

베이저안 추론 기반 니켈기 초합금의 열화도 정량화 방법과 열화도 및 크리프 수명 예측의 방법

유준상^{1,2} · 오하영^{3*}

Degradation Quantification Method and Degradation and Creep Life Prediction Method for Nickel-Based Superalloys Based on Bayesian Inference

Junsang Yu^{1,2} · Hayoung Oh^{3*}

¹Manager, Data Analytics Team, Doosan Enerbility, Changwon, 51711 Korea

²Graduate Student, Department of Applied Data Science, Sungkyunkwan University, Suwon, 16419 Korea

^{3*}Associate Professor, College of Computing and Informatics, Sungkyunkwan University, Seoul, 03063 Korea

요 약

본 연구의 목적은 고온부품의 소재로 사용하는 니켈기 초합금인 DA-5161 SX에 대한 크리프시험으로 얻은 시편의 주사전자현미경으로 촬영한 미세조직 단면의 이미지로부터 인공지능 기반 열화인덱스(Degradation Index)로 정량화 하는 새로운 방법을 제시하고 운전 중인 기기의 고온부품의 구성품을 파괴하지 않고 베이저안 추론 기반 열화도를 예측하는 모델과, Larson-Miller Parameter(LMP)를 예측하여 크리프 수명 예측 모델을 제안하는 것이다. 니켈기 초합금 미세조직인 감마프라임 상(γ')의 기하학적 특징 및 베이저안 추론 기반으로 소량의 이미지로 일관성 있는 대푯값을 추론하는 새로운 열화인덱스 방법과 고온부품을 파괴하지 않고 소재의 환경조건 정보만으로 열화인덱스와 LMP를 예측할 수 있는 방법을 제안한다.

ABSTRACT

The purpose of this study is to determine the artificial intelligence-based degradation index from the image of the cross-section of the microstructure taken with a scanning electron microscope of the specimen obtained by the creep test of DA-5161 SX, a nickel-based superalloy used as a material for high-temperature parts. It proposes a new method of quantification and proposes a model that predicts degradation based on Bayesian inference without destroying components of high-temperature parts of operating equipment and a creep life prediction model that predicts Larson-Miller Parameter (LMP). It is proposed that the new degradation indexing method that infers a consistent representative value from a small amount of images based on the geometrical characteristics of the gamma prime phase, a nickel-base superalloy microstructure, and the prediction method of degradation index and LMP with information on the environmental conditions of the material without destroying high-temperature parts.

키워드 : 가스터빈, 극한소재, 특성공학, 수명예측, 베이저안 모델

Keywords : Gas turbine, extreme materials, feature engineering, life prediction, Bayesian model

Received 20 November 2022, Revised 29 November 2022, Accepted 2 December 2022

* Corresponding Author Hayoung Oh (E-mail: hyoh79@gmail.com, Tel:+82-2-583-8585)

Associate professor, College of Computing and Informatics, Sungkyunkwan University, Seoul, 03063 Korea

Open Access <http://doi.org/10.6109/jkiice.2023.27.1.15>

print ISSN: 2234-4772 online ISSN: 2288-4165

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

1.1. 연구 배경

머신러닝, 딥러닝 등 인공지능 기술이 방대해진 데이터와 빅데이터 플랫폼 등의 풍부해진 컴퓨팅 자원의 결합으로 발전이 가속화되고 있다. 근래에는 빅데이터뿐만 아니라 미디어 또는 스몰 데이터에 대한 분석을 위해 발달된 컴퓨팅 기술을 이용하여 베이지안 접근방법이 더욱 용이 해졌다. 적은 데이터로 의사결정을 하기 어려운 것은 불확실성 때문이다. 베이지안 접근방법은 불확실성을 확률로 정량화하고, 가용한 증거들을 기반으로 알려지지 않은 모수의 사후확률분포를 추론함으로써 직관적인 설명이 가능하여 의사결정에 적합한 방법이다.[1]

최근 제조업 분야에서는 Prognostic Health Management (PHM)이 전통적인 신뢰성 분석의 한계를 극복할 수 있는 핵심 요소로서 자리 잡고 있다. PHM은 기기의 건전성을 모니터링하고, 이상치를 발견하고 결함을 진단하고, 무엇보다 중요하게도 잔존 사용 수명(RUL; Remaining Useful Life)을 예측하기 위해 기기의 센서 신호를 사용하는데 주력한다. 이는 기기의 소유자들에게도 강력한 경제적 이득을 갖게 하는데, PHM 기반의 정비전략은 기존의 고장이 발생하면 정비하거나, 시간이 경과하면 정비하는 전략과 다르게 기기의 고장이 언제 일어날지 예측하고 기기가 계속 정상 동작할 수 있도록 정비할 수 있게 하기 때문이다.[2]

에너지와 환경의 문제를 해결하기 위한 핵심 장비로서 가스터빈의 수요 증가가 예상되며, 가스터빈의 정교하고 정확한 진단 결과가 요구됨에 따라 데이터 기반의 가스터빈 진단 솔루션들이 개발될 것으로 예상된다. 가스터빈과 같이 극한 환경 소재로 구성된 중장비의 건전 상태에 대한 정확한 진단은 기기의 효율을 높이고 적절한 정비 시점을 예측함으로써 운영유지 비용을 줄일 뿐 아니라 파손에 의한 사고 위험성을 줄이고 안전에 기여할 수 있게 된다. 터빈 블레이드는 극한 환경에 놓여 있는 가스터빈의 핵심 구성품으로 소재의 열화도(degree of degradation)와 수명을 정량화하여 예측할 수 있다면 진단 솔루션의 정확도 향상에 기여 할 것으로 예상된다. 현행의 소재 열화도 평가 방법으로는 고온부품을 파괴하여 절단면의 미세조직을 주사전자현미경(SEM; Scanning Electron Microscope)등의 촬영장비로 촬영하여 소재 전문가가 γ/γ' 상의 형상 정보 등을 가지고 정성평가를

진행한다. 운행 중인 구성품의 보수를 위해 레플리카(replica)를 추출하는 경우에도 소재의 미세조직 이미지를 얻을 수 있으며 유사한 방법으로 소재 전문가가 열화도에 대한 정성평가 후 구성품의 폐기, 보수, 연장사용 등의 의사결정을 하게 된다. 크리프 수명과 잔여 수명은 Life Fraction Rule을 통해 계산할 수 있으며 이때 Larson-Miller Parameter(LMP)는 실제 작동 조건에 대한 응력-파단시험의 결과를 외삽 추론하는데 쓰이는 파라미터로 크리프 수명 예측에 주요한 파라미터이다.[3] LMP는 (1)의 식과 같이 크리프 시험에서 얻어진 파단시간과, 온도를 입력으로 얻을 수 있다.[4]

$$LMP = T(C + \log t) \quad (1)$$

이때, T는 온도(K), t는 응력-파단 시간, C는 상수로써 보통 20이 사용된다.

본 연구를 진행하게 된 배경은 니켈기 초합금을 소재에 대해 고비용으로 취득한 적은 수량의 미세조직 이미지 데이터를 기반으로 열화도를 설명력 있게 정량화하고, 고온부품을 파괴하지 않고 측정 가능한 환경정보를 통해 열화도와 크리프 수명을 예측하며 불확실성을 설명할 수 있는 모델을 찾기 위함이었다. 니켈기 초합금과 같이 내열성이 강한 소재는 크리프시험을 통해 파단까지 관측을 하기 위해 시간과 비용이 많이 투입되어 다양한 환경조건에 대한 데이터를 다량으로 취득하기 어렵다. 관측된 소량의 미세조직 이미지로부터 특성공학(Feature Engineering)을 통해 설명 가능하고 이상치에 강건한 대표 특징을 찾아 열화인덱스를 구성할 수 있는 방법이 필요하다. 과거의 시험 정보를 사전정보로 활용하여 미세조직 이미지가 입력되었을 때 이미지에서 열화 상태를 대표할 수 있는 유의 γ' 상을 자동으로 찾아내고, 유사 γ' 상의 특징들을 갖는 군집을 찾아내며, 정량화된 열화도를 잘 설명할 수 있어야한다. 이미지로부터 열화도를 정량화 후에는 적용을 위해 추가 파괴 검사 없이 측정 가능한 고온부품 소재의 환경 정보만으로 소재의 열화도 및 LMP를 예측할 수 있는 모델이 필요하며, 결과에 대한 직관적인 설명력을 갖춰야 소재 전문가에게는 소재 열화 평가의 시간을 단축시켜주고, 비전문가에게는 일관성 있고 직관적인 예측 결과를 제공함으로써 사업적인 활용도를 높일 수 있다. 소량의 데이터 및 다양하지 않은 시험 환경 정보로도 열화도 및 LMP를 예측해야 되는 목표를 달성하기 위해서 과거에 시험하지

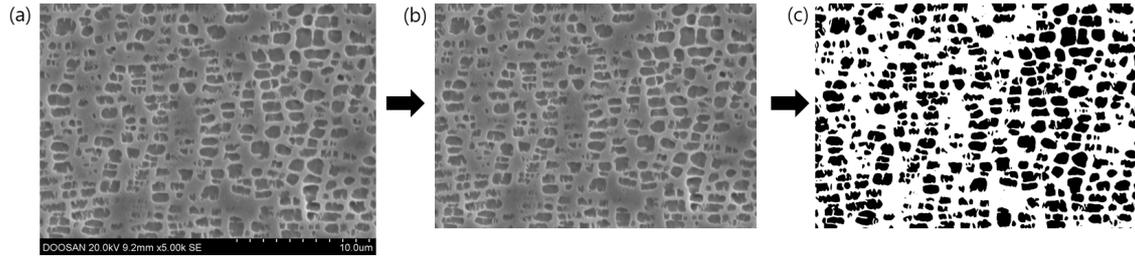


Fig. 1 Prediction process of gamma-prime phase region through U-Net (a) Microstructure original image (b) Region-of-interest image (c) Gamma-prime region prediction image

않아 관측되지 않은 환경정보에 대한 열화도와 LMP를 예측할 수 있고, 불확실성에 대한 설명으로 직관적인 해석을 가능하게 해야 한다. LMP 예측 모델의 경우 기존의 파괴시험 데이터로부터 확정적으로 계산되는 기존의 방법에서 벗어나 운영 중인 기기를 완전히 파괴하지 않고 레플리카 이미지만으로 LMP와 불확실성과 예측할 경우 크리프 수명을 입체적으로 해석할 수 있다.

II. 데이터 획득 및 전처리

2.1. 데이터 획득

시험에 사용된 소재는 DA-5161 SX로서 터빈 블레이드 제작에 사용되는 니켈기 초합금이다. ASTM E139-11 [5]에 제시된 금속 재료에 대한 크리프, 크리프-파단, 응력-파단 시험의 표준 시험 방법을 기준으로, 길이 120mm, 두께 6.25mm인 시편을 기준으로 16가지 환경조건(응력, 온도) 하에 파단이 되는 시간을 측정하였다. 파단이 된 시험편은 소재에 대한 환경조건의 영향도를 관찰하기에 적절한 위치를 절단하여 주사전자현미경(SEM; Scanning Electron Microscope)으로 Hitachi 모델 S-3400N을 사용하여 촬영하였다. 촬영 전 시편 가공의 방법으로는 연마지 #200, #400, #600, #800, #1000를 순차적으로 사용하여 그라인딩하고, 최종 폴리싱으로 Colloidal silica 1 μ m를 사용하여 1시간 이상 연마하였다. 에칭은 Marble's reagent(10g CuSO₄ + 50ml HCL + 50ml 증류수)로 진행하였다. 16가지 환경조건에서 각각 시험한 동일 시편에서 위치를 다르게 촬영하여 한 시편당 약 30~120장씩, 총 991장의 이미지를 촬영하였다. 촬영 조건으로 가속전압(Accelerating Voltage)은 20kV, 배율(Magnification)은 x5.00k, 작동거리(Working

Distance)는 9.2mm이며, 이미지 크기는 1280x960이다.

2.2. 데이터 전처리

그림 1과 같이 취득한 환경조건 별 이미지 데이터는 하단의 촬영조건을 배제한 조직상의 이미지만 보일 수 있도록 관심영역(ROI; Region of Interest)을 추출한다. 이때 이미지의 크기는 1280x960에서 1280x896로 줄어든다. 관심영역 이미지에 대응하는 γ' 의 Semantic Segmentation 예측 모델을 위한 52장의 어노테이션(Annotation)을 수행하여 Segmentation용 학습이미지를 구성했다. 학습이미지로 딥러닝 모델 중 U-Net을 통해 학습모델을 만들었다. 이때 입력 이미지는 224x224x3이며, 출력 이미지는 224x224x1의 형태를 갖는다. 총 파라미터 수는 약 22백만개이며 그 중 학습에 사용된 파라미터는 약 8백만개이다. 본 연구를 위해 사용된 991장의 크리프 시험 이미지의 관심영역은 학습된 U-Net의 입력이 되어 γ' 의 영역이 예측된 이진화 이미지(γ' : 검정색, γ : 흰색)가 출력되어 딥러닝 모델에 의해 γ' 을 빠르게 특징화 할 수 있다.

시험 환경 조건은 총 16가지로 각 조건별 온도, 응력, 파단시간, LMP는 Min-Max 정규화를 통해 0.1에서 0.9 사이 값을 갖도록 처리하였다.

III. 연구방법 및 실험결과

3.1. 연구 방법

본 연구에서는 새로운 열화 인덱스의 정량화 방법을 제안하고 열화 인덱스와 LMP 모델의 학습방법에서는 γ' 의 기하학적인 특징별 다양한 회귀모델의 성능을 비교하고, 유의 설명변수의 선택여부에 따른 모델 성능이

향상되는지 실험하였다. 특히 베이지안 회귀모델과 가우시안 프로세스 회귀모델은 MCMC를 활용한 Metropolis-Hastings Sampler를 통해 회귀식을 구성하는 파라미터(회귀계수, 오차항 등)의 모수를 추론하고 예측의 불확실성을 확률로서 직관적으로 설명하였다.

데이터셋은 991장의 이미지에 대해 각 환경조건별 랜덤으로 80%는 학습데이터(792장)로, 20%는 테스트 데이터(199장)로 분리하였다. 학습데이터는 열화 인덱스 정량화 및 예측모델을 학습하는데 사용되었으며, 테스트데이터는 예측모델의 성능을 검증하기 위해 사용되었다.

3.1.1. 열화 인덱스의 정량화 방법

샘플링 기반 베이지안 추론 방법을 활용한 열화 인덱스 정량화 방법으로는 그림 2와 같은 순서로 진행된다. SEM 이미지를 촬영 후 U-Net을 통해 미세조직인 γ' 을 빠른 특징화(Fast Characterization)하고 γ/γ' 영역을 이진화하여 파괴시험 데이터베이스에 축적 및 저장된다. 이를 Bag-of-Image(BoI)라고 명명하였다. 추후 크리프 시험과 같은 파괴 시험이 연구용으로 진행되면 BoI에 실험에 사용된 환경조건들과 특징화된 이미지 데이터가 쌓이게 된다. 이런 과정은 연구과정에서 γ/γ' 영역을 구분하기 위해 소요되는 시간을 절약할 뿐 아니라 데이터가 축적됨에 따른 열화 인덱스의 모수에 가까운 추론

을 할 수 있다.

BoI가 구성되면 Bag-of-Feature(BoF)를 구성하게 된다. 여기서 Feature가 의미하는 바는 이미지 내에 존재하는 γ' 영역의 기하학적 특징(Geometric Feature)를 의미한다. γ' 의 기하학적 특징을 사용함으로써 이후 생성되는 열화 인덱스를 γ' 의 기하학적으로 설명할 수 있게 된다. 사용한 기하학적인 특징은 총 12가지이며, BoI에 있는 모든 이미지에서 관측되는 γ' 의 기하학적 특징을 수치로 정량화하여 Bag-of-Feature(BoF)를 구성한다. 이 과정은 γ' 의 형상으로부터 수치화된 정량화하는 것으로 특성공학(Feature Engineering)이 적용되는 과정이다. 본 연구에서 사용한 기하학적 특징 중 Contour perimeter는 γ' 의 둘레의 길이를 의미하며, Contour area는 γ' 내부 면적을 의미한다. Convexhull perimeter는 [6]에서 제안한 γ' 을 블록하게 감싸는 둘레의 길이이며, Convexhull area는 그 넓이를 의미한다. Aspect ratio는 식 (2)와 같이 γ' 을 둘러싸는 bounding box의 종횡비를 의미하며 전통적으로 래프팅의 정도를 판단할 때 많이 사용한다. 본 연구에서 종횡비는 다음과 같이 정의하였다.

$$aspect\ ratio = \frac{width}{height} \quad (2)$$

Elongation은 식 (3)과 같이 γ' 을 둘러싼 bounding box의 늘어난 정도로 짧은 변의 길이를 긴 변의 길이로

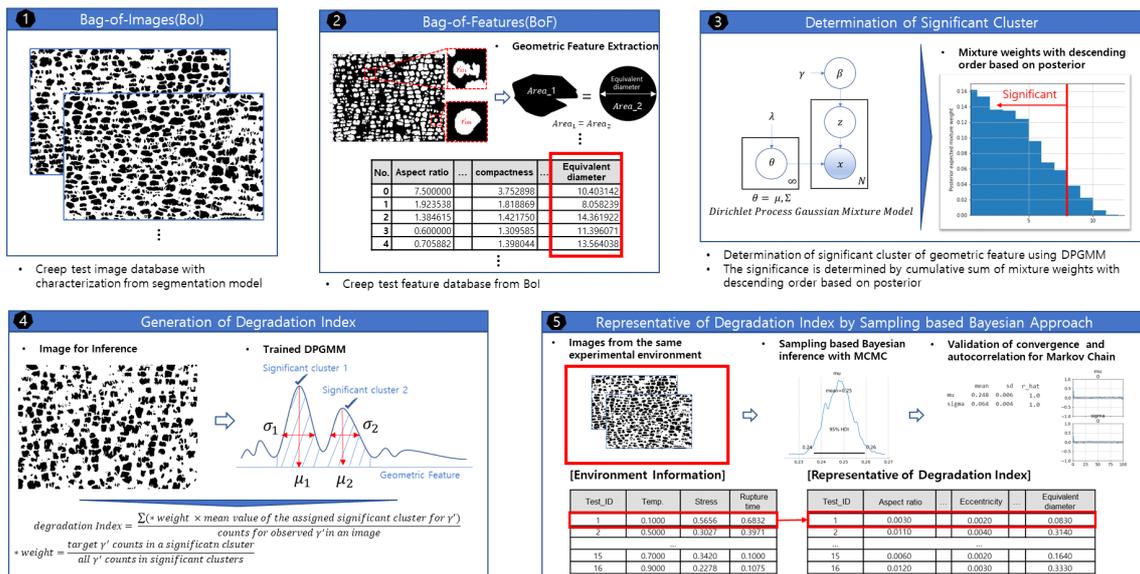


Fig. 2 A method for quantifying degradation index using a sampling-based Bayesian inference method

나는 값이다.[7]

$$elongation = \frac{length_1}{length_2} (length_1 < length_2) \quad (3)$$

본 연구에서는 Compactness는 식 (4)과 같이 Contour Area에 대해 얼마나 조밀한 Contour Perimeter를 갖는지 재정의하였으며 Contour Perimeter가 원에 비교하여 얼마나 변형되었는지 알 수 있다. 원의 경우 1의 값을 가지며, 복잡한 형태이면 Compactness의 값은 1보다 증가하게 된다.[7]

$$compactness = \frac{contour\ perimeter^2}{4\pi \times contour\ area} \quad (4)$$

Solidity는 식 (5)와 같이 오목도(concavity)를 측정할 수 있는 지표로서 원이나 타원의 경우에는 1의 값을 가진다.[7]

$$solidity = \frac{contour\ area}{convex\ hull\ area} \quad (5)$$

Modified eccentricity는 식(6)과 같이 물체를 감싸는 타원의 major axis의 길이와 minor axis의 길이의 비로 정의하였다. 이것은 기존의 eccentricity와는 다른 식으로서 타원의 축 길이를 이용한 점을 고려하여 Modified eccentricity로 정의하였다.[7]

$$modified\ eccentricity = \frac{minor\ axis\ length}{major\ axis\ length} \quad (6)$$

Angle은 물체를 둘러싼 bounding box의 기울어진 각도를 의미하며, Equivalent diameter는 식 (7)과 같이 물체의 면적과 동일한 면적을 갖는 원의 지름을 의미한다.

$$equivalent\ diameter = \sqrt{\left(\frac{4 \times contour\ area}{\pi}\right)} \quad (7)$$

본 연구에서는 792장의 학습데이터를 기반으로 BoF를 구성하였으며 각 기하학적 특징 12가지에 대한 156,862개의 정량화된 특징값을 얻었다. 각 기하학적 특징별로 Dirichlet Process Gaussian Mixture Model (DPGMM)로 비지도학습을 진행하여 사후분포에 기반한 Mixture weight들을 내림차순으로 나열하여 누적합이 0.9이 되기 전의 클러스터를 유의한 클러스터로 판단하였다.

열화 인덱스를 구하고자 하는 이미지에 대해서 각각

의 γ' 영역에 대해 기하학적인 특징이 계산되며, 이미지 내에 존재하는 γ' 개수만큼 특징값을 얻을 수 있다. 이미지의 열화를 설명하는 특징의 차원이 γ' 개수만큼 늘어나는 것을 회피하기 위해 이미지 하나에 대해 하나의 기하학적 특징이 추출될 수 있는 방법을 제안한다. 기하학적 특징별 학습된 DPGMM의 클러스터의 평균 벡터와 공분산 행렬을 활용하여 다변량정규분포의 가능도에서 가장 높은 확률값을 가지는 클러스터를 γ' 기하학적 특징값의 클러스터로 정하고 그 중 결정된 유의 클러스터에 해당하는 기하학적 특징값을 이용하여 다음 식 (8)과 같이 열화 인덱스를 정의할 수 있다.

$$열화\ 인덱스 = \frac{\sum(*weight \times \gamma')\text{이 할당된 클러스터의 평균}}{\text{이미지 내 존재하는 } \gamma'\text{의 총 개수}} \quad (8)$$

$$*weight = \frac{\text{해당 클러스터 위치의 } \gamma'\text{ 개수}}{\text{이미지 내 할당된 클러스터가 있는 } \gamma'\text{ 총 개수}}$$

식 (8)은 유의 클러스터에 해당하는 γ' 만 선택하고 선택하지 않은 γ' 특징값은 제거한다. 그 효과로서 Segmentation 예측의 오류의 영향도, 이미지 노이즈에 대해 강건하고 유의미한 γ' 의 영향을 잘 반영할 수 있다. 이미지 내 주류를 이루는 γ' 의 클러스터의 영향을 강화하기 위해 각 클러스터에 할당된 γ' 개수와 이미지내 총 유의 γ' 개수의 비율로 가중평균을 구한다. 마지막으로 이미지 내 존재하는 γ' 의 총 개수로 나누어서 동일 배율에서 관측되는 γ' 의 개수 정보를 열화 인덱스에 반영하고자 하였다.

미세조직 이미지별 열화 인덱스로서 하나의 기하학적 특징이 대응이 된 후에는 동일한 시험조건 하에 촬영된 수십장의 이미지의 열화 인덱스들을 대표할 수 있는 하나의 대표값을 만든다. MCMC를 기반한 샘플링 기반 베이지안 추론 방법을 사용하여 열화 인덱스의 대푯값을 설정하였다. 그림 3과 같은 베이지안 모델을 기반으로 MCMC 및 Metropolis-Hastings Sampler를 활용하여 샘플링을 통한 열화 인덱스의 평균에 대한 사후분포의

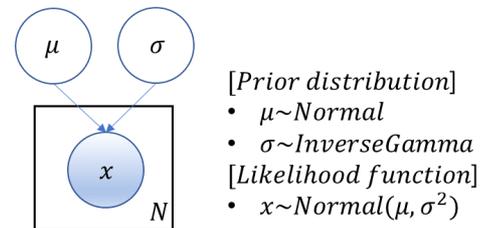


Fig. 3 Bayesian plate diagram for inferring the representative value of degradation index

기댓값을 열화 인덱스의 대푯값으로 정의하였다. 열화 인덱스의 대푯값 열화 인덱스의 평균에 대한 사후분포가 신뢰성을 확보하기 위해서는 수렴성 검정과 자기상관성 검증을 해야하는데, 샘플링 결과에 대한 수렴성과 자기상관성을 정성적으로 평가할 수 있으며, 정량적으로는 자기수렴성 검정으로 \hat{R} 값이 1 ± 0.05 범위 안에 있으면 수렴한 것으로 판단, Effective Sample Size(ESS)가 500 이상이면 자기상관성(Autocorrelation)이 낮은 신뢰할만한 Markov chain으로 판단하였다. 수렴성과 자기상관성의 기준을 충족하기 위해 Metropolis-Hastings 샘플링의 iteration은 10,000회, 4개의 Markov Chain을 가동하였으며, \hat{R} 값을 만족하기 위해 5,000개에 대한 샘플을 burn-in, 자기상관성 기준을 만족하기 위해 5개의 샘플마다 thinning을 진행하였다. burn-in 및 thinning을 진행한 샘플들의 분포를 사후분포로 평가하였고 그 중 μ 의 사후분포의 기댓값으로 열화 인덱스의 대푯값을 최종 추론하였다. 열화 인덱스에 대한 대푯값 추론은 학습데이터와 테스트데이터 각각 수행되었으며, 각 동일 환경조건에 대한 학습데이터와 테스트데이터가 추론한 열화 인덱스의 변동성은 식 (9)에 의해 정량화했으며 결과는 표 1과 같다.

$$\left[\left(\sum_i^N \frac{|DI_i^{train} - DI_i^{test}|}{DI_i^{train}} \right) / N \right] \times 100(\%) \quad (9)$$

Table. 1 Variability of the representative value inference result of the degradation index of the training versus test image data

Degradation Index by geometric feature	Variation ratio(%)
contour_perimeter	11.41%
contour_area	12.95%
convexhull_perimeter	11.74%
convexhull_area	12.34%
aspect_ratio	14.10%
elongation	24.93%
compactness	11.01%
solidity	12.10%
extent	12.96%
modified_eccentricity	20.16%
angle	6.85%
equivalent_diameter	11.77%

여기서 DI_i^{train} 는 학습데이터에서 i 번째 실험 환경조건에 대한 이미지를 대표하는 열화 인덱스를 의미하며, DI_i^{test} 는 학습데이터에서 i 번째 실험 환경조건에 대한 이미지를 대표하는 열화 인덱스를 의미하고, N 은 환경조건 수 본 연구에서는 16이 된다. 학습과 테스트 이미지 데이터로부터 추론되는 12개의 기하학 특징별 열화 인덱스의 변동성은 평균적으로 13.53%로 나타났으며 변동성이 가장 큰 것은 elongation이었고, 변동성이 작은 순서로 열화 인덱스 종류를 기하학적 특징으로 5가지를 나열하면 angle이 6.85%로 가장 작으며 compactness, contour perimeter, convexhull perimeter, equivalent diameter 순으로 작았다.

3.1.2. 열화 인덱스 예측 모델

학습 이미지 데이터로부터 16가지 실험 환경조건에 대해 하나씩 대응되는 열화 인덱스 대푯값을 추론한 결과를 반응변수로, 소재가 처한 환경 조건(온도, 응력, 시간)을 설명변수로 하는 회귀모델을 구성함으로써 열화 인덱스 예측 모델을 학습할 수 있다. 회귀모델로서는 Ordinary Least Square(OLS) 회귀모델, Generalized Linear Model(GLM), Bayesian Regression(BR), Gaussian Process Regression(GPR)을 사용하였으며 각 모델에 대해 유의 설명변수를 선택하는 과정의 유무에 따라 성능 변화를 관찰하였다. BR과 GPR 모델의 경우 OLS 및 GLM 모델에 비해 갖는 특징점으로는 베이지안 통계기반의 모델로서 학습데이터에 관측되지 않은 설명변수에 대해서도 불확실성의 설명과 함께 예측이 가능하다. BR과 GPR은 각 모델의 파라미터를 랜덤변수로 보고 MCMC 및 Metropolis-Hastings(MH) Sampler를 활용하여 추출된 샘플로부터 파라미터의 사후분포를 예측하고, 사후분포의 통계량으로부터 모델을 설명할 수 있다.

설명변수로서 환경조건(온도, 응력, 시간)을 주성분 분석(PCA)을 통해 하나의 주성분(PC)으로 차원 축소하는 방법과 라쏘(LASSO) 회귀를 통해 유의변수를 중요도가 높은 순서로 2개를 선택하는 방법을 비교하였다. 설명변수를 1개 또는 2개 차원으로 축소하는 것은 회귀 모델의 시각화시 직관적인 해석을 지원하기 위함이다.

하나의 PC로 설명변수를 차원축소 하는 경우 학습데이터 대상으로 PC를 설명변수, 기하학 특징에 따른 열화 인덱스를 반응변수로 하여 1에서 5차수까지의 다항 회귀를 5-fold cross validation 결과 RMSE값이 가장 낮

은 차수를 적정 차수로 결정한다.

2개의 유의 설명변수를 선택하는 과정으로는 온도, 응력, 시간 3개의 변수를 standard scaler로 정규화 후 설명변수, 기하학적 특징에 따른 열화 인덱스를 반응변수로 하여 1에서 4차수까지 교호변수를 포함하여 라쏘 회귀를 정규화 강도를 나타내는 하이퍼파라미터 α 를 변형시키면서 5-fold cross validation을 진행하여 negative mean squared error가 가장 큰 회귀 차수 및 α 의 조합에 대해 회귀계수의 절대값이 큰 순서로 2개를 최종 유의 변수로 선택하며, 그때의 차수를 다항회귀의 차수로 정한다.

학습 데이터에 대해 설명변수는 온도, 응력, 시간을 하나의 주성분으로 구성한 값 또는 유의변수로 선택한 변수로 정하고 반응변수는 기하학적 특징을 기반에 둔 각 열화 인덱스로 하여 네가지 회귀모델(OLS, GLM, BR, GPR)을 구성하였다.

OLS는 최소제곱법을 활용하여 회귀모델의 예측값과 실제값 사이의 잔차를 최소화 하는 회귀계수를 찾는 방법이며, 다항회귀의 경우 식 (10)과 같다.

$$\arg \min_{\Theta} E(\Theta) = \arg \min_{\Theta} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (10)$$

where $\Theta = (\theta_0, \theta_1, \dots, \theta_k)$,

$$\hat{y}_i = \theta_0 + \theta_1 x_i + \dots + \theta_k x_i^k$$

이때 $E(\Theta)$ 는 에러항, \hat{y}_i 은 i 번째 데이터의 예측값, y_i 는 i 번째 데이터의 실제값, n 은 데이터의 개수, x_i 는 i 번째 데이터의 설명변수에 해당하는 입력값, k 는 다항회귀의 차수를 의미한다. 회귀계수 θ_k 로 $E(\Theta)$ 를 편미분한 식이 0이 될 때 $E(\Theta)$ 가 최소화하는 회귀계수 θ_k 를 찾을 수 있다. 본 연구에서는 식 (11)과 같이 설명변수로서 하나의 주성분을 활용한 2차 OLS 다항회귀 모델과 식 (12)와 같이 기하학적 특징별 열화 인덱스에 대해 유의 설명변수 2개를 선택한 회귀모델의 두 경우에 대한 모델을 구성하였다.

$$y = \theta_0 + \theta_1 PC_1 + \theta_2 PC_1^2 + \epsilon \quad (11)$$

where $\epsilon \sim N(0, \sigma^2)$

$$y = \theta_0 + \theta_1 \cdot selectedvariable_1 + \theta_2 \cdot selectedvariable_2 + \epsilon \quad (12)$$

where $\epsilon \sim N(0, \sigma^2)$

GLM은 선형 모형의 반응변수의 데이터 타입이 연속형이나, 범주형인 경우에도 사용할 수 있도록 식 (13)과 같이 일반화한 선형 모형이다.

$$g(\mu) = g(E(Y | x_1, \dots, x_n)) \quad (13)$$

$$= g(\theta_0 + \theta_1 x_1 + \dots + \theta_n x_n)$$

where $g(\cdot) = link\ function$

선형예측값의 범위와 μ 의 범위가 다를 때의 한계점을 연결함수(link function)을 통해 해결할 수 있다. 본 연구에서는 식 (14)와 같이 설명변수로서 하나의 주성분을 활용한 GLM과 식 (15)와 같이 기하학적 특징별 열화 인덱스에 대해 유의 설명변수 2개를 선택한 GLM의 두 경우에 대한 모델을 구성하였다.

$$\ln(\mu) = \ln(E(Y | PC_1)) \quad (14)$$

$$= \ln(\theta_0 + \theta_1 PC_1)$$

where link function = $\ln(\cdot)$

$$\ln(\mu) = \ln(E(Y | selectedvariable_1, selectedvariable_2)) \quad (15)$$

$$= \ln(\theta_0 + \theta_1 selectedvariable_1 + \theta_2 selectedvariable_2)$$

where link function = $\ln(\cdot)$

BR 및 GPR은 베이저안 통계에 기반한다. 베이저안 통계에서 가장 근간이 되는 이론은 베이즈 이론(Bayes' Theorem)은 다음 식 (16)과 같이 알려져 있다.

$$P(w | D) = \frac{P(D | w)P(w)}{P(D)} \quad (16)$$

where $P(D) = \int P(D | w)P(w)dw$

이때, $P(D | w)$ 는 가능도 함수이며 각각의 매개변수 벡터 w 에 대해 관측된 데이터 집합이 얼마나 그렇게 나타날 가능성이 있는지를 표현한다. $P(w)$ 는 사전확률분포로서 이전의 지식을 활용한 믿음의 정도로 볼 수 있으며, 데이터를 관측하기 전의 w 에 대한 가정으로 할 수 있다. $P(w | D)$ 는 사후확률분포로서 사전확률분포와 가능도의 곱에 비례하여 사전 확률 분포를 업데이트한다.

베이저안 추론시 매개변수의 집합에 대한 결합사후분포를 계산해야하나 적분의 계산이 실제 어렵고, 차원이 높아질수록 그 난이도는 증가하므로 analytic equation을 통해 문제 해결이 어려워지게 된다. 이 문제를 해결하기 위해 본 연구에서는 Metropolis-Hastings Sampling 알고

리즘에 기반하여 베이지안 추론을 진행했다. Metropolis-Hastings Sampling 방법은 다음과 같다: (1) 제안(후보) 샘플 $x^{candidate}$ 을 제안분포 $q(x^{(i)} | x^{(i-1)})$ 로부터 샘플을 추출한다. (2) 수락함수(acceptance function) $\alpha(x^{candidate} | x^{(i-1)})$ 로부터 제안분포와 결합분포 $\pi(\cdot)$ 에 기반한 수락확률(acceptance probability)을 계산한다. (3) 제안(후보) 샘플 $x^{candidate}$ 을 수락확률 α 의 확률로 채택하며, $1 - \alpha$ 의 확률로 거절한다. 거절된 샘플은 제거하고 이전 샘플 $x^{(i-1)}$ 로 대체한다.[8]

MCMC에서 iteration을 수행한 뒤 결과로 얻은 추론하고자 하는 모수의 사후분포에 대한 Markov Chain이 충분히 수렴했는지를 평가하기 위해 sampling iteration을 시각화한 그래프를 통해 정성평가, \hat{R} 값이 1 ± 0.05 범위 안에 있으면 수렴한 것으로 판단하였다. 자기상관성은 정성적으로 자기상관성 플롯을 통해 Lag에 따라 자기상관성이 특정 방향으로 패턴이 없고 충분히 작은지 평가하고, 정량적으로 Effective Sample Size(ESS)가 500 이상인지 기준으로 신뢰할만한 Markov chain으로 판단하였다. 본 연구에서는 열화도 예측 모델 학습을 위해 BR 및 GPR은 Metropolis-Hastings 샘플링의 iteration은 20,000회, 4개의 Markov Chain을 가동하였으며, \hat{R} 값을 만족하기 위해 10,000개에 대한 샘플을 burn-in, 자기상관성 기준을 만족하기 위해 8개의 샘플마다 thinning을 진행하였다. burn-in 및 thinning을 마친 샘플들의 분포를 최종 사후분포로 평가하였다. 모수에 대한 사후분포가 생성되면 새로운 데이터 x^* 에 대한 예측을 위해 식 (17)과 같이 사후예측분포를 활용하여 새로운 설명변수에 대한 반응변수의 값을 예측할 수 있다.

$$p(y^* | x^*, y) = \int p(y^* | x^*, w) p(w | y) dw \quad (17)$$

이때 y 는 관측된 반응변수이며, w 는 추론하고자 하는 매개변수이다. x^* 는 예측하고자 하는 새로운 설명변수 입력이며, y^* 는 설명변수 x^* 에 대한 예측 반응변수이며 $p(w | y)$ 는 매개변수에 대한 사후분포이다. 사후예측 분포도 적분의 계산이 어려운 관계로 샘플링기반의 근사를 통해 analytic solution을 회피할 수 있는데, 직관적인 해석으로는 사후분포에서 w 의 샘플을 생성해내고, 새로운 설명변수 x^* 및 샘플링된 w 과 정밀도 β 를 고려한 y^* 를 샘플링 함으로서 예측한다.

BR은 그림 4와 같이 모델이 구성되며 회귀계수 및 오차항을 랜덤변수로 간주하여 MCMC 및 Metropolis-Hastings Sampling을 통해 각 그 사후분포를 추론하였다. 그림 4(a)는 설명변수로서 하나의 주성분을 사용한 BR이며, 그림 4(b)는 기하학적 특징별 열화 인덱스를 설명하기에 유의한 설명변수 2개를 선택하여 열화 인덱스를 예측한 BR의 graphical representation이다.

GPR에서 입력 x 로 하는 함수 f 의 출력 y 은 식 (18)과 같이 “시그널”(f)과 “노이즈”(ε)로 이뤄져 있고 시그널 부분은 특정 분포를 따르는 랜덤변수가 된다.

$$y = f(x) + \epsilon \quad (18)$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$

이 분포는 f 의 불확실성(uncertainty)을 나타내며 각기 다른 입력값에 대한 함수의 출력이 관측됨으로서 불확실성이 낮아질 수 있다. GPR에서 $f(x)$ 는 식 (19)와 같이 가우시안 프로세스(Gaussian Process)를 따른다.

$$f(x) \sim GP(m(x), k(x, x')) \quad (19)$$

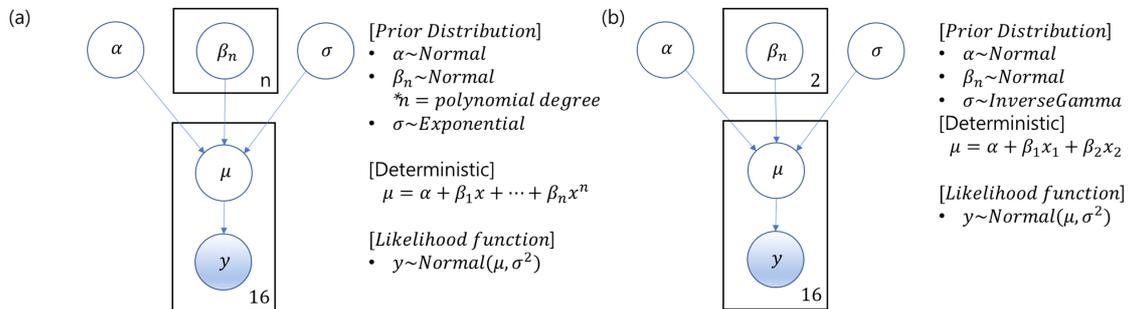


Fig. 4 Plate Notation of Bayesian regression of the degradation index prediction model (a) When explanatory variable is used as a principal component (b) When the top two significant explanatory variables are selected

가우시안 프로세스(GP)는 함수에 대한 분포이고 평균(mean)과 공분산(covariance) 함수로 구성되어있다. 평균 함수와 공분산함수는 식 (20), (21)과 같이 정의된다:

$$m(x) = E[f(x)] \quad (20)$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x')))] \quad (21)$$

이때, x, x' 는 각기 다른 입력값을 의미한다. k 는 가우시안 프로세스의 커널(kernel)이라고 불리며 평활도와 데이터로부터 기대되는 비슷한 패턴과 같은 가정에 기반한 적절한 커널을 선택하게된다. 가장 널리 알려져 있는 커널은 RBF(Radial Basis Function) 커널이며 식 (22)와 같이 정의된다:

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{\|x - x'\|^2}{2\lambda^2}\right) \quad (22)$$

이 커널에서 λ 는 길이척도, σ_f^2 는 신호 분산을 의미하며 데이터포인트 사이 사전 상관관계를 낮추거나 높이며 결과적으로 결과로 나오는 함수의 변동성에 따라 달라진다. X_* 는 각 행이 새로운 입력인 $x_i, i = 1, \dots, n$ 로 이루어져 있는 행렬이다. 함수를 샘플링하기 위해 X_* 상간의 공분산을 계산하여 $n \times n$ 행렬을 다음과 같이 만든다:

$$K(X_*, X_*) = \begin{bmatrix} k(x_1^*, x_1^*) & \dots & k(x_1^*, x_n^*) \\ \vdots & \ddots & \vdots \\ k(x_n^*, x_1^*) & \dots & k(x_n^*, x_n^*) \end{bmatrix}$$

사전 평균 함수 $m(x) = 0$ 으로 선택하여 행렬 계산을 단순화 할 수 있다. 다변량 정규분포로부터 샘플링을 함으로서 GP로부터 X_* 입력으로 하는 f 의 샘플을 추출할 수 있다:

$$f_* \sim N(0, K(X_*, X_*))$$

where $f_* = [f(x_1^*), \dots, f(x_n^*)]^T$

f_* 는 함수값들 중의 하나의 샘플이다. 예측 대상인 y_* 를 샘플링 하기 위해서 노이즈 항인 ϵ 를 추가해야 한다. 관측 데이터가 $D_t = \{X_t, y_t\}$ 이고 사후분포 $p(f | D_t)$ 로부터 f_* 를 샘플링하여 새로운 입력인 X_* 에 대한 예측을 해야한다. 정의에 의해서 이미 관측된 y_t 와 함수값 f_* 는 결합 다변량 정규분포를 따르며 다음과 같다:

$$\begin{bmatrix} y_t \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X_t, X_t) + \sigma_\epsilon^2 I & K(X_t, X_*) \\ K(X_*, X_t) & K(X_*, X_*) \end{bmatrix}\right)$$

여기서 $K(X_t, X_t)$ 는 관측된 값들 사이의 공분산 행렬, $K(X_*, X_*)$ 는 새로운 값들에 대한 공분산 행렬, $K(X_*, X_t)$ 는 새로운 값들과 이전에 관측된 값들 사이의 공분산 행렬, $K(X_t, X_*)$ 는 이전에 관측된 값들과 새로운 값들 사이의 공분산 행렬이고 I 는 항등행렬이다. 최종적으로 새로운 값들에 대한 사후분포는 다음 식 (23)을 가지게 되며 함수값을 추론할 수 있다.[9, 10]

$$p(f_* | X_t, y_t, X_*) \sim N(K(X_*, X_t)[K(X_t, X_t) + \sigma_\epsilon^2 I]^{-1}y_t, K(X_*, X_*) - K(X_*, X_t)[K(X_t, X_t) + \sigma_\epsilon^2 I]^{-1}K(X_t, X_*)) \quad (23)$$

본 연구에서는 GPR의 RBF 커널에서의 하이퍼파라미터인 σ_f, λ 와 σ_ϵ 를 랜덤변수로 두고 MCMC 및 Metropolis-Hastings Sampling을 통해 사후분포를 근사하였다.

3.1.3. LMP 예측 모델

LMP 예측 모델 학습 방법은 열화도 예측 모델 학습 방법과 유사하나 설명변수를 미세조직 이미지로부터 얻어진 12가지의 기하학적 특징별 열화 인덱스로, 반응변수를 크리프 수명을 예측하는데 사용할 수 있는 LMP로하는 회귀모델을 학습하는 방법이다. 회귀모델로서는 OLS 회귀, GLM 및 MCMC 샘플링 기반의 BR, GPR을 사용하여 다음 두 가지 경우에 대해 학습하였다: 1) 미세조직 이미지로부터 추출되는 12가지의 기하학적 특징을 가지는 열화 인덱스를 하나의 주성분으로 차원 축소하고 반응변수를 LMP로 연결시키는 방법과 2) 기하학적 특징별 열화 인덱스 각각을 설명변수로하고 LMP를 반응변수로 연결하는 방법을 사용하였다. 1)의 방법은 고온부품의 레플리케이션을 입력으로 모델에 의해 신속한 LMP 예측을 함으로서 비전문가도 크리프 수명을 정량화하여 직관적으로 평가하는데 사용하기 위함이다. 2)의 방법은 학습된 각 열화 인덱스별 LMP 예측 모델을 열화 예측모델 후단에 연결함으로써 열화 시뮬레이션 시 소재의 환경정보를 입력하면 12가지 기하학적 특징별 열화 인덱스가 출력되고 이것이 LMP 예측모델의 입력이 되어 열화 인덱스와 LMP를 동시에 예

측할 수 있게 하기 위함이다.

1)의 방법으로 열화 예측모델 학습 방법과 마찬가지로 12가지 열화인덱스를 하나의 주성분으로 차원축소하여 설명변수로, LMP를 반응변수로 하여 1에서 5차수까지의 다항회귀를 5-fold cross validation 결과 RMSE 값이 가장 낮은 차수를 적정 차수로 결정한다. 본 연구의 결과로는 3차(degree)가 적정차수로 결과를 얻었다. 이후 3차 다항 OLS 회귀, GLM, MCMC 샘플링 기반의 BR, GPR을 학습을 수행하였다. BR 및 GPR은 사후예측분포로부터 샘플들을 생성하여 새로운 설명변수값에 대한 예측을 진행했다. 추가적으로 라쏘(LASSO) 회귀를 통해 기하학 특징 중 반응변수인 LMP를 설명하는 중요도가 높은 순서로 상위 5개의 유의변수를 선정하고 1개의 주성분으로 차원축소하여 회귀모델 학습하여 12가지 열화 인덱스를 하나의 주성분으로 차원축소한 LMP 예측 모델 대비 성능변화가 있는지 실험하였다.

3.2. 실험결과

3.2.1. 평가지표

3.2.1.1. Root Mean Squared Error(RMSE)

RMSE는 모든 데이터에 대한 예측값과 실제값의 차를 제공하여 평균을 구하여 제곱근을 취한 값으로, 회귀모델의 성능 지표로서 많이 사용되며 단위(unit)가 데이터의 단위와 같아 직관적으로 모델의 오차가 얼마나 되는지 알기 쉽다. RMSE가 작은 모델일수록 모델의 예측값과 실제값이 차이가 크지 않다는 것을 의미한다. RMSE는 식 (24)과 같이 구할 수 있다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

이때 \hat{y}_i 는 i 번째 데이터의 예측값, y_i 는 i 번째 데이터의 실제값을 의미하며, n 은 데이터의 개수를 의미한다.

3.2.1.2. R-Squared(R^2)

R^2 는 결정계수이며 설명변수가 설명할 수 있는 반응변수의 분산량의 정도로 해석하고, 회귀 모델의 적합도(goodness of fit)를 평가할 수 있으며 식 (25)와 같이 구할 수 있다.

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

이때 \hat{y}_i 는 i 번째 데이터의 예측값, y_i 는 i 번째 데이터의 실제값을 의미하며, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ 이며, n 은 데이터의 개수를 의미한다.

3.2.2. 열화 인덱스 예측모델 성능

열화 인덱스 예측모델은 OLS, GLM, BR, GPR 4개의 회귀모델과 설명변수로서 환경조건(온도, 응력, 시간)에 대해 주성분 1개로 차원축소하여 입력, 선택된 유의변수 2개 입력하는 경우에 대해 반응변수를 열화 인덱스로하여 각 기하학 특징별 열화인덱스에 대해 8개 모델 성능을 비교하였다. 실험 결과 elongation 열화 인덱스를 제외한 모든 종류의 열화인덱스가 2개의 유의변수를 설명변수로 GPR 모델로 예측할 때 테스트 데이터에 대한 우수한 성능을 보였다.

테스트 데이터에 대해 열화인덱스 예측 모델의 설명변수를 하나의 주성분으로 할 때와 2개의 유의변수를 선택할 때의 성능 비교를 하였다. 모든 열화 인덱스에 대해 평균적으로 OLS는 RMSE가 28%, GLM은 29%, BR은 18%, GPR은 36% 낮아지는 개선 효과를 보였다. 특히 Modified Eccentricity 열화 인덱스는 유의변수를 선택하여 설명변수로 사용시 64%의 RMSE가 낮아지는 효과가 있었다. Modified Eccentricity 열화 인덱스가 유의설명변수를 선택한 OLS 모델에서는 R^2 가 0.9588로 나타나 가장 설명력이 높은 열화 인덱스였다. 이외에도 Elongation, Angle 열화 인덱스도 R^2 가 0.9 이상으로 열화를 잘 설명하는 기하학적 특징임을 확인할 수 있었다.

설명변수를 하나의 주성분으로 할 때와 2개의 유의변수를 선택할 때 RMSE 기준 학습성능 대비 테스트 성능이 어떻게 변화하는지 실험하였다. 학습시 RMSE 대비 테스트시 RMSE가 차이가 크면서도 높게 나타나다면 모델의 과적합(overfitting)이 이뤄진 것으로 볼 수 있으므로 모델의 적합도를 검증할 수 있다. 설명변수를 하나의 주성분으로 할 때는 BR이 평균적으로 학습대비 테

스트 RMSE가 4%만 증가하여 가장 적게 증가했다. 2개의 유의 설명변수를 입력하는 경우에는 GLM의 경우 평균적으로 학습대비 테스트 RMSE가 11%가 감소하였으며, BR의 경우 평균적으로 9% 증가한 것을 확인하였다. 기하학적 특징별 열화도 인덱스 관점에서의 적용 회귀모델의 종류와 관계없이 학습대비 테스트 RMSE가 낮게 결과가 나온 열화 인덱스는 설명변수로 하나의 주성분 입력시 Contour perimeter, Convexhull perimeter, Convexhull area, Equivalent diameter이었고, 설명변수로 열화인덱스를 설명하는데 중요도가 높은 변수 2개를 입력시 Elongation, Modified Eccentricity이었다. 각 열화인덱스의 기하학적 특징별 적합한 모델의 종류가 달라짐을 확인하였다. 회귀모델의 학습성능 대비 테스트 성능의 변동이 크지 않은 실험결과는 본 연구의 새로운 열화 인덱스 생성 방법의 적합성이 검증됨을 설명할 수 있다.

3.2.3. LMP 예측모델 성능

OLS, GLM, BR, GPR 4개의 회귀모델과 설명변수로써 각 기하학적 열화 인덱스를 하나를 입력하는 경우, 두가지 방법에 대해 반응변수를 LMP로 하는 모델 성능을 비교하였다. LMP 예측모델은 회귀모델로서 GPR이 테스트 데이터에 대한 RMSE값이 모든 설명변수로서의 열화 인덱스에 대해 제일 낮게 도출되어 적합함을 확인하였다.

미세조직 이미지로부터 추출된 모든 열화 인덱스를 사용하여 하나의 주성분을 설명변수로 할 경우와 LMP를 설명하기에 유의한 상위 5개 열화 인덱스를 찾고 하나의 주성분으로 설명변수로 할 경우 LMP 예측 모델의 성능을 비교하였다. 라쏘 회귀를 통한 변수 중요도를 판단하여 LMP를 잘 설명하는 열화 특징 5개는 aspect ratio, angle, contour area, convexhull area, equivalent diameter임을 확인하였다. 추가적으로 12개의 열화 인덱스를 하나의 주성분으로 설명변수로 하는 것이 5개의 유의 열화 인덱스를 하나의 주성분으로 설명변수로 하는 것보다 4개의 회귀모델에서 모두 RMSE가 낮은 것으로 나왔으나, 그 변동율은 4개 모델 평균 8%로 나타났으며 적은 설명변수로 하나의 주성분을 구성시에도 성능의 변동이 크지 않음을 확인하였다.

LMP 예측 모델에서 모델의 적합도를 검증하기 위하여 학습 성능 대비 테스트 성능변화의 변동을 확인하였

다. 모든 또는 5개의 유의 설명변수를 하나의 주성분으로 하는 경우나, 각 기하학적 특징별 열화인덱스를 설명변수로하는 모든 경우에 대해서 RMSE 기준 GPR이 학습성능 대비 테스트 성능의 변동이 1% 내로 가장 적은 것을 확인할 수 있었다. LMP 예측모델 또한 과적합되지 않음을 확인하였다. 기하학적 특징별 열화도 인덱스 관점에서의 적용 회귀모델의 종류와 관계없이 학습대비 테스트 RMSE의 변동폭이 낮게 결과가 나온 열화 인덱스 중 하위 5개는 Convexhull area, Equivalent diameter, Compactness, Convexhull perimeter, Contour perimeter임을 확인하였다.

IV. 결 론

본 연구에서는 니켈기 초합금의 미세조직 이미지로 딥러닝을 활용한 빠른 특징화(Fast Characterization)를 시작으로 소재 환경조건을 입력하면 기하학적 특징으로 정량화된 열화도 인덱스를 예측하고 LMP를 예측함으로써 크리프 수명예측을 가능하게 하였다. 미세조직 이미지의 입력으로부터 열화도 및 수명예측까지 딥러닝과 기계학습을 활용한 자동화를 통해 소재 분야의 전문가에게는 분석시간과 노력을 절감시키고, 비전문가라고 할지라도 신속하고 직관적이면서 일관된 결과를 얻을 수 있게 하였다. 고온 부품을 파괴하지 않고도 기존의 고비용으로 얻어진 실험데이터를 기반으로 소재가 처한 환경조건을 입력으로 하여 열화도와 크리프 수명의 중요 파라미터를 예측할 수 있게 되면서 실험비용을 절감하면서도 환경조건에 따른 열화와 수명을 평가할 수 있게 되었다. 특히 미세조직 이미지의 γ 의 상태를 기하학적으로 설명할 수 있고, 잡신호 및 이상치에 강건하고, 저차원의 설명자로서 열화 인덱스를 구성하는 방법을 제안함으로써 기존 연구의 제한사항들을 개선하였다. 동일 환경조건에 대한 이미지들을 학습 및 테스트 데이터로 분리하여 열화 인덱스 생성 시 변동성을 측정하였고 테스트 이미지가 학습 이미지의 1/4 양으로 적음에도 불구하고 열화 인덱스의 변동량이 크지 않음을 검증함으로써 적은 이미지로도 열화 인덱스가 모수와 가까운 추론을 함을 검증할 수 있었다.

파괴시험 데이터로 구축한 열화 인덱스 데이터베이스로부터 열화 인덱스 예측모델, LMP 예측모델을 구성

하고 유기적으로 연결함으로써 파괴검사 없이도 환경 조건의 입력만으로 운영 중인 기기의 열화 및 수명을 예측 시뮬레이션을 가능하게 하였으며, 정비 시 레플리케이션 미세조직 이미지 데이터를 취득 시에도 열화도와 수명을 예측할 수 있게 되었다. 특히 여러 종류의 기하학적 특징에 기반한 열화 인덱스를 활용한 예측모델을 구성함으로써 γ '의 특징이 열화 및 수명의 영향을 미치는 정도를 입체적으로 해석할 수 있도록 직관적이면서도 다양한 관점을 제시하였다. 본 연구에서 제안한 예측 모델은 가스터빈과 같은 기기의 고온 부품의 유지, 보수, 폐기의 의사결정 모델 개발을 가능하게 하는 기반 기술로 사용될 것을 기대한다.

ACKNOWLEDGEMENT

This work was supported by the Doosan Enerbility and the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2022R1F1A1074696).

REFERENCES

[1] A. O'Hagan, "Bayesian statistics: principles and benefits," *Frontis*, pp. 44, May. 2004.

[2] B. Rezaeianjouybari and Y. Shang, "Deep learning for prognostics and health management: State of the art, challenges, and opportunities," *Measurement*, vol. 163, pp. 107929, Oct. 2020. DOI: 10.1016/j.measurement.2020.107929.

[3] G. Marahleh, A. R. I. Kheder, and H. F. Hamad, "Creep-life prediction of service-exposed turbine blades," *Materials Science and Engineering*, vol. 42, no. 4, pp. 476-481, Jul. 2006. DOI: 10.1007/s11003-006-0103-8.

[4] F. R. Larson and J. Miller, "Closure to "Discussions of 'A Time-Temperature Relationship for Rupture and Creep Stresses,'" *Transactions of the American Society of Mechanical Engineers*, vol. 74, no. 5, pp. 774-775, Jul. 1952. DOI: 10.1115/1.4015915.

[5] American Society for Testing and Materials std. E139-11, *Standard test methods for conducting creep, creep-rupture, and stress-rupture tests of metallic materials*, ASTM International, 2006.

[6] R. L. Graham and F. F. Yao, "Finding the convex hull of a simple polygon," *Journal of Algorithms*, vol. 4, no. 4, pp. 324-331, Dec. 1983. DOI: 10.1016/0196-6774(83)90013-5.

[7] University of Guelph CIS 6320 (2001). Shape analysis and measurement [Internet]. Available: <http://www.cyto.purdue.edu/cdroms/micro2/content/education/wirth10.pdf>.

[8] Dept. of Brain and Cognitive Sciences, Univ. of Rochester, Rochester, NY(2012). Bayesian Inference: Metropolis-Hastings Sampling [Internet]. Available: https://www2.bcs.rochester.edu/sites/jacobslab/cheat_sheet/MetropolisHastingsSampling.pdf.

[9] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of Mathematical Psychology*, vol. 85, pp. 1-16, Aug. 2018. DOI: 10.1016/j.jmp.2018.03.001.

[10] M. Ebden. (2015, August), Gaussian processes: A quick introduction. arXiv preprint arXiv:1505.02965, [Online]. Available: <https://arxiv.org/pdf/1505.02965.pdf>.



유준상(Junsang Yu)

성균관대학교 데이터사이언스융합학과
두산에너지빌리티 데이터분석팀
*관심분야: 베이지안 추론, 통계적 데이터분석, 컴퓨터 비전



오하영(Hayoung Oh)

성균관대학교 소프트웨어융합대학 부교수
*관심분야: 소셜정보망 분석, 추천시스템, 데이터분석 및 인공지능