

SVM 기반 자동 품질검사 시스템에서 상관분석 기반 데이터 선정 연구

송동환 · 오영광 · 김남훈[†]

울산과학기술원 제어및설계공학과

Study on Correlation-based Feature Selection in an Automatic Quality Inspection System using Support Vector Machine (SVM)

Donghwan Song · Yeong Gwang Oh · Namhun Kim

Department of System Design and Control Engineering, Ulsan National Institute of Science and Technology

Manufacturing data analysis and its applications are getting a huge popularity in various industries. In spite of the fast advancement in the big data analysis technology, however, the manufacturing quality data monitored from the automated inspection system sometimes is not reliable enough due to the complex patterns of product quality. In this study, thus, we aim to define the level of trusty of an automated quality inspection system and improve the reliability of the quality inspection data. By correlation analysis and feature selection, this paper presents a method of improving the inspection accuracy and efficiency in an SVM-based automatic product quality inspection system using thermal image data in an auto part manufacturing case. The proposed method is implemented in the sealer dispensing process of the automobile manufacturing and verified by the analysis of the optimal feature selection from the quality analysis results.

Keywords: Support Vector Machine, Feature Selection, Quality Monitoring System, Data Analysis, Correlation Analysis

1. 서론

최근 이슈가 되고 있는 4차 산업혁명(Industry 4.0)은 사이버 물리시스템(Cyber-Physical System)을 기반으로 제조 공정 안에서 유연한 생산체계를 지향하는 자동화 개념이다. 지금까지 공장 자동화는 생산 자동화 연구에 집중되었지만, 최근 센서 및 데이터 해석 기술이 비약적으로 발달함에 따라 품질 자동화에 대한 연구와 적용이 활발하다(Chen *et al.*, 2013; Ngai *et al.*, 2014). 품질 자동화 시스템이 풍부한 현장 경험 및 동종 산업분야의 지식과 분류 기술을 갖춘 전문가의 역할을 대체하기 위해서는 제한된 시간 내에 품질을 분석하여야 하고, 사용자가 자동화 시스템의 분석 결과를 신뢰할 수준이 되어야 한다(Oh *et al.*, 2015). 따라서, 데이터 분석 전에 목적에 부합하는

적정 수준의 데이터만을 추출하여 관리하기 위하여 공정 데이터의 분석 효율을 높이는 특징 선택(Feature Selection)에 대한 연구가 필요하다(Hansson *et al.*, 2016). 특징 선택의 결과는 데이터의 잡음을 제거하고 데이터 차원 수를 줄여서 효율적으로 데이터 저장 및 관리를 할 수 있도록 한다.

제조 시스템 분석 결과의 정확도를 향상시켜 사용자가 데이터 분석 시스템에 신뢰도를 가지도록 제조 데이터 분석 관련 연구가 진행 되어 왔다(Lee *et al.*, 2014). 품질 검사를 위한 기계학습 알고리즘 중에서 SVM은 지도 학습 모델 중 하나로 주로 데이터를 분류 및 회귀 분석을 위해 사용하며 현재 다양한 분야에 적용되고 있다. 대표적인 연구 사례로는, 반도체 생산 라인에서 반도체 데이터를 전처리 후 SVM(Support Vector Machine) 기반의 제품 품질 결과를 예측(Jang and bae, 2009),

[†] 연락저자 : 김남훈 교수, 44919 울산광역시 울주군 언양읍 유니스트길 50, 울산과학기술원 제어및설계공학과, Tel : 052-217-2715, Fax : 052-217-2709, E-mail : nhkim@unist.ac.kr

2016년 5월 2일 접수; 2016년 8월 26일 1차 수정본 접수; 2016년 11월 3일 2차 수정본 접수; 2016년 11월 29일 게재 확정.

원자력발전소의 센서 측정값을 사용하여 SVM 기반의 운전 매개변수 참값 예측(Kim *et al.*, 2008), LS-SVM을 이용한 TFT-LCD 패널 내의 결함 검사(Choi *et al.*, 2009) 등이 있다.

본 연구에서는 SVM 기반의 품질 분석 시스템을 제안하고, 선루프 실러 도포 공정에 적용을 통한 그 효용성을 검증하고자 한다. 이를 위해, 먼저, 제조공정 데이터에서 제품 품질에 영향을 주는 데이터를 선별하고 관리하기 위해서 상관관계 분석을 실시하고 데이터 차원을 축소한다. 차원이 축소된 데이터를 통해서도 충분히 사용 가능한 수준의 품질 평가 체계 구축이 가능함을 실제 공정 실험 데이터를 분석하여 검증한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 SVM과 특징 선택기법에 관한 알고리즘을 소개하고, 제 3장에서는 본 연구에서 제안하는 열화상 데이터를 얻는 과정과 불량 판정 과정에 대해 단계별로 설명한다, 제 4장에서는 제안하는 프로세스를 자동화 시스템에 적용하여 시스템 분석 결과를 비교하고 최적화를 검증한다. 마지막으로 제 5장에서는 결론 및 향후 연구 방향을 제시한다.

2. 관련 연구

2.1 SVM

SVM은 마진(Margin)이 최대가 되는 결정평면(Decision Surface)을 찾고, 훈련에 사용되는 각 클래스의 샘플을 이용하여 데이터를 분류하는 기법이다. 여기서, 마진은 서로 다른 클래스를 분리하는 기준이 되는 결정평면과, 서로 다른 클래스의 샘플인 서포트 벡터(Support Vector) 사이의 거리이다. 입력데이터가 x_i 이고 결과 데이터는 $y_i \in \{1, -1\}$ 인 N 개의 샘플 $\{(x_i, y_i)\}_{i=1}^N$ 을 고려할 때, 결정평면은 $w \cdot x + b = 0$ 로 나타낸다. w 는 결정평면과 수직인 법선 벡터이고, b 는 바이어스를 의미한다. w 의 크기는 $\|w\|$ 로 표현한다. 결정평면과 결과 데이터의 곱이 양수라 가정하면 식 (1)에서 비용함수가 최소가 되는 w 와 b 를 구하여 최적화된 결정평면을 얻을 수 있다. 이때 두 분류 집합을 분리하는 마진은 $2/\|w\|$ 이고, 이를 최대화하기 위해서는 $\|w\|$ 가 최소가 되어야 한다(Vapnik, 2013).

$$\arg \min_{(w, b)} \frac{1}{2} \|w\|^2 \quad (1)$$

$$y_i (w \cdot x_i - b) \geq 1, \text{ for all } 1 \leq i \leq n$$

식 (2)에서 C 는 두 그룹간 차이와 분류 오류 사이를 조율(Trade-off)해주는 상수이다. 최적화 문제는 라그랑지승수(Lagrange Multiplier)를 이용하여 해결한다.

$$\arg \min_{w, \xi, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (2)$$

$$\text{Subject to } y_i (w \cdot x_i + b) \geq 1 - \xi_i$$

$$\xi \geq 0, \text{ for all } i$$

하지만, 일반적으로 데이터의 선형분리는 가능하지 않기 때문에, 식 (3)과 같은 커널 함수를 사용하여 비선형 문제를 해결한다(Hsu and Lin, 2002).

$$K(x, x') = \langle \phi(x), \phi(x') \rangle \quad (3)$$

$$x, x' \in R^p$$

커널 함수 선택은 데이터 유형과 커널 선택에 따라 분석 결과가 다르므로 적용하고자 하는 데이터에 알맞은 커널 함수를 선택하여 사용하거나 변형하여 사용한다(Cristianini and Shawe-Taylor, 2000). 본 연구에서는 가장 대표적으로 쓰이는 RBF, Polynomial, Sigmoid를 비교하고 데이터에 적절한 커널을 선택한다(Widodo and Yang, 2007).

SVM 기반으로 데이터를 분석하는 연구는 다양한 분야에서 진행되고 있다(Chen and Lin, 2006; Pal and Foddy, 2010; Kim *et al.*, 2013, Oh *et al.*, 2015). 특히, 제조공정에서 Oh *et al.*(2013)는 SVM을 도어 트림 공정에 적용하여 실시간 공정 모니터링 시스템을 제안하였다. 그리고 Kim *et al.*(2013)은 SVM 알고리즘을 자동차 선루프 부품 유무 비전 검사 시스템에 적용하였다.

2.2 특징 선택기법 및 상관분석

데이터 전처리 단계에서는 수집한 데이터를 사용하기 위해 데이터 형식을 변경하거나, 수집한 데이터 중에서 필요한 데이터를 추출하여 데이터의 차원을 축소한다. 특히, 산업현장에서 수집하는 데이터는 노이즈(Noise)가 심하고 고차원적 특성을 가지는 불필요한 독립변수로 구성되어 있다(Kadlec *et al.*, 2009). 불필요한 데이터를 분석시스템에 포함하여 판정하면 판정 시스템의 효율을 저하시키고, 불필요한 관리비용 증가를 야기한다. 따라서, 특징 선택 방법으로 관리 데이터를 축소하면서 분석 정확도 감소를 최소화하여 품질 정책에 유의한 결과를 도출할 필요가 있다.

특징 선택기법 중 데이터 상관관계 분석은 데이터 상관관계를 분석하여 품질정책의 유의수준을 만족하면서 관리가 필요한 데이터를 축소하는 방법이다(Yu and Liu, 2003; Yu and Liu, 2004; Jung *et al.*, 2010; Son *et al.*, 2012; Jang *et al.*, 2015). 상관관계 분석에 사용되는 상관계수(Correlation Coefficient)는 두 변수 간에 어떤 선형적 관계(방향성 및 크기)가 있는지 나타낸다. 상관계수는 두 변수 간 선형관계의 절대적 크기를 나타내는 공분산(Covariance)을 두 변수의 표준편차로 나눔으로써 방향성의 나타낸다. 두 변수를 x, y 라 할 때, 피어슨(Pearson) 상관계수를 수식으로 나타내면 다음과 같다(Jung, 2009).

$$r = r_{xy} = \frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

상관계수 기법을 이용하여 특징선택을 시행할 수 있는데 대표적으로 Yu and Liu(2003)는 상관계수를 이용하여 최적 특징선택을 찾는 Fast Correlation Based Filter Approach 알고리즘을 제안하고, 그 결과를 비교하였다.

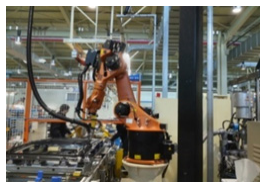
데이터 분석에서 데이터차원을 축소하기 위해 특징선택을 적용 후, SVM 기반으로 결과를 비교 분석한 연구도 진행되고 있다(Chen and Lin, 2006; Pal and Foody, 2010). Chen and Lin (2006)은 서로 다른 특징선택 기법을 사용하여 동일한 데이터에 적용 후 SVM 결과를 비교하였다. Pal and Foody(2010)는 특징선택 기반으로 축소한 데이터 차원에 따라 SVM 분석 결과가 다르게 나타나는 것을 실험을 통해 증명하였다. 따라서, 특징선택기법으로 데이터 차원을 축소하면 원본 데이터를 사용하는 것과 다른 SVM 기반 분석 결과가 다르게 나타남을 알 수 있다.

본 연구에서는 상관분석으로 독립변수(특징)들 간의 상관계수를 도출하고, 유효인자를 분리한다. 그리고 상관분석으로 분리한 유효인자를 SVM 기반 자동 품질검사 시스템에 적용한다.

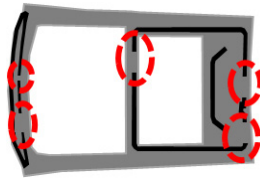
3. 판정적용 공정 및 데이터 설명

3.1 적용 대상 공정 : 선루프 실러 도포

본 논문에서 제안하는 방법을 적용할 공정은 자동차 선루프 생산 공정 중, 선루프 프레임에 실러를 도포하는 공정이다. 이 공정에서는 <Figure 1>(a)와 같이 선루프 프레임의 유리 부착면에 로봇이 접착제인 실러를 균일하게 도포한다. 하지만, 이 과정에서 도포된 실러가 고객사에서 요구하는 품질기준을 만족하지 못할 경우, <Figure 1>(b)와 같이 실러의 끊김이나 도포량 미달 등의 불량 발생한다. 불량으로 판정된 제품은 유리 압착시 유격을 발생 시키고, 제품 누수 또는 파손의 원인이 된다.



(a) Sealer dispensing process using robot arm



(b) Example of defective position

Figure 1. Overview of Sealer Application Process

기존에는 품질 검사자 또는 작업자가 눈으로 검사를 실시하였다. 공정의 품질 합격 기준은 존재 하였으나, 사람의 판단에 의해 제품의 품질이 결정되므로 잘못된 판단을 하는 경우가 빈번히 발생하였다. 제안하는 SVM 기반 자동 품질 검사 시스템은 열화상 이미지를 사용하여 제품의 품질을 검사하는 시스템으로 전수검사를 수행 한다. 제안하는 SVM 기반 자동 품

질 검사의 특징은 다음과 같다. 먼저, 주변 환경 변화에 따른 오차를 줄이기 위해 열화상 카메라 촬영시점은 제품 생산공정에서 실러 도포 전 한번, 실러 도포 완료 후 한번이다. 촬영된 이미지는 통신 카드를 통해 KIOSK 장비로 전송한다. KIOSK 장비에서 이미지에서 공정의 온도를 분석하고 실러 도포의 불량 판정을 한다. 그리고 불량제품 판정 또는 설비에 문제가 발생하는 경우, 경고메시지를 작업자에게 전달한다.

열화상 카메라로 촬영한 이미지는 <Figure 2>와 같이 도포된 실러와 주변 정보를 포함한다. 파노라마 선루프와 같은 제품의 경우는 <Figure 2>와 같이 왼쪽에 보이는 앞쪽(Front) 부분과 오른쪽에 보이는 뒤쪽(Rear)에 실러가 도포 된다.

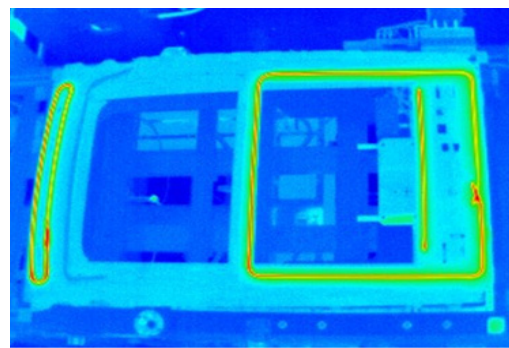


Figure 2. An Example of Thermal Camera Image

다음 장에서는 상관관계 분석으로 수집한 데이터의 상관도를 분석하고 특징선택을 실시하여 수집한 데이터 중에서 SVM 기반 자동 품질 검사 시스템의 입력 데이터를 선정하는 방법을 소개한다.

3.2 유효인자 분석을 통한 분석 결과 정확도 최적화 방법

본 연구에서 제안하는 SVM 기반 자동 품질 분석 시스템 프레임워크는 <Figure 3>과 같다. 먼저 열화상 카메라로부터 데이터를 수집하고, 데이터 전처리를 실시한다. 데이터 전처리 단계에서 데이터 분석 및 변환은 다음과 같은 과정으로 이루어진다. 먼저, 미가공 데이터의 특징을 분석하기 위해 평균, 표준편차 등 통계분석 기반의 방법으로 독립변수를 생성한다. 그 후, SVM 기반의 품질검사 시스템으로 제품의 품질을 판정할 수 있도록 양품은 1로 불량은 -1로 데이터 형식을 변환하고, 데이터 불균형 해소를 위한 Under-Sampling을 적용한다. 예를 들어 공정에서 불량 데이터가 양품 데이터 관측치보다 현저히 적게 생성되면 불량 데이터의 관측치 개수에 양품 데이터의 관측치 개수를 동일한 비율로 낮추는 Under-Sampling이 필요하다. 그리고, 누락값 보정을 실시한다. 누락값 보정은 수집한 온도 데이터에 누락된 데이터가 있을 경우 관측치를 제거하는 방법이다. 또한, Cross Validation으로 부분집합 나눌 때 특정 판정 결과가 테스트집합에 물리는 것을 방지하기 위해 관측치 행을 무작위로 섞는 Randomize를 적용한다.

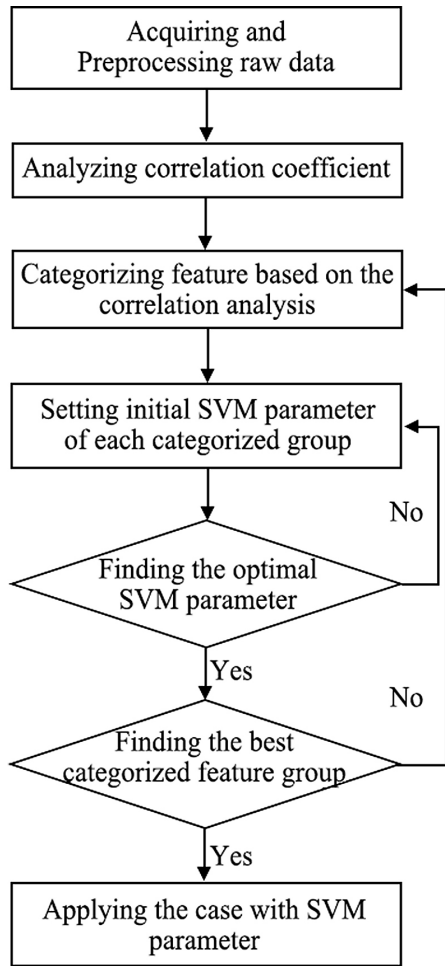


Figure 3. The Flow Chart of SVM Feature Selection and Optimal Parameter Calculation Based on Correlation Analysis

상관 관계분석 단계에서는 피어슨 상관분석을 통하여 독립 변수들의 상관계수를 도출하고, 데이터 연관성이 큰 유효 인자를 선별하여 중복 특징을 가지는 데이터를 제거 한다. 그리고, SVM 매개변수 최적화 단계에서는 Cross Validation기반으로 SVM 매개변수 C와 Gamma값을 변경하면서 매개변수의 최적값을 찾는다.

SVM 테스트 및 정확도 분석 단계에서는 커널에 따른 성능을 비교 분석하여 정확도가 가장 높은 커널을 선정한다. 제안하는 방법의 마지막 단계에서는 가장 높은 정확도를 가지는 커널과 SVM 매개변수를 시스템에 적용한다.

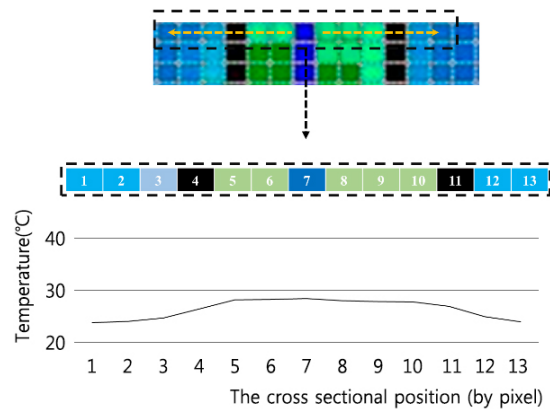
4. 실험 및 결과분석

4.1 제조 공정 데이터

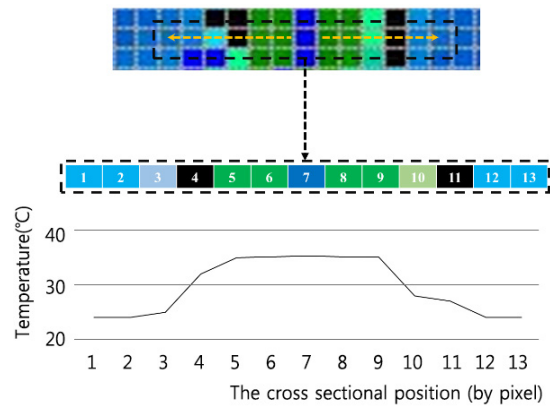
도포된 실러의 품질을 판정하기 위한 입력 데이터는 열화상 카메라로부터 수집한 온도 데이터이다. 전체 온도 데이터에서 제품 품질에 영향을 주는 데이터는 도포된 실러의 중심을 기

준으로 좌우 각각 6개 픽셀의 온도 데이터 이다. 따라서, 중심 선 픽셀이 변할 때마다 13개의 픽셀 데이터 값이 변한다. 제품이 불량인 경우, <Figure 4>(a)와 같이 도포된 실러의 픽셀온도가 정상인 경우보다 낮은 온도를 가지거나, 실러가 도포되지 않은 픽셀과 비슷한 온도 분포를 가진다. 도포된 실러가 양품인 경우, <Figure 4>(b)와 같이 도포된 실러 온도가 도포 되지 않은 픽셀의 온도보다 높은 온도 분포를 가진다.

품질 판정을 위한 변수들은 <Table 1>과 같다. 열화상 카메라로부터 수집한 미가공 데이터와 1차 유도 변수는 연속형 변수이다. 종속변수인 품질 판정 결과는 SVM 기반 품질 검사를 위해 양품을 1(Good)로 불량을 -1(NG)로 변환한다.



(a) An example of defective part



(b) An example of good part

Figure 4. An Example Data Group Having 13 Pixels around Dispensed Sealer

생산 공정으로부터 수집한 데이터 결과값은 불량으로 판정된 데이터보다 양품으로 판정된 데이터가 많은 불균형 데이터 집합의 형태를 나타낸다. 이에, 데이터 불균형 문제를 해소하기 위해, 불량으로 판정된 온도 데이터의 총 개수에 맞도록 양품 데이터를 Under-Sampling 하여 관측치 수를 축소한다. 분석에 사용한 제조 공정 데이터 그룹은 2,400개(OK 1,200, NG 1,200)이다. 각 데이터 그룹에서 입력변수들은 실러 도포 전의 선루

프 상면 온도분포(before_T[])와 실러 도포 후 선루프 상면 온도분포(after_T[])이고, 유도 입력변수로는 단면방향으로의 실러의 폭(width), 평균온도(avg_T), 표준편차온도(std_T), 상대 온도(rel_T[]), 상대 최소온도(rel_min_T), 그리고 상대 최대온도(rel_max_T)이다.

Table 1. An Illustration of Input, Derived Input and Output Variables

	Feature	Data Type
Raw Variables	before_T[1]~before_T[13], after_T[1]~after_T[13]	Continuous
Derived Input Variables	width, avg_T, std_T, rel_max_T, rel_min_T, rel_T[1]~rel_T[13]	
Output Variable	Quality Result	Discrete

4.2 상관계수에 따른 요인 설정

본 실험에서는 피어슨 상관관계 분석을 통해 변수들 간의 상관도를 도출하고 특징선택을 실시한다. 피어슨 상관계수는 1에 가까울수록 양의 상관관계를 나타내고, -1에 가까울수록 음의 상관관계를 나타낸다. 만약 피어슨 상관계수가 0에 가깝다면, 상관관계가 희박함을 의미한다. 상관계수의 유의 확률(p-value)이 설정된 유의수준(Significant level) 이하일 때, 상관계수가 유의미하다고 할 수 있다.

<Table 2>는 도포된 실러의 시작지점인 1번 픽셀에서의 상대온도(rel_T[1])는 다른 입력변수들 사이의 피어슨 상관계수 값을 보여준다. 피어슨 상관계수 값이 0.7 이상 또는 -0.7 이하를 가지면 상관도가 크며, 0.9 이상 또는 -0.9 이하면 매우 강한 상관도를 나타낸다(Mukaka, 2012; O'Rourke and Hatcher, 2013).

<Table 2>와 같이 본인 자신과 rel_T[2]의 상관계수는 0.938로 매우 강한 상관도를 가진다. 이와 마찬가지로 모든 변수들의 상관계수를 계산하고 상관도를 분석한다.

상관도 분석을 바탕으로 <Table 3>과 같이 세 가지 타입의 입력데이터를 설정하고 SVM 기반의 품질 검사를 실시하여 판정 정확도를 비교한다. Case 1은 공정에서 수집한 미가공 데이터 집합이고 Case 2는 모든 입력변수들과 통계기법으로 찾은 유도입력변수들의 집합이며, <Table 3>의 Case 3은 피어슨 상관계수 값이 0.9보다 큰 변수를 기준으로 특징 선택을 진행한다. 변수들 사이의 상관계수가 0.9보다 크면 두 변수는 불량 판정 시스템에 거의 동일한 영향력을 끼치는 독립변수라고 할 수 있기 때문에 중복 변수로 간주하여 제거한다. 예를 들어 rel_T[1]과 rel_T[2]의 경우 0.938로서 높은 상관도를 가지기 때문에 중복 입력변수로 간주하며 둘 중 한 변수를 제거 하였으며 많은 변수들과 중복되는 변수를 우선적으로 선택된 변수로 설정하였다. 변수의 중복의 수가 동일할 경우 두 변수 중 임의로 한 변수를 선택하는 방식을 사용하였다. 다음과 같은 방식

Table 2. Pearson Correlation Coefficient between rel_T[1] and Other Features

Feature	Correlation	Feature	Correlation
width	0.403	after_T[5]	0.329
avg_T	0.495	after_T[6]	0.273
std_T	0.182	after_T[7]	0.260
rel_max_T	0.294	after_T[8]	0.264
rel_min_T	0.681	after_T[9]	0.335
rel_T[1]	1.000	after_T[10]	0.374
rel_T[2]	0.938	after_T[11]	0.336
rel_T[3]	0.793	after_T[12]	0.272
rel_T[4]	0.548	after_T[13]	0.235
rel_T[5]	0.350	before_T[1]	-0.056
rel_T[6]	0.294	before_T[2]	-0.040
rel_T[7]	0.280	before_T[3]	-0.046
rel_T[8]	0.288	before_T[4]	-0.051
rel_T[9]	0.359	before_T[5]	-0.048
rel_T[10]	0.401	before_T[6]	-0.042
rel_T[11]	0.378	before_T[7]	-0.048
rel_T[12]	0.313	before_T[8]	-0.052
rel_T[13]	0.273	before_T[9]	-0.045
after_T[1]	0.861	before_T[10]	-0.051
after_T[2]	0.838	before_T[11]	-0.050
after_T[3]	0.738	before_T[12]	-0.047
after_T[4]	0.522	before_T[13]	-0.038

Table 3. Set Experiment Feature Case based on Correlation Coefficient

Type	Feature	Description
Case 1	after_T[1]~[13], before_T[1]~[13]	Raw Variables
Case 2	avg_T, std_T, width rel_T[1]~[13], rel_Max_T, rel_Min_T, after_T[1]~[13], before_T[1]~[13]	All Variables
Case 3	avg_T, std_T, rel_Min_T, rel_T[2], rel_T[3], rel_T[4], rel_T[5], rel_T[7], rel_T[8], rel_T[9], rel_T[10], rel_T[11], rel_T[12], rel_T[13], after_T[1]	Feature Selected Variables

으로 모든 변수들 간의 상관계수 검증을 통해 데이터 차원을 축소하여 Case 3과 같은 최종 입력변수들을 선정하였다. 이는 관리 데이터의 양과 SVM 기반 품질 검사 시간을 줄인다.

4.3 SVM 기반 품질 검사 결과 및 성능 분석

본 장에서는, 제 4.2장에서 분류한 세 가지 Case에 대해 SVM 기반의 검사를 실시하고 그 결과를 비교한다. 정확도 비교에 사

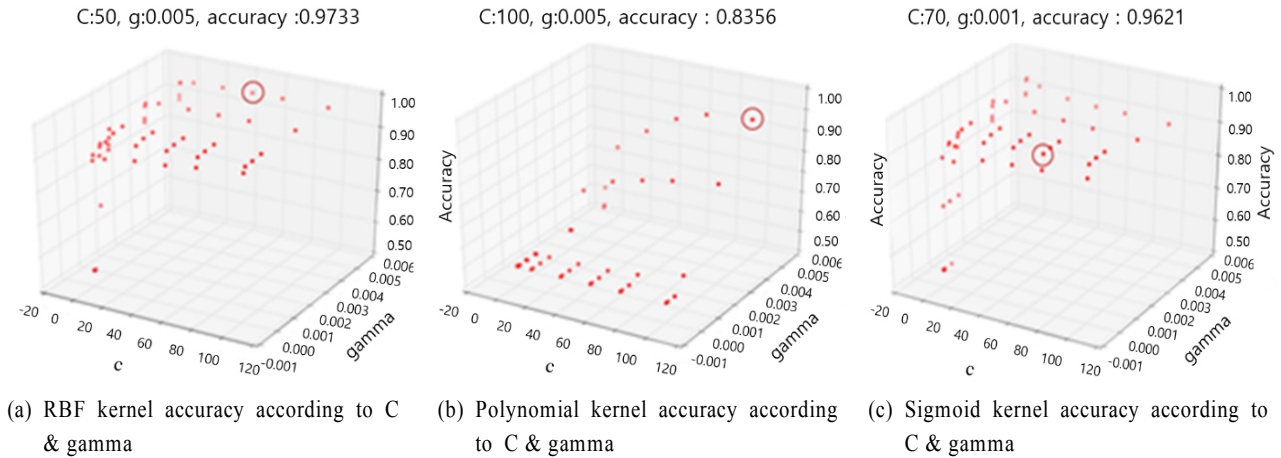


Figure 5. The Comparison of SVM Results of Case 3 Depend on Different Kernel Selection

용하는 SVM 커널은 RBF, Polynomial Sigmoid이다. LIBSVM(A Library for Support Vector Machines) API를 참조하여 각 커널의 변수는 gamma로 표현한다(Chang and Lin, 2011). SVM 매개변수는 k-Fold Cross-Validation(k = 5)을 사용하여 Case와 커널 별로 가장 높은 정확도를 가지는 값으로 설정한다. <Figure 5>는 Case 3의 입력데이터를 SVM 커널과 매개변수에 따라 변하는 정확도 분포를 표현한 그래프이다. 여기서 C값의 범위는 [0.1, 1, 10, 30, 50, 70, 100]이고, Gamma값의 범위는 [0.005, 0.003, 0.001, 0.0005, 0.0001, 0.00001]이다. <Figure 5>는 SVM 매개변수인 C와 gamma의 변화에 따른 품질판정의 예측 정확도 분포와 최적값을 각 커널별로 나타내고 있으며 각 Case별 최적값은 <Table 4>와 같다. 각 Case별 SVM 기반 품질 검사 정확도를 비교해 볼 때, SVM의 입력 변수를 열화상 카메라로부터 수집한 미가공 데이터만을 사용했을 때 보다 통계적 기법을 통해 찾은 유도 변수를 함께 활용하였을 때 정확도가 증가됨을 알 수 있다. 또한, Case 3과 같이 특징 선택을 실시하고 SVM 기반 품질 검사를 하면 입력 데이터 개수를 26개에서 15개로 42.3%가까이 줄일 수 있다. RBF 커널 적용 시, Case 3의 불량 판정 성능은 Case 1과 Case 2에 비해 소폭 떨어지지만 품질 정책 유의 수준을 만족하므로 Case 3을 적용할 수 있다.

Table 4. The Accuracy of SVM Results and SVM Parameters, C and Gamma by Kernel Types

Kernel	RBF Accuracy (C, Gamma)	Polynomial Accuracy (C, Gamma)	Sigmoid Accuracy (C, Gamma)
Case 1	0.9821 (100,0.003)	0.8834 (100,0.005)	0.9413 (100,0.001)
Case 2	0.9841 (70,0.005)	0.9654 (100,0.005)	0.9650 (70,0.0005)
Case 3	0.9733 (50,0.005)	0.8356 (100,0.005)	0.9621 (70,0.001)

5. 결론 및 향후 연구

본 논문에서는 제조 공정의 자동화를 위한 SVM 기반 자동 품질 검사 시스템을 제안하고 선루프 실러 도포 공정에 적용한다.

제안하는 시스템 적용을 위해, 데이터 상관분석을 통하여 특징 선택을 실시하고 품질검사를 위한 입력 데이터를 선정한다. 그리고 커널 함수에 따른 SVM 매개변수 값을 비교하여 제안하는 방법을 적용하는 시스템에 적절한 커널 함수와 가장 높은 정확도를 가지는 SVM 매개변수 값을 찾는다. 선정된 입력 데이터와 SVM 매개변수 값을 사용하여 SVM 기반 자동 품질 검사를 실시한다.

제안하는 방법은 상관계수를 바탕으로 한 특징선택을 적용하여 관리 데이터의 수를 줄이면서 품질정책의 유의수준을 만족하는 입력데이터를 선정하므로 데이터 관리 및 서버 유지보수에 효율성 제고에 도움이 될 것으로 기대된다. 또한, 품질 분석에 SVM 시스템 판정을 사용하여 실제 작업자의 육안 검사와 병행하여 품질 판정 시간을 매우 감소시킬 수 있으며 불량 판정률도 낮출 수 있게 된다.

하지만, 특징 선택 할 때 사용한 데이터 차원의 개수가 상대적으로 적고, NG 판정 샘플 데이터가 불충분한 것은 본 연구의 한계점이다. 따라서 향후 연구에서는 본 연구의 방법을 확장하여 열화상 이미지의 화질을 높이며 구간을 더욱 세분화하여 나눠 특징수를 높은 상태로 특징선택을 하여 판정 정확도를 높이며 유지비용을 최소화하는 연구가 필요할 것이다.

참고문헌

Chang, C. C. and Lin, C. J. (2011), LIBSVM : a library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 27.
 Chen, Y. W. and Lin, C. J. (2006), Combining SVMs with various feature selection strategies, *In Feature extraction*. Springer Berlin Heidelberg,

- 315-324.
- Chen, J., Zhang, W., Yu, Z., and Feng, Z. (2013), Automated Spot Weld Inspection using Infrared Thermography, *In Trends in Welding Research 2012 : Proceedings of the 9th International Conference*, 148-151.
- Choi, H. H., Lee, G. H., Kim, J. G., Joo, Y. B., Choi, B. J., Park, K. H., and Yun, B. J. (2009), A Defect Inspection Method in TFT-LCD Panel Using LS-SVM, *Journal of Korean Institute of Intelligent Systems*, **19**(6), 852-859.
- Cristianini, N. and Shawe-Taylor, J. (2000), An introduction to support vector machines and other kernel-based learning methods, *Cambridge university press*, 149-160.
- Hansson, K., Yella, S., Dougherty, M., and Fleyeh, H. (2016), Machine Learning Algorithms in Heavy Process Manufacturing, *American Journal of Intelligent Systems*, **6**(1), 1-13.
- Hsu, C. W. and Lin, C. J. (2002), A comparison of methods for multiclass support vector machines, *Neural Networks, IEEE Transactions on*, **13**(2), 415-425.
- Jang, D.-Y. and Bae, S.-J. (2009), Hybrid Datamining Algorithm for Monitoring Input Variables in Semiconductor Manufacturing Process, *In Conference of the Korean Institute of Industrial Engineers*, 563-569.
- Jung, H. S. (2009), An empirical study on the characteristics of the correlations between industry indexes in the Korean stock market, *Korea Advanced Institute of Science and Technology, KGSF-Theses_Master*.
- Kadlec, P., Gabrys, B., and Strandt, S. (2009), Data-driven soft sensors in the process industry, *Computers and Chemical Engineering*, **33**(4), 795-814.
- Kim, G. S., Lee, S., and Cho, J. S. (2013), A Learning-based Visual Inspection System for Part Verification in a Panorama Sunroof Assembly Line using the SVM Algorithm, *Journal of Institute of Control, Robotics and System*, **19**(12), 1099-1104s.
- Kim, S. J., Seo, I. Y., and Shin, H. C. (2008), An SVM based On-line Monitoring for Sensor Calibrations in the Nuclear Power Plant, *Journal of the Korean Institute of Industrial Engineers*, **5**, 1137-1145.
- Lee, J., Kao, H. A., and Yang, S. (2014), Service innovation and smart analytics for industry 4.0 and big data environment, *Procedia CIRP*, **16**, 3-8.
- Mukaka, M. M. (2012), A guide to appropriate use of Correlation coefficient in medical research, *Malawi Medical Journal*, **24**(3), 69-71.
- Ngai, E. W. T., Peng, S., Alexander, P., and Moon, K. K. (2014), Decision support and intelligent systems in the textile and apparel supply chain : An academic review of research articles, *Expert Systems with Applications*, **41**(1), 81-91.
- Oh, Y. G., Park, H. S., Yoo, A., Kim, N. H., Kim, Y. H., Kim, D. C., Choi, J. U., Yoon, S. H., and Yang, H. J. (2013), A Product Quality Prediction Model Using Real-Time Process Monitoring in Manufacturing Supply Chain, *Journal of the Korean Institute of Industrial Engineers*, **39**(4), 271-277.
- Oh, Y. G., Ju, I. C., Lee, W. Y., and Kim, N. H. (2015), Modeling and Implementation of the Affordance-based Human-Machine Collaborative System, *Journal of the Korean Institute of Industrial Engineers*, **41**(1), 34-42.
- O'Rourke, N. and Hatcher, L. (2013), A step-by-step approach to using SAS for factor analysis and structural equation modeling, *Sas Institute*, 379-392.
- Pal, M. and Foody, G. M. (2010), Feature selection for classification of hyperspectral data by SVM, *IEEE Transactions on Geoscience and Remote Sensing*, **48**(5), 2297-2307.
- Vapnik, V. (2013), The nature of statistical learning theory, *Springer Science and Business Media*, 181-223.
- Widodo, A. and Yang, B. S. (2007), Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical systems and signal processing*, **21**(6), 2560-2574.
- Yu, L. and Liu, H. (2003), Feature selection for high-dimensional data : A fast correlation-based filter solution, *Paper presented at the ICML*.
- Yu, L. and Liu, H. (2004), Efficient feature selection via analysis of relevance and redundancy, *The Journal of Machine Learning Research*, **5**, 1205-1224.