

Regular Article

Federated Learning of Creep Rupture Time and High Temperature Tensile Strength Prediction Models

Junya SAKURAI¹⁾ , Keisuke TORIGATA²⁾, Manabu MATSUNAGA³⁾, Naoto TAKANASHI⁴⁾ , Shinya HIBINO⁴⁾ , Kenichi KIZU⁵⁾, Akira MORITA⁵⁾, Masahiro INOMOTO⁶⁾, Nobuaki SHIMOHATA⁷⁾, Kodai TOYOTA⁸⁾, Tadaaki NAKAMURA⁹⁾, Keita HASHIMOTO⁹⁾, Tatsuya OKUBO¹⁰⁾, Loic BEHESHTI¹⁰⁾, Vincent RICHARD¹⁰⁾ and Masahiko DEMURA^{1)*} 

1) Research Network and Facility Services Division, National Institute for Materials Science

2) Materials & Structural Dept., Technology Platform Center, Corporate Research and Development Division, IHI Corporation

3) Engineering G., Aero Engine, Space & Defense Business Area Rocket Development Dept., IHI Corporation

4) Kawasaki Heavy Industries, Ltd.

5) Maintenance Coordination Group, Operations Department, Thermal Power Division, The Kansai Electric Power Co., Inc.

6) Materials Research Laboratory, Technical Development Group, Kobe Steel, Ltd.

7) Electric Power Development Co., Ltd.

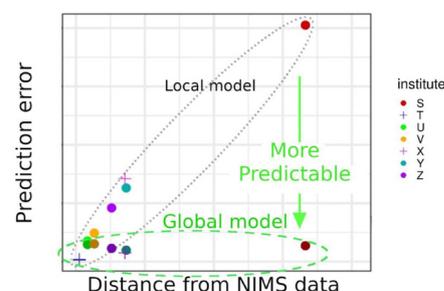
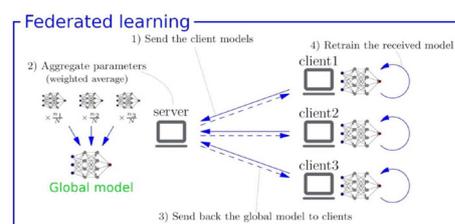
8) Japan Atomic Energy Agency

9) Research & Innovation Center, Mitsubishi Heavy Industries, Ltd.

10) Elix, Inc.

Abstract: Creep testing is time-consuming and costly, leading institutions to limit the number of tests conducted to the minimum necessary for their specific objectives. By pooling data from each institution, it is anticipated that predictive models can be developed for a wide range of materials, including welded joints and degraded materials exposed to service conditions. However, the data obtained by each institution is often highly confidential, making it challenging to share with others. Federated learning, a type of privacy-preserving computation technology, allows for learning while keeping data confidential. Utilizing this approach, it is possible to develop creep life prediction models by leveraging data from various institutions. In this paper, we constructed global deep neural network models for predicting the creep rupture life of heat-resistant ferritic steels in collaboration with eight institutions using the federated learning system we developed for this purpose. Each institution built a local model using only its own data for comparison. While these local models demonstrated good predictive accuracy for their respective datasets, their predictive performance declined when applied to data from other institutions. In contrast, the global model constructed using federated learning showed reasonably good predictive performance across all institutions. The distance between each institution's data was defined in the space of explanatory variables, with the NIMS data, which had the largest dataset, serving as the reference point. The global model maintained high predictive accuracy regardless of the distance from the NIMS data, whereas the predictive accuracy of the NIMS local model significantly decreased as the distance increased.

Keywords: ferritic heat resistant steel; federated learning; creep rupture time; high temperature tensile strength; machine learning; prediction.



Received November 20, 2024; Accepted January 30, 2025; Advance online published February 6, 2025; Published April 1, 2025

* Corresponding author. E-mail: DEMURA.Masahiko@nims.go.jp, Address: National Institute for Materials Science, 1-1 Namiki Tsukuba Ibaraki 305-0044



© 2025 The Iron and Steel Institute of Japan. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

クリープ破断時間および高温引張強度 予測モデルの連合学習

櫻井 惇也¹⁾ ・鳥形 啓輔²⁾・松永 学³⁾・高梨 直人⁴⁾ ・日比野 真也⁴⁾ ・木津 健一⁵⁾・
森田 聡⁵⁾・井元 雅弘⁶⁾・下畠 伸朗⁷⁾・豊田 晃大⁸⁾・中村 忠暉⁹⁾・橋本 憩太⁹⁾・
大久保 達矢¹⁰⁾・ベヘシティ ロイック¹⁰⁾・リチャル ヴァンサン¹⁰⁾・出村 雅彦¹⁾ * 

Federated Learning of Creep Rupture Time and High Temperature Tensile Strength Prediction Models

Junya SAKURAI, Keisuke TORIGATA, Manabu MATSUNAGA, Naoto TAKANASHI, Shinya HIBINO, Kenichi KIZU, Akira MORITA,
Masahiro INOMOTO, Nobuaki SHIMOHATA, Kodai TOYOTA, Tadaaki NAKAMURA, Keita HASHIMOTO,
Tatsuya OKUBO, Loic BEHESHTI, Vincent RICHARD and Masahiko DEMURA

1. 緒言

クリープ試験の実施には非常にコストと時間がかかる。そのため、すでに取得済みのデータを活用することが重要となる。例えば、我々はNIMSが長年に亘って蓄積してきたクリープデータシート¹⁾のデータ(以下、「NIMSデータ」という。)に基づいてフェライト耐熱鋼のクリープ寿命予測を試みている²⁾。その結果、機械学習を活用することで、クリープデータシートが対象とする規格鋼に対して十分に精度の高い予測が可能であることを示してきた。しかし、機械学習による予測は学習データの制約を強く受けるので、学習データとかけ離れたケースや材料に対する予測は難しい。例えば、実プラントで使用している材料の余寿命予測や溶接材料などの特殊な鋼種に対する寿命予測は、実用上、重要であるものの、NIMSデータだけでは予測が困難である。

NIMS以外の機関においてもそれぞれの目的でクリープ試験データを取得しており、これらを合わせて活用できれば、さらに幅の広い対象に対して適用可能な予測モデルを構築できると期待される。特に、企業においては実用上重要なケースについて試験データを有していることが期待さ

れる。これはNIMS等の公的機関が基盤的なデータを保有していることと補完的な関係にあると言える。しかし、各機関がデータを互いに開示しあうことは難しい。特に、企業データは競争力の源泉であり一般に開示が難しいことに加え、サプライチェーンの中で複数のステークホルダーが当該データの管理に関わることになる難しさも存在する。実際のデータの取り扱いの観点でも、中央サーバに各々のデータを集約して学習する方法はデータ管理上、様々なリスクを抱えることになる。

データを互いに開示することなく統合的に学習する方法として、近年、連合学習³⁾という手法が提案されている。この手法を用いることでデータの秘匿性を保ったまま協調して学習できる。連合学習はデータ保有者各自の元で学習したモデルのみを中央のサーバに送ってモデルを統合するものであり、データを集約する必要がない。各機関のデータ数のアンバランスや統合されたモデルの取り扱いなどを検討する必要はあるものの、データを開示することなく、かつ、データ管理上のリスクも抱えることなく実施できることから合意が得られやすい方法であると思われる。

本論文では、合意が得られた8機関が共同して実施した連合学習によるフェライト耐熱鋼のクリープ寿命予測モ

2024年11月20日受付 2025年1月30日受理 2025年2月6日早期公開 2025年4月1日発行 (Received November 20, 2024; Accepted January 30, 2025; Advance online published February 6, 2025; Published April 1, 2025)

- 1) 国立研究法人物質・材料研究機構技術開発・共用部門 (Research Network and Facility Services Division, National Institute for Materials Science)
- 2) (株) IHI 技術開発本部技術基盤センター 材料・構造技術部 (Materials & Structural Dept., Technology Platform Center, Corporate Research and Development Division, IHI Corporation)
- 3) (株) IHI 航空・宇宙・防衛事業領域 ロケット開発事業推進部技術グループ (Engineering G., Aero Engine, Space & Defense Business Area Rocket Development Dept., IHI Corporation)
- 4) 川崎重工業 (株) 技術開発本部技術研究所 (Kawasaki Heavy Industries, Ltd.)
- 5) 関西電力 (株) 火力事業本部火力運営部門保全グループ (Maintenance Coordination Group, Operations Department, Thermal Power Division, The Kansai Electric Power Co., Inc.)
- 6) (株) 神戸製鋼所技術開発本部材料研究所 (Materials Research Laboratory, Technical Development Group, Kobe Steel, Ltd.)
- 7) 電源開発 (株) 技術開発部若松研究所 (Electric Power Development Co., Ltd.)
- 8) 国立研究開発法人日本原子力研究開発機構大洗原子力工学研究所 (Japan Atomic Energy Agency)
- 9) 三菱重工業 (株) 総合研究所 (Research & Innovation Center, Mitsubishi Heavy Industries, Ltd.)
- 10) (株) Elix (Elix, Inc.)

* Corresponding author. E-mail: DEMURA.Masahiko@nims.go.jp, Address: National Institute for Materials Science, 1-1 Namiki Tsukuba Ibaraki 305-0044



© 2025 The Iron and Steel Institute of Japan. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives license (<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.ja>).

デル構築の結果を報告する。データを秘匿するという目的から各機関データに含まれる鋼種を明確にはできないが、NIMSデータについては既報の機械学習によるクリープ寿命予測の研究²⁾で使用したものと同一データを用いており、炭素鋼 (STB410) から高Cr鋼 12Cr-2W-0.4Mo-1Cu-Nb-V (KA-SUS 410J3 DTB) まで広範である。連合学習のためのシステムは連合学習機能を有したAI創薬向け機械学習ライブラリ kMoL⁴⁾ をベースにNIMSが今回の目的のために独自に開発したものであり、オープンソースとして公開する (<https://github.com/nims-federated-learning/NIMS-FL>)。連合学習は2016年に提案されているものの、材料分野に適用した報告はわずかである。具体的には、材料データに連合学習手法を適用するコンセプトの提案⁵⁾、放射光施設の μ XRDの画像スクリーニングへの応用⁶⁾、組織写真のセグメンテーションに関する連合学習適用可能性の検証^{†1)}が存在するが、いずれも連合学習の可能性を検証する研究である。実際に複数機関が共同してデータを秘匿した状態で実施した例は、筆者の知る限り、本報告が初めてである。本研究では比較のために高温引張強度についても予測対象とした。

2. 連合学習システムの開発

2.1 連合学習

連合学習をFig.1に図説する。連合学習システムの構成はモデル統合用のサーバと連合学習に参加する各機関のクライアント端末からなり、サーバ-クライアント間の通信がある。連合学習では次のようなステップでグローバルモデルを作成し一連のサイクルを繰り返して更新する。まずクライアントは自身が保有する訓練データを用いてモデルを学習する。学習後に各クライアントはサーバへモデルを送り、サーバは各モデルのパラメータを統合したグローバル

ルモデルを作成する。そしてサーバは各クライアントにグローバルモデルを送り、各クライアントはグローバルモデルを初期パラメータとして自身が保有する訓練データにより学習を進める。クライアントとサーバ間でこのような学習と統合を繰り返すことによりグローバルモデルが更新されていく。

このように連合学習ではクライアントからサーバへデータを送ることなく各クライアントによる学習を反映したグローバルモデルを構築できる。

連合学習におけるパラメータ統合方法には複数の手法があるが、本プログラムでは以下4つから選択できる。

- Plain aggregation : データ数比によらず各クライアントを等価に扱う
 - Weighted aggregation : 各クライアントのデータ数比に応じて重み付けする
 - Benchmark aggregation : クライアントのテストセットの評価スコアに応じて重み付けする
 - Fedprox aggregation : 他の統合方法と組み合わせ、各roundの初期モデルからの乖離を防ぐ正則化を加える
- 本研究では、各機関のデータ数にアンバランスがあることを考慮して、Weighted aggregationを選択した。

2.2 実行環境

連合学習システムを動作させるための環境と必要なライブラリをTable 1に示す。

- OS : Linux
- python-3.8

連合学習システムはpythonのパッケージとなっており、インストール時に上記依存ライブラリもインストールされる。連合学習を行うには統合用サーバおよびクライアント用端末それぞれに連合学習システムをインストールする必要がある。

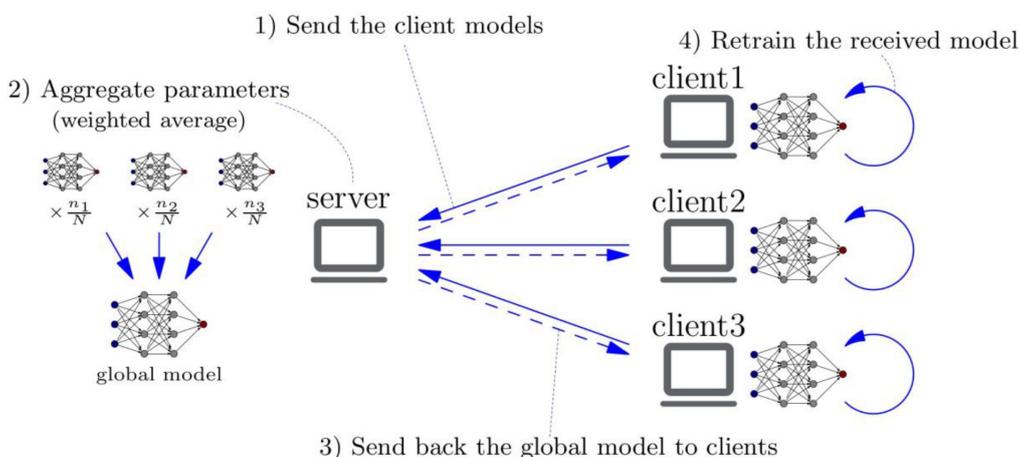


Fig. 1. Schematic of federated learning. (Online version in color.)

†1 B.Ma, X.Yin, X.Ban, H.Huang, N.Zhang, H.Wang and W.Xue, <https://doi.org/10.48550/arXiv.2111.07892>

2・3 プログラム仕様

2種類のモードを備え、通常の学習と連合学習を行うことができる。

通常の学習では通信を介さず単独でモデルを学習する。実施時にはファイルから読み込まれたデータを指定割合で訓練セット・テストセットへと分割し、このうち訓練セットが学習に使われる。機械学習モデルは多層パーセプトロンで構成されるディープニューラルネットワークモデルを用いる。モデル構造を、入出力ノード数や使用する変数、パラメタのノード数や層数などを指定して定義する。学習はミニバッチ学習で行われ、訓練データから指定したバッチサイズのサブセットを取り出してモデルの学習を進める。訓練データすべてが学習に現れるサイクルを1 epochとし、指定したepoch数だけ学習を繰り返す。学習は学習データに対する説明性を高めるようにパラメタが最適化されていく。最適化手法や学習率の初期値や最大値、学習の進行に応じた学習率の調整手法などによって学習の効率や収束性などが異なる。学習が終了するとテストセットを用いて予測精度を評価する。モデルの構成、説明変数の指定、テストデータと訓練データの比率、学習を調整する各種変数などはモデルコンフィグファイルで指定する。詳細には、Table 2に示すような項目がモデルコンフィグに記述される。

連合学習では統合サーバとクライアントがそれぞれ協調し、通信しながら学習を進めていく。サーバ-クライアント間の通信は Secure Sockets Layer (SSL) プロトコルにより暗号化している。実施開始後、サーバは待機し、その

Table 1. Required libraries and versions.

cairosvg (==2.5.2)
grpcio (==1.31.0)
grpcio-tools (==1.31.0)
mordred (==1.2.0)
numpy (==1.19.1)
opacus (==0.11.0)
optuna (==2.10.0)
pandas (==1.3.2)
pyhumps (==3.0.2)
rdkit-pypi (==2021.3.5.1)
scikit-learn (==0.24.2)
torch (==1.8.2+cpu)
torch-cluster (==1.5.9)
torch-geometric (==1.6.0)
torch-scatter (==2.0.7)
torch-sparse (==0.6.12)
torch-spline-conv (==1.2.1)
torchvision (==0.9.2+cpu)

間にクライアントが接続することで、連合学習に参加するクライアントが確定する。連合学習の概要は2・1で述べたとおりである。初めにサーバからクライアントへモデルコンフィグファイルが送られ、各クライアントは通常の学習と同様に自らのデータの訓練セットを用いて指定されたepoch数の学習を進め、最終epochのモデルをサーバへと送る。サーバは指定された統合方法により各クライアントのモデルのパラメタを統合したグローバルモデルを作成し、グローバルモデルおよび更新したモデルコンフィグファイルをクライアントへ送る。このサイクルをroundと呼ぶ。グローバルモデルのパラメタ更新を指定round数、繰り返すことによって連合学習が完了する。サーバおよびクライアントの起動時にサーバコンフィグファイルおよびクライアントコンフィグファイルを与えることで上述の連合学習に関わる設定や使用するモデルコンフィグファイルの指定などを行う。Table 3に指定する項目の一部を示す。

2・4 NIMSデータを用いた性能検証

開発した連合学習システムについてNIMSデータを用いて種々の性能検証を行った。性能検証のための実験は、クライアント数、クライアントのデータ数間のバランス、統合手法の選択の3つの観点で行った。

検証は10-fold交差検証で行った。交差検証ではfoldごとに重ならないようにテストセットを選り分けておく。各

Table 2. Example of items in model configuration.

Key	Description
model	Model Structure
loader	How to load data file and specify input/output variables.
featureizer	features to use
splitter	How to split training/test data
criterion	Loss function
optimizer	Optimization method
scheduler	Way to tune learning rate
epochs	Number of epochs
batch_size	Batch size
test_metrics	Evaluation criterion for test data

Table 3. Contents of server config file and client config file.

Server config	Client config	Key	Description
	○	task_configuration_file	Model config file
	○	aggregator_type	Model aggregation method
	○	aggregator_options	Options for model aggregation. (e.g. weights)
	○	rounds_count	Number of rounds
	○ ○	ssl_private_key	Path of private key for SSL
	○ ○	target	Server IP address

foldにおいて訓練セットを擬似的に分割してクライアントの訓練セットとする。その上で、連合学習を行い、得られたグローバルモデルに対してテストセットを用いて予測性能を評価する。具体的にはRMSEを評価のスコアとして用いた。交差検証のための分割およびクライアント訓練セットの分割にはランダム分割を採用しシードを固定して行った。

この検証ではTable 4に示す入力変数を用い、前処理として平均=0、標準偏差=1となるよう正規化を施して用いた。出力変数は対数破断時間でありこれも同様の方法で正規化して用いた。

モデルは全結合型の多層パーセプトロンで構成されるディープニューラルネットワーク (DNN) を用いた。詳細には、5つの隠れ層がそれぞれ512のノードをもつDNNである。最適化はAdamW^{†2}で行い、MSE損失を用い、重み減衰係数は0.00056を指定した。過学習を抑制しつつ効率的に学習を進めるためにスケジューラによって学習率を調整した。具体的にはOneCycleLRスケジューラを用い、そのパラメタはpytorchのデフォルト設定を用いた。デフォルト値pct_start=0.3でありサイクルの30%で学習率の上昇が行われその後減少に転じる挙動で調整が行われる。round数は50、各roundにおけるepoch数は10とした。

Table 4. Input variables used when validating federated learning.

C [mass%]
Si [mass%]
P [mass%]
S [mass%]
N [mass%]
B [mass%]
Mn [mass%]
Ni [mass%]
Cr [mass%]
Mo [mass%]
Cu [mass%]
Al [mass%]
Ti [mass%]
Nb+Ta [mass%]
V [mass%]
W [mass%]
Co [mass%]
Fe [mass%]
Nb [mass%]
Test temperature [K]
Test stress [MPa]
0.2% proof stress [MPa]

2・4・3では異なる統合手法間で比較を行い、2・4・1および2・4・2に述べる2つの実験についてはFedAvgを使用しデータ数に応じて統合した。

ベンチマークとして、クライアントに擬似的分割せず連合学習で用いるDNNと同じモデルで、連合学習のトータルepoch数と同じ500epochの訓練で予測モデルを構築した。得られたモデルの予測精度は、各foldのRMSEの平均±標準偏差が 0.1871 ± 0.0096 であった。これは約1.5 (= $10^{0.1871}$)のfactorの誤差に該当するもので、先行研究²⁾で得られた機械学習モデルの性能に匹敵する。

2・4・1 クライアント数に関する実験

クライアント数を2-10まで変更して連合学習の精度を比較する。各クライアントが均一なデータ数を持つよう配分する条件でクライアント数を変更した。Table 5に、各分割ケースで得られた連合学習の結果(各foldのRMSEの平均±標準偏差)を示す。クライアント数が2の場合ではベンチマークと遜色ない結果が得られており、連合学習が十分に機能していることがわかる。クライアント数が多いほど性能が落ち、今回の機関数である8ではスコアは 0.231 ± 0.0151 であった。

2・4・2 クライアントのデータ数間のバランスに関する実験

各クライアントが不均衡なデータ数を有している種々のケースを想定して比較する。ここではデータ数の割合のほかクライアント数も変化させて実験する。Table 6に示す各条件の表記方法は例として30-30-15-15-10の場合は5つのクライアントのうち2つがそれぞれデータ数全体のうち30%を占め、2つが15%を占め、1つが10%を占めている場合を表す。統合手法にPlain aggregationを用いて単純にクライアントのパラメタを平均すると、2クライアント、3クライアント、5クライアントのいずれの条件群でも、データ数の不均衡が大きい場合に性能が落ちる傾向が見られた(Table 6)。

Table 5. Relationship between number of clients and prediction accuracy.

Number of clients	score
2	0.184 ± 0.0216
3	0.2 ± 0.0208
4	0.203 ± 0.0186
5	0.215 ± 0.0221
6	0.218 ± 0.0192
7	0.225 ± 0.0196
8	0.231 ± 0.0151
9	0.239 ± 0.0218
10	0.261 ± 0.0148

†2 I.Loshchilov and F.Hutter, <https://doi.org/10.48550/arXiv.1711.05101>

Table 6. Prediction accuracy depending on the balance of the number of data among clients and the aggregation method.

Balance	Plain aggregator	Weighted aggregator
60-40	0.182 ± 0.0205	0.177 ± 0.0205
67-33	0.191 ± 0.0198	0.175 ± 0.0195
75-25	0.2 ± 0.0206	0.178 ± 0.0165
80-20	0.202 ± 0.0246	0.179 ± 0.0174
40-30-30	0.192 ± 0.0209	0.185 ± 0.0197
60-20-20	0.2 ± 0.0181	0.18 ± 0.017
60-30-10	0.2 ± 0.0203	0.18 ± 0.0186
80-10-10	0.209 ± 0.0249	0.181 ± 0.0172
30-30-15-15-10	0.224 ± 0.0231	0.203 ± 0.0192
60-10-10-10-10	0.233 ± 0.0149	0.19 ± 0.0183

2・4・3 統合手法に関する実験

2・1で使用可能な統合手法には複数の選択肢があることを示したが、それらのうちPlain aggregationとWeighted aggregationの間で比較する。Plain aggregationは各クライアントのモデルのパラメタを単純に平均する。他方、Weighted aggregationは各クライアントのモデルのパラメタをクライアントのデータ数に応じた重みを付けて平均する。クライアント間でデータ数の割合が等しい場合には両者の違いはないため、ここではデータ数が不均衡な条件について比較する。2・4・2でTable 6に示したPlain aggregationと同様の各条件についてWeighted aggregationの実験を行った結果を同表に示した。Plain aggregationに比べ、データ数が不均衡な条件でもWeighted aggregationでは良い結果が得られた。特に、クライアント数が3, 4においても、ベンチマークに匹敵する性能が得られている点は、特筆すべきである。

3. 複数機関による連合学習の実施方法

3・1 データフォーマット

連合学習を行う際には、データ形式を定めそれに従ってデータを準備する必要がある。しかし、各機関は様々な点で異なるデータを有しており、最終的な共通の入力フォーマットに至るまでに種々の擦り合わせを要した。擦り合わせ前の試作フォーマット時点では、クリープ破断試験に関するデータフォーマットに関して先行研究²⁾で使用したものと同一変数を設けた。高温引張試験に関するデータフォーマットについてはクリープ破断試験と同じ組成項目、引張試験温度、試験温度における0.2%耐力に加えて引張強度と伸びを設けた。

各機関のデータに存在する変数のセットはまちまちである。そのため、データフォーマットに設ける変数について実測値が不在である場合があり、その取り扱いを検討す

る必要がある。不在の場合にその値を空とするか何らかの値を埋めるかについて、以下のような点を複合的に検討した。具体的には、学習に利用可能なデータ数、プログラム上で空の扱いが通常の数値と異なる点、特定の値を除く等の前処理に利用できる可能性、未測定の対象に特定の値を仮定することが合理的か否かといった点である。検討の結果、0.2%耐力についてはデータ数が想定より少なく必須とすると学習データが不足するため実測値がない場合には空とした。これについては後述する。次に、実測値がない元素の組成については0を仮定して入力するものとした。ただし、Fe [mass%]については今回フェライト鋼に限定しているために組成が記録されていない場合が多いとみられることから、100- (他の元素組成の合計値) により値を与えた。また、先行研究²⁾にはない幾つかの元素 (O [mass%], Ru [mass%], Re [mass%]) を追加した。また、溶接金属の酸素については実測値がない場合、溶接施工法に応じた経験的な推定値を与え、別に推定値の列 (O_est [mass%]) を設けた。

筆者らの先行研究²⁾では、材料情報として化学組成のみを用いていた。しかし、化学組成が同じであってもプロセスや使用履歴が異なるとマイクロ組織が影響を受け、クリープ寿命は当然、変化する。したがって、化学組成だけでは不足する。先行研究では規格鋼のまだ使用に供されていない新材を対象としていたため化学組成だけで十分であったが、今回のようにプラントで使用された履歴のある材料が含まれる場合には化学組成だけで予測することは困難と考えられる。実際、先行研究²⁾においても、化学組成が同一で熱処理のみが異なるCDS53の2ヒートに対しては、化学組成のみでは予測は困難であることを報告している。さらに、このような場合でも、試験温度における0.2%耐力を説明変数に加えることにより予測が改善できることを見出している。熱処理の違いを吸収する予測改善は0.2%耐力がマイクロ組織の変化を間接的に表現したことによってもたらされたと解釈できる。

本研究でも、当初、同様に0.2%耐力を用いて、複数機関の様々なマイクロ組織を有するデータが混在する状況での予測に役立てることを予定していた。しかし、想定より0.2%耐力のデータが少なかったことから、代替としてよりデータの多い室温ビッカース硬さが利用可能かを検討した。予備的にNIMSクリープデータを用いて0.2%耐力の代わりに室温ビッカース硬さを加えて学習したところ、上述のCDS53の2ヒートに関して、0.2%耐力を使用したモデルと同程度に予測可能となることを確認できた。このため、本研究のクリープ破断時間予測モデルの学習には室温ビッカース硬さを材料情報の説明変数として加えた。

各機関のデータには規格鋼だけでなく、溶接金属、溶接のワイヤーをdeposited metalとした試験片、再現HAZ材、使用材 (抜管材など) が含まれる可能性があった。これら

種類ごとの部分データで学習する可能性や材料種の情報を入力としても用いる可能性を考慮して、材料種の情報を組み込むこととした。具体的には、材料種ごとの変数を用意して、該当する場合は1を該当しない場合は0を与えるOne-hot表現を採用した。

以上の擦り合わせの結果得られたクリープ破断データ

用のフォーマットおよび高温引張試験データ用のデータフォーマットに関して、それぞれに含まれる変数をTable 7に示す。後述のモデル作成条件それぞれで使われる入力変数 (in) および目的変数 (out) も同表中に示した。

3・2 前処理

データを学習に用いるにあたり変数のスケール変換を

Table 7. Variables included in the input data format for creep rupture and high temperature tensile testdata and the handling of each value and the variables used in each modeling condition.

Variable name	Type	If inexist	Specific handling	creep_std creep_weld creep_used	creep _all	tensile_std tensile_weld tensile_used	tensile _all
C [mass%]	Numeric	Fill 0		in	in	in	in
Si [mass%]	Numeric	Fill 0		in	in	in	in
P [mass%]	Numeric	Fill 0		in	in	in	in
S [mass%]	Numeric	Fill 0		in	in	in	in
N [mass%]	Numeric	Fill 0		in	in	in	in
B [mass%]	Numeric	Fill 0		in	in	in	in
Mn [mass%]	Numeric	Fill 0		in	in	in	in
Ni [mass%]	Numeric	Fill 0		in	in	in	in
Cr [mass%]	Numeric	Fill 0		in	in	in	in
Mo [mass%]	Numeric	Fill 0		in	in	in	in
Cu [mass%]	Numeric	Fill 0		in	in	in	in
Al [mass%]	Numeric	Fill 0		in	in	in	in
Ti [mass%]	Numeric	Fill 0		in	in	in	in
Nb+Ta [mass%]	Numeric	Fill 0	Input only one of Nb+Ta [mass%] of Nb [mass%]	in	in	in	in
V [mass%]	Numeric	Fill 0		in	in	in	in
W [mass%]	Numeric	Fill 0		in	in	in	in
Co [mass%]	Numeric	Fill 0		in	in	in	in
Fe [mass%]	Numeric		Calculate 100 - (Sum of other compositions)	in	in	in	in
Nb [mass%]	Numeric	Fill 0	Input only one of Nb+Ta [mass%] of Nb [mass%]	in	in	in	in
Sn [mass%]	Numeric	Fill 0		in	in	in	in
O [mass%]	Numeric	Fill 0		in	in	in	in
Ru [mass%]	Numeric	Fill 0				in	in
Re [mass%]	Numeric	Fill 0		in	in	in	in
O_est [mass%]	Numeric	Empty or Estimated value					
Vickers hardness at RT [HV]	Numeric	Empty	Allow conversion from other hardness	in	in	in	in
Test temperature [K]	Numeric	(Neccesary)	Absolute temperature $t + 273.15$ K for t °C.	in	in	in	in
Test stress [MPa]	Numeric	(Neccesary)		in	in		
0.2% proof stress [MPa]	Numeric	Empty					
Time to rupture [h]	Numeric	(Neccesary)					
log10 timetorupture [h]	Numeric	(Neccesary)		out	out		
Tensile strength [MPa]	Numeric	(Neccesary)				out	out
Elongation [%]	Numeric	(Neccesary)					
Standard material (matrix)	0 or 1		One-hot (one of them is 1, 0 for others)		in		in
Weld metal	0 or 1		One-hot (one of them is 1, 0 for others)		in		in
Deposited metal	0 or 1		One-hot (one of them is 1, 0 for others)				
Simulated HAZ	0 or 1		One-hot (one of them is 1, 0 for others)				
Used sample	0 or 1		One-hot (one of them is 1, 0 for others)		in		in
Inner pressure creep test	0 or 1						
Hoop stress [MPa]	Numeric	Empty	Input only if Inner pressure creep test is 1				

行った。まず、各機関はそれぞれ前述のデータフォーマットに沿ってデータを準備する。学習時には各変数に対してスケールの変換を施して用いることで分布の偏りがニューラルネットワークによる学習の妨げとならないようにする。データ間で変換方法は同一でなければいけないが、データは秘匿されているので全機関のデータに関する平均と標準偏差を取り出して正規化に用いることはできない。ここでは、厳密な正規化を行う代わりに、殆どの変数に関してはデータ数の大きいNIMSデータの変数の平均と標準偏差を代用することとした。NIMSデータに値が存在しない一部の変数については当該データを有する各機関の平均と標準偏差の平均を用いた。このように、データは秘匿しているが、場合により平均と標準偏差の情報は開示する必要が生じる点は留意する必要がある。

3.3 学習

3.3.1 モデルの構造

モデルは連合学習システムの性能評価で用いたのと同様、全結合多層パーセプトロンによるDNNモデルを用いる。入力変数の数 n および出力変数の内容は後述のモデル作成条件に応じて異なるが、共通して隠れ層のノード数を512、出力変数を1とする5層の全結合層とした (Fig.2)。各全結合層のパラメタの数はそれぞれ、 $n \times 512$, 512×512 , 512×512 , 512×512 , 512×1 となっている。

3.3.2 連合学習のプロセス

連合学習との比較のために各機関がサーバとやり取りせずに自機関のデータのみで行う通常の学習 (個別学習) によるローカルモデル作成も行った。連合学習および個別学習の処理を表すフローをFig.3 (a), (b) に示す。いずれも

学習前に3・2で述べた前処理を元データに対して行い、後述のモデル作成条件に応じたデータを作成する。予測精度評価にホールドアウト検証を用いるため、この時点で8割の訓練データをランダムで抽出し学習に使い、2割のテストデータを予測精度評価用に残した。

学習はミニバッチ学習であり、データ全体はバッチサイズごとに分けられ各イテレーションで重み更新に使われる。データ全体が学習に使われるまでの全イテレーションが1 epochを成し、各epochのモデルが保存される。連合学習の場合はまず指定epoch数だけ各機関でモデルの学習を進めた後に最終epochのモデルをサーバに送る。受け取っ

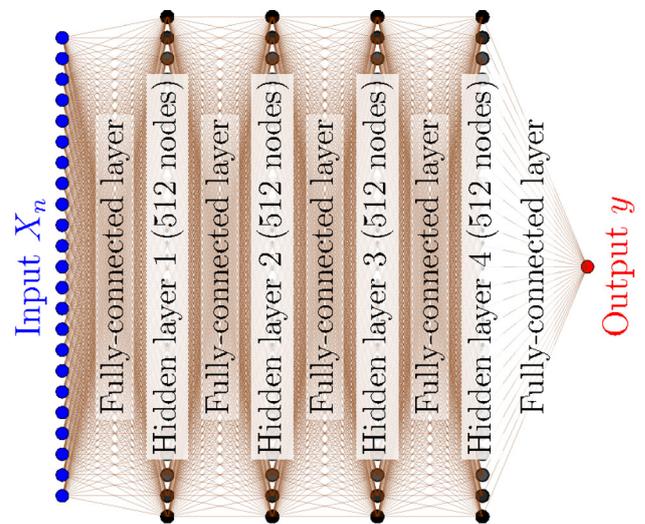


Fig. 2. Structure of the model consisting of 5 fully connected layers used for training. (Online version in color.)

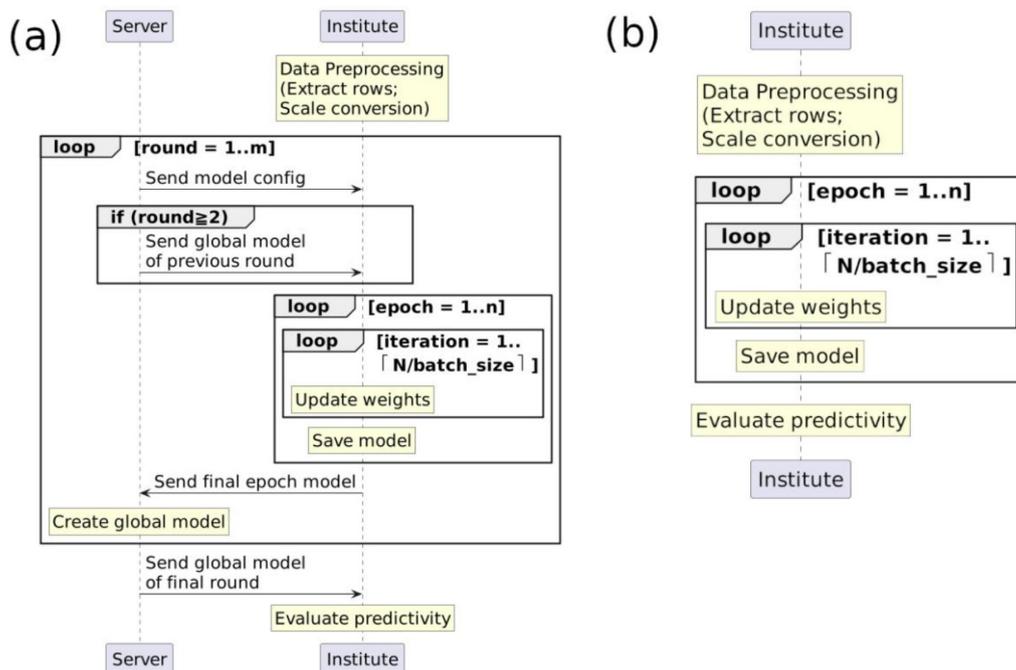


Fig. 3. Flowchart for federated and individual learning. (a) Federated learning; (b) Individual learning. (Online version in color.)

たサーバは各モデルのパラメタを予め設定した訓練データで加重平均したグローバルモデルを作成する (Federated Averaging (FedAvg)³⁾)。次のroundは各機関は受け取ったグローバルモデルを元に学習を継続し、このような各機関による学習とサーバ上でのモデル統合のサイクルが指定round数だけ繰り返される。最終roundで作成されたグローバルモデルを予測精度評価に使用した。個別学習の場合は指定epoch数だけモデルの学習が進み、最終epochで作成されたモデルが予測精度評価に使用される。

予測精度の評価はグローバルモデル、各機関自身のローカルモデルおよびNIMSローカルモデルに対して行なった。それぞれのモデルについて各機関のデータおよびNIMSデータを用いて評価し、種々の組み合わせで結果を検討した。

学習の詳細は以下の通りである。過適合を防ぐ効果があるドロップアウトについては割合を0.1に設定した。すなわち10%のノードを毎回ランダムに選出し学習に参加させない。パラメタの重みの最適化にはAdamW¹²⁾を使用し、学習率のスケジューリングにはOneCycleLR¹³⁾を用いた。このスケジューラを用いて、初期の学習率 $4e-4$ から一旦アニーリングによりepoch全体の30%をかけて最大の学習率 $1e-2$ まで増加させ、そこから最小の学習率 $4e-7$ へと減少するよう調整している。

予測精度評価の指標にはRMSEを用い、スケール変換されたままの目的変数に関して計算した。

3・4 モデル作成条件

Table 8に示すクリープ/高温引張と試験片種類と連合学習/個別学習の組み合わせからなる複数のモデル作成条件そ

れぞれに関してモデルの作成を行う。連合学習と個別学習で対応したモデル作成条件を設けるがepoch数のみが異なり、連合学習ではround数とepoch数の積が個別学習のepoch数と同数になるよう設定している。モデル作成条件にはクリープ/高温引張の別、連合学習/個別学習の別、入力変数の組、目的変数、バッチサイズ、epoch数、行抽出条件が含まれる。ただし学習可能であるかは条件に該当するデータがあるか否かによる。加えて、連合学習に参加する機関も、当該条件のデータを有するかに依存する。

Table 8に各モデル作成条件の内容を示す。行抽出条件は $3 \cdot 1$ で述べた材料種を表すone hot表現に対応したものとになっている。batch_size, epochs, roundsに関しては $3 \cdot 3 \cdot 2$ で述べた。in_featuresは入力変数の数であり、各条件における変数の詳細はTable 7に示した。学習の実行可能性は行抽出条件とデータに依存し、今回各モデル作成条件で個別学習を実施あるいは連合学習に参加した機関数を示した。

材料種として規格鋼 (Standard material (matrix)), 溶接金属 (Weld metal), 使用材 (Used sample) があり、それぞれone hot表現の変数群とした。これらは次の2通りの方法で利用した。creep_std, creep_weld, creep_used, tensile_std, tensile_weld, tensile_usedでは名称末尾で示される材料種ごとにデータを絞り特化したモデルを作るために使用し、creep_all, tensile_allでは説明変数として利用した。

3・5 主成分分析 (Principal component analysis)

本研究では対象をフェライト系耐熱鋼に限定しているものの、それでもかなり材料組成や材料種に広がりがある。この広大な空間の中で、各機関が偏りなくデータを保有している状況は考えにくい。連合学習では互いのデータを秘

Table 8. Modeling Conditions and Number of Institutions.

Modeling condition	local/global	Row extraction condition	batch_size	epochs	rounds	in_features	Number of institutes
creep_std	local	"Standard material (matrix)" == 1	260	500		25	6
creep_std	global	"Standard material (matrix)" == 1	260	10	50	25	6
creep_weld	local	"Weld metal" == 1	260	500		25	1
creep_weld	global	"Weld metal" == 1	260	10	50	25	1
creep_used	local	"Used sample" == 1	260	500		25	1
creep_used	global	"Used sample" == 1	260	10	50	25	1
creep_all	local		260	500		28	8
creep_all	global		260	10	50	28	8
tensile_std	local	"Standard material (matrix)" == 1	140	500		25	5
tensile_std	global	"Standard material (matrix)" == 1	140	10	50	25	5
tensile_weld	local	"Weld metal" == 1	140	500		25	1
tensile_weld	global	"Weld metal" == 1	140	10	50	25	1
tensile_used	local	"Used sample" == 1	140	500		25	1
tensile_used	global	"Used sample" == 1	140	10	50	25	1
tensile_all	local		140	500		28	7
tensile_all	global		140	10	50	28	7

†3 L.N.Smith and N.Topin, <https://doi.org/10.48550/arXiv.1708.07120>

匿するため、各機関のデータについて、どの程度の偏りや重なりがあったのかを検討できない。データの秘匿性を可能な限り保ったまま各機関のデータの広がりに関する手がかりを得るため、Principal component analysis (PCA) を利用した。

PCAでは各変数の線形結合により主成分と呼ばれる変数を合成する。第1主成分はデータの分散を最大化するような変数であり、第k主成分は第1主成分から第(k-1)主成分までのすべてと直交しかつ分散を最大化するような変数である。PCAは次元圧縮を目的として下で述べる寄与率という値が小さい主成分を除外するために用いられる。今回は第1、第2主成分のみを取り出すことにより元データを再現不能なデータへと加工するためにPCAを利用した。これによって、各機関のデータのもつ広がりやデータの位置を2軸への射影で表現し、他機関と比較できるようになる。

PCAでは特異値分解を用いている。すなわちn個のデータ X 、左特異ベクトルをならべた行列 U 、右特異ベクトルをならべた行列 W として $X=U\Sigma W^*$ という式で表される。対角行列 $\Sigma=diag(\sigma)$ の要素である特異値 σ は各主成分軸に関する標準偏差 s_i と関連がある($s_i=\sigma_i/\sqrt{n-1}$)。そして、標準偏差の2乗である分散の比として、各主成分の寄与率 $s_i^2/\Sigma_j s_j^2$ が表される。データ点 X を主成分スコア(主成分空間上の点を表す各軸成分) $T=XW$ へと変換し、そのうち寄与率が上位2つに対応する主成分スコアを抽出する。

実施手順としては、まずNIMSデータ X を用いてPCAを行い主成分スコアの計算に必要な情報 W, U, Σ を得た。各機関はそれを受け取り、機関 a のデータ X_a に主成分スコア計算と同様の変換 $T_a=X_a W$ を施し、第一主成分と第二主成分に関する列(t_1, t_2)を抽出して変換後のデータとした。ここに連合学習のホールドアウト検証における訓練・テストセットの別を示す情報を加え、各機関の変換後データを収集して第一主成分と第二主成分の2軸のプロットを重ねて各機関のデータの広がりを比較する。

4. 結果と考察

4.1 ローカルモデルとグローバルモデルの各自機関データによる評価

グローバルモデルおよび各ローカルモデルに関して各機関自身のデータで評価した結果をTable 9に示す。各モデル作成条件において個別学習と連合学習の訓練データは同じであり、訓練データのフィッティングによる評価およびテストデータの予測精度評価をそれぞれtrain, testに分けて記載した。比較のため評価結果をFig.4にモデル作成条件別にプロットした。縦軸はRMSEによる評価を示し、値が小さいほど精度が良い。横軸は機関名とテストデータ数・訓練データ数の組を表す。グローバルモデルと各ローカルモデルの別および訓練データ(train)とテストデータ(test)

の別の組み合わせについて異なる色を与えた。

この評価方法では殆どの場合でローカルモデルに比べグローバルモデルのテストデータに関するRMSEは大きい。各機関が有する学習データの範囲内では、連合学習による予測精度の向上は無く予測性能は悪化していると思われる。NIMSの場合はローカルモデルとグローバルモデルのRMSEの差が比較的小さい。これはモデル統合時の重み付けを訓練データ数に比例した重みとしたためNIMSデータで学習したモデルのパラメタが大きく反映されたためと思われる。このように、連合学習によって自身のデータを予測する精度は向上してはいない。

creep_weld, creep_used, tensile_weld, tensile_usedの各モデル作成条件では1機関のみで連合学習を行っている。その場合でも個別学習の評価結果との差があり悪化している場合が多い。他の複数機関が参加するモデル作成条件の場合には機関間のデータ数のアンバランスやデータの分布の違いが存在するが、上記4条件の場合には1機関のみのためそれらを考慮する必要はない。1クライアントで行う場合の個別学習と連合学習で条件が異なる要素はサイクル構成とそれに付随するepochやroundの設定のみであり、それらがスケジューラの挙動の違いを生じ学習過程の主要な相違点であったことについて付録Aで述べるとともにNIMSデータを用いた検証を行った。

4.2 NIMSデータに関するNIMSローカルモデルとグローバルモデルの評価

他機関に比べNIMSデータに関してはグローバルモデルとローカルモデルの予測精度の差が小さかった。実際にどの程度の予測精度が得られているかを観測値-出力値プロットにより示す(Fig.5(a)-(d))。学習時の観測値や出力値はスケール変換された値だが、ここでは元のスケールで表示している。いずれのプロットにおいてもローカルモデルのほうが観測と出力が良く合っていることが確認できる。グローバルモデルは全体的に分散が増しており、ローカルモデルと比較すると予測性能は低下しているものの、極端に予測性能が低下している条件は見られない。つまり、多くの条件である程度良い予測が出来ていると思われる。ローカルモデル側で大きく予測が外れた条件はグローバルモデル側でも外れている場合が多い。値が小さい側と大きい側がそれぞれ外れるよう全体的に対角線から傾く傾向がグローバルモデルにおいて見られる。事前にバッチサイズを調査した際に、個別学習試験でしばしば同様の傾向に遭遇していた。このことから、連合学習の学習過程やデータに直接的に由来するものではなくバッチサイズに関係して現れた現象であると思われる。

4.3 各クライアントのデータに関するNIMSローカルモデルとグローバルモデルの評価

モデル統合時の重み付けを訓練データ数に比例するものとしたがNIMSと他機関ではデータ数に一桁程度の違いが

Table 9. Evaluation results based on each institution's own data for local model and global model.

Modeling condition	local/global	Institution	Training data			Test data		
			N	MSE	RMSE	N	MSE	RMSE
creep_std	local	A	180	0.2024	0.4499	46	0.2613	0.5111
creep_std	local	B	292	0.0479	0.2189	74	0.1065	0.3263
creep_std	local	C	116	0.2223	0.4715	29	0.1721	0.4148
creep_std	local	D	38	0.0251	0.1585	10	0.0604	0.2458
creep_std	local	E	109	0.0249	0.1579	28	0.017	0.1303
creep_std	local	NIMS	4525	0.0032	0.0568	1132	0.0479	0.2189
creep_std	global	A	180	0.7275	0.8529	46	0.5938	0.7706
creep_std	global	B	292	0.4448	0.6669	74	0.5499	0.7415
creep_std	global	C	116	0.8152	0.9029	29	0.5891	0.7675
creep_std	global	D	38	0.655	0.8093	10	0.3609	0.6008
creep_std	global	E	109	0.3692	0.6076	28	0.4453	0.6673
creep_std	global	NIMS	4525	0.0386	0.1966	1132	0.075	0.2738
creep_weld	local	G	173	0.003	0.0543	44	0.0184	0.1355
creep_weld	global	G	173	0.0366	0.1914	44	0.0529	0.23
creep_used	local	F	64	0.0173	0.1317	16	0.0215	0.1465
creep_used	global	F	64	0.0616	0.2482	16	0.071	0.2664
creep_all	local	A	180	0.0542	0.2329	46	0.1204	0.347
creep_all	local	B	292	0.0753	0.2744	74	0.12	0.3464
creep_all	local	C	116	0.8013	0.8951	29	0.5188	0.7202
creep_all	local	D	38	0.0139	0.1179	10	0.0422	0.2055
creep_all	local	E	109	0.0233	0.1527	28	0.0172	0.1313
creep_all	local	F	68	0.009	0.0948	17	0.0073	0.0853
creep_all	local	G	173	0.0024	0.0493	44	0.0158	0.1256
creep_all	local	NIMS	4525	0.0034	0.0579	1132	0.0414	0.2036
creep_all	global	A	180	0.2992	0.547	46	0.2452	0.4952
creep_all	global	B	292	0.5059	0.7113	74	0.5401	0.7349
creep_all	global	C	116	0.4626	0.6801	29	0.4686	0.6845
creep_all	global	D	38	0.496	0.7043	10	0.325	0.5701
creep_all	global	E	109	0.5748	0.7581	28	0.5801	0.7616
creep_all	global	F	68	0.0914	0.3023	17	0.0053	0.0729
creep_all	global	G	173	0.1212	0.3482	44	0.1437	0.379
creep_all	global	NIMS	4525	0.0483	0.2197	1132	0.0861	0.2935
tensile_std	local	B	387	0.0023	0.0476	97	0.005	0.0709
tensile_std	local	C	15	0.0714	0.2672	4	0.237	0.4868
tensile_std	local	D	26	0.0142	0.1192	7	0.0238	0.1543
tensile_std	local	E	20	0.0061	0.078	6	0.0456	0.2135
tensile_std	local	NIMS	1720	0.001	0.0311	431	0.0125	0.112
tensile_std	global	B	387	0.0572	0.2392	97	0.0468	0.2164
tensile_std	global	C	15	6.7107	2.5905	4	8.3064	2.8821
tensile_std	global	D	26	1.3612	1.1667	7	1.0897	1.0439
tensile_std	global	E	20	0.3387	0.582	6	0.4768	0.6905
tensile_std	global	NIMS	1720	0.0211	0.1452	431	0.0322	0.1796
tensile_weld	local	G	73	0.0028	0.0532	19	0.0013	0.0356
tensile_weld	global	G	73	0.0097	0.0986	19	0.0075	0.0865
tensile_used	local	F	64	0.0825	0.2873	16	0.0603	0.2455
tensile_used	global	F	64	0.1057	0.3251	16	0.0933	0.3055
tensile_all	local	B	387	0.0245	0.1564	97	0.0227	0.1505
tensile_all	local	C	15	0.0256	0.1602	4	0.1395	0.3735
tensile_all	local	D	26	0.0017	0.0417	7	0.0039	0.0627
tensile_all	local	E	20	0.0072	0.0848	6	0.0466	0.2158
tensile_all	local	F	68	0.0873	0.2954	17	0.1022	0.3197
tensile_all	local	G	73	0.002	0.0446	19	0.0152	0.1232
tensile_all	local	NIMS	1720	0.001	0.0323	431	0.0131	0.1145
tensile_all	global	B	387	0.0532	0.2306	97	0.0488	0.2209
tensile_all	global	C	15	11.8621	3.4441	4	15.2857	3.9097
tensile_all	global	D	26	1.1415	1.0684	7	0.9217	0.96
tensile_all	global	E	20	0.2985	0.5464	6	0.549	0.7409
tensile_all	global	F	68	3.0349	1.7421	17	2.857	1.6903
tensile_all	global	G	73	0.3775	0.6144	19	0.4735	0.6881
tensile_all	global	NIMS	1720	0.0184	0.1355	431	0.0256	0.1601

ある。もしグローバルモデルにNIMS以外の機関のモデルのパラメタが殆ど反映されていないならば連合学習の意義が疑わしくなる。そのため、実際に得られたNIMSローカルモデルとグローバルモデルにはどの程度の違いがあるか検討する。NIMSのデータではcreep_std, creep_all, tensile_std, tensile_allの作成条件が学習可能であった。これらのモデル作成条件に関してNIMSローカルモデルと対応するグローバルモデルについて各機関のテストデータに対して精度を計算し、対応するモデル間で精度を比較した。モデル作成条件毎に各軸でNIMSローカルモデルとグローバルモデルによるテストデータのRMSEをそれぞれ表したプロットをFig.6に示す。

各プロットの上部にモデル作成条件名を示し、横軸でNIMSローカルモデルによる精度、縦軸でグローバルモデルによる精度を示す。左下にあるほど、グローバルモデル、NIMSローカルモデルともに精度がよく、右であるほどNIMSローカルモデルによる精度が悪く、上であるほどグローバルモデルの精度が悪いことに相当する。両モデルの精度が等しくなる箇所に45度の黒線を描いた。45度線より右下の領域は、NIMSローカルモデルでは精度が悪いがグローバルモデルでは精度が良い場合を示す。ほとんどの点が45度線上かあるいは45度線の右下の領域に存在しており、グローバルモデルはNIMSローカルモデルと比べて性能が改善されていることがわかる。特に、creep_stdにおけるD, creep_allにおけるA, C, D, Gでは改善が顕著である。他に、creep_allのE, tensile_stdのD, tensile_allにおけるDなどではグローバルモデルは僅かに精度が良い。このように、各機関のデータを学習することによってNIMSデータではカバーできなかった領域のデータに対して予測性を獲得できたといえ、連合学習の効果を確認できた。

ただし、cについては、例外的に、tensile_stdとtensile_allにおいて予測性能が低下した。データを秘匿していること、データやモデルが複雑であることからこの原因を探ることは難しい。

4・4 NIMSデータに関するローカルモデルとグローバルモデルの評価

各機関自身の持つデータに近い条件であれば自身のローカルモデルを用いて予測すれば良い。しかし、機関自身の持つデータ範囲にない外側の条件に関しての予測性を高めることが、今回、連合学習を行う目的である。NIMSデータ

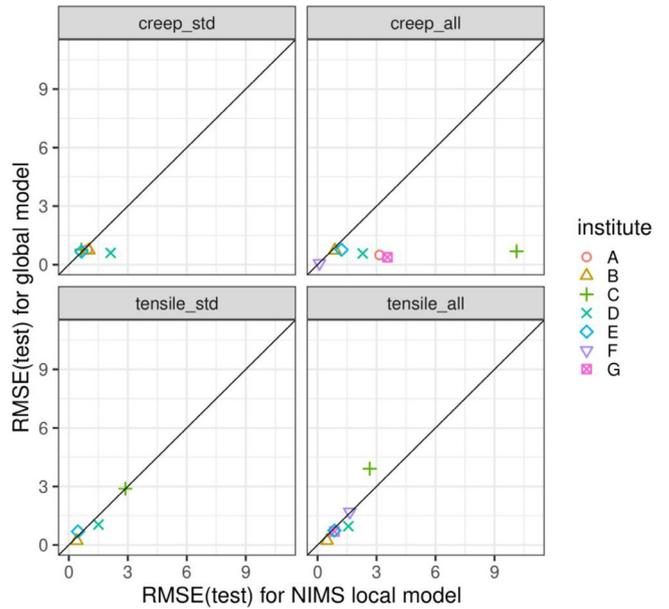


Fig. 6. Comparison of prediction accuracy between the global model and the NIMS local model by data from respective institutes. (Online version in color.)

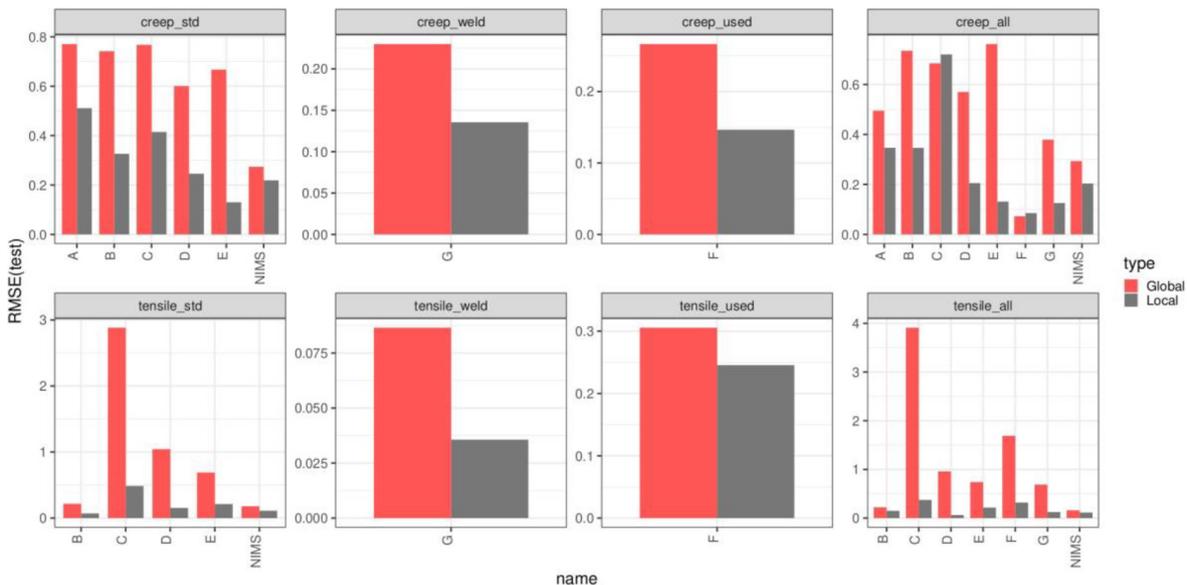


Fig. 4. Comparison of evaluations by each institution's own data between local model and global model. (Online version in color.)

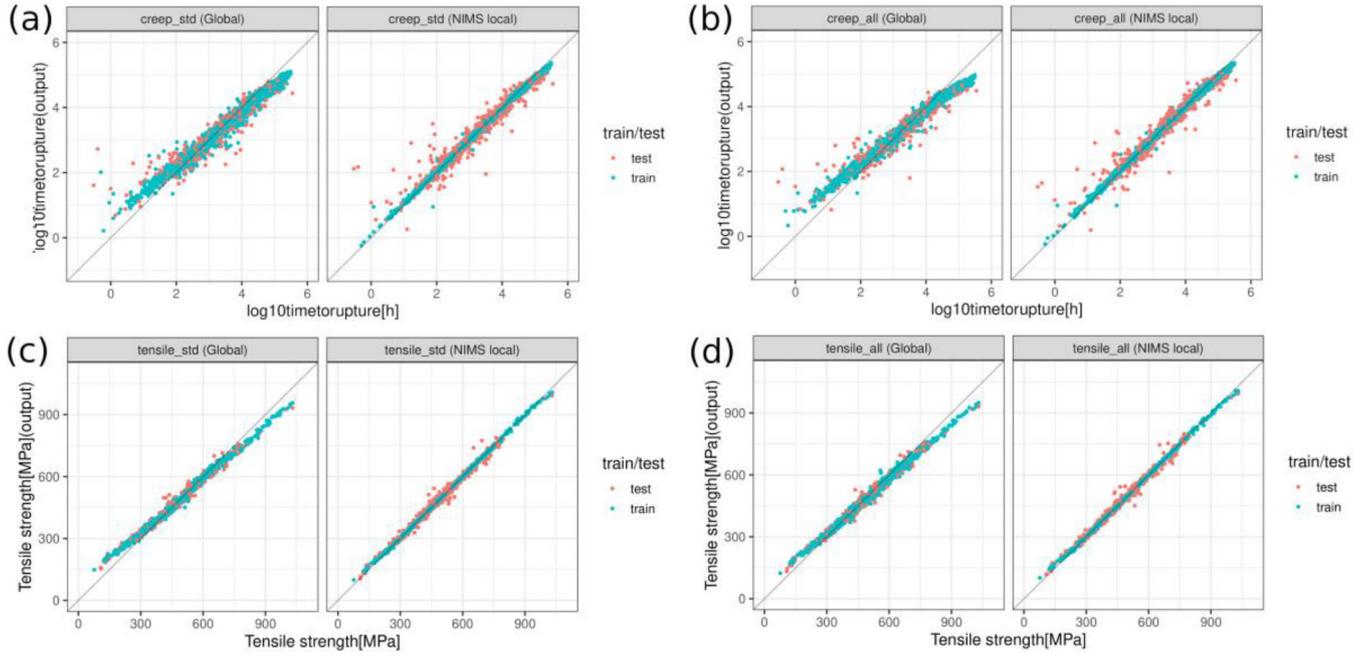


Fig. 5. Comparison between global model and local model for each modeling condition by plotting measured vs. predicted values (in original scale) of NIMS data: (a) creep_std; (b) creep_all; (c) tensile_std; (d) tensile_all.

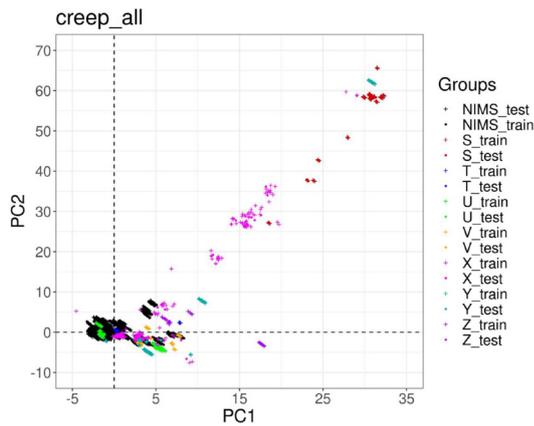


Fig. 8. Principal component scores for the first and second principal components of creep_all. Colors represent institutions and symbols represent training/testing.

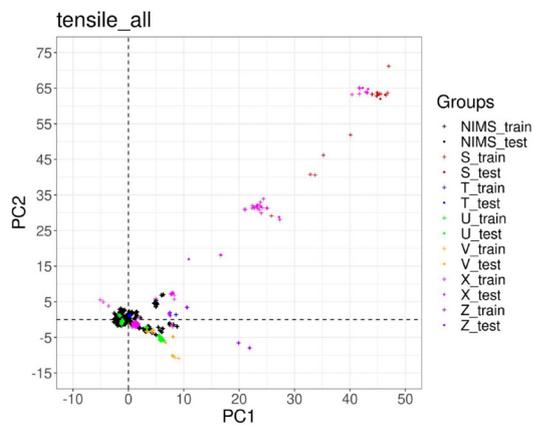


Fig. 10. Principal component scores for the first and second principal components of the tensile_all. Colors represent institutions and symbols represent training/testing.

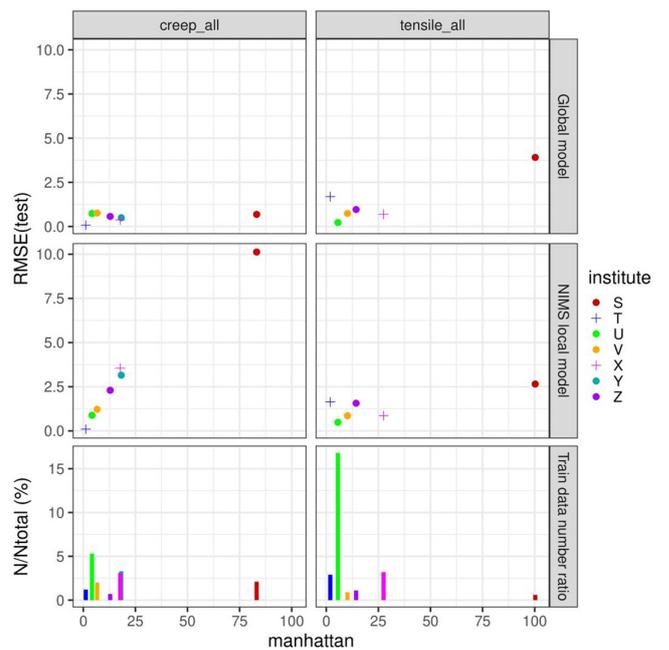


Fig. 12. Comparison of prediction accuracy between the global Relationships between Manhattan distance and prediction accuracies for test data of the global model and the NIMS local model, respectively, and relationship between Manhattan distance and the number of training data, on creep_all and tensile_all.

と各機関のデータがどの程度異なっているのかを知ることは出来ず厳密な比較が難しい。以下では各機関のローカルモデルおよびグローバルモデルについてNIMSデータに対する予測性能を評価することで、模擬的に外側の条件に関する予測性能を調査する。

Fig.7にモデル作成条件毎に各機関のローカルモデルおよびグローバルモデルによる精度をバープロットで示す。いずれのモデル作成条件においてもグローバルモデルの精度は他機関のローカルモデルによる精度よりNIMSデータの条件に関して優れた予測性能を有していることがわかる。このように連合学習によって、各機関のローカルデータにない外側の条件に関する予測性能を向上させることが示されたといえる。

4・5 主成分分析 (Principal component analysis)

説明変数空間に着目して、各機関のデータの違いを主成分分析で表現することを試みた。モデル作成条件 creep_all について、Fig.8に各機関の主成分スコアを第一主成分 (PC1)-第二主成分 (PC2) 軸上にプロットした。各機関を異なる色で表し、訓練・テストの別をシンボルの違いで表した。機関名は精度との対応を伏せるために別名とした。

NIMSのデータを表す黒い点は原点付近の比較的密集した集団とPC1側に10程度まで伸びる領域、PC1, PC2とも5程度の位置にある領域にみられる。それぞれの中で左上か

ら右下に向かう連続した点が見られるが、それらは試験温度や試験応力の条件のみ異なるデータ群だと思われる。機関 T, U, V, X, Y にはNIMSデータと重なっている、あるいは、近い点が存在する。ただし、PC1, PC2以外の主成分の情報がないため、実際に近いデータ点であるかは不明である。他方、機関 X, S, Y では大きく右上にかけて広がっており、これらの点はPC1, PC2という上位2主成分について異なる値を有している。これらの点は、左下の領域の点と大きく離れた条件であることは明らかである。他に、PC1が10程度までの領域やPC1, PC2が5付近の領域の周囲を埋めるような条件もある程度見られる。テストデータ周囲に訓練データが存在しないケースは2軸上でしか判断できないが稀であるとみられ、テストデータの選ばれ方が予測精度を極端に乱す状況にはなっていないと想像される。

このように様々に広がった各機関の主成分スコアと予測精度の間に関係があるかを検討するため、各機関の主成分スコアの統計量を求めて creep_all のテストデータによる予測精度との間のプロットを行った。統計量として、ある機関の主成分スコア全点の重心 ($\overline{PC1}$, $\overline{PC2}$) からNIMS訓練データの重心 (0,0) までのユークリッド距離、マンハッタン距離、および、各軸のみの差 $\overline{PC1}$, $\overline{PC2}$ を計算した。これらの統計量とグローバルモデルとNIMSローカルモデルによるテストデータの予測精度をFig.9にプロットした。

NIMSローカルモデルでは特にマンハッタン距離やユークリッド距離と予測精度の間に線形に強い正の相関が見られた。各機関のデータ範囲がNIMSの訓練データから遠いほどNIMSローカルモデルによる予測性は悪化する傾向を示しており、直感に適う。これに対してグローバルモデルでは、最もNIMSデータ重心から遠いSに至るまでいずれの機関のデータに対しても良い予測精度を維持している。このように、データを互いに秘匿したまま溶接継手や使用損傷材を含むかなり広範なフェライト鋼に対して予測性能を維持するグローバルモデルを構築できたことは驚くべき

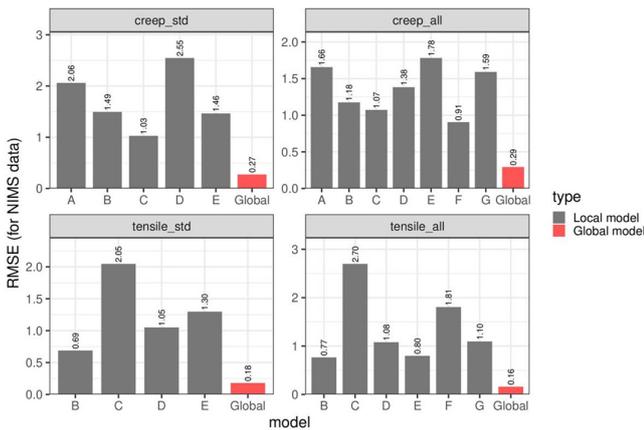


Fig. 7. Comparison of prediction accuracy between the global model and the local models of the institutions using NIMS data. (Online version in color.)

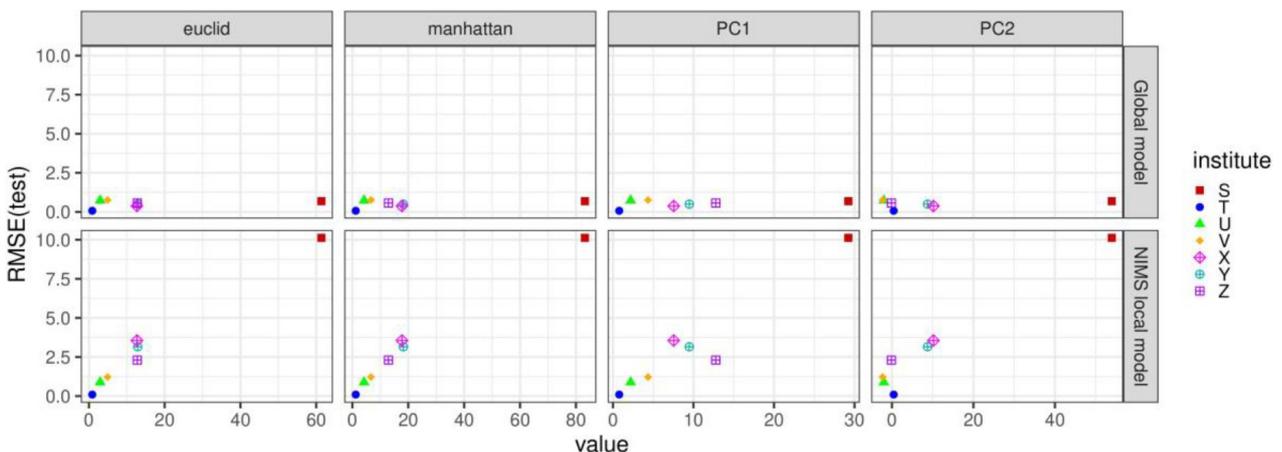


Fig. 9. Relationship between predictive accuracy for test data and each principal component score statistics on global model and NIMS local model for creep_all. (Online version in color.)

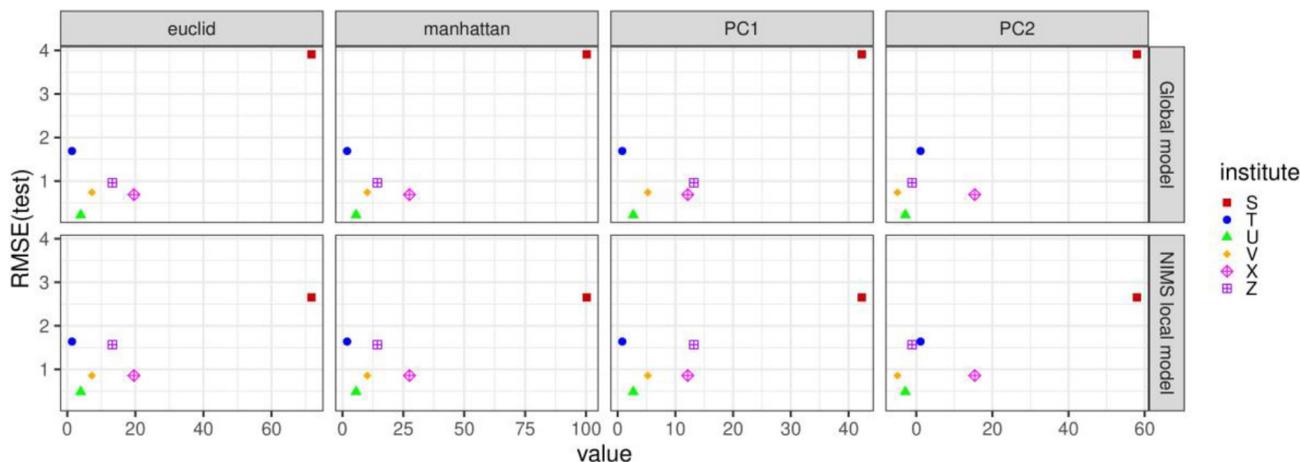


Fig. 11. Relationship between predictive accuracy for test data and each principal component score statistics on global model and NIMS local model for tensile_all. (Online version in color.)

結果である。今回開発した連合学習システムと学習の手順によって、データ数が最多の NIMS モデルのパラメタに偏らずに、各機関のデータで学習したモデルパラメタをよく反映できた結果といえる。

creep_allの場合と同様に、モデル作成条件tensile_allについて、Fig.10に各機関の主成分スコアを第一主成分(PC1)-第二主成分(PC2)軸上にプロットした。creep_allと同様にNIMSのデータを表す黒い点は原点付近とPC1側に伸びる領域、および、PC1、PC2とも5程度の領域に見られる。機関T、U、V、XにはNIMSデータと重なっているか、あるいは近い点が存在する。また、X、Sは大きく右上にかけて広がっている。

tensile_allに関して主成分スコアの統計量とグローバルモデルとNIMSローカルモデルによるテストデータの予測精度をFig.11にプロットした。

上記プロットそのままでは予測精度と各統計量の明らかな関係が見られない。しかし、試験片種類が異なるTとXを除外すると、creep_allと同様にマンハッタン距離やユークリッド距離と予測精度の間に強い正の相関があると見ることが出来る。NIMSデータには試験片種類を示す変数についてStandard material (matrix) が1でWeld metalとUsed sampleが0のデータしか存在しない。そのため、PCAで得られたWにおけるこれらの変数の重みは0であった。すなわち試験片種類の情報が主成分スコアおよび上記プロット中のデータの遠さとして反映されていないためにTとXが上記関係に合わなかった可能性が考えられる。

tensile_allのグローバルモデルではNIMSローカルモデルと比べて明確に予測が良いとは言えず離れた機関Sでは悪化している。これについて検討するためFig.12にマンハッタン距離と予測精度の関係をプロットしtensile_allとcreep_allの結果および訓練データ数に関して比較した。なお試験片種類の異なるT、Xについてはシンボルを変えて示した。

creep_allでは、すでに述べたように、NIMSローカルモデルの予測精度はデータが遠いほど悪化しているがグローバルモデルで改善されている。一方でtensile_allではNIMSローカルモデルでも既に十分に小さいRMSEが得られており、グローバルモデルでこれ以上の改善ができていないことがわかる。最下行に各機関の訓練データ数割合(%)をバープロットで示した。パラメタの統合はデータ数割合に応じた加重平均を用いたため、データが多いほど反映されやすい。NIMSデータから最も遠いSの訓練データ数割合はcreep_allの場合に2.1%と少数であったが、グローバルモデルはある程度良い予測を与えることができた。tensile_allの場合は機関Sの訓練データ数割合は0.6%と更に僅かであり、これがグローバルモデルによる改善が難しい理由であると考えられる。

5. 結言

本研究では連合学習システムを開発し、フェライト系耐熱鋼のクリープ破断時間および高温引張強度の予測モデルの作成を目的としてNIMS含め8機関の連携によってデータを秘匿した状態で連合学習を実施した。データ毎に試験片種類の扱いの異なるモデル作成条件を設けて全8種のグローバルモデルを作成した。比較のため各機関のローカルモデルも作成した。加えて、主成分分析の主成分スコアを利用しNIMS訓練データ重心から各機関訓練データ重心へのマンハッタン距離と予測精度の関係を調査した。

自機関データ範囲の条件の予測は自機関のローカルモデルが特化しており殆どのケースでグローバルモデルより良い予測精度であった。これに対して、NIMSデータを用いて各機関データ条件外の予測性能を評価するとグローバルモデルで予測精度の改善が確認された。

クリープ破断時間予測ではNIMS訓練データ重心からの距離とともにNIMSローカルモデルの予測精度が低下

したがグローバルモデルでは最も遠い機関も含め良い予測性能が維持された。高温引張強度予測では試験片種類が Standard material (matrix) である機関に限定すればNIMS ローカルモデルの傾向は同様だがグローバルモデルでは最も遠い機関で予測が改善されていない。当該機関の訓練データ数割合はクリープで2.1%、高温引張で0.6%であり、ある程度のデータ数があれば訓練に用いる効果が確認できたといえる。以上の結果から、各機関データを秘匿した状態でも広くカバーできるグローバルモデルを得ることができたと結論される。

利益相反に関する宣言

本論文に関して開示すべき利益相反関連事項はない。

謝辞

本研究は、内閣府総合科学技術・イノベーション会議の戦略的イノベーション創造プログラム (SIP)「統合型材料開発システムによるマテリアル革命」(管理人：JST), 文部科学省データ創出・活用型マテリアル研究開発プロジェ

クト事業JPMXP1122684766の助成、および、NIMS 構造材料DX-MOPの粹組みのもと実施されました。

付録A 連合学習におけるスケジューラの影響

1クライアントの場合に個別学習と連合学習で予測性能が異なった原因について検討する。両者の違いはサイクルの構成とそれに伴う epoch数やround数といった条件設定であり、これが学習率を調整するスケジューラの挙動の違いを生じさせる。OneCycleLRスケジューラによる学習率調整の過程は3・3・2に書いた通り epoch全体の30%をかけて学習率を増加させアニーリングした後に減少させる。例えば、個別学習の場合には500 epochにわたり1回のアニーリングを行い、Fig.A1 (a) に示すように、30% (150 epoch) までは学習率を上げるが、その後過学習を抑制するために学習率を次第に低減させていく。他方、連合学習の場合には各roundで1回のアニーリングが行われることになり、Fig.A1 (b) のように50回アニーリングが行われる。各roundには10 epochしか含まれないので、小刻みに学習率の増加と低減が行われることになる。

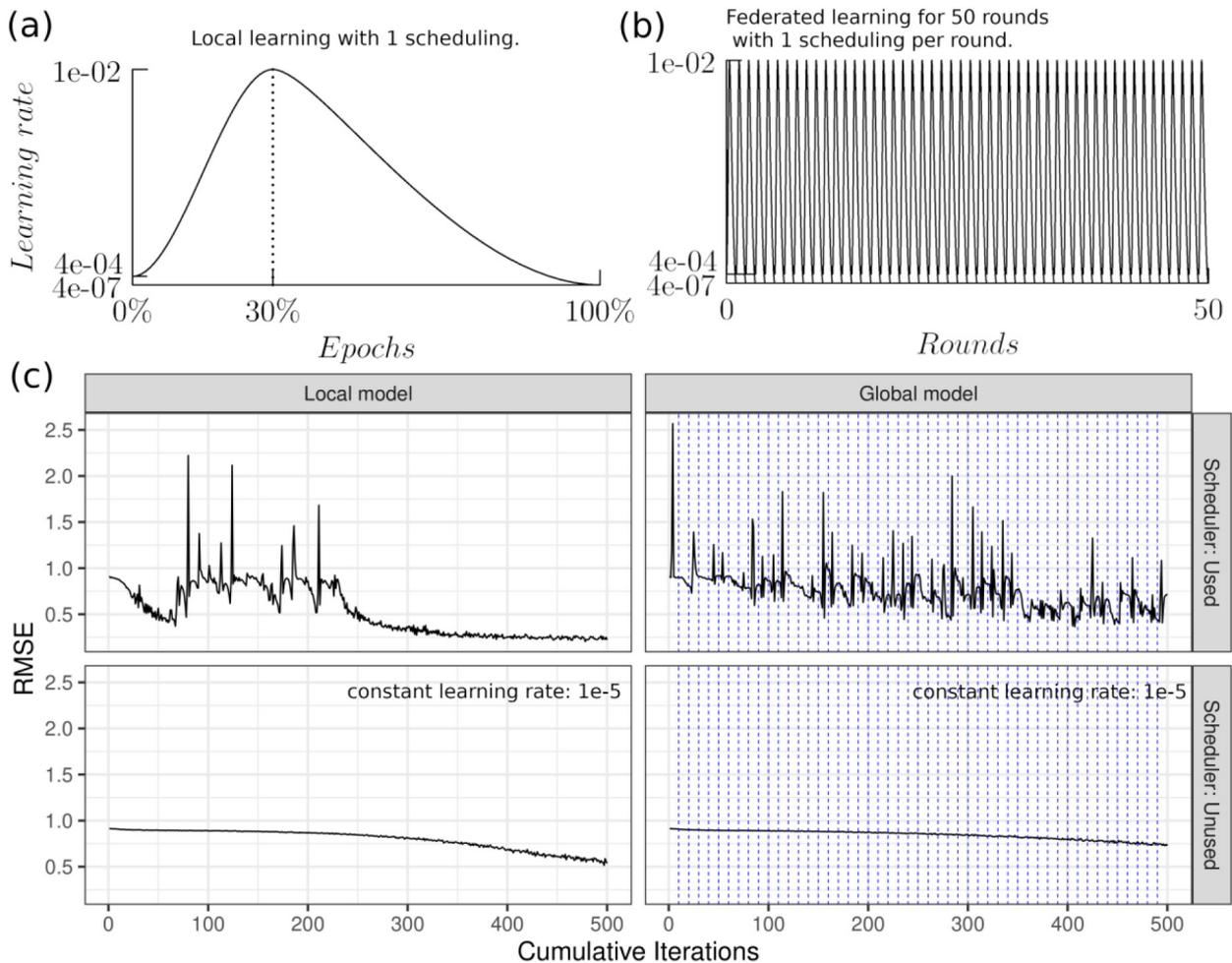


Fig. A1. Schematics of learning rate curves scheduled with OneCycleLR for (a) local learning and (b) federated learning.; (c) RMSE changes on local learning and federated learning held with 1 client until 500 iterations with and without OneCycleLR scheduling. Blue vertical dotted lines separate rounds of federated learning. (Online version in color.)

実際の学習過程を比較するため、他機関と同程度のデータ数に絞ったNIMSデータ (CDS19Bのクリープ破断データから採取した訓練データ数218およびテストデータ数55) を用いて1クライアントによる連合学習と個別学習を行なった。総イテレーション数 (1 epoch内でパラメタの更新回数はイテレーション数*i*だけ存在し、データ数*n*とバッチサイズ*b*を用いて $i = \lceil n/b \rceil$ と表される) 時点でのMSE損失をRMSEに変換してFig.A1 (c) にプロットした。まず、スケジューラを使わず一定の学習率 ($1e-5$) で学習を行った場合 (Fig.A1 (c) 下段 Scheduler: Unusedプロット), 学習曲線の挙動は個別学習と連合学習で大きな違いはなく、徐々に損失が減少していく。OneCycleLRを用いた場合 (Fig.A1 (c) 上段 Scheduler: Usedプロット) には個別学習と連合学習で学習曲線の挙動に違いが現れた。詳細に見ると、個別学習では学習率が高いアニーリング中は大きくパラメタを動かすため不安定な挙動をとり、その後は安定して損失が減少し収束に向かう様子が分かる。連合学習では不安定な挙動が最後まで繰り返し現れ、最終roundのアニーリング後に十分に損失の減少が進まない段階で学習が終了している。以上の結果から、1クライアントにおける個別学習と連合学習の違いはスケジューラの導入による

ものと推定される。本研究では、過学習を防ぎながら効率よく学習を進めるためにスケジューラを導入しているが、データ数が少ない場合には不安定な挙動を引き起こしている可能性がある。

文 献

- 1) K.Sawada, K.Kimura, F.Abe, Y.Taniuchi, K.Sekido, T.Nojima, T.Ohba, H.Kushima, H.Miyazaki, H.Hongo and T.Watanabe: *Sci. Technol. Adv. Mater.*, **20**(2019), 1131. <https://doi.org/10.1080/14686996.2019.1697616>
- 2) J.Sakurai, M.Demura, J.Inoue and M.Yamazaki: *Tetsu-to-Hagané*, **108**(2022), 424 (in Japanese). <https://doi.org/10.2355/tetsutohagane.TETSU-2022-003>
- 3) H.B.McMahan, E.Moore, D.Ramage, S.Hampson and B.A.y Arcas: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, PMLR Vol. 54, PMLR, Fort Lauderdale, FL, (2017), 1273.
- 4) kMoL, a machine learning library for AI drug discovery with federated learning functionality, has been released., (February 8, 2022), PR TIMES Corporation, <https://prtimes.jp/main/html/rd/p/000000007.000027687.html>, (accessed 2024-10-28), (in Japanese).
- 5) J.Guo, Z.Chen, Z.Liu, X.Li, Z.Xie, Z.Wang and Y.Wang: *Sci. Rep.*, **12**(2022), 15326. <https://doi.org/10.1038/s41598-022-19426-8>
- 6) B.Chen, K.Xu, Y.Zhu, L.Tian and V.Chang: *IEEE Trans. Ind. Inform.*, **19**(2023), 2228. <https://doi.org/10.1109/TII.2022.3205372>