웹 클릭 스트림의 효율적 분석을 위한 시간 간격 제한을 활용한 관심 순차패턴 탐색[†]

(Mining Interesting Sequential Pattern with a Time-interval Constraint for Efficient Analyzing a Web-Click Stream)

장 중 혁*

(Joong-Hyuk Chang)

요약 웹 관련 기술의 발달 및 스마트폰과 같은 지능형 모바일 서비스 기기의 사용 증가로 인해 오늘날 많은 분야에서 다양한 웹기반 서비스들이 널리 활용되고 있다. 이러한 환경에서 개인화 및 지능화된 웹 서비스를 제공하기 위한 연구들이 활발히 진행되고 있으며, 웹 서비스 이용 기록으로부터 생성되는 웹 클릭 스트림에 대한 분석 기술은 관련 기술 중 핵심 기술의 하나이다. 본 논문에서는 순차정보 형태로 발생하는 웹 클릭 스트림에 대한 효율적 분석을 위해서 데이터 스트림 처리에 대한 기본적인 요구사항을 만족하면서 정제된 결과를 얻기 위한 순차패턴 마이닝 방법을 제시한다. 이를 위해서 먼저 순차패턴에 포함되는 단위항목들의 단순 발생 순서뿐만 아니라 발생 시간 정보를 추가로 활용하는 시간 간격 제한 관심 순차패턴을 정의하고, 이어서 웹 클릭 스트림과 같은 데이터 스트림에서 이를 효율적으로 탐색하기 위한 마이닝 방법을 제안한다. 해당 연구 결과는 웹 클릭 스트림뿐만 아니라 전자상거래, 생물정보학 및 USN 환경 등과 같이 데이터 스트림 형태로 정보를 발생시키는 여러 컴퓨터 응용 분야에서 유용하게 활용될 수 있을 것이다.

핵심주제어 : 웹 클릭 스트림, 시간 간격 제한, 순차패턴, 순차정보 데이터 스트림

Abstract Due to the development of web technologies and the increasing use of smart devices such as smart phone, in recent various web services are widely used in many application fields. In this environment, the topic of supporting personalized and intelligent web services have been actively researched, and an analysis technique on a web-click stream generated from web usage logs is one of the essential techniques related to the topic. In this paper, for efficient analyzing a web-click stream of sequences, a sequential pattern mining technique is proposed, which satisfies the basic requirements for data stream processing and finds a refined mining result. For this purpose, a concept of interesting sequential patterns with a time-interval constraint is defined, which uses not only the order of items in a sequential pattern but also their generation times. In addition, A mining method to find the interesting sequential patterns efficiently over a data stream such as a web-click stream is proposed. The proposed method can be effectively used to various computing application fields such as E-commerce, bio-informatics, and USN environments, which generate data as a form of data streams.

Key Words : Web-click streams, Time-interval constraint, Sequential patterns, Sequencae data streams

1. 서 론

인터넷과 웹 관련 기술의 발달로 인해 오늘날 많은 분야에서 다양한 웹기반 서비스들이 활발히 제공되고 있고, 스마트폰과 같은 지능형 모바일 서비스 기기의 사용 증가는 이러한 현상을 더욱 확대시키고 있다. 이에 따라 근래 들어 웹에서 관리되는 데이터 및 웹 사용자의 행동 패턴 등을 효과적으로 분석하기 위한 웹 데이터 마이닝에 대한 관심이 증가되어 왔으며, 웹 데이터 마이닝을 통해 얻어진 다양한 분석 결과를 활용하여 개인화 및 지능화된 웹 기반 서비스를 제공하기 위한 연구들도 활발히 진행되어 왔다[1-3].

대표적인 웹 데이터 마이닝 분야 중의 하나인 웹 사용 기록(web usage logs) 분석은 웹 사이트에 접근하는 사용자들의 사용 기록을 분석하여 흥미로운 사용 패턴을 탐색하는 것으로서 일반적으로 웹 사용 기록으로부터 얻어진 웹 클릭 스트림이 데이터 마이닝의 처리 대상이 된다. 하나의 웹 사이트에서 발생하는 웹 클릭 스트림에 대한 분석을 통해 해당 웹 사이트를 이용하는 사용자의 개인적인 선호도 및 이용 성향 등을 반영한 개인화 및 지능화 서비스를 제공할 수 있다. 한편, 웹 클릭 스트림은 구성요소가 지속적으로 확장된다. 즉, 다수의 사용자들이 하나의 웹 사이트에 접근함에 따라 해당 웹 클릭 스트림을 구성하는 단위 구성 요소 및 이들의 조합으로 이뤄진 순차정보 또는 트랜잭션(transaction) 정보 등이 지속적으로 발생된다. 이러한 특성으로 인해 이전의 한정적인 데이터 집합에 대한 분석 및 데이터 마이닝 기술로는 웹 클릭 스트림을 효율적으로 분석하는데 많은 어려움이 있으며, 이를 보완하기 위한 방법으로 구성 요소가 지속적으로 생성되는 데이터 스트림에 대한 효율적 분석 기술들[4,5]이 제안되어 왔다.

순차패턴 마이닝(sequential pattern mining)은 순차 데이터 집합으로부터 흥미로운 순차패턴(sequential pattern)을 탐색하는데 목적을 두고 있으며, 이는 주요한 데이터 마이닝 분야 중의 하나로서 다양한 컴퓨터 응용 분야에서 널리 활용되고 있다. 순차패턴 탐색은 분석 대상 데이터 집합 및 출현빈도 수 임계값이 주

어졌을 때 해당 임계값 이상의 출현빈도 수를 갖는 모든 순차패턴을 찾는 작업이다. 일반적으로 마이닝 수행 결과로 얻어지는 순차패턴의 수가 매우 많으며, 이를 바로 응용 분야의 특성을 이해하기 위해서 활용하는데 어려움이 있다. 따라서 관심도가 큰 순차패턴을 얻기 위해서는 일반적인 순차패턴 탐색으로 얻어진 결과를 다시 분석하는 등 후처리 작업을 필요로 한다.

이러한 문제를 보완하고 보다 정제된 형태의 순차패턴 마이닝 결과를 얻기 위하여 순차패턴 탐색 과정에서 제한 조건 등을 활용하는 방법들[6-10]이 활발히 연구되어 왔다. 일반적으로 순차패턴 마이닝에서는 단위항목(atomic value)의 단순 발생 순서만을 고려하여 마이닝 결과를 구한다. 하지만 하나의 순차정보 또는 순차패턴에 있어서 이를 구성하는 단위항목의 시간 간격도 매우 중요한 의미를 갖는다. 예를 들어 인터넷 쇼핑몰 사이트에서 접근한 다음의 두 사용자의 접근 기록을 고려해 보자.

User_1 :

홈 (2분후) 컴퓨터 (30분후) 가전제품 (1분후) 가구

User_2 :

홈 (1분후) 컴퓨터 (2분후) 가전제품 (1분후) 가구

두 사용자는 서로 동일한 범주의 제품들에 관심을 갖고 해당 웹페이지를 순차적으로 접근하였다. 하지만, 각 웹페이지를 접근한 시점은 큰 차이를 보인다. 즉, User_1의 경우 컴퓨터 관련 웹페이지에 접근한 다음 30분 경과 후 가전제품 관련 웹페이지에 접근하였다. 반면 User_2의 경우 컴퓨터 관련 웹페이지에 접근한 다음 2분 경과 후 가전제품 관련 웹페이지에 접근하였다. 두 경우에 있어서 접근된 웹페이지의 단순 순서만을 고려하는 경우 서로 동일한 것으로 간주되나 웹페이지 접근에 대한 시간 간격을 고려하는 중요성 혹은 관심도가 서로 다른 것으로 간주할 수 있다. 일반적으로 짧은 시간 내에 접근된 웹페이지들이 서로 연관성이 높은 것으로 간주할 수 있기 때문이다. 이러한 상황에서 순차패턴 마이닝 과정에서 단위항목의 발생 시간 간격을 고려하여 최대 허용시간간격 또는 최소 허용시간간격 등을 적용하는 경우 일반적인 순차패턴 탐색에 비해 보다 정제된 마이닝 결과를 얻을 수 있

* 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2009-0064582)

* 대구대학교 컴퓨터IT공학부

다.

본 논문에서는 웹 클릭 스트림에 대한 순차패턴 마이닝에서 관심도가 높은 순차패턴을 얻기 위한 접근법을 제안한다. 즉, 지속적으로 확장되는 웹 클릭 스트림의 특성을 고려하여 데이터 스트림 처리의 기본적인 요구 조건들[11]을 만족하면서 시간 간격에 기반한 제한 조건 설정을 통해 관심도나 흥미도가 큰 순차패턴을 얻을 수 있는 데이터 스트림에 대한 순차패턴 탐색 방법을 제안하며, 해당 순차패턴을 얻기 위해서 순차패턴 마이닝 과정에서 순차패턴을 구성하는 단위항목들간의 시간 간격을 일정 정도로 제한하는 방법을 활용한다. 이를 통해서 웹 클릭 스트림에 대한 순차패턴 탐색 과정에서 메모리 사용량 및 수행 시간 최소화 등과 같은 데이터 스트림 처리를 위한 기본적인 요구 조건을 만족하면서 관심도나 흥미도가 큰 순차패턴을 효율적으로 얻을 수 있다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 웹 클릭 스트림에 대해서 기술하고, 본 논문에서 다루고자 하는 문제를 명확히 정의한다. 이와 연관된 이전의 연구 및 데이터 스트림에서 순차패턴 탐색을 위한 기본적인 접근 방법은 3장에서 기술되며, 4장에서는 본 논문에서 제안하는 주된 내용인 시간 간격 제한 기반 순차패턴의 기본 개념을 기술하고, 웹 클릭 스트림에서 이를 효율적으로 탐색할 수 있는 마이닝 방법을 제시한다. 5장에서는 웹 로그 데이터를 활용한 실험을 통해 제안된 방법의 성능을 검증하고, 6장에서 논문의 결론을 맺는다.

2. 기본 정리

본 절에서는 논문에서 제안하는 시간 간격 제한을 활용한 빈발 순차패턴 탐색의 처리 대상이 되는 웹 클릭 스트림에 대해서 기술한다. 이를 위해서 먼저 순차정보 데이터 스트림을 명확히 기술하고, 더불어 하나의 웹 사이트에서 발생하는 사용 기록으로부터 순차정보 웹 클릭 스트림을 정의하는 방법을 간략히 기술한다. 이어서 논문에서 다루고자 하는 문제인 시간 간격 제한을 활용한 순차패턴 마이닝을 명확히 정의한다.

2.1 순차정보 데이터 스트림 및 웹 클릭 스트림

순차패턴 탐색을 위한 분석 대상이 되는 순차정보 데이터 스트림은 시간적으로 정렬된 단위항목들의 집합인 순차정보가 지속적으로 생성되는 무한집합으로 간주할 수 있으며 다음과 같이 정의된다[5]. **단위항목 집합** $I = \{i_1, i_2, \dots, i_n\}$ 는 데이터 스트림에서 현재까지 생성된 단위항목들의 집합으로서 응용 분야에서 발생한 개별 정보를 의미한다. 하나의 **순차패턴**(sequential pattern) s 는 단위항목들이 순차적으로 정렬된 리스트(ordered list)로서 $\langle e_1 e_2 \dots e_l \rangle$ 와 같이 나타낸다. 여기서 $e_j (1 \leq j \leq l)$ 는 단위항목을 나타낸다. 즉, $e_j \in I (1 \leq j \leq l)$ 관계를 만족하며, 일반적으로 순차패턴의 구성요소라 지칭하기도 한다. 하나의 순차패턴 s 에 대해서 **순차패턴의 길이** $|s|$ 는 해당 순차패턴을 구성하는 단위항목들의 개수를 의미하며, n 개의 단위항목들로 구성된 순차패턴은 **n -순차패턴**이라 지칭한다. 한편, 순차패턴 $s_1 = \langle a_1 a_2 \dots a_n \rangle$ 와 $s_2 = \langle b_1 b_2 \dots b_m \rangle$ 에 있어서 $a_1 = b_{j_1}, a_2 = b_{j_2}, \dots, a_n = b_{j_n}$ 관계를 만족하면서 $1 \leq j_1 < j_2 < \dots < j_n \leq m$ 관계를 만족하는 정수 $j_1, j_2, \dots, j_n$ 이 존재할 때, s_1 은 s_2 의 **부분-순차패턴**(sub sequential pattern)이라 하고 s_2 는 s_1 의 **확대-순차패턴**(super sequential pattern)이라 한다. 즉, $s_1 \subseteq s_2$ 관계를 만족하며, s_1 은 s_2 에 포함된다고 하고 s_2 는 s_1 을 포함한다고 표현한다. **순차정보**(sequence) S 는 하나 이상의 단위항목들이 정렬된 집합으로서 각 순차정보는 다른 순차정보와 구별되는 **식별자**(sequence identifier) SID 를 갖는다. 본 논문에서는 하나의 데이터 스트림에서 k 번째 생성된 순차정보를 S_k 로 나타내며, 이때 해당 순차정보의 식별자 SID 는 k 가 된다. 순차정보 S_k 가 새롭게 생성되었을 때, **현재 데이터 스트림** D_k 는 현재까지 생성된 모든 순차정보들로 구성된다. 즉, $D_k = \langle S_1, S_2, \dots, S_k \rangle$ 이며, 해당 데이터 스트림에 포함된 **순차정보의 총 수**를 $|D|_k$ 로 나타낸다.

한편, 하나의 웹 사이트에서 생성된 사용 기록으로부터 얻어지는 순차정보 데이터 스트림 형태의 웹 클릭 스트림은 다음과 같이 정의될 수 있다. 해당 웹사이트를 구성하는 각 웹페이지를 단위항목으로 간주하고 연속적으로 접근된 웹 페이지들을 하나의 순차정보로 구성한다. 이때 일정 시간 동안 사용자 입력이

없는 경우 하나의 순차정보가 생성된 것으로 간주하며, 또한 사용자가 접속을 종료한 경우에도 동일하게 간주한다. 이러한 방법을 통해 하나의 웹 사이트에서 접근하는 여러 사용자의 지속적인 접근 기록을 순차정보 데이터 스트림 형태를 띄는 웹 클릭 스트림으로 변환할 수 있다. 즉, 웹 클릭 스트림은 하나의 웹 사이트에 대한 접근 기록으로부터 생성된 것으로서 앞서 정의된 순차정보 데이터 스트림의 형태를 갖는다. 따라서 논문의 나머지 부분에서 순차정보 데이터 스트림이라 함은 웹 클릭 스트림을 지칭한다.

2.2 문제 정의

하나의 순차정보 데이터 스트림 D_k 에 대한 일반적인 순차패턴 마이닝에서는 해당 데이터 스트림에서 순차패턴 s 가 발생되었을 때 s 의 출현빈도 수 $C_k(s)$ 는 D_k 에 포함되는 k 개의 순차정보들 중에서 해당 순차패턴 s 를 포함하는 순차정보의 수를 나타낸다. 마찬가지로 해당 시점에서 순차패턴 s 의 지지도 $S_k(s)$ 는 D_k 에 포함되는 총 순차정보의 수에 대한 출현빈도 수 $C_k(s)$ 의 비율로 정의된다. 따라서 데이터 스트림 D_k 에 대한 일반적인 순차패턴 탐색에서 사전에 주어진 최소 지지도 S_{min} 보다 크거나 같은 지지도를 갖는 순차패턴은 해당 시점에서의 빈발 순차패턴으로 정의된다[5,12]. 하지만 본 논문에서 제안하는 시간 간격 제한을 활용하는 순차패턴 탐색 문제에서는 순차패턴의 출현빈도 수에 대한 정의가 시간 간격 제한을 적용하여 다음과 같이 변경된다. **최소지지도 S_{min} 과 최대 허용시간간격 T_{max}** 가 주어졌을 때, 순차패턴 s 의 시간 간격 제한을 적용한 출현빈도 수는 최대 시간 간격이 T_{max} 보다 작거나 같으면서 해당 순차패턴과 동일한 순차패턴을 부분 패턴으로 포함하는 순차정보의 수를 나타낸다. 지지도 및 빈발 순차패턴에 대한 정의는 일반적인 데이터 스트림에서의 순차패턴 탐색과 동일하다. 결론적으로 본 논문에서 다루고자 하는 문제는 하나의 순차정보 데이터 스트림에 대해서 최소지지도 S_{min} 과 최대 허용시간간격 T_{max} 가 주어졌을 때, 해당 조건을 만족하는 모든 빈발 순차패턴을 탐색하는 것이다.

3. 관련 연구

순차패턴 마이닝에서 보다 정제된 마이닝 결과를 얻기 위한 다양한 연구들이 활발히 진행되어 왔다. 대표적인 연구에는 순차패턴 탐색 과정에서 일정한 제한 조건을 제시하여 관심도가 큰 순차패턴을 마이닝 결과로 얻기 위한 연구[7,8], 하나의 순차정보에서 인접한 두 단위항목들 간의 발생 시간 차이 정보를 활용하여 관심도가 보다 높은 순차패턴을 찾는 연구[13,14] 및 closed 순차패턴이나 maximal 순차패턴 탐색에 대한 연구[9,10] 등이 있다. [7] 및 [8]에서 단위항목들의 발생 시간이나 발생 간격을 순차패턴 탐색 시 제한 조건으로 활용하여 정제된 형태의 마이닝 결과를 얻고자 시도하였다. [13] 및 [14]에서는 발생 시간 정보를 갖는 순차 데이터 스트림에서 하나의 순차정보에서 인접한 두 단위항목들 간의 발생 시간 차이 정보를 활용하여 관심도가 높은 순차패턴을 찾기 위한 마이닝 방법이 제안되었다. [9] 및 [10] 등에서 연구된 closed 순차패턴이나 maximal 순차패턴 탐색은 부분-순차패턴 및 확대-순차패턴 관계로 서로 포함관계에 있는 두 순차패턴에 대해서 이들의 지지도를 비교하여 보다 관심도가 큰 순차패턴을 탐색하기 위한 방법들이 제안되었다. 하지만 이들 연구들은 마이닝 수행 과정에서의 분석 대상 데이터 집합에 대한 탐색 횟수, 메모리 사용량 및 처리 시간 등의 문제로 데이터 스트림 환경에서는 효과적으로 적용되는데 어려움이 있다.

데이터 스트림에 대한 순차패턴 탐색 방법으로 제안된 eSeq[5] 방법은 처리 대상이 되는 순차정보 데이터 스트림에서 발생하는 순차패턴들중 일정 기준 이상의 지지도를 갖는 중요 순차패턴만의 출현빈도수를 메모리 상에 유지되는 순차패턴의 출현빈도 수 관리 구조인 모니터링 트리에서 관리함으로써 마이닝 수행 과정에서의 메모리 사용량을 최소화하고 마이닝 수행 시간도 단축한다. 해당 방법은 매개변수 갱신 과정, 출현빈도수 갱신 과정, 순차패턴 추가 과정, 빈발 순차패턴 탐색 과정 및 강제 전지 과정으로 구성된다.

메모리 사용량 감소 및 수행 시간 단축을 위한 핵심 과정인 순차패턴 추가 과정에서는 하나의 데이터 스트림 D_k 에 순차패턴 S_k 가 생성되었을 때 모니터링 트리의 각 노드 중에서 하나의 순차정보 S_k 에서 발생된 순차패턴과 연관된 노드에 대한 출현빈도 수 갱신 작업이 종료된 후 S_k 에 출현하였으나 모니터링 트리에 관리되고 있지 않는 중요 순차패턴을 모니터링 트리

에 추가하기 위한 작업을 수행한다. 이를 위해서 S_k 에서 비중요 단위항목을 제외하고 정제된 순차정보 $\bar{S}_k$ 를 구한다. 하나의 순차패턴이 비중요 단위항목을 하나 이상 포함하는 경우 해당 순차패턴은 중요 순차패턴이 될 수 없기 때문에 비중요 단위항목들은 미리 제거하고 정제된 순차정보를 이용하여 새로운 중요 순차패턴 발생 여부를 판단한다. 정제된 순차정보 $\bar{S}_k$ 를 얻기 위해서는 모니터링 트리를 탐색하여 순차정보 S_k 를 구성하는 각 단위항목의 출현빈도 수를 확인하고 중요 지지도 S_{sig} 보다 작은 지지도를 갖는 단위항목을 제거한다. 이때 모니터링 트리에 관리되고 있지 않는 새로운 단위항목은 이를 나타내는 노드를 모니터링 트리에 추가한다. 정제된 순차정보 $\bar{S}_k$ 가 구해지면 이를 이용하여 모니터링 트리를 한 번 더 탐색하면서 S_k 에 새롭게 발생한 중요 순차패턴을 찾고 이를 나타내는 연관 노드를 모니터링 트리에 추가한다. 추가된 노드의 각 구성요소(element) 정보는 [5]에서 기술된 바와 같은 방법으로 값이 계산되어 초기화된다. 한편, 각 시점에서 최신의 마이닝 결과 집합을 얻기 위한 과정인 순차패턴 탐색 과정에서는 전위트리를 기반으로 하는 일반적인 데이터 마이닝 방법에서와 동일한 과정으로 모니터링 트리의 각 노드를 탐색하여 해당 노드의 지지도가 최소 지지도 S_{min} 보다 크거나 같은 경우 해당 노드가 나타내는 순차패턴은 빈발 순차패턴으로 구해진다.

4. 시간 간격 제한 관심 순차패턴 탐색

기존의 일반적인 순차패턴과 달리 시간 간격 제한을 활용한 관심 순차패턴은 순차패턴 탐색시 순차패턴에 포함되는 단위항목들의 시간 정보를 고려한다. 즉, 단위항목들간의 시간 간격이 일정 값 이하인 경우에만 관심 순차패턴으로 간주하고, 이보다 큰 시간 간격을 갖는 순차패턴은 관심대상에서 제외된다. 이러한 내용을 보다 명확히 정의하기 위해서 본 절에서는 시간 간격 제한을 활용한 빈발 순차패턴의 기본 개념을 정의하고 이를 탐색하기 위한 데이터 마이닝 방법인 *TiConS(Time-interval Constrained Sequential Pattern Mining over Data Streams)* 방법을 제시한다. 해당 방법은 데이터 스트림 처리를 위한 기본적인 요

구사항을 만족하는 것은 물론이거니와 시간 간격 제한을 활용하여 관심도나 큰 순차패턴들로 구성되는 정제된 마이닝 결과를 얻을 수 있도록 지원한다.

4.1 시간 간격 제한 관심 순차패턴의 정의

순차패턴 마이닝에서 시간 간격 제한을 활용하여 관심 순차패턴을 탐색하기 위해서는 하나의 순차패턴에서 단위항목들간의 발생 시간 차이를 활용하여 시간 간격을 명확히 정의할 필요가 있다. 순차정보 데이터 스트림에서 발생하는 하나의 순차정보 S 에 대해서 이에 포함되는 단위항목들의 발생 시간 정보 리스트를 $TSL(S)$ 라 할 때 해당 순차정보에 발생한 하나의 순차패턴 $s = \langle a_1 a_2 \dots a_l \rangle$ 는 이와 연관된 단위항목들의 발생 시간 정보 리스트 $TSL(s) = \langle t_1 t_2 \dots t_l \rangle$ 를 갖는다. 이때 해당 순차패턴에는 총 $(l-1)$ 개의 인접 단위항목 조합이 존재하며, 각 인접 단위항목 조합에 존재하는 시간 간격을 정의한다. 즉, 하나의 순차패턴에 포함되는 인접한 두 단위항목 s_i 및 s_j ($1 \leq i \leq l-1, j=i+1$) 사이의 시간 간격 TI_{ij} 는 다음과 같이 정의된다.

$$TI_{ij} = t_j - t_i$$

이러한 정의에 따라 해당 순차패턴에 존재하는 각 인접 단위항목 조합의 시간 간격을 구할 수 있으며, 이들 중 가장 큰 시간 간격을 해당 순차패턴 s 의 시간 간격 $T(s)$ 으로 정의하며, 다음과 같이 구해진다.

$$T(s) = \max(\{TI_{ij} \mid 1 \leq i \leq l-1, j=i+1\})$$

이와 같이 정의된 시간 간격을 바탕으로 순차정보 데이터 스트림에서 발생하는 순차패턴의 시간 간격 제한 출현빈도 수 및 해당 순차패턴의 지지도를 정의할 수 있다. 순차패턴의 일반적인 출현빈도 수 및 지지도에 대한 정의와 유사하나 각 순차정보에서 해당 순차패턴의 출현 여부를 판단하는데 있어서 시간 간격 제한을 고려하여 다음과 같이 정의된다. 하나의 순차 데이터 스트림 D_k 에서 발생한 하나의 순차패턴 s 에 대해서 최대 허용시간간격 T_{max} 가 주어졌을 때 해당 순차패턴의 시간 간격 제한 출현빈도 수 $TiC_k(s)$ 는 다음과 같이 정의된다.

$$TiC_k(s) = |\{s | (s \subseteq S) \wedge (T(s) \leq T_{max}) \wedge (S \subseteq D_k)\}| \quad (1)$$

따라서 D_k 에서 해당 순차패턴 s 의 시간 간격 제한 지지도 $TiS_k(s)$ 는 다음과 같이 정의된다.

$$TiS_k(s) = \frac{|\{s | (s \subseteq S) \wedge (T(s) \leq T_{max}) \wedge (S \subseteq D_k)\}|}{|D_k|} \quad (2)$$

이어서 순차패턴의 시간 간격 제한 지지도가 정의되면 이를 바탕으로 순차 데이터 스트림에 대해서 최소지지도 및 최대 허용시간간격이 주어졌을 때 시간 간격 제한 관심 순차패턴을 정의할 수 있다. 즉, 하나의 순차 데이터 스트림 D_k 에 대해서 최소지지도 S_{min} 과 최대 허용시간간격 T_{max} 가 주어졌을 때, 해당 데이터 스트림에서 발생한 하나의 순차패턴 s 의 시간 간격 제한 지지도 $TiS_k(s)$ 가 S_{min} 보다 크거나 같은 값을 가질 때 해당 순차패턴을 시간 간격 제한 관심 순차패턴이라 정의한다.

하나의 순차패턴에 대해서 해당 순차패턴의 시간 간격을 앞서 기술한 바와 같이 인접 단위항목들간의 시간 간격에서 구하는 방법 이외에도 다른 방법을 고려할 수 있다. 대표적인 예로서 하나의 순차패턴을 구성하는 첫 번째 단위항목과 마지막 단위항목간의 시간 차이를 구하여 해당 순차패턴을 대표하는 시간 간격으로 지정할 수도 있다. 이와 같이 응용 분야의 특성에 따라 순차패턴을 대표하는 시간 간격을 달리 구할 수 있으며, 이들을 적용하는 경우에도 앞서 기술한 기본적인 개념은 동일하게 적용될 수 있다. 한편, 최소 허용시간간격을 적용하여 관심 순차패턴을 고려할 수도 있다. 하지만, 일반적으로 하나의 순차패턴에서 단위항목들간의 시간 간격이 작은 경우는 중요한 의미를 갖는 것으로 간주되며 순차패턴 탐색 과정에서 이를 배제하고자 하는 경우는 극히 드물다. 따라서 본 논문에서는 최소 허용시간간격을 적용하는 경우에 대해서는 상세한 설명은 생략한다. 필요시 앞서 기술한 수식 (1)과 (2)에서 $T(s) \leq T_{max}$ 를 최소 허용시간간격에 대한 조건식으로 변경하여 적용함으로써 최소 허용시간간격을 적용한 결과를 얻을 수 있다.

4.2 시간 간격 제한 지지도의 anti-monotone 속성 및 TiConS 방법

순차패턴 마이닝에서 동일한 순차정보 데이터 집합에서 발생한 두 개의 순차패턴 s_1 과 s_2 에 대해서 s_2 가 s_1 의 확대-순차패턴($s_1 \subseteq s_2$)일 때 순차패턴 s_2 의 지지도 $S(s_2)$ 는 순차패턴 s_1 의 지지도 $S(s_1)$ 보다 항상 작거나 같음을 의미하는 순차패턴 지지도의 anti-monotone 속성은 마이닝 수행 과정에서 분석 대상 데이터 집합에 대한 탐색 횟수나 탐색 범위를 감소시키는 중요한 역할을 한다. 따라서 앞서 기술한 순차패턴의 시간 간격 제한 지지도에 있어서도 anti-monotone 속성을 만족하는 경우 시간 간격 제한 관심 순차패턴 탐색 시 마이닝 수행 속도 및 메모리 사용량 등을 크게 줄일 수 있다.

순차 데이터 스트림 D_k 에서 발생한 두 개의 순차패턴 s_1 과 s_2 에서 s_2 는 s_1 의 확대-순차패턴($s_1 \subseteq s_2$)이라고 가정하자. 이때 순차패턴 s_1 의 시간 간격 제한 지지도는 앞서 기술한 수식 (2)에 의해 다음과 같이 구해진다.

$$TiS_k(s_1) = \frac{|\{s_1 | (s_1 \subseteq S) \wedge (T(s_1) \leq T_{max}) \wedge (S \subseteq D_k)\}|}{|D_k|} \quad (3)$$

한편, $s_1 \subseteq s_2$ 인 관계를 만족하기 때문에 다음의 관계를 만족한다.

$$\frac{|\{s_1 | (s_1 \subseteq S) \wedge (T(s_1) \leq T_{max}) \wedge (S \subseteq D_k)\}|}{|\{s_2 | (s_2 \subseteq S) \wedge (T(s_2) \leq T_{max}) \wedge (S \subseteq D_k)\}|} \geq \quad (4)$$

따라서 수식 (3)과 (4)에 의해 다음의 관계가 만족함을 알 수 있다.

$$\frac{|\{s_1 | (s_1 \subseteq S) \wedge (T(s_1) \leq T_{max}) \wedge (S \subseteq D_k)\}|}{|D_k|} \geq \frac{|\{s_2 | (s_2 \subseteq S) \wedge (T(s_2) \leq T_{max}) \wedge (S \subseteq D_k)\}|}{|D_k|} \quad (5)$$

즉, $TiS_k(s_1) \geq TiS_k(s_2)$ 관계를 만족한다. 따라서, 순차패턴 s_2 의 시간 간격 제한 지지도 $TiS_k(s_2)$ 는 항상 순차패턴 s_1 의 시간 간격 제한 지지도 $TiS_k(s_1)$ 보다 작거나 같다. 즉, 시간 간격 제한 지지도가 anti-monotone 속성을 만족함을 알 수 있다.

이어서 해당 속성을 활용하며 데이터 스트림 처리의 기본 요건을 만족하면서 시간 간격 제한을 활용한 관심 순차패턴 탐색 방법인 *TiConS* 방법을 기술한다.

TiConS 방법은 3장의 관련 연구에서 기술된 데이터 스트림에서 순차패턴 탐색을 위한 기본적인 접근 방법을 기반으로 시간 간격 제한을 적용하기 위한 과정을 추가하여 구현되었다.

TiConS 방법에서는 분석 대상이 되는 순차정보 데이터 스트림에서 하나의 순차정보가 새롭게 생성되었을 때, 이를 처리하기 위해서 3장에서 기술된 일련의 과정들이 순차적으로 수행된다. 일정 시점에서 마이닝 결과를 얻기 위한 빈발 순차패턴 탐색 과정 및 마이닝 수행 과정에서 메모리 사용량을 줄이기 위한 강제 전지 작업도 필요시 또는 주기적으로 수행된다. 이때

시간 간격 제한을 활용하기 위해서는 출현빈도 수 갱신 과정 및 순차패턴 추가 과정에서 처리 대상이 되는 순차패턴의 시간 간격을 구하고 해당 시간 간격이 사전에 설정된 T_{max} 보다 큰 경우에는 처리 과정에서 더 이상 고려되지 않는다. 이를 통해서 T_{max} 보다 작은 시간 간격을 갖는 관심도가 큰 순차패턴들만 마이닝 수행 과정에서 다뤄지도록 한다. 하나의 순차 데이터 스트림 D_k 에 대해서 최소지지도 S_{min} 과 최대 허용시간 간격 T_{max} 가 주어졌을 때, 시간 간격 제한을 활용한 관심 순차패턴 탐색을 위한 TiConS 방법의 수행 과정은 <그림 1>과 같다.

Method : TiConS

입력: 순차정보 데이터 스트림 D_k , 최소지지도 S_{min} , 중요지지도 S_{sig} , 최대 허용시간간격 T_{max}

출력: 시간 간격 제한 관심 순차패턴 집합 $TiConSP_k$

```

ML = ∅; // 중요 순차패턴의 출현빈도 수 관리 구조, (cnt, cnt_r, sid, sid_r) 정보 유지
for each  $S_k$  in  $D_k$  {
     $|D_k|$  및 관련 매개변수 값 갱신;
    for each 순차패턴  $s \in S_k$  {
         $s$ 의 시간 간격, 즉  $\Pi(s)$  계산하기;
        if  $\Pi(s) \leq T_{max}$  {
            if 'ML상에서  $s$ 와 연관된 노드(cnt, cnt_r, sid, sid_r)가 존재' &&  $sid < k$  {
                cnt 및 sid 값 순차적으로 갱신;
                if (cnt /  $|D_k| < S_{sig}$ 
                     $s$ 와 연관된 노드는 ML에서 제거됨; //  $s$ 는 비중요 순차패턴
                if ( $s \subseteq R(S_k)$ ) && ( $sid_r < k$ ) //  $R(S_k)$  is a remaining-sequence of  $S_k$  [5]
                    cnt_r 및 sid_r 값 순차적으로 갱신;
            }
            if ' $s$ 와 연관된 노드(cnt, cnt_r, sid, sid_r)가 ML에 미존재' && ' $s$ 가 중요 순차패턴'
                 $s$ 와 연관된 노드(cnt, cnt_r, sid, sid_r)가 ML에 추가되고 cnt, cnt_r, sid 및 sid_r 값이
                [5]에서 기술된 바에 따라 초기화됨;
        }
    }
}
GetMiningResult(ML,  $TiConSP_k$ ); // 빈발 순차패턴 탐색, 필요시 수행
}

```

Function : GetMiningResult(ML, $TiConSP_k$) {

```

     $TiConSP_k = \emptyset$ ;
    for all ML상에 관리되고 있는 순차패턴  $s$ 
        if  $TiS_k(s) \geq S_{min}$ 
             $TiConSP_k = TiConSP_k \cup \{s\}$ ;
}

```

<그림 1> TiConS 방법

5. 실험 결과 분석

웹 클릭 스트림에 대한 분석에서 제안된 방법의 효용성을 검증하기 위해서 SD_{Web} 데이터 집합을 사용하여 검증 실험을 수행하였다. 해당 데이터 집합은 국내 유명 SNS서비스 제공 사이트중의 하나인 C

사이트의 사용자 접근 기록으로부터 생성된 것으로서, 사용자 접근 기록에 대한 원천 데이터를 가공하여 순차정보 데이터 집합을 생성하였다. 2.1절에서 기술한 바와 같이 사용자 접근 로그에서 각 웹페이지를 단위항목으로 간주하고 연속적으로 접근된 웹 페이지들을 하나의 순차정보로 구성하였으며, 일정 시

간 동안 사용자 입력이 없거나 사용자가 접속을 종료한 경우 하나의 순차정보가 생성된 것으로 간주하였다. 실험에 사용된 데이터 집합은 순차정보 구분을 위한 시간 간격을 60초로 설정하여 생성되었다. 즉, 동일 사용자가 60초 동안 다른 웹 페이지를 접근하지 않는 경우 하나의 순차정보가 종료된 것으로 간주되며, 해당 데이터 집합에서는 인접한 두 단위항목들간의 발생 시간 간격은 0이상 60초 미만의 범위에 있다. 해당 데이터 집합은 545 개의 단위항목을 가지며 1,000,000 개의 순차정보로 구성된다. 또한 가장 긴 순차정보는 30 개의 단위항목으로 구성된다.

논문에서 제안된 *TiConS* 방법 수행을 위한 매개변수들 중 S_{sig} 및 메모리 사용량 감소를 위한 강제 전지 작업 주기는 시간 간격 제한 활용에 따른 순차패턴 마이닝 방법의 성능 변화에는 영향을 미치지 않는 것으로서 본 논문의 각 실험에서 동일하게 설정하였다. 즉, S_{sig} 는 S_{min} 의 30%로 설정하고 강제 전지 작업은 1,000 개의 순차정보마다 반복하여 수행되도록 설정하였다. 한편, 웹 클릭 스트림 환경에서는 분석 대상 순차정보를 순차적으로 처리하게 되며, 이를 위해서 실험에서는 각 데이터 집합을 구성하는 순차정보를 하나씩 차례로 탐색하여 처리한다.

제안된 방법의 효용성을 평가하기 위해서 마이닝 결과로 얻어지는 빈발 순차패턴 집합에 대한 분석 실험을 수행하였다. 본 실험에서 빈발 순차패턴을 정

<그림 2> 마이닝 결과 순차패턴 수 변화
(인접 단위항목간 시간 간격 기반)

의하기 위한 최소 지지도는 0.005로 설정되었다. 먼저, <그림 2>와 <그림 3>은 마이닝 결과로 얻어지는 순차패턴의 개수 변화를 보여주며, 실험 데이터

<그림 3> 마이닝 결과 순차패턴 수 변화
(순차패턴 전체의 시간 간격 기반)

집합에 속하는 순차정보가 지속적으로 발생되는 상황에서 순차정보가 매번 100,000 개씩 처리된 시점에서의 마이닝 결과를 구하여 비교하였다. <그림 2>와 <그림 3>은 시간 간격 제한 활용 방법을 달리한 것으로서, 전자는 하나의 순차패턴에 존재하는 모든 인접항목간의 시간 간격에 대해서 제한 조건을 적용한 결과이며 후자는 하나의 순차패턴에서 최초 단위항목과 마지막 단위항목의 발생 시간 간격에 대해서 제한 조건을 적용한 결과이다. 두 그림의 결과를 보면 제한 조건 적용 방법에 따라 마이닝 결과로 얻어지는 순차패턴 개수의 감소 정도에 다소 차이가 있으나 대체로 비슷한 경향을 보인다. 즉, T_{max} 값에 따라 마이닝 결과로 얻어지는 순차패턴의 수가 크게 달라짐을 알 수 있다. 그 이유는 T_{max} 값이 작게 설정된 경우에는 순차정보들에 존재하는 순차패턴들 중에서 인접 단위항목들간의 시간 간격이 큰 순차패턴은 관심도가 낮은 순차패턴으로 간주되어 마이닝 결과로 구해지지 않기 때문이다. 한편, T_{max} 값이 증가되어 충분히 큰 값으로 설정되는 경우 집합이 시간 간격 제한을 적용하지 않는 일반적인 순차패턴 마이닝에서 얻어진 결과 집합에 매우 근접한 결과를 얻게 된다.

시간 간격 제한을 적용했을 때 얻어지는 순차패턴 집합을 보다 상세히 분석하기 위해서 결과로 얻어지는 순차패턴의 길이별 변화를 분석하였다. <그림 2> 및 <그림 3>에서 확인한 바와 두 가지의 시간 간격 정의 방법에 있어서 T_{max} 값 변화에 따른 마이닝 결과 순차패턴의 수 변화가 대체로 비슷한 경향을 보이므로 본 실험에서는 하나의 순차패턴에 존재하는 모든 인접항목간의 시간 간격에 대해서 제한 조건을

<표 1> 마이닝 결과 집합의 길이별 순차패턴 개수 분석

T_{max}	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8 이상	계
미적용	90	387	605	492	249	56	1	0	1880
10 min	90	385	605	491	249	56	1	0	1877
7 min	90	377	602	483	245	55	1	0	1853
5 min	90	368	561	458	229	48	1	0	1755
3 min	90	328	441	341	171	34	1	0	1406

적용한 실험 결과판을 제시하였으며, 논문의 나머지 실험에서도 동일하다. <표 1>은 몇가지 T_{max} 값에 대한 길이별 순차패턴의 개수를 분석한 결과를 제시하고 있다. 표에서 $L_k(k=1,2,\dots)$ 는 k 개의 단위항목으로 구성된 빈발 순차패턴을 의미한다. 하나의 단위항목으로 구성되는 L_1 의 경우는 순차패턴 내에 시간 간격이 존재하지 않으므로 모든 경우에 얻어지는 결과가 동일하다. $L_k(2 \leq k \leq 7)$ 에 있어서는 앞서 제시된 실험 결과와 마찬가지로 T_{max} 값이 증가하는 경우 마이닝 결과로 얻어지는 각 길이별 순차패턴의 수도 대체로 감소됨을 알 수 있다. 즉, 앞서 설명한 바와 같이 상대적으로 관심도가 낮은 시간 간격이 큰 순차패턴들은 제외되고 시간 간격 제한을 만족하는 관심도가 큰 순차패턴들로 구성된 정제된 마이닝 결과를 얻게 된다. 따라서 시간 간격 제한을 적용하여 마이닝 수행 단계에서 시간 간격이 큰 순차패턴들을 마이닝 결과 집합에서 제외함으로써 보다 효율적인 순차패턴 마이닝을 수행할 수 있다.

웹 클릭 스트림과 같이 구성요소가 매우 빠른 속도로 생성되며 지속적으로 확장되는 데이터 스트림에 대한 마이닝에서는 마이닝 수행 시간 및 수행 과정에서의 메모리 사용량도 중요한 이슈가 되며, 이 둘 가지 측면에서 제안된 방법의 성능을 검증하기 위한 실험을 수행하였다. 일반적으로 데이터 스트림에 대한 처리 방법에 있어서 수행 시간 측면의 유용성을 검증하기 위해서는 순차정보 등과 같은 하나의 단위 정보를 처리하는데 소요되는 시간을 측정하여 비교한다. 본 실험에서 빈발 순차패턴을 정의하기 위한 최소 지지도 등의 매개 변수는 이전의 실험에서와 동일하게 설정되었으며, 마이닝 수행시간 및 수행 과정에서의 메모리 사용량을 <그림 4>에서 보여주고 있다. 먼저 <그림 4>-(a)는 마이닝 수행 시간을 비교하여 보여주는 것으로서, 실험 데이터 집합에 포함

(a) 마이닝 수행 시간

(b) 메모리 사용량

<그림 4> 기본 성능 평가

되는 전체 순차정보를 발생 순서에 따라 200,000 개씩 5개의 구간으로 나누어 각 구간별 평균 수행 시간을 제시하고 있다. 그림에서 볼 수 있듯이 기본적으로 하나의 순차정보를 처리하기 위한 수행시간이 1 msec 미만으로 매우 작으며, 시간 간격 제한을 적용하는 경우 미적용시보다 마이닝 수행시간이 더욱 감소됨을 알 수 있다. <그림 4>-(b)는 수행 과정에서의 메모리 사용량을 측정된 결과를 보여준다. 일반적으로 메모리 사용량은 최대 사용량이 중요한 척도가 될 수 있으며, 본 실험은 앞서와 같이 5개의 구간으로 나누어 각 구간별 최대 메모리 사용량을 측정하였다. 그림에서 알 수 있듯이 순차정보가 지속적으로 발생되는 경우에도 메모리 사용량은 큰 변화 없이 일정 수준 이하로 유지되고 있으며, 특히 시간 간격 제한을 적용하는 경우 메모리 사용량이 크게 감소됨을 알 수 있다. 이러한 실험 결과를 통해 본 논문에서 제안된 방법은 웹 클릭 스트림에 대한 순차패턴 마이닝 과정에서 시간 간격 제한을 적용하여 보다 정제된 마이닝 결과를 제공할 뿐만 아니라 마이닝 수행 시간 및 메모리 사용량 측면에서도 매우 유용

한 방법임을 알 수 있다.

6. 결 론

스마트폰을 비롯한 다양한 지능형 모바일 서비스 기기의 사용 증가는 웹 서비스 이용 빈도를 급격히 증가시키고 있다. 이러한 환경에서 개인화 및 지능화된 웹 서비스를 제공하기 위한 연구들이 활발히 진행되고 있으며, 웹 서비스 이용 기록으로부터 생성되는 웹 클릭 스트림에 대한 분석도 이러한 연구들 중의 하나이다. 일반적으로 다수의 사용자들이 하나의 웹 사이트에 접근함에 따라 해당 웹 클릭 스트림을 구성하는 단위 구성 요소 및 이들의 조합으로 이뤄진 순차정보 또는 트랜잭션 정보 등이 지속적으로 발생되며, 이러한 특성으로 인해 이전의 한정적인 데이터 집합에 대한 분석 및 데이터 마이닝 기술로는 웹 클릭 스트림을 효율적으로 분석하는데 많은 어려움이 있다. 이를 보완하기 위해서 구성요소가 지속적으로 확장되는 데이터 스트림에서 효율적 순차패턴 탐색을 위한 연구들이 활발히 진행되고 있다.

웹 클릭 스트림에서 시간 간격 제한을 활용한 관심 순차패턴 탐색 방법도 이러한 연구들 중의 하나로, 순차정보에 대한 분석 과정에서 순차정보 구성 요소의 단순 발생 순서뿐만 아니라 발생 시간 간격에 대한 제한 조건을 적용하여 사용자의 관심도가 보다 큰 순차패턴을 마이닝 결과로 얻는데 목적을 두고 있다. 이를 통해 기존의 단순 순차패턴 탐색 과정에서 지나치게 많은 수의 순차패턴들이 마이닝 결과로 탐색되는데 따른 단점을 줄일 수 있다.

본 논문에서는 순차정보 데이터 스트림에 대한 순차패턴 탐색 과정에서 하나의 순차패턴을 구성하는 단위항목들의 발생 시간 정보를 활용하여 정제된 마이닝 결과 집합을 얻기 위한 방법으로 시간 간격 제한을 활용한 데이터 스트림 순차패턴 탐색 방법을 제안하였다. 해당 방법에서는 하나의 순차패턴을 구성하는 단위항목들 사이의 발생 시간 간격으로부터 해당 순차패턴을 대표하는 시간 간격을 구하고, 이에 대한 최대 허용 값을 설정함으로써 관심도가 큰 빈발 순차패턴을 탐색한다. 웹 사이트의 사용자 접근 기록으로부터 생성된 웹 클릭 스트림을 대상으로 한 실험을 통해 제안된 방법이 데이터 스트림에 대한

순차패턴 탐색 과정에 효과적으로 적용될 수 있으며 시간 간격 제한 값에 따라 다양한 형태의 정제된 순차패턴 집합을 마이닝 결과로 구해줄 수 확인하였다. 제안된 방법은 웹 클릭 스트림뿐만 아니라 데이터 스트림 형태로 정보를 발생 시키는 다양한 컴퓨터 응용 환경에서 유용하게 활용될 수 있을 것으로 판단된다.

참 고 문 헌

- [1] D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. D. Spyropoulos, "Web Usage Mining as a Tool for Personalization: A Survey," *User Modeling and User-Applied Interaction*, Vol. 13, pp. 311-372, 2003.
- [2] 강신재, "데이터의 웹을 위한 상호연결된 대규모 온톨로지 네트워크 구축," *한국산업정보학회논문지*, 15권 1호, pp. 15-23, 2010.
- [3] 이흥희, 이충권, 유상진, "제품 비교 웹서비스의 고객 검색 패턴에 대한 연구," *한국산업정보학회논문지*, 14권 3호, pp. 98-105, 2009.
- [4] J.H. Chang and W.S. Lee, "A Sliding Window Method for Finding Recently Frequent Itemsets over Online Data Streams," *Journal of Information Science and Engineering*, Vol. 20, pp. 753-762, 2004.
- [5] J.H. Chang and W.S. Lee, "Efficient Mining Method for Retrieving Sequential Patterns over Online Data Streams," *Journal of Information Science*, Vol. 31, pp. 420-432, 2005.
- [6] E. Chen, H. Cao, Q. Li, and T. Qian, "Efficient Strategies for Tough Aggregate Constraint-based Sequential Pattern Mining," *Information Sciences*, 178(6), pp. 1498-1518, 2008.
- [7] X. Ji, J. Bailey, and G. Dong, "Mining Minimal Distinguishing Subsequence Patterns with Gap Constraints," *Knowledge and Information Systems*, 11(3), pp. 259-296, 2007.
- [8] J. Pei, J. Han, and W. Wang, "Mining Sequential Patterns with Constraints in Large Databases," *Proc. of the 2002 ACM Int'l Conf.*

on Information and Knowledge Management (CIKM '02), pp. 18-25, 2002.

- [9] P. Tzvetkov, X. Yan, and J. Han, "TSP: Mining Top-K Closed Sequential Patterns," *Knowledge and Information Systems*, 7(4), pp. 438-457, 2005.
- [10] J. Wang and J. Han, and C. Li, "Frequent Closed Sequence Mining without Candidate Maintenance," *IEEE Transactions on Knowledge and Data Engineering*, 19(8), pp. 1042-1056, 2007.
- [11] M. Garofalakis, J. Gehrke, and R. Rastogi, "Querying and Mining Data Streams: You Only Get One Look," in The tutorial notes of the 28th Int'l Conf. on Very Large Data Bases, 2002.
- [12] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M.-C.- Hsu, "Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, pp. 1424-1440, 2004.
- [13] Y.-L. Chen, M.-C. Chiang, and M.-T. Ko, "Discovering Fuzzy Time-Interval Sequential Patterns in Sequence Databases," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 35(5), pp. 959-972, 2005.
- [14] Y.-L. Chen and T. C.-H. Huang, "Discovering Time-Interval Sequential Patterns in Sequence Databases," *Expert Systems with Applications*, 25(1), pp. 343-354, 2003.

장 중 혁 (Joong-Hyuk Chang)

- 1996년 2월 연세대학교 컴퓨터과 학과 (이학사)
- 1998년 8월 연세대학교 컴퓨터과 학과 (공학석사)
- 2005년 8월 연세대학교 컴퓨터과학과 (공학박사)
- 2006년 1월~2008년 7월 : UIUC, Wright State University 박사후 연구원
- 2008년 9월~현재 : 대구대학교 컴퓨터IT공학부 교수

논문접수일 : 2011년 04월 01일
 1차수정완료일 : 2011년 04월 22일
 게재확정일 : 2011년 05월 25일