

# 機械学習を用いたプリンターインクヘッド交換時期の予測

中村伊吹<sup>\*1</sup>, 増田武史<sup>\*1</sup>, 左座祐之助<sup>\*1</sup>, 広瀬啓雄<sup>\*</sup>

<sup>\*1</sup> 公立諏訪東京理科大学

## Predicting when to replace printer ink heads using machine learning

Ibuki Nakamura<sup>\*1</sup>, Takeshi Masuda<sup>\*1</sup>, Yuunosuke Zouza<sup>\*1</sup>, Hiroo Hirose<sup>\*1</sup>

<sup>\*1</sup> Suwa University of Science

大判インクジェットプリンタの保守作業において点検する部位やタイミングは作業者の経験や勘によるものが大きく、熟練技能者から新人への技術伝承に多くのコストを要する。本研究では機械学習を利用して大判インクジェットプリンタのインクヘッド交換を予測することで保守作業への技術支援を行うシステムの構築を目的とする。予測にはプリンター機器から送信されてくる印刷枚数等のプロファイルデータを使用する。プロファイルデータは次元数が膨大であるため、Random Forest, Tab Net から算出される重要度を用いて有効な特徴量を抽出し、Gated Recurrent Unit, Auto Encoder を用いて故障の予測を行う。結果として、11 件の故障記録を学習させ 6 回中 6 回の予測を行うことができた。また、予測の問題点として検出期間が一定ではなく、検出期間があまりに長いものについては無駄なメンテナンスにつながる可能性があげられる。今後は予測精度の向上や作成した予測モデルを利用したシステムの開発に取り組んでいく。

キーワード: 技術支援システム, 機械学習, メンテナンス, Random Forest, Gated Recurrent Unit

### 1. はじめに

#### 1.1 研究の背景

A 社では大判インクジェットプリンタを製造し、世界各国に販売とサービスの提供を行っている。A 社ではプリンターインクヘッドの故障が発生した場合、保守担当者が早急にお客様先へ出向き、プリンターインクヘッドの交換を行っている。この時、お客様先では故障が発生してから A 社へ連絡をし、A 社から保守担当者が現場に到着し修理を行うまでに業務中断期間が発生してしまう。また、保守担当を行う A 社においても早急な保守対応が求められるため、急な業務が発生してしまう。(図 1)

プリンター機器が故障することを事前に知ることができれば A 社側からお客様先へプリンター機器の保守作業の呼びかけを行うことや A 社では計画的な保守業務を行うことが可能である。

近年の機械学習技術は急速に発展しており異常検知、

故障予測の分野においても深層学習の登場により自動的に特徴を抽出して機器の故障を判断することができるものまである。深層学習を用いた故障予測のほとんどは画像データを用いたものがほとんどでありセンサーといった数値データを用いたものはあまりない。故障予測技術が活用される製造業の現場においてはセンサーを設置しているものの取得したデータを有効に活用できている企業はあまりなく、製造業においてセンサーデータの活用は保守コストの削減や稼働率の上昇の観点からも重要度が高いものである。また、センサーデータの活用機械学習を用いることの利点として統計的な予測と比較して統計分布を仮定する必要がないため取得したセンサーデータをより有効に扱えることや統計的な予測は説明性を重視していることに対して機械学習は予測の精度を重視しているため機器の故障を見逃すことも少なくなる。

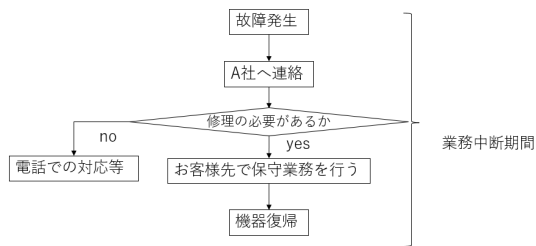


図 1.A 社における保守対応

## 1.2 研究の目的

本研究の目的は、各プリンター機器から送信されてくるプロファイルデータと機械学習を用いてプリンターインクヘッドの故障を予測することで的確なタイミングでの保守対応に利用可能な故障予測モデルを構築することを目的とする。また、研究の最終段階としては作成した予測モデルを利用した保守作業向けの技術支援システムの構築を目指す。

## 2. 対象データ

分析に使うデータはプリンター機器から不定期に送られてくるプロファイルデータを用いる(表 1)。プロファイルデータにはインクの使用量やクリーニング状況、印刷枚数、カッター使用回数などの変数が存在する。データの取得期間は機体によって変動するもののおよそ 3~4 年でその期間に 1,2 回インクヘッドの故障によるヘッド交換が行われている。

表 1.データの概要

	次元数	データ件数	故障件数	機体数
前処理前	125	3748	18	39
前処理後	302	15877	17	14

また、今回の分析に用いるデータには予測を行う際に以下のような問題点が存在する。

1 点目はデータの取得間隔である。各プリンター機器から送られてくるプロファイルデータの取得間隔が不定期であるため予測を行う際に補完が必要であることがあげられる。本研究では質的変数には近傍補間を行い、量的変数には線形補間を行うことで対応をしている。機体によっては 1 週間に 1 度の間隔でしかデータが送られていない機体も存在するため補間による予測精度の低下が起きてしまうことが考えられる。

2 点目は予測に用いる変数が膨大であることがあげられる。今回のデータは予め予測に不必要と思われる

変数を落とした状態で 300 次元近く存在しており、故障の予測を行う上で重要な変数を選択することや高次元のデータを低次元に削減する様な前処理が重要である。

## 3. 前処理

### 3.1 サンプル間隔の変更

A 社より提供されたデータの取得間隔が不定期であるため、機械学習を行うためにはサンプル間隔を一定にする必要がある。本研究では質的変数に対しては近傍補完を行い、量的変数に対しては線形補完を行うことでアップサンプリングを行い、サンプル間隔を 1 日毎に変更した(図 2)。

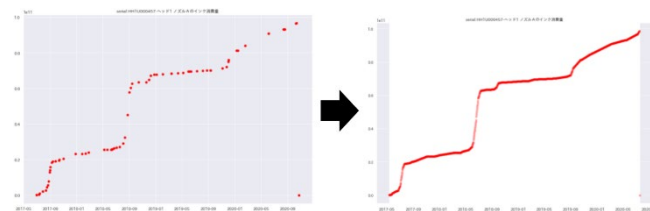


図 2.サンプル間隔の変更前(左) 変更後(右)

### 3.2 データ数の少ない機体の削除

取得したデータの中にはデータ数が少ない機体で 2 件、多い機体で 300 件以上といった差が見られた(図 3)。データ数の少ない機体では上記のサンプル間隔を変更する際に行ったデータ補完が上手く行えないと考え、補完前のデータ数が 100 件以下の機体は学習に使用せず、100 件より多い機体を学習に使用した。



図 3.データ数の少ない機体におけるデータ取得間隔

### 3.3 カウンター系変数の初期化

大判インクジェットプリンタには印刷枚数や使用したインクの量、カッターの使用回数といったカウンタ

一変数が多く存在している。これらのカウンター系変数をヘッド交換毎に初期化することで、ヘッド交換からのプリンターの使用状況をモデルに学習させ、経年劣化のような予測を行うことが出来ると考えた(図4)。

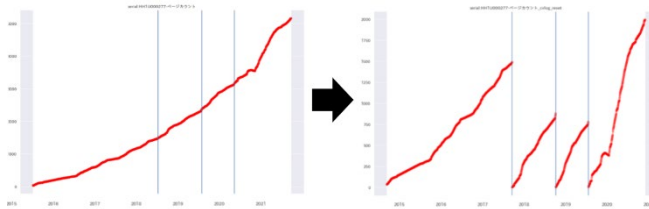


図4.カウンター系変数の初期化前(左),初期化後(右)

### 3.4 移動平均,階差系列

移動平均とは、一定区間の平均をデータの区間を移動しながら求めていくことで短期的なデータの変動を除去し、長期的なデータの傾向であるトレンドを抽出することができる。

階差系列とは、データのひとつ前の値との差をとることでデータから長期的な傾向であるトレンドが除去され、短期的な変動を求めることができる。

本研究では、移動平均と階差系列を各モータ寿命、カッターなどのカウンター系変数、各ヘッドのインク消費量において適用し特徴量として加えた

### 3.5 次元圧縮

本研究では Isomap を用いて次元削減を行った(図5)。Isomap とは非線形の次元削減手法の一つで K 近傍グラフを用いて多様体上の測地戦距離を求め、多次元尺度構成法を使い近似的に低次元空間に射影を行う。

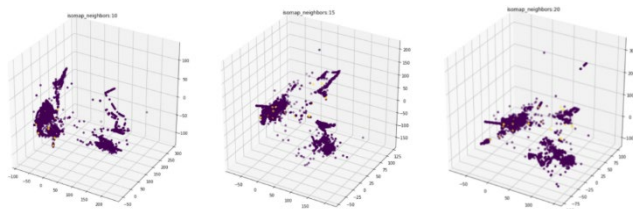


図5.次元削減後の特徴量

本研究で使用するデータは 300 次元以上の高次元なものであるため Isomap といった次元削減手法を行うことで高次元のデータを少ない次元の変数で表すことが期待できる(図5)。

## 4. 実験

### 4.1 実験目的

各プリンター機器から送信されてくるプロフィールデータから機械学習を用いてプリンターインクヘッドの故障予測を行い、的確なタイミングでのメンテナンスに活用できるかを検証する。

### 4.2 実験方法

Auto Encoder 以外の各機械学習アルゴリズムではヘッドの交換日にクラス 1 それ以外の日にクラス 0 としてラベルを作成し、作成したラベルを教師データとしてクラス分類を行う。

データの次元数が膨大であるため、予測に重要と思われる特徴量を抽出する。予測に重要と思われる特徴量の選定には、機械学習手法である Random Forest と Tab Net を用いて全特徴量で予測を行い、各変数の重要度を求める。Random Forest, Tab Net から計算された各変数の重要度の上位 10 個ずつ、最大で 20 個の特徴量を用いて Auto Encoder, Gated Recurrent Unit で故障予測を行う。最後に、各機械学習アルゴリズムでの予測結果を表 5 にまとめる。

使用するデータは、前処理後データの 70% を学習用データ、30% をテストデータとする。学習用データを用いて学習を行った後、構築した予測モデルを用いてテストデータを予測することでモデルの汎化性能を評価する。また、学習用データには 11 件の故障記録とその間の正常データ、テストデータには 7 件の故障記録とその間の正常データが含まれている。

### 4.3 実験結果

各機械学習手法による予測結果を示す。

#### 4.3.1 Random Forest による特徴量選択

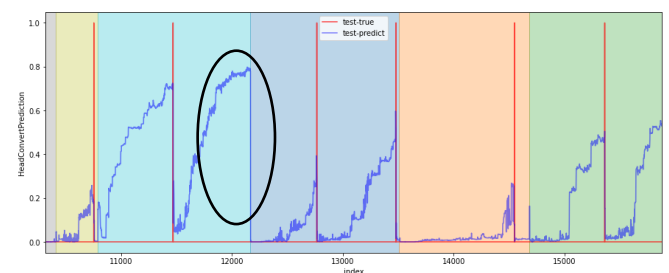


図6. Random Forest による予測結果縦軸：故障確率、横軸：日数の経過、背景色：機体の変化、赤線：ヘッド交換日（ヘッド交換の行われた日を 1, それ以外の日

を 0) 青線 : 予測モデルによって得られた故障確率

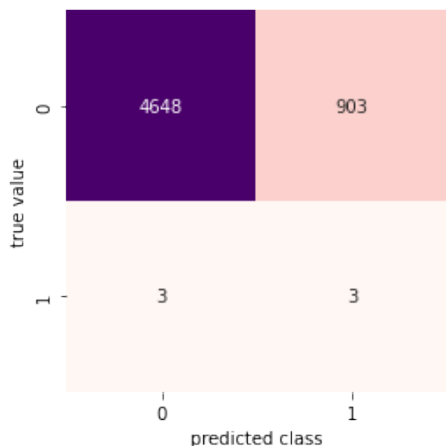


図 7.Random Forest による予測結果 (混同行列)

テストデータにおいて 6 回中 3 回の予測を行うことができた.故障確率がヘッド交換日に向けて上昇する傾向がみられるため故障の予兆をとらえることができたと考えられる.しかし,上図における黒丸の期間で多くの誤検知が発生していることがわかった(図 6,7).

以下の表 2 に Random Forest による重要度の上位 10 個の特徴量を示す.

表 2.Random Forest から算出された重要度

変数	重要度
データ長 (JobStatus)_cvlog_reset	0.024096257904726973
カラー番号 5 のインク消費量_cvlog_reset	0.023637793299075227
カラー番号 4 のインク消費量_cvlog_reset	0.02334422169381352
timestamp_cvlog_reset	0.022832627448841753
キャリッジ (Y) 方向モーター_cvlog_reset	0.017998170053829542
ポンプモーター A のステップ数_cvlog_reset	0.017136965297676738
集積印字タイマ (秒)_cvlog_reset	0.0170907458976037
原点検測液のインク量_cvlog_reset	0.016963623388771602
カラー番号 2 のインク消費量_cvlog_reset	0.016645316457560552
カラー番号 5 のインク消費平均_cvlog_reset	0.016606551530761814

※cvlog\_reset : ヘッド交換日に初期化したカウンター系変数,mean:一週間ごとの移動平均

Random Forest での予測の際に重要度が高いと思われる変数には前処理 4-3 で行ったヘッド交換日に初期化したカウンター系変数が多くみられた.

#### 4.3.2 Tab Net による特徴量選択

Tab Net による予測結果を以下に示す.

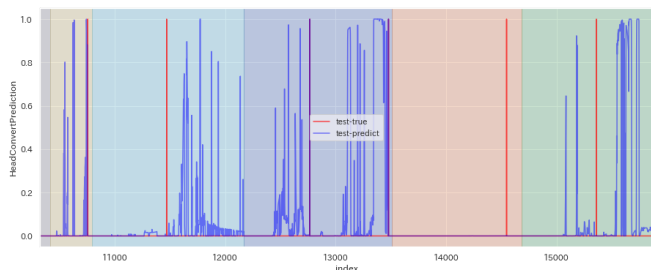


図 8.Tab Net による予測結果縦軸 : 故障確率,横軸 : 日数の経過,背景色 : 機体の変化,赤線 : ヘッド交換日 (へ

ッド交換の行われた日を 1,それ以外の日を 0) 青線 : 予測モデルによって得られた故障確率

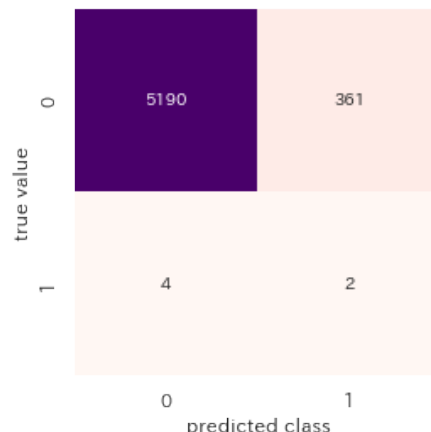


図 9.Tab Net による予測結果 (混同行列)

テストデータにおいて 6 回中 2 回の予測を行うことができた.予測モデルから得られた故障確率はヘッド交換日に向けて徐々に上昇する傾向はみられなかったが,Random Forest と比較して故障検出回数に対する誤検知の数を 121 回抑えることができた(図 8,9).

以下の表 3 に Tab Net による重要度の上位 10 個の特徴量を示す.

表 3.Tab Net から算出された重要度

変数	重要度
カラー番号インク消費合計_cvlog_reset	0.2815989992801387
LLE_neighbors10_0	0.17211420112100614
ポンプモーター寿命: 集積_cvlog_reset	0.1536578446929551
ポンプモーター寿命: 集積_mean	0.10725626287168133
ヘッド1-4インク消費合計_mean	0.10494808558028465
ヘッド2インク消費合計_mean	0.06932142774725947
カラー番号 5 の充填されているインクの種類	0.0564163924825021
isomap_neighbors15_1	0.041481126666020614
原点検測液のインク量_mean	0.004375840440321316
カラー番号 5 のインク消費量_mean	5.106580410194818e-05

※cvlog\_reset : ヘッド交換日に初期化したカウンター系変数,mean:一週間ごとの移動平均

Tab Net での予測の際に重要度が高いと思われる変数は Random Forest と同様にヘッド交換日に初期化した変数に加えて,一週間ごとの移動平均をとった変数が重要であることが分かった.

#### 4.3.3 Auto Encoder

Auto Encoder による故障予測を行う.予測に用いる変数は Random Forest から得られた重要度の上位 10 個と Tab Net から得られた重要度の上位 10 個を用いて行う.

Auto Encoder による予測結果を以下に示す.

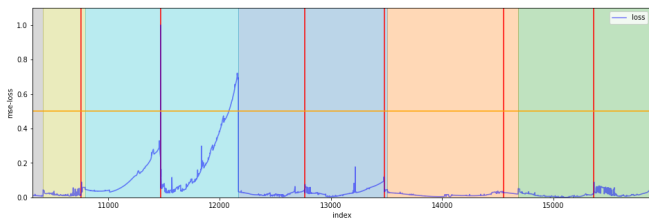


図 10. Auto Encoder による予測結果縦軸：異常度,横軸：日数の経過,背景色：機体の変化,赤線：ヘッド交換日（ヘッド交換の行われた日に 1,それ以外の日を 0）青線：予測モデルから得られた異常度

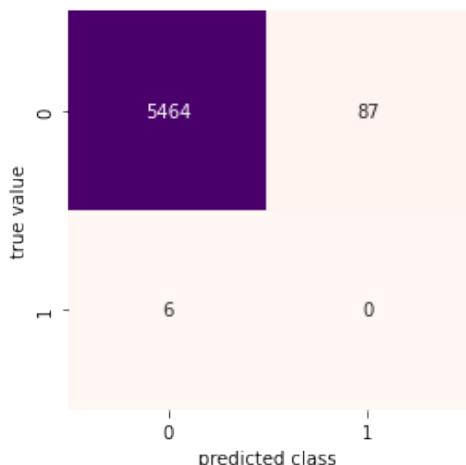


図 11.Auto Encoder による予測結果（混同行列）

Auto Encoder では 6 回中 0 回の故障検出であった。しかし,2 回目の故障においては前日に予測モデルが故障を検出していることが分かった.前日に予測を行うことのできた 2 回目の故障以外の故障記録においては異常度が上昇せず適切な予測を行えていない結果となった(図 10,11).

#### 4.3.4 Gated Recurrent Unit による予測

Gated Recurrent Unit による故障予測を行う.予測に用いる変数は Random Forest から得られた重要度の上位 10 個と TabNet から得られた重要度の上位重要度の上位 10 個を用いて行う.

Gated Recurrent Unit による予測結果を以下に示す.

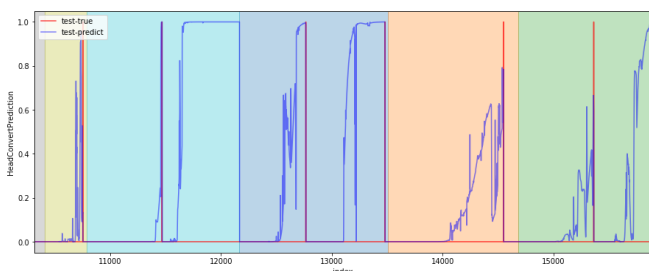


図 12.Gated Recurrent Unit による予測結果

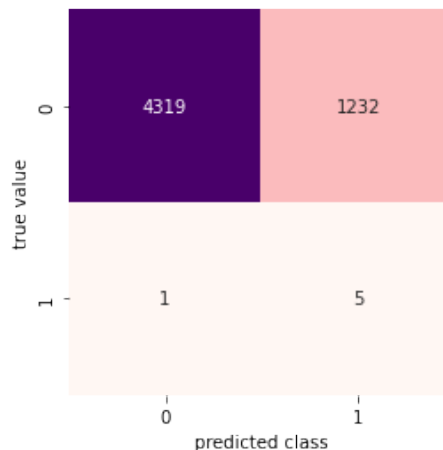


図 13.Gated Recurrent Unit による予測結果(混同行列)

故障の閾値を 50%とすると 6 回中 5 回の予測を行うことができた.ヘッド交換日に向けて故障の確立が上昇している傾向がみられ,Random Forest と比較しても故障検出回数に対する誤検知の数を 55 回抑えられていることがわかった.一回目の故障について,ヘッド交換日に予測モデルから出力されるクラスが 1（故障と判定されるクラス）にはならなかったものの前日に故障の予測が行えていた.また,本研究で使用した機械学習手法の中でヘッド交換日の見逃しが最も少ない結果となった(図 12,13).

#### 4.3.5 各予測結果の比較

各予測結果を下記表にまとめる.

今回の実験で行った故障予測は使用する機械学習手法によって前日では故障と判断されていたが,次の日には故障しないと判断されてしまうといったブレが発生している.そのため,ヘッド交換が行われるより前に予測モデルが一度でも故障すると判断した場合,ヘッド交換が行われるまで常に故障が発生する予測が出ているものとする(図 14).

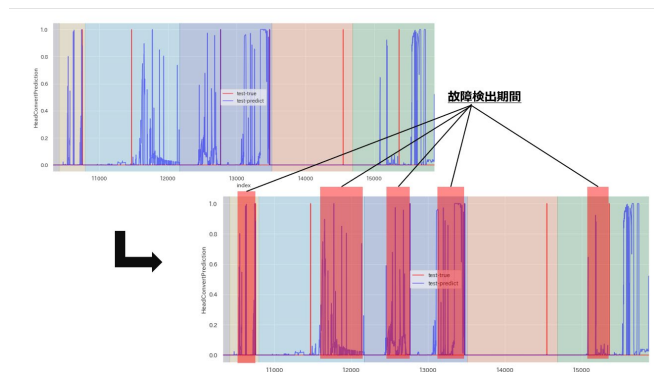


図 14.故障検出期間の例

以下の表 5 に各機械学習手法によるインクヘッド交換の検知期間を表す.

表 4 各機械学習手法による予測結果の比較

	1回 目	2回 目	3回 目	4回 目	5回 目	6回 目
Random Forest	×	403	×	3	×	1
TabNet	215	×	310	372	×	180
Auto Encoder	×	1	×	×	×	×
Gated Recurrent Unit	23	11	201	354	135	5

## 5. 考察

重要だと思われる特徴量について、インクの消費量やモータの寿命といったヘッド交換ごとにカウントをリセットした変数は Random Forest と Tab Net の両者の機械学習アルゴリズムにおいて重要度が高い結果となった。このような結果になった理由として以下の2点が考えられる。1 点目は、今回使用したデータに欠損値が大きく存在していたため、補間を行いやすいカウンター系変数の重要度が高いと判断されたこと。2 点目は、カウンター系変数をヘッド交換日ごとに初期化したことで各インクの消費量の値やモータの寿命といった変数がそれぞれの形でインクヘッドの経年劣化を現したのではないかと考えることができる。

Auto Encoder では 6 回中 1 回の交換予測を行うことができた。期待通りの予測結果が得られなかった原因について、Auto Encoder は今回の研究で使用した機械学習手法の中で唯一の教師なし学習で、クラス分類ではなく異常検知をもとにした故障予測を行っていることがあげられる。教師なし学習を行う場合、予測結果をある特定の目標に向けて最適化することは難しく、各種パラメータやモデルの構造、学習に使用する特徴量の選択をより慎重に行う必要があったことが考えられる。

Gated Recurrent Unit では 6 回中 6 回の予測を行うことができ、他の機械学習手法と比較してインクヘッド交換日の見逃しが少なく、正常時に故障と判断する誤検知についても最も少ない結果となった。Gated Recurrent Unit での予測が最も良くなった理由とし

て、Gated Recurrent Unit には時系列を考慮した学習が可能であり、今回の予測に使用したプリンター機器から送信されてくるプロファイルデータも各センサーの記録やインクの消費量といった時系列データであるため予測をうまく行うことができたことが考えられる。

予測精度の向上を図るには以下の3点が重要であると考えられる。1 点目は、データの取得間隔の改善である。今回の予測に使用したデータは各プリンター機器から不定期に取得しているため、機体によっては一か月に一度しかデータを取得していない機体も存在している。データの取得間隔をより短くすることができれば、今回のデータで補間がうまくいかなかった短期的に変動するような変数を扱えるようになるため、より直近の予測を行うことができると考えられる。

2 点目は特徴量選択である。今回の研究では特徴量選択を Random Forest と Tab Net の二つの機械学習手法に依存している。そのため、実際に保守業務を行っている A 社の職員に故障原因のヒアリングを行い、特徴量選択に適用することで予測精度の向上が期待できる。

3 点目は、今回の実験で使用していない機械学習手法の検討である。今回の実験では使用していない他の手法として、k 近傍法や One Class SVM、予測精度の高かった時系列を考慮した予測が可能な手法である Convolutional Neural Network や Long Short Term Memory といった機械学習手法で実験することで予測精度の向上につながる可能性がある。

本研究の目的は各機械学習手法による予測結果が的確なタイミングでのメンテナンスに活用できるかということである。データ提供をいただいた A 社で働く社員の方に今回の予測に対する評価をいただいた結果、Gated Recurrent Unit を用いた故障予測は本研究で使用したデータに関してはメンテナンスに有効であると判断された。しかし、今回の実験で作成した予測モデルを使用したメンテナンスを実際に実施するためには、今回の実験では使用していない新たに取得したデータを用いて予測モデルの精度を調査していく必要があることがあげられた。

## 6. まとめ

本研究では、プリンター機器から送信されてくるプ

## 参 考 文 献

- (1) 梅津里香,杉江卓哉,長瀬雅之,et al. ”機械学習を用いた宇宙機の故障の予兆検知,宇宙科学情報解析論文誌”,JAXA-RR-18-008,8号,p.11-20 (2019)
- (2) 奥 武憲,水野 哲志,林 俊輔,荒井 仁,et al. MFP のビッグデータを活用した故障予測”, Ricoh technical report,,43号,p. 56-62 (2018)
- (3) 今村 誠, “予知保全のための機械学習”, システム／制御／情報, 2021, 65 巻, 4 号, p. 119-125 (2021)
- (4) Arik, Sercan O.Pfister, Tomas,TabNet: “Attentive Interpretable Tabular Learning,eprint arXiv”:1908.07442, (2019)
- (5) 毛利拓也 ,北川廣野 ,澤田千代子 ,谷一徳”scikit-learn データ分析 実装ハンドブック”,秀和システム(2020)
- (6) Ankur A. Patel, “Python ではじめる教師なし学習——機械学習の可能性を広げるラベルなしデータの利用”,オライリー・ジャパン (2020)
- (7) 農学情報科学, “gated recurrent unit / RNN (参照日 2021 年 12 月 13 日)”<https://axa.biopapyrus.jp/deep-learning/rnn/gru.html>
- (8) 井出剛,杉山将,”異常検知と変化検知”,講談社(2015)

ロファイルデータと機械学習技術を利用することでプリンター機器のインクヘッド交換を予測し,的確なタイミングでのメンテナンスに役立てることを目的とした. Random Forest, Tab Net, Auto Encoder, Gated Recurrent Unit の 4 種類の機械学習手法を用いてインクヘッド交換の予測モデルを構築した.Random Forest, Tab Net を用いて 300 次元以上のデータから重要度の高い特徴量を 10 個ずつ,最大で 20 個の特徴量を選択し,Gated Recurrent Unit で予測することで 11 件の故障記録からテストデータにおいて 6 回中 6 回のインクヘッド交換の予測を行うことができた.また,Gated Recurrent Unit による予測をプリンター機器の製造を行っている A 社の職員に評価していただき,今回の研究で使用したデータにおいてはメンテナンスに有効であることが考えられるという結果となった.しかし,Gated Recurrent Unit による予測結果では検出期間が短いもので 5 日前,長いもので 354 日前と一定でなく,検出期間があまりに長いものについては無駄なメンテナンスにつながる可能性があげられる.予測精度を向上させるためには,今まで不定期に取得していたデータの間隔を数時間毎や半日毎に切り替えることや,故障の原因となりそうな要素を A 社職員とのヒアリングを通じて調査すること,今回の実験では使用しなかった機械学習手法を利用して予測モデルの構築を行うことがあげられる.

今後の課題として,新たなデータを取得して本研究で作成した予測モデルの精度を検証することや,故障の予測を行うだけではなく,機械学習技術を用いて故障の原因を突き止めることで保守業務へのコストダウンを図ること,以下の図 15 に示すような予測モデルの出力結果を A 社の保守対応を行う作業員に伝えることで保守作業の技術支援を行うシステムの構築があげられる.

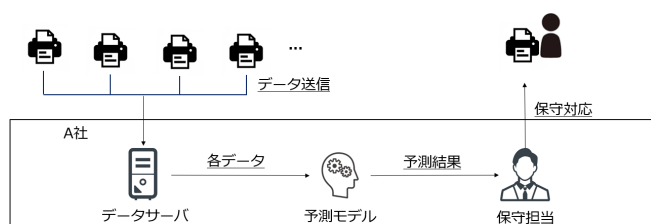


図 15.予測モデルを利用したメンテナンスシステム図