

스피어만 상관계수를 이용한 사용자 상황 및 특성 처리 개선

안찬식[†], 오상엽^{**}

요 약

CRUMPET과 같은 모바일 단말기 서비스 시스템은 사용자의 상황과 특성이 사용자마다 서로 달라 정보가 매우 적고 이 또한 연관성을 찾기가 매우 어렵다. 사용자의 상황과 특성에 따른 정보 선호도를 제공하기 힘든 단점과 사용자가 원하는 정확한 정보를 추천하기 어렵기 때문에 개념적인 정보를 제한적으로 추천한다. 따라서 본 논문에서는 스피어만 상관 계수를 이용하여 사용자 상황과 특성에 적합한 정보를 추천하는 시스템을 제안하였다. 사용자 정보로부터 연관성을 찾아 서열화하고 사용자의 상황과 특성에 적합한 정보를 목록으로 제공하여 제한적인 개념적 정보 추천의 단점을 해결하였다. 성능 실험은 기존의 서비스 시스템들과 비교하여 효과성인 정확도와 재현율로 측정하였으며, 성능 실험 결과 정확도는 92.3%, 재현율은 73.8%로 나타났다.

Improvement of User's Context Aware and Characteristic Process using spearman correlation coefficients

Chan-Shik Ahn[†], Sang-Yeob Oh^{**}

ABSTRACT

There is very little information on mobile terminal service systems such as CRUMPET because the all the users have different situations and characteristic, and so it is also difficult to find correlations. Because of the difficulty of customizing and recommending information based on preference stemming from the users' various situations and characteristics, they usually provide limited, conceptual information. This paper will recommend a system that recommends information tailored to the user's situation and characteristics, using the Spearman correlation coefficients. It finds correlations from users' information and sequences information that is suitable to the user's situation and characteristics into a list, thereby solving the problem of limited, conceptual information. Performance tests have revealed when compared to existing service systems, this system is more effective in terms of precision and recall, with a 92.3% precision rate, and a 73.8% recall rate.

Key words: Spearman correlation coefficients(스피어만 상관 계수), characteristic process(특성 처리), user's context aware(사용자 상황), recommendation service system(추천 서비스 시스템)

1. 서 론

유비쿼터스 기술의 발전으로 멀티미디어 콘텐츠

의 질적 양적 성장에 힘입어 모바일 콘텐츠 시장이 확대되고 성장하였으며 이에 따라 모바일 단말기의 보급으로 인하여 무선 인터넷이 확대되어 실시간으

* 교신저자(Corresponding Author): 오상엽, 주소: 경기도 성남시 수정구 복정동 산65(461-702), 전화: 031)750-5798, FAX: 02)426-9159, E-mail: syoh@kyungwon.ac.kr
접수일: 2010년 4월 4일, 수정일: 2010년 9월 28일

완료일: 2010년 10월 20일

[†] 정회원, 광운대학교 컴퓨터공학과 박사수료
(E-mail: absoluti@kw.ac.kr)

^{**} 종신회원, 경원대학교 IT대학 컴퓨터소프트웨어 교수
* 본 연구는 2010년도 경원대학교 지원에 의한 결과임.

로 정보를 교환하고 시간과 공간적으로 사용자의 상황을 인지하여 사용자의 상황에 맞는 서비스를 제공한다. 무선 인터넷으로 인한 모바일 단말기 사용은 시간적, 공간적 제한을 받지 않기 때문에 사용자의 환경적 상황에 따라 선호하는 콘텐츠가 변화한다[1].

사용자의 개별적인 취향과 선호에 따른 콘텐츠를 추천하여 유용한 서비스를 제공하는 인식 기술이 이용되며 사용자의 선호 정보를 입력 받아 취향과 선호에 맞는 콘텐츠 추천의 역할이 중요하게 작용한다. 검색에 의한 결과를 확인하기 위해서는 사용자로부터 정보를 입력 받아야 하며 정확한 정보의 제공을 위해 사용자로부터 많은 정보를 입력 받게 되어 입력에 할애되는 시간이 상당히 많이 소요된다. 효율적인 콘텐츠 제공을 위한 콘텐츠 검색 기법은 사용자의 개별적인 취향과 선호에 따른 콘텐츠를 추천하기보다 개념적인 정보를 제한적으로 추천하게 되며 이 또한 검색을 위한 시간이 필요하게 된다[2].

모바일 단말기를 이용한 서비스 시스템은 사용자의 상황과 특성이 시간적, 공간적, 상황적으로 변화하여 많은 정보가 필요하게 된다. 하지만 사용자는 정보를 입력하기 위한 시간과 정보를 할애하기보다 빠른 시간에 정확한 정보를 제공받기 원한다. 사용자마다 상황과 특성이 서로 다르기 때문에 추천을 위한 정보가 매우 적고, 적은 정보에서 연관성을 찾기가 매우 어렵다. 따라서 다음과 같은 단점이 존재한다. 첫 번째는 사용자의 상황과 특성에 따른 정보 선호도를 제공하기 힘들고 두 번째는 사용자가 원하는 정확한 정보를 추천하기가 어려우며 세 번째는 개념적인 정보만을 제한적으로 추천한다[3].

따라서 본 논문에서는 스피어만 상관 계수를 이용하여 사용자 상황과 특성에 적합한 정보를 추천하는 시스템을 제안한다. 적은 정보와 연관성을 스피어만 상관 계수를 이용하여 서열화하고 사용자의 상황과 특성에 적합한 정보를 서열화한 목록으로 제공하므로 모바일 단말기 서비스 시스템에서 존재하는 제한적인 개념적 정보 추천의 단점을 해결하였다.

제안된 방법의 수행 과정에서는 특성을 추출하여 특성 처리와 특성 클러스터링 처리를 수행하고 스피어만 상관 계수를 이용하여 예측하고자 하는 사용자와 유사한 선호도를 가지는 사용자들의 정보를 기반으로 사용자의 상황별 선호도를 예측하여 유사 선호도를 구하고 유사 선호도로부터 서열화된 최종 추천

객체 목록을 생성한다.

성능 실험은 기존의 서비스 시스템들과 비교하여 효과성을 측정하여 정확도와 재현율로 표현하였으며, 성능 실험 결과 정확도는 92.3%, 재현율은 73.8%로 나타났다

본 논문의 구성으로는 2장에서 상황 인식 정보와 선호도 추천 시스템의 관련 이론에 대해서 살펴보고, 3장에서는 본 논문에서 제안한 k-means 알고리즘 기반 사용자 상황 선호도 추천 개선 시스템에 대하여 설명한다. 그리고 4장에서는 실험 환경과 실험 결과, 5장에서는 결론을 다룬다.

2. 관련연구

2.1 선호도 선정 상관 계수

상황 정보는 입력되어진 정보의 상호 작용을 통하여 사용자의 상황을 특성화한다. 사용자의 현재 상황 정보의 입력으로부터 데이터의 매칭에 의한 적절한 정보를 서비스하기 위해 이러한 상황을 이용하며 이를 상황 인식이라 한다. 상황 인식 서비스 시스템에서 서비스의 가능 여부와 서비스의 종류를 결정하는 방법은 사용자로부터 입력된 정보와 사용자의 상황 정보를 고려하여 선호도 선정 기법을 이용한다[4].

선호도 선정 기법을 사용하기 위해 데이터를 모델링하는 방법은 상관 계수를 이용한다[5]. 상관 계수는 연속된 자료 값을 갖는 두개의 특성 사이의 관계를 객관적인 수치로 표현하기 위한 방법이다. 두 특성의 모집단을 x 와 y 변수로 나타내면 다음 식 (1)과 같다.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (1)$$

\bar{x} 는 x 의 평균값을 나타내고 \bar{y} 는 y 의 평균값을 나타낸다. 식 (1)로부터 x , y 의 모집단 전체를 표현하여 분산 값을 구하기 위한 형태로 표현하면 식(2) 같이 나타난다.

$$S_{xx}^2 = \sum_{i=1}^n (x_i - \bar{x})^2, \quad S_{yy}^2 = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2)$$

S_{xx}^2 은 x 로부터 평균값의 차이에 대한 제곱 값의 평균을 낸 분산을 나타내고 S_{yy}^2 은 y 로부터 평균값의 차이에 대한 제곱 값의 평균을 낸 분산을 나타낸다. 식 (2)로부터 특성의 분산 값을 구하는 식은 다음 식

(3) 같이 나타난다.

$$S_{xy}^2 = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (3)$$

식 (3)에서 구해진 분산 값을 이용하여 상관 계수를 계산하며 식 (4)과 같다.

$$r = \frac{S_{xy}^2}{S_{xx} \cdot S_{yy}} \quad (4)$$

r은 상관 계수를 나타내고 상관계수 r은 두 특성의 관련성을 나타내는 척도로 사용되어지며 연관성이 양으로 나타나면 상관 계수의 값도 양의 값을 갖는다. 연관성의 표현은 -1과 1사이의 값으로 표현되며 -1인 경우 음의 직선상에 나타나고 1인 경우 양의 직선상에 나타나게 된다.

상관 계수를 이용한 선호도 측정은 특정 사용자와 선호도가 비슷한 객체를 선정하는 기법으로 사용되며 이를 활용한 클러스터링 방법과 k-최대근접 방법을 이용한다[6].

미리 결정되어진 군집 수에 의하여 전체 데이터를 상대적으로 유사한 군집으로 구분하여 사용하는 방법이 클러스터링 방법이며[7], 과거 구매 기록을 통하여 특정 고객과 선호도가 가장 비슷한 k명의 고객들을 선택하는 방법이 k-최대근접 방법이다[8].

상관 계수를 이용한 선호도 측정 방법은 비슷한 선호도를 가지는 사용자들의 객체에 대한 평가를 바탕으로 추천하는 방법이기 때문에 높은 예측 추천 능력을 가지는 장점이 있다.

2.2 사용자 상황 처리 시스템

상황 인식은 사용자의 위치, 행위, 표정과 현재 시간, 주변 사물 등을 수집하여 사용자에게 유용한 정보를 추천한다. 모바일 환경의 단말기에서 제공되는 사용자 상황을 수집하여 위치와 특성, 현재의 시간 및 계절, 이동 수단, 개인 취향 정보 등을 이용하여 최신 영화, 주변 맛집, 주변 관광지 등을 추천한다. 상황 인지 기반 개인 맞춤형 서비스는 유비쿼터스 컴퓨팅 환경에서 의료, 교육, 재난, 구조, 쇼핑 등에 응용되고 있다[9].

추천 서비스 시스템은 자동화된 정보 필터링 기법을 적용하여 사용자의 맞춤 정보를 추천 서비스하며 선호도 추천 시스템은 사용자에게 개인 신상, 관심

분야, 선호도 등의 정보를 입력받아 사용자의 선호도에 알맞은 정보를 추천 서비스한다[10].

선호도 추천 서비스 기반의 시스템은 데이터를 이용하는 방법에 따라 분류하며 시스템에 저장되어 있는 기존 데이터와 사용자에 의해 입력되어지는 입력 데이터, 기존 데이터와 입력 데이터를 조합하여 형성된다[11].

사용자 선호도 추천 서비스 시스템에는 단일 협업 기법과 내용 기반 기법, 지식 기반 기법, 유용성 기반 기법 등이 있다.

단일 협업 기법은 사용자의 기초 정보와 고객들의 선호도를 선별하여 비슷한 패턴을 보이는 고객들을 하나의 그룹으로 묶어 추천 서비스를 결정하며 기존 데이터와 입력 데이터를 각 항목에 대한 점수로 표현하고 k-최대 근접 방법을 이용하여 점수가 유사한 고객을 찾아 처리하므로 기존 데이터의 의존도가 높으며 사용자 상황을 고려하기가 어려운 단점이 있다. 내용 기반 기법은 기존 데이터의 특징을 추출하고 추출된 특징으로 부터의 사용도와 입력 데이터의 점수에 따라 클러스터링을 이용하여 분류하므로 클러스터링 방법에 의존해야 한다.

지식 기반 기법은 고객 요구에 충족하는 지식의 기존 데이터의 필요와 흥미에 대한 입력 데이터를 이용하여 욕구를 충족시키는 항목을 추정하므로 확률에 의존하게 되고 기존 데이터가 많이 필요한 단점이 있다. 유용성 기반 기법은 기존 데이터의 특징을 추출하고 추출된 특징과 입력되는 데이터로부터 선호 항목의 유용성을 클러스터링 방법을 이용하여 분류한 후 선택하므로 클러스터링 방법에 의존해야 한다[12].

3. 시스템 모델

추천 서비스 시스템은 사용자와 추천 객체간의 상호 관계만을 표현하는 문제점과 상황 인식 서비스 시스템의 상황별 객체 선호도를 고려하지 않는 단점을 개선하기 위하여 다양하게 변화하는 상황 정보와 사용자의 기존 데이터를 이용하여 모델링된 정보로부터 사용자의 상황을 인식하여 사용자의 상황별 객체 선호도 처리를 수행하는 시스템 모델을 그림 1과 같이 구성하였다.

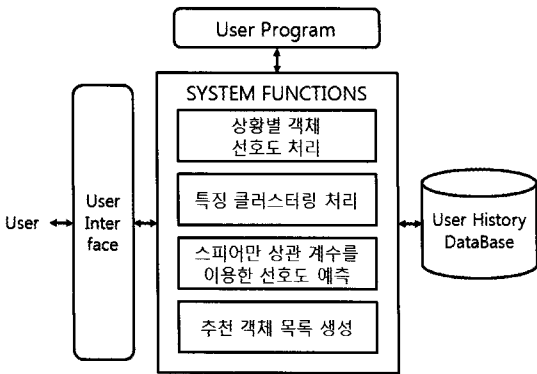


그림 1. 사용자의 특성을 이용한 추천 서비스 시스템

구성된 시스템에서 상황별 객체 선호도 처리에서는 사용자의 상황에 따라 변화하는 선호도를 처리하기 위해 단말기로부터 전송되어지는 정보를 객체별로 처리하기 위한 분류를 수행하며 특성 클러스터링 처리는 분류된 정보로부터 특성을 추출하여 단계별로 특성 클러스터링 처리를 수행한 후 스피어만 상관 계수를 이용하여 사용자의 상황별 선호도를 예측한다. 예측하고자 하는 사용자의 정보와 유사한 선호도를 가지는 사용자들을 선별하여 사용자들의 기존 데이터를 기반으로 객체에 대한 사용자의 상황별 선호도를 예측하고 유사 선호도를 구한다. 유사 선호도로부터 사용자의 입력 정보와 히스토리 데이터를 이용하여 최종적으로 추천 객체의 목록을 생성한다.

3.1 상황별 객체 선호도 처리

다양하게 변화하는 사용자의 상황에 대해 단말기로부터 사용자의 정보와 단말기 위치를 전송받아 주변의 정보와 상황 및 특성을 사용자 데이터로 분류하여 상황에 대한 모든 추천 객체로 이용한다. 사용자의 상황을 분석하고 특성을 인식하여 사용자의 상황에 맞는 추천 가능 범위와 사용자의 특성에 따라 추천 형태를 결정하여 사용자의 위치에서 가까운 범위 내의 사용자의 특성에 적합한 추천 객체를 생성한다.

각 객체의 사후 확률 $P(\omega_1|x)$ 이 가장 큰 값을 가진 객체를 결정하기 위해 사용하였으며 식 (5)와 같이 표현된다[13].

$$P(\omega_1|x) > P(\omega_2|x) \quad (5)$$

x 는 특징 벡터를 나타내고 ω_1 는 특징 벡터들이 포함된 클래스를 나타낸다. $P(\omega_1|x)$ 는 특징 벡터가 포함된 클래스의 확률을 나타내며, 클래스 ω_1 이 클래스 ω_2 보다 크면 클래스 ω_1 이 선택되어지며 그렇지 않으면 클래스 ω_2 가 선택되어진다. 식 (5)에 대하여 베이시안 정리를 적용하여 정리하면 식 (6)처럼 나타내어진다.

$$\frac{P(x|\omega_1)P(\omega_1)}{P(x)} > \frac{P(x|\omega_2)P(\omega_2)}{P(x)} \quad (6)$$

식 (6)에서 $P(x)$ 특징 벡터 x 의 확률 값을 나타내며 결정 규칙에 영향을 미치지 않아 제거하여 특징 벡터를 포함함 항과 클래스 항으로 정리하면 식 (7)와 같이 표현된다.

$$\Lambda(x) = \frac{P(x|\omega_1)}{P(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \quad (7)$$

특징 벡터를 포함한 클래스 ω_1 의 확률 값이 특징 벡터를 포함한 클래스 ω_2 의 확률 값보다 크면 특징 벡터를 포함한 클래스 ω_1 의 확률 값이 선택되어지며 그렇지 않으면 특징 벡터를 포함한 클래스 ω_2 의 확률 값이 선택되어진다.

식 (7)에서 $\Lambda(x)$ 는 우도비(likelihood)의 계산을 위한 수식으로 표현되며 확률 밀도 함수의 예측이 가능하고 선호도 예측 처리를 위해 사용하였다.

3.2 특성 클러스터링 처리

특성 클러스터링 처리는 유사한 선호도를 가지는 사용자들을 몇 개의 의미 있는 군집으로 나누기 위하여 사용한다. 객체의 각 속성에 대한 선호도를 각기 다른 차원으로 구성하여 좌표 평면의 점으로 표현하고 기존 사용자들을 k 개로 클러스터링 한다. 클러스터링을 통하여 나누어진 k 개의 각 군집의 대푯값들과 테스트 사용자의 각 속성의 선호도 값에 대한 거리를 계산하여 가장 최소의 값을 가지는 클러스터를 선택한다. 결정된 클러스터에 속하는 다른 사용자들은 테스트 사용자에 대한 새로운 클러스터로 구성된 다[14].

클러스터링을 위한 알고리즘은 다음과 같다.

```

Initialize the Clustering
Decision cluster k
Select cluster k = 1 to N
Select specimen k = 1 to N
{
    int k;
    for(k=1;k<=N;k++){
        m_mean[i]=0;
        for(k=0;k<m_cacheN;k++){
            for(int j=1;j<=m_dimN;j++){
                m_data[i][j]=0;
            }
        }
        new cluster calculate
    }
}
form collect data to classification
    
```

클러스터 k 를 결정하여 k 개의 초기 클러스터 중심을 선택하고 일반적으로 주어진 표본 집합의 처음 k 개의 표본을 임의로 선택한다. 각 사용자들을 각 클러스터의 중심과 가장 가까운 거리에 있는 군집 영역에 분배하여 모든 군집에 대한 해당 군집에 포함된 모든 사용자들의 선호도들로부터 새로운 클러스터 중심을 계산한다. 모든 군집에 대하여 기존의 중심과 새로운 중심의 차이가 없을 때까지 분배를 반복하고 그렇지 않으면 수렴하여 데이터로 분류한다.

3.3 스피어만 상관 계수를 이용한 선호도 예측

스피어만 상관 계수는 데이터가 순서에 의한 척도인 경우 사용된다. 측정형 변수나 순서형 분류형 변수들의 상관관계를 자료의 순위 값에 의하여 계산하는 방법으로 순서형 분류형 변수들의 상관관계를 계산한다.

데이터의 값 대신 순위를 이용하는 상관 계수이며 데이터를 작은 것부터 차례로 순위를 정하여 서열 순서로 바꾸어 순위를 이용해 상관 계수를 구하는 방법이다. 두 클러스터 간의 연관 관계가 있는지 없는지를 밝혀주며 데이터에 이상점이 있거나 표본크기가 작을 때 유용하게 사용된다[15].

사용자간의 유사도를 계산하기 위하여 사용자의 객체에 대한 선호도를 테이블로 형성한 후 식 (8)과 같이 스피어만 상관계수를 이용하여 선호도 점수를 변환한다.

$$r_s = 1 - \frac{\sum_{i=1}^n (R(x_i) - R(y_i))^2}{n(n^2 - 1)} \quad (8)$$

$R(x_i)$ 는 변수 x 의 i 번째 관측치의 순위를 표현하며 $R(y_i)$ 는 변수 y 의 i 번째 관측치의 순위를 표현하며 전체 서열 순위는 r_s 의 순위로 결정된다.

스피어만 상관 계수를 개선한 알고리즘은 다음과 같다.

```

void PhoneNode::Spearman()
{
    m_lastFrameX=-2;
    m_maxLike=MIN_LIKELIHOOD;
    m_prevID = NGRAM_NO_ID;
    m_curLM = MIN_LIKELIHOOD;
    for(int i=0;i<m_likeA.size();i++){
        m_likeA[i]=MIN_LIKELIHOOD;
        m_pathA[i]=0;
        m_pathPhoneA[i]=0;
    }
    if (IsHead())
    {
        for(i=0;i<m_amodelA.size();i++){
            m_amodelA[i] = m_amodelP;
        }
    }
}
    
```

데이터로 구성되어진 클러스터들에 대해 객체 목록을 생성하기 위하여 클러스터링된 데이터들의 연관 관계를 구하고 이를 서열 순서로 구성하여 스피어만 상관 계수를 이용하였다. 최초의 순위를 구하여 1순위로 부터의 상관 계수를 1과 -1로 나타내어 순위를 조절한다. 두 데이터의 순위가 완전히 일치하면 +1로 표현하고 두 데이터의 순위가 완전히 반대이면 -1로 표현된다.

3.4 추천 객체 목록 생성

모바일 단말기로부터 제공되는 사용자의 위치와 특성, 현재 시간 및 계절, 이동 수단, 개인 취향 정보 등을 수집하여 데이터를 구축하며 사용자의 입력 정보와 함께 사용자의 정보로 보존된다. 사용자 정보와 사용자로부터 선호하는 정보를 이용하여 최신 영화, 주변 맛집, 주변 관광지 등을 데이터로 분류한다. 사용자 정보와 선호하는 정보로부터 추천 서비스한 목

록을 정리하여 그림 2와 같이 사용자 히스토리 데이터를 구축한다.

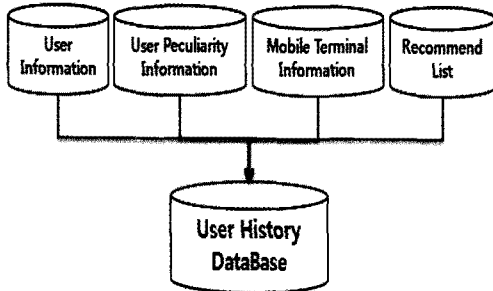


그림 2. 사용자 히스토리 데이터베이스

구축되어진 데이터베이스는 사용자로부터 서비스 요청을 받게 되면 재 추천을 위해 사용자의 히스토리 데이터로 활용하게 되며 또 다른 상황과 특성의 사용자 입력과 함께 선호도를 계산을 위해 사용된다.

사용자 상황 특성 선호도는 사용자 정보에 따라 선호도가 달라 질 수 있기 때문에 사용자의 히스토리 데이터와 입력 정보를 최대한 반영하여 선호도를 계산한다.

히스토리 데이터 결정을 위한 의사코드는 다음과 같다.

```

Begin
  Initialize the Database
Job1 : Get Input UserInfo
  If UserInfo is exist then next
  else goto end
  Clear RecomList
  Process the HistoryDB,
  Add the PeculInfo to Evaluation(n)
  Process the MobileInfo transition and
  Add the PeculInfo to RecomList
  Insert RecomList to HistoryDB
  Get the RecomList for transition
  Process HistoryDB
  Replace RecomList
  Goto Job1
Job2 : Create RecomList having
  max likelihood of HistoryDB
  
```

사용자의 특성에 따른 추천 객체의 결정과 적용 가능한 범위를 결정하여 추천 객체의 대상을 적합하

게 구성하고 사용자의 위치와 특성을 고려하여 생성된 추천 객체를 대상으로 사용자의 상황에 따른 입력 조건으로부터 파라미터의 값을 조정하여 조건에 적합하지 않은 추천 객체들을 제거한다. 사용자의 상황과 특성을 고려하여 서비스를 제공할 장치를 결정하고 사용자가 사용하는 단말기로 제공할 서비스 방법을 결정하여 전송한다.

4. 실험 결과 및 분석

본 논문에서 제안한 시스템의 성능을 평가하기 위해 효과성을 측정하여 정확도와 재현율로 나타낸다.

정확도(Precision)는 검색된 데이터들 가운데 연관된 데이터의 비율을 계산하여 연관된 데이터만을 선택하였을 경우 높은 결과를 보인다. 실제 선택한 데이터와 추천한 데이터의 교집합을 추천한 데이터로 나누어 표현한다. 재현율(Recall)은 검색된 데이터를 관련된 모든 데이터의 수로 나눈 수치이다. 선택한 데이터와 추천한 데이터의 교집합을 선택한 데이터로 나누어 표현한다[16].

다양한 특성 정보를 이용하기 위하여 식품, 맛집 [17], 관광, 도서, 인터넷 쇼핑물 등을 기존의 연구에서 분석하였으나 본 연구에서는 총 200개의 음식점 정보를 이용한 사용자 상황 특성 정보의 선호도 예측을 수행하였다. 표 1에 실험 데이터의 일부를 나타내었다.

표 1. 실험 데이터

음식점명	지역	가격	종류	전화번호	추천음식
이즈미	서울	3만원	일식	621-6543	히레까스
마루이	부산	2만원	한식	417-5035	정식
개화	대구	1만원	한식	643-2145	비빔밥
바우고개	충남	1만원	한식	806-1482	국화주
고센	광주	2만원	한식	244-0567	대나무통밥

실험 데이터는 지역별, 음식 종류별, 가격별, 추천 음식별로 구분하여 정리하였으며 음식점명과 전화번호를 추천할 수 있도록 하였다.

다음 그림 3은 시스템 탐색 화면을 나타내며, 어휘 관리와 음성 인식을 처리하기 위한 인터페이스 화면으로 구성된다.

그림 4는 시스템에서 처리되는 과정의 결과를 보인다. 인 것이다.

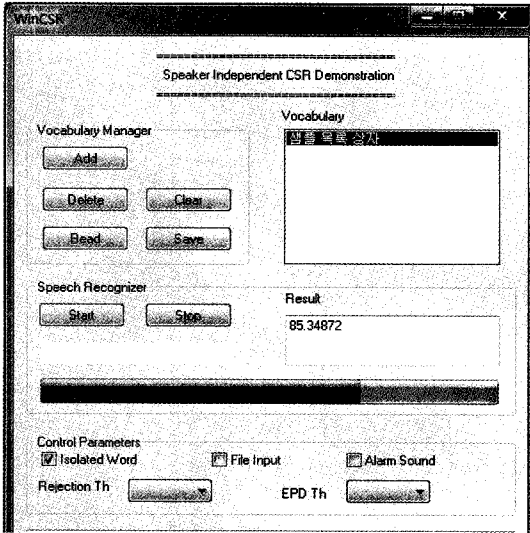


그림 3. 시스템 탐색 화면

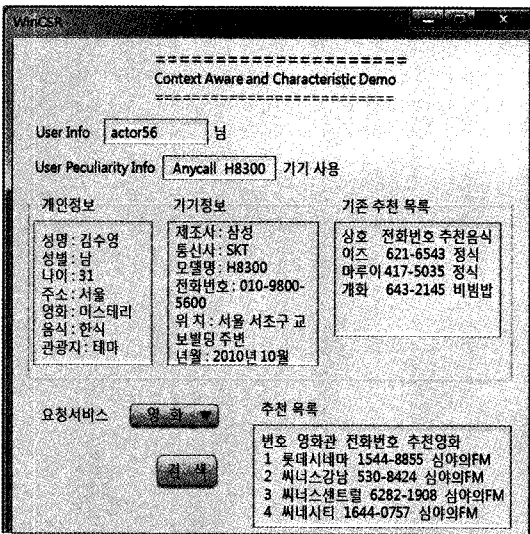


그림 4. 탐색 실행 결과

예측 값의 정확성을 평가하기 위해 MAE(Mean Absolute Error)를 사용하였다. MAE의 대한 수식은 식(9)와 같이 나타난다[18].

$$MAE = \frac{\sum_{i=0}^N |\epsilon_i|}{N} \quad (9)$$

N은 총 예측 회수, ϵ_i 는 예측 값과 실제 값의 오차, i 는 각 예측 단계를 나타낸다. 식 (9)를 이용하여 예측 값의 정확성 평가를 수행하였으며 표 2는 기존 연구와 비교한 결과이다.

표 2. 예측의 정확성 비교 평가 결과

	MAE
단일 협업	0.236
내용 기반	0.247
지식 기반	0.192
유용성 기반	0.186
제안 방법	0.183

제안 방법과 기존 연구의 비교 결과 기존 연구보다 정확성의 예측에서 좋은 예측 결과를 보였다.

성능 평가의 또 다른 방법은 추천 리스트를 평가하는 방법이다. 추천 리스트를 평가하기 위한 방법으로 정확도와 재현율을 사용하고 있으며 정확도는 추천 리스트 중 몇 개의 음식점을 실제로 선택했는지를 평가하고 재현율은 사용자가 선호하는 음식점 중 얼마나 많은 음식점이 추천되었는지를 평가한다. 추천 리스트를 평가하기 위해 정확도와 재현율을 이용하여 기존 연구와 비교 실험을 수행하였다.

표 3의 실험 결과 기존 추천 서비스 시스템들과 비교하여 정확도는 92.3%, 재현율은 73.8%의 성능을 보였다.

표 3. 효과성의 비교 실험 결과

	정확도(%)	재현율(%)
단일 협업	88.5	75.2
내용 기반	89.9	74.7
지식 기반	86.2	67.4
유용성 기반	90.5	76.1
제안 방법	92.3	73.8

k-means를 이용한 방법, Pearson 상관 계수를 이용한 방법과의 사용자가 선호하는 대상의 처리율과 선호대상의 정확성, 그리고 추천 목록 생성의 정확성을 각각 선호도 처리, 선호도 예측, 추천 목록 생성으로 표현하여 비교 실험을 수행하여 표 3에 나타내었으며 평가하기 위한 수식은 식 (10)과 같이 표현한다.

$$F = \frac{2 \times precision \times recall}{precision + recall} \quad (10)$$

표 4와 같이 실험 결과 선호도 처리에서 k-means를 이용한 방법보다 성능이 높게 나타났으나

Pearson 상관 계수를 이용한 방법과는 비슷한 결과를 보였으며 선호도 예측에서는 성능이 가장 높게 나타났다. 추천 목록 생성 비교 실험에서는 k-means를 이용한 방법보다 성능이 높게 나타났으나 Pearson 상관 계수를 이용한 방법과는 비슷한 결과를 보였다. 따라서 Spearman 상관 계수 이용하여 선호도 예측이 가능함을 실험을 통하여 확인하였다.

표 4. 추천 시스템의 비교 실험 결과

	k-means	Pearson	Spearman
선호도 처리(%)	95.3	97.6	97.2
선호도 예측(%)	96.5	98.7	98.8
추천 목록 생성(%)	98.2	99.1	98.8

비교 실험 결과 본 논문에서 제안한 사용자 정보 특성과 선호도를 이용한 정확도에서 개선된 성능을 보였으며 본 논문에서 200개의 제한된 지역적 음식점 정보와 적은 수의 사용자를 대상으로 실험하였지만 더 많은 음식점 정보를 수집한다면 특성에 따른 사용자의 선호도에 맞는 다양한 음식점을 추천할 수 있고, 여러 사용자를 대상으로 실험할 경우 유사도가 높은 사용자의 선호도를 이용할 수 있기 때문에 더 정확한 추천 서비스가 가능하다.

5. 결론

사용자의 상황과 특성에 따른 정보 선호도를 제공하기 힘들고 사용자가 원하는 정확한 정보를 추천하기가 어려우며 개념적인 정보만을 제한적으로 추천하는 시스템을 개선하기 위하여 본 논문에서는 스피어만 상관 계수를 이용하여 사용자 상황과 특성에 적합한 정보를 추천하는 시스템을 제안한다.

제안된 방법의 수행 과정은 단말기로부터 전송되어지는 정보를 사용자의 상황에 따라 변화하는 선호도를 처리하기 위해 객체별로 처리하기 위한 분류를 수행하며 분류된 정보로부터 특성을 추출하여 단계별로 특성 클러스터링 처리를 수행한 후 스피어만 상관 계수를 이용하여 사용자의 상황별 선호도를 예측한다. 예측하고자 하는 사용자의 정보와 유사한 선호도를 가지는 사용자들을 선별하여 사용자들의 기존 데이터를 기반으로 유사 선호도를 구한다. 사용자의 입력 정보와 히스토리 데이터를 이용하여 최종적

으로 추천 객체의 목록을 생성한다.

본 논문에서 제안한 시스템의 성능 실험은 기존의 서비스 시스템들과 비교하여 효과성을 측정하여 정확도와 재현율로 표현하였으며, 성능 실험 결과 정확도는 92.3%, 재현율은 73.8%로 나타났다.

참고 문헌

- [1] 정현만, 이정현, "유비쿼터스 컴퓨팅 환경에서의 상황 인식을 위한 확률 확장 온톨로지 모델," 한국컴퓨터정보학회논문지, 제11권, 제3호, 239-248쪽, 2006년 7월.
- [2] 한동조, 박대영, 최기호, "사용자 상황을 이용한 추천 서비스 시스템의 필터링 기법에 관한 연구," 한국ITS학회논문지, 제8권, 제1호, 119-126쪽, 2009년 2월.
- [3] 박지선, 김택현, 류영석, 양성봉, "추천시스템을 위한 2-way 협동적 필터링 방법을 이용한 예측 알고리즘," 정보과학회논문지, 제29권, 제9호, 669-675쪽, 2002년 10월.
- [4] 안명환, 권준희, "개념 계층 모델을 이용한 온톨로지 기반 상황 인식 추천 서비스," 인터넷정보학회, 제8권, 제5호, 81-89쪽, 2007년 10월.
- [5] 조광수, 최수영, 김화중, "유비쿼터스 컴퓨팅 환경에서의 효율적인 상황 정보 이용을 위한 데이터 관리 구조에 대한 연구," 2005 대한전자공학회 통신소사이어티 하계학술대회 논문집, 제28권, 제1호, 2005년.
- [6] S. Poslad, H. Laamanen, R. Malaka, A. Nick, P. Buckle, and A. Zipf, "CRUMPET: Creation of User-friendly Mobile Services Personalised for Tourism," Second International Conference on 3G Mobile Communication Technologies, pp.28-32, 2001.
- [7] G. Chen and D. Korz, "Context Aggregation and Dissemination in Ubiquitous Computing System," Proc. of the Fourth IEEE Workshop on Mobile Computing Systems and Applications, pp.105, 2005.
- [8] A. Ferscha, C. Holzmann, and S. Oppl, "Context Awareness for Group Interaction Support," Proc. of the Second International Work-

shop on Mobility Management and Wireless Access Protocols, pp.88-97, 2004.

[9] Norman Sadeh, "A Sematic Web Enviroment for Context-Aware Mobile Compus Services," Proc. of Wireless World Research Forum Conference, 2001.

[10] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," Accepted for publication at the WWW10 Conference, May, 2001.

[11] Harry Chen, Sovrin Tolia, Craig Sayers, Tim Finin, and Anupam Joshi, "Creating Context-Aware Software Agents," Proc. of the first GSFC/JPL Workshop on Radical Agent Concepts, pp.186-200, 2001.

[12] Badrul M. Sarwar, George Karypis, Joseph A. Konstan, and John T. Riedle, "Application of Dimensionality Reduction in Recommender System-A Case Study," ACM WebKDD 2000 Web Mining for E-Commerce Workshop, 2000.

[13] 한학용, "패턴인식 개론," 한빛미디어, 130-167쪽, 2005년 7월.

[14] Herlocker, J., "Understanding and Improving Automated Collaborative Filtering Systems," Ph. D. Thesis, Computer Science Dept., University of Minnesota, 2000.

[15] 김민호, 김진흠, "회귀나무에서 변수선택 편의에 관한 연구," 한국통계학회:학술대회논문집, 한국통계학회 2003년도 추계 학술발표회 논문집, 263-268쪽, 2003년 10월.

[16] Latifur Khan, Dennis McLeod, Eduard H. Hovy, "Retrieval effectiveness of an ontology-based model for information selection," VLDB J. 13(1): pp.71-85, 2004.

[17] 손창환, "Web상에서 개인화된 상품추천을 위한 Hybrid 추천시스템에 관한 연구," 박사학위논문, 영남대학교 대학원, 2006년.

[18] Jonathan L. Herlocker, Joseph A. Kosran, Al Borchers, and John Riedl, "An Algorithm Framework for Performing Collaborative Filtering," Proceedings of the 1999 Conference on Research and Development in Information Retrieval, 1999.



안 찬 식

2002년 광운대학교 컴퓨터공학과 석사
 2004년 광운대학교 컴퓨터공학과 박사수료
 관심분야: 음성인식, 분산처리, 음성/음향 신호처리



오 상 엽

1999년 광운대학교 전자계산학과 박사
 1993년~현재 경원대학교 IT대학 컴퓨터소프트웨어 교수
 관심분야: 소프트웨어공학, 버전관리, 소프트웨어 재사용, 형상관리, 객체지향, 음성인식, 분산 처리, 음성/음향 신호처리