

2018 발전산업 빅데이터 분석
AI 활용 공모전

(트랙2)제안 계획서

1.	참여 과제	터빈 Anomaly Detection with GAN
2.	참 가 자	<개인/조형배>

제안 과제명 : 터빈 Anomaly Detection with GAN

계획서 요약



항목	내용	비고
1. 추진배경 및 필요성	<ul style="list-style-type: none">•15.12.5 일, 시험운전중이던 당진화력 9호기 터빈서 화재 발생.-피해 규모 천문학적 규모. 하루 미가동으로 인한 피해 규모: 약 4 억원으로 추산.•터빈은 복잡한 형상과 블레이드 간 밀집된 구조로 비파괴검사 수행이 어려움.•서부발전, GE, 지멘스,,,	
2. 제안 내용	<ul style="list-style-type: none">•Anomaly detection•GAN(Generative Adversarial Network)	
3. 실행방안 및 기대효과	<ul style="list-style-type: none">•의료 등 다른 영역에서 선행 연구 사례가 존재하며 오픈소스코드가 있는 방법으로서 대회 기간 내 실현가능성이 있다고 판단.•학습에 사용되지 않은 실제 고장 데이터에 대해 이상을 감지할 수 있는지 시연이 가능할 것으로 생각됩니다.•향후에 드론 영상 등 다른 유형의 데이터셋 구축 시, 응용 및 발전 가능.	
4. 기타	<ul style="list-style-type: none">•터빈의 고장 예지가 동서발전에 기여할 수 있는 기술인지 궁금합니다.•터빈의 데이터가 필요합니다.	

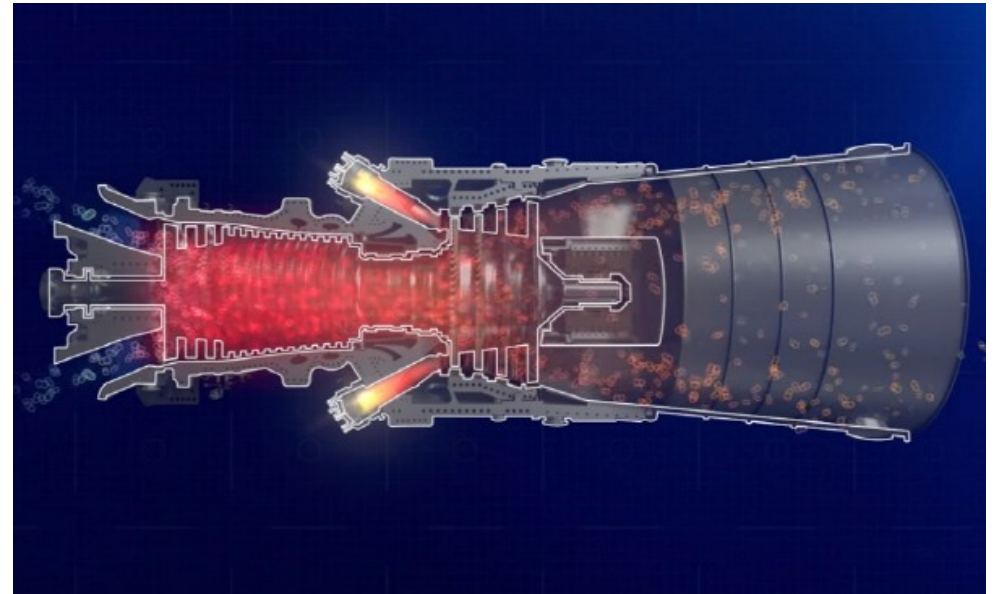
제안 과제명 : 터빈 Anomaly Detection with GAN

1. 추진 배경 및 필요성

HOME > 뉴스 > 종합

당진9호기 화재로 1,400여억 손실 예상

한운승 기자 | 승인 2015.12.05 22:35 | 댓글 0



15.12.5일, 시험 운전 중이던 당진화력 9호기 터빈서 화재 발생.

-피해 규모 천문학적 규모.

하루 미가동으로 인한 피해 규모: 약 4 억원으로 추산.

화력 발전의 핵심 설비 터빈.

복잡한 형상과 밀집된 구조.

AI로 고장을 진단할 수는 없을까?

제안 과제명 : 터빈 Anomaly Detection with GAN

2. 선행 사례: GE, 서부발전



인프라 산업에 인공지능(AI)을 붙여넣고 있는 GE

GE는 프레딕스 플랫폼에 인공지능을 적용해 작업자가 정보를 바탕으로 결정을 내리기 위해 필요한 지식을 제공한다. 예를 들어 터빈 블레이드를 폐기하거나 수리해야할 경우 일반적으로 작업자는 본인의 판단을 바탕으로 최종 결정을 내렸다. 그러나 딥러닝을 기술을 이용한다면 판단착오를 줄일 수 있다



서부발전 인공지능(AI) 터빈 예측 진단 솔루션 개발
서부발전 관계자 “현재까지 4차산업 기술은 스마트한 운영 분야에 집중되고 있지만, 앞으로는 빅데이터를 기반으로 인공지능을 접목한 예측 진단 등의 기술이 급속히 발전할 것”

3. 이론적 배경



한국동서발전은 이미 이와 관련한 인공지능 과제를 수행 중. 울산과학기술원에서 자체 개발한 ‘군집 시계열 딥러닝 모델’을 활용해 발전소 고장 데이터를 학습하고 상호 인과관계를 분석하여 고장원인과 차기 정비시기를 예측하는 것이 목적.

이미 관련 연구를 수행중이지만...
다른 측면에서의 접근을 제안하고자 합니다.

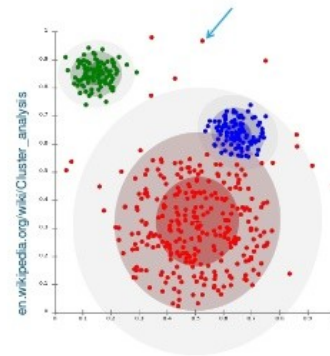
고장 예지의 어려움

-이상하다고 규정할 사례가 매우 적다.

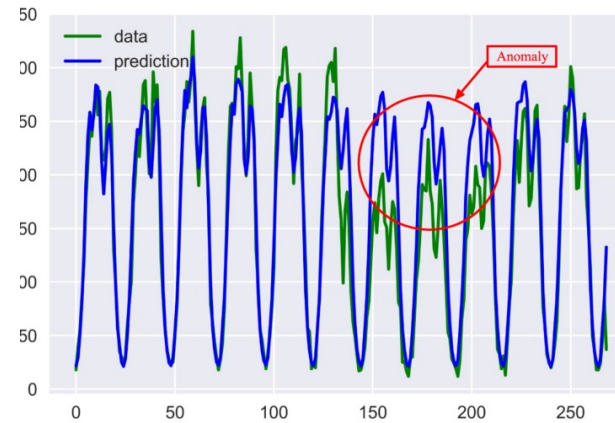
즉, 지도 학습 분류 모델을 만들기 위한 데이터셋이 부족
따라서 비지도 학습 **Anomaly detection** 모델이 활용된다

Clustering

- Find areas dense with data (conversely, areas without data)
- Anomaly = far from any cluster
- **Unsupervised** learning
- Supervise with labels to improve, interpret



cloudera



3. 이론적 배경 : GAN 열풍



2014년 등장한 한 논문은 인공지능(AI) 업계를 뒤집어 놓았다.

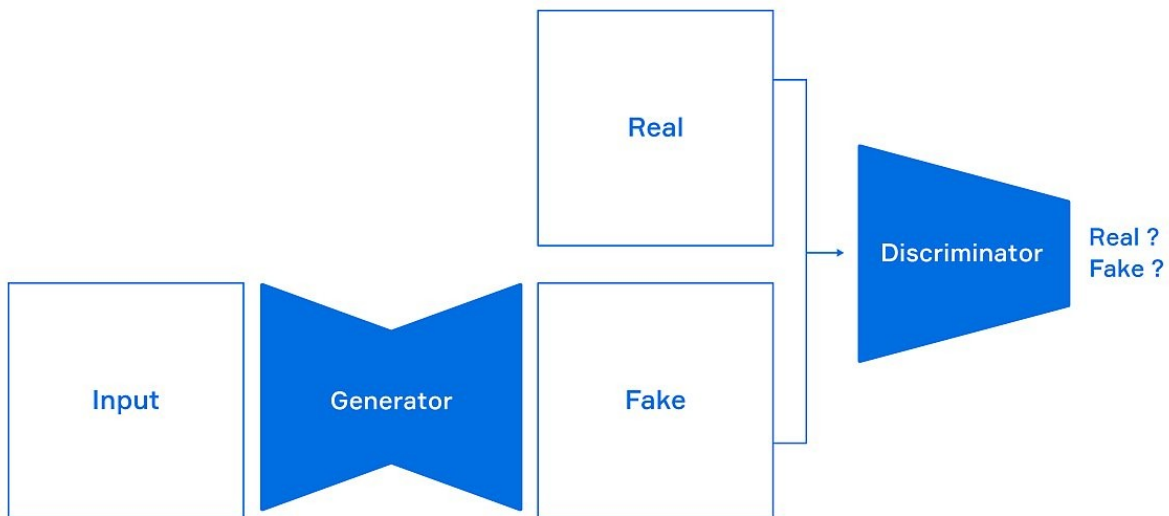
...

<중략>

...

딥러닝의 대가이자 페이스북 AI 연구팀 리더인 얀 르쿤(Yann Lecun) 교수는 GAN을 최근 10년간 머신러닝 연구 중 가장 혁신적인 아이디어로 꼽았다. 무엇보다 진짜 같은 가짜를 쉽고 빠르게 만들 수 있다는 점에서 기술이 널리 확산되고 있다.

출처: 네이버캐스트



$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim P_{data}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim P_{\mathbf{z}}(\mathbf{z})} [\log (1 - d(G(\mathbf{z})))]$$

GAN은 생성적 적대 신경망이라는 이름처럼 두 신경망 모델의 경쟁을 통해 학습하고 결과물을 만들어낸다. 이안 굿펠로우는 생성자를 위조지폐범에, 감별자를 경찰에 비유했다. 생성자는 감별자를 속이지 못한 데이터를, 감별자는 생성자에게 속은 데이터를 입력받아 학습한다. 이 과정이 반복되면서 위조지폐가 정교해지듯 점점 더 실제에 가까운 거짓 데이터를 만들 수 있게 되는 셈이다.

4. 제안 내용

Anomaly Detection with GAN

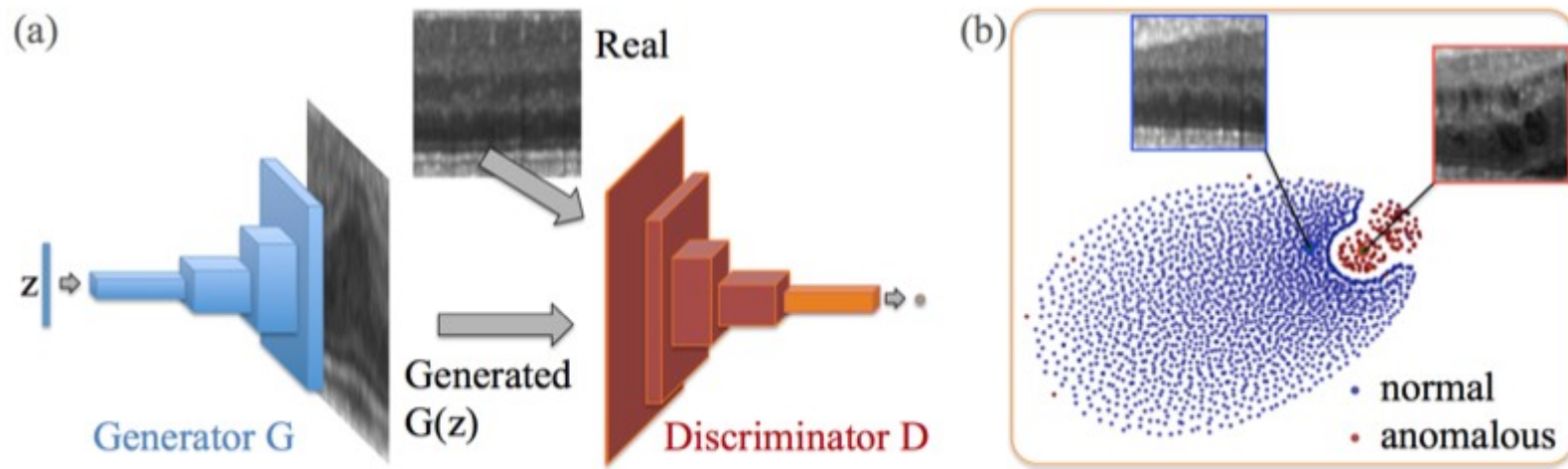


Fig. 2. (a) Deep convolutional generative adversarial network. (b) t-SNE embedding of normal (blue) and anomalous (red) images on the feature representation of the last convolution layer (orange in (a)) of the discriminator.

왼쪽의 그림은 “Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery”라는 논문에서 GAN을 이용하여 Anomaly Detection을 수행한 예입니다. 이 연구에서는 정상 의료 영상 데이터를 이용하여 학습한 모델로 새로운 의료 영상 데이터에 대하여 이상을 감지하는 것을 목표로 합니다. 본 과제와 이 논문은 에너지와 의료라는 도메인이 다르고 영상과 수치라는 데이터의 성격은 다르지만 대부분 정상 데이터로 이루어지는 도메인에서 이상을 검출하는 시스템을 개발한다는 점에서 이 연구는 중요한 참고가 됩니다.

5. 실행 방안

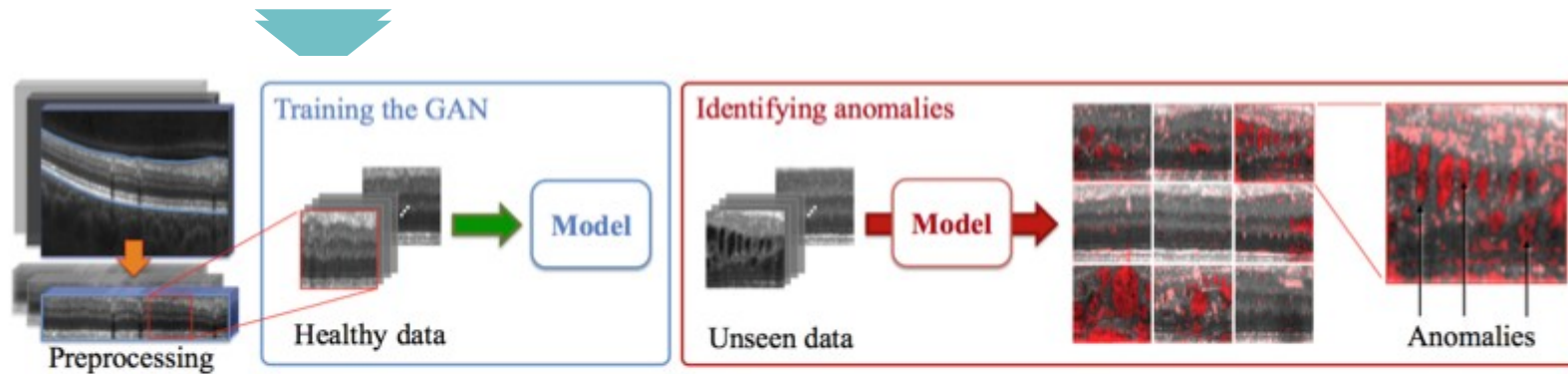


Fig. 1. Anomaly detection framework. The preprocessing step includes extraction and flattening of the retinal area, patch extraction and intensity normalization. Generative adversarial training is performed on healthy data and testing is performed on both, unseen healthy cases and anomalous data.

5. 목표 및 기대효과

- ① 학습에 사용하지 않은 실제 이상 데이터에 대해서 이상 감지 여부를 시연할 수 있는 모델을 대회 기간 내에 개발하는 것을 목표로 합니다.
- ② 향후에 드론 영상 등 다른 형식의 데이터셋이 주어졌을 때 적용가능하며 더욱 성능이 향상하는 시스템이 될 것이라 기대합니다.

- ① 대부분 정상 데이터로 구성되어있을 것으로 예상되는 터빈 데이터로 GAN 모델을 학습
- ② 다음과 같이 구성되는 이상 점수를 계산하는 모델을 개발

$$\text{Residual Loss } L_R(z) = \sum_k \|x - G(z)\|_1$$

$$\text{Discrimination Loss } L_D(z) = \sum_k \|f(x) - f(G(z))\|_1$$

$$\tilde{A}(x) = (1 - \lambda)L_R(z_T) + \lambda L_D(z_T).$$

제안 과제명 : 터빈 Anomaly Detection with GAN

개인/조형배

1. 참가자(팀원) 역량 (팀원의 전문성, 과제해결 역량 등)

제가 자신 있게 내세울 수 있는 역량은 텍스트, 이미지 등과 같은 비정형 데이터를 정제하고 분석할 수 있는 역량입니다. 저는 이 역량을 갖추기 위해 3가지 노력을 하였습니다. 첫째, 저는 포항공과대학교의 빅데이터, 인공지능 인재양성 심화 과정에 참여하여 컴퓨터 비전, 자연어처리에 대한 수업을 듣고 자기소개서를 분석하고 요약하는 팀 프로젝트를 수행하였습니다. 이를 통해 딥러닝 기술을 이용하여 비정형 데이터를 정제하고 분석하는 훈련을 할 수 있었습니다. 우수한 역량을 인정받아 최우수교육생 표창을 받기도 했습니다. 둘째, 한국어 자연어 처리 기법을 제안하는 논문을 쓰고 학회에 참석하여 포스터 발표하였습니다. 이 논문을 쓰면서 한국어에 적합한 자연어 처리 기법을 연구하고 발전시킬 수 있었습니다. 셋째, 9월까지 재직한 포항공과대학교 정보통신연구소에서 얼굴인식 연구 및 개발 프로젝트를 통해 사진 데이터를 분석하는 경험을 하였습니다. 이 프로젝트를 통해 최신의 이미지 처리 기술을 적용해볼 수 있었습니다.

2. 대회 참여 배경 (대회를 통해 얻고싶은 바, 향후 계획/포부 등)

저의 인생의 목표는 세상을 이롭게 하는 가치를 창출하는 것입니다. 이에 다양한 봉사활동에 참여하였고 경기도민의 복지증진에 기여한 공으로 표창을 받기도 했습니다. 데이터과학자라는 저의 진로도 이 가치관으로부터 결정되었습니다. 저는 빅데이터 분석이 세상을 지금보다 더 나아지게 할 수 있는 잠재력이 있다고 믿기 때문입니다. 특히 학교에서 데이터 분석을 통해 얻은 금연에 대한 통찰을 프로젝트 발표를 통해 학우들과 공유하였을 때 큰 뿌듯함을 느꼈고 진로에 대한 확신이 더욱 커졌습니다. 이 대회는 이러한 저의 가치관과 진로에 잘 부합합니다. 이 대회를 통해 데이터 과학자로서 전문성을 더욱 향상하는 한편 발전소의 효율을 높이고 미세먼지 배출을 줄여서 세상을 이롭게 하는 가치를 창출하는데 이바지하고 싶습니다.