

Naive Bayesian 알고리즘을 이용한 P2P 모바일 에이전트의 필터링 기법

Filtering Technique of P2P Mobile Agent using Naive Bayesian Algorithm

이세일, 이상용*

공주대학교 컴퓨터공학과, 공주대학교 정보통신공학부*

Se-Il Lee, Sang-Yong Lee*

Dept. of Computer Engineering, Division of Information & Communication
Engineering*, Kongju National University

E-mail : lsilhr@hanmail.net

요 약

유비쿼터스 컴퓨팅에서 사용자에게 필요한 서비스를 지능적으로 제공하기 위해서는 컨텍스트 정보의 효과적인 필터링이 필요하다. 현재까지 사용되고 있는 필터링 기술은 온라인상에서 사용되는 사용자 정보를 기준으로 서비스를 제공하고 있다. 하지만 휴대용 유·무선기기에서 컨텍스트 인식에 기반을 둔 서비스를 제공하기 위해서는 복잡한 필터링과정과 큰 저장 공간이 요구된다.

따라서 본 논문에서는 사용자 주변에 널려 있는 센서를 통해 입력된 컨텍스트 정보들을 효율적으로 필터링하여 사용자에게 필요한 서비스만을 제공하도록 하였다. 이를 위해서 기존의 P2P 모바일 에이전트에서 사용되는 협력적 필터링 기술에 Naive Bayesian 알고리즘을 혼합한 컨텍스트 협력적 필터링 알고리즘을 제안한다.

1. 서론

유비쿼터스 컴퓨팅(Ubiquitous Computing)은 우리의 일상생활 속에서 언제 어디서나 자유롭게 개인별 맞춤 서비스를 제공하는 컴퓨터 환경을 말한다. 유비쿼터스 컴퓨팅 환경에서 사용자에게 적합한 서비스를 제공하기 위해서는 주변 환경 정보를 인식할 수 있는 컨텍스트 인식 과정(Context-Awareness)과 사용자의 위치에 따라 취할 수 있는 행동이 필요하다. 컨텍스트 인식을 위한 자료 수집은 네트워크, 센서 및 단말기와 같은 장치들을 통해서 수집된다. 만약 인식된 사용자 컨텍스트가 현재 서비스에 사용되고 있다면, 사용되는 서비스는 종속적이기 때문에 다른 사용자가 사용하려면 컨텍스트를 변경하거나 새로운 컨텍스트를 필요로 한다[1][2]. 이러한 종속성 문제를 해결하기 위한 컨텍스트 구조는 5W1H(Who, When, Where, What, Why, How)를 중심으로 처리하고 있다. 5W1H는 사용자의 제약조건과 선호도를 분석하기에 충분하며 사용자의 욕구도 파악이 가능하다[3].

현재 MAUCA와 같은 P2P 모바일 에이전트 시스템에서는 기준치 이하의 기댓값을 갖는 서비스는 제공될

수 없는 문제점을 가지고 있다[4]. 이를 해결하기 위한 방법은 사용자의 선호도에 따른 서비스 필터링 과정을 적용하는 것이다. 그리고 서비스 필터링을 위해서는 사용자의 위치와 시간에 따라 적용된 서비스 항목이 필요하다. 이를 위해서는 제공된 서비스를 공유시키는 방법이 연구되고 있다. 하지만 컨텍스트를 저장할 수 있는 공간이 부족하고, 협력적 필터링에 대한 예측성이 떨어지는 문제점을 가지고 있다.

본 논문에서는 P2P 모바일 에이전트인 MAUCA의 개선된 구조에 필요한 협력적 필터링 구조를 제안한다. 제안한 협력적 필터링 구조는 Naive Bayesian 알고리즘을 사용하여 서비스의 종속성 문제를 해결하고 서비스에 대한 예측성을 높게 하였다.

본 논문의 구성은 2장에서 본 연구를 위한 관련연구를 기술하고, 3장에서는 제안된 협력적 필터링 구조를 설명하며, 4장에서는 실험 및 평가한 내용을 기술하였다. 마지막으로 5장에서는 결론을 기술하였다.

2. 관련 연구

2.1 협력적 필터링

협력적 필터링은 초기 평가 문제(First Rater Problem), 희박성 문제(Sparsity Problem) 및 확장성 문제(Scalability Problem) 등의 문제점을 가지고 있다. 하지만 이러한 문제점은 추천 에이전트와 전자상거래 분야에서 성공적으로 개선되어 사용되고 있다[5].

협력적 필터링의 일반적인 처리 과정은 사용자들의 상황을 분석하여 다른 사용자들의 상황 정보를 가지고 선호 벡터를 구성한 뒤, 이 선호 벡터와 비슷한 선호도를 가지는 값을 예측할 수 있다. 이 예측값을 구하는 방법으로는 피어슨 상관 계수(Pearson Correlation Coefficient), 보안 코사인 유사도(Adjust Cosine Similarity) 방법 등이 있으나 가장 많이 사용하는 방법은 피어슨 상관 계수이다[6]. 유사도를 구하는 식은 아래 <식 1>과 같다.

$$W_{x,y} = \frac{\sum_{a=1}^n (r_{x,a} - \bar{r}_x)(r_{y,a} - \bar{r}_y)}{\sqrt{\sum_{a=1}^n (r_{x,a} - \bar{r}_x)^2} \sqrt{\sum_{a=1}^n (r_{y,a} - \bar{r}_y)^2}} \dots\dots\dots <식 1>$$

a는 사용자 x와 사용자 y가 모두 평가한 항목이며 \bar{r}_x 는 사용자 x의 전체 항목에 대한 평가 평균값이다. $r_{x,a}$ 는 사용자 x가 항목 a에 대하여 평가한 값이고, n은 항목의 총 개수이다.

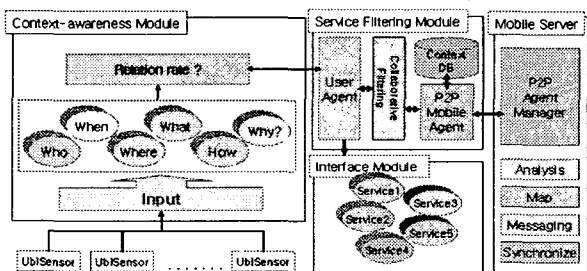
유사도 식과 이웃의 수를 사용하여 값을 예측하는데 그 예측값은 <식 2>와 같다.

$$P_{x,b} = \bar{r}_x + \frac{\sum_{y=1}^n w(x,y)(r_{y,b} - \bar{r}_y)}{\sum_{y=1}^n w(x,y)} \dots\dots\dots <식 2>$$

$P_{x,b}$ 는 사용자 x와 항목 b에 대한 선호도를 예측한 값이고, \bar{r}_x 는 사용자 x의 선호도 평균값이다. $w(x,y)$ 는 사용자 x와 사용자 y의 유사도 가중치이고, $r_{y,b}$ 는 사용자 y가 항목 b에 대하여 평가한 값이다. \bar{r}_y 는 사용자 y의 선호도 평균값이고, n은 결정된 이웃의 수이다.

2.2 P2P 모바일 에이전트

P2P 모바일 에이전트 구조인 MAUCA(Mobile Agents for Users' Context-Awareness)는 모바일 에이전트에 협력적 필터링과 P2P 에이전트를 결합했다 [4].



[그림 1] MAUCA 프레임워크

MAUCA는 구조는 [그림 1]과 같이 컨텍스트-인식

모듈(Context-awareness Module), 서비스 필터링 모듈(Service Filtering Module), 인터페이스 모듈(Interface Module), 모바일 서버(Mobile Server)로 구성된다.

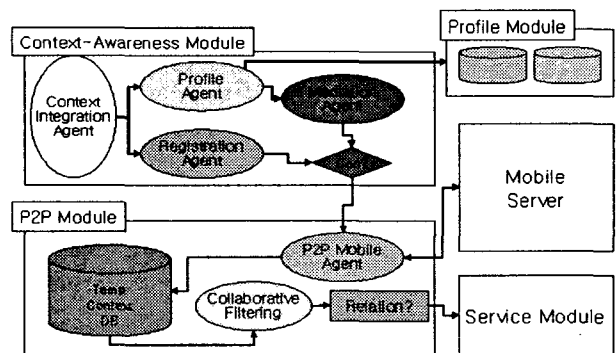
MAUCA의 구조는 동일 지역내에서 각 사용자의 컨텍스트 정보를 수집하고 분석하여 비슷한 선호도를 갖는 사용자들로 그룹핑한다. 그룹핑된 사용자 그룹은 P2P 모바일 에이전트를 이용하여 이전에 서비스된 정보를 공유한다. 또한 협력적 필터링 기법에서는 공유된 사용자의 컨텍스트 정보에 맞게 서비스 정보를 분석하고, 우선순위를 부여하여 제공된다. 그리고 이 구조에서는 사용자들의 행위와 서비스를 지속적으로 관찰 및 학습하여 새로운 상관 관계를 유지하도록 설정하였다.

3. IMAUCA의 협력적 필터링 구조

본 논문에서는 모바일 에이전트에 협력적 필터링 기법과 P2P 에이전트를 혼합한 P2P 모바일 에이전트 구조를 가진 IMAUCA 구조를 설명한다. MAUCA의 서비스 필터링 모듈의 복잡성을 줄이고 컨텍스트 DB(Context DB)의 액세스 방식을 개선하기 위하여 구조를 개선하였다. 또한 협력적 필터링 모듈에 Naive Bayesian 알고리즘을 이용하여 예측성을 증대하였다.

3.1 개선된 협력적 필터링 구조

개선된 구조인 IMAUCA(Intelligent Mobile Agent for Users' Context-Awareness)는 다중 에이전트 구조를 갖도록 설계하였으며, 컨텍스트 및 서비스 공유를 위해 P2P 모바일 에이전트의 구조를 강화하였다. [그림 2]은 개선된 IMAUCA의 협력적 필터링 구조이다.



[그림 2] IMAUCA의 협력적 필터링 구조

IMAUCA의 서비스 처리 흐름을 살펴보면, 먼저 컨텍스트 통합 에이전트(Integration Agent)는 센서로부터 읽어들이는 사용자 컨텍스트 정보를 통합하며, 프로파일 에이전트(Profile Agent)와 등록 에이전트(Registration Agent)에게 통합된 정보를 전달한다. 프로파일 에이전트는 사용자 컨텍스트 정보를 기록하고, 비교하여 조정 에이전트(Mediation Agent)에게 전달하여 서비스 항목을 생성한다. 그리고 등록 에이전트는 지역 내에 사용자를 등록시키기 위하여 P2P 모바일 에이전트를 생성한다. 조정 에이전트는 생성한 서비스

항목을 검사하여 유사성이 낮은 경우 P2P 모바일 에이전트에게 서비스 항목의 수집 신호를 전달한다. P2P 모바일 에이전트는 지역(Cell)내에 사용자들의 컨텍스트 정보 일부와 서비스 항목을 공유하여 임시 기억장소(Temp Context DB)에 기록한다. 기록된 정보는 협력적 필터링을 거쳐 사용자에게 가장 적합한 서비스를 추천받게 된다. 그리고 추천된 서비스는 관계율이 재조정되어 프로파일 모듈(Profile Module)에 기록된다.

3.2 Prefiltering

유비쿼터스 컴퓨팅에서 효과적인 서비스 제공은 인식된 컨텍스트와 이전 정보, 그리고 변화된 정보를 바탕으로 분석된 결과를 통해서 서비스되어야 한다. 이를 위해 조정 에이전트는 프로파일 에이전트(컨텍스트 인식 정보)의 값과 다이나믹 컨텍스트 DB(Dynamic Context DB, 서비스 정보) 값을 공분산 한다.

공분산은 센서에 의해 입력된 컨텍스트 인식 정보 (X)와 사용자 에이전트가 가지고 있는 서비스 정보 (Y)의 기댓값으로 계산되며 <식 3>과 같다.

$$Cov(x, y) = \frac{\sum_{i=1}^n (x_i - u_x)(y_i - u_y)}{n} = \frac{\sum_{i=1}^n (x_i y_i - u_x u_y)}{n} \dots\dots\dots <식 3>$$

공분산 값이 0.5이상이면 직접 서비스를 실행하지만 0.5 미만이면 모바일 에이전트에서 가지고 온 다른 그룹 정보와 필터링 한 후 서비스를 하게 된다.

서비스 정보는 주변 확률 밀도 함수(Marginal Probability Density Function)를 이용하여 사용자에게 서비스된 항목(y)과 사용자 컨텍스트 인식 정보(x)의 관계값을 측정하여 이전에 제공된 서비스의 만족도를 평가하여 순위를 결정하게 된다. 관계물의 평가값은 <식 4>와 같다.

$$f_x(x) = \sum_y f(x, y) \dots\dots\dots <식 4>$$

관계값을 계산하기 위해서는 서비스 정보가 필터링 후에 측정된 값인지 또는 직접 서비스된 값인지에 따라 다르게 계산된다. 필터링에 의해 측정된 관계값은 <식 5>같지만 직접 서비스된 값은 변함이 없다. <식 5>의 pmv는 프로파일 모듈 값이며, cfv는 필터링에 의해 측정된 값이다.

$$\text{관계값} = (cfv + pmv)/2 \dots\dots\dots <식 5>$$

다이나믹 컨텍스트 DB는 관계값에 의해 내림차순 정렬되므로 이전에 발생한 서비스는 능동적으로 서비스를 지원하고 서비스에 불필요한 항목들은 자연히 제거된다.

3.3 Naive Bayesian을 이용한 협력적 필터링

Naive Bayesian은 기계학습 방법 중에서 확률적 이론을 근거로 한 분류 방법이다. 이것은 임의의 항목이 특정 클래스에 속할 확률을 계산하여, 그 중 가장 높은 확률을 보이는 클래스를 선택하여 분류된다. 항목

이 {item₁, item₂, item₃, ..., item_n}라고 하였을 때 <식 6>을 이용하여 클래스 {class₁, class₂, class₃, ..., class_n} 중 하나로 분류된다.

$$P(class_i | item_i) = \frac{P(item_i | class_i)P(class_i)}{P(item_i)} \dots\dots <식 6>$$

P(class_i|item_i)은 분류 item_i에 속하는 항목집합에서 임의로 추출한 item이 class_i가 될 확률이다[7].

Naive bayesian은 자료의 양이 많을 경우 학습법의 정확도가 다른 학습 방법보다 높기 때문에 협력적 필터링에 많이 사용된다.

협력적 필터링 실험을 위한 가정은 영화관에서 모바일 서비스를 사용한다고 가정하고 영화별 장르를 클래스로 구분하였다. 또한 모바일 에이전트에서 가지고 온 서비스 정보는 장소별로 클래스를 형성하고 훈련 집합을 만들도록 하였다. 형성된 클래스는 Naive Bayesian 학습 알고리즘을 적용하여 훈련시켰으며 서비스의 사용 회수도 측정한다.

컨텍스트 인식 정보는 인식한 서비스 정보의 선호도를 장르[8]와 나이별로 다르게 추정치를 적용하도록 하였다. 선호도를 평가한 서비스 정보는 학습 집단에 추정치를 부여하고 클래스별 서비스 정보인 P(Service_i|Class_i)와 B_{x,z}를 곱하여 <식 7>과 같이 구한다. B_{x,z}는 컨텍스트 인식 정보 x와 서비스 정보 z를 평가한 선호도 값이다. aWeight는 선택된 서비스 정보에 대한 사용자의 연령 차를 계산한 값으로 유사도 계산의 가중치를 위해 사용된다. 예를 들어, 내가 20대라고 가정하고 40대의 정보가 들어오게 되면 40대들 모두 그 영화를 좋아하는 것이 아니다. 그러므로 40대가 선택한 정보에 가중치를 부여하기 위하여 1 - (40-20) * 0.01와 같이 식을 세울 수 있다.

$$R_{x,z} = aWeight * P(Sevice_i | Class_i) * B_{x,z} \dots\dots\dots <식 7>$$

위 식을 피어슨 상관계수에 적용하면 유사도 가중치가 <식 8>와 같이 정의 된다.

$$W_{x,y} = \frac{\sum_{z=1}^n (R_{x,z} - \bar{R}_x)(R_{y,z} - \bar{R}_y)}{\sqrt{\sum_{z=1}^n (R_{x,z} - \bar{R}_x)^2} \sqrt{\sum_{z=1}^n (R_{y,z} - \bar{R}_y)^2}} \dots\dots\dots <식 8>$$

W_{x,y}는 컨텍스트 인식 정보 x와 컨텍스트 인식 정보 y의 유사도 가중치이다. R_{x,z}는 컨텍스트 인식 정보 x와 서비스 정보 y에 대하여 가중치가 부여된 선호도이며, \bar{R}_x 는 x의 선호도 평균값이다. z는 x와 y가 모두 선택한 서비스 정보가 된다.

새로운 컨텍스트 인식 정보는 Naive Bayesian 분류자[9]를 이용하여 <식 9>와 같이 장소별로 분류한다.

$$Class_i = \underset{class_i}{\operatorname{argmax}} P(Class_i) IIR_x, z P(Sevice_i | Class_i) \dots\dots\dots <식 9>$$

Service_i는 새로운 컨텍스트 인식 정보가 선호도를 표시한 서비스 정보들이며 Class_i에 분류된다. P(Service_i|Class_i)는 컨텍스트 인식 정보가 선호도를 평가한 서비스 정보들이 클래스에 포함될 확률이다.

새로운 컨텍스트 인식 정보가 구해지면 선호도를 예측하기 위해서 Naive Bayesian 알고리즘을 이용한 사용자 유사도를 구하고 이웃을 선정한다. 이웃 선정 방법은 K-nearest Neighbor 방법을 이용하여 상위 50%에 포함되는 이웃만을 선정하여 선호도를 예측한다.

4. 평가

본 논문에서 제안한 시스템은 C#과 J2ME, WIPI를 이용하여 펜티엄 IV, 2.8Ghz, 512MB의 환경에서 설계하고 실험하였다.

사용자의 컨텍스트 인식과 서비스 정보를 위한 시나리오는 다음과 같다.

A씨는 모처럼 한가하여 일요일에 멀티영화관을 찾아 갔다. A씨는 영화에 대한 정보가 없기 때문에 나와 유사한 나이에 나와 비슷한 취향을 가진 사람들의 어떤 영화를 많이 보고 있는지 알고 싶어 한다.

A씨는 P2P 모바일 에이전트를 이용하여 사용자들에게 사생활 보호에 영향이 없는 정보만 추천 받는다.

협력적 필터링을 평가하기 위한 사용자의 컨텍스트 정보 및 서비스 항목은 장르별로 그룹핑하여 평가하였다. [그림 3]은 에플레이터를 이용하여 서비스가 제공된 화면이다.

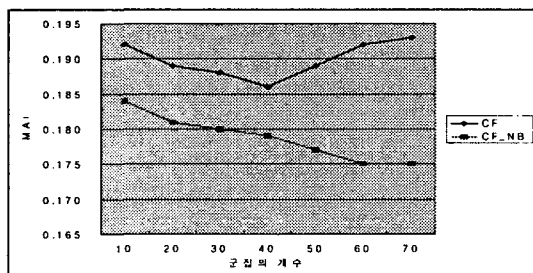


[그림 3] 서비스 실행화면

아래 <식 10>은 예측의 정확성을 위하여 MAE(Mean absolute Error)를 사용하였다. e_i 는 실제 선호도와 예측 선호도의 오차이며, N은 총 예측 회수이다.

$$|E| = \frac{\sum e_i}{N} \dots\dots\dots <식 10>$$

[그림 4]는 MAUCA의 고전적 협력적 필터링 결과와 개선된 IMAUCA의 Naive Bayesian 협력적 필터링 결과를 비교한 그래프이다.



[그림 4] 군집의 개수에 따른 MAE

실험결과 CF_NB는 군집의 개수가 40여개 이하이면 CF보다 0.008정도 좋았다. 특히 40여개 이상부터는 2배정도 차이가 났다. 위 결과로 보아 제안한 Naive Bayesian과 함께 사용한 협력적 필터링 방법이 고전적 협력적 필터링 방법보다 우수하였다.

5. 결론

본 논문에서는 MAUCA의 구조를 변경하여 다중 에이전트 구조를 갖도록 설계하였으며, 컨텍스트 및 서비스 공유를 위해 P2P 모바일 에이전트의 구조를 강화하였다. 그리고 사용자에게 우수한 서비스를 제공하기 위해 협력적 필터링 영역을 개선하였다. 개선된 협력적 필터링은 나이와 다중 선택에 대한 가중치를 결합한 Naive Bayesian 알고리즘을 적용하였다. 제안된 방법은 WIPI 에플레이터를 사용하여 실험하였다. 그리고 기존 P2P 모바일 에이전트인 MAUCA의 협력적 필터링과 비교한 결과 그룹 설정 및 서비스 지원 성능 평가 면에서 평균적으로 6.5% 정도 우수함을 보였다.

참고문헌

- [1] T.D.Hodes, R.H.Katz, E.Servan-Scriber, and L.Rowe. "Composable ad-hoc mobile services for universal interaction", In Mobicom'97, pp.1-12, 1997
- [2] A.K.Dey and G.D.Abowd, "Towards a Better Understanding of Context and Context-Awareness", Gvu Technical Report GIT-Gvu -99-22. Submitted to the 1st International Symposium on Handheld and Ubiquitous Computing (HUC '99), June 1999
- [3] ubiHome을 위한 컨텍스트 기반 응용 서비스 모형, 정보과학회논문지, vol 30, no 5_6, pp.550-558, 2003,6
- [4] 윤효근, 이상용, "협력적 필터링 기법을 이용한 P2P 모바일 에이전트 기반 사용자 컨텍스트 인식 및 서비스 처리구조", 퍼지 및 지능시스템학회 논문지, Vol.15, No.1, pp. 104-109, 2005
- [5] Sarwar, B. et al., "Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System", Proc. ACM CSCW 98, pp.345-345, 1998
- [6] N.Good, B. Schafer, J.Konstan, A. Borchers, B.Sarwar, J. Riedl, "Combining Collaborative filtering with personal Agents for Better Recommendation", AAAI/IAAI, pp.439-446, 1999
- [7] 행렬 전치를 이용한 효율적인 NaiveBayes 알고리즘, 정보처리학회논문지, B 제11-B권 제1호, 2004.2
- [8] 선호도 재계산을 위한 연관 사용자 군집 분석과 Representative Attribute -Neighborhood를 이용한 협력적 필터링 시스템의 성능향상, 정보처리학회논문지, v.10B, n.003, pp.287-296, 2003.6
- [9] <http://www.free-definition.com/Naive-Bayesian-classification.html>