

사례기반추론 코스트 모델의 정성변수 속성가중치 산정방법

A Method of Assigning Weight Values for Qualitative Attributes in CBR Cost Model

이 현 수* 김 수 영** 박 문 서*** 지 세 현**** 성 기 훈***** 편 재 호*****
Lee, Hyun-Soo Kim, Sooyoung Park, Moonseo Ji, Sae-Hyun Seong, Kihoon Pyeon, Jae-Ho

요 약

건축 프로젝트는 그 다양성과 특수성으로 인해 많은 불확실성을 갖고 있다. 이러한 불확실성을 해소하기 위해 공사비 예측은 건축 프로젝트의 전 과정에 걸쳐 반복적으로 이루어져야 하며 특히 초기단계의 공사비 예측은 효과적인 사업 추진을 위해 매우 중요한 과정이다. 통상 초기단계 공사비 예측은 과거에 수행되었던 실적공사와의 비교를 기반으로 하며, 이러한 원리를 이용한 기계학습방법이 사례기반추론이다. 사례기반추론은 해결하고자 하는 문제와 유사한 사례를 데이터베이스에서 검색, 수정하여 해답을 얻는 방법으로 이를 위해서는 속성 유사도와 속성 가중치의 정의가 필요하다. 그러나 속성 가중치를 결정하는 문제에 있어서, 기존의 방법들은 정성변수의 속성 가중치 결정이 불가능하다는 단점이 있으며, 이는 사례기반추론에 사용할 수 있는 변수를 한정시키기 때문에 공사비 예측의 정확성을 저하시키는 요인이 되고 있다. 따라서 본 연구는 최적화 문제를 해결하는 기법의 하나인 유전 알고리즘을 이용하여 정성변수의 속성 가중치 결정 방법을 제안하고, 이를 국방 병영생활관과 공공아파트에 적용하여 그 유효성을 검증하였다.

키워드 : 공사비예측, 사례기반추론, 유전알고리즘

1. 서론

1.1 연구의 배경 및 목적

건설사업의 완성에는 많은 시간과 자원이 소요되기 때문에 사업의 전 단계에 걸쳐서 공사비 예측이 반복적으로 이루어진다. 공사비 예측은 건설공사에 투입되는 자원의 구체적인 내용과 작업 프로세스 등 기술적인 사항을 금액 단위로 환산하는 과정을 포함한다. 구체적인 공사 정보가 확정되지 않은 초기단계에서는 보다 제한적인 공사의 내용 및 특성 정보만을 이용하여 공사비를 예측한다. 특히 초기단계 공사비 예측은 사업의 성패를 좌우하는 중요 의사결정에 큰 영향을 미친다(Trost and Oberlender 2003, 김수영 외 4인 2009). 또한 단계가 진행될수록 변경에 많은 비용이 소요되어 비용효율이 낮아지기 때문에 이는 매우 중

요하다(Duverlie and Castelain, 1999).

이러한 초기 공사비의 예측방법 중 최근 부각되는 것이 사례기반추론(Case-Based Reasoning; CBR)이다. 사례기반추론은 주어진 문제(예측대상 프로젝트)와 유사한 실적자료들을 찾고 이것을 이용하여 해결책(예측결과)을 제시한다(Kolodner, 1993).

사례기반추론에서의 중요한 쟁점 중 한 가지는 어떻게 유사한 사례를 조회할 것인가에 대한 것이다(Aamodt and Plaza 1994, Luu et al. 2005). 사례의 조회는 성공적 문제 해결에 결정적인 영향을 미치고, 단지 유사사례의 검색 뿐 아니라 사례기반추론을 이용한 시스템의 전체 품질에까지 영향을 미치게 된다(Goh and Chua, 2009). 이때, 조회의 정확도는 각 속성들 간의 중요도를 반영한 속성 가중치에 크게 의존한다(Doğan et al., 2006).

기존의 사례기반추론을 이용한 예측 모델들에서 속성 가중치를 산정하기 위한 방법으로는 동일가중치, 경사하강법, 회귀분

* 중신회원, 서울대학교 건축학과 정교수, 공학박사, hyunslee@snu.ac.kr

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

*** 중신회원, 서울대학교 건축학과 부교수, 공학박사, mspark@snu.ac.kr

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

***** 일반회원, 서울대학교 건축학과 대학원 석사, emperorskh@hotmail.com

***** 일반회원, Assistant Prof., Civil and Environmental Engineering, San Jose State University, jae.pyeon@sjsu.edu

석, 계층화 분석법 등이 사용되어져 왔다. 그러나 기존의 방법들은 정성변수에 대한 속성가중치 산정이 주관적이거나 불가능하였다. 이는 사용 가능한 속성의 범위를 한정시켜서, 결과적으로 공사비 예측의 정확성을 떨어뜨릴 수 있다는 한계를 갖는다(김수영 외 4인, 2009).

이러한 문제를 해결하고자, 본 연구는 유전 알고리즘을 이용한 사례기반추론 코스트 모델의 속성가중치 산정 방법을 제시한다. 이 방법은 정량변수와 정성변수의 속성가중치 값을 함께 할당할 뿐 아니라 정성변수의 속성값 자체를 정량화된 값으로 변환하여, 유사사례 검색 시 고려되는 속성의 차원을 증가시킨다. 따라서 주어진 문제와 보다 유사한 사례를 추출할 수 있으며, 결과적으로 더 정확한 공사비 예측을 기대할 수 있다.

1.2 연구의 범위 및 방법

본 연구는 건설사업의 초기단계 공사비 예측을 위해 유전 알고리즘을 이용한 사례기반추론 코스트 모델의 정성변수 속성가중치 산정 방법을 제시하고 그 유효성을 검증하고자 다음의 절차와 방법에 따라 연구를 진행한다.

- (1) 연구의 범위는 건설사업의 초기공사비 예측을 위한 사례기반추론 코스트 모델을 대상으로 한다.
- (2) 문헌조사로 사례기반추론과 유전 알고리즘의 이론적 내용을 고찰하고 선행연구를 분석하여 기존 속성가중치 산정 방법의 시사점 및 한계점을 도출한다.
- (3) 분석된 한계점을 개선하기 위한 유전알고리즘 기반의 속성가중치 산정 방법을 제시한다.
- (4) 제시된 방법을 국방 병영생활관 사업과 공공아파트 사업에 적용하여 그 유효성을 검증한다.

2. 예비적 고찰

2.1 사례기반추론

Aamodt and Plaza(1994)는 사례기반추론이란 새로운 문제를 해결하기 위해 과거의 유사한 상황을 기억하고, 그 상황의 정보와 지식을 재사용하는 것이라고 정의하였다. Watson(1994)은 사례기반추론은 과거 유사사례의 성공적인 해결책을 적용하여 새로운 문제를 해결하는 것이라고 정의하였다.

사례기반추론이 다른 인공지능 기법과 구별되는 특징은 첫째로, 새로운 문제해결을 위해 과거사례의 구체적 지식을 사용한다는 점이다. 둘째로, 새로운 문제가 해결되면 과거사례로 저장되기 때문에 미래에 발생할 문제해결에 사용될 수 있다는 점이다

(Aamodt and Plaza, 1994). 사례기반추론은 의사결정이 필요한 다양한 현실 문제 해결에 적용되고 있으며, 특히 건축 설계, 공법 선정, 공사비 예측 등과 같이 과거의 경험과 지식을 중요시하는 건설 분야에서 널리 사용되고 있다.

2.1.1 유사사례 조회

사례기반추론에서 유사사례를 조회하는데 가장 일반적으로 사용되는 방법에는 최근이웃 조회 방법(nearest neighbor retrieval)과 귀납적 조회 방법(inductive retrieval), 지식기반 조회 방법(knowledge-based retrieval)이 있다. 이 중에서 최근이웃 조회 방법은 저장되어 있는 사례들 중에서 일정한 유사성 척도에 의해 새로운 사례와 유사한 사례를 추출한 후 가장 합계하여 대응시키는 방법이며, 이를 위해서는 속성값을 비교할 수 있는 규칙 및 수식과 속성 가중치의 정의가 필요하다. 귀납적 조회 방법은 조회 목표가 잘 정의된 경우에 가장 좋은 방법으로 사례는 데이터 자체에서 귀납된 결과에 대한 중요 영향요소에 의해 색인화된다. 지식기반 조회 방법은 관련 사례를 추출하는데 기존의 분야별 지식을 적용하는 것으로 규칙기반 시스템과 유사하며, 모든 지식이 완벽히 수집되거나 표현되지 못한다는 단점 때문에 단독으로 쓰이지 않고, 다른 조회 방법과 함께 쓰인다(이지은 1998, Barletta 1991, Brown and Gupta, 1994).

건설 프로젝트는 일회성, 비반복성의 특성을 지니고, 현재와 동일한 과거의 사례가 존재하지 않는다. 또한 공사비 예측에 사용되는 변수는 주로 수치의 형태로 표현된다. 최근이웃 조회 방법은 이러한 자료들을 바탕으로 유사 사례를 추출하는데 효과적이다. 또한 귀납적 조회 방법에 비해 잘못되거나 누락된 데이터의 영향을 적게 받기 때문에 코스트 모델의 안정성을 향상시킬 수 있다. 따라서 본 연구에서는 사례기반추론의 유사사례 조회 방법으로 최근이웃 조회 방법을 사용한다. 그림1은 최근이웃 조회 방법의 프로세스를 나타낸 것이다.

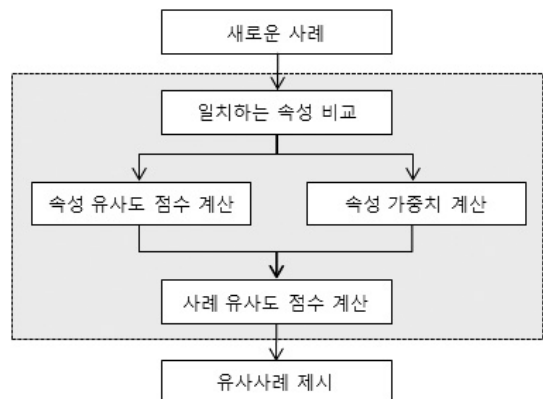


그림 1. 최근이웃 조회 방법 프로세스 (박문서 외 4인, 2010)

2.1.2 사례 유사도

최근이웃 조희 방법을 이용하여 유사사례를 추출하기 위해서는 먼저 사례 유사도를 정의하여야 한다. 사례 유사도는 예측하고자 하는 대상과 기존 사례간의 유사도를 표현한 것으로, 사례 유사도가 얼마나 잘 정립되었느냐에 따라 사례기반추론의 정확도가 좌우된다(Burkhard, 2001). 사례 유사도의 평가는 연구자에 따라 다른 방법이 적용되고 있으나, 공통적으로 최근이웃 조희 방법을 사용할 때에는 사례를 설명하는 각 속성의 유사도와 그것의 가중치를 곱한 후 합하는 방식이 사용된다. 즉, 유사도의 평가는 측정된 유사도를 계산하는 함수와 속성의 가중치에 의해 결정된다고 할 수 있다.

본 연구에서는 Ji et al.(2009)이 제시한 사례의 거리를 측정하는 방법인 유클리디안 거리측정원리를 기반으로 하는 방법론을 사용하였으며, 이는 식 (1)과 같다.

$$SIM(x_i, x_j) = \left[1 - \frac{\sqrt{\sum_{r=1}^n w_r^2 \{a_r(x_i) - a_r(x_j)\}^2}}{\sum_{r=1}^n w_r^2} \right] \dots \dots \dots \text{식 (1)}$$

- 여기서, w_r : r 번째 속성의 가중치
- $a_r(x_i)$: i 번째 사례의 r번째 속성값
- $a_r(x_j)$: 예측하고자 하는 사례의 r번째 속성값

2.1.3 속성 가중치

전술한 바와 같이 속성 가중치란 유사 사례를 찾기 위해 중요한 속성에 높은 가중치를 할당하여 보다 정확한 탐색을 유도하는 것이다. 일반적인 사례기반추론은 출력 변수에 미치는 입력 변수들의 영향도 차이를 고려하지 않고 동일한 중요도를 반영하여 예측 정확도가 떨어지는데, 속성가중치가 정확하게 평가될수록 예측 정확도는 높아진다(이장희, 2008).

2.2 유전 알고리즘

유전 알고리즘은 생물학적인 진화의 원리를 문제 해결에 이용하는 진화 연산의 대표적인 갈래이다. 자연 선택의 원리와 자연계의 생물 유전학에 기초를 두며 병렬적이고 전역적인 탐색 알고리즘으로서, 모든 생물은 주어진 환경 속에 적응함으로써 살아남는다는 적자생존의 이론을 기본 개념으로 한다.

유전 알고리즘은 선택(Selection), 교차(Crossover), 변이(Mutation), 대체(Replacement) 등의 주요 연산으로 구성된다.

먼저 선택에서는 한 세대에서 다음 세대로 전해지는 해의 후보가 되는 해들을 선택한다. 교차 연산은 두 부모 염색체의 교차점을 기준으로 서로의 다른 부분을 가지고 새로운 염색체를 조합해내며, 변이 연산은 염색체의 어떤 자리를 확률에 의한 값으로 변화를 주게 된다. 대치는 교차와 변이를 거쳐 만들어진 새로운 해를 해집단에 추가하고 기존 해 중 열등한 해를 가려내서 제외시키는 연산이다.

사례기반추론에서 속성 가중치를 산정하기 위해서는 공사비에 영향을 미치는 속성을 찾고, 속성과 공사비와의 관계를 바탕으로 속성 가중치의 해를 구해야 한다. 본 연구에서는 이러한 문제를 해결하기 위해 사례기반추론 코스트 모델의 속성가중치 산정에 유전 알고리즘을 사용하여 속성들과 공사비의 관계를 가장 잘 설명해 줄 수 있는 속성 가중치의 최적 해를 도출하였다.

3. 선행연구 분석

3.1 기존 속성가중치 산정 연구

기존의 사례기반추론의 속성가중치 산정 방법으로는 동일 가중치, 경사하강법, 회귀분석법, 계층화 분석법, 유전 알고리즘 등의 방법들이 사용되어져 왔다.

김중환 외 3인(1996)은 철강제품의 품질설계를 위한 사례기반추론 모델에서 각 속성에 동일 가중치를 적용하였다. 그러나 동일 가중치의 경우는 모든 속성에 동일한 가중치를 부여하기 때문에 속성간의 중요도 차이를 반영하지 못한다.

Yau and Yang(1998)은 사례기반추론을 이용한 건설 공기 및 공사비 예측 모델에서 경사하강법을 이용하였으며, 김광희와 강경인(2004)은 공동주택 초기 공사비 예측을 위한 사례기반추론 모델에서 경사하강법을 이용하여 속성가중치를 산정하였다. 경사하강법은 미리 정한 출발점에서 시작하여 함수의 구배에 따라 반복적으로 새로운 점을 탐색하는 방법으로, 시작점에 따라 그 값이 좌우되며, 지역 최적점에 빠질 가능성이 크다는 단점이 있다.

Chun and Park(2006)은 금융지수 예측 모델의 가중치 산정에 회귀분석을 이용하였으며, 지창윤(2009)은 회귀분석에서 도출된 표준화계수를 바탕으로 공공아파트 사업초기단계의 공사비 예측을 위한 속성가중치를 산정하였다. 회귀분석을 이용한 방법은 다른 방법론들에 비해 비교적 빠른 시간 내에 가중치를 구할 수 있다. 그러나 회귀분석은 종속변수와 독립변수의 관계가 선형적이며, 독립변수 상호간의 상관관계가 없다는 가정을 전제로 하나, 실제로는 비선형이며 독립변수들 간의 상관관계로 인한 다중공선성의 문제로 인하여 그 신뢰성이 떨어지게 된다.

Park and Han(2002)은 파산 예측 모델의 속성가중치 산정

에 계층화 분석법(Analytic Hierarchy Process; AHP)을 적용하였으며, 안성훈과 강경인(2005)은 계층화 분석법을 이용하여 전문가지식을 활용하여 속성가중치를 산정하였다. 그러나 계층화 분석법의 경우 전문가의 경험과 지식이라는 주관적 요소를 근거로 하기 때문에 전문가에 따라 속성 가중치가 변하는 문제점이 있다.

Ji et al.(2009)은 유전 알고리즘 기반의 사례기반추론 방법을 제안하였으며, 박문서 외 4인(2010)은 사례기반추론을 이용한 초기단계 공사비 예측 모델의 속성 가중치 산정을 위해 유전 알고리즘을 이용하였다. 유전 알고리즘을 이용한 방법은 수학적 모델에 기초하여 신빙성 있는 결과를 얻을 수 있으며 전역적 해를 찾기 때문에 경사하강법과는 달리 지역 최적점에 빠질 가능성은 낮지만, 매개 변수를 설정해야 하고 시간이 오래 걸린다는 단점이 있다.

위와 같이 사례기반추론의 속성 가중치 산정과 관련하여 많은 연구가 진행되어 오고 있지만, 기존의 연구들은 정성변수의 속성가중치 산정이 주관적이거나 불가능하여 사례기반추론에 사용할 수 있는 속성의 종류를 정량적인 것들로만 한정시킨다는 단점이 있다.

표 1은 사례기반추론의 속성 가중치 산정과 관련한 연구들을 정리한 것이다.

표 1. 사례기반추론의 속성 가중치 산정 연구

가중치 산정 방법	연구내용	저자	한계
동일 가중치	철강제품의 품질설계를 위한 사례기반 추론 모델	김종한 외 3인 (1996)	속성간 차이 반영 불가
경사 하강법	사례기반추론을 이용한 건설 공기 및 공사비 예측 모델	Yau and Yang (1998)	지역 최적점 수렴 문제
	공동주택 초기단계의 사례기반추론 공사비 예측	김광희, 강경인 (2004)	
회귀분석	사례기반추론을 이용한 금융지수 예측 모델	Chun and Park (2006)	비선형성 및 다중공선성 문제
	공공아파트 공사비 예측을 위한 사례기반추론 보정모델	지형윤 (2009)	
계층화 분석법	사례기반추론을 이용한 파산 예측 모델	Park and Han (2002)	주관적 요소에 의존
	전문가지식을 활용한 공동주택 초기단계 공사비 예측	안성훈, 강경인 (2005)	
유전 알고리즘	사례기반추론을 이용한 초기단계 공사비 예측 모델	Ji et al. (2009)	매개변수 설정 필요 및 긴 계산시간
	유전 알고리즘을 이용한 속성 가중치 산정 연구	박문서 외 4인 (2010)	

3.2 유전 알고리즘 기반의 속성가중치 산정 방법

Ji et al.(2009)은 유전 알고리즘을 이용하여 사례기반추론의 속성 가중치를 산정하는 방법을 제시하였으며, 이는 다음과 같다.

데이터분석을 통해 분석한 영향요인에 적당한 계수를 곱하여

더하면 공사비 정보를 표현할 수 있다. 각 사례의 속성을 X , 해당 열의 공사비를 C , 가중치를 W 라 하면 다음과 같은 식 (2)을 도출할 수 있다.

$$C_i = X_{i1}W_1 + X_{i2}W_2 + \dots + X_{ij}W_j \dots\dots\dots \text{식 (2)}$$

여기서, C_i : i 번째 사례의 공사비
 X_{ij} : i 번째 사례의 j 번째 속성의 속성값
 W_j : j 번째 속성의 가중치

이 식을 모든 사례에 만족하는 일반해를 구하기 위해 행렬식으로 나타내면 식 (3)과 같이 표현할 수 있다.

$$\begin{pmatrix} X_{11} & \dots & X_{1j} \\ \vdots & \ddots & \vdots \\ X_{i1} & \dots & X_{ij} \end{pmatrix} \times \begin{pmatrix} W_1 \\ \vdots \\ W_j \end{pmatrix} = \begin{pmatrix} C_1 \\ \vdots \\ C_i \end{pmatrix} \dots\dots\dots \text{식 (3)}$$

그러나 위의 행렬식을 만족시키는 일반해는 무수히 많이 존재한다. 구하고자 하는 것이 속성의 가중치이므로 W_j 의 제약조건을 $0 \leq W_j \leq 1$ 로 만들기 위해 우선 정량변수에 해당하는 영향요인(속성)과 공사비가 정규분포를 따른다고 가정하고 표준화하고 이를 0에서 1 사이의 값을 갖도록 누적정규분포로 변환하였다. 변환된 각 속성과 가중치, 공사비를 소문자 x_{ij}, w_j, c_j 라 하면 다음의 식 (4)과 같이 표현된다.

$$\begin{pmatrix} x_{11} & \dots & x_{1j} \\ \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ij} \end{pmatrix} \times \begin{pmatrix} w_1 \\ \vdots \\ w_j \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_i \end{pmatrix} \dots\dots\dots \text{식 (4)}$$

여기서, c_i : i 번째 사례의 공사비(표준화), $0 \leq c_i \leq 1$
 x_{ij} : 사례 i 의 j 번째 속성의 속성값(표준화),
 $0 \leq x_{ij} \leq 1$
 w_j : j 번째 속성의 가중치, $0 \leq w_j \leq 1$

표준화를 통하여 모든 속성의 단위영향을 제거하였으며, 각 속성은 각각의 가중치를 가지고 공사비를 표현하게 된다. 식 (3)을 만족하는 일반해 w_j 을 구하기는 불가능 하므로 표준화된 공사비와 각 해가 도출하는 수치와의 절대거리(distance) d_i 의 합을 최소화 하도록 다음의 식 (5)을 세우고 이를 최적화 하는 해를 유전 알고리즘을 통해 도출한다.

$$\begin{pmatrix} c_1 \\ \vdots \\ c_i \end{pmatrix} - \begin{pmatrix} x_{11} & \dots & x_{1j} \\ \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ij} \end{pmatrix} \times \begin{pmatrix} w_1 \\ \vdots \\ w_j \end{pmatrix} = \begin{pmatrix} d_1 \\ \vdots \\ d_i \end{pmatrix} \dots\dots\dots \text{식 (5)}$$

여기서, d_i : 사례 i 의 절대거리

즉, 식 (5)에서 $\sum_{k=1}^i |d_k|$ 을 최소화 하는 w_j 을 구한다.

이 방법은 전역적 탐색 알고리즘인 유전 알고리즘에 기초하여 신빙성 있는 결과를 얻을 수 있으나, 마찬가지로 정성변수에 대해서는 적용이 불가능하다는 단점이 있다.

4. 유전 알고리즘 기반 정성변수 속성가중치 산정 방법

본 연구는 유전 알고리즘을 이용한 최적화 과정에서 정성변수에 임의의 변수를 할당하고 속성가중치와 함께 최적화하는 방법을 사용하여 Ji et al.(2009)이 제시한 방법을 정성변수의 속성가중치 산정이 가능하도록 개선한 것이다.

즉, j개의 속성이 m개의 정량적 속성(p)과 l개의 정성적 속성(q)으로 이루어진 경우, 식 (3)은 다음과 같이 다시 정리할 수 있다.

$$\begin{pmatrix} p_{11} & \dots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{i1} & \dots & p_{im} \end{pmatrix} \times \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix} + \begin{pmatrix} q_{11} & \dots & q_{1l} \\ \vdots & \ddots & \vdots \\ q_{i1} & \dots & q_{il} \end{pmatrix} \times \begin{pmatrix} w'_1 \\ \vdots \\ w'_l \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_i \end{pmatrix}$$

..... 식 (6)

여기서, c_i : i 번째 사례의 공사비(표준화), $0 \leq c_i \leq 1$

p_{im} : 사례 i의 m 번째 정량적 속성값(표준화),

$$0 \leq p_{im} \leq 1$$

w_m : m 번째 정량적 속성의 가중치,

$$0 \leq w_m \leq 1$$

q_{il} : 사례 i의 l 번째 정성적 속성값

w'_l : l 번째 정성적 속성의 가중치, $0 \leq w'_l \leq 1$

정성변수의 속성가중치 산정을 위해, 다음과 같이 정성변수의 동일 속성값에 임의의 변수를 할당한다.

$$Q_l = \{x | x = q_{il}\} = \{a_{1l}, a_{2l}, \dots, a_{nl}\}, f(a_{nl}) = b_{nl} \text{ 이고,}$$

$$S(q_{il}) = b_{nl} \text{ when } q_{il} = a_{nl} \text{ 일 때,}$$

$$r_{il} = S(q_{il}) \text{ 식 (7)}$$

여기서, Q_l : l 번째 속성의 동일 속성값을 원소로 가지는 집합

a_{nl} : l 번째 정성적 속성의 n 번째 동일 속성값

b_{nl} : a_{nl} 에 해당하는 임의의 변수, $0 \leq b_{nl} \leq 1$

r_{il} : q_{il} 에 해당하는 정량화된 속성값

예를 들어 ‘지붕형태’ 이라는 정성변수가 a_{11}, a_{12}, a_{13} 에 해당하는 ‘평지붕’, ‘박공지붕’, ‘경사지붕’ 의 세 가지의 속성값을 가질 경우, 각각의 속성값에 0에서 1 사이의 값을 가지는 변수 b_{11}, b_{12}, b_{13} 을 할당한다. 이렇게 정해 놓은 임의의 변수를 속성가중치와 함께 유전 알고리즘으로 최적화한다.

$$\begin{pmatrix} c_1 \\ \vdots \\ c_i \end{pmatrix} = \begin{pmatrix} p_{11} & \dots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{i1} & \dots & p_{im} \end{pmatrix} \times \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix} + \begin{pmatrix} r_{11} & \dots & r_{1l} \\ \vdots & \ddots & \vdots \\ r_{i1} & \dots & r_{il} \end{pmatrix} \times \begin{pmatrix} w'_1 \\ \vdots \\ w'_l \end{pmatrix} = \begin{pmatrix} d_1 \\ \vdots \\ d_i \end{pmatrix}$$

..... 식 (8)

여기서, d_i : 사례 i 의 절대거리

즉, 식 (8)에서 $\sum_{k=1}^i |d_k|$ 을 최소화하는 w_j, w'_l 및 r_{il} 을 구한다.

이상의 프로세스를 요약하여 정리하면 다음과 같다(그림 2).

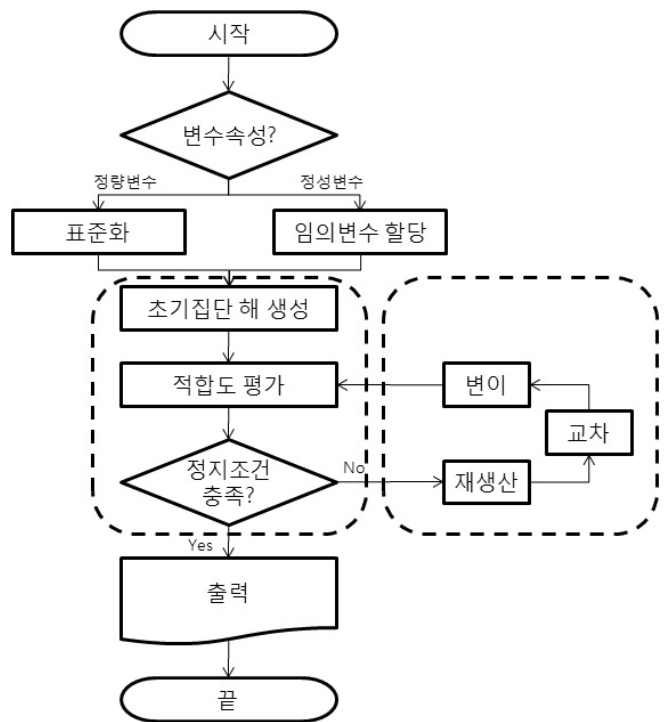


그림 2. 속성가중치 산정 프로세스

- (1) 변수의 속성에 따라 정량변수는 표준화, 정성변수에는 임의의 변수를 할당한다.
- (2) 초기집단 해를 생성하고, 적합도를 평가한다.
- (3) 정지조건이 아니면 해집단을 재생산하여 교차와 변이를 통해 또 다른 해집단을 만들어내고, 이를 다시 평가하여 적합도가 낮은 해는 도태된다.
- (4) 이를 반복하여 생성된 결과들 중 가장 최적화된 결과를 도출한다.

5. 사례 적용 및 검증

본 연구에서 제시한 정성변수 속성가중치 산정 모델의 활용성을 평가하기 위해, 관련 프로젝트 수행으로 자료 접근이 용이한 국방부 병영생활관과 A공사 공공아파트의 두 가지 대상에 대해 사례 적용 및 검증을 실시하였다.

각각의 검증 대상에 대해 단순무작위추출법을 이용하여 무작위로 10개의 사례를 선정하였다. 이를 본 연구에서 제시한 방법론, 동일 가중치, 회귀분석, 정성변수를 고려하지 않은 기존 유전알고리즘을 이용한 방법론을 적용한 각각의 사례기반추론 모델에 입력하고, 예측하고자 하는 대상의 직접공사비와 유사도 점수 상위 5개의 직접공사비 평균을 비교하였다. 이때의 오차율의 절대평균편차를 비교함으로써 어떤 모델의 예측율이 우수한지 비교할 수 있다. 또한 오차율의 절대값의 표준편차를 비교함으로써 공사비 예측 모델의 안정성을 비교하였다.

정성변수에 적용이 불가능한 기존의 방법론을 사용한 모델들은 정량변수만을 바탕으로 가중치를 산정하였다. 본 연구 방법론과 기존 GA-CBR 모델은 유전 알고리즘을 지원하는 상용 프로그램인 'Evolver 4.0'을 이용하여 사례기반추론의 속성 가중치를 산정하였다. 유전 알고리즘의 조건은 교차확률=0.05, 변이 확률=0.1을 이용하였으며 초기집단 해는 50개, 정지조건은 n=5,000,000번으로 설정하였다. 회귀분석의 경우 상용 프로그램인 'SPSS v12.0'을 사용하였다.

5.1 사례 1 : 국방 병영생활관

5.1.1 데이터 분석

본 연구에서 사용한 자료는 2005년부터 2008년까지 4년 간 시공되었던 국방 병영생활관 3개 군의 직접공사비 자료이다. BTL사업과 설계시공일괄입찰사업과 같은 특수한 경우를 제외한 본 연구에서 사용된 자료의 수는 115개이다. 분석대상 사업 개요는 표 2와 같다.

표 2. 분석대상 사업 개요(병영생활관)

연도	수요기관	A군	B군	C군	계
2005		30	1	5	36
2006		6	3	11	20
2007		16	1	10	27
2008		28	1	3	32
계		80	6	29	115

수집된 자료 중 검증용 사례를 제외한 105개 사례를 바탕으로 정성변수 속성가중치 산정을 위한 데이터베이스를 구축하였다.

이를 바탕으로 사업 초기단계에서 얻을 수 있는 영향요인들 중 전문가와 면담을 통해 병영생활관의 직접공사비에 영향을 미칠 것으로 예상되는 6가지 영향요인을 도출하였으며, 이는 표 3과 같다.

표 3. 변수의 유형과 내용(병영생활관)

유형	영향요인	단위	속성값
정량변수	수용인원	명	8 - 656
	층수	층	1 - 4
	연면적	m ²	83 - 9,160
	건축면적	m ²	83 - 3,378
정성변수	수요기관	-	A군, B군, C군
	구조형식	-	RC, 철골, RC+철골

또한 프로젝트의 수행년도가 2005년에서 2008년까지 분포되어 있으므로 이를 건설기술연구원에서 발행하는 건설공사비 지수 중 국방 병영생활관에 해당하는 '55 건축건설'의 건설공사비 지수를 이용하여 2008년도 기준으로 변환하였다.

5.1.2 가중치 산정

앞에서 설명한 4가지 방법론에 대해 각각 속성가중치를 산정하였다. 이를 통해 구한 속성가중치의 값은 표 4와 같다.

표 4. 속성가중치 계산값(병영생활관)

모델	속성	정량변수				정성변수	
		수용인원	층수	연면적	건축면적	수요기관	구조형식
본 연구 방법론		0.0013	0.0943	0.7898	0.0985	0.0204	0.1125
동일가중치		0.2500	0.2500	0.2500	0.2500	-	-
회귀분석		0.0021	0.0840	0.7317	0.1822	-	-
기존 GA-CBR		0.0055	0.0335	0.9453	0.0171	-	-

5.1.3 결과 및 분석

표 5는 정확도를 비교한 결과이다. 개별 사례에 따라서는 조금씩 차이가 있지만, 본 연구에서 제시한 정성변수를 고려하여 유전 알고리즘을 이용하여 산정한 속성가중치를 적용하였을 때 평균 오차율이 감소하였다. 또한 표준편차의 경우에도 본 연구의 방법을 적용하였을 때 더욱 감소하였다. 이는 정성변수를 이용하여 사례기반추론을 진행할 경우 기존의 방법에 비해 예측 정확도 및 공사비 예측 모델의 안정성이 상승한다는 것을 보여주는 결과로, 본 연구에서 제시한 방법이 타당하며 적용 가능하다는 것을 보여준다.

표 5. 사례별 정확도 오차율(병영생활관)

(절대편차, 단위 : %)

모델	본 연구 방법론	동일 가중치	회귀분석	기존 GA-CBR
사례1	12.0	28.3	12.4	4.4
사례2	4.4	5.1	12.6	10.5
사례3	3.3	14.3	7.2	2.6
사례4	7.4	17.0	18.9	18.9

표 5. 사례별 정확도 오차율(병영생활관)(계속)

(절대편차, 단위 : %)

모델 사례	본 연구 방법론	동일 가중치	회귀분석	기존 GA-CBR
사례5	9.3	11.1	9.3	9.3
사례6	3.2	3.2	10.3	16.0
사례7	7.8	7.8	6.2	6.2
사례8	4.0	4.0	11.2	11.2
사례9	4.1	6.1	11.0	5.2
사례10	11.9	51.6	11.9	13.0
평균	6.7	14.8	11.1	9.7
표준편차	3.5	15.0	3.5	5.2

5.2 사례 2 : A공사 공공아파트

5.2.1 데이터 분석

본 연구에서 사용한 자료는 2006년부터 2008년까지 3년 간 시공되었던 A공사 4개 지구의 84면적형 공공아파트의 동별 공사비 자료이다. 사용된 자료의 수는 86개이다.

표 6. 변수의 유형과 내용(공공아파트)

유형	영향요인	단위	속성값
정량변수	세대수	명	5 - 56
	연면적	㎡	546 - 6,189
	세대구성	EA	1 - 6
	층수	층	3 - 15
	엘리베이터수	EA	1 - 3
	엘리베이터 1대당 기준층 세대수	EA	2 - 4
정성변수	필로티세대수	EA	0 - 8
	공용형태	-	복도형, 계단형
	지붕형태	-	평지붕, 박공지붕, 경사지붕
	구조형식	-	L형, -형

수집된 자료 중 검증용 사례를 제외한 76개 사례를 바탕으로 정성변수 속성가중치 산정을 위한 데이터베이스를 구축하였다. 이를 바탕으로 사업 초기단계에서 얻을 수 있는 영향요인들 중 전문가와 면담을 통해 공공아파트의 직접공사비에 영향을 미칠 것으로 예상되는 10가지 영향요인을 도출하였으며, 이는 표 6과 같다. 또한 프로젝트의 수행년도가 2006년에서 2008년까지 분포되어 있으므로 이를 건설기술연구원에서 발행하는 건설공사비 지수 중 공공아파트에 해당하는 '123_주택건축'의 건설공사비 지수를 이용하여 2008년도 기준으로 변환하였다.

표 7. 속성가중치 계산값(공공아파트)

모델 속성	정량변수							정성변수		
	세대수	연면적	세대구성	층수	엘리베이터수	엘리베이터 1대당 기준층 세대수	필로티 세대수	공용형태	지붕형태	주동형식
본 연구 방법론	0.000	0.8413	0.0020	0.000	0.0042	0.000	0.043	0.0087	0.0988	0.0018
동일가중치	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	-	-	-
회귀분석	0.0193	0.6593	0.0410	0.056	0.1192	0.0782	0.0270	-	-	-
기존 GA-CBR	0.000	0.8122	0.000	0.0901	0.0490	0.000	0.0487	-	-	-

표 8. 사례별 정확도 오차율(공공아파트)

(절대편차, 단위 : %)

모델사례	본 연구 방법론	동일가중치	회귀분석	기존 GA-CBR
사례1	1.4	1.7	6.3	1.7
사례2	3.4	8.5	13.9	13.9
사례3	25.1	36.0	34.5	31.4
사례4	2.3	2.8	2.8	2.8
사례5	3.6	15.6	2.1	1.9
사례6	9.8	22.8	24.8	24.8
사례7	8.7	7.4	6.9	4.2
사례8	0.4	2.3	11.1	3.6
사례9	7.3	6.9	6.9	6.9
사례10	10.0	18.3	5.9	5.9
평균	7.9	16.2	11.7	9.7
표준편차	6.9	13.5	10.3	10.4

5.2.2 가중치 산정

사례 1과 동일한 방법으로 속성가중치를 산정하였으며, 이는 표 7과 같다.

5.2.3 결과 및 분석

표 8은 정확도를 비교한 결과이다. 공공아파트의 경우에도 병영생활관과 마찬가지로 본 연구에서 제시한 정성변수 속성가중치 산정 모델을 적용하였을 경우 평균 오차율 및 표준편차가 감소하는 것을 확인할 수 있다. 이는 본 연구의 방법이 타 방법에 비해 정확도 및 안정성 측면에서 가장 우수한 것을 나타내고 있다.

6. 결론

지금까지 사례기반추론의 속성가중치 산정을 위해 많은 방법들이 제시되어 왔다. 그러나 기존의 방법들은 정성변수에 대한 속성가중치 산정이 주관적이거나 때로는 불가능하여, 이로 인하여 사례기반추론에 사용할 수 있는 속성의 종류가 한정된다는 단점이 있었다.

본 연구에서는 이러한 단점을 보완하기 위해 기존의 유전 알고리즘 기반의 속성가중치 산정 방법을 개선하여 정성변수 속성가중치 산정이 가능한 새로운 방법을 제안하였다. 제안한 방법은 기존의 방법과 비교해 볼 때 그 정확도 및 안정성이 향상된 것으로 나타났으며, 이는 제안한 방법이 충분히 활용 가능하다는 것을 나

타내고 있다.

그러나 본 연구는 기존의 방법에 비해 더 많은 변수를 사용함에 따라 이러한 것들이 정확도에 미치는 영향에 대한 검증이 부족하며, 정성변수와 속성값의 수가 많아질수록 계산시간이 오래 걸린다는 단점이 있으며, 이는 사용성을 저해하는 요인이 되고 있다.

본 연구는 기존에 불가능하였던 정성변수의 속성가중치 산정을 가능하게 함으로써 사례기반추론에 적용 가능한 속성의 범위를 확장하였다는데 그 의의가 있다. 본 연구에서는 명목척도를 가지는 정성변수의 예를 들었지만, 마감등급이나 정보통신등급 등과 같은 서열척도를 가지는 정성변수에도 적용 가능하다.

본 연구는 국방부 병영생활관과 A공사 공공아파트를 대상으로 연구를 진행하였기 때문에 다른 대상에 적용시킬 경우 추가적인 연구가 필요하다. 향후 연구에서는 사례기반추론의 4가지 단계 중 수정 및 저장 단계에 대한 연구를 통해 유전 알고리즘 기반 사례기반추론의 사용성을 높이는 연구가 필요하다.

감사의 글

본 연구는 국토해양부 R&D정책인프라(06기반구축A03) 결과의 일부임

참고문헌

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

김수영 외 4인(2009), “사례기반추론의 정성변수 속성가중치 산정방법 : 건축공사 초기 공사비 예측을 중심으로, 한국건설관리학회 학술발표대회 논문집: (2009-11), pp. 297-302

김종한 외 3인(1996), “사례기반 추론에서 Fuzzy 개념을 이용한 유사사례 추출”, 한국전문가시스템학회 학술발표논문집, 제 2권 1호, pp. 153-160

박문서 외 4인(2010), “사례기반추론을 이용한 초기단계 공사비 예측 방법 : 속성 가중치 산정을 중심으로”, 한국건설관리학회 논문집, 제 11권 4호, pp. 22-31

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

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

이지은(1998), “사례기반 시스템의 이론적 고찰 및 구현 현황”, 경영학연구 제 6호, 상명대학교 경영연구소

지창윤(2009), “공동주택 건설공사의 공사비 예측정확도 향상을 위한 CBR-Revision Model”, 서울시립대학교 석사학위논문

Aamodt, A. and Plaza, E. (1994). “Case-based reasoning: Foundational issues, methodological variations and system approaches”, *AI Communications*, 7(1), pp. 35-39

Barletta, R. (1991). “An introduction to case-based reasoning”, *AI Expert*, 6(8), pp. 42-49

Brown, C. E. and Gupta, U. G. (1994). “Applying case-based reasoning to the accounting domain”, *Intelligent Systems in Accounting Finance and Management*, 3, pp. 205-221.

Burkhard, H. D. (2001). “Similarity and Distance in Case-based Reasoning”, *Fundamenta Informatica*, 47, pp. 201-215.

Chun, S. H. and Park, Y. J. (2006). “A new hybrid data mining technique using a regression case based reasoning: Application to financial forecasting”, *Expert Systems with Applications*, 31(2), pp. 329-336

Dogan, S. Z., Arditi, D. and Gnaydin, H. M. (2006). “Determining attribute weights in a CBR model for early cost prediction of structural systems”, *ASCE Journal of Construction Engineering and Management*, 132(10), pp. 1092-1098

Duverlie, P and Castelain, J. M. (1999), “Cost Estimation During Design Step: Parametric Method versus Case Based Reasoning Method”, *The International Journal of Advanced Manufacturing Technology*, 15, pp. 895-906

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), ASCE, pp. 1181-1189

Ji, S. H. et al. (2009). “Cost Estimation Model for Building Projects Using Case-Based Reasoning”, Submitted to the *Canadian Journal of Civil Engineering* at oct 21st. 2009.

Kolodner, J. (1993). *Case-Based Reasoning*. Morgan Kaufmann Publishers Inc.

- Luu, D. T., Ng, S. T. and Chen, S. E., (2005). "Formulating procurement selection criteria through case-based reasoning approach". *Journal of Computing in Civil Engineering*, 19(3), pp. 269-276
- Park, C. S. and Han, I. (2002). "A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction", *Expert Systems with Applications*, 23(3), pp. 255-264
- Trost, S. M and Oberlender, G. D. (2003). "Predicting Accuracy of Early Cost Estimates Using Factor Analysis and Multivariate Regression", *Journal of Construction Engineering and Management*, 129(2), ASCE, pp. 198-204.
- Watson, I. and Marir, F. (1994). "Case-Based Reasoning: A review", *the Knowledge Engineering Review*, 9(4), pp. 355-381
- Yau, N. J. and Yang, J. B. (1998). "Case-Based Reasoning in Construction Management", *Computer-Aided Civil and Infrastructure Engineering*, 13, pp. 143-150

논문제출일: 2010.08.31

논문심사일: 2010.09.03

심사완료일: 2010.10.29

Abstract

For construction projects, the importance of early cost estimates is highly recognized by the project team and sponsoring organization because early cost estimates are frequently a foundation of business decisions as well as a basis for identifying any changes as the project progresses from design to construction. However, it is difficult to accurately estimate construction cost in the early stage of a project due to various uncertainties in construction. To deal with these uncertainties, cost estimates should be made several times over the course of the project. In particular, early cost estimates are essential process for successful project management. For accurate construction cost estimates, it is necessary to compare cost estimates with actual costs based on historical project data. In this context, case-based reasoning (CBR), which is the process of solving new problems based on the solutions of similar past problems, can be considered as an effective method for cost estimating. To obtain this, it is also required to define the attribute similarities and the attribute weights. However, no existing method is capable of determining attribute weights of qualitative variables. Consequently, it has been a well-known barrier of accurate early cost estimates. Using Genetic Algorithms (GA), this research suggests the method of determining the attribute weight of qualitative variables. Based on building project case studies, the proposed methodology was validated.

Keywords : *Cost Estimating, Case-Based Reasoning, Genetic Algorithm*
