

사례기반추론을 이용한 공동주택의 월간투입비용 예측모델 개발에 관한 연구

A Study on Developing a Case-based Forecasting Model for Monthly Expenditures of Residential Building Projects

이 준 성*

Yi, June Seong

요 약

건설업은 타산업, 특히 제조업 분야와 비하여 시장구성이 분절적이며 시장상황의 변동이 지극히 유동적이다. 그간 많은 연구자들이 기존의 비용예측 기법들을 통해 이러한 건설부문의 특성을 반영, 그 예측력을 높이기 위하여 많은 노력을 경주해 오고 있다. 본 연구에서는 공동주택 비용투입 형태를 예측함에 있어서 사례기반추론(Case-Based Reasoning : CBR)을 이용, 기존 사업에 대한 실적데이터를 활용함에 있어서 유사한 프로젝트를 일정기준에 의하여 선별, 해당 비용정보만을 참조하여 향후 비용투입을 예측함으로써 그 정확도를 향상시키고자 하였다. 비용예측모델의 정확도를 제고하기 위해 비교 프로젝트간의 유사성을 비교함에 있어서, 비용정보는 세 가지 수준의 공종분류, 즉 전체프로젝트 수준, 7개 대공종 분류, 총 20개 세부공종별 분류에 따라 분석하였다. 본 연구의 결론은 데이터베이스화된 자료 중, 유사성을 계량화한 후 유사성평가 상위 12~19 %의 프로젝트의 정보만을 참조하는 것이 그 예측도를 극대화시킬 수 있는 것으로 판명되었다.

키워드: 비용예측, 사례기반추론(Case-based reasoning), 공동주택, 워크패키지(Work package)

1. 서론

1.1 연구의 배경 및 목적

프로젝트를 수행함에 있어서 사업관리자로 하여금 투입비용에 대한 정확한 예측 및 변이점의 조기발견은 전체 사업성패에 매우 중요한 역할을 한다. 공사진행에 필요한 비용조달과 더불어 기간별 투입비용을 예측하는 것은 사업의 성패는 물론 해당 기업의 경영안정을 위해서 필수적인 요소가 되고 있다. 건설시장의 경쟁이 더욱 심화되고 있는 현 상황에서 현금흐름을 예측하고 이에 대한 대비를 하는 것은 기업의 생존에 있어서 가장 필수적인 전략이자 생존조건이 될 것이다. 그러나 이러한 중요성에도 불구하고 건설사에서는 각개 프로젝트의 기간별 비용예측 정확성을 향상시키고자 하는 적극적 노력이 부족한 실정으로서,

대부분 본사에서 각 현장으로의 유선연락을 통한 보고를 통하여 산별적으로 이루어지고 있는 형편이다. 본 연구에서는 이렇듯 그 중요성에 비하여 체계적이지 못한 기간별 투입비용예측모델을 개발하고자 하며, 건설업의 특성을 충분히 반영하여 과거의 실적데이터 중 유사한 프로젝트의 비용정보만을 선별적으로 참조하여 그 정확성을 제고하고자 한다.

최근 들어 여러 통계모델, 데이터베이스, 기계학습기법(machine learning) 등의 도움으로 연구자들은 대량의 정보를 분석하는 노력에 결실을 맺고 있으나, 축적된 데이터의 유효성과 그 효율적 이용에 의문이 제기되고 있다. 이용자로 하여금 원하는 정보를 적절하게 획득할 수 있는 방안을 제시하는 것이 이러한 도구 및 기법의 활용을 의미있는 연구로 완성시켜줄 수 있을 것이다.

사업관리자는 프로젝트 진행에 따라 새롭게 생산·획득되는 신규 데이터들과 더불어 기 축적된 정보 및 경험을 기반으로 하여 문제해결을 위한 결정을 하도록 끊임없이 요구되어진다. 대부분의 경우 현재에서 맞닥뜨리는 문제와 과거의 해결방안이 정확하게 일치되는 경우는 없을 것이며, 이에 대한 응용은 전적으로 담당자의 문제인식 및 해결능력에 좌우된다고 할 수 있다. 이

* 중신회원, 이화여자대학교 건축학부 전임강사, 공학박사(교신저자),
jsyi@ewha.ac.kr

이 연구는 2005학년도 이화여자대학교 교내연구비 지원에 의한 연구임. 과제번호 2005-0991-1

러한 관점에서 프로젝트 초기단계에서의 월간투입비용에 대한 예측 또한 사업관리자의 주관적 경험에 크게 의존한다고 볼 수 있으며, 관련정보가 점차 축적되어진 후 이러한 주관적 판단이 차차 보정되어진다고 할 수 있을 것이다. 기존의 많은 기법들은 과거의 실적데이터를 활용함에 있어서 현재 프로젝트가 진행되는 상황을 반영한 후 유사사례의 선별참조를 하고자 하는 노력에 많은 관심을 기울이지 못하고, 기 실적자료의 평균치나 주관적 판단에 근거하여 비용을 예측하는 경우가 일반적이다. 이러한 상황에서 과거의 유용정보를 활용함에 있어서 현재의 상황을 심분 반영한 후, 이와 유사한 과거 유사프로젝트의 정보를 선별할 수 있는 체계적 기법의 필요성이 요구되어진다.

투입비용 예측모델을 개발함에 있어서 시장변화의 역동성 및 구조적 분절성이라는 건설산업의 특성으로 인하여 제기되는 장애요인은 다음과 같다. 첫째, 프로젝트 중심(project-based)의 시스템으로 인하여 과거의 사업으로부터 획득된 정보와 지식이 공유되지 못하고 있으며, 둘째로는 프로젝트 진도측정에 대한 기준이 모호하다는 것이다. 대부분 프로젝트의 진행사항은 작업의 난이도를 고려하기보다는 단순히 기간 투입(결제)된 비용을 통하여 단순계량되고 있다고 할 수 있다. 마지막으로 공종별 투입비용 패턴분석의 중요성이 간과되고 있다는 점이다. 대부분 사업진행의 추이를 모니터링함에 있어서 전체 프로젝트 단계에서의 비용투입만을 주목함으로써 공종간 상호상쇄에 따라 노출되지 않는 문제점을 파악하지 못하고 있는 것으로 보인다.

1.2 연구의 범위 및 방법

본 연구에서는 당면한 문제를 해결함에 있어서 과거의 유사사례에서 그 해결방안을 도출하는 사례기반추론(Case-based reasoning)을 근간으로 하고 있다. 이는 전술하였던 문제점들을 극복하고 효율적인 비용예측모델을 구축하는데 있어서 가장 적합한 방안으로 평가되어지기 때문이다. 본 논문의 연구범위는 국내건설업에서 상당한 비중을 차지하고 있는 공동주택 사업에 있어서의 월간투입비용 예측으로 설정하였다. 연구진행은 각 세부공종별로 정리된 비용투입 정보를 수집하여 과거실적 프로젝트 데이터베이스를 구축한 후, 일부 진행중인 현행 프로젝트와 축적된 사례프로젝트를 현재 진행시점까지 비용투입 유형의 유사성에 의거하여 순위를 부여한다. 이와 같은 유사성 순위에 의해 구성된 과거실적 프로젝트 목록에서 유사성이 높다고 평가된 상위 몇 % 정도의 사례를 선별·참조하는 것이 현재진행 공사의 월간 비용투입 예측에 있어서 최상의 정확도를 제공하는지 그 결과를 확인한다. 이러한 과정을 통해 도출된 참조비율이 적정인지 그 유효성을 검증하기 위해서는 구축된 데이터베이스와

별도로 준비된 과거 실적 프로젝트의 사례를 활용하여 예측 정확도를 기존의 방식과 비교하도록 한다. 수집된 총 98개의 비용 정보 자료 중에서 88개의 자료로 사례 데이터베이스를 구축하였고, 제외된 10개의 사례를 이용하여 모델의 유효성을 검증하였다.

2. 이론적 고찰

2.1 기간별 투입비용(monthly expenditure) 예측

프로젝트 진행에 있어서 기간별 투입비용을 예측하는 것은 전체 사업단계에 걸쳐 사업 수익성분석, 협력업체 기성지급, 공사기간 추정, 계획대비 실적비교 관리 등에 기초자료로 사용되므로 사업의 성공에 영향을 미치는 핵심적인 관리요소로 평가되고 있다. 이렇듯 투입비용예측은 사업관리자로 하여금 현금흐름 파악에 있어서 큰 변수요인으로 간주되는 바 이에 대한 예측정확도를 높임으로써 수행프로젝트의 성공가능성을 제고시킨다고 할 수 있다. 이와 관련된 기존의 연구들을 고찰해보면, 우선 Peterman(1972)은 공정 바차트(bar chart)와 계약단가에 근거하여 도급업체의 현금흐름을 예측하는 전산화 도구를 제안하였고, Ashley와 Teicholz (1977)는 공정진행정도에 근거하여 전체 사업기간에 걸친 현금흐름을 분석하고자 하였다. Peer(1982)는 시공전 계획단계에서 실적자료의 회귀분석을 기초로하여 현금투입을 예측하였다. Pattern(1982)은 경영도구로서 현금흐름파악의 중요성을 강조하였는데, 과거의 실적치를 현행사업 진행시 판단기준으로 준용하고자하는 노력을 보여주었다.

한편 Singh 과 Woon(1984)은 그들의 연구에서 프로젝트 특성과 투입비용 경향사이에 깊은 연관성이 있음을 주장하였는데, Oliver(1984)와 Kenley 등(1986)은 각자의 연구에서 별개의 프로젝트는 각각 고유한 공정 진행율을 보여주며 결과적으로 개별적인 비용투입 경향을 나타낸다는 것을 제시하였다. 이유섭(2003)은 공사비 구성 주요항목 분석을 통하여 공사비를 예측하고자 하였으며, 양용철 등(2003)은 대단지 공동주택 건설공사의 계획기성 산출프로세스의 개선방안을 제시하였다. 전재열(2003)은 실적자료 분석에 따라 적정 공사비를 선정하는 시스템을 구축하고자 하였다.

2.2. 사례기반추론(Case-Based Reasoning)

2.2.1 이론개요

프로젝트 수행은 관리자에게 있어서는 무수한 의사결정과정의 연속이라 할 수 있으며, 이러한 의사결정의 기초가 되는 방대

한 정보들을 처리하기 위해서는 효율적인 전산기법의 도움이 필수 불가결한 요소가 되었다. 특히 비용예측 분야처럼 그 결과치에 대하여 정량적인 값을 도출하는 부문에서는 통계적 방법이나 신경망 이론과 같은 인공지능(AI: Artificial Intelligence)기법을 활용하여 사용자의 의사결정을 지원하는 많은 연구 성과물들이 도출되고 있다. 이중 사례기반추론은 인간이 사고하는 과정과 가장 유사한 방식으로 인지되고 있는바, 현재의 문제를 해결함에 있어서 과거의 문제해결에 적용되었던 해결방안을 참조한 후, 상호간의 차이점을 분석하여 이를 현재의 당면문제에 적절하도록 수정 후 적용하는 것이다. 그동안 다수의 전문가시스템(expert system)에서 활용되는 규칙기반추론(rule-based reasoning)에서는 축적된 모든 지식을 규칙으로 구축한 후, 규칙베이스(rule base)에 따라서 문제의 해를 도출하지만, 이는 문제해결을 위한 모든 사안들을 규칙화한다는게 현실적으로 불가능함에 따라 복잡한 사안에 현안에 있어서는 그 실효성에 의문이 생길 수밖에 없다. 사례기반추론(case-based reasoning)에서는 과거사례를 참조, 현상황에 맞도록 수정·조치함으로써 기존의 방식에서 제기되었던 난관을 극복하고 있다. 아래 [그림 1]은 사례기반추론기법의 근간이 되는 절차로써 흔히 '4REs'라고 표현되고 있으며, 각 단계별 대략적인 설명이다.(Aamodt, 1994)

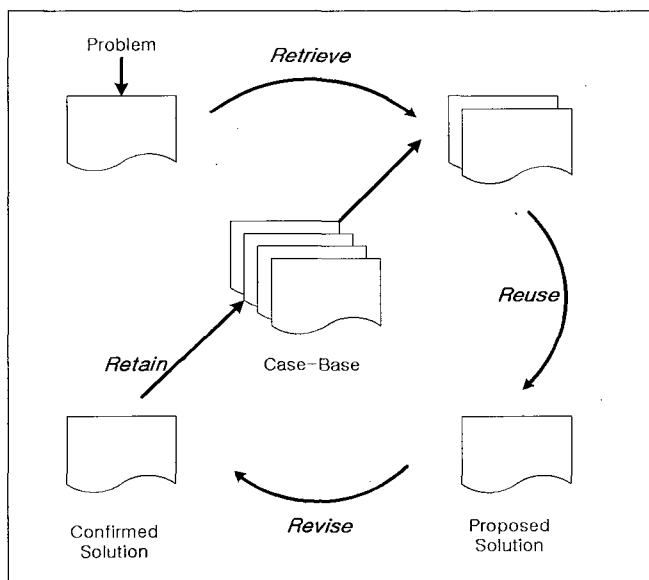


그림 1. 사례기반추론 사이클(CBR Cycle)

(1) 조회(REtrieve)

당면한 새로운 문제(problem)와 가장 유사한 과거사례(들)를 구축된 사례 데이터베이스로부터 찾아낸다. 이 단계에서 가장 중요한 관점은 상호간의 유사성을 어떠한 관점에서 매치(match)시킬 것인가에 대한 논의일 것이다.

(2) 재사용(REuse)

당면 문제에 대한 해를 도출하기 위하여 유사사례로서 조회된 과거사례의 데이터와 지식을 재사용하도록 한다.

(3) 수정(REvise)

현문제와 조회된 사례사이의 차이점을 분석, 필요에 따라 적절하게 해결방안을 수정한다.

(4) 저장(REtain)

문제해결에 적용된 금번의 해결책을 향후 새로운 문제제기 시에 참조·사용할 수 있도록 사례 데이터베이스에 저장한다.

2.2.2 CBR을 활용한 비용예측에 관한 연구

비용예측과 관련한 기존의 연구를 살펴보면, 정량적 결과물 도출에 적합성을 나타내는 각종 통계적 기법 및 신경망 이론 등을 활용한 연구 성과물들이 주류를 이루고 있으나, Stotiller(1992)는 불확실한 정보제공 상황에서 경영 수익률을 확보하기 위한 원가 및 판매고를 예측하는 모델을 사례기반추론을 활용하여 개발하였다. 그는 해당연구에서 과거의 유사사례의 정보를 기초로 하여, 판매조직구성, 생산량 및 재고량 설정, 금융조달과 같은 경영계획을 수립하고자 하였다. Lee 등(1995)은 사례기반추론을 적용하여 현금흐름(cash flow)을 예측하는 시스템을 개발하였는데, 기존의 통상적인 가중평균값(weighted mean)의 적용대신에 퍼지함수(fuzzy integral)를 이용하여 종합적인 유사성 평가방식을 제안하였다. 본 연구에서는 그간 정성적인 정보, 혹은 텍스트정보처리에 그 적합성이 강조되었던 사례기반추론(CBR)이 비용예측과 같은 정량적 결과치를 도출함에 있어서도 효율적으로 활용될 수 있음을 제시하였다.

3. CBR을 활용한 투입비용 예측모델 개발

3.1 개발모델의 논리전개 구성

본 연구의 주요 관심사는 현재 수행중인 프로젝트의 기간별 투입비용을 예측함에 있어서 구축된 전체 데이터베이스에서 유사한 과거 사례들만을 선별하여, 필요정보만을 선택적으로 참조하여 그 예측정확도를 향상시키는 방안을 도출하는데 있다. 이를 위한 유사사례의 선정 및 투입비용예측을 위한 논리전개는 다음과 같은 네 가지 단계를 따른다: 1) 투입비용 실적데이터의 표준화, 2)비용정보의 전환, 3)현재 프로젝트와 과거사례프로젝트간의 유사성 평가작업, 4)유사프로젝트로 선별된 과거프로젝트들의 투입비용정보를 기초로 한 새로운 투입비용 곡선의 작성

및 이에 따른 월간투입비용 산출.

(1) 데이터 표준화

개발모델의 첫 단계로서 먼저 수집된 과거 프로젝트의 월간실적 비용정보는 해당프로젝트의 전체공사비에 대한 백분율로 변환된다. 이러한 투입비용정보는 차후에 예측하고자하는 현행 프로젝트의 공사기간을 기준으로 하여 재조정된다.

(2) 데이터 전환

사례 데이터베이스에 저장된 각 프로젝트는 각각 공사기간 진행에 따른 투입비용 분포곡선을 생성함에 있어서 현재 예측하고자하는 프로젝트의 여건(공기)을 감안하여 재조정되며, 이러한 분포곡선은 다음의 유사사례 조희단계에서 프로젝트 상호간의 유사성 판정을 위한 기초자료로 이용된다.

(3) 유사사례 조희

현재 프로젝트와 과거사례프로젝트간의 유사성을 평가함에 있어서, 그 계량척도로서는 비교대상 프로젝트에 있어서 각각 전체공사비 대비 월간투입비용 비율간의 편차제곱합(SSD: sum of squares of differences)을 준용하는데, 그 합계치가 작을수록 비교프로젝트는 현재 비용예측을 수행하고자하는 프로젝트와 유사하다고 평가된다.¹⁾ 데이터베이스에 저장된 모든 과거사례들은 이러한 평가과정을 통해, SSD 수치가 작은(유사성이 크다고 판명된) 프로젝트부터 순차적으로 나열된다.

(4) 투입비용 예측을 위한 투입진도율곡선 생성

상기 과정에서 데이터베이스에 있는 모든 사례 프로젝트들은 현행 프로젝트와의 유사정도에 따라 순위가 부여된다. 차후 모델검증을 통해서 결정된 적정 유사사례 선택률을 적용하여 향후 투입비 동향을 예측하는데 참조할 사례프로젝트를 선정한 후, 해당 자료를 기초로 하여 기간별 투입비용 예측을 위한 진도율곡선을 생성한다.

1) 본 논문에서 유사성 평가기준으로 준용되는 편차제곱합(SSD)은 각 세부공종별 월별 투입금액을 전체공사비대비 투입율로 환산한 후, 비교대상 프로젝트간 대응되는 행렬(동일공종에서 동일비교기간에 해당되는 위치)의 수치편차를 제곱한 후 그 총합을 산출하였다. 만일 비교대상 프로젝트간 SSD값이 0이라 한다면, 해당프로젝트들은 비용투입형태 측면에 있어서 완벽하게 일치한다고 할 수 있을 것이다. 이러한 접근은 현재 진행중인 공사가 공사착공 후 비용예측시점까지 보여주는 각 세부공종별 비용투입형태와 유사성을 보여주는 과거사례 프로젝트를 선별하고자하는 시도로써, 각 연관 공종별 착수시점과 각각의 공정진행율이 반영된다. 이를 위한 비용데이터의 표준화 및 세부계산과정은 본문 [그림 5]와 식(1)에서 상세히 기술하였다.

3.2 데이터베이스 구축

3.2.1 수집데이터 특성요약

본 연구수행을 위해 수집된 데이터는 공동주택 프로젝트로 그 범위가 한정되어 있으며, 총 99개의 프로젝트 정보가 수집되어, 이중 모델검증을 위해 제외된 10개 프로젝트를 제외한 88개의 공동주택 프로젝트들이 사례 데이터베이스로 활용되었다. 전체 데이터베이스에 대한 특성분석은 아래 [표 1]과 같으며, 공사도금액 및 공사수행기간을 기준으로 한 구성비는 아래 [그림 2]와 같다.

수집된 사례 프로젝트로 데이터베이스를 구축함에 있어서 다음의 세 가지 논리적 가정이 전제되었는데, 첫 번째로는 프로젝트 수행년도의 차이에 따른 공사비 할증을 고려하지 않았다. 수집된 사례들은 1995년과 2003년 사이에 수행된 프로젝트로서 각 기간별로 자재비, 인건비, 장비비 등에 대한 개별적인 물가변동요인이 반영되어야 하나, 본 연구의 분석에서는 이에 대한 고려가 없었으며, 향후 연구에서 이러한 물가변동요인을 적절하게 반영한다면 그 예측정확도가 제고된 더욱 신뢰성 높은 모델을 구축할 수 있을 것으로 기대된다. 두 번째로는 모든 사례프로젝트에 있어서 월간 투입비용은 일정한 시점 즉, 매월 말일에 결산되는 것으로 가정하였다. 마지막으로 사례프로젝트의 장소적

표 1. 구축된 데이터베이스 특성요약

Variable	N	Mean	Median	TrMean	StDev	SE Mean
공기(월)	88	29.386	29.000	29.288	4.290	0.457
연면적(m ²)	88	907270	638573	790516	897288	95651
지하층수	88	1.6364	2.0000	1.6000	0.6640	0.0708
지상층수	88	20.307	20.000	20.263	4.499	0.480
동수	88	7.091	6.000	6.538	5.364	0.572
세대수	88	591.9	452.5	563.9	401.6	42.8
공사비(백만)	88	41829	34154	39568	27557	2938

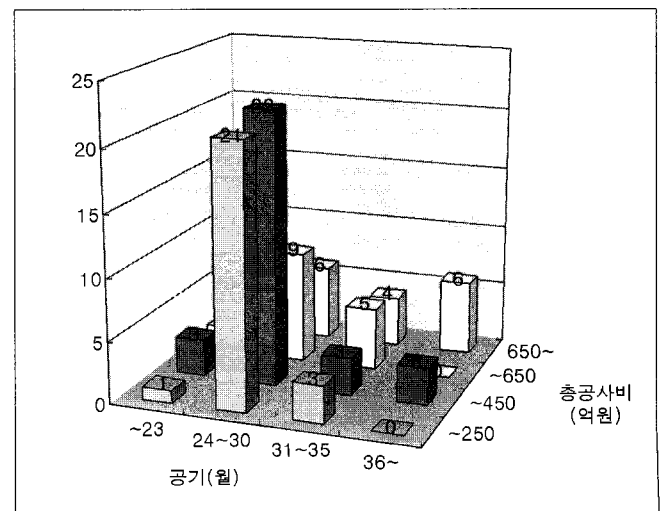


그림 2. 사례프로젝트 구성요약(공기 및 공사비)

차이(수도권 혹은 지방)에 대한 보정은 고려되지 않았다. 향후 관련 비용정보가 축적된다면 언급된 세 가지 사안들을 반영하여 더욱 개선된 비용예측모델이 얻어질 수 있을 것으로 기대된다.

3.2.2 데이터 형식 및 구성

연구를 위해 수집된 각 프로젝트별 투입비용정보는 월별투입 금액이 공종별로 구분되어 있다. 공동주택 사업수행에 따른 공종구성은 프로젝트정보 제공사의 분류체계에 준하고 있는데, 세분화정도는 7개 대공종(공통비, 건축, 전기, 설비, 토목, 부속동공사, 조경)으로 분류되며, 이는 다시 총 20개 세부공종(6개 대공종 + '건축'의 14개 세부공종)까지 세분화가 가능하도록 구성되어 있다. 비용정보는 각 데이터베이스 구성사례들의 통일화 및 표준화를 위해서 각각 전체공사비에 대한 백분율로 변환될 것이다. 향후 축적되는 사례들이 지속적으로 증가된다면 더욱 안정적이고 신뢰성 있는 예측이 가능한 데이터베이스 구축이 가능할 것으로 기대된다.

3.2.3 표준화된 투입비용 진도율곡선

공사진도율 혹은 투입비용 누적분포곡선을 도출하기 위하여 통계적 기법을 이용한 노력이 계속되고 있으며, 충분한 데이터 세트(data set)확보를 통한 통계적 안정성을 유지한 결과물들이 제공되고 있다. 본 연구의 산출물 중 하나로서 88개의 사례프로젝트에서 도출된 각 세부공종 및 대공종별, 혹은 전체프로젝트 수준에서의 진도율곡선이 얻어지는데, 아래 [그림 3]은 그 일례로 20개 세부공종 중 하나인 콘크리트공종의 기간별 투입비용 비율 및 누적분포도를 보여주고 있다. 데이터베이스를 구성하는 사례프로젝트들은 표준화를 위해서 횡축의 공사기간 및 종축의 공사비 모두 각각 백분율로 표시되었다.

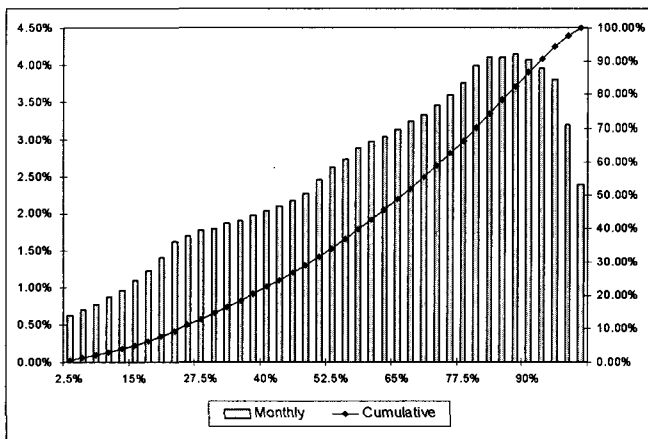


그림 3. 기간별 투입비용 진도율 및 누적분포도(콘크리트공종)

3.3 모델개발

사례기반추론을 이용한 비용예측모델을 개발함에 있어서 그 시작은 데이터세트(data set) 구조에 대한 명확한 정의 및 표준화라 할 수 있다. 본 연구의 근간이 되는 88개 프로젝트의 비용데이터들은 각각 20개 세부공종수준에서의 월간투입비용으로서, 그 구성은 기본적으로 자료제공 업체의 회계관리상 구분에 준하고 있다. 본 연구에서 제시한 투입비용 예측모델은 1)데이터 표준화, 2)데이터 전환, 3)유사사례 조회 및 4)기간별 투입비용예측의 4가지 모듈로 구성되어 있으며 그 논리전개는 [그림 4]와 같다.

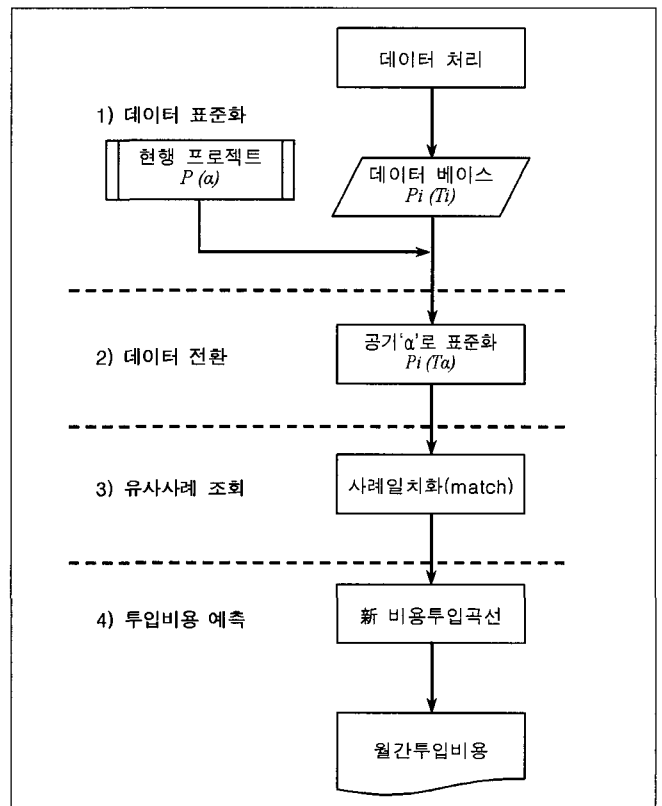


그림 4. 월간투입비용 예측모델 논리전개도

3.3.1 데이터표준화 모듈

데이터 전환시점에서 고려되어야 할 주안점은 데이터베이스를 구성하는 데이터세트(data set) 제공원의 현행 시스템과 비교, 급격한 변화를 피해야 하는데 이는 데이터입력 및 비용예측을 수행하는 사용자의 이용편리성을 위해서도 필요한 고려사항이다. 개발모델의 첫 단계로서 먼저 수집된 과거 프로젝트의 월간실적 비용정보는 해당프로젝트의 전체공사비에 대한 백분율로 변환된다. 이러한 투입비용정보는 차후 데이터전환 모듈에서 예측하고자 하는 현행 프로젝트의 공사기간을 기준으로 하여 재조정된다. 본 모듈에서의 또 다른 기능은 유사성 평가시점에서 어느 공종수준(전체프로젝트 레벨, 7개 대공종 레벨, 20개 세부공종 레벨)의 SSD 수치를 평가기준으로 선택할 것인가를 결정

하도록 지원한다. 전체 데이터베이스에서의 유사성 상위사례 참조비율과 공중수준 결정은 각 프로젝트 유형에 따라 모델검증 단계에서 도출된 수치를 적용하게 될 것이다.

3.3.2 데이터전환 모듈

데이터전환 모듈에서는 선행단계에서 전체공사비 대비 백분율로 표준화된 각 사례프로젝트의 비용투입 진도곡선을, 예측하고자하는 현행프로젝트의 공기에 대응하도록 재조정한다. 일례로 현행 프로젝트의 예정공기가 30개월로 추정되고 있다면 각 데이터세트의 횡축은 보간법을 이용, 총 30개로 분할되어 각 분기점별로 구간별 비용투입율이 도출될 것이다. 이때 근접하는 구간에서 비용투입은 선형적 특성을 나타낸다는 가정을 전제로 하고 있다.

3.3.3 유사사례조회(case matching) 모듈

유사사례조회 모듈은 구축된 데이터베이스에서 현행 프로젝트의 비용투입 패턴과 가장 유사한 기존사례들을 추출하는 기능을 하는 전체 모델 구성상 가장 중추적인 부분이라고 할 수 있다. 유사성 순위는 데이터베이스에 저장된 과거의 모든 프로젝트들의 기간별 비용정보가 현재 진행중인 프로젝트의 공기에 준하여 동기화(synchronized) 된 후, 각각 대응되는 셀(cell) 상호간의 편차제곱합(SSD: sum of squares of differences)에 의해서 결정되어진다. 이때 SSD를 산출함에 있어서 공중구분은 3단계(전체 프로젝트, 7개 대공중, 20개 세부공중) 중, 모델검증의 결과에 의거하여 선택적으로 정해질 수 있다. 공식(1)과 [그림 5]는 7개 대공중 수준에서의 SSD 계상의 일례를 보여주고 있다.

$$SSD = \sum_{i=1}^I \sum_{j=1}^n (W_f ij - W_d ij)^2 \dots\dots\dots(1)$$

- W_f : 현재 진행중인 프로젝트의 투입비용 진도백분율
- W_d : 데이터베이스내 사례프로젝트 투입비용 진도백분율
- i : 워패키지 ($i=1, 2, \dots, I$), 7개 대공중 분석에서 $I=7$.
- j : 개월 ($j=1, 2, \dots, n$)
- n : 비용예측 실행월 (착공 후 n 번째 월)

비용예측 대상인 현재 수행중인 프로젝트 W_f

i	Description	Month			
		1	2	n-1	n
1	Building	$W_{f1,1}$	$W_{f1,2}$	$W_{f1,n-1}$	$W_{f1,n}$
2	Electrical	$W_{f2,1}$	$W_{f2,2}$	$W_{f2,n-1}$	$W_{f2,n}$
3	Mechanical	$W_{f3,1}$	$W_{f3,2}$	$W_{f3,n-1}$	$W_{f3,n}$
4	Earth Work	$W_{f4,1}$	$W_{f4,2}$	$W_{f4,n-2}$	$W_{f4,n}$
5	Annex Building	$W_{f5,1}$	$W_{f5,2}$	$W_{f5,n-2}$	$W_{f5,n}$
6	Planting	$W_{f6,1}$	$W_{f6,2}$	$W_{f6,n-1}$	$W_{f6,n}$
7	General Expenses	$W_{f7,1}$	$W_{f7,2}$	$W_{f7,n-1}$	$W_{f7,n}$

데이터베이스내의 과거 사례프로젝트 W_d

i	Description	Month			
		1	2	n-1	n
1	Building	$W_{d1,1}$	$W_{d1,2}$	$W_{d1,n-1}$	$W_{d1,n}$
2	Electrical	$W_{d2,1}$	$W_{d2,2}$	$W_{d2,n-1}$	$W_{d2,n}$
3	Mechanical	$W_{d3,1}$	$W_{d3,2}$	$W_{d3,n-1}$	$W_{d3,n}$
4	Earth Work	$W_{d4,1}$	$W_{d4,2}$	$W_{d4,n-2}$ $W_{d4,n}$
5	Annex Building	$W_{d5,1}$	$W_{d5,2}$	$W_{d5,n-2}$	$W_{d5,n}$
6	Planting	$W_{d6,1}$	$W_{d6,2}$	$W_{d6,n-1}$	$W_{d6,n}$
7	General Expenses	$W_{d7,1}$	$W_{d7,2}$	$W_{d7,n-1}$	$W_{d7,n}$

그림 5. 유사사례조회를 위해 표준화된 세부정보 (7대공중 수준에서의 SSD 산정)

3.3.4 기간별 투입비용예측 모듈

그 유사성이 인정되는 사례프로젝트들에서 도출된 비용정보들만의 평균치를 참조하여 현재 진행 중인 프로젝트의 향후 월간 투입비용을 예측하는데 그 근거자료로 삼는다. 투입비용을 예측하고자 하는 프로젝트의 성격(본 논문에서는 공동주택) 및 축적된 사례프로젝트의 규모에 따라 전체프로젝트 대비 유사 사례프로젝트의 선택비율이 달라질 수 있는데, 그 값은 모델검증 단계에서 결정되어질 것이다.

모델검증 과정에서는 이처럼 전체 데이터베이스에서 선별적으로 참조할 유사 사례프로젝트의 전체 데이터베이스 대비 적용비율 뿐 아니라, 사례조회(case matching)과정에서 어느 정도 세부공중까지 분류하여 비교·검토하는 것이 더욱 향상된 예측정확성을 보여주는가에 대한 검증도 이루어지게 되며, 본 연구에서는 20개 세부공중, 7개 대공중, 전체 프로젝트 수준의 3단계로 구분하여 비교되었다.

4. 개발모델의 검증

본장에서는 금번 연구에서 제시된 기법이 공동주택의 월간투입비용을 예측함에 있어서 어느 정도 예측정확도를 향상시킬 수 있는지 검증하고자 한다. 검증은 크게 두 가지 관점에서 수행되었는데, 첫째는 데이터베이스에 구축된 전체 사례프로젝트 중 그 유사성이 인정되어 현행프로젝트의 예측에 참조된 사례프로젝트의 선택비율의 결정이다.

이 과정에서 도출된 선택비율에 의거, 선정된 프로젝트의 비용정보를 기반으로 한 투입비용 예측값과 데이터베이스내 모든 프로젝트를 근거로 하여 산출된 평균치를 근거로 한 예측치와 비교평가 될 것이다.

두 번째로는 유사성 산정의 기준이 되는 SSD를 계산함에 있어서 공중구분 수준 선택의 경우, 3가지 분석(전체 프로젝트, 7개 대공중, 20개 세부공중)중 어느 수준에서의 분석이 더욱 향상

된 예측정확도를 제공하는가를 검증할 것이다.

4.1 예측정확도 성능측정

본고에서 제시된 모델의 예측정확도는 데이터베이스구축에 이용되지 않은 별도의 10개 프로젝트를 활용하여 검증하였다. 모델성능은 비용예측 시점 후 사업 잔여기간동안에 대한 예측치 오차율(% error)을 통하여 평가되었으며, 다음의 두 단계로 수행된다.

첫 단계로는 착공 후 i 번째 달에서의 투입비용 예측치(X_{Fi})를 산출하는 것으로서, 이는 아래 식(2)에서 제시되었듯이 현행프로젝트의 비용예측을 위한 사례참조로 선택된 유사프로젝트들의 평균치에 의해서 구해진다.

$$X_{Fi} = \frac{\sum_{j=1}^n X_{Fij}}{n} \dots\dots\dots (2)$$

j : 비용예측을 위해 채택된 사례프로젝트
n : 전체DB에서 사례참조를 위해 선택된 프로젝트 갯수

두 번째 단계로는 모델예측성능평가 검증을 위해 이용되는 프로젝트의 i 번째 달에서의 실제투입비용과 모델에서 도출된 예측값 사이의 차이인 오차율(% error)을 산출하는 것으로서, 아래 식(3)의 순서에 따른다.

$$\% \text{ error at } i^{\text{th}} \text{ month} = \frac{|X_{Fi} - X_{Ai}|}{X_{Ai}} \times 100 \dots\dots (3)$$

X_{Ai}: i월 실투입비용
X_{Fi}: i월 예측투입비용

위의 식(2),(3)에서 얻어지는 오차율(% error)값은 잔여 공사기간 동안 지속적인 예측정확도 성능평가를 위해 아래 식(4)의 과정을 통하여 재검증된다.

$$\% \text{ error} = \frac{\sum_{i=k}^N \left| \frac{\sum_{j=1}^n \frac{X_{Fij}}{n} - X_{Ai}}{X_{Ai}} \right|}{(N - k)} \times 100 \dots (4)$$

X_{Ai}: i월 실투입비용
X_{Fi}: i월 예측투입비용
N : 총공사기간 (월)
k : 비용예측 시점(월)
n : 전체DB에서 사례참조를 위해 선택된 프로젝트 갯수

4.2 예측정확도 성능평가

4.2.1 공정진행에 따른 예측정확도 성능평가

본절에서 다루어질 주요 관심사는 공동주택프로젝트의 월간 투입비용을 예측함에 있어서 데이터베이스에 구축된 전체 사례 프로젝트들 중, 본고에서 제시된 유사성 기준에 의거하여 상위 몇 %의 사례들을 선택하여 현행 프로젝트의 투입비용예측의 근거자료로 참조 할 것인가이다.

현행 프로젝트에 대한 예측정확도를 측정함에 있어서, 전체 88개의 데이터세트(data set) 중 유사성 순위에 따라 정렬된 사례프로젝트들을 1개에서 88개까지 그 선택 범위를 변화시켜 가면서 예측오차율(% error)의 변화를 모니터링한 후, 최고의 예측정확도를 보여주는 전체 데이터베이스 대비 사례프로젝트 선택률을 도출하였다. 이러한 검증은 각각 공정을 35%, 50%, 70%에서 별도로 실행되었으며, 또한 각 예측시점에서는 세 가지 분석방법(전체프로젝트 수준, 7개 대공종 수준, 20개 세부공종 수준)의 선택에 따른 차별화된 결과의 발생여부를 확인하였다.

다음 [그림 6]에서 황축은 현행프로젝트의 비용예측을 위하여 참조되어지는 데이터베이스내 사례프로젝트의 개수를 나타내며, 그 범위는 유사성 평가시 최소 SSD값(최고의 유사성)을 나타낸 프로젝트 1개만의 참조에서 데이터베이스에 구축된 전체 사례의 참조를 의미하는 88개까지 증가된다. 본 사례에서 SSD 산출은 20개 세부공종수준에서 실행되었다. 검증용 프로젝트에 대한 비용예측정확도에 있어서 예측시점인 공정을 35%, 50%, 70% 모두에서 유사한 경향을 보여주고 있다. 즉, 오차율(% error)은 사례프로젝트 선택의 숫자가 증가됨에 따라 감소되어 어느 지점을 지나면 다시금 증가된다.

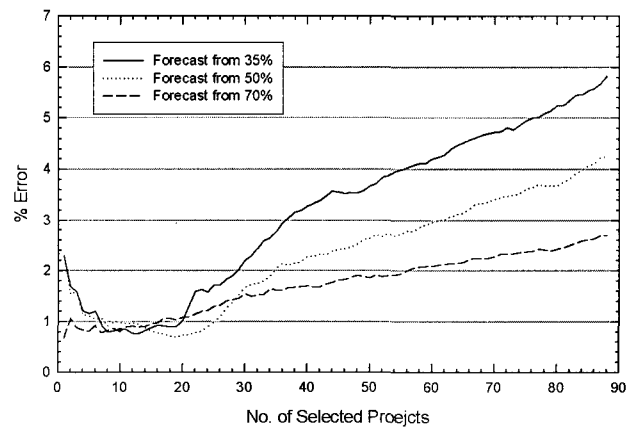


그림 6. 참조프로젝트 선택 개수 변화에 따른 오차율(% error) (20개 세부공종 수준에서의 SSD분석)

따라서 본 연구에서는 과연 최상의 예측정확도를 보여주는 변

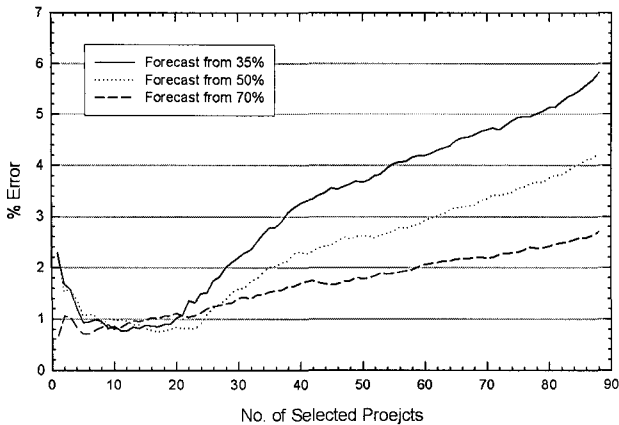


그림 7. 참조프로젝트 선택 개수 변화에 따른 오차율(% error) (7개 대공정 수준에서의 SSD분석)

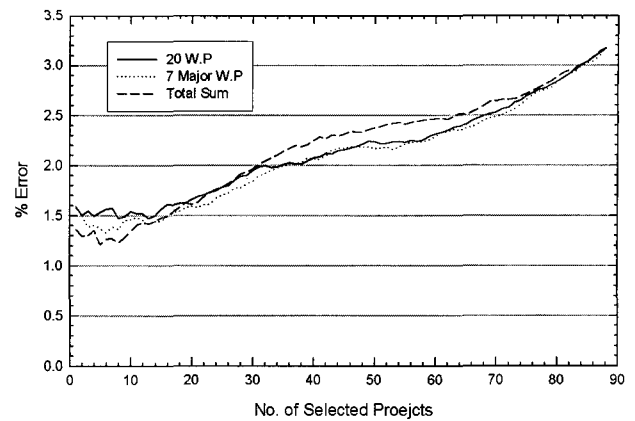


그림 9. SSD산출시 공중세분화 정도에 따른 오차율 대비 (공정진행율 50% 시점에서 예측)

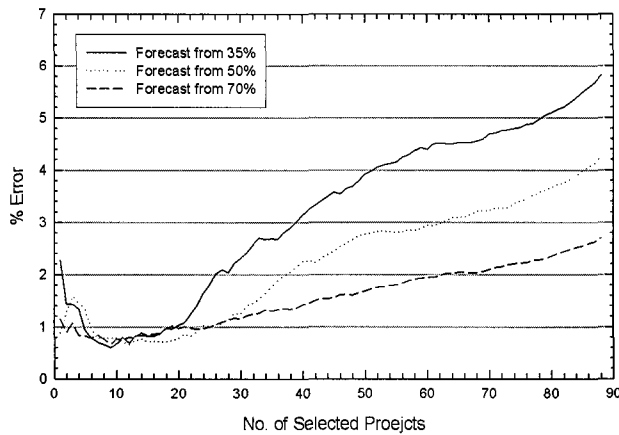


그림 8. 참조프로젝트 선택 개수 변화에 따른 오차율(% error) (전체 프로젝트 수준에서의 SSD분석)

사한 결과가 얻어지는데, 참조프로젝트로서 9개를 선택하면 그 오차율이 2.67%에서 0.78%로 개선된 결과를 얻게 된다.

20개 세부공중 수준에서 산출된 SSD를 기초로하여 얻어진 이러한 경향은 7개 대공정 수준과 전체프로젝트 수준에서의 분석에서도 동일하여, 그 결과는 아래 [그림 7]과 [그림 8]에서 그래프를 통하여 보여지고 있다.

4.2.2 공중구분 세분화도에 따른 비용투입예측

본 절에서는 본 연구에서 제시된 비용예측방안을 적용함에 있어서, 유사프로젝트의 순위선정에 기준이 되는 SSD치를 계산함에 있어서 비교프로젝트 상호간의 공중세분화 정도(전체프로젝트 수준, 7개 대공중 수준, 20개 세부공중 수준)에 따른 차별성이 유의수준에 만족할 만큼 존재하는가에 대한 검증을 하였다. 이러한 검증은 세 가지 상이한 예측시점(35%, 50%, 70%)에서 동일하게 실행되었으며, 분산분석(ANOVA)을 통하여 그 유의성을 살펴보았다.

[그림 9]는 공정진행율 50% 시점에서 예측값들에 대한 오차율(% error) 변동을 나타내고 있는데, 공중세분화 차이에 의한 SSD산출이 예측력에 큰 차이를 생성하지는 않는 것으로 결과도출 되었다. 즉, 분산분석(ANOVA)에서 얻어진 F 값은 0.64(표 2 참조)로서 이는 SSD를 산출함에 있어서 사용된 세 가지 방안이 비용투입 예측정확도에 있어서 큰 차이점을 나타내는 원인이 되기에는 불충분함을 설명하고 있다.

다른 두 시점(공정율 30%, 70%)에서도 이와 유사한 결과도출 되었는바, 이는 공동주택 프로젝트의 경우 세부공중 분류 정도에 따른 모니터링이 좀 더 향상된 예측정확도를 획득하는데 필수적인 요소가 아님을 보여준다. 이와 같은 분석결과는 해당 정보제공원이 공동주택 사업수행에 있어서 충분한 많은 경험과 회계처리 등에 대하여 일정한 규칙을 견지하고 있는 것으로 원

곡점을 찾는 것이라 정리할 수 있으며, [그림 6]의 결과에서는 공정을 35%시점에 투입비용 예측을 수행함에 있어서, 전체 88개의 데이터들 중에서 12개의 최상위 유사성 프로젝트만을 선별하여 참조하는 것이 최고의 결과를 나타내었다. 그 정량적 수치는 88개 전체프로젝트의 평균치를 이용한 예측오차율이 5.82%였음에 반하여 12개의 유사프로젝트만을 참조하였을 경우에는 0.76%로 현저하게 향상됨을 보여주고 있다.

공정을 50%와 70%에서의 비용예측에 대한 오차율 검증 또한 동일한 방법으로 실행되었다. 공사진행율 50%인 시점에서 검증 프로젝트에 대한 비용투입예측을 한 결과, 총 88개의 사례프로젝트 중 유사성 순위 상위 17개의 프로젝트들에서 추출된 비용정보의 평균치를 이용하여 예측값을 산정하면 오차율 0.69%의 결과를 얻게 되는데, 이는 전체 88개 프로젝트에 대한 평균치에 근거한 예측의 오차율 4.21%에 비교하여 월등하게 향상된 성과이다.

공정진행 70%인 시점에서의 비용예측 정확도 검증에서도 유

표 2. 분산분석(ANOVA): 20 WP, 7WP, 전체 프로젝트 (공정진행율 50% 시점에서 예측)

Source	DF	SS	MS	F	P
Factor	2	0.323	0.162	0.64	0.527
Error	261	65.734	0.252		
Total	263	66.058			

Individual 95% CIs For Mean Based on Pooled StDev					
Level	N	Mean	StDev		
20 WP	88	2.1449	0.4683	(------*-----)	
7 WP	88	2.0961	0.4929	(------*-----)	
Total	88	2.1816	0.5416	(------*-----)	

Pooled StDev = 0.5019				
	2.00	2.10	2.20	2.30

인추정되며, 공동주택 이외의 여타 프로젝트의 경우에는 상이한 결과를 도출할 개연성이 큰 것으로 보이는데 이에 대한 검증은 향후 연구에서 수행하고자 한다.

5. 결론

본 연구에서는 공동주택 프로젝트의 공사진행 중기간별 월간 투입비용을 예측함에 있어서 그 정확성을 제고하기위한 방안을 제시하였으며 주요내용은 다음과 같다.

첫째, 예측정확도 개선을 위하여 사례기반추론(CBR: Case-based reasoning) 기법을 적용하였는데, 이는 축적된 데이터베이스내 사례프로젝트 중 그 유사성이 인정되는 선별된(공동주택의 경우 데이터베이스내 전체프로젝트 개수 대비 상위 12%~19% 정도) 프로젝트의 비용정보만을 참조하여 진행 프로젝트의 향후 기간별 투입비를 예측하는 것이다. 본고에서 제시된 방안을 통한 예측정확도 향상을 살펴보면, 공정진행율 35% 시점에서는 당초 예측오차율 5.82%에서 0.76%로 개선되었다. 또한 공정진행율 50%와 70%에서도 당초 4.21%와 2.67%을 나타내던 예측오차율이 각각 0.69%와 0.78%로 현저하게 향상되었다.

둘째, 현행 프로젝트와 과거의 사례프로젝트의 유사성 평가 기준으로는 SSD(sum of squares of difference) 수치를 적용하였으며, SSD를 산출함에 있어서 상호비교 공중구분 수준은 전체프로젝트 수준, 7개 대공중 수준, 20개 세부공중 수준의 세 가지로 구분하였다. 그러나 분산분석(ANOVA) 결과는 각각 세 가지 SSD 산출방식에 따른 예측정확도의 차이는 미미한 것으로 그 결과가 도출되었다. 이는 공동주택 프로젝트의 경우에는 과거의 실적이 풍부한 바, 투입비용의 회계분기 및 예산수립 표준화가 상당부분 구축되어 나타난 결과로서 추정되며 여타 프로젝트에서는 상이한 결과가 얻어질 수 있을 것으로 예상되어지는 바, 향후 연구에서 이에 대한 추가검증이 필요할 것으로

사료된다.

본 연구는 제안된 월간 투입비용 예측정확도 개선을 통해 국내 민간건설업에서 그 비중이 큰 공동주택분야의 업무개선에 기여할 뿐 아니라, 현재 건설업 전반에서 축적되어지고 있는 방대한 데이터 정보를 효율적으로 활용할 수 있는 사례를 제시하고자 하였다. 이러한 노력은 다양한 건설참여주체들이 과거 사업에서 습득된 정보와 경험을 발전적인 지식체계로 재구축하는데 일조할 것으로 기대된다.

참고문헌

1. 양용철, 김상희, 김재준(2003). “공정-공사비 통합 프로세스 개선을 통한 대단지 아파트 건설공사의 계획기성 산출 및 관리”, 건설관리, 제4권 4호, 한국건설관리학회, pp. 123-136
2. 이윤희(2003). “코스트 중요항목 분석을 통한 공사비 예측모델 연구”, 건설관리, 제4권 4호, 한국건설관리학회, pp. 212-219
3. 전재열(2003). “실적자료 분석에 의한 적정 공사비 산정방법의 전산화 알고리즘 구축에 관한 연구”, 건설관리, 제4권 4호, 한국건설관리학회, pp. 192-200
4. Aamodt, A. and E. Plaza (1994). “Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches.” AI Communications 7, pp. 39-59.
5. Ashley, D.B. and P. Teicholz (1977). “Pre-Estimate Cash Flow Analysis.” Journal of Construction Division 10(3), pp. 111-113.
6. Kenley, R. and O. Wilson (1986). “A Construction Project Cash Flow Model ?an Ideographic Approach.” Construction Management and Economics 4, pp. 213-232.
7. Lee, R.W., Barcia, Ricardo Miranda, and Khator, Suresh K. (1995). “Case-Based Reasoning for Cash Flow Forecasting Using Fuzzy Retrieval.” First International Conference (ICCBR ? 95).
8. Oliver, J.C. (1984). “Modeling Cash Flow Projections Using a Standard Micro Computer Spreadsheet Program.” Construction Management Projects, Loughborough University of Technology.
9. Pattern, W.N. (1982). “Application of Cost Flow Forecasting Models.” Journal of Construction Division 17128, 108 C02(3), pp.226-232.

10. Peterman, G.G. (1972). "Construction Company Financial Forecasting." Arizona State University.
11. Singh, S. and P.W. Woon (1984). "Cash Flow Trends for High-rise Building Projects." Proceedings of the 4th International Symposium on Organization and Management of Construction, University of Waterloo, Canada.
12. Stottler, R.H. (1992). "Case-Based Reasoning for Bid Preparation." AI Expert(March), pp. 44-49.1.

논문제출일: 2006.02.23

심사완료일: 2006.04.04

Abstract

The objective of this research is to explore a more precise forecasting method by applying Case-based Reasoning (CBR). The newly suggested method in this study enables project managers to forecast monthly expenditures with less time and effort by retrieving and referring only projects of a similar nature, while filtering out irrelevant cases included in database. For the purpose of accurate forecasting, 1) the choice of the numbers of referring projects and 2) the better selection among three levels? which include a 20-work package level, a 7-major work package level, and a total sum level analysis, were investigated in detail. It is concluded that selecting similar projects at 12~19 % out of the whole database will produce a more precise forecasting.

The new forecasting model, which suggests the predicted values based on previous projects, is more than just a forecasting methodology; it provides a bridge that enables current data collection techniques to be used within the context of the accumulated information. This will eventually help all the participants in the construction industry to build up the knowledge derived from invaluable experience.

Keywords : expenditure forecasting, cost data, CBR (Case-based reasoning), residential building, work package
