

모바일 장비에서 사용자의 선호도를 반영한 베이지안 네트워크 기반 추천 시스템

박문희^o 조성배

연세대학교 컴퓨터과학과

moonypark@sclab.yonsei.ac.kr^o, sbcho@yonsei.ac.kr

A BN-based Recommendation System Reflecting User's Preference in Mobile Devices

Moon Hee Park^o, Sung-Bae Cho

Dept. of Computer Science, Yonsei University

요 약

무선통신의 발달에 따라 모바일 장비 기반의 이동성을 고려한 서비스에 관한 연구가 활발하다. 모바일 장비는 제한된 화면크기, 부족한 리소스 등의 한계와 함께 사용자의 이동 중에 발생하는 이벤트를 처리해야 한다는 문제가 있기 때문에, 사용자에게 친숙한 인터페이스와 개별화된 추천 서비스가 요구된다. 본 논문에서는 사용자의 선호도를 반영한 베이지안 네트워크를 이용하여 모바일 장비에서 개인화된 추천 시스템을 개발한다. 실시간으로 변화하는 환경에 적응하도록 네트워크를 설계하기 위하여 전문가에 의해 구조를 설계하고, 수집된 사용자 로그를 바탕으로 파라미터를 학습하여 베이지안 네트워크 모델을 생성한 후, 학습된 모델 기반의 추천결과를 실제 콘텐츠와 비교하여 시스템에 매핑시킴으로써 사용자에게 추천한다. 실제 신촌지역 음식점 추천을 대상으로 실험한 결과 그 가능성을 확인할 수 있었다.

키워드: 추천시스템, 컨텍스트 로그, 베이지안 네트워크, 베이지안 파라미터 학습

1. 서 론

최근 사용자 개개인의 취향이나 특성에 맞는 정보를 자동으로 찾거나 추천해 주는 개인화 기술이 크게 요구된다[1]. 개인화 추천 시스템은 자동화된 정보 필터링 기술을 적용하여 사용자의 취향에 맞는 상품을 추천해 주는데, 모바일 컴퓨팅의 발전으로 시간과 장소에 구애받지 않고 컴퓨팅 자원을 자유롭게 이용할 수 있게 되었다. 특히 위치정보를 이용하여 서비스를 제공하는 LBS는 국내외에서 지속적인 성장 가능성을 가진 기술로 위치 정보에 대한 정확한 추적 및 활용 방안에 대하여 활발히 연구가 진행되고 있다.

본 논문에서는 모바일 장비가 갖는 제한된 화면크기와 같은 인터페이스의 한계점을 극복하기 위해 지도기반의 시스템을 구현한다. 그리고 불확실한 정보를 바탕으로 확률적인 추론이 가능한 모델 기반의 대표적인 협업필터링 방법인 베이지안 네트워크를 기반한 사용자의 선호도 반영 추천시스템을 제안한다.

2. 배경 및 관련연구

2.1 모바일 환경에서 컨텍스트와 추천

컨텍스트는 환경이나 장소 등의 상황을 특징화하기 위해서 사용될 수 있는 정보를 의미한다. 상황인지 컴퓨팅에 관한 연구는 위치, ID, 시간정보를 컨텍스트 정보로 사용하여 레스토랑 추천[2], 여행가이드[3], 광고[4] 등의 서비스를 제공한다. 위치는 모바일 사용자가 어디에 있는지 지리적으로 나타내는 정보이며[5], 시간은 시스템이나 서비스 요청 시간에 따른 정보이다[6]. ID는 장비의 현재 사용자와 그의 선호도에 관련한 정보이다.

모바일 폰은 SMS나 다른 상호교환적인 채널을 이용한 'push' 서비스를 통해 더 직접적인 추천을 제공한다[7]. 많은

연구와 시스템에서 사용자 선호도에 근거한 추천방법을 사용해서 모바일 웹에서 추천과 개인화가 연구되고 있다[5]. 많은 연구에서 모바일 웹에서의 컨텍스트를 정의했고, 서비스와 콘텐츠의 추천에 컨텍스트 정보를 적용했다. 표 1은 대표적인 상황인지 추천 시스템을 보여준다.

표 1. 상황인지 추천 시스템

	MIT Media Lab	National Tsing Hua Univ.	Telematica Institute	Fu-Jen Univ.
추천타입	레스토랑	레스토랑	여행정보	광고
컨텍스트 정보	위치	위치, 시간, 날씨	위치, 시간, 날씨, 일정, 쇼핑리스트	위치, 시간, 요금, 콘텐츠 속성
추천 방법	에이전트 상호작용	요청에 따른 검색	다양한 예측 전략	신경망 학습
장비	PDA, GPS	Pocket PC, GPS	모바일, GPS	모바일

2.2 추천 시스템

전통적인 추천 시스템은 협업 필터링 방법을 사용하여 같은 취향이나 취미를 가진 사람들의 정보를 바탕으로 추천할 때 이용한다. 이러한 협업 필터링은 두 가지로 분류할 수 있다.

2.2.1 메모리 기반의 협업 필터링

사용자 기반의 협업 필터링으로 알려져 있는 알고리즘으로 사용자의 평가 패턴에 근거한 사용자 사이의 상관관계를 생성한다. 이 상관관계는 서로 다른 사용자들 사이의 동적 계산으로 계산량이 많아지고, 시스템의 메모리 적재에 많은 자원을

소비하게 된다. 이로 인하여 실시간 예측을 필요로 하는 시스템이나 많은 데이터 집합을 바탕으로 추천하기에는 부적절하다. 대표적으로 상관계수와 벡터유사도 방법이 있다

2.2.2 모델 기반의 협업 필터링

아이템 기반의 협업 필터링으로 알려져 있는 알고리즘으로 오늘날 널리 사용될 뿐만 아니라, 많은 데이터 집합을 바탕으로 추천하기에 무리가 없다. 시스템은 사용자 사이의 유사도가 아닌 사용자가 평가한 아이템 사이의 유사도를 계산한다. 대표적으로 베이지안 네트워크와 클러스터링 방법이 있다

[8]에서는 마이크로소프트 웹 사이트 방문여부, 2주간 님스 네트워크 텔레비전 시청, 2년간 EachMovie(0~5의 평가값)와 같은 3종류의 데이터 집합을 베이지안 네트워크, 상관계수, 베이지안 클러스터, 벡터 유사도 방법을 적용하였다 실험을 통해 베이지안 네트워크와 상관계수 방법이 최상의 성능을 보임을 확인하였다. 따라서 본 논문에서는 실시간 데이터 처리를 요구하는 모바일 환경에 적합하도록 모델기반의 협업 필터링 방법인 베이지안 네트워크 기반의 추천 시스템을 구현한다.

2.3 베이지안 네트워크

베이지안 네트워크(Bayesian Network, BN)는 변수들 간의 원인과 결과 관계를 확률적으로 모델링하는 도구로서 불확실한 환경에서 좀 더 신뢰성 있는 결과를 추론하기 위한 대표적인 방법이다.

베이지안 네트워크의 구조를 모델링한다는 것은 전문가가 변수집합에 속한 변수들 사이의 의존관계 중에서 특히 중요하다고 생각되는 결합 확률에 큰 영향을 미칠 것으로 생각되는 것들만 선택적으로 결정해 주는 과정을 말한다 구조가 결정되면 전문가는 해당 변수들의 조건부 확률 테이블을 채워 베이지안 네트워크의 설계 과정을 마무리하게 된다. 일반적으로 해당 도메인에 대한 전문적인 지식을 가진 사람이 필요하다 그러나 해당 도메인의 전문지식을 획득하기 어렵다면 데이터로부터 직접 학습하는 방법도 있다[9]. 데이터로부터 베이지안 네트워크의 구조와 파라미터를 학습할 수 있는데 파라미터 학습은 각 노드의 확률을 데이터로부터 학습하는 것이다

3. BN 기반의 사용자 선호도 반영 추천 시스템

3.1 개요

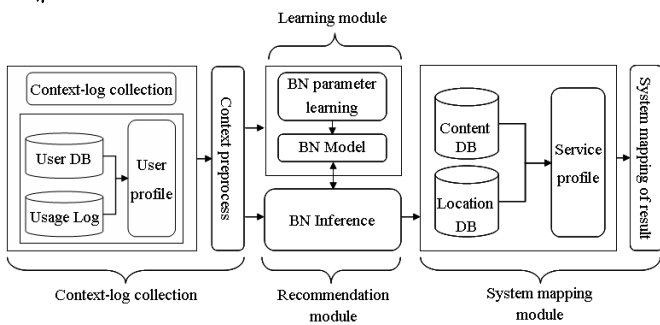


그림 1. 시스템 개요

본 논문에서 제안하는 추천 시스템의 구조는 그림 1과 같다. 우선 사용자의 개인 프로파일 정보와 모바일 장치를 바탕으로 컨텍스트 정보를 수집한다. 수집된 로그 정보를 베이지안 네트워크의 파라미터 학습을 위해 각 노드의 이름과 상태에 맞게 변경하고 전처리를 수행한다 전처리된 컨텍스트 데이터를 바탕으로 각 노드는 파라미터 학습을 진행하여 조건부 확률 테이블을 얻는다. 파라미터 학습된 베이지안 네트워크 모델을 이용하여 사용자의 새로운 요청을 받으면, 각 항목별 최상위 확률 값을 가진 파라미터를 선택한다. 이 최상위 데이터 집합과 일

치하는 실제 정보를 콘텐츠 데이터베이스의 검색을 통해 구성하고, 이를 모바일 장비를 통해 사용자에게 제공한다.

3.2 컨텍스트 정보

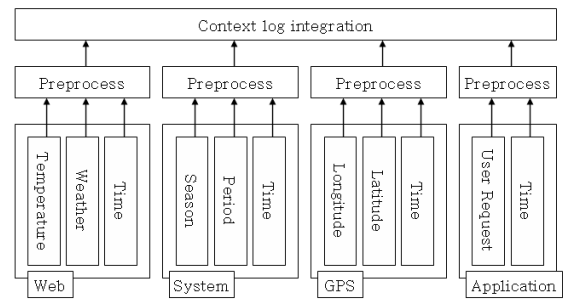


그림 2. 컨텍스트 로그정보

본 논문에서 사용하는 컨텍스트 로그 정보는 그림 2와 같다. 날씨 정보를 제공하는 웹사이트를 통해서 기온과 기상정보를 얻는다. 시스템은 날짜와 시간을 바탕으로 계절과 아침 점심, 저녁으로 구분된 정보를 제공하고 GPS를 통해서 위도, 경도로 표현된 사용자의 현재 위치를 얻는다. 어플리케이션에서는 사용자의 선호도가 반영된 직접적인 요청을 통해 컨텍스트 정보를 얻는다. 각 장치별로 수집된 컨텍스트 정보는 전처리 연산 후, 시간정보를 기준으로 하나의 데이터 파일로 통합한다 개인 프로파일 정보는 사용자의 직접적인 입력을 통해서 선호식당 종류, 성별, 연령, 차량소유, 월 소득, 혈액형 등의 정보를 얻는다.

3.3 베이지안 네트워크 모델 파라미터 학습

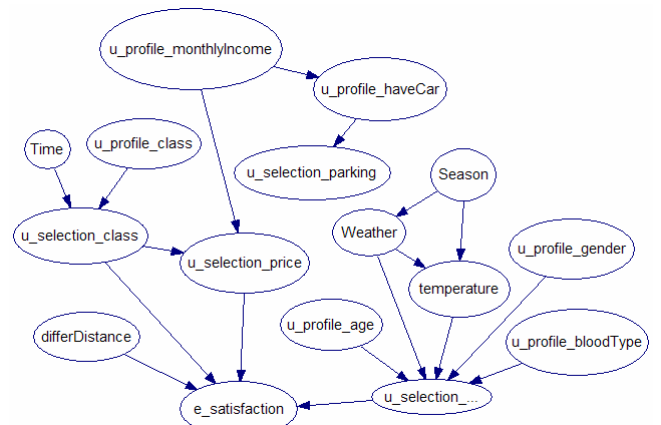


그림 3. BN 구조

기존의 베이지안 네트워크 설계는 전문가의 지식을 바탕으로 하고 있으나, 설계에 많은 시간이 소요되고 네트워크의 정확성을 보장할 수 없다는 문제가 있다. 또한 설계된 베이지안 네트워크는 환경의 변화를 반영할 수 없다. 이런 단점을 극복하기 위해, 파라미터 학습방법을 적용한다 본 논문에서 베이지안 네트워크의 구조는 그림 3과 같이 전문가에 의해 설계하고 파라미터는 수집된 데이터를 바탕으로 학습을 진행하였다 파라미터 학습은 EM 알고리즘(expectation Maximization Algorithm)을 이용하여 조건부 확률 테이블을 생성한다

3.4 추론과 결과의 시스템 매핑

베이지안 네트워크의 결과로 도출된 확률 값만으로 직접 이

편 서비스를 선택하거나 행동을 취하는 결정을 하기는 어렵다 따라서 본 논문에서는 추론의 결과로 얻은 식당종류 확률 값 집합을 $Class = \{c_1, c_2, \dots, c_l\}$, 식당 가격대 확률 값 집합을 $Price = \{p_1, p_2, \dots, p_m\}$, 식당 분위기 확률 값 집합을 $Mood = \{m_1, m_2, \dots, m_n\}$ 라 하고, 속성에 대한 이름은 각각 $c_i^{name}, p_j^{name}, m_k^{name}$ 라고 하고, 각 노드에 대한 가중치는 $Weight = \{w_{class}, w_{price}, w_{mood}\}$ 라고 가정한다.

$$X_{ijk} = (c_i \times w_{class}) + (p_j \times w_{price}) + (m_k \times w_{mood}) \quad (1)$$

$$Recommend\ Value = \max_{i=1 \dots l, j=1 \dots m, k=1 \dots n} (X_{ijk}) \quad (2)$$

가중치를 가진 각 속성조합에 대한 확률 값은 식(1)과 같이 계산할 수 있다. 식(2)와 같이 모든 속성조합에 대한 확률 값 중에서 가장 큰 확률 값을 Recommendation Value 라고 하고, 이때 RecommendSet = $\{c_i^{name}, p_j^{name}, m_k^{name}\}$ 과 일치하는 실제 음식점 정보를 지도와 함께 모바일 장비에서 확인할 수 있다. 또한 본 논문에서는 모바일 장치의 제한된 화면이 갖는 한계점을 극복하기 위해서 사용자에게 직관적인 지도 인터페이스를 제공하는 시스템을 구현한다. 그림 4의 모바일 추천 시스템은 실제 모바일 장비에서 개별 사용자의 요청에 따른 BN추론결과로 얻어지는 추천정보를 텍스트뷰어와 지도뷰어에 매핑해 사용자가 직관적으로 정보를 획득할 수 있다. 사용된 지도는 Naver Map Open API이다. 개인별로 학습된 모델 선택을 위한 로그인 과정을 거쳐 직접적으로 요청한다

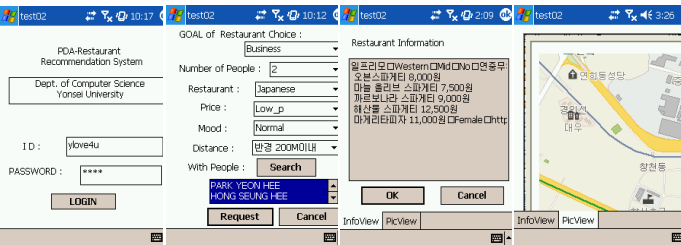


그림 4. 모바일 추천 시스템

4. 실험 및 결과

4.1 실험환경

실험을 위해서 사용된 콘텐츠 데이터로는 신촌지역의 870×500m² 내에 있는 약 50개의 음식점 정보[10]를 이용하였고, 사용자 데이터는 7일(11/1~11/7)동안 각각 2명의 남녀 사용자에게 의해 수집된 데이터를 7-fold cross validation하여 파라미터 학습을 진행하였다. 사용자 프로파일 정보로 표 2와 같이 연령, 성별, 혈액형, 차량소유 여부, 월 소득, 선호음식을 수집하였다. 일반적으로 음식에 대한 기호도 조사 연구에서 조사 대상자에 대한 특성은 성별, 연령, 직업, 경제력 및 교육수준을 포함한다 [11]. 본 논문에서는 추천하는 음식점의 주차장 여부에 대한 판단을 위해 차량소유 정보도 수집하였다. 표 3은 수집된 컨텍스트 정보를 시간, 사용자 요청, 사용자 프로파일, 위치, 날씨, 평가 항목으로 전처리한 데이터 집합의 형식이다. 시간항목은 다시 계절과 시간으로 구분하였다. 계절은 12개월을 3개월 단위로 봄(3월~5월), 여름(6월~8월), 가을(9월~11월), 겨울(12월~2월)로 구분하였고, 시간은 하루를 3등분하고 일반적인 식사 시간을 기준으로 아침(3:00~11:00), 점심(11:00~18:00), 저녁(18:00~3:00)으로 구분하였다. 사용자 요청 정보는 식당의 종류, 분위기, 가격대, 주차 공간의 유무를 직접적인 입력을 통해서 획득

한다. 사용자 프로파일 정보는 시스템 초기 사용 시, 사용자의 특징을 입력받는다. 위치는 실험지역을 3등분하여 사용자의 현재 위치를 바탕으로 반경 100m내의 거리는 Near, 반경 200m내의 거리는 Mid, 반경 300m내의 거리는 Far로 나타낸다. 날씨정보는 기상청 홈페이지의 시간대별 자료를 바탕으로 Sunny, Rain, Cloudy, Snow로 나타내고, 기온은 기상청의 30년간 서울 지방의 평균기온에 관한 통계자료를 바탕으로 Warm(13~20°C), Hot(20~30°C), Cool(7~13°C), Cold(-5~7°C)로 구분하였다[12].

표 2. 사용자 프로파일 정보

	사용자-1	사용자-2	사용자-3	사용자-4
연령	20대 중반	30대 초반	20대 후반	30대 초반
성별	Female	Female	Male	Male
혈액형	B	AB	B	O
차량소유	No	Yes	No	Yes
월 소득 (만원)	150~200	250~300	100~150	250~300
선호음식	한식, 일식, 중식, 양식	한식, 일식, 양식	한식, 일식, 중식	한식, 일식, 중식, 양식

표 3. 수집된 로그 데이터 집합

	노드	값
시간	계절	Spring, Summer, Fall, Winter
	시간	Breakfast, Lunch, Dinner
사용자 요청	식당종류	Korean, Japanese, Chinese, American, Italian, Etc
	식당 분위기	Romantic, Tidy, Exotic, Normal
	가격대	Low, Low_Mid, Mid, Mid_High, High
사용자 프로파일	주차 공간	Yes, No
	선호식당 종류	Korean, Japanese, Chinese, American, Italian, etc
	성별	Male, Female
	연령	10대, 20대, 30대
	차량소유	Yes, No
	월 소득 (만원)	100~150, 150~200, 200~250, 250~300
	혈액형	A, B, AB, O
위치	거리차	Near, Mid, Far
날씨	날씨	Sunny, Rain, Cloudy, Snow
	기온	Warm, Hot, Cool, Cold

4.2 실험결과

수집된 데이터를 바탕으로 파라미터 학습된 베이지안 네트워크 모델은 사용자의 선호 식당 카테고리, 가격대, 식당 분위기를 추론한다. 사용자 프로파일 정보, 시스템 정보, 웹으로부터 획득되는 날씨정보는 증거변수로 채택되고 나머지 노드들은 추론되어 각 노드별 속성의 확률분포를 얻게 된다. 표 4는 사용자의 각 노드별로 추론된 속성의 확률분포의 결과를 나타낸다. 사용자-1의 경우에 아침은 일식, 중저가, 깔끔한 분위기, 점심은 한식, 중저가, 로맨틱한 분위기, 저녁은 양식, 저가, 로맨틱한 분위기에 대한 추론 확률 값이 높다. BN추론확률결과와 음식점 종류, 가격대, 분위기의 순서로 설정된 가중치에 따른 연산을 통해 시스템은 확률분포가 가장 높은 추천속성집합을 구성한다. 표 5는 추천속성집합을 바탕으로 실제 콘텐츠 데이터와 유사도가 높은 실제 추천된 상위 2개에 랭킹된 음식점 데이터들의 속성 값이다.

표 4. 사용자별 추천 결정된 항목과 확률 (R: Restaurant)

		R_Class		R_Price		R_Mood	
user 1	아침	Japanese	0.76	Low_Mid	0.54	Tidy	0.46
	점심	Korean	0.58	Low_Mid	0.84	Romantic	0.33
	저녁	American	0.42	Low	0.51	Romantic	0.33
user 2	아침	Japanese	0.75	Low	0.64	Normal	0.57
	점심	Korean	0.57	Low	0.45	Normal	0.29
	저녁	American	0.45	Low	0.88	Normal	0.29
user 3	아침	Japanese	0.72	Low_Mid	0.30	Exotic	0.45
	점심	Korean	0.52	Low_Mid	0.65	Tidy	0.39
	저녁	American	0.38	Low_Mid	0.42	Tidy	0.39
user 4	아침	Japanese	0.76	Low	0.65	Normal	0.55
	점심	Korean	0.58	Low	0.45	Romantic	0.43
	저녁	American	0.42	Low	0.87	Romantic	0.43

표 5. 사용자별 실제 추천된 음식점 리스트 결과 (상위 2개 랭킹만) R: Restaurant

		R_name	R_Class	R_Price	R_Mood
user1	아침	쿠마짱	Japanese	Low_Mid	Tidy
	점심	요이시스	Japanese	Low_Mid	Tidy
	저녁	놀부부대찌개	Korean	Low_Mid	Normal
		홍초불닭	Korean	Low_Mid	Normal
user2	아침	넛바	Italian	Mid	Romantic
		일프리모	American	Mid	Romantic
	점심	쿠마짱	Japanese	Low_Mid	Tidy
		요이시스	Japanese	Low_Mid	Tidy
저녁	닭한마리	Korean	Mid	Normal	
	굽터	Korean	Low_Mid	Normal	
user3	아침	아웃백 스테이크	American	Mid	Exotic
		우노	American	Mid	Exotic
	점심	쿠마짱	Japanese	Low_Mid	Tidy
		요이시스	Japanese	Low_Mid	Tidy
저녁	호박이 넝쿨찌	Korean	Low_Mid	Tidy	
	놀부부대찌개	Korean	Low_Mid	Normal	
user4	아침	아웃백 스테이크	American	Mid	Exotic
		우노	American	Mid	Exotic
	점심	쿠마짱	Japanese	Low_Mid	Tidy
		요이시스	Japanese	Low_Mid	Tidy
저녁	굽터	Korean	Low_Mid	Normal	
	호박이 넝쿨찌	Korean	Low_Mid	Tidy	

표 4의 결과와 표 5를 비교해보면, 음식점 종류 항목은 베이지안 네트워크 확률 추천결과를 충실히 이행하고 있다 하지만, 가격대와 분위기는 추천결과와 실제 데이터와의 매핑 사이에 차이가 있다. 이는 추천된 속성과 일치하는 콘텐츠 데이터의 양이 적기 때문이다. 표 6는 상위랭킹을 고려한 추천결과와 실제 데이터 사이의 정확률을 보여준다. 음식점 종류 항목은 랭킹 2까지 고려했을 경우, 정확률이 더 높아졌다. 반면에 분위기 항목은 랭킹 2까지 고려를 하면 정확률이 감소하는데 이는 음식점 종류, 가격대, 분위기의 순서로 우선순위를 갖기 때문이다.

표 6. 추천결과와 실제 데이터 매핑 사이의 정확률

	R_Class	R_Price	R_Mood
상위 1랭킹	83.33%	33.33%	41.66%
상위 2랭킹	91.66%	33.33%	37.5%

5. 결론 및 향후연구

본 논문에서는 모바일 장비의 제한된 화면크기, 부족 리소스의 문제점을 극복하기 위해서 사용자에게 친숙한 지도 기반의 인터페이스를 구현하였고, 사용자의 프로필 정보와 모바일 장비로부터 얻을 수 있는 컨택스트 정보를 이용하여 사용자 선호도를 반영한 베이지안 네트워크 기반 추천 시스템을 제안하였다.

향후 과제로는 본 논문에서 수행한 실험의 다각적인 분석과 사용자 테스트를 통해 유용성을 검증할 필요가 있다. 또한 베이지안 네트워크의 추천결과와 실제 콘텐츠 데이터 사이에 정확한 매핑을 위해서 콘텐츠 데이터의 양적 보충이 요구되며 유사한 성향을 갖는 사용자 집합 모델 형성을 위해서 더 많은 양의 사용자 데이터 수집이 필요하다.

참고문헌

- [1] M. J. Pazzani, "A framework for collaborative, content-based and demographic filtering," *Artificial Intelligence Review*, vol. 13, no. 5-6, pp. 393-408, 1999.
- [2] G. Tewari, J. Youll and P. Maes, "Personalized location-based brokering using an agentbased intermediary architecture," *Decision Support Systems*, vol. 34, no. 2, pp. 127-137, January 2003.
- [3] M. C. Setten, S. Pokraev and J. Koolwaaij, "Context-aware recommendations in the mobile tourist application COMPASS," *AH 2004*, pp. 235-244, 2004.
- [4] S. T. Yuan and Y. W. Tsao, "A recommendation mechanism for contextualized mobile advertising," *Expert Systems with Applications*, vol. 24, no. 4, pp. 399-414, 2003.
- [5] I. Lee, J. Kim and J. Kim, "Use contexts for the mobile internet: A longitudinal study monitoring actual use of mobile internet services," *International Journal of Human-Computer Interaction*, vol. 18, no. 3, pp. 269-292, 2005.
- [6] T. Hofer, W. Schwinger, M. Pichler, G. Leonhartsberger and J. Altmann, "Context-awareness on mobile devices: The hydrogen approach," *Proceeding of the 36th Hawaii International Conference on System Sciences*, 2003.
- [7] S. Y. Ho, and S. H Kwok, "The attraction of personalized service for users in mobile commerce: An empirical study," *ACM SIGecom Exchanges*, 2003.
- [8] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," *Proceedings of the 14th UAI*, 1998.
- [9] G. F. Cooper and E. A. Herskovits, "A bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309-347, 1992.
- [10] <http://www.cocofun.co.kr/>
- [11] K. J. Cho, "A study on the relationship between lipid intake style and cardiovascular disease of the elderly," *KHEA*, vol. 40, no. 5, pp.1-14, 2002.
- [12] <http://www.kma.go.kr/intro.html>