Week 1
Data Mining Overview

Seokho Chi
Professor | Ph.D.
SNU Construction Innovation Lab

서울대학교 건설혁신연구실
Construction Innovation Lab SNU

Course Information

- Title: 457.658 Construction IT and Automation
- Timetable
  - Monday 2-5pm @ 35-223
- Instructor: Prof. Seokho Chi
  - shchi@snu.ac.kr, 35-304
  - TA: Daeyoun Won, wdh91@snu.ac.kr, 35-219, 880-7370

Course Information

- *Yourself?*
- *Why are you taking? What do you want to learn?*

Course Objectives

- Understand the fundamentals of data mining and knowledge discovery in database
- Apply data management techniques for data classification, prediction, clustering, and mining association rules
- Demonstrate how knowledge discovery in database can be used to support construction management
- Recognize the design, analysis, and implementation issues for data management in civil engineering

Course Materials

- Required
  - Lecture slides and handouts
  - eTL: Update correct contact info

- References
  - Tan, P., Steinback, M., and Kumar, V. (2005) Introduction to Data Mining, Addison-Wesley

Note

- Group Assignment
  - Teamwork is important.
  - Active participation is required.

- Cheating and Plagiarism
  - 0% for the given assessment item without any excuse
  - Penalty by SNU's regulations

Assessment

Item	Weight	Due
Attendance	10%	
Group Assignment		
Interim Report	15%	11/2
Final Report	20%	12/14
Final Presentation	5%	12/14
Individual Assignment	20%	6 times
Final Exam	30%	12/7
TOTAL	100%	

Course Schedule (1)

Week	Date	Contents
1	9.7	Course Introduction Data Mining Overview
2	9.14	Data Types Data Pre-Processing Data Exploration and Visualization
3	9.21	Classification
4	9.28	Classification
5	10.5	Prediction Deep Learning
6	10.12	Computer Lab (1)

Course Schedule (2)

Week	Date	Contents
7	10.19	Natural Language Processing
8	10.26	Computer Lab (2)
9	11.2	**Interim Group Presentation**
10	11.9	Keywords and Network Analysis Computer Lab (3)
11	11.16	Cluster Analysis
12	11.23	Mining Association Rules
13	11.30	Mining Complex Data Types Trends and Construction Applications
14	12.7	**Final Exam**
15	12.14	**Final Group Presentation**

Group Project Brief

- For this project, each group will mine a database to analyze/solve a construction engineering problem. Each group must identify a data set for this project.
- Examples include: productivity, safety performance, pavement management, environmental remediation, project disputes, soil characterization, structural monitoring, schedule control, property appraisals, quality control, among others.

- On Phase I, each team must submit a project proposal. The proposal must describe the problem that will be investigated, justify the need to conduct a data mining study to analyze/solve this problem, provide a short background review on related topics, specify the specific project objectives and scope, identify the target data set, and describe the proposed data mining approaches.
- Each team should use **at least four** different algorithms/methods.

Group Project Brief

- On the Final Phase, each team must submit a project report, including the results, discussion, conclusions, and recommendations.

- Each group must meet **at least two** times with me until the end of the course to discuss about the project proposals, progress, and results → Each group should meet **at least once** before the due data of each deliverable. Groups should contact me to schedule these meetings.

- The data mining should be conducted using WEKA, SAS or other software of your choice.

Group Project Brief

- **DELIVERABLES**
  - Deliverable 1 (11/2) – Project Proposal
    - Problem definition, background, literature review, need, objectives, scope, target data set, and proposed data mining approaches
  - Deliverable 2 (12/14) – Project Report
    - Summary of items included on deliverable 1, final results, discussion, conclusions, and recommendations.

- **PRESENTATIONS**
  - Phase 1 (11/2) – Deliverable 1
  - Final (12/14) – Deliverable 2

Introduction

What is Site Video Analysis?

CCTV installation legislation by Korean Government since 2016

(Source: Pradhananga and Teizer 2015)

Visual inspection *is one of the most effective tools to understand on-site performance*

Manual Inspection (Human Vision)

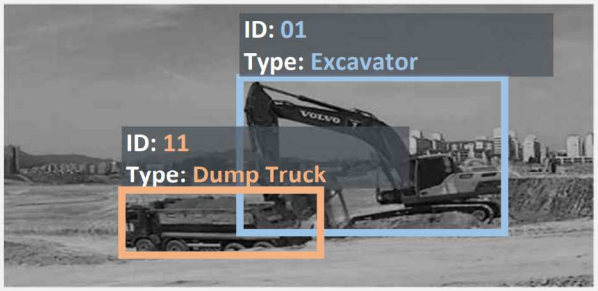
Automated Inspection (Computer Vision)

Application: Earthmoving Productivity Monitoring

Vision-based Productivity Monitoring of Earthmoving Operations through Operational Context Analysis

Site Videos

Equipment Tracking

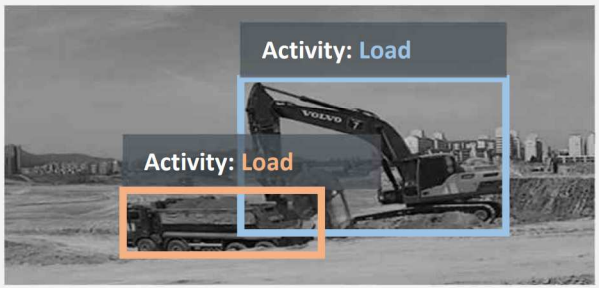


Equipment Types, Local IDs, Locations

Action Recognition

Action Types
(e.g., hauling, moving, stopping)

Activity Identification



Activity Types
(e.g., loading, traveling, idling)

Productivity Monitoring

Work/Idle Time, Cycle Time,
Soil Quantity, Match-factor, ..

Application: Earthmoving Productivity Monitoring

Video Stream Data

- Collected from three different earthmoving sites
  - Asan highway construction site, Chungcheongnam-do, South Korea
  - Namyangju new town development site, Gyeonggi-do, South Korea
  - Brisbane QUT campus building construction site, Queensland, Australia
- Total 164,968 image frames (257 min of operations)

Step 1: Equipment Tracking by Online Learning

Experimental results

- Precision: 94.3%, Recall Rate: 95.4%

Step 2: Action Recognition by Sequential Pattern Analysis

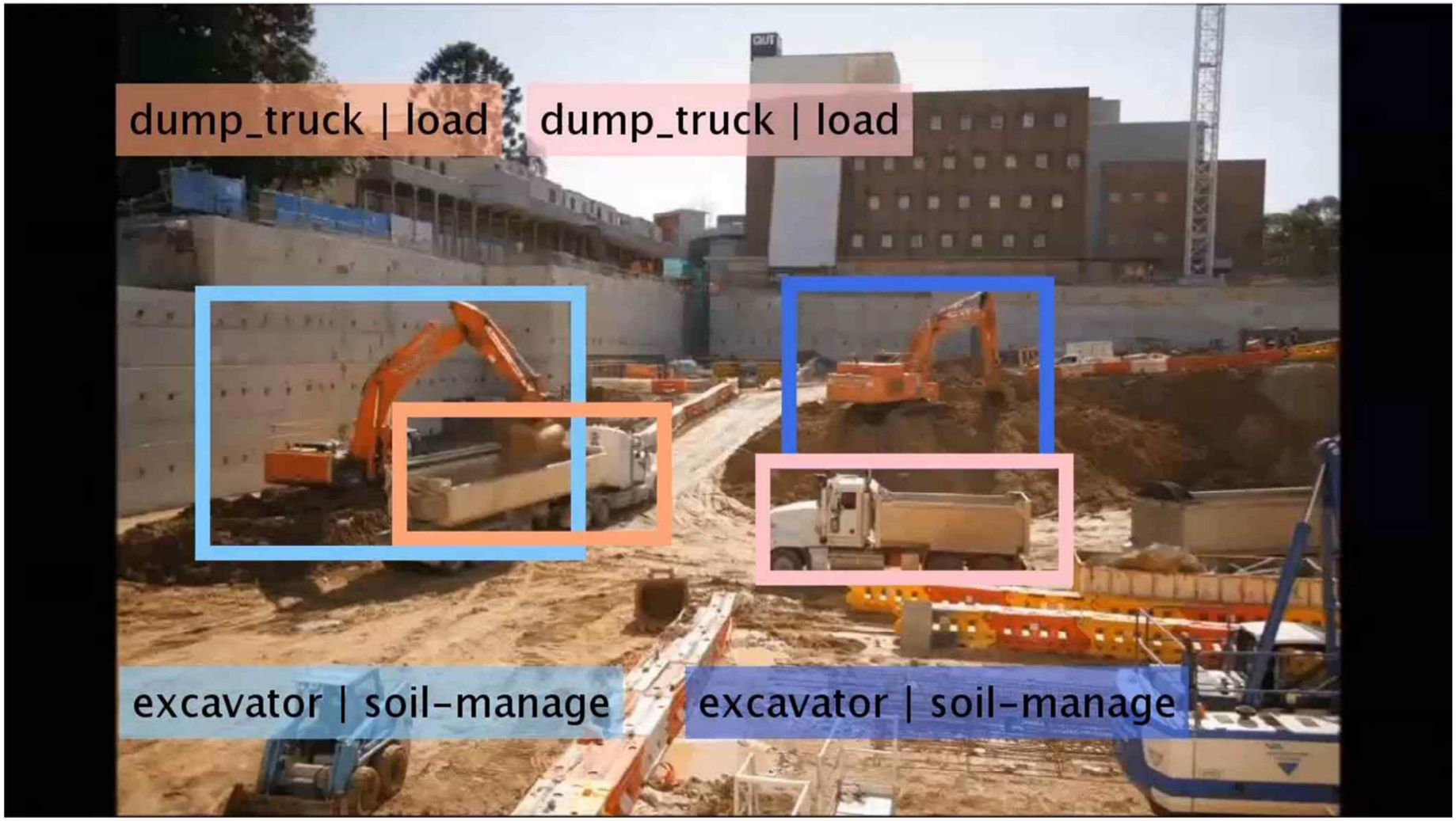
Experimental Results

- Precision: 92.4%, Recall rate: 92.0%

Step 3: Activity Recognition by Interaction Analysis

Experimental Results

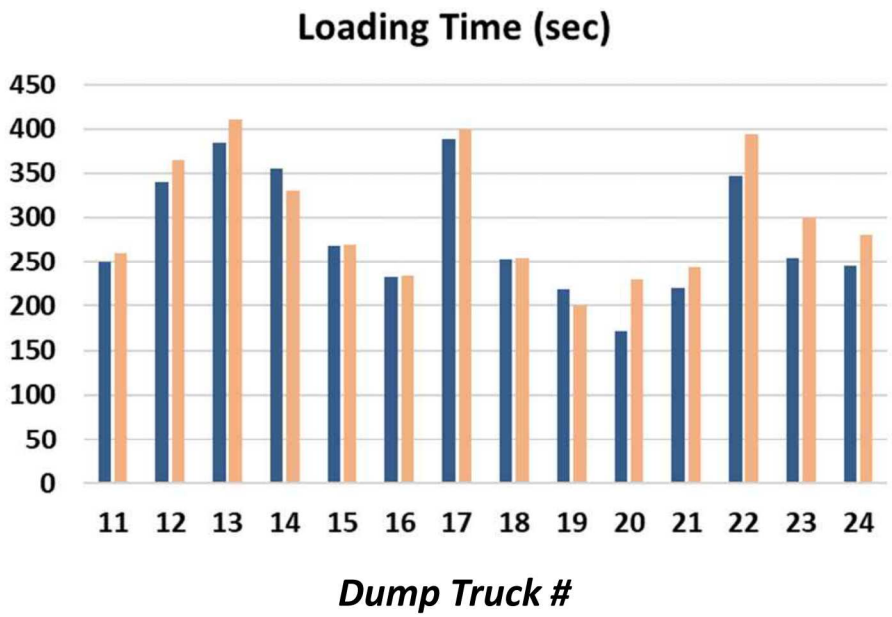
- Precision: 92.0%, Recall rate: 93.1%

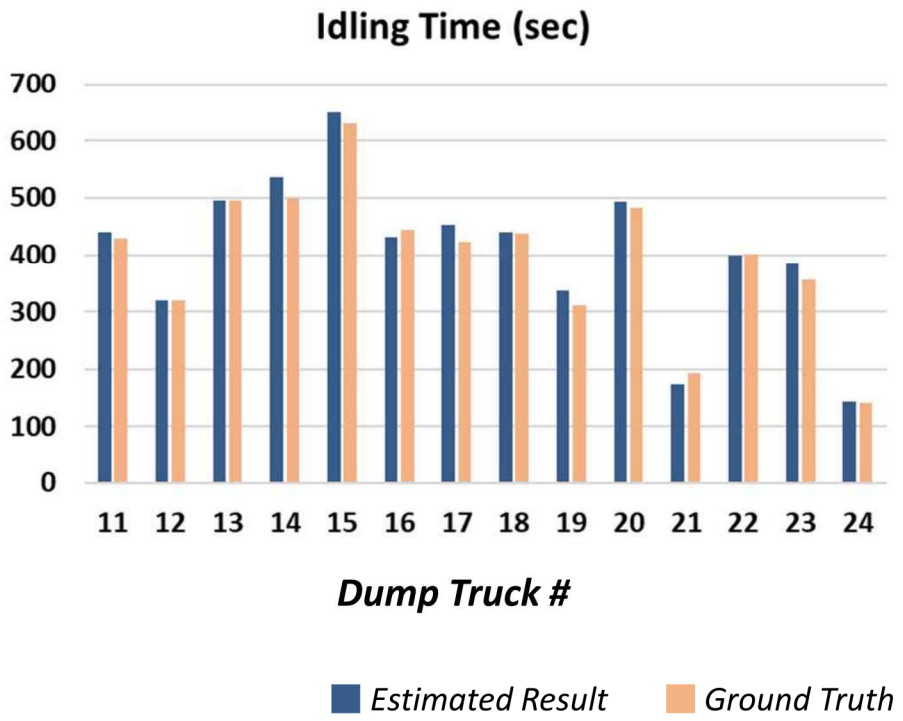


Step 4: Productivity Monitoring

Experimental Results: Dump Trucks

- Loading time: 8.1% error rate
- Idling time: 4.8% error rate

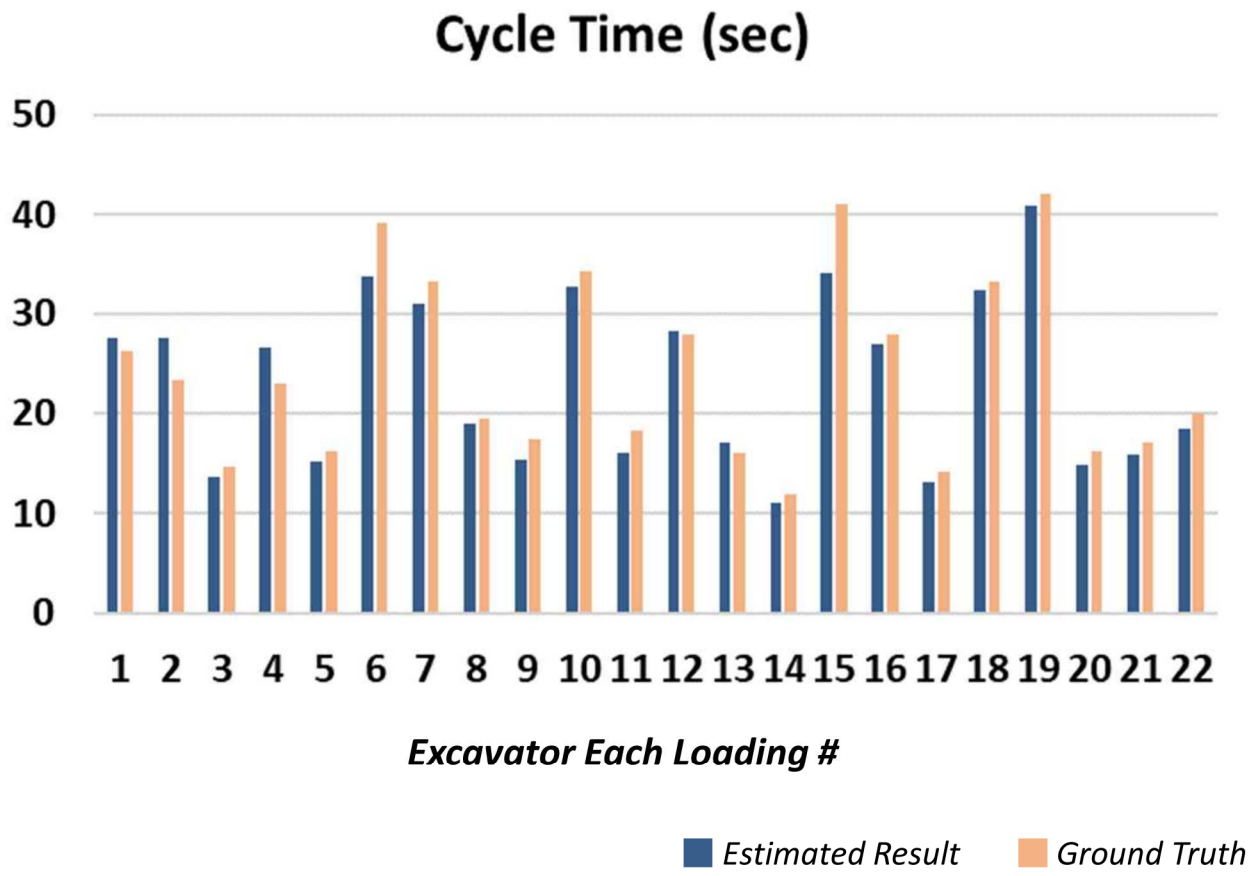




Step 4: Productivity Monitoring

Experimental Results: Excavators

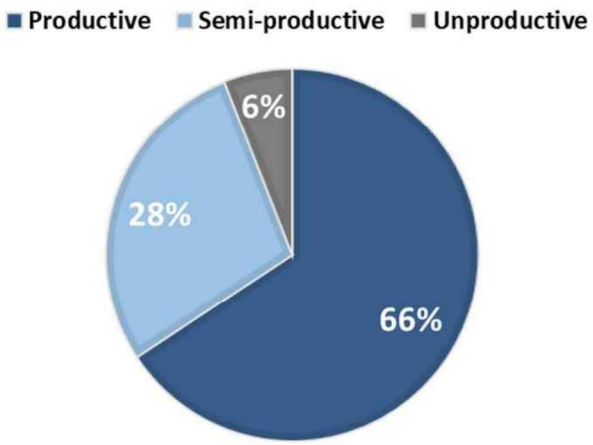
- Cycle time: 9.2% error rate



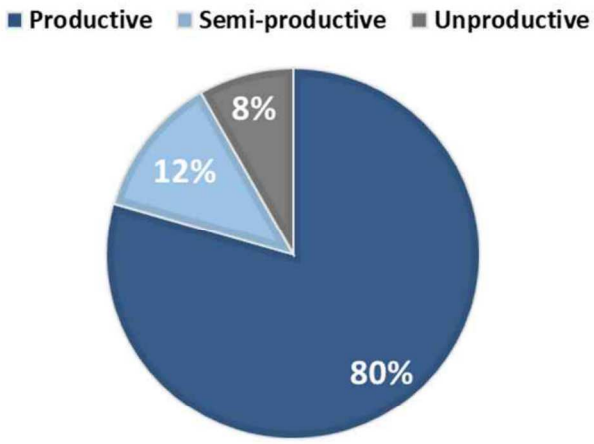
Step 4: Productivity Monitoring

Experimental Results

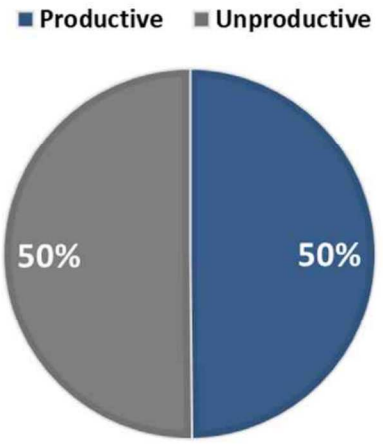
- Direct work rate analysis



Total: 66%



Excavator: 80%

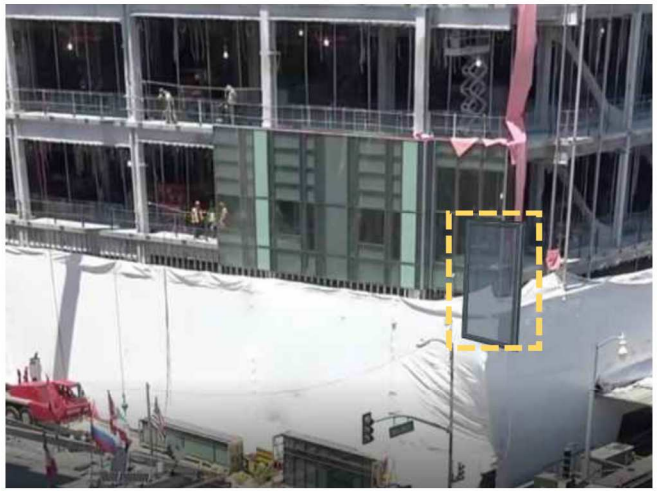


Dump Truck: 50%

On-going Research: DB-free Vision-based Monitoring

Experimental Results: Construction Materials

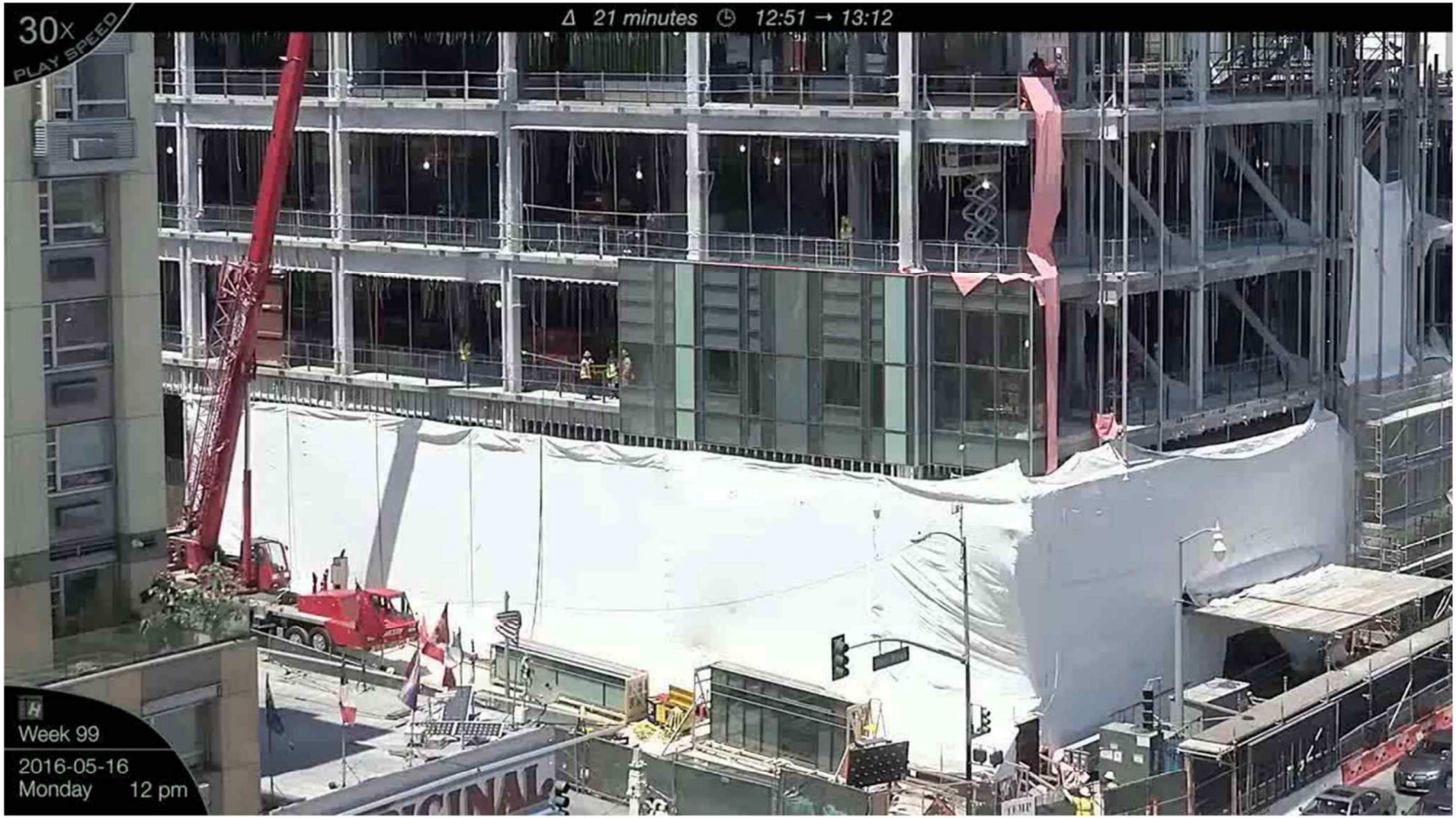
- **Class:** 마감공사 커튼월 패널
- **Train Data:** 1,270 Image Frames (73%, WEB-VR) **| Test Data:** 480 Image Frames (27%, Real)
- **Model:** Faster R-CNN Resnet (TensorFlow) & IoU Tracker



On-going Research: DB-free Vision-based Monitoring

Experimental Results: Construction Materials

- **Precision:** 93.2%, **Recall Rate:** 91.3%



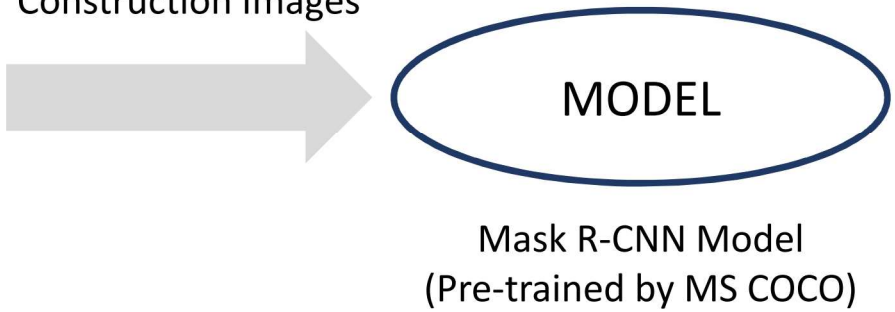
On-going Research: DB-free Vision-based Monitoring

Experimental Results: Construction Workers

- **Class:** Worker
- **Model:** Mask R-CNN
  - Pre-trained by MS-COCO dataset (81 classes)

MS COCO Dataset (Daily Images)

Without Using Construction Images



Mask R-CNN Model
(Pre-trained by MS COCO)

On-going Research: DB-free Vision-based Monitoring

Experimental Results: Construction Workers

- **Test dataset:** NAVER 1784 Construction Site

Introduction

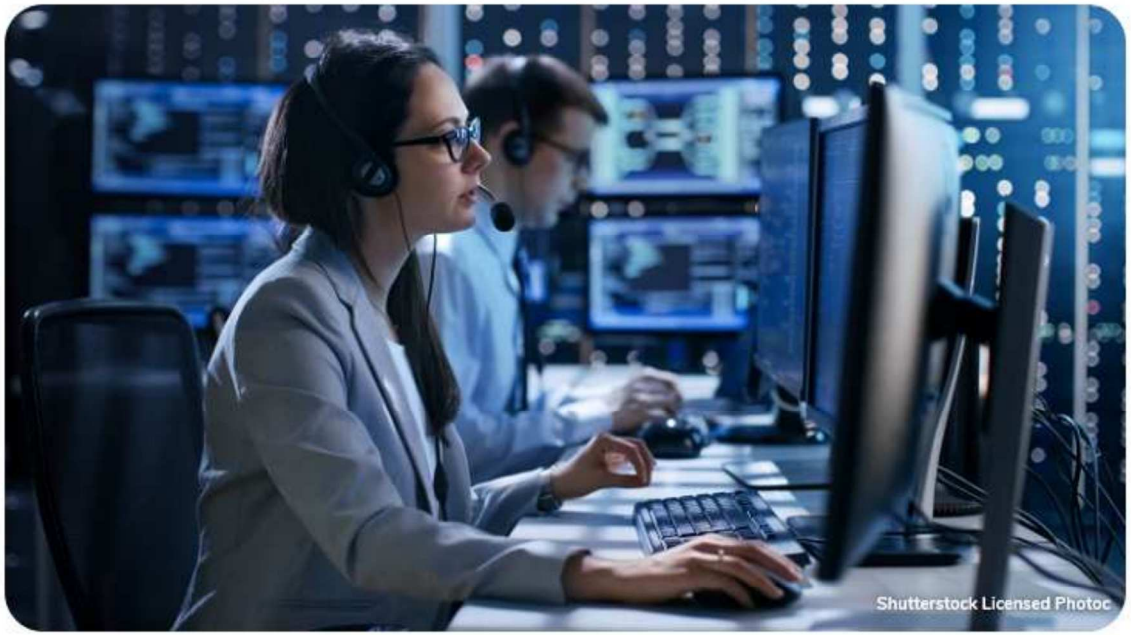
Police Are Using Big Data To Predict Future Crime Rates

Some police are starting to use big data to predict crime circumstances, and when and where illegal acts could happen. Here's what to know about it.

Andrej Kovacevic
November 7, 2018

44 Shares 5,370 Views

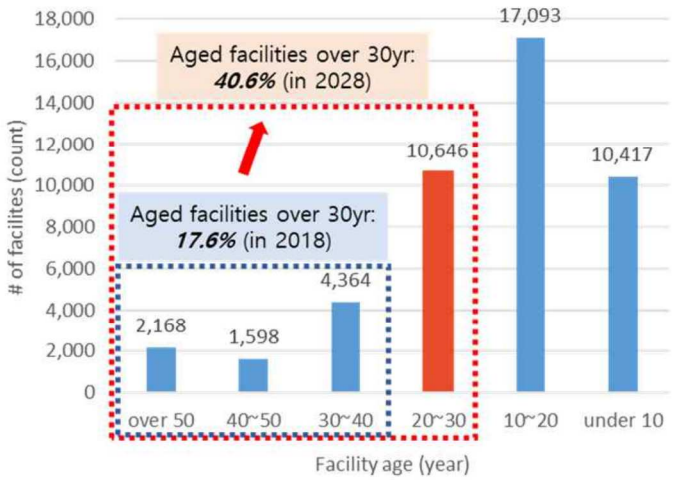
Share on Facebook Tweet on Twitter +



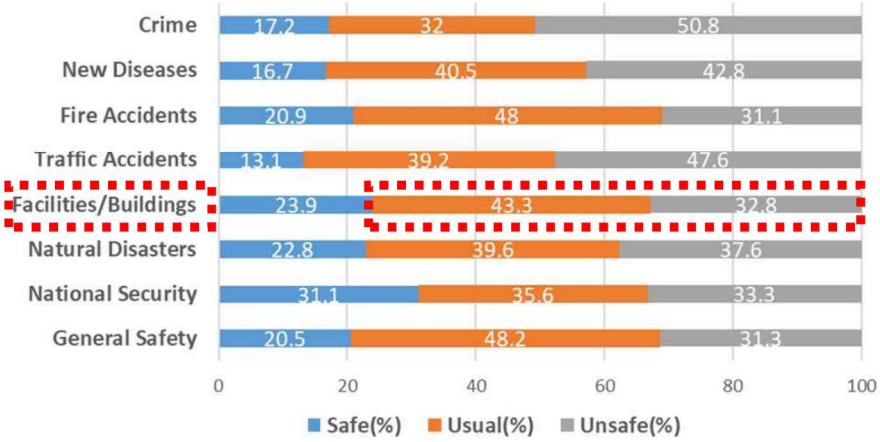
(Source: SmartData Collective, 2018)

Introduction

A Need for Preventive Facility Maintenance based on Big-data Analysis

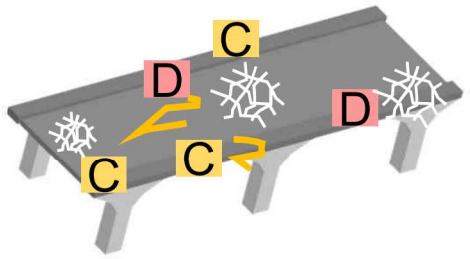


Increasing Number of Aged (≥30) Facilities
(KISTEC, 2018)



Increasing Public Concerns about Facility Safety and Serviceability
(Statistics Korea, 2019)

Big-data Analysis



Previous Condition Data

Manager Perspective

Predict Damage Locations, Types, Severities, …

Explore Unsafe or Low-service Factors for User

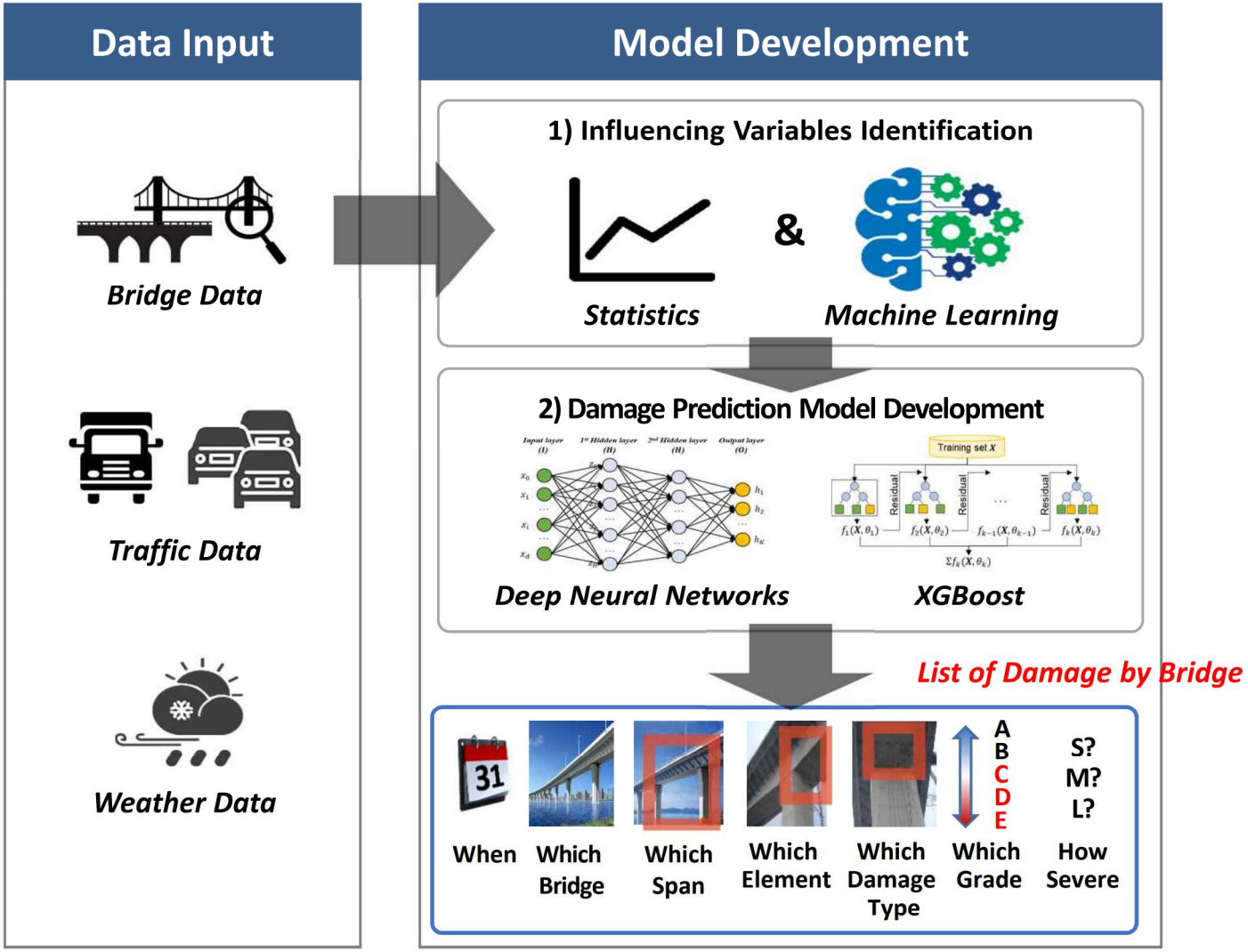
User Complaints/Opinions Data

User Perspective

Facilities include infrastructures(e.g., bridge, tunnel, river, harbor, dam), public buildings, and private buildings.

Bridge Damage Prediction

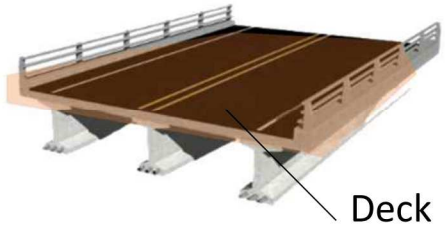
Bridge Damage Prediction Model Development



Bridge Damage Prediction

Data Characteristics

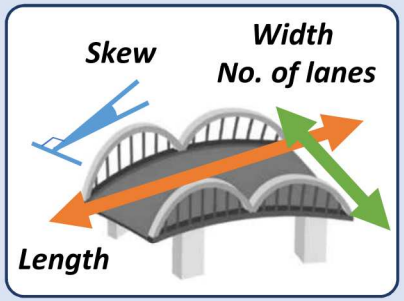
- Korean Bridge Management System Data
- Scope: Pre-stressed Concrete I-type (PSCI) bridges, Deck damage
- 2,388 bridges, 10,187 inspection records, 142,439 data



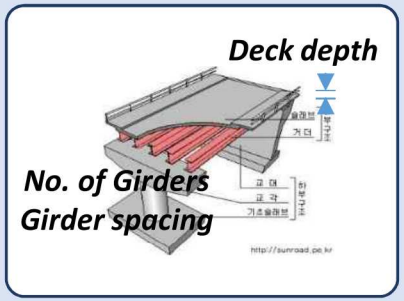
- 61 Variables (52 numerical, 9 categorical) → 59 dependent variables, 2 independent variables

From Bridge Management System (BMS), Korea Transport Database (KTDB), and Korean Meteorological Administration (KMA)

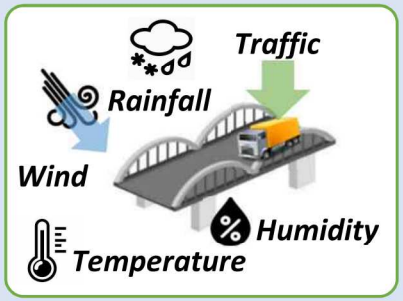
Identification factors (22)



Structural factors (17)



Environmental factors (18)



Inspection factors (2)

Bridge No.	Location (Span#)	Inspection Date	Construction Date	Age
000078	1	20070919	19870101	21
033054	5	20151218	19991230	16
072235	3	20131015	20101229	3

Inspection factors (2)

Element	Damage	Condition
Deck	Cracking	C
Deck	Corrosion	B
Deck	Breakage	B

Types of Damage (7)


Cracking


Map Cracking


Scaling


Breakage


Leakage


Efflore-scence


Corrosion of exposed rebar

Damage Prediction Model Development

Prediction Results Example

Damage = No L1 Severity Level 1~4

Damage Type	Bridge No.	Inspection Date	Age	Location	Predict					Actual					Total Difference
					A	B	C	D	E	A	B	C	D	E	
Cracking	030039	2015-06-24	19	Middle Left		L4	L3				L4	L3			0
	032058	2014-08-30	15	Right					L1					L1	0
	028151	2012-07-15	9	Middle Right				L4					L3		1
	033150	2004-06-15	2	Middle Right	L2	L1				L1	L2				2
Efflorescence	001342	2015-06-09	12	Left				L4					L4		0
	028191	2016-06-21	12	Middle Right			L3					L2			1
	032005	2006-06-26	10	Right		L1									1
	002545	2007-10-23	21	Middle Left		L1	L1				L3	L2			2

Prediction Accuracy by Damage Types

Model	Cracking	Honeycombing	Scaling	Breakage	Efflorescence	Leakage	Corrosion of Exposed Rebar	Average
DNN	68.1	74.7	77.3	71.2	71.7	62.9	69.0	**73.6**
XGBoost	89.56	93.60	95.23	91.45	89.83	91.53	93.97	**94.48%**

(1) Imbalanced data: "No Damage" is predominant → XGBoost weights more for misclassified samples

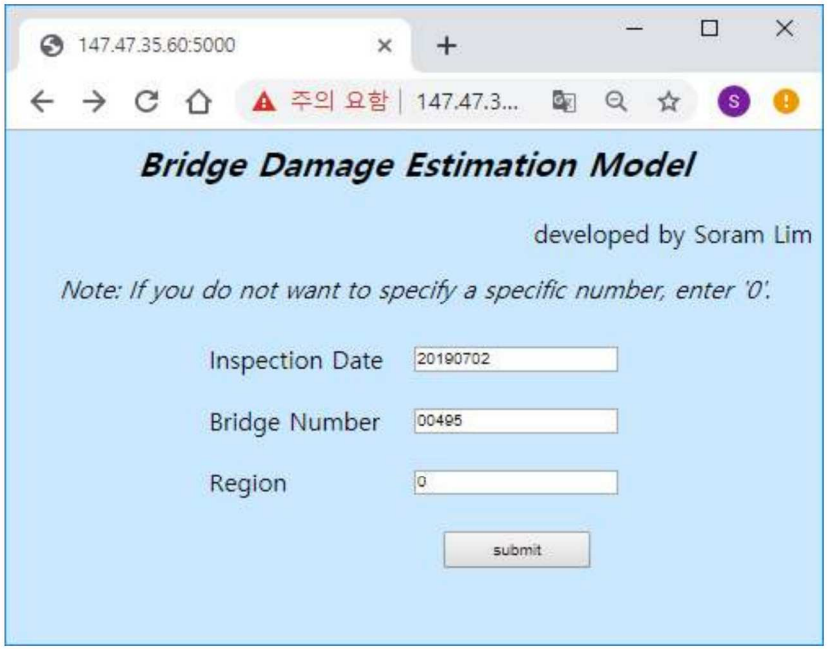
(2) Lack of enough data for every 35 submodels → Difficult to train DNN fully

Bridge Inspection Support

바닥판으로 시작해서 거더(2019), 교량 하부구조 (받침, 교대, 교각, 신축이음, 2020)로 확대 → 2021 년 운영시스템 탑재

Example: Portfolio Generation for Each Bridge

- Expected Inspection: 2019 2nd Half
- Bridge Number: 000495
- Region: Gangwon-Do
- Superstructure: Pre-Stressed Concrete I type
- Age: 28



	Number of Damage
Level 1	1
Level 2	2
Level 3	3 – 5
Level 4	6 and more

CR	Cracking
HC	Honeycombing
SC	Scaling
BR	Breakage
EF	Efflorescence
LE	Leakage
CE	Corrosion of Exposed Rebar

Location		Left							Middle Left							Center							Middle Right							Right						
Span Number		1							2							3							4							5						
Damage Type		CR	HC	SC	BR	EF	LE	CE	CR	HC	SC	BR	EF	LE	CE	CR	HC	SC	BR	EF	LE	CE	CR	HC	SC	BR	EF	LE	CE	CR	HC	SC	BR	EF	LE	CE
Condition Grade	A		L1																																	
	B									L3			L1							L1				L1		L3	L1						L3			
	C		L1			L1															L1						L1	L1						L1		L1
	D		L2					L1												L4		L4							L1					L4		L1
	E																																			

Understanding User Experience and Satisfaction on Facility

Research Methodology

Data Input

신고일
2018-12
신고분야
시설
신고발생지역
대전광역시 유성구
신고내용
어은교 교량 난간이 파손되어
인도에 방치되어 있어 야간이동시
안전사고 우려됨

Complaints or Opinions Text Data

Safety Reports and Open Data(SNS Posts, Newspaper Comments, …)

Text Preprocess

User Dictionary (Synonym, Domain-specific keywords)

Customized Tokenization

Pos-tagging & Extracting Noun

Stopwords Removal

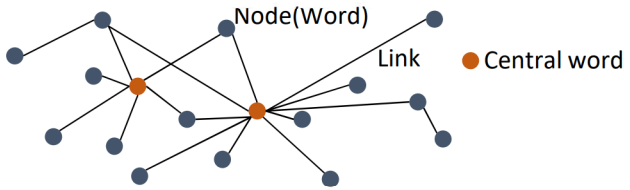
Exploration of User-experienced Factors

***1) Keywords Extraction** (e.g., unsafe or low-service factors)*

Word	TF	DF	TF-IDF
파손	600	300	1.8
난간	550	280	1.9
교량	1200	990	1.2
…	…	…	…
인도	400	300	1.3

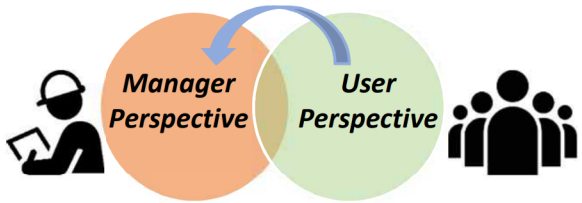
TF-IDF

Word Cloud for Visualization



Analysis of the Association between Keywords (Word Network for Visualization)

2) Verification of Applicability to Facility Inspection Site



Introduction

Necessity of Text Mining for Construction Specifications Review

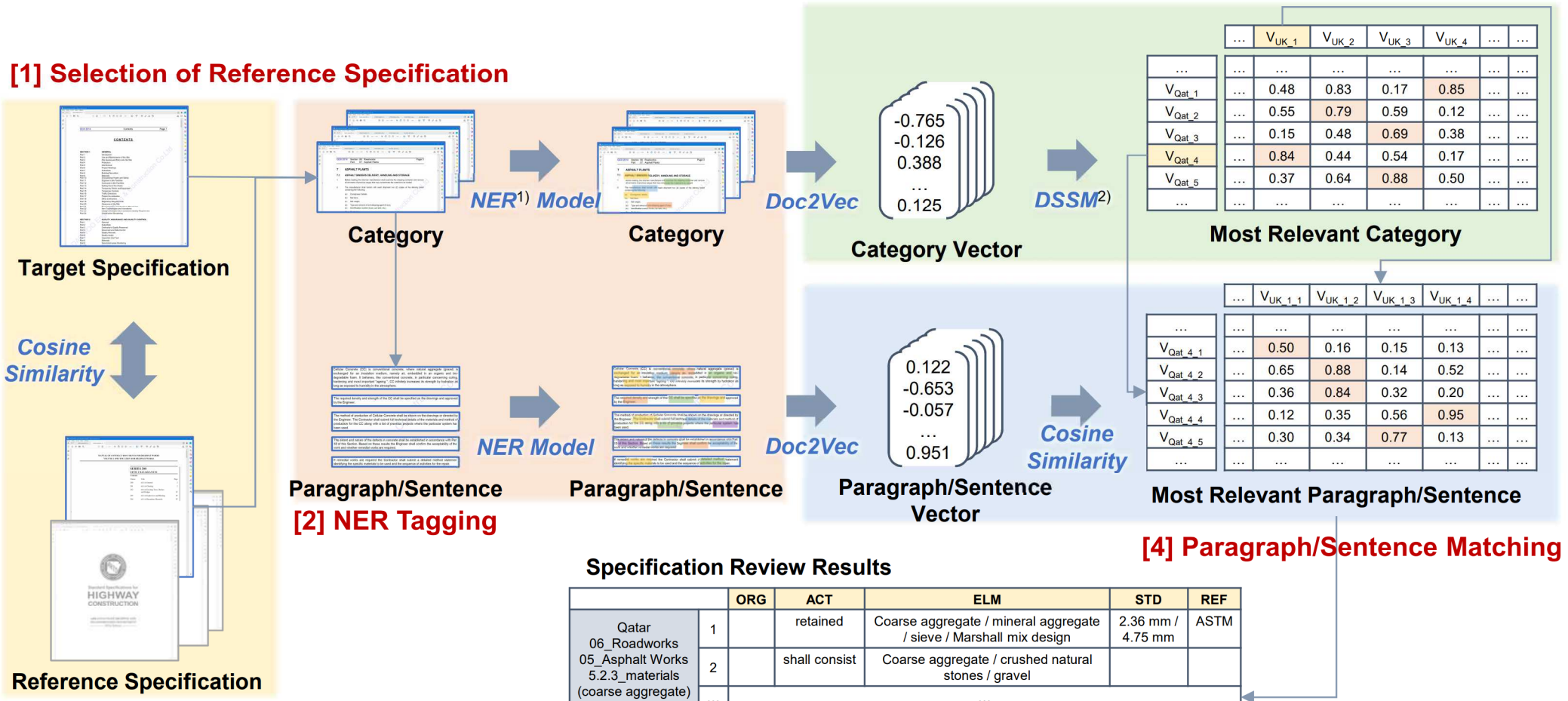
- ***Clients' requirements*** are specified in the construction specifications
- Not satisfy → Rework, resource waste, project delay, …

In Qatar Road Construction Case …

- Performed by Korean Contractor
- Construction specification integrated from standard specifications from US, BS, …
- ***Errors on construction specification*** found during construction
  (e.g., asphalt pavement standard for cold weather in the US, which was not suitable for the Qatar's hot weather condition)
- 4790 pages

Automatic Specifications Review

Automatic Road Construction Specifications Review using Natural Language Processing and Text Mining



Specification Review Results

		ORG	ACT	ELM	STD	REF
Qatar 06_Roadworks 05_Asphalt Works 5.2.3_materials (coarse aggregate)	1		retained	Coarse aggregate / mineral aggregate / sieve / Marshall mix design	2.36 mm / 4.75 mm	ASTM
	2		shall consist	Coarse aggregate / crushed natural stones / gravel		
	…	…				
United Kingdom 09_Road Pavements 9.1_Bituminous Pavement Mixtures	1		retained	Coarse aggregate / mineral aggregate / sieve / Marshall mix design	4.78 mm / 2.21 mm	ASTM
	2		must consist	Fine aggregate / crushed natural stones / gravel		
	…	…				

1) Named Entity Recognition
2) Deep Structured Semantic Model

Paragraph/Sentence Matching

Match Corresponding Sentences

- Embedding every sentence in both of sections matched
- For each sentence in Qatar, find the most corresponding sentence from U.K.

Sentence Match Results

Qatar 06_Roadworks 05_Asphalt Works 5.2.3_materials (coarse aggregate)		United Kingdom 09_Road Pavements 9.1_Bituminous Pavement Mixtures	
1	Coarse aggregate is that portion of the mineral aggregate retained on the **2.36 mm** ASTM sieve for the Marshall mix design and retained on the **4.75 mm** ASTM sieve for the Superpave mix design	1	Coarse aggregate is that portion of the mineral aggregate retained on the **4.78 mm** ASTM sieve for the Marshall mixture design and retained on the **2.21 mm** ASTM sieve for the Superpave mixture design
2	**Coarse aggregate shall** consist of crushed natural stones and gravel	2	**Fine aggregate must** consist of crushed natural stones and gravel
3	Crushed particles shall be cubic and angular in shape and shall not be thin, flaky or elongated	3	Crushed particles shall be cubic and angular in shape and shall not be thin, flaky or elongated
4	The gradation shall be such that when combined with other aggregate fraction in proper proportions, the resultant mixture will meet the required gradation	4	The gradation shall be such that when combined with other aggregate fraction in proper proportions, the restaurant mixture will meet the required gradation
…	…	…	…

Paragraph/Sentence Matching

Experimental Results: Standard Comparison

		ORG	ACT	ELM	STD	REF
Qatar 06_Roadworks 05_Asphalt Works 5.2.3_materials (coarse aggregate)	1		retained	Coarse aggregate / mineral aggregate / sieve / Marshall mix design	2.36 mm / 4.75 mm	ASTM
	2		shall consist	Coarse aggregate / crushed natural stones / gravel		
	…	…				
United Kingdom 09_Road Pavements 9.1_Bituminous Pavement Mixtures	1		retained	Coarse aggregate / mineral aggregate / sieve / Marshall mix design	4.78 mm / 2.21 mm	ASTM
	2		must consist	Fine aggregate / crushed natural stones / gravel		
	…	…				