

사례기반추론을 이용한 초기단계 공사비 예측 방법 : 속성 가중치 산정을 중심으로

Schematic Cost Estimation Method using Case-Based Reasoning
: Focusing on Determining Attribute Weight

박문서* 성기훈** 이현수*** 지세현**** 김수영*****
Park, Moonseo Seong, Kihoon Lee, Hyun-soo Ji, Sae-Hyun Kim, Sooyoung

요약

프로젝트 초기단계에서 산정된 공사비는 발주자의 중요한 의사결정에 영향을 미치므로 그 중요성이 강조되고 있지만, 정보의 부족으로 인하여 주로 견적전문가의 경험과 지식에 의존하여 진행된다. 이것은 현재 문제와 가장 유사한 과거 사례를 선택하여 사용하는 사례기반추론으로 발전되었다. 사례기반추론 모델의 예측 성능은 속성 가중치의 산정 결과에 많은 영향을 받으므로, 정확한 속성 가중치의 산정이 요구된다. 기존의 연구는 수학적 방법 또는 전문가의 주관적 판단을 이용하는 방법을 사용한다. 본 연구는 기존 연구의 문제점을 보완하기 위해 유전자 알고리즘을 이용한 사례기반추론 공사비 예측 모델을 제안한다. 공사비 예측 모델은 최근이웃 조희 방법의 과정에 의해 추출한 사례의 공사비 정보를 이용하여 예측 대상의 공사비를 산정한다. 검증 결과, AACE에서 정의한 견적시기별 예측 정확도와 표준화 회귀계수, 동일가중치를 사용한 방법보다 높은 오차율을 나타내었다. 따라서 본 연구는 유전자 알고리즘을 도입하여 예측 성능을 향상시키고, 사례기반추론 방법을 사용하여 사용자가 이해하기 용이한 해결책 도출과정을 제시하였는데 그 의미가 있다.

키워드 : 공사비 예측, 사례기반추론, 속성 가중치, 유전자 알고리즘

1. 서론

1.1 연구의 배경 및 목적

프로젝트 초기단계의 공사비 예측은 프로젝트 예산, 규모 및 공사기간과 같은 주요사항에 대한 발주자의 의사결정에 영향을 미치므로, 그 중요성이 강조되고 있다. 하지만, 초기단계에서의 공사비 산정은 도면과 시방서 등 설계도서가 확정되지 않아 정보가 부족한 상태에서 주로 전문가들의 지식과 경험에 의존하여 진행되며(안성훈 외 2인 2005), 견적전문가들은 과거 사례를 기반으로 한 공사비 예측방법을 선호하는 경향이 있다.

이러한 과거의 경험과 지식을 기반으로 현재 문제와 가장 유사한 과거 사례를 선택하여 사용하는 방법은 사례기반추론(Case-Based Reasoning, CBR)이라는 기계학습(Machine Learning)으로 발전되었다. 사례기반추론은 인간이 문제를 해결하는 과정에서 과거 문제의 해결책을 현재 문제의 해결에 적용하는 것을 모방한 방법으로, 공사비 예측 과정을 견적전문가에게 보다 쉽게 이해시킬 수 있으며, 해결책의 신뢰도를 향상시킬 수 있다.

사례기반의 각 사례는 결과값의 특성을 설명하는 몇 개의 속성으로 표현되며, 유사 사례의 추출을 위해서는 속성 유사도와 가중치의 정의가 필요하다. 특히, 사례기반추론 모델의 성능은

* 종신회원, 서울대학교 건축학과 교수, 공학박사, mspark@snu.ac.kr

** 일반회원, 서울대학교 대학원, 석사과정(교신저자), emperorskh@hotmail.com

*** 종신회원, 서울대학교 건축학과 교수, 공학박사, hyunslee@snu.ac.kr

**** 일반회원, 서울대학교 대학원, 박사과정, oldclock@snu.ac.kr

***** 일반회원, 서울대학교 대학원, 석사과정, finalwing@naver.com

속성 가중치의 산정 결과에 많은 영향을 받으므로, 신뢰성 있는 사례기반추론 모델 구축을 위해서는 정확한 속성 가중치의 산정이 요구된다(Dogan et al 2008).

기존의 연구는 속성 가중치 산정을 위해 경사하강법(Yau and Yang 1998; 김광희와 강경인 2004; 김광희 외 2인 2004), 회귀 분석법(김광희 외 2인 2004)과 같은 수학적 방법과 전문가 지식을 활용한 계층분석과정(안성훈과 강경인 2005) 등의 방법을 이용한다. 하지만, 이들의 방법은 다수의 입출력 변수를 최적화하는 과정에서 수학적 오류를 발생시키거나, 전문가의 주관적 판단에 따라 속성 가중치가 변화되는 단점이 있다.

따라서 본 연구는 기존의 속성 가중치 산정 방법의 한계를 보완하기 위하여 최적화 문제를 해결하는 기법인 유전자 알고리즘(Genetic Algorithm, GA)을 이용한 사례기반추론 공사비 예측 모델을 제안하고, 모델의 유효성을 검증하는 것을 목적으로 한다.

1.2 연구의 범위 및 방법

건설 프로젝트의 시공 전 단계는 기획, 계획설계, 기본설계, 실시설계로 구분할 수 있다. 본 연구의 공사비 예측 모델은 기본설계/실시설계에 비하여 상대적으로 정보가 부족한 기획/계획설계와 같은 초기단계를 대상으로 한다. 또한, 모델 구축과 검증은 국내 공공아파트 데이터를 사용하여 실행한다. 연구는 다음과 같은 절차에 따라 진행된다.

첫째, 선행연구 분석을 통하여 시사점을 도출하고, 사례기반추론과 유전자 알고리즘의 원리와 적용을 고찰한다.

둘째, 공공아파트 건축부문의 초기단계 공사비 예측을 위하여 사례기반추론과 유전자 알고리즘을 요소기술로 하는 공사비 예측 모델을 구축한다.

셋째, 발전 모델의 정확성을 검증하여 본 연구에서 제시한 공사비 예측 방법의 유효성을 확인한다.

2. 예비적 고찰

2.1 사례기반추론(Case-Based Reasoning)

2.1.1 사례기반추론 개요

Aamodt와 Plaza(1994)는 사례기반추론이란 새로운 문제를 해결하기 위하여 과거의 유사한 사례를 저장하고, 그 사례의 정보와 지식을 재사용하는 것이라고 정의한다.

사례기반추론이 다른 인공지능 기법과 구별되는 특징은 첫째로, 새로운 문제해결을 위해 과거사례의 구체적 지식을 사용한다는 점이다. 둘째로, 해결된 새로운 문제는 사례기반의 과거사

례로 저장되기 때문에, 미래에 발생할 문제 해결에 사용될 수 있다는 점이다(Aamodt and Plaza 1994).

사례기반추론은 문제 해결을 위해 그림1과 같이 조회(Retrieve), 재사용(Reuse), 수정(Revise), 저장(Retain)의 4단계 과정을 거친다(Aamodt and Plaza 1994).

조회란 새로운 문제와 가장 일치하는 과거사례를 찾아내는 과정이며, 재사용은 조회단계에서 추출된 과거사례의 해결책을 사용하는 것이다. 수정은 필요에 따라서 과거사례의 해결책을 새로운 문제에 적절하게 변화시키는 작업이고, 저장은 새로운 문제를 해결한 후 수정되어 얻어진 해를 사례로 저장하는 것을 의미한다.

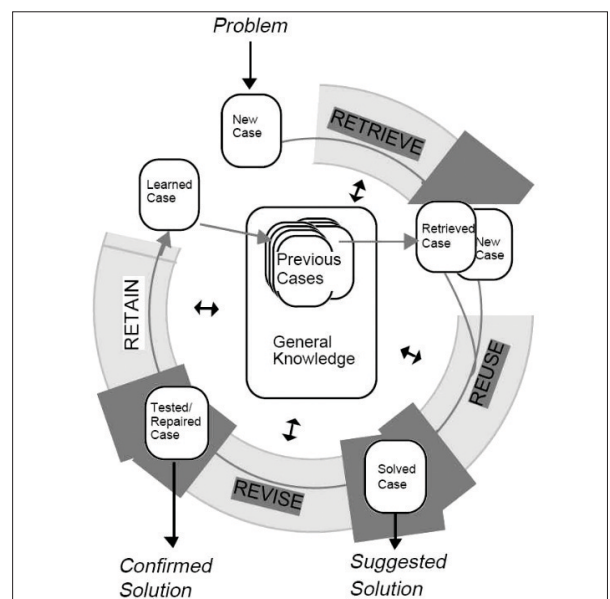


그림 1. 사례기반추론의 사이클 (Aamodt and Plaza 1994)

2.1.2 유사사례 조회 방법

사례기반추론의 과정은 순환적으로 수행되므로, 각 단계는 이전 단계 혹은 다음 단계에 영향을 미치게 된다. 따라서 사례기반추론을 사용해 문제 해결을 하기 위해서는 첫 번째 단계인 과거 유사사례의 조회가 성공적으로 이루어져야 한다. 사례기반추론의 대표적인 유사사례 조회 방법으로 귀납적 조회 방법(Inductive Retrieval)과 최근이웃 조회 방법(Nearest Neighbor Retrieval)이 일반적으로 사용된다.

귀납적 조회 방법은 유사사례 조회의 목표가 잘 정의될 때 효과적인 방법으로, 새로운 사례가 과거 사례 자체에서 귀납된 결과에 대한 주요 영향요인에 의해 색인화 되는 방법이다(이지은 1998). 이 방법은 수집된 사례들로부터 일반화된 규칙 또는 수식으로 표현된 분류 방법을 미리 도출한 후, 새로운 사례가 입력되

면 정의된 분류 방법을 적용하여 유사한 과거 사례를 찾게 된다.

최근이웃 조회 방법은 저장되어 있는 사례들 중에서 일정한 유사성 척도에 의해 새로운 사례와 과거 사례를 대응시켜 유사 사례를 조회하는 방법이다(김광희 외 2인 2004). 그림2는 최근이웃 조회 방법의 알고리즘을 나타낸 것이다. 첫 번째 과정은 조회 대상에 해당하는 새로운 문제와 과거 사례의 일치하는 속성(Attribute)을 비교하는 것이다. 이어서 속성값을 비교할 수 있는 규칙 또는 수식을 이용하여 과거 사례의 속성 유사도(Attribute Similarity) 점수를 계산하고, 과거 사례 분석을 통해 속성 가중치(Attribute Weight)를 결정한다. 마지막으로 모든 속성들의 유사도 점수와 가중치를 곱한 값을 합계하여 사례 유사도(Case Similarity) 점수를 계산하여, 높은 점수를 획득한 사례를 새로운 문제의 유사사례로 선택한다.

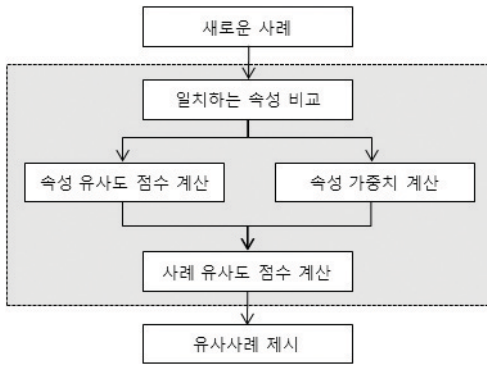


그림 2. 최근이웃 조회 방법 알고리즘

귀납적 조회 방법은 사례기반에 현재 사례와 일치하는 사례가 존재하지 않을 경우, 유사한 과거 사례를 조회하지 못하는 단점이 있다. 하지만, 최근이웃 조회 방법은 사례가 부족하거나 현재 사례와 동일한 과거 사례가 존재하지 않아도 유사한 과거 사례를 조회할 수 있다. 그리고 최근이웃 조회 방법은 숫자와 같은 구체적이고 명확한 정보를 이용하여 과거 사례를 조회하는데 적합한 방법이다.

공사비 예측 대상이 되는 건설 프로젝트는 단품생산 방식을 취하고 사용목적·규모·사용재료 등의 변화요인이 다수 존재하기 때문에, 현재 수행하려는 프로젝트와 동일한 과거 프로젝트 사례는 존재하지 않는 것이 일반적이다. 또한, 공사비 예측의 입력력 변수는 숫자의 형태로 표현되는 경우가 많다. 따라서 본 연구는 건설 프로젝트와 공사비 예측의 특성을 고려하여 사례기반추론의 조회 방법으로 최근이웃 조회 방법을 사용한다.

2.2 속성 가중치

속성 가중치란 유사한 사례를 탐색할 때 중요한 입력 속성에 높은 가중치를 할당하여 정확한 탐색을 유도하는 방법이다(이장희 2008). 표1에서 보여지 듯 기존의 연구는 사례기반추론의 속성가중치 산정을 위해 동일가중치, 경사하강법, 회귀분석법, 계층분석과정 등 다양한 방법을 사용한다.

표 1. 사례기반추론의 가중치 산정 방법 연구

가중치산정방법	연구내용	관련문헌
경사하강법	사례기반추론을 이용한 건설프로젝트 공기 및 공사비 예측	Yau, Yang (1998)
경사하강법	공동주택 초기단계의 사례기반추론 공사비 예측방법 제시	김광희, 강경인 (2004)
경사하강법	사례기반추론과 인공지능경험을 이용한 공동주택 공사비 예측 모델의 정확도 비교	김광희 외 2인 (2004)
계층분석과정	전문가 지식 기반의 속성가중치 산정을 통한 프로젝트 초기단계의 공사비 예측	안성훈, 강경인 (2005)
동일가중치	사례기반추론 기반의 객체지향형 예측모델을 통한 벽돌공사 생산성 결정방법	Karshenas Tse (2002)
동일가중치	철강제품의 품질설계를 위한 사례기반추론 모델 제시	김중환 외 3인 (1996)
경험적 방법	초고층건물 구조시스템 선정 위한 귀납적 방법과 유사사례 선정 위한 최근이웃 방법 제시	송화철 외 2인 (2007)

동일가중치(Feature Counting)는 모든 속성에 동일한 가중치를 부여하는 방법이며, 속성간의 중요도 차이를 반영하지 못하는 단점이 있다(김중환 외 3인 1996; Karshenas and Tse 2002).

경사하강법(Gradient Descent Method)은 함수의 구배를 사용하여 탐색의 방향을 결정짓고, 현재 위치에서 새로운 점을 반복적으로 탐색하는 방법이다. 하지만 경사하강법은 서로 다른 방향에 대한 곡률이 상이할 경우, 지역 최소점(Local Minimum)에 수렴하기 위해 불필요한 탐색과정을 수행하는 문제점이 있다(Yau and Yang 1998; 김광희와 강경인 2004; 김광희 외 2인 2004).

회귀분석법(Regression Analysis)은 회귀분석에서 도출된 표준화계수를 속성가중치로 활용하는 방법이다. 회귀분석은 종속변수와 독립변수의 관계를 추정할 때 선형성을 가정하지만, 대부분의 회귀모형이 가정을 충족시키지는 못하기 때문에 결과값의 신뢰성이 떨어진다(김태근 2006). 또한, 독립변수간 매우 강한 선형관계가 존재하는 경우 회귀계수의 통상적인 해석은 공선성(Collinearity)의 문제를 야기한다(김광희 외 2인 2004).

계층분석과정(Analytic Hierarchy Process, AHP)은 전문가의 지식과 경험이라는 주관적 요소를 근거로 하기 때문에 전문가에 따라서 속성 가중치가 변하는 문제점이 있다(안성훈과 강경인 2005).

이와 같이 속성 가중치 결정에 있어 많은 연구가 이루어지고 있으나, 기존의 방법은 지역 최소점, 비선형성, 공선성, 객관성 부족 등의 문제점을 가지고 있다.

본 연구에서는 속성 가중치 산정을 위해 기존의 연구에서 사용한 경사하강법, 회귀분석법, 계층분석과정 등의 방법을 사용하지 않고 유전자 알고리즘을 도입한다. 유전자 알고리즘은 적자생존 원리를 기반으로 시행착오를 통해 최적화된 해를 찾기 때문에 지역 최소점과 공선성 등의 문제를 발생시키지 않는다. 또한, 유전자 알고리즘은 수학적 방법론이므로 도출한 해의 객관성을 확보할 수 있다.

2.3 유전자 알고리즘(Genetic Algorithm)

2.3.1 유전자 알고리즘 개요

유전자 알고리즘은 부모의 유전자로부터 자손의 유전자를 형성하는 유성식식과 다윈의 적자생존 원리를 컴퓨터 알고리즘과 결합시킨 것이다(진강규 2004).

최적화 문제란 존재하는 수많은 해 중에서 가장 좋은 해를 찾는 것인데, 유전자 알고리즘은 최적화 문제를 해결하기 위해 사용된다. 이 방법은 많은 입출력 변수를 다루기 때문에 결정론적인 방법으로 최적화된 해를 구하지 못하는 경우에 사용된다(문병로 2008). 특히, 입력 변수와 출력 변수간 강한 상관성이 존재하고 최적화할 목표값이 함수식의 가정과 연관된 경우에 적합하다.

유전자 알고리즘이 기존의 방식과 비교하여 가장 두드러진 차이점은 알고리즘이 다루는 해의 개수에 있다. 기존의 최적해 탐색 알고리즘은 하나의 초기 해를 생성한 후 적절한 연산을 사용하여 이것을 점차적으로 개선해 나간다. 하지만 유전자 알고리즘은 초기 해를 여러 개 생성하여 해 집단을 구성하고 개선해 나가는 방식을 사용한다. 생물종이 많은 개체로 구성되며 개체들이 교차(Crossover)와 변이(Mutation)를 통해 세대를 거치면서 조금씩 진화하는 방식과 유사하게 동작하는 것이다.

2.3.2 유전자 알고리즘 과정

유전자 알고리즘은 초기화, 재생산, 교배, 변이의 4가지 단계를 통해 해의 적합도(Fitness)를 평가하고 최적해를 구하게 된다.

초기화 단계에서는 주어진 문제의 해가 될 가능성이 있는 개체들의 해 집단이 형성된다. 초기의 해 집단은 해공간 내에 무작위로 분포되도록 선택되거나 경험적인 방법으로 선택된다.

재생산(Reproduction) 단계는 최적개체의 유전적 특성이 다음 세대에 전달되는 자연선택과 유사한 과정이다. 이 단계에서는 개체들의 강점과 약점이 적합도 항목으로 평가되며, 이들의 적합도 값에 따라 현재 집단 내의 개체들이 재생산을 위해 선택된다.

성능이 좋은 개체들은 더 많이 선택되어 복제되고, 반면에 성능이 나쁜 개체들은 집단으로부터 소멸된다.

교배(Crossover)는 재생산 과정에서 선택된 개체들이 재결합하는 단계이다. 적용된 교배확률에 따라 무작위로 두 개체를 선택하여 개체의 일부분을 서로 교환함으로써 새로운 두 개의 개체를 생성하여 개체군에 새롭게 삽입한다.

변이(Mutation)는 선택된 개체의 하나 이상의 유전정보를 임의로 변경하여 집단에 새로운 정보를 도입하는 단계이다. 그러나 이러한 변화는 매우 낮은 확률을 가지고 이행된다. 이와 같은 과정을 반복함에 따라 해 집단의 품질을 향상시킨다. 그림3는 유전자 알고리즘의 최적해 탐색 과정을 나타낸 것이다.

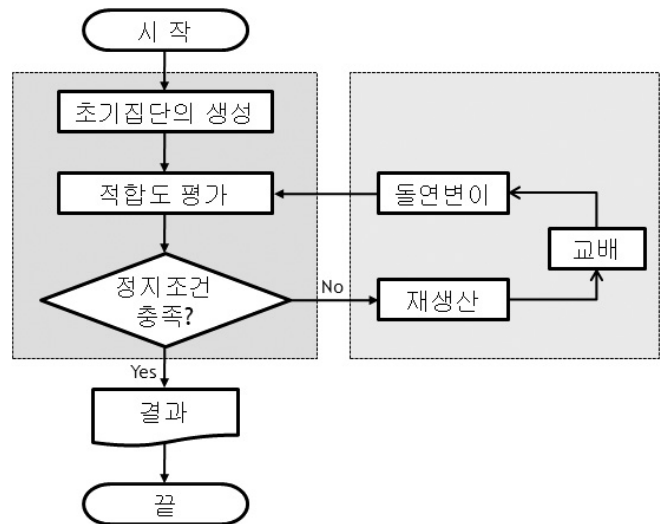


그림 3. 유전자 알고리즘 프로세스

사례기반추론의 속성 가중치 산정의 경우, 공사비 변화에 영향을 미치는 다수의 속성이 존재하고, 속성은 공사비와의 상관성이 큰 특징이 있다. 또한, 각 속성이 갖아야 할 속성 가중치의 범위가 정해져 있지 않기 때문에 수많은 해 집단이 존재한다. 이와 같은 특징으로 인해 본 연구는 사례기반추론 모델의 속성 가중치 산정을 위해 유전자 알고리즘을 사용한다.

3. 공사비 데이터 분석

본 연구는 공사비 예측 모델 구축과 검증을 위해 국내의 공공 아파트 프로젝트 데이터를 수집하고, 프로젝트 구성요소 중 아파트동의 건축부문을 분석대상으로 한다. 또한, 아파트의 다양한 평면형태, 세대조합의 변화를 반영하기 위해서 면적형별 공사비를 산출한다.

3.1 분석대상 사업개요

수집된 데이터는 10~15층 규모의 3개 지구, 14개 단지, 143개 동의 공공아파트 공사로서 49㎡, 59㎡, 84㎡, 114㎡의 4가지 연면적형으로 구성된다. 또한, 공공아파트 프로젝트의 구성요소에서 지하주차장과 부속동을 제외한 아파트동의 건축부문을 대상으로 한다. 표2는 수집된 프로젝트의 시행년도, 단지수, 동수, 연면적형별 데이터수를 나타낸 것이다.

표 2. 분석대상 사업개요

프로젝트	A지구	B지구	C지구	합계
단지	9	2	3	14
동	90	24	29	143
49㎡	16	4	3	23
59㎡	20	11	12	43
84㎡	47	12	18	77
114㎡	12	6	7	25
합계	95	33	40	168

3.2 공사비 분석 방법

분석대상 프로젝트는 다양한 평면형태, 세대조합의 변화를 보인다. 그림4와 같이 한 개의 아파트동이 두 가지 이상의 연면적형으로 이루어져 있기 때문에 수많은 연면적형 조합이 발생하게 된다. 따라서 아파트동의 동별 공사비를 산출하지 않고, 동일 아파트동에 포함되어 있더라도 연면적형을 구분하여 공사비를 산출한다(지세현 외 4인 2008). 이를 위해 식(1)과 같이 해당 아파트동의 총공사비를 각 연면적형의 연면적비로 분개하고 세대수로 나누어 단위세대 공사비를 분석한다. 이 방법은 아파트동의 가변적 연면적형 조합과 평면구성 등에 탄력적으로 대응하여 공사비를 예측할 수 있는 장점을 가진다.

$$\text{단위세대 공사비} = \frac{\text{단위건물(동) 총공사비}}{\text{연면적형 연면적비} \times \text{연면적형 세대수}} \dots \text{식 (1)}$$

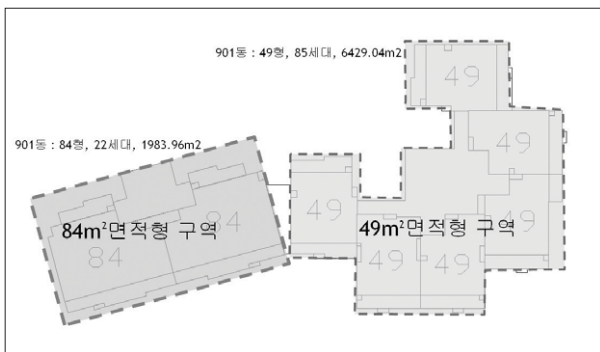


그림 4. 단위건물의 연면적형 구역별 구분 (지세현 외 4인 2008)

아파트 동별 혹은 타입별로 구분하여 공사비 정보를 데이터베이스에 입력해야 하므로, 내역서와 골조·마감 물량산출서를 사용한다. 내역서에는 1개 단지 전체에 사용된 공종별 아이템의 재료비, 노무비, 경비만 제시되어 있을 뿐 동별 공사비는 나타나 있지 않고, 골조·마감 물량산출서에는 각 동에 사용된 아이템의 물량이 나타나 있지만 공사비 정보는 제시되어 있지 않다. 따라서 골조·마감 물량산출서의 동별 아이템 총물량과 내역서의 아이템별 단가를 이용하여 모든 아이템의 비용을 합계하여 동별 공사비를 구한다.

4. 공사비 예측 모델 개발

본 연구의 공사비 예측 모델은 사례기반추론을 이용하여 데이터베이스에서 예측 대상과 유사한 과거 사례를 추출하고, 그것의 공사비 정보를 이용하여 예측 대상의 공사비를 산정하는 과정을 나타낸 것이다. 공사비 예측 모델 개발은 다음과 같은 프로세스에 의해 진행된다. 공사비 예측 모델의 범위를 정의한 후, 설계도면과 내역서를 기반으로 영향요인과 공사비를 분석하여 데이터베이스를 구축한다. 이어서 공사비 예측모델의 요소기술인 사례기반추론과 유전자 알고리즘을 사용하여 과거 유사사례를 추출한다.

4.1 공사비 예측 모델 범위 정의

본 연구는 공공아파트 기획 및 계획설계 단계의 공사비 예측 모델 개발을 목표로 한다. 공공아파트 프로젝트는 아파트 동, 지하주차장, 근린생활시설, 부대복리시설, 경비실로 구성된다. 수집 자료의 분석결과, 각 구성 건축물의 총공사비 대비 공사비 비율은 아파트동 81.2%, 관리노인정 0.97%, 근린생활시설 0.85%, 경비실 0.13%, 지하주차장 16.84%으로 나타난다. 공사비 예측 모델은 공사비 구성 비율이 가장 높은 아파트 동을 대상으로 한다.

공공아파트는 일반적으로 49㎡, 59㎡, 84㎡, 114㎡의 4가지 연면적형으로 분류할 수 있다. 수집 자료에 따르면 사용된 마감 자재 수준은 49㎡, 59㎡ 연면적형과 84㎡, 114㎡ 연면적형간 차이를 보인다. 이를 반영하기 위하여 49㎡, 59㎡와 84㎡, 114㎡의 공사비 예측 모델을 구분하여 개발한다.

4.2 데이터베이스 구축

데이터베이스는 수집된 사례의 결과값에 해당하는 공사비와 공사비 변화에 영향을 미치는 영향요인을 구분하여 구축한다.

영향요인은 데이터의 특성을 반영하는 요소로서 속성(Attribute)이라 정의한다. 기획 및 계획설계 단계는 프로젝트 초기단계에 해당하기 때문에 설계도면이 없고 스페이스 프로그램 정도의 개략 정보만 획득가능하다. 기획 및 계획설계 단계의 정보 수준을 고려하여 세대수, 연면적, 세대구성, 층수, 엘리베이터 수, 엘리베이터 1대당 세대수, 필로티세대수의 7가지 속성을 추출할 수 있다.

공사비 정보(Ci)는 3.2에서 설명한 방식으로 아파트 타입별로 구분하여 데이터베이스에 입력한다. 분석대상 프로젝트의 시행연도가 상이하므로 동일한 시점을 기준으로 공사비 변환이 필요하다. 이에 따라, 건설공사에 투입되는 자원의 직접공사비 가격변동을 측정하는 건설공사비 지수를 사용하여 공사비를 보정한다. 3.2에서 설명한 방법에 따라 분석한 공사비에 2009년 10월의 건설 공사비 지수 126.9를 적용하여 공사비를 변환한다.

다음의 표3은 각 프로젝트별 시행연도, 평균 건설공사비 지수, 보정계수를 나타낸 것이다.

표 3. 프로젝트별 건설공사비 지수

프로젝트	A지구	B지구	C지구
시행연도	2005. 6	2007. 4	2008. 2
건설공사비 지수	99.8	105.7	109.9
보정계수	1.27	1.20	1.15

3개 프로젝트를 분석하여 구축한 데이터베이스는 속성(Attribute, Xij) 정보와 공사비(Ci) 정보로 구성되어, 신규 데이터(사례)의 저장 작업을 용이하게 하고 유사 과거사례의 조회 시간을 단축한다. 아래의 표 4는 데이터베이스 일부분을 나타낸 것이다.

표 4. 데이터베이스 예시

사례	연면적형	속성 (Attribute, Xij)							공사비 (Ci)
		X1	X2	X3	X4	X5	X6	X7	
KI101	59	46	12	5065.7	4	1	4	4	3,275,446,483
KI101	59	22	12	2452.9	2	1	2	4	1,586,034,592
KI102	84	46	12	5065.7	4	1	4	4	3,250,117,033
KI103	59	20	11	1654.1	2	1	2	4	1,110,217,630
KI103	84	42	11	4629.7	4	1	4	4	3,107,397,207
KI104	59	38	10	3103.3	4	2	2	0	2,186,822,930
KI201	84	24	12	2638.6	2	1	2	0	1,798,949,424
KI202	84	48	12	5277.1	4	2	2	0	3,222,867,508
KI203	84	48	12	5277.1	4	2	2	0	3,200,369,629
KI204	59	46	12	3749.9	4	2	2	2	2,476,526,104
KI205	84	40	11	4448.1	4	1	4	4	2,918,912,107
KI206	84	48	13	5320.1	4	1	4	4	3,362,661,185
KI206	114	24	13	3426.2	2	1	2	2	2,165,620,120
KI207	84	42	11	4629.7	4	1	4	4	2,964,984,173
KI207	114	22	12	3144.0	2	1	2	2	2,013,463,343
KI208	84	44	11	4840.0	4	2	2	0	2,912,696,483

X1 = 세대수, X2 = 층수, X3 = 연면적, X4 = 세대구성
 X5 = 엘리베이터수, X6 = 엘리베이터 1대당 세대수, X7 = 필로티세대수

4.3 유사사례 추출

조회 과정은 사례기반추론의 첫 번째 단계로써, 이의 결과는 이어지는 과정에 지속적인 영향을 미치므로 현재 문제와 유사한 과거 사례의 추출이 요구된다. 본 연구는 거리 개념을 도입하여 속성 유사도 점수를 구하고, 유전자 알고리즘을 사용하여 속성 가중치를 결정한다. 정의된 속성 유사도 점수와 속성 가중치를 가중합계하여 사례 유사도 점수를 산정하여, 이를 유사사례 추출의 기준으로 한다.

4.3.1 속성 유사도(Attribute Similarity)

데이터는 정성적 데이터와 정량적 데이터로 구분한다. 본 연구에 사용된 데이터는 정량적 데이터로써, 적합한 수식을 이용하여 속성 유사도 점수를 측정할 수 있다. 아래의 식(2)는 거리 측정의 원리를 이용하여 속성유사도 점수를 구한 것이다. 식(2)는 데이터베이스에 저장된 속성 j의 최대 차이에 대한 조회사례 속성과 과거사례 속성의 차이를 구함으로써 속성간의 상대적 거리를 측정한다. 속성 유사도 점수는 조회대상인 현재사례 혹은 비교대상인 과거사례가 바뀌면 변화하는 상대적인 값이다. 반면, 속성j의 최대값과 최소값은 데이터베이스 사례가 추가 혹은 삭제되지 않은 이상 변화하지 않는다.

$$AS(j) = 1 - \frac{\sqrt{(X_{Nj} - X_{Cj})^2}}{\sqrt{(X_{Hj} - X_{Lj})^2}} \dots \text{식 (2)}$$

- 여기서, AS(j) : 속성의 유사도
- X_{Nj} : 조회하는 현재사례의 속성 j 값
- X_{Cj} : 비교하는 과거사례의 속성 j 값
- X_H : 속성 j 최대값 (모든 데이터 대상)
- X_L : 속성 j 최소값 (모든 데이터 대상)

4.3.2 속성 가중치(Attribute Weight)

사례 유사도를 판단하기 위해서는 속성 유사도와 함께 속성 가중치를 결정하여야 한다. 본 연구는 기계학습방법 중 유전자 알고리즘을 사용하여 속성 가중치를 결정한다.

공사비는 추출한 7개 속성값과 계수의 곱으로 표현 가능하다. 계수는 각 속성값이 공사비 변화에 미치는 영향을 나타내므로 속성 가중치에 해당한다. 따라서 사례 i의 공사비에 대하여 식(3)과 같이 표현한다. 공사비가 속성값과 속성 가중치 곱의 합과 완벽히 일치하지 않지만, 실제 공사비와 가장 유사한 값을 갖게 하는 속성 가중치 W_j를 결정하는 것을 목표로 한다.

$$C_i = X_{i1} \times W_1 + X_{i2} \times W_2 + \dots + X_{ij} \times W_j \dots \text{식 (3)}$$

- 여기서, C_i : 사례의 공사비
- X_{ij} : 사례의 속성의 값
- W_j : 속성의 가중치 (모든 사례 동일 적용)

모든 사례를 만족시키는 일반해 W_j 를 구하기 위해 사례1부터 사례*i*의 공사비를 행렬식으로 나타내면 식(4)와 같이 행렬 형태로 표현 가능하다.

$$\begin{pmatrix} X_{11} & \dots & X_{1j} \\ \vdots & \ddots & \vdots \\ X_{i1} & \dots & X_{ij} \end{pmatrix} \times \begin{pmatrix} W_1 \\ \vdots \\ W_j \end{pmatrix} = \begin{pmatrix} C_1 \\ \vdots \\ C_i \end{pmatrix} \dots \text{식 (4)}$$

위의 행렬식을 만족시키는 $W_1 \sim W_j$ 의 해는 $-\infty$ 에서 $+\infty$ 의 범위에서는 무수히 많은 해가 존재하므로, W_j 의 제약조건을 $0 \leq W_j \leq 1$ 으로 한다. 식(4)를 만족시키기 위해서는 좌변과 우변의 단위가 일치하여야 한다. 속성 가중치 W_j 는 단위가 없는 무차원의 값이지만, 속성값 X_{ij} 과 공사비 C_i 는 서로 다른 단위를 갖는다. 이에 따라, 7개 속성과 공사비의 단위 영향을 제거하기 위하여 속성값과 공사비 정보는 표준화 $N \sim (0,1)$ 한 값을 사용한다.

위에서 기술한 바와 같이 식 (4)를 정확히 만족하는 W_j 는 존재하지 않으므로, 식(4)의 좌변과 우변은 차이 D_i 가 발생한다. 식 (5)는 공사비(C_i)와 속성값과 속성가중치의 곱의 합($X_{ij} \times W_j$)과 그것의 차(D_i)를 나타낸 것이다.

$$\begin{pmatrix} C_1 \\ \vdots \\ C_i \end{pmatrix} - \begin{pmatrix} X_{11} & \dots & X_{1j} \\ \vdots & \ddots & \vdots \\ X_{i1} & \dots & X_{ij} \end{pmatrix} \times \begin{pmatrix} W_1 \\ \vdots \\ W_j \end{pmatrix} = \begin{pmatrix} D_1 \\ \vdots \\ D_i \end{pmatrix} \dots \text{식 (5)}$$

데이터베이스의 모든 사례에 대하여 D_i 의 값이 최소화되도록 최소화하는 W_j 를 도출한다. 이를 위해 Excel 기반의 유전자 알고리즘 상용화 프로그램인 Evolver를 사용하며, 초기 개체집단 50 · 교배율 0.05 · 변이율 0.1로 실행한다. 다음의 표5는 연면적 형별 속성 가중치를 나타낸 것이다. 속성 가중치는 조희대상인 현재사례가 변화하여도 동일하게 적용되고, 데이터베이스의 사례가 추가되거나 삭제되는 경우에만 변화하는 값이다.

표 5. 연면적형별 속성 가중치

연면적형 (㎡)	속성 가중치 (Attribute Weight, W)						
	W1	W2	W3	W4	W5	W6	W7
49, 59	0.017	0.051	0.748	0.003	0.102	0.003	0.075
84, 114	0.120	0.049	0.714	0.100	0.003	0.009	0.005

4.3.3 사례 유사도(Case Similarity)

사례 유사도 점수는 각 속성의 가중치와 유사도를 곱하여 더한 값을 속성 가중치 합으로 나누어 구한다. 식 (6)은 이를 수식으로 나타낸 것이다.

$$\text{사례 } i \text{의 유사도} = \frac{\sum_{l=1}^j (AS_l \times W_l)}{\sum_{l=1}^j W_l} \dots \text{식 (6)}$$

사례 유사도 점수는 속성 유사도 점수와 같이 조희대상 혹은

비교대상인 과거사례가 바뀌면 변화하는 상대적인 값이다. 이는 유사사례 추출의 기준이 되는 값으로, 점수가 높은 순으로 단수 혹은 복수의 사례를 추출하게 된다. 단수 사례의 추출은 가장 높은 점수를 획득한 사례를 선택하여 사용하게 된다. 복수의 사례를 추출하는 경우, 상위의 유사도 점수 순으로 일정 개수의 사례를 사용하거나 일정 점수 이상을 획득한 사례를 사용할 수 있다. 본 연구에서는 사례 유사도 점수 0.9 이상을 획득한 복수의 사례를 추출한다.

표6은 검증 사례1을 조희대상으로 하여 사례 유사도 점수를 산정한 것의 일부이다. 과거사례 중 B4이 0.964의 가장 높은 사례 유사도 점수를 획득하였으며, 18개의 사례가 0.9이상의 사례 유사도 점수를 나타내었다. 따라서 검증 사례1에 대한 유사사례는 B4를 포함한 18개의 사례를 추출한다.

5. 모델 검증

5.1 검증 방법

모델 혹은 시스템의 검증은 사용된 전체 데이터의 10%를 사용하는 것이 일반적이다(Dogan et al 2008). 본 연구는 수집한 168개의 데이터 중 모델 개발에 사용한 151개 데이터를 제외한 17개 데이터를 사용하여 모델 검증을 실시하였다. 검증 대상의 공사비는 사례 유사도 점수 0.9 이상을 갖는 사례의 평균 연면적당 공사비와 검증 대상의 연면적을 곱하여 구하였다.

공사비 예측 모델의 성능은 검증 대상의 실제 공사비와 예측 공사비의 차이를 실제 공사비로 나누어 구한 오차율로 나타내었다. 또한, 유전자 알고리즘을 이용한 속성 가중치 산정 방법의 유효성을 판단하기 위하여 동일가중치와 표준화 회귀계수를 이용한 경우의 오차율과 비교하였다.

5.2 검증 결과 및 분석

다음의 표7은 사례기반추론 모델의 속성 가중치 산정을 위해 유전자 알고리즘, 표준화 회귀계수, 동일가중치를 이용한 경우의 검증 결과를 나타낸 것이다.

유전자 알고리즘을 이용한 사례기반추론 공사비 예측 모델은 평균 3.57%, 표준편차 2.64%의 오차율을 나타내었다. 특히, 17개 사례 중 12개 사례에 대하여 5% 미만의 오차율이 도출되었다. 이를 AACE에서 프로젝트 정보량에 따라 견적단계를 5단계로 분류하여 정의한 예상 견적 정확도(AACE 1997)와 비교할 수 있다. 프로젝트가 10%~40% 진행되었을 경우, AACE의 예상 견적 정확도는 과소측정은 -10%~-20% / 과대측정은 +10%~+30%

표 6. 사례 유사도 점수 산정 예시

조희사례	연면적형 (m²)	속성 (Attribute, X _i)														공사비(C _i)
		X ₁	W ₁	X ₂	W ₂	X ₃	W ₃	X ₄	W ₄	X ₅	W ₅	X ₆	W ₆	X ₇	W ₇	
Test No. 1	114	26	0.120	14	0.049	3688.03	0.714	2	0.100	1	0.003	2	0.009	2	0.005	
사례	연면적형 (m²)	속성 (Attribute, X _i)														사례유사도 (Case Similarity)
		X ₁	AS(X ₁)	X ₂	AS(X ₂)	X ₃	AS(X ₃)	X ₄	AS(X ₄)	X ₅	AS(X ₅)	X ₆	AS(X ₆)	X ₇	AS(X ₇)	
B4	114	28	0.95	14	1.00	3929.89	0.96	2	1.00	1	1.00	2	1.00	0	0.75	0.964
A90	114	28	0.95	14	1.00	3966.9	0.95	2	1.00	1	1.00	2	1.00	0	0.75	0.960
A79	114	28	0.95	14	1.00	3967.74	0.95	2	1.00	1	1.00	2	1.00	0	0.75	0.960
C38	114	28	0.95	15	0.80	3923.94	0.96	2	1.00	1	1.00	2	1.00	2	1.00	0.956
C39	114	28	0.95	15	0.80	3924.13	0.96	2	1.00	1	1.00	2	1.00	2	1.00	0.956
C40	114	28	0.95	15	0.80	3932.46	0.96	2	1.00	1	1.00	2	1.00	2	1.00	0.955
A12	114	24	0.95	13	0.80	3426.23	0.96	2	1.00	1	1.00	2	1.00	2	1.00	0.953
C3	114	24	0.95	12	0.60	3341.76	0.94	2	1.00	1	1.00	2	1.00	0	0.75	0.932
C1	114	23	0.92	12	0.60	3290.79	0.93	2	1.00	1	1.00	2	1.00	1	0.88	0.924
A18	84	28	0.95	15	0.80	3108.12	0.90	2	1.00	1	1.00	2	1.00	2	1.00	0.916
A19	84	28	0.95	15	0.80	3108.12	0.90	2	1.00	1	1.00	2	1.00	2	1.00	0.916
A53	84	28	0.95	15	0.80	3108.11	0.90	2	1.00	1	1.00	2	1.00	2	1.00	0.916
A58	84	28	0.95	15	0.80	3103.31	0.90	2	1.00	1	1.00	2	1.00	2	1.00	0.915
B2	84	28	0.95	15	0.80	3091.49	0.90	2	1.00	1	1.00	2	1.00	2	1.00	0.914
A91	114	30	0.89	15	0.80	4249.17	0.91	2	1.00	1	1.00	2	1.00	0	0.75	0.910
A14	114	22	0.89	12	0.60	3143.96	0.91	2	1.00	1	1.00	2	1.00	2	1.00	0.904
A75	114	22	0.89	12	0.60	3141.83	0.91	2	1.00	1	1.00	2	1.00	2	1.00	0.903
C2	114	22	0.89	12	0.60	3107.9	0.90	2	1.00	1	1.00	2	1.00	2	1.00	0.900
C33	84	34	0.79	12	0.60	3660.87	1.00	3	0.50	1	1.00	3	0.50	2	1.00	0.897
C28	84	34	0.79	12	0.60	3646.74	0.99	3	0.50	1	1.00	3	0.50	2	1.00	0.896
C32	84	37	0.71	13	0.80	3980.66	0.95	3	0.50	1	1.00	3	0.50	2	1.00	0.867
C29	114	20	0.84	11	0.40	2950.29	0.88	2	1.00	1	1.00	2	1.00	2	1.00	0.865
B3	84	24	0.95	13	0.80	2655.47	0.83	2	1.00	1	1.00	2	1.00	2	1.00	0.862
C27	84	31	0.87	11	0.40	3355.52	0.95	3	0.50	1	1.00	3	0.50	2	1.00	0.861
B5	84	22	0.89	14	1.00	2541.74	0.81	2	1.00	1	1.00	2	1.00	6	0.50	0.850
A7	84	24	0.95	12	0.60	2638.56	0.83	2	1.00	1	1.00	2	1.00	0	0.75	0.849
A82	84	24	0.95	12	0.60	2637.72	0.83	2	1.00	1	1.00	2	1.00	0	0.75	0.849
C6	84	24	0.95	12	0.60	2529.48	0.81	2	1.00	1	1.00	2	1.00	0	0.75	0.837
C21	84	26	1.00	10	0.20	3055.56	0.90	3	0.50	1	1.00	3	0.50	4	0.75	0.831
C4	84	23	0.92	12	0.60	2493.57	0.80	2	1.00	1	1.00	2	1.00	1	0.88	0.830
A67	84	22	0.89	12	0.60	2453.89	0.80	2	1.00	1	1.00	2	1.00	4	0.75	0.821
A70	84	22	0.89	12	0.60	2453.89	0.80	2	1.00	1	1.00	2	1.00	4	0.75	0.821
A27	84	22	0.89	11	0.40	2420.01	0.79	2	1.00	1	1.00	2	1.00	0	0.75	0.808
C31	84	22	0.89	12	0.60	2396.45	0.79	2	1.00	0.5	0.67	4	0.00	2	1.00	0.806
C7	84	22	0.89	11	0.40	2325.06	0.78	2	1.00	1	1.00	2	1.00	0	0.75	0.796
C10	84	22	0.89	11	0.40	2365.12	0.78	2	1.00	0.5	0.67	4	0.00	0	0.75	0.791
C12	84	22	0.89	11	0.40	2365.12	0.78	2	1.00	0.5	0.67	4	0.00	0	0.75	0.791
A64	84	44	0.53	12	0.60	3367.97	0.95	4	0.00	2	0.33	2	1.00	4	0.75	0.783
C8	84	21	0.87	11	0.40	2228.28	0.76	2	1.00	1	1.00	2	1.00	1	0.88	0.782
C24	84	20	0.84	11	0.40	2220.35	0.76	2	1.00	0.5	0.67	4	0.00	2	1.00	0.769
A51	84	20	0.84	10	0.20	2200.62	0.75	2	1.00	1	1.00	2	1.00	0	0.75	0.766

에 해당하므로, 본 연구의 공사비 예측 모델은 AACE의 기준보다 우수한 오차율을 보였다. 또한, 표준화 회귀계수와 동일가중치를 이용한 방법은 각각 5.80%와 8.42%의 평균 오차율을 나타내었다.

표 7. 공사비 예측 결과

순서	실제공사비 (Actual Ci)	유전자 알고리즘		표준화 회귀계수		동일가중치	
		예측공사비 (Estimated Ci)	오차율 (%)	예측공사비 (Estimated Ci)	오차율 (%)	예측공사비 (Estimated Ci)	오차율 (%)
1	2,014,329,211	2,055,294,242	2.03	2,037,428,034	1.15	2,142,225,080	6.35
2	1,062,969,235	1,051,891,877	1.04	1,185,378,673	11.52	1,003,192,995	5.62
3	1,837,897,103	1,694,468,702	7.80	1,667,654,438	9.26	1,750,123,050	4.78
4	1,110,472,157	1,169,806,242	5.34	1,053,498,698	5.13	1,226,425,209	10.44
5	1,562,291,611	1,472,317,133	5.76	1,496,646,305	4.20	1,461,314,863	6.46
6	1,129,212,506	1,140,055,084	0.96	1,057,644,745	6.34	1,179,166,160	4.42
7	1,346,734,711	1,319,600,602	2.01	1,377,742,293	2.30	1,270,384,193	5.67
8	861,420,273	868,034,995	0.77	960,896,976	11.55	954,029,949	10.75
9	1,701,524,831	1,762,274,972	3.57	1,747,653,653	2.71	1,953,102,450	14.79
10	1,225,123,890	1,233,374,071	0.67	1,204,461,298	1.69	1,311,435,116	7.05
11	1,433,878,455	1,510,669,663	5.36	1,543,848,335	7.67	1,530,032,708	6.71
12	1,087,484,176	1,085,452,766	0.19	1,111,595,146	2.22	1,181,434,461	8.64
13	2,825,638,466	2,685,307,104	4.97	2,619,413,016	7.30	2,525,549,385	10.62
14	2,931,593,728	2,819,367,852	3.83	2,829,103,720	3.0	2,688,871,589	8.28
15	2,861,861,186	2,762,558,821	3.47	2,799,993,278	2.16	2,609,869,690	8.81
16	1,540,050,738	1,685,498,321	9.44	1,740,591,626	13.02	1,714,700,786	11.34
17	1,137,042,207	1,098,038,511	3.43	1,215,603,788	6.91	1,277,584,834	12.36
평균			3.57		5.80		8.42
표준편차			2.64		3.81		2.92

이어서, 분산분석(Analysis of Variance, ANOVA)을 사용하여 유전자 알고리즘 · 표준화 회귀계수 · 동일가중치를 이용한 방법의 평균 오차율의 차이를 검증하였다. 유의수준 0.05로 분산분석을 실행한 결과 P값은 0.05보다 작은 0.000231로 도출되었기 때문에, 세 가지 방법이 나타낸 평균 오차율은 유의미한 차이이다. 따라서 유전자 알고리즘을 사용하여 속성 가중치를 산정한 공사비 예측 모델이 가장 높은 정확도를 나타낸다고 판단할 수 있다.

6. 결론

사례기반추론 모델의 정확성은 속성 가중치 산정 결과에 많은 영향을 받는다. 기존의 연구는 속성 가중치 산정을 위해 경사하강법, 회귀분석법, 계층분석과정 등의 방법을 사용하지만, 이들의 방법은 수학적 오류 혹은 전문가의 주관적 판단에 따라 속성 가중치가 변화하는 단점을 가진다. 따라서 본 연구는 속성 가중치 산정을 위해 유전자 알고리즘을 이용한 사례기반추론 공사비 예측 모델을 제안하였다.

공사비 예측 모델은 기획/계획설계단계와 같은 초기단계를 대상으로 하였으며, 공공아파트 건축부문의 데이터를 이용하였다. 유사사례 추출은 최근이웃 조회 기법의 과정에 따라 속성 유사도

와 속성 가중치를 정의하고 사례 유사도 점수를 구하였다. 0.9이 상의 사례 유사도 점수를 획득한 과거 사례의 평균 연면적당 공사비를 사용하여, 검증 대상의 공사비를 계산하였다. 모델의 유효성을 검증한 결과, 평균 3.57%, 표준편차 2.64%의 오차율을 나타내었다. 이는 AACE가 정의한 견적시기별 예측 정확도보다 우수한 정확도로 볼 수 있다. 또한, 표준화 회귀계수와 동일가중치를 이용한 경우보다 높은 예측 성능을 나타내었다.

본 연구는 다음과 같은 차별성을 갖는다.

(1) 속성 가중치 산정에 있어서 다수의 속성 변수와 수많은 해 집합으로 인하여 기존의 결정론적 방법론이 가지는 문제점을 보완하기 위해, 최적화 문제 해결 기법인 유전자 알고리즘을 도입하였다. 이를 통해 표준화 회귀계수, 동일가중치를 사용한 경우보다 향상된 예측 결과를 얻을 수 있다.

(2) 사례기반추론 방법을 사용하여 인공지능망 등 타 인공지능 방법에 비해 해결책의 도출과정을 사용자가 이해하기 쉽고, 유사 사례 조회 알고리즘을 통해 빠른 시간에 과거 유사사례를 추출할 수 있다.

본 연구는 사례의 속성으로 정량적 변수만을 다루고 있는 한계를 가지고 있다. 이를 보완하여 정성적 변수를 측정하는 방법을 제시한다면 사용성을 증가시킬 수 있을 것으로 판단된다. 또한, 속성 가중치 산정 방법의 타당성 확보를 위해 동일가중치, 표준화 회귀계수 외에 경사하강법, 계층분석과정 등을 이용한 방법과의 추가적 비교가 필요하다.

감사의 글

본 연구는 건설교통 R&D정책인프라사업(06기반구축 A03과제) 건축공사 적정공사비 산정 및 관리시스템 개발 연구결과의 일부임.

참고문헌

김광희, 강경인 (2004). “사례기반추론 기법을 이용한 공동주택 초기 공사비 예측에 관한 연구”. 대한건축학회논문집 구조계, 20권 5호, 대한건축학회, pp.83~92

김광희 외 2인 (2004). “공동주택 공사비 예측 정확도 비교에 관한 연구, 사례기반추론 기법과 신경망을 중심으로”. 대한건축학회논문집 구조계, 20권 5호, 대한건축학회, pp.93~102

김광희 외 2인 (2006). “신경망과 유전자알고리즘을 이용한 공사비예측 모델의 예측정확도 비교에 관한 연구; 공동주택 공사비를 중심으로”. 대한건축학회논문집 구조계, 22권 3호, 대

한건축학회, pp.111~118

김광희, 안성훈 (2007). “공사단계별 공사비 영향 변수의 선택방법과 데이터 개수와의 상관관계에 관한 연구; 인공지능망을 활용한 공동주택 공사비 예측”. 대한건축학회논문집 구조계, 23권 4호, 대한건축학회, pp.129~137

김종한 외 3인 (1996). “사례기반 추론에서 Fuzzy 개념을 이용한 유사사례 추출”, 한국전문가시스템학회 학술발표논문집, 제2권 1호, 한국전문가시스템학회, pp. 153~160

김태근 (2006). u-Can 회귀분석, 초판, 인간과복지, 서울, pp.53~88

노형진 (2008). SPSS에 의한 다변량분석 기초에서 응용까지, 초판, 한울출판사, 서울, pp.13~45

문병로 (2008). 쉽게 배우는 유전 알고리즘, 초판, 한빛미디어, 서울, pp.23~34

박문서 외 3인 (2008). “건설사업관리를 위한 지속가능한 개선견적 모델 개발”. 대한건축학회논문집 구조계, 24권 10호, 대한건축학회, pp.167~178

손보식 외 2인 (2007). “영향변수에 따른 수량변화 분석을 이용한 기본설계단계의 개선견적 모델 개발”. 한국건설관리학회 논문집, 제8권 제2호, 한국건설관리학회, pp.155~166

송화철 외 2인 (2007). “국내 초고층 건물의 사례 데이터베이스를 이용한 사례기반추론기법”. 한국공간구조학회지, 제7권 제6호, 한국공간구조학회, pp.75~82

안성훈, 강경인 (2005). “전문가지식을 활용한 공동주택 초기단계 공사비 예측에 관한 연구”. 대한건축학회논문집 구조계, 21권 6호, 대한건축학회, pp.81~88

안성훈 외 2인 (2005). “판별분석을 위용한 개선견적 평가모델 구축에 관한 연구”. 대한건축학회논문집 구조계, 21권 9호, 대한건축학회, pp.169~176

이동원, 김성호 (2006). 엑셀을 활용한 현대통계학, 초판, 서울경제경영, pp.267~275

이장희(2008). “자기조직화신경망을 활용한 속성가중치를 사용하는 사례기반추론의 적용에 관한 연구”, 대한경영학회지, 제 21권 1호, 대한경영학회, pp. 417~437

지세현 외 4인 (2008). “공사비 예측을 위한 수량기반 데이터 분석방법 : 공공아파트 수장공사 중심으로”. 한국건설관리학회 논문집, 제9권 제6호, 한국건설관리학회, pp.20~28

진강규 (2004). 유전자 알고리즘과 그 응용, 제2판, 교우사, 서울, pp.61~63

AACE Recommended Practice No. 17R-97 (1997) : Cost Estimate Classification System, AACE, Inc.

- Agnar Aamodt & Enric Plaza (1994), "Case-Based Reasoning: Foundational Issues", Methodological Variations, and System Approaches, AI Communications. IOS Press, Vol.7:1, pp.39~59
- Ian Watson & Farhi Marir (1994), Case-Based Reasoning: A Review, The Knowledge Engineering Review, Vol.9 No.4, pp.6~13
- Janet Kolodner (1993), Case-Based Reasoning. Morgan Kaufmann Publishers, Inc., San Mateo, pp.289~389
- Ji et al. (2009). "Cost Estimation Model for Building Projects Using Case-Based Reasoning". Canadian Journal of Civil Engineering, National Research Council of Canada, under review
- Nie-Jia Yau & Jyh-Bin Yang (1998), "Case-Based Reasoning in Construction Management". Journal of Computer-Aided Civil and Infrastructure Engineering, 13, pp.143~150
- P. Duverile & J. M. Castelain (1999), "Cost Estimation During Design Step: Parametric Method versus Case Based Reasoning Method", Journal of Advanced Manufacturing Technology, 15, pp.895~906
- Sevgi Zeynep Dogan, David Arditi & H.Murat Gunaydin, "Determining Attribute Weights in a CBR Model for Early Cost Prediction of Structural Systems". Journal of Construction Engineering and Management, Vol.132 No.10, 2006, pp.1092~1098
- Sevgi Zeynep Dogan, David Arditi & H.Murat Gunaydin, "Using Decision Trees for Determining Attribute Weights in a Case-Based Model of Early Cost Prediction". Journal of Construction Engineering and Management, Vol.134 No.2, 2008, pp.146~152

논문제출일: 2009.12.11
 논문심사일: 2009.12.18
 심사완료일: 2010.01.20

Abstract

Because the estimated cost at early stage has great influence on decisions of project owner, the importance of early cost estimation is increasing. However, it depends on experience and knowledge of the estimator mainly due to shortage of information. Those tendency developed into case-based reasoning(CBR) method which solves new problems by adapting previous solution to similar past problems. The performance of CBR model is affected by attribute weight, so that its accurate determination is necessary. Previous research utilizes mathematical method or subjective judgement of estimator. In order to improve the problem of previous research, this suggests CBR schematic cost estimation method using genetic algorithm to determine attribute weight. The cost model employs nearest neighbor retrieval for selecting past case. And it estimates the cost of new cases based on cost information of extracted cases. As the result of validation for 17 testing cases, 3.57% of error rate is calculated. This rate is superior to accuracy rate proposed by AACE and the method to determine attribute weight using multiple regression analysis and feature counting. The CBR cost estimation method improve the accuracy by introducing genetic algorithm for attribute weight. Moreover, this makes user understand the problem-solving process easier than other artificial intelligence method, and find solution within short time through case retrieval algorithm.

Keywords : *Cost Estimation, Case-Based Reasoning, Attribute Weight, Genetic Algorithm*
