

# 공사일지의 텍스트 마이닝을 통한 우천 공기지연 리스크 정량화

박종호\* · 조민건\*\* · 엄세호\*\*\* · 박선규\*\*\*\*

Park, Jongho\* · Cho, Mingeon\*\* · Eom, Sae Ho\*\*\* · Park, Sun-Kyu\*\*\*\*

## Quantification of Schedule Delay Risk of Rain via Text Mining of a Construction Log

### ABSTRACT

Schedule delays present a major risk factor, as they can adversely affect construction projects, such as through increasing construction costs, claims from a client, and/or a decrease in construction quality due to trims to stages to catch up on lost time. Risk management has been conducted according to the importance and priority of schedule delay risk, but quantification of risk on the depth of schedule delay tends to be inadequate due to limitations in data collection. Therefore, this research used the BERT (Bidirectional Encoder Representations from Transformers) language model to convert the contents of a construction log, which comprised unstructured data, into WBS (Work Breakdown Structure)-based structured data, and to form a model of classification and quantification of risk. A process was applied to eight highway construction sites, and 75 cases of rain schedule delay risk were obtained from 8 out of 39 detailed work kinds. Through a K-S test, a significant probability distribution was derived for four kinds of work, and the risk impact was compared. The process presented in this study can be used to derive various schedule delay risks in construction projects and to quantify their depth.

**Key words :** Construction log, Unstructured data, Text mining, Schedule delay risk, Quantification

### 초록

건설공사에서의 공기지연은 공사금액 증가, 발주처 클레임, 무리한 공기단축에 따른 건설공사의 질 하락 등 건설프로젝트에 악영향을 끼치는 주요 리스크 요인이다. 기존 연구에서는 공기지연 리스크의 중요도 및 우선순위를 파악하고 중요도에 따라 공정을 관리하였으나, 공기지연 리스크의 심도는 데이터 수집의 한계 등으로 정량화 연구가 미흡하다. 따라서 본 연구에서는 BERT (Bidirectional Encoder Representations from Transformers) 언어 모델을 활용하여 비정형데이터로 저장된 공사일지의 작업내용을 분석 가능한 WBS (Work Breakdown Structure) 기반의 정형데이터로 변환하고 리스크 분류 및 도출 체계, 공정계획에 사용가능한 리스크 발생확률, 리스크 확률분포(심도)의 정량화 방안을 제시하였다. 제안된 프로세스를 고속도로공사 8개 공구에 적용하여, 39개 세부 공중 중 8개의 세부 공중에서 75건의 우천 공기지연 리스크를 도출하였다. K-S 검정을 통해 4개 공중에서 유의미한 확률분포를 도출하였으며 위험도를 비교하였다. 향후 본 연구에서 제시된 프로세스는 시공단계에서 발생하는 다양한 공기지연 요인의 도출 및 심도 정량화에 적용될 수 있을 것으로 기대된다.

**검색어 :** 공사일지, 비정형데이터, 텍스트마이닝, 공기지연 리스크, 정량화

\* 종신회원 · 성균관대학교 리질리언트·에코스마트시티 연구교수 (Sungkyunkwan University · rhapsode@skku.edu)

\*\* 정회원 · 성균관대학교 글로벌스마트시티융합공학과 박사과정 (Sungkyunkwan University · raonik6713@naver.com)

\*\*\* 정회원 · 성균관대학교 글로벌스마트시티융합공학과 석사과정 (Sungkyunkwan University · alalsp767@g.skku.edu)

\*\*\*\* 종신회원 · 교신저자 · 성균관대학교 건설환경공학부 교수, 공학박사 (Corresponding Author · Sungkyunkwan University · skpark@skku.edu)

Received September 23, 2022/ revised November 11, 2022/ accepted December 12, 2022

## 1. 서론

### 1.1 연구의 배경 및 목적

건설프로젝트는 다양하게 투입되는 자원이 복잡하게 상호작용하며, 외부 환경요인에 노출되어 있기 때문에 원활한 공사를 진행하는데 있어서 다양한 리스크가 존재하고 있다. 건설프로젝트 리스크는 불확실성의 결과로서 프로젝트 목적에 악영향을 주는 사건의 발생 가능성에의 노출로 정의될 수 있다(AI-Bahar, 1989). 건설 리스크는 미국의 PMI(미국 사업관리 협회) 등 프로젝트 관리 체계에 관한 전문 기관들에 의하여 관리 프로세스가 상당 부분 표준화되어있으며, 각 사업 특성을 반영한 리스크 관리 프로세스 가이드라인이 존재한다(PMI, 2021).

다양한 건설 리스크 요인 중 공기지연 리스크는 공기 준수 및 공정계획의 합리화 측면에서 건설프로젝트의 성공과 밀접하다. 공기지연으로 인하여 준공일 내에 건설 프로젝트가 완료되지 못할 경우 시공자는 공사금액 증가, 발주처 클레임 등으로 인한 손해가 발생하며, 이를 방지하기 위한 무리한 공기단축은 건설공사의 질을 떨어뜨린다(Park, 2012).

건설프로젝트에서 나타나는 공기지연 리스크는 다양한 방법으로 연구되었다. Kang et al.(2002)은 정치·사회적, 기획단계, 입찰단계, 계약단계, 설계단계, 시공단계인자로 구분하여 리스크 인자의 중요도를 분석하고 건설공사 리스크 관리 시스템에서의 리스크분류체계를 제시하였다. 국내 건설사의 해외건설공사 리스크 저감을 위하여 계약문서 분석을 통한 계약단계에서의 공기지연 리스크가 분류되었으며(Lee, 2018; Yoon, 2017), 실제 공기지연 사례를 기반으로 요인과 사례를 연관하여 리스크 분류가 수행되었다(Choi, 2015; Kim, 2022; Yang, 2020). 건설프로젝트 전체공정에 가장 큰 영향을 미치는 철근콘크리트 및 철골 공사를 대상으로 공기 및 작업 지연요인의 항목 간 정량적 분석을 통해 우선순위를 산정하고 리스크 평가 프로세스가 제안되었다(Lee et al., 2009; Yoon et al., 2008). Park(2012)은 공기지연 리스크 요인을 관리하기 위하여 공공과 민관의 공기산정 방식을 조사하여 절충안을 마련하였으며, Yoon(2020)은 적정 공기산정 시 비작업일을 정형화하여 반영하는 방법을 개발하여 기온, 비, 눈, 바람, 미세먼지에 대한 비작업일을 산정하였다. Yu et al.(2017) 및 Ha et al.(2018)은 확률론적 추정 개념을 적용하여 건설 공사 현장 사고 및 인명피해 요인의 발생 빈도와 피해금액을 정량화하고 몬테카를로 시뮬레이션을 활용하여 리스크를 예측하였으나, 비교적 데이터를 확보하기 쉬운 피해금액을 활용하였다. Park et al.(2010)은 실적 공정표와 계획 공정표의 비교를 통해 외생 변수가 포함된 공기변화 확률분포를 도출하여 몬테카를로 시뮬레이션을 통한 공기 예측 모델을 개발하였으나, 공기지연 리스크의 상세한 분류가 수행되지 못하였

으며, 철근콘크리트 구조에 한정되어있다.

대부분의 공기지연 리스크 연구는 이미 발생한 리스크의 사례를 분석하고 전문가 설문조사를 통해 개별 리스크 요인의 중요도를 분석하여 우선순위를 제시할 수 있으나, 리스크 요인의 영향정도를 정량화하여 시뮬레이션 기반의 리스크 관리를 위한 DB를 구축하는 등의 방안 마련은 미흡하다.

건설 분야에서는 4차 산업혁명에 따른 빅데이터, 머신러닝, 인공지능 등의 기술들을 점차 확대, 적용하면서 다양한 센서를 활용한 디지털 트윈 개념의 데이터 수집 및 분석, 시뮬레이션 기반 분석을 통해 현실 문제 해결을 위해 노력하고 있다. 건설 리스크 분야에서는 건설 프로젝트의 결과물로서 발생하는 건설 문서를 텍스트 마이닝함으로써 리스크 요인을 분석하는 연구가 수행되고 있다. 건설사고에 따른 인터넷 기사 및 사고 보고서를 Word2Vec 알고리즘, 빈도 분석 및 중심성 분석, 주성분 분석 및 군집분석 등을 통하여 사고 원인 및 특징을 도출하였다(Kang and Yi, 2018; Kim and Kim, 2019; Park and Kim, 2021). 이외에 해외건설프로젝트 지연사례를 분석하여 지연사유와 대책을 분석하였으며(Kim, 2022), 계약 문서 분석을 통해 입찰 리스크를 도출하여 사업의 타당성을 검토하기 위한 기반을 마련하고(Lee and Yi, 2017; Marzouk and Enaba, 2019; Son and Lee, 2019), 건설 프로젝트 감독일지 분석을 통해 새로운 형태의 리스크를 정의하였다(Wang et al., 2020). 다만, 텍스트 마이닝을 활용한 건설 리스크 요인 분석도 기존의 요인 도출 과정을 스마트화 하거나, 새로운 형태의 리스크 요인을 도출하는데 활용되었을 뿐 공기지연 리스크 요인의 심도를 정량화하기 위한 분석 방안은 제시하지 못하고 있다.

따라서 본 연구에서는 건설 프로젝트 수행 과정에서 공기지연 요소를 가지고 있을 것으로 판단되는 비정형데이터를 도출하고 이를 텍스트 마이닝 기법을 통해 정형데이터로 변환한 후 리스크 요인을 정량화하고자 한다. 건설 현장에서의 공정정보는 PMIS를 활용하여 체계적으로 관리되고 있으나 상세한 건설 정보를 포함하는 공사일지 등의 현장정보는 형식적인 수준으로 저장된 비정형 데이터로써 활용성이 낮다. 다만, 공사일지가 포함하는 비정형데이터에 다양한 리스크 정보가 포함되어 있을 것으로 판단되며, 본 연구에서는 공사일지를 텍스트 마이닝을 활용하여 분석 가능한 정형데이터로 변환한 후 공기지연 리스크의 분류 및 도출 프로세스를 제안 및 적용하여 공기지연 리스크를 정량화하였다.

### 1.2 연구의 범위 및 방법

본 연구는 한국도로공사가 진행한 고속도로 공사 10개 공구의 공사일지 및 감독일지의 텍스트 정보를 대상으로 한다. 공사일지 및 감독일지는 당일 공사내역이 상세하게 기록되어 있기 때문에 공사 진행의 연속적 흐름을 파악할 수 있으며, 현장의 정보를 파악할

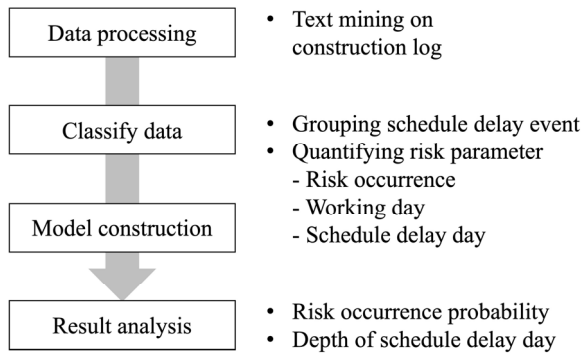


Fig. 1. Flow Chart Showing the Research Process

수 있는 유일한 정보이다. 공종분석 단계는 여러 공종 중 모든 건설 현장에서 공통적으로 적용 가능한 토공단계를 선정하였다. 다양한 공기지연 리스크 요인 중 공기산정 시 비작업일로써 공시기간에 적용되는 기후요인 중(Yoon, 2020), 토공단계에 많은 영향을 끼칠 것으로 예상되는 우천 리스크를 중심으로 연구 범위를 한정하였다.

Fig. 1에 전체 연구방법을 나타내었다. 텍스트 마이닝 기술을 이용하여 공사일지 중 당일공사내용을 WBS 기준으로 분류하여 토공 작업 내용을 추출하고, 우천을 중심으로 공기지연 이벤트가 나타난 공종을 분류하였다. 이를 통해 실제 작업일과 우천으로 인한 공기지연 일수를 분석함으로써 우천으로 인한 공기지연의 발생확률과 심도(공기지연 일)를 정량화하였다.

## 2. 공기지연 리스크 도출

### 2.1 공사일지 수집 및 데이터 전처리

공사일지는 공사일 동안의 당일 공사 내역과 감독일지 등 현장의 정보를 담고 있으나 당일공사내역의 공종 구분이 되어있지 않으며 구간별 분류가 수행되어 있지 않은 단순한 텍스트 문치의 비정형데

이터로써 분석을 위해 정형데이터로 변환되어야 한다. 특히 각 공정은 일정한 공종의 집합으로 이루어져 있으나, 공사일지가 WBS (Work Breakdown Structure)를 기준으로 구분되어 있지 않아 연속된 공종의 파악이 불가능하다. 따라서 본 연구에서는 PLM (Pretrained Language model)인 BERT (Bidirectional Encoder Representations from Transformers) 알고리즘을 활용하여 공사 내역을 학습시키고 이를 통해 WBS를 구분하였다.

#### 2.1.1 BERT

BERT는 구글이 공개한 자연어 모델로써 위키피디아 25억 단어와 BooksCorpus 8억 단어와 같은 레이블이 없는 텍스트 데이터로 사전 훈련된 언어 모델이다(Vaswani et al., 2017). Fig. 2와 같이 사전학습(Pre-training)과 파인튜닝(Fine-tuning) 과정을 통해 자연어처리에 우수한 성능을 보이며(Devlin et al., 2019), 양방향 학습구조로 문장의 문맥을 고려하고 적은 양의 텍스트 데이터만으로도 우수한 성능을 보인다(Lee et al., 2020). 사전학습은 모델의 성능을 높이기 위해 레이블이 없는 대량의 텍스트 데이터를 사전에 학습하는 것으로 문장의 일부분을 마스킹(Masking)한 후 해당 부분을 예측하는 Masked Language Model과 문장별 관계를 학습하는 Next Sentence Prediction 방식이 적용되었다. 파인튜닝은 사전학습된 BERT 모델을 새로운 모델에 적용시키기 위하여 하이퍼파라미터를 조정하고 새로운 데이터를 학습하는 과정으로 Epoch, Batch size, Learning rate, Maximum length, Optimizer 등의 파라미터를 튜닝하는 것이 최적화된 모델을 찾기 위한 대표적인 방법이다(Lee et al., 2020).

본 연구에서는 BERT 모델을 한국어 텍스트에 적합하도록 개발한 SK T-Brain의 Ko-BERT 모델을 사용하였다. 당일작업내용을 기반으로 WBS를 분류하기 위하여 공사일지에서 추출한 5,516개 작업내용 중 토공 208개의 작업내용을 WBS를 매칭하여 학습하였다. Table 1에 반복적인 하이퍼파라미터 튜닝을 통해 결정된 파라미

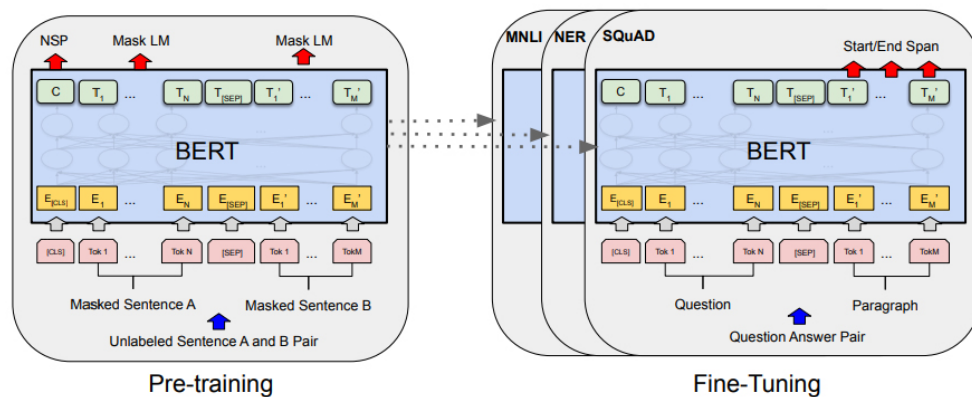


Fig. 2. Transformer Procedures for the BERT Algorithm

터 값을 나타내었으며, 작업단위1 정확도 80 %, 작업단위2 정확도 71 %의 분류 모델을 생성하였다. 최종 모델은 한정된 토공단계 안에서 WBS 분류 작업을 수행하므로 본 연구에 적용 가능한 정확도를 확보한 것으로 판단된다.

2.1.2 한국도로공사 WBS

본 연구에서는 국내 고속도로공사의 공기지연 리스크 관리를 위한 리스크 정량화를 목표로 하며, 분석 대상의 공사 구간이 한국도로공사에서 수행한 고속도로공사임을 고려하여 한국도로공사의 WBS 체계를 적용하였다. 공사현장, 도로시설, 공중, 시설물, 방향공간, 확장공간, 작업관리의 7레벨 계층 구조로 구성된다. 공사현장, 도로시설, 공중은 동일한 건설프로젝트에서 고정되며 시설물과 방향공간은 고속도로 프로젝트 종류에 상관없이 고정되는 값이며, 확장공간은 지역위치 및 구조물 이름에 따라 변하는 속성을 가지고 있다. 따라서 본 연구에서는 Table 2와 같이 공중분류체계가 변하지 않고 고정되는 공중, 시설물, 작업관리단위를 Ko-BERT 모델의 학습 범위로 선택하였다. 방향공간의 경우 속성이 고정될 수 있으나 건설공사의 측면에서 시설물 단위에 포함될 수 있을 것이라 판단하여 학습 데이터에서 제외하였다. 또한 작업관리단위는 하나의 계층으로 구분되어 있으나, 본 연구에서는 공중의 세분화를 통한 리스크

관리 체계의 고도화를 위하여 2개로 구분하였다.

2.2 공기지연 리스크 분류 및 도출

공중별 작업내용 분류가 완료된 공정정보에서 공기지연 이벤트가 발생한 조건을 찾기 위한 기준을 Fig. 3에 나타내었다. 공사일지에서 공기지연 이벤트가 발생한 것으로 판단할 수 있는 조건은 2가지이다. 첫 번째는 평일에 작업을 수행하지 않은 경우, 두 번째는 휴일에 작업을 수행한 경우이다. 이중 두 번째 조건은 예정된 공정을 기한 내에 수행하지 못하여 휴일에 대체근무를 진행한 것으로 보아 최종 공기지연 리스크 이벤트로 선정하지 않고 실제 작업일로 포함시켰다. 공기지연 이벤트가 감지된 당일의 4일 전후의 감독일지 내용을 기반으로 리스크 요인을 도출하였다. 또한 공기지연 이벤트 발생과 공중의 종류를 구분하기 위하여 동일한 공중이 7일 연속 이어지지 않은 경우에는 해당 공중이 완료되었다고 판단하였다. 실제 작업일은 공기지연 이벤트가 발생하고 비작업일이 7일 이상 발생하지 않은 연속된 공중의 일자를 계산하며, 공기지연 발생횟수는 연속으로 발생한 이벤트를 1회로 간주하였다.

2.3 공기지연 리스크 정량화 모델

2.3.1 공기지연 리스크 발생확률

우천 공기지연 리스크의 발생확률을 가정하기 위하여 Eq. (1)과 같이 리스크 발생횟수를 실제 작업 기간으로 나누어 특정 기간

Table 1. Predictive Performance by Hyperparameter of Ko-BERT

Parameter	Value
Max length	64
Batch size	64
Epochs	15
Learning rate	5E-05
Optimizer	AdamW
Accuracy for activity 1	0.80
Accuracy for activity 2	0.71

Table 2. Work Breakdown Structure (WBS) of Earthwork for Fine-Tuning of Ko-BERT

Construction kind	WBS		
	Facility	Activity 1	Activity 2
Earthwork	Interchange/ Junction (IC/JC)	Soil cut	Soil, Reaping, Blast, Soil cut, etc.
		Soil fill	Subsoil, Subgrade, Soil fill, etc.
	Main line  Side/ Access road	Earth work, etc.	Slope protection Soft ground improvement Earth retaining wall Concrete retaining wall Retaining wall, etc. Earth work, etc.

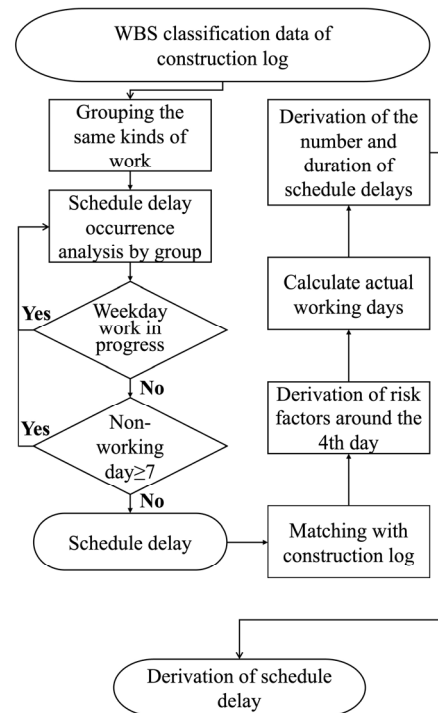


Fig. 3. Risk Classifying Algorithm

동안의 발생확률을 도출하였다. 이때 도출된 개별 리스크 값은 분류기준에 따라 합산하였다.

$$P = \frac{RO_1 + RO_2 + \dots + RO_n}{WD_1 + WD_2 + \dots + WD_n} \quad (1)$$

여기서 RO (Risk Occurrence)은 연속된 공중 작업 내에서 공기지연 이벤트가 발생한 횟수, WD (Working day)는 같은 공중이 연속적으로 수행된 실제 작업기간이며, n은 분류기준에 따른 그룹의 개수이다.

### 2.3.2 공기지연 리스크 심도

우천으로 인한 공기지연 리스크의 심도는 Kolmogorov-Smirnov Test (K-S 검정)를 활용하여 확률분포의 형태를 결정하였다. K-S 검정은 가정한 확률밀도함수와 경험적 확률밀도함수를 비교하여 분포적합도를 판단하는 검정 방법이다. 표본 자료들의 적합 정도를 크기에 관계없이 전체적으로 동일하게 고려하고 전체적인 양상에 따른 적합도를 비교적 정확하게 검정하기에 표본의 개수가 적을 것으로 판단되는 본 연구에 적용하였다(Shin et al., 2010). K-S 검정은 P-value가 0.05 이상이 되면 두 분포가 적합하다고 가설한 귀무가설을 기각할 수 없기에 통계적으로 유의하다고 판단할 수 있으며, 정규분포, t-분포, 균일분포, 카이제곱분포의 4가지 분포를 검정하였다.

## 3. 우천 공기지연 리스크 정량화

### 3.1 우천 공기지연 리스크 도출

10개 공구 중 공사일지 작성 양식에 따라 구간 및 구조물 분류가 불가능한 2개 공구를 제외하고 8개 공구를 분석하였다. 총 작업일은 8,086일이며, 세부 구간은 116개, 분석문장은 9,633개이다. 4단계 WBS(공중-시설물-작업단위1-작업단위2)를 학습한 BERT 모델을 활용하여 공사일지에서 도출된 문장을 WBS 단위로 변환하고, Fig. 3의 절차에 따라 우천으로 인한 공기지연 이벤트가 발생한 지점에서 리스크 요인을 도출하였다. 이러한 공사일지 데이터 전처리 및 리스크 도출 과정을 거쳐 75개의 우천으로 인한 공기지연 이벤트가 도출되었으며, 우천으로 인한 공기지연의 도출 근거를 Table 3에 나타내었다. 도출 근거는 Fig. 3에 따라 도출된 공기지연 이벤트의 해당 날짜에 해당되는 감독일지 내용이다. 감독일지 내용에 따라 정비/점검(improvement/inspection), 작업취소(job cancellataion), 대비(preparing)의 3가지 타입으로 구분하였으며, 정비/점검의 경우 우천으로 인한 시설물의 점검 및 현장의 확인 등의 작업이 진행되었다.

공사일지 분석을 통해 도출된 우천 공기지연 리스크의 특성을

살펴보기 위하여 Table 4와 Fig. 4에 우천 공기지연 발생 월에 따른 리스크 도출 결과를 나타내었다. 1월과 12월은 우천으로 인한 공기지연이 발생하지 않았다. 장마철인 6월부터 8월까지의 발생한 우천 공기지연은 총 13회로 전체의 17.3 %, 이로 인한 기간은 총 21일로 전체의 20.2 %를 차지하였다. 2월-5월, 9월~11

Table 3. Supervisor Comments for Determining Schedule Delay by Rain

Type	Comment
Improvement/ Inspection	Improvement waterway and work waiting
	Work waiting
	Access road improvement
	Workplace inspection / Field inspection / Field safety check
	Field cleaning
Job cancellation	Work waiting / canceling
Preparing	Pre-inspection of the vulnerability
	Field inspection
	Disaster training

Table 4. Data of Schedule Delay Risk by Month

Month	Risk occurrence (RO, n)	Delay day (DD, day)	Working day (WD, day)	Risk probability (P, %)
2	9	14	220	4.1
3	12	12	180	6.7
4	7	9	272	2.6
5	4	4	404	1
6	5	7	461	1.1
7	4	10	571	0.7
8	4	4	194	2.1
9	12	14	637	1.9
10	10	17	794	1.3
11	8	13	271	3
Total	75	104	-	-

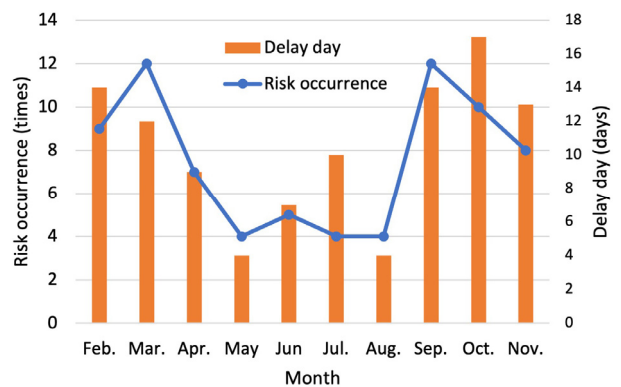


Fig. 4. Number of Risks and Delay Days by Month

월의 우천 공기지연은 장마철 대비 높은 빈도와 지연기간을 차지하였다. 이는 6월부터 8월까지 장마철에 대비하여 공정 진행 상 충분한 여유일 또는 비작업일을 확보하였기 때문에 공정관리에 유리하였고, 이외에 기간에서는 우천에 대한 충분한 대응 부족에 따라 우천 공기지연이 다수 발생한 것으로 판단된다.

Eq. (1)에 따른 리스크 발생확률은 3월이 6.7 %로 가장 높았으며, 5월과 7월이 각 1 %, 0.7 %로 가장 낮은 확률을 나타내었다. 장마기간인 6, 7, 8월의 경우 7월이 가장 낮은 0.7 % 확률로 리스크가 발생되었고, 6월과 8월은 각각 1.1 %, 2.1 % 확률을 나타내었다.

Table 5에 공종별 우천으로 인한 공기지연 리스크 발생 특성 나타내었다. 우천으로 인해 공기지연이 발생한 공종은 39개 중 8개 공종이었으며, 작업단위2를 기준으로 흙깎기기타(soil cut) 3개, 흙쌓기기타(soil fill, etc.) 1개, 노상(subgrade), 노체(subsoil), 토공기타(earth work, etc.)가 각각 1개씩으로 나타났다. 흙깎기기타 공종의 우천 공기지연 발생횟수가 타 공종대비 높게 나타났으며, Ko-BERT 모델을 활용한 WBS 분류 시 세부 작업내용의 분류가 모호하여 기타 공종으로 분류된 경우가 많았기 때문인 것으로 판단된다. 시설물 기준으로는 IC/JC, 본선(main line), 지선/부체도로

(side/access road) 순으로 우천으로 인한 공기지연 이벤트가 많이 나타났다. IC/JC의 경우 Ramp 공사를 위한 긴 작업기간을 가지기에 상대적으로 우천 공기지연 발생횟수가 높게 나온 것으로 판단된다.

Eq. (1)에 따라 공종별 우천 공기지연 리스크의 발생확률을 계산하였으며, 발파(blast) 공종의 경우 철저한 사전계획에 따라 공정을 진행하므로 갑작스런 우천 공기지연 요인에 영향을 덜 받아 가장 낮은 발생확률을 가진 것으로 판단된다.

### 3.2 우천 공기지연 리스크 심도

K-S 검정에 따른 우천 리스크의 공기지연 확률분포(리스크 심도)를 Table 6에 나타내었다. P-value가 0.05 이상으로 유효한 확률분포를 가지고 있는 공종은 IC/JC의 흙쌓기-노상(이하, 노상), 본선의 흙깎기-발파(이하, 발파), 지선/부체도로의 흙쌓기-흙깎기기타(이하, 흙깎기기타), 흙쌓기-흙쌓기기타(이하, 흙쌓기기타)의 4개 공종으로 나타났다. 표본에서의 모든 공종의 최소 지연기간은 1일이다. IC/JC의 흙깎기-흙깎기기타 공종은 가장 많은 표본 크기임에도 정규분포의 형태를 따르지 못하는 것으로 나타났으며, Ko-BERT 분류기의 성능 향상 및 더 많은 표본의 확보가 지속적으로 필요할 것으로 판단된다.

Table 5. Data of Schedule Delay Risk by Work Breakdown Structure (WBS) in Earthwork

Facility	WBS		Risk occurrence (RO, n)	Delay day (DD, day)	Working day (WD, day)	Risk probability (P, %)
	Activity 1	Activity 2				
IC/JC	Soil cut	Soil cut, etc.	30	42	451	6.7
	Soil fill	Subgrade	8	11	56	14.3
		Subsoil	3	3	52	5.8
Main line	Earth work, etc.	Earth work, etc.	7	8	105	6.7
	Soil cut	Blast	6	9	283	2.1
		Soil cut, etc.	13	20	356	3.7
Side /Access road	Soil cut	Soil cut, etc.	5	6	81	6.2
	Soil fill	Soil fill, etc.	3	5	10	30
Total			75	104	-	-

Table 6. Probability Distribution of Schedule Delay Risk

Work breakdown structure (WBS)			Risk probability (P, %)	Sample size	Distribution	P-value	Mean	Std.	Maximum
IC/JC	Soil cut	Soil cut, etc.	6.7	30	norm	3.8E-4	1.4	0.72	4
		Subgrade	14.3	8	t	0.106	1.38	0.52	2
	Soil fill	Subsoil	5.8	3	uniform	-	1	0	1
Main line	Earth work, etc.	Earth work, etc.	6.7	7	norm	0.030	1.14	0.38	2
	Soil cut	Blast	2.1	6	norm	0.400	1.5	0.55	2
		Soil cut, etc.	3.7	13	norm	0.034	1.54	0.78	3
Side/Access road	Soil cut	Soil cut, etc.	6.2	5	norm	0.124	1.2	0.45	2
	Soil fill	Soil fill, etc.	30	3	chi-square	0.762	1.67	0.58	2

### 3.3 우천 공기지연 기간의 누적확률분포

제시된 리스크 발생확률 및 심도 데이터는 건설 공정계획에 적용하여 개별 공종의 리스크 위험도를 계산할 수 있다. 다만, 건설 프로젝트의 규모, 기간, 공정계획 방법 등에 따라 적용 방법이 상이하다. 또한 본 연구의 범위가 공사일지를 기반으로 한 리스크의 분류 및 도출 프로세스를 제안하는 것이기 때문에, 도출된 리스크 심도 데이터를 활용하여 우천으로 인한 공기지연 리스크가 발생하였을 때의 심도의 누적확률분포를 Fig. 5에 나타내어 공종별 위험도를 비교하였다. 누적확률 95%를 기준으로 노상 공종은 1.72일, 발파 공종은 2.41일, 흙막이기타 공종 1.93일, 흙쌓기타 공종 2.48일로 나타났다. 흙쌓기타 공종의 경우 표본의 개수가 적고 우천 공기지연 기간 평균일이 1.67일로써 가장 크게 나타났기 때문에 낮은 누적확률에서도 높은 우천 공기지연이 발생하는 것이 확인되었다. 다만, 비교 대상 공종에서 가장 적은 표본을 가지고 있기 때문에 우천으로 인한 공기지연 심도의 직접적인 비교는 불가능하다.

발파 공종은 다른 3개 공종보다 누적확률 증가에 따른 우천으로

인한 공기지연 기간 증가 폭이 다른 2개 공종보다 크게 나타나 흙쌓기타 공종을 제외한 최대 지연 기간인 2.41일 나타났다. 노상 및 흙막이기타 공종은 표본의 최대치인 2일 이내에서 누적확률 95%의 확률을 가지고 있는 반면에 발파 공종은 표본의 최대치인 2일을 초과하는 구간에서 누적확률 95%를 보이고 있어 타 공종 대비 발생확률은 낮지만, 우천 공기지연 리스크 발생 시 위험도는 클 것으로 예상된다.

### 4. 결론

본 연구에서는 형식적으로 저장되어 데이터로써의 활용성이 낮은 공사일지 및 감독일지 등의 비정형데이터를 텍스트 마이닝 기술을 활용하여 분석가능한 문서(정형데이터)로 변환하고, 이를 분석하여 우천에 따른 공기지연 리스크를 정량화하였다. 공사일지에서 Ko-BERT 알고리즘을 활용한 공종 정보 분류 모델 및 공기지연 리스크를 도출하기 위한 알고리즘과 리스크 정량화 모델(공기지연 발생확률 및 심도)을 제시하였다. 고속도로공사 8개 공구의

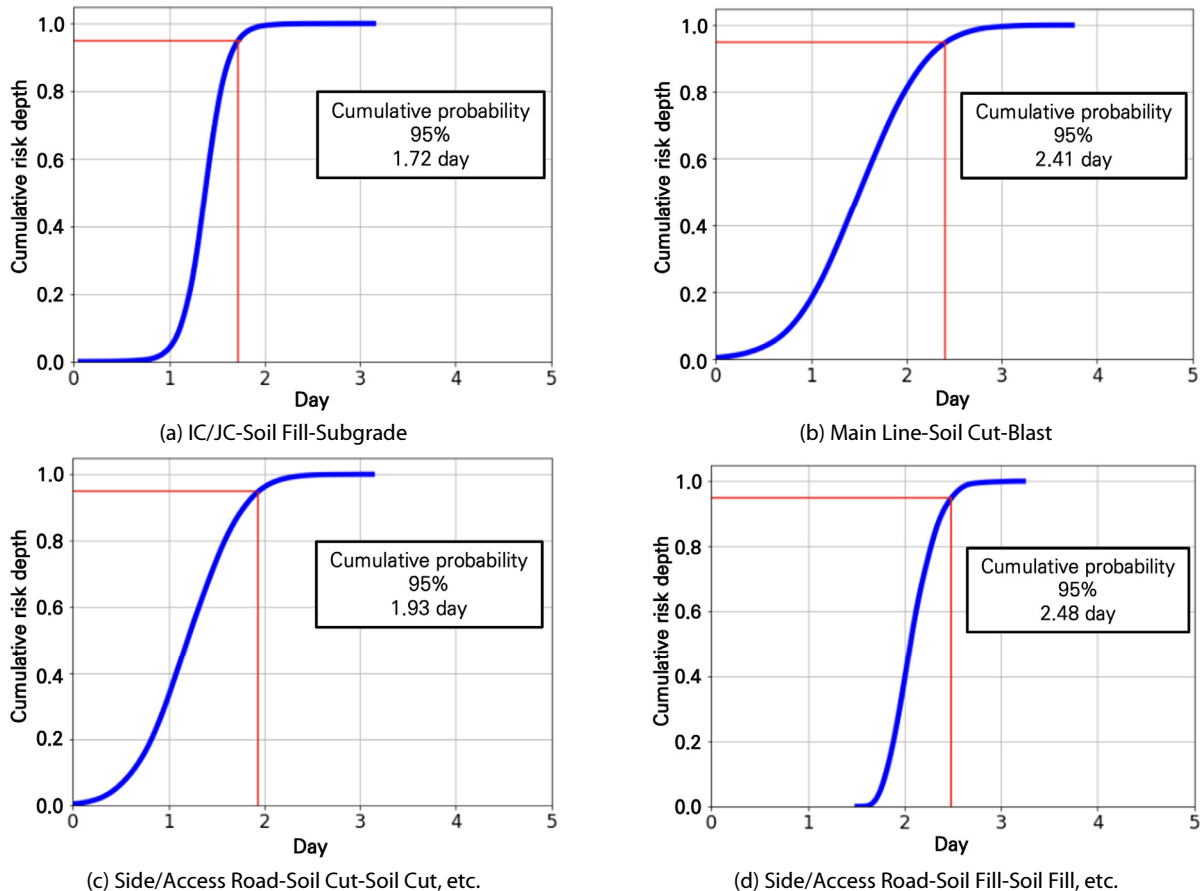


Fig. 5. Cumulative Distribution of Schedule Delay Days for Different Construction Activities

토공 작업을 대상으로 제시한 모델을 적용하였다.

분석 결과, 장마철에 따라 우천 공기지연이 많이 발생할 것으로 예상되었던 6월부터 8월까지의 초기에 충분한 여유일 또는 비작업 일 등 우천에 대한 대응이 잘되어 우천 공기지연 발생 및 기간이 적게 나타나는 것이 확인되었다. 즉, 실제 시공단계에서는 초기에 계획된 공정 계획과 다른 우발적인 리스크에 대한 대비가 필요하다. K-S 검정을 통해 유의미한 확률분포를 가진 4개 공종에 대한 우천 공기지연 리스크 심도를 누적확률 95 %를 기준으로 비교하였다. 표본의 개수가 작은 흙쌓기타 공종을 제외하면, 리스크 발생 확률이 가장 낮은 발파 공종의 위험도가 2.41일로써 가장 높게 나타났다. 따라서 다른 공종보다 철저한 대비 및 대응방안의 마련이 필요하다.

본 연구에서 제시한 공종 정보 분류 모델, 공기지연 리스크 도출 및 정량화 모델은 건설 프로젝트에 내재된 공기지연 정보를 새롭게 발굴함으로써 체계적인 공정계획 예측을 위한 기초자료를 확보할 수 있다. 추후 지속적인 데이터 확보를 통해 공기지연 정량화 DB가 구축된다면 공정계획 시 체계적인 디지털트윈 시뮬레이션 등에 활용될 수 있을 것이다.

## 감사의 글

이 연구는 국토교통부/국토교통과학기술진흥원이 시행하고 한 국토로공사가 총괄하는 “스마트건설기술개발 국가R&D사업(과제 번호 22SMIP-A158708-03)”의 지원으로 수행하였습니다.

## References

Al-Bahar, J. F. (1989). *Risk management in construction projects: a systemic analytical approach for contractors*, Ph.D. Dissertation, University of California Berkeley, Berkeley, California, USA.

Choi, J. W. (2015). *An analysis on regional differences of major delay factors in overseas architectural projects*, Master. Dissertation, Hanyang University, Seoul, Korea (in Korean).

Devlin, J., Chang, M. W., Lee, K. and Toutanova, K. (2019). “BERT: Pre-training of deep bidirectional transformers for language understanding.” *arXiv:1810.04805v2*, pp. 1-16.

Ha, S. G., Kim, T. H., Son, K. Y., Kim, J. M. and Son, S. H. (2018). “Quantification model development of human accidents on external construction site by applying probabilistic method.” *Journal of the Korea Institute of Building Construction*, Vol. 18, No. 6, pp. 611-619 (in Korean).

Kang, H. B. and Yi, J. S. (2018). “An analysis of public text data in construction disaster cases using Word2Vec-based data visualization.” *Autumn Annual Conference of AIK, 2018, Architectural Institute of Korea*, Vol. 38, No. 2, pp. 567-570 (in Korean).

Kang, L. S., Kim, C. H. and Kwak, J. M. (2002). “Analysis for the importance of risk factors through the project life cycle.” *Journal*

*of the Architectural Institute of Korea Structure & Construction*, Vol. 17, No. 8, pp. 103-110 (in Korean).

Kim, J. S. (2022). *Analysis of project delay using big data*, Master. Dissertation, Hanyang University, Seoul, Korea (in Korean).

Kim, J. S. and Kim, B. S. (2019). “Characteristics analysis of seasonal construction site fall accident using text mining.” *Korean Journal of Construction Engineering and Management*, Vol. 20, No. 3, pp. 113-121 (in Korean).

Lee, C. H., Lee, Y. J. and Lee, D. H. (2020). “A study of fine tuning pre-trained korean BERT for question answering performance development.” *Journal of Information Technology Services*, Vol. 19, No. 5, pp. 83-91.

Lee, H. C., Lee, J. H. and Go, S. S. (2009). “A study on the priority analysis of work delay factors in steal-frame work using FMEA.” *Construction Engineering and Management*, Vol. 10, No. 1, pp. 91-101 (in Korean).

Lee, H. K. (2018). *A risk assessment model for delay management protocol in the contract conditions of international construction projects*, Ph.D. Dissertation, Ajou University, Suwon, Korea (in Korean).

Lee, J. H. and Yi, J. S. (2017). “Predicting project’s uncertainty risk in the bidding process by integrating unstructured text data and structured numerical data using text mining.” *Applied Sciences*, Vol. 7, No. 11, pp. 1-15.

Marzouk, M. and Enaba, M. (2019). “Text analytics to analyze and monitor construction project contract and correspondence.” *Automation in Construction*, Vol. 98, pp. 265-274.

Park, G. S. (2012). *A study on the estimation of an appropriate construction duration of apartment*, Master. Dissertation, Seoul National University of Science and Technology, Seoul, Korea (in Korean).

Park, J. H., Choi, H. S., Cho, Y., Bang, K. S., Yun, S. H. and Paek, J. H. (2010). “A study on the development of probabilistic duration estimation module using monte carlo simulation.” *Journal of the Architectural Institute of Korea Structure & Construction*, Vol. 26, No. 5, pp. 101-108 (in Korean).

Park, K. C. and Kim, H. K. (2021). “Analysis of seasonal importance of construction hazards using text mining.” *KSCE Journal of Civil and Environmental Engineering Research*, KSCE, Vol. 41, No. 3, pp. 305-316 (in Korean).

Project Management Institute (PMI) (2021). *A guide to the project management body of knowledge (PMBOK Guide)*, Project Management Institute, Newton Square, PA.

Shin, H. J., Sung, K. M. and Heo, J. H. (2010). “Derivation of modified anderson-darling test statistics and power test for the gumbel distribution.” *Journal of Korea Water Resources Association*, KWRA, Vol. 43, No. 9, pp. 813-822 (in Korean).

Son, B. Y. and Lee, E. B. (2019). “Using text mining to estimate schedule delay risk of 13 oshore oil and gas EPC case studies during the bidding process.” *Energies*, Vol. 12, No. 10, pp. 1-25.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017). “Attention is all you need.” *arXiv:1706.03762v5*, pp. 1-15.



- Wang, G., Liu, M., Cao, D. and Tan, D. (2020). "Identifying high-frequency-low-severity construction safety risks: An empirical study based on official supervision reports in Shanghai." *Engineering, Construction and Architectural Management*, Vol. 29, No. 2, pp. 940-960.
- Yang, H. S. (2020). *Comparison of recognition on the risks affecting schedule delays and cost overruns in overseas civil construction projects*, Master. Dissertation, Hanyang University, Seoul, Korea (in Korean).
- Yoon, J. K. (2017). *A study to derive delay factors based on international standard contract at building construction in the south-east Asia*, Master. Dissertation, University of Seoul, Seoul, Korea (in Korean).
- Yoon, J. S. (2020). *Development of method for quantifying and applying non-working day to planning and scheduling to calculate normal project duration*, Ph.D. Dissertation, Ajou University, Suwon, Korea (in Korean).
- Yoon, Y. S., Suh, S. W., Park, M. S. and Jang, M. H. (2008). "Construction process based schedule risk management system." *Construction Engineering and Management*, Vol. 9, No. 4, pp. 101-110 (in Korean).
- Yu, Y. J., Son, K. Y., Kim, T. H. and Kim, J. M. (2017). "A basic study on cost prediction model based on repair and replacement causes by using probabilistic forecasting concept - focused on educational facilities." *Architecture & City in JEJU, Architectural Institute of Korea*, Vol. 37, No. 1, pp. 887-890 (in Korean).