

GA-CBR 공사비 예측 모델의 정확도 향상을 위한 Dual-optimization 방법



김수영 한양대학교 ERICA 건축학부 연구교수

KICEM

1. 서론

건설 프로젝트의 완성에는 많은 시간과 자원이 소요되며, 이에 공사비 예측은 프로젝트의 전 단계에 걸쳐서 지속적으로 이루어져야 한다. 특히 초기단계 공사비 예측은 프로젝트의 성공 여부를 결정하고, 프로젝트가 진행될수록 공사비를 줄일 수 있는 가능성이 낮아지기 때문에 이러한 초기단계 공사비 예측은 매우 중요하다(Duverlie and Castelain 1999, Trost and Oberlender 2003, Koo et al. 2011). 과거의 유사 사례를 바탕으로 새로운 문제를 해결하는 기법인 사례기반추론(Case-based Reasoning; CBR)은 인간의 문제해결 프로세스와 유사하며, 설득력있고 정확한 해답을 빠르게 제시할 수 있으며, 유지관리가 쉽고, 사용할수록 사례 저장을 통해 정확도가 향상된다는 장점을 가지고 있으며(Arditi and Tokdemir 1999, Ji et al. 2011), 이에 많은 연구자들이 사례기반추론을 공사비 예측에 적용하여 예측의 정확도를 향상시키기 위한 연구를 진행해 오고 있다.

사례기반추론 공사비 예측 모델의 정확도를 향상시키기 위해, 기존의 연구들은 크게 두 가지의 이슈에 대해 다루고 있다. 첫 번째 이슈는 어떻게 유사한 사례를 조회(retrieve)하느냐이다(Aamodt and Plaza 1994, Leake 1996). 유사사례 조회는 사례기반추론의 성능에 큰 영향을 미치는 과정이며, 대부분의 사례기반추론 공사비 예측 모델은 수치화된 자료의 검색에 적합하며, 누락된 데이터의 영향을 적게 받는 최근접 이웃 조회(Nearest Neighbor Retrieval)를 사용하고 있다. 조회의 정확도는 사례를 표현하는 속성(attribute), 그리고 속성의 중요도를 나타내는 속성 가중치가 어떻게 적용되었느냐에 영향을 받게 된다. 그러나 기존의 사례기반추론 방법들은 정성적 속성의 영향을 무시하거나 정성적 속성값간의 차이를 반영하지 못하고

있으며, 이는 공사비 예측에 활용할 수 있는 속성의 수를 감소시키거나 유사사례의 유사도 순서를 왜곡하게 되며, 공사비 예측의 정확도를 감소시킬 수 있다.

두 번째 이슈는 조회된 사례를 어떻게 보정(adapt)하느냐이다(Goh and Chua 2009, Hu et al. 2015). 문제와 과거 사례가 완벽하게 일치할 수 없기 때문에, 과거 사례의 해결책(solution)은 새로운 상황에 맞도록 보정되어야 한다. 문제와 유사사례간의 차이가 미치는 영향을 계산하고 이를 완화하기 위해 많은 연구들이 수행되어 왔지만, 기존의 방법들은 추가적인 분석 혹은 데이터베이스, 학습 등이 필요하다는 단점이 있다. 또한 하나의 사례만 문제해결에 사용하는 것은 다른 유사사례들의 좋은 특성을 반영하지 못하게 되기 때문에, 여러 개의 사례를 통해 해결책을 도출해내는 방법(Multiple Cases Adaptation)이 활용되고 있으며, 사례 유사도를 가중치로 반영하는 가중 평균 방법(Weighted mean method)이 주로 사용된다. 그러나 이 방법은 유사사례간의 가중치의 비가 상대적으로 적게 계산되어, 유사도 차이에 따른 보정 효과가 작다는 단점이 있다.

이러한 두 가지 이슈 중 조회 문제를 해결하기 위해, 본 연구는 글로벌 해를 찾을 수 있는 유전 알고리즘(Genetic Algorithm; GA)을 활용하여 사례기반추론 공사비 예측 모델에서 정성적 속성을 고려하기 위한 최적화 방법인 Dual-optimization을 제안하였다. 이는 속성 가중치와 함께, 정성적 속성의 속성값에 랜덤 변수를 할당하고 이를 함께 최적화함으로써 가중치 산정과 정성적 속성의 정량화를 동시에 가능하게 하는 알고리즘이다. 또한 공사비 예측의 정확도를 더욱 향상시키기 위해, 사례 보정 문제를 해결하기 위한 조회 오차 보정 방법(Retrieving Error Adaptation Method)과 개선된 가중 평균 방법(Improved Weighted Mean Method)을 함께 제안하였다.

II. 예비적 고찰

1. 초기단계 공사비 예측 방법

공사비 예측은 그 단계에 따라 크게 개념전적, 상세전적, 확정전적으로 나눌 수 있으며, 프로젝트의 예산을 맞추기 위해 지속적으로 수행된다. 프로젝트가 진행될수록 얻을 수 있는 정보의 양 및 공사비는 증가하나, 의사결정이 전체 공사비에 미칠 수 있는 영향은 감소하게 되며, 이에 제한된 정보로 공사비를 예측해야 하는 초기단계의 공사비 예측은 매우 중요하다 (Barrie and Paulson 2000). 다른 단계와 달리, 초기단계 공사비 예측에서 기대하는 정확도는 상대적으로 매우 낮으며, 그만큼 초기단계 공사비 예측은 매우 어렵다고 할 수 있다. 공사비 예측의 정확도 범위가 넓으면, 그만큼 더 많은 불확실성을 가지고 있다는 것이며, 이는 더 많은 예비비를 필요로 하게 된다. 이에 초기단계 공사비 예측의 정확도를 높이기 위한 많은 연구가 진행되어 오고 있다.

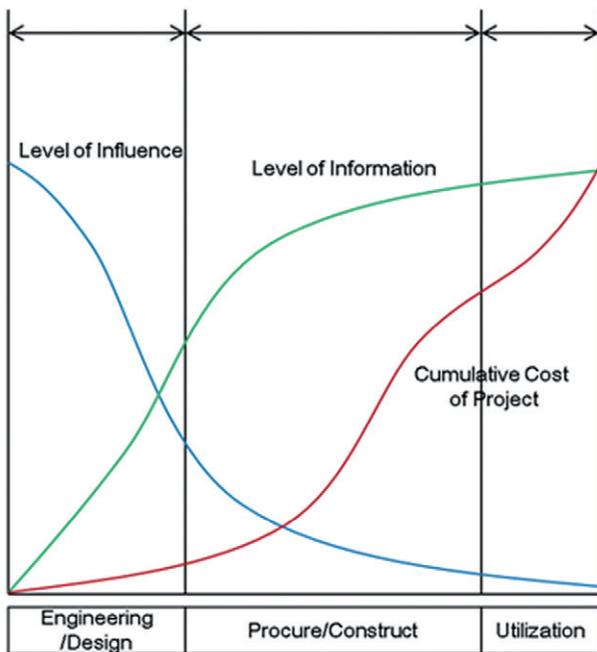


그림 1. Level of Influence and Information (Paulson 1976)

초기단계 공사비 예측을 위한 전통적인 방법에는 비용지수법, 비용용량법, 기본단가법, 매개변수비용법 등이 있으며, 이러한 방법들은 비록 간편하나, 공사비에 영향을 미치는 다양한 변수들을 고려하지 못하여 그 정확도가 상대적으로 낮다는 단점이 있다. 이에 인공지능 기법을 이용하여 초기단계의 공사비를 예측하는 방법들이 시도되어 오고 있다. 가장 기본적인 방법

으로는 여러 개의 규칙을 생성하고 이를 활용하여 추론하는 규칙기반추론(Rule-based Reasoning)이 있다. 그러나 규칙기반추론은 해당 규칙이 없으면 문제를 해결할 수 없고, 이러한 규칙을 생성하는 것은 매우 어려우며, 규칙의 수가 증가할수록 추론 속도가 감소한다는 단점이 있다. 이러한 문제를 해결하기 위해, 다른 인공지능 기법들이 공사비 예측에 활용되어져 왔다. 인공신경망(Artificial Neural Networks)의 경우, 비교적 높은 정확도로 공사비를 예측할 수 있기 때문에 많은 연구가 이루어져 왔지만, 인풋과 아웃풋의 관계를 설명하기 힘든 블랙박스 모델이며, 과적합 문제로 인한 일반화의 어려움 등의 단점이 있다. 이러한 방법은 공사비 예측보다는 문제 해결을 위한 지식이 부족한 분야에 더욱 적합한 방법이다. 반면, 사례기반추론은 문제와 과거 사례간의 유사도를 통해 가장 유사한 사례를 조회하여 그 해결책을 활용하는 방법으로, 인공신경망 방법에 비해 업데이트가 용이하고, 정보가 부족한 상황에서도 잘 작동하기 때문에 경험에 기반을 둔 건설분야의 문제 해결에 적합한 방법이다.

2. 사례기반추론(Case-based Reasoning)

사례기반추론은 과거사례의 정보와 지식을 재사용하여 새로운 문제를 해결하는 방법이다. 다른 인공지능 방법들과 비교할 때, 사례기반추론은 규칙을 새로 생성하기보다는 과거의 문제 해결 경험을 활용하고, 그 결과는 다시 저장되어 다른 문제를 해결하는데 쓰인다는 차이점이 있다(Kolodner 1993, Aamodt and Plaza 1994). 사례기반추론은 조회(Retrieve), 재사용(Reuse), 수정(Revise), 저장(Retain)의 네 가지 프로세스로 구성되어 있다. 문제를 해결하기 위해, 새로운 문제는 케이스 베이스에 저장된 사례들과 매칭되게 되고, 하나 혹은 여러 개의 유사사례를 조회한다. 조회된 사례는 보정을 거쳐 문제 해결에 재사용되고, 재사용 결과 문제 해결에 실패한 경우 그 해를 수정한다. 최종적으로 문제해결 결과는 사례 베이스에 다시 저장되게 된다. 이처럼 사례기반추론은 문제해결이 반복될수록 사례 베이스에 저장되는 사례가 많아지게 되고, 추론의 정확도 역시 사용할수록 증가하게 된다.

사례기반추론에서 사례 조회를 위해 쓰이는 방법에는 최근접 이웃 조회 방법과 귀납적 조회 방법이 있다. 이중 최근접 이웃 조회 방법은 문제와 사례간의 유사도를 바탕으로 유사한 사례를 조회하는 방법으로, 귀납적 조회 방법에 비해 수치로 표현된 자료를 다루는데 효과적이며, 잘못되거나 누락된 데이터의 영향을 적게 받기 때문에 공사비 예측에 주로 사용된다. 최근접

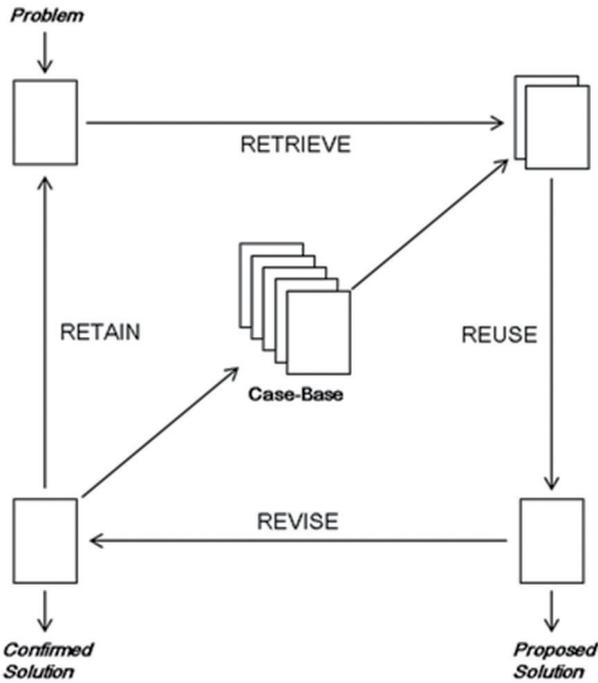


그림 2. CBR Cycle (Watson and Marir 1994)

이웃 조회 방법은 거리 함수에 의존하며, 본 연구는 속성간의 중요도를 고려해야 할 때 효과적인 방법인 Weighted Euclidean distance를 활용하여 유사도를 계산한다.

사례기반추론의 속성가중치 산정을 위한 다양한 방법이 제시되어 오고 있지만, 각각 다음과 같은 문제를 가지고 있다. 경사하강법은 최적화 과정에서 지역 최적점에 빠질 가능성이 크고, 회귀분석은 공사비 변수들이 가지는 비선형성 및 다중공선성에 영향을 받는다. AHP의 경우 개인의 지식 및 경험 등 주관적인 요소에 의존하고, 순위 역전과 같은 문제가 있다. 유전 알고리즘은 전역 최적화 방법으로 지역 최적점에 빠질 가능성은 낮지만, 다른 방법에 비해 계산시간이 오래 걸린다는 단점이 있다. 정성변수의 속성가중치 산정과 관련한 문제를 살펴보면, 정량화 없이 속성의 일치 여부를 0과 1로 표현하는 방법과 전문가의 의견을 반영하는 AHP 방법이 주로 사용되었으며, 이러한 두 방법은 정성변수의 속성값들의 차이를 반영하기 어렵다는 단점이 있다.

III. Dual-optimization을 활용한 GA-CBR 공사비 예측 모델

본 장에서는 GA-CBR 공사비 예측 모델의 정확도 향상을 위한 세 가지 방법을 제안하고, 이를 활용한 공사비 예측 프로세

스에 대해 살펴보고자 한다.

1. Dual-optimization

사례 조회 문제를 해결하기 위해, 본 연구는 유전 알고리즘을 이용하여 속성가중치 산정과 함께 정성변수의 정량화가 가능한 Dual-optimization을 제안한다. 사례 i 의 속성을 X_i , 가중치를 W_i 라고 하면 공사비 C_i 는 다음과 같이 표현할 수 있다.

$$\begin{pmatrix} X_{i1} & \cdots & X_{ij} \\ \vdots & \ddots & \vdots \\ X_{i1} & \cdots & X_{ij} \end{pmatrix} \times \begin{pmatrix} W_1 \\ \vdots \\ W_j \end{pmatrix} = \begin{pmatrix} C_1 \\ \vdots \\ C_i \end{pmatrix}$$

각각의 속성들이 다른 범위를 가지고 있기 때문에, 이를 표준화하는 것이 필요하다. 이에 각각의 속성 및 공사비는 표준 정규 누적분포를 통해 0과 1 사이의 값으로 변환된다. 변환 후, 위의 행렬식은 다음과 같이 표현된다.

$$\begin{pmatrix} x_{11} & \cdots & x_{1j} \\ \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} \end{pmatrix} \times \begin{pmatrix} w_1 \\ \vdots \\ w_j \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_i \end{pmatrix}$$

그러나 정성적 속성은 표준화가 불가능하기 때문에 위 방법을 적용할 수 없다. 이에 본 연구는 최적화 대상에 정성적 속성의 정량화된 속성값을 추가하여 이를 해결하였다. 정량적 속성 p 와 정성적 속성 q 를 구분하였을 때, 위 행렬은 다음과 같이 표현될 수 있다.

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

정성적 속성값의 정량화를 위해, 각각의 동일 속성값에 해당하는 랜덤변수를 할당한다. 예를 들어, '지붕형태'라는 정성적 속성이 '평지붕', '박공지붕', '경사지붕'의 세 가지의 속성값을 가질 경우, 각각의 속성값에 0에서 1 사이의 값을 갖는 변수를 할당한다. 표준화된 공사비와 각 해가 도출하는 수치와의 절대거리 d 를 세우고 유전 알고리즘을 통해 모든 거리의 합을 최소화하도록 함으로써 이를 최적화한다. 이를 통해 실제 사례들을 가장 잘 설명해줄 수 있는 가중치의 조합과 더불어, 앞서 할당하였던 정성적 속성의 각 속성값들을 정량화할 수 있다.

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

2. 조회 오차 보정 방법

문제와 유사사례간의 차이에 의해 발생하는 오차를 조회 오차(Retrieving Error)라고 정의하였다. 하나의 속성에 대해, 속성값의 차이가 공사비에 미치는 영향은 다음의 식과 같이 표현될 수 있다.

$$e_{ij} = w_j(x_{ti} - x_{ij})$$

이를 모든 속성에 대해 확장하면 아래의 식과 같으며, 이때 어느 하나의 케이스의 조회 오차가 된다. 유사사례의 공사비에 조회 오차를 더하여 보정된 공사비를 구하게 되며, 이는 표준화된 값이기 때문에 표준정규누적분포의 역함수를 통해 보정된 공사비를 원단위로 변환한다.

$$E_i = \sum_{j=1}^n e_{ij} = \sum_{j=1}^n w_j(x_{tj} - x_{ij})$$

3. 개선된 가중 평균 방법

가중 평균 방법은 사례유사도를 통해 더 유사한 사례에 더 높은 가중치를 부여하는 방법이지만, 유사사례간의 유사도의 비가 상대적으로 작기 때문에 차이를 반영하기 어렵다는 단점이 있다. 이에 유사도를 직접 활용하는 대신, 유사도의 표준정규누적분포값을 가중치로 활용하는 개선된 가중 평균 방법을 제안한다. 아래와 같이 표준정규누적분포 함수를 통해, 기존의 유사도를 0에서 1 사이의 값으로 변환한다. 이처럼 유사사례의 유사도의 상대적인 위치를 가중치로 활용함으로써, 더 유사한 사례의 상대적인 영향력을 향상시킬 수 있다.

$$\text{when } F(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] dt,$$

$$s_k = F^{-1}(s_k|\mu, \sigma)$$

4. 공사비 예측 프로세스

제안된 방법을 활용한 공사비 예측 프로세스는 다음과 같다. 문제가 주어지면, 유사도 계산을 통해 유사사례를 도출하고, 이를 보정하여 해결책을 도출하게 된다. 도출된 해결책은 실제 문제 해결에 적용되고, 적합하지 않을 경우 이를 수정하게 된다.

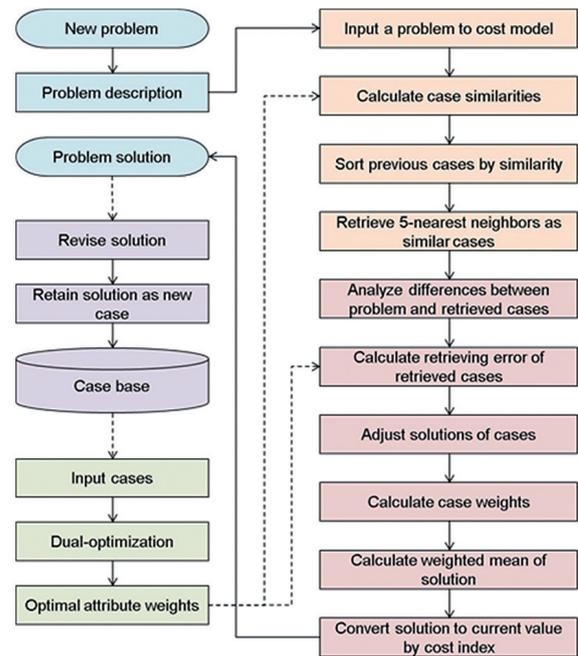


그림 3. Cost Estimation Process

도출된 해결책은 새로운 사례로 사례 베이스에 저장되고, 주기적 혹은 일정 수 이상의 사례가 추가되었을 때 Dual-optimization을 통해 가중치를 업데이트하게 된다. 시스템이 사용될수록 사례의 수가 증가하고, 저장된 사례들에 더 적합한 가중치가 업데이트되면서 사례기반추론 공사비 예측 모델의 정확도가 향상된다.

IV. 결론

본 연구는 기존의 방법들의 단점을 보완한 새로운 세 가지 방법을 제안하였다는 데에 그 의미가 있다. Dual-optimization 방법은 정성적 속성의 속성가중치 계산 및 속성값의 정량화를 가능하게 하여 공사비 예측에 활용 가능한 속성의 수를 증가시킬 수 있으며, 조회 에러 보정 방법은 추가적인 분석이나 데이터베이스 없이 문제와 사례간의 차이에 대한 보정이 가능하다는 장점이 있다. 개선된 가중 평균 방법은 기존의 방법에 비해 더 유사한 사례의 영향을 증가시켜 더욱 정확한 해결책 도출을 가능하게 한다. 결과적으로 이를 적용한 새로운 GA-CBR 공사비 예측 모델은 더 정확한 공사비 예측을 가능하게 함으로써, 건설 프로젝트의 참여자들에게 더 정확한 의사결정을 지원하는 도구로 활용될 수 있다. 또한 Dual-optimization은 범용적인 방법이기 때문에, 가중치 산정이 필요한 다른 분야의 문제에도

적용이 가능하다.

정성변수의 수가 많아질수록 계산시간이 오래 걸린다는 점은 사용성을 저해하는 요인이다. 그러나 이러한 문제는 컴퓨터 과학의 발전, 알고리즘의 개선 등을 통해 충분히 해결 가능할 것으로 기대된다. 향후 연구에서는 사례기반추론의 수정 및 저장 단계에 대한 추가적인 연구를 통해, 본 연구에서 제시한 모델의 사용성을 증대시키는 것이 필요하다.

참고문헌

김수영 (2017). "Dual-optimization Method for Improving Accuracy in GA-CBR Cost Estimating Model." 서울대학교 박사학위논문

Aamodt, A., & Plaza, E. (1994). "Case-based reasoning: Foundational issues, methodological variations, and system approaches." *AI communications*, 7(1), 39-59.

Arditi, D., & Tokdemir, O. B. (1999). "Comparison of case-based reasoning and artificial neural networks." *Journal of computing in civil engineering*, 13(3), 162-169.

Barrie, D. S., & Paulson Jr, B. C. (2000). *Professional construction management: including CM, design-construct, and general contracting*. McGraw-Hill, 3rd edition.

Duverlie, P., & Castelain, J. M. (1999). "Cost estimation during design step: parametric method versus case based reasoning method." *The international journal of advanced manufacturing technology*, 15(12), 895-906.

Goh, Y. M., & Chua, D. K. H. (2009). "Case-based reasoning approach to construction safety hazard identification: adaptation and utilization." *Journal of Construction Engineering and Management*, 136(2), 170-178.

Hu, J., Qi, J., & Peng, Y. (2015). "New CBR adaptation method combining with problem-solution relational analysis for mechanical design." *Computers in Industry*, 66, 41-51.

Ji, S. H., Park, M., & Lee, H. S. (2011). "Cost estimation model for building projects using case-based reasoning." *Canadian Journal of Civil Engineering*, 38(5), 570-581.

Kolodner, J. (1993). *Case-based reasoning*, Morgan

Kaufmann Publishers Inc., New York, USA.

Koo, C., Hong, T., & Hyun, C. (2011). "The development of a construction cost prediction model with improved prediction capacity using the advanced CBR approach." *Expert Systems with Applications*, 38(7), 8597-8606.

Leake, D. B. (1996). *Case-Based Reasoning: Experiences, lessons and future directions*, MIT press.

Paulson Jr, B. C. (1976). "Designing to reduce construction costs." *Journal of the construction division*, 102(C04).

Trost, S. M., & Oberlender, G. D. (2003). "Predicting accuracy of early cost estimates using factor analysis and multivariate regression." *Journal of Construction Engineering and Management*, 129(2), 198-204.

Watson, I., & Marir, F. (1994). "Case-based reasoning: A review." *Knowledge Engineering Review*, 9(4), 327-354.

■ 본 논고는 저자의 박사학위논문(김수영, 2017)에서 발췌, 요약되었음.