

# システムインテグレーションプロジェクトでのAIによるコスト悪化の予兆検知

江上侑希\*  
Yuuki Egami  
吉田賢吾\*  
Kengo Yoshida  
白浜広彬\*  
Hiroaki Shirahama

鶴田季丸\*  
Tokimaru Tsuruta  
西ノ宮弘一\*  
Koichi Nishinomiya

Predicting Cost Overrun Levels in System Integration Projects Using AI

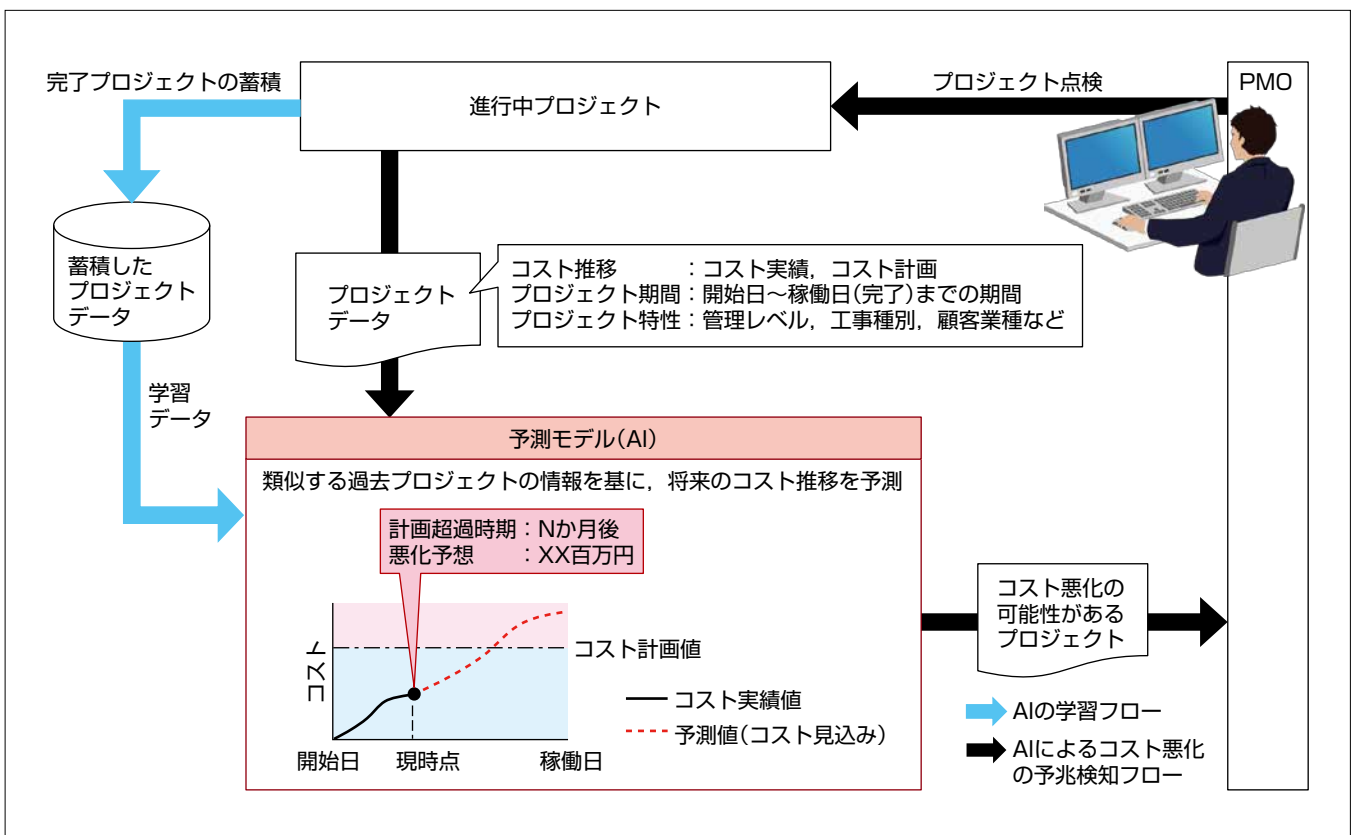
## 要旨

三菱電機インフォメーションシステムズ株式会社(MDIS)では、プロジェクトのコスト計画達成を目的としてプロジェクト管理の強化やプロジェクト点検等の対策を図り、一定の成果を上げてきた。しかし、プロジェクト点検でコスト悪化の予兆を検知するのは点検者のスキルに依存し、年間数百件ある全プロジェクトを網羅的に点検するには、リソースのひっ迫と過大な負荷がかかる状況であった。そこで、スキル依存の解消と全プロジェクト点検の負荷低減を実現するためにAIを利用し、コスト悪化の予兆を検知する予測モデルの構築に取り組んだ。

予測モデルの学習データには、プロジェクトの予実管理として蓄積してきたプロジェクトデータ(コスト推移、プ

ロジェクト期間、プロジェクト特性)を用いた。プロジェクトデータをコスト推移の形状が類似するパターンで分類し、プロジェクト特性とともに予測モデルに学習させた。これによって、進行中プロジェクトのコスト推移の形状とプロジェクト特性が類似する過去プロジェクトの情報を基に、稼働日までのコスト推移の予測を可能にした。

構築した予測モデルは、試行運用としてプロジェクトの点検を行う組織(PMO: Project Management Office)によるプロジェクト点検で活用しており、コスト悪化の予兆検知に一定の効果があることが確認できた。今後も予測モデルの改善を継続していくことで、プロジェクト計画の確実な達成に貢献していく。



## AIによるコスト悪化の予兆を検知する仕組み

プロジェクトのコスト計画達成を狙って、蓄積したプロジェクトデータをAIに学習させることで、コスト悪化の予兆を検知する予測モデルを構築した。予測モデルは進行中プロジェクトのデータを入力することで、類似する過去プロジェクトの情報を基にコスト悪化の可能性のあるプロジェクトを一覧を出力する。その結果を踏まえて、PMOが対象プロジェクトを点検することで問題がないか確認し、必要に応じて対策立案・フォローを実施する。

## 1. ま え が き

MDISではプロジェクト計画を達成するために、PMOによるプロジェクト点検等の対策を図り、一定の成果を上げてきた。しかし、プロジェクト点検でコスト悪化の予兆を検知できるかは点検者のスキルに依存し、全プロジェクトを網羅的に点検するには、リソースのひっ迫と過大な負荷がかかる状況であった。

一方、AIの発展・浸透とともに、故障予兆診断、需要予測等、蓄積してきたデータを利活用することで業務を支援する事例がAI白書<sup>(1)</sup>で報告されている。

本稿では、AIによるコスト悪化の予兆を検知する仕組みとその推進上の課題と対策について述べる。

## 2. AI導入の経緯

### 2.1 コスト悪化撲滅に向けたMDISの取組みと課題

一般的に大規模プロジェクトほどプロジェクト計画達成の難易度が上がり、計画未達時の影響度も高くなる。そのため、MDISでは、プロジェクトの規模、難易度、重要性、リスク等を勘案して管理レベルを設定し、管理レベルが高いプロジェクトに対して、PMOによる重点的なプロジェクト点検を実施してきた。その成果もあり、PMOが重点的に点検しているプロジェクトでは、大幅なコスト悪化が減少してきた。しかし、PMOの管理が十分に行き届かないプロジェクトでは小規模悪化が発生しており、全プロジェクトで計画を達成させるために対策が必要であった。

コスト悪化を防止するためには、PMOによるプロジェクト点検のカバー範囲を広げる必要があったが、悪化予兆を見つけるには点検者のスキルに依存するという問題がある。また、年間数百件のプロジェクトが進行している中で、全プロジェクトを網羅的に点検するには、リソースのひっ迫と過大な負荷がかかる状況であった。

### 2.2 蓄積してきたプロジェクトデータの利活用

MDISでは、プロジェクト管理として月次で予実管理を実施している。そのため、月次で変化していくプロジェクトデータを利用できる状況であった。

そこで、プロジェクト点検の課題解決として全プロジェクトを点検し、コスト計画の達成率を上げるため、点検負荷低減の観点からAIに着目し、蓄積した過去のプロジェクトデータを利活用することでコスト悪化の予兆を検知する仕組みづくりに取り組んだ。

## 3. AIによるコスト悪化の予兆を検知する予測モデルの構築

### 3.1 構築方針

- (1) 大量の時系列データを機械的に処理するために、AIの要素技術の一つである機械学習で予測モデルを構築する。
- (2) プロジェクトの開発形態は、システム開発、パッケージソフトウェアを利用したシステム開発、システム基盤構築等多岐にわたるため、プロジェクトの特徴に対応した予測を行える予測モデルを構築する。
- (3) コスト悪化プロジェクトを漏れなく検知できたかの指標である“再現率”とコスト悪化と予測したものが正しかったか(実際に悪化したか)の指標である“適合率”を用いて評価する。二つの指標はトレードオフの関係にあるため、コスト悪化プロジェクトの検知を目的とする再現率を優先する。
- (4) データの“質”が予測モデルの精度を左右するため、学習に利用するデータは、統計分析に基づいて選定する。また、データの定義が統一されていない場合、精度に悪影響を及ぼすため、プロジェクト間で定義が一定であるデータを優先する。

### 3.2 構築アプローチ

- (1) 過去の知見を踏まえて、仮説を立案する。
- (2) 仮説に基づいて、データが使えるものか確認するために、データ量、外れ値や欠損値の割合、相関関係等の基礎分析を実施する。
- (3) 学習データの準備として、学習に適したデータに変換する。データクレンジングによって欠損値の補完や修正を実施し、データの品質を高める。
- (4) 予測モデルを作成する。
- (5) 検証評価によって改善点を見極めて、繰り返し改善を図る。

### 3.3 目的変数と説明変数の設定

予測対象になる目的変数と、目的変数の特徴づける説明変数が必要である。今回構築した予測モデルでは、次のように目的変数と説明変数を設定した。

#### (1) 目的変数

コスト悪化プロジェクトの検知を目的としているため、目的変数はプロジェクトがコスト計画を達成したか否かとした(表1)。品質・工程を指標に設定する方法も考えられるが、品質悪化や工程遅延は最終的にコスト悪化につながるため、コスト計画を達成したか否かだけで判断した。

表 1. 目的変数の定義

データ定義	ラベル
コストが計画値を超過	計画未達(コスト悪化)
コストが計画値の範囲内	計画達成

表 2. 説明変数

データ分類	データ項目
コスト推移	コスト実績, コスト計画
プロジェクト期間	開始日～稼働日(プロジェクト完了)までの期間
プロジェクト特性	管理レベル(プロジェクトの規模), 工事種別(システム開発, システム基盤構築等), 顧客業種など

(2) 説明変数

プロジェクトデータから、比較検証した中で最も精度が高かったデータ項目を説明変数とした(表2)。

3.4 予測モデルの構築

機械学習のライブラリである“scikit-learn<sup>(2)</sup>”とコスト推移の将来予測を行う独自方式を組み合わせ、コスト推移とプロジェクト特性に基づいて予測する予測モデルを構築した(図1)。

3.4.1 類似パターンに分類

コスト推移とプロジェクト特性に基づく予測を行うため、コスト推移を類似パターンのグループに分類し、その分類結果とプロジェクト特性を学習させる。

コスト推移である時系列データの分類には、クラスタリング<sup>(3)</sup>を用いた。プロジェクト期間の長さによってコスト実績のデータ数が異なること、コスト推移が変動する時期が重要であることから、クラスタリングを実施する前にデータ変換を実施した。データ変換では、各プロジェクト

でのコスト実績のデータ数を同一のデータ数になるようにした。また、クラスタリングでのデータ間の距離測定は比較検証した手法の中で最も精度が高かったコサイン類似度を採用した。

3.4.2 類似パターンの推定

予測では進行中プロジェクトのコスト推移と類似するパターンを推定する。プロジェクトが進行するにつれてコスト推移の波形は変化していくため、予測をする直前の時点の波形とプロジェクト特性を基に、波形の形状が類似するグループを推定する。グループの推定には、機械学習を用いた。複数の機械学習手法を比較検証し、最も精度の高かった勾配ブースティング<sup>(4)(5)</sup>を採用した。

3.4.3 類似パターンに基づく予測

推定されたグループ内からコスト推移の形状が最も類似するプロジェクトデータを更に変更する。選択する件数は、予測精度を向上させるために、予測対象になるプロジェクトの進行状況によって変化させる。プロジェクトの進行序盤では選択する件数を多くし、プロジェクトの終盤では少ない件数を選択する。

取得したプロジェクトのコスト推移を基に、予測対象になるプロジェクトとの近傍データ(距離に近いデータ)を取得する。取得したコスト推移の平均値を利用して、将来のコスト推移を予測する。

現在から最終的なコスト見込みを予測するため、コスト計画からの逸脱度合いが高いプロジェクトから優先的に点検する等の活用が可能になる。

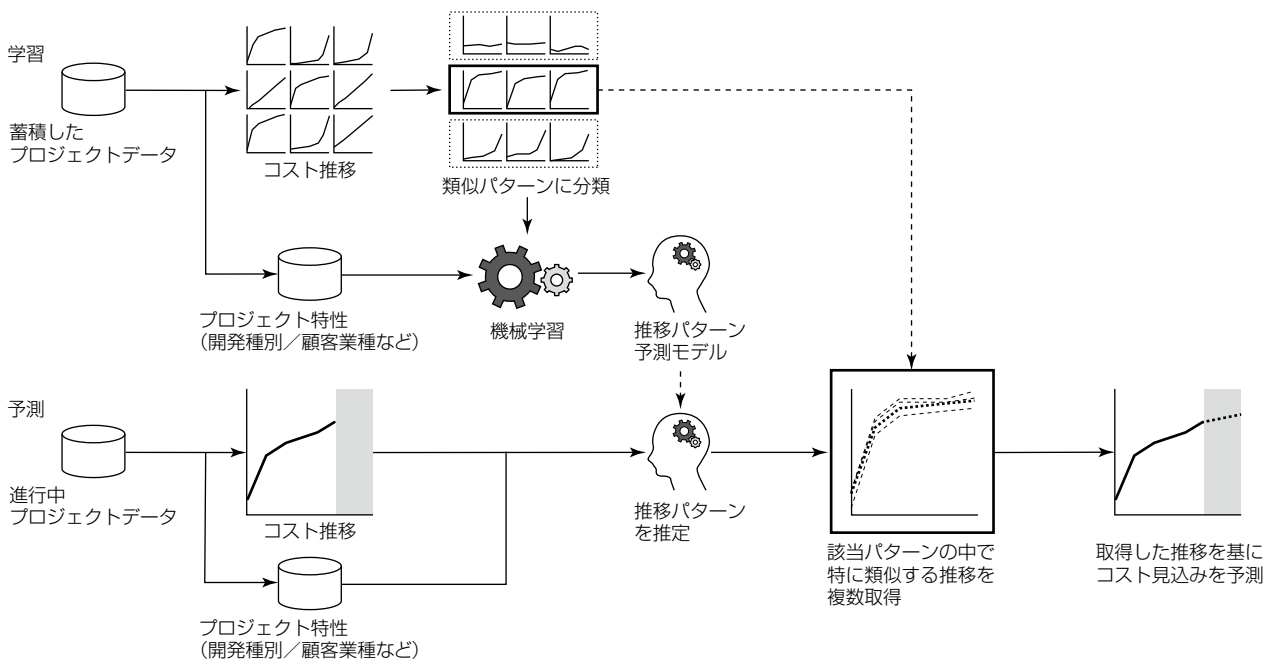


図 1. コスト推移とプロジェクト特性に応じた学習と予測

## 4. 有効性検証

### 4.1 検証評価

検証では過去のプロジェクトデータを学習用データ80%、検証用データ20%に分割し、検証用データを用いて予測モデルの汎用性を確認した。

検証データは、コスト悪化する前に検知できたかを確認するために、悪化する1時点前(≒1か月前)、3時点前(≒3か月前)までのデータを準備した。なお、コストが計画内に収まったプロジェクトは、完了時点前までのデータを準備した。これらのデータを用いて、1時点先予測、3時点先予測の“再現率”と“適合率”を測定した。

検証の結果、予測精度は表3のとおりである。予測する先が近くなるほど精度が高くなることが確認できる。また、適合率は、1時点先予測に比べて、3時点先予測で大きく下がっている。この結果は、3時点前に悪化の予兆をつかみ対策を図ったことで計画達成になったプロジェクトが含まれているため、精度を下げたと推察する。

また、コスト悪化の度合いには幅があり、経営への影響度合いが異なってくる。そこで、直近5年でコスト悪化額が高かったプロジェクトに対して確認した結果、再現率は75.0%になり一定の効果があることを確認した。

### 4.2 試行評価

構築した予測モデルがプロジェクト点検で有効か評価するために、PMOによるプロジェクト点検で試行した。試行の結果、PMOがアラートを上げた悪化予兆があるプロジェクトのうち、80%は予測モデルもアラートを上げており、一定の効果を確認できた。ただし、プロジェクト点検に活用する上での課題があり、その内容を5章で述べる。

## 5. 課題と対策

### 5.1 予測結果に対する説明性

機械学習の性質上、一般的に予測結果を導き出す過程はブラックボックスになるため、プロジェクト点検から対策につなげるのに困る場面が確認された。判断を伴う業務では予測結果に対する説明性が重要になってくる。

対策として、二つのアプローチで取組みを進めている。一つ目は予測モデルを活用したプロジェクト点検でAIが

表3. 予測精度

データ定義	1時点先予測	3時点先予測
再現率(%)	72.1	65.9
適合率(%)	67.0	43.5

悪化と予測した場合の点検観点を蓄積し、ガイド化する。二つ目は、予測結果に対する解釈性を高めるために、予測を導く上で重視した項目を判別できるようにした予測モデルを統計手法で構築し、今回構築した予測モデルと並行して運用する。

### 5.2 上流フェーズの予兆検知

予測モデルでは、時系列でのコスト推移を用いて予測するため、コスト推移の実績が少ない状況では適切に予測ができない。

対策として、上流フェーズでの予兆検知を検討する。上流フェーズでは不確定要素が多くリスクの潜在率が高い。リスクの顕在化はプロジェクト目標の達成に影響を与えて、コスト悪化につながる可能性があるため、リスクのキャッチアップに着目する。PMOではプロジェクトの概要・状況等の報告文書をリスクキャッチアップの情報の一つにしている。そこで、上流フェーズの予兆検知として、報告文書等からリスクをキャッチアップすることでコスト悪化の予兆を検知する新たな予測モデルを検討する。

## 6. むすび

プロジェクト点検業務にAIを適用することで、コスト悪化の予兆検知に一定の効果があることが確認できた。ただし、定常的なプロジェクト点検で使用するには“予測結果に対する説明性”や“上流フェーズの予兆検知等の予測精度向上”に課題がある。スキルに依存しない網羅的な点検・対策を実施するために、説明性の確保や予測精度の向上に向けて、活動を継続していく。

### 参考文献

- (1) 独立行政法人 情報処理推進機構 (IPA) : AI白書2020, 角川アスキー総合研究所 (2020)
- (2) scikit-learn Machine Learning in Python  
<https://scikit-learn.org/stable/>
- (3) 宮本定明: クラスタ分析入門, 森北出版 (1999)
- (4) Chen, T., et al.: Xgboost: A scalable tree boosting system, Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785~794 (2016)
- (5) Ke, Guolin., et al.: Lightgbm: A highly efficient gradient boosting decision tree, Advances in neural information processing systems, 30, 3146~3154 (2017)