

사출 성형품의 품질향상을 위한 이상 탐지 기반 불량품 예측

김재윤¹, 김태영¹, 손재성¹
광운대학교 정보융합학부

초 록

사출성형이란 금형을 이용한 성형법 중 하나로, 플라스틱 등의 재료를 가열해서 녹이고 금형에 주입한 뒤 냉각시켜 원하는 성형품을 제조하는 공정이다. 사출성형 과정에서는 여러 공정 조건의 영향으로 불량품 발생은 피할 수 없다. 본 보고서는 사출 공정 현장의 직접적인 문제 상황을 소개하며, 불균형이 극심한 사출 공정 데이터 셋에 알맞은 다양한 이상탐지 알고리즘을 적용한다. 이상탐지 알고리즘에 기반한 사출 공정의 불량 예측 실험을 진행하며, 사용된 알고리즘에 대하여 분류 성능 평가 지표로 비교 분석을 진행한다. 최종적으로 공정품의 품질 예측을 위한 적합한 모델을 발견하고자 한다. 현재 자동차 부품을 생성하는 많은 사출 공정에서는 여전히 육안으로 제품의 양품 선별을 진행한다. 기계학습 기반의 방법론을 통해 사출 공정에서 더 나은 품질의 균일성 확보를 보장할 수 있으며, 실질적인 품질 및 비용 절감 개선에 기여할 수 있는 점에 있어서도 시사하는 바가 크다고 판단된다.

1. 서론

사출성형은 성형 주기가 짧고, 능률이 좋아서 대량 생산이 가능하다. 또한 열가소성 수지, 열경화성 수지가 가능하여 광범위하게 수지의 성형이 가능하다. 자동 제어화 또한 가능하며, 정밀도가 높아 복잡한 모양의 대형 성형품의 성형이 가능하다. 여러 단계에 거쳐서 복잡한 공정 과정이 진행되는데, 본 프로젝트는 사출 성형기로부터 수집된 계량, 온도, 압력, 시간, 속도와 관련된 데이터를 기반으로 다양한 이상탐지 알고리즘을 학습시키고 성능을 평가한다. 최종적으로는 본 데이터셋에 적합한 최적의 이상탐지 알고리즘을 선정한다.

2. 문제 설명

2.1. 사출 공정의 문제 상황

대부분의 제조 공정 단에서 발생하는 대표적인 문제 상황으로는 ‘경험에 의존한 장비 운영’, ‘생산기술 암묵지 현상’, ‘불명확한 불량원인’, ‘제조 원가의 증가’ 등이 있다. 본 보고서는 제조 공정 중에서도 사출 공정의 문제 상황을 소개하며, 이를 기계학습의 이상탐지 알고리즘에 기반하여 접근한다. 사출 성형품의 품질에 영향을 주는 주요한 성형 조건은 ‘수지 온도’, ‘압력’, ‘균일한 냉각’이다. 이들의 관계가 어떠한지에 따라 문제가 야기되며, 성형품의 형상과 크기, 그리고 사용 재료 또한 품질에 영향을 줄 수 있다. 다양한 성형 조건들을 잘 조합하는 것은 무척 중요한데, 일반적인 방법으로는 아래와 같은 순으로 조합을 이루어 사출 조건을 정하게 된다.

- 온도를 일정하게 하고 저압(고압) 저속(고속)
- 온도를 일정하게 하고 고온(저온) 고속(저속)
- 온도를 일정하게 하고 고온(저온) 고압(고압)

성형 조건의 조합이 잘못되면, 공정 자체의 큰 문제로 이어질 수 있다. 따라서 품질을 결정하는 여러 변수 간 상관관계를 고려하여, 이상탐지 알고리즘을 이에 적용한다면 사출 공정 불량 예측에 큰 기여를 할 수 있다.

2.2. 데이터 정의 및 소개

본 프로젝트에 사용된 데이터는 “인공지능 중소벤처 제조 플랫폼”에서 제공하고 있으며, 데이터셋에 대한 정의 및 소개는 다음과 같다.

실험에 사용된 데이터는 자동차 앞 유리 사이드 몰딩을 사출하는 공정 과정에서 생성되었다. 해당 공정에서 생성되는 데이터는 1차 가공 데이터와 2차 가공 데이터로 크게 두 가지 형태로 볼 수 있다. 본 프로젝트에 사용된 데이터 셋은 1차 가공으로 약간의 처리가 이루어진 2차 가공 데이터이다. 수집 기간은 2020년 10월 16일부터 2022년 11월 19일까지 1개월간 수집되었다.

데이터셋의 변수들에 대한 설명은 아래와 같다. 독립 변수는 사출 공정에서의 품질 문제의 원인이 될 수 있는 온도, 압력, 시간, 속도, 양과 관련된 특징을 지니고 있다.

공정 변수 조건	내용	
독립 변수	온도 관련	수지, 금형, 유압 주변 환경 등
	압력 관련	충진 압력, 보압, 배압, 형체 압력 등
	시간 관련	충진 시간, 보압 시간, 냉각 시간 등
	속도 관련	사출 속도, 스크류 회전 속도 등
양 관련	계량	계량, 이형량, 쿠션량 등
	불량 여부	Y : 양품, N : 불량품

모든 독립 변수는 수치형 변수이며, 변수 각각에 대한 구체적인 설명은 아래와 같다.

Attribute	Description
PassOrFail	양품 혹은 불량품
Injection_Time	고압 + 사출시간 (사출압)
Filling_Time	충진시간
Plasticizing_Time	계량시간
Cycle_Time	제품 생산의 소요 시간
Clamp_Close_Time	금형을 사출기가 닫아주는 시간
Cushion_Position	보압 하기 위한 위치
Plasticizing_Position	계량 완료 위치
Clamp_Open_Position	금형 열리고 난 위치
Max_Injection_Speed	용융수지가 흘러가는 최대 속도
Max_Screw_RPM	스크류 최대 속도
Average_Screw_RPM	스크류 평균 속도
Max_Injection_Pressure	용융수지가 흘러갈때 최대 압력
Max_Switch_Over_Pressure	사출에서 보압으로 변환 압력
Max_Back_Pressure	스크류 밀림 현상 저지 최대압력
Average_Back_Pressure	스크류 밀림 현상 저지 평균압력
Barrel_Temperature_1	배럴 온도 (온도가 일정해야 된다)
Barrel_Temperature_2	
Barrel_Temperature_3	
Barrel_Temperature_4	
Barrel_Temperature_5	
Barrel_Temperature_6	
Hopper_Temperature	재료주입구의 온도
Mold_Temperature_3	
Mold_Temperature_4	

2.3. 분석 목표

이상탐지(Anomaly detection)는 전체 데이터에 대해 비정상 값을 찾아내는 방법론으로 제조업에서는 특히나 많이 사용된다. 대부분의 제조 공정 과정에서 나오는 데이터의 경우, 정상 데이터에 비해 비정상 데이터의 수가 현저히 적다는 점에 착안하여, 이상 탐지 방법론으로 해당 문제를 해결하는 것이 목표이다. 다양한 이상 탐지 알고리즘을 활용하여, 사출 공정에서의 최적화된 불량 예측 모델을 개발하는 것을 최종 목표로 한다.

3. EDA 및 전처리

이상탐지 알고리즘 적용 전, 필수적인 탐색적 데이터 분석을 진행한다. 주요 결과는 아래와 같다.

- Sample 수 : 6,736 (정상 : 6,697 / 비정상 : 39)
- 결측치 : 0
- 독립 변수의 특성 : klib 라이브러리를 활용하여 모든 변수의 분포를 시각화하였다. 공정과 관련된 영역이 동일한 변수들은 비슷한 분포를 띄는 것을 확인하였고, 다중공선성 파악을 위한 변수 간 상

관 분석 또한 진행하였다.

이상탐지 알고리즘으로써의 모델 학습과 Confusion Matrix를 통한 분류 모델의 성능 평가를 진행하기 위한 train set과 test set을 구축하였다. 이상탐지 모델의 학습은 정상 데이터만으로 진행하며, 비정상에 해당하는 관측치는 모델의 성능 평가 및 비교 시 사용된다. test set은 정상 관측치와 비정상 관측치의 비율을 동일하게 설정하여 구축하였으며, 아래와 같이 요약할 수 있다.

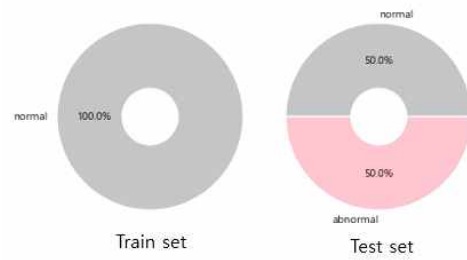


Figure 1. Train set / Test set의 Label 비율

더욱 강건한 학습을 위해서 학습에 사용될 Train set을 더욱 정상에 가깝게 구성하기 위해 IQR 방식을 사용하여 해당 데이터 셋에 포함된 이상치를 제거하였다. PCA 차원축소 결과, 주성분의 설명력이 충분함을 확인하였고, 아래와 같이 scatter plot과 box plot을 통해 정상 데이터 내 이상값의 존재 여부를 파악하였다.

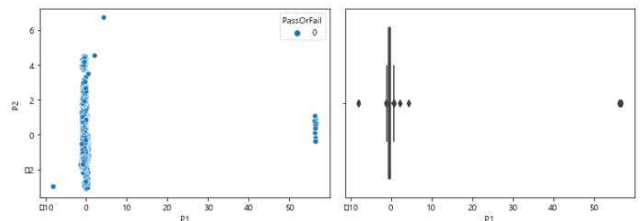


Figure 2. PCA(train set) 결과 scatter, box plot

위 분석 결과 $[Q_1 - IQR * 1.5, Q_3 + IQR * 1.5]$ 에 해당하는 데이터만 모델 구축에 사용하였으며, 이상값 제거 후 PCA(n_component=1) 차원 축소 후 데이터의 분포는 아래와 같다.

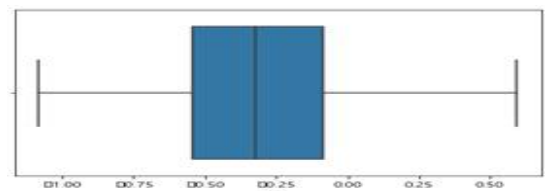


Figure 3. 이상치 제거 후 Train set의 box plot

4. 사용한 이상탐지 알고리즘 소개

이상탐지 알고리즘을 다양하게 적용하여 모델 학습 및 비교 평가를 진행하였으며 각 알고리즘에 대한 설명은 이어지는 내용과 같다.

4.1 밀도 기반 이상탐지 기법 (GDE, LOF)

밀도 기반 이상탐지 기법은 데이터의 분포를 고려하여 이상치를 탐지한다. 첫 번째로 사용한 알고리즘은 GDE(Gaussian Density Estimation)이다. 이는 데이터가 정규분포를 따른다는 가정하에 사용되는 기법이다. 두 번째로 사용한 알고리즘은 LOF(Local Outlier Factor)이다. 이는 데이터가 가지는 상대적인 밀도까지 고려한 이상탐지 기법이다.

4.2 거리 기반 이상탐지 기법 (KNN)

KNN은 대표적인 Instance-based learning 알고리즘이다. 이는 학습 모델을 구축하지 않고, 인접 관측치로부터 예측을 수행한다. 그러나 추론 속도가 느리다는 단점이 있다.

4.3 재구축 기반 이상탐지 기법 (PCA, AutoEncoder)

PCA는 사영 후 분산을 최대화하는 목적을 가지고 있다. PCA 기반의 이상탐지는 기존에 사영했던 것을 다시 복원함으로써 기존의 값과 차이가 큰 것을 이상치로 간주한다.

AutoEncoder는 모델의 출력값과 입력값이 비슷해지도록 학습이 수행된다. AutoEncoder도 기존의 Input으로 복원을 수행했을 때, 기존의 값과 차이가 큰 값을 이상치로 간주한다.

4.4 새로운 알고리즘 (VAE)

추가적인 알고리즘으로 VAE(Variational Auto Encoder)를 적용하였다. 해당 모델은 기존 Auto Encoder 모델을 조금 수정하여 이상탐지에 활발하게 활용된다. 알고리즘을 요약하자면 다음과 같다. 입력 데이터의 정보를 요약하는 잠재 벡터가 정규분포와 유사하도록 특정 제약식을 추가한다. 제약식을 구성하기 위해, 데이터의 정보를 요약한 이후, 요약된 정보의 평균과 표준편차를 계산하며, 이를 정규분포에서 나온 임의의 input값과 연산하여 잠재 벡터로 계산한다. 이렇게 계산된 잠재 벡터를 다시 데이터로 복원하며 모델을 학습시킨다.

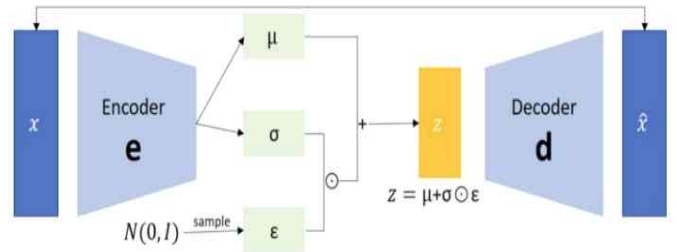


Figure 4. Variational Autoencoder(VAE) 구조

4.4 새로운 알고리즘 (DBSCAN)

DBSCAN은 밀도 기반의 클러스터링 기법이다. DBSCAN의 장점은 비선형의 클러스터링이 가능하다는 것이다. 이 후 반경에 포함되지 않는 데이터 포인트는 이상치로 간주한다.

위 설명의 이상탐지 기법들을 포함한 여러 알고리즘의 학습 및 성능 평가를 진행하였으며, 결과는 아래와 같았다.

5. 실험 결과 및 결론

	Accuracy	Recall	Precision	F1
Isolation Forest	0.91	0.92	0.9	0.91
LOF	0.82	0.71	0.90	0.8
GDE	0.91	0.95	0.88	0.91
AutoEncoder	0.80	0.67	0.93	0.78
OCSVM	0.72	0.90	0.66	0.76
PCA	0.91	0.95	0.88	0.91
VAE	0.95	0.95	0.95	0.95
DBSCAN	0.5	1.0	0.5	0.67

Table 3. 결과

본 데이터셋에 대한 대부분의 이상탐지 알고리즘은 좋은 성능을 보였다. 일반적으로 강건한 Decision Tree 기반의 Isolation Forest는 본 데이터셋에 대해서도 준수한 성능을 보였고, 밀도 기반, 재구축 기반의 모델도 충분히 좋은 성능을 보였다. 추가적인 모델로써 적용한 VAE는 모든 성능 지표에 대해서 좋은 결과를 보였다. 해당 모델은 추가된 제약식을 통해 본 데이터의 특성을 더욱 효과적으로 잘 파악한다고 볼 수 있다.

최종적으로 이상탐지 기반의 접근을 통해 사출 공정을 수행하는 제조 현장의 실질적인 품질 향상 및 비용 절감을 기대할 수 있다.

참고문헌

[1] “사출성형의 품질 향상을 위한 다양한 공정조건에 대한 해석적 고찰“