

이산 속성 컨텍스트를 위한 시퀀스 매칭 기반 컨텍스트 예측

Context Prediction based on Sequence Matching for Contexts with Discrete Attribute

최영환* · 이상용**†

Young-Hwan Choi* and Sang-Yong Lee**†

* 공주대학교 컴퓨터공학과

** 공주대학교 컴퓨터공학부

요 약

지금까지 컨텍스트 예측 방법들은 이산 속성 컨텍스트를 대상으로 예측을 수행한 경우와 연속 속성 컨텍스트를 대상으로 예측을 수행한 경우로 나뉘어서 발전되어 왔다. 대부분의 예측 방법들은 컨텍스트의 획득 환경이나 특성에 맞게 특정 도메인에서 각각 예측 알고리즘을 작성하여 사용하여 왔기 때문에, 다양한 환경과 특성을 갖는 사용자의 컨텍스트를 대상으로 예측을 수행하기가 어렵다.

본 논문에서는 특정 도메인이나 컨텍스트의 특성에 국한되지 않고 이산 속성이나 연속 속성 컨텍스트들에 모두 적용 가능한 컨텍스트 예측 방법을 제안한다. 이를 위해 컨텍스트 속성간의 연관규칙을 고려하여 컨텍스트를 시퀀스로 생성하고, 컨텍스트 속성별 가변 가중치를 적용시켜 시퀀스 매칭 기반의 컨텍스트 예측을 수행한다. 제안한 방법을 평가하기 위해 이산 속성 컨텍스트와 연속 속성 컨텍스트에 각각 시뮬레이션한 결과 이산 속성 컨텍스트에서 80.12%, 연속 속성 컨텍스트에서 81.43%의 예측 정확도로 기존 예측 방법들과 비슷한 성능을 보였다.

키워드 : 컨텍스트 예측, 컨텍스트 속성, 연관규칙, 가변 가중치, 시퀀스 매칭

Abstract

Context prediction methods have been developed in two ways - one is a prediction for discrete context and the other is for continuous context. As most of the prediction methods have been used with prediction algorithms in specific domains suitable to the environment and characteristics of contexts, it is difficult to conduct a prediction for a user's context which is based on various environments and characteristics.

This study suggests a context prediction method available for both discrete and continuous contexts without being limited to the characteristics of a specific domain or context. For this, we conducted a context prediction based on sequence matching by generating sequences from contexts in consideration of association rules between context attributes and by applying variable weights according to each context attribute. Simulations for discrete and continuous contexts were conducted to evaluate proposed methods and the results showed that the methods produced a similar performance to existing prediction methods with a prediction accuracy of 80.12% in discrete context and 81.43% in continuous context.

Key Words : Context Prediction, Context Attribute, Association Rules, Variable Weights, Sequence Matching

1. 서 론

컨텍스트 인식 기반 서비스의 대표적인 분야는 추천 서비스이다. 컨텍스트 인식 기반 추천 서비스는 사용자의 다양한 컨텍스트 정보들을 자동으로 획득한 후 사용자의 상황에 맞는 최적의 개인화 된 서비스를 실시간으로 제공하는 것을 목적으로 한다[1]. 효과적으로 컨텍스트 인식 기반 개인화 추천 서비스를 위해서는 사용자의 컨텍스트를 정확하게 인식하고 인식된 컨텍스트 정보들로부터 사용자의 성향이나 선호도를 분석하여 사용자의 발생 가능한 다음 상황을 예측할 수 있어야 한다[2].

하지만 다양한 환경과 각종 센서들로부터 획득되는 컨텍스트들을 이용하여 추론과 예측을 할 경우, 컨텍스트들의 다양한 속성과 환경 상의 요인 등으로 인해 예측의 정확도를 떨어뜨릴 수 있기 때문에 회귀분석, 사례기반 추론, 인공지능망 등의 기존의 예측을 위해 사용되었던 방법들을 그대로 적용하기가 어렵다[3].

또한, 기존의 컨텍스트 예측에서 주로 사용되는 시퀀스 매칭에서는 사용자의 이동경로, 시간의 흐름에 따른 속성의 변화 등 연속 속성을 가지는 컨텍스트들을 대상으로 시퀀스 매칭을 수행하여 예측을 하게 된다. 시퀀스 매칭 방법은 어느 정도의 예측 정확도를 유지 한 채로 수행시간을 단축시킬 수 있는 장점은 있지만 예측의 대상이 되는 컨텍스트들의 속성들이 연속, 이산, 다중 속성 등 다양한 형태들로 존재하며 이러한 컨텍스트들을 대상으로 시퀀스 매칭을 수행하기가 어렵다.

접수일자 : 2010년 12월 10일

완료일자 : 2011년 4월 28일

+ 교신저자

본 논문에서는 이러한 컨텍스트들의 특성으로 인해 발생 되는 문제점들을 해결하기 위해 이산 속성, 연속 속성, 다중 속성을 가지는 컨텍스트들을 각 컨텍스트 속성별 인과관계를 이용하여 시퀀스로 생성하고, 시퀀스 매칭 기법으로 컨텍스트를 예측하는 방법을 제안한다.

2. 관련연구

컨텍스트 예측을 위하여 그 동안 Bayesian Networks, Markov Model, Sequence Matching 등의 각종 예측 방법들이 제안되어 성공적으로 사용되어 왔지만, 컨텍스트 인식 환경의 다양성과 변동성, 컨텍스트 속성의 다양성(이산속성/연속속성/다중속성), 개인 정보 획득의 어려움, 신속한 처리의 필요성 등으로 인하여 실시간 모바일 환경에 그대로 적용하기에 적합하지 않다. 또한 컨텍스트 간의 인과관계는 잘 알려져 있지 않고, 알려졌다고 하더라도 특정 도메인 내로 한정되기 때문에 일반성이 결여되어 있어 이를 고려한 예측 규칙을 획득하는 것도 매우 어렵다. 컨텍스트 예측 기술과 관련된 국내의 연구들은 이산 속성 컨텍스트를 다루는 예측 알고리즘과 시계열 속성 컨텍스트를 다루는 예측 알고리즘으로 나눌 수 있으며, 특징과 문제점들은 다음과 같다.

2.1 이산 속성 컨텍스트를 다룬 예측 알고리즘

이산 속성 컨텍스트를 대상으로 한 예측 알고리즘에는 Markov Models[4], Neural Networks[5], Bayesian Networks[6], Dynamic Bayesian Networks[7][8] 등이 있지만 학습을 통한 예측 알고리즘들은 실시간 컨텍스트 인식 환경을 고려하지 않았고, 특정 목적의 컨텍스트 예측 환경에서의 정확도, 신뢰도, 컨텍스트 정보와 부합 여부는 미지수이며, 구현상의 어려움이나 처리시간의 비효율성 등과 같은 문제점들을 가지고 있기 때문에 하이브리드 형식이나 수정된 알고리즘 형태로 사용되는 경우가 대부분이다. 특히 예측 시스템에서 가장 많이 사용하고 있는 Bayesian Networks는 이산 속성을 갖는 컨텍스트 간의 인과관계를 설명하는 부분에 있어서는 탁월한 성능을 가지지만, 시간이나 위치와 같은 연속 속성을 갖는 컨텍스트의 예측에 취약한 단점을 보인다.

2.2 연속 속성 컨텍스트를 다룬 예측 알고리즘

이산 속성 컨텍스트를 다루는 예측 알고리즘들이 가지는 문제점을 해결하기 위해 Sequential Matching 알고리즘[9], Semantic Caching[10], Context-Aware Replacement 알고리즘[11], 시계열 분석 기반 알고리즘[12] Alignment prediction[13] 등의 Matching 기반 알고리즘과 온톨로지를 활용한 RDF 기반의 알고리즘[14] 등이 연구되었다. 컨텍스트 인식 환경에서의 예측 알고리즘들도 기존의 예측 알고리즘들에 비해 향상된 능력을 보이지만 다음과 같은 문제점들이 발생한다.

Matching 기반 알고리즘들은 주로 연속 속성을 갖는 컨텍스트를 다루기 때문에 다양한 환경변수를 고려하기가 어렵고, 시퀀스가 생성되지 않는 환경에서는 사용하기 어렵다. 또한, 하나의 컨텍스트 정보를 예측하기 위해 요청이 올 때마다 데이터베이스를 처음부터 순차적으로 매칭하기 때문에 실시간 서비스를 제공하기 위해서는 비효율적이다. 또한 매치 자체가 없을 경우 전체 예측 정확도가 감소하는 단점이 있어, 상황이 반복적으로 발생하는 성향이 높아서 일치

하는 매치가 나타날 가능성이 높은 도메인에는 적합하지만 그렇지 않은 경우에는 정확도가 급격하게 떨어지게 된다.

온톨로지를 활용하는 RDF 기반의 알고리즘은 예측이나 추론을 위해 방대한 과거 자료를 필요로 하지 않고 현재의 컨텍스트만으로 예측이 가능하다는 장점이 있지만, 예측 및 추론을 위한 규칙을 사전에 정의할 수 있어야 한다는 적용상의 한계점과 구축비용이 증가하는 단점이 있다.

본 논문에서는 컨텍스트 인식환경이 다르고 공개된 필드 데이터가 없어 다른 방법들과 직접적인 비교는 어렵지만, 이산속성 컨텍스트와 연속 속성 컨텍스트를 이용하여 각각 시뮬레이션함으로써 기존 예측 알고리즘의 예측 정확도를 유지할 수 있음과 처리시간에서의 성능 향상을 보여준다.

3. 시퀀스 매칭 기반 컨텍스트 예측

3.1 다양한 형태의 컨텍스트 속성 분류 및 정의

컨텍스트 예측을 위해서는 먼저 컨텍스트의 획득 상황이나 활용 도메인에 따라 구분이 달라지므로 다양한 형태의 컨텍스트들의 속성을 분류하고 정의하는 과정이 필요하다.

컨텍스트들은 다양한 센서와 디바이스로부터 획득할 수 있으며, 성별, 나이, 위치, 접근 시간, 온도, 날씨 등 다양한 형태와 속성을 가지게 되는데, 본 논문에서는 컨텍스트를 다음과 같이 연속 속성을 가지는 컨텍스트[표 1], 이산 속성을 가지는 컨텍스트[표 2], 다양한 형태로 분류할 수 있는 다중 속성을 가지는 컨텍스트[표 3]로 분류한다.

표 1. 연속 속성 컨텍스트

Table 1. Time series attribute context

종류	구분
이동경로	A - C - B - F - E
체중변화	75kg - 79kg - 81kg
온도변화	25℃ - 28℃ - 31℃ - 27℃

표 2. 이산 속성 컨텍스트

Table 2. Discrete attribute context

종류	구분
성별	남성, 여성
날씨	맑음, 흐림, 눈, 비
계절	봄, 여름, 가을, 겨울
직업	전문직, 공무원, ..., 기타
종교	기독교, 불교, 천주교, ..., 기타
결혼여부	기혼, 미혼

연속 속성 컨텍스트들은 시퀀스 매칭을 위해 시퀀스를 생성하기가 쉽고, 시퀀스 매칭을 이용하여 예측을 수행할 경우 예측 정확도도 높고 예측을 위한 수행 시간도 짧아 컨텍스트 예측 시스템에 적용하기가 수월하다. 하지만 이산 속성이나 다중 속성을 가지는 컨텍스트들은 시간의 흐름에 따른 시퀀스로 생성하기가 어렵고, 서로 다른 속성들을 가지기 때문에 시퀀스 매칭을 이용하여 예측을 수행하기가 어려워 좀 더 복잡한 알고리즘을 이용하여 예측을 수행해야 한다. 이러한 경우 정확도 측면에서는 어느 정도 성능을 보일 수 있지만 수행 시간 측면에서의 이득을 보기 어려워

실시간 예측에 적용시키기 어렵다.

표 3. 다중 속성 컨텍스트

Table 3. Multiple attribute context

종류	구분
나이	10대, 20대, ..., 60대
	유아, 어린이, 청소년, 청년, 노인
시간	01시, 02시, ..., 24시
	오전, 점심, 오후, 저녁
온도	덥서, 화서
	추움, 시원함, 따뜻함, 더움

또한, 이산 속성을 가지는 컨텍스트들은 속성들 간의 관계가 없기 때문에 이산형태로 속성의 표현이 가능하지만, 다중 속성이나 연속 속성을 가지는 컨텍스트들을 이산형태로 표현하게 되면 컨텍스트 속성간 인과관계를 정확히 표현할 수 없어 예측 정확도를 떨어뜨리는 요인이 될 수 있으므로 다중 속성이나 이산 속성을 가지는 컨텍스트들을 효과적으로 시퀀스 형태로 표현하고 이를 이용하여 시퀀스 매칭을 통한 예측을 수행할 수 있어야 한다.

3.2 시퀀스 매칭 기반 컨텍스트 예측

본 논문에서 제안하는 시퀀스 매칭 기반 컨텍스트 예측 방법(CPSM : Context Prediction based on Sequence Matching)의 전체적인 프로세스는 다음 그림 1과 같다.

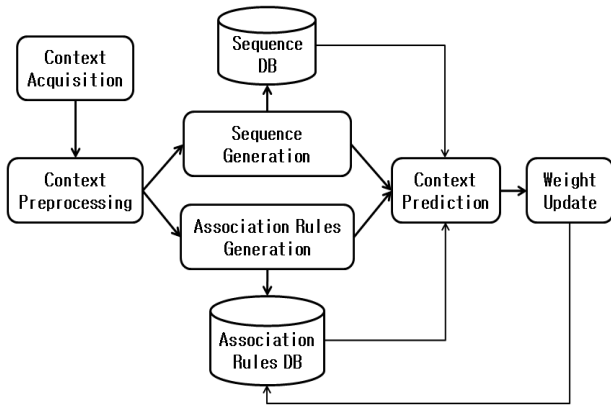


그림 1. 컨텍스트 예측 프로세스
Fig. 1. Context Prediction Process

3.2.1. 시퀀스 생성

컨텍스트 예측을 위한 컨텍스트들은 하나의 속성을 가지지 않고, 대부분 연속 속성, 이산 속성, 다중 속성들이 혼합되어 있다. 이러한 컨텍스트들로부터 시퀀스를 생성하는 방법은 다음과 같다.

먼저, 현재 획득된 컨텍스트 집합에서 각 속성별로 정규화하는 과정을 거친다. 예를 들어 성별인 경우 <남성, 여성>의 속성을 가지게 되는데, 이를 <1, 2>와 같은 형태의 컨텍스트 속성으로 바꾸는 과정이다. 나이의 경우 <유아, 어린이, 청소년, 청년, 노인>에서 <1, 2, 3, 4, 5, 6>으로 바꾼다. 하지만 현재 상황에서 각 속성에 대한 수치는 의미를 갖지 않고 속성을 구분하기 위한 분류이다.

획득된 컨텍스트 집합에서 각 개별 컨텍스트 속성에 대해서 속성간 인과관계를 계산한다. 이 때 각 개별 컨텍스트가 모두 명목 척도일 경우에는 확률분석을 통해 가장 비율이 높은 값을 채택한다.

예를 들어 다음과 같이 나이대별로 해당 장소에서 특정 콘텐츠를 제공 받은 데이터가 있는 경우는 다음 표 4와 같이 추정한다.

표 4. 연령별 콘텐츠 접근 정보

Table 4. Information content approach by ages

연령	위치	콘텐츠	비율(%)
10대	B	D	48%
10대	B	B	23%
10대	B	C	22%
10대	B	A	7%

위의 경우에 사용자가 연령이 10대이고 B의 위치에서 콘텐츠 D를 제공받은 비율이 48%로 가장 크므로 “연령 10대 and 위치 B -> 콘텐츠 D”와 같이 연관규칙을 정의할 수 있다. 위와 같은 방법을 통해서 각 속성별로 N개의 인과관계를 획득하고, 모든 얻어진 속성간 연관규칙을 저장한다.

3.2.2 시퀀스 매칭

다음은 위의 단계를 거쳐 얻어진 속성간 연관규칙을 기반으로 생성된 시퀀스가 {1, 2, 4, 1, 2, 2}(성별-남, 연령-20대, 직업-개인사업, 요일-일요일, 시간대-정오, 위치-No2)과 같이 생성되었다고 가정했을 때, 일치하는 시퀀스가 존재할 경우에는 표 5와 같이 해당 시퀀스를 후보 시퀀스로 설정하고 결과를 예측한다.

이때, 생성된 시퀀스와 일치하는 후보 시퀀스가 3개가 검색될 경우 출현 빈도가 가장 높은 시퀀스의 결과를 예측 결과로 선택한다. 만약, 같은 빈도를 가지는 시퀀스가 두 개 이상 존재할 때는 가장 최근에 나타난 시퀀스를 예측 결과로 선택한다.

표 5. 일치하는 시퀀스 존재시 매칭 방법

Table 5. Matching method when a same sequence exist

현재 생성된 시퀀스 : {1, 2, 4, 1, 2, 2} {성별, 연령, 직업, 요일, 시간대, 위치}		
시퀀스	빈도	결과
패턴 : {1, 2, 4, 1, 2, 2}	4	1(콘텐츠 A)
패턴 : {1, 2, 4, 1, 2, 2}	3	2(콘텐츠 B)
패턴 : {1, 2, 4, 1, 2, 2}	1	3(콘텐츠 C)
예측 결과		1(콘텐츠 A)

일치하는 시퀀스가 존재하지 않을 때는 표 6처럼 패턴 a, 패턴 b, 패턴 c처럼 3개의 후보 패턴 중 각 5개의 속성이 일치하게 되는데, 이 때 일반적으로 가장 최근에 생성된 시퀀스나 출현 빈도수가 가장 많은 시퀀스를 선택할 수 있다. 그러나 패턴 a의 네 번째 속성, 패턴 b의 첫 번째 속성, 패턴 c의 세 번째 속성이 결과에 미치는 영향을 고려하지 않으면 예측 정확도가 떨어지게 된다.

표 6에서 각 컨텍스트 속성간 연관규칙을 해당 위치(No2)에서 제공받은 콘텐츠에 관한 연관규칙이 각각 성별

(50%), 직업(30%), 요일(25%) 라면 성별이 결과에 가장 많은 영향을 미친 컨텍스트 속성으로 판단하여 패턴 B를 예측 후보로 선택함으로써 예측의 정확도를 향상 시킬 수 있게 된다.

표 6. 일치하는 시퀀스 미 존재시 매칭 방법
Table 6. Matching method when a same sequence dose not exist

현재 생성된 시퀀스 : {1, 2, 4, 1, 2, 2} {성별, 연령, 직업, 요일, 시간대, 위치}	
시퀀스	결과
패턴 a : {1, 2, 4, 3, 2, 2}	1(콘텐츠 A)
패턴 b : {2, 2, 4, 1, 2, 2}	2(콘텐츠 B)
패턴 c : {1, 2, 3, 1, 2, 2}	3(콘텐츠 C)
예측 결과	2(콘텐츠 B)

다음 그림 2는 생성된 시퀀스를 기반으로 매칭을 하는 과정을 보여준다. 처음 매칭을 시도할 때 사용자의 후보 시퀀스를 기준으로 매칭을 시도하고, 만약 후보 시퀀스들 중에 매치되는 시퀀스가 없는 경우에는 컨텍스트 히스토리 중에 가장 최근의 시퀀스들부터 매칭을 시도하여 매치되는 시퀀스를 찾게 되면 이를 예측후보로 설정하고 가중치를 업데이트하여 규칙 DB에 저장하게 된다.

```

BEGIN
  Input pre_data[], i, j, context_count,
        seq_match, context-sequence size
  Generate context-sequence
  FOR j=0 to (context-sequence size) DO
    initialize each context information
    FOR i=0 to ( context-sequence size ) DO
      IF (context[i] == pre_data[i]) &&
         (context[i+1] == pre_data[i+1]) THEN
        Increase seq_match by 1
      ENDIF
    ENDFOR
    IF (context-sequence >= 1) THEN
      increase context_count by 1
      reset context_count
    ELSE
      Do original matching
    ENDIF
  ENDFOR
END
    
```

그림 2. 시퀀스 매치
Fig. 2. Sequential matching

고정된 가중치를 갖는 패턴 규칙은 예측의 정확도를 떨어뜨릴 수 있기 때문에 다양한 컨텍스트 속성이 사용자의 행동변화에 영향을 미치는 정도를 수치적으로 표현하고 능동적으로 적용시키기 위해 가변 가중치를 사용한다.

시간의 흐름이나 환경의 변화에 따른 사용자의 행동양식이나 습관의 변화에 적응할 수 있도록 주기적으로 업데이트하는 time-driven 방식과 상황의 변화가 발생할 때마다 업데이트하는 event-driven 방식을 사용하여 가중치를 업데이트한다.

이트한다.

각각의 상황에 따라 컨텍스트 속성의 우선순위를 분석하여 가중치를 부여하고, 예측 결과를 바탕으로 예측에 성공하였을 경우에는 가중치를 증가시키고, 실패하였을 경우에는 가중치를 감소시킴으로써 각각의 결과를 컨텍스트 속성에 반영시킬 수 있도록 한다.

가중치는 일치하는 시퀀스가 다수 존재할 때, 출현 빈도와 가장 최근에 일어난 시퀀스 확률을 이용한다.

4. 실험 및 평가

본 논문에서 제안한 시퀀스 매칭 기반 컨텍스트 예측 방법(CPSM)의 성능을 평가하기 위해 독일의 Augsburg 대학에서 실제 구축한 컨텍스트 인식 환경 기반의 공개된 컨텍스트 필드데이터(Augsburg Indoor Location Tracking Benchmarks : AILTB)[15]와 본 연구실에서 이전의 연구 결과물인 전시관 관람 컨텍스트 데이터 셋[16]을 시뮬레이션을 통해 수행하였다.

4.1 시뮬레이션 1

AILTB는 네 명의 연구원에 의해 여름과 가을 두 차례에 걸쳐 생성하였고, 그림 3과 같이 날짜, 시간, 섹터, 사용자, 시간스탬프의 포맷으로 이루어져 있다. 그림 4는 위치와 위치정보를 나타낸다. 본 논문에서는 가을의 데이터를 트레이닝 데이터로 여름의 데이터를 테스트 데이터로 하여 시뮬레이션을 수행하였다.

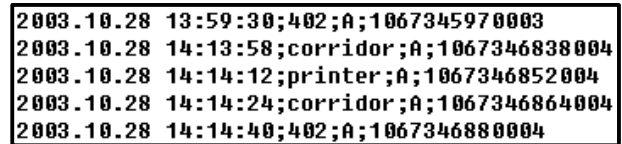


그림 3. AILTB 데이터의 구조
Fig. 3. Structure of AILTB

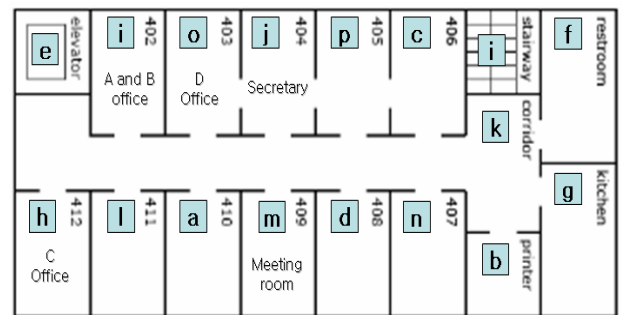


그림 4. 섹터와 섹터정보
Fig. 4. Sector and sector information

표 7은 제안한 방법(CPSM)과 J. Petzold[15]가 평가한 기존 알고리즘의 예측정확도 측면에서의 성능평가 결과를 보여준다. 제안된 방법의 예측정확도는 동일한 조건하에서 평균 81.43%로 기존의 예측 알고리즘과 거의 비슷한 성능을 가짐을 보여준다.

표 7. 알고리즘들의 예측 정확도 비교(단위 : %)

Table 7. Prediction accuracy comparison of algorithms

User	Prediction Algorithm				
	MLP	BN	MP	SM	CPSM
A	87.39	85.58	90.18	85.23	85.41
B	75.66	86.54	78.97	83.73	81.23
C	68.68	86.77	75.17	75.39	78.38
D	74.06	69.78	78.05	82.45	80.72
Ave	76.45	82.17	80.59	81.70	81.43

- MLP : Multi Layer Perceptron
- BN : Bayesian Network
- MP : Markov Predictor
- SM : Sequence Matching
- CPSM : Context Prediction based on Sequence Matching

4.2 시뮬레이션 2

컨텍스트 데이터 500셋 중, 70%는 트레이닝 데이터로 나머지 30%는 테스트 데이터로 하여 성능을 테스트하였다. 표 8은 사용자 컨텍스트를 위한 프로파일 객체이며, 표 9는 구축된 사용자들의 컨텍스트이다.

표 8. 사용자 프로파일 객체

Table 8. Object of user's profiles

Column Name	Data Type	Comments	Type
user_id	char(20)	사용자 ID	String
gender	char(1)	성별	1, 2
age	varchar(3)	나이	1~6
job	varchar(20)	직업	1~6
view_date	datetime	관람일자	String
view_time	datetime	관람시간	1~3
location	varchar(5)	위치	1~12

표 9. 사용자들의 컨텍스트

Table 9. Contexts of users'

1	user_id	gender	age	job	view_date	view_time	location
2	aquablue	female	26	5	2009-08-03	16:02:58	No3
3	aquablue	female	26	5	2009-08-03	16:04:21	No9
4	aquablue	female	26	5	2009-08-03	16:06:02	No10
5	aquablue	female	26	5	2009-08-03	16:02:11	No2
6	aquablue	female	26	5	2009-08-03	16:03:24	No5
7	aquablue	female	26	5	2009-08-03	16:04:22	No11
8	aquablue	female	26	5	2009-08-03	16:24:03	No9
9	aquablue	female	26	5	2009-08-03	16:24:59	No10
10	after_school	male	19	8	2009-07-30	11:42:00	No2
11	after_school	male	19	8	2009-07-30	11:47:45	No3

다음 표 10은 생성된 사용자들의 시퀀스이며, 6개의 항목들로 이루어져 있다.

성별은 <남, 여>, 연령은 <10대, 20대, 30대, 40대, 50대, 60대>, 직업은 <학생, 일반직, 전문직, 개인사업, 주부, 기타>, 요일은 <일, 월, 화, 수, 목, 금, 토>, 시간대는 <오전, 정오, 오후>, 위치는 <1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12> 값을 가진다. 여기에서 정수의 크기는 의미가 없으며 위의 각 객체가 갖는 구분을 위해 사용하였다.

표 10. 생성된 사용자들의 시퀀스

Table 10. Generated users' sequence

gender	age	job	view_date	view_time	location
1	2	4	1	2	2
2	3	3	1	2	5
2	2	2	7	1	7
1	1	1	1	1	7
1	1	1	1	2	3
1	2	3	7	1	2
2	4	5	3	3	5
2	5	5	4	3	4
1	6	6	2	1	4
2	2	5	5	3	11

평가 결과 제안된 방법의 예측정확도는 80.12%로 기존 알고리즘의 예측정확도와 비슷한 성능을 보였다.

각각의 컨텍스트를 대상으로 시뮬레이션을 한 결과 시뮬레이션 1은 81.43%, 시뮬레이션 2는 80.12%의 예측 정확도를 보였는데, 이러한 결과는 시뮬레이션 1의 데이터가 4명의 사용자들 대상으로 몇 개월간 지속적으로 컨텍스트를 수집하였기 때문에 컨텍스트 특성이 시퀀스 매칭에 좀 더 적합하였기 때문에 시뮬레이션 2보다 좋은 결과를 보인 것으로 판단된다.

가중치는 시뮬레이션 결과, 시뮬레이션 1은 평균 출현빈도가 높은 시퀀스가 0.8, 가장 최근의 시퀀스가 0.2로 측정되었으며, 시뮬레이션 2는 평균 출현빈도가 높은 시퀀스가 0.3, 가장 최근의 시퀀스가 0.3으로 측정되었다. 이는 지속적으로 획득되지 않는 환경에서의 가중치가 개인의 컨텍스트가 지속적으로 획득되는 환경보다 출현빈도가 조금 더 예측 정확도에 영향을 미치는 것으로 이해할 수 있다.

5. 결 론

본 논문에서는 다양한 환경과 컨텍스트들의 특성으로 인해 발생하는 문제점들을 해결하기 위해 시퀀스 매칭 기반 컨텍스트 예측을 이용하여 이산속성 컨텍스트와 연속속성 컨텍스트에 모두 적용 가능한 방법을 제안하였다. 이를 위해 컨텍스트 속성간의 연관규칙과 컨텍스트 속성별 가변가중치를 이용하여 사용자들의 컨텍스트들을 시퀀스로 생성하고, 시퀀스 매칭 방법을 이용하여 컨텍스트를 예측하였다.

컨텍스트 예측 시 가장 최근의 사용자 시퀀스와 주기적으로 나타나는 사용자 시퀀스를 대표 시퀀스로 두고 새롭게 생성되는 사용자 컨텍스트로부터 생성된 시퀀스와 우선 비교하도록 하여 가능성이 가장 높은 시퀀스와 비교함으로써 비교횟수를 줄일 수 있었다. 실험결과를 통해 본 연구에서 제안한 시퀀스 매칭 기반 컨텍스트 예측은 실시간 서비스 환경에서 컨텍스트 속성에 상관없이 컨텍스트 예측 알고리즘 수행 시 기존의 예측 알고리즘들과 비슷하게 예측 정확도를 유지할 수 있음을 보여줌으로써, 각각의 도메인이나 특성에 맞는 컨텍스트 예측 방법을 구현할 필요 없이 컨텍스트 예측을 통한 개인화 서비스가 가능함을 보여주었다. 또한 개인의 프라이버시 침해 문제로 인해 정보 제공을 원치 않는 경우나, 놀이공원이거나 관광지처럼 개인의 컨텍스트를 지속적으로 획득할 수 없는 환경에서도 제한적인 컨텍스트만으로도 예측이 가능함을 보였다.

참 고 문 헌

- [1] D. Tennenhouse, "Proactive computing", *Communications of the ACM*, vol. 43, pp. 43-50, 2000.
- [2] D. Patterson, L. Liao, D. Fox, and H. Kautz, "Inferring High-Level Behavior from Low-Level Sensors," *Proc. of the Fifth International Conference on Ubiquitous Computing*, pp. 73-89, Seattle, WA, October, 2003.
- [3] R. Mayrhofer, "An Architecture for Context Prediction," *Trauner Verlag, Schriften der Johannes-Kepler-Universität Linz*, volume C45, 2005.
- [4] E. Behrends, "Introduction to Markov Chains," *Advanced Lectures in Mathematics Vieweg*, 1999
- [5] K. Gurney, "An Introduction to Neural Networks," Routledge, 2002.
- [6] F. V. Jensen, *An Introduction to Bayesian Networks*, UCL Press, 1996.
- [7] A. Raj, A. Subramanya, D. Fox, J. Bilmes, "Rao-blackwellized Particle Filters for Recognizing Activities and Spatial Context from Wearable Sensors", *Experimental Robotics*, vol.39, pp. 211-221, 2008.
- [8] 양성익, 홍진혁, 조성배, "동적 베이지안 네트워크를 이용한 멀티모달센서기반 사용자 행동인식", *정보과학회논문지*, 제15권, 1호, pp. 72-76, 2009.
- [9] V. Guralnik, K. Z. HAIGH, "Learning Models of Human Behaviour with Sequential Patterns," *AAAI Workshop on Automation as Caregiver*, pp. 24-30, July 2002.
- [10] Q. Ren, M. H. Dunham, "Using Semantic Caching to Manage Location Dependent Data in Mobile Computing," *In Proc. of the International Conference on Mobile Computing and Networking*, pp. 210-221, 2000.
- [11] 황우민, 임상석, 박규호, "웨어러블 컴퓨터를 위한 사용자 컨텍스트 인식 페이지 캐시," *한국차세대PC학회논문지*, 제2권, 1호, pp.46-55. March, 2006.
- [12] 최영환, 권일경, 김진태, 서효석, 이상용, "k-means 클러스터링과 시계열 분석을 이용한 컨텍스트 예측 기반 개인화 서비스", *한국지능시스템학회*, 제20권, 1호, pp. 442-445, 04. 2010.
- [13] S. Sigg, S. Haseloff, K. David, "A Novel Approach to Context Prediction in Ubicomp Environments," *The 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2006.
- [14] Projects of Industrial Ontologies Group, <http://www.cs.jyu.fi/ai/OntoGroup/projects.htm>.
- [15] J. Petzold, "Augsburg Indoor Location Tracking Benchmarks", *Institute of Computer Science University of Augsburg Germany*, April, 2004.
- [16] 최영환, 이상용, "유비쿼터스 컴퓨팅 환경에서 컨텍스트 예측을 위한 시계열 분석 기반 사용자 모델링", *한국지능시스템학회논문지*, 제19권, 5호, pp. 655-660, 2009.

저 자 소 개



최영환(Young-Hwan Choi)

1999년 : 공주대학교 전자계산학과(학사)
 2001년 : 공주대학교 대학원 전자계산학과 (이학석사)
 2001년~현재 : 공주대학교 대학원 컴퓨터 공학과 박사과정

관심분야 : 인공지능, 유비쿼터스 컴퓨팅, 컨텍스트 예측, 개인화 서비스 등
 E-mail : cyhmad@kongju.ac.kr



이상용(Sang-Yong Lee)

1984년 : 중앙대학교 전자계산학과(공학사)
 1988년 : 일본동경대학대학원 총합이공학 연구과(공학석사)
 1988년~1989년 : 일본 NEC 중앙연구소 연구원
 1993년 : 중앙대학교 일반대학원 전자계산학과(공학박사)

1996년~1997년 : University of Central Florida 방문교수
 1993년~현재 : 공주대학교 컴퓨터공학부 교수

관심분야 : 인공지능, 컨텍스트 예측, 컴퓨터 게임 등
 E-mail : sylee@kongju.ac.kr