

디지털 TV에서 시멘틱 환경의 유헬스 서비스를 위한 나이브 베이지안 필터링 기반 개인화 서비스 추천 방법

김재권*, 이영호**, 김종훈***, 박동균****, 강운구**

Semantics Environment for U-health Service driven Naive Bayesian Filtering for Personalized Service Recommendation Method in Digital TV

Jae-Kwon Kim*, Young-Ho Lee**, Jong-Hun Kim***, Dong-Kyun Park****, Un-Gu Kang**

요약

디지털 TV에서 시멘틱 환경의 유헬스 개인화 서비스 추천은 개인의 신체조건, 질병, 건강상태를 평가해서 이루어져야 한다. 기존의 시멘틱 환경의 유헬스 개인화 추천 방법은 온톨로지에 의존하여 의미 분석으로 추천을 하기 때문에 사용자 만족도가 떨어진다. 이에 본 논문에서는 디지털 TV에서 시멘틱 환경의 유헬스 서비스를 위한 나이브 베이지안 필터링 기반 개인화 서비스 추천 방법을 제안한다. 제안하는 방법은 온톨로지를 이용하여 상황데이터를 추론하여 트랜잭션을 저장하고, 선호도 정보를 이용한 나이브 베이지안 필터링 기법을 사용하여 온톨로지로부터 생성된 트랜잭션과 사용자 선호도 정보를 이용하여 추론하여 서비스를 제공한다. 나이브 베이지안 필터링 기반으로 추론된 서비스는 기존의 필터링 방법 보다 콘텐츠 추천의 높은 정확도와 재현율을 보인다.

▶ Keywords : 디지털TV, 유헬스 서비스, 개인화 서비스 추천, 나이브 베이지안 필터링

Abstract

For digital TV, the recommendation of u-health personalized service of semantic environment

• 제1저자 : 김재권 • 교신저자 : 강운구

• 투고일 : 2012. 6. 19, 심사일 : 2012. 7. 1, 게재확정일 : 2012. 7. 20.

* 인하대학교 컴퓨터공학과(Dept. of Computer Science & Information Engineering, Inha University)

** 가천대학교 IT 대학(College of IT, Gachon University)

*** 비트컴퓨터(Bit Computer Co.)

**** 가천대 길병원 유헬스케어 센터(u-Healthcare Center, Gachon University Gil Hospital)

※ 본 연구는 지식경제부 및 한국산업기술평가관리원의 산업원천기술개발사업의 일환으로 수행하였음[과제번호 : 10032115, 과제명 : 디지털 TV 기반의 AI를 활용한 u-Health 시스템 개발]

should be done after evaluating individual physical condition, illness and health condition. The existing recommendation method of u-health personalized service of semantic environment had low user satisfaction because its recommendation was dependent on ontology for analyzing significance. We propose the personalized service recommendation method based on Naive Bayesian Classifier for u-health service of semantic environment in digital TV. In accordance with the proposed method, the condition data is inferred by using ontology, and the transaction is saved. By applying naive bayesian classifier that uses preference information, the service is provided after inferring based on user preference information and transaction formed from ontology. The service inferred based on naive bayesian classifier shows higher precision and recall ratio of the contents recommendation rather than the existing method.

▶ Keywords : Digital TV, U-health Service, Personalized Service Recommend, Naive Bayesian Filtering

I. 서론

디지털 TV는 기존의 아날로그 TV에서는 구현할 수 없었던 다양한 기능과 서비스들을 수용하면서 발전하고 있다. 디지털 TV를 통해 여러 기능과 서비스를 제공할 수 있게 된 것은 TV에 고성능 컴퓨팅 OS환경의 결합으로 가능하게 되었다 [1]. 유 헬스(Ubiquitous Health) 서비스의 중요성이 점차적으로 증가하고 있는 가운데, 기존의 웹 환경에서 제공했던 서비스가 스마트폰, IPTV 등 여러 디바이스를 활용한 멀티 플랫폼 환경[2]으로 발전하면서, 최근에는 디지털 TV에 유헬스 서비스를 제공하는 방식이 주목을 받고 있다.

디지털 TV기반의 유헬스 서비스를 제공하기 위해서는 개인 데이터를 통해 사용자의 상황에 알맞은 서비스를 제공해야 하며, 최근에는 상황 데이터의 추론을 위해 온톨로지를 이용하여 서비스 제공에 높은 효과를 보이고 있다[3]. 개인화 서비스 제공을 위한 시멘틱 환경의 유헬스 서비스는 개인의 질병 데이터, 위치, 날씨 등의 상황데이터를 기반으로 하여 추론을 하는 것으로서, 맞춤형 서비스 제공이 가능하다[4]. 유헬스 서비스의 개인화 추천은 사용자의 질병, 생체 등의 정보와 추천 정보의 로그를 바탕으로 지능형 개인화 추천서비스가 이루어진다[5]. 기존의 시멘틱 환경의 유헬스 서비스는 규칙 기반 필터링 방법인 온톨로지를 이용하여 지능형 서비스를 제공한다. 규칙 기반 필터링 방법은 입력 데이터에 대해 의미 분석과 룰 베이스를 이용하여 추천한다. 하지만 규칙 기반 필터링 방법은 사용자의 선호도를 고려하는 개인화 서비스를 제공하기에는 한계가 있으며, 현재까지도 개인화 서비스 추천의

성공적 사례가 없다[6].

의료 서비스 분야는 다른 분야와는 달리 개인 정보, 질병 정보 등의 다양한 속성을 체계적으로 분류하기 어렵고, 개개인마다 원하는 서비스 성향이 다르다. 따라서 기존의 온톨로지를 이용한 규칙 기반 필터링 방법은 개개인의 성향을 적용하기 어려우며, 사용자가 선호하지 않는 콘텐츠도 추천이 될 수 있기 때문에 정확도가 떨어진다. 이를 해결하기 위해 협력적 필터링을 이용하여 서비스의 정확도를 높여야 한다[7]. 따라서 기존의 온톨로지를 이용한 추천 방법에서 서비스 제공의 정확도를 높이기 위한 하이브리드 필터링 방법이 필요하다.

이에 본 논문에서는 디지털 TV에서 시멘틱 환경의 유헬스 서비스를 위한 나이브 베이지안(Naive Bayesian)[8] 필터링 기반 개인화 서비스 추천 방법을 제안한다. 제안하는 방법은 온톨로지를 이용하여 상황데이터를 추론하여 트랜잭션을 저장하고, 서비스 콘텐츠의 필터링을 위해 나이브 베이지안 기법을 사용하여 온톨로지[9]로부터 생성된 트랜잭션과 사용자 선호도 정보를 필터링하여 서비스를 제공한다.

본 논문의 구성은 2장에서는 관련연구를 기술하며, 3장에서는 추천 방법에 대해 기술을 하며, 4장에서는 구현 및 실험을 기술하고, 마지막으로 5장에서는 결론을 기술한다.

II. 관련 연구

1. 유헬스 서비스

유헬스 서비스의 중요성이 높아지는 이유는 현대인들의 생활의 변화, 활동량의 감소 등의 변화로 만성 질환 환자가



그림 1. 디지털 TV 기반 유헬스 서비스 시스템 구성도
 Fig. 1. Digital TV Driven U-health Service System Architecture

급증하고 있기 때문이다. 이에 따라 고혈압, 당뇨병의 합병증이 발생할 가능성이 높으며, 질환자는 자기관리 소홀로 인해 질병으로 인한 큰 고통과 회복이 불가능한 상태로 초래할 수 있다[2]. 이에 여러 가지 디바이스를 통한 유헬스 서비스 모델의 관심이 높아지고 있다.

웹을 통한 유헬스 서비스는 개인별 신체정보, 질병정보의 입력을 바탕으로 사람의 특성에 따라 맞춤형 식단 및 운동 처방을 통해 건강을 관리하는 서비스로, 현재 가장 많이 상용화되고 있다[10]. IPTV를 통한 유헬스 서비스는 기존의 방문 처방과 웹 서비스방식이 아닌 IPTV 플랫폼 기반 서비스를 제공하며, 사용자들에게 친숙한 TV를 통해 서비스를 제공하기 때문에 사용자의 활용도가 높아질 수 있다[11].

2. 온톨로지와 나이브 베이지안 기반 필터링

개인화 추천방법을 위한 필터링 기법은 대표적으로 협력적 필터링, 규칙 기반 필터링이 있다. 협력적 필터링 기법[12]은 사용자들의 서비스 선호 정보에 대한 데이터베이스를 구축하고, 선호 정보에 기반을 두어 새로운 사용자의 유사한 취향을 가진 선호정보를 찾아내어 추천하는 방법이다. 규칙 기반 필

터링[13]은 협업 전문가가 사용자의 성향을 일정한 규칙으로 정의하고 이를 서비스의 특성과 연결시켜 추천하는 방법이다.

위와 같은 필터링 기법들을 이용한 개인화 추천 서비스에 많은 연구가 진행되고 있다. 규칙 기반 필터링 방법인 시멘틱 웹 기술을 이용한 연구는 불필요한 검색시간이 줄어들고 피드백을 통해 점차 만족도가 향상되는 결과를 보인다[14]. 하지만 규칙기반 필터링에 의존하기 때문에 사전에 추천 서비스 만족도에 대한 평가가 필요하다. 따라서 규칙기반 필터링은 선호 정보를 고려하지 않기 때문에 사용자의 만족도를 높이기 위해서는 선호도 정보를 이용하는 협력적 방법이 필요하다. 이에 복잡도가 높은 필터링 대신 능동적인 분석 소요 시간을 줄이기 위한 나이브 베이지안 기반 필터링 방법은 구현이 용이하며 오버헤드가 적다.

협력적 분류 기법인 나이브 베이지안은 속성 집합과 클래스 변수 사이의 확률적 관계를 모델링하는 접근 방법이다. 클래스의 사전 지식과 데이터로부터 획득한 새로운 증거를 사전 확률로 정의하고, 이에 따라 새로운 지식을 나타내기 위해 조건부 확률을 이용하여 한 확률의 변수가 특정 값을 가질 확률을 계산한다[8].

이에 본 논문에서는 디지털 TV를 통한 사용자 서비스 추천을 위해 규칙 기반 필터링으로 구성된 시멘틱 환경에서 사용자 선호도 정보를 이용하는 협력적 기반 필터링을 이용하는 나이브 베이지안 추론 방법을 이용하여 콘텐츠 추천의 정확도를 높여 준다.

III. 디지털 TV기반 시멘틱 유헬스 서비스

디지털 TV에서 시멘틱 유헬스 서비스를 위한 나이브 베이지안 추론 기반 개인화 서비스 추천 방법을 제안하기 위해 서비스를 제공하기 위한 시스템 구성도를 설명하며, 온톨로지 구성에 대해 설명한다.

1. 디지털 TV 서비스 시스템 구성도

유헬스 기반의 디지털 TV 서비스를 제공하기 위한 시스템 구성도는 그림 1과 같다.

사용자는 디지털 TV를 통해 서비스를 이용 받을 수 있으며, 개인의 건강데이터를 전송하기 위해 무구속, 무자각 지향의 센서를 사용하여 데이터를 건강관리 센터에 제공한다.

건강관리 센터는 사용자 관리와 서비스를 제공하며, 디지털 TV 서비스와 웹 서비스를 제공한다.

디지털 TV 건강관리 플랫폼은, 사용자의 데이터와 센서로부터 받은 데이터를 이용하여 맞춤형 서비스를 제공하기 위한 부분이며, AI(Artificial Intelligence)기반 알고리즘과 데이터 마이닝을 통해 건강을 분석 및 평가하여 개인화 콘텐츠 서비스를 제공한다.

의료기관은 사용자의 질병력 데이터를 PHR 표준에 의하여 데이터를 전송 받으며, 서비스 제공자는 외부 전문가로부터 구성된 콘텐츠 데이터를 사용자에게 제공하기 위한 협업 체계이다. 이와 같이 디지털 TV기반 서비스 시스템 구성도를 구성하였으며, 이를 통해 사용자에게 유헬스 서비스를 제공한다.

2. 온톨로지 구성

사용자로부터 데이터를 획득하여 추천 서비스를 추천하기 위해 그림 2와 같이 운동 서비스 콘텐츠를 제공하는 온톨로지를 구성하였다. 온톨로지는 최상위 클래스인 Thing과 5개의 상위 클래스인 Device, Healthy, Disease_History, Individual로 구성되어있으며 각 클래스는 하위 클래스를 갖는다. Device는 건강정보 측정 디바이스의 종류를 구성 하였으며, 각 디바이스마다 측정 할 수 있는 결과 값을 Instance로 구성하였다. Healthy는 사용자 건강 상태를 의미하며,

EQ-VAS와 EQ-5D는 경제성 평가에 사용되는 대표적인 건강관련 삶의 질의 도구이며 가치평가에 대한 점수를 의미한다. Disease는 병원에서 측정한 임상 결과를 통한 상황 데이터이며, 각 질병에 대한 상태와 치료내용, 약물 및 치료 처방 방법에 대한 데이터를 의미한다. Individual은 사용자의 이름, 나이 등의 기본정보와 생활 방식을 포함하고 있으며, 결혼 및 현재 혼인 여부, 흡연 유무 및 흡연 시기, 하루 흡연량 등의 정보와 주량과 관련된 정보를 포함한다. 모든 Instance은 Service에 참조되어 룰 베이스를 참조하여 서비스 콘텐츠를 추천한다. 서비스 콘텐츠는 운동과 식단에 대한 운동 종류와 선호하는 운동 종류를 포함하고 있으며, 이밖에도 한의약 추천, 약물 추천등에 대한 상황 정보를 포함한다.

위와 같이 운동 서비스 콘텐츠 제공을 위한 온톨로지를 구성하였으며, 해당 콘텐츠의 전문 지식에 따르는 콘텐츠 추천을 위해서 표 1과 같이 질환자의 운동 처방을 통해 콘텐츠 추천을 위한 룰 베이스를 ACSM Guide Line2010[16]을 참조하여 규칙을 생성한다.

표 1. 룰 베이스
Table 1. Rule Base

구분	규칙
체질량 지수	(?In-body ?biosignal)∩(?Customer hasWeight ?Weight)∩ (?Customer hasHeight ?Height) -> (?Customer hasBMI ?BMI)
고혈압	(?Customer hasDisease ?Hypertention) ∩ (?MaxBP < 120 or MinBP < 80) -> (?Hypertention hasHypertention ?Safety) ((?MaxBP >= 120 && < 139) or (MinBP >= 80 && <89)) -> (?Hypertention hasHypertention ?PreHypertention) ...
건강수준	(?Customer hasBMI ?BMI)∩ (?BMI hasHealthLevel ?Health_Level)∩ (?Customer hasAge ?Age)∩ (?Age hasHealthLevel ?Health_Level)∩ (?Customer hasSex ?Sex)∩ (?Sex hasHealthLevel ?Health_Level) -> (?Customer hasRecommend_HealthLevel ?Health_Level)
고혈압환자 운동 추천	(?Customer hasHealthlevel ?Health_Level5)∩ (?Customer hasDisease_History ?Hypertension) ->(?ExerciseRecommand hasExerciseRecommand ?Exer5_Hypertension) ... (?Customer hasHealthlevel ?Health_Level1)∩ (?Customer hasDisease_History ?Hypertension) ->(?ExerciseRecommand hasExerciseRecommand ?Exer1_Hypertension)

IV. 온톨로지와 나이브 베이지안 기반 필터링 추천 방법

1. 개인화 서비스 콘텐츠 추천 방법

디지털 TV 기반 유헬스 서비스를 제공하기 위한 개인화 서비스 추천 방법은 개인 데이터와 질병 데이터를 상황 데이터로 구성한 후, 온톨로지를 이용한 추론을 통해 건강 상태 따른 서비스 콘텐츠를 제공할지 결정한다. 다음으로, 서비스 콘텐츠 내용은 복수개가 나올 수 있으니 맞춤형 서비스를 제공하기 위해 학습트랜잭션에서 선호도 정보와 실시간 질병정보를 통해 나이브 베이지안 필터링을 사용하여 서비스를 제공한다. 상황데이터를 이용한 온톨로지 기반 추론을 통해 개인화 서비스 추천을 하기 위한 구성도는 그림 3과 같다.

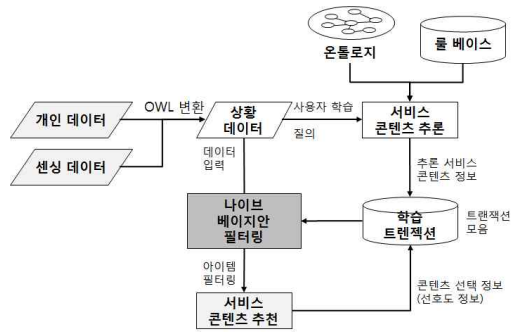


그림 3. 서비스 추천 방법
Fig. 3. Service Recommend Method

온톨로지와 조건부 확률 추론 기반 개인화 서비스를 제공하기 위해서는 첫 번째로, 개인 데이터와 센서로부터 측정된 센싱 데이터를 온톨로지 기반 추론이 가능한 OWL(Ontology Web Language)[15]형식으로 변환을 하여 상황데이터로 구성한다. 두 번째로, 상황데이터를 기반으로 온톨로지와 룰

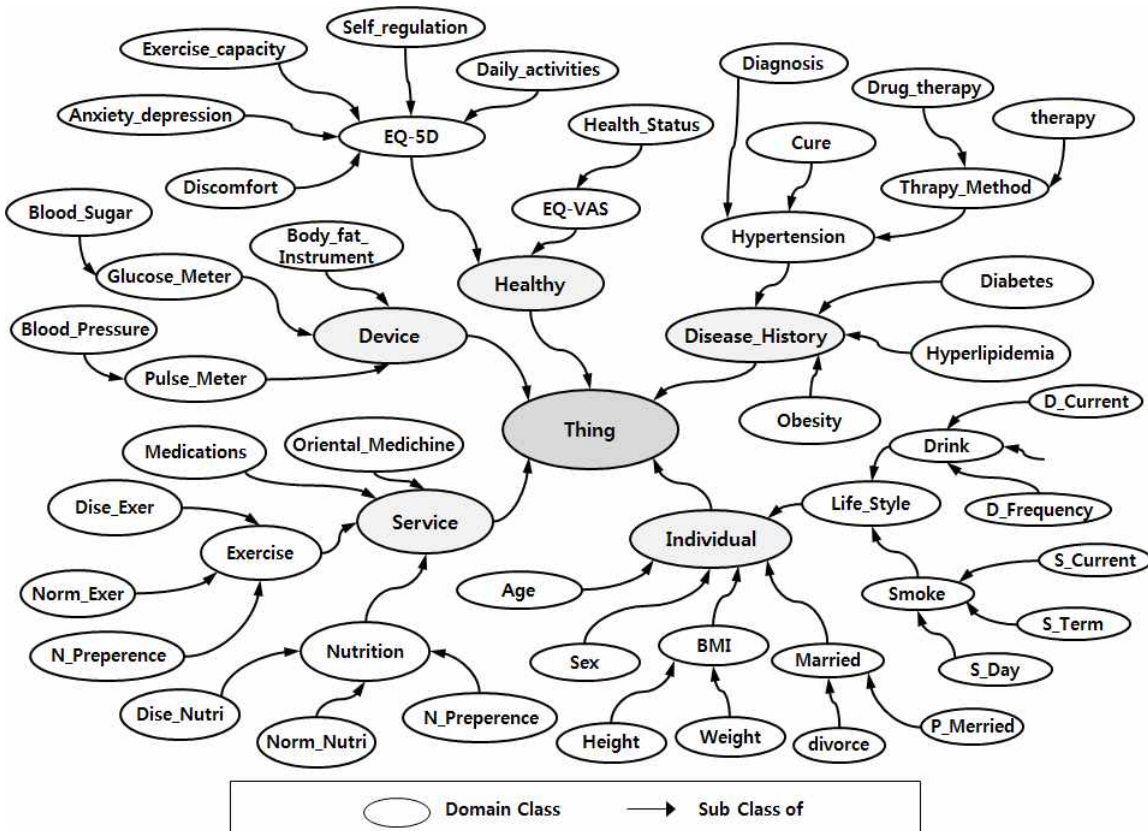


그림 2. 운동 추천 온톨로지
Fig. 2. Exercise Recommend Ontology

베이스를 이용하여 서비스 콘텐츠 추론을 하여 생성된 추론 서비스 콘텐츠 정보를 학습 트랜잭션에 저장한다. 마지막으로, 기존의 로그 데이터인 학습 트랜잭션을 이용하여 입력받은 상황 정보와 서비스 추론 정보를 필터링 하기 위해 나이브 베이지안 추론을 하여 맞춤형 추천 서비스를 제공한다. 사용자는 이로서 맞춤형 서비스 콘텐츠를 추천 받을 수 있으며, 서비스 콘텐츠 선택 정보에 따라 학습 트랜잭션에 입력이 된다.

2. 학습 트랜잭션 구조

온톨로지를 이용한 학습과 사용자의 선호 정보를 이용하여 나이브 베이지안 필터링으로 서비스 콘텐츠를 추천하기 위해서는 학습된 레코드들이 필요하며, 이를 통해 조건부 확률 추론이 가능하다. 학습 트랜잭션은 나이브 베이지안으로 필터링을 하기 위한 개인 레코드 정보이다. 학습 트랜잭션의 레코드와 그 관계에 대한 구조는 그림 4와 같다.

$C1 \sim Ci$ 는 상황정보 입력 부분이며, $I1 \sim Ij$ 는 상황정보로부터 추론된 콘텐츠 정보, $S1 \sim Sk$ 는 사용자가 선택한 선호도 정보이다. 즉 Ij 와 Sk 는 같은 콘텐츠를 의미한다. 또한 각 데이터의 정보는 0(False), 1(True)로 구성한다.

상황데이터로부터 얻은 정보는 $C1 \sim Ci$ 로 레코드에 입력이 되며, 0일 경우는 상황데이터가 입력받지 않은 것이고 1일 경우는 해당 상황데이터가 입력된 경우이다.

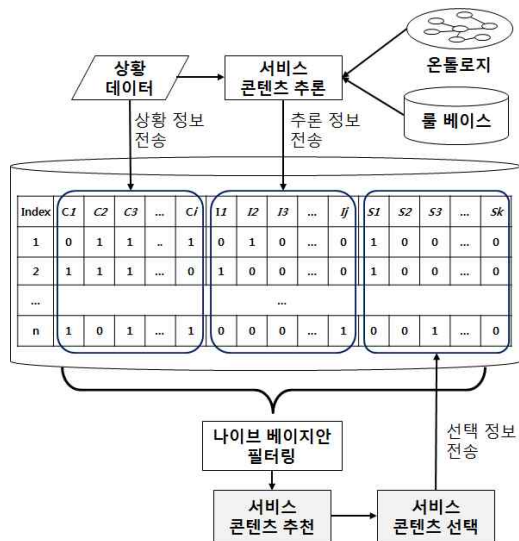


그림 4. 학습 트랜잭션 레코드 구조
Fig. 4. Learning Transaction Record Structure

상황정보를 이용하여 온톨로지로부터 추론된 서비스 콘텐

츠에 대한 정보는 $I1 \sim Ij$ 에 입력이 되며, 0일 경우는 콘텐츠를 추론하지 않은 것이며, 1일 경우는 서비스를 추론하여 나온 결과이다.

선호도 정보는 사용자가 나이브 베이지안으로 부터 추론된 콘텐츠 중 실제 사용자가 선택한 콘텐츠를 저장하는 것이며, $S1 \sim Sk$ 에 저장이 된다.

위와 같이 학습 트랜잭션의 레코드가 구성이 되며, 나이브 베이지안 메소드를 통해 추론하게 된다.

3. 나이브 베이지안 필터링 방법

베이지안 기반 추론은 불충분한 정보를 가진 환경을 표현하고 추론하는 대표적인 기법이며, 조건부 확률을 통한 의사 결정에 많은 장점이 있다. 학습 트랜잭션을 이용한 나이브 베이지안의 필터링 방법은 그림 5와 같다.

나이브 베이지안은 3단계로 추론을 하게 되며 첫 번째로, 상황데이터와 서비스 콘텐츠 추론 $P(C/I)$ 을 한다. 그리고 두 번째로, 상황데이터와 선호도 정보를 추론 $P(C/S)$ 을 한다. 마지막으로 추론된 $P(C/I)$ 와 $P(C/S)$ 에 대해서 사후확률 $P(I/S)$ 을 계산한다. 사후확률은 각각의 콘텐츠에서 계산된 선호도 추론 결과와 서비스 콘텐츠 추론결과를 계산하여 서로 다른 콘텐츠 중에서 선호도 값과 콘텐츠 값의 결과가 가장 높은 콘텐츠를 추천하기 위해서이다.

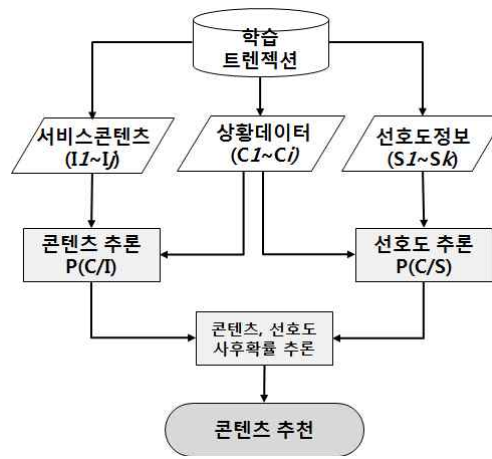


그림 5. 나이브 베이지안 필터링 방법
Fig. 5. Naive Bayesian Filtering Method

훈련된 데이터는 온톨로지의 추론과 콘텐츠 선택 정보로부터 수집된 학습 트랜잭션을 사용하고, 상황정보에서 콘텐츠 추론 방법은 식 1과 같으며, 선호도 추론 방법은 식 2와 같다.

$$P(I_j) = P(C|I_j) = \prod_{i=1}^{Context} P(C_i|I_j) \quad (1)$$

$$P(S_k) = P(C|S_k) = \prod_{i=1}^{Context} P(C_i|S_k) \quad (2)$$

C는 상황 데이터를 의미하며, Context은 입력된 상황 데이터의 종류를 의미한다. 식 1의 Ij는 각각의 서비스 콘텐츠의 종류를 의미하며, P(C|Ij)는 상황 데이터가 콘텐츠 추론 정보에 포함될 확률이다. 식 2의 Sk는 각각의 서비스 콘텐츠의 선호도 종류를 의미하며, P(C|Sk)는 상황데이터가 선호도 추론 정보에 포함될 확률이다.

또한, 나이트 베이지안은 속성 중 하나의 상황데이터의 조건부 확률 값이 0이면, 전체 추론 값이 사라진다. 예를 들어, 학습 트랜잭션의 C1의 값이 모두 0이거나 1일 경우, 베이지안 분류기는 이 항목을 분류할 수 없다. 따라서 이 문제를 해결하기 위해 식 3과 같이 m-추정치 기법을 사용한다[8].

$$P(x_i|y_i) = \frac{n_c + mp}{n + m} \quad (3)$$

n은 상황 데이터 yi의 사례수, nc는 xi 값을 갖는 상황데이터 yi 내의 훈련 사례수, m은 등가샘플 크기(equivalent sample size)라 불리는 매개변수, p는 사용자 정의 매개변수이다. 즉, 학습 집합이 없다면(n=0), P(xi|yi)=p 이다. 등가 샘플 크기는 사전확률 p와 관측확률 nc와 n 사이의 타협점을 결정한다. m-추정치 기법을 사용하면 일반적으로 학습 데이터의 수가 적을 때 확률 값을 추정하는 방법보다 확실한 방법을 제공한다.

각각의 서비스 콘텐츠 추론 정보 P(C|Ij)와 콘텐츠 추론 정보 P(C|Sk)가 생성이 되면 마지막으로 서비스 콘텐츠 추천을 위해 사후확률 P(I|S)을 계산한다. 사후확률의 추론 공식은 식 4과 같다.

$$P(I_1|S) = \frac{P(I_1)P(C|S_1)}{P(I_1)P(C|S_1) + \dots + P(I_j)P(C|S_k)} \quad (4)$$

서비스 콘텐츠 P(I|S)의 확률을 구하기 위해서는 각각의 콘텐츠 확률인 P(I)P(I|S1), P(I|S2), ..., P(I|Sk)의 값을 모두 더한 후에, P(I)P(C|S1)의 값을 나누어서

P(I|S1)의 확률을 추론하게 된다. 마지막으로, 추론된 각 P(I|S)중에서 확률값이 가장 높은 서비스 콘텐츠를 사용자에게 추천하게 된다.

V. 실험

1. 실험 환경

본 논문에서 제안하는 시멘틱 유헬스 서비스에서 나이트 베이지안 필터링 기반 개인화 서비스 방법의 검증을 위해 서비스 제공의 정확도를 측정한다.

실험을 위한 데이터는 2009년도 지역사회건강조사 결과보고서[17]를 이용하였다. 입력 변수로는 나이, 성별의 기본 데이터와 고혈압과 당뇨 질환을 가지고 있는 환자에 대한 데이터, EQ-5D 데이터, 운동, 흡연, 주량 등의 데이터를 사용하였다. 출력 변수로는 처방방법에 대한 콘텐츠의 선호도 내용이 포함되어 있는 데이터를 사용하였다. 레코드의 갯수는 1,500개를 이용하였으며, 이 중 레코드 1,000개는 트레이닝 셋으로 구성하였으며, 나머지 500개는 실험으로 이용한다.

실험을 위한 H/W환경은 Intel Core(TM) i3 CPU 2.93GHz, 4GB RAM, 500GB HDD를 이용하였다. S/W 환경은 Windows 7의 OS환경이며, 시스템 구동을 위해서 Java SDK환경을 이용하였다. 또한, 온톨로지 표현을 위해 Protege 4.0을 이용하였으며, 추론 엔진은 Jena2.0을 이용하였다.

2. 성능 평가

본 논문에서 제안하는 필터링 방법의 정확도(Precision)와 재현율(Recall)의 성능을 평가한다. 온톨로지를 이용한 추론은 상황 데이터에 대해 모든 콘텐츠를 제공해주기 때문에 높은 정확도를 나타낸다. 따라서 온톨로지를 이용하여 추론된 콘텐츠에 대해서 필터링 방법의 성능을 평가한다. 평가 방법은 추천된 모든 데이터 중에서 사용자가 선택하는 콘텐츠와 추천되는 콘텐츠를 1:1 비교매칭을 하여 정확도와 재현율을 평가한다[18].

필터링 성능을 평가하기 위한 비교 대상은 온톨로지와 나이트 베이지안을 기반으로 하는 협력적 필터링 기법 기반의 추천 방법(CFRM: Collaborative Filtering based Recommend Method)과 본 논문에서 제안하는 온톨로지와 나이트 베이지안을 기반으로 하는 규칙 기반과 협력적 필터링 기반의 추천 방법(RCFRM: Rule and Collaborative

Filtering based Recommend Method)을 비교한다. 실험을 위한 정확도와 재현율의 식은 식 5와 같이 구성한다.

$$Precision = \frac{|선택콘텐츠 \cap 추천콘텐츠|}{|추천 콘텐츠|} \quad (5)$$

$$Recall = \frac{|선택콘텐츠 \cap 추천콘텐츠|}{|선택콘텐츠|}$$

시나리오는 학습 트랜잭션에 학습된 레코드의 개수에 따라 추천 정보의 정확률과 재현율을 비교한다.

첫 번째 측정 방법으로, 정확도는 추천된 콘텐츠 중에서 연관된 콘텐츠의 비율을 계산한 것으로 입력 데이터와 선호도가 연관된 콘텐츠를 추천했을 경우 높은 수치를 보인다. 여기서는 실제 사용자가 선택한 콘텐츠와 추천한 콘텐츠의 교집합을 추천한 콘텐츠로 나누어 나타낸다. 실험에 대한 결과는 그림 6과 같다.

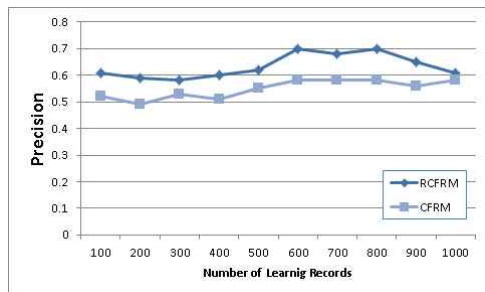


그림 6. 정확도 실험
Fig. 6. Precision Experiment

정확도의 실험 결과, 각 실험군의 평균 정확도는 RCFRM은 0.634이며, CFRM은 0.548이다. 따라서 RCFRM이 더 높은 정확도를 보인다고 할 수 있다. RCFRM은 선호도 값과 추천 값을 함께 계산하기 때문에 높게 나타나고 있으며, CFRM은 단순히 입력 데이터만을 이용했기 때문이다. 그러나 실험 레코드의 개수가 800개 이상일 경우 RCFRM의 정확도가 떨어지는 것으로 나타나고 있으며, 이는 사용자가 추천된 서비스 콘텐츠를 선택하지 않고, 상황 데이터와 맞지 않는 다른 콘텐츠를 선택할 경우가 있기 때문이다.

두 번째 측정 방법으로, 재현율은 추천된 콘텐츠에 대해 관련된 모든 콘텐츠의 수로 나눈 수치로 사용자가 선택한 콘텐츠를 추천하는 경우 높은 수치를 보인다. 여기서는 사용자가 선택한 콘텐츠와 추천된 콘텐츠의 교집합을 선택한 콘텐츠로 나누어 계산한다. 실험에 대한 결과는 그림 7과 같다.

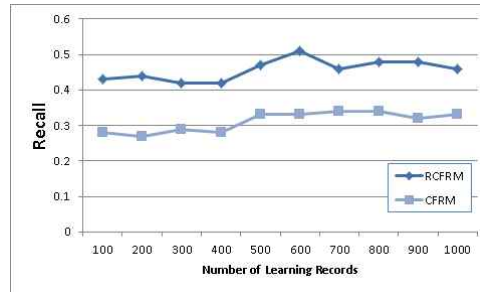


그림 7. 재현율 실험
Fig. 7. Recall Experiment

재현율의 실험결과, 각 실험군의 평균 재현율은 RCFRM은 0.457이며, CFRM은 0.311이다. 따라서 RCFRM이 더 높은 재현율을 보인다고 할 수 있다.

RCFRM은 사용자의 선호도를 반영했기 때문이며, 그렇지 않은 CFRM보다 높은 재현율을 보이고 있다. 그러나 0.5 이상의 재현율이 나타나지 않는 이유는 사용자가 선호하는 콘텐츠가 추천되는 콘텐츠를 모두 만족하지 않기 때문이다. 따라서 사용자가 선호하는 콘텐츠를 제공하기 위해서는 사용자 선호도 정보의 학습 방법에 대한 연구가 필요하다.

VI. 결론

본 논문에서는 디지털 TV에서 시멘틱 환경의 유헬스 서비스를 위한 나이브 베이지안 필터링 기반 개인화 서비스 추천 방법을 제안하였다. 기존의 시멘틱 환경의 유헬스 서비스는 사용자의 선호도 정보를 반영하지 못하는 규칙기반 필터링이기 때문에 사용자의 만족도가 떨어진다. 이에 본 논문에서는 협력적 필터링인 나이브 베이지안 기법을 혼용하여, 규칙기반 필터링의 정확성과 사용자의 선호 정보를 고려하는 필터링을 구현하였다.

실험 결과 본 논문에서 제안하는 온톨로지 기반의 나이브 베이지안 필터링 방법은 규칙기반과 협력적 기반 방법을 모두 고려했기 때문에 기존의 협력 필터링보다 높은 정확도와 재현율을 나타내고 있다. 제안하는 필터링 방법은 디지털TV에서 유헬스 서비스를 받는 사용자에게 콘텐츠 추천의 높은 만족도를 나타낼 것으로 기대한다.

참고문헌

- [1] Hong-Jin Cho, and Sang-Min Bae, "Improving Digital TV boot-up time", 38th Conference on KIISE, Vol. 32, No. 2(A), pp. 165-167, South Korea, Seoul, November 2011.
- [2] Dong-Kyun Park, Jong-Hun Kim, Jae-Kwon Kim, Eun-Young Jung, and Young-Ho Lee, "U-Health Service Model for Managing Health of Chronic Patients in Multi-platform Environment", Journal of The Korea Contents Association, Vol. 11, No. 8, pp. 23-32, August 2011.
- [3] Byung-Mun Lee, Jaekwon Kim, JongHun Kim, Youngho Lee, and Un-Gu Kang, "A Customized Exercise Service Model based on the Context-Awareness in u-Health Service", Journal of Korean Institute of Information Technology, Vol. 9, No. 2, pp. 141-152, January 2011.
- [4] Joong-Kyung Ryu, Jong-Hun Kim, Jae-Kwon Kim, Jung-Hyun Lee, and Kyung-Yong Chung, "Context-aware based U-health Environment Information Service", Journal of Korean Institute of Information Technology, Vol. 11, No. 7, pp. 21-29, April 2011.
- [5] J. H. Kim, D. S. Lee and K. Y. Chung, "Context-Aware Based Item Recommendation for Personalized Service", ICISA International Conference on Information Science and Applications, pp.1-6, 2011.
- [6] J. Ben Schafer, Dan Frankowski, Jon Herlocker and Shilad Sen, "Collaborative filtering recommender systems", Lecture Notes In Computer Science, pp.291-324, 2007.
- [7] K. Jung, Y. Na ,and J. Lee, "FDRAS: Fashion Design Recommender Agent System Using the Extraction of Representative Sensibility and the Two-Way Combined Filtering on Textile", Lecture Notes in Computer Science, pp.631-640, 2003.
- [8] Pang Ning Tan., Michel Steinbach, Vinpin Kumar, Introduction to Data Mining, Addison Wesley, 2007.
- [9] Stuart Russell, "Artificial Intelligence : A Modern Approach(3rd Edition, Paperbak)", Pearson Education, 2010.
- [10] Weltizen, <http://www.weltizen.com>
- [11] Bit Computer co., <http://www.bit.kr>
- [12] P. Resnick, et. al., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", In Proc. of ACM CSCW'94 Conference on Computer Supported Cooperative Work, pp.175-186, 1994.
- [13] J. S. Breese and D. Heckerman and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", In Proc. of the 14th Conference on Uncertainty in Artificial Intelligence, 1998.
- [14] Chae-Soo Eun, Dong-Ju Cho, Kyung-Yong Jung, and Jung-Hyun Lee, "Development of Apparel Coordination System Using Personalized Preference on Semantic Web", Journal of Korean Institute of Information Technology, Vol. 7, No. 4, pp. 66-73, April 2007.
- [15] Owl Web Ontology Language Over View, <http://www.w3.org>
- [16] ACSM's Resource Manual for Guideline for Exercise Testing and Prescription Six Editon, 2010.
- [17] Community Health Survey, <http://chs.cdc.go.kr>
- [18] Myoung-hwan Ahn, and Joon-hee Kwon, "Ontology based Context-Aware Recommendation System using Concept Hierarchy", Journal of Korean Society for Internet Information, Vol. 8, No. 5. pp. 81-89, July 2007.

저 자 소 개



김 재 권
2011: 가천의과대 정보처리과 공학사.
현 재: 인하대학교 컴퓨터정보과 석사
과정
관심분야: 인공지능, 데이터마이닝,
분산처리
Email : jaekwonkorea@naver.com



이 영 호
1996: 한국외국어대학교 응용전산학
과 이학석사.
2005: 이주대학교 의료정보학과 이학
박사.
2002: IBM Korea BI & CRM EM
현 재: 가천대학교 IT 대학 교수
현 재: ISO/TC215 전문위원
현 재: 가천대학교 u-헬스케어연구소
책임연구원
현 재: 수송물류분야 단체표준 전문위원
현 재: 스마트의료정보표준포럼 사무총장
관심분야: 메디컬인포메틱스, 유헬스케어
Email : lyh@gachon.ac.kr



김 종 훈
2001: 인천대학교 물리학과 이학사
2003: 인하대학교 전자계산공학과 공
학석사
2010: 인하대학교 컴퓨터정보공학과
공학박사
2008: 대림대학 컴퓨터정보계열 전임
강사
2011: 가천대학교 u-헬스케어연구소
선임연구원
현 재: (주)비트컴퓨터 유헬스케어 사
업부 책임연구원
관심분야: 유헬스케어, 인공지능, 메디
컬 데이터마이닝
Email : kimjh@bit.kr



박 동 군
1992: 충북대학교 의과대학 의학사.
2003: 인하대학교 의과대학 의학석사.
2010: 인하대학교 의과대학 의학박사.
현 재: 가천대학교 길병원 교수 소화기
내과분과 전문의, 내시경 전문의
현 재: 가천대학교 길병원 유헬스케어
센터장
관심분야: Medical Informatics,
RFID/USN Application,
System Medicine
Email : pdk66@gilhospital.com



강 운 구
2001: 인하대학교 전자계산공학과 공
학박사.
2006: 뉴미디어연구소장
현 재: 가천대학교 IT대학 교수
현 재: 가천대학교 유헬스케어 연구소장
현 재: IT융합산업정책위원
현 재: 한국산학연합회 이사
관심분야: 의료정보, 유헬스케어,
소프트웨어공학
Email : ugkang@gachon.ac.kr