

사용자의 행동 분석을 위한 과거 기록의 협력 필터링 적용

김용준¹, 박정은², 오경환³

서강대학교^{1 2 3}

{adeuxist¹, fayemint², kwoh³}@sogang.ac.kr

Applying Collaborative Filtering for Analysis of User's behavior

Yong Jun Kim¹, Jung Eun Park², Kyung Hwan Oh³
Sogang Univ.^{1 2 3}

요약

모든 곳에 존재하는 네트워크 환경을 의미하는 '유비쿼터스' 시대와 최신 기술로 구현되어 인간을 도와주는 '지능형 로봇'의 시대가 도래하고 있다. 기술의 흐름은, 이제 우리에게 공장과 공원 등의 공공 장소뿐 만이 아니라, 생활의 기본이 되는 가정 안에서의 로봇을 받아들일 준비를 요구하고 있다.

로봇과 사용자는 실제 생활 속에서 많은 상호 작용을 하게 되며, 필연적으로 여러 가지의 불확실성을 내포하게 되는데, 각각의 요청들과 상황들은, 미리 정해진 규칙에 의거해 처리하기에는 너무 다양하다. 그 어려움을 극복하는 방법으로, 어떤 상황에 적응하는 방법으로 기억을 사용 하는 인간과 마찬가지로, 로봇은 새로운 요청을 처리하기 위해 과거의 기록을 사용할 수 있다. 여러 가지 과거의 기록들을 잘 정리해서 분류하여 저장해둔 후, 현재의 요청에 대한 답으로, 가장 가능성 있는 과거의 기록을 찾아내는 것이다.

본 논문에서는 사용자와 로봇 사이에서 상호 작용에서 발생할 수 있는 불확실성을 과거기록의 탐색을 통해 해결하고자 하였다. 과거 기록은 시간, 장소, 대상 물건, 행동 유형으로 구분되어 저장하였으며, 각각의 유사 가능성(Possibility)들의 합을 기준으로, 전체 기록을 K-Means 알고리즘을 통하여 군집화하고 협력 필터링을 기반으로 현재의 요청이 담고 있는 불확실성에 대한 가능성 있는 값을 추천해 주었다. 제한된 공간과 제한된 자료의 수에 의한 실험 결과로서의 한계를 가지고 있지만, 실제 가정용 로봇에서의 적용 가능성을 보여주었다.

Keyword : Robot, K-Means Algorithm, Clustering, Collaborative Filtering

1. 서론

모든 곳에 존재하는 네트워크 환경을 의미하는 '유비쿼터스' 시대와 최신 기술로 구현되어 인간을 도와주는 '지능형 로봇'의 시대가 도래하고 있다. 기술의 흐름은, 이제 우리에게 공장과 공원 등의 공공 장소뿐 만이 아니라, 생활의 기본이 되는 가정 안에서의 로봇을 받아들일 준비를 요구하고 있다.

본 논문에서는 가정에서 인간의 심부름을 하는 간단한 로봇을 대상으로 하여, 그 로봇의 지능을

이루고 있는 부분 중 하나인 계획 수립기(planner)에서의 특정 불확실성(uncertainty)을 처리하는 부분을 다룬다. 계획 수립기란 간단히 말하여, 로봇의 행동 계획을 수립하는 역할을 하는 것으로, 주어진 제약조건하에서 주어진 목표까지의 경로에 있는 개개의 연산처리들을 하나의 순서로 만든 계획(plan)을 생성하고 선택하는 작업을 수행하게 된다. 이 작업 수행에 있어서 수 많은 어려움이 있지만, 그 중 하나는 사용자의 요청이 완벽한 형태

가 아닌 모호하고 불완전한 형태로 로봇에게 전해지고, 그 부족한 정보를 기반으로 목표를 수행해야 할 경우에 발생하게 된다. 이것을 불확실성이라고 하는데, 다시 한번 정의하자면 불확실성이란 판단이나 의사결정에 필요한 적절한 정보의 부족이라고 할 수 있다.

사용자와 로봇 사이에서 상호 작용에서 발생할 수 있는 불확실성을 과거기록의 탐색을 통해 해결하고자 한다. 과거 사용자의 요청과 로봇의 행동 결과를 데이터베이스로 저장하고, 불확실성을 담고 있는 요청이 들어왔을 경우, 가장 유사한 과거 기록을 탐색하여 사용자에게 추천해주는 것을 목표로 한다. 과거 기록은 크게 시간, 장소, 대상 물건, 로봇의 행동 유형으로 구분되어 저장되며, 각각의 기록들의 가능성들의 합을 기준으로 전체 과거 기록을 K-Means 알고리즘을 통하여 군집화한다. 그 후, 현재의 요청과 유사한 군집을 선택하고, 그 군집내의 기록들의 상호 협력 필터링(collaborative filtering)을 기반으로 가장 유사한 값을 사용자의 요청에서의 불확실성 값으로 추천한다.

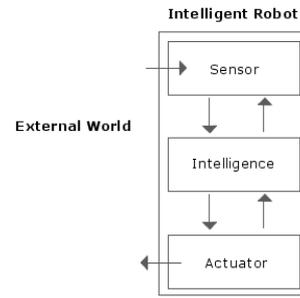
본 논문은 다음과 같이 구성되어 있다. 2 장에서는 지능형 로봇 연구와 K-Means 알고리즘, 협력 필터링의 기본적인 정의와 특성에 대하여 살펴본다. 3 장에서는 위의 방법들이 사용자의 요청에서의 불확실성 값의 가능성을 추론하는 것에 적용된 방법에 대하여 설명하며, 4 장에서는 구성된 데이터를 이용한 실험을 통하여 결과를 확인한다. 마지막으로 5 장에서는 결론 및 향후 연구과제에 대하여 논의하였다.

2. 연구 배경

2.1 지능형 로봇의 프레임워크

지능형 로봇을 이루는 구성요소는 크게 세가지로 이루어져 있는데, 센서(sensor), 지능(intelligence), 작동기(actuator)가 그것이다. 센서는 외부의 환경으로부터 필요한 정보를 가져오는 역할을 하며, 지능은 정보를 가지고 목표에 필요한 학습, 추론, 인지 연산을 하는 역할을 한다. 그리고, 작동기는 외부환경으로의 직접적인 로봇의 행

동을 제어하는 역할을 한다.



<그림 2-1> 지능형 로봇의 프레임워크

이중 '지능'의 역할을 수행하고 있는 중요한 부분 중 하나인 계획 수립기(planner)는 로봇의 행동 계획을 수립하는 부분이다. 계획 수립기는 주어진 제약조건하에서 주어진 목표까지의 경로에 있는 개개의 연산처리들을 하나의 순서로 만든 계획을 생성하고 선택하는 작업을 수행하게 된다.

2.2 K-평균 알고리즘 (K-Means Algorithm)

실제로 가장 보편적으로 이용되는 군집화 방법 중 하나인 K-평균 알고리즘은 각 객체를 가장 가까운 중심점에 할당하는 방법으로, 아래와 같은 절차에 따라서 군집을 형성한다.

각 개체들은 다음과 같이 벡터로 표현된다.

$$x = [x_1, \dots, x_n]^T$$

각 개체(벡터)의 Euclidean Norm 은 다음과 같이 정의된다.

$$\|x\| = \left[\sum_{i=1}^n x_i^2 \right]^{1/2}$$

같은 차원의 두 벡터 x 와 z 의 차를 나타내는 식은 다음과 같다.

$$\|x-z\| = \left[\sum_{i=1}^n (x_i - z_i)^2 \right]^{1/2}$$

1) 초기화 단계

생성할 군집의 개수 K 를 정하고, K 개의 각 군집에 대하여 군집의 초기 중심점을 선택한다.

$$\{z_1(l), z_2(l), \dots, z_k(l)\}$$

2) 객체 분산 단계

각 개체를 가장 가까운 중심점을 갖는 군집에 할당한다.

$$x^{(p)} \in S_j^{(l)} \quad \text{if} \quad \|x^{(p)} - z_j(l)\| < \|x^{(p)} - z_i(l)\|$$

for all $i = 1, 2, \dots, K, \quad i \neq j$

3) 새로운 군집의 중심 계산 단계

객체 분산 단계의 결과로 각 객체들이 군집에 할당이 되면, 이제 각 군집의 중심을 새로 계산한다. 군집의 중심은 군집의 중심에서 그 군집에 속한 객체들까지의 거리 합이 최소가 되도록 정해야 한다. 즉, 아래의 식을 최소화 시키는 군집의 중심을 찾는다.

$$J_j = \sum_{x^{(p)} \in S_j^{(l)}} \|x^{(p)} - z_j(l+1)\|^2$$

$j = 1, 2, \dots, K$

J_j 를 최소화시키는 $z_j(l+1)$ 는 간단히 클러스터 j 에 속하는 모든 개체들의 평균값을 취함으로써 얻을 수 있다. 즉,

$$z_j(l+1)_j = \frac{1}{N_j} \sum_{x^{(p)} \in S_j^{(l)}} x^{(p)}$$

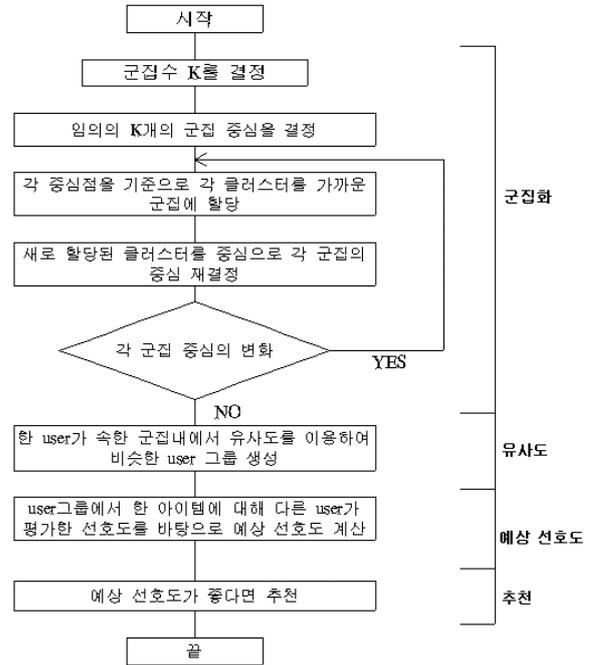
4) 수렴 여부의 확인

K-Means 알고리즘은 더 이상 각 군집의 중심에 변화가 생기지 않을 때까지 위의 2), 3) 단계를 반복하여 최종적으로 K 개의 군집을 형성한다.

2.3 협력 필터링(collaborative filtering)

협력 필터링(collaborative filtering)은 추천시스템에서 채택하고 있는 대표적 추천 방법이다.[1] 협력 필터링은 과거에 아이템을 선택하는데 있어 유사한 성향을 보였던 사용자들은 새로운 아이템에 대해서도 유사한 성향을 보일 것이라는 가정을 바탕으로 추천을 수행하는 시스템이다. 사용자의 로그 등의 모아진 정보를 사용하여 사용자를 비슷한 선호도를 가진 집단으로 나누어 그 집단 내에서 서로에게 추천하는 방식을 사용한다. 이러한 그룹 형성의 과정과 교차추천(cross-recommendation)의

과정은 자동으로 이루어지게 된다.[2]



<그림 2-2> 협력 필터링의 과정

협력 필터링의 장점으로는 내용분석의 용이성에 관계없이 어떠한 아이템에 대해서나 분석이 가능하다는 점이 있다. 사용자가 처음 사용하는 경우라 하더라도 충분한 자료가 축적되어 있을 경우 사용자에게 만족할 만한 서비스가 가능하다. 이런 특성으로 인하여, 특히 웹 기반의 사용자 추천 시스템으로서 많은 연구가 이루어졌으며, 사용자들의 즉각적인 성향을 만족시켜줄 수 있는 방법으로 적용되고 있다.[3] 또한 다른 특징으로는, 본질적으로 사용자의 개인정보를 공개하지 않아도 서비스 제공이 가능하기 때문에 최근의 개인정보 보호를 우선시하는 익명 개인화(anonymous personalization) 추세에도 잘 맞는 방법이다. 그리고 모바일 환경에서는 휴대 단말기의 인터페이스(interface)상의 제약으로 말미암아 사용자들이 직접 정보를 검색하는 것이 더욱 어려워지게 될 것으로 예상되는 데, 이러한 제한된 환경하에서 협력 필터링의 중요성은 더욱 증대될 것이다.

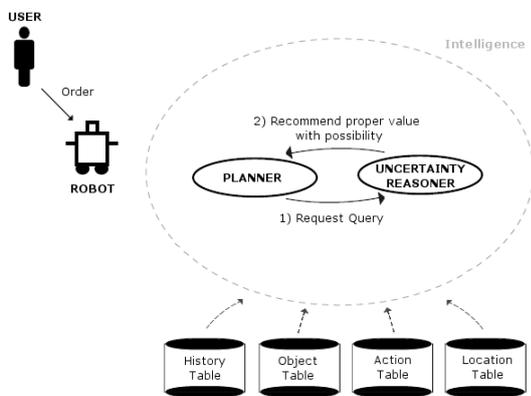
그러나, 협력 필터링은 새로운 사용자 또는 아이템이 끊임없이 등장하는 상황에서 이를 즉각적으로 반영하기가 어렵다는 문제점을 지니고 있다.

이는 새로운 사용자가 등장하게 되면, 보관하고 있던 기존 사용자의 데이터와 새로운 사용자의 데이터를 모두 이용하여 예측을 위한 계산을 실행하고, 그 결과를 바탕으로 추천을 수행하는 데에서 비롯된다. 즉, 끊임없이 새로운 사용자가 등장하는 상황에서, 기존 사용자 관련 데이터를 저장하는 데 드는 비용뿐 아니라 계산을 수행하는 데 걸리는 시간도 급격히 증가하게 된다.[4] 아울러, 사용자의 데이터는 많은 결측치를 포함하고 있다. 협력 필터링에서는 이와 같은 결측치를 처리하기 위해 주로 상관계수와 같은 통계기법을 사용하고 있다.

3. 협력 필터링을 이용한 불확실성 값의 추천

3.1. 추천 시스템의 구조

지능형 로봇은 사용자의 요청을 받아 해당 목적을 달성하는 것을 목표로 한다. 지능형 로봇은 크게 동작기(actuator), 센서(sensor), 그리고 지능(intelligence)으로 구성되어 있으며, 지능 부분에서의 여러 가지 역할 중 하나는 해당 목적을 달성하기 위해 순차적인 세부 행동 목록을 만드는 것인데, 이를 담당하는 부분을 계획 수립기(planner)라고 부른다. 계획 수립기는 종종 판단이나 의사결정에 필요한 적절한 정보가 부족한 상황을 겪게 되는데, 이를 보조해 주는 것이 불확실성 추론기(uncertainty reasoner)의 역할이다.



<그림 3-1> 추천 시스템의 구조

본 논문에서 불확실성 값의 추천 시스템으로 제안

된 불확실성 추론기는 데이터베이스에 저장되어 있는 과거기록들과, 현재 도메인내의 장소 정보, 로봇 행동 정보, 대상 물건 정보, 시간 정보를 이용하여 과거 기록을 군집화한다. 사용된 군집화 알고리즘은 비 계층적인(non-hierarchical) 구조를 가지고, 한 객체가 단 하나의 군집에 소속되는 특성을 갖는 K-Means 알고리즘이다. 그 결과를 기반으로, 협력 필터링을 사용하여 새롭게 요청된 질의와의 유사 가능성 관계를 계산하여 가장 가능성 있는 값을 추천하여 준다.

3.2. 추천 시스템에 사용된 과거 기록

과거 기록은 크게 로봇의 행동, 대상 물건, 장소, 시간으로 구분되어 저장되어 있다. 전체 과거 기록은 각각의 목록을 참조하여 구분 가능한 식별 번호로 저장된다. 각각의 데이터를 최소한의 길이로 표현하는 것은 저장공간의 효율성을 가져올 수 있다.

- 로봇의 행동 : 사용자가 로봇에게 주문할 수 있는 전체 로봇의 행동을 카테고리에 의해 구분하고, 식별 번호를 부여하여 사용한다. 로봇의 기능은 하나의 행동 또는 몇 가지의 로봇 행동의 조합으로 이루어진다고 가정하고, 계획 수립기는 이 행동의 목록을 이용하여 계획을 생성한다고 가정한다.

- 대상 물건 : 사용자와 로봇이 있는 공간에서 로봇의 행동의 대상이 되는 물건들을 모두 등록하여 식별 번호를 부여하여 사용한다.

- 장소 : 사용자와 로봇이 있는 공간을 몇 개의 의미 있는 구간으로 나누어 식별 번호를 부여하여 사용한다. 사용자가 직접, 로봇에게 장소의 중심 위치의 X-Y 좌표와 원으로 설정된 범위의 반지름을 설정한다.

- 시간 : 사용자에게 의미 있는 시간 구분인, 년, 월, 일, 시, 분을 사용한다. 매년, 매월, 매일 등의 비슷한 시간대의 행동 패턴을 구분하는 기준으로

사용한다.

3.3 각 속성의 유사 가능성 측정

두 시점의 과거 기록을 비교하기 위해 위의 각 속성간의 유사 가능성을 계산하고, 그 수치들간의 평균을 구하여 최종 유사 가능성으로 사용한다. 앞서 배경연구에서 보았듯이 K-Means 알고리즘을 이용한 군집화의 결과는 두 데이터 사이의 거리 또는 유사성을 어떻게 정의하는가의 의해 크게 좌우된다. 본 논문에서는 특정시간에서의 한정된 공간내의 한정된 행동에 관한 유사도를 측정하여야 함으로, 기존의 거리 측정 방법은 적용할 수 없었으며, 각각의 속성에 맞는 휴리스틱한 방법으로 계산하였다. 전체 유사 가능성은 각 속성들의 유사 가능성의 평균으로 계산된다.

$$\text{Possibility}_{\text{average}} = \frac{\text{Possibility}_{\text{action}} + \text{Possibility}_{\text{object}} + \text{Possibility}_{\text{location}} + \text{Possibility}_{\text{time}}}{4}$$

- Possibility_{action}

로봇의 행동의 유사가능성은 과거 기록에서의 대상물건, 장소의 분포를 고려하여 계산하였다. 이는 같은 대상물건을 사용하는 빈도가 높고, 사용자가 로봇에게 요청한 장소와 탐색할 것을 요청한 장소가 같다면, 두 가지의 행동이 서로 유사하다는 가정으로부터 시작하였다. 예를 들어, A 라는 행동이 대상물건으로 신문을 요청하는 행동이 9 번이고, B 라는 행동이 대상물건으로 신문을 요청하는 행동이 8 번 일 때, C 이라는 행동이 대상 물건으로 신문을 요청하는 행동이 3 번이라면, A 와 B 행동의 유사가능성이 A 와 C, 그리고 B 와 C 의 유사가능성 보다 높다고 판단할 수 있다. 이를 활용하기 위해 모든 행동을 각각 서로 비교하여 행동-행동 유사가능성 테이블을 계산하였다. 각각의 행동에 대하여 과거기록을 모두 탐색하여, 요청되었던 대상 물건과, 장소에 대하여 분포를 비교하여 테이블을 만들어 두고, 과거 기록간의 유사 가능성 계산에 활용하였다. 행동-행동 유사 가능성 테이블의 값들은, 0 부터 1 사이의 값을 가지도록 하였다.

- Possibility_{object}

대상물건의 유사 가능성의 경우, 요청 되었던 행동과, 장소를 고려하여 계산되었다. 동일한 행동이 많고, 동일한 장소가 많다면, 두 대상물건은 서로 유사하다는 가정으로 시작하였다. 예를 들어, 신문과 책, 그리고 지팡이 사이에는 요청되어지는 행동의 분포가 다르며, 또한 장소의 분포도 다르게 된다. 신문과 책을 요청하는 행동은 주로, ‘돋보기 기능’이나, ‘문자를 읽어주기’와 같은 행동이 요청되게 되며, 거실과 안방에서 요청되는 경우가 잦을 것이다. 반면에, 지팡이의 경우 ‘물건의 위치 알려주기’, 혹은 ‘물건을 가져오기’의 행동의 분포가 상대적으로 많게 되며, 장소로서는 현관이 높은 분포를 가지게 될 것이다. 이런 가정으로부터 출발한, 대상물건-대상물건 유사 가능성은, 역시 과거 기록을 모두 탐색하여, 요청되었던 행동과, 장소의 분포를 비교하여 테이블을 만들고, 과거 기록간의 유사 가능성 계산에 활용하였다.

- Possibility_{location}

장소의 유사가능성의 경우는, 제한된 도메인에서 그 구분 개수가 많지 않아, 큰 차이를 보이지 않으나, 위와 동일한 가정에서, 행동과 대상 물건의 분포에 기반하여 계산하였다. 예를 들어 안방과 작은방의 경우 요청된 행동과, 대상 물건의 분포가, 안방과 주방의 경우의 요청된 행동과 대상 물건의 분포와 비교하여 더욱 유사할 가능성이 크다.

- Possibility_{time}

시간의 경우 각각 년, 월, 일, 시간, 분 값으로 나누어, 두 시점의 시간 데이터들의 년, 월, 일, 시, 분간의 각각의 차이를 계산하고 그것의 평균을 사용하였다.

3.4 협력 필터링 적용

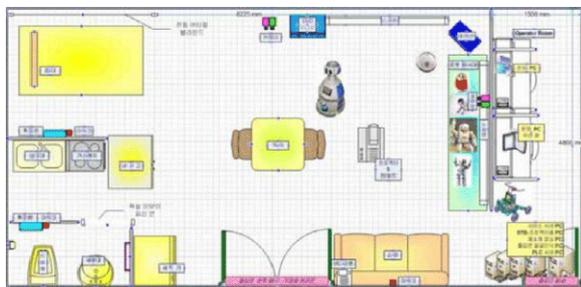
협력 필터링을 위하여, 우선 과거 기록을 K-Means 알고리즘을 이용하여 군집화를 수행한다. 군집화가 수행된 후에는, 서로 유사한 과거 기록

들로 하나의 군집을 형성하게 된다. 군집의 의미는 알지 못하더라도, 과거기록의 유사가능성 계산이 효율적이라면, 서로 구분 지어진 군집들로 전체 과거기록은 군집화가 된다. 그렇게 군집화가 끝난 후, 한 질의가 전달되었을 때, 우선 그 질의와 가장 유사한 군집을 찾아내게 된다. 가장 유사한 군집을 찾아내는 작업은 각각의 군집의 중심값들과 비교하는 것이다. 이 방법은 전체 과거기록과 일일이 비교하지 않아도 되므로, 속도 면에서 많은 절약을 하게 된다. 가장 유사한 군집을 찾아 낸 후에는, 질의에서 요청된 필드를 기준으로 군집내의 각각의 기록들을 탐색하게 된다. 전체 기록 중에 추천 값의 대상이 되는 값들의 목록을 뽑아낸 후에, 각각의 값들의 유사 가능성 평균과 분산을 구한다. 그리고 평균에 비례하고, 분산에 반비례하는 추천 값이 가장 큰 값을 선택하여, 질의어의 필드로 추천 해 주게 된다. 단, 추천 값이 임계최소치(threshold)를 넘지 않는다면, 추천해 주지 않는다.

4. 구현 및 실험

4.1. 실험 환경

사용자와 로봇이 함께 거주하는 공간을 <그림 4-1>과 같은 가상공간으로 임의로 설정하여, 실험에 사용하였다. <그림 4-1>에서와 같이 거실, 침실, 주방, 욕실 등을 가진 공간으로서, 가구 및 생활용품 등의 물건들을 임의로 배치하였다



<그림 4-1> 가상 실험 공간

4.2. 실험 데이터

본 실험에 사용된 데이터는 의미에 따라, 로봇의 행동 목록, 대상 물건의 목록, 장소 목록, 과거 기록의 목록으로 구분되어 저장되어 사용 되었다.

[표 4-1] 전체 데이터 수

구분	데이터 수
로봇의 행동 목록	32
대상 물건 목록	30
장소 목록	7
과거 기록 목록	100

1) 로봇의 행동 목록 (action table)

사용자가 로봇에게 주문할 수 있는 행동들의 전체 목록을 등록하여 사용하였다. 실험에 사용된 로봇의 행동의 예는 [표 4-2] 과 같다.

[표 4-2] 로봇의 행동 목록

Act_ID	Category	Act_name
1	신체보조	물건 옮기기
2		책을 읽어주는 기능
3		마사지 기능

2) 대상 물건 목록 (object table)

도메인 내의 대상 물건들의 전체 목록을 등록하여 사용 하였다. 실험에 사용된 대상 물건의 예는 [표 4-3]와 같다.

[표 4-3] 대상 물건 목록

Obj_ID	Obj_name	Info
1	소파	3 인용
2	소파	1 인용
3	탁자	

3) 장소 목록 (location table)

사용자가 로봇에게 주문을 내릴 때의 사용자와 로봇의 장소, 사용자가 로봇에게 주문을 내린 장소, 로봇의 실제 행동이 발생한 장소들을 등록하여 사용하였다. 실험에 사용된 장소 목록의 예는 [표 4-4]와 같다.

[표 4-4] 장소 목록

Loc_ID	Loc_name	Pos_X	Pos_Y	Rad
1	거실	655	327	100
2	주방	105	284	100
3	식당	455	232	150

4) 과거 기록 목록 (history table)

사용자가 로봇에게 요청할 수 있는 주문 시나리오의 여러 가지가 존재한다. 제한된 수용 가능한 요청과 위에서 임의로 설정한 장소, 물건 들을 조합하여 시나리오를 만들고, 그 시나리오를 위의 세가지 목록을 이용하여 기록하였다. 아래 [표 4-5]는 과거 기록 목록의 추가적인 몇 가지 예를 보여준다.

[표 4-5] 과거 기록 목록

Order No.	Time	Location				Action	Object
		user	agent	object	act		
1	2005/11/25 11:59:00	식당	식당	식당	식당	밀다	의자
2	2005/11/25 15:26:13	거실	거실	작은 방	작은 방	화상 채팅 연결	컴퓨터

예를 들어 사용자가 “냉장고에 가서 음료수를 가지고 오너라”라는 요청을 내렸을 경우, 이는 식별번호인 Order No 와, 11 월 10 일 오후 9:10 라는 시각, Location user 값으로 거실을, Location agent 값으로 거실을, Location object 값으로 냉장고를, Location act 값으로 냉장고를, Action 값으로 ‘가져오다’ 값을, Object 값으로 음료수를 각각 갖는다.

4.3 시나리오

4.3.1 시나리오 1: 대상 물건의 위치 찾기

사용자는 안방의 침대에서 잠이 들기 전, 독서를 위해 안경을 찾는다. 저녁시간에 사용자가 안방에서 안경을 찾았을 경우, 주로 안경은 TV 시청을 마친 후, 거실에 놓아진 경우가 많았다. 로봇은 과거 기록들의 탐색을 통해 안경을 찾을 공간으로 거실을 추천한다.

4.3.2 시나리오 2: 대상 물건 찾기

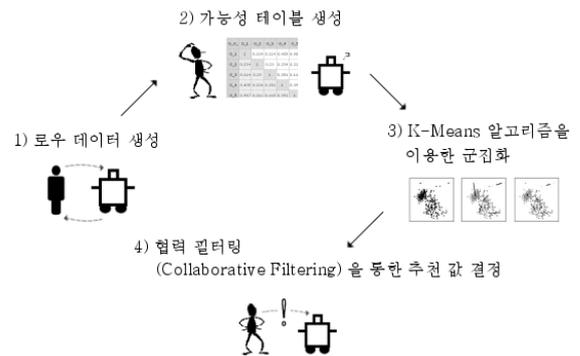
사용자는 잠에서 깨어, 안방에서 로봇을 부른다. 아침마다 신문을 보는 사용자는 로봇에게 이전과 같은 물건을 요청한다. 로봇은 과거 기록들의 탐색을 통해 대상 물건으로 신문을 추천한다.

4.3.3 시나리오 3: 로봇의 행동 문의

식사를 마친 후 30 분쯤이 지난 시각에, 사용자는 자신이 잊고 있는 계획이 없는지 로봇에게 문의한다. 로봇은 당뇨병 환자인 사용자가, 매일 식사 후 규칙적으로 건강 체크를 하는 것을 과거 기록으로부터 탐색하여 추천해준다.

4.4 실험

4.4.1 실험 과정



<그림 4-2> 실험 순서

총 100 개의 과거 기록 목록을 이용하여, 앞서 언급한 시나리오들에 대한 실험을 각각 수행하였다. 우선, 총 3 일간의 가상 공간 내에서의 사용자와 로봇간의 있음직한 상호 작용을 각각 과거 기록 목록에 직접 입력 하였으며, 이를 기반으로 <그림 4-3>과 같은 전체 과거 기록들간의 가능성 수치를 100x100 행렬로 계산하여 저장하였다. 그 후, 가능성 수치 행렬을 바탕으로 K-Means 알고리즘을 이용하여 각각의 과거기록들의 총 군집의 수는 10 개로 설정한 후, 군집화를 수행하였다. 마지막으로, 각각의 시나리오에 맞는 질의를 보내어 협력 필터링에 의한 추천 값을 확인하였다.

	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	O_9	O_10	O_11		O_100
O_1	1	0.224	0.114	0.405	0.987	0.227	0.119	0.401	0.218	0.219	0.22		0.104
O_2	0.224	1	0.23	0.234	0.211	0.992	0.232	0.237	0.625	0.626	0.627		0.226
O_3	0.114	0.23	1	0.251	0.118	0.237	0.994	0.247	0.228	0.23	0.231		0.258
O_4	0.405	0.234	0.251	1	0.392	0.234	0.253	0.995	0.194	0.196	0.197		0.224
O_5	0.987	0.211	0.118	0.392	1	0.218	0.122	0.388	0.23	0.232	0.233		0.108
O_6	0.227	0.992	0.237	0.234	0.218	1	0.239	0.23	0.633	0.634	0.635	...	0.233
O_7	0.119	0.232	0.994	0.253	0.122	0.239	1	0.249	0.226	0.227	0.229		0.256
O_8	0.401	0.237	0.247	0.995	0.388	0.23	0.249	1	0.19	0.191	0.193		0.22
O_9	0.218	0.625	0.228	0.194	0.23	0.633	0.226	0.19	1	0.998	0.997		0.256
O_10	0.219	0.626	0.23	0.196	0.232	0.634	0.227	0.191	0.998	1	0.998		0.257
O_11	0.22	0.627	0.231	0.197	0.233	0.635	0.229	0.193	0.997	0.998	1		0.258
						:						1	0.259
O_100	0.104	0.226	0.258	0.224	0.108	0.233	0.256	0.22	0.256	0.257	0.258	0.259	1

<그림 4-3> 과거 기록들 간의 가능성 수치 행렬

4.4.2 실험 1: 군집화 평가

100 개의 실험 데이터들을 총 10 개의 군집으로 각각 군집화 한다. 군집화의 평가 척도로는 한 객체와 가장 유사 가능성이 높은 객체를 판별한 후, 두 객체가 동일한 군집에 포함되어 있는지 여부로 군집화 평가를 수행하였다.

[표 4-6] 군집화의 정확성

	전체	0.5 이상의 유사가능성 대상을 가진 데이터
데이터의 수	100	66
정확도	68%	84.8%

[표 4-6]에서 보는 바와 같이 전체 100 개의 데이터 중, 의미 있는 유사 가능성(0.5 이상) 대상을 가진 자료 66 개중, 가장 높은 유사 가능성을 동일 군집 내에 함께 포함하고 있는 객체가 총 56 개로 조사되어 84.8%의 정확도를 보였다. 이는 앞서 제시한 유사 가능성 판단의 기준이 주어진 데이터를 군집화 하는데, 84.8%의 정확도를 보였다고 생각할 수 있으며, 반대로 나머지 15.2%의 경우에 대해서는 적절한 군집으로 분류되지 못하였다고 생각할 수 있다. 제시된 시스템에서는 어떻게 군집이 이루어 지느냐가, 곧 불확실성 값의 추천 값이 어떻게 결정되느냐를 결정하므로, 군집화의 정확도는 곧 전체 시스템의 성능을 나타낸다. 결국, 주어진 데이터의 환경에서는 84.8%의 경우, 시스템

이 올바른 추천 값을 추출해준다고 할 수 있다.

[표 4-7] 데이터 수에 따른 군집화 처리 시간 비용

데이터의 수	처리 시간 (초)	중심 재계산 반복 횟수
25	2.47	4
50	16.45	7
75	73.71	8
100	109.74	12

[표 4-7]에서 보는 바와 같이 군집화 수행 시간과 중심 재계산 반복 횟수(iteration)은 데이터의 수에 비례하여 증가하는 것을 알 수 있다. 결국 실시간으로 반응해야 하는 불확실성 값 추천은, 군집화가 수행된 시점 이전의 기록들을 이용해야 할 것이며, 새로운 자료의 증가로 인한 새로운 군집화 수행은 로봇의 유휴 시간(idle time)에 처리하여야 할 것이다. 다시 말해, 로봇의 실시간 학습은 현재 시스템으로는 불가능하며, 가장 최근에 군집화가 된 시점부터 새롭게 군집화가 수행될 때까지의 기록에 대해서는 고려를 하지 못하게 된다는 단점이 있다.

4.4.3 실험 2: 시나리오의 추천 값 결과

시나리오 1에서의 질의와 동일한 군집내의 과거기록들의 유사 가능성은 <그림 4-4>와 같다. 명령(order) 11은 시나리오를 수행하기 위한 질의이며, 우선 가장 근접한 군집을 결정하게 된다. 해당 군집에는 가장 높은 유사가능성을 보인 명령 97을 포함하여 총 4개의 과거 기록들이 있으며, 이들은 질의와 동일한 장소에서, 동일한 로봇 행동을 수행하도록 요청되었으며 시간 차이가 근소한 사건의 의미를 가진다. 해당 군집을 찾은 다음 과정으로, 추천 값을 찾기 위해 시나리오 1의 질의에서 요구한, 대상 물건의 장소의 가능성을 계산한다. 대상 물건의 장소에 해당하는 주방과 식당을 대상으로 추천 값(평균에는 비례, 분산에 반비례)을 계산한 결과, 가장 높은 값을 가진 주방이 추천 값으로 결정된다.

Order	Possibility	Loc_Object	Object
11	0.834		보성녹차
97	0.843	주방	보성녹차
12	0.841	주방	보성녹차
64	0.696	식당	복숭아농장
23	0.69	식당	복숭아농장

<그림 4-4> 시나리오 1: 질의와 동일 군집 내의 사건들의 유사 가능성

시나리오 2에서는 대상 물건을 불확실성 값으로 가지는 질의를 요청하였다. 근소한 시간 차이와, 동일한 장소, 동일한 행동을 가진 과거의 기록으로 추천 값을 계산한 결과, 가장 높은 값을 가진 '신문'이 추천 값으로 결정된다.

시나리오 3에서는 로봇의 행동을 불확실성 값으로 가지는 질의를 요청하였다. 동일한 장소, 대상물건, 근소한 시간 차이를 가진 과거의 기록으로 추천 값을 계산한 결과, 가장 높은 값을 가진 '건강체크'가 추천 값으로 결정된다.

위의 결과는 가정용 로봇에서 발생하는 요청의 불확실성을 과거 기록을 통해 탐색하는 아이디어 적용의 가능성을 보이고 있다. 실험 결과 100개의 데이터 내에서, 위의 각각의 시나리오들을 성공적으로 수행하였다. 그러나, 제한된 공간에서의 제한된 자료의 수에 의한 실험 결과로서의 한계를 가지고 있다. 직접 입력한 데이터들로 K-Means 알고리즘을 이용한 군집화와 협력 필터링을 적용한 추천 값 계산을 수행하였기에, 실제 데이터가 적용되었을 경우 오차범위가 커질 수 있다.

5. 결론

본 논문에서는 사용자와 로봇 사이에서 상호 작용에서 발생할 수 있는 불확실성을 과거기록의 탐색을 통해 해결하고자 하였다. 과거 기록은 시간, 장소, 대상 물건, 행동유형으로 구분되어 저장하였으며, 각각의 필드들의 가능성들의 합을 기준으로, 전체 기록을 K-Means 알고리즘을 통하여 군집화하고 협력 필터링을 기반으로 현재의 요청이 담고 있는 불확실성에 대한 가능성 있는 값을 추천해 주도록 하였다. 그리고 실험을 통하여 주어진 시나리오를 올바르게 수행하는 것을 확인하다.

향후 연구 과제로는 과거의 기록을 위하여 더

많은 속성들을 기록하고, 각각의 속성들의 효과적인 가중치(weight)의 조정을 학습할 수 있도록 하는 것이 있다. 또한, 가상으로 입력된 데이터들이 아닌 실제 데이터들을 이용하는 시스템의 구현과 실험이 필요하며, 더 다양하고 복잡한 시나리오를 만족할 수 있는 풍부한 데이터들을 이용하여 실험을 수행해야 하는 것이 필요하다. 그리고 데이터의 증가에 따른 계산의 복잡도 증가를 극복하기 위한 알고리즘의 개선이 필요하다.

감사의 글

이 연구(논문)는 산업자원부 지원으로 수행하는 21세기 프론티어 연구개발사업(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었습니다.

참고 문헌

- [1] P. Resnick, et al., "GroupLens : An Open Architecture for Collaborative Filtering of Netnews", In Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work, 1994
- [2] D. Fisher et al, "SWAMI: A Framework for Collaborative Filtering Algorithm Development and Evaluation.", SIGIR 2000, Athens, Greece, July 2000
- [3] Michelle Anderson et al, "RACOFI: A Rule-Appling Collaborative Filtering System", In Proceedings of IEEE/WIC COLA'03, Halifax, Canada, October 2003
- [4] J.S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for Collaborative Filtering.", Technical report, Microsoft Research, 1998
- [5] Salton and McGill, "Introduction to Modern Information Retrieval", mcgraw-Hill, New York, 1983.
- [6] Ayse Yasemin Seydim, "Intelligent Agent : A Data Mining Perspective", Department of Science and Engineering, Southern Methodist University, May 1999.
- [7] 박현근, "지능 로봇의 구성요소", MicroSoftware magazine, 2003.9.