

오프라인 쇼핑몰에서 개인화된 상품 추천을 위한 사용자의 이동패턴 분석

Users' Moving Patterns Analysis for Personalized Product Recommendation in Offline Shopping Malls

최영환* · 이상용**

Young-Hwan Choi and Sang-Yong Lee

* 공주대학교 컴퓨터공학과

** 공주대학교 컴퓨터공학부 (교신저자)

요 약

유비쿼터스 컴퓨팅에서 대부분의 시스템들이 개인화된 추천을 위하여 사용자와 성향이 비슷한 사람들의 컨텍스트 정보를 분석하는데 인구통계학적 방법이나 협력적 필터링을 주로 사용한다. 서비스 추천 시스템들은 컨텍스트 정보 중에서 성별, 나이, 직업, 구매이력 등의 정적 컨텍스트를 주로 사용하고 있다. 그러나 이러한 시스템은 이동경로 같은 사용자의 상황을 고려하기가 어렵기 때문에 개인의 성향을 정확하게 분석하여 실시간으로 개인화된 추천 서비스를 제공하는데 한계가 있다. 본 논문에서는 사용자의 상황을 고려하기 위해 동적 컨텍스트 중에서 사용자의 이동경로를 이용한다. 이동경로의 예측 정확도를 높이기 위해 RSOM의 입력으로 들어가는 이동경로를 경로보정 알고리즘을 사용하여 보정한다. 그리고 보정된 경로를 RSOM으로 학습시켜 사용자의 이동패턴을 분석하고 향후 이동경로를 예측한 후, 사용자의 선호도가 높은 상품들 중에서 예측 경로 상에 있는 가장 가까운 상품을 실시간으로 추천한다. 제안한 방법의 예측 정확도를 측정된 결과 MAE가 평균 0.5 이하로 측정됨으로써 사용자의 이동경로를 올바르게 예측할 수 있음을 확인하였다.

Abstract

Most systems in ubiquitous computing analyze context information of users which have similar propensity with demographics methods and collaborative filtering to provide personalized recommendation services. The systems have mostly used static context information such as sex, age, job, and purchase history. However the systems have limitation to analyze users' propensity accurately and to provide personalized recommendation services in real-time, because they have difficulty in considering users situation as moving path.

In this paper we use users' moving path of dynamic context to consider users situation. For the prediction accuracy we complete with a path completion algorithm to moving path which is inputted to RSOM. We train the moving path to be completed by RSOM, analyze users' moving pattern and predict a future moving path. Then we recommend the nearest product on the prediction path with users' high preference in real-time. As the experimental result, MAE is lower than 0.5 averagely and we confirmed our method can predict users moving path correctly.

Key words : Ubiquitous Computing, Path Completion, Personalized Recommendation, Moving Path Analysis, Context Awareness

1. 서 론

최근 유비쿼터스 컴퓨팅이 IT의 새로운 패러다임으로 등장했다. 유비쿼터스 컴퓨팅은 언제, 어디서나 사용하는 컴퓨팅환경을 지칭하며, 다양한 종류의 컴퓨터가 사람, 사물, 환경 속에 내재되어 있고, 이들이 서로 연결되어 필요한 곳에서 컴퓨팅을 구현할 수 있는 환경을 의미한다. 즉, 유비쿼터스 컴퓨팅은 인간의 실제 세계에 산재해 있는 컴퓨팅 장치들과 인간을 자연스럽게 상호작용하도록 하는 것이다[1]. 이를 위해서는 사용자의 상황정보를 포함하고 있는 컨텍스트를 인식하고 이를 활용하여 인간에게 알맞은 서비스를 제공하는

것이 매우 중요하다.

컨텍스트란, 인간이나 장소, 객체 같은 개체들을 정의하는데 사용되는 모든 정보들을 의미하며, 이 정보들은 사용자와 응용 사이의 상호작용과 관련이 있는 것이어야 한다[2]. 컨텍스트는 사용자의 프로파일과 관련된 정보인 정적 컨텍스트와 시간이나 장소에 따라 자주 변화하는 동적 컨텍스트로 나누어지며 사용자에게 알맞은 서비스를 제공하기 위해서는 정적 컨텍스트와 동적 컨텍스트를 모두 고려해야 한다. 하지만, 현재 서비스 대부분의 시스템이 사용자에게 개인화된 추천 서비스를 제공하기 위해 정적 컨텍스트 정보만을 사용함으로써 개인의 성향을 정확하게 파악하지 못하는 경우가 많다.

예를 들어, 사용자의 선호도를 파악하기 위해 정적 컨텍스트인 구매이력과 개인 프로파일 정보만을 이용하게 되면, 사용자가 관심이 있는 상품인데도 불구하고 단지 구매를 하지 않았다는 이유만으로 그 상품에 관한 선호도가 없다고 판단

접수일자 : 2006년 2월 3일

완료일자 : 2006년 4월 13일

하게 될 수도 있다.

또한 실시간으로 사용자에게 선호도가 가장 높은 상품을 먼저 추천하는 경우, 사용자의 현재 위치와 이동경로를 고려하지 않게 되면, 사용자의 현재 위치에서 멀리 떨어져 있는 상품을 추천하게 될 수도 있다. 그렇게 되면 사용자는 추천된 상품을 구매하거나 보기 위해서는 현재 위치에서 멀리 떨어진 곳까지 이동을 해야 하는 불편함을 가져올 수 있다.

본 논문에서는 이런 문제를 해결하기 위해 협력적 필터링을 사용하여 사용자를 그룹으로 분류하고, 사용자의 프로파일을 토대로 사용자가 관심을 가질만한 상품을 예측한다. 그리고 사용자의 이동경로를 경로보정 알고리즘과 RSOM을 이용하여 분석하고 향후 이동경로를 예측한 후, 사용자의 현재 위치를 파악하여 사용자의 주변에 있는 상품 중에서 사용자가 이동할 예측경로 중에 있는 가장 가까운 상품 정보를 효율적으로 추천할 수 있는 방법을 제안한다.

2. 관련연구

2.1 컨텍스트 인식

유비쿼터스 컴퓨팅에서 사용되는 컨텍스트에 대하여 Schilit 등은 사용자의 위치, 사용자 주변에 있는 사람 정보, 그리고 사용 가능한 자원 등의 정보라고 정의하였다[3]. Dey 등은 기존 정의들을 종합하여 컨텍스트를 사용자와 응용서비스 사이의 상호작용을 위해 필요한 사용자, 장소, 대상물 등의 개체 상태를 나타내는 정보라고 정의하였으며[4], Jang 등은 컨텍스트란 5W1H : Who, What, Where, When, Why, How 이며, 응용 서비스에 따라 5W1H의 조합으로 나타난다고 정의하였다[5].

컨텍스트가 사용자의 상황을 나타내는 정보라고 하면, 컨텍스트 인식 시스템은 이러한 정보를 기반으로 적절한 서비스를 제공하는 시스템을 의미하며 크게 두 종류로 나뉜다. 첫째, 주변 환경의 변화를 감지하여 컨텍스트를 파악하고, 컨텍스트 정보에 따라 적절한 서비스를 제공하는 것이다 [3][4][6]. 둘째, 컨텍스트에 맞추어 시스템의 실행 조건이나 주변 환경을 스스로 변경하는 것이다.

현재 컨텍스트 인식 시스템과 응용 서비스에 대한 개발이 활발하게 진행되고 있다. 사무환경, 가정환경, 공동작업환경, 교육환경, 실내환경 등의 분야에서 연구가 많이 진행되고 있지만, 아직까지는 쇼핑물이나 극장, 식당 등 사용자에게 개인화된 서비스를 추천하는 분야에서는 사용자의 행위에 관련된 요소를 제외한 정적 컨텍스트만을 사용하는 경우가 대부분이다.

2.2 개인화 추천

개인화 추천이란 사용자 프로파일을 모델링하여 각 사용자의 특성에 따라 각각 다른 서비스를 제공하는 기술이다. 일반적으로 사용자 프로파일은 이름이나 주민등록번호, 기존 활동 내역 등의 변하지 않는 정적인 특성과 현재 사용자의 상태(위치, 기호도, 감정 등)를 나타내는 동적인 특성을 통해 모델링 될 수 있다. 이러한 프로파일은 맞춤형 서비스를 위한 가장 중요한 기반 자료로써 활용되며 분석을 통해 개인화된 서비스를 제공한다.

유비쿼터스 컴퓨팅 환경에서는 사용자가 직접 컴퓨터 장비 등을 통하여 정보 및 서비스를 찾는 것이 아니라 사용자의 위치 이동에 따라 서비스 장치들이 동적으로 사용자의 최

소한의 개입으로 이루어져야 한다. 즉, 개인마다 각각 다른 개인화된 서비스를 제공한다는 것이며 사용자의 욕구에 가장 근접한 맞춤형 서비스가 되어야 한다.

개인화된 서비스 추천을 위해 주로 데이터 마이닝 기법들이 사용되고 있으며, 연관규칙[7], 분류규칙, 클러스터링, 유사성 탐색 등이 있다. 본 논문에서는 클러스터링 기법 중 협력적 필터링을 사용하여 사용자가 선호하는 패턴과 유사한 다른 사용자들의 선호도를 기반으로 사용자들을 그룹화 한다 [8].

2.3 이동패턴 분석

유비쿼터스 컴퓨터의 등장으로 센서를 이용하여 사용자의 상태를 예측하기 위한 사용자 모델링 방법에 관한 연구가 활발히 진행 중이다. 그 중에서도 위치 정보는 사용자의 상태를 예측하여 이에 맞는 서비스를 제공하기 위한 중요한 정보이기 때문에 이를 활용하기 위한 연구가 중요하다. 위치 정보의 활용에 대한 연구로는 크게 정확한 위치 추적을 위한 방법에 대한 연구와 위치 정보를 이용한 다른 정보를 추론하는 연구가 있다. 이렇게 추적된 위치정보는 사용자의 다음 위치를 예측하는데 사용될 수 있다. 사용자의 다음 위치를 예측하는 것은 사용자에게 필요한 서비스를 효율적으로 제공하는데 매우 유용하게 활용될 수 있다.

Patterson 등은 이를 위해 GPS 신호로 사용자 위치를 추적하고, 베이지안 네트워크로 GPS 신호를 모델링 하여 현재 사용자가 이용하고 있는 이동수단을 예측하는 방법을 제안하였다[9]. Ashbrook 등은 GPS 데이터를 k-means 알고리즘을 사용해 클러스터링 하여 사용자가 오래 머무르는 유효한 위치를 찾아내고, 이 위치간의 이동을 마코프 모델을 사용하여 모델링 하여 사용자 위치를 예측하도록 하였다[10]. 하지만 사용자가 머무르는 위치만을 사용하기 때문에 이동 중 경로에 따른 예측은 불가능하다. 또한 마코프 모델은 매우 결정적이기 때문에 예측하는데 있어 유연성이 떨어지는 단점이 있다.

3. 개인화된 상품 추천을 위한 사용자의 이동패턴 분석

본 논문에서 제안하는 방법의 목적은, 사용자가 쇼핑물에 들어오게 되면 센서로 사용자를 확인하고 이동경로를 예측하여 선호도가 높은 상품을 추천하는데 있다.

이를 위해 센서로부터 사용자의 컨텍스트를 입력받아 ID로 사용자를 구분하고, 컨텍스트를 이력데이터와 경로데이터로 분류한다. 이력데이터는 협력적 필터링 과정을 거쳐 각 쇼핑물의 상품과 사용자를 분류하고 유사성이 높은 사용자와 상품을 클러스터링 한다. 경로데이터는 경로보정 알고리즘(path completion algorithm)으로 이동경로를 보정한 후, RSOM(recurrent self organization map)으로 학습하여 지역모델을 생성하고, 새로운 데이터를 지역모델로 평가하여 이동경로를 예측하고, 사용자가 관심을 가지는 상품 중에서 사용자의 향후 이동경로 중 가장 가까운 곳에 위치한 상품을 실시간으로 추천하게 된다.

그림 1에서는 오프라인 쇼핑물에서의 개인화된 상품 추천의 처리 흐름을 보여준다.

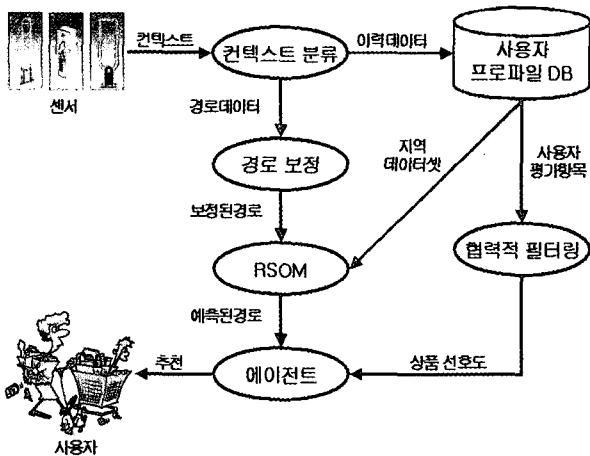


그림 1. 상품 추천의 처리 흐름
Fig 1. Process of the Product Recommendation

3.1 사용자 프로파일

사용자 프로파일은 사용자의 컨텍스트를 표현한 것으로, 본 논문에서는 초기 사용자 프로파일의 생성을 위해 쇼핑몰에 처음 사용자가 등록을 할 때 생성되는 정적 컨텍스트 정보와 센서를 통해 획득한 동적 컨텍스트 정보를 이용한다. 생성된 초기 프로파일과 이미 그룹화 되어 있는 그룹을 비교하여 그룹 중 가장 유사한 그룹을 찾아내어 사용자를 그룹에 포함시킨 후, 생성되는 구매이력, 이동경로 등을 분석하여 사용자 프로파일을 갱신하게 된다. 표 1은 사용자 프로파일 생성과 갱신에 사용되는 컨텍스트 항목들이고, 표 2는 사용자 프로파일의 내용이다.

표 1. 컨텍스트 항목
Table 1. Context Items

형태	항목	범위
정적 컨텍스트	ID	1 ~ 525
	성별	1 ~ 2
	직업	1 ~ 12
	나이	20 ~ 62
	결혼여부	1 ~ 2
	자녀의 수	0 ~ 4
	월평균 소득	1 ~ 5
동적 컨텍스트	구매이력	상품 ID
	구매액수	1 ~ 5
	쇼핑평균시간	1 ~ 3
	쇼핑시간대	1 ~ 8
	이동경로	A-A-B- ... -E-D

표 2. 사용자 프로파일
Table 2. Users Profile

번호	성별	나이	소득	결혼	직업	시간	자녀	시간대	지출
1	1	53	5	1	6	1	2	6	4
2	2	23	1	0	2	2	0	7	2
3	2	35	2	1	3	2	1	4	3
4	1	33	2	0	3	1	0	4	3
5	1	44	4	1	6	1	1	5	2
...

3.2 사용자 이동패턴 분석

사용자의 이동경로를 획득하기 위해 쇼핑몰을 20개의 위치로 구분한다. 그리고 사용자가 쇼핑몰에 들어오는 순간부터 10초 단위로 사용자의 위치를 모니터링 한다. 10초 단위로 측정하는 이유는 사람이 붐비는 경우 상품에 관심이 있어 한 장소에 머무는 것이 아니고, 움직이지 못하는 경우를 고려해야 하기 때문이다. 설문조사 결과 관심이 있는 상품 앞에 머무는 시간이 평균 20초 이상인 것으로 나타났다. 그러므로 특정 위치에 머문 시간이 20초를 넘게 되면 사용자가 그 위치에 있는 상품에 관심이 있는 것으로 판단하게 된다.

사용자의 컨텍스트 정보 중 특정 위치의 체류시간은 상품을 구매하지 않았다 하더라도 잠재적으로 구매를 할 확률이 높은 것으로 예측할 수 있기 때문에 상품 구매 이력과는 별도로 가중치를 주게 된다. 또 이동패턴의 정확한 분석을 위해 사용자와 비슷한 성향을 가지는 그룹의 이동패턴을 분류하여 지역모델을 만들고, 새로운 데이터를 지역모델로 평가해 사용자의 이동패턴을 예측하도록 하였고, 이를 위해 RSOM을 사용한다[11].

3.2.1 경로보정 알고리즘

사용자에게 효율적으로 추천하기 위해서는 얼마나 정확하게 사용자의 다음 이동경로를 예측할 수 있는지가 중요하다. RSOM의 입력으로 사용되는 시퀀스의 정확도를 향상시키기 위해 사용자의 이동경로를 다음의 경로보정 알고리즘으로 보정한다.

```

Let  $U_A$  denote a User  $A$ 
Let  $L_A$  denote a resent location of User  $A$ 
Let  $stk$  denote a stack
Let  $T$  denote a fixed time interval
Let  $H_A = \{L_{in}, \dots, L_{out}\}$  denote a time ordered path history
Let  $L_{in}$  denote a away in
Let  $L_{out}$  denote a away out
Let  $S_{ij} \in S$  denote a set of Sequence  $L_i$  to  $L_j$ 

for each  $U_A$  do
  if  $L_A = L_{in}$  then
     $stk.push(0, L_{in})$ 
  while( $L_A \neq L_{out}$ )
    check  $L_A$  for each  $T$ 
    set  $j=1$ 
     $stk.push(L_{in+j}, L_A)$ 
    increment  $j$ 
  if  $L_A = L_{out}$  then
     $stk.push(L_{i+j}, L_{out})$ 

for  $k=L_{in}$  up to  $L_{out}$  do
  if( $L_k = L_{k+1}$ ) then
    set  $L_k = L_k'$ 
    insert  $L_k'$  into a  $H_A$ 
  else
    insert  $L_{k+1}$  into a  $H_A$ 
    set  $S_{ij} = S_{k,k+1}$ 
    insert  $S_{k,k+1}$  into  $S$ 

end;
```

① 각각의 사용자에게 대하여 사용자의 현재 위치를 컨텍스트 인지 센서에서 10초 단위로 모니터링 하여 입구에서부터 출구까지의 이동경로를 생성한다.

② 20초 이상 같은 위치에 머물렀을 경우 그 위치에 있는 상품에 관심이 있는 것으로 판단하고, 해당하는 경로만 추출하여 경로를 생성한다. 이 경로는 쇼핑물 내에서의 관심 있는 상품들로 이루어진 경로이다.

③ ②에서 생성된 경로에서 위치가 중복되는 경우 중 20초 이상 체류했을 때, 그 장소에 있는 상품에 관심이 있는 것으로 파악하여 가중치를 증가시킨다.

④ 위 과정을 거친 경로를 기반으로 구매가 일어난 경로에 가중치를 부여하여 weighted-path를 생성하고 최종 경로로 만든다.

⑤ 생성된 사용자들의 최종 경로들을 RSOM의 입력으로 쓰일 입력 시퀀스로 만든다. 사용자의 위치이동의 시퀀스들이 RSOM의 입력으로 사용되며 한 개의 시퀀스는 한 번의 이동을 나타낸다.

3.2.2 지역모델 생성과 이동경로 예측

사용자의 이동경로를 예측하여 상품을 추천하기 위해서는 사용자가 쇼핑물에 들어올 때부터 센서로 움직임을 계속 모니터링하여 사용자와 이동패턴이 비슷한 사람들의 이동패턴을 바탕으로 사용자의 이동경로를 예측해야한다. 즉, 새로운 경로데이터가 들어오면 이를 이미 분석된 이동패턴으로 평가하여 가장 유사한 패턴을 선택함으로써 이동경로를 예측해야한다.

하나의 분류기를 이용해 모든 사용자의 이동패턴을 분석하면 시간이 많이 걸리기 때문에 실시간으로 사용자의 위치를 예측하는 것이 어렵다. 이 문제를 해결하기 위해, 획득된 경로데이터를 클러스터링하여 작은 지역모델로 만들어 평가해야 한다. 본 논문에서는 이를 위해 RSOM을 사용한다.

기본적으로 SOM은 비감독학습 신경망으로 클러스터링이나 벡터 양자화 문제에 많이 사용되어 왔다. RSOM은 시퀀스 데이터의 처리에 적합하도록 변형된 것으로 일반적인 SOM이 가지고 있는 특징들을 모두 계승한다. 기본적으로 SOM은 비감독학습을 하기 때문에 사전지식이 없이도 유사한 이동패턴들의 집합을 발견할 수 있어 사용자의 간섭 없이 이동패턴을 클러스터링 할 수 있는 장점이 있다.

그림 2는 지역모델로 이동경로를 예측하는 과정으로, 점선 부분은 보정된 경로로부터 지역 모델을 생성하는 과정이고, 실선부분은 이동경로를 예측하는 과정이다.

지역모델의 생성은 입력 시퀀스인 보정된 경로를 RSOM으로 학습시켜 여러 개의 유사한 시퀀스 별로 사용자의 이동패턴을 클러스터링하여 생성한다. 클러스터링된 사용자의 이동패턴이 각 위치에서의 지역모델이 된다. 입력된 시퀀스는 식(1)로 계산하여 가장 마지막 최정합 노드(best matching node)에 따라 클러스터링 한다. 그리고, 식(2)에 의하여 계산된 최정합 노드 변화 시퀀스의 집합을 지역모델로 생성한다.

이동경로의 예측은 새로운 경로 데이터가 들어왔을 때 시퀀스를 RSOM에 입력시키켜 학습시키면 그에 해당하는 최정합 노드가 출력된다. 현재까지 출력된 노드 시퀀스는 각 지역모델에 의해 평가되어 다른 지역모델에 비해 특히 높은 확률을 보이는 지역모델이 있을 경우 즉시 선택되어 사용자의 다음 이동경로를 예측한다. 평가를 위한 가중치는 식 (3)으로 계산하여 가장 유사한 지역모델을 선택한다.

지역모델을 만들고 이동경로를 예측하기 위한 RSOM의 학습 알고리즘은 다음과 같다.

$x(n)$ 을 시퀀스의 n 번째 데이터, α 를 현재 데이터의 가중치, $w_i(n)$ 을 i 번째 노드의 가중치라고 할 때, 시퀀스의 n 번째 데이터에서 출력층의 i 번째 노드의 출력값 $y_i(n)$ 은 다음 식 (1)과 같이 계산한다.

$$y_i(n) = (1 - \alpha)y_i(n-1) + \alpha(x(n) - w_i(n)) \quad (1)$$

이 때 바로 전 데이터 값의 출력 노드의 값 $y_i(n-1)$ 도 반영함으로써 데이터 사이의 상태정보를 유지하게 한다. 시퀀스의 n 번째 단계에서 입력 시퀀스와 가장 가까운 최정합 노드 y_b 는 다음 식 (2)와 같이 계산한다.

$$y_b = \min_i \{ |y_i(n)| \} \quad (2)$$

y_b 가 계산된 후 다음 단계에서의 가중치 $w_i(n+1)$ 은 $\gamma(n)$ 을 학습률, $h(b(n), i)$ 를 이웃 함수라고 할 때 다음 식 (3)과 같이 계산한다.

$$w_i(n+1) = w_i(n) + \gamma(n)h_{b(n)}(x(n) - w_i(n)) \quad (3)$$

이러한 과정을 거쳐 한 개의 입력 시퀀스의 처리가 끝나면 모든 출력 노드의 값을 0으로 설정한 후 새로운 데이터로 위의 과정을 반복한다.

3.3 사용자의 선호도 계산

사용자의 상품에 대한 선호도를 계산하기 위해 사용자의 구매이력과 체류시간을 사용한다.

체류시간은 사용자가 해당하는 위치에 들어간 시간부터 나온 시간을 측정하며 일정 시간(20초) 이상을 체류하게 되면 해당 상품에 관심이 있는 것으로 판단하여 가중치 값을 증가시켜 준다. 또, 추천 모듈로부터 추천받은 상품을 사용자가 구매했을 경우에는 긍정적 피드백으로 가중치를 증가시키고, 구매하지 않은 경우에는 부정적 피드백으로 가중치를 감소시킨다. 이후 협력적 필터링을 이용하여 사용자들을 보다 유사한 취향을 가진 그룹으로 다시 그룹화 하고, 그룹 내 사용자들의 컨텍스트를 이용하여 협력적으로 추천한다.

사용자 a 와 k 의 유사도 $w_{a,k}$ 는 식(4)의 Pearson 상관관계 계수식으로 계산한다. 유사도는 -1.0~1.0 사이의 값을 가지

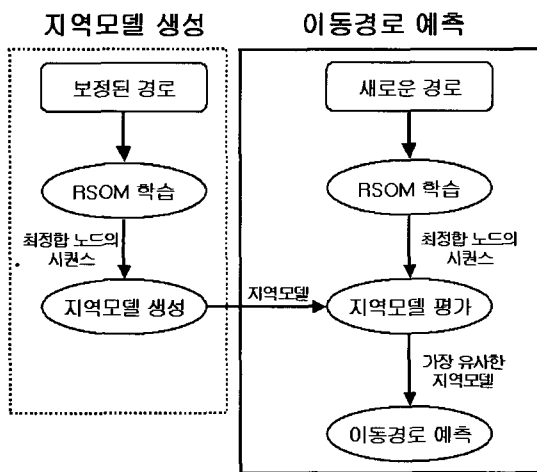


그림 2. 이동경로 예측 처리과정
Fig 2. Process of the Moving Path Prediction

며, 큰 값일수록 유사도가 높음을 의미한다. 계산된 유사도를 식(5)에 적용하여 새로운 상품 i 에 대한 선호도를 예측한다.

$$w_{a,k} = \frac{\sum_j (r_{a,j} - \bar{r}_a)(r_{k,j} - \bar{r}_k)}{\sqrt{\sum_j (r_{a,j} - \bar{r}_a)^2 \times \sum_j (r_{k,j} - \bar{r}_k)^2}} \quad (4)$$

$$P_{a,i} = \bar{r}_a + \frac{\sum_k \{(r_{k,i} - \bar{r}_k) \times w_{a,k}\}}{\sum_k |w_{a,k}|} \quad (5)$$

여기에서 j 는 사용자 a 와 k 가 동시에 선호도를 부여한 아이템을 의미하고, $r_{a,j}$ 는 사용자 a 의 상품 j 에 대한 선호도를 말하고, 그리고 \bar{r}_a 는 사용자 a 의 평균 선호도를 나타낸다.

3.4 사용자 프로파일 갱신

사용자의 프로파일은 식 (6)에서 각각의 상품에 대한 가중치를 곱한 값을 계산한 후 각각의 max 값을 이용, 정규화 하여 갱신할 값을 계산하며 하루에 한 번 사용자 프로파일 DB를 갱신한다.

프로파일에 저장되어 있지 않은 상품에 대한 체류시간이 생겼거나, 새로운 상품에 대한 구매가 일어나면 프로파일에 새로 추가하고 이미 존재하는 내용에 대해서는 기존 값에 갱신 값을 더한다. 갱신 후의 선호도가 0 이하이면 프로파일에서 상품에 대한 선호도를 삭제한다.

$$V_{a,j} = \sum_{j \in F(+)} pr_{a,j} + \sum_{j \in F(-)} nr_{a,j}$$

$$V_{a,j} = \frac{V_{a,j}}{\max(V_{a,j}, j)} \quad (6)$$

$$r_{a,j} = r_{a,j} + V_{a,j} \cdot j$$

여기에서 a 는 사용자, j 는 상품을 의미하고, $F(+)$ 는 긍정적 피드백을 얻은 상품의 집합, $F(-)$ 는 부정적 피드백을 얻은 상품의 집합을 말하고, 그리고 $pr_{a,j}$ 는 긍정적 피드백을 얻은 상품의 가중치, $nr_{a,j}$ 는 부정적 피드백을 얻은 상품의 가중치를 나타낸다.

4. 실험 및 평가

4.1 실험환경

실험을 위한 환경은 펜티엄 IV 3.0GHz, 1GB RAM 시스템으로, MySQL 4.0을 사용하였고, Windows XP 환경에서 실험하였다.

사용자의 이동경로를 파악하기 위해 C마켓의 구조를 바탕으로 단순화시켰고, 이동할 수 있는 위치는 20개로 하였다. 사용자의 선호도를 측정하기 위한 데이터들은 설문지에 포함된 구매이력을 사용하였다. 설문은 525명을 대상으로 하였고, 이동경로는 쇼핑 시 주된 이동경로를 사용자가 설문지에 직접 작성하도록 하였다.

4.2 평가 기준

평가 기준은 제안한 방법으로 사용자의 이동경로를 예측한 후 예측의 정확도를 평가하는 것으로 한다. 정확도는 k-fold cross-validation 방법을 이용하여 학습과 테스트를 k 번 반복 수행하여 측정한다. 이 방법은 데이터 셋을 k 개의 상호 배반 부분 집합으로 분할한 후 i 번째 집합을 테스트 집합으로 나머지를 학습 집합으로 사용한다[12]. k 번의 실험에서 MAE(Mean Absolute Error)를 측정하여 예측의 정확도를 평가한다. MAE는 예측된 평가 값이 작을수록 정확도가 높다고 할 수 있다.

대상 집합의 실제 평가 값을 $\{r_1, \dots, r_n\}$ 이라 한다면, 예측 값은 $\{p_1, \dots, p_n\}$ 으로 표현하고, 오차 $E = \{e_1, \dots, e_n\} = \{p_1 - r_1, \dots, p_n - r_n\}$ 이라하면 MAE의 계산식은 다음 식(7)과 같다.

$$|\bar{E}| = \frac{\sum_{i=1}^N |e_i|}{N} \quad (7)$$

실험을 통하여 최적화된 임계값을 발견하고 임계값 미만의 데이터를 분석단계에서 제외시킨다. 이와 같이 관련성이 적은 데이터를 제거함으로써 학습데이터의 신뢰도와 추천의 정확성을 향상시킬 수 있다.

4.3 실험결과 및 분석

제안한 방법으로 사용자의 이동경로를 예측했을 때의 정확성을 측정하기 위한 실험을 하였다. 최적의 임계값을 구하기 위해 데이터를 학습데이터와 실험데이터로 분류하고 k-fold cross-validation 방법을 이용하였다. k 값은 10으로 설정하였으며, 임계값 별로 MAE를 구하였다. 실험 결과는 그림 3과 같다.

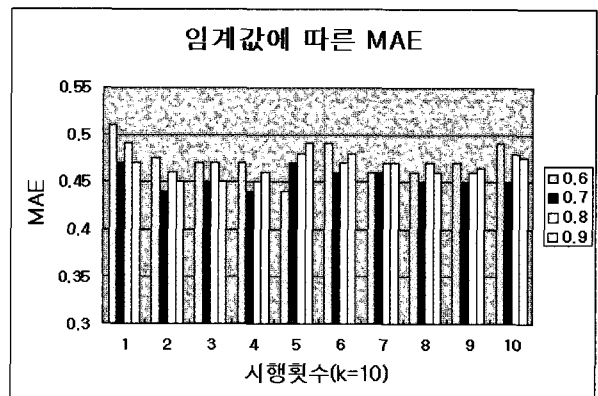


그림 3. 임계값에 따른 MAE
Fig 3. MAE according to the Threshold

그림 3의 실험 결과를 보면 추천을 위한 임계값이 0.7일 때 다른 임계값에서의 MAE보다 평균적으로 좋은 것을 알 수 있다. 임계값이 0.6 이하인 경우는 불필요한 학습데이터가 많은 경우이고, 임계값이 0.8 이상인 경우는 분석하기에 불충분한 양의 학습데이터가 있을 경우이다.

테스트 결과 임계값을 0.7로 했을 경우, 이동경로 예측 MAE가 0.5 이하로 측정됨으로써 사용자의 이동경로를 올바르게 예측할 수 있음을 확인하였다.

5. 결 론

본 논문에서는 오프라인 쇼핑몰에서 사용자의 이동패턴을 분석하고 예측하여 개인화된 상품 추천을 하는 방법을 제안하였다. 이 방법은 사용자의 이동경로를 경로보정 알고리즘으로 보정하여 RSOM을 위한 입력 시퀀스로 만들고, 생성된 입력 시퀀스들을 RSOM으로 학습하여 지역모델을 생성한다. 그 후 새로운 데이터가 입력되면 이를 지역모델로 평가함으로써 이동경로를 예측한다. 그리고 사용자의 예측 이동경로 상에 있는 상품 중에서 사용자의 선호도가 높은 상품을 추천하게 된다.

이동경로의 예측 정확도는 k-fold cross-validation을 이용하여 학습과 테스트를 k번 반복 수행하여 측정하였다. 최적의 임계값을 구하기 위해 데이터를 학습데이터와 실험데이터로 분류하고 k값은 10으로 설정하였으며, 임계값 별로 MAE를 구하였다. 실험 결과 추천을 위한 임계값이 0.7일 때 다른 임계값에서의 MAE보다 평균적으로 좋은 것을 알 수 있었다. 또한, 임계값을 0.7로 했을 경우, 이동경로 예측 MAE가 0.5 이하로 측정됨으로써 사용자의 이동경로를 올바르게 예측할 수 있어 사용자의 이동경로에 따른 상품의 추천이 가능함을 확인하였다.

향후 실제 환경에서 더 많은 데이터를 이용하여 추천의 정확도를 평가하고, 결과를 분석하는 연구를 수행할 예정이다.

참 고 문 헌

[1] M. Weiser, "Some computer science issues in ubiquitous computing," Communication of the ACM, Vol.36, Issue 7, pp.75-84, July 2003.

[2] A. K. Dey, and G. D. Abowd, "A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications," Human Computer Interaction, Vol.16, pp.97-166, 2001.

[3] B. Schilit, N. Adams, and R. Want, "Context-Aware Computing Applications," 1st International Workshop on Mobile Computing Systems and Application., pp.85-90, 1994.

[4] A. K. Dey, and G. D. Abowd, "Towards a Better Understanding of Context and Context-Awareness," In Proceedings of the CHI 2000 Workshop on The What, Who, Where, When, and How of Context-Awareness (The Hague, Netherlands), Apr. 2000.

[5] 장세이, 우운택, "유비쿼터스 컴퓨팅 환경을 위한 컨텍스트 기반 애플리케이션 구조", 한국정보과학회 HCI 논문집, 제2권, pp.346-351, 2003.

[6] J. Pascoe, "Adding Generic Contextual Capabilities to Wearable Computers," 2nd International Symposium on Wearable Computers, pp. 92-99, 1998.

[7] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Analysis of Recommendation Algorithms for E-Commerce," The ACM E-Commerce 2000 Conference, 2000.

[8] Lyle H. Unger and Dean P. Foster, "Clustering

Methods for Collaborative Filtering," Proceeding of the 1998 Workshop on Recommendation Systems, pp. 114-129, 1998.

[9] 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.

[10] D. Ashbrook and T. Starmer, "Learning Significant Locations and Predicting User Movement with GPS," Proc. of IEEE Sixth Int. Symp. on Wearable Computing, Seattle, WA. October, 2002.

[11] T. Koskela, M. Varsta, J. Heikkonen, and K. Kaski, "Temporal Sequence Processing using Recurrent SOM," Proc. of Second Int. Conf. on Knowledge-Based Intelligent Engineering Systems, vol 1, pp. 290-297, Adelaide, Australia, April 1998.

[12] J. Han, M. Kamber, "Data Mining : Concepts and Techniques," Morgan Kaufman, 2001.

저 자 소 개



최영환(Young-Hwan Choi)

1999년 공주대학교 전자계산학과(학사)
 2001년 공주대학교 대학원 전자계산학과(이학석사)
 2001년~현재 공주대학교 대학원 컴퓨터공학과 박사과정 수료

관심분야 : 인공지능, 유비쿼터스 컴퓨팅, 에이전트, 개인화 서비스 등

E-mail : cyhmad@kongju.ac.kr



이상용(Sang-Yong Lee)

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

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

관심분야 : 인공지능, 에이전트, 컴퓨터게임, RFID 시스템 등
 E-mail : sylee@kongju.ac.kr