

Tree-based Navigation Pattern Analysis¹⁾

Hyun Jip Choi²⁾

Abstract

Sequential pattern discovery is one of main interests in web usage mining. The technique of sequential pattern discovery attempts to find inter-session patterns such that the presence of a set of items is followed by another item in a time-ordered set of server sessions. In this paper, a tree-based sequential pattern finding method is proposed in order to discover navigation patterns in server sessions. At each learning process, the suggested method learns about the navigation patterns per server session and the discovered patterns are summarized into the modified Rymon's tree.

Keywords : Sequential pattern, Rymon's tree, Web usage mining, Navigation pattern.

1. 서론

웹 마이닝(Web mining)은 웹에서 발생하는 자료로부터 의미 있는 패턴(pattern)을 찾기 위한 데이터 마이닝의 응용분야이다. 웹 마이닝 기법들 중에서 웹 서버의 로그자료를 이용하여 사용자들의 네비게이션 패턴(navigation pattern)을 분석하기 위한 응용기법을 웹 사용 마이닝(web usage mining)이라고 하며, Srivastava등(2000)은 웹 사용 마이닝의 정의와 분석단계, 분석기법 그리고 최근 연구결과 등에 관하여 자세히 정리하고 있다.

웹 사용 마이닝은 사용자의 네비게이션 행위를 분석하고, 분석된 패턴으로부터 사용자들을 개인화(personalization)하여 웹 기반 상거래의 마케팅을 효율적으로 수행하기 위한 목적이 있다. 개인화를 위한 패턴 발견 기법들 중에서 많은 연구가 이루어지고 있는 방법들로는 연관성 규칙(association rule), 군집분석(clustering), 분류(classification) 그리고 열 패턴(sequential pattern) 분석방법 등이 있다.

웹 사용 마이닝은 주 분석대상 자료가 웹 서버의 로그에 있기 때문에 로그 파일에 영향을 미치는 모든 사용 현황 즉, 트랜잭션(transaction)들로부터 의미 있는 패턴을 얻기 위하여 많이 이용된다. 따라서 앞의 각 분석방법들은 의미 있는 패턴을 찾는 데 유용한 방법들이나 웹 서버 로그자료의 성격에 맞게 기존의 방법들을 수정하여 분석할 수 있는 방법에 관한 연구가 활발히 진행되고

1) This work was supported by the Research Support Program of Kyonggi University.

2) Assistant Professor, Department of Applied Information Statistics, Kyonggi University, Seoul, 442-760, Korea.

E-mail : hjchoi@stat.kyonggi.ac.kr

있다. 특히, Han등(1997)은 트랜잭션 자료로부터 발견된 연관성 규칙을 군집분석 자료로 이용하여 유사한 트랜잭션을 보이는 사용자 집단을 식별하는 방법을 제안하였다. 또한 Agrawal등(1993, 1996)이 제안한 Apriori 알고리즘을 직접 적용하여 트랜잭션 순서에 의해 생성되는 열 패턴을 발견하기 위한 방법 등이 제안되고 있다.

사용자의 네비게이션에 의해 생성되는 트랜잭션 자료는 순서가 부여된 URL들의 열로 정리될 수 있다. Mobasher등(1996)은 이러한 열로부터 순서를 가진 연관성 규칙을 찾기 위하여 알고리즘을 제안하고 있다. 또한 Schechter등(1998)은 사용자들의 네비게이션에 의해 생성되는 열들로부터 생성 가능한 모든 경로(path)들을 나무(tree)로 정리하는 알고리즘을 제안하여 현재 사용자에게 제공된 자원(resource) 이후에 사용 가능한 자원을 예측하기 위한 방법을 제안하였다. 그러나 이들 방법들은 주어진 특정기간 내에 발생한 트랜잭션 자료에 의해 분석이 이루어지기 때문에 분석에 이용될 총 열들의 수가 고정된 상태에서 학습(learning)이 이루어지고, 또한 자료가 가진 모든 규칙을 발견하기 위해서는 전체 자료에 대하여 여러 번의 반복학습이 이루어져야 한다. 따라서 열의 수가 고정되어 있지 않은 동적인 자료에 대해서는 이들 방법들을 직접 적용하기에는 어려움이 있다.

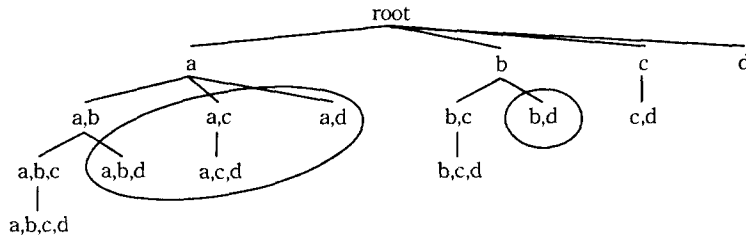
본 논문에서는 서버 세션(server session)에 의해 생성된 네비게이션 열들로부터 열 패턴을 찾기 위한 동적인 알고리즘을 제안하고자 한다. 제안된 알고리즘은 총 열들의 수가 고정되지 않은 상태에서 이전 학습에 의해 발견된 패턴에 새로운 열이 추가될 때마다 추가된 열에 대해서만 학습이 이루어지기 때문에 학습이 상당히 효율적으로 이루어진다. 또한 발견된 열 패턴들은 Wang(1997)이 제안한 열 패턴 분석방법과 같이 동적으로 변하는 나무에 의해 주어진 열들로부터 발생 가능한 모든 경로에 관한 정보가 정리되고 학습이 이루어진다. 따라서 이를 위해 Rymon(1992)이 제안한 나무에 열 패턴에 관한 정보를 노드에 부여한 동적인 나무를 제안하고자 한다.

2절에서는 열 패턴을 정리하기 위한 나무를 제안하고, 제안된 나무를 통해 학습을 수행하는 알고리즘을 제안하고 있다. 3절에서는 제안된 알고리즘을 평가하기 위해 Wang(1997)이 제안한 모형을 이용하여 모의실험을 수행한 결과를 정리하였다. 마지막으로 4절에서는 본 논문의 내용을 정리하였다.

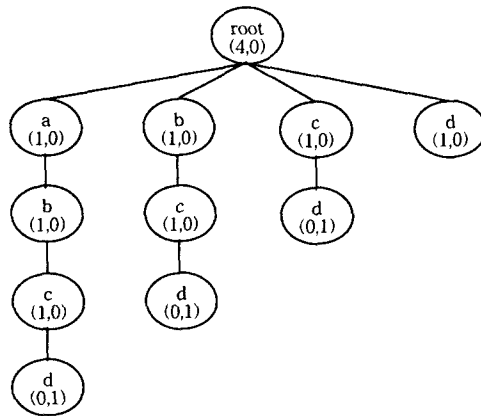
2. 동적인 나무에 의한 열 패턴 분석

순서에 의존하지 않는 J 개 원소로 구성된 집합 $R = \{r_1, r_2, \dots, r_J\}$ 을 고려하기로 한다. 또한 집합 R 의 원소들로 구성되고 순서가 부여된 열(sequence) $S = \{u_1, u_2, \dots, u_J\}$ 를 고려하기로 하자. 여기서 집합 R 은 특정 사이트(site)에서 보유하고 있는 자원을 나타내며, 열 S 는 사용자들에 의해 생성되는 W3C/WCA(World wide web committee web usage characterization activity ; <http://www.w3c.org/WCA>)에 정의한 서버세션인 URL 열을 의미한다. 이러한 열 S 로부터 열을 구성하는 각 URL들의 순서에 의해 생성되는 점프(jump)를 허용하지 않는 모든 가능한 경로(path)들에 의한 집합 $P = \{p_k; k=1, 2, \dots, K\}$ 를 정의하기로 한다. 집합 P 에는 길이 1인 경로가 J 개 존재하며, 길이 2인 $J-1$ 개 경로 그리고 길이 3인 $J-2$ 개 경로 등이 포함되며 길이 J 인 경로는 한 개뿐이므로 집합 P 는 $J^2/2$ 개 원소로 구성된다.

점프를 허용하는 주어진 열 S 의 모든 가능한 경로들의 집합 P 는 Rymon(1992)이 제안한 나무 (tree)로 표현할 수 있다. 예를 들어, $S = \{a, b, c, d\}$ 라고 하면 열 S 로부터 생성되는 모든 가능한 경로들에 대한 Rymon의 나무는 <그림 1>과 같다.



<그림 1> $S = \{a, b, c, d\}$ 에 대한 Rymon의 나무



<그림 2> 경로 도수를 고려한 나무

<그림 1>은 순서에 의존하는 4개 원소들로 구성된 열에서 생성할 수 있는 모든 가능한 경로를 나무로 나타내고 있으며, 뿌리노드(root node)를 포함한 16개 노드들로 구성되어 있는 것을 알 수 있다. 그러나 만일 점프를 허용하지 않는다면 <그림 1>에서 원안에 포함된 점프를 허용하는 경로들을 제외한 11개 노드들로 구성된 나무에 의해 모든 가능한 경로가 정리될 수 있을 것이다. 따라서 점프를 허용하지 않는 순서가 부여된 열로부터 발생 가능한 모든 경로들은 Rymon의 나무를 수정하여 나타낼 수 있다. 그러나 이러한 나무는 주어진 열에서 순서가 부여된 발생 가능한 모든 경로들의 집합을 나타내기 위한 방법으로 나무에 속한 각 경로들이 나타난 횟수에 대한 정보를 포함하지 않는다. 따라서 나무를 구성하는 각 노드마다 현재노드까지 이르는 경로들의 도수를 부여한 <그림 2>와 같은 수정된 나무를 고려하기로 한다.

<그림 2>의 나무는 Rymon의 나무와는 달리 노드 이름에 현 노드까지의 경로전체를 표시하지 않고 경로의 마지막 노드만을 표현하였다. 노드 이름 아래 괄호 안의 값 중에서 왼쪽 값은 해당 노드까지의 경로 도수 그리고 오른쪽 값은 해당 노드가 경로의 잎 노드(leaf node)인 경우의 도수를 나타낸다. 예를 들어, 제일 왼쪽 가지(branch)의 노드 c 는 도수 쌍이 $(1,0)$ 이므로 주어진 열 S 에서 발생 가능한 모든 경로들 중에서 경로 $a \rightarrow b \rightarrow c$ 는 한번 나타났으며 노드 c 는 잎이 아닌 것을 알 수 있다. 또한 뿌리노드의 도수는 열 S 에 속한 총 원소의 수를 나타내며, 뿌리노드의 자손노

드(child node)들의 도수 쌍의 합이 4인 것을 알 수 있다. 즉, 열 S 의 길이 $|S|$ 가 4인 것을 식별하게 해준다.

```

•  $S \leftarrow S_i, i=1$ 
  •  $j=1$ 
    •  $n \leftarrow u_0$ , 여기서  $u_0$ 는 뿌리
    •  $k=j$ 
      • 만일  $n$ 이 잎이 아니고  $(count(n) + leaf(n)) > threshold$ 이면
        •  $n \leftarrow u_k$ 인 노드 생성,  $u_k \in S$ 
      • 만일  $n$ 의 자손  $c$ 가 존재하면
        •  $n \leftarrow c$ 
      •  $J$ 에 이를 때까지  $k = k+1$ 
      •  $J$ 에 이를 때까지  $j = j+1$ 
      • 경로 도수 쌍 갱신
    •  $i = i+1$ 
  
```

<그림 3> 열 패턴 학습 알고리즘

제안된 나무는 Wang(1997)이 정의한 것과 같은 열 패턴을 얻기 위한 각 경로에 대한 지지도(support)와 신뢰도(confidence)를 나무로부터 직접 얻을 수 있다. 즉, 노드 a 의 도수 쌍의 합은 1이므로 노드 a 의 지지도 $sup(a) = 1/|S| = 1/4$ 인 것을 나무에서 직접 식별할 수 있고, 노드 a 로부터 시작되는 경로 $a \rightarrow b$ 의 지지도 $sup(a \rightarrow b) = 1/4$ 를 얻을 수 있다. 따라서 노드 a 에서 노드 b 에 이르는 열 패턴의 신뢰도 $conf(a \rightarrow b) = sup(a)/sup(a \rightarrow b) = 1$ 인 것 또한 쉽게 얻을 수 있다. 결국 제안된 나무는 노드에 경로 도수에 관한 정보를 갖게 함으로써 주어진 열에서 생성 가능한 모든 열 패턴에 관한 정보를 쉽게 얻을 수 있게 해 준다.

이제 서버세션을 나타내는 URL열들에 의한 순서가 고려된 무한 열 S_1, S_2, \dots 를 고려하기로 한다. 이들 무한 열에 의해 생성될 수 있는 점프를 허용하지 않는 모든 가능한 경로들을 정리한 제안된 나무를 생성하기 위하여 <그림 3>과 같은 알고리즘을 고려하기로 한다.

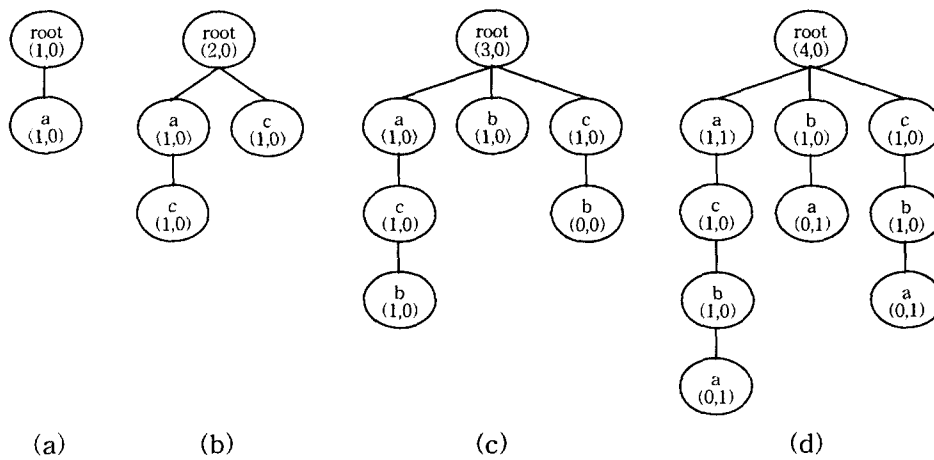
<그림 3>의 알고리즘에서 n 은 노드를 $count(n)$ 은 현재 노드에 이르는 경로의 도수, $leaf(n)$ 은 잎 노드에 이르는 경로의 도수를 나타낸다. 그리고 $threshold$ 는 현재노드의 자손노드를 생성하기 위한 기준값을 나타낸다. 즉, 만일 $threshold$ 가 '1'이라면 주어진 열의 모든 경로가 나무안에 포함되는 것을 의미한다. 따라서 $threshold$ 가 커지면 자손노드의 생성이 주어진 $threshold$ 에 이를 때까지 멈추게 됨을 의미한다. 이때 노도도수는 현재 노드에 이르기까지의 경로의 지지도를 나타내므로 $threshold$ 의 결정은 열 패턴을 찾기 위한 최소지지도에 의해 결정될 수 있다. 그러므로 주어진 알고리즘에 의해 생성되는 나무는 가지치기에 의해 열 패턴을 식별할 수 있을 뿐만 아니라 주어진 열 패턴에 의해 나무를 성장시키는 방법으로 열 패턴을 식별 할 수 있다. 마지막으로 알고리즘에서 "경로 도수 쌍 갱신"단계는 현재 학습에서 사용된 열에 포함된 모든 경로들에 의한 도수를 이전 단계까지 학습된 나무의 도수를 추가하여 경로 도수를 새롭게 얻는 단계를 의미한다. 이 단계는 이전 단계의 반복과 같으나 단지 새로운 노드를 추가하지 않는다는 차이만을 가진다.

위 알고리즘을 쉽게 설명하기 위하여 간단한 예를 들기로 한다. 이를 위해 집합 $R = \{a, b, c\}$

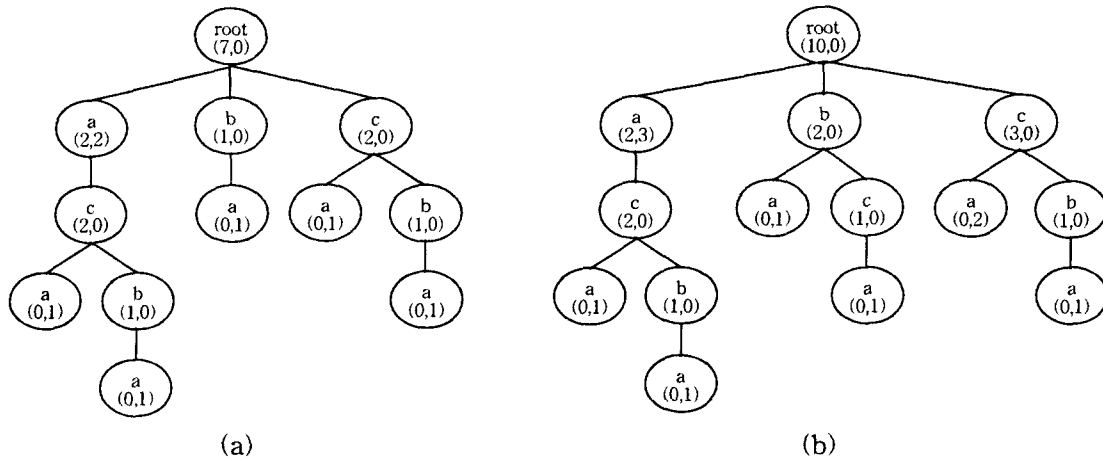
에서 발생 가능한 3개 열 $S_1 = \{a, c, b, a\}$, $S_2 = \{a, c, a\}$, $S_3 = \{b, c, a\}$ 를 고려하기로 하자.

<표 1> 모든 가능한 경로

S_1	S_2	S_3
$a \rightarrow c$	$a \rightarrow c$	$b \rightarrow c$
$a \rightarrow c \rightarrow b$	$a \rightarrow c \rightarrow a$	$b \rightarrow c \rightarrow a$
$a \rightarrow c \rightarrow b \rightarrow a$	$c \rightarrow a$	$c \rightarrow a$
$c \rightarrow b$		
$c \rightarrow b \rightarrow a$		
$b \rightarrow a$		



<그림 4> S_1 을 위한 반복에 의해 생성된 나무



<그림 5> S_2 와 S_3 에 의한 나무

3개 열에서 생성될 수 있는 모든 가능한 경로들을 정리하면 <표 1>과 같다. 이들 모든 가능한 경로를 주어진 알고리즘에 의해 식별하기 위해 threshold를 '1'로 결정하기로 한다. 먼저, 열 S_1 의

각 원소들에 의해 순차적으로 생성되는 나무의 형태를 <그림 4>의 (a)~(d)에서 보여주고 있다. 열 S_1 에 대한 학습이 끝난 최종 나무를 나타내는 (d)에서 <표 1>에 정리된 모든 가능한 경로가 정리되어 있음을 알 수 있다. 특히, 노드 a의 도수 쌍이 (1,1)인 것은 열 S_1 에는 a가 두 번 포함 되어 있고, 그 중에 한 번은 앞에 위치 한다는 것을 알려준다. 따라서 열 S_1 에서 발견 가능한 패턴 중에서 노드 a의 신뢰도 $conf(a)=2/4$ 인 것을 쉽게 알 수 있다.

<그림 5>의 나무 (a)는 열 S_2 가 주어졌을 때 <그림 4>의 나무 (d)에 추가적으로 학습이 이루어진 나무를 보여주고 있다. 그리고 <그림 5>의 나무 (b)는 열 S_3 가 주어졌을 때 나무 (a)에 추가적으로 학습이 이루어진 나무이다. 즉, 새로운 열이 추가될 때마다 모든 열에 대하여 재학습이 이루어지는 것이 아니라 이전 열들에 의해 학습된 나무에 새로운 열에서 얻어진 패턴이 추가되는 것으로 새로운 열에 관한 패턴이 학습되고 있음을 보여준다. 따라서 제안된 나무는 새로운 열이 추가될 때마다 나무 구조가 변하는 동적인 나무가 된다.

3. 모의실험

Wang(1997)은 주어진 단일 열에서 의미 있는 패턴을 얻기 위하여 나무를 이용한 학습 알고리즘을 제안하였다. 제안된 알고리즘은 열 안의 일부가 바뀌었을 때에도 나무가 쉽게 갱신될 수 있도록 하는데 초점을 두고 있다. 그는 제안된 알고리즘의 평가를 위해 길이 2,000,000인 단일 열에 대하여 모의실험을 수행하였고, 그가 모의실험에 사용한 패턴은 <표 2>와 같다. 본 논문에서 제안된 알고리즘은 Wang의 알고리즘과 적용 대상이 다르지만 실제 자료를 얻는데 많은 제한점이 있기 때문에 비교를 위해 <표 2>의 패턴을 고려하여 모의실험을 수행하였다.

<표 6> 모의실험에 사용된 패턴

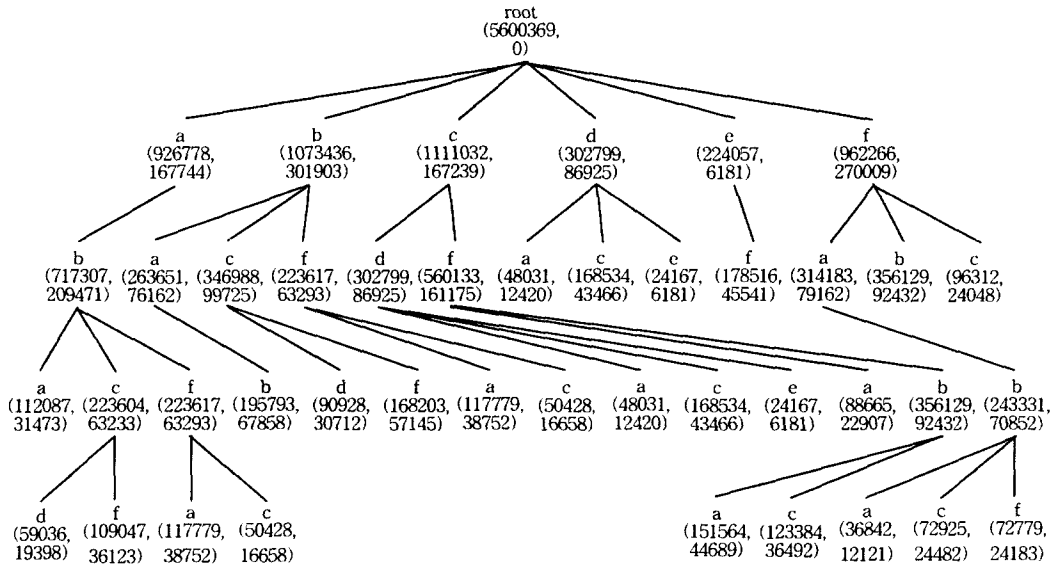
a $\xrightarrow{1}$ b	bf $\xrightarrow{0.7}$ a	cd $\xrightarrow{0.2}$ a	cf $\xrightarrow{0.8}$ b	fb $\xrightarrow{0.55}$ a
ab $\xrightarrow{0.2}$ a	bf $\xrightarrow{0.3}$ c	cd $\xrightarrow{0.7}$ c	e $\xrightarrow{1}$ f	fb $\xrightarrow{0.45}$ c
ab $\xrightarrow{0.4}$ c	c $\xrightarrow{0.35}$ d	cd $\xrightarrow{0.1}$ e	ef $\xrightarrow{0.7}$ a	
ab $\xrightarrow{0.4}$ f	c $\xrightarrow{0.65}$ f	cf $\xrightarrow{0.2}$ a	ef $\xrightarrow{0.3}$ c	

모의실험은 $R=\{a, b, c, d, e, f\}$ 으로 6개 원소로 구성되어 있으므로 각 열의 길이가 1~10인 임의의 열 S 를 <표 2>에 의해 총 1,000,000개를 생성하여 수행하였다. 모의실험에서 사용된 전체 열이 가지고 있는 모든 가능한 경로를 식별하기 위하여 threshold를 '1'로 하였고, 이 경우 학습이 완료된 후의 총 노드는 3,434개 그리고 총 열의 길이는 5,600,369 이었다.

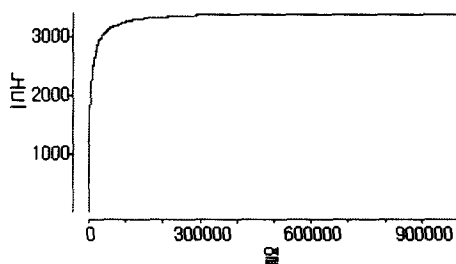
<그림 6>은 제안된 방법에 의해 생성된 동적인 나무의 깊이 3인 최종 결과를 보여주고 있다. 이 결과는 지지도 5%에서 가지치기(pruning)가 이루어진 결과로 <표 3>에 <그림 6>에서 발견된 패턴들을 정리하였다. <표 3>에서 발견된 패턴에 (*) 표시가 된 것은 <표 2>의 원 패턴을 의미한다. 즉, 최소 지지도 약 1%에서 원 패턴이 모두 발견된 것을 알 수 있다. 이러한 결과는 Wang의 결과와 매우 유사한 결과이다.

<표 3> 제안된 알고리즘에 의해 발견된 패턴

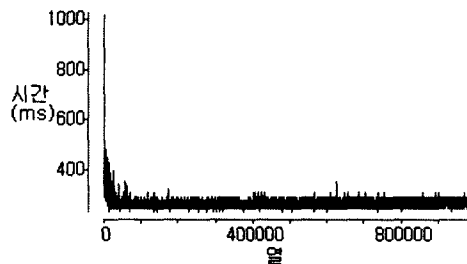
패턴	신뢰도(%)	지지도(%)	패턴	신뢰도(%)	지지도(%)
a→b (*)	1.0000	0.1655	c→f→a (*)	0.1992	0.0200
a→b→a (*)	0.2001	0.0256	c→f→b (*)	0.8008	0.0800
a→b→c (*)	0.3999	0.0512	d→a	0.1996	0.0108
a→b→f (*)	0.3999	0.0512	d→c	0.7001	0.0379
b→a	0.3166	0.0607	d→e	0.1003	0.0054
b→c	0.4162	0.0800	e→f (*)	1.0000	0.0400
b→f	0.2673	0.0512	e→f→a (*)	0.7000	0.0224
b→f→a (*)	0.7000	0.0279	e→f→c (*)	0.3000	0.0095
b→f→c (*)	0.3000	0.0200	f→a	0.4088	0.0702
c→d (*)	0.3508	0.0696	f→b	0.4662	0.0800
c→d→a (*)	0.1996	0.0108	f→c	0.1250	0.0215
c→d→c (*)	0.7001	0.0379	f→b→a (*)	0.5499	0.0350
c→d→e (*)	0.1003	0.0054	f→b→c (*)	0.4501	0.0285
c→f	0.6492	0.1288			



<그림 6> 5%에서 가지치기 된 나무



<그림 7> 노드의 변화



<그림 8> 학습시간의 변화

<그림 7>은 주어진 열의 수에 따른 노드수의 변화를 나타내고 있다. 수행된 모의실험에서는 약 100,000 열의 학습 후에 노드수의 변화가 거의 없음을 보여주고 있다. 또한 <그림 8>은 열의 수에 따른 100개 열을 학습하는데 걸리는 시간의 변화를 나타내고 있다. <그림 7>의 노드 수 변화와 비교하면 노드수가 안정되는 시점에서 학습시간이 안정되는 모습을 볼 수 있다. 참고로 1,000,000개 열 전체를 학습하는데 소요된 시간은 약 15분이었고 100개 열의 학습시간은 약 0.272초 정도가 소요되었다.

4. 결론 및 토의

본 논문에서는 특정 사이트 내에서 발생하는 웹 사용자들의 네비게이션 패턴을 분석하기 위한 나무에 기반한 동적인 알고리즘을 제안하였다. 사용자들의 네비게이션 경로는 순서가 부여된 URL들의 열로 구성되며 다수 사용자들의 네비게이션 경로들은 각각 순서가 부여된 URL 열들의 집합이 된다. 이러한 열들의 집합은 웹 로그파일에서 식별되며 HTTP(Hyper Text Transfer Protocol)의 특성상 웹 로그파일로부터 정확한 사용경로를 식별하기 위해서는 Cooley등(1999)이 제안한 방법을 이용할 수 있다.

로그파일로부터 식별된 이들 열들의 집합으로부터 사용자들의 네비게이션 패턴을 분석하기 위해서 Mobasher등(1996)이 제안한 순서가 부여된 Apriori 알고리즘을 확장한 방법을 고려할 수 있다. 그러나 이 방법은 Apriori 알고리즘의 특성상 candidate set에 포함된 각 규칙의 신뢰도와 지지도를 얻기 위해 매번 첫 열부터 마지막 열까지 다시 계산을 수행하여야 하며, 이것은 분석에 사용될 총 열들의 수가 제한되어야 함을 의미한다. 또한 Schechter등(1998)이 제안한 나무기반 알고리즘 역시 주어진 열들에 포함된 모든 가능한 경로를 나무의 가지로 표현하기 위해서는 분석에 사용된 전체 열들에 대하여 나무구조가 수렴될 때까지 반복계산을 수행하여야 한다. 따라서 새로운 열이 추가되면 새로운 열이 포함된 전체 자료에 대하여 새롭게 나무를 생성하여야 한다.

반면에 제안된 방법은 첫 열에 의해 생성된 나무에 열들이 순차적으로 추가될 때마다 이전 나무에 해당 열에 의한 정보가 추가되어 갱신되는 동적인 방법에 의해 열 패턴을 찾기 때문에 이미 나무에 포함된 열은 다시 계산에 이용되지 않는다. 즉, 분석에 이용될 총 열들의 수에 제한을 받지 않는 특징을 갖는다. 따라서 제안된 방법은 총 열들의 수가 제한된 경우의 열 패턴을 찾기 위해 이용될 수 있을 뿐만 아니라 새로운 열들이 지속적으로 추가되는 경우에 더욱 유용하게 사용될 수 있다. 보다 구체적인 열 패턴 분석의 응용사례들은 Wang(1997) 그리고 웹 사용 마이닝의 목적 중에 하나인 개인화에 관한 응용사례는 Mobasher등(1999)과 Srivastava등(2000)을 참고할 수 있다.

References

- [1] Agrawal, R., Imielinski, T., and Swami, A. (1993). Mining Associations between Sets of Items in Massive Databases, *Proceedings of the 1993 ACM-SIGMOD International Conference. on Management of Data*, 207-216.
- [2] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A. I. (1996). Fast Discovery of Association Rules, *Advances in Knowledge Discovery and Data*

- Mining*, AAAI Press, 307-328.
- [3] Cooley, R., Mobasher, B., and Srivastava, J. (1999). Data Preparation for Mining World Wide Web Browsing Patterns, *Journal of Knowledge and Information Systems*, Vol. 1, No. 1, 1999.
 - [4] Han, E. H., Karypis, G., Kuram, V., and Mobasher, B. (1997). Clustering Based On Association Rule Hypergraphs, *SIGMOD'97 Workshop on Research Issues on Data Mining and Knowledge Discovery*.
 - [5] Mobasher, B., Cooley R., and Srivastava, J. (1999). Automatic Personalization Through Web Usage Mining, *Technical Report TR99-010*, Department of Computer Science, Depaul University, 1999.
 - [6] Mobasher, B., Jain, N, Han, E. H., and Srivastava, J. (1996). Web Mining : Pattern Discovery from World Wide Web Transactions, *Technical Report TR96-050*, Department of Computer Science, University of Minnesota.
 - [7] Rymon, R. (1992). Search through Systematic Set Enumeration, *Proceedings of Third International Conference of Principles of Knowledge Representation and Reasoning*, 539-550.
 - [8] Schechter, S., Krishnan, M., and Smith, M. D. (1998). Using Path Profiles to Predict HTTP Requests, *WWW7/Computer Networks*, Vol. 30. 457-467.
 - [9] Srivastava, J., Cooley, R., Deshpande, M., and Tan, P. N. (2000). Web Usage Mining : Discovery and Applications of Usage Patterns from Web Data, *SIGKDD Explorations*, Vol. 1, 12-23.
 - [10] Wang, Ke. (1997). Discovering Patterns from Large and Dynamic Sequential Data, *Journal of Intelligent Information Systems*, Vol 9. 33-56.