

Automatic Algorithm for Cleaning Asset Data of Overhead Transmission Line

가공송전 전선 자산데이터의 정제 자동화 알고리즘 개발 연구

Sung-Duk Mun, Tae-Joon Kim, Kang-Sik Kim, Jae-Sang Hwang

Abstract

As the big data analysis technologies has been developed worldwide, the importance of asset management for electric power facilities based data analysis is increasing. It is essential to secure quality of data that will determine the performance of the RISK evaluation algorithm for asset management. To improve reliability of asset management, asset data must be preprocessed. In particular, the process of cleaning dirty data is required, and it is also urgent to develop an algorithm to reduce time and improve accuracy for data treatment. In this paper, the result of the development of an automatic cleaning algorithm specialized in overhead transmission asset data is presented. A data cleaning algorithm was developed to enable data clean by analyzing quality and overall pattern of raw data.

Keywords: Asset Data, Asset Management System, Data Cleaning, Legacy System, Missing Data Treatment, Outlier Data Treatment, Preprocessing, Overhead Transmission Lines

I. Introduction

빅데이터 분석 기술은 전 세계적으로 활발하게 이용되고 있으며 빅데이터를 활용한 플랫폼 등의 새로운 가치가 창출되고 시장 규모가 점차 확대될 것으로 전망된다. 전력분야에서도 빅데이터 분석을 통해 에너지 수요와 전력망 공급관리, 설비의 고장 및 정전 예측, 유지보수 예산의 최적화, 설비운영 효율성 향상 등에 활용 중에 있다. 또한 전력설비에 대한 데이터 정보 기술의 발전으로 인공지능 기반의 설비의 다양한 상태 평가, 오류와 예측 분석에 활용되고 있으며 데이터 품질 향상을 위한 알고리즘도 적용 중이다 [1][2].

현재 한국전력공사에서는 고장확률 산출 및 고장파급의 경제성 평가를 복합적으로 고려한 자산관리를 도입 및 구축하고 있으며, 이는 설비 운영 시 발생하는 다양한 빅데이터를 과학적 기반에 근거하여 설비의 생애주기를 관리하는 방식인 자산설비 관리 기법이다. 한국전력공사에서는 운영, 진단, 고장 등 다양한 시스템을 통해 데이터를 수집하고 있으며 이를 Legacy 시스템이라 한다. 자산관리를 위해 각 Legacy 시스템으로부터 데이터를 수집하고 통합하여 별도의 자산관리시스템으로 연계 및 구축이 필요하다. 한국전력공사에서는 송전, 변전, 배전설비를 대상으로 자산관리시스템 구축을 위해 설비별 고장확률 알고리즘 개발을 진행 중이다. 고장확률 알고리즘은 데이터가 기반이 되기 때문에 데이터가 누락되거나 이

상이 의심된다면 이를 추려낼 뿐 아니라 전처리 과정이 반드시 수반되어야 한다. 기존에는 사람이 수십만 개의 데이터를 일일이 눈으로 확인하는 방법을 이용하였는데 이는 데이터의 양이 많고 복잡하기 때문에 오랜 시간이 소요될 뿐 아니라 인력 낭비가 되며 수작업으로 확인하는 과정으로 진행되기 때문에 실수가 발생할 여지가 있다. 또한 데이터는 지속적으로 수집되므로 새로 수집된 데이터에 대한 전처리 작업이 행해져야 한다 [3].

가공송전 전선 관련 자산데이터로는 송전운영시스템(TOMS)으로부터 수집한 운영데이터와 송변전통합설비관리시스템(STOM)으로부터 취득한 점검, 진단데이터 등이 있다. 본 논문에서는 자산데이터의 전처리를 통한 신뢰성 있는 데이터 확보의 중요성에 따른 한국전력공사에서 운용중인 가공송전 전선에 특화된 데이터 자동 정제 알고리즘 개발 연구를 수행하였다. 가공송전 전선 자산관리에 필요한 자산데이터를 대상으로 결측 및 이상데이터에 대한 정제 처리기법을 제시하였다.

II. Trend of Data Preprocessing Technology

빅데이터를 보유하고 있는 글로벌 기업들은 빅데이터 수집, 저장관리와 전처리까지 다방면에 걸쳐 기술을 주도하고 있으며 정

Article Information

Manuscript Received January 13, 2021, Revised January 21, 2021, Accepted February 07, 2021, Published online June 30, 2021

The Authors are with KEPCO Research Institute, Korea Electric Power Corporation, 105 Munji-ro Yuseong-gu, Daejeon 34056, Republic of Korea.

Correspondence Author: Sung-Duk Mun (moonsd1640@kepcoco.kr)

ORCID: 0000-0002-7394-5318 (J. S. Hwang)



This paper is an open access article licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Public License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0>
This paper, color print of one or more figures in this paper, and/or supplementary information are available at <http://journal.kepcoco.kr>.

TABLE 1

Data Classifications and Detailed Data Lists in Overhead Transmission Line		
Classification	Detailed data lists	Number of data lists
Line information	Number of circuits, Length of line, Number of towers, etc.	48
Circuit information	Construction date of circuits, Length of circuit, Rated current, etc.	46
Phase conductor information	Span, Rated voltage, Type of conductor, Cross section, Number of subconductors, etc.	52
Tower information	Type of towers, Pollution class, Wind pressure, etc.	118
Ground conductor information	Span, Type of conductor, Cross section, etc.	55
Detailed inspection information	Judgement result in inspection, Condition of conductor strands and fittings, etc.	43
Infrared measurement information	Inspected circuit, Temperature of conductors, Temperature of joints. etc.	38

TABLE 2

Status of Missing on the Asset Data in Overhead Transmission Line		
Classification	Number of asset data (Approximately)	Percentage of missing data (Approximately)
Line information	70,000	0%
Circuit information	100,000	25%
Phase conductor information	9,000,000	25%
Tower information	1,800,000	15%
Ground conductor information	1,300,000	10%
Detailed inspection information	5,300,000	5%
Infrared measurement information	12,000,000	5%

제 및 통합이 가능한 툴을 개발하여 활용 중에 있다.

설비운영 시스템과 같은 빅데이터시스템에는 다양한 형태의 정보를 포함한 데이터가 지속적으로 축적되고 있으며 그 중 분석에 필요한 데이터를 활용하기 위해서는 데이터의 전처리 과정이 반드시 필요하다. 데이터의 전처리 단계는 일반적으로 데이터의 여과(filtering), 데이터의 형태 변환(transforming), 그리고 정제(cleaning) 과정으로 구분되며 최종적으로 데이터의 형태로 통합(integration)되어 활용된다 [4].

데이터의 정제는 원시 데이터(raw data) 중 오염된 데이터를 정제된 데이터로 처리하는 것을 의미한다. 원시 데이터 내에는 오류 데이터(wrong data)와 결측 데이터(missing data) 등 다양한 유형의 오염 데이터(dirty data)가 존재할 수 있으므로 분석용 데이터로 변환하기 위해서는 원시데이터를 가공하는 정제가 필요하다. 오염된 데이터의 정제 방법은 회귀모형, 전문가 자문, 추정치 대체 등의 다양한 방법이 있으며 오염된 데이터의 유형에 맞는 정제 방법을 적용하여야 한다. 이를 위해서는 원시데이터로부터 오염데이터를 분류하고 전반적인 데이터의 패턴 분석을 통해 정제 값을 판단하고 검증하는 방법으로 진행되어야 한다 [5].

기존에는 데이터의 전처리 작업이 사람에 의한 수작업으로 진행되어 상당한 시간이 소요되고 실수나 의도하지 않은 잘못된 정보가 기입되는 등의 문제가 존재하였다. 또한 시스템에는 새로운 데이터가 생성되어 데이터의 양은 지속적으로 증가하는데 그때마다 정제를 위한 전처리 작업이 행해져야 한다. 정제 작업의 작업 소요 기간은 전처리 과정 중 약 60%가 넘는 것으로 나타났으며, 작업 소요 시간 뿐만 아니라 작업강도, 지루한 작업 진행 방식도 문제로 존재한다. 그때마다 정제를 위한 전처리 작업이 행해져야 하는데 시간 단축과 작업 신뢰성, 그리고 데이터의 품질 향상을 높일 수 있도록 데이터의 정제 자동화가 필요하다 [6].

III. Analysis of Asset Data in Overhead Transmission line

A. 가공송전 자산데이터 현황

가공송전 전선 자산관리에 필요한 데이터 확보를 위해 4개의

레거시시스템으로부터 데이터 수집을 진행하였다. 레거시시스템은 송전운영시스템(TOMS), 송변전통합설비관리시스템(STOM), 변전소 운전실적관리시스템(SOMAS), 정전고장관리시스템 등이다. 그 중에 TOMS는 송전설비의 운영정보를 관리하는 시스템이며 STOM은 설비의 점검, 진단을 통한 상태데이터를 관리하는 시스템이다. 본 연구에서는 가공송전 전선 자산관리시스템 구축에 필요한 주요 자산데이터를 TOMS와 STOM시스템을 통해 조사 및 수집하였다. 데이터 상세 수집 항목은 TABLE 1과 같이 송전선로 전반적인 현황 및 전력선 자산설비 관련 데이터 항목이 존재한다.

B. 오염데이터 탐색

상기 가공송전 전선 자산데이터의 현황에서 알 수 있듯이 송전운영시스템 항목만으로도 데이터의 항목은 다양한 것으로 확인되었으며 이외 부하, 고장 등의 데이터를 포함한다면 다뤄야 할 데이터 수량은 수천만개에 다다르며 수집하여 검사하는 데만 상당한 소요시간이 걸릴 것으로 예상된다. 이러한 Legacy 데이터를 자산관리시스템에 연계하여 활용하는 것이 가장 큰 목적인데 사전에 필요 사항으로 데이터의 품질 분석이 이뤄져야 한다. 따라서 수집된 데이터에 대해 데이터 프로파일링 기법 및 데이터 시각화 기법을 활용하여 오염탐색을 수행하였다. 데이터 프로파일링은 데이터에 대한 요약 통계로 품질을 전반적으로 이해하는데 용이하며 데이터 값이 어떤 표준이나 패턴을 갖는지 확인하는 방법이다. 또한 데이터 시각화 방법은 평균, 표준편차, 데이터 값의 범위, 분위수와 같은 통계적 방법을 이용하여 시각적으로 표현하는 방법이다.

C. 오염데이터 현황

가공송전 자산데이터 중 오염으로 판단되는 데이터는 “결측 데이터”, “오기입 의심데이터”, “중복 데이터” 3가지 유형으로 분류할 수 있다. 결측데이터는 데이터 항목별로 그 편차가 다르지만 시스템에 데이터를 직접 입력하는 방식이기 때문에 미입력으로 인해 누락된 데이터와 어떤 자산에는 해당되지 않아 입력되지 않는 데이터 항목도 존재하는 것으로 판단된다. TABLE 2는 가공송전 전선 관련으로 취득한 자산데이터를 대상으로 미 입력된 결측데이터 현황을 나타냈다. 오기입 의심데이터는 결측 데이터와 마찬가지로 직접 입력하는 과정에서 발생하는 것으로 판단된다. 중복데이터는 전력선, 지지물 등 동일한 자산 설비가 1회 이상 반복 입력되어 시

Item	Circuit Name	Line Segment Code	Tower (Start)	Tower (End)	Rated Voltage [kV]	Type of Conductor	Cross Section (㎡)	Number of Conductors	Date of Installation
Raw Data	A-B #1	BTC0000001	No.1	No.2	345	ACSR	480	4B	2011.11.11
	A-B #1	BTC0000002	No.2	No.3	Missing	Missing	Missing	Missing	2011.11.11
	A-B #2	BTC0000003	No.1	No.2	345	ACSR	480	4B	2011.11.11
	A-B #2	BTC0000004	No.2	No.3	345	ACSR	480	4B	2011.11.11
Clean Data	A-B #1	BTC0000001	No.1	No.2	345	ACSR	480	4B	2011.11.11
	A-B #1	BTC0000002	No.2	No.3	345	ACSR	480	4B	2011.11.11
	A-B #2	BTC0000003	No.1	No.2	345	ACSR	480	4B	2011.11.11
A-B #2	BTC0000004	No.2	No.3	345	ACSR	480	4B	2011.11.11	

Fig. 1. Missing data cleaning algorithm example for conductor information.

Item	Circuit Name	Rated Voltage	Tower Code	Tower Segment NO.	Pollution Class	Class of Wind Pressure	Date of Installation
Raw Data	A-B	154	BTA0000004	No.1	A	2	2011.11.11
	A-B	154	BTA0000005	No.2	Missing	2	2011.11.11
	A-B	154	BTC0000006	No.3	A	2	2011.11.11
Clean Data	A-B	154	BTA0000004	No.1	A	2	2011.11.11
	A-B	154	BTA0000005	No.2	A	2	2011.11.11
	A-B	154	BTC0000006	No.3	A	2	2011.11.11

Fig. 3. Missing data cleaning algorithm example for pollution class.

Item	Circuit Name	Rated Voltage	Line Segment Code	Tower (Start)	Tower (End)	Type of Conductor	Date of Installation
Raw Data	A-B #1	154	BTC0000001	No.1	No.2	ACSR	2011.11.11
	A-B #1	154	BTC0000002	No.2	No.3	ACSR	2011.11.11
	A-B #1	154	BTC0000003	No.3	No.4	ACSR	Missing
Clean Data	A-B #1	154	BTC0000001	No.1	No.2	ACSR	2011.11.11
	A-B #1	154	BTC0000002	No.2	No.3	ACSR	2011.11.11
	A-B #1	154	BTC0000003	No.1	No.2	ACSR	2011.11.11

Fig. 2. Missing data cleaning algorithm example for installation date.

시스템에 중복으로 등록된 경우로 판단된다. 이러한 오염데이터는 결측데이터를 먼저 정제 후 이상데이터 처리가 필요하다.

IV. Cleaning Algorithm of Asset data in Overhead Transmission line

A. 결측데이터 처리 알고리즘

가공송전 전선 자산데이터는 결측데이터의 비중이 크에 따라 이러한 결측데이터에 대한 정제가 필수적이다. 가공송전 전선 자산 설비에 대한 데이터 수는 상당한 양을 보유하고 있어 사람이 직접 정제하기 위해서는 오랜 기간이 소요되는 부담이 있어 정제를 위한 작업 소요기간이 단축되어야 한다. 따라서 신속하고 정확하게 데이터를 정제 가능한 알고리즘 개발을 수행하였다.

한국전력공사는 송전선로 전체를 하나의 자산으로 설정하는 것이 아니라 선로를 구간 단위로 구분하여 관리하고 있으며 각 구간에 해당하는 전력선에 대해 데이터 정보를 입력하여 관리하고 있다.

따라서 선로의 일부인 구간별 전력선은 공통적으로 선로 정보를 공유하게 된다. 즉, 선로 내 구간에 해당되는 전력선은 공통 사항인 동일한 정보를 포함하게 된다. 이러한 특성을 반영하여 어떤 구간에 일부 결측데이터가 존재했을 때 공통사항이 있는 경우 동일한 정보를 채우는 방식으로 정제가 가능하며 이를 통해 기존 대비 결측률을 효과적으로 낮추는 결과를 얻을 수 있다.

이와 같이 설비에 공통적으로 저장된 정보를 통해 결측 구간에 대한 답을 추적하는 것처럼 정해진 규칙기반의 결측데이터 정제 알고리즘을 개발하였다.

1) 전력선 정보 [전압, 선종, 도체 단면적, 도체수]

전력선의 전압, 선종, 도체 단면적, 도체수는 수기방식으로 입

력됨에 따라 동일한 회선 내에 구간 정보가 동일한 정보로 입력되어야 함에도 불구하고 명칭이나 타입을 달리 입력하여 오탈자가 존재하거나 누락된 데이터가 존재하였다.

Fig. 1은 A-B #1 회선 및 A-B #2 회선 2개로 구성되어 운용 중인 선로로, A-B #1 회선 중 No. 2~No. 3 구간에 대한 전압, 선종, 도체 단면적 및 도체수 4가지 전력선에 대한 정보가 누락된 사례다.

A-B #1 회선의 No. 1~No. 2 구간과 A-B #2 회선의 No. 1~No. 2 구간 및 No. 2~No. 3 구간을 확인해 보니 구간으로 구분은 되어 있으나 공통 정보 사항으로 데이터가 존재하는 것을 확인하였다. 만약 A-B #1 회선의 No. 2~No. 3 구간이 변경되거나 없어진 경우라면 본 시스템인 TOMS에 해당 전력선 구간에 대한 정보는 완전히 제거되어야 한다. 또한 A-B 선로의 경우 1회선 운용 선로가 아닌 2회선 운용 중인 선로로서 A-B #1 회선의 1개 구간을 제거한다면 #1 회선은 계통 연계가 되지 않는 상태가 되는 것이다. 또한 시스템 상 A-B #1 No. 2~No. 3 구간의 전력선 설비코드 데이터가 존재하므로 해당구간의 정보는 제거된 것이 아니라 입력되지 않고 누락된 결측데이터로 판단된다.

A-B #1 회선과 A-B #2 회선은 설치날짜가 일치하는데 이는 일괄 설치한 경우이므로 결측데이터는 다른 구간 정보를 통해 정제 가능하다. 결측데이터 정제결과 전압(kV) "345", 선종 "ACSR", 도체 단면적 "480" 그리고 도체수 "4B"로 처리 가능하다.

2) 전력선 설치일자

전력선은 지지물 시/종점 구간으로 구분하여 관리하는데 신규건설 또는 교체로 인해 일부구간의 설치일자가 달라질 수 있으나 대부분 선로의 전력선 설치일자는 동일하게 적용하고 있다. 이러한 전력선 설치정보는 운영년수(age)를 알 수 있어 설비 자산관리 관점에서 중요한 데이터로서 잘못된 정보가 입력되는 오류가 발생하지 않도록 해야 하나, 시스템에 수기로 입력을 하는 과정에서 Fig. 2와 같이 누락된 사례가 존재한다. A-B #1 No. 3~No. 4 구간에 대한 설치날짜 데이터가 누락된 사례로 전체 회선의 패턴을 분석하여야 하며 대부분 나머지 회선 내 구간 정보를 통해서 데이터 정제가 가능하다.

3) 지지물 오손등급/풍압 정보

지지물 오손등급은 지지물이 설치되는 지역에 대한 등가염분 부차밀도에 따른 청정구간, 그리고 A~D로 등급으로 분류되며 이는 환경정보를 반영하는 데이터로서 해당 지지물에 가선되는 전력선 도체에도 동일하게 적용하여야 하는 중요한 정보라 할 수 있다.

일반적으로 345 kV 이상의 가공 송전선로는 100기에 다다른 지지물로 구성되며 300기 이상의 지지물을 포함하는 장거리 선

Item	Circuit Name	Line Segment Code	Tower (Start)	Tower (End)	Phase	Type of Conductor	Date of Installation
Raw Data	A-B #1	BTC0000001	No.1	No.2	Top	ACSR	2011.11.11
	A-B #1	BTC0000001	No.1	No.2	Top	ACSR	2011.11.11
	A-B #1	BTC0000002	No.1	No.2	Middle	ACSR	2011.11.11
	A-B #1	BTC0000002	No.1	No.2	Middle	ACSR	2011.11.11
	A-B #1	BTC0000003	No.1	No.2	Bottom	ACSR	2011.11.11
Clean Data	A-B #1	BTC0000001	No.1	No.2	Top	ACSR	2011.11.11
	A-B #1	BTC0000001	No.1	No.2	Top	ACSR	2011.11.11
	A-B #1	BTC0000002	No.1	No.2	Middle	ACSR	2011.11.11
	A-B #1	BTC0000002	No.1	No.2	Middle	ACSR	2011.11.11
	A-B #1	BTC0000003	No.1	No.2	Bottom	ACSR	2011.11.11

Fig. 4. Outlier data cleaning algorithm example for conductor phase.

Item	Circuit Name	Line Segment Code	Tower (Start)	Tower (End)	Phase	Type of Conductor	Date of Installation
Raw Data	A-B #1	BTC0000001	No.1	No.2	Top	ACSR	2011.11.11
	A-B #1	BTC0000002	No.2	No.3	Top	ACSR	2011.11.11
	A-B #1	BTC0000003	No.3	No.4	Top	ACSR	2011.11.11
	A-B #1	BTC0000004	No.4	No.6	Top	ACSR	2011.11.11
	A-B #1	BTC0000005	No.6	No.7	Top	ACSR	2011.11.11
	A-B #1	BTC0000006	No.7	No.8	Top	ACSR	2011.11.11
Clean Data	A-B #1	BTC0000001	No.1	No.2	Top	ACSR	2011.11.11
	A-B #1	BTC0000002	No.2	No.3	Top	ACSR	2011.11.11
	A-B #1	BTC0000003	No.3	No.4	Top	ACSR	2011.11.11
	A-B #1	BTC0000004	No.4	No.5	Top	ACSR	2011.11.11
	A-B #1	BTC0000005	No.5	No.6	Top	ACSR	2011.11.11
	A-B #1	BTC0000006	No.6	No.7	Top	ACSR	2011.11.11

Fig. 5. Outlier data cleaning algorithm example for tower start/end of conductor.

로 존재하기도 한다. 1개의 선로 내 지지물에 대한 구간별 오손 등급은 설치되는 지역에 따라 적용하므로 대체적으로 동일한 등급으로 산정된다. Fig. 3과 같이 154 kV A-B 선로 No. 1~No. 3 지지물 구간 중 No.2와 같이 일부 구간에 대한 지지물 오손등급 데이터가 누락될지라도 No. 1 및 No. 3와 같은 동일 선로에 대한 정보를 확인함으로써 결측 정보를 정제 처리가 가능하다.

지지물에 대한 오손등급 외에 설치 지역의 환경정보로서 풍압정보가 존재하는데, 공기 중에 노출돼 설치되는 가공송전 전력선에 대해 오손등급과 함께 대표적인 환경정보로서 결측데이터가 존재하는 경우 오손등급 정제와 동일하게 정제 처리가 가능하다.

B. 이상데이터 처리 알고리즘

1) 전력선 상 정보

전력선 상 정보는 보통 “상(Top)”, “중(Middle)”, “하(Bottom)”로 명칭하여 3상 시스템으로 구성하며 데이터 입력은 상-중-하 순서대로 입력하여 관리되고 있다. 전체 상별로 개수는 서로가 일치해야 하나, 오기입 또는 데이터 중복하여 입력하는 경우 개수의 차이가 발생할 수 있다. 전구 선로에 대한 전력선 상 정보를 분석한 결과 “상”, “중”, “하” 개수가 일치한다는 것을 확인하였으나 일부 데이터에서 중복 기입된 사례가 존재하는 것을 확인하였다. 설비등록 시스템에 전력선 설비 자산을 등록할 때 설비코드가 부여되는데 자산 설비당 코드가 달라야 하나, Fig. 4을 보면 A-B #1의 No. 1~No. 2 구간의 상 정보가 “상”, “상”, “중”, “중”, “하”, “하”로 데이터가 입력되었으며 각 상별로 동일한 코드를 중복 등록한 것으로 판단된다. 따라서 동일한 코드가 부여된 자산의 존재여부를 추적하고 시스템 상에서 제거될 수 있도록 관리되어야 한다.

2) 전력선 지지물 구간 전/후 정보

전력선은 지지물 전과 후로 구간을 구분하는데 보통 규칙을 기반으로 지지물 전과 후를 “1(左上)”, “2(左上)”과 같은 형태로 표기하며 표기 숫자의 순서대로 정보를 입력하여 관리한다. Fig. 5을 보면 전력선 구간 No. 3~No. 4이 존재하며 다음 구간으로 순서상 No. 4~No. 5 으로 입력이 되어야 하나, No. 4~No. 6 으로 입력되어 이는 오기입 의심이 되는 사례가 된다. 대부분의 선로에서 구간별 순서대로 관리하고 있으므로 회선 전체적으로 순서 패턴을 확인하여 이상여부를 판단할 필요가 있으며 순서 정제가 필요한 구간 이후부터 정제하여야 한다.

Item	Circuit Name	Rated Voltage	Tower Code	Tower Segment NO.	Pollution Class	Class of Wind Pressure	Date of Installation
Raw Data	A-B	154	BTA0000004	No.1	A	2	2011.11.11
	A-B	154	BTA0000005	No.2	B	2	2011.11.11
	A-B	154	BTC0000006	No.3	A	2	2011.11.11
Clean Data	A-B	154	BTA0000004	No.1	A	2	2011.11.11
	A-B	154	BTA0000005	No.2	A	2	2011.11.11
	A-B	154	BTC0000006	No.3	A	2	2011.11.11

Fig. 6. Outlier data cleaning algorithm example for pollution class.

3) 지지물 오손등급/풍압 정보

지지물 오손등급에 대한 이상데이터 즉 정확한 정보가 입력되어야 하나 수기로 입력하는 과정에서 오기입되어 기본정보와 다른 값이 반영된 사례가 존재한다. 앞서 지지물 오손등급에 대한 결측데이터를 정제 처리와 유사한 방법으로 이상데이터도 정제 처리가 가능하다. 지지물 오손등급은 청정등급과 A~D로 등급으로 분류되는데 선로 내 지지물에 대한 구간별 오손등급은 설치되는 지역에 따라 적용하므로 대체적으로 동일한 등급으로 산정된다. Fig. 6과 같이 154 kV A-B 선로 No.1~No.3 지지물 구간 중 No.2와 같이 일부 구간에 대한 지지물 오손등급 데이터가 다른 사례가 존재하며 이는 회선 전체 패턴을 분석함으로써 이상데이터를 추출한 후 No.1과 No.3와 같은 다른 구간에서 필요로 하는 정보와 동일한 데이터 항목을 확인하여 정제 처리가 가능하다.

지지물에 대한 오손등급과 마찬가지로 환경정보인 풍압 정보 중에도 선로 내 동일한 패턴의 데이터가 입력이 되어야 하는데 수기입력으로 인해 오기입이 되어 다른 풍압 등급이 입력된 사례가 존재한다. 이는 오손등급의 이상여부 정제 과정과 같으며 데이터 항목을 달리하여 정제가 필요하다. 우선 선로 내 전혀 다른 풍압 정보가 입력된 구간을 찾아내어 동일 선로 내 필요 데이터 항목을 확인후 정제 처리가 가능하다.

4) 전선접속개소 열화상진단 정보

한국전력공사 가공송전운영업무기준에 따르면 전선접속개소 열화상진단은 압축형인류클램프, 압축슬리브, 보수슬리브 등 전선 접속개소에 대하여 적외선 열상장비로 과열여부를 확인하는 진단 방법으로 1회/년 또는 2회/년의 점검주기로 시행된다. 진단방법은 상부, 중부, 하부에 전력선에 대한 전선과 접속점의 온도를 각각 측정하며 전선과 접속개소 간 온도차를 확인하여 진단결과 판정기

Item	Circuit Name	Line Segment Code	Tower Segment NO.	Conductor Temperature (A)	Joint Temperature (B)	Difference of Temperature (A-B)	Judgement Results
Raw Data	A-B #1	BTC0000001	No.1	30	32	2	Good
	A-B #1	BTC0000002	No.2	29	31	2	Good
	A-B #1	BTC0000003	No.3	31	60	30	Good
Clean Data	A-B #1	BTC0000001	No.1	30	32	2	Good
	A-B #1	BTC0000002	No.2	29	31	2	Good
	A-B #1	BTC0000003	No.3	31	60	29	Fault

Fig. 7. Outlier data cleaning algorithm example for infrared inspection information.

준에 따라 양부판정을 하게 된다. 양부판정은 정상, 요주의, 이상 3 가지 단계로 구분하여 판단한다. 이러한 과정을 통해 취득한 측정 온도와 온도차 그리고 판정결과를 데이터로 STOM 시스템에 입력 하는 방식으로 관리하고 있다. 이러한 자산설비에 대한 진단결과는 고품질의 데이터로서 제공되어 자산관리에 반영되어야 하며 이를 위해서는 반드시 데이터의 이상여부를 확인한 후 데이터의 정제 처리가 필수적이다.

전선접속개소에 열화상진단 데이터 분석결과 기본적으로 정제되어야 할 데이터 항목은 전선과 접속점 온도 데이터가 된다. 입력데이터를 기준으로 측정온도 중 오기입 의심데이터가 존재하는 것으로 판단된다. 이러한 이상데이터로 인해 판정결과는 “정상”인데 “이상”으로 확인되는 다른 결과를 초래할 수 있다. 이처럼 상태 판정에 미치는 데이터의 경우 시각화 방법을 통해서 데이터 전반적인 분포를 확인하는 과정이 필요하다. Fig. 7과 같이 전선온도(A)가 “31°C”이며, 접속점의 온도(B)는 “60°C”로 값을 입력하였으나 온도차는 “29°C”가 되어야 하는데 “30°C”로 오기입한 것으로 판단된다. 또한 판정결과도 “10°C” 이상이면 “이상(Fault)”으로 판정되어야 하나 “양호(Good)”으로 입력되었다. 이러한 오기입된 진단데이터는 설비의 상태판정 알고리즘 인자로 적용될 수 있으므로 측

정값과 판정결과 모두 정제를 함으로써 정확한 데이터가 적용되어야 한다.

V. Conclusion

본 논문은 송전설비의 자산관리 알고리즘 개발에 필요한 자산데이터의 품질을 분석하고 정제 알고리즘 개발 결과를 제시하였다. 가공송전 전선 관련 자산데이터를 대상으로 오염데이터를 추출하고 분류한 결측데이터와 이상데이터에 대한 정제 알고리즘을 개발하였다. 정제처리는 결측데이터 선 처리 후 이상데이터를 처리하는 알고리즘으로 구성하였다.

정제 알고리즘은 가공전선 자산데이터에 특화된 규칙기반의 데이터 자동정제 알고리즘으로 구현하였다. 이는 자동정제 시스템에 탑재하여 활용하기 위해 개발 중에 있으며 자동정제시스템을 통한 고품질의 데이터 제공이 가능할 것으로 보이고 나아가 빅데이터 기반의 자산관리 RISK 평가 알고리즘의 신뢰성 측면에서 기존 대비 향상이 가능할 것으로 기대된다.

References

- [1] Ravi V Angadi, “Role of Big Data Analytics in Power System Application,” 2020.
- [2] Betsy Barry, “Big Data Issues and Opportunities for Electric Utilities,” 2015.
- [3] JIEJIE DAI, “Cleaning Method for Status Monitoring Data of Power Equipment Based on Stacked Denoising Autoencoders,” 2017.
- [4] D. Chapman, “PRINCIPLES AND METHODS OF DATA CLEANING,” 2005.
- [5] Salvador García, “Big data preprocessing: methods and prospects,” 2016.
- [6] Fakhitah Ridzuan, “A Review on Data Cleansing Methods for Big Data,” 2019.