Journal of the Korean Society of Safety
Vol. 37, No. 1, pp. 41-48, February 2022
https://doi.org/10.14346/JKOSOS.2022.37.1.41

ISSN 1738-3803 (Print)
ISSN 2383-9953 (Online)
http://www.kosos.or.kr/jkosos

교량 건설 문서의 강화된 XML 스키마 매칭을 위한 인공신경망 기반의 요소 가중치 선정 방안

박상일* · 권태호** · 박준원*** · 서경완** · 윤영철****†

Artificial Neural Network-based Weight Factor Determination Method for the Enhanced XML Schema Matching of Bridge Engineering Documents

Sang I. Park* · Tae-Ho Kwon** · Junwon Park*** · Kyung-Wan Seo** · Young-Cheol Yoon****†

†Corresponding Author
Young-Cheol Yoon
Tel : +82-2-300-1135
E-mail : ycyoon@mjc.ac.kr

Received : January 18, 2022
Revised : February 18, 2022
Accepted : February 21, 2022

Copyright@2022 by The Korean Society of Safety All right reserved.

Abstract : Bridge engineering documents have essential contents that must be referenced continuously throughout a structure's entire life cycle, but research related to the quality of the contents is still lacking. XML schema matching is an excellent technique to improve the quality of stored data; however, it takes excessive computing time when applied to documents with many contents and a deep hierarchical structure, such as bridge engineering documents. Moreover, it requires a manual parametric study for matching elements' weight factors, maintaining a high matching accuracy. This study proposes an efficient weight-factor determination method based on an artificial neural network (ANN) model using the simplified XML schema-matching method proposed in a previous research to reduce the computing time. The ANN model was generated and verified using 580 data of document properties, weight factors, and matching accuracy. The proposed ANN-based schema-matching method showed superiority in terms of accuracy and efficiency compared with the previous study on XML schema matching for bridge engineering documents.

Key Words : bridge engineering document, XML, schema matching, artificial neural network, weight factor, matching accuracy

1. 서 론

구조물과 관련된 설계, 시공 및 운영/유지관리 등에 필요한 각종 지침, 설계서 등의 건설 문서는 건설 프로젝트 참여자에 의해 구조물의 생애주기 각 단계뿐만 아니라 폐기 단계를 포함하는 전 생애주기에 걸쳐 문서의 내용을 지속적으로 참조된다. 특히, 건설 프로젝트에서는 그 특성상 하나의 문서 내에서 깊이 있는 내용을 참조하기보다는 여러 문서에서 특정 내용을 다양하게 접근하여 검토하는 경우가 대부분이다[1]. 이러한 과정이 원활하게 진행되도록 하기 위한 중요 사항은 사용자가 필요로 하는 컨텐츠에 정확하고 빠르게 접근하고 추출하는 데 있다. 깊은 계층 구조 형태로 많은 내용을 담고 있는 건설 문서의 특성상 위의 과정은 컴퓨터를 활용한 자동화가 필수적인 요소이다. 컴퓨터를 활용한 건설 문서의 활용은 컴퓨터의 문서 정보의 인지가 가장 핵심적인 부분이다. 이를 위해서 문서 내용 간의 계층 또는 관계 정보의 생성, 의미의 인지 등에 대한 연구가 활발히 수행되었다. 초기 연구들은 주로 정보를 공유 및 교환하는 과정에서 발생하는 문서 정

*국토안전관리원 안전성능연구소 책임연구원 (Research Institute for Safety Performance, Korea Authority of Land & Infrastructure Safety)
**연세대학교 건설환경공학과 박사과정 (School of Civil & Environmental Engineering, Yonsei University)
***명지전문대학 토목공학과 겸임교수 (Department of Civil Engineering, Myongji College)
****명지전문대학 토목공학과 교수 (Department of Civil Engineering, Myongji College)

보들을 정해진 체계에 따라 저장하고 이를 응용 분야에서 활용하는 방식에 대한 것에 초점을 맞추었다[2-4]. Suh[5]는 신기술의 위험도를 안전공학의 관점에서 정량적으로 평가하기 위해서 신기술 관련 문서에서 추출한 키워드를 계층분석과정(AHP: Analytic Hierarchy Process)에 적용하였다. 건설안전 분야에서 Kim 등[6]은 사고조사보고서의 내용을 디지털 연산이 가능한 데이터로 치환하여 추출한 키워드의 내용을 확률적으로 추론함으로써 사고 발생유형을 정량적으로 분석하는 방법을 제안하였다. 이런 형태에서 발전하여 건설 문서를 자동으로 분류하여 지식 자원으로 활용할 수 있는 방안을 제시하거나[7,8], 다중관점으로 문서 구조를 정의하여 문서 내 특정 정보를 사용자에게 제공하는 방식[1] 등도 제시되었다. 그러나 이러한 방식은 문서 내 정보 항목 간의 계층이나 관계에 대한 정보는 다루지 않거나 이들 정보가 미리 생성되어 있다는 가정에 기초한다. 즉, Liu 등[9]의 연구에서 강조한 바와 같이 문서 정보를 효율적이고 효과적으로 관리하고 공유하기 위해서는 문서 자체에 대한 자동화된 분석이 선행되어야 한다. 과거에는 쉽게 접근하기 어려웠던 문서에 대한 자동화된 분석은 최근 인공지능의 발전과 확산으로 인해 그 효율성과 정확도가 점차 향상되어 가고 있다. 이러한 접근 방식은 보통 자연어 처리 방식으로의 접근, 규칙 기반의 접근, 의미론적인 접근 또는 이들의 조합을 통한 접근 방법 등으로 구분할 수 있다[10]. 건설 문서는 일반 문서보다 모호성이 적기 때문에 자연어 처리를 통한 문서 정보의 추출 및 활용에 대한 향후 기대가 높으나, 현재는 추출 정확도에 대한 개선의 여지가 많다[11,12]. 규칙 기반 접근은 자연어 처리 방식의 접근에 비해 더 많은 사전 노력이 필요하지만, 영역에 특화된 규칙을 생성하거나 규칙의 지속적 보완을 통해 문서 구조 분석이나 정보 추출에 있어 효과적이다. Kim 등[13]이 제시한 문장의 상대적 비교를 통한 건설 문서의 구조화나 Zhang과 El-Gohary[10]이 제시한 PSG 기반의 통사분석(syntactic analysis)을 통한 문서 정보 추출이 대표적인 사례이다. 의미론 기반의 접근은 영역(domain) 온톨로지를 사용하여 건설 문서 또는 문서가 포함하는 내용 분류의 정확도를 높이고 있다[14,15]. 그러나 이를 위해서는 문서에 대한 온톨로지 모델을 구축하는 사전 작업이 요구되며, 문서 정보 추출의 정확도는 온톨로지 모델의 영향을 받는다. 스키마 매칭은 두 개 이상의 스키마 구성요소의 의미적 관계를 식별하는 프로세스로 사용되거나[16], 별도의 사전 작업 없이 의미론을 기반으로 문서 구조 분석에 사용되고 있다[17-19]. 여기서 스키마는 데이터베이스를 기술하기 위해 사용되며, 데이터의 구조를 정의하는 개념이다. 본 연구에서는 디지털화된 건설 문서의 참조 측면에서의 활용을 위해 이를 이용하였다. 특히, 고품질의 스키마와의 비교를 통해 기존에 생성되어 있는 건설 문서 구조 분석 및 정보 추출의 정확도를 향상시킬 수 있다[20,21]. Park과 Lee[22]의 연구에서는 교량 건설 문서의 특징인, 컨텐츠의 양이 많고 계층 깊이가 깊은 경우에 효과적인 스키마 매칭 방법을 제안하였다. 그러나 매칭 결과의 정확도를 유지하기 위해서는 문서별로 달라지는 최적의 매칭 가중치를 선정해야 하는데, Park과 Lee[22]의 연구에서 제시한 의사결정나무의 역연산 방식은 기존 데이터를 활용하여 결과를 예측하는 방식이기 때문에 새로운 입력 데이터에 대해서는 특정된 값이 아니라 범주를 예측 한다.

본 연구에서는 건설 문서 분석에 있어 의미론적 접근 방식인 스키마 매칭에 기계 학습 중 대표적인 인공신경망(Artificial Neural Network, ANN)을 이용한 회귀분석 기법을 접목하여 데이터 품질을 보장하면서 건설 문서의 자동 분석을 지원하는 방안을 제시하였다. 선행연구[22]에서 제시된 교량 건설 문서에 특화된 스키마 매칭 방안을 간략히 소개하고 이를 활용하여 ANN 모델을 생성하였다. 개선된 방법을 선행 연구 결과와 비교하여 교량 건설 문서에 대해 향상된 효율성을 검증하였다.

2. 교량 건설 문서를 위한 스키마 매칭 기법

스키마 매칭은 두 개의 스키마 모델 사이에서 의미론적인 관계에 기인한 상호 매핑 기술로서 표준화된 구조에 따른 정보 변환을 통해 저장 및 추출 데이터의 질적 향상을 가져올 수 있다. 자동화된 스키마 매칭은 요소의 이름이나 설명 등 언어적 특징을 이용하는 것에서부터 요소의 제약조건, 요소의 계층정보 활용, 스키마에 따라 생성된 실제 데이터 활용 등 여러 가지 방법이 있다[23-26]. 본 연구에서는 교량 건설 문서의 방대한 데이터 양, 이에 따른 데이터 형식 정의에 소모되는 시간 비용, 그리고 깊은 계층 구조의 특징을 잘 반영하기 위하여 Extensible Markup Language (XML) 기반의 스키마 매칭을 사용하였다[20]. 이 XML 응용 스키마 매칭은 유사도 측정(similarity measuring)과 요소들 간의 관계를 고려하여 이를 반영하는 완화 라벨링(relaxation labeling)의 두 과정으로 구성된다. 유사도 측정은 식 (1)과 같이 나타낼 수 있다.

$$SM(s,t) = \sum_{e} (w_e \, N_e(s,t)) \tag{1}$$

여기서, s와 t는 비교하고자 하는 두 스키마에 속한 요소 하나를 나타낸다. e에 들어갈 수 있는 요소는 하나의 기준 요소(NE), 그 기준요소의 부모 요소(P), 형제 요소(S) 및 자식 요소(C)이다. N은 두 요소의 정량적인 유사도 값을 나타내는 함수이며, w는 N에 대응하는 유사도의 가중치이다. Park 등[21]의 연구에 따르면 XML 스키마 매칭 대상 문서에 따라 높은 정확도를 얻기 위한 w의 최적값은 변하지만 w와 매칭 정확도 사이의 직접적인 연관 관계를 규정하는 것은 어렵다.

XML 응용 스키마 매칭의 두 번째 과정인 완화 라벨링 과정은 요소 사이의 연결 관계나 구조적 제약을 반영하는 과정으로서 매칭 신뢰도는 식 (2a)-(2c)와 같이 표현할 수 있다.

$$R^{d+1}(m,k)=\frac{R^{t}(m,k)q^{t}(m,k)}{\sum_{u=1}^{v}P^{t}(m,u)q^{t}(m,u)} \tag{2a}$$

$$q(m,k)=\sum_{n=1}^{y}\sum_{u=1}^{v}\mu_{mn}(k,u)P(n,u) \tag{2b}$$

$$\psi_{mn}(k,u)=\begin{cases}1, \text{if } m=n\wedge k=u\\ 0, others\end{cases} \tag{2c}$$

여기서 m과 k는 두 스키마 요소를 나타내며, d는 반복되는 차수, y와 v는 두 스키마의 요소 수이다. q는 구조적 제약 거리를 정량화하기 위한 함수이며, μ는 구조적 연결 관계에 따른 유사도를 나타낸다. 구조적 연결 관계는 스키마의 크기가 작은 경우에는 전체 연산 비용에 큰 영향을 미치지 않지만, 건설 문서와 같이 요소 수가 많고 깊이가 깊은 경우에는 연산 시간이 과도하게 소모되는 치명적인 단점이 있다. 이에 따라 Park과 Lee[22]의 연구에서는 μ를 대신해서 사용할 수 있는 ψ를 제안하여 연산 속도를 대폭 상승시켰다(Fig. 1 참조). 이때 정확도의 유지를 위해 가중치(w) 조정의 방법을

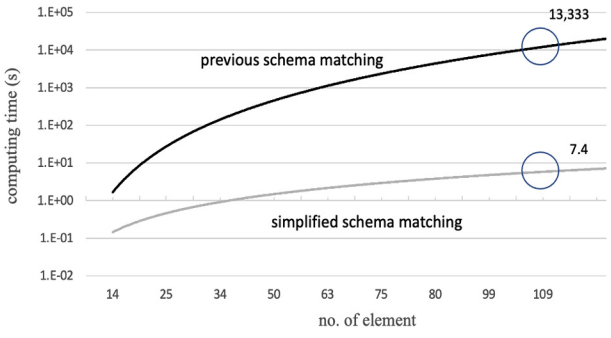

Fig. 1. Comparison of computing time of XML schema matching between the previous study and simplified model.

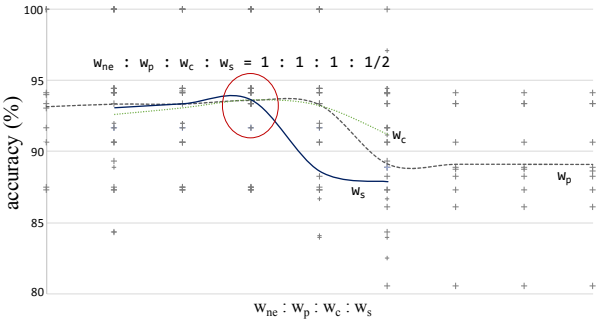

Fig. 2. Process for determining the proper XML schema matching weight factors for bridge circuit construction documents.

활용하였다(Fig. 2 참조). 가중치 조정의 방법은 Lin[27]이 제안한 equality constraint method의 개념을 활용하여 하나의 가중치를 제외한 나머지 가중치는 일정한 값으로 고정하고, 실험적으로 가장 높은 정확도를 보이는 가중치를 선택하는 방법이다.

3. 교량 건설 문서 XML 스키마 매칭 지원을 위한 인공신경망 모델 생성

3.1 효율적 스키마 매칭 적용을 위한 인공신경망 알고리즘의 선택

최적 설계는 주어진 조건을 만족시키는 한도 내에서 원하는 최적의 결괏값을 끌어내는 과정이라고 할 수 있다. 본 연구는 건설 문서를 대상으로 정확한 XML 스키마 매칭을 위해 유사도 측정 과정에서 요소 유사도의 가중치에 대한 최적값을 선정하는 것을 목표로 한다는 점에서 최적 설계와 유사한 과정을 거친다. 그러나 전술한 바와 같이 XML 스키마 매칭의 정확도와 이때 사용되는 유사도 가중치 간의 관계는 정량화할 수 없으며, 이것은 요소 유사도 가중치들이 XML 스키마 매칭의 직접적인 제약조건으로 사용될 수 없다는 것을 의미하여 최적 설계의 방식을 본 연구에 직접적으로 적용하는 것에는 어려움이 있다. 데이터 마이닝은 새로운 지식의 발견의 한 부분 과정으로 수집한 기존 데이터를 기반으로 통계학적 기술 등 여러 기법을 활용하여 데이터 간의 새로운 패턴을 찾아내는 과정이다[28,29]. 데이터 마이닝의 가장 중요한 목적은 예측이라고 할 수 있으며, 군집, 분류, 회귀 등의 방법을 사용하여 귀납적으로 지식을 구성할 수 있다. 즉, 데이터 마이닝은 주어진 제약조건 하에서 최적해를 구하는 문제에 활용할 수 있다. 인공신경망은 하드웨어적인 컴퓨터의 발전에 힘입어 최근 급격하게 활용도가 높아진 대표적인 데이터 마이닝의 한 방법이다. 인공신경망은 인간의 뉴런 구조를 본떠 만든 기계학습 모델로서 몇

개의 층을 만들어 그 안에 세포들을 넣고 다른 층간의 세포들을 연결해 놓은 상태에서 여러 입력 신호를 받아 가중치를 곱한 후 모두 더하여 일정 수준이 넘어가면 활성화되어 결과를 내보낸다. 인공신경망의 가장 중요한 특징은 적절한 데이터가 주어지면 연속적인 비선형 함수를 모델링하여 근삿값을 예측할 수 있다는 것인데[30], 이는 본 연구에서 목표로 하는 정량화가 어려운 요소 가중치를 선정하기에 매우 적절하다. 이 중 역전파(backpropagation) 학습 알고리즘이 계산의 효용성과 편의성으로 인해 널리 사용되는데, 역전파 알고리즘은 출력값을 측정값과 비교하여 평균 자승 오차를 최소화한다[30]. 이러한 인공신경망 모델을 적용하는 경우, 훈련에 의해 생성된 모델이 실제 데이터에서는 큰 오차를 나타내는 과적합(overfitting)의 문제가 발생할 수 있어서 이를 최소화하기 위해서는 훈련 데이터가 많을수록 좋다. Bayesian regularized artificial neural network (BRANN)은 자체적으로 정규화를 진행하기 때문에 ANN 모델의 검증 과정이 필요 없어 상대적으로 훈련 데이터가 부족한 경우에도 검증용 데이터도 훈련 데이터로 활용할 수 있는 장점이 있다[31,32]. Kim 등[33]의 연구에서 확인할 수 있는 바와 같이 대표적 교량 설계 문서인 구조계산서는 다른 교량 설계문서에 비해 컨텐츠의 양이 많고, 문서의 깊이가 깊으며, 일관되지 않는 컨텐츠가 다량 포함되어 있는 특징이 있다. 이로 인해서 문서 분석 결과가 가장 낮게 나오기 때문에 본 연구에서는 BRANN을 구조계산서 스키마의 매칭에 적용하였다.

3.2 교량 건설 문서 XML 스키마 매칭을 위한 인공신경망 모델

Park과 Lee[22]는 문서를 만든 기관, 구조물의 형식 및 문서의 양에 따라 건설 문서를 대상으로 하는 XML 스키마 매칭의 가중치 변화는 유의미한 관계가 있음을 보였다. 이에 따라 본 연구에서는 언급한 정보들과 XML 스키마 매칭에 사용하는 요소별 가중치들을 이용해서 총 580개의 매칭 정확도를 산정한 데이터를 활용하여 인공신경망 모델을 생성하였다. 2장에서 서술한 바와 같이 XML 스키마 매칭에 사용하는 가중치는 총 네 개(w_{ne}, w_s, w_c, w_p)로서, 이들의 합은 1이기 때문에 인공신경망 모델 생성을 위해 사용한 가중치는 세 개(w_{ne}, w_s, w_c)만 적용하면 된다. 모델 생성에 사용한 문서는 다섯 개의 다른 기관에서 생성되었으며, 구조물의 형식 역시 다섯 종류이다. 인공신경망 모델 생성을 위한 학습 과정은 Fig. 3과 같다.

전술한 바와 같이 BRANN에서는 검증을 위한 데이터가 필요 없으며, 이에 따라 본 연구에서는 총 580개의 데이터 중에서 학습을 위한 데이터는 검증용 데이터와 합하여 전체의 85%인 493개, 학습 모델의 테스트를 위한 데이터는 전체의 15%인 87개를 사용하였다. Fig. 4는 교량 건설 문서에 대한 XML 스키마 매칭을 위한 인공신경망 모델링 결과를 나타낸다. 회귀 모델과의 근접도를 나타내는 결정계수(coefficient of determination)인 R^2는 0.9621의 높은 값을 보였다. Fig. 5는 매칭 수행 후 직접 비교를 통해 획득한 정확도(exact accuracy)와 인공신경망을 통해 추정한 정확도(predicted accuracy)를 비교한 것으로서 두 방법이 거의 동일한 결과를 나타냄을 확인할 수 있다.

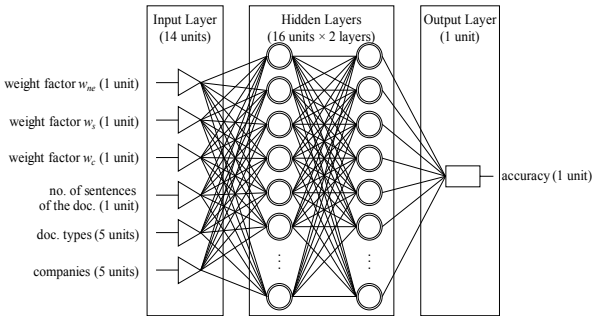

Fig. 3. The learning process for the XML schema matching for the bridge construction documents.

4. 인공신경망 모델을 이용한 교량 건설 문서의 XML 스키마 매칭 결과 추정 및 가중치 선정

본 연구에서 얻고자 하는 결과물은 교량 건설 문서의 XML 스키마 매칭의 정확도가 최대로 나오도록 하는 요소 가중치이다. 다음과 같은 절차를 통해 인공신경망 모델에서 요소 가중치를 얻어내어 실제 결과와 비교하였다.

i) 문서를 만든 기관, 구조물의 형식 및 문서의 양에 대한 정보 취득
ii) 스키마 매칭에 대한 요소 가중치(w_{ne}, w_s, w_c)에 대한 난수 생성. 이때, 세 가중치의 합은 1보다 작거나 같아야 한다.
iii) i)과 ii)의 과정을 통해 생성한 데이터를 입력 데이터로 사용하여, 생성해 놓은 인공신경망 모델에 적용하여 XML 스키마 매칭 정확도 추정
iv) iii)에서 가장 높은 정확도를 보이는 경우의 요소 가중치 선정(최대 정확도가 동일한 여러 경우가 나타나면 이 값들의 평균값 사용)
v) 선정한 요소 가중치를 이용하여 교량 건설 문서를 대상으로 XML 스키마 매칭 수행

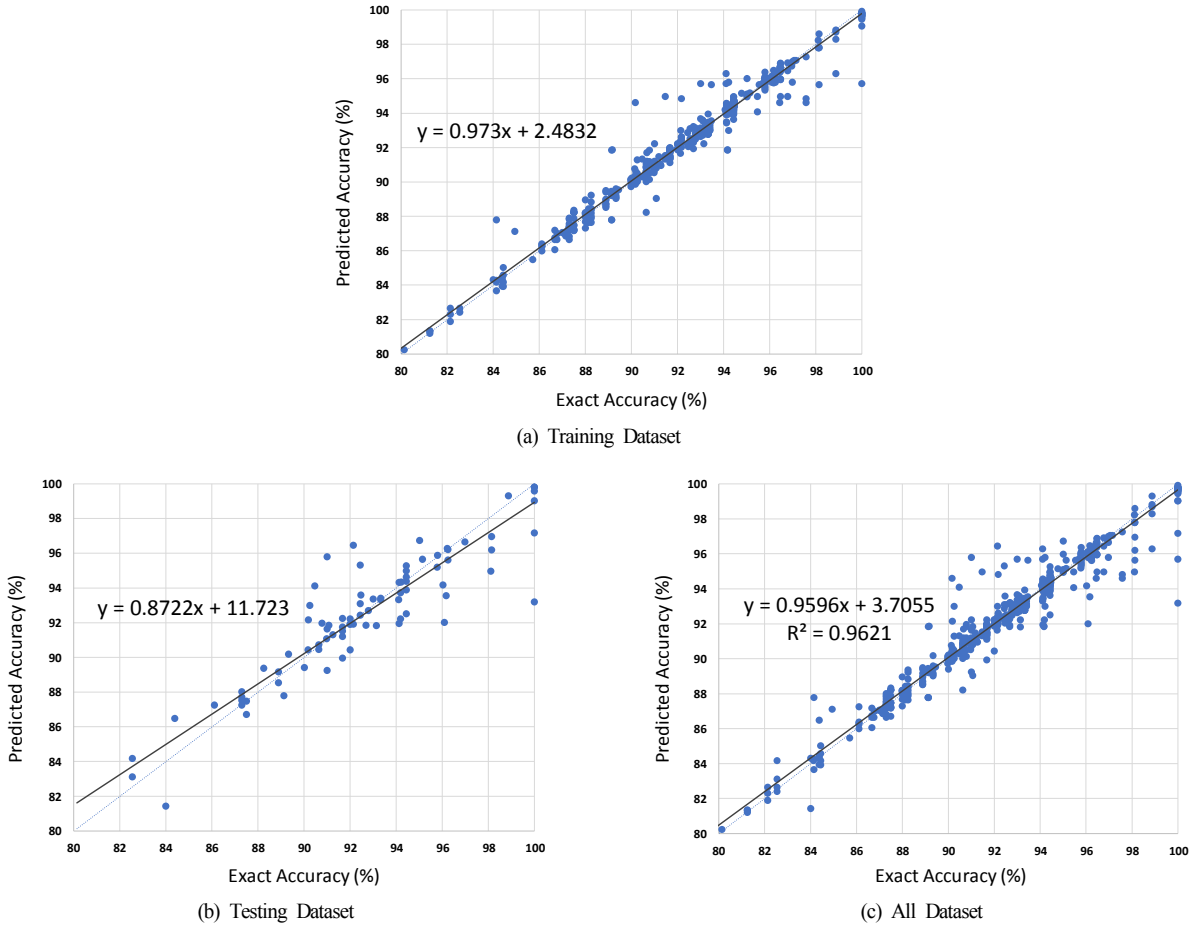

Fig. 4. ANN model for the XML schema matching for the bridge construction documents.

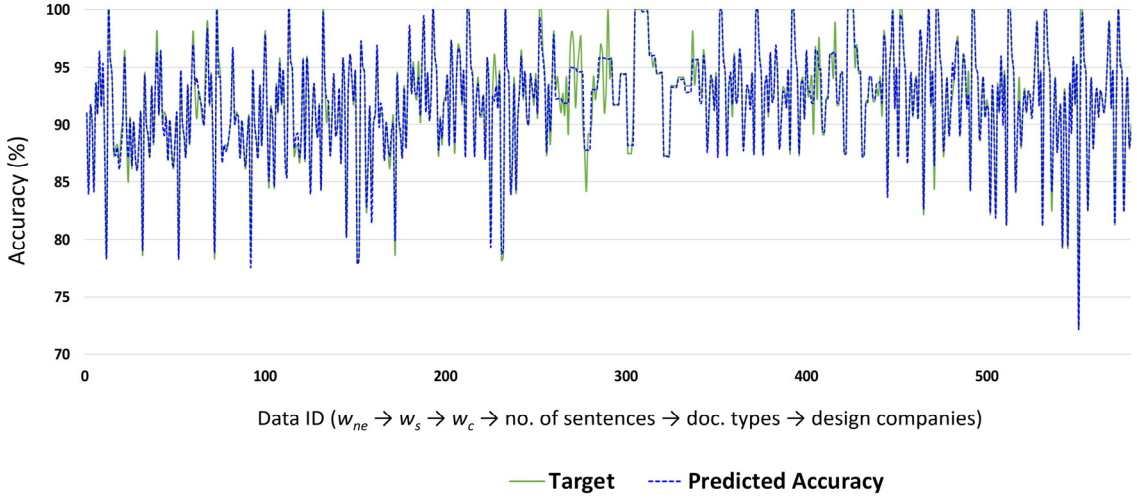

Fig. 5. Comparison of the target and predicted accuracy of the XML schema matching for the bridge construction documents.

Fig. 6은 전술한 과정을 따라 강플레이트 교량 형식과 'M Eng.' 회사에서 생성한 1,500문장을 갖는 건설 문서를 대상으로 인공신경망을 활용하여 추정한 정확도를 나타낸다. 요소 가중치 선정을 위해 1,000개의 난수를 생성하였다. 이때 매칭 정확도 추정치 100%를 나타내는 경우는 12가지가 생성되었고 이 경우의 w_{ne}, w_s, w_c, w_p는 각각 0.41, 0.17, 0.39, 0.03이 선정되었다. 이러한 방식으로 5회 반복하여 산정한 스키마 매칭의 평균

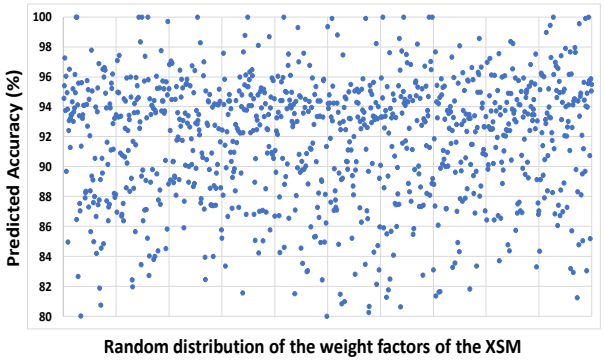

Fig. 6. Predicted accuracies of the XML schema matching (XSM) based on the random weight factors.

정확도는 98.23%로 나타났다. Park과 Lee[22]의 연구에서는 대표적인 5가지의 사례에 관한 결과를 제시하고 있는데, 본 연구에서 제안한 방법과 비교한 결과는 Table 1과 같다.

Table 1에서 알 수 있듯이, 본 연구에서 제안하는 가중치는 이전 연구에 비해 정확도는 모두 상승하고 있음을 확인할 수 있으며, 가중치에 대한 특정한 패턴은 나타나지 않았다. Fig. 7은 Park과 Lee[22]의 연구에서 최종 검증을 위해 추가적으로 사용한 26개의 문서에 대해 본 연구에서 제안한 방법과 스키마 매칭의 정확도를 비교한 것이다. 이때 막대그래프는 두 연구에 대한 정확도를 비교한 것으로 본 연구 방법에 대한 막대그래프는 점 데이터들의 평균값이다. 여기서 알 수 있듯이 3가지의 경우(case 9, 19, 24)를 제외하면 모두 건설 문서를 대상으로 하는 XML 스키마 매칭의 정확도가 상승하였음을 확인할 수 있다.

Table 1. Comparison of XML schema matching accuracy based on decision tree and ANN

Case	Input variable			Previous study Park과 Lee[22]		This study	
	bridge type (doc. type)	company	no. of sentences	weight factors	accuracy (%)	weight factors	accuracy (%)
1	cable stayed bridge	S Eng.	1028	w_{ne}=0.26, w_s=0.21, w_c=0.27, w_p=0.26	95.08	w_{ne}=0.27, w_s=0.19, w_c=0.50, w_p=0.04	98.38
2	steel plate bridge	D Eng.	845	w_{ne}=0.21, w_s=0.40, w_c=0.33, w_p=0.06	97.26	w_{ne}=0.28, w_s=0.27, w_c=0.43, w_p=0.02	99.07
3	v-type substructure	K Eng.	549	w_{ne}=0.32, w_s=0.13, w_c=0.38, w_p=0.17	96.71	w_{ne}=0.47, w_s=0.01, w_c=0.31, w_p=0.21	99.83
4	steel box girder bridge	Y Eng.	1826	w_{ne}=0.19, w_s=0.39, w_c=0.18, w_p=0.24	94.58	w_{ne}=0.46, w_s=0.01, w_c=0.19, w_p=0.34	99.87
5	cable stayed bridge	M Eng.	1933	w_{ne}=0.21, w_s=0.21, w_c=0.20, w_p=0.38	98.65	w_{ne}=0.28, w_s=0.20, w_c=0.35, w_p=0.17	98.69

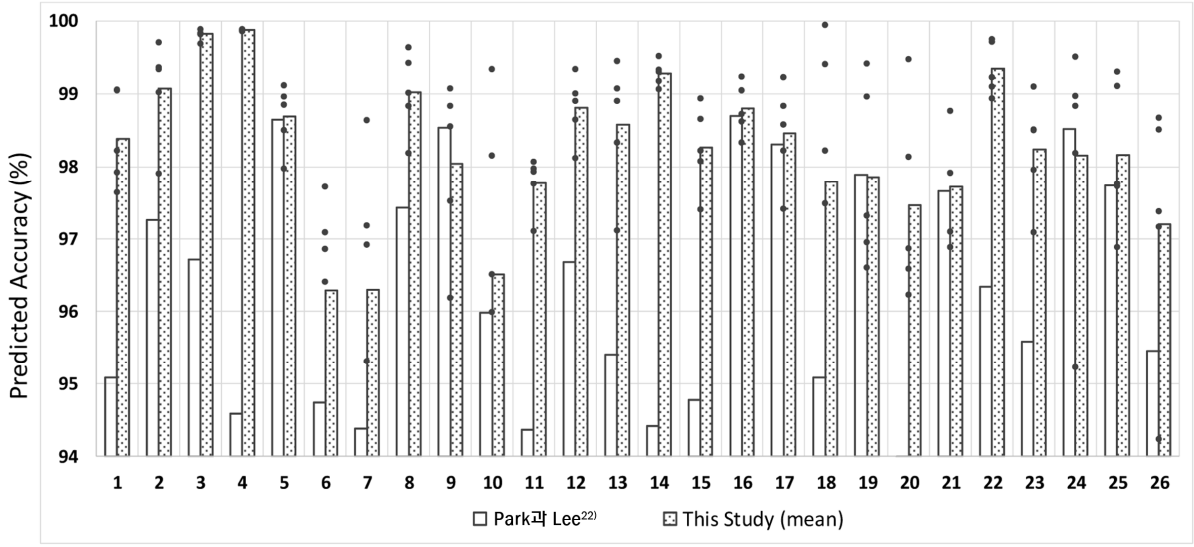

Fig. 7. Comparison of XML schema matching accuracies of bridge engineering documents between the previous study and this study.

5. 결 론

교량 건설 문서는 구조물의 전 생애주기에 있어 지속적으로 참조해야 하는 중요 자료임에도 불구하고 이 문서가 포함하고 있는 내부 데이터 질은 문서 생성자에 크게 의존하며, 한번 생성된 이후에 이들 내용에 대한 검토가 진행되는 경우는 극히 드물다. 본 연구자는 교량 건설 문서를 대상으로 문서의 질을 향상하기 위한 방법으로 XML 스키마 매칭을 성공적으로 적용한 바 있다. 이때, 건설 문서와 같이 문서의 구조가 깊고 문서가 포함하고 있는 요소의 수가 많은 경우 이를 그대로 적용하기에는 연산 시간이 과도하게 소요됨을 확인하였다. 이에 따라 간략화된 XML 스키마 매칭을 사용하는 경우 연산 시간은 성공적으로 줄였으나, 매칭 정확도를 높게 유지하기 위해서는 매칭에 사용되는 요소별 가중치를 조정해야 하는 필요성이 대두되었다. 따라서 본 연구에서는 인공신경망 모델을 활용하여 높은 정확도를 유지하는 XML 스키마 매칭 요소 가중치를 선정할 수 있는 방안을 제시하였다. 가중치 변화에 따른 건설 문서에 대한 XML 스키마 매칭을 위해 580개의 데이터를 활용하여 인공신경망 모델을 구성하였다. 이때 모델 구성을 위한 입력 변수는 문서를 만든 기관, 구조물의 형식 및 문서의 양과 스키마 매칭에 사용할 요소 가중치를 활용하였고 출력 변수는 스키마 매칭의 정확도로 설정하였다. 이를 활용하여 결정계수 0.9621의 값을 가지는 인공신경망 모델을 생성할 수 있었다. 또한 새로운 가중치를 선정하기 위해서 인공신경망 모델에 난수를 통해 생성한 매칭 가중치와 건설 문서에 대한 정보를 입력변수로 사용하여 가장 높은 매칭 정확도를 보이는 매칭 가중치를 선택하였다. 결과적으로 본 연구에서 제시한 방법을 통해 실제 건설 문서를 대상으로 매칭 정확도를 산정하고 이전 연구와의 비교를 통해 인공신경망 모델 사용의 효율성을 확인하였다. 또한, 건설안전 분야에서 자연어 기반 문서 자료에서 안전관리를 위해 알아야 할 위험요소를 정량적이고 시각적으로 추출할 수 있는 것처럼 건설안전 분야의 다른 다양한 영역으로의 활용을 기대해 볼 수 있다.

Acknowledgement: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1F1 A1048259).

References

1) S. Liu, C. A. McMahon, M. J. Darlington, S. J. Culley and P. J. Wild, “A Computational Framework for Retrieval of Document Fragments based on Decomposition Schemes in Engineering Information Management”, Adv. Eng. Inform., Vol. 20, No. 4, pp. 401-413, 2006.
2) Y. Zhu, R. R. Issa and R. F. Cox, “Web-based Construction Document Processing via Malleable Frame”, J. Comput. Civil. Eng., Vol. 15, No. 3, pp. 157-169, 2001.
3) M. Zhiliang, L. Heng, Q. Shen and Y. Jun, “Using XML to Support Information Exchange in Construction Projects”, Autom. Constr., Vol. 13, No. 5, pp. 629-637, 2004.
4) M. Zhiliang, K. D. Wong, L. Heng and Y. Jun, “Utilizing Exchanged Documents in Construction Projects for Decision Support based on Data Warehousing Technique”, Autom. Constr., Vol. 14, No. 3, pp. 405-412, 2005.
5) Y. Suh, “Data Analytics for Social Risk Forecasting and Assessment of New Technology”, J. Korean Soc. Saf., Vol. 32, No. 3, pp. 83-89, 2017.
6) B. S. Kim, S. Chang and Y. Suh, “Text Analytics for Classifying Types of Accident Occurrence using Accident Report Documents”, J. Korean Soc. Saf., Vol. 33, No. 3, pp. 58-64, 2018.
7) Y. Rezgui, “Ontology-centered Knowledge Management using Information Retrieval Techniques”, J. Comput. Civil. Eng., Vol. 20, No. 4, pp. 261-270, 2006.
8) L. Soibelman, J. Wu, C. Caldas, I. Brilakis and K.-Y. Lin, “Management and Analysis of Unstructured Construction Data Types”, Adv. Eng. Inform., Vol. 22, No. 1, pp. 15-27, 2008.
9) S. Liu, C. A. McMahon and S. J. Culley, “A Review of Structured Document Retrieval (SDR) Technology to Improve Information Access Performance in Engineering Document Management”, Comput. Ind., Vol. 59, No. 1, pp. 3-16, 2008.
10) J. Zhang and N. M. El-Gohary, “Semantic NLP-based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking”, J. Comput. Civil. Eng., Vol. 30, No. 2, p. 04015014, 2016.
11) M. Al Qady and A. Kandil, “Concept Relation Extraction from Construction Documents using Natural Language Processing”, J. Constr. Eng. Manage., Vol. 136, No. 3, pp. 294-302, 2010.
12) D. A. Salama and N. M. El-Gohary, “Automated Compliance Checking of Construction Operation Plans

using a Deontology for the Construction Domain", J. Comput. Civil. Eng., Vol. 27, No. 6, pp. 681-698, 2013.
13) B.-G. Kim, S. I. Park, H.-J. Kim and S.-H. Lee, "Automatic Extraction of Apparent Semantic Structure from Text Contents of a Structural Calculation Document", J. Comput. Civil. Eng., Vol. 24, No. 3, pp. 313-324, 2010.
14) N. M. El-Gohary and T. E. El-Diraby, "Domain Ontology for Processes in Infrastructure and Construction", J. Constr. Eng. Manage., Vol. 136, No. 7, pp. 730-744, 2010.
15) K. Liu and N. El-Gohary, "Ontology-based Semi-supervised Conditional Random Fields for Automated Information Extraction from Bridge Inspection Reports", Autom. Constr, Vol. 81, pp. 313-327, 2017.
16) E. Rahm and P. A. Bernstein, "A Survey of Approaches to Automatic Schema matching", the VLDB Journal, Vol. 10, No. 4, pp. 334-350, 2001.
17) J. Pouliot, S. Larrivée, C. Ellul and A. Boudhaim, "Exploring Schema Matching to Compare Geospatial Standards: Application to Underground Utility Networks", Intl. Arch. Photogramm, Remote Sens. Spat. Inf. Sci., Vol. 42, No. 4/W10, pp. 157-164, 2018.
18) F. Troncoso-Pastoriza, P. Eguía-Oller, R. P. Díaz-Redondo and E. Granada-Álvarez, "Generation of BIM Data based on the Automatic Detection, Identification and Localization of Lamps in Buildings", Sust. Cities Soc., Vol. 36, pp. 59-70, 2018.
19) S. Yang, R. Wei, and A. Shigarov, "Semantic Interoperability for Electronic Business through a Novel Cross-context Semantic Document Exchange Approach", in Proceedings of the ACM Symposium on Document Engineering 2018, 2018, pp. 1-10.
20) S. Yi, B. Huang and W. T. Chan, "XML Application Schema Matching using Similarity Measure and Relaxation Labeling", Inf. Sci., Vol. 169, No. 1-2, pp. 27-46, 2005.
21) S. I. Park, B.-G. Kim and S.-H. Lee, "An Efficient Application of XML Schema Matching Technique to Structural Calculation Document of Bridge", KSCE J. Civ. Eng., Vol. 32, No. 1D, pp. 51-59, 2012.
22) S. I. Park and S.-H. Lee, "Heuristic Solution using Decision Tree Model for Enhanced XML Schema Matching of Bridge Structural Calculation Documents", Front. Struct. Civ. Eng., Vol. 14, No. 6, pp. 1403-1417, 2021.
23) W.-S. Li and C. Clifton, "SEMINT: A Tool for Identifying Attribute Correspondences in Heterogeneous Databases using Neural Networks", Data Knowl. Eng., Vol. 33, No. 1, pp. 49-84, 2000.
24) J. Madhavan, P. A. Bernstein and E. Rahm, "Generic Schema Matching with Cupid", in vldb, 2001, Vol. 1: Citeseer, pp. 49-58.
25) A. Doan, J. Madhavan, P. Domingos and A. Halevy, "Learning to Map between Otologies on the Semantic Web", Proceedings of the 11th International Conference on World Wide Web, pp. 662-673, 2002.
26) S. Melnik, H. Garcia-Molina and E. Rahm, "Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching", Proceedings 18th International Conference on Data Engineering, IEEE, pp. 117-128, 2002.
27) J. Lin, "Multiple-objective Problems: Pareto-optimal Solutions by Method of Proper Equality Constraints", IEEE Trans. Autom. Control, Vol. 21, No. 5, pp. 641-650, 1976.
28) U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery in Databases", AI Mag., Vol. 17, No. 3, pp. 37-37, 1996.
29) P. Adriaans and D. Zantinge, "Data Mining Addison Wesley Longman Limited", Edinbourgh Gate, Harlow, CM20 2JE, England, 1996.
30) F. Burden and D. Winkler, "Bayesian Regularization of Neural Networks", Artif. Neural Netw., pp. 23-42, 2008.
31) D. J. MacKay, "A Practical Bayesian Framework for Backpropagation Networks", Neural Comput., Vol. 4, No. 3, pp. 448-472, 1992.
32) D. M. Hawkins, S. C. Basak and D. Mills, "Assessing Model fit by Cross-validation", J. Chem. Inf. Comput. Sci. Vol. 43, No. 2, pp. 579-586, 2003.
33) B. G. Kim, S. I. Park, H. J. Kim and S. H. Lee, "Automatic Extraction of Apparent Semantic Structure from Text Contents of a Structural Calculation Document", J. Comput. Civil. Eng., Vol. 24, No. 3, pp. 313-324, 2010.