

# 데이터 그룹화를 이용한 상호진화연산 기반의 추천 시스템

## A Recommendation System Based-on Interactive Evolutionary Computation with Data Grouping

김 현 태, 안 창 옥, 안 진 웅\*  
(Hyun-Tae Kim<sup>1</sup>, Chang Wook Ahn<sup>1</sup>, and Jinung An<sup>2</sup>)

<sup>1</sup>Sungkyunkwan University

<sup>2</sup>DGIST (Daegu Gyeongbuk Institute of Science & Technology)

**Abstract:** Recently, recommender systems have been widely applied in E-commerce websites to help their customers find the items what they want. A recommender system should be able to provide users with useful information regarding their interests. The ability to immediately respond to the changes in user's preference is a valuable asset of recommender systems. This paper proposes a novel recommender system which aims to effectively adapt and respond to the immediate changes in user's preference. The proposed system combines IEC (Interactive Evolutionary Computation) with a content-based filtering method and also employs data grouping in order to improve time efficiency. Experiments show that the proposed system makes acceptable recommendations while ensuring quality and speed. From a comparative experimental study with an existing recommender system which uses the content-based filtering, it is revealed that the proposed system produces more reliable recommendations and adaptively responds to the changes in any given condition. It denotes that the proposed approach can be an alternative to resolve limitations (e.g., over-specialization and sparse problems) of the existing methods.

**Keywords:** recommender system, interactive evolutionary computation, content-based filtering, data grouping

### I. 서론

의사결정(decision making)은 인간의 일상생활에서 중요한 부분을 차지한다[1]. 대부분의 사람들은 의사결정을 내리는데 있어서 부족한 지식과 경험으로 인해 만족할 만한 결론 도출에 실패하는 경우를 자주 경험한다. 또한 선택 가능한 대안(alternatives)이 너무 많아 의사결정을 내리는데 어려움을 겪게 된다. 예를 들어 의류, 음반 및 영화를 선택하고자 할 때 대부분의 사람들(또는 사용자)은 해당 주제에 대해 충분한 경험과 지식을 보유한 친구나 주위 사람에게 문의하거나, 인터넷 또는 신문기사를 참고하여 도움이 되는 정보를 얻고자 한다. 이와 같은 일련의 과정을 통해 추출된 정보를 추천(recommendation)이라 칭하고[1], 이 때 얻어진 다양한 추천(정보)을 기반으로 사용자는 자신에게 유용한 의사결정을 할 수 있게 된다. 인터넷이 생활의 수단이 된지 이미 오래이고 수많은 인터넷 쇼핑 등과 같은 전자상거래가 보편화된 현대 사회에서 사용자의 취향(preference)에 부합하는 정보 및 대안을 제공하는 추천 시스템(recommender system)의 필요성이 대두되고 있다[1,2].

1990년 후반 World Wide Web (WWW)의 출현은 방대한 지식 정보사회의 발전을 가속화 시켜, 사람들로 하여금 인터넷을 통해 많은 정보와 자원을 쉽고 빠르게 접근하고 선택할 수 있게 되었다[2]. 이러한 정보와 자원의 다양성과 방대함은

오히려 인터넷 사용자들이 자신에게 맞는 최선의 정보를 빠른 시간 안에 취사선택하는 데 있어 매우 많은 수의 대안을 검토해야만 하는 정보 과부하 문제(information overload problem)를 초래하고 있다[2,3]. 이러한 문제를 해결하고자 많은 전자 상거래 회사 (Amazon.com, eBay 및 CDNow 등)들은 고객들이 쉽게 정보를 접근해서 그들이 요구하는 항목들을 제안할 수 있는 추천 시스템을 활용해왔다[3,4].

추천 시스템을 구성하는 기본요소 중의 하나인 추천 기법에 대표적으로 내용-기반 필터링(content-based filtering)과 협업 필터링(collaborative filtering)이 널리 사용되고 있다[5-8]. 내용 기반 필터링은 사용자로부터 입력 받은 프로파일(profile)과 시스템 내 존재하는 비슷한 속성을 갖는 항목들을 찾아 추천하는 기법이다[5,6]. 이 기법에서 프로파일은 시스템 내 존재하는 항목에 대한 사용자의 취향을 나타내는 일종의 평가 점수를 의미한다. 반면, 협업 필터링은 앞서 언급한 사용자의 프로파일을 수집하여 데이터베이스를 구축하고 새로운 추천 항목을 구성하고자 할 때 과거에 구축된 데이터베이스에서 비교대상인 사용자의 프로파일과 부합하는 다른 사용자들의 프로파일을 참조하여 항목을 구성한다[7,8].

이외에도 두 가지 기법을 결합하거나 효율적인 다른 기법을 적용하여 기존 기법들의 한계점을 극복하기 위한 연구가 진행되고 있다[6-8]. 이러한 맥락에서 현재의 추천 시스템은 앞으로 더욱 발전할 수 있는 가능성을 지니고 있다.

본 논문에서는 내용-기반 필터링에 상호진화연산을 결합하여 사용자의 취향을 효과적으로 반영할 수 있는 새로운 추천 시스템을 설계하며, 양질의 추천 기능을 구현하기 위해 데이터 그룹화를 추가적으로 적용한다. 제안 시스템의 성능을 평가하기 위해 다수의 곡(track)으로 구성된 음악 데이터 집합

\* 책임저자(Corresponding Author)

논문접수: 2011. 5. 20., 수정: 2011. 6. 5., 채택확정: 2011. 6. 20.

김현태, 안창옥: 성균관대학교 정보통신공학부

(arkii@skku.edu/cwan@skku.edu)

안진웅: 대구경북과학기술원(robot@dgist.ac.kr)

※ 본 연구는 교육과학기술부의 대구경북과학기술원 일반사업에 의해 수행되었음(11-BD-01).

을 실험에 사용한다. 또한 테스트 에이전트 기반의 객관적인 실험을 통해 제안 시스템이 사용자의 취향에 적합한 유용한 정보(음악 데이터)를 제공함을 검증한다.

**II. 관련 연구**

**1. 추천 시스템(recommender system)**

추천 시스템은 사용자의 요구를 충족시켜 줄 수 있는 정보를 제공하는 것을 목적으로 한다. 이러한 관점에서 추천 시스템과 정보검색시스템(information retrieval system) 사이의 차이점을 구분하는 것은 쉽지 않다. 그 이유는 정보검색시스템 또한 사용자의 요구에 부응하는 정보를 제공하는 것을 목적으로 설계되기 때문이다.

예를 들어 정보검색시스템은 사용자가 입력하는 질문(또는 질의어)에 정확도가 높은 일련의 정보를 제공한다. 그렇기 때문에 사용자는 추천 시스템과 마찬가지로 자신이 정보검색시스템을 통해 얻은 결과가 유용한 정보이기를 기대한다. 이렇듯 유사한 속성을 지닌 두 시스템 사이에도 각 시스템을 구분 지을 수 있는 차이점을 각 시스템에서 처리하는 정보의 의미에서 찾을 수 있다. 즉, 정보검색시스템에서는 오직 사용자가 요구하는 검색 조건과 시스템 내에 존재하는 정보들의 유사도를 기준으로 처리하는 반면 추천 시스템은 각 사용자에 최적화, 개인화된 정보를 제공하며 그 정보가 사용자들에게 흥미롭고 유용한 정보임을 인식시키는 것에 의미를 둔다.

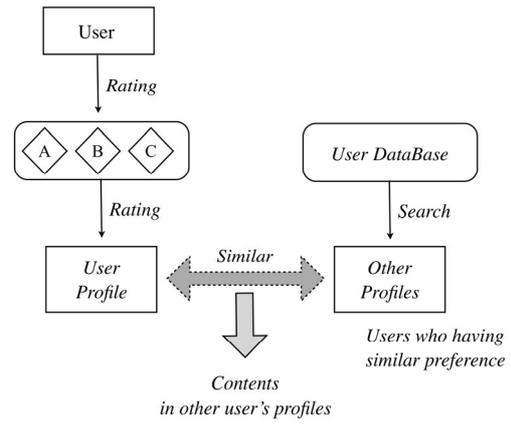
이러한 추천 시스템 분야에서의 주요 쟁점은 앞서 언급한 정보 과부화 문제(information overload problem) 해결이며, 이에 대한 해결책으로 수많은 자원 중에서 사용자의 취향에 맞는 정보를 제공하는 추천 시스템의 개발에 대한 요구가 증가하고 있다[5,6]. 이를 위해 추천 시스템 분야에서는 다양한 추천 기법들이 개발되었으며, 협업 필터링 기법과 내용-기반 필터링 기법이 널리 사용되고 있다[1,5,6].

**2. 협업 필터링과 내용-기반 필터링(collaborative filtering and content-based filtering)**

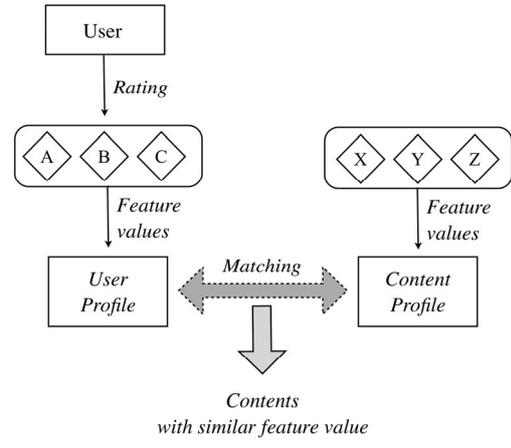
일반적으로 추천 시스템에서 항목의 유용성은 사용자의 평가를 기준으로 나타낸다. 즉, 추천 시스템은 사용자의 선호도에 초점을 맞추어 설계되는 시스템이기 때문에 추천항목을 생성할 때 사용자의 평가를 기준으로 삼는다. 이를 위해 추천 시스템은 각 사용자들의 평가를 수집하여 프로파일로 시스템 내에 저장한 후, 이를 활용하여 추천 항목을 생성한다. 그림 1은 기존의 추천 시스템에서 대표적으로 사용되는 두 가지 추천 기법들을 도시한다.

먼저 협업 필터링 기반 추천 시스템은 저장되어 있는 각 사용자의 프로파일 간 상관관계(correlation)를 정의하고 이를 통해 사용자들 사이의 프로파일 유사도를 기준으로 일정한 수의 그룹으로 분류하여 저장한다. 마지막으로 사용자가 이전에 평가 했던 항목과 동일한 평가 기록을 지니고 있는 그룹 사용자들의 프로파일을 참조하여 다른 사용자 프로파일 내에 있는 항목을 바탕으로 새로운 추천 항목을 구성한다[8-10].

반면 내용 기반 필터링을 이용한 추천 시스템은 사용자가 평가한 항목들로부터 성분을 추출하여 저장한 후 이를 기준으로 시스템 내에 남아 있는 모든 다른 항목들과 유사도를 측정하여 높은 유사도의 항목을 새로운 추천 항목으로 구성한다[7,8].



(a) Collaborative filtering.



(b) Content-based filtering.

그림 1. 기존 추천 기법들의 구성.  
Fig. 1. Description of existing recommendation methods.

하지만 기존 추천 기법은 다음과 같은 한계점들이 있다. 먼저, 협업 필터링 시스템에 새로운 항목이 새롭게 추가될 경우, 추가된 항목들은 아직 사용자에게 노출 되지 않은 상태이기 때문에 평가 값이 존재하지 않는다. 즉, 사용자의 평가 값을 바탕으로 추천 항목을 구성하는 협업 필터링 기법 기반 추천 시스템은 새로운 항목에 대한 접근이 불가능하므로 새롭게 추가된 항목을 추천 받을 수 없다. 이와 같이 새로운 일련의 데이터가 시스템이 추가되어 전체적으로 시스템 내 자원(데이터)이 분산 됨으로 인해 추천 시스템이 새로운 항목을 정상적으로 처리하지 못하는 것을 데이터 분산 문제(sparse problem)라 정의 된다.

그리고 내용-기반 필터링 적용 추천 시스템은 사용자의 프로파일을 분석하여 높은 평가 값을 갖는 항목과 유사한 항목만을 추천한다. 그러므로 사용자가 이전에 평가했던 항목들과 비슷한 항목들만 계속해서 추천하도록 하는 과도특정화 문제(overspecialization problem)를 유발한다. 또한 협업 필터링 문제점인 데이터 분산 문제 역시 내용 기반 필터링 기법에서도 발생한다. 즉, 새로운 항목의 속성을 추출하여 프로파일로 저장하지 않을 경우 시스템은 그 항목을 처리하지 못한다.

본 논문에서는 이러한 기존 기법들의 한계점들을 효율적으로 극복하기 위하여 내용 기반 필터링 기법에 상호진화연산 기법을 결합한 새로운 기법을 제안한다.

3. 성분 추출(feature extraction)

성분 추출은 특정 데이터(문서, 음악 혹은 사진)에서 고유한 속성을 추출하는 기술을 뜻하며, 유한 개의 자원 내에서 그룹화를 구현할 때 유용하게 쓰인다[11]. 내용-기반 필터링을 적용함에 있어 주어진 항목으로부터의 성분 추출 과정이 선행 되어야 함으로 제안 시스템에서는 성분 추출 도구 (CLAM [23])를 사용하였다. CLAM은 음성과 음악 분야에서 연구와 어플리케이션 개발을 위한 일종의 소프트웨어 프레임워크이며 기본적인 음악 성분의 분석은 물론 복잡한 음성 신호에 대한 분석과 변환 등을 제공한다.

4. 상호진화연산(IEC: Interactive Evolutionary Computation)

유전연산(evolutionary computation)은 자연세계 생물학의 진화 원리를 기반으로 하는 여러 컴퓨팅 기술을 대표하는 기술이다. 그 중에서도 유전 알고리즘(GAs: Genetic Algorithms)은 가장 널리 쓰이는 유전 연산을 대표하는 기술 중의 하나이다. 유전 알고리즘은 자연 선택(natural selection) 및 유전 원리(genetic inheritance)를 기반으로 하여 최적의 해를 찾는 확률적 탐색 기법(stochastic search)이다. 유전 알고리즘은 문제의 해결 가능한 해로 이루어진 개체군(population)을 바탕으로 구동된다. 개체군을 구성하는 각 해(개체)는 문제의 최적해에 근접한 정도를 나타내는 적합도(fitness value)를 갖고 있다. 그리고 다른 해 보다 높은 적합도를 나타내는 해법은 선택되며 다음 세대까지 살아남게 된다. 유전 알고리즘은 이러한 선택 해법들의 결합(combination) 과정을 통해 더 나은 새로운 해법을 생성하게 된다[12-15].

상호진화연산은 유전 연산과 같은 최적화 기법 중 하나이다. 하지만 상호진화연산에서는 사용자가 직접 각 개체(추천될 항목들)에 대한 적합도를 평가하게 된다. 이러한 점에서 기존 유전연산과는 다른 구조를 갖고 있으며 사용자는 항목들에 대한 평가를 통해 쉽게 자기 자신의 취향을 나타낼 수 있다[16,17]. 그림 2는 유전 연산과 상호진화연산의 전체 동작 과정을 비교하여 묘사하고 있다.

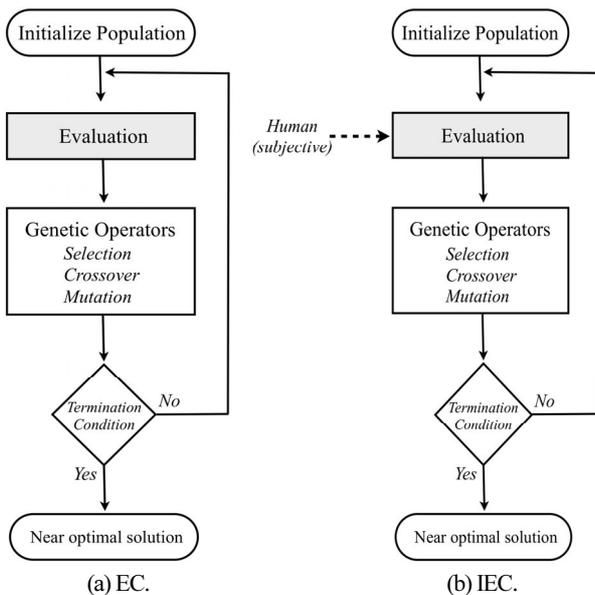


그림 2. 유전연산 과 상호진화연산의 구조.  
Fig. 2. Difference of structure between EC and IEC.

**k-means clustering Algorithm**

Perform clustering (or grouping)  $N$  data points into  $k$  disjoint subset.

1. Randomly choose  $k$  points from  $N$  data points (to divide  $N$  data into  $k$  groups)
  2. Set  $N$  points to the centroid of each group
  3. Apply below pseudo code to each data points of groups  
For no more changes in each group do
    - (1) Calculate distance between item( $I_i$ ) to each centroid ( $C_i$ )  
 $D_i = |I_i - C_i|, i \leq k$
    - (2) If the  $D_i$  is smaller than distance between  $I_i$  and  $C_i$   
Change group which has distance  $D_i$
    - (3) Else  
There is no change
- End do

그림 3. k-means 클러스터링 알고리즘.  
Fig. 3. Whole processes of k-means clustering algorithm.

본 논문에서는 사용자의 선호도를 빠르고 정확하게 인식하기 위해 제안 시스템에 상호진화연산을 사용 하였으며, 사용자와 추천 시스템 간의 지속적인 상호 작용(interaction)을 통해 사용자의 취향을 정확히 파악하고 빠르게 반응할 수 있도록 하였다.

5. 데이터 그룹화(data grouping)

제안 시스템이 질적으로 우수한 추천 서비스를 제공하고 동시에 추천 서비스에 대한 정확도 향상을 위하여 데이터 그룹화 기법을 적용하였다. 이러한 그룹화 작업은 항목 간 구조적 성분의 유사도를 바탕으로 추천 시스템 내에 있는 모든 항목들을 일정한 수의 그룹을 구성하게 된다.

제안 시스템은 그룹화 후 사용자의 평가를 바탕으로 추천 항목을 구성하며 이를 통하여 생성된 그룹화 정보를 활용한다. 그러므로 제안 시스템이 추천 항목을 구성하는데 소요되는 시간을 줄일 수 있다.

본 연구에서는 효율적으로 그룹화 작업을 위하여 k-means 클러스터링 알고리즘[18,19]을 사용하였다. 이 알고리즘은 널리 분포 되어 있는 데이터 집합을 주어진 크기( $k$ )로 그룹을 분할하는 기법이다. 그림 3은 k-means 클러스터링 알고리즘의 전체 동작 과정에 대한 의사 코드를 보여준다.

**III. 제안 추천 시스템**

본 논문에서 제안하는 추천 시스템은 사용자의 성향을 인식하고 그에 적절한 항목을 추천 하는 것을 목적으로 설계되었다. 이를 위해 기존 추천 기법 중 대표적인 내용-기반 필터링 기법에 상호진화연산을 결합한 새로운 추천 기법을 제안 하였다. 먼저 상호진화연산의 초기 세대수(initial population)는 구성 성분 추출 과정을 통해 이루어 진다. 이는 상호진화연산의 반복적인 실행을 통해 사용자의 취향에 알맞은 추천 항목을 생성한다.

제안 시스템은 MP3(MPEG-3 Layer) 포맷의 음악 데이터(총 5개 장르, 400곡)로 사용하였다. 모든 데이터는 고유한 음악 속성을 지니고 있으며 추천 항목 생성시 사용된다.

제안 추천 시스템은 총 3가지 단계들(전처리, 사용자 평가, 상호진화연산)로 구성된다.

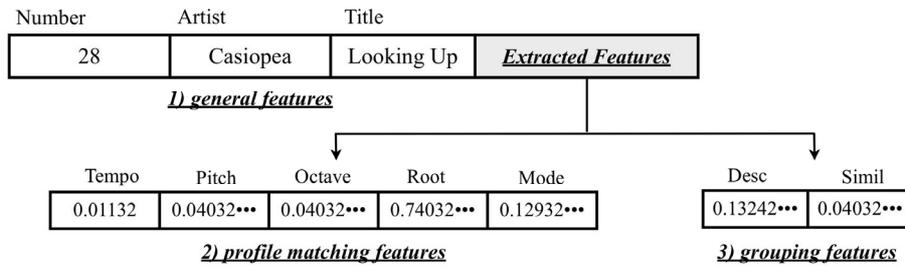


그림 4. 개인개체의 구성 예.  
Fig. 4. Example of individual.

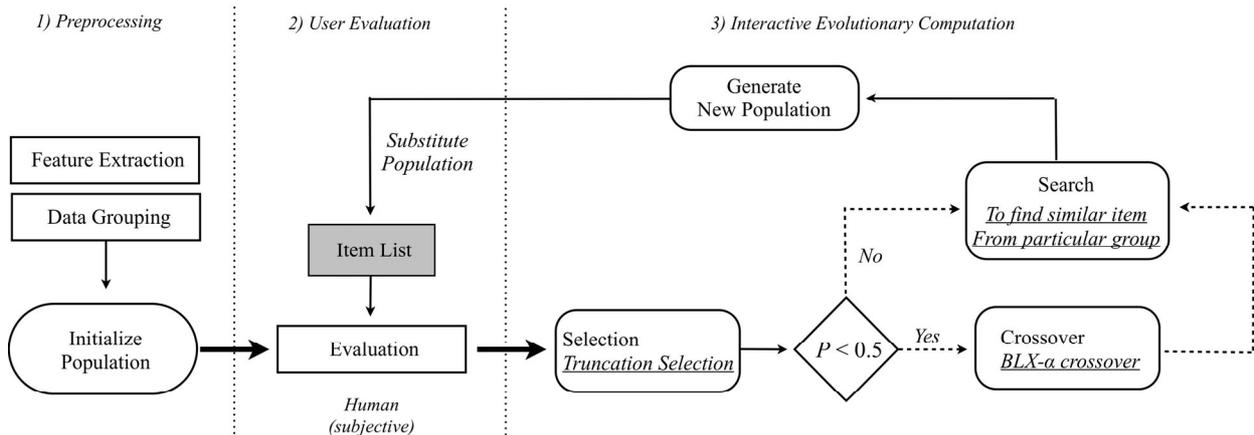


그림 5. 상호진화연산 단계의 구성.  
Fig. 5. Process of IEC phase.

1. 전처리 단계(preprocessing phase)

전처리 단계에서는 시스템 내 데이터의 고유한 성분을 추출한다. 이를 이용하여 그룹화를 수행한다. 이러한 일련의 모든 동작 과정을 전처리 단계로 정의한다

먼저 각 데이터(곡 혹은 트랙)에 대해서 성분 추출 도구인 CLAM을 이용하여 데이터 고유한 성분인 음악적 성분을 추출한다. 이를 통하여 음악적 성분과 곡의 기본정보가 포함되어 있는 XML형태의 결과 파일이 생성되고 파싱(parsing)을 통하여 상호진화연산의 초기 개체(initial population)를 구성한다.

제안 기법에서는 CLAM을 통하여 실수 형태의 총 7가지의 추출된다. 이들은 쓰임새에 따라 총 3가지로 분류되어 저장된다. 그림 4는 CLAM을 통해 추출된 성분들의 종류와 개인개체의 구성을 도식화 한다.

성분 추출 과정을 거친 음악 데이터들은 k-means 클러스터링 알고리즘을 통하여 k개의 그룹을 구성된다. 이러한 그룹화 정보는 제안 시스템에 저장되어 추후 상호진화연산 단계에서 새로운 추천 항목을 구성하는데 사용된다.

2. 사용자 평가 단계(user evaluation phase)

제안 시스템은 사용자가 각 데이터의 적합도를 직접적으로 평가 할 수 있도록 설계하였다. 다시 말해서 사용자는 각자의 취향에 따라 평가 점수를 부여함으로써 항목을 평가한다. 이렇게 평가된 적합도들은 향후 제안 시스템을 동작하는 구간이 된다.

3. 상호진화연산 단계(IEC phase)

상호진화연산 단계는 제안 시스템에 있어 가장 중요한 구

성 요소이며 이전 단계인 사용자 평가 단계에서 주어진 항목별 적합도를 바탕으로 새로운 추천 항목을 생성한다.

제안 시스템에 적용한 추천 기법은 내용 기반 필터링 기법에 상호진화연산을 혼합한 기법이다. 이는 기존 추천 기법들의 단점들인 특정화 문제 및 자원 분산 문제를 해결함으로써 질적으로 향상된 사용자 추천 항목의 생성 서비스 제공하고 자 한다.

이를 위하여 상호진화연산이 가지고 있는 유전 상속 성질(genetic inheritance)을 활용하였다. 즉, 기존 항목의 성질을 이어받아 새로운 항목을 생성함은 물론 질적으로 우수한 항목들(제안 시스템의 경우, 사용자로부터 높은 평가를 받은 항목의 성질)은 계속 유지하면서 그것을 이용하여 새로운 항목을 생성하였다. 이러한 추천 항목을 구성함에 있어 내용 기반 필터링을 적용하여 각 항목의 고유한 속성값을 기준으로 새로운 추천 항목을 검색하도록 하였다.

제안 기법에서 사용된 상호진화연산은 유전연산과 유사한 구조로 동작한다. 그러나 제안 시스템에서는 사용자 평가 항목의 음악적 성질을 파괴할 가능성이 있기 때문에 인위적으로 해의 변화를 주는 돌연변이 연산자는 사용하지 않았다. 그림 5는 제안 기법에서 사용된 상호진화연산의 전체적인 동작 과정을 보여준다.

3.1 선택(selection)

일반적으로 유전 연산자들(genetic operator) 중 선택 연산자(selection operator)는 가장 신중히 설계되어야 한다. 왜냐하면 선택 기법에 따라 전체적인 유전연산의 성능(해를 찾아가는

과정)에 영향을 끼치지 때문이다.

제안 시스템에서는 높은 적합도를 갖는 항목을 효율적으로 선택하는 기법을 사용하였다. 즉, 사용자로부터 낮은 평가를 받은 항목은 사용자의 취향에서 만족하지 못함을 나타낸 것은 물론 추천 항목을 구성 하는데 유용한 정보로 취급되지 않기 때문에 제안 시스템에서는 이를 무시한다. 이를 바탕으로 하여 주어진 선택 가능한 해들 중에서 높은 적합도를 기준으로 정렬한 후 그 중 일정한 비율( $T\%$ )의 해를 선택하고 다른 나머지 해들은 버리는 절단 선택기법의 적용 하였다 [20,21].

선택된 항목들은 선택적으로 교배 단계와 검색 단계에 사용된다. 이는 제안 시스템이 추천 항목을 생성함에 있어 사용자의 기존 선호도(교배 단계를 거치지 않고 검색 단계로 진행되는 항목)를 고려하기 위함이다.

3.2 교배(crossover)

제안 시스템에서 교배 단계는 기존의 항목의 성질을 혼합하여 새로운 항목을 생성하는 중요한 가교 역할을 하며 사용자 취향의 최적화를 위한 중요한 요소이다.

제안 시스템에서 개인 개체는 실수 형태로 구성된 음악적 성분으로 구성 되어있기 때문에 실수로 이루어진 개체에 적용 가능한 BLX- $\alpha$  교배 기법[22]을 사용 하였다. 이 교배 기법의 전체 동작 과정을 그림 6에서 보여 준다.

3.3 검색(search)

검색 단계에서는 앞선 단계들에서 처리된 항목들과 추천 시스템에 남아 있는 다른 항목 들간의 유사도를 측정하여 그 중 가장 낮은 계산 값을 갖는 항목을 찾아내는 것을 목적으로 한다. 또한 제안 시스템은 비교되는 대상의 그룹 정보를 바탕으로 같은 그룹에 속해있는 항목을 대상으로 가장 유사한 항목을 검색한다.

제안 시스템은 앞서 전처리 단계에서 개체의 속성을 기준으로 그룹화를 미리 실행하였기 때문에 각 항목의 그룹 정보를 저장하고 있으며, 같은 그룹에 속한 항목들은 서로 유사한 속성을 지니고 있기 때문에 검색 단계에서 시스템 내의 모든 항목들과 비교할 경우에 비해 검색 시간을 단축할 수 있다.

제안 시스템에서 교배 과정을 수행하지 않은 항목은 일반 (generic) 항목이라 지칭하며, 교배 단계를 거쳐 검색 단계로

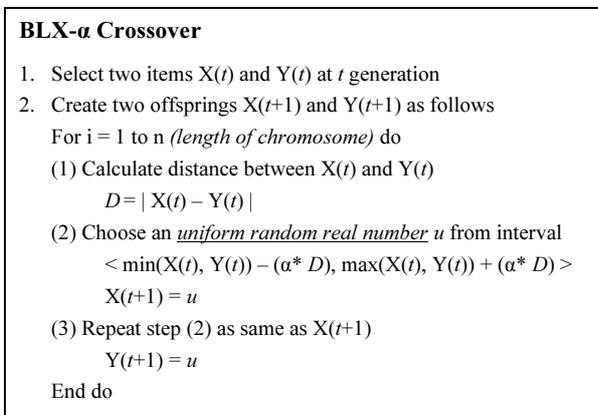


그림 6. BLX- $\alpha$  교배 기법의 단계.

Fig. 6. Process of BLX- $\alpha$  crossover method.

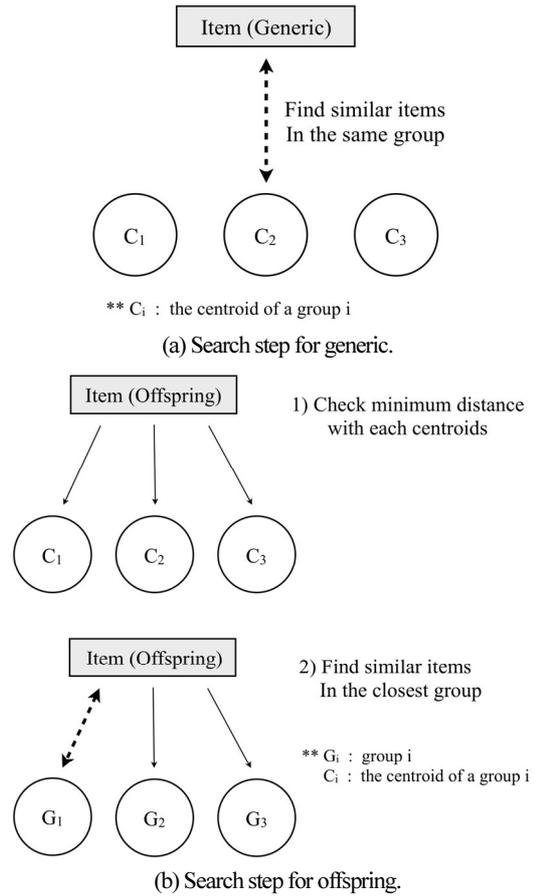


그림 7. 항목 종류에 따른 검색 단계의 구성.

Fig. 7. Different search step for item type.

전달된 항목을 후손(offspring)이라 정의한다. 그림 7은 이 두 가지의 항목별 검색 단계를 보여준다. 여기서 교배 단계를 거치지 않은 일반 항목은 이미 전처리 과정에서 그룹 정보를 저장 하였기 때문에 제안 시스템은 항목 간 유사도를 비교할 대상을 같은 그룹에서 찾을 수 있으나, 교배 단계를 거친 후손 항목은 기존 항목들의 성질을 이어 받은 전혀 새로운 항목으로 시스템에 이와 관련된 그룹 정보가 존재하지 않기 때문에 검색 대상 그룹이 불분명하다. 따라서 제안 시스템은 이러한 문제점을 해결하기 위해 먼저 각 그룹의 중심 항목 (centroid)들 간의 거리(유사도)를 측정하여 가장 근접한 그룹을 선택한 후 선택된 그룹에서 검색 단계를 진행 하였다

또한 검색 단계에서 실제 항목간의 유사도를 측정은 수식 (1)과 같이 정의 하였으며, 낮은 값을 갖는 항목 일수록 높은 유사도를 갖는다.

$$distance(s, t) = \sqrt{\sum_{i=1}^n \frac{1}{m} \sum_{j=1}^m (s_j - t_j)^2} \quad (1)$$

여기서  $distance(s, t)$ 는 두 항목간의 유클리드 거리(euclidean distance)를 나타낸다. 수식 (1)의 변수  $n$ 은 각 항목이 갖고 있는 성분의 수를 뜻하며 변수  $m$ 은 개인 개체의 각 성분을 구성하는 실수 값의 길이를 뜻한다. 또한 2개의 항목의 유사도를 측정 함에 있어 0과 1사이로 평준화(normalization)을 실행 하였다.

IV. 실험 및 분석

제안 시스템의 수율 성능을 컴퓨터 시뮬레이션을 통하여 검증한다. 이를 위하여 총 400개의 MP3 음악파일을 사용하였으며, 각 파일이 지니고 있는 고유 성분은 CLAM을 이용하여 추출하였다. 이렇게 추출된 성분들은 제안 시스템에 저장되며, 이를 기반으로 하여 제안 시스템은 추천 항목을 생성한다.

1. 제안 시스템 성능 검증을 위한 실험 설계

효율적인 제안 시스템의 성능 검증을 위하여 실험용 에이전트(test agent)를 설계하였다. 이전 수행된 연구 [17]에서는 사용자가 웹사이트 방문을 통하여 실험을 진행하였다. 그러나 장시간 동안의 진행되는 실험으로 인하여 참가자들의 감정변화와 집중력 감소 등은 의미 있는 실험 결과를 도출함에 있어 어려움을 있었다. 이러한 문제점을 해결하기 위하여 본 연구에서는 에이전트 기반의 실험을 수행하였다(참조 그림 8). 이를 통하여 결과를 수집 및 분석에 있어 수행 시간을 절약할 수 있었다.

이번 장에서 모든 실험은 공정한 성능 평가를 위하여 100번의 반복실험을 통하여 수행되며, 그 결과는 평균값으로 산출된다. 제안 시스템의 성능 평가를 위해 두 가지 실험을 구성하였다. 각 실험 조건은 표 1에서 구체적으로 설명한다.

2. 실험 결과 및 분석

2.1 실험 #1

제안 추천 시스템은 내용 기반 필터링 기법에 상호 유전 연산을 결합한 새로운 추천 기법으로 구성된다. 또한 그룹화 기법을 적용하여 추천 항목의 질적 향상 및 시간 효율성 또한 증대하였다. 이러한 제안 시스템의 특징들을 효율적으로

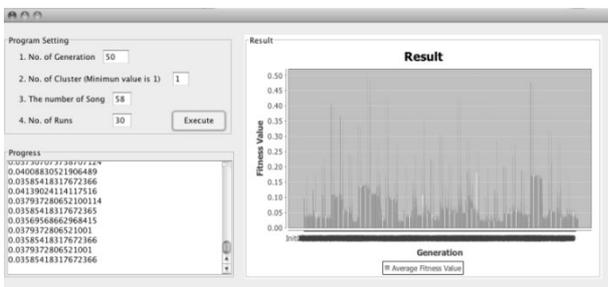


그림 8. 제안된 실험용 에이전트.

Fig. 8. The proposed test agent.

표 1. 제안 시스템 평가를 위한 두 가지 실험.

Table 1. Two test cases to evaluate the proposed system.

[실험 #1]

목적	그룹화에 따른 제안 시스템의 성능 분석
평가 기준	다음 기준들의 최소화 - 평균 적합도(average fitness value) - 실행 시간(execution time), 단위: ms

[실험 #2]

목적	시스템 내 저장된 개체 수의 증가 (데이터 분산 문제)에 따른 제안 시스템의 성능 분석
평가 기준	- 평균 적합도(average fitness value) 최소화 - 특정 지점(개체 수의 증가 지점) 이후 평균 적합도의 변화

평가하기 위하여 첫 번째 실험을 설계하였다. 수행 실험의 조건은 다음과 같다.

- 1) 목록 당 10개의 항목
- 2) 100번의 반복 실행
- 3) 매 실행 시 50세대(generation)까지 진행
- 4) 그룹의 수: 1~10개

여기서 그룹의 수가 1인 경우는 그룹화 없이 실험을 수행한 것을 의미한다.

그림 9는 각 그룹수에 따른 평균 적합도(Average fitness value) 및 실행 시간(Execution time)을 나타낸다. 그림 9(a)에서는 그룹 수가 증가함에 따라 평균 적합도가 같이 증가함을 알 수 있다. 또한 그림 9(b)를 통하여 그룹의 수가 증가함에 따라 실행 시간이 감소하는 것을 확인할 수 있다. 특히 두 가지의 그룹 수들 2와 3에서 실행 시간이 급격하게 감소함을 알 수 있다. 이에 해당하는 수치형 실험 결과들을 표 2에서 보여준다. 이를 통하여 비그룹화(그룹의 수가 1) 경우에 비하여 2, 3개의 그룹을 생성하였을 때 평균 적합도 값이 높음을 알 수 있다. 반면 평균 실행시간의 경우 그룹화시 53% 감소하였음을 알 수 있다. 즉, 그룹화 기법을 이용한 제안 시스템의 성능이 향상됨을 확인 하였다.

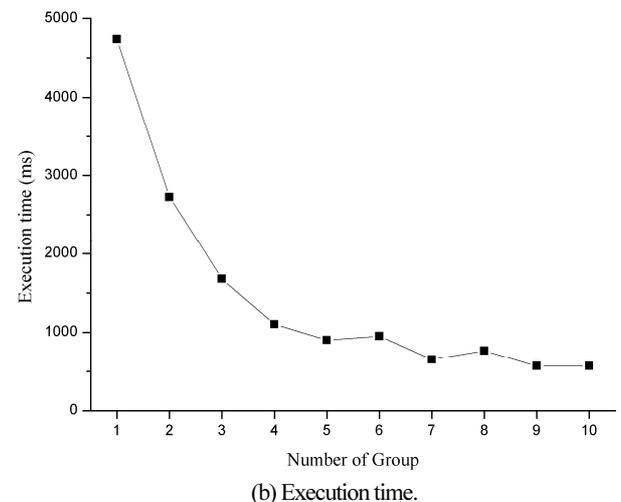
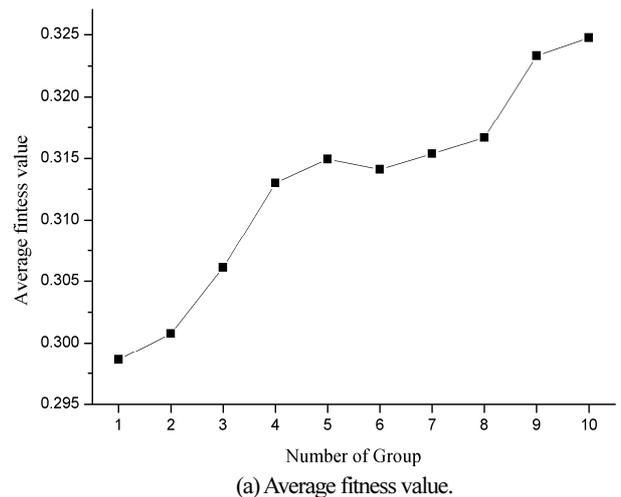


그림 9. 그룹화에 따른 평균 적합도 및 실행시간.

Fig. 9. Average fitness value and execution time for grouping.

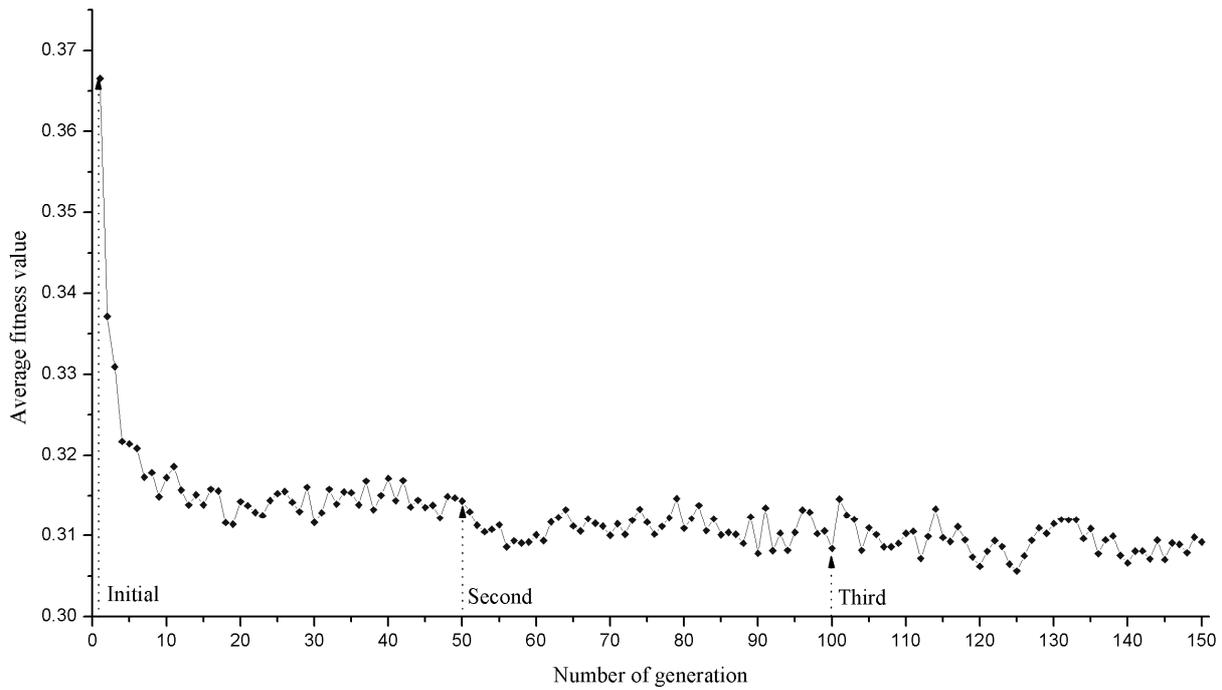


그림 10. 시스템 내 개체 수 변화에 따른 평균 적합도.

Fig. 10. Average fitness of the system when changing the number of data sets.

표 2. 실험 #1 결과.

Table 2. Numerical simulation results on the test case #1.

그룹수	평균 적합도 (Average fitness value)	실행 시간 (Execution time)
1	0.29861	4739
2	0.30076 (+1%)	2725 (-42%)
3	0.30611 (+3%)	1677 (-65%)

( ) 그룹 수(1개) 대비 증가/감소 비율

표 2에서 그룹화를 실행하지 않은(그룹의 수가 1)경우에 비해 2, 3개의 그룹을 생성하였을 때 평균 적합도 값이 높게 형성된 것을 알 수 있다. 반면 실행 시간의 경우 그룹화를 실행했을 때 결과 값이 평균 53% 감소 하였음을 알 수 있다.

2.2 실험 #2

두 번째 실험에서는 기존 기법의 한계점인 데이터 분산 문제(참조 2.2절)에 대하여 제안 시스템의 성능을 평가 및 분석하는 것을 목적으로 한다. 성능 검증 조건은 다음과 같이 정의 된다.

- 1) 음악 데이터 수의 변화
- 2) 목록 당 10개의 항목
- 3) 100번의 반복 실행
- 4) 150세대(generation)까지 진행
- 5) 그룹의 수: 3개

여기에서 선행 수행 실험의 결과를 바탕으로 하여 그룹 수를 3으로 정하였다. 또한 일정한 시간 경과 후 시스템에 새로운 데이터 집합을 추가하는 방법을 통하여 음악 데이터수 변화 시킴으로써 데이터 분산 문제가 추천 시스템의 미치는 영향을 측정하였다. 표 3에서 이러한 음악 데이터 수 변화와 그 시점을 묘사한다.

그림 10에서는 개체수 증가에 따른 제안시스템의 추천 목

표 3. 실험 #2 결과.

Table 3. Experimental result of the test case #2.

시점(세대 수)	시스템 내 개체 수	각 시점 별 명칭
0	200	Initial
50	300 (+100)	Second
100	400 (+100)	Third

( ) 개체수 증가치

록의 평균적합도를 나타내고 있다. 이를 통하여 새로운 개체(음악데이터)가 추가되는 시점 이후 이전 시점의 평균적합도가 일정한 경향을 유지하는 것을 알 수 있다.

V. 결론

본 논문에서는 상호진화연산을 내용-기반 필터링 기법에 결합한 새로운 추천 기법을 제안하였다. 기존의 추천 시스템에서 쓰이던 내용 기반 필터링 기법을 기반으로 개체에 대한 분석 데이터와 상호유전 연산을 통해 사용자가 개인화된 평가 지표를 피드백 하도록 하여 추천과정에 시스템이 유동적으로 구동될 수 있도록 하였고, 데이터 그룹화를 통해 추천 항목을 구성하는데 있어서 시간 효율성을 높이고자 하였다.

제안 시스템의 성능 평가와 효율적인 데이터 수집을 위해 테스트 에이전트를 이용한 실험을 진행 하였으며, 이를 통해 사용자의 평가 지표 값과 실행 시간의 변화를 기준으로 시스템의 성능을 평가 하였다. 모든 실험 결과는 객관적인 수집을 위해 각 실험마다 반복적 실행과 그 결과의 평균 값을 사용하였으며 제안 시스템이 본 논문에서 목표로 하고 있는 “추천”이라는 목적에 맞추어 최적의 사용자의 취향을 파악함은 물론 빠른 시간 효율성까지 보장할 수 있음을 확인하였다. 또한 기존 기법의 문제점으로 지적되던 데이터 분산 문제를 효율적으로 해결할 수 있는 가능성을 확인하였다.

## 참고문헌

- [1] P. Resnick and H. R. Varian, "Recommender system," *Communications of the ACM*, vol. 40, no. 3, pp. 56-58, 1997.
- [2] L. Terveen and W. Hill, "Beyond recommender systems: helping people help each other," *HCI in the New Millennium*, Addison-Wesley, pp. 487-509, 2001.
- [3] J. Schafer, J. Konstan, and J. Riedl, "Recommender systems in E-commerce," *Proc. of the 1st ACM conference on Electronic commerce*, pp. 158-166, 1999.
- [4] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of recommendation algorithms for e-commerce," *Proc. of the 2nd ACM conference on Electronic commerce*, pp. 158-167, Oct. 2000.
- [5] M. Balabanovi'c and Y. Shoham, "FAB: content-based, collaborative recommendation," *Communications of the ACM*, vol. 40, no. 3, pp. 66-72, 1997.
- [6] R. Burke, "Hybrid web recommender systems," *Lecture Notes in Computer Science*, vol. 4321, pp. 377-408, 2007.
- [7] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, pp. 734-749, June 2005.
- [8] M. J. Pazzani, "A framework for collaborative, content-based and demographic filtering," *Artificial Intelligence Review*, vol. 13, no. 5-6, pp. 394-408, 1999.
- [9] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," *Proc. of the Conference on Supported Cooperative Work Conf.*, pp. 175-186, 1994.
- [10] D. Goldberg, D. Nichols, B. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, 35, 12 1992.
- [11] K. B. Sim and K. E. Ko, "Study of emotion recognition based on facial image for emotional rehabilitation biofeedback," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 16, no. 10, pp. 957-962, Oct. 2010.
- [12] D. W. Lee, K. R. Cho, and S. W. Baek, "Fitness change of mission scheduling algorithm using genetic theory according to the control constants," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 16, no. 6, pp. 572-578, June 2010.
- [13] D. Fogel, *Evolutionary Computation: Towards a New Philosophy of Machine Intelligence*, IEEE Press, 1996
- [14] M. Mitchell, *An Introduction to Genetic Algorithm*, MIT Press, 1998.
- [15] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine Learning*, vol. 3, no. 2-3, pp. 95-99, 1988.
- [16] H. Takagi, "Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation," *Proc. of the IEEE*, vol. 89, pp. 1275-1296, 2001.
- [17] H. T. Kim, E. G. Kim, J. H. Lee, and C. W. Ahn, "A recommender system based on genetic algorithm for music data," *Proc. of International Conference of Computer Engineering and Technology*, vol. 6, pp. 414-417, 2010.
- [18] M. David, "Chapter 20. an example inference task: clustering," *Information Theory, Inference and Learning Algorithm*, 2003.
- [19] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl, "Constrained k-means clustering with background Knowledge," *Proc. of 18th International Conference on Machine Learning*, pp. 577-584, 2001.
- [20] D. Thierens and D. Goldberg, "Elitist recombination: an integrated selection recombination GA," *Proc. of the First IEEE Conference on Evolutionary Computation*, pp. 508-512, 1994.
- [21] J. F. Crow and M. Kimura, "Efficiency of truncation selection," *Proc. of the National Academy of Sciences of the United States of America*, vol. 76, no. 1, pp. 396-399, 1979.
- [22] F. Herrera, M. Lozano, and J. L. Verdegay, "Tackling real-coded genetic algorithms: operators and tools for behavioural analysis," *Artificial Intelligence Review*, vol. 12, pp. 265-319, 1998.
- [23] CLAM, <http://clam-project.org>



김 현 태

2002년 건국대학교 컴퓨터학과(공학사). 2011년 성균관대학교 정보통신공학부(공학석사). 2011년~현재 성균관대학교 정보통신공학부 박사과정 재학. 관심분야는 최적화 기법, 진화 알고리즘, 상호추천시스템.



안 창 욱

2000년 고려대학교 전파공학과(공학석사). 2005년 광주과학기술원 정보통신공학과(공학박사). 2005년~2007년 삼성중합기술원전문연구원. 2008년~현재 성균관대학교 정보통신공학부 조교수. 관심분야는 진화알고리즘, 지능형네트워크, 자기-제조립 군집 로봇.



안 진 응

1999년 한국과학기술원 기계공학과(공학석사). 2005년 동 대학원(공학박사). 2005년~2008년 광운대학교 정보제어공학과 교수. 2008년~현재 대구경북과학기술원 실용로봇연구소 소장. 관심분야는 군집지능, HRI, 진화알고리즘, 뇌-기계인터페이스, 재활로봇, 군집로봇.