기계학습을 이용한 전철전력설비 이상상태 데이터 분석

신승권, 조규정, 정호성, 김형철
한국철도기술연구원 스마트전기신호본부
e-mail:skshin@krri.re.krkr

Anomaly Data Analysis of Electric Railway Facility using Machine Learning

Seung-Kwon Shin, Kyu-Jung Cho, Ho-Sung Jung, Hyung-Chul Kim
Metropolitan Transportation Research Center,
Korea Railroad Research Institute

요 약

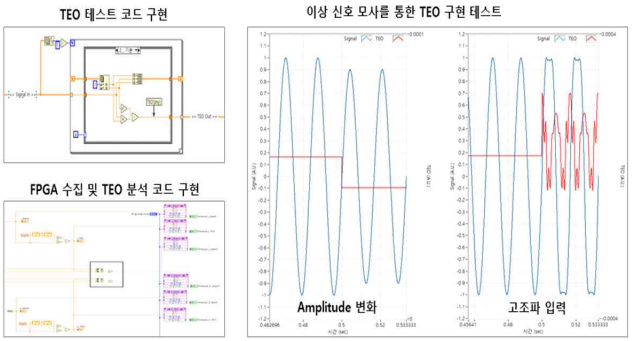
철도전기시설물의 노후화가 진행되어 다수의 전기장애가 발생하고 있으며, 2022년까지 전철거리를 4,303km(전철화율 86%) 증설할 계획이다. 2018년 기준 경부고속철도 전철분야 노후화율은 변전설비 87.6%, 전력설비 89.4%, 장애건수 76건으로 기존 인력중심의 점검방식으로는 한계가 있어 새로운 진단기법이 요구되고 있다. 철도운영기관에서는 노후시설의 장애를 사전에 탐지하고 예측하기 위해 전철전력분야의 육안검사 비중을 최소화하고, 운영 중에도 이상상태를 빠르고 정확하게 자동으로 검출하는 기술을 요구하고 있다. 본 논문에서는 순간적인 전압강하, 파형왜곡 등의 전기장애의 전조현상을 파악하기 위해 순시치 파형에 포함되어 있는 미세신호로부터 데이터마이닝 기술을 통해 데이터 모델을 추출하고 사전에 장애를 탐지하고자 한다.

1. 서론

현재 전기철도에서 전기분야 장애는 지락, 단락 등 급격한 변화의 이상신호가 나타나면 전기를 차단하여 열차운행을 중단하여 인명·재산 피해를 최소화하나, 전기장애의 전조현상을 알지 못하고 있다. 또한 철도운영기관은 전기장애 전조현상을 미리 파악하여 후속조치가 가능함으로서 열차운행 중단 없이 전력공급이 가능한 시스템을 요구 요구하고 있다. 다양한 형태로 나타나는 전기장애 전조현상은 사고 직전뿐만 아니라 몇일 전에도 나타날 수 있으므로 사전예측이 가능한 검출은 특히 어려움이 있다. 따라서 본 논문에서는 순간적인 전압강하, 파형왜곡 등의 전기장애의 전조현상을 파악하기 위해 TEO(Teasor Energy Operator)연산자와 STFT(Short Time Fourier Transform)를 이용하였다. 또한 순간적인 파형을 16.7ms이내 검출이 가능하도록 초당 최대 50k 샘플을 Linux기반 실시간 OS와 FPGA로 실시간 감지모듈을 구현하였다. 또한 순시치 파형에 포함되어 있는 미세신호로부터 데이터마이닝 기술을 통해 데이터 모델을 추출하고 예측하여 사전에 장애를 탐지하고자 한다.

2. 실시간 전력감시시스템 개발

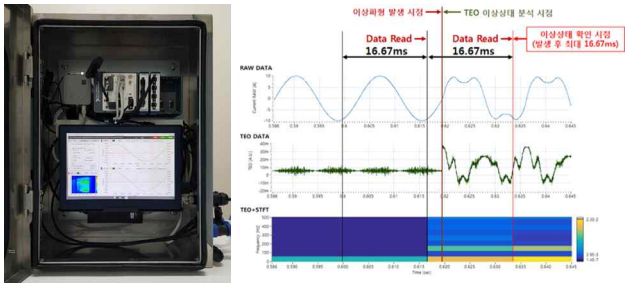
기존의 실효치 데이터 방식의 고장감시 및 보호시스템은 철도부하의 노이즈와 가선전압의 변동으로 인해 빠른 시간내에 이상신호를 분석하기 어렵다. 따라서, 개발된 전철전력 이상상태 실시간 감시시스템은 실효치 데이터가 아닌 순시치 데이터를 기반으로 동작한다. 이를 위하여 미소신호, 순간적인 전압강하, 파형왜곡 등의 이벤트를 16.7ms(1cycle) 이내에 검출하도록 TEO 연산자와 STFT분석을 적용한 감시 알고리즘을 구현하였다.



[그림 1] TEO와 STFT분석을 적용한 알고리즘 구현

그림 1은 이상상태 검출 알고리즘을 Labview로 구현 화면

이며 TEO 연산값이 이상상태 검출 역치를 초과하면 STFT 분석이 수행되어 이상상태가 발생한 시점과 주파수 분석결과를 알 수 있다. 개발된 알고리즘은 초당 최대 50,000개의 샘플을 처리할 수 있는 고속용 Linux 기반 실시간 OS와 FPGA로 구현하였으,며 그림 2은 제작된 전철전력 이상상태 실시간 감시시스템과 1 cycle이내 이상상태가 검출되고 있음을 나타낸다.



[그림 269] 실시간 전력감시시스템 제작

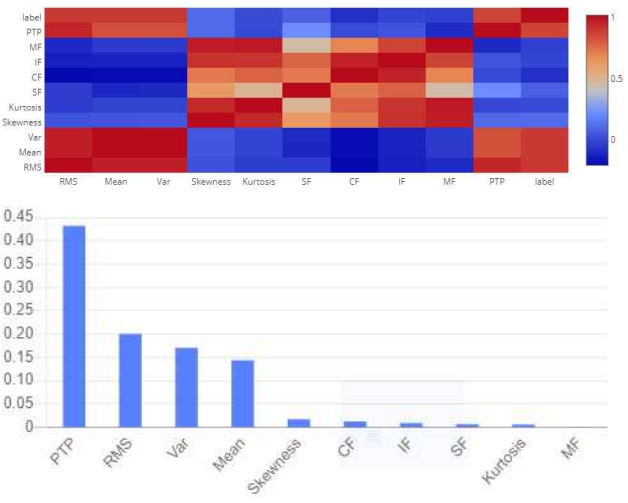
실시간 감시시스템은 동작 검증을 위하여 그림 3과 같이 오송 철도종합시험선로에 설치하였다. 설치된 감시시스템은 설정된 전압 및 전류의 TEO 연산자 값을 초과하는 이상상태가 발생하면 발생 시점을 기준으로 특정 기간 동안의 데이터를 저장하며 계산된 TEO 연산자 값에 따라 심각도를 구분하여 알람으로 운영자에게 경고한다. 알람과 기록된 데이터는 서버를 통하여 운영자(Client)가 실시간으로 감시할 수 있다.

[그림 270] 실시간 전력감시시스템 현장설치

실시간 감시시스템의 전압, 전류 측정성능 검증을 위하여 전력품질분석기인 Dewetron DEWE-571을 통해 동시간에 측정한 전압, 전류 기록 데이터와 비교를 수행하였으며, TEO값의 STFT를 통한 이상상태 시점이 기록된 특정 시간대의 데이터를 비교하였으며, 검증 결과, 종합시험선로에서 운영 중인 열차부하의 이동에 따라 급격한 전압, 전류 변동과 고조파 등이 발생하였을 때 TEO 및 STFT 연산을 적절히 수행하는 것을 확인하다.

3. 전철전력설비 이상상태 분석

본 논문에서는 실시간 전력감시시스템을 오송종합시험선로에 설치하여 전력설비의 이상상태 1667개의 Dataset을 취득하였으며 Dataset당 60,000개의 데이터로 구성된다. 측정된 데이터로부터 데이터마이닝 기법을 이용하여 4단계(관심/주의/경계/심각)로 클러스터링 군집화하고, 라벨링된 데이터를 이용하여 예측하고자 한다. 이때 고려하는 변수로는 TEO값의 RMS, Variance, Skewness, Shape factor, Crest factor 등 시간파형 데이터의 특징을 이용하였다. 측정된 Dataset으로부터 TEO값의 RMS, Variance, Skewness, Shape factor, Crest factor 등 시간파형 데이터의 특징의 상관도 분석을 수행하였다. 또한 상관도 분석의 정확도를 향상시키기 위해 스케일링을 통한 데이터의 정규화를 하였으며, overflow나 underflow를 방지하고 독립 변수의 공분산 행렬의 condition number를 감소시켜 안정성 및 수렴 속도를 높였다. 또한 Outlier값에 영향을 최소화하기 위해 상관계수가 0.95 이상인 변수를 제거하였다. 그림 4은 상관분석 결과를 나타내었다.

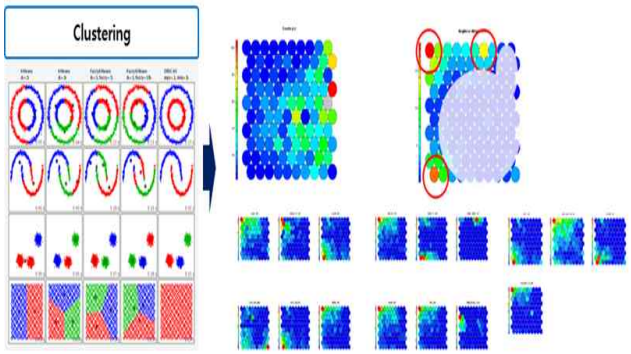


[그림 4] 상관도 분석결과

클러스터링 분석에서 클러스터 내에 속한 패턴들의 특성은 동질적이고, 서로 다른 클러스터에 속한 패턴들 간의 특성은 서로 이질적이 되도록 각각의 패턴들을 분류한다. 비슷한 특성을 가진 4등급(관심/주의/경계/심각)으로 군집화하고 각 클러스터별로 변수 통계량을 분석하여 클러스터의 특성을 알아보고자 분석을 수행하였다.

클러스터링 분석을 위한 데이터로는 관심/주의/경계/심각의 각 TEO값 분류를 기준으로 RMS, Variance, Skewness, Kurtosis, Shape factor, crest factor등 시간파형 데이터을 활

용하였다. 먼저 클러스터링을 하기 전에 데이터마다 값의 범위가 차이가 크므로 데이터 스케일링을 통해 정규화하였다. 클러스터링 알고리즘 결과 비지도 학습인 클러스터링을 이용하여 4개의 군집을 생성하고, 이때 고려하는 변수로는 TEO 값의 RMS, Variance, Skewness, Kurtosis, Shape factor, crest factor등 시간파형 데이터의 특징을 이용하였다.



[그림 5] 클러스터링 결과

전철전력설비 이상상태 데이터 예측분석을 위해 랜덤포레스트, SVM, XGBoost 알고리즘을 이용하여 전철전력설비 고장데이터 분류를 예측하였다. 첫 번째로 랜덤포레스트는 의사결정나무의 앙상블 기법으로 모델을 구성하는 기본 요소로 의사결정나무를 사용한다. 의사결정나무 모델의 앙상블을 통해 예측 성능은 유지되면서 과대 적합이 줄어드는 장점이 있다. 두 번째로 SVM은 분류, 회귀, 특이점 판별에 쓰이는 지도학습 머신러닝 방법 중 하나이다. SVM은 명백한 이론적 근거에 기반하므로 결과 해석이 용이하고, 실제 응용에 있어서 인공신경망 수준의 높은 성과를 내고, 적은 학습자료만으로 신속하게 분별학습을 수행할 수 있다. 마지막으로 XGBoost는 약한 분류기를 세트로 묶어서 정확도를 예측하는 기법이다. 전철전력설비 이상상태 데이터로부터 관심/주의/경계/심각으로 예측하기 위해 그림7과 같이 3가지의 알고리즘 중 최적의 알고리즘을 선택하였다.

PTP	RMS	Var	Mean	Skewness	CF	IF	SF	Kurtosis	MF	predict_y	real_y
2495423744	294661072.6	8.55076E+16	36298030.52	0.21	4.8	6.74	1.41	1.37	0	0	0
2465004544	317449604.1	9.92022E+16	39648775.65	0.3	4.6	6.48	1.41	1.56	0	0	0
2651230208	284746971	7.95745E+16	38811741.7	0.04	4.95	6.83	1.38	1.47	0	0	0
2600349440	3018756596	8.96615E+16	38306604.32	0.19	4.77	6.78	1.42	1.77	0	0	0
2754224128	306105985.8	9.21837E+16	38950283.2	0.21	4.7	6.63	1.41	1.77	0	0	0
2430565376	304858401.1	9.14099E+16	39099767.64	0.18	4.61	6.31	1.37	1.03	0	0	0
2245271552	282161177.6	7.81959E+16	37669774.02	0.04	4.32	5.92	1.37	0.96	0	0	0
319593984	53000459.45	2.31898E+15	22137588.29	0.11	3.42	4.4	1.28	0.63	0	0	0
2442299904	291599160.1	8.36315E+16	37397432.01	0.12	4.49	6.16	1.37	1.18	0	0	0
151634688	29045808.38	3.70428E+14	21753880.41	0.11	3.38	4	1.18	1.38	0	0	0
203615232	33690232.7	6.58665E+14	21825818.36	0.21	3.77	4.65	1.23	2.05	0	0	0
8757942784	1713928565	2.64135E+18	544245718.6	0.43	3.36	4.43	1.32	0.08	0	3	3
2217037824	289168650	8.20331E+16	39817275.32	0	3.87	5.19	1.34	0.85	0	0	0
8687195904	1769637309	2.78109E+18	592051444.4	0.37	3.21	4.17	1.3	-0.08	0	3	3
2309128960	275458412.7	7.47046E+16	34245146.79	0.22	4.44	6.31	1.42	1.55	0	0	0
2547041280	3179329129	9.94422E+16	40486848.81	0.17	4.55	6.34	1.39	1.26	0	0	0
2695431680	293191618.7	8.45373E+16	37736410.25	0.24	5.41	7.55	1.4	1.51	0	0	0
260363264	41822672.2	1.26747E+15	21946977.94	0.08	3.58	4.56	1.27	1.06	0	0	0
5130981120	611091512	3.64842E+17	92684551.29	0.53	4.92	8.81	1.79	2.96	0	1	2

[그림 6] 이상상태 데이터 분류 예측결과

예측모델 학습을 위해 Train, Test 데이터를 7:3 비율로 학습과 테스트데이터를 분할하고, 분할된 데이터는 데이터 스케일링을 통해 정규화하였다. 예측결과로는 모든 데이터셋에서 대부분 97%이상으로 RandomForest가 가장 정확도가 높게 나타나고, SVC가 가장 낮게 나타났다. 예측 결과, 어느 변수가 결과에 영향을 미쳤는지 알아보기 위해 변수중요도를 구하였다. PTP, RMS, Var, Mean, MF 순으로 등급을 예측하는데 중요 인자임을 알 수 있었다.

4. 결론

전철전력설비 이상상태 데이터는 오송종합시험선로에서 측정하고 TEO값을 기준으로 4등급(관심/주의/경계/심각)으로 분류하였다. 해당 연구결과를 토대로 시험선로가 아닌 실제선로에서 전력장치에서 실시간으로 데이터가 확보되면 더욱 정교하고 빠른 검측으로 유지 보수의 효율화와 안전사고 예방할 수 있는 실시간 모니터링에 기계학습 적용 가능성을 확인하였다.

참고문헌

[1] No-Geon Jung, Chul-Min Park, Jae-Bum Lee, Young Park, 1Seung-kwon Shin, “A Study on the Application of TEO and STFT Signal Processing Techniques for Detection of Electric Railway Contact Loss,” The Transactions of the Korean Institute of Electrical Engineers, Vol. 67, No. 11, pp. 1530-1535, 2018. 11.

[2] Je-Ho Yoo, Seung-Kwon Shin, Jong-young Park, Soo-Hwan Cho, “Advanced Railway Power Quality Detecting Algorithm Using a Combined TEO and STFT Method,” Journal of Electrical Engineering & Technology, Vol. 10, No. 6, pp. 2442~2447, 2015. 11.

[3] Chanho Jung, Seungkwon Shin, Jiwon Lee, Wonjun Kim, “Objective Evaluation of Image Decomposition Algorithms for DepthMap Upsampling,” Journal of Electrical Engi- neering & Technology, 14:2619~2624, 2019.

[4] Soo-Hwan Cho, Hoon-Chul Shin, Jae-Bum Lee, Seung-Kwon Shin “An Effective Detection Method of Voltage And Frequency Fluctuations Based on a Combination of TEO/ DESA and STFT Analysis,” Journal of Electrical Engineering & Technology, Vol. 14, No. 2, pp. 985-992, 2019. 2.