을 위한 빈발항목집합을 찾기 위해 Apriori 알고리즘은 최소 지지도 $\text{minimum support}=0.4(\text{minimum count}=2)$ 로 하고, 웹 페이지 방문 시간을 고려한 알고리즘에는 최소 가중값 $\text{minimum weight}=0.4$ 로 하였다.

<표 4> 방문 페이지에 따른 가중값

User ID	(이동 및 방문 페이지, 방문 시간 가중값)
ID001	(A, 3/12), (B, 1/12), (D, 1/12), (E, 5/12), (I, 2/12)
ID002	(A, 0/18), (B, 1/18), (C, 1/18), (D, 1/18), (G, 3/18), (H, 2/18), (J, 1/18), (M, 9/18)
ID003	(A, 2/20), (B, 3/20), (C, 1/20), (F, 1/20), (H, 2/20), (K, 6/20), (L, 5/20)
ID004	(A, 1/22), (B, 1/22), (D, 2/22), (F, 2/22), (J, 2/22), (K, 4/2), (M, 10/22)
ID005	(A, 1/21), (C, 3/21), (D, 1/21), (H, 7/21), (I, 4/21), (L, 5/21)

<표 5>에서 보이는 결과는 다음의 순서로 진행되어 나타난 결과이다.

단계 1 : 전처리 된 웹 로그 데이터는 사용자들이 방문한 웹 페이지와 방문 시간을 기록한 자료를 구성하는 과정이다(<표 1>).

단계 2 : 단계 1에서 전처리 된 웹 로그 데이터를 정제하기 위해 사용자들이 방문한 웹 페이지가 2회 이상인 경우에는 최대 방문 시간을 할당하고, 1회 방문인 경우에는 1회 방문 시간을 할당하는 과정이다(<표 2>, <표 1>에서 고딕체로 표시된 내용이 할당된다).

단계 3 : 단계 2에서 정제된 웹 로그 데이터의 방문 시간에 따른 방문 시간 단위 가중값을 이용하여, 각 웹 페이지에 방문 시간 단위를 부여하는 과정이다. 방문 시간 단위는 1분으로 하였으며 방문 시간이 10초 미만인 경우는 방문 시간 단위

를 0으로 하였다(<표 3>).

단계 4 : 단계 3에서 할당된 웹 페이지 방문 시간 단위를 전체 웹 페이지 방문 시간 단위 합으로 나누어 방문 페이지에 따른 가중값을 부여하는 과정이다(<표 4>).

단계 5 : 단계 4에서 방문 페이지에 따른 가중값을 이용하여 빈발항목집합을 구성하는 과정이다(<표 5>).

4. 결 론

대부분의 연관 규칙 탐색 알고리즘들 특히 Apriori 알고리즘은 빈발항목집합을 생성하여 연관 규칙을 탐색한다. Apriori 알고리즘을 웹 페이지간 연관 규칙 탐색에 적용하게 되면 빈발하지 않은 웹 페이지는 연관 규칙이 생성되지 않게 된다. 이러한 결과는 자주 방문하지는 않지만 오랜 시간동안 방문하는 웹 페이지의 경우 빈발하지 않은 결과로 나타나기 때문에 사용자는 실제 의미 있는 웹 페이지이지만 빈발하게 나타나지 않기 때문에 연관 규칙으로 생성될 수 없다. 그러므로 사용자에 따른 웹 페이지 방문시간을 고려하여 웹 페이지간 연관 규칙을 파악하여야 한다. Apriori 알고리즘과 제안한 방법을 비교하기 위해 <표 1>의 가상 데이터를 구성한 후 두 알고리즘을 비교해 보면 <표 5>에서 보이는 바와 같이 Apriori 알고리즘의 경우 방문 웹 페이지횟수를 고려하여 빈발항목집합을 구성하기 때문에 방문 횟수가 적지만 방문 시간이 긴 페이지는 빈발항목집합으로 생성되지 않지만, 제안한 알고리즘은 사용자의 전체 웹 페이지 방문 시간을 고려하여 상대적으로 오랫동안 머무른 웹 페이지는 방문 횟수가 적더라도 빈발항목집합으로 선택됨을 볼 수 있다.

본 논문에서 제안한 웹 페이지 방문 시간을 고려한 연관 규칙 생성은 웹 사용자들이 원하는 정보를 찾기 위해 불필요한 웹 페이지를 방문하는 경우 또는 다른 웹 페이지로 이동하는 도중에 방문하는 웹 페이지는 방

<표 5> <표 4>에 대한 결과

항목 집합	Apriori 알고리즘 $\text{min_sup}=0.4(\text{min_count}=2)$	웹 페이지 방문 시간을 고려한 알고리즘 $\text{min_weight}=0.4$
1-항목 집합	A, B, C, D, F, H, I, J, K, L, M	A, B, E, H, I, K, L, M
2-항목 집합	AB, AC, AD, AF, AH, AI, AJ, AK, AL, AM, BC, BD, BF, BH, BJ, BK, BM CD, CH, CL, DH, DI, DJ, DM, FK, HL, JM	AB, AE, AH, AI, AK, AL, AM, BE, BH, BK, BL, BM, EI, HI, HK, HL, HM, KL, KM
3-항목 집합	ABC, ABD, ABF, ABH, ABJ, ABM, ABK, ACD, ACH, ACL, ADM, ADH, ADI, ADJ, AFK, AHL, AJM, BCH, BDJ, BDM, BFK, BJM, CDH, CHL, DJM	ABE, ABH, ABI, ABK, ABL, ABM, AEI, AHI, AHK, AHL, AIK, AIL, AKL, AKM, BEI, BHK, BHL, BHM, BKL, BKM, HIL, HKL

문 시간 단위 가중치를 0 또는 적은 값으로 두고 빈발하게 방문한 웹 페이지라 할지라도 의미 있는 규칙으로 생성하지 않으며 비록 빈발하게 방문하지 않은 웹 페이지라 할지라도 오랜 시간동안 방문한 웹 페이지는 방문 시간 단위 가중치를 큰 값으로 두어 의미 있는 규칙을 생성하고자 하는데 목적이 있다.

참고문헌

[1] 강현철, 한상태, 선영수, "A Clustering Algorithm Considering Structural Relationships of Web Contents", *The Korean Communications in Statistics*, 12(1) : 191-197, 2005.

[2] 김근형, 황병웅, 김민철, "중요지지도를 고려한 연관 규칙 탐사 알고리즘", *정보처리학회 논문지*, 11-D(3) : 545-552, 2004.

[3] 김정현, 김재련, "시간을 고려한 연관규칙을 이용한 웹 사용자 접근패턴 분석", *한국경영과학회/대한산업공학회 춘계공동학술대회 발표논문집*, 관동대학교, 양양, pp. 852-855, 2001.

[4] 최승배, 김규곤, 강만기, 웹 마이닝, 자유아카데미, pp. 146-149, 2004.

[5] Mannila, H., Toivonen, H., and Verkanmo, A. I., "Discovering frequent episodes in sequences," In Proc. of the First Int'l Conference on Knowledge Discovery and Data Mining, pp. 210-215. 1995.

[6] Jiawei Han, "Miclne Kamber, Data Mining : Concepts and Techniques," Simon Fraser University, pp. 227-236, 2001.

[7] Agrawal, R. and Srikant, R., "Fast algorithms for mining association rules," In Proceeding of the 20th VLDB Conference, pp. 487-499, Santiago, Chile, 1994.

[8] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava, "Data preparation for mining world wide web browsing patterns," *Knowledge and Information Systems*, 1, 1999.