

실험계획과 머신러닝을 활용한 CNC 절삭공정 개선과 품질예측모델 개발 사례

주혜진¹ · 서호진¹ · 김영일² · 김수진³ · 이건명³ · 김상현⁴ · 정윤현⁴ · 변재현^{1*}

¹경상국립대학교 산업시스템공학부 / ²경상국립대학교 빅데이터혁신융합대학사업단

³경상국립대학교 기계공학부 / ⁴일광금속

A Case Study of CNC Machining Process Improvement and Quality Prediction Model Development Using Design of Experiments and Machine Learning

Hyejin Ju¹ · Hojin Seo¹ · Yeoung-Il Kim² · Su-Jin Kim³ · Gun-Myung Lee³ · Sang-Hyeon Kim⁴
Yoon-Hyeon Jeong⁴ · Jai-Hyun Byun¹

¹Department of Industrial and Systems Engineering, Gyeongsang National University

²Bigdata Convergence and Open Sharing System, Gyeongsang National University

³Department of Mechanical Engineering, Gyeongsang National University

⁴Ilkwangmetal

This paper presents a case study of systematically obtaining feature data and applying machine learning methods for a small CNC machining company that cannot obtain big data using sensors. In order to obtain the feature data, an experiment is planned and conducted using the 2-level 4-factor fractional factorial design with four machining process variables, and then 1) the outer diameter dimensional data is obtained using an automatic measurement tool and 2) appearance defects are visually inspected. An improved process conditions are determined to enhance productivity, to reduce tool wear, and to prevent defects. By analyzing the dimensional data and the number of non-defective/defective items obtained through observation, quality prediction models are also developed. This paper is expected to be used as a reference for small and medium-sized enterprises to improve the manufacturing processes in the future.

Keywords: Machining Process, Machine Learning, Design of Experiments, Process Improvement, Quality Prediction Model, Evolutionary Operation

1. 서론

품질개선 활동은 데이터 수집, 분석 및 의사결정 과정으로 이루어지는데, 최근 가공공정에서 수집한 데이터에 머신러닝 방

법을 적용하여 의사결정을 수행하는 연구가 활발하다. CNC 절삭가공에 한정해 보면, 대부분 공구에 온도, 진동, 소음, 위치 감지 등의 센서를 통해 피쳐 데이터를 모으고 이를 활용한 머신러닝 분석을 하는 것 위주로 연구가 진행되고 있다(Lee et

본 논문은 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 3단계 산학연협력 선도대학 육성사업(LINC 3.0)과 자원순환(재제조) 산업 고도화 인력양성사업(No. 20214000000520)의 연구 결과입니다.

* 연락저자 : 변재현 교수, 52828 경남 진주시 진주대로 501 경상대학교 산업시스템공학부, Tel: 055-772-1692, Fax: 055-772-1699,

E-mail: jbyun@gnu.ac.kr

2023년 4월 11일 접수; 2023년 5월 6일 수정본 접수, 2023년 5월 26일; 2023년 6월 20일 게재 확정.

al., 2019). Lee et al.(2019)은 공구 마모도를 예측하고자 센서를 이용하여 CNC 기계에서 48개의 피쳐 데이터를 얻고, 랜덤 포레스트 기법을 활용하여 마모도의 예측 정확도가 높은 모델을 도출하였다. Han(2022)은 5일 동안 CNC 장비에서 센서를 통해 수집한 데이터를 이용하여 가공 불량 예측모델을 구축하고 스핀들의 위치, 속도 등 불량 예측에 영향을 미치는 변수를 확인하였다.

제조공정 개선을 위해 머신러닝 모델을 개발하고자 한다면, 선행 연구에서 확인한 바와 같이 가능한 한 많은 수의 데이터가 필요하다. 공정 진행 과정에서 센서를 사용하여 미세하게 변화하는 공정변수의 값을 피쳐 데이터로 수집하고, 레이블 데이터로는 가공품의 품질 특성 측정치나 양·불량 여부를 이용한다. 만일 가공품의 품질 특성을 측정하는 데 시간과 비용이 과다하게 소요된다면 샘플링을 통하여 생산되는 가공품 일부의 품질 특성만을 측정한다(Kulahci, 2020). 대부분의 연구와 달리 아직 센서에 투자할 만한 여력이 없는 중소기업에서는 많은 데이터를 수집하기 쉽지 않기 때문에 머신러닝 방법을 이용하여 품질예측 모델을 구하는 데에 어려움을 겪고 있다. 일정한 가공조건에서 공정이 진행되고, 주요 공정변수 값의 변화를 나타내는 센서를 활용한 피쳐 데이터가 없으므로 머신러닝 알고리즘을 적용할 수가 없다. 이런 상황에서 머신러닝을 이용하여 품질예측 모델을 구하기 위해서는 별도의 피쳐 데이터를 생성해야 한다.

본 사례연구는 센서 데이터로부터 피쳐를 생성할 수 없는 중소기업의 CNC 가공공정 최적화를 통하여 불량을 줄이고, 품질비용을 절감하며, 생산성을 향상도를 도모하고, 머신러닝을 이용하여 CNC 가공품 품질을 예측할 수 있는 모델을 개발하고자 한다. 이를 위하여 레이블 데이터에 해당하는 가공품 치수는 자동검사기를 통해, 외관 불량 데이터는 육안검사로 수집할 수 있으므로 레이블 데이터는 확보되어 있다. 하지만 가공변수의 값은 항상 고정되어 있고, 가공 중 변수값의 움직임을 포착할 수 있는 센서도 없으므로 현 상태에서는 피쳐 데이터를 구할 수 없다. 품질을 예측하기 위한 머신러닝 모델을 개발하려면 레이블 데이터뿐만 아니라 피쳐 데이터도 필요하므로, 본 연구에서는 능동적으로 피쳐 데이터를 수집하기 위해 실험계획법을 이용한 체계적 실험을 진행한다. 이를 위하여 황삭과 정삭의 절삭속도와 이송속도 4개의 공정변수를 대상으로, 반복이 있는 2수준 4인자 2_{IV}^{4-1} 부분요인배치를 이용하여 8개의 공정조건에서 2회 반복 실험을 시행한다. 1회 실험 당 180~200개의 부품을 가공하였는데, 이들 가공품의 레이블인 외경(Outer Diameter)은 자동검사기로, 외관 불량은 육안검사로, 공구 플랭크 마모도는 현미경으로 측정한다. 가공품의 피쳐는 4개의 공정변수 외에 가공품 번호, 오프셋(선삭 직경 보정), 오프셋 후 가공품 번호의 7가지이다. 수집한 실험데이터를 분석하여 가공변수의 최적 조건을 산정하고, 품질예측모델을 개발한다. 기존 연구들은 CNC 가공공정 개선을 위해 CNC 가공불량의 원인을 공구 마모로 한정하거나, 센서를 활용한 데이터 수집 및 분석에 초점을 맞춰 공구수명

예측모델을 구축하는 데 집중되어 있다. 이와 달리 본 연구는 다양한 조건에서 체계적 실험을 통하여 데이터를 수집하고 실험 데이터를 분석하여 예측모델을 개발하는 것 외에도 설비 가동의 최적 조건을 도출하여 품질비용을 절감하고 생산성 향상을 도모한다.

제2장에서는 연구 대상이 되는 CNC 가공공정, 품질예측 알고리즘에 관해 간략하게 설명하고, 제3장에서는 피쳐 데이터를 생성하기 위해 부분요인배치법을 이용한 실험점 선정, 실험 진행 절차, 실험 결과 분석, 최적 가공조건 선정 과정을 기술한다. 수집한 데이터를 기반으로 가공품 품질예측모델의 개발과 분석 내용은 제4장에서, 중소기업이 머신러닝을 적용하면서 필요한 공정개선사항을 제5장에 제안한다. 제6장에서는 본 연구의 내용을 요약하고, 추후 연구 방향을 제시한다.

2. 연구 배경

2.1 CNC 가공공정

본 연구 대상인 CNC 가공공정은 엔진의 동력을 바퀴에 전달하는 삼각대 조인트(Tripod Joint)(<Figure 1>)의 부품인 T/P Spider(<Figure 2>)를 생산하는 공정이다. T/P Spider 제작에 사용된 CNC 기계는 수평형 터닝센터(Turning Center)(<Figure 3>)로서, 주요 사양은 척 직경 8inch, 최대 가공경 355mm, 최대 가공길이 440mm이며, 최대 스핀들 속도는 4,000rpm이다.

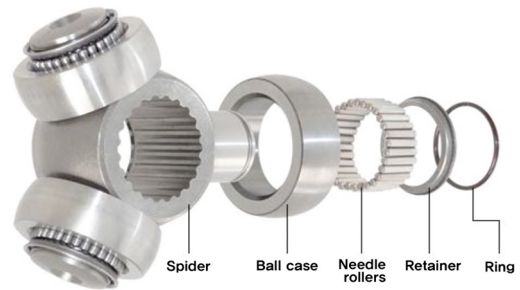


Figure 1. Tripod Joint

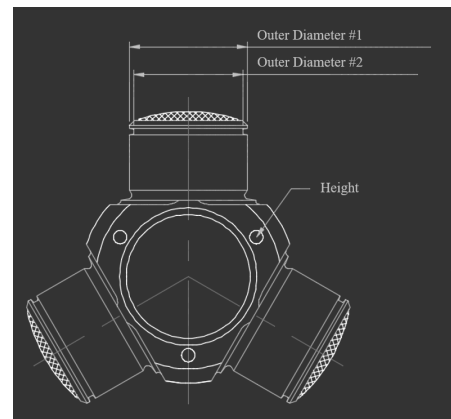


Figure 2. T/P Spider



Figure 3. Turning Center

CNC 터닝센터에서 T/P Spider를 가공하는 공정은 냉간단조, 선삭 1차, 선삭 2차, 브로칭의 단계를 거치는데, 선삭 2차 공정이 끝나면 자동검사기로 가공품의 주요 치수를 검사한다. 가공이 끝난 부품들은 한곳에 모아 육안으로 외관 불량을 확인하는 검사가 진행된다. 치수 검사에서는 외경#1, 외경#2, 높이, 단차, 높이평행도, 단차평행도의 치수를 확인하고, 외관 검사에서는 찍힘, 후피, 떨림, 녹, 버(Burr) 등이 발생했는지 확인한다.

CNC 가공공정 선삭 1차는 내경황삭, 외경/발부, 내경정삭, 가공단면 순으로 공정이 진행되고, 선삭 2차는 외경황삭, 외경정삭, 홈 가공 순으로 공정이 진행된다. 선삭 1차의 내경황삭은 가공품 내경의 가공 여유가 클 때 행해지는 거친 절삭가공, 외경/발부는 가공품 외경에서 돔 모양 부분을 가공하는 공정이다. 선삭 1차의 내경정삭은 내경황삭 후 내경의 규격 및 가공면을 얻기 위해 마무리하는 공정이며, 가공 단면은 가공품 양면의 단면을 가공하는 공정이다. 선삭 2차의 외경황삭과 외경정삭은 각각 선삭 1차의 내경황삭, 내경정삭과 유사한 공정이지만 가공품 외경을 가공하는 공정이고, 홈 가공은 가공품 외경에서 움푹 들어간 부분을 가공하는 공정이다.

본 연구의 대상인 CNC 가공공정에서 불량에 영향을 많이 주는 공정은 선삭 2차의 외경 황삭과 정삭 공정이므로, 황삭절삭속도(Roughing Speed; Vr), 황삭이송속도(Roughing Feed Rate; Fr), 정삭절삭속도(Finishing Speed; Vf), 정삭이송속도(Finishing Feed Rate; Ff)를 주요 공정변수로 삼아 연구를 진행하고자 한다.

2.2 품질예측 알고리즘

머신러닝에서 각 피쳐 데이터의 정답에 해당하는 레이블을 학습해 새로운 피쳐 데이터의 레이블을 예측하는 방법을 지도 학습이라 하는데, 지도학습은 예측해야 하는 레이블이 연속적인 값을 가지는 ‘회귀(Regression)’와 연속되지 않는 몇 개의 카테고리를 구분하는 ‘분류(Classification)’로 나뉜다. 본 사례연구에서는 회귀모델을 적합하기 위해 선형회귀, 랜덤포레스트

회귀, 서포트벡터 회귀를 적용하고, 분류모델을 위해서는 로지스틱 회귀, 랜덤포레스트, 서포트벡터머신, 의사결정나무를 활용한다. 지도학습 알고리즘은 이외에도 여러 가지가 있으나, 본 사례는 머신러닝을 활용하기 쉽지 않은 중소기업에서도 이를 이용하여 가공품 품질예측 모델 개발 및 적용하는 것이 목적이다. 따라서 추후 다른 제품의 품질개선 과정에서도 어렵지 않게 예측모델을 활용할 수 있도록 복잡하지 않으면서 가장 흔하게 사용하는 기본적인 알고리즘을 사용하고자 한다.

학습한 회귀모델의 우수성을 평가하기 위한 지표로서 평균 제곱오차(Mean Squared Error; MSE)와 평균절대오차(Mean Absolute Error; MAE)를 이용한다. MSE는 실제 값과 모델에 의해 예측된 값의 차이 제곱을 평균한 것으로, 식 (1)과 같다. MAE는 실제 값과 모델에 의해 예측된 값의 차이 절댓값을 평균한 것으로, 식 (2)와 같다. 식 (1)과 식 (2)에서 n 은 데이터의 개수, Y_i 는 실제 값, \hat{Y}_i 은 모델에 의해 예측된 값을 의미한다.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (2)$$

분류모델 성능평가 지표 중 본 연구에서 사용한 지표는 정확도(Accuracy), 특이도(Specificity), F1 점수이다. 정확도는 실제 값과 모델에 의해 예측된 값이 같게 나오는 데이터의 비율이고, 특이도는 실제 불량품인 값 중 예측값이 불량인 데이터의 비율을 나타낸 지표이다. F1 점수는 식 (3)과 같이 정밀도와 재현율을 결합하여 나타낸 지표이다. 여기서 정밀도(precision)는 양품이라고 예측한 값 중 실제 값이 양품인 비율을, 재현율(recall)은 실제 양품인 값 중 예측값이 양품인 비율을 의미한다.

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

제조공정에서는 양품에 비해 불량품의 수가 현저히 작으며, 이런 상황에서 수집한 데이터를 있는 그대로 이용하여 분석할 경우, 결과가 양품으로 편향되어 나타난다. 본 연구에서도 데이터 불균형이 발생하였고, 이로 인해 편향된 결과가 나오는 것을 막기 위해 리샘플링을 한다. 리샘플링 방법은 다수 클래스 데이터 개수를 소수 클래스의 데이터 개수에 맞추어 줄이는 언더샘플링과 소수 클래스의 데이터 개수를 다수 클래스 개수에 맞추는 늘리는 오버샘플링이 있다. 소수 클래스에 속하는 치수 불량품의 개수가 전체 실험 수량 3,998개 중에서 9개로 매우 작으므로, 오버샘플링 방법 중 우수한 성능을 보이는 ADASYN을 사용하여 소수 클래스를 오버샘플링 한다(Kim and Kwahk, 2022).

본 사례 연구에서는 가공품 품질에 영향을 미치는 가공변수를 피쳐 데이터로 삼고, 연속형 레이블 데이터와 이산형 레이블 데이터를 가지는 모델을 개발하기 위해 지도학습의 회귀모델과 분류모델 기법을 모두 사용한다. 치수 데이터를 레이블

로 하여 가공품의 치수를 예측하는 모델은 회귀 알고리즘을, 가공품의 양·불량 여부 데이터를 레이블로 하여 예측하는 모델은 분류 알고리즘을 적용한다. 가공품의 치수가 규격을 벗어나 불량 발생하는 공정조건을 도출하기 위해서는 예측 정확도가 낮지만, 시각적 표현을 잘할 수 있는 의사결정나무 알고리즘을 채택한다.

3. 실험계획을 활용한 데이터수집

연구 대상인 기업에서는 센서로 피쳐 데이터를 얻을 수 없으므로, 본 연구에서는 공정을 개선하고 머신러닝을 활용한 품질예측 모델을 개발하기 위해 양산공정을 대상으로 실험을 진행하여 피쳐 데이터를 생성한다. 여러 가지 잡음 요소가 반영되는 양산공정을 대상으로 시행하는 실험이라서 같은 조건에서 반복 실험하는 것이 필요한데, 4개의 인자를 고려하면서 2수준 완전요인배치법을 쓰면 실험점의 수가 너무 많아져서 2수준 4인자 2^{4-1} 부분요인배치를 반복 2회 실험을 시행하기로 한다. 2^{4-1} 부분요인배치를 적용하면 해가 IV이므로 각 인자의 주효과는 파악이 되지만 2인자 교호작용효과는 교락이 되는 단점이 있다(Box *et al.*, 2005; Montgomery, 2013). 각 인자의 수준은 공구 제작사에서 추천한 공구별 절삭속도(V), 이송속도(F) 적정 범위의 최대, 최소로 값으로 설정하고(<Table 1>), 이를 바탕으로 실험점을 도출하여 실험을 진행한다. 인자별로 3수준 실험을 하면 실험영역 전반에 걸쳐 더 많은 정보를 얻을 수 있지만, 기업 생산설비의 가용성 측면을 고려하여 부득이 2수준 실험을 시행한다.

Table 1. Factors and Levels

Factors	Low Level	High Level	Current Level	Unit
Vr	200	250	200	rpm
Fr	0.20	0.26	0.25	m/min
Vf	180	250	180	rpm
Ff	0.15	0.20	0.20	m/min

3.1 부분요인배치법을 활용한 실험

8개의 실험점을 대상으로 반복 2회, 총 16회의 실험을 양산공정에서 수행하기 위하여 공장의 한 생산라인을 2022년 11월 중순부터 12월 초순에 걸쳐서 전용으로 사용하였다. 실험의 설계행렬은 <Table 2>에, 설계공간은 <Figure 4>에 나타낸다. 각 실험점의 공정조건에서 선삭 2차 가공을 진행하면서 치수 데이터를 수집하고, 가공이 끝난 후 육안으로 가공물의 외관 불량 검사와 배율이 200배인 현미경으로 공구 플랭크 마모도 측정을 한다(<Figure 5>). 외관 불량 선별은 검사자가 모든 가공물을 눈으로

Table 2. Design Matrix

No	Actual Level				Coded Level			
	Vr	Fr	Vf	Ff	Vr	Fr	Vf	Ff
1	200	0.20	180	0.15	-	-	-	-
2	250	0.20	180	0.20	+	-	-	+
3	200	0.26	180	0.20	-	+	-	+
4	250	0.26	180	0.15	+	+	-	-
5	200	0.20	250	0.20	-	-	+	+
6	250	0.20	250	0.15	+	-	+	-
7	200	0.26	250	0.15	-	+	+	-
8	250	0.26	250	0.20	+	+	+	+

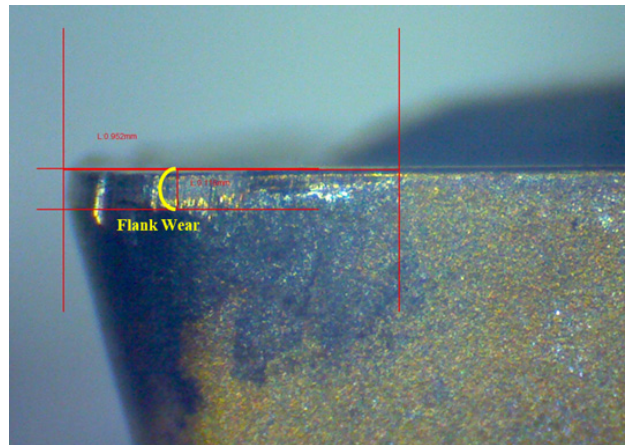


Figure 5. Flank Wear of Finishing Tool

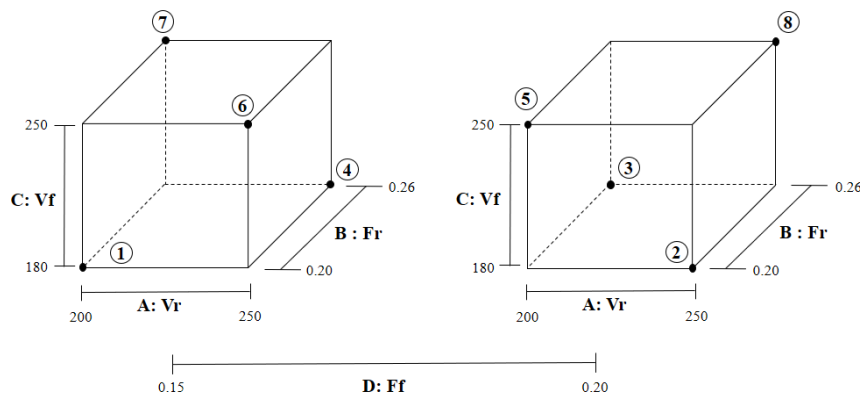


Figure 4. Design Space

Table 3. Averages and Standard Deviations of the Outer Diameters from the Design Points

No	Vr	Fr	Vf	Ff	#1-1		#1-2		#1-3	
					Averages	Standard Deviations	Averages	Standard Deviations	Averages	Standard Deviations
1	-	-	-	-	25.9128	0.0101	25.9138	0.0100	25.9140	0.0113
					25.9146	0.0116	25.9174	0.0130	25.9145	0.0115
2	+	-	-	+	25.9049	0.0104	25.9060	0.0106	25.9064	0.0103
					25.9248	0.0091	25.9140	0.0097	25.9155	0.0108
3	-	+	-	+	25.9150	0.0114	25.9175	0.0118	25.9185	0.0126
					25.9161	0.0113	25.9133	0.0100	25.9144	0.0091
4	+	+	-	-	25.9134	0.0123	25.9147	0.0110	25.9133	0.0117
					25.9126	0.0107	25.9150	0.0113	25.9139	0.0111
5	-	-	+	+	25.9097	0.0117	25.9109	0.0116	25.9109	0.0119
					25.9036	0.0113	25.9072	0.0125	25.9038	0.0118
6	+	-	+	-	25.9147	0.0100	25.9152	0.0090	25.9164	0.0094
					25.9029	0.0088	25.9046	0.0102	25.9016	0.0093
7	-	+	+	-	25.9181	0.0080	25.9182	0.0083	25.9180	0.0086
					25.9135	0.0099	25.9119	0.0090	25.9131	0.0094
8	+	+	+	+	25.9262	0.0106	25.9255	0.0113	25.9249	0.0113
					25.9161	0.0082	25.9167	0.0092	25.9168	0.0098

살펴서 공구 불량과 관련된 후피와 버(Burr) 불량 여부만 조사한다. 후피는 열간 압연 가공 등으로 만들어진 철강 재료의 산화 피막으로서 가공 시 제거되어야 하는 부분이다. 버는 가공 후 잔여 부위로서 대부분 날카로운 형상을 지닌다. 치수 데이터는 자동검사기를 통해 T/P Spider의 외경#1-1, #1-2, #1-3을 측정된 것이다.

8개의 실험점 별로 세 군데 위치에서 외경 치수의 평균과 표준편차를 계산한 것을 <Table 3>에, 외경 치수 표준편차의 평균을 <Figure 6>에 나타내었는데, 모든 점에서 표준편차의 평균이 0.012 미만이며, 2, 6, 7, 8번 조건의 표준편차가 비교적 작다. <Table 4>에 각 실험점의 총생산수량, 공구 플랭크 마모도, 외관 불량인 버가 나타난 가공품의 수, 치수 규격을 벗어난 가공품의 수, 불량률을 제시한다. 후피로 인한 외관 불량은 나타나지 않아서 표시하지 않는다. 불량률은 외관과 치수로 인한 불량품 수량의 합을 총생산수량으로 나눈 것이다. <Figure 7>

은 각 실험점의 불량률 그래프이다. 4번, 8번 조건에서 불량이 발생하지 않으며, 5번 조건은 불량률이 가장 높다. <Figure 8>은 실험점 별 공구 플랭크 마모 데이터를 나타낸 그래프이다.

실험데이터를 통계적 방법으로 분석하였는데, 주효과와 교호작용효과가 유의하지 않아서 실험점 중에서 가장 좋은 성과를 내는 실험점을 찾는다. 4번, 8번 실험점이 불량률이 없고, 표준편차가 크지 않으며, 공구 마모도 크지 않아서 전반적으로 우수하다. <Table 2>를 보면 4번과 8번 가공조건은 (Vr, Fr, Vf, Ff)가 각각 (250, 0.26, 180, 0.15), (250, 0.26, 250, 0.20)이다. 이들을 비교하면, (Vr, Fr)는 둘 다 (250, 0.26)으로 같고, (Vf, Ff)는 서로 다르다. Vf나 Ff가 크다는 것은 그만큼 생산을 신속하게 할 수 있어서 생산성이 높다는 것이므로 (Vf, Ff)가 둘 다 큰 8번 가공조건, 즉, (Vr, Fr, Vf, Ff) = (250, 0.26, 250, 0.20)을 최종 가공조건으로 선정한다.

Table 4. Flank Wear of Tool, Number of Defects, and Defectives from the Design Points

No	Vr	Fr	Vf	Ff	# of Machined Items	Flank Wear of Tool		# of Defects		Defectives (%)
						Roughing	Finishing	Burr	Outer Diameter	
1	-	-	-	-	374	0.275	0.097	1	0	0.267
2	+	-	-	+	392	0.204	0.110	1	0	0.255
3	-	+	-	+	390	0.228	0.120	1	0	0.256
4	+	+	-	-	380	0.202	0.096	0	0	0.000
5	-	-	+	+	387	0.176	0.110	1	9	2.584
6	+	-	+	-	379	0.236	0.110	3	0	0.792
7	-	+	+	-	362	0.203	0.105	1	0	0.276
8	+	+	+	+	396	0.214	0.127	0	0	0.000

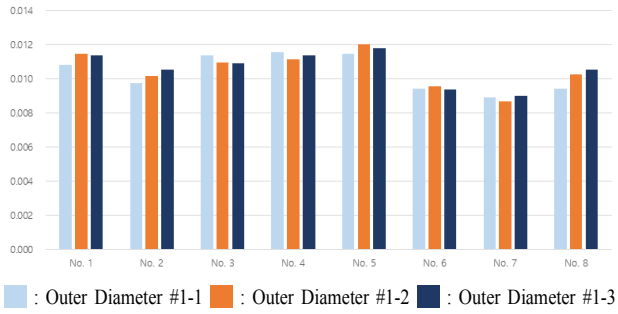


Figure 6. Outer Diameter Standard Deviations from 8 Design Points

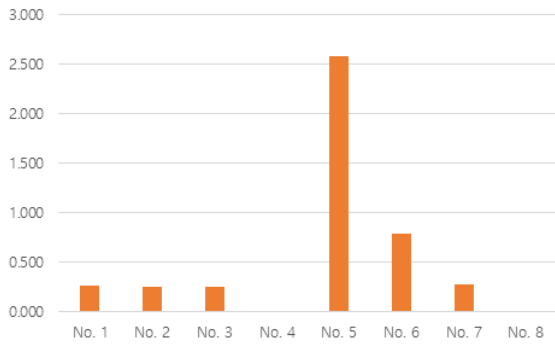


Figure 7. Defectives from the Design Points

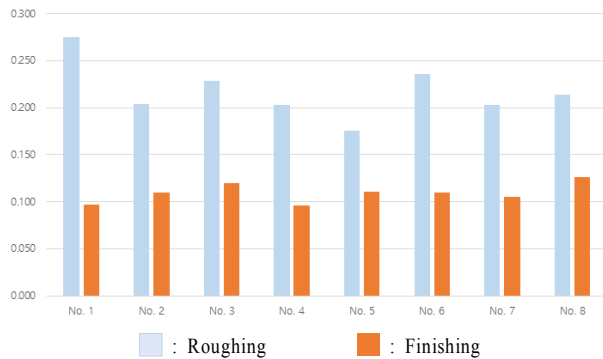


Figure 8. Flank Wear of Tool from the Design Points

최적 가공조건에서 확인 실험을 두 번 시행하여 나온 가공물 치수의 표준편차, 불량 여부, 공구 마모도를 구한다. <Table 5>는 (Vr, Fr, Vf, Ff) = (250, 0.26, 250, 0.20) 공정 조건에서 시행한 4차례의 실험결과를 정리한 것이다.

3.2 기존 조건과 최적 조건 비교

<Table 6>은 최적 가공조건과 기존 조건을 외경 위치별 표준편차, 생산성, 불량률, 품질손실 측면에서 비교한 것이다. 첫째, 외경 위치별 표준편차는 최적 조건에서 다소 감소했다. 둘째, CNC 가공공정의 생산성 관점에서 기존과 최적 조건을 비교한다. CNC 공정의 가공시간은 절삭속도(V)와 이송속도(F)에 반비례한다. 그러므로 절삭속도(V)와 이송속도(F)가 클수록 가공시간이 줄어들어 생산성이 높아진다. <Table 6>을 보면, 두 가공조건에서 이송속도는 같거나 거의 차이가 없고, 절삭속도는 최적 조건에서 더 빠르다. 기존 조건과 최적 조건의 절삭속도를 비교하여 최적 조건에서 가공시간의 단축, 즉 생산성 증가분을 알아본다. 최적 조건에서 황삭절삭속도(Vr)가 200에서 250으로 증가하여 단축된 가공시간과 정삭절삭속도(Vf)가 180에서 250으로 높아짐에 따라 발생한 시간 단축분은 각각 식 (4)와 식 (5)에 나타내었다. 최적 조건에서 생산성은 기존 조건에 비해 황삭공정에서 20%, 정삭공정에서 28%씩 증가한다.

$$\frac{1}{200} - \frac{1}{250} = \frac{250 - 200}{250 \times 200} = \frac{250 - 200}{250} = 0.20 \quad (4)$$

$$\frac{1}{180} - \frac{1}{250} = \frac{250 - 180}{250 \times 180} = \frac{250 - 180}{250} = 0.28 \quad (5)$$

셋째, 불량률 관점에서 보면, 기존 조건은 총 977개의 가공부품에서, 흑피 1개, 버 3개의 외관 불량이 발생하여 불량률이

Table 5. Performance Results from the Optimal Design Points

No	# of Machined Items	Standard Deviations of the Outer Diameters		# of Defects		Flank Wear of Tool		
				Burr	Outer Diameter	Roughing	Finishing	
Opt Pt	First Rep.	198	#1-1	0.0106	0	0	0.184	0.120
			#1-2	0.0113				
			#1-3	0.0113				
	Second Rep.	198	#1-1	0.0082	0	0	0.244	0.133
			#1-2	0.0092				
			#1-3	0.0098				
First Conformation Exp.	189	#1-1	0.0096	0	0	0.192	0.132	
		#1-2	0.0104					
		#1-3	0.0107					
Second Conformation Exp.	192	#1-1	0.0084	0	0	0.192	0.112	
		#1-2	0.0087					
		#1-3	0.0086					

Table 6. Comparison of the Current and Optimal Conditions

Cond.	Vr	Fr	Vf	Ff	Standard Deviations of the Outer Diameters		Machining Time Reduction	Defective	Expected Quality Loss
					#1-1	#1-2			
Current	200	0.25	180	0.20	#1-1	0.0097		0.4%	367
					#1-2	0.0099			
					#1-3	0.0097			
Optimal	250	0.26	250	0.20	#1-1	0.0091	Roughing: 20% Finishing: 28%	0.0%	206
					#1-2	0.0098			
					#1-3	0.0099			

0.4%인데, 최적 조건에서는 총 777개 부품 모두 양품이라서 불량률은 0%이다. 넷째, 규격 내에 있더라도 제품의 특성치가 목표값에서 멀어지면 품질손실이 증가한다는 다구치의 손실함수 개념을 이용하여 기대손실(Expected Loss) 측면에서 품질비용을 비교해 본다. 기대손실(Expected Loss) EL은 다음의 식(6)과 같이 구한다(Phadke, 1989).

$$EL = \frac{A_0}{\Delta_0^2} \{(\bar{X}-t)^2 + s^2\} \quad (6)$$

식에서 A_0 는 불량이 발생할 때 나타나는 개당 부품비용으로서 대상 부품의 가격은 1,844원이고, Δ_0 은 허용공차로서 0.035mm이다. \bar{X} 는 데이터의 평균을, t 는 목표값을, s 는 데이터의 표준편차이다. 본 사례에서 부품의 규격은 $t \pm \Delta_0 = 25.915 \pm 0.035$ 이다. 식(6)에 $A_0 = 1840$, $\Delta_0 = 0.035$, $t = 25.915$, 기준과 최적 조건에서 치수 데이터 평균과 표준편차 $(\bar{X}, s) = (25.903, 0.010)$ 과 $(25.919, 0.011)$ 을 대입하면, 기준과 최적 가공조건의 부품 단위당 기대손실은 각각 367원, 206원이다. T/P Spider 제품의 월평균 생산량이 5만 개이므로 최적 조건으로 가공할 경우, 연간 기대손실 절감액은 $(367 - 206) \times 50,000 \times 12 = 96,600,000$ 원이다.

4. 가공품 품질예측을 위한 머신러닝 모델

다양한 조건에서 실험을 진행하여 피쳐 데이터와 레이블 데이터를 수집하여, 이를 기반으로 품질예측모델을 개발하기 위하여 파이썬의 Sklearn, Numpy, Pandas, Matplotlib 등의 라이브러리를 사용한다. 가공품 치수 데이터의 치수 값과 양-불량 여부를 레이블 데이터로 사용하여 회귀모델과 분류모델을 각각 적용한다. 가공품 각각에 대한 외관 검사 불량 여부를 파악하지 못하여 외관 불량에 대한 예측모델은 개발하지 못했다.

4.1 데이터 구성

실험을 진행하고, 각 실험조건에서 구한 결과를 하나의 파일로 통합해 레이블 데이터와 피쳐 데이터를 구성한다. <Table 7>에 7개의 피쳐를 나열하였다. Vr, Fr, Vf, Ff는 실험에서 설정한 가공공정 변수인 황삭과 정삭의 절삭속도와 이송속도이다. ‘가공품 번호(Number)’는 해당 실험조건의 몇 번째 가공품인

지 나타내는 값인데, 최소값이 1, 최대값이 199이다. ‘오프셋 보정 값(Off_value)’은 실험 중 작업자가 오프셋 보정 당시 수기로 기록한 좌표의 조정값으로, 최소값은 -0.06, 최대값은 0.07이다. 공구 마모 등으로 외경 치수가 관리 치수보다 커지게 되면 보정 값을 그 차이만큼 양수로 설정하며, 작으면 그만큼 음수로 설정한다. ‘오프셋 보정 후 가공 수량(Off_count)’은 작업자가 오프셋 보정 후 몇 번째에 가공하는 제품인지 기록한 값으로, 최소값이 1, 최대값이 85이다. 레이블 데이터는 외경 치수이며, 외경 치수 값들과 치수의 불량 여부는 {양품, 불량품} = {1, 0}으로 구성되어 있다(<Table 8>). 네 가지 공정변수를 이용하여 8개의 실험조건에서 반복 2회 실험을 진행하였고, 각 실험에서 180~200개의 치수 데이터를 얻었다. 치수 데이터에 대응하는 피쳐 데이터로는 공정변수 외에 ‘가공품 번호’, ‘오프셋 보정 값’, ‘오프셋 보정 후 가공 수량’도 수집하였다. 이들 세 개의 추가 피쳐 데이터는 실험조건과 무관하게 변하는 값이므로, 같은 공정조건에서 가공된 부품일지라도 다른 값을 가진다. 피쳐 데이터와 레이블 데이터를 구성하고 난 후, 외경 치수의 값을 레이블로 하는 가공품 치수 예측모델과 치수 불량여부를 레이블로 하는 가공품 치수 불량여부 예측모델을 각각 개발하고, 이를 구현한다.

Table 7. Feature Data

No	Data	Definition
1	Vr	Roughing Speed
2	Fr	Roughing Feed Rate
3	Vf	Finishing Speed
4	Ff	Finishing Feed Rate
5	Number	Part Number
6	Off_value	Offset Compensation Value
7	Off_count	Serial Part Number after Offset

Table 8. Label Data

No	Data	Definition	Note
1	OD1_1	Outer Diameter Measurement	#1-1
	OD1_2		#1-2
	OD1_3		#1-3
			Specification: [25.880, 25.950]
2	OK/NG	Outer Diameter Defectiveness Judgment	Defect: 0, Non-Defect: 1

4.2 가공품 치수 예측모델

가공품 치수 예측모델은 7가지 피처에 대응하는 레이블 데이터 외경#1-1, 외경#1-2, 외경#1-3의 값을 예측하는 모델이다. 가공품 치수 예측모델의 알고리즘은 다음의 순서를 따른다.

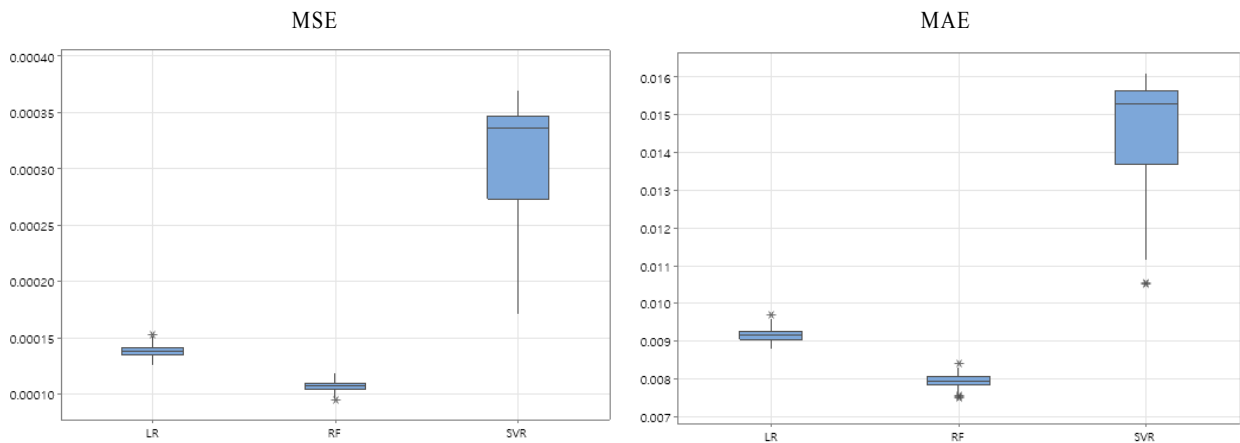
- (1) 훈련 데이터, 테스트 데이터 분할 : 모델의 과적합(Overfitting)을 방지하기 위해 훈련 데이터와 테스트 데이터를 7:3으로 나눈다.
- (2) 모델 학습 : 해당 모델은 연속적인 값을 가지는 외경 치수 데이터를 예측하는 것인데, 주요 회귀모델 기법인 선형회귀, 랜덤포레스트 회귀, 서포트벡터 회귀를 사용한다.
- (3) 모델 성능평가 : 학습한 모델의 성능을 평가하기 위한 지표로서 평균제곱오차(Mean Squared Error; MSE)와 평균절대오차(Mean Absolute Error; MAE)를 활용한다. 학습한 모델에 테스트 데이터를 입력하여 예측값을 구하고, 모델별 MSE와 MAE를 계산하여 성능을 확인한다.
- (4) 데이터 분할, 모델 학습과 평가의 반복 : 데이터 분할을 랜덤하게 100번 시행하여, 모델 학습, 성능평가를 매번 다른 데이터로 수행하고, MSE, MAE의 평균과 상자그림(Box Plot)을 확인하여 최종 모델을 선정한다.

세 가지 외경별로 가공품 치수 예측모델은 데이터를 훈련 데이터와 테스트 데이터로 나눈 후, 훈련 데이터를 이용하여 선형회귀와 랜덤포레스트 회귀, 서포트벡터 회귀모델을 학습한다. 선형회귀의 경우, 모델의 절편 계산 여부와 훈련 데이터

복사 여부 모두 'True'로 설정하고, 랜덤포레스트 회귀의 경우, 결정트리 개수는 100개, 결정트리 노드를 분지할 때 사용하는 불순도 측정 방식은 'Squared_error', 결정트리 최대 깊이는 5로 하이퍼파라미터를 설정한다. 또한 서포트벡터 회귀의 경우, 커널 유형은 'RBF', 커널 계수는 'Scale'로, 정규화 매개변수는 1로 설정한다. 선형회귀와 서포트벡터 회귀 모델은 하이퍼파라미터를 따로 설정하지 않고, 파이썬 함수의 기본 값을 사용한다. 학습한 모델에 테스트 데이터의 피처 데이터를 이용하여 레이블 데이터 예측값을 도출하고, 도출한 예측값과 테스트 데이터의 실제값을 바탕으로 성능평가 지표인 MSE와 MAE를 계산한다. 전체 데이터를 훈련 데이터와 테스트 데이터로 랜덤하게 나누는 단계부터 테스트 데이터로 모델 성능을 평가하는 단계까지 총 100번 반복을 진행한다. <Table 9>에 외경 위치별 모델의 MSE와 MAE 평균값을 계산하여 나타내고, 100번 시행한 성능평가 결과값의 분포는 <Figure 9>, <Figure 10>, <Figure 11>에 상자그림으로 표현한다. <Table 9>에서 치수예측 성능평가의 평균을 보면, 모든 외경에 대해 MSE와 MAE 측면에서 랜덤포레스트 회귀모델이 가장 우수한 성능을 보인다. 3개의 상자그림은 다른 두 모델에 비해 랜덤포레스트 회귀모델의 성능평가 결과값의 산포가 크지 않고, MSE와 MAE의 평균이 낮다는 것을 시각적으로 보여준다. 따라서, (외경#1-1, 외경 #1-2, 외경#1-3)의 MSE와 MAE가 각각 (0.00011, 0.00010, 0.00010)과 (0.00795, 0.00772, 0.00788)인 랜덤포레스트 회귀를 최종 모델로 선정한다.

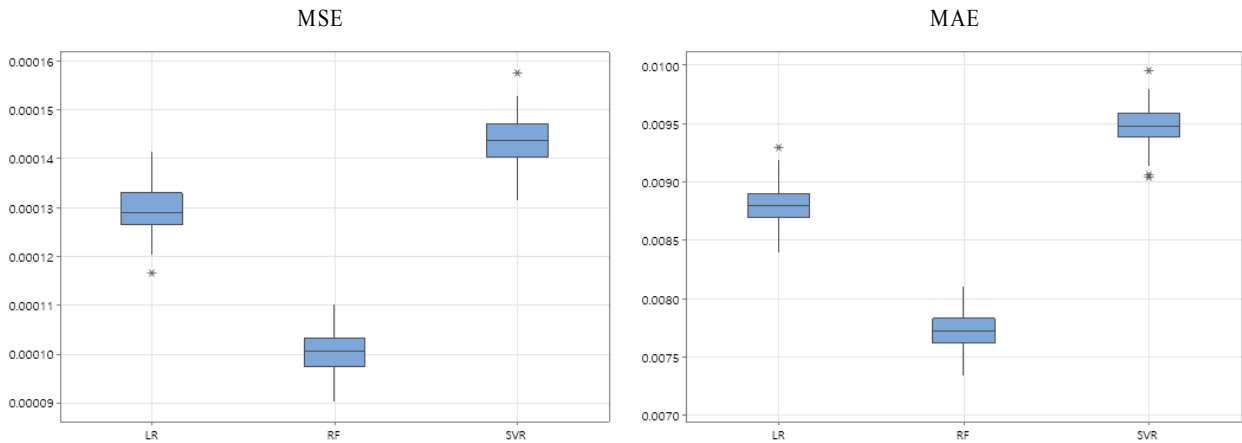
Table 9. Average Performance Indices of the Three Outer Diameters Prediction Models

		Linear Regression		Random Forest Regression		Support Vector Regression	
		MSE	MAE	MSE	MAE	MSE	MAE
Outer Diameter	#1-1	0.00014	0.00918	0.00011	0.00795	0.00031	0.01463
	#1-2	0.00013	0.00881	0.00010	0.00772	0.00014	0.00948
	#1-3	0.00013	0.00090	0.00010	0.00788	0.00015	0.00987



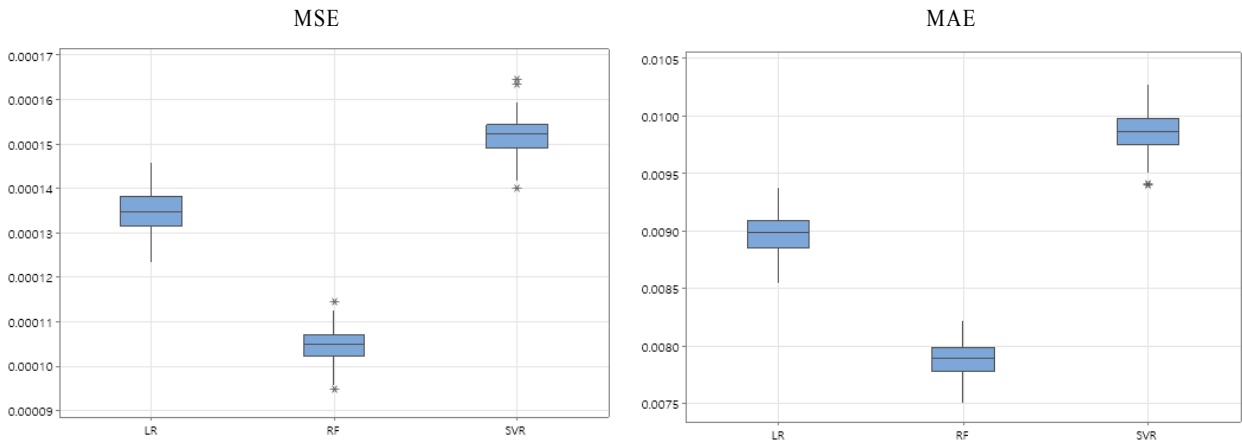
LR: Linear Regression, RF: Random Forest Regression, SVR: Support Vector Regression

Figure 9. Performance Comparisons of the Three Algorithms for Outer Diameter #1-1



LR: Linear Regression, RF: Random Forest Regression, SVR: Support Vector Regression

Figure 10. Performance Comparisons of the Three Algorithms for Outer Diameter #1-2



LR: Linear Regression, RF: Random Forest Regression, SVR: Support Vector Regression

Figure 11. Performance Comparisons of the Three Algorithms for Outer Diameter #1-3

4.3 가공품 치수 불량여부 예측모델

가공품 치수 불량여부 예측모델은 제품의 치수 데이터가 규격 내에 들어와 양품으로 또는 규격을 벗어나 불량품으로 판정되는지를 예측하는 모델이다. 모델 학습에 사용한 피쳐 데이터는 가공품 치수 예측모델과 같고, 레이블 데이터는 치수 불량 여부로 구성되어 있다. 가공품 치수 불량 여부 예측모델을 구축하는 순서는 다음과 같다.

- (1) 훈련 데이터, 테스트 데이터 분할 : 모델의 과적합(Overfitting)을 방지하고자 훈련 데이터와 테스트 데이터를 7:3의 비율로 데이터를 나눠 각각 다른 데이터로 모델을 학습하고 성능을 평가한다.
- (2) 데이터 불균형 처리 : 양품 데이터에 비해 불량품 데이터의 수가 현저히 작으면, 모델이 모두 양품만 예측하게 될 가능성이 커 불량품을 예측하는 모델을 개발하기 어렵다. 이런 경우에 유용한 ADASYN을 적용해 소수 클래스를 오버샘플링 한다.

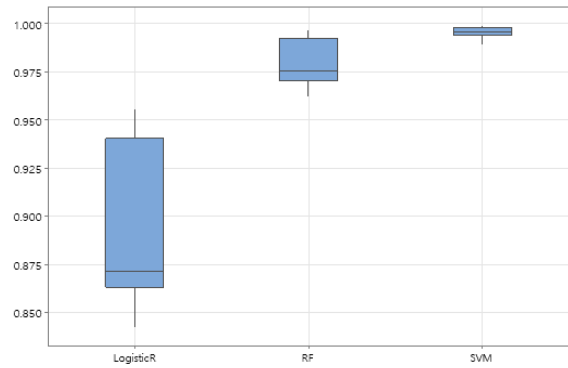
- (3) 모델 학습 : 레이블이 치수 불량 여부를 나타내는 0(불량품)이나 1(양품)의 값만 가지므로 훈련 데이터를 이용하여 분류모델을 학습하는데, 로지스틱 회귀, 랜덤포레스트, 서포트벡터머신을 사용한다.
- (4) 모델 성능평가 : 테스트 데이터를 학습한 모델에 입력하여 모델의 예측 결과를 확인하고, 이를 실제 테스트 데이터의 레이블과 비교한다. 비교한 결과를 바탕으로 정확도, 특이도, F1 점수를 계산하여 학습한 모델의 성능을 평가한다.
- (5) 데이터 분할, 불균형 처리, 모델 학습과 평가의 반복 : 실험데이터 3,998개 중 불량 데이터는 9개로 양품 데이터에 비해 매우 적으므로, 테스트 데이터에 불량이 몇 개 할당되는지에 따라 성능지표의 결과가 큰 폭으로 변화한다. 예를 들어, 테스트 데이터에 불량품 1개가 포함되어 있고, 그것을 예측했다면 특이도는 1.0인데, 테스트 데이터의 불량품 개수가 2개일 때 불량품 1개를 맞출 경

우, 특이도는 0.5이다. 따라서, 데이터 분할을 랜덤하게 100번 시행하고, 각 시행에서 얻은 서로 다른 훈련 데이터와 테스트 데이터를 이용하여 (2) ~ (4) 단계를 반복하여 수행한다. 각 모델의 정확도, 특이도, F1 점수의 평균과 산포 정도를 감안하여 최종 모델을 선정한다.

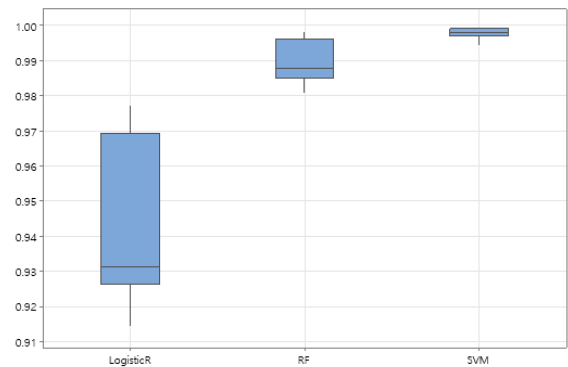
ADASYN을 이용하여 데이터 불균형 처리를 한 데이터로 모델 학습을 진행한다. 훈련 데이터를 이용하여 로지스틱 회귀, 랜덤포레스트, 서포트벡터머신 순서로 모델을 학습한다. 모델 학습 과정에서 로지스틱 회귀의 경우, 최대 반복 횟수를 5,000으로 설정한다. 랜덤포레스트의 경우, 결정트리 개수 150개, 결정트리 학습에 사용할 샘플 수 50개로 하이퍼파라미터를 지정한다. 서포트벡터머신의 경우는 따로 지정하지 않고 파이션 함수의 기본 값을 사용하여, 정규화 매개변수는 1, 커널 유형은 'RBF', 커널 계수는 'Scale'로 설정한다. 학습한 모델 각각에 테스트 데이터의 피쳐 데이터를 이용하여 레이블 데이터의 예측값을 구하고, 구한 예측값과 테스트 데이터의 실제값을 이용하여 성능평가 지표인 정확도, F1 점수, 특이도를 계산한다. 4.2절과 같이 훈련 데이터와 테스트 데이터 분할을 100회 시행하여 얻은 서로 다른 데이터 세트로 모델을 학습하고 성능을 평가하여 구한 세 가지 지표의 평균값을 <Table 10>에 제시하고, 성능평가 지표 값의 분포를 상자그림으로 나타낸다(<Figure 12> ~ <Figure 14>). <Table 10>을 보면 서포트벡터머신 모델의 정확도와 F1 점수는 가장 높으나, 특이도는 가장 낮다. 품질 예측을 할 때 특이도는 '실제 불량인 것을 불량으로 판단하는 지표'로서 가장 중요한 지표이므로 서포트벡터머신은 최종 모델로 선정하지 않는다. 한편, 로지스틱 회귀 모델의 특이도는 가장 높지만, 정확도와 F1 점수가 가장 좋지 않은 성능을 가진다. 랜덤포레스트 모델은 서포트벡터머신 모델과 비교하면 정확도와 F1 점수는 0.01 ~ 0.02 정도 작지만, 특이도는 0.1 정도 높고, 로지스틱 회귀모델과 비교하면 특이도는 0.005 정도 작지만, 정확도와 F1 점수는 0.04 ~ 0.09 정도 높다. 3개의 성능지표를 종합적으로 고려할 때, 랜덤포레스트 모델을 쓰는 것이 가장 바람직하다고 판단한다. 또한 <Figure 14>로부터 알 수 있듯이, 랜덤포레스트 모델의 특이도가 0이 되는, 즉 불량 탐지에 실패하는 경우가 서포트벡터머신 모델보다 적게 나타난다. 분석 결과를 바탕으로 본 사례에서는 (정확도, F1 점수, 특이도) = (0.9796, 0.9896, 0.8846)인 랜덤포레스트를 최종 모델로 선정한다.

Table 10. Average Performance Indices of the Three Outer Diameters Defectiveness Judgment Prediction Models

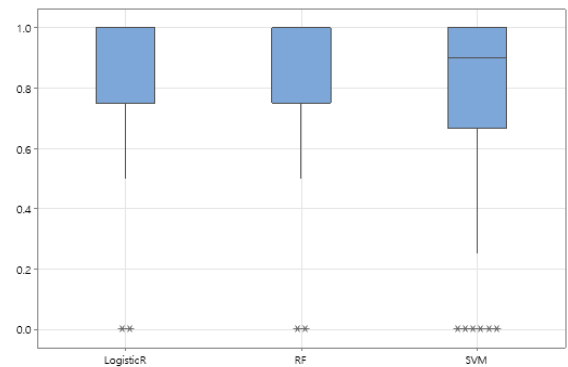
	Logistic Regression	Random Forest	Support Vector Machine
Accuracy	0.8896	0.9796	0.9959
F1 Score	0.9411	0.9896	0.9979
Specificity	0.8899	0.8846	0.7784



Logistic R: Logistic Regression, RF: Random Forest, SVM: Support Vector Machine
Figure 12. Box Plots of Defectiveness Judgment Accuracy for Three Prediction Models



LogisticR: Logistic Regression, RF: Random Forest, SVM: Support Vector Machine
Figure 13. Box Plots of Defectiveness Judgment F1 Score for Three Prediction Models



Logistic R: Logistic Regression, RF: Random Forest, SVM: Support Vector Machine
Figure 14. Box Plots of Defectiveness Judgment Specificity for Three Prediction Models

4.4 가공품의 치수 불량 발생 조건

제3장에서 최적 가공조건을 제시하였지만, 공정의 운영담당자는 치수 불량이 발생하는 공정조건에도 관심이 있어서 가공품의 치수가 규격을 벗어나는 조건을 파악해 보았다. 치수 불량이 발생하는 공정조건을 도출하기 위하여 의사결정나무

Table 11. Outer Diameter Defect Generating Conditions

Common Conditions		Additional Conditions		
		Number	Off_value	Off_count
Vr:	175 ~ 225	100 ~ 111	-0.015 ~ -0.011	
Fr:	0.23 ↑	111 ↑		11 ↑
Vf:	215 ↑			
Off_count:	10 ~ 21	111 ↑	-0.015 ~ -0.006	

알고리즘을 이용한다. 우선 훈련 데이터와 테스트 데이터를 7:3의 비율로 분할하고, ADASYN을 이용하여 데이터 불균형 처리를 진행한다. 그 후, Sklearn 라이브러리를 사용하여 의사결정나무 모델을 학습하고, Graphviz 라이브러리를 이용하여 모델을 시각화한다. 의사결정나무는 모델 시각화를 위해 사용한 알고리즘이므로, 하나의 시각화 결과만을 확인하기 위해 데이터 분할부터 시각화까지 1회 진행한다.

의사결정나무를 이용한 모델 시각화 결과를 정리하면 다음과 같다. 불량 발생 하는 공통조건은 Vr: 175 ~ 225, Fr: 0.23 이상, Vf: 215 이상, 오프셋 보정 후 가공수량: 10 ~ 21이다. 이러한 공통조건과 다음 중 하나의 조건이 일어나면 불량이 발생한다. (1) 가공품 번호: 100 ~ 111, 오프셋 보정 값: -0.015 ~ -0.011, (2) 가공품 번호: 111 이상, 오프셋 보정 후 가공수량: 11 이상, 오프셋 보정 값: -0.015 ~ 0, (3) 가공품 번호: 111 이상, 오프셋 보정 값: -0.015 ~ -0.006. 불량이 발생하는 조건을 정리하여 <Table 11>에 나타낸다.

5. 공정개선과 품질예측의 방향

5.1 현행 공정의 개선사항

실험을 진행하고 결과를 분석하면서 현행 공정 운영의 문제점을 찾아본다. 먼저 같은 조건에서 반복 2회 가공할 때, 부품의 외경 치수가 다르게 나타나는 현상을 살펴본다. <Figure 15>는 같은 실험조건에서 생산한 가공품 외경#1-1의 치수 데이터인데, 반복 1회와 반복 2회에서 나온 데이터의 평균값에

차이가 크다. 이러한 차이는 가공 시작 전에 작업자가 절삭 깊이를 다르게 세팅했기 때문에 나타난 것으로 판단한다. <Figure 15>로 알 수 있는 또 다른 사실은 각 반복 회차에서 데이터의 평균이 목표값(Target)에서 벗어난 것이다. 식 (6)에서 알 수 있듯이 이렇게 평균값이 목표값에서 멀어지면 품질의 기대손실은 벗어난 편차의 제곱에 비례하여 커지고 이는 결국 기업의 품질비용으로 나타난다. 따라서, 공정의 책임자는 가공 전에 주요 공정변수를 세팅하는 방법 등에 관한 표준운영절차(Standard Operating Procedure; SOP)를 수립해야 한다.

두 번째는 치수 데이터의 급격한 변화에 관한 사항이다. <Figure 16>은 특정 조건에서 외경#1-1, 외경#1-2, 외경#1-3 치수의 시계열 변동을 나타낸 그래프이다. 가공번호 50번과 140번에서 치수가 한꺼번에 급격히 높아지는 경향을 볼 수 있다. 이러한 급격한 치수의 변화는 작업장 온도기 가동 등 작업장의 대기 온도 변화의 영향으로 예상된다. Yang(1999)은 CNC 선반의 온도는 위치에 따라 10도까지 변하는데, 온도의 변화로 발생하는 열팽창으로 인해 가공물의 직경 오차가 35μm까지 늘어나는 사실을 밝혀냈다. 과연 작업장 온도가 가공물의 외경 치수에 영향을 미치는지 확인하기 위하여 같은 시간대에 외경 치수와 작업장 온도를 측정하기로 했다. <Figure 17>의 (a)는 가공품 외경의 시계열 데이터이다. 치수 데이터의 전반적 경향을 더 선명하게 파악하기 위해 (a)의 치수 데이터 10개씩 이동평균(Moving Average)을 계산하여 <Figure 17>의 (b)에 나타낸다. 이 그래프를 보면, 치수 데이터가 시간의 흐름에 따라 상승하는 경향을 알 수 있다. 작업장 온도는 연속적으로 젤 수가 없어서 작업자가 CNC 기계 주변 온도를 3차례 측정된 것을 연결하여 <Figure 17>의 (c)와 같이 나타낸다. <Figure 17>의 (b)와 (c)를 비교해 보면, 두 그래프가 비슷한 추세를 따르므로 온도가 부품의 치수에 영향을 미친다는 점을 확인할 수 있다. 치수 데이터가 지속해서 증가하는 원인은 온도 외에도 공구 마모로 인해 부품의 직경이 느리게 증가하는 현상 때문일 수도 있으나, 본 사례의 경우, 중간에 작업자가 오프셋 보정을 수행하였음에도 그래프가 계속 증가하므로 작업장 온도가 주원인이라고 판단한다. 또한, <Figure 17>의 외경#1-1, 외경#1-2, 외경#1-3 치수 데이터와 온도 데이터 간

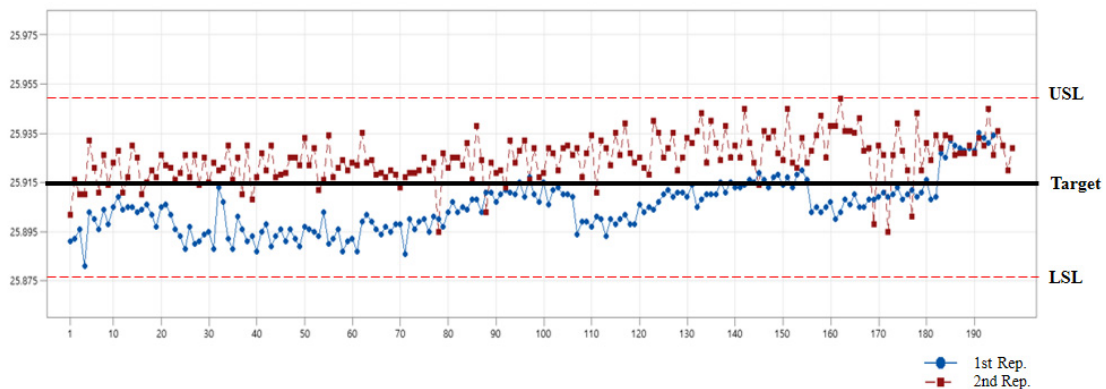


Figure 15. Outer Diameter #1-1 Time Series Data of Two Replications from the Same Design Point

상관계수는 각각 0.842, 0.803, 0.826로 높게 나와 상관성이 높음을 확인하였다. 치수 데이터에서 이러한 온도 영향을 제외하고 싶다면, ‘(원래 데이터 값) - (이동평균을 계산한 값)’을 확인하면 된다(<Figure 18>). <Figure 18>을 보면, 잡음의 변동이 잘 관리

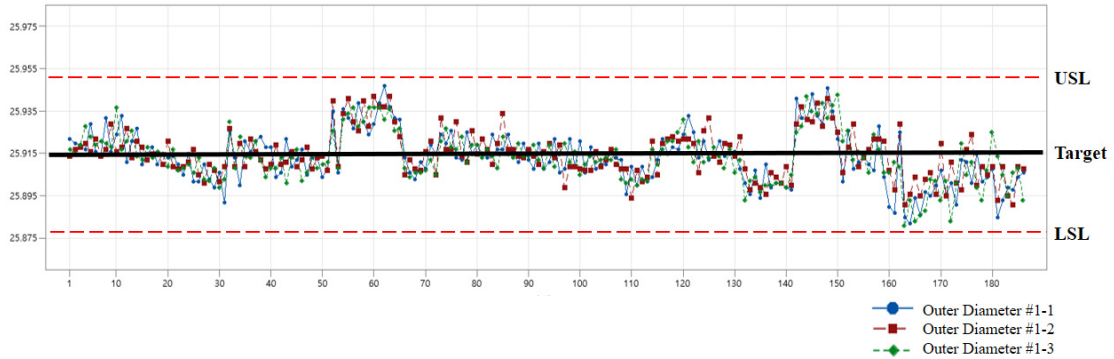
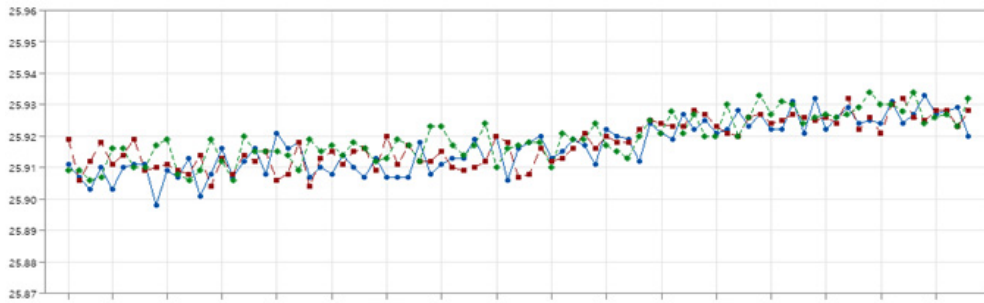
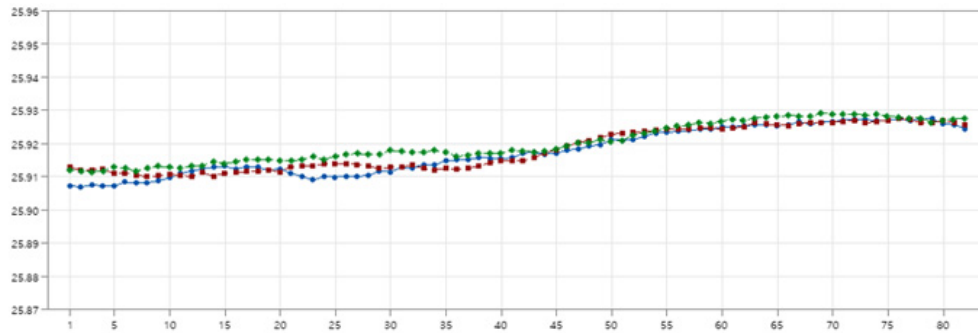


Figure 16. Time Series Data of the Outer Diameter #1-1, #1-2, and #1-3

(a) Outer Diameter Time Series Data:



(b) Outer Diameter Time Series Moving Average Data:



(c) Temperature Data:

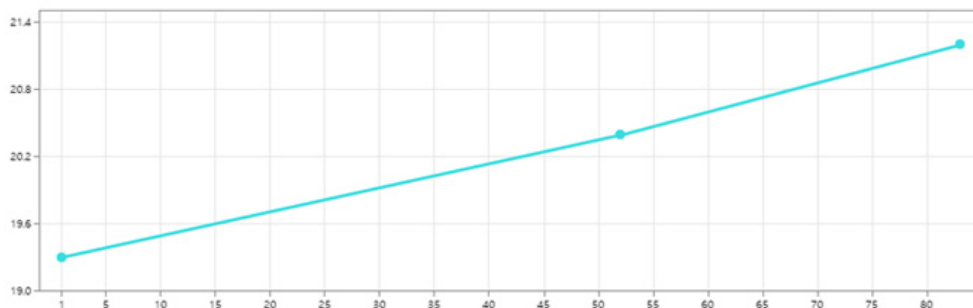


Figure 17. An Illustration of Outer Diameter Time Series and Temperature Data

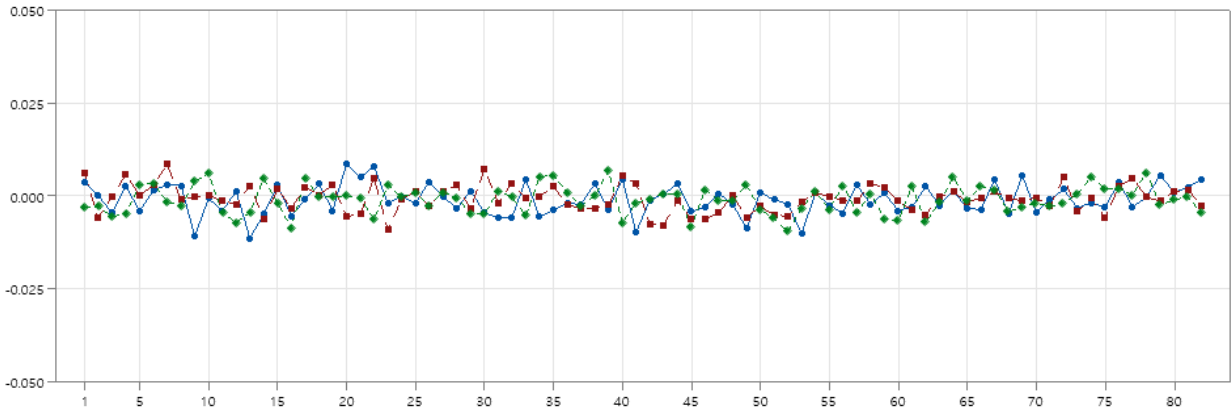


Figure 18. Outer Diameter Time Series Data Without Moving Average Effect

되고 있다고 판단된다. 추가로, <Figure 16>에서 후반부에 나타난 높은 치수 변동은 공구 마모로 인한 현상으로 보인다.

이 외에도 CNC 공정 및 추후 타 공정에서 품질예측 모델 관련 연구를 진행할 때는 데이터의 추적이 필요하다고 판단된다. 제4장에서 언급했다시피 가공품 각각에 대한 불량 여부 및 유형을 파악하지 못해 피쳐 데이터와 레이블 데이터를 연결하는 것이 어려워서 외관 불량 유형 예측모델을 개발하지 못했다. 기존의 연구들은 레이블 데이터로 양·불량 여부만 수집하나, 불량은 공정의 각 특성치가 어떤 값을 가지는지에 따라 각기 다른 불량 유형이 나오므로, 공정을 실제로 개선하기 위해서는 불량 유형을 알고 불량 유형별로 데이터를 추적할 수 있어야 한다.

5.2 피쳐 데이터의 확보

본 연구에서는 센서를 활용하여 다양한 피쳐 데이터와 대용량의 데이터를 구하기 어려운 중소기업에서 실험계획 방법을 이용하여 별도의 피쳐 데이터를 생성하고 머신러닝 알고리즘을 적용하여 품질예측 모델을 구하는 방법을 제시하였다. 실험계획법을 이용하여 피쳐의 수를 늘리는 데에는 한계가 있으므로 피쳐를 추가로 확보하기 위한 노력이 필요하다. 우선

<Figure 19>에 제시한 바와 같이 제조공정은 재료, 선행가공품, 정보를 투입물(Input)로 받아들여 통제할 수 있는 공정변수(Controllable Variables)의 값을 설정한 다음에 가공하여 산출물(Output)을 얻는다. 이때 산출물의 중요한 품질특성(Critical-To-Quality; CTQ)을 측정 또는 평가하여 문제가 있으면 공정변수의 값을 조정한다. 그런데 산출물의 CTQ에 영향을 미치는 변수에는 통제할 수 없거나 통제하기 어려운 잡음변수(Noise Variables)가 있는데, 이들은 공정에서 통제하지 않는 조건, 환경변수, 입력 재료나 선행가공품의 물성 또는 표면 상태의 차이 등이다. 이러한 잡음변수 중에서 측정할 수 있는 것들의 값을 구하면 이들도 피쳐 데이터로 확보할 수가 있다. 본 사례에서는 절삭유 농도, 팁 재질, 작업장 온도와 습도, 단조품 치수, 단조품 경도 등이 이에 해당한다. 두 번째 고려할 수 있는 것은 단위공정이 아니라 몇 개의 연결된 공정을 대상으로 데이터를 수집하여 분석하는 것이다. 본 사례의 경우에도 T/P Spider를 가공하기 위한 냉간단조, 선삭 1차, 선삭 2차, 브로칭 공정을 모두 대상으로 삼으면 더 많은 피쳐 데이터를 얻어서 분석을 할 수 있었을 것이다. 이때에는 물론 각 단위공정의 데이터를 모아 분석하기 쉽도록 통합을 하는 것과 데이터의 추적성(traceability)을 확보하는 것에 주의를 기울여야 한다.

생산을 수행해 가면서 변수의 값을 체계적으로 조금씩 변화시켜 나가는 방법인 진화적 조업법(Evolutionary Operation, EVOP)을 이용하여 공정을 개선하고, 품질예측을 위한 피쳐 데이터를 확보할 수 있다(Box and Draper, 1969). EVOP은 공정의 상태를 수동적으로 관찰하여 얻은 데이터를 분석하는 것이 아니라, 공정의 변수를 의도적으로 조금씩 바꾸어가면서 피쳐의 값을 만들어내는 능동적인 방법이다. 양산공정의 공정조건은 최적이라는 보장이 없고, 설사 최적의 공정조건을 구했다고 하더라도 생산을 수행하는 과정에서 그 최적조건은 이동하기 마련이다. Angelopoulos *et al.*(2020)은 머신러닝 분석 결과의 유효성을 확보하기 위해서는 실제 현장에서 알고리즘의 성능을 확인하는 것이 필요하고, 이를 위하여 실제 실험을 수행할 것을 강조하였다. 시간이 지남에 따라 최근 데이터를 사용하여 머신러닝 알고리즘을 주기적으로 수정하고 다시 실행해

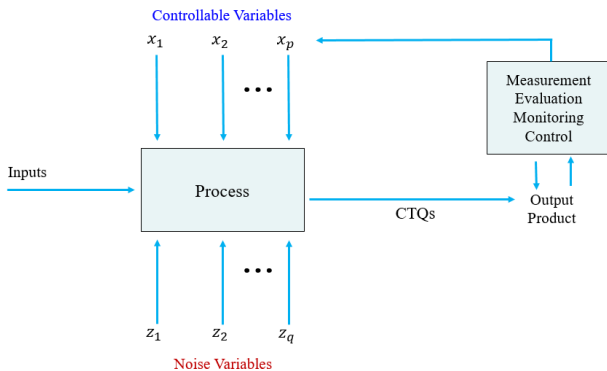


Figure 19. Block Diagram of a Manufacturing Process (Montgomery, 2013)

야 하는 것이다. 과거에 EVOP를 적용하는 데 큰 어려움 중의 하나는 고려하는 공정변수 외에 환경변수 등 다른 요인들은 움직이지 않도록 관리를 해야 하는 것이었다. 그런데 머신러닝을 활용함에 있어서는 다른 조건들이 움직여도 그것을 측정할 수만 있다면 피쳐로 포함하여 분석할 수 있으므로 과거보다 현시점에서 오히려 EVOP을 더욱 용이하게 적용할 수 있다. EVOP은 기본적으로 2수준 2인자 또는 3인자 기반의 요인배치 방법을 이용한다. 하지만 가능한 한 신속하게 공정을 개선하고자 하면, 다른 형태의 EVOP을 이용할 수 있다. Byun and Kim(1999)은 공학적 지식이나 과거 경험을 통하여 불량 유형과 공정변수 간에 불량 발생에 관한 사전 정보가 있을 때 활용할 수 있는 방법을 제시하였다. 실험 수를 줄이기 위해서는 Box *et al.*(2005)이 제안한 각 공정변수의 한쪽으로는 탐색할 수 있는 비대칭합성계획(Asymmetric Composite Design)을 이용할 수도 있다. 과거에는 생산공정의 변동요인들이 크게 작용하면 공정변수의 수준 변화로 인한 효과가 파악할 수 없어서 EVOP 사용을 꺼리기도 했다. 하지만 지금은 이러한 변동요인을 파악하여 측정할 수만 있다면 그것들까지 피쳐로 포함하여 분석할 수 있다.

6. 결론

본 연구의 목표는 실험계획법을 활용하여 실험을 진행하고 실험 결과 데이터를 분석함으로써 CNC 가공공정을 최적화하여 품질비용을 절감하고, 기업의 디지털 전환에 기초가 되는 품질예측모델을 개발하는 것이다. 개선된 가공조건으로 가공 시간을 줄이고, 다구치의 기대손실 측면에서 품질비용을 절감하였다. 또한, 본 사례에서는 가공품 치수 데이터 예측과 치수 불량여부 예측을 위한 모델로서 랜덤포레스트 기법이 우수하다는 것을 확인하고, 의사결정나무 기법으로 가공품의 불량 발생 조건을 확인하여 불량이 발생할 수 있는 피쳐 데이터의 영역을 확인하였다. 이러한 최적 공정 조건 탐색방법과 품질예측모델 개발 과정은 추후 제품개발 및 다른 제품의 공정 최적화에도 활용할 수 있어 향후 기업의 경쟁력 향상에 이바지할 것으로 기대된다.

빅데이터와 인공지능 기법이 발전함에 따라 중소기업은 적극적으로 이런 기법들을 도입할 필요가 있다. 이를 위하여 우선 생산공정의 표준운영절차를 수립하고, 공정 환경의 지속적 변화에 대응하기 위한 공정변수와 환경변수의 피쳐 데이터를 구하여 해당 기업에 맞는 머신러닝 알고리즘을 개발하여야 한다. 공정 개선에 유용한 피쳐 데이터를 확보하기 위해서는 해당 공정의 전문가의 식견과 공정운영자의 경험이 모두 포함되어야 한다. 본 사례연구는 아직 센서 장비 등에 투자가 이루어지지 않은 중소기업이 실험계획 등 체계적인 방법으로 피쳐 데이터를 얻은 다음에 머신러닝 방법을 이용하여 공정을 개선할 수 있는지에 관한 하나의 예를 제시한 것이다.

참고문헌

Angelopoulos, A., Michailidis, E. T., Nomikos, N., Trakadas, P., Hatziefremidis, A., Voliotis, S., and Zahariadis, T. (2020), Tackling Faults in the Industry 4.0 Era: A Survey of Machine-Learning Solutions and Key Aspects, *Sensors*, **20**, 1-34.

Box, G. E. P. and Draper, N. R. (1969), *Evolutionary Operation*, John Wiley & Sons, NY.

Box, G. E. P., Hunter, J. S., and Hunter, W. G. (2005), *Statistics for Experimenters*, 2nd Ed., John Wiley & Sons, NJ.

Byun, J. H. and Kim, Y. (1999), An EVOP Procedure Using the Relationship Between Defect Types and Process Variables of Injection Molding, *IE Interfaces*, **12**(1), 26-31.

Han, Y. H. (2022), Prediction Model of CNC Processing Defects Using Machine Learning, *Journal of Korea Convergence Society*, **13**(2), 249-255.

Lee, K., Park, S., Sung, S., and Park, D. (2019), A Study on the Prediction of CNC Tool Wear Using Machine Learning Technique, *Journal of the Korea Convergence Society*, **10**(11), 15-21.

Kim, J. H. and Kwahk, K. Y. (2022), Class Imbalance Resolution Method and Classification Algorithm Suggesting Based on Dataset Type Segmentation, *Journal of Intelligence and Information System*, **28**(3), 23-43.

Kulahci, M., Frumosu, F. D., Khan, A. R., Ronsch, G. O., and Spooner, M. P. (2020), Experiences with Big Data: Accounts from a Data Scientist's Perspective, *Quality Engineering*, **32**(4), 529-542.

Montgomery, D. C. (2013), *Statistical Quality Control: A Modern Approach*, 7th Ed., John Wiley & Sons, NJ.

Phadke, M. S. (1989), *Quality Engineering Using Robust Design*, Prentice-Hall International, London.

Yang, J., Yuan, J. and Ni, J. (1999), Thermal Error Mode Analysis and Robust Modeling for Error Compensation on a CNC Turning Center, *International Journal of Machine Tools and Manufacture*, **39**(9), 1367-1381.

저자소개

주혜진 : 경상국립대학교에서 산업공학 학사학위를 받았고, 산업시스템공학과 석사과정 학생이다. 관심 분야는 품질공학, 실험계획법, 품질빅데이터 분석이다.

서호진 : 경상국립대학교에서 산업공학 학사와 석사학위를 받았고, 현재 해성디에스에서 근무하고 있다. 관심분야는 품질빅데이터 분석, 실험계획법이다.

김영일 : 경상국립대학교 산업시스템공학과에서 공학박사 학위를 취득하였으며 LG전자, 두산에너지빌리티 등에서 근무하였다. 관심분야는 인공지능 기반의 제조시스템 최적화이다.

김수진 : 서울대학교에서 학사, KAIST에서 기계공학 석사 및 박사학위를 취득하였고, 현재 경상국립대학교 기계공학부에서 교수로 근무하고 있다. 관심 분야는 절삭가공, CAM, 가공 시뮬레이션이다.

이진명 : 기계공학 전공으로 서울대학교 학사, KAIST 석사, Pennsylvania State University 박사학위를 취득했으며, 현재 경상국립대학교 기계공학과 명예교수이다. 관심분야는 기계진동학이다.

김상현 : 경상국립대학교 산업공학과에서 학사학위를 받았고, 현재 일광금속 경영기획실장으로 근무하고 있다. 관심분야는 작업분석과 설계, 품질경영이다.

정윤현 : 경상국립대학교에서 산업공학 학사학위를 받았고, 현재 일광금속 연구소 개발팀에 근무하고 있다. 관심분야는 제품 3차원 모델링, 공정설계이다.

변재현 : 서울대학교에서 산업공학 학사, KAIST에서 산업공학 석사 및 박사학위를 취득하였고, 현재 경상국립대학교 산업시스템공학부에서 교수로 근무하고 있다. 관심 분야는 실험계획법, 품질경영, 데이터 분석공학이다.