

## 複合機の故障における特徴抽出への機械学習の適用 Fault Feature Extraction Method in Multifunction Peripheral

稲木 達哉<sup>†</sup> 高明淑<sup>†</sup>  
Tatsuya Inagi Myungsook Ko

### 1. はじめに

複合機メーカーでは、顧客の業務負荷低減と複合機のダウンタイム削減を目的に、消耗品の手配や故障発生時の修理依頼、印刷枚数の検針、アップデートなどを自動化するため、複合機の各種センサーデータやカウンターデータ、ファームウェア、機体構成情報、エラーデータなどをインターネット経由で取得する遠隔保守サービスを提供している。また、複合機は保守エンジニアによって定期メンテナンスや故障および障害発生時の対応作業などが日々行われており、それら保守作業についても作業内容や交換したパーツなどの履歴データが保存されている。近年、これら遠隔保守サービスと保守作業履歴に蓄積されてきたデータを活用し、更なる価値を提供することを目的とした故障分析や故障診断、故障予測の技術開発が進められている[1][2][3]。

複合機の故障は、大きく 3 つに分類される。1 つ目は画質に関する問題で、例えば印刷した用紙に黒いスジが出る現象である。2 つ目は紙の搬送に関する問題で、給紙から排紙までに発生する紙詰まりが代表的な現象である。3 つ目は、前述を除く電装や機械動作に関する問題である。1 つ目と 2 つ目については、主な原因がパーツの劣化や汚れであるため、市場に導入してから時間が経つにつれ、徐々に発生回数が増加する傾向にある。一方、3 つ目の電装や機械動作に関する問題は、主な原因が設計や製造に起因するため、新製品の市場導入直後から発生する傾向がある。そのため、電装や機械動作に関する問題は、早期の故障原因の特定と対応を行わなければならない。よって、新製品導入後の市場から収集した遠隔保守サービスのデータを分析する必要がある。しかし、遠隔保守サービスで収集された数千種類のデータの中から、故障の原因や原因部位、故障が再現する条件や組み合わせに関する情報を人の手で絞り込むには時間がかかる。

そこで本研究では、エンジニアの故障調査の効率化に向けて、複合機の遠隔保守サービスで収集されたデータから故障機体に共通する特徴を抽出し提示することで、故障の原因や原因部位、故障が再現する条件や組み合わせに関する情報を絞り込む作業を支援することを目的に、故障機体に共通する特徴を抽出する手法に機械学習の手法を適用し、過去の調査済み故障での評価を行った。

### 2. アプローチ

遠隔保守サービスで収集されるデータには、センサー値などの数量データと、印刷時の設定情報および機体構成情報などのカテゴリカルデータが存在する。そのため本研究

では、カテゴリカルデータに関してはダミー変数化を行った上で数量データと共に入力データとする。その入力データから目的変数として対象故障を 1、その他故障を 0 と判別するモデルを機械学習の手法を用いて構築する。その後、構築したモデルに含まれる説明変数から故障機体に共通する特徴を抽出する。

モデル構築の手法には、ランダムフォレストと Elastic Net[4]の正則化項付きロジスティック回帰を用いた。ランダムフォレストは、全入力変数から重複なしでランダムに選択した入力変数をもって構築した決定木を弱識別器とし、複数決定木の多数決によって判別を行うモデルであり、説明変数の重要度を得ることができる。正則化項付きロジスティック回帰は、複雑なモデルにペナルティを科す正則化項をロジスティック回帰のコスト関数に追加することで説明変数の選択が行われ、故障機体に共通する特徴の絞り込みができる。構築したモデルより、ランダムフォレストについては説明変数の重要度 TOP10 に入った特徴を、正則化項付きロジスティック回帰については正負の重み係数各 TOP10 に入った説明変数を、故障機体に共通する特徴として抽出し、エンジニアに提示した。

### 3. 過去事例での評価

#### 3.1 評価データと評価方法

東芝テック株式会社の遠隔保守サービスにおける機種 A 約 2 万台の約 1 年間分のデータを用いた。この期間に発生した故障 5 種類を対象に、エンジニアの調査結果で明らかになった故障原因に関する特徴を抽出できるか評価を行った。具体的には、過去の故障 5 種類毎に、共通特徴を抽出する対象の故障レコードと、そのレコード件数と同数となるように対象故障を除いたその他の故障レコードを重複なしで抽出し、入力データを作成した。そして、各入力データから構築した各手法のモデルで、2. で示した共通特徴としてエンジニアに提示する範囲に、既知となった故障原因に関する特徴が入っているか評価した。

過去の故障 5 種類における既知の故障原因に関する特徴の数は表 1 である。

表 1 既知となった故障原因に関する特徴数

	特徴数 (特徴)
故障 1	1 (低気温)
故障 2	1
故障 3	2
故障 4	1
故障 5	1

<sup>†</sup>株式会社 東芝 研究開発センター システム技術ラボラトリー, System Engineering Laboratory Corporate Research & Development Center TOSHIBA Corporation

### 3.2 モデルの判別精度の検証

判別精度の低いモデルでは、故障発生機体に共通する特徴が上手くモデルで表現できていないと考えられる。そこで、共通特徴を抽出する前に各手法の判別精度を 10 分割交差検証で求めた。表 2 に過去の故障 5 種類でのランダムフォレスト (RF) と正則化項付きロジスティック回帰 (LR) における F 値の平均と標準偏差の結果を示す。故障 5 種類において、各手法で F 値の平均が 0.8 を超えることを確認した。

表 2 各事例における各手法での F 値

	F 値 平均 (標準偏差)	
	RF	LR
故障 1	0.885 (0.051)	0.876 (0.028)
故障 2	0.896 (0.026)	0.872 (0.028)
故障 3	0.904 (0.019)	0.863 (0.048)
故障 4	0.862 (0.005)	0.814 (0.035)
故障 5	0.878 (0.006)	0.860 (0.008)

### 3.3 共通特徴の抽出結果と考察

過去の故障 5 種類において、どちらの手法でも表 1 に示す既知の故障原因に関する特徴を全て抽出できることを確認した。

表 3 は、故障原因に関する特徴が低気温である、故障 1 の共通特徴を抽出した結果例の一部である。抽出結果については、ランダムフォレストおよび正則化項付きロジスティック回帰ともに、ジョブ設定やオプションパーツ情報といった複合機ならではの機体に共通する特徴が一目で分かるような抽出結果となっている。故障 1 の故障原因の特徴である低気温については、ランダムフォレストおよび正則化項付きロジスティック回帰の両方で、気温が特徴として抽出できている。正則化項付きロジスティック回帰では気温の重みが負である。これは、高い温度だと故障 1 が発生しやすく、低い温度の方がより故障 1 が発生しやすい状況であることを示している。

抽出結果に対してエンジニアにヒアリングを行った。故障 1 では、故障が無い平時には抽出されない低温度といった特異な原因特徴が抽出されるので、理解しやすいとの意見であった。一方、カテゴリカルデータの変数については、ダミー変数化を行っているため、例えば連続した順序関係のある変数が歯抜けで出てきた時など、何を表しているのか分かりづらい場合もあると意見が出た。これに対しては、遠隔保守サービスの入力変数となるデータ毎に適切な特徴表現方法の検討が必要であると考えられる。また、正則化項付きロジスティック回帰に対しては、カテゴリカルデータの変数毎にグループを設定した上で、グループ単位で変数選択を行う Group lasso[5]を正則化項に用いる手法が考えられる。Group lasso を用いることで、ダミー変数化する前の変数自体の絞り込みを行うことができ、またダミー変数化後のグループ内の重み係数の比較確認ができる。よって、連続した順序関係のあるカテゴリカルデータの変数についても、変数選択後も歯抜けにならず前後との関係も確認できる提示が行えると考えられる。

これらエンジニアにヒアリングした結果より、複合機の遠隔保守サービスで収集されたデータから故障機体の特徴

を抽出する手法としては、正則化項付きロジスティック回帰を用いた方がランダムフォレストより、故障調査支援の観点ではより好ましい情報が提示できると考えられる。

### 4. おわりに

本研究では、遠隔保守サービスのデータから故障発生時の機体に共通する特徴抽出にランダムフォレストと正則化項付きロジスティック回帰を適用し、各手法において過去の故障 5 種類に対して原因特徴を抽出できることを確認した。しかし、今回の過去の故障では、原因に関する特徴の数が多くても 2 つと少ない故障での評価であったため、今後は故障原因に関する特徴がより多い故障に対しても評価していく。

表 3 故障 1 におけるエンジニアに提示した結果例の一部

A. ランダムフォレストにおける特徴の重要度 TOP10			B. 正則化項付きロジスティック回帰における負の重み係数 TOP10		
順位	特徴	重要度	順位	特徴	重み
1	カウンター A 値	0.072	1	ジョブ設定 J	-1.078
2	ジョブ設定 B	0.067	2	ジョブ設定 H	-0.988
3	ジョブ設定 C	0.045	3	製造情報 K	-0.979
4	ジョブ設定 D	0.043	4	ROM_L 情報	-0.810
5	ジョブ設定 E	0.042	5	FW_M バージョン	-0.777
6	センサー F 値	0.040	6	ジョブ設定 N	-0.663
7	気温	0.039	7	給紙情報 O	-0.620
8	ジョブ設定 G	0.034	8	ジョブ設定 C	-0.599
9	ジョブ設定 H	0.032	9	オプションパーツ P 情報	-0.597
10	ジョブ設定 I	0.031	10	気温	-0.573

### 参考文献

- [1] 足立 康二, 山田 紀一, 上床 弘毅, 安川 薫, “ベイジアンネットワークによる複合機故障診断技術”, 富士ゼロックス テクニカルレポート, No.19 (2010).
- [2] 北原 洋一, “障害発生前にメンテナンスを促す MFP 障害予測技術”, 東芝レビュー, Vol.67, No.12 (2012).
- [3] 奥 武憲, 水野 哲志, 林 俊輔, 荒井 仁, 西村 拓, 西田 浩, 古賀 文晴, 松島 妙利, 畑中 諭志, “MFP のビッグデータを活用した故障予測”, Ricoh Technical Report, No.43 (2018).
- [4] Hui Zou, Trevor Hastie, “Regularization and Variable Selection via the Elastic Net”, Journal of the Royal Statistical Society. Series B, Vol.67, No.2 (2005).
- [5] Ming Yuan, Yi Lin, “Model Selection and Estimation in Regression with Grouped Variables”, Journal of the Royal Statistical Society. Series B, Vol.68, No.1 (2006).