

사례기반추론을 이용한 개략비용 예측시스템 개발 - 토양정화법 정화비용사례를 중심으로 적용가능성 검토 -

김상태* · 심진아¹ · 김흥래¹

¹(주)평화엔지니어링 기술연구원

Development of Approximate Cost Estimation System Based on CBR Technique; Applicability Study for Landfarming Soil Remediation Technology

Kim, Sang-Tae*, Shim, Jin-Ah¹, Kim, Heung-Rae¹

¹R&D Institute, Pyunghwa Engineering Consultants

Abstract : This study proposes a approximate cost estimation system based on Excel with VBA using weighted CBR(Case Based Reasoning). One characteristic of this system is that it generates the sheet automatically as many as the number of similar case and new estimation when it performs a case learning and a new estimate and cell formula is automatically entered into each sheet. User can be free to compose a combination of attribute factors because they can select up to ten attribute factors. This paper presents an applicability of estimation model for estimating the soil remediation cost when it use a landfarming method. When compared to a estimation model by using average unit cost and optimum multiple regression, this model shows a better result. This study was aimed at landfarming method, but it is expected that a cost estimation model using CBR will be more likely to apply in soil remediation technologies which various remediation technologies and pollutant species exist.

Keywords : Case based reasoning, Soil remediation cost, Cost estimation system, Mutiple regression model

1. 서론

1.1 연구의 배경 및 목적

2005년 토양정화법의 등록제 시행 이후, 토양오염지역에 대한 정화사업이 지속적으로 진행되고 있고 오염토양 정화기술이 다양하게 개발되고 있지만 정화공법은 현장 또는 제안에 따라 일관성이 결여되어 예산수립 등에도 많은 어려움이 따른다고 한다(환경부 2010).

이처럼 사업 계획단계에서의 적절한 사업비용 추정은 초기 사업방향의 의사결정에 미치는 영향이 상당히 크며 효과적인 사업비 관리 및 예산수립에도 중요한 사안으로 평가되고 있다. 특히, 도로나 교량 등 SOC시설물이나 주택의 개략공사비 예측에 대한 연구는 다수의 연구가 이루어져 왔으며, 특히 사업계획단계의 가용정보가 제한된 조건하에서의 예측모델의

연구는 사례기반추론 (Case-Based Reasoning)을 이용한 예측모델 연구가 주로 수행되어 왔다.

교량 및 도로와 관련한 CBR기반 기획단계 개략공사비 예측모델 연구로, 강찬성(2008)의 PSC빔 교량 개략공사비 예측모델,곽수남(2009)의 도로 개략공사비 예측모델, 정민선(2011)의 강박스거더교 개략공사비 예측모델, 김민지(2013)의 교량 초기공사비 예측모델, 전진영(2013)의 수로교 개보수 개략공사비 예측모델 등이 있으며, 주택공사비와 관련해서는 안성훈(2005)의 전문가지식을 영향인자 속성에 반영한 공동주택 초기공사비 예측모델, 김광희(2006)의 신경망과 유전자 알고리즘을 이용한 공동주택 초기공사비 예측모델 현창택(2009)의 메가프로젝트사업의 계획단계 건축공사비 예측모델 등이 있으며, 강병주(2010)의 공공청사를 대상으로한 계획설계단계의 물량기반 철근콘크리트 철조공사비 산정시스템, 서정규(2010)의 공공청사를 대상으로 한 기획단계 공사비계획 및 의사결정지원모델, 이현수(2012)의 공공아파트 기획단계 적정공사비 산정시스템 개발 등 상당수의 연구가 수행되었다.

본 연구에서는 건설사업이나 환경사업 등 다양한 사업의

* Corresponding author: Kim, Sang-Tae, R&D Institute, Pyunghwa Engineering Consultants, Anyang, 431-755, Korea
E-mail: stkim99@pec.kr
Received April 30, 2015; revised October 19, 2015
accepted December 19, 2015

계획단계 뿐만 아니라 개략비용 산정이 필요한 경우에, 한정된 정보를 활용하여 간편하면서도 정확한 개략비용 추정이 가능한 사례기반추론기반 비용예측시스템을 개발하고, 이를 토양경작법을 중심으로 한 토양정화사례에 적용시켜 토양정화비용 예측모델의 적용가능성을 검토하였다.

1.2 연구의 범위 및 방법

본 연구에서는 사업 계획의 초기단계에 한정된 정보로 사업비를 개략으로 추정할 수 있는 사례기반추론을 이용한 Excel VBA기반의 비용예측 시스템을 개발하고, 이를 토양정화비용 예측에 적용하였다. 신규추정 결과를 평균단가모델, 다중회귀모델과 비교하여 예측모델의 신뢰성을 검증하였다.

(1) 기존 문헌을 통해 사례기반추론 비용예측 알고리즘을 이용한 연구의 현황과 한계점을 파악한다.

(2) 특정분야에 국한된 개략비용 예측시스템이 아닌 활용성 향상을 고려한 Excel VBA기반의 셀수식입력방식의 시트 자동생성 기능을 구현한 개략비용 예측시스템을 개발한다.

(3) 사례기반추론 비용예측 프로그램의 적용사례로 토양정화사업 계획단계의 개략정화비용 추정에 적용하여 예측모델의 적용가능성을 검토한다.

(4) 사례기반추론 비용예측 프로그램의 신규추정 결과를 검증하기 위하여 평균단가모델, 다중회귀모델, 사례기반추론 모델의 비교분석한다.

2. 사례기반추론 비용예측모델 개발

2.1 가중사례기반추론 비용예측 알고리즘

2.1.1 가중사례기반추론 알고리즘 개요

사례기반추론을 활용한 초기단계 비용예측모델은 공사비를 예측하고자하는 사업의 영향인자를 바탕으로 유사한 특성을 지닌 과거 사례의 공사비 정보를 활용하여 신규사업의 공사비를 파악한다. 이러한 신규사업의 공사비를 추정하기 위해서는 신규사업과 유사한 사례를 검색하기 위하여 유사도 점수의 개념을 활용하며, 유사도가 가장 높은 몇 개 사례의 공사비정보를 가중평균하여 신규사업의 공사비를 산정한다. 이는 한 개의 사례를 사용하거나 여러개의 유사사례를 산술 평균하는 방식에 비해 예측의 정확도를 높이고 오차율 줄일 수 있다는 장점이 있다(곽수남 2009).

2.1.2 비용예측 알고리즘 절차

가중 사례기반추론의 절차를 요약하면 다음과 같다.

(1단계) 영향인자 유사도점수 산정

공사비를 산정하고자 하는 신규 사업과 가장 유사한 사례를 파악하기 위하여 신규사업의 영향인자와 과거 기존사례들의 영향인자를 비교하여 유사도점수를 계산한다. 유사도점수

는 문자열의 경우 영향인자별로 신규사례의 영향인자와 비교하여 일치하는 경우 100점을 부여하며, 일치하지 않는 경우에는 0점을 부여한다. 한편 수치일 경우에는 일치범위 수준에 따라 차등점수를 부여하는 방식(김민지 2013)으로 신규사업과 과거 사례의 차이가 10%이하인 경우 100점을 부여하며, 10~20%이하인 경우는 80점, 20~30% 이하인 경우는 60점, 30~40% 이하인 경우는 40점, 40% 이상인 경우는 0점을 부여한다(Table 1). 오차율이 10% 범위안에 들어올 경우에만 점수를 부여하는 방식을 일반적으로 활용하지만(곽수남 2009, 김광희 2004), 본 연구에서는 일반적인 방식을 적용할 경우 영향인자 가중치를 학습사례의 평균오차율이 최소가 되는 최적해로 산정할 경우 특정 영향인자를 반영하지 못하는 문제점이 발생되었다. 이는 기존사례DB의 신뢰성에서 기인한 것으로 판단하고 이를 개선하기 위해 일치범위 수준에 따라 차등점수를 부여하였다.

Table 1. Calculation of attributes similarity score

Type	Calculation method
String	SSC = 100 ; Identical = 0 ; Not identical
Numeric	SSC = 100 ; E < 10% = 80 ; 10% ≤ E < 20% = 60 ; 20% ≤ E < 30% = 40 ; 30% ≤ E < 40% = 0 ; 40% ≤ E $E = \left \frac{V_{Test\ case} - V_{Retrieved\ case}}{V_{Retrieved\ case}} \right \times 100(\%)$

(2단계) 기존사례 유사도점수 산정

각 사례의 영향인자의 유사도 점수에 각 영향인자별 가중치를 곱하여 기존사례의 유사도점수를 산정한다.

$$SS_k = \sum_{i=1}^A w_i \times SSC_i \tag{1}$$

w_i = i번째 영향인자 가중치, $i=1\sim A$ (영향인자 수)

SSC_i = i번째 영향인자 유사도점수, $i=1\sim A$ (영향인자 수)

SS_k = k번째 기존사례 유사도점수, $k=1\sim n$ (기존사례 수)

(3단계) 기존사례 가중치 산정

일정 순위 안에 진입하는 기존사례를 대상으로 유사도점수에 기초하여 각 기존사례의 가중치를 도출한다.

$$W_k = \frac{SS_k}{\sum_{i=1}^n SS_i} \tag{2}$$

W_k = k번째 기존사례 가중치, $k=1\sim n$ (기존사례 수)

(4단계) 예측사례 공시비 산정

예측사례(학습사례 또는 신규사례)의 공사비는 기존사례의

가중치와 공사비를 이용하여 산출한다.

$$PC_i = \sum_{k=1}^n W_k \times TC_k \quad (3)$$

TC_k = k번째 기준사례 실제값, $k=1\sim n$ (기준사례 건수)
 PC_i = i번째 예측사례 예측값, $i=1\sim N$ (예측사례 건수)

한편, 가중사례기반추론 모델에서 적용된 영향인자의 가중치를 산정하기 위하여 본 연구에서는 학습사례 예측공사비의 평균오차율이 최소가 되는 가중치를 구하기 위해 다음의 목적함수와 제약조건을 설정하여 최적해를 구하도록 하였다.

$$\text{목적함수} : \text{Min.} \sum_{i=1}^N \frac{ER_i}{N} = \sum_{i=1}^N \frac{1}{N} \left(\frac{|TC_i - PC_i|}{TC_i} \right) \quad (4)$$

$$\text{제약조건} : w_i \geq 0, \sum_{i=1}^A w_i = 1$$

ER_i = i번째 예측사례의 오차율
 PC_i = i번째 예측사례의 예측값
 TC_i = i번째 기준사례의 실 비용
 N = 예측사례 건수

본 연구에서는 상기의 최적해를 도출하기 위해 Excel 2010 해찾기 기능을 활용하였다. 개발한 프로그램에서는 학습사례의 평균오차율을 설정한 '사례학습'시트 G4셀의 최소화를 목적함수로, D8~D17셀을 변수셀로 설정하고 식(4)의 제약조건을 설정하여 Evolutionary 엔진을 이용하여 영향인자의 가중치를 산정하였다.

2.2 CBR기반 비용예측 시스템 개발

2.2.1 시스템 구성

가중사례기반추론 알고리즘을 이용하여 개발한 비용예측 시스템은 Excel기반의 VBA 프로그램으로 3개의 메인시트('사례DB'시트, '사례학습'시트, '신규추정'시트)로 구성되어 있다(Fig. 1).

본 시스템의 특징은 사례학습과 신규추정시 학습사례건수와 신규추정건수 만큼의 시트를 자동생성하며, 각각의 시트는 셀수식이 자동으로 입력되어 사용자의 편의성을 제고하는데 역점을 두었다. 또한 영향인자를 최대 10개까지 선택이 가능하도록 하여 영향인자의 조합을 자유롭게 구성할 수 있도록 하였다.

입력단계에서는, '사례DB'시트에 사례데이터를 입력하고, '사례학습'시트에 학습사례를 입력하고, 마지막으로 '신규추정'시트에 신규로 추정하고자 하는 데이터를 입력한다. 입력이 완료된 후에는 영향인자를 선택하고 편차기준 및 순위기준의 유사사례 기준을 설정하는 것으로 입력단계가 완료된다.

'사례학습시트작성' 기능을 통하여, 학습사례건수만큼의 시트를 생성하게 되어 영향인자 가중치 산정을 준비단계를 완료하게 된다. 다음으로 Excel 해찾기 기능을 이용하여 영향인자 가중치를 찾게 되고 최종 도출된 가중치를 적용한 학습사례의 비용추정결과가 '사례학습'시트에 집계된다.

신규추정단계에서는 사례학습단계에서 산정된 가중치를 적용한 신규추정이 가능하게 되고, '신규사례추정' 기능을 통

Fig. 1. Cost estimation system based on CBR

하여 신규사례를 추정하게 된다.

2.2.2 시스템 기능제한

현재까지는 사용자가 학습사례를 선택하여 '사례학습'시트에 입력하는 방식을 채택하고 있으며, 사례학습의 선택에 따라 영향인자의 가중치는 영향을 받게 되므로 시행착오를 통한 최적의 선택이 필요하게 된다.

3. 토양정화비용 예측모델 적용

3.1 사례데이터 수집

토양정화비용 예측모델의 적용가능성을 검토하기 위하여 2006년부터 2008년까지 3년간 토양정화 실적을 수집한 기존 사례데이터를 활용하였다(환경부 2010). 총 1,046건의 토양정화사업 사례 중 토양정화방법은 단독 또는 복합공법으로 가장 많이 사용되고 있으며 약 32% 차지한다. 그 다음으로는 토양증기추출법, 생물학적 통풍법, 화학적산화법 순으로 많이 적용되었다.

토양정화방법 적용사례 가운데 TPH를 정화 대상으로 한 133건 중, 결손 데이터 22건 및 이상치 8건을 제외한 103건을 사례대상으로 하였다. 토양정화사례별로 오염물질, 정화 전 농도, 정화 후 농도, 정화방법, 정화위치(부지 내, 부지 외), 오염규모, 정화기간, 총정화비용 등 총 8개의 정보가 활용 가능하였으나(Table 2), 정화전후의 농도차인 제거농도와 오염규모와 제거농도의 곱으로 산정되는 제거오염량을 조합변수(예를 들어, 전건영 2013)로 추가하여 사례DB를 재구성하여 활용하였다.

Table 2. Case database sample of soil remediations

A	B	C	D	E	F	G	H	I
TPH	8,425	0	LF	On	3	2,100	755.4	2
TPH	11,824	1,592	LF	Off	105	34,000	322.9	13
TPH	13,107	134	LF	On	600	49,500	82.5	2
TPH	4,083	267	LF	On	13	8,000	615.4	4
TPH	14,978	130	LF	On	4,410	268,557	60.9	19
TPH	12,397	0	LF	Off	159	38,000	239	5
TPH	19,563	1,410	LF	Off	354	80,360	227	8
TPH	8,611	0	LF	Off	259	59,840	231	6
TPH	2,206	136	LF	On	461	80,800	175.1	4
TPH	3,056	229	LF	On	23	14,000	619.6	2
TPH	3,889	613	LF	On	53	20,000	377	5
TPH	10,302	65	LF	On	128	60,000	469.6	2
TPH	3,784	756	LF	On	12,600	660,000	52.4	2
TPH	9,591	0	LF	Off	195	78,600	403.1	4
TPH	5,785	754	LF	Off	86	14,900	173.3	3
...

A: Contaminants [-], B: Concentration before remediation [mg/kg], C: Concentration after remediation [mg/kg], D: Remediation technology [-], E: On/off site [-], F: Remediation volume [m³], G: Remediation cost [10³ won], H: Remediation cost per volume [10³ won/m³], I: Remediation period [Month]

이상치는 수집데이터의 범위에서 상당히 벗어난 아주 작은 값이나 아주 큰 값을 의미하며, 단위 총정화비용(=총정화비용/오염규모)이 “제1사분위수-1.5×사분위범위” 보다 작거나 “제3분위수+1.5×사분위범위” 보다 큰 경우를 이상치로 판단하였다.

3.2 영향인자 선정 및 가중치 산정

3.2.1 영향인자 선정

영향인자를 선정하기 위하여 정화 전 농도, 정화 후 농도, 정화위치(부지 내, 부지 외), 오염규모, 정화기간, 제거농도, 제거오염량, 총정화비용에 대한 상관분석을 실시하였다 (Table 3).

Table 3. Results of correlation analysis for attributes

Correlation Coeff. (p-value)	A	B	C	D	E	F	G	H
A	1.000							
B	.370 (.000)	1.000						
C	.021 (.830)	-.060 (.545)	1.000					
D	.858 (.000)	.259 (.008)	.092 (.358)	1.000				
E	.246 (.012)	.134 (.176)	.086 (.388)	.125 (.209)	1.000			
F	-.027 (.785)	.051 (.607)	.063 (.530)	-.156 (.115)	.076 (.447)	1.000		
G	.368 (.000)	.999 (.000)	-.096 (.334)	.255 (.009)	.131 (.188)	.049 (.623)	1.000	
H	.730 (.000)	.633 (.000)	-.001 (.990)	.583 (.000)	.186 (.061)	-.043 (.669)	.631 (.000)	1.000

A: Remediation cost [10³ won], B: Concentration before remediation [mg/kg], C: Concentration after remediation [mg/kg], D: Remediation volume [m³], E: Remediation period [Month], F: On/off site [-], G: Treatment concentration [mg/kg], H: Treatment mass [g]

상관분석 결과, 총정화비용과 정화전농도간의 상관계수는 0.37, 정화후농도간의 상관계수는 0.21, 오염규모간의 상관계수는 0.858, 정화기간간의 상관계수는 0.246, 정화위치간의 상관계수는 -0.027, 제거농도간의 상관계수는 0.368, 제거오염량간의 상관계수는 0.730으로 나타났으며, 정화후농도, 정화기간, 정화위치의 유의확률은 0.05이상으로 총정화비용과의 상관관계가 0이라는 귀무가설을 기각할 수 없어 유의미한 상관관계가 없는 것으로 분석되었다. 하지만 토양정화사업 기간단축에 정화비용 증가와 사업부지내외에 따른 부대시설비용과 운송비용을 감안하여 정화기간과 정화위치(부지 내, 부지 외)는 영향인자로 채택하기로 하였다.

한편 정화전농도는, 유의확률로는 대립가설이 기각되어 총정화비용과의 상관관계가 유의미하나 정화전농도 대신 조합변수인 제거농도(=정화전농도-정화후농도)를 채택하여 실

제 정화되는 오염물질 농도차를 반영하도록 하였다. 또한 상관분석 결과에서도 정화전농도와 제거농도의 상관계수는 0.999로 양의 상관관계를 확인할 수 있었다. 최종적으로는 오염규모, 정화기간, 정화위치, 제거농도, 제거오염량을 선정하여 총 5개의 영향인자를 채택하였다(Table 4).

Table 4. Attributes for landfarming soil remediation

Attributes	Unit	Type	Note
Remediation volume	m ³	Numeric	
Remediation period	Month	Numeric	
On/off site	-	String	
Treatment concentration	mg/kg	Numeric	Concentration difference after remediation
Treatment mass	g	Numeric	Treatment concentration × Remediation volume

3.2.2 영향인자 가중치 산정

앞서 설명한 바와 같이 학습사례 예측공사비의 평균오차율이 최소가 되는 영향인자의 가중치를 구하는 문제는 학습사례의 선정방식에 밀접한 영향을 받게 된다. 본 사례연구에서는 시행착오 방법을 통하여 학습사례 25건에 대한 예측공사비의 평균오차율이 최소가 되는 가중치를 산출하였다(Table 5).

Table 5. Weights of attributes

Remediation volume	Remediation period	On/off site	Treatment concentration	Treatment mass
0.4481	0.1651	0.0371	0.1163	0.2335

3.3 정화비용 예측모델 비교

3.3.1 평균단가 모델

오염토양 단위체적당 토양경작법의 정화비용 단가를 산출하여 예측모델 비교분석에 활용하였다. 대상부지내에서 처리하는 토양경작법의 평균단가는 247천원/m³, 부지외의 평균단가는 266천원/m³으로 집계되었다(Table 6).

Table 6. Unit cost of landfarming soil remediation

	On-site	Off-site	Total
Number of case	68	35	103
Unit cost [10 ³ won/m ³]	247.24	265.96	253.60

3.3.2 다중회귀 모델

토양정화비용의 영향인자를 설명변수로 하는 다중회귀모델에서 설명변수의 최적조합을 찾아내어 최적 다중회귀모델을 도출하였고, 평균단가모델과 더불어 예측모델 비교분석에 활용하였다.

Ueda(2004)는 다중회귀분석에서 설명변수의 최적조합을 선택하는 문제에 대해 시행착오를 거쳐 다음과 같이 설명변수선택기준(Ru)을 제시하였다.

$$Ru = 1 - (1 - R^2) \frac{n + k + 1}{n - k - 1} \quad (5)$$

여기서 R은 다중상관계수, n은 데이터 수(관측수), k는 설명변수 개수(회귀분석 자유도)를 나타내며, Ru이 최대가 되는 조합이 최적의 회귀모델식이라고 제안하였다. 설명변수의 최적조합을 찾는 알고리즘은 회귀분석 결과 얻어지는 설명변수의 유의확률(P값)을 설명변수 도입에 따라 오차가 커질 확률로 이해하고 P값이 큰 설명변수부터 제외시켜 회귀분석을 실행하여, 최종적으로는 설명변수 조합별로 Ru를 계산하여 최대가 되는 설명변수 조합을 찾게 된다.

토양정화비용의 영향인자 5개를 대상으로 최적 다중회귀분석을 한 결과, 오염규모, 정화기간, 정화위치, 제거오염량 4개의 영향인자 조합일때가 가장 최적의 다중회귀모델로 분석되었다. 한편, 다중회귀분석에서 독립변수간의 다중공선성을 검토하기 위하여 각 단계별로 각각의 영향인자에 대해 분산팽창인자(Variation Index Factor, VIF)를 계산하였다. 그 결과 5변수 회귀모델의 제거오염량만 VIF가 2.73이고 나머지는 모두 2미만으로 분석되었고 다중공선성이 없는 것으로 판단하였다(Table 7).

최종 다중회귀모델식은 다음과 같다.

Table 7. Process of optimal multiple regression

Step	A	B	C	D	E	Number of Var.	R	Ru	Result
1	● (1.61)	● (1.05)	● (1.05)	● (1.95)	● (2.72)	5	0.9243	0.8343	
2	● (1.56)	● (1.05)	● (1.05)		● (1.57)	4	0.9242	0.8375	Optimal
3	● (1.52)	● (1.03)			● (1.56)	3	0.9218	0.8362	
4	● (1.52)				● (1.52)	2	0.9138	0.8240	
5	●					1	0.8609	0.7298	

A : Remediation volume [m3], B: Remediation period [Month], C: On/off site [-],
D : Treatment concentration [mg/kg], E: Treatment mass [g]
(.): VIF

$$Y_{\text{정화비용}} = -3,575.8 + 54.1 \times X_{\text{오염규모}} + 4,506.1 \times X_{\text{정화기간}} + 20,184.9 \times X_{\text{정화위치}} + 0.8 \times X_{\text{제거오염량}}$$

각 설명변수(영향인자)의 회귀계수와 t통계량을 정리하였다(Table 8). 오염규모가 토양정화비용에 가장 큰 영향을 미치며 그 다음 순으로는 제거오염량, 정화기간, 정화위치 순이다.

Table 8. Result of optimal multiple regression

Variable	Unit	Coefficient	t statistic
Y-intercept		-3,575.8	-0.2784
Remediation volume	m ³	54.1	13.0354
Remediation period	Month	4,506.1	2.7442
On/off site	-	20,184.9	1.6110
Treatment mass	g	0.8	6.8235

3.3.3 정확비용 신규추정 비교검증

앞서 설명한 평균단가모델, 다중회귀모델, 사례기반추론모델을 이용한 토양정화비용 예측모델의 검증을 위하여 토양정화작업 사례DB 103중 10건을 무작위 추출하여 이를 대상으로 정화비용 예측결과를 비교분석하였다.

사례기반추론을 이용하여 추정한 신규사례 예측모델의 결과와 더불어 평균단가모델, 다중회귀모델의 결과를 제시하였다(Table 9). 신규사례추정 10건에 대한 평균오차율은 평균단가모델이 94%, 다중회귀모델이 54%, 사례기반추론모델이 49%로, 사례기반추론모델이 평균단가모델보다 45%p, 다중회귀모델보다 5%p의 평균오차율 개선을 나타내었다. 또한 최대오차율과 표준편차에 있어서도 사례기반추론모델이 단위단가모델 뿐만 아니라 평균단가모델보다 다소 개선된 결과를 나타내었다.

3. 결론

본 연구에서는 개략비용 예측시스템의 범용성을 염두에 두고 셀수식 입력방식의 사례시트 자동생성 기능을 구현한 Excel VBA기반 개략비용 예측시스템을 개발하였다. 개발된 개략비용 예측시스템은 사례학습과 신규추정시에 셀수식이 입력된 시트를 각각의 사례건수만큼 자동생성하여 시스템의 활용성을 고려하였으며, 영향인자의 선택을 최대 10개까지

가능하도록 하여 사용자의 편의성을 향상시켰다.

또한, 개발시스템의 적용성을 검토하기 위하여 토양정화작업을 대상으로 토양정화기술의 사례기반추론 비용예측에 활용하였다. 수집된 사례데이터의 상관분석을 통하여 영향인자를 선정하고 이를 개략비용 예측시스템에 적용한 결과, 신규사례추정 10건에 대해 사례기반추론모델이 평균단가모델보다 45%p, 다중회귀모델보다 5%p의 평균오차율 개선효과를 확인할 수 있었으며, 또한 최대오차율과 표준편차에서도 개선된 결과를 확인할 수 있었다.

감사의 글

본 연구는 한국환경산업기술원 토양·지하수오염방지기술 개발사업(GAIA project) (RE201402087)에 의해 일부 수행되었습니다. 이에 감사드립니다.

References

- An, S. H., and Kang, K. 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.
- Hyun, C. T., Hong, T. H., Son, M. J., Kim, Y. S., and Jang, D. W. (2009). "Development of the Construction Cost Prediction Model Based on Case-Based Reasoning in the Planning Phase of Mega-Project", *Journal of the Architectural Institute of Korea Structure & Construction*, 25(9), pp. 181-190.
- Jeon, G. Y., Cho, J. Y., and Huh, Y. (2013). "Development of Approximate Cost Estimate Model

Table 9. Error rate of cost estimation models for verification

No	A	B	C	D	E	F	Error rate of estimation models(%)		
							Unit cost	Optimal regression	CBR
1	4,410	19	On	14,848	65,480	268,557	337	39	24
2	86	3	Off	5,031	433	14,900	43	136	48
3	258	4	On	3,326	858	73,000	6	60	46
4	400	3	Off	65,578	26,231	46,200	114	58	67
5	190	3	On	5,420	1,030	47,580	6	56	1
6	503	4	Off	4,757	2,391	38,000	227	68	110
7	291	4	Off	23,754	6,912	70,000	3	20	24
8	2,250	5	Off	1,900	4,275	374,186	49	56	38
9	83	2	On	6,674	554	11,000	101	6	91
10	1,156	7	Off	5,778	6,679	185,702	54	37	39
Average error rate							94%	54%	49%
Standard deviation							109%	35%	33%
Maximum error rate							337%	136%	110%

A: Remediation volume [m³], B: Remediation period [Month], C: On/off site [-], D: Treatment concentration [mg/kg], E: Treatment mass [kg], F: Remediation cost [103 won]

- for Aqueduct Bridges Restoration –Focusing on Comparison between Regression Analysis and Case-Based Reasoning–", *Journal of the Korean Society of Civil Engineers*, 33(4), pp. 1693–1705.
- Jung, M. S., Kyung, K. S., Jeon, E. K., and Kwon, S. C. (2011). "Computation and Verification of Approximate Construction cost of Steel Box Girder Bridge by using Case-Based Reasoning", *International Journal of Steel Structure*, 23(5), pp. 557–568.
- Kang, B. J., Cho, J. H., Lee, J. S., Lee, J. S., and Chun, J. Y. (2010). "Development of Quantity-based Reinforced Concrete Structural Construction Cost Estimating System using CBR+GA Model – Focused on Government Office Building", *Journal of the Architectural Institute of Korea Structure & Construction*, 26(3), pp. 139–146.
- Kang, C.S., Lee, G.H., Kim, K.M., and Kim, K.J. (2008). "Approximate Estimating Model Using the Case Based Reasoning – PSC BEAM Bridge", Proceedings of KICEM Anunual Conference, pp. 445–448
- 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, 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, M. J., Moon, H. S., and Kang, L. S. (2013). "Development of an Approximate Cost Estimating Model for Bridge Construction Project using CBR Method", *Korean Journal of Construction Engineering and Management*, KICEM, 14(3), pp. 42–52.
- Kwak, S. N., Kim, D. Y. , Kim, B. I. , Choi, S. J., and Han, S. H. (2009). "Cost Prediction Models in the Early Stage of the Roadway Planning and Designbased on Limited Available Information", *Korean Journal of Construction Engineering and Management*, KICEM, 10(4), pp. 87–99.
- Lee, H. S., Lee, H. K., Park, M. S., Kim, S. Y., and Ahn, J. (2012). "Conceptual Cost Estimating System Development for Public Apartment Projects", *Korean Journal of Construction Engineering and Management*, KICEM, 13(4), pp. 152–163.
- Ministry of Environment (2010). Plans of Standardization and Competitiveness for the activation of Soil remediation industry.
- Park, Moon-Soo, Seong, Ki-Hoon, Lee, Hyun-Soo, Ji, Sae-Hyun, and Kim, Soo-Young (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. 22–30.
- Park, Y., Shin, J., Lee, B., Hwang, H., and Lee, S. (2013). *Statistical Data Analysis on Excel*, Kyungmoon Publishers, pp. 295–329.
- Seo, J. K., Kim, S. M., Cho, J. H., and Chun, J. Y. (2010). "Supporting Model of Cost Planning using Case-Based Reasoning in Pre-design Phase – Focused on the Case of Public Office Building", *Journal of the Architectural Institute of Korea Structure & Construction*, 26(4), pp. 159–168.
- Ueda, T., Kobayashi, M., and Huchigami, M. (2004). *Introduction of Regression Analysis on Excel*, Ohmsha, pp. 102–132.

요약 : 가상사례기반추론 알고리즘을 이용하여 Excel VBA 기반의 개략비용 예측시스템을 개발하였다. 개발 시스템의 특징은 사례 학습과 신규추정시 학습사례건수와 신규추정건수 만큼의 시트를 자동생성하며, 각각의 시트는 셀수식이 자동으로 입력되어 시스템의 활용성을 고려하였으며, 또한 영향인자를 최대 10개까지 선택이 가능하도록 하여 영향인자의 조합을 자유롭게 구성할 수 있도록 하였다. 개발한 비용예측시스템을 토양경작법 정화비용예측에 이용하여 토양정화 예측모델의 적용가능성을 검토하였다. 평균단가 예측모델은 물론 최적다중회귀 예측모델보다 개선된 결과를 확인하였다. 본 연구에서는 토양경작법을 대상으로 검토 하였으나 토양정화기술과 오염물질 종류 등 다양한 시나리오가 나타나는 토양정화사업분야에 사례기반추론을 이용한 비용예측모델은 향후 사례데이터 축적과 더불어 적용가능성이 커질 것으로 기대된다.

키워드 : 사례기반추론, 토양정화비용, 비용예측시스템, 다중회귀모델
