Week 1

Data Mining Overview

Seokho Chi
Associate Professor I Ph.D.
SNU Construction Innovation Lab

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 Information

- Title: 457.658 Construction IT and Automation
- Timetable
  - Monday 3-7pm @ 35-223
- Instructor: Prof. Seokho Chi
  - shchi@snu.ac.kr, 35-304
  - TA: Jinwoo Kim, jinwoo92@snu.ac.kr, 35-429, 880-4146

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

Course Information

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

Note

- English Lecture, Presentation, and Assignment
- 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%	10/31
Final Report	20%	12/12
Final Presentation	5%	12/12
Individual Assignment	20%	6 times
Final Exam	30%	12/5
TOTAL	100%	

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 perform **at least two** data mining tasks (e.g., classification and clustering) and use **at least three** different algorithms/methods (e.g., decision tree, neural network, and naïve bayes).

Course Schedule (1)

Week	Date	Contents
1	9.5	Course Introduction Data Mining Overview
2	9.12	Data Types Data Pre-Processing Data Exploration
3	9.19	Data Visualization Classification
4	9.26	Classification
5	10.3	개천절
6	10.10	Journal Presentation (1) Computer Lab (1)

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.

Course Schedule (2)

Week	Date	Contents
7	10.17	Classification Prediction
8	10.24	Computer Lab (2)
9	10.31	**Interim Group Presentation**
10	11.7	Cluster Analysis
11	11.14	Mining Association Rules
12	11.21	Journal Presentation (2) Computer Lab (3)
13	11.28	Mining Complex Data Types Trends and Construction Applications
14	12.5	**Final Exam**
15	12.12	**Final Group Presentation**

Group Project Brief

- **DELIVERABLES**
  - Deliverable 1 (10/31) – Project Proposal
    - Problem definition, background, need, objectives, scope, target data set, and proposed data mining approaches
  - Deliverable 2 (12/12) – Project Report
    - Summary of items included on deliverable 1, final results, discussion, conclusions, and recommendations.
- **PRESENTATIONS**
  - Phase 1 (10/31) – Deliverable 1
  - Final (12/12) – Deliverable 2

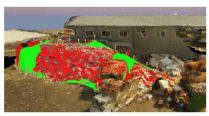
1 IMAGE DATA

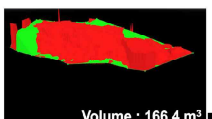
Worker and Equipment Tracking on Construction sites

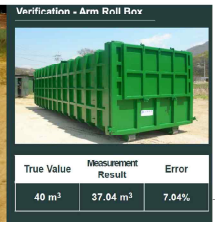
Worker and Equipment Tracking on Construction sites

How much disaster waste is in this picture?

Mesh Model

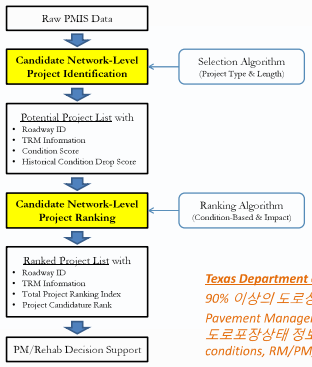






2 SYSTEM DATA

TxDOT Pavement Maintenance Decision Support



CS	CS_Drop	Roadway	TRM	TRM_DISP
18	-24	BI0035LK	422	1.5
58	33	BI0035LK	424	0
53	5	BI0035LK	424	0.5
56	26	BU0079BK	456	0.2
52	15	FM0112 K	556	0
20	-5	FM0112 K	556	0.5
20	1	FM0112 K	556	1
47	31	FM0112 K	556	1.5
58	24	FM0112 K	558	1.5
52	27	FM0112 K	562	1
35	-55	FM0112 K	562	1.5
53	11	FM0112 K	564	0.5
47	7	FM0112 K	564	1
48	16	FM0112 K	566	1.5
27	7	FM0112 K	568	1
28	-6	FM0112 K	568	1.5

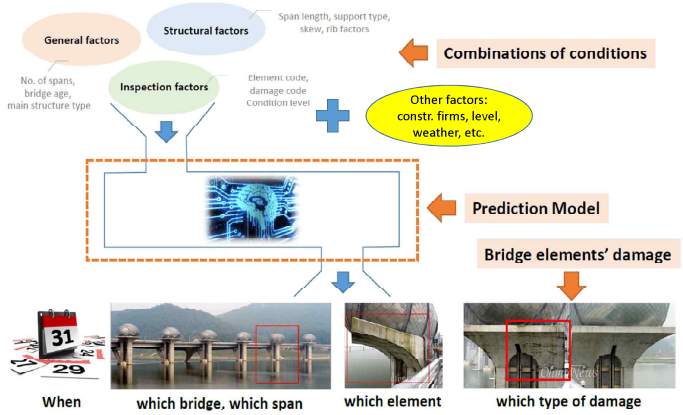
Texas Department of Transportation

90% 이상의 도로상태를 "Good or Better"로 목표

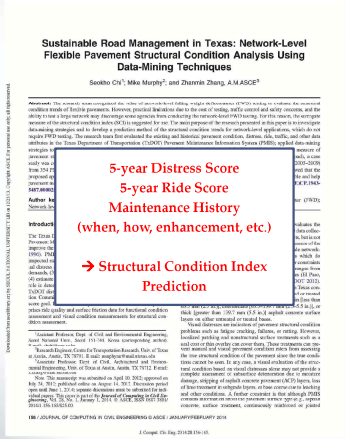
Pavement Management Information Systems (PMIS): 4개월에 걸친 도로표장상태 정보수집(distress score, ride score, actual distress conditions, RM/PM/Rehab 예산관련정보 등)

참조: 텍사스 도로 총 연장: 314,000 km (2008) VS 한국: 106,414km (2013)

Bridge Damage Prediction



TxDOT Pavement Maintenance Decision Support

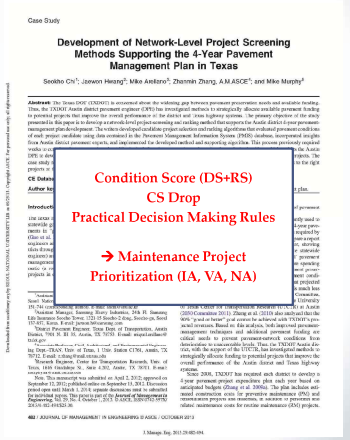


Sustainable Road Management in Texas: Network-Level Flexible Pavement Structural Condition Analysis Using Data-Mining Techniques

Seokho Chi[1]; Mike Murphy[2]; and Zhanmin Zhang, A.M.ASCE[3]

5-year Distress Score
5-year Ride Score
Maintenance History
(when, how, enhancement, etc.)

→ Structural Condition Index Prediction



Case Study

Development of Network-Level Project Screening Methods Supporting the 4-Year Pavement Management Plan in Texas

Seokho Chi[1]; Jaewon Hwang[2]; Mike Arellano[3]; Zhanmin Zhang, A.M.ASCE[4]; and Mike Murphy[5]

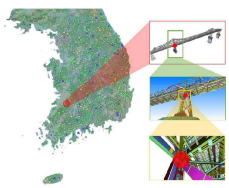
Condition Score (DS+RS)
CS Drop
Practical Decision Making Rules

→ Maintenance Project Prioritization (IA, VA, NA)

Bridge Damage Prediction

Expected Results

- Find combinations of specific conditions related to damages and build portfolios
- Suggest probability of damage occurrence



1) Bridge number: 03447
2) Bridge name: ○○
3) Span number: 3

Combination of conditions	Element	Damage	Level	Pro.
{(320 ≤ chloride level < 360) (970.0 ≤ Bridge length < 1289.3), (1829 ≤ ADT < 22064), (7.75 ≤ Pavement thickness < 8.50)}	Slab	Crack	C	70 ~80%
{Location = △△, (80 ≤ Precipitation < 100), Constr. Level = high, (17.85 ≤ wide < 20.08), (23.7 ≤ Deck thickness < 27.5),}	Deck	Crack	D	60 ~70%
{Start date = 1998, Location = XX, Constr. firm = A, Number of lanes = 4, Girder type = PSCB}	Cross beam	Exposed rebar	C	50 ~60%

3 TEXT DATA

두산의 타선, '봇물 터졌다'...kt 상대로 대승
(2016-06-21, kt 1勝 12 두산, 잠실)

21일 잠실구장에서 열린 2016 타이어뱅크 KBO리그 두산과 kt의 경기에서 두산이 포문을 연 에반스의 1루타 이후 타선이 폭발하면서 kt를 상대로 파죽의 대승을 거두었다. 두산은 16안타 3홈런을 뽑아내며 탁월한 경기력을 보여줬다. 두산은 0:0으로 경기 중이던 3회 말, 에반스가 1루타를 때리며 1점을 획득했다. 이후, 허경민, 박세혁이 활약해서 두산의 승리에 크게 기여했다.

1회 말 두산은 2사 2루 상황에서 김재환의 볼넷으로 2사 1, 2루 상황을 만들었으나 이후 에반스의 3루수 땅볼로 공수교대가 이루어지며 점수를 내지 못했다. 그 후 2회 말에는 2사 1, 2루 상황에서 박건우의 중견수 플라이로 공격권을 넘겨주며 득점에 실패했다. 또 3회 말에는 무사 1, 2루 상황에서 에반스의 1타점 적시타로 1점 앞서나가기 시작했고, 허경민의 1타점 적시타로 2점을 달아났다. 4회 말에는 무사 1, 3루 상황에서 정수빈의 1타점 적시타로 1점을 달아났고, 이중도루로 무사 2, 3루 , 김재환의 볼넷으로 1사 만루 상황을 만들고 에반스의 4점 홈런으로 1점을 달아났다. 박세혁의 2점 홈런으로 2점을 달아났다. 5회 말에는 박건우의 1점 홈런으로 1점을 달아났다. 무사 1, 3루 상황에서 김재환의 희생플라이로 점수 차를 벌렸다. 2사 1, 3루 득점찬스를 맞이하였으나 허경민의 좌익수 플라이로 이닝이 종료되며 달아나지 못했다.

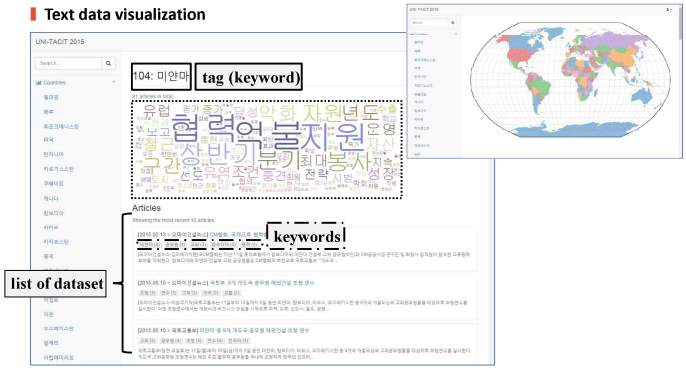
8회 초 kt는 김상현의 1점 홈런으로 1점을 따라잡았다.

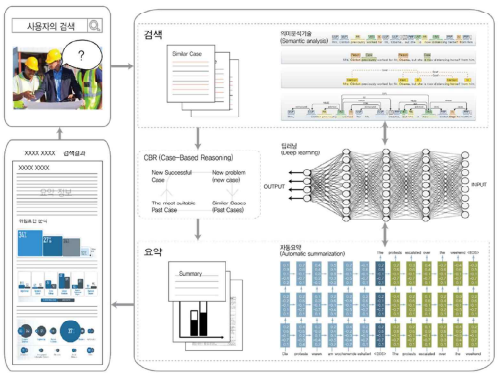
8회 말 두산은 무사 2루 상황에서 양의지의 데드볼로 무사 1, 2루 상황을 만들고 허경민의 1타점 적시타로 1득점하며 경기결과를 확정지었다.

끝내 두산은 kt에게 대패를 안겨줬다. 오늘 경기 결과에 따라 두산은 3연승을 올렸고 현재 1위(승률 0.727)이다. 한편 kt는 4연패의 수렁에 빠졌고 현재 한화와 공동9위(승률 0.406)이며 3안타를 때려 다소 아쉬운 플레이를 보여줬다.

UNI(User Needed Information)-Tacit

Text data visualization



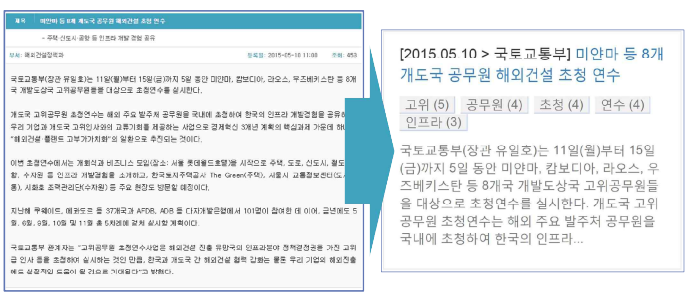


E.g. 1) Accident reports → when, where, what type of accidents, why?

E.g. 2) SOC inspection reports → maintenance characteristics based on type, design, specification, methods, etc.?

UNI(User Needed Information)-Tacit

Automated keyword extraction and document tagging



[2015.05.10 > 국토교통부] 미얀마 등 8개 개도국 공무원 해외건설 초청 연수

고위 (5) 공무원 (4) 초청 (4) 연수 (4) 인프라 (3)

국토교통부(장관 유일호)는 11일(월)부터 15일(금)까지 5일 동안 미얀마, 캄보디아, 라오스, 우즈베키스탄 등 8개국 개발도상국 고위공무원들을 대상으로 초청연수를 실시한다. 개도국 고위공무원 초청연수는 해외 주요 발주처 공무원을 국내에 초청하여 한국의 인프라...