

# 혼합형 유전자 알고리즘을 적용한 사례기반추론 공사비에측 - 상관분석을 이용한 지역탐색 기법을 중심으로 -

정상선<sup>1</sup> · 박문서<sup>2</sup> · 이현수<sup>3</sup> · 윤인석<sup>4\*</sup>

<sup>1</sup>서울대학교 건축학과 석박통합과정 · <sup>2</sup>서울대학교 건축학과 교수 · <sup>3</sup>서울대학교 건축학과 교수 · <sup>4</sup>서울대학교 건축학과 석박통합과정

## Cost Estimation of Case-Based Reasoning Using Hybrid Genetic Algorithm - Focusing on Local Search Method Using Correlation Analysis -

Jung, Sangsun<sup>1</sup>, Park, Moonseo<sup>2</sup>, Lee, Hyun-Soo<sup>3</sup>, Yoon, Inseok<sup>4\*</sup>

<sup>1</sup>Integrated M.S & Ph.D Student, Department of Architectural Engineering, Seoul National University

<sup>2</sup>Professor, Department of Architectural Engineering, Seoul National University

<sup>3</sup>Professor, Department of Architectural Engineering, Korea University

<sup>4</sup>Integrated M.S & Ph.D Student, Department of Architectural Engineering, Seoul National University

**Abstract :** Estimates of project costs in the early stages of a construction project have a significant impact on the operator's decision-making in important matters, such as the site's decision or the construction period. However, it is difficult to carry out the initial stage with confidence because information such as design books and specifications is not available. In previous studies, case-based reasoning was used to predict initial construction costs, and genetic algorithms were used to calculate the weight of the inquiry phase among them. However, some say that it is difficult to perform better than the current year because existing genetic algorithms are calculated in random numbers. To overcome these limitations, correlation numbers using correlation analysis rather than random numbers are reflected in the genetic algorithm by method of local search, and weights are calculated using a hybrid genetic algorithm that combines local search and genetic algorithms. A case-based reasoning model was developed using the weights calculated and validated with the data. As a result, it was found that the hybrid GA-CBR applied with local search performed better than the existing GA-CBR.

**Keywords :** Cost Estimation, Case-Based Reasoning, Hybrid Genetic Algorithm, Local Search, Correlation Analysis

## 1. 서론

### 1.1 연구의 배경 및 목적

건설공사에서 프로젝트 초기 단계 공사비에측은 프로젝트 예산책정이나 공사 기간산정과 같은 중요한 업무에서 발주자의 의사결정에 큰 영향을 미친다. 초기 공사비에측은 도면과 시방서 등 설계도서가 완전히 갖춰지지 않은 상태에서 개략적인 정보만을 가지고 진행된다. 초기 단계 공사비에측이 건설 사업의 성패를 결정하는 요인임 에도 불구하고 신뢰도 있게 공사비에측을 수행하는 것에 어려움이 있다 (An and Kang, 2005). 이에 초기 공사비에측의 정확성 향상

을 위한 기존의 연구들은 다양한 방법의 예측모델들을 개발하였다.

대표적으로 공사비에측을 위해 과거의 사례를 기반으로 현재 문제(예측대상 프로젝트)의 해결책(예측 결과)을 제시하는 방법인 사례기반추론(Case-Based Reasoning, CBR)을 활용한 연구가 수행되었다(Kolodner, 1993). 사례기반추론은 조회(retrieve), 재사용(reuse), 수정(revise), 저장(retain)의 네 단계로 구성된다. 이 중 첫 단계인 조회는 추론 과정에 있어 프로젝트 사례 데이터 중 유사한 사례의 탐색을 위한 단계이다(Goh and Chua, 2009). 조회를 위해서는 데이터의 속성 가중치가 산정되어야 하며 가중치가 어떻게 산정되느냐에 따라 도출되는 유사한 사례가 달라지며, 이는 사례기반 추론의 정확도에 큰 영향을 준다(Dogan et al., 2006). 사례기반 추론을 활용한 많은 선행연구에서 속성 가중치 산정을 위한 방법으로 유전자 알고리즘(Genetic Algorithm, GA)을 활용했다(Park et al., 2010).

\* **Corresponding author:** Yoon Inseok, Department of Architectural Engineering, Seoul National University, Seoul 08826, Korea  
E-mail: yoon92411@snu.ac.kr

**Received** October 11, 2019; **revised** November 29, 2019

**accepted** December 29, 2019

유전자 알고리즘은 진화의 원리를 모방한 대표적인 최적화 알고리즘으로써 다양한 분야의 문제 해결을 위해 활용되어왔다. 유전자 알고리즘은 선택, 교차, 변이 등을 연산자로 하여 최적화를 수행하게 된다. 하지만 정해진 규칙과 수식에 따라 연산할 뿐, 해당 연산자와 데이터의 특징 사이에는 관련성이 부족하다(Lee et al., 2005). 일반적인 유전자 알고리즘 프로세스에서는 연산 시 임의의 값을 사용하여 새로운 해를 생성하여 해집단을 구성한다. 하지만 단순히 임의의 값을 활용하여 연산을 반복하는 방법으로는 현재 해보다 더 좋은 해를 발견하기 어렵다는 단점이 있다(Hwang and Kim, 2010). 이러한 한계를 극복하고 보다 좋은 성능을 얻기 위해 유전자 알고리즘과 지역탐색기법(Local Search)을 결합하여 더 좋은 예측 결과를 얻고자 하는 연구가 이루어지고 있고(Hwang, 2010; Kang et al., 2002; E.K. Burke et al., 2007; R. Qu et al., 2008), 이를 혼합형 유전자 알고리즘(Hybrid Genetic Algorithm)이라고 정의한다(Oh, 2004).

한편 공사비의 영향 요인 도출을 위해 통계 기법인 상관분석(Correlation Analysis)을 이용하는 연구가 수행되어 왔다. Song et al. (2006)은 공사비의 중 여러 영향요인이 총경비에 미치는 영향도와 우선순위를 규명하기 위해 ISM(Interpretive Structural Modeling) 기법을 사용하여 상관관계에 대한 해석을 진행했다. 또한, Kim et al. (2013)은 상관분석을 다중회귀분석에 적용함으로써 개략공사비 산정 모델을 구축했다. 이처럼 속성 간 상관분석을 통해 공사비에 영향을 미치는 정도를 상관계수로 표현할 수 있으며, 이는 공사비예측 모델 개발에 있어 중요하게 활용되었다.

따라서 본 연구에서는 상관계수를 활용한 지역탐색 기법과 유전자 알고리즘을 결합한 혼합형 유전자 알고리즘으로 속성 가중치를 산정한 후 사례기반추론을 이용해 공사비예측을 수행한다. 피어슨 상관분석을 이용하여 상관계수를 산출한 후 그 값을 지역탐색방법으로 활용하여 각 세대의 모집단 생성 과정에 적용해 그 프로세스를 개선한다. 이후 개선된 프로세스를 기반으로 실제 데이터를 통해 사례기반추론 공사비예측 모델의 유효성을 검증하는 것을 목적으로 한다.

### 1.2 연구의 방법

연구는 다음 절차에 따라 진행된다.

첫째, 선행연구 분석을 통해 시사점을 도출한다.

둘째, 개선된 유전자 알고리즘 개발을 위해 사례기반추론, 유전자 알고리즘, 지역 탐색(Local Search)의 원리와 적용을 고찰한다.

셋째, 아파트, 군 병영생활관 시설, 사무용 건물 세 가지 사례 데이터에 대해 상관분석을 진행하고 그에 따른 각 속성별 상관계수를 도출한다.

넷째, 상관계수를 사용하여 지역탐색을 적용한 혼합형 유전자 알고리즘으로 속성별 가중치를 도출하고 그것을 이용해 사례기반추론 공사비예측 모델을 구축한다.

다섯째, 본 연구의 혼합형 GA-CBR 모델과 다른 가중치 산정방법을 사용한 모델의 예측 정확도 비교를 통해 유효성을 검증한다.

## 2. 관련 연구 분석

사례기반추론에 있어 속성 가중치 산정은 예측 성능에 있어 중요한 역할을 한다. 이에 본 연구에서는 속성 가중치 산정의 품질을 높이기 위해 유전자 알고리즘에 지역탐색방법을 결합한 혼합형 유전자 알고리즘으로 최적화를 수행한다. 본 장에서는 사례기반추론의 개념을 정의하고 유전자 알고리즘 및 지역탐색에 대한 원리를 고찰한다.

### 2.1 사례기반 추론

사례기반추론이란 새로운 문제를 해결하기 위하여 과거의 유사한 사례를 저장하고, 그 사례의 정보와 지식을 재사용하는 것이라고 정의한다. 그 과정은 (Fig. 1)과 같으며 사례기반추론이 다른 인공지능 기법과 구별되는 특징은 첫째, 새로운 문제 해결을 위해 과거 사례의 구체적 지식을 사용한다는 점이다. 둘째, 해결된 새로운 문제는 사례기반의 과거 사례로 저장되기 때문에, 미래에 발생할 문제 해결에 사용될 수 있다는 점이다(Aamodt and Plaza, 1994).

본 연구에서는 사례기반추론의 과정 중 조회단계를 이용해 건설공사비 데이터의 유사사례를 추출해서 새로운 데이터의 공사비를 추정하고자 한다. 조회단계는 문제와 과거 사례의 일치하는 속성을 비교함으로써 유사사례를 추출하는

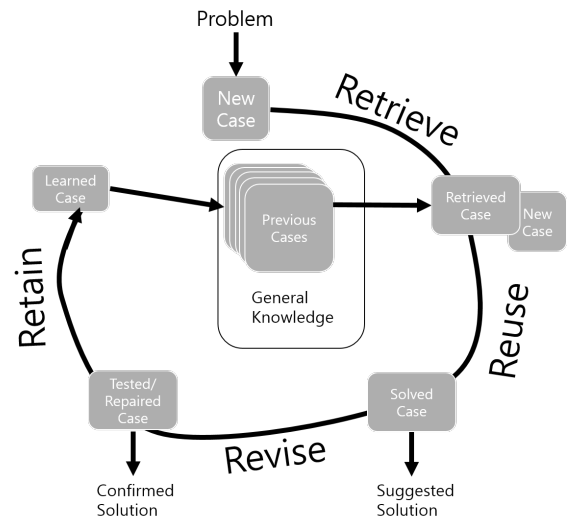


Fig. 1. Case-Based Reasoning cycle (Aamodt and Plaza, 1994)

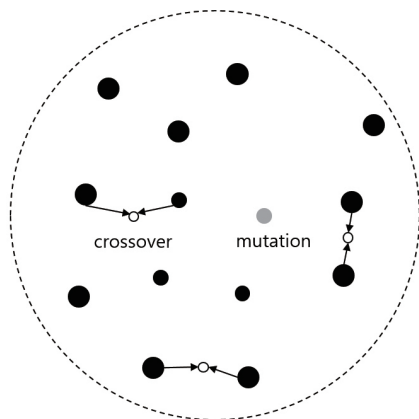
것이다. 유사사례의 판단은 사례 유사도 점수로 수행하고 사례 유사도는 속성 유사도 점수와 속성 가중치를 산정해 두 가지를 연산한 값으로 결정한다. 속성 유사도는 과거 사례의 속성값과 문제사례의 속성값의 차이를 수식 혹은 규칙을 이용하여 계산한다. 속성 가중치는 과거 사례분석을 통해 결정하는데, 속성 가중치란 유사한 사례를 조회할 때 중요한 입력 속성에 높은 가중치를 할당하여 정확한 조회를 유도할 수 있도록 도와주는 값이다(Lee, 2008). 이러한 속성 가중치를 산정하는 방법에는 <Table 1>의 선행연구들과 같이 동일 가중치법, 회귀분석법, 유전자 알고리즘 등이 있다. 본 연구에서는 속성 가중치 산정을 위해 상관분석과 지역탐색을 활용해 혼합형 유전자 알고리즘을 사용한다.

**Table 1. Methods for determining the weight of Case-Based Reasoning**

Weighting Method	Research Content	Author	year
Uniform weight	Extraction of similar cases using Fuzzy concepts in Case-Based Reasoning	Kim et al.	1996
Regression analysis	A cost estimating model for construction projects in the early stage using Support Vector Regression	Park and Kim	2007
	Financial index forecasting model using Case-Based Reasoning	Chun and Park	2006
Genetic algorithm	A method of assigning weight values for qualitative attributes in Case-Based Reasoning cost model	Lee et al.	2004
	Comparison of the accuracy between cost prediction models based on neural network and genetic algorithm	Kim et al.	2006

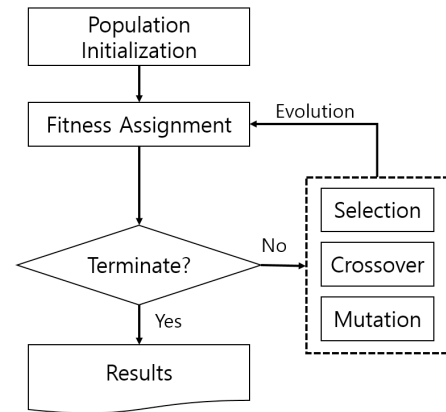
**2.2 유전자 알고리즘**

유전자 알고리즘은 대표적인 전역 최적화(Global Optimization) 알고리즘으로 다윈의 진화론을 토대로 해의 값들을 자연적인 모집단 초기화, 선택, 교차, 변이 등의 연산



**Fig. 2. Search scope of global optimized solution**

자들의 단계를 통한 진화로 최적 해를 찾아가는 진화 알고리즘이다(D.E. Goldberg, 1989). <Fig. 2>와 같이 해집단 전체 탐색을 통해 최적의 해집단을 찾게 되며, 그 과정에 선택, 교차, 변이 등의 연산자들이 활용된다. 해들을 나타내는 자료구조는 유전자(gene)라고 하고, 이들을 변형함으로써 점점 더 좋은 해를 만들어내어 세대(population)를 넘어가는 과정은 진화(evolution)로 표현할 수 있다.



**Fig. 3. General genetic algorithm process**

일반적인 유전자 알고리즘의 프로세스는 <Fig. 3>과 같다. 연산자를 이용해 최적 해를 찾기 위해서는 초기 해의 집단이 필요하다. 일반적으로는 유전자를 임의의 난수로 생성하여 초기 해집단을 구성한 후 해집단에 포함되어있는 모든 해의 품질을 적합도 함수를 통해 산정하게 된다. 적합도를 기반으로 선택, 교차, 변이를 통해 다음 세대의 해의 집합을 생성하게 된다. 이 과정을 세대를 거듭하면서 반복하면, 해들은 점점 최적 해에 가까워지게 된다. 유전자 알고리즘이 최적 해를 구하기 위해 많은 인구를 유지하면서 많은 세대를 내려갈 필요가 있으며 세대가 일정 수준 진화되었거나, 해가 특정 범위에 들게 되면 알고리즘을 종료한다.

건설 분야의 공사비예측에도 GA를 사용하는 연구들이 진행되고 있다. Park et al. (2010)은 공사비예측 분야에 있어서 유전자 알고리즘을 사용해 가중치를 산정한 모델과 표준화 회귀계수, 동일가중치 등의 방법을 사용한 모델을 비교함으로써 유전자 알고리즘의 성능이 뛰어난 것을 검증하였다. Lee et al. (2010)은 정성변수의 속성 가중치 산정방법을 제시함으로써 사례기반추론을 위한 데이터에 정성 속성이 포함된 경우의 예측 정확도를 향상하였다. Kim et al. (2006)은 오류 역전과 신경망의 각 파라미터를 유전자 알고리즘에 의해서 결정하고, 신경망의 학습을 유전자 알고리즘을 이용해 시행하는 등 신경망과 유전자 알고리즘을 결합해 공사비를 예측하는 연구를 진행했다. 이러한 기존 연구들은 유전자 알고리즘의 기본 개념인 임의의 난수를 이용하여 연산을 수행

하는 방법을 적용한다. 이것은 최적화된 해를 구함에 있어서 건설프로젝트 속성 간의 특징이 반영되지 않고, 임의의 수들의 연산을 통해 해를 구하게 된다는 것을 의미한다. 하지만 해집단을 임의의 값으로 생성하는 방법으로는 현재 해보다 더 좋은 해를 발견하기 어렵다는 단점이 있다(Hwang and Kim, 2010). 이러한 단점을 보완하기 위해 본 연구에서는 임의의 값이 아닌 공사비 데이터 각 속성 간의 상관관계를 분석하여 그 결과값을 유전자 알고리즘이 진행되는 세대마다 지역탐색 방법으로 적용해 최적화를 수행한다.

### 2.3 지역 탐색(Local Search)

지역탐색은 최적화 알고리즘에서 지역탐색 절차의 상위에 위치하여 지역 최소점에서 탈출할 수 있도록 도와주는 일반적인 메타휴리스틱이다. 기본 원리는 <Fig. 4>와 같이 해의 탐색영역 내에서 현재 해를 기준으로 이웃해를 탐색해 목적함수 결과에 대한 비교 평가를 통해 최적해로 개선해가는 방식이다. 지역탐색법은 전역탐색법과 같이 해집단의 모든 해를 대상으로 최적해를 탐색하는 것이 아닌 현재해를 그 지역 내에서 목적함수에 근접하게 변경시키는 것을 의미한다(Hwang and Kim, 2010).

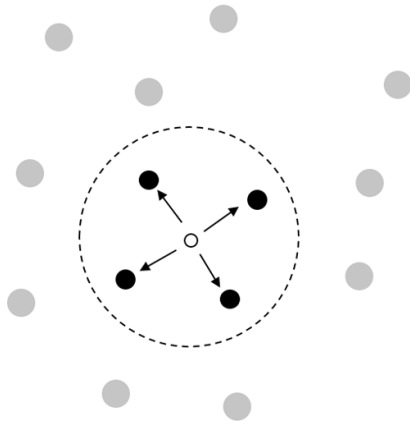


Fig. 4. Search scope of local optimized solution

이후 현지 검색을 위해 목적함수를 사용하여 계속 최적해 탐색을 진행하며 목적함수의 사용을 통해 다음의 탐색이 지역 최소값에서 벗어나 더 좋은 최적해로의 수렴을 목표로 한다. 지역탐색에서 목적함수를 설정하는 것은 각각의 문제마다 다르며 이에 따른 문제 해결방법도 다르다. 지역탐색은 차량 경로 문제(Bruno De Backer et al., 1997), 인력스케줄링 문제(Tsang, E.P.K. et al., 1997) 무선링크 주파수 할당 문제(Vourmouris, C. et al., 1998) 등과 같은 여러 최적화 및 탐색문제에 성공적으로 적용되었다.

더불어 지역탐색과 다른 최적화 방법을 결합하여 사용하는 선행연구도 진행되었다. Hwang and Kim (2010)은 지역

탐색 기법의 하나인 언덕 오르기 탐색방법과 정수계획법을 결합해 더 좋은 해를 도출할 수 있음을 검증했다. 언덕 오르기 탐색은 차이감소법(Difference reduction)을 사용한다. 차이감소를 언덕 오르기(hill climbing)라고 부르는데, <Fig. 5>와 같이 목표가 높은 곳에 있을 때 도달하는 방법은 위로 올라가는 것이다. 언덕 오르기는 길을 계속 따라가다 보면 가장 높은 지점(global maximum)보다 낮은 어떤 언덕의 꼭대기에 도달하게 될 것이라는 개념이다.

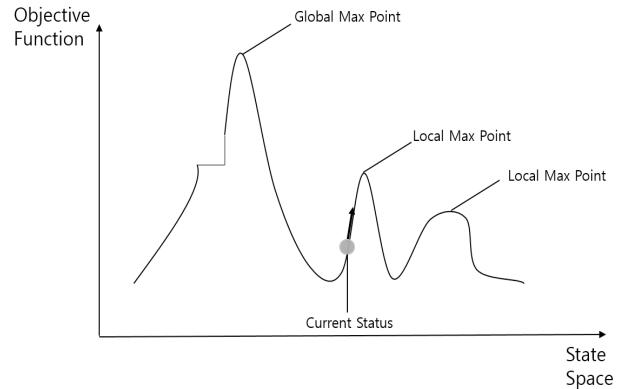


Fig. 5. Hill climbing search

이는 현재 상태와 목표상태의 차이를 줄여서 목표에 접근하는 방법이다. 구체적으로는 차이를 줄여서 현재 상태보다는 목표상태와 더 가깝게 새 상태로 현재 상태를 바꾸는 것을 의미한다. 이 연구에서는 초기해 생성 후 일반적인 단순 언덕 오르기 탐색과 마찬가지로 종료 조건이 충족될 때까지 이웃해 하나를 생성하고 이를 현재 해로 재설정하는 과정을 반복 실행하게 된다. 지역탐색방법이 적용된 점은 이웃해 생성을 위한 목적함수로서 정수계획법을 적용한 부분이다. 적용한 모델을 N-Queens 최대화 문제(J. Paredis, 1993)를 대상으로 실험한 후, 지역탐색과 정수계획법을 적용한 모델이 다른 탐색 기법들보다 더 좋은 해를 산출했다는 유효성을 검증했다.

또한, Kim and Choi (2009)는 지역탐색 기법인 최고우선 탐색(Best First Search)의 한 예로 A\*알고리즘을 사용하여 필요한 작업의 교착 현상을 방지함과 동시에 총 작업수행시간을 최소화하는 등 스케줄링 문제의 해결방안을 제시하였다. 최고우선탐색이란 문제의 특성에 대한 정보인 휴리스틱에 따라 목표까지의 가장 좋은 경로상에 있다고 판단되는 노드를 우선 방문하도록 진행되는 것이다. 이 연구에서는 사용된 도달 가능 그래프(Reachability Graph)의 노드 수가 많아 최고우선탐색 기법으로 최소의 노드만을 탐색하는 방법을 사용했고 이를 통해 최적의 스케줄을 산정했다. 또한, 기존 알고리즘을 이용한 대상보다 43.7% 이상의 노드 탐색 수를 감소했다는 유효성을 검증했다.

이렇듯 지역탐색은 좋은 해를 탐색하는 방법으로서 최적화 알고리즘 내에서 적용되었을 때 알고리즘의 성능을 향상시킬 수 있다(R. Qu et al., 2008). 또한, 지역탐색방법은 지역 최적해 산출을 위한 방법이고, 전술한 유전자 알고리즘은 대표적인 전역 최적해 산출을 위한 방법이다. 본 연구는 유전자 알고리즘의 전역 해 공간 탐색 능력과 상관분석을 이용한 지역탐색의 강점을 결합하여 혼합형 유전자 알고리즘으로 가중치를 산정하고 사례기반추론을 통해 공사비에측을 수행하며 다음 장에서 그 방법에 대해 구체적으로 설명한다.

### 3. 지역탐색을 적용한 혼합형 유전자 알고리즘 가중치 산정

본 연구는 GA-CBR의 가중치를 산정하기 위해 앞서 언급한 개념인 지역탐색의 방법에서 각 속성의 상관계수를 사용한다. 이를 위해 상관분석을 통해 상관계수를 도출할 필요가 있다. 유전자 알고리즘에서 지역탐색방법을 적용하는 부분은 두 가지이다. 첫째, 기존 유전자 알고리즘이 임의의 수로 모집단을 초기화(Initialization)하는 것에 대해 본 연구는 기존의 모집단에 각 속성의 상관계수를 반영해 모집단을 초기화 방법을 개선한다. 둘째, 유전자 알고리즘 내 세대(Population)가 진화하기 직전 단계에서 각 유전자에 산출한 상관계수를 반영해 다음 세대의 진화를 수행한다. 이 두 가지는 기존 유전자 알고리즘과 달리 가중치 산정 시 상관계수를 적용함으로써 건설 속성의 특성을 알고리즘에 반영할 수 있으며 임의의 수가 아닌 각 속성의 상관계수를 적용함으로써 좋은 성능을 기대할 수 있다.

#### 3.1 상관분석(Correlation Analysis)

본 연구에서 제시한 상관분석을 적용한 공사비에측 모델 구축과 검증에 위해 공공아파트, 군 시설(병영생활관), 사무용 건물 데이터를 수집했다. 공사비에측에 사용한 데이터의 속성(Attribute)과 데이터에 대한 정보는 각 사례에 따라 <Table 2>, <Table 3>, <Table 4>과 같으며 속성들은 설계 단계에서 활용 가능한 정보들을 수집했다.

##### 3.1.1 피어슨 상관분석

지역탐색을 수행하기 위해 데이터의 각 속성 중 독립변수(총공사비)와 종속변수(그 외 속성) 간 상관분석을 수행한다. 산출한 상관계수를 이용하여 지역탐색을 진행하고 이를 유전자 알고리즘에 결합하여 최적화된 가중치를 산정하며, 상관계수는 피어슨 상관계수를 사용한다. 피어슨 상관계수는 두 변수의 공분산을 표준편차의 곱으로 나눈 값이다. 프로젝트의 총공사비를 x축, 그 외 다른 속성을 y축으로 설정하고 상관분석을 수행했다.

#### 3.1.2 사례 1 : 공공아파트

본 연구에서 사용된 사례 1 공공아파트 데이터는 국내 공사에서 발주한 아파트 공사 9개 단지를 사용하였다. 데이터 속성은 세대 수, 연면적, 엘리베이터 수, 공사비 등 12개이고 프로젝트 데이터의 총 개수는 165개이다. 각 속성 정보 및 그에 따른 상관계수는 <Table 2>와 같다.

Table 2. Information of case 1 attributes

No	Attributes	Type	Correlation coefficient
X1	Number of households	Numeric	0.8336
X2	Gross floor area	Numeric	0.9701
X3	Number of unit floor households	Numeric	0.6952
X4	Number of elevators	Numeric	0.3865
X5	Number of floors	Numeric	0.7290
X6	Number of piloti with household scale	Numeric	0.4854
X7	Number of households of unit floor per elevator	Numeric	0.4556
X8	Height between stories	Numeric	0.5171
X9	Depth of pit	Numeric	0.0166
X10	Roof type	Flat or inclined(1 or 0)	0.4296
X11	Hallway type	Hall or corridor(1 or 0)	0.4135
X12	Cost	Numeric	1

분석 결과 X1(아파트당 세대 수), X2(연면적), X3(층수), X12(구조형식) 네 가지의 속성이 총공사비와의 상관계수가 0.5 이상으로 나타났으며, 약 0.83, 0.97, 0.72, 0.70으로 상관관계가 높은 것을 확인하였다. 이와 반대로 X4(엘리베이터 수), X7(엘리베이터 한 대당 가구 수), X9(피트의 깊이), X11(복도 종류)은 상관계수가 0.5 이하였으며 각각 약 0.38, 0.45, 0.016, 0.41을 나타냈고 상관관계가 상대적으로 낮은 것을 확인하였다.

#### 3.1.3 사례 2 : 군 시설(병영생활관)

사례 2의 데이터는 3개 군의 국방병영생활관 직접공사비 자료이며 데이터의 속성은 수용인원, 연면적, 내부면적, 사무실 면적, 공사비 등 9개이고 프로젝트 데이터의 총 개수는 117개이다. 각 속성정보 및 그에 따른 상관계수는 <Table 3>과 같다.

Table 3. Information of case 2 attributes

No	Attributes	Type	Correlation coefficient
X1	Number of capacity	Numeric	0.8263
X2	Number of floors	Numeric	0.6830
X3	Gross floor area	Numeric	0.9814
X4	Building area	Numeric	0.9306
X5	Room area	Numeric	0.0295
X6	Office area	Numeric	0.2108
X7	Basement floor status	Existence or Non(1 or 0)	0.2936
X8	Pit status	Existence or Non(1 or 0)	0.4160
X9	Cost	Numeric	1

분석 결과 X1(수용인원), X3(연면적), X4(건축면적) 세 가지의 속성이 총공사비와의 상관관계가 0.7 이상으로 나타났으며, 약 0.82, 0.98, 0.93으로 상관관계가 높은 것을 확인하였다. 이와 반대로 X5(실 면적), X6(사무실 면적), X7(지하층 유무)는 상관관계가 0.3 이하였으며 각각 약 0.02, 0.21, 0.29를 나타냈고 상관관계가 상대적으로 낮은 것을 확인하였다.

**3.1.4 사례 3 : 사무용 건물**

사례 3의 데이터는 조달청 공공공사 데이터의 프로젝트 유형 중 일반청사 및 대형청사의 데이터를 수집했다. 건물 데이터의 속성은 대지면적, 지하층 수, 지상층 수, 건폐율, 용적률, 공사비 등 10개이고 프로젝트 데이터의 총 개수는 52개이다. 각 속성정보 및 그에 따른 상관계수는 <Table 4>와 같다.

**Table 4. Information of case 3 attributes**

No	Attributes	Type	Correlation coefficient
X1	Lot Area	Numeric	0.1905
X2	Gross floor area	Numeric	0.2512
X3	Building Coverage Ratio	Numeric	0.0057
X4	Floor Area Ratio	Numeric	0.0333
X5	Number of Underground Floor	Numeric	0.2278
X6	Number of Ground Floor	Numeric	0.2434
X7	Structure Type(RC)	Existence or Non(1 or 0)	0.2717
X8	Structure Type(SRC)	Existence or Non(1 or 0)	0.0965
X9	External Material	Metal or Stone(1 or 0)	0.0518
X10	Cost	Numeric	1

분석 결과 X2(연면적), X5(지하층 수), X6(지상층 수), X7(구조형식:RC) 네 가지의 속성이 총공사비와의 상관관계가 0.2 이상으로 나타났으며, 약 0.25, 0.22, 0.24, 0.27으로 상관관계가 상대적으로 높은 것을 확인하였다. 이와 반대로 X3(건폐율), X4(용적률)가 0.04 이하였으며 각각 약 0.005, 0.033을 나타냈고 상관관계가 상대적으로 낮은 것을 확인하였다.

사례 1, 사례 2, 사례 3의 데이터에 대해 상관분석을 진행한 결과 공통적으로 연면적의 계수가 높게 나타났다. 산출한 속성 간 상관계수는 지역탐색기법의 목적함수로서 사용되며 유전자 알고리즘 프로세스 내의 모집단 및 세대의 해집단(가중치)에 곱해주는 형태로 반영한다.

**3.2 모집단 초기화/각 세대 지역탐색적용**

유전자 알고리즘의 모집단 초기화는 최초염색체 세대를 만들어내는 과정으로 임의의 난수 생성기를 이용하는 경우가 대부분이다. 이러한 모집단 초기화 과정은 유전자 알고리즘이 탐색할 문제의 공산을 형성하는 과정으로써 공간에 각 염색체의 배치 구성에 따라 유전자 알고리즘의 성능과 효율

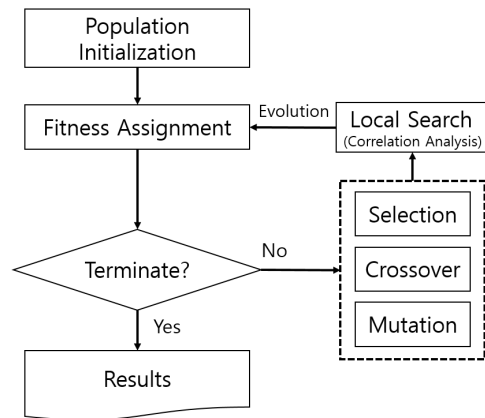
성에 영향을 줄 수 있다(Heilkki Maaranen, 2007). 이에 본 연구는 유전자 알고리즘의 성능을 높이기 위해 기존 초기화된 모집단에 각 속성의 상관계수를 반영한 개선된 모집단 초기화를 진행한다.

모집단 초기화에 대한 적용은 식 (1)과 같다.

$$P_i = R_i * C_i \tag{1}$$

$P_i$ : 새로운 모집단의  $i$ 번째 값  
 $R_i$ : 임의의 난수 집합 중  $i$ 번째 값  
 $C_i$ : 상관계수집합 중  $i$ 번째 값

이렇게 만들어진 첫 세대로 유전자 알고리즘을 수행한다. 이후 세대마다 지역탐색을 진행하기 위한 본 연구의 혼합형 유전자 알고리즘은 <Fig. 6>과 같다. 각 연산자(선택, 교차, 변이)를 거쳐 추출된 유전자(속성별 가중치)에 상관계수를 지역탐색의 방법으로 반영함으로써 해 집단을 개선한다.



**Fig 6. Genetic algorithm with local search in this research**

세대별 지역탐색방법을 적용하는 과정은 다음과 같다. <Fig. 6>과 같이 적합도산정 및 선택, 교차, 변이 등의 연산자를 거치게 되면 한 세대를 거친 가중치 세트가 결과물로서 생성된다. 그 생성된 가중치 세트에 지역탐색을 적용하는 것인데 그 순서는 먼저, 데이터의 각 속성의 상관계수를 추출하고, 세대에서 생성된 속성의 가중치 값에 속성의 상관계수 값을 곱해서 지역탐색을 적용한다. 이후 다음 세대로 진화하고 같은 과정을 반복함으로써 최적의 가중치를 생성한다.

상관계수를 세대별 가중치에 반영하기 위한 계산식은 식 (2)와 같다.

$$P_i = X_i * C_i \tag{2}$$

$P_i$ : 새로운 모집단의  $i$ 번째 값  
 $X_i$ : 이전 세대 유전자집합 중  $i$ 번째 값  
 $C_i$ : 상관계수집합 중  $i$ 번째 값

초기화된 모집단을 기반으로 유전자 알고리즘이 실행되며 같은 방식으로 매 세대에 각 속성의 상관계수를 반영해 진화하고 해가 일정 범위 내로 수렴할 때까지 세대를 진화시킨다. 모집단과 각 세대는 데이터 속성 상관계수를 적용함으로써 프로젝트 속성정보들을 반영할 수 있게 개선된다. 이렇게 개선된 혼합형 유전자 알고리즘을 이용하여 가중치를 산정하고 사례기반추론 공사비에측모델에 적용한다.

### 4. 공사비에측모델 개발

본 연구의 공사비에측 모델은 사례기반추론을 이용하여 데이터 세트에서 예측대상과 유사한 과거 사례를 추출하고 해당 사례의 공사비 정보를 이용하여 공사비를 예측하는 과정을 나타낸 것이다. 본 연구에 사용한 사례기반추론에서 유사사례를 조회하는데 K-최근접 이웃 조회 방법(K-nearest neighbor, KNN)을 이용했다. KNN은 새로운 데이터가 주어졌을 때 기존 데이터 가운데 가장 가까운 k개 이웃의 정보로 새로운 데이터를 예측하는 방법론이다(Park et al., 2010). 공사비에측 모델 개발의 프로세스는 다음과 같다. 첫째, 상관계수를 적용한 혼합형 유전자 알고리즘으로 가중치를 산정한다. 둘째, 유사도와 속성 가중치를 결합해 사례 유사도를 산출해 공사비에측모델을 개발한다.

#### 4.1 가중치 산정

사례의 유사한 정도를 판단하려면 사례 간 차이의 정도와 속성 가중치를 결정하여야 한다. 3장에서 기술한 바와 같이

본 연구에서는 최적화된 가중치 산정을 위해 공사비 데이터의 각 속성 상관계수를 적용해 혼합형 유전자 알고리즘을 사용한다. 혼합형 유전자 알고리즘의 연산자(적자생존, 선택, 교차, 변이)의 비율은 각각 5%, 40%, 50%, 5%로 적용했고, 세대는 100세대를 반복해 알고리즘을 수행한다.

#### 4.2 사례 유사도

본 연구에 사용된 데이터는 정성적 데이터와 정량적 데이터로 구분하고, 속성 간의 유사한 정도를 정량적으로 판단하기 위해 유클리디안 거리(Euclidean Distance) 측정방식에 기반한 사례의 거리를 측정하는 방법을 이용했다(Ji et al., 2008). 유클리디안 거리측정 방식은 인공지능 분야에서 두 가지 개체의 속성값들이 여러 개일 경우 이 속성값들에 의한 두 개체 사이의 유사도를 구할 때 자주 사용한다. 전술한 바와 같이 사례 유사도의 점수는 속성 가중치와 이 거리공식을 이용해 산출한 각 속성 간 유사도를 곱하여 더한 값을 속성 가중치 합으로 나누어 구하게 되며 사용된 식은 식 (3)과 같다.

$$\text{사례 } i \text{의 유사도}(x_i, x_j) = \left[ 1 - \frac{\sum_{r=1}^n w_r^2 a_r(x_i) - a_r(x_j)}{\sum_{r=1}^n w_r^2} \right] \quad (3)$$

여기서,  $w_r$ : r번째 속성의 가중치  
 $a_r(x_i)$ : i번째 사례의 r번째 속성값  
 $a_r(x_j)$ : 예측하고자 하는 사례의 r번째 속성값

Table 5. Weight set and error rate according to method

Methodology	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	Error rate(%)	
Case1 (APT)	Hybrid GA-CBR	0.1051	0.8741	0.0039	0.00032	0.0049	0.0006	0.0005	0.0008	2.571 E-08	0.0004	0.0004	0.0086	4.73
	Existing GA-CBR	0.1254	0.1479	0.0412	0.01447	0.1420	0.0079	0.1417	0.1325	0.0077	0.0420	0.1335	0.0631	8.25
	Uniform Weight	0.0833	0.0833	0.0833	0.0833	0.0833	0.0833	0.0833	0.0833	0.0833	0.0833	0.0833	0.0833	11.18
	Regression Analysis	0.002	0.0005	0.0903	0.2543	0.0742	0.0240	0.080	0.0483	0.0237	0.0829	0.0971	0.2210	8.76
Case2 (Military)	Hybrid GA-CBR	0.0024	0.0003	0.9245	0.0726	2.607 E-15	1.785 E-08	2.600 E-07	4.485 E-06					8.72
	Existing GA-CBR	0.1632	0.1540	0.1811	0.1931	0.0995	0.0712	0.0289	0.1091					14.87
	Uniform Weight	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125					19.24
	Regression Analysis	0.0014	0.3112	0.0014	0.0015	0.216	0.1917	0.0721	0.204					9.03
Case3 (Office)	Hybrid GA-CBR	0.0717	0.1146	2.865 E-06	0.0004	0.1116	0.2454	0.4309	0.0149	0.0004				7.67
	Existing GA-CBR	0.0773	0.2155	0.0759	0.0732	0.2291	0.0739	0.1167	0.1379	0.0005 377				8.00
	Uniform Weight	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.111				10.94
	Regression Analysis	2.99 E-06	5.42 E-06	8.53 E-05	0.0895	0.0254	0.3710	0.4305	0.0214	0.0618				8.82

각 사례의 유사도를 측정한 후 점수가 높은 순으로 단순 혹은 복수의 유사사례를 추출하게 된다. 복수의 사례를 추출하는 경우에는 상위의 유사도 점수순으로 일정 개수의 사례를 선택할 수 있다. 본 연구에서는 상위점수 3개의 복수의 유사사례를 추출해 공사비예측모델을 학습한다.

## 5. 모델 검증

### 5.1 검증방법

본 연구의 공사비예측 모델의 검증을 위해 수집한 3개의 사례 데이터 중 모델 개발에 이용한 70%의 데이터를 제외하고 총 데이터 개수 중 30%의 데이터를 사용하여 모델을 검증하였다. 공사비예측 모델의 성능은 검증 대상의 실제 공사비와 예측공사비의 차이를 실제 공사비로 나누어 구한 오차율로 나타내었다. 또한, 해당 모델의 유효성을 판단하기 위해 본 연구의 혼합형 유전자 알고리즘, 기존 유전자 알고리즘, 동일가중치법, 회귀분석법으로 수행한 각 공사비 예측모델의 오차율을 비교하였다.

### 5.2 검증 결과 및 분석

〈Table 5〉는 사례 1, 사례 2, 사례 3의 각 방법론으로 가중치 산정의 결과인 속성 가중치 값과 그에 따라 사례기반추

론으로 공사비예측을 수행한 사례의 평균 오차율이다.

사례 1에서의 본 연구의 방법론을 적용했을 때, 최적화된 속성 중 X2(연면적)에 대한 가중치가 약 0.87로 가장 높은 비중을 나타냈다. 최적화 가중치로 사례기반추론을 이용해 유사사례를 추출하고 공사비예측을 진행한 결과 각 사례의 평균 오차율 4.73%가 도출 되었다. 같은 방법으로 사례 2의 경우도 X3(연면적)의 가중치 값이 약 0.92로 가장 높은 값을 나타냈고, 공사비예측을 진행한 결과 8.72의 평균 오차율을 나타내었다. 사례 3의 경우 X7(구조형식:SRC)의 속성 가중치 값이 약 0.43으로 가장 높았고 평균 오차율은 7.67%가 도출되었다.

이를 AACE (American Association of Cost Engineers)에서 프로젝트 견적단계를 해당 프로젝트가 가지고 있는 정보량에 따라 5단계로 분류하여 정의한 예상 견적 정확도와 비교하였을 때 프로젝트가 10%~40% 진행되었을 경우, AACE의 예상 견적 정확도는 과소측정 시 -10% ~ -20% / 과대측정 시 +10% ~ +30%에 해당하기 때문에, 본 연구의 공사비예측 사례별 모델의 예측 정확도는 AACE의 기준보다 우수한 오차율을 보였다(AACE, 1997).

또한 본 연구의 가중치 산정방법에 대한 유효성 검토를 위해 혼합형 GA-CBR 뿐만 아니라 〈Table 6〉과 같이 기존 GA-CBR, 동일가중치법, 회귀분석법 등 네 가지 방법

Table 6. Each case the comparison of actual construction cost and estimating models

Case	Actual Cost (천원)	Hybrid GA-CBR		Existing GA-CBR		Uniform Weight		Regression Analysis	
		Estimated Cost	Error (%)	Estimated Cost	Error (%)	Estimated Cost	Error (%)	Estimated Cost	Error (%)
A1	4,707,437	4,031,304	14.36	3,721,210	20.95	3,853,584	18.14	4,319,676	8.24
A2	2,889,350	2,758,302	4.54	3,415,602	18.21	3,800,988	31.55	2,677,657	7.33
A3	3,047,139	3,141,282	3.09	4,331,251	42.14	4,331,251	42.14	3,003,531	1.43
A4	3,504,668	3,570,193	1.87	3,476,952	0.79	3,402,633	2.91	3,864,799	10.28
A5	3,507,199	3,570,193	1.80	3,476,952	0.86	3,402,633	2.98	3,864,799	10.20
A6	3,418,988	3,283,438	3.96	3,476,952	1.70	3,615,971	5.76	4,148,977	21.35
A7	2,735,539	2,758,302	0.83	3,541,578	29.47	3,541,578	29.47	2,177,965	20.38
Error mean		4.73		8.25		11.18		8.76	
M1	2219,882	1,960,750	11.67	2,129,966	4.05	2,121,152	4.45	2,023,655	8.84
M2	977,369	995,926	1.90	1,009,753	3.31	995,926	1.90	867,252	11.27
M3	774,032	818,789	5.78	719,388	7.06	657,058	15.11	719,388	7.06
M4	1,053,153	1,097,053	4.17	1,013,964	3.72	995,926	5.43	867,252	17.65
M5	582,516	589,644	1.22	579,274	0.56	579,274	0.56	579,274	0.56
M6	609,386	602,510	1.13	601,984	1.21	588,231	3.47	668,621	9.72
M7	571,788	589,644	3.12	579,274	1.31	579,274	1.31	579,274	1.31
Error mean		8.72		14.87		19.24		9.03	
O1	552,285	655,661	18.72	595,408	7.81	599,447	8.54	686,552	24.31
O2	763,098	561,256	26.45	571,724	25.08	561,642	26.40	574,050	24.77
O3	510,173	483,344	5.26	530,690	4.02	672,515	31.82	621,040	21.73
O4	570,691	520,654	8.77	540,492	5.29	520,654	8.77	520,654	8.77
O5	455,386	534,179	17.30	471,901	3.63	492,742	8.20	514,745	13.04
O6	492,610	574,050	16.53	580,446	17.83	608,799	23.59	574,050	16.53
O7	627,670	688,703	9.72	688,703	9.72	631,038	0.54	688,703	9.72
Error mean		7.67		8.00		10.94		8.82	

A : APT case M : Military case, O : Office case



의 예측모델을 실제 공사비와 비교했다. 그 결과 공사비에 예측 모델의 평균 오차율이 혼합형 GA-CBR / 기존 GA-CBR / 동일가중치법 / 회귀분석법 순서로 사례 1(APT)에서는 4.73 / 8.25 / 11.18 / 8.76의 오차율(%)을 나타냈고, 사례 2(Military)에서는 8.72 / 14.87 / 19.24 / 9.03의 오차율(%), 그리고 사례 3(Office)에서는 7.67 / 8.00 / 10.94 / 8.82의 오차율(%)을 나타냈다. 이를 통해 본 연구의 예측모델이 다른 가중치 산정방법을 사용한 예측모델보다 높은 정확도를 나타낸다고 판단할 수 있다.

## 6. 결론

사례기반추론 모델의 정확도는 각 속성의 가중치 할당에 많은 영향을 받는다. 기존 GA-CBR 공사비예측 모델에서는 가중치 산정 시 유전자 알고리즘 내에 임의의 난수와 연산자를 사용했다. 하지만 이 방법은 가중치 산정 시 현재 해보다 좋은 해를 발견하기 어렵다는 단점을 가진다. 따라서 본 연구는 유전자 알고리즘을 활용한 속성 가중치 산정 프로세스에 있어 상관계수를 활용한 지역탐색 방법을 제시하고 결합함으로써 프로세스를 개선하였다. 이후 실제 데이터를 기반으로 개선한 혼합형 유전자 알고리즘을 통해 속성 가중치를 산정하고 사례기반추론 공사비예측 모델을 개발하였다. 모델의 유효성 검증 결과 일반적인 GA-CBR, 동일가중치법, 회귀분석법과 같은 기존 모델보다 좋은 성능을 나타내었고, 이는 AACE의 견적 예측 정확도보다 우수한 정확도로 판단되었다.

본 연구는 기존 사례기반추론 관련 연구들이 가장 기본적인 유전 알고리즘을 활용했던 것에 비해, 지역탐색을 결합한 혼합형 유전자 알고리즘을 제시함으로써 그 성능을 향상했다는 것에 의의가 있다. 더불어 지역탐색 과정에 각 속성과 공사비의 상관계수를 활용하여 최적 가중치 산정에 영향을 주도록 함으로써 기존 도메인 지식이 반영되도록 하였다.

본 연구는 지역탐색 방법을 유전자 알고리즘과 함께 활용함으로써 가중치 산정 시 건설데이터 속성들의 특징을 반영하여 정확하고 설명력이 있는 해를 산정할 수 있지만, 제시한 지역탐색 방법의 특성으로 인해 수렴의 속도가 빨라 지역최적해에 빠질 수 있다는 한계를 가지고 있다. 향후 이러한 한계 보완을 위해 다른 지역탐색 방법을 개발하거나 타 모델을 함께 활용하는 연구가 필요하다.

## 감사의 글

본 연구는 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(과제번호:10077606).

## References

- AACE Recommended Practice No. 17R-97 (1997). Cost Estimate Classification System, AACE, Inc.
- Aamodt, A., and Plaza, E. (1994). "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches." *AI Communications*, 7(1), pp. 35-39.
- An, S.H., and Kang, J.I. (2005). "A Study on Predicting Construction Cost of Apartment Housing Using Experts' Knowledge at the Early Stage of Projects." *JOURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Structure & Construction*, 21(6), pp. 81-88.
- Bruno, De Backer., Vincent, Furnon, Philip, Kilby, Patrick Prosser, and Paul, Shaw. (1997). "Local Search in Constraint Programming: Application to the Vehicle Routing Problem." *Proc. CP-97 Workshop Indust. Constraint-Directed Scheduling*. Schloss Hagenberg Austria.
- Christensen, P., and Dysert, L. R. (1997). "Cost estimate classification system." *AACE international recommended practice 17R-97*.
- Christos, Voudouris, and Tsang, E.P.K. (1998). "Solving the radio link frequency assignment problem using guided local search." *Proceedings NATO Symposium on Radio Length Frequency Assignment, Sharing and Conservation Systems (Aerospace)*, Aalborg, Denmark.
- Chun, S.H., and Park, Y.J. (2006). "A new hybrid data mining technique using a regression case based reasoning: Application to financial forecasting." *Korea Advanced Institute of Science and Technology*, 31, pp. 329-336.
- E.K. Burke, T. Curtois, G. Post, R. Qu, and B. Veltman. (2007). "A Hybrid Heuristic Ordering and Variable Neighbourhood Search for the Nurse Rostering Problem." *European Journal of Operational Research*, 188(2), pp. 330-341.
- Goh, Y.M., and Chua, D.K.H. (2009). "Case-Based Reasoning for Construction Hazard Identification: Case Representation and Retrieval." *Journal of Construction Engineering and Management*, 135(11), pp. 1181-1189.
- Goldberg, David E., Bradley Korb, and Kalyanmoy Deb. (1989). "Messy genetic algorithms: Motivation, analysis, and first results." *Complex systems* 3.5, pp. 493-530.
- Heikki Maaranen., Kaisa Miettinen., and Antti Penttinen

- (2007). "On initial populations of a genetic algorithm for continuous optimization problems." *Journal of Global Optimization*, 37(3), p 405.
- Hwang, J.H., and Kim, S.Y. (2010). "Integer Programming-based Local Search Technique for Linear Constraint Satisfaction Optimization Problem." *Journal of the Korea Society of Computer and Information*, 15(9), pp. 47-55.
- Hwang, J.H. (2010). "An Integration of Local Search and Constraint Programming for Solving Constraint Satisfaction Optimization Problems." *Journal of the Korea Society of Computer and Information*, 15(5), pp. 39-47.
- Ji, S.H., Park, M.S., Lee, H.S., Seong, K.H., and Yoon, Y.S. (2008). "Method of Quantity Data Analysis for Building Construction Cost Estimation : Focusing on Finish Work of Public Apartment Project." *Korean Journal of Construction Engineering and Management*, KICEM, 9(6), pp. 235-243.
- Kang, M.G., Park, S.W., Im, S.J., and Kim, H.J. (2002). "Parameter Calibrations of a Daily Rainfall-Runoff Model Using Global Optimization Methods." *Journal of Korea Water Resources Association*, 35(5), pp. 541-552.
- Kim, G.H., An, S.H., and Cho, H.K. (2006). "Comparison of the Accuracy between Cost Prediction Models based on Neural Network and Genetic Algorithm - Focused on Apartment Housing Project Cost." *JOURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Structure & Construction*, 22(3), pp. 111-118.
- Kim, G.H., and Kang, K.I. (2003). "A Study on Model of Neural Networks Training by Genetic Algorithms for Predicting Cost Estimates of Apartment Projects at the Early Project Stage." *JOURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Structure & Construction*, 19(10), pp. 133-142.
- Kim, G.H., and Kang, K.I. (2004). "A Study on Predicting Construction Cost of Apartment Housing Projects Based on Case Based Reasoning Technique at the Early Project Stage." *JOURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Structure & Construction*, 20(5), pp. 83-92.
- Kim, H.H., and Choi, J.Y. (2009). "An Efficient Search Algorithm for Flexible Manufacturing Systems (FMS) Scheduling Problem with Finite Capacity." *IE interfaces* 22(1), pp. 10-16.
- Kim, S.G., Lee, U.K., Cho, H.H., Kang, K.I. (2006). "Decision Support System for Slab Form-work Selection of High-rise Building Construction." *JOURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Structure & Construction* 22(11), pp. 207-214.
- Kolodner, J. (1993). *Case-Based Reasoning*. Morgan Kaufmann Publishers Inc.
- Lau, T.L., and Tsang, E.P.K. (1997). "Solving the Processor Configuration Problems with a Mutation-Based Genetic Algorithm." *International Journal on Artificial Intelligence Tools*, 6.04, pp. 567-585.
- Lee, H.S., Kim, Euntai, and Kim, D. (2005). "Pattern Recognition System Combining KNN rules and New Feature Weighting algorithm." *The Institute of Electronics Engineers of Korea - Computer and Information*, 42(4), pp. 43-50.
- Lee, H.S., Kim, S.Y., Park, M.S., Ji, S.H., Seong, K.H., and Pyeon, J.H. (2011). "A Method of Assigning Weight Values for Qualitative Attributes in CBR Cost Model." *Korean Journal of Construction Engineering and Management*, 12(1), pp. 53-61.
- Lee, J.H. (2008). "A Study on the application of Case-based Reasoning with Feature Weighting using Self-Organizing Map." *Korean Journal of Business Administration*, 21(1), pp. 417-437.
- Oh, I.S., Lee, J.S., and Moon, B.R. (2004). "Hybrid genetic algorithms for feature selection." *IEEE Tr. Pattern Analysis and Machine Intelligence*, 26(11), pp. 1424-1437.
- Paredis, J. (1993). "Genetic State-Space Search for Constrained Optimization Problems." *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pp. 967-972.
- Park, M.S., Seong, K.H., Lee, H.S., Ji, S.H., and Kim, S.Y. (2010). "Schematic Cost Estimation Method using Case-Based Reasoning: Focusing on Determining Attribute Weight." *Korean Journal of Construction Engineering and Management*, KICEM, 11(4), pp. 22-31.
- Park, U.Y., and Kim, G.H. (2007) "A Study on Predicting Construction Cost of Apartment Housing Projects Based on Support Vector Regression at the Early Project Stage." *JOURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Structure & Construction*, 23(4), pp. 165-172.

- Qu, R., and He, F. (2008). "A Hybrid Constraint Programming Approach for Nurse Rostering Problems." *Applications and Innovations in Intelligent Systems XVI : Proceedings of AI-2008*, pp. 211-224.
- Sevgi Zeynep Doğan., and David Arditi., and H. Murat Günaydın. (2006). "Determining Attribute Weights in a CBR Model for Early Cost Prediction of Structural Systems." *Journal of Construction Engineering and Management*, 132(10), pp. 1092-1098.
- Song, K.R., Jung, E.K., and Im, C.S. (2006). "A Study on the Correlation Analysis of the Item of Expenditure using ISM Method." *JOURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Structure & Construction*, 5, pp. 153-160.

---

**요약 :** 건설프로젝트 초기 단계의 사업비 추정치는 사업지의 결정이나 공사 기간과 같은 중요한 사항에서 사업자의 의사결정에 상당한 영향을 미친다. 그러나 초기 단계는 설계도서나 시방서 등의 정보가 부족한 채로 진행되기 때문에 신뢰도 있게 수행하기 어렵다. 기존 연구에서는 초기 공사비를 예측하기 위해 사례기반추론을 사용했으며, 그 중 조회 단계의 가중치를 계산하는 방법으로 유전자 알고리즘을 사용했다. 그러나 기존 유전자 알고리즘은 임의의 수로 연산하기 때문에 현재해보다 좋은 성능을 내기 힘들다는 한계가 있다. 이러한 한계를 극복하기 위해 임의의 수가 아닌 상관분석을 이용한 상관계수를 지역탐색의 방법으로 유전자 알고리즘에 반영하고, 지역탐색과 유전자 알고리즘을 결합한 혼합형 유전자 알고리즘으로 가중치를 산정한다. 산정한 가중치를 사용하여 사례기반추론 모델을 개발하고 데이터를 통해 유효성을 검증하였다. 그 결과, 지역탐색을 적용한 혼합형 GA-CBR이 기존 GA-CBR보다 우수한 성능을 보인 것으로 확인되었다.

**키워드 :** 공사비 예측, 사례기반추론, 혼합형 유전자 알고리즘, 지역탐색, 상관분석

---