

개선된 k-means 알고리즘을 적용한 사용자 특성 선호도 추천 시스템

안 찬 식*, 오 상 엽**

User's Individuality Preference Recommendation System using Improved k-means Algorithm

Chan-Shik Ahn*, Sang-Yeob Oh**

요 약

모바일 단말기에서 사용자의 상황을 고려하고 사용자의 취향이나 특성을 반영하여 정보를 찾아주거나 추천하는 서비스 시스템은 개념적인 정보만을 제한적으로 추천한다. 또한 사용자의 특성에 따른 정보 선호도를 제공하지 않으므로 정확한 정보 추천의 어려운 단점이 있다. 따라서 본 논문에서는 사용자 특성에 따른 선호도를 고려하여 정확한 상황 정보를 추천 할 수 있는 개선된 k-means 알고리즘을 적용하여 사용자 특성에 따른 선호도 추천 시스템을 제안하였다. 본 연구에서는 사용자 특성에 따른 선호도를 상관 계수를 이용하여 구하고 사용자의 특성 선호도를 개선된 k-means 알고리즘을 이용하여 추천하였다. 제한적인 개념의 정보만을 제공하던 시스템에서 사용자의 특성에 따른 정보 선호도를 제공하여 정확한 정보를 추천하므로 제한된 정보 추천의 단점을 해결하였다. 성능 실험은 기존의 서비스 시스템들과 비교하여 정확도와 재현율로 대변되는 효과성을 측정하였으며, 성능 실험 결과 정확도는 85%, 재현율은 68%로 나타났다.

Abstract

In mobile terminal recommend service system has general information restrictive recommend that individuality considering to user's information find and recommend. Also it has difficult of accurate information recommend bad points user's not offer individuality information preference recommend service. Therefore this paper is propose user's information individuality preference considering by user's individuality preference recommendation system using improved k-means algorithm. Propose method is correlation coefficients using user's information individuality preference when user's individuality preference recommendation using improved k-means algorithm. Restrictive information recommend to fix a problem, information of restrictive general recommend that user's information individuality preference offer to accurate information recommend. Performance experiment is existing service system as compared to evaluating the effectiveness of precision and recall, performance experiment result is appear to precision 85%, recall 68%.

• 제1저자 : 안찬식 교신저자 : 오상엽

• 투고일 : 2010. 03. 30, 심사일 : 2010. 05. 10, 게재확정일 : 2010. 05. 26.

* 광운대학교 컴퓨터공학과 박사과정 ** 경원대학교 IT대학 컴퓨터소프트웨어

※ 본 연구는 2010년도 경원대학교 지원에 의한 결과임.

- ▶ Keyword : 정보 추천(information recommendation), 서비스 시스템(service systems), 사용자 특성 선호도(user individuality preference), k-means 알고리즘(k-means algorithm)

I. 서론

멀티미디어 콘텐츠의 성장과 유비쿼터스 기술의 발전으로 인하여 모바일 콘텐츠 시장이 성장하였으며 모바일 단말기들은 네트워크로 연결되어 정보를 교환하고 시간, 공간에서 사용자의 특성을 인지하여 사용자에게 특성에 맞는 서비스를 제공한다[1]. 모바일 단말기들은 고정된 장소만이 아닌 사용자에게 따라 다른 장소와 서로 다른 시간에 이용될 수 있기 때문에 사용자의 환경적 특성에 따라 선호하는 콘텐츠가 달라질 수 있다[2]. 다양한 콘텐츠들 중에 사용자의 선호 정보를 파악하여 사용자의 개별적인 취향에 따른 콘텐츠 추천의 역할이 중요하다. 이에 사용자에게 유용한 서비스를 제공해 주는 인식 기술이 사용된다[3].

하지만, 사용자가 찾고자 하는 콘텐츠에 관한 정보를 직접 입력하기 때문에 입력 시간과 검색하기 위한 시간이 많이 필요하게 되고 효율적인 콘텐츠 제공을 가능하게 하기 위한 콘텐츠 검색 기법은 개념적인 정보만을 제한적으로 추천하고 있다[4].

기존의 특성 인식 기반 추천 서비스 시스템에는 사용자와 친화적이고 개인화를 적용시킨 관광 추천 서비스 시스템 CRUMPET[5]이 있다. 이는 추천을 이용하여 사용자에게 의미적으로 가장 가까운 정보를 추천하지만 사용자의 입력이 필요하고 개념 수준의 정보만을 추천하는 단점이 있으며 사용자의 특성에 따른 정보를 고려하지 않고 선호도를 추천하여 서비스를 제공하므로 사용자의 특성에 따른 정확한 정보 추천이 어렵다[6].

따라서, k-means 알고리즘을 개선한 사용자 특성 선호도 추천 시스템을 제안한다. 제안된 기법은 제한적인 개념의 정보만을 제공하던 것을 사용자 특성 정보를 모바일 단말기로부터 입력받아 사용자 특성 정보를 최대한 반영하여 선호도를 계산하므로 제한적인 개념의 정보를 제공하는 단점을 사용자 특성 정보로 제공하여 해결하였으며 사용자의 특성별 객체를 구하여 정확한 정보를 추천하므로 사용자의 특성에 따른 정보를 고려하지 않고 서비스를 제공하는 어려움을 극복하였다.

제안된 방법의 수행 과정에서는 특성을 추출하여 특성 처리와 특성 클러스터링을 개선된 k-means 알고리즘을 이용하여 수행하는 특성 클러스터링 처리를 수행하고 상관 계수를 이용하여 예측하고자 하는 사용자와 유사한 선호도를 가지는 사용자들을 선별한다. 선별된 사용자들의 평가를 기반으로 객체에 대한 사용자의 특성별 선호도를 예측하여 유사 선호도를

구하고 유사 선호도로부터 최종적으로 추천 객체의 목록을 생성한다. 성능 실험은 기존의 서비스 시스템들과 비교하여 정확도와 재현율로 대변되는 효과성을 측정하였다. 성능 실험 결과 정확도는 85%, 재현율은 67%로 나타났다.

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

II. 기존 연구

2.1 특성 인식 정보

특성은 실세계에 존재하는 특징화된 개체의 상태를 정의한 정보이다. 특성 정보는 정보들이 상호작용을 통해 참여자의 특성을 특성화한 것이다. 사용자의 현재 특성 정보에 따라 적절한 정보 혹은 서비스를 제공하기 위해 이러한 특성을 이용하는 것을 특성 인식이라 한다[7]. 특성 인식 시스템에서는 입력된 정보와 사용자의 특성 정보를 고려하여 선정 기법을 이용한 서비스의 가능 여부 또는 서비스의 종류를 결정한다[8]. 특정 사용자와 비슷한 선호도를 가지는 객체를 선정하는 기법에는 클러스터링, k -최대근접과 같은 방법을 이용한다[9]. 클러스터링 방법은 사전에 결정된 군집 수 k 에 기초하여 전체 데이터를 상대적으로 유사한 k 개의 군집으로 구분하는 방법이며 각 데이터는 좌표평면의 점으로 표현하여 사용한다[10]. k -최대근접 방법은 과거 구매 기록을 통하여 특정 고객과 선호도가 가장 비슷한 k 명의 고객들을 선택한다[11].

사용자가 좋아할 만한 객체를 예측하기 위하여 비슷한 선호도를 가지는 다른 사용자들의 객체에 대한 평가를 근거로 사용하여 추천하는 방법이므로 높은 예측 능력과 추천 능력을 가지는 장점이 있다. 특성 인지 기반 개인 맞춤형 서비스는 유비쿼터스 컴퓨팅 환경에서 의료, 교육, 재난, 구조, 쇼핑 등에 응용된다. 특성인식은 사용자의 위치, 행위, 표정과 현재 시간, 주변 사물 등을 수집하여 서비스된다. 모바일 환경의 단말기에서 제공되는 사용자 특성을 수집하여 위치와 위치의 특성, 현재의 시간 및 계절, 이동수단, 개인 취향 정보 등을 이용하여 최신 영화, 주변 맛집, 주변 관광지 등을 추천한다[12].

2.2 선호도 추천 시스템

개인화된 추천 시스템은 자동화된 정보 필터링 기법을 적용하여 사용자의 취향과 특성에 맞는 정보를 추천해주는 시스템이다. 선호도 추천 시스템은 사용자에게 개인 신상, 관심 분야, 선호도 등의 질문을 하여 사용자의 선호도 정보에 알맞은 정보 및 서비스를 추천하는 것이다[13]. 선호도 추천은 시스템 기반의 데이터를 이용하는 방법에 따라 분류할 수 있으며 시스템에 저장되어 있는 기존 데이터와 사용자에 의해 입력되어지는 입력 데이터, 그리고 기존 데이터와 입력 데이터를 조합하는 알고리즘 등으로 형성된다[14].

선호도 추천 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 혹은 서비스를 결정한다. 기존 데이터와 입력 데이터를 각 항목에 대한 점수로 표현하고 k -최대근접 방법을 이용하여 점수가 유사한 고객을 찾아 처리하므로 기존 데이터의 의존도가 높으며 사용자 특성을 고려하기가 어렵다. 내용 기반 기법은 기존 데이터의 특징을 추출하고 추출된 특징으로 부터의 사용도와 입력 데이터의 점수에 따라 클러스터링을 이용하여 분류하므로 클러스터링 방법에 의존한다. 지식 기반 기법은 고객 요구에 충족하는 지식의 기존 데이터와 필요와 흥미에 대한 입력 데이터를 이용하여 욕구를 충족시키는 항목을 추정하므로 확률에 의존하게 되고 기존 데이터가 많이 필요하다. 유용성 기반 기법은 기존 데이터의 특징을 추출하고 추출된 특징과 입력되는 데이터로부터 선호 항목의 유용성을 클러스터링 방법을 이용하여 분류한 후 선택하므로 클러스터링 방법에 의존한다.

III. 시스템 모델

일반적인 추천 서비스 시스템은 사용자와 추천 객체간의 상호 관계만을 표현하는 문제점을 가지고 있으며 특성 인식 서비스 시스템도 특성별 객체 선호도를 고려하지 않고 있다. 본 연구에서는 그림 1과 같이 다양한 특성 정보와 모델링된 정보로부터 사용자의 특성을 인식하여 사용자 특성 정보를 모바일 단말기로부터 전송받는다. 전송 받은 사용자의 특성 정보를 이용하여 사용자의 특성별 객체 선호도 처리를 수행한다. 특성 클러스터링 처리는 특성을 추출하여 특성 처리와 특성 클러스터링을 개선된 k-means 알고리즘을 이용하여 수행하는 특성 클러스터링 처리를 수행한 후 상관 계수를 이용하여 사용자의 특성별 선호도를 예측한다.

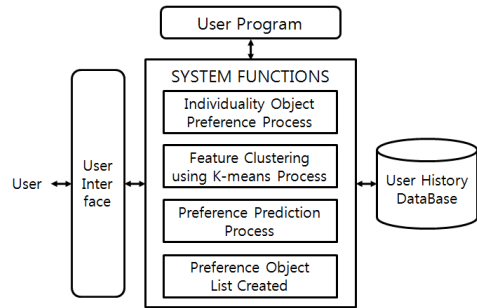


그림 1. 사용자의 특성을 이용한 추천 서비스 시스템
Figure 1. Recommendation system using user's context

예측하고자 하는 사용자와 유사한 선호도를 가지는 사용자들을 선별하여 사용자들의 평가를 기반으로 객체에 대한 사용자의 특성별 선호도를 예측하고 유사 선호도를 구한다. 유사 선호도로부터 최종적으로 추천 객체의 목록을 생성한다.

3.1 특성별 객체 선호도 처리

사용자의 위치를 인식하여 위치에 존재하는 모든 추천 객체를 선택하고, 사용자의 상태를 인식하여 사용자의 상태에 따라 추천 가능한 범위를 결정하여 사용자의 위치에서 가까운 범위 내의 추천 객체를 생성한다. 생성된 추천 객체를 대상으로 사용자의 특성에 따른 조건 파라미터의 값을 조정하여 조건에 맞지 않는 추천 객체를 삭제한다. 사용자의 특성을 인식하여 서비스를 제공할 장치를 결정하여 해당 장치에 제공할 서비스 방법을 결정한다. 사용자간의 유사도를 계산하기 위하여 사용자의 객체에 대한 선호도를 테이블로 형성한 후 식 (1)과 같이 상관계수를 이용하여 선호도를 계산한다[15].

$$Sim_{ij} = \frac{\sum (S_{ik} - \overline{S}_i)(S_{jk} - \overline{S}_j)}{\sqrt{\sum (S_{ik} - \overline{S}_i)^2 \cdot (S_{jk} - \overline{S}_j)^2}} \dots\dots\dots (1)$$

S_{ik} 는 사용자 i 가 객체의 종류 k 에 대해 평가한 선호도, S_{jk} 는 사용자 j 가 객체의 종류 k 에 대해 평가한 선호도, \overline{S}_i 는 사용자 i 의 선호도 평균, \overline{S}_j 는 사용자 j 의 선호도 평균을 나타내며, Sim_{ij} 는 두 사용자의 선호도가 유사한 경우 1에 가까운 값을 가지며, 선호도가 상반되는 경우 -1에 가까운 값을 가진다. 특성 처리 기능은 사용자로부터 입력되어진 특성을 추출하여 클러스터링 하기 위한 전 처리과정으로써 특성들을 그룹으로 나누는 과정을 수행한다. 처리된 특성들은 클러스터링 과정에서 근접한 거리에 있는 특성들을 묶는 과정을 수행한다.

3.2 개선된 특성 클러스터링 처리

특성 클러스터링 처리는 기존 k-means 알고리즘을 개선하여 이용하였다. 유사한 선호도를 가지는 사용자들을 몇 개의 의미 있는 군집으로 나누기 위하여 개선된 k-means 알고리즘을 사용한다. 객체의 각 속성에 대한 선호도를 각기 다른 차원으로 구성하여 좌표 평면의 점으로 표현하고 개선된 k-means 알고리즘을 이용하여 기존 사용자들을 k개로 클러스터링 한다.

개선된 k-means 알고리즘 통하여 나누어진 k개의 각 군집의 대표값들과 테스트 사용자의 각 속성의 선호도 값에 대한 거리를 계산하여 가장 최소의 값을 가지는 클러스터를 선택한다. 결정된 클러스터에 속하는 다른 사용자들은 테스트 사용자에게 대한 새롭게 구성된 클러스터들이 된다.

기존의 k-means 알고리즘은 임의의 클러스터 k를 정한 후, k개의 초기 클러스터 중심을 선택하고 일반적으로 주어진 표본 집합의 처음 k개의 표본을 임의로 선택하여 각 사용자들을 각 클러스터의 중심과 가장 가까운 거리에 있는 군집 영역에 분배하며 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 수렴하게 된다.

클러스터링을 위한 개선된 k-means 알고리즘은 다음과 같다.

```

Featurek-means
InitOncePerUtterance(const string& filename)
{
    m_localFrameX=0;
    m_bufN=0;
    int k;
    for(k=1;k<=m_dimN;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;
        }
    }
    return true
}
    
```

개선된 k-means 알고리즘에서는 클러스터 k를 선택할시 모든 데이터를 2차원 배열에 저장한 후 킥정렬을 통하여 데이터를 정렬하여 정렬된 상태에서 클러스터 k를 선택하여 클러스터링을 수행한다.

3.3 선호도 예측 처리

상관 계수는 두 변량 사이의 상관관계의 정도를 나타내는 수치로 사용된다. 데이터 사이의 상관 계수를 작은 것부터 차례로 순위를 정하여 상관 계수를 구하며 두 클러스터 간의 상관관계를 유무를 확인하며 데이터 간의 특징이 속한 클래스를 결정하는데 사용한다. 각 클래스의 사후 확률 $P(\omega_i|x)$ 가 가장 큰 값을 가진 클래스를 결정하기 위해 사용하며 식 (2)와 같이 표현한다.

$$P(\omega_1|x) \begin{matrix} \omega_1 P(\omega_2|x) \\ > \\ < \\ \omega_2 \end{matrix} \dots\dots\dots (2)$$

ω_1 이 ω_2 보다 크면 ω_1 을 선택하고 그렇지 않으면 ω_2 를 선택한다. 식 (2)에 대해서 베이시안 정리를 적용하면 식 (3)처럼 나타낸다.

$$\frac{P(x|\omega_1)P(\omega_1)}{P(x)} \begin{matrix} \omega_1 \frac{P(x|\omega_2)P(\omega_2)}{P(x)} \\ > \\ < \\ \omega_2 \end{matrix} \dots\dots\dots (3)$$

식 (3)에서 $P(x)$ 는 같은 값으로 결정 규칙에 영향을 미치지 않으므로 제거하여 다시 정리하면 식 (4)와 같이 표현된다.

$$\Lambda(x) = \frac{P(x|\omega_1)}{P(x|\omega_2)} \begin{matrix} \omega_1 \frac{P(\omega_2)}{P(\omega_1)} \\ > \\ < \\ \omega_2 \end{matrix} \dots\dots\dots (4)$$

식 (4)에서 $\Lambda(x)$ 는 우도비(likelihood)로 표현되며 확률 밀도 함수의 예측이 가능하며 선호도 예측처리를 위해 사용하였다. 선호도 예측 처리를 위해 확률 밀도 함수로 구성되어진 상관계수를 이용하였다. 데이터로 구성되어진 클러스터들에 대해 객체 목록을 생성하기 위하여 클러스터링된 데이터들의 연관 관계를 구하고 이를 확률 밀도 함수로 구성하여 상관계수로 활용하였다. 최초의 순위를 구하여 1순위로 부터의 상관계수를 1과 -1로 나타내어 순위를 조절하였다.

선호도 예측 처리를 위해 상관 계수를 이용한 알고리즘은 다음과 같다.

```

void
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;
    }
}
    
```

3.4 추천 객체 목록 생성

모바일 단말기로부터 제공되는 사용자의 위치와 그 위치의 특성, 현재 시간 및 계절, 이동 수단, 개인 취향 정보 등을 수집하여 데이터를 구축한다. 구축되어진 데이터는 사용자의 입력 정보와 함께 구성되며 사용자의 히스토리 데이터로 보존된다. 사용자로부터 입력된 정보를 이용하여 선호하는 최신 영화, 주변 맛집, 주변 관광지 등을 분류하여 데이터로 구성하고 선호하여 추천 서비스한 목록 또한 분류하여 그림 2와 같이 정리한다.

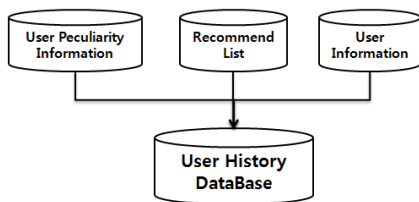


그림 2. 사용자 히스토리 데이터베이스
Figure 2. User History Database

정리하여 구축되어진 데이터는 재 추천 시 히스토리 데이터로 활용하고 입력되어진 정보와 함께 사용자 정보 특성에 따른 선호도를 계산하기 위해 사용된다. 사용자 정보 특성 선호도는 사용자 정보 특성에 따라 선호도가 달라질 수 있기 때문에 사용자의 정보를 최대한 반영하여 선호도를 계산한다.

사용자의 특성에 따른 추천 객체의 결정과 적용 가능한 범위를 결정하여 추천 객체의 대상을 줄이고 사용자의 특성을

인식하여 추천 객체 목록을 제공할 장치에 따라 추천 객체의 서비스 제공 방법을 다르게 한다. 추천 대상이 되는 객체를 선별하는 과정은 사용자의 특성을 고려하여 추천 대상이 되는 객체를 결정하여 목록을 생성하고 이를 사용자의 이용기기에 서비스한다.

IV. 실험결과 및 분석

본 논문에서 제안한 시스템의 성능을 평가하기 위해 [16]의 정확도와 재현율로 표현한다. 정확도(Precision)는 검색된 모든 데이터 중에서 연관된 데이터의 비율을 계산한 것으로 오직 연관된 데이터만을 가져왔을 경우 높은 수치를 보인다. 실제 선택한 데이터와 추천한 데이터의 교집합을 추천한 데이터로 나누어 나타낸다. 재현율(Recall)은 검색된 데이터에 대해 관련된 모든 데이터의 수로 나눈 수치를 보인다. 선택한 데이터와 추천한 데이터의 교집합을 선택한 데이터로 나누어 나타낸다[17].

기존의 연구에서 분석을 위한 식품, 맛집[18], 관광, 도서, 인터넷 쇼핑몰 등이 있으나 본 연구에서는 다양한 특성 정보를 이용하기 위한 샘플 중에서 총 200개의 음식점 정보를 이용한 사용자 정보 특성과 선호도 예측을 수행하였다. 표 1에 실험 데이터의 일부를 나타내었다.

표 1. 실험 데이터
Table 1. Experimentation data

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, 추천음식별로 구분하여 정리하였으며 음식점명과 전화번호를 추천할 수 있도록 하였다.

본 논문에서 제안한 시스템의 성능을 평가하기 위한 실험 환경을 표 2에 표시하였으며 운영체제는 Windows XP를 사용하였으며 데이터베이스는 MySQL 5.0, 시멘틱 웹 라이브러리는 Jess를 사용하였다.

표 2. 실험 환경

Table 2. Experiment environment

운영체제	Windows XP
데이터베이스	MySQL 5.0
온톨로지 모델링 도구	Protege 3.3.1
시멘틱 웹 라이브러리	Jess
온톨로지를 위한 쿼리 언어	RDF

예측 값의 정확성을 평가하기 위해 MAE(Mean Absolute Error)를 사용하였고 식(5)와 같이 구하였다.[19]

$$|E| = \frac{\sum_{i=1}^N |\varepsilon_i|}{N} \dots\dots\dots (5)$$

N은 총 예측 회수를 나타내고, ε_i는 예측 값과 실제 값의 오차를 나타내며 i는 각 예측 단계를 나타낸다.

표 3은 식 (5)를 이용하여 예측 값의 정확성 평가를 수행한 결과이다.

표 3. 예측의 정확성 평가 결과

Table 3. Experiment result accuracy of Prediction

	MAE
단일 협업	0.236
내용 기반	0.247
지식 기반	0.192
유용성 기반	0.188
기존 K-means	0.185
제안 방법	0.183

또한 추천 리스트를 평가하기 위한 방법으로 precision, recall을 사용하여 precision은 추천 리스트 중 몇 개의 음식점을 실제로 선택했는지를 평가하였으며 recall은 사용자가 선호하는 음식점 중 얼마나 많은 음식점이 추천되었는지를 평가하였다. precision과 recall은 동등한 중요도로 하나의 평가방법으로 사용하는 방법이다. 평가하기 위한 수식은 식 (6)과 같이 표현한다.

$$F = \frac{2 \times precision \times recall}{precision + recall} \dots\dots\dots (6)$$

표 4. 추천 시스템의 실험 결과

Table 4. Experiment result of recommendation service system

	Precision	Recall
단일 협업	76	78
내용 기반	79	74
지식 기반	61	64
유용성 기반	82	78
기존 K-means	83	65
제안 방법	85	68

제안 방법의 성능을 평가하기 위해 기존 단일 협업 방법, 내용 기반 방법, 지식 기반 방법, 유용성 기반 방법과 비교 실험을 수행하였다.

표 4의 비교 실험 결과를 보면 본 논문에서 제안한 사용자 정보 특성과 선호도를 이용한 제안 방법이 기존 필터링 방법 보다 정확도에서 우수한 성능을 보였다.

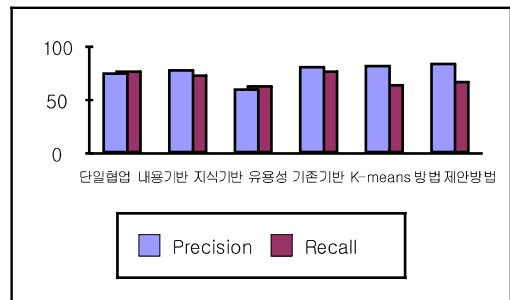


그림 3. 추천 시스템의 실험 결과

Figure 3. Experiment result of recommendation service system

실험결과 기존 추천 서비스 시스템들과 비교하여 precision의 경우 각각 11%, 2% 향상되었고, recall의 경우 각각 8%, 4% 향상되었으며, 전체적으로 precision 77%, recall 53%의 성능을 보였다.

본 논문에서는 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.

V. 결론

본 논문은 모바일 단말기 추천 서비스 시스템에서 개선된 k-means 알고리즘 기반 사용자 특성 선호도 추천 시스템을 제안한다.

제안된 기법은 제한적인 개념의 정보를 제공하는 시스템에서 모바일 단말기로부터 사용자 특성 정보를 입력받아 선호도를 계산하여 제공하므로 제한적인 개념의 정보를 제공하는 단점을 해결하였으며 사용자의 특성별 객체를 모델링하여 정확한 정보를 추천하기 위한 히스토리 데이터로 활용하여 사용자의 특성에 따른 정보를 고려하지 않고 서비스를 제공하는 어려움을 극복하였다.

제안된 방법의 수행 과정에서는 특성을 추출하여 특성 처리를 수행하고 특성 클러스터링을 개선된 k-means 알고리즘을 이용하여 수행하였다. 상관 계수를 이용하여 예측하고자 하는 사용자와 유사한 선호도를 가지는 사용자들을 선별할 수 있으며 확률 밀도 함수의 예측이 가능하여 선호도 예측처리를 수행한다. 선별된 사용자들의 히스토리 데이터를 활용하여 객체에 대한 사용자의 특성별 선호도를 예측하고 유사 선호도를 구하여 유사 선호도와 히스토리 데이터로부터 최종적으로 추천 객체의 목록을 생성한다.

본 논문에서 제안한 성능 실험은 기존의 서비스 시스템들과 비교하였으며 정확도와 재현율로 대변되는 효과성으로 측정하였다. 성능 실험 결과 정확도는 85%, 재현율은 67%로 나타났다.

참고문헌

[1] 윤선희, "소프트웨어기반 상황인식활용 인터넷쇼핑몰의 설계 및 구현," 한국컴퓨터정보학회논문지, 제14권, 제1호, 183-190쪽, 2009년 1월.

[2] 정현만, 이정현, "유비쿼터스 컴퓨팅 환경에서의 상황 인식을 위한 확률 확장 온톨로지 모델," 한국컴퓨터정보학회논문지, 제11권, 제3호, 239-248쪽, 2006년 7월.

[3] 조광수, 최수영, 김화중, "유비쿼터스 컴퓨팅 환경에서의 효율적인 상황 정보 이용을 위한 데이터 관리 구조에 대한 연구," 2005 대한전자공학회 통신소사이어티 하계학술대회 논문집, 제28권, 제1호, 2005년.

[4] 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.

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

[6] 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.

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

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

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

[11] 한학용, "패턴인식 개론," 한빛미디어, 2005년 7월.

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

[13] 박지선, 김택현, 류영석, 양성봉, "추천시스템을 위한 2-way 협동적 필터링 방법을 이용한 예측 알고리즘," 정보과학회논문지, 제29권, 제9호, 669-675쪽, 2002년 10월.

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

- [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] 안명환, 권준희, "개념 계층 모델을 이용한 온톨로지 기반 상황 인식 추천 서비스," 인터넷정보학회, 제8권, 제5호, 81-89쪽, 2007년 10월.
- [18] 한동조, 박대영, 최기호, "사용자 상황을 이용한 추천 서비스 시스템의 필터링 기법에 관한 연구," 한국ITS학회 논문지, 제8권, 제1호, 119-126쪽, 2009년 2월.
- [19] 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.

저자 소개



안 찬 식

2002 : 광운대학교 컴퓨터공학과 공학석사.

2004 : 광운대학교 컴퓨터공학과 박사수료.

관심분야 : 음성인식, 분산처리, 음성/음향 신호처리



오 상 엽

1991 : 광운대학교 전자계산학과 이학석사.

1999 : 광운대학교 전자계산학과 이학박사.

현 재 : 경원대학교 IT대학 컴퓨터 소프트웨어 교수

관심분야 : 소프트웨어공학, 버전관리, 소프트웨어재사용, 형상관리, 객체지향, 음성인식, 분산처리, 음성/음향 신호처리