

## 철도 위험관리 데이터 연계 분석을 위한 기준 데이터 매핑 연구

### A Study on Data Mapping for Integrated Analysis of Railway Safety Data

변현진 · 이용상\*

Hyun-Jin Byun · Yong-Sang Lee

**Abstract** The railway system is an interface industry that can be safely operated by organically operating the lines, vehicles, controls, etc. Various data are generated in the operation and maintenance activities of the railway system. These data are utilized in cooperation with safety and maintenance activities in each field, but amount of data is insufficient for data analysis of safety management due to relevant data being produced without any synchronous criteria such as time or space. In particular, reference data such as location and time of failure data for each field are set to different criteria according to the work characteristics in each field. So, it is not easy to analyze data integrally based on location and time. Therefore, mapping of reference data can be required for integrated analysis of data defined in different formats. By selecting data mapping tools and verifying the results of safety relevant data with the same criteria, the purpose of this paper is to enable integrated analysis of railway safety management data occurring in different fields based on location and time.

**Keywords** : Railway, Risk management, Data mapping, Location information, Risk factor

**초 록** 철도는 인터페이스 산업으로 철도차량·선로·관제 등이 유기적으로 맞물려야 움직일 수 있다. 이를 정상적으로 운영하기 위해서는 많은 유지보수 활동이 필요한데 이러한 활동과정에서 발생하는 데이터들은 각 분야별로 관리하고 있어 업무특성에 따라 위치, 시간 등 기준 데이터가 각각 상이하여 시·공간 기반의 통합 분석을 할 수 없는 실정이다. 각 분야별로 상이한 데이터를 통합 분석하기 위해서는 데이터마이닝 기법을 통해 유의미한 정보를 추출하여 분석하는데, 이러한 분석을 위해서는 데이터 유형의 상호관련성을 인지하기 위한 기준정보의 추출과 매핑이 반드시 필요하다. 본 연구는 서로 다른 분야에서 발생하는 안전과 관련된 데이터에 대한 매핑 툴 선정과 결과에 대하여 검증을 수행하고자 한다.

**주요어** : 철도, 위험관리, 데이터 매핑, 위치정보, 위험요인

## 1. 서 론

철도는 철도차량, 선로, 관제 등이 유기적으로 결합하여 정상적으로 운영될 때 비로소 열차가 안전하게 운행되는 시스템이다. 열차를 안전하게 운행하기 위해서는 다양한 유지보수 활동이 필요한데 이 과정에서 수집되는 정보는 하루에도 엄청난 양의 정보가 누적된다. 이러한 정보는 한국철도공사의 경우 KOVIS<sup>1)</sup> 및 XROIS<sup>2)</sup> 등 분야별 시스템에 주요 정보를 관리하고 주관부서에서 실시간 감시 또는 정기적인 분석을 시행하고 있으나, 정보 상호간의 연관관계와 철도사고·장애 예방 차원에서 어떻게 관리되어야 하는지 과학적인 분석이 이루어지지 않고 있다.

최근 국내·외 교통분야에서는 대부분 GPS좌표 기반의 위치정보에 대한 연구가 활발하게 진행 중이며, 선진국의 주요 연구사례로 철도 유지보수 및 데이터 관리를 위하여 빅데이터를 활용하여 실 상황에 맞는 관리 방법을 연구하여 진행 중이다. 예를 들면, 미국의 유니온 퍼시픽 사는 철도 궤도이탈을 포함한 안전사고 관리 및 배기가스 최소화 등 철도사업 운영개선에 빅데이터를 활용하였으며, [1] 필라델피아의 SEPTA Regional Rail은 정확한 열차 도착과 출발 및 승객의 흐름을 모니터링하고 최적화

<sup>1)</sup> KOVIS : KORail Vision Innovation System의 두문 문자로 전사적 자원관리 시스템

<sup>2)</sup> XROIS : EXtended Railroad Operating Information System의 두문 문자로 차세대 철도운영 정보시스템

\*Corresponding author. E-mail: yongsang@wsu.ac.kr.

© 2017 The Korean Society for Railway. All rights reserved.

<https://doi.org/10.7782/JKSR.2017.20.5.703>

해를 도출하기 위하여 빅데이터를 활용하여 분석연구를 진행 중이다.[2] 데이터 매핑 관련하여 고영만 등(2005)은 문헌정보에 대해 데이터 매핑을 위한 데이터 요소의 명명규칙에 대하여 연구를 하였고[3], 서태설 등(2007)은 메타 데이터 레지스트리 표준에 따라 기존 데이터 모델과 메타데이터 레지스트리를 혼합한 의미적 데이터 모델링 프로세스를 제시하였다.[4] 김학기 등(2016)은 의료분야의 서로 다른 형식으로 정의된 데이터를 활용하여 통합분석을 위한 데이터 매핑 연구를 지속적으로 진행해 왔다.[5]

데이터베이스는 개발환경 및 조건에 따라 같은 의미의 데이터라도 다른 이름으로 다른 값을 표현하고 있어 데이터의 상호용성에 많은 어려움을 가진다. 이러한 문제해결을 위해 같은 의미의 데이터를 같은 스키마와 속성을 가지게 하는 것을 데이터 매핑이라 한다.[6] 이와 같이, 매핑은 정보의 수집주기 및 저장방식이 분야별 업무특성에 따라 각각 상이한 데이터간 의미적 불일치(semantic heterogeneities)가 발생하는 분야에서 필요하다. 특히 데이터간의 상호관련성을 가지는 연계 정보를 분석하는데 기준데이터를 설정하여 매핑하는 것이 필수적이라 할 수 있다. 이러한 정보를 시간과 거리정보를 기준데이터로 KDD (Knowledge Discovery in Databases)[7-8] 데이터마이닝 기법을 활용하여 데이터 기반 위험관리 정보의 상호관련성을 파악하고자 하였다. 이를 위해서는 다음 Fig. 1과 같이 데이터의 수집과 선택, 전처리, 변환 데이터 마이닝, 결과의 해석과 평가를 통해 달성될 수 있다.

본 논문에서는 한국철도공사의 경부고속철도 노선을 중심으로 2011년 이후 최근까지 발생한 운영 및 유지관리 데이터를 이용하였으며, 차량·시설·전기 세 분야로 각자의 데이터가 관리되어 운영 및 유지보수 되어왔다. 하지만 철도 위험관리를 위해서는 분야간 위험요인들이 상호연관성이 있을 것이라는 가정과 데이터 매핑을 통한 복합적 고장 및 장애가 발생하는 구간을 분석할 수 있다고 판단하였다. 또한, 현재 한국철도공사 내 분야별 데이터 관리는 되고 있으나, 통합적으로 파악할 수 있는 연구 및 협업에는 한계가 존재한다. 이러한 문제점을 해결하기 위하여 분야별 데이터를 매핑하여 상호연관성을 파악하고 패턴분석을 실시하고자 한다. 따라서 시·공간적 기준을 중심으로 철도시스템 운영 및 유지관리 활동 중 다양하게 발생하는 선로, 전차선, 차량분야에서 수집된 데이터의 유형을 2.1에서 살펴보고, 이러한 데이터의 상호관련성을 확인하기 위하여 필요한 데이터를 시간과 공간을 기준으로 정렬시켜 분석하는 전처리 프로세스인 매핑방법의 선택과 매핑 툴의 선정을 2.2에서 소개하고자 한다. 그리고 2.3에서 매핑된 결과 등 처리과정 연구를 통해 데이터 신뢰성을 검증하는 것으로 본 연구의 범위를 한정하였다.

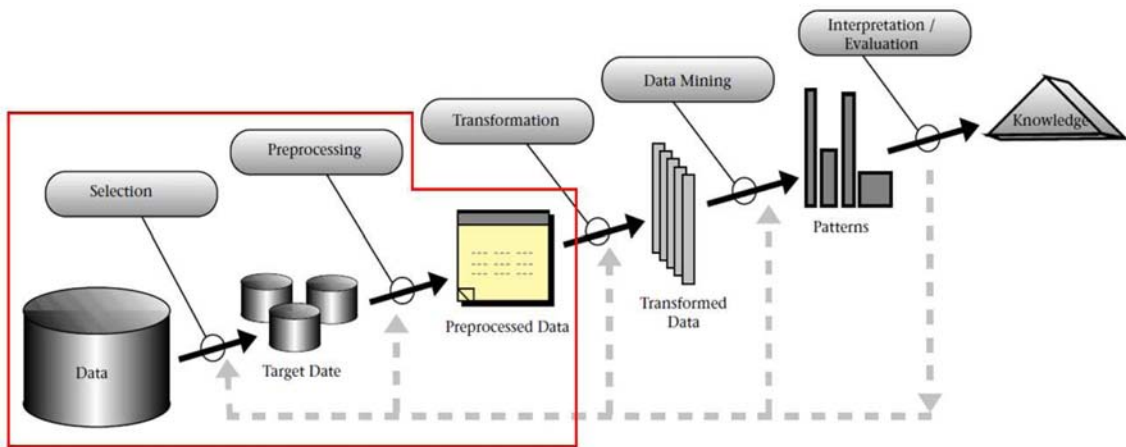


Fig. 1. Overview of the steps that compose the KDD (Knowledge discovery in databases) process.

## 2. 본 론

### 2.1 철도 분야별 데이터 유형

고속철도시스템의 운영 및 유지관리 과정에서 발생하는 데이터는 KOVIS(KOrail Vision Innovation System)내 D/B로 저장되고 분석된다. KTX의 경우 열차 운행 중에 발생하는 차량고장을 차상컴퓨터(OBCS) 장치로 검지하여 CDMA 방식의 무선전송 장치를 통해 데이터를 실시간으로 전송하고 있으며, 또한 레일, 침목, 전차선 등 시설물은 레일점검차, 전기검측차 등 검측차량의 주기적 운영을 통해 결함을 검지하여 데이터를 저장하고 있다. 이러한 데이터들은 각 담당분야의 안전 및 유지보수 활동과 연계하여 활용하고 있다.

**Table 1.** Data structure for railway vehicle field.

Classification	Format	Example
Fleet number	Char	1C1001
Car number	Float	100102
Computer	Char	P1
Date	Date	2011-10-08
Time	Time	8:23:31PM
Faultcode	Char	C1-28-08
Fault type	Char	MF
Code name	Char	Unstable bogie(ch1 pc5)
Object sub-system	Char	Running equipment
System name	Char	KTX-drive
System group	Char	KTXSYS06
System code	Float	030
Action	Char	Yes
Action description	Char	Inspection

**Table 2.** Data structure for a track inspection car.

Classification	Format	Example
Year	Float	2011
Date	Date	2011-04-14
Equip. number	Char	TIC0702
Location code	Char	H-H1-H11895B-TR0-MM201
Number	Float	159
Defect class	Float	3
Line	Char	H-H1
Department	Float	58400
Up/down	Char	MM201
Start point	Float	193.801
End point	Float	193.801
Width	Float	21
Length	Float	31
Area	Float	433
Defect group	Char	Compact Defect et al.
Rail(left/right)	Char	Left

**Table 3.** Data structure for a track alignment monitoring car.

Classification	Format	Example
Equipment	Char	A track alignment monitoring car
Operation	Char	regular(1/4)
Location code	Char	R-01-010000A-TR0-MM002
Inspection date	Date	2011-04-19
Plant	Float	2100
Work place	Float	41110
Type	Char	point out
Overlapping	Float	0
Start point	Float	0.301
End point	Float	0.319
Defect code	Float	18
Defect type	Char	Alignment Light et al.
Irregularity(max)	Char	24
Location(max)	Float	0.316
Speed type	Float	3
Speed limit	Float	5

**Table 4.** Data structure for railway electric field.

Classification	Format	Example
Date	Date	2011-02-23
Location code	Char	H-H1-H12723B-GT-01-T001-M082
Location name	Char	Catenary pole[FL]: Jicheon IEC-Daegu North:T1():274-1
Planner group	Float	1400
Group name	Char	Osong
Work place	Float	71040
Work place	Char	Yakmock
Doc. name	Float	537602121
Sight	Float	4379920
Location	Char	Catenary 274-1
Measurement value	Float	5050
Unit	Char	MM
Noice number	Float	30000483609
Notice descript.	Char	Integrated inspection(Feb) catenary height
Speed type	Float	3
Speed limit	Float	5

**2.1.1 차량분야**

한국철도공사 전사자원관리시스템(KOVIS)에 저장되는 차량분야 안전관련 데이터는 고속열차 내 차상컴퓨터(OBCS)에서 고장정보가 실시간으로 수집된다. 차상컴퓨터에 기록된 정보는 무선전송장치를 통해 실시간 전송되거나 차량기지에서 일괄 업로드 되어 서버에 저장된다. 지상시설과의 인터페이스 위험요인을 파악하기 위하여 발생하는 고장정보 중 대차현팅으로 한정하여 데이터를 추출하였고, 차상컴퓨터에 기록되는 데이터 유형은 Table 1과 같다. 또한, 각 노선별로 상-하행에 따른 해당 차량기지

별로 입력되는 데이터를 활용하였다.

### 2.1.2 시설분야

시설분야 데이터는 선로점검차, 궤도점검차 등 점검차량의 측정자료를 주로 수집하였다. 선로점검차의 경우 레일표면결함, 체결구결함(절손 및 탈락), 침목결함(균열)의 데이터가 수집되며, 궤도-종합점검차는 고저, 방향, 궤간, 수평, 뒤틀림 등 궤도틀림 데이터가 수집된다. Table 2는 선로점검차, Table 3은 궤도-종합점검차의 데이터 구조를 나타낸다. 각 분야간 인터페이스 위험요인 및 시설분야 데이터간 상관관계를 파악하기 위하여 발생하는 점검정보를 모두 활용하였다.

### 2.1.3 전기분야

전기분야에서 위험관리를 위한 데이터는 주로 궤도점검차, 전철시험차, 종합점검차, KTX-36호에서 수집된다. 궤도점검차는 전차선로의 높이, 편위 불량을 측정하고 전철시험차는 전차선로의 높이, 편위, 마모에 관한 불량 정보가 측정되며 종합점검차는 전차선로의 편위, 높이, 마모, 무선통신, 궤도회로에 관한 불량 정보가 수집된다. KTX-36호의 경우에는 전사자원관리시스템(KOVIS)에서 2017년 3월 이후의 전차선 집전상태와 열차무선설비(전계강도 등), 불연속정보 등에 관한 정보가 수집되고 있다. 각 분야간 인터페이스 위험요인을 파악하기 위하여 노선별 궤도점검차, 전철시험차, 종합점검차, KTX-36호로부터 수집되는 정보 중 전차선로 편위, 높이 불량 데이터로 한정하여 데이터를 추출하였고, 전사자원관리시스템(KOVIS)에 기록되는 데이터 구조는 Table 4와 같다. 또한, 차량분야와 같이 경부고속선 대상의 동일한 발생시점 기준으로 중복되는 데이터는 제거하였다.

## 2.2 철도 분야별 시·공간 데이터 매핑

철도의 각 분야별로 상이한 데이터를 동일한 시간, 공간기준으로 통합하는 방법인 데이터 매핑을 통해 시공간기반으로 기준 정보를 Fig.1[2]과 같이 매핑하는 과정인 전처리(preprocessing) 프로세스에 대한 연구로 한정하였다. 연구의 시간적 배경으로 2011년부터 2016년까지의 과거 데이터 중 기준 데이터가 상이한 유형의 데이터를 대상으로 하였다. 본 연구의 대상범위는 차량분야의 고속차량 고장코드, 시설분야의 점검차 측정기록, 전기분야의 점검차 및 점검장비가 장착된 KTX-36호의 측정기록 데이터를 활용하였다.

데이터 매핑의 의미는 시스템 별로 산재해 있는 데이터 정보 요소에 대한 명칭, 정의, 규칙에 대한 원칙을 수립하여 모든 사용자가 동일한 의미로 사용하고 일관된 관점으로 정보를 제공하기 위해 필요한 매핑 데이터를 생성하는 일련의 프로세스와 매핑 데이터를 전사적으로 적용하는 것을 말한다.[6]

한국철도공사 전사자원관리시스템(KOVIS)에 저장되는 차량분야 위험관리 데이터는 2.1에서 소개한 바와 같이 차량, 시설, 전기분야의 데이터들은 각 분야 담당자 및 전문가에 의한 주관적 정보의 입력방식으로 관리되어 왔다. 이러한 각 분야별 데이터들을 조합하여 위험관리 분석을 위한 데이터로 활용하기 위해서는 데이터의 매핑이 반드시 필요하다. 2011년부터 2016년까지의 각 노선별 시간적 데이터와 공간적 데이터의 매핑을 통하여 안전 데이터 통합분석을 수행하였다.

### 2.2.1 데이터 매핑시 접근 툴(Tool)

본 논문에서 데이터 매핑을 위한 툴로는 첫째, Oracle, MySQL, MS-SQL 등 많은 데이터베이스 관리 프로그램(DBMS: DataBase Management System)들 중 MS-SQL을 활용하여 데이터베이스를 구축하였다. MS-SQL은 오픈소스 기반의 MySQL에 비해 UI(User Interface)가 탁월하며, 타 컴퓨터 프로그래밍용 프로그램과 연동하여 사용할 경우 접근성이 용이하다. 또한, 본 논문에서 사용할 통계 분석 프로그램인 R과의 호환성이 용이하다. 둘째로 다양한 데이터 분석 프로그램들 중 통계적 접근에 용이하고 오픈소스를 활용하여 다양한 오픈 데이터들과 연계, 매핑, 코딩, 분석이 가능한 R 통계 프로그램을 채택하여 데이터 병합, 필터링, 통계분석, 그래프 등을 수행하였다. Table 5는 데이터 매핑을 위하여 사용된 R 통계 프로그램 패키지의 목록이다.

Table 5. R packages used for standardization.

Package	Explanation
Xlsx	Read/write/format Excel 2007 and Excel 97/2000/XP/2003 files
Xlsxjars	Package required POI jars for the package
RJava	Low-level R to Java interface
RODBC	ODBC database access
MASS	Support function and datasets for Venables and Ripley's MASS
Reshape2	Flexibly reshape data: a reboot of the reshape package
Ggplot2	Create elegant data visualisations using the grammar of graphics

### 2.2.2 시·공간적 데이터 전처리 과정

차량분야 데이터의 경우 편성코드, 차량번호, 컴퓨터, 발생일자, 발생시간, 고장코드, 장애구분, 코드명칭을 추출하였고, 장애구분의 MF(Main Fault) 중 대차현팅 감지 데이터를 추출하여 시·공간적 데이터 전처리를 시행하였다. 시설·전기분야간 인터페이스 위험요인을 파악하기 위하여 차량의 흔들림과 관련이 되는 대차현팅 감지정보를 차량분야 데이터의 대표로 사용하였다. 차량에서 발생하는 고장데이터의 경우 현재 위치정보가 제공되지 않으나, 발생일자 및 시간을 운행정보기록장치(ATESS)와 결합하여 열차위치정보를 연계하였다.

시설분야에서 선로점검차로부터 추출한 데이터는 운행년도, 운행일자, 기능위치, 결함등급, 노선, 상하구분, 결함시작위치, 결함종료위치, 결함폭, 결함길이, 결함면적, 결함종류, 레일의좌우측이며 위치정보의 경우 각 노선별 상·하행에 대한 결함시작위치와 결함종료위치의 중간 값을 활용하였다. 또한, 궤도점검차로부터 추출한 데이터는 기능위치, 검측일자, 시점, 중점, 불량, 불량 유형, 최대 틀림치, 최대 틀림위치이고, 위치정보의 경우 각 노선별 상·하행정보를 기능위치에서 별도로 추출하였으며, 최대 틀림치 정보를 사용하였다. 그리고 면틀림 좌우, 줄틀림 좌우, 비틀림, 궤간틀림 등 결함유형에 따라 별도로 데이터베이스화 하였다.

전기분야의 경우는 일자, 기능위치, 기능위치명, 측정지점, 측정위치, 측정값, 측정단위, 통지번호를 추출하였다. 전기분야의 위치정보는 전신주 번호 기반으로 구축되어 있어서 정확한 위치를 파악하기 어렵다. 또한, 발생하는 불량 데이터의 경우 현재 위치정보가 기능위치명에서 간헐적으로 제공되어서 본 연구에서는 기능위치에서 나타난 전신주 코드마다 해당 위치 값을 별도로 도출하여 데이터베이스화 하였다. 그리고 높이불량과 편위불량에 대한 별도의 구분이 없어서 측정값에 따라 불량구분을 별도로 데이터베이스화 하였다.

### 2.2.3 시·공간적 데이터 매핑 방법

차량분야의 대차현팅 감지데이터, 시설분야 선로점검차의 레일표면결함, 체결구결함, 침목결함, 궤도점검차의 면틀림 좌·우, 줄틀림 좌·우, 비틀림, 궤간틀림, 그리고 전기분야의 높이불량, 편위불량의 데이터들을 통합하여 발생일자, 운행일자, 검측일자, 일자에 정보들을 년, 월, 일별 시간적 데이터로 매핑 하였다. 위치정보의 경우 각 분야별 기능위치로부터 고속·일반, 상·하행, 노선별 데이터를 추출하였고 Table 6과 같이 R 프로그램으로 reshape2 패키지 내 melt함수와 cast함수를 이용하였다. reshape2 패키지는 데이터의 모양을 바꾸거나 그룹별 요약 값을 계산하는 함수를 담고 있는 패키지로 그 중 melt함수는 여러 컬럼(column)으로 구성된 데이터 식별자(id), 측정 변수(variable), 측정값(value)이라는 3개 컬럼으로 변환하는 함수이고, cast함수는 melt된 데이터를 다시 여러 컬럼으로 변환하는 함수이다.

본 연구에서는 운행년도, 발생위치, 결함등급을 melt함수를 활용하여 추출하였고, cast함수를 활용하여 일자별 발생위치에 대한 빈도를 도출하고 데이터들을 통합하였다. 그 결과 값을 ggplot2 패키지를 사용하여 x축을 km로 표기하고 y축을 년/월/일별 발생 빈도로 누적그래프를 도출하여 시·공간적 분야별 결함유형을 도출하였다.

Table 6. R packages(function) used for data mapping.

Function Name	Application method
Cbind()	Cbind(year, month = month(year\$date))
Melt()	Melt(year, id.vars = c("date", "dist", "defect class", "month"), measure.vars = "location")
Cast()	Cast(melt function, year+dist-month+variable, lenth)

### 2.2.4 시·공간 데이터 매핑 결과

Table 7~Table 10은 시·공간적으로 통합 분석을 위하여 안전관련 데이터의 위치정보를 매핑하는 일련의 과정 및 결과를 나타낸다. 일련의 과정들은 모두 MS-SQL로부터 받아온 데이터를 R 프로그램을 활용하여 필터링, 데이터 병합, 통계분석, 그래프 작성 등을 수행하였다. Table 7는 선로점검차의 일자별 데이터를 추출한 결과를 보여주는 예시로 년/월/일별로 추출하였다. 이와 같이 시설, 차량, 전기분야의 데이터들을 동일하게 처리하였다. Table 8과 Table 9는 melt함수와 cast함수로 위치정보를 추출하고 이를 매핑하는 과정의 예시를 R 프로그래밍 소스 코드를 통하여 나타난 결과 화면이다. Table 10은 Table 7의 결과를 위치정보에 따른 월별 결함 및 불량빈도를 나타낸 결과의 예시를 보여주는 화면이다. 이렇게 추출한 시설, 차량, 전기분야의 병합된 최종 데이터를 활용하여 시각화 할 수 있었으며, 시각화된 결과를 통하여 각 분야별 데이터 간 상관관계를 검증할 수 있었다.

**Table 7.** Example of data extraction.

Date	Location code	Defect class	Up/down	Start point	End point	Width	Length	Area	Defect group	Rail (left/right)
2011-04-11	H-H1-H11895B-TR0-MM201	3	MM201	193.801	193.801	21.00	31.00	433.00	Compact Defect	Left
2012-06-12	R-03-031374B-TR0-MM101-	3	MM101	139.067	139.067	23.00	58.00	401.00	Compact Defect	Right
2013-05-07	R-61-610137B-TR2-MM001	3	MM001	18.242	18.242	31.00	41.00	412.00	Compact Defect	Left
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

**Table 8.** Example of location data mapping.

Date	Absolute location	Defect class	Month	Real locate
2011	144	3	4	144.21
2012	143	3	6	143.17
2013	131	3	5	131.02
⋮	⋮	⋮	⋮	⋮

**Table 9.** Example of location data mapping(monthly).

Date	Absolute location	Jan	Feb	Mar	Apr	...	Sep	Oct	Nov	Dec
2011	144	0	0	0	0	...	2	2	0	0
2012	143	1	0	0	8	...	4	8	1	0
2013	131	2	0	2	7	...	45	12	2	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

**Table 10.** Example of location data mapping result.

date	Km	Jan. fastening defect	...	Dec. fastening defect	Jan. rail defect	...	Dec. rail defect	Jan. bougie hunting	...	Dec. bougie hunting	Jan. alignment	...	Dec. alignment	...
2011	1	2	...	0	0	...	0	1	...	0	2	...	1	...
2012	3	1	...	1	0	...	2	2	...	1	0	...	0	...
2013	7	2	...	0	1	...	3	1	...	2	3	...	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

### 2.3 철도 분야별 시·공간 매핑 결과 분석

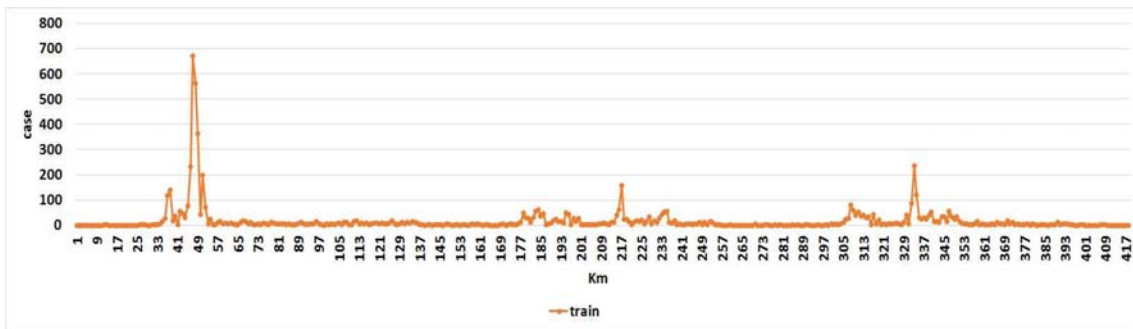
본 연구에서 적용한 매핑방법에 대한 검증을 위하여 데이터 매핑 결과를 그래프로 시각화하여 분석하고 분석결과 특정구간에서 발생하는 복합장애 요인에 대해 실제로 측정할 결과를 비교하여 서울기점 40~60km 구간에서 유사한 형태로 차량과 지상 시설의 장애가 발생하고 있음을 확인할 수 있었다.

#### 2.3.1 시공간 데이터 매핑결과의 시각화

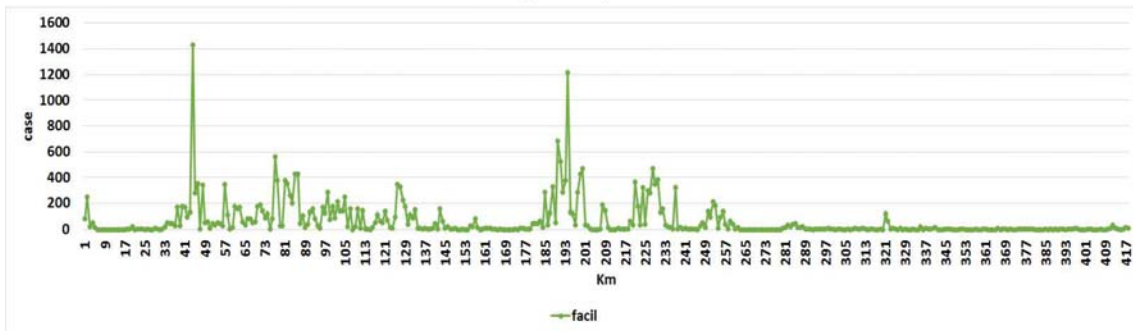
앞에서 설명한 바와 같이, 한국철도공사 내 분야별로 수집되는 데이터들 간 상호연관성을 알아볼 수 있는 별도의 연구나 관리된 바 없으며, 일반적으로는 선로검측차와 궤도검측차로부터 얻는 시설·전기분야의 시·공간적 데이터로 한정적이다. 특히, 차량분야의 전처리리를 하기 위해서는 공간적 정보와 장애정보 간 연계는 더욱이 어렵다. 분야별 데이터로 취약개소를 관리하는 형태의 분석은 해당부서에서 진행된 바가 있다. Fig. 2는 분야별로 동일노선에 결함 데이터들을 전처리하여 그래프 형태로 시각

화 하였으며, 분야간 상호연관성을 알아보기에는 한계가 존재한다. 또한, Fig.3과 같이 분야별로 결합하여 상호연관성을 알아보기 위한 시도는 시행된 바가 없으며, 분야별 담당자에 의하여 해당 분야의 결합에 관해서만 보고 되어왔다. 이러한 분야별 상호연관성을 파악하고 인과관계를 알아보기 위하여 본 연구에서는 2011년 이후 최근까지 차량의 고장과 시설물 유지관리 데이터를 시공간 기준으로 매핑하여 확인한 결과 Fig.3과 같이 각 분야별로 경부고속철도 노선에서 발생한 장애·결함을 분야별로 확인한 결과, 동일 구간인 서울기점 47km구간에서 시설분야는 1,428건, 차량분야는 최대 672건의 대차현팅 정보가 발생한 것으로 나타났다. 또한, 매년 반복적으로 유사구간에서 시설과 차량의 결합이 발생하는 것으로 확인되었다. 전기분야는 경부 노선에서 최대 중대결함이 특정구간 3건으로 상호 관련성을 파악하기에는 부족한 것으로 확인되었다.

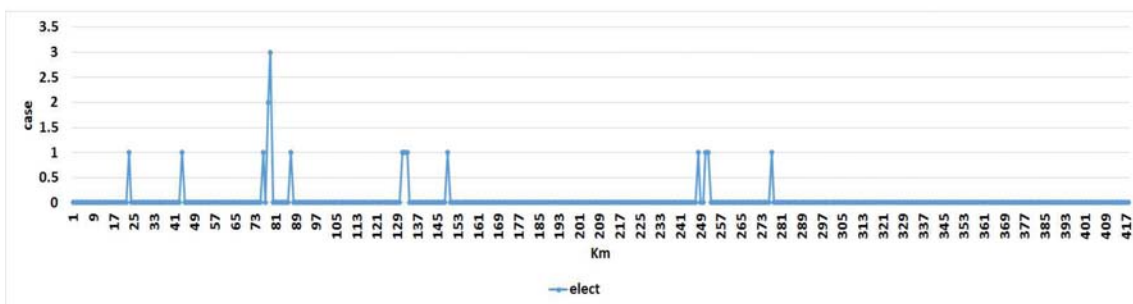
본 연구의 데이터 매핑 방법을 검증하기 위해서 2.2.4 시공간 데이터 매핑결과를 기준으로 경부고속선에서 공통적으로 발생하는 결합구간이 차량과 지상시설이 결합되면서 발생하는 장애가 특정구간에서 반복되고 있음을 확인하는 것으로 판단할 수 있었다. 매핑방법을 적용하여 분석한 결과 동일노선 분야별 데이터는 Fig.3과 같이 시설, 차량, 전기분야의 결합데이터를 추출하였고, Fig.2와 같이 경부고속선 상행을 기준으로 40~60km, 180~200km, 220~240km, 320~340km 구간에서 차량, 시설, 분야의 복합장애가 실제 발생하였던 것으로 나타났다.



(a) railway vehicle



(b) railway facilities



(c) railway electric

Fig. 2. Graphs before mapping for each field.

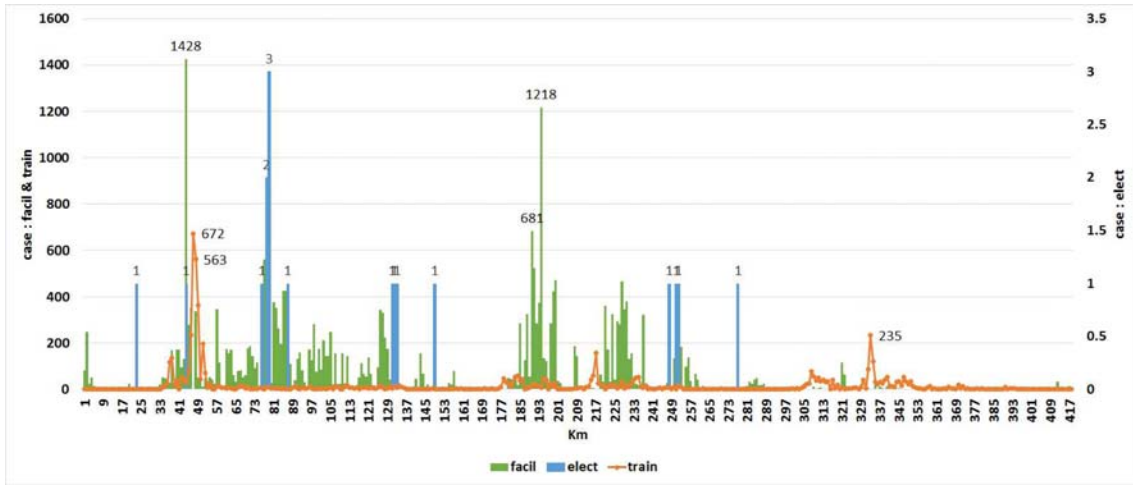


Fig. 3. A graph after mapping for each fields.

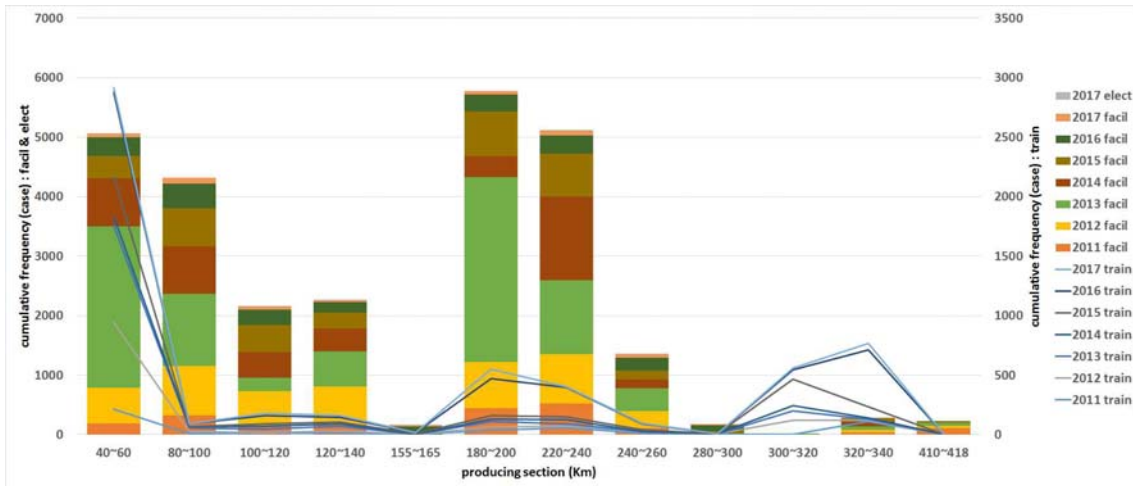


Fig. 4. Results of pattern analysis for each field on Kyungbu high speed line.

둘째로 80~100km, 100~120km, 120~140km, 240~260km 구간에서는 대부분 시설분야 결함과 일부 차량분야 결함이 동일 시점에 발생하였고, 그 외 구간은 분야별 독립적인 유형으로 장애가 발생하는 패턴을 확인할 수 있었다. 특히, 복합적으로 발생하는 40~60km구간의 경우는 선로조건이 급곡선, 구배변화(변곡점), 설빙 발생에 의한 레일표면 결함 등 복합적 장애현상을 연차별 결함보고서 검토결과를 통하여 확인할 수 있었다. 특히 40~60km, 180~200km, 220~240km, 320~340km 구간의 경우 차량, 시설분야의 밀접한 상호연관성을 확인할 수 있었다.

Fig. 4는 2011년 이후 동일 구간에서 장애 빈도가 반복되고 있는 정도를 확인하기 위해 연차별 결함 및 장애 요인에 따라 반복되고 있음을 분야별 그래프 중첩을 통하여 보여주고 있다. Fig. 3의 결과를 구간별로 묶어서 특정 구간에 대하여 연차별, 분야별 결함을 확인하였다. 그 결과 그래프에서 차량, 시설분야가 상호연관성을 가지고 선로조건이 취약한 경우에 대차현상이 함께 나타나는 것을 확인할 수 있었다. 이와 같이, 40~60km 외 복합적으로 장애 및 불량이 발생하는 구간에 대하여 동일하게 검증 실시하였고, 그 결과 실제로 각 분야의 유지보수 이력과 전문가 분석 그리고 현장조사를 통하여 패턴분석 결과를 확인할 수 있었다. 경부고속선 상행을 대상으로 40~60km, 180~200km, 220~240km, 320~340km 구간에 차량, 시설, 전기분야의 전문가 협조를 받아 검증을 실시하였다. 그 결과, 40~50km의 경우, 평면곡선이 연속으로 발생하고 구배변화(변곡점)도 차이가 많이 나는 구간이었다. 또한, 2013년 3월에 설빙 발생에 의한 레일표면결함이 집중적으로 발생하였고 2014년에 하반기 폭설주의보 발령 시, 감속운행 등 선제적 안전관리 시행으로 레일표면결함 발생량이 급격히 감소됨을 확인하였으며, 차량에서는 이로 인한 대차현상이 많이 발생하게 되었다. 180~200km의 경우, 선로조건이 평면곡선(원화곡선)이 나타나는 구간으로 구배가 연속으로 발생하여 콘크리트 도상에서 자갈도상으로 접속하는 구간이며, 40~50km구간과 마찬가지로 같은 시기에 설빙이 발생하였다. 그



래서 시설적 결함과 대차현팅 발생 가능성이 큰 구간으로 판단할 수 있다. 220~240km의 경우, 평면곡선이 두 번 발생하고 구배변화(변곡점) 역시 두 번 나타났다. 이는 분기기 사이에 신축부 형태의 구조로 열차의 연속 통과에 따른 진동발생의 잠재적 요인을 내재하고 있다. 마지막으로, 320~340km의 경우도 평면곡선이 두 번, 구배변화(변곡점) 두 번 나타나는 구간으로 건천연 결선 분기기 및 신경주역 구내 분기기 구간으로 열차 진동이 많이 발생하는 구간으로 확인되었다. 실제 해당 구간의 현장 담당 자들도 시설과 차량의 결함이 충분히 발생할 수 있다는 의견을 수렴할 수 있었다.

### 2.3.2 차량과 지상시설의 복합장애 실측 결과

2017년 2.10~5.31일까지 경부 고속선 상선에서 실측한 구간의 결함개소와 비교해 본 결과, KTX차량의 대차현팅 검지정보와 궤간틀림, 고저틀림, 전차선 편위불량 등의 복합 결함이 다음 그림 Fig. 5와 같이 경부고속선 상행을 기준으로 20~90km구간에서 유사하게 측정되고 있음을 확인 할 수 있었다.

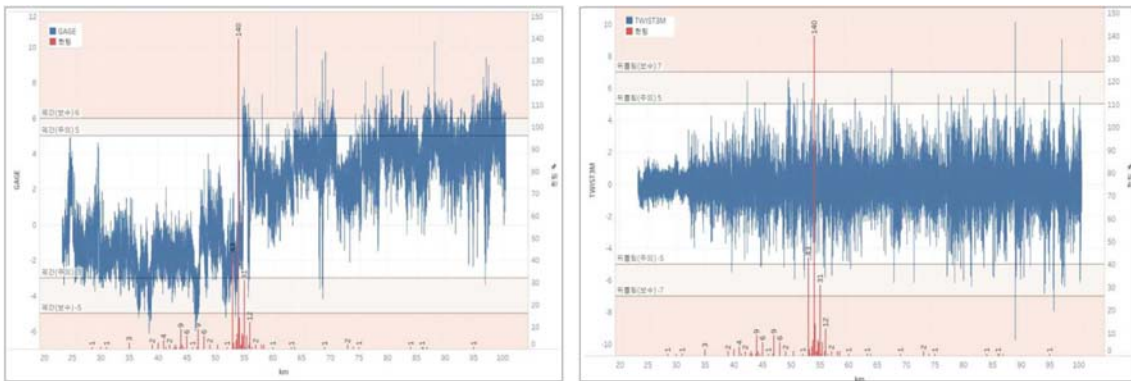


Fig. 5. Data measurement results by failure factor from rolling stock and facilities(Kyung-bu line 20~90km).

## 3. 결 론

본 연구의 최종 목적은 선제적 안전관리 실현을 위해 실시간 운영정보를 통합·분석하여 이상상황 판단, 예측 및 조기대응 할 수 있는 데이터 기반 위험관리 분석시스템을 구축하는 것이다. 사고는 시점, 장소, 원인을 예측할 수 없다. 그러나 분야별로 상이한 기준 데이터 매핑을 통해 위험관리 데이터의 시·공간기반 통합분석이 이루어지도록 하여 위험분석 알고리즘 구현한다면 위험관리 분석시스템을 구축할 수 있을 것이다. 기존에는 해당 시설의 상태에 따른 열차의 결함요인을 파악할 수 없었고, 과거에 유지보수 전과 후에 열차에서 발생하는 결함요인 역시 확인할 수 없었고 한국철도공사 내 전사자원관리시스템(KOVIS) 데이터는 크게 차량·시설·전기 세 분야로 각자의 데이터가 관리되어 운영 및 유지보수 되어왔다. 철도 위험관리를 위해서는 분야간 위험요인들이 상호연관성이 있는지에 대한 분석이 필요하고, 현재 분야별 데이터 관리는 되고 있으나 통합적으로 파악할 수 있는 경로의 한계가 존재했다. 이러한 문제점을 해결하기 위하여 분야별 데이터를 매핑하여 상호연관성을 파악하고 패턴분석을 실시하였다. 이를 위해 차량, 시설, 전기 각 철도분야별 연관성이 있는 데이터를 식별하여 데이터 유형 및 구조를 검토하였고, MS-SQL를 활용하여 데이터베이스를 구축하였다. 또한, 각 데이터별 발생시간, 노선구분, 노선종류, 상·하구분, 기능위치 등 관련 기준 데이터를 추출하였고 R 프로그램 내 melt 및 cast함수를 활용하여 통합 변경·처리하였다. 기존 데이터 기반의 통합된 데이터를 바탕으로 ggplot2 패키지를 사용하여 x축을 km로 표기하고 y축을 년/월/일별 발생 빈도로 누적그래프를 도출하여 시·공간적 분야별 결함발생 패턴을 도출하였다. 이 결과 값을 토대로 각 분야의 유지보수 이력과 전문가 검토, 그리고 현장조사를 통하여 결함이 복합적으로 발생하는 구간에 대한 선형조건, 취역사항 등을 검토하여 분야간 인터페이스 위험요인에 관한 상호연관성을 도출할 수 있었으며, 이는 각 분야별로 구축된 안전관련 데이터를 이용하여 통합분석 시 발생할 수 있는 문제점을 해결함으로써 시·공간기반 위험관리 데이터 분석의 기반을 구축한 것에 의미를 둘 수 있다. 추후 전국 96개 노선으로 확대 적용하여 주요 위험구간의 랭킹을 관리하고, 해당 구간에 사전 유지보수와 열차 위험요인에 대하여 선제적 안전관리가 가능하다.

이후 본 연구에서 적용한 데이터 매핑을 기초로 기관사 운행정보기록, 기상정보 등 데이터 확장을 통하여 시·공간기반 통합 분석 결과의 신뢰성을 더욱 높일 수 있도록 해야 할 것이다. 이러한 매핑 작업은 실시간 위험관리를 위한 위험분석 알고리즘 개발에 기초로 활용될 수 있음을 확인하였다. 이와 같이 각 분야별 데이터 정보에 대한 상호연관관계를 분석하여 데이터들의 상호관련성, 패턴, 규칙 분석을 할 수 있도록 데이터마이닝을 한다면 선제적 안전관리 의사결정을 지원해 줄 수 있는 실시간 위

험관리 분석시스템이 구축될 것이다.

## References

- [1] N. Attoh-Okine (2014) Big data challenges in railway engineering, *IEEE International Conference on Big Data*, Washington DC, pp. 7-9.
- [2] H. Devid, K. Jason (2013) Big data and railroad analytics, *Newsletter of the Railway Applications Section*, p. 12.
- [3] Y. Ko, T. Seo (2005) A Study on the naming Rules of metadata based on ontology, *Journal of the Korean Society for Information Management*, 22(4), pp. 97-109.
- [4] T. Seo, D. T. Pham (2007) Data modeling process to ensure semantic interoperability of data, *Information Management Research*, 37(1), pp. 59-73.
- [5] H. Kim, Y. Jeong, H. Kang, D. Park (2016) Data mapping and development guideline for PHR system establishment using CMD, *Journal of Korea Institute of Information Technology*, pp. 133-141.
- [6] C. Pluempitwiriyawej, J. Hammer (2000), A classification scheme for semantic and schematic heterogeneities in XML data sources, Technical Report TR00-004, University of Florida, Gainesville, FL, pp. 36.
- [7] J. Han, M. Kamber (2011) Data Mining: Concepts and Techniques (Third Edition), *Burlington, MA: Morgan Kaufmann*.
- [8] U. Fayyad, G. Piatesky-shapino, P. Smyth (1996) From data mining to knowledge discovery in databases, *AI magazine*, 17(3), pp. 37-54.
- [9] S.D. Nelson (2016) Informatics and interoperability: Speaking the same language, *ISPOR 20<sup>th</sup> Annual International Meeting*.
- [10] Korea Data Agency (2016) Data Quality Management Guidelines ver 2.1.

(Received 4 September 2017; Revised 20 September 2017; Accepted 29 September 2017)

---

**Hyun-Jin Byun:** bhj@korail.com

Safety Innovation Headquarters, Korea Railroad Corp., 240, Jungang-ro, Dong-Gu, Daejeon 34618, Korea

**Yong-Sang Lee:** yongsang@wsu.ac.kr

Department of Railroad Management, Woosong University, 171 Dongdaejon-ro, Dong-Gu, Daejeon 300-718, Korea