Rule 기반 AI 모델의 지속운용을 위한 프레임워크

박영지[1], 이태진[2]
[1]호서대학교 컴퓨터공학부 학부생
[2]호서대학교 컴퓨터공학부 교수
yj010102@gmail.com, kinjecs0@gmail.com

A Framework for Continuous operational techniques of AI Model based on Rule

Yeong-Ji Park[1], Tae-Jin Lee[2]
[1]Dept. of Computer Science, Hoseo University
[2]Dept. of Computer Science, Hoseo University

요 약

오늘날 AI 기술은 다양한 분야에서 활용되며 발전해나가고 있다. 하지만 AI 모델의 복잡도가 증가하며 AI의 산출 결과의 해석이 불가능한 Black-box 성격을 지니게 되었고, 이는 실 환경에서 AI 도입의 커다란 걸림돌로 작용하고 있다. 이에 따라 AI 판단 결과에 대한 Interpretation을 제공하는 AI Decision Support의 중요성이 커지는 추세이다. 본 논문에서는 Reference 기반 Rule을 통해 AI 모델의 판단 결과에 대한 해석을 제공하고 입력된 데이터에 관한 Rule 적합도를 산출하여 AI Decision Support를 제공하고자 한다. 또한, Rule 적합도 정보를 기반으로 기존의 모델보다 정확한 산출 결과를 통해 수집된 데이터의 Label을 확정시킨다. 이를 토대로 AI 모델의 업데이트를 실행하여 지속적으로 AI의 성능을 개선하면서도 지속 운용이 가능한 AI 운용 프레임워크를 제안한다.

1. 서론

현재 AI 기술은 의료, 산업시설 등 다양한 측면에서 사용되고 있으며 보안 분야에서도 많은 연구가 이뤄지고 있다. 그러나 AI 모델 성능이 향상됨과 동시에 복잡도도 증가하였고 AI 산출 결과의 과정을 알 수 없는 Black-box 문제는 실 환경에서 AI 도입의 커다란 걸림돌로 작용되고 있다. Gerlach 등[1]은 AI 판단 결과에 대한 Interpretation을 제공하는 XAI 기법 등이 활용되면서 AI Decision Support의 중요성이 커지는 추세라고 말한다. 본 논문에서는 학습된 AI 모델로부터 AI 모델의 판단 결과 해석을 위한 RuleSet을 자동적으로 생성하고 입력된 데이터에 대한 Rule 적합도를 산출한다. 이후 산출된 Rule 적합도 정보를 기반으로 AI 모델의 판단 결과에 대한 Decision Support를 제공하여 발생할 수 있는 오탐지를 감소시키고, 기존 모델보다 정확한 산출 판단 결과를 기반으로 수집된 데이터의 Label을 확정하여 AI 모델의 업데이트를 수행함과 동시에 지속적으로 AI의 성능을 개선해나가는 지속 운용이 가능한 Rule 기반 AI 운용 프레임워크를 제안한다.

2. 관련 연구

Han 등[2]의 연구에서는 비지도 학습에서의 Interpretation 제공을 위해 anomaly 데이터와 가장 가까우면서도 정상 데이터 범주에 있는 값을 나타내는 Reference를 제안했다. Reference는 Threshold의 MSE를 기준 잡아 원본과의 거리를 체크하며 Optimizer를 통해 산출한다. 비지도 학습은 데이터에 대한 label이 없어 선행 연구된 XAI 기법을 적용하기 어렵다. Reference는 이러한 문제를 해결하기 위해 anomaly 데이터와 비교할 수 있는 적절한 Reference 값을 제시하여 두 값의 오차를 산출함으로써 label이 없는 상황에서도 비지도 학습 모델의 판단에 대한 Interpretation을 제공한다.

3. 제안 프레임워크

3.1 Dataset & AI Model

데이터셋은 4개의 feature를 사용하며 normal data와 anomaly data로 구성된다. normal data는 각 feature가 1~10의 값을 가지는 2000개의 data이며, anomaly data는 각 feature가 11~20의 값을 지니는 20개의 data로 총 2020개의 data로 구성된다. normal과 anomaly 데이터의 비율은 약 99:1이며, 본 논문에서는 원본과 복원값 간의 차이, MSE를 통해 Anomaly Detection을 수행하는 AutoEncoder를 사용한다.

3.3 Proposed Framework

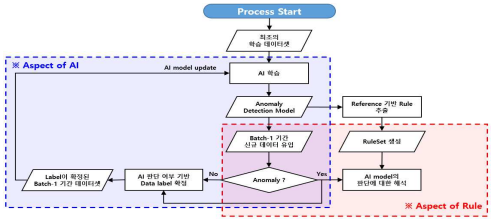

그림 1 Proposed Framework

본 논문에서 제안하는 Framework는 그림 1과 같다. 최초의 학습 데이터 셋을 통해 AI 모델을 학습 및 생성한 후, AI 모델로부터 Reference 기반의 RuleSet을 자동적으로 생성한다. 이후 다음 Batch 기간 동안 유입된 신규 데이터는 기존에 학습된 AI 모델의 판단과 RuleSet의 적합도에 따라 Label을 확정 짓는다. 동일 기간 동안 유입된 Anomaly 데이터를 대상으로 자동적으로 RuleSet이 생성되고, 다음 Batch 기간에 AI 모델의 판단에 대한 Decision Support를 제공받는다.

3.4 Reference 기반 Rule 설정

Reference는 정상 데이터 범주에 속하면서도 Anomaly 데이터와 가까이 위치한 값이기 때문에 Normal과 Anomaly의 사이에 있는 경곗값으로 산출된다. 즉, 산출된 Reference Value의 값을 초과하면 Anomaly라는 의미이며 해당 Value를 Rule로써 추출하게 되면 Anomaly 판단에 대한 Rule로 사용할 수 있게 됨을 의미한다.

4. 실험 결과

4.1 Reference 산출 결과

그림 2를 통해 Reference 산출 결과를 확인할 수 있다. Reference의 값이 normal 데이터 경계인 10 부근에 생성되었고, anomaly의 값과 가깝게 산출되었으므로 값이 적절하게 나왔음을 알 수 있다.

4.2 Reference 기반 Rule 추출 결과

Anomaly 데이터로 설정해둔 Index 2000번~2019번의 데이터들이 Anomaly로 잘 산출이 되어 해당 데이터의 Reference를 Rule로 설정하였고 해당 결과를 그림 3과 같이 확인할 수 있다.

4.3 Ruleset 기반 Decision Support 결과

생성한 RuleSet에 Normal 데이터와 Anomaly 데이터를 넣어 적합도를 확인한 결과는 그림 4와 같다. Normal 데이터인 1057번째 데이터의 Rule 적합

Feature Description	Value in Anomaly	comp.	Value in Reference
feature1	20.0	>	10.496
feature3	17.0	>	9.882
feature2	14.0	>	10.447
feature4	11.0	>	10.224

그림 2 Result of Calculation Reference

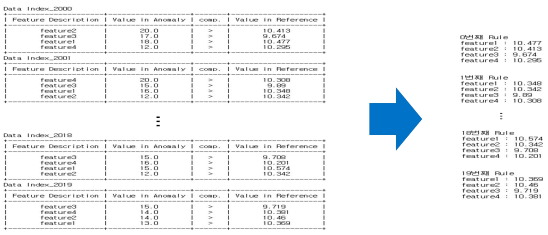

그림 3 Extraction of RuleSet from Reference Result

그림 4 RuleSet based Interpretation (Normal vs Anomaly)

도 결과는 모두 0이며, Anomaly 데이터인 2017번째 데이터의 Rule 적합도 결과는 모두 1의 결과가 나옴을 확인할 수 있었다.

5. 결론

본 논문에서는 RuleSet과 AI 모델의 업데이트를 통해 AI의 성능을 개선해나가면서 지속적으로 운영할 수 있는 프레임워크를 제안했다. Reference 기반 RuleSet을 사용하여 적합도를 산출한 결과 AI 모델의 판단 결과 해석 제공이 가능하였다. 이를 기반으로 Rule을 통해 제공되는 AI 모델 판단의 해석과 더욱 정확한 데이터 라벨링 기반 AI 모델의 업데이트를 통해 지속적으로 AI의 성능을 증가시키면서 지속 운용이 가능한 AI 모델을 지향하고자 한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부와 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음 (2019-0-01834)

참고문헌

[1] Gerlach, Jana, et al. "Decision support for efficient XAI services-A morphological analysis, business model archetypes, and a decision tree." Electronic Markets (2022): 1-20.

[2] Han, Dongqi, et al. "DeepAID: interpreting and improving deep learning-based anomaly detection in security applications." Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security. 2021.