

# 서포트 벡터 머신을 이용한 건설업 안전보건관리비 예측 모델

신성우<sup>†</sup>

부경대학교 안전공학과  
(2017. 2. 3. 접수 / 2017. 2. 10. 채택)

## Construction Safety and Health Management Cost Prediction Model using Support Vector Machine

Sung Woo Shin<sup>†</sup>

Department of Safety Engineering, Pukyong National University  
(Received February 3, 2017 / Accepted February 10, 2017)

**Abstract :** The aim of this study is to develop construction safety and health management cost prediction model using support vector machine (SVM). To this end, theoretical concept of SVM is investigated to formulate the cost prediction model. Input and output variables have been selected by analyzing the balancing accounts for the completed construction project. In order to train and validate the proposed prediction model, 150 data sets have been gathered from field. Effects of SVM parameters on prediction accuracy are analyzed and from which the optimal parameter values have been determined. The prediction performance tests are conducted to confirm the applicability of the proposed model. Based on the results, it is concluded that the proposed SVM model can effectively be used to predict the construction safety and health management cost.

**Key Words :** construction safety and health management, safety management cost estimation, support vector machine

### 1. 서론

건설업은 동일한 공종이라도 공법이나 현장의 여건, 인력의 수준 등에 따른 작업 위험도가 다르고, 또한 필요한 안전시설물의 종류 및 수량도 매우 큰 차이가 나기 때문에, 안전관리에 필요한 비용도 건설프로젝트의 제반 조건에 따라 차이가 날 수 밖에 없는 것이 현실이다. 그러나, 이러한 현실에도 불구하고 건설공사의 안전관리에 필요한 비용은 “산업안전보건관리비 계상 기준”에 따라 공사의 종류와 대상액(재료비와 직접노무비의 합)만을 기준으로 일률적으로 계상하고 있으며, 이러한 일률적 계상의 불합리성이 자주 지적되고 있는 실정이다<sup>1-5)</sup>. 일례로 건설현장의 안전보건관리비 사용 실태에 관한 여러 연구에서, 안전보건관리비가 조기에 소진되어 부족하거나 또는 과다하게 책정되어 미집행 잔액이 발생하는 경우가 매우 빈번한 것으로 보고하고 있으며, 이러한 문제점을 해결하기 위해서는 보다 실제적인 안전보건관리비의 책정 방법이 필요하다고 제

시하고 있다.<sup>1-3)</sup> 건설공사에 실제로 필요한 안전보건관리비를 책정하기 위해서는, 해당 건설공사에 필요한 안전보건관리비를 정확하게 예측할 수 있는 건설공사 안전보건관리비 예측 모델이 필요하다. 그러나, 아직까지 건설공사 안전보건관리비를 예측할 수 있는 모델에 대한 연구는 국내/외에서 거의 이루어지지 않은 실정이어서 이에 대한 연구가 필요하다.

한편, 건설프로젝트를 실행하고자 하는 발주자에게 프로젝트에 소요되는 예상 비용과 그 추정 비용의 정확도는 프로젝트의 수행 여부를 결정하는 중요한 요소이다. 뿐만 아니라 프로젝트 초기에 견적된 비용은 실제 공사에 필요한 예산 수립과 조정의 기반이 된다<sup>6)</sup>. 이에 따라 건설 프로젝트의 초기 단계에서 제한된 정보를 이용하여 공사비를 정확하게 추정 예측할 수 있는 방법에 대한 많은 연구가 이루어져 왔으며, 최근에는 딥러닝 등 머신러닝(Machine Learning)기술의 발전에 따라 공사비 예측 모델 개발에 이를 활용하는 연구가 매우 활발하게 이루어지고 있다<sup>7-10)</sup>. 머신러닝은 인

<sup>†</sup> Corresponding Author : Sung Woo Shin, Tel : +82-51-629-6473, E-mail : shinsw@pknu.ac.kr  
Department of Safety Engineering, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan 48513, Korea

과관계가 불분명하여 이론적으로 예측 모델을 구현하기 어렵거나, 모델의 구조가 매우 복잡하여 단순 회귀 분석으로는 정확한 모델링이 어려운 경우에 통계학적 데이터나 관측 자료 등을 이용하여 예측 모델을 구현할 수 있는 방법이며, 대표적인 머신러닝 기법으로는 인공신경망(Artificial Neural Network) 기법, 서포트 벡터 머신(Support Vector Machine; 이하 SVM) 기법 등이 있다<sup>11)</sup>. Kim 등은 공동주택 공사비 예측에 인공신경망 기법을 이용한 예측 모델과 회귀분석기법을 이용한 예측모델의 예측 성능을 비교하는 연구를 수행하였으며, 인공신경망 모델이 회귀분석 모델보다 더 정확하게 공사비를 예측할 수 있음을 보였다.<sup>7)</sup> 그러나, Kim 등은 인공신경망 기법을 이용할 경우 공사비를 정확하게 예측할 수 있는 모델을 구현할 수는 있으나, 구현 과정에서 입력 패턴의 분포를 추정하기 위해 다량의 학습 데이터가 필요하고, 특히 입력데이터에 이상 데이터(outlier) 등이 포함되어 있을 경우 이로 인해 과적합된(over-fitted) 모델을 생성할 수 있어 결과를 일반화하기 까다로운 문제점이 있음을 지적하였다<sup>8)</sup>. 한편, An 등은 공사비 예측 모델 개발에 있어서 인공신경망 기법의 이러한 단점을 극복할 수 있는 방법으로, 일반화 능력이 뛰어난 SVM 기법을 이용한 공사비 예측 모델을 개발하였으며, 인공신경망에 비해 소수의 학습데이터 만으로도 충분히 정확한 예측 모델을 생성할 수 있음을 보였다<sup>9)</sup>.

본 연구에서는 건설공사에 실제로 필요한 안전보전관리비의 책정을 위해, 건설공사의 안전보전관리비를 예측할 수 있는 모델을 SVM 기법을 이용하여 개발하고, 그 적용 가능성을 검토하고자 한다. 이를 위해 본 연구에서는 (1) SVM 기법의 이론에 대해 고찰하고, (2) SVM 기법을 이용한 안전보전관리비 예측 모델을 개발하며, (3) 마지막으로 실제 사례를 이용하여 개발된 모델의 예측 성능을 평가하고, 이로부터 개발된 모델의 실적용 가능성을 검토하고자 한다.

## 2. SVM 기법

SVM은 Vapnik이 제안한 머신러닝 기법으로, 경험적 위험 최소화 원칙을 기반으로 하는 다른 통상적인 머신러닝 기법과는 달리 구조적 위험 최소화를 기반으로 하여 일반화 오류의 상한을 최소화할 수 있는 머신러닝 기법이다<sup>12,13)</sup>. 특히, SVM은 인공신경망 기법의 문제점으로 지적되는 과적합 문제를 벌칙(Penalty)항을 이용하여 피할 수 있으며, 또한 함수 근사에 있어서 이상데이터(outlier)에 둔감하기 때문에 높은 일반화 성능

을 가진다. 따라서, 만약 동일한 데이터를 활용할 경우, 데이터의 특성에 따라 인공신경망 기법에 비해 상대적으로 예측력이 우수한 모델의 구현이 가능한 장점이 있다<sup>13)</sup>. 한편, SVM 기법은 자료의 분류(Classification)에 이용하는 기법과 회귀(Regression)에 이용하는 기법으로 나눌 수 있으며, 본 장에서는 안전보전관리비 예측 모델에 맞는, 회귀에 대한 SVM 기법의 이론적인 개념을 간단하게 소개하고자 한다. SVM 기법의 보다 상세한 내용은 참고문헌에서 찾을 수 있다<sup>12)</sup>.

Vapnik에 의해 제안된 가장 단순한 형태의 SVM 회귀 문제는,  $m$ 개의 학습용(training) 데이터가 벡터  $x_i$ 와  $y_i$ 의 쌍 $\{(x_i, y_i), \dots, (x_m, y_m) | x_i \in R^n\}$ 으로 주어질 경우, 주어진 입력벡터값  $x_i$ 에 대한 실제의 목표값  $y_i$ 와 최대  $\delta$ 만큼의 편차 내에 가능한 좁은 범위의 가중치벡터값  $w$ 를 갖는 다음과 같은 함수  $f(x)$ 를 찾는 문제이다<sup>12)</sup>.

$$f(x) = w^T x_i + b \tag{1}$$

여기서 상첨자  $T$ 는 행렬의 전치(transpose)이며,  $b$ 는 바이어스(bias)이다. 이러한 함수  $f(x)$ 를 찾는 문제는,  $w$ 의 유클리드 놈(Euclidean norm)인  $\|w\|^2$ 을 최소화하는 볼록 최적화(convex optimization) 문제로 간주할 수 있다.

$$\begin{aligned} &\text{minimize } \frac{1}{2} \|w\|^2 \\ &\text{subject to } \begin{cases} y_i - w^T x_i - b \leq \delta \\ w^T x_i + b - y_i \leq \delta \end{cases} \end{aligned} \tag{2}$$

그러나, 식(2)와 같은 형태의 볼록 최적화 문제는 일반적으로 해를 찾기가 매우 까다롭다. 이러한 문제점을 해결하기 위해 SVM 기법에서는 새로운 slack 변수  $\xi_i$ 와  $\xi_i^*$ 를 도입하여, 식(2)를 다음과 같은 등가의 최적화 문제로 치환한다.

$$\begin{aligned} &\text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \\ &\text{subject to } \begin{cases} y_i - w^T x_i - b \leq \delta + \xi_i \\ w^T x_i + b - y_i \leq \delta + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{3}$$

여기서  $C$ 는 추정 오차에 대한 페널티로서 모델의 복잡도를 결정하는 모수(parameter)이다. 즉,  $C$ 가 크면 오차에 대해 더 큰 페널티를 할당하여 낮은 일반화 수준으로 오차를 최소화시킨다. 반면  $C$ 가 작으면 오차에

대해 작은 페널티를 할당하여 높은 일반화 수준을 갖게 된다. 따라서 적절한 C를 선택하는 것은 모델의 복잡도를 조정하는 효과뿐만 아니라 SVM 회귀의 일반화 성능도 높일 수 있다. 식 (3)에서 slack 변수  $\xi_i$ 와  $\xi_i^*$ 는 각각 출력에 대한 상한조건과 하한조건을 나타낸다. SVM 기법에서는 일반화 능력을 최대화하기 위해 목표값으로부터 일정한 거리 이내의 오차를 무시하는  $\delta$ -insensitive loss function을 slack 변수  $\xi$ 에 대해 다음과 같이 도입한다.

$$\xi = \begin{cases} 0 & \text{if } |\xi| \leq \delta \\ |\xi| - \delta & \text{otherwise} \end{cases} \quad (4)$$

SVM 기법에서는 Lagrange multiplier 방법을 이용하여, 식 (4)의 조건을 포함한 식(3)의 최적화 문제를 다음 식 (5)와 같이 Lagrange 함수로 변환한다.

$$\begin{aligned} \text{minimize } L = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \quad (5) \\ & - \sum_{i=1}^m \alpha_i (\delta + \xi_i - y_i + w^T x_i + b) \\ & - \sum_{i=1}^m \alpha_i^* (\delta + \xi_i^* + y_i - w^T x_i - b) \\ & - \sum_{i=1}^m (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned}$$

한편, 식(5)의 쌍대 최적화 문제는 다음과 같이 표현된다.<sup>12)</sup>

$$\begin{aligned} \text{maximize } & -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) x_i^T x_j \quad (6) \\ & - \delta \sum_{i=1}^m (\alpha_i + \alpha_i^*) + \sum_{i=1}^m y_i (\alpha_i - \alpha_i^*) \\ \text{subject to } & \begin{cases} \sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned}$$

여기서 제약 조건을 만족하는 방정식을 푸는 것은 식(6)을 최대화하는  $\alpha_i, \alpha_i^*$ 를 구하는 것이며, 이를 만족하는  $\alpha_i, \alpha_i^*$ 를 구하면 최초로 구하고자 했던 가중치 벡터  $w$ 를 다음 식을 이용하여 구할 수 있다.

$$w = \sum_{i=1}^m (\alpha_i - \alpha_i^*) x_i \quad (7)$$

한편, SVM 기법에서는 학습용 데이터의 비선형성으로 인해 선형 예측 모델을 구현하기 어려운 경우에는,

입력벡터를 고차원의 특성 공간으로 사상시켜(mapping) 선형 모델로 변환한 후 이를 다시 사상하여 최적 모델을 구하는 방법을 이용하며, 이때 입력벡터의 사상에 커널함수(kernel function)가 주로 이용된다. 즉, 비선형 SVM 기법은, 식(6)에 커널함수를 적용하여 다음 식(8)과 같이 변환하고, 앞서 선형의 경우와 동일한 방법으로 식(8)의 최적해를 구한다.

$$\begin{aligned} \text{maximize } & -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) \quad (8) \\ & - \delta \sum_{i=1}^m (\alpha_i + \alpha_i^*) + \sum_{i=1}^m y_i (\alpha_i - \alpha_i^*) \\ \text{subject to } & \begin{cases} \sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned}$$

일반적으로 많이 사용되는 커널함수  $K(x_i, x_j)$ 는 다항 함수, Radial Basis Function(RBF) 함수, 시그모이드 함수 등이 있다<sup>12)</sup>.

### 3. SVM을 이용한 안전보건관리비 예측 모델링

#### 3.1. 입력 및 출력 변수의 선정

SVM을 이용한 안전보건관리비 예측 모델을 개발하기 위해서는, (1) 입력변수와 이에 대응하는 출력변수를 결정하고, (2) 결정된 입력변수와 출력변수에 대한 학습용 데이터 쌍을 획득하고, (3) 마지막으로 예측 모델(즉,  $f(x)$ )을 구현하기 위해 학습용 데이터 쌍에 대응하는 가중치벡터( $w$ )를 구하여야 한다. 본 연구에서는 건설현장에 필요한 안전보건관리비를 보다 실제적으로 책정하기 위해, 완성된 공사에 대한 실제 공사비 원가 결산 자료를 바탕으로 예측 모델을 개발하고자 한다. 공사비 원가 결산 시의 회계 항목은 예산 수립시의 회계 항목과 대부분 동일하며, 크게 재료비, 노무비, 외주비, 현장경비의 4가지 대항목으로 분류된다. 또한 대항목 중 노무비와 현장경비는 각각 2개와 29개의 하위 항목으로 재분류되며, 안전보건관리비는 현장경비의 하위 항목에 속한다<sup>14)</sup>.

본 연구는 안전보건관리비의 예측이 목적이기 때문에 출력변수는 안전보건관리비로 선정하였다. 한편, 입력변수의 경우 실제 안전보건관리비 사용에 영향을 미치거나 또는 상관관계가 높은 비용 항목을 선정하는 것이 예측 모델의 정확성 측면에서 타당하다고 볼 수 있다. 그러나, 아직까지 안전보건관리비의 사용과 높은 연관성을 갖는 비용 항목에 대해서는 밝혀진 바가 없는 실정이다. 따라서, 본 연구에서는 실무 전문가의 의견을

Table 1. Model variables

Variables	Unit	Remarks
Safety and health management cost	Million won	Output (y)
Construction type	1: Civil 2: Architectural 3: Industrial facility 4: Landscape	Input (x)
Total construction cost (contracted)	Million won	
Material cost		
Labor cost (direct)		
Outsourcing cost		
Transportation cost		
Machine and equipment cost		
Quality control cost		
Temporary work cost		
Safety inspection cost		

조회하고, 대한건설협회와 국가통계포털에서 공개하는 완성공사원가통계의 항목을 분석하여, 안전보건관리비 사용과 연관성이 있다고 판단되는 10개의 항목을 입력 벡터를 구성하는 변수로 선정하였다 (Table 1 참조). 특히, 공종별 특성을 반영하기 위해, 공종(Construction Type)을 입력 변수로 설정하고 이에 대한 자료 입력시에는 Table 1에 제시된 바와 같이 공종별로 미리 부여한 숫자를 입력하도록 설정하였다.

3.2. 커널함수와 모수의 선정

일반적으로 입력 변수로 구성된 항목에 대한 데이터의 선형성을 미리 알 수 없기 때문에, 본 연구에서는 커널함수를 이용한 비선형 SVM 기법을 예측 모델 개발에 적용하였다. 비선형 SVM 기법의 커널함수는 여러 적용 가능한 커널함수 중에 일반화 성능이 가장 우수한 것으로 알려져 있는 RBF 함수로 이용하였다<sup>15)</sup>. RBF 함수는 다음 식과 같이 표현되며, 여기서  $\lambda$ 는 RBF 함수의 모수(parameter)이다.

$$K(x_i, x_j) = \exp(-\lambda |x_i - x_j|^2) \tag{9}$$

비선형 SVM 기법은 선형 SVM 기법과는 달리, 오차에 대한 페널티에 해당하는 상수 C의 값 뿐 아니라 사용하는 커널함수의 모수에 해당하는 값에 따라서도 모델의 예측력에 영향을 받는다. 따라서 예측력이 우수한 모델을 개발하기 위해서는 최적의 C값과 커널함수 모수 값의 결정이 필요하다. 본 연구에서는 기존의 SVM 관련 연구를 참고하여, C값은 0.1, 1.0, 50, 100의

값을 분석 대상으로 하고, RBF 커널함수의 모수 값은 0.1에서 50 사이의 여러 값을 최적 예측 모델구현을 위한 분석 대상으로 하였다. 연산을 종료하는 조건인  $\delta$  값은 0.01을 적용하였다<sup>9,13,15)</sup>. C값과 커널함수 모수 값의 변화에 따른 예측 모델의 오차율을 분석하여 가장 우수한 예측력을 보이는 C값과 모수 값을 구하고 이를 최적 모델 구현에 이용하였다.

마지막으로 수집된 입/출력 데이터를 이용하여 SVM 예측 모델을 구현하기 위해서는, 식 (8)에 대한 최적해를 구할 수 있는 Solver가 필요하다. 본 연구에서는 Mathworks<sup>®</sup>사에서 제공하는 MATLAB기반 SVM Solver인 ‘svmtrain’을 이용하여<sup>16)</sup>, SVM 예측 모델을 구현하였다.

4. 개발 모델의 검증 및 적용성 검토

4.1. 데이터의 수집 및 구성

본 연구에서는 대한건설협회의 완성공사원가통계 분석 시에 사용된 원가 결산 자료를 수집하여 모델 개발 및 검증을 위한 데이터로 활용하였다. Table 2에 나타난 바와 같이 수집된 자료는 2013년에 완공된 공사에 대한 공사원가 결산 자료이며, 공사계약액 기준 50억 이상인 공사 150건을 수집하였다.

Table 2. Data descriptions

Construction types	Number of data	Remarks
Civil	53 (10)	Use
Architectural	81 (10)	Use
Industrial facility	12	Discard
Landscape	4	Discard
Total	150	

본 연구에서는 학습용 데이터로 충분하지 않은 숫자로 판단되는 산업설비 공사 및 조경 공사에 대한 자료 16건은 제외하고, 토목 및 건축 공사 자료 134건만 이용하였다. 총 134건의 자료 중 예측 모델의 구현 및 검증과정에서는 토목공사 10건과 건축공사 10건을 제외한 114건의 자료만을 이용하였으며, 제외된 나머지 데이터는 최적으로 구현된 모델의 예측 성능 평가에 이용하였다.

4.2. 결과 및 분석

본 연구에서는 k-중첩교차조사(k-fold cross validation) 방법을 이용하여 개발 모델의 정확도를 검증하였다<sup>13,15)</sup>. k-중첩교차조사 방법은 수집된 데이터를 크기가 비슷한 k개의 부분 집합( $S_1, \dots, S_k$ )으로 임의 분할한 후, 이들 부분

집합을 이용하여 학습 및 검정을 k번 반복 수행하는 검증 방식이다. 즉, 첫 번째 반복에서는  $S_1$ 을 제외한 나머지 부분 집합  $S_2, \dots, S_k$ 를 이용하여 학습을 수행하고, 학습된 모델에 대한 정확도를  $S_1$ 을 이용하여 검정한다. 본 연구에서는 전체 데이터의 개수를 고려하여, 6-중첩교차조사를 수행하였다. 교차조사 시 예측 모델의 정확도는, 실제 안전보건관리비 금액과 SVM에 의해 예측된 금액과의 오차율을 이용하여 분석하였다.

우선 예측 모델에 적용할 C값과 커널함수의 모수값을 결정하기 위해, 6-중첩교차조사의 검정 과정에서 구한 결과에 대한 오차율의 평균을 C값과 커널함수의 모수값에 따라 비교하였으며, 그 결과를 Fig. 1에 나타내었다. 그림에서 확인할 수 있듯이 C값과 커널함수의 모수 값에 따라 최대 약 20%, 최소 약 5% 정도의 오차율을 보임을 알 수 있다. 일부 결과에서 오차율의 요동(fluctuation)이 발생하는 경우가 있으나, 전반적으로 커널함수의 모수 값이 증가함에 따라 오차율은 감소하다가 다시 증가하는 경향이 나타남을 알 수 있다. 오차율의 변동이 가장 작게 나타나는 경우는 C값이 50인 경우이며, 가장 큰 변동을 보이는 것은 C값이 1.0인 경우임을 알 수 있다. 한편, C값이 1.0인 경우는 변동성은 상대적으로 크지만, 전체에서 가장 낮은 오차율인 4.85%의 오차율을 보임을 알 수 있다. 이러한 결과를 바탕으로, 본 연구에서는 최적 예측 모델의 구현에 적용할 C값으로, 여러 모수 값에 대해 오차율의 변동성이 상대적으로 가장 작은 C=50을 선정 하였으며, 커널함수의 모수 값은 C=50인 경우 중에서 최소 오차율이 나타난 30을 선정하였다.

한편, 본 연구에서의 최적 예측 모델인 C=50이고 커널함수의 모수값이 30인 SVM 예측 모델에 대한 예측 성능을 평가하기 위해, 최적 예측 모델의 구현에 사용

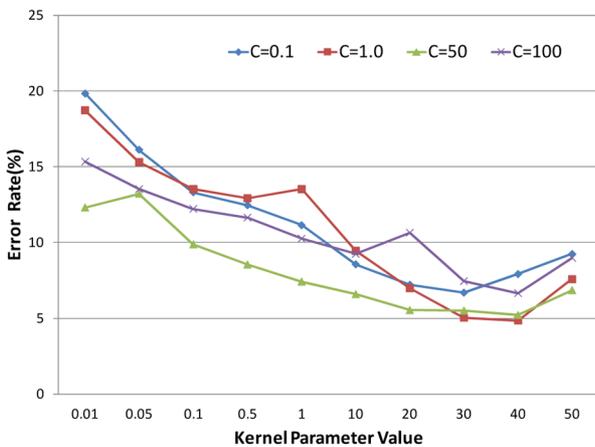


Fig. 1. Error rate for various parameter values.

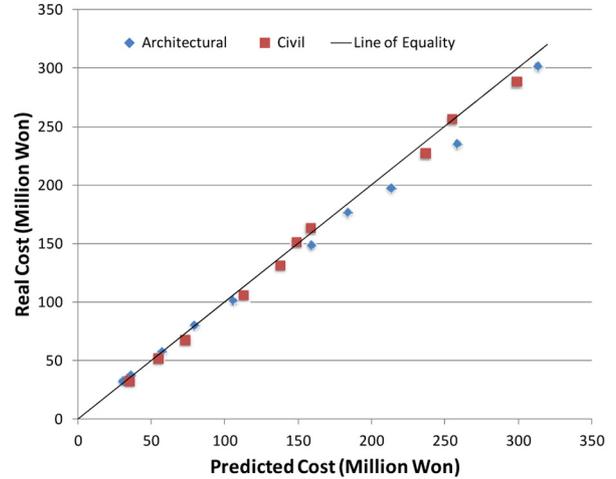


Fig. 2. Prediction performance.

Table 3. Prediction error

Civil (Million won)			Architectural (Million won)		
Real	Predicted	Error (%)	Real	Predicted	Error (%)
29.9	33.2	11.03	33.3	34.5	3.60
35.2	38.8	10.22	52.2	53.8	3.06
56.4	58.2	3.19	68.7	72.3	5.24
78.9	81.2	2.91	107.2	112.1	4.57
105.3	102.1	3.03	132.3	137.3	3.78
158.3	149.5	5.55	152.4	148.2	2.75
183.2	177.8	2.94	164.2	157.6	4.02
213.2	198.1	7.08	228.1	236.3	3.60
258.1	236.3	8.44	257.3	254.1	1.24
313.3	302.8	3.35	289.1	298.2	3.14
Average		5.77	Average		3.50

하지 않은 데이터를 이용하여 예측 성능을 평가하였다. Fig. 2는 예측에 사용하지 않은 토목공사에 대한 데이터 10개와 건축공사에 대한 데이터 10개에 대한 안전보건관리비의 예측값과 실제값을 비교한 결과를 나타내며, Table 3은 이에 대한 오차 분석 결과를 나타낸다.

전반적으로 예측값이 실제값과 큰 차이가 나지 않는 것을 그림에서 확인할 수 있다. 그러나, Table 3에서 확인할 수 있듯이, 세부적으로는 토목공사의 경우 최대 11% 정도의 오차가 발생하며, 안전보건관리비가 40백만원 이하인 구간과 200백만원 이상인 구간에서 오차가 크게 나타나는 것을 알 수 있다. 반면에 건축공사의 경우는, 최대 오차가 약 5.2% 정도이고, 전체 범위에서 고른 오차율을 보일 뿐만 아니라 평균 오차도 3.5%로 토목공사에 비해 작은 값을 보임을 알 수 있다. 이러한 차이가 나타나는 이유는 여러 가지가 있

을 수 있으나, 모델 구현에 사용된 데이터의 개수가 건축공사에 비해 토목공사가 적은 것이 주된 이유로 사료된다. 한편, 전체의 평균 오차는 토목공사가 건축공사에 비해 상대적으로 크지만, 이를 적중률로 환산해 보면 약 94%의 적중률로 매우 우수한 적중률을 보임을 알 수 있다. 이러한 결과는 개발된 SVM 예측 모델이 안전보건관리비의 예측에 유효하게 적용 가능하다는 것을 입증한다.

### 5. 결론

본 연구에서는 SVM 기법을 이용한 건설공사 안전보건관리비 예측 모델을 개발하고자 하였으며, 이를 위해 (1) 예측 모델 구현을 위한 SVM 기법의 이론을 고찰하고, (2) SVM 기법을 이용한 안전보건관리비 예측 모델을 구현하였으며, (3) 실제 데이터를 이용하여 구현된 예측 모델의 예측 성능을 평가하여 실 적용 가능성을 검토하였다. 본 연구의 결과로부터, 개발된 SVM 예측 모델은 건설공사 안전보건관리비를 상대적으로 정확하게 예측하는 것을 알 수 있었으며, 또한 개발된 SVM 예측 모델이 실제 안전보건관리비의 예측에도 유효하게 적용이 가능하다는 것을 알 수 있었다. 한편, 본 연구에서 개발된 SVM 예측 모델은 토목공사와 건축공사에 대해서만 그 적용성이 검토되었으며, 타 공종의 공사에 대해서는 이에 대한 검토가 아직 이루어지지 않은 한계점이 있다. 따라서, 보다 보편적으로 사용할 수 있는 안전보건관리비 예측 모델을 구현하기 위해서는, 추후 다양한 공종의 건설공사에 대한 자료를 확보하여 적용성을 검토하는 연구가 필요할 것으로 사료된다.

### References

- 1) S. W. Oh, Y. S. Kim, S. H. Choi and J. W. Choi, "A Study on the Estimation of Occupational Safety and Health Expense Rate by Safety Environment Change in Construction Industry", *Journal of Korea Institute of Construction Engineering and Management*, Vol.14, Issue 4, pp. 97-107, 2013.
- 2) K. T. Jung, "Developing Criteria for Standard Safety Management Cost", *Korea Industrial Safety Corporation*, pp. 51, 1997.

- 3) K.S. Son, W.M. Gal and H.S. Yang, "A Study on the Estimating Rate of Safety Management Cost in Building Work", *Journal of Korean Society of Safety*, Vol.22, Issue 5, pp. 33-40, 2007.
- 4) M. G. Lee, "The Ways to Enhance the Efficiency of the Occupational Safety and Health Expenses Operation System", *Research Report, KOSHA*, 2009.
- 5) K. S. Son, "Establishing Appropriate Rate for Standard Safety and Health Management Cost", *Research Report, KOSHA*, 2005.
- 6) A. Ashworth, "Cost Studies of Buildings - 3rd Edition", *Harlow Longman, USA*, 1999.
- 7) G. H. Kim, S.Y. Kim and K. I. Kang, "Comparing Accuracy of Prediction Cost Estimation using Case-Based Reasoning and Neural Networks", *Journal of Architectural Institute of Korea-Structure*, Vol.20, Issue 5, pp.93-103, 2004.
- 8) G. H. Kim, S. H. An and H. K. Cho, "Comparison of the Accuracy between Cost Prediction Models based on Neural Network and Genetic Algorithm", *Journal of Architectural Institute of Korea-Structure*, Vol.22, Issue 3, pp.111-118, 2006.
- 9) S. H. An, U.Y. Park, K.I. Kang, M.Y. Cho and H.H. Cho, "Application of Support Vector Machines in Assessing Conceptual Cost Estimates", *ASCE Journal of Computing in Civil Engineering*, Vol.21, Issue 4, pp.259-264, 2007.
- 10) A. O. Elfaki, S. Alatawi and E. Abushandi, "Using Intelligent Techniques in Construction Project Cost Estimation: 10-Year Survey", *Advanced in Civil Engineering*, Vol.2014, Article ID 107926, 2014.
- 11) M. Mohri, A. Rostamizade and A. Talwalkar, "Foundations of Machine Learning", *MIT Press, USA*, 2012.
- 12) V. N. Vapnik, "The Nature of Statistical Learning Theory", *Springer-Verlag, Germany*, 1995.
- 13) A. Smola and B. Scholkopf, "A Tutorial on Support Vector Regression", *Technical Report, Royal Holloway College, UK*, 1998.
- 14) S. G. Noh, "Construction Cost Pricing", *Korea Corporation Management Research Institute*, 2013.
- 15) J. S. Park, "Comparison on Support Vector Regression and Artificial Neural Network Techniques in Data Mining", *Doctoral Dissertation, Dongguk University*, 2006.
- 16) <http://www.mathworks.com>