ダイヤ乱れ時を対象とした列車混雑予測手法

信号技術研究部　運転システム研究室
副主任研究員　中挾 晃介

背景・目的

- 現状、各事業者は混雑情報を提供しているが、多くは平常時が対象。
- ダイヤ乱れ時に、利用列車変更の判断ができる情報は未提供。

- 目的：ダイヤ乱れ時を対象とした混雑予測モデルを複数の手法を用いて構築。
  各モデルの得意不得意（精度特性）を把握。→状況毎の使い分け方を検討。

本日の発表

- 混雑予測モデルの構築
  - –混雑予測に用いる手法の選定
  - –各手法を用いたモデルの入出力データの検討
- 精度特性分析
- まとめ

混雑予測に用いる手法の選定

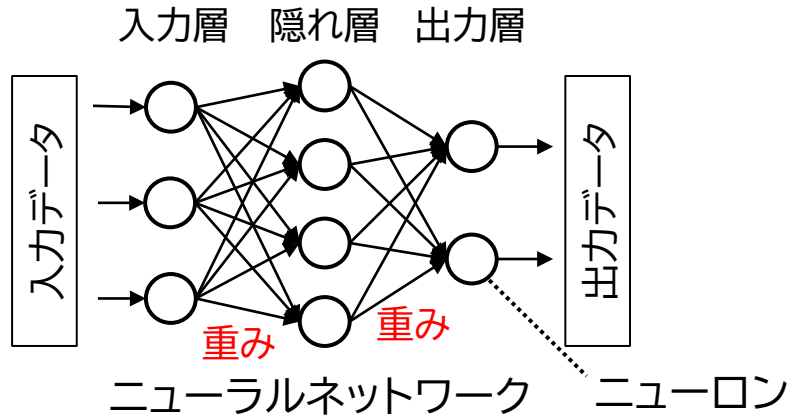
- ダイヤ乱れ時に列車単位で混雑を予測する先行研究は無し。
- 平常時も含め、混雑予測に適用されている予測手法を、予測対象や利用データ等の観点で整理。

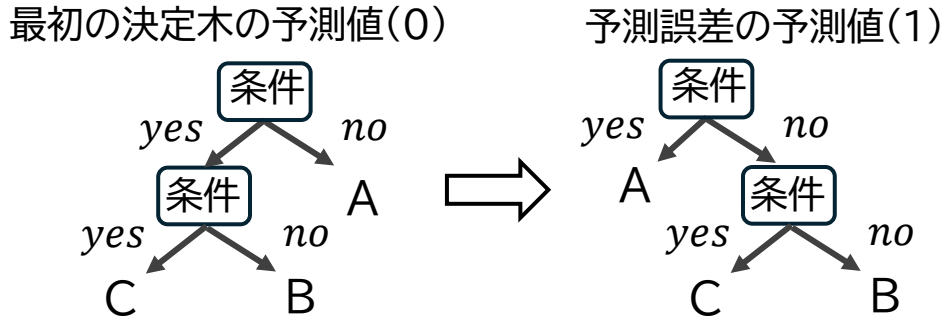
→ダイヤ乱れ時にも適用可能な、NN、LightGBM、SVM、ARIMAX（ARIMAの拡張モデル）、重回帰分析を候補に選定。

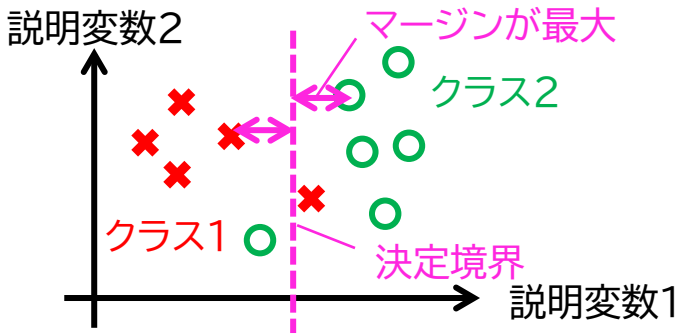
予測手法の種類	予測手法	予測場面	予測対象	利用データ
機械学習手法	ニューラルネットワーク（NN）	平常時	乗車率	応荷重，運行実績
	LightGBM	平常時	混雑区分	実績乗車率，運行実績，気象，暦など
		乱れ時	断面輸送量	応荷重，運行実績
	サポートベクターマシン（SVM）	平常時	乗車率	実績乗車率，運行実績，列車番号，駅名，日付など
時系列データ分析手法	ARIMA	平常時	乗車率	計画ダイヤ，座席販売実績
統計的手法	線形回帰	平常時	乗車率	応荷重

混雑予測に用いる手法(機械学習手法)

- NN（Neural Network）
  - 脳の神経回路網をモデル化。
  - 出力誤差が小さくなるよう、ニューロン間のエッジの重みを調整。
- LightGBM
  - 決定木をベースとした手法。
  - 「予測誤差を予測する決定木」の作成を繰り返す。
- SVM（Support Vector Machine）
  - 異なるクラスのデータ間に境界線を設定。
  - データがどのクラスに属するかを予測。







混雑予測に用いる手法(機械学習以外の手法)

- ARIMAX
  - 自己回帰和分移動平均モデル(ARIMA)に外生変数を組み込むよう拡張。
  - 外生変数:予測に影響を与えうる外部要因。

$$\Delta(y_t) = \underbrace{y_t - y_{t-1}}_{\text{階差時系列}}, \quad y_t = c + \underbrace{\beta x_t}_{\text{外生変数}} + \underbrace{\sum_{s=1}^{p} \phi_s y_{t-s}}_{\text{自己回帰モデル}} + \underbrace{\varepsilon_t + \sum_{s=1}^{q} \theta_s \varepsilon_{t-s}}_{\text{移動平均モデル}}$$

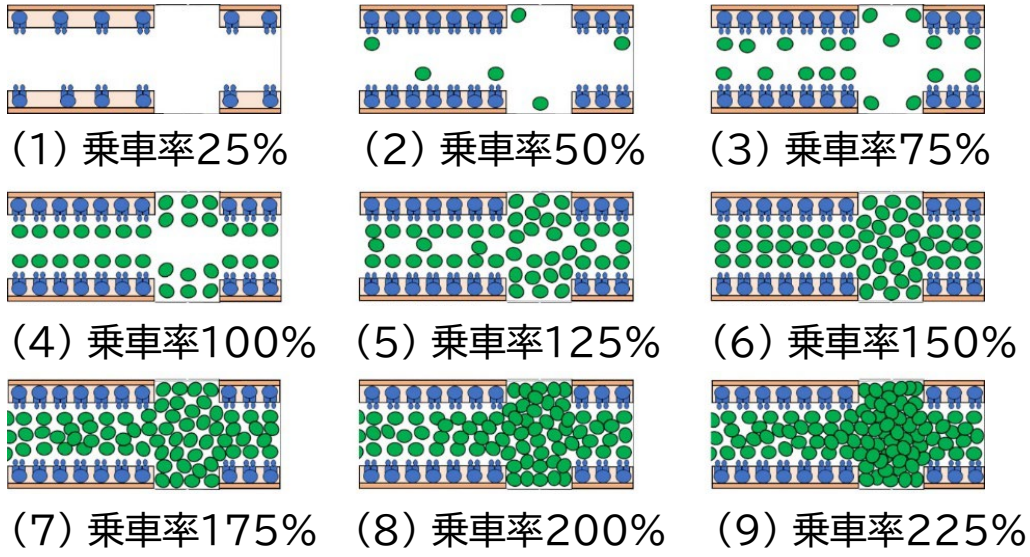
- 重回帰分析
  - 複数の説明変数から目的変数を予測する回帰分析手法。

各手法を用いたモデル構築の方針

- NN
  - 平常時が対象の乗車率（連続値）予測モデルを既構築。
- LightGBM
  - 分類問題を高精度に予測。
- SVM
  - 分類問題を高精度に予測。
- ARIMAX
  - 時系列データの予測に特化。
- 重回帰分析
  - 連続的な数値データの予測に特化。

- 乗車率（連続値）予測モデルを構築。

- 乗車率25%区切りの9段階の混雑区分を予測する、区分予測モデルを構築。



- 乗車率（連続値）の時系列変化の予測モデルを構築。

モデルの入出力データの検討（1/2）

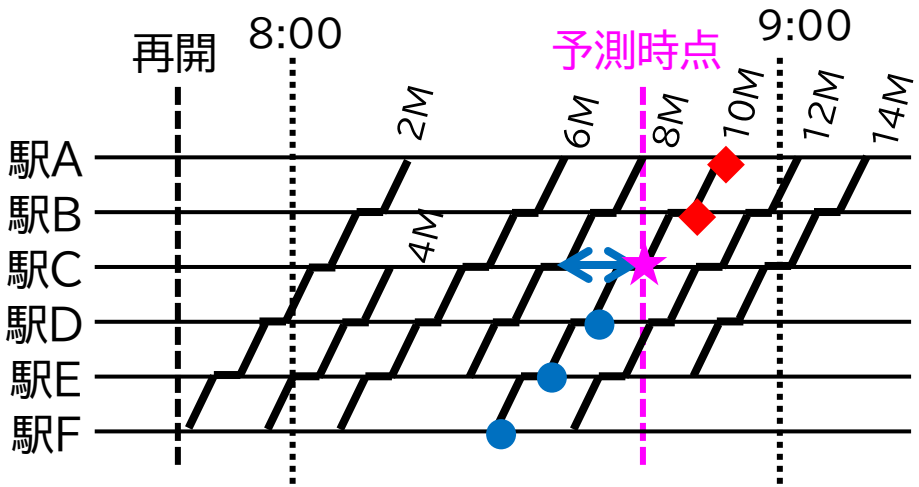
- NN、重回帰分析

複数の組合せから、少ない種類数で高精度に予測できる入出力を選定。

– 入力データ
  - ● 自列車の乗車率、
  - ⇔ 前列車との時間間隔

– 出力データ
  - ◆ 終着駅までの乗車率と現時点の乗車率の増減量

乗車率、混雑区分は、編成内の平均値

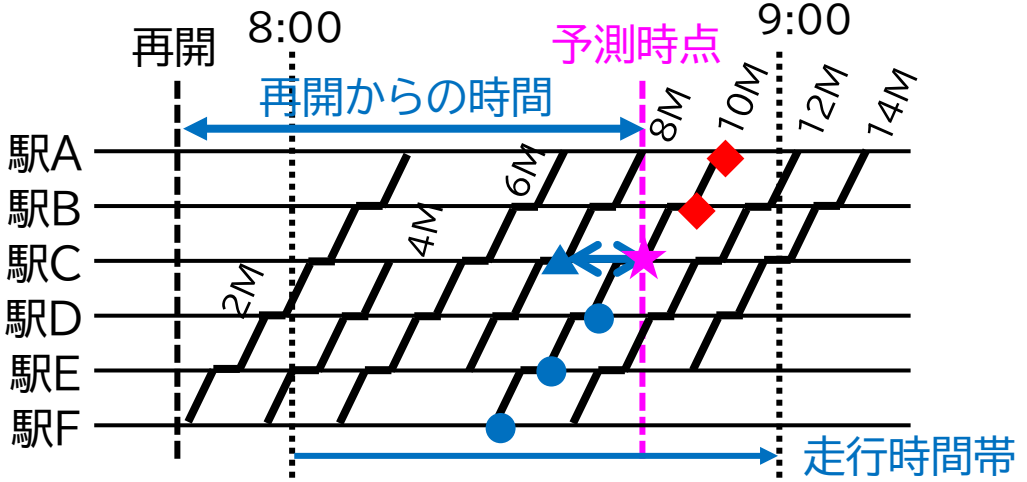


- LightGBM、SVM

「再開からの時間」のような、ダイヤ乱れ時特有の、混雑に影響する特徴量を含める。

– 入力データ
  - ● 自列車の混雑区分
  - ▲ 前列車の混雑区分
  - ⇔ 前列車との発発時隔
  - ➡ 走行時間帯
  - ↔ 再開からの時間　など

– 出力データ
  - ◆ 終着駅までの9段階の混雑区分



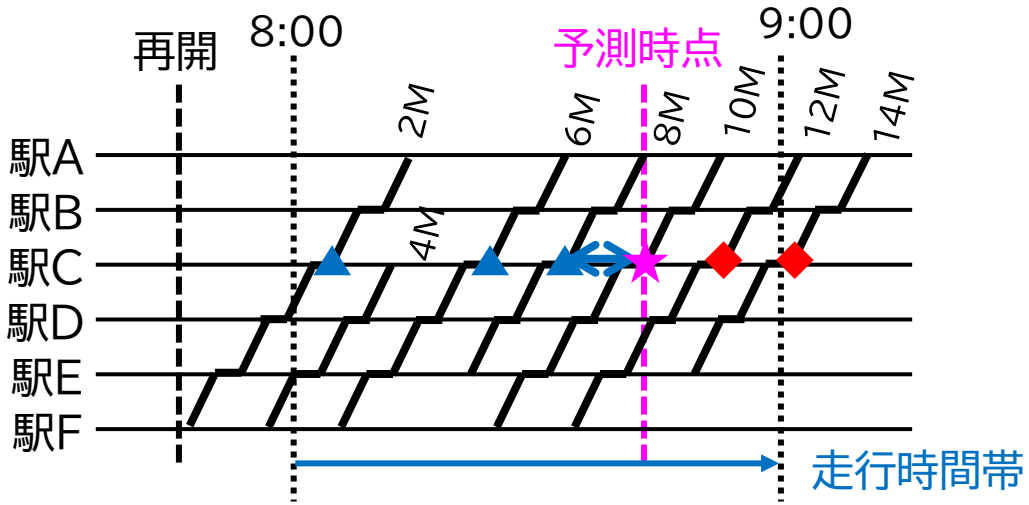
モデルの入出力データの検討（2/2）

- ARIMAX

ある駅を順に発車する列車の混雑（下図の▲, ★, ◆）を時系列データと捉えてモデル構築。

– 入力データ
  - ▲ 数列車前から自列車までの乗車率
  - ↔ 前列車との時間間隔（**外生変数**）
  - → 走行時間帯（**外生変数**）

– 出力データ
  - ◆ 最終列車までの乗車率



本日の発表

- 混雑予測モデルの構築
  - –混雑予測に用いる手法の選定
  - –各手法を用いたモデルの入出力データの検討

- 精度特性分析

- まとめ

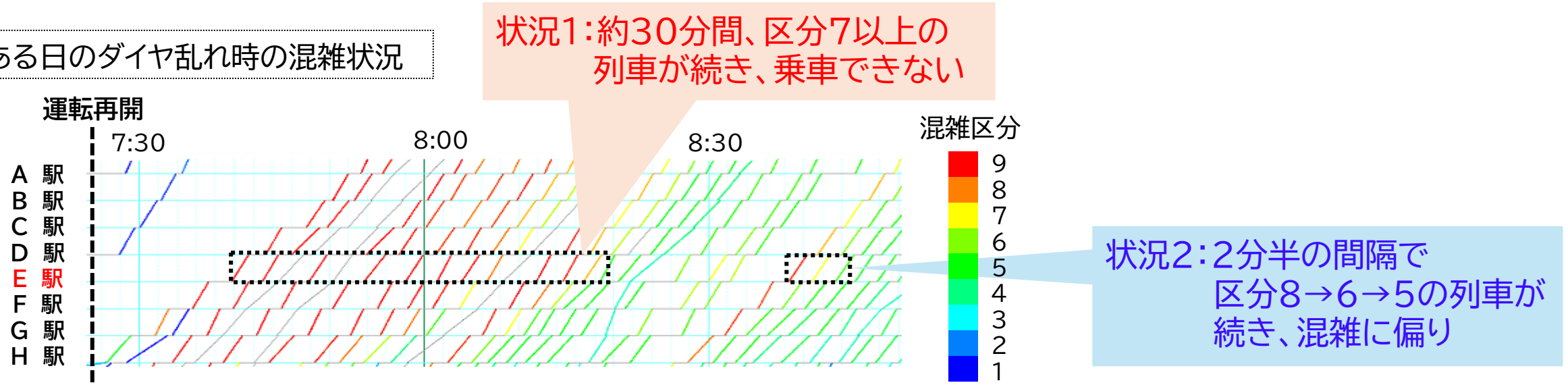
精度特性分析の概要

- 目的
  - ダイヤ乱れ時に特有の状況を対象に、複数の予測断面（走行時間帯、予測対象駅、再開からの時間など）毎に精度を確認し、各モデルの精度特性を把握。
- 検証対象路線
  - 大都市近郊の全24駅、終日の列車本数が600本程度（平日）の通勤路線。
- 学習・予測データ
  - ダイヤ乱れが発生した平日30日分のデータ。
- 検証方法
  - 学習・予測データの内、29日分を学習、1日分を予測に用いる。これを30日分繰り返す。（leave-one-out交差検証）
  - NN、ARIMAX、重回帰のモデルは、出力した乗車率を9段階の混雑区分に変換。

精度特性分析で対象とする混雑状況

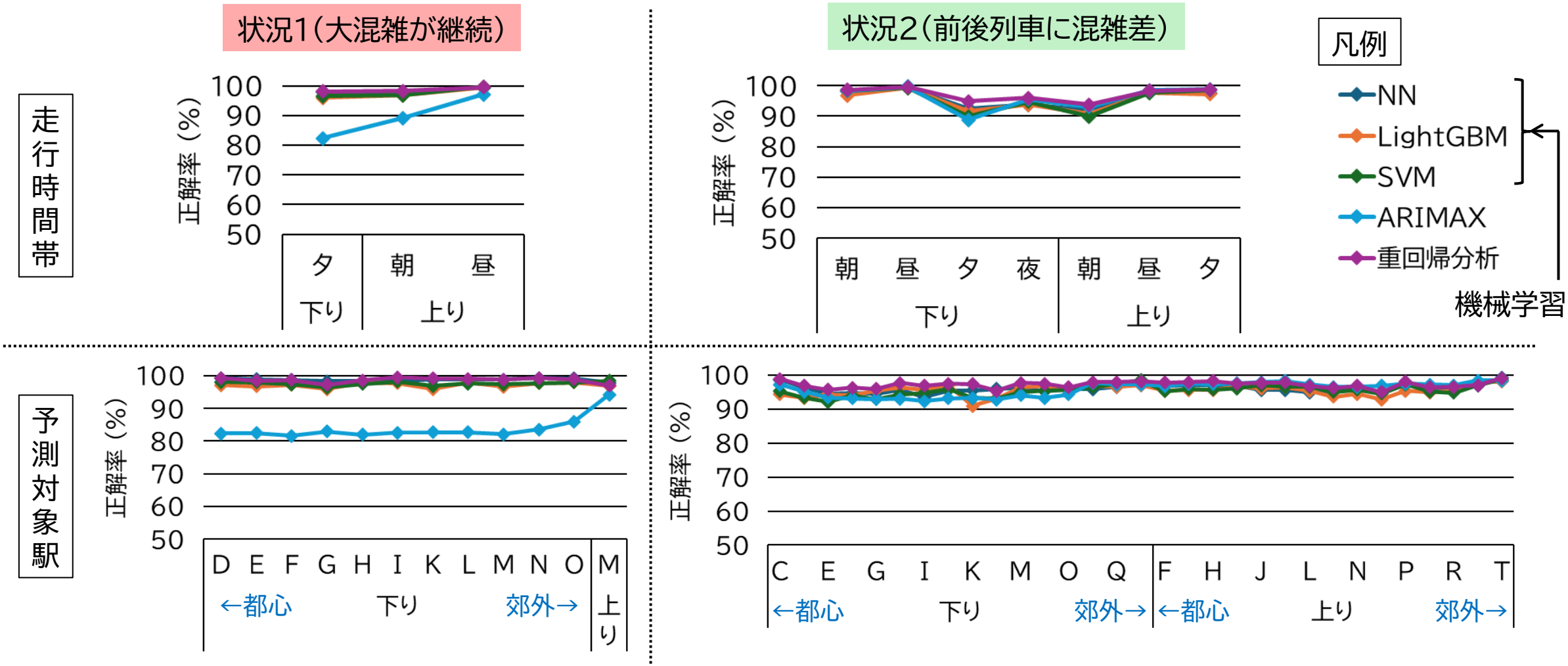
- ダイヤ乱れ時に実際に発生している混雑状況から分析対象を検討。

ある日のダイヤ乱れ時の混雑状況



- 状況1：ある駅で、20分以上の間、発車する全列車の混雑区分が7以上。
- 状況2：ある駅で、後続列車の混雑区分が6以上、さらに次の後続が5以下。
  →2つの状況について、走行時間帯や予測対象駅など複数の予測断面で比較・分析。

予測状況・予測断面毎の精度検証



- ARIMAXは精度が劣るが、他モデルの得意不得意の傾向無し。→評価指標を変更

評価指標毎の精度検証

- 正解率とは異なる評価指標を用いて、予測断面毎に精度を確認し、
  各モデルの精度特性の把握を試みる。

「大混雑が継続するか否か」を予測する場合の結果の組み合わせ

	大混雑が継続すると予測	大混雑が継続しないと予測
実際は大混雑が継続する	大混雑が継続すると予測し、正解した件数（*TP*）	大混雑は継続しないと予測し、間違えた件数（*FN*）
実際は大混雑が継続していない	大混雑が継続すると予測し、間違えた件数（*FP*）	大混雑は継続しないと予測し、正解した件数（*TN*）

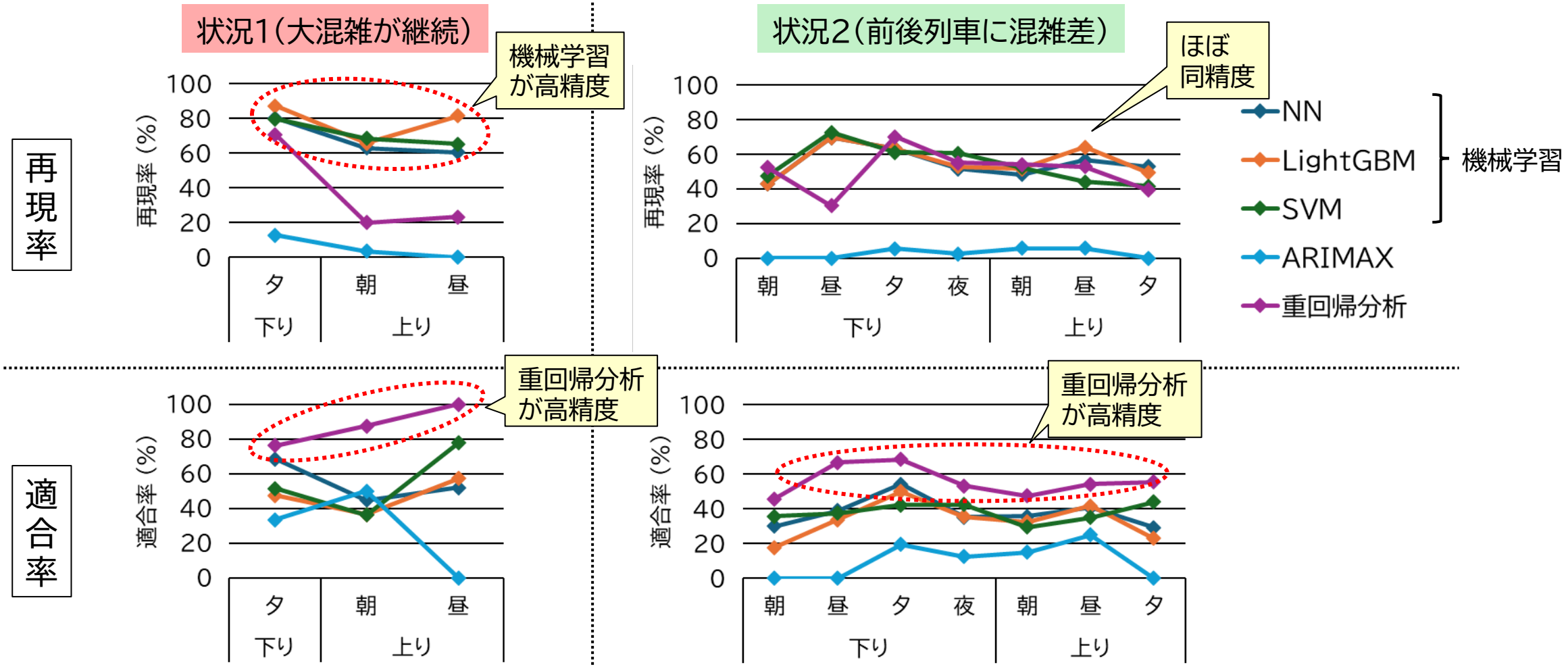
再現率：$TP/(TP+FN)$

適合率：$TP/(TP+FP)$

※正解率：$(TP+TN)/(TP+FN+FP+TN)$

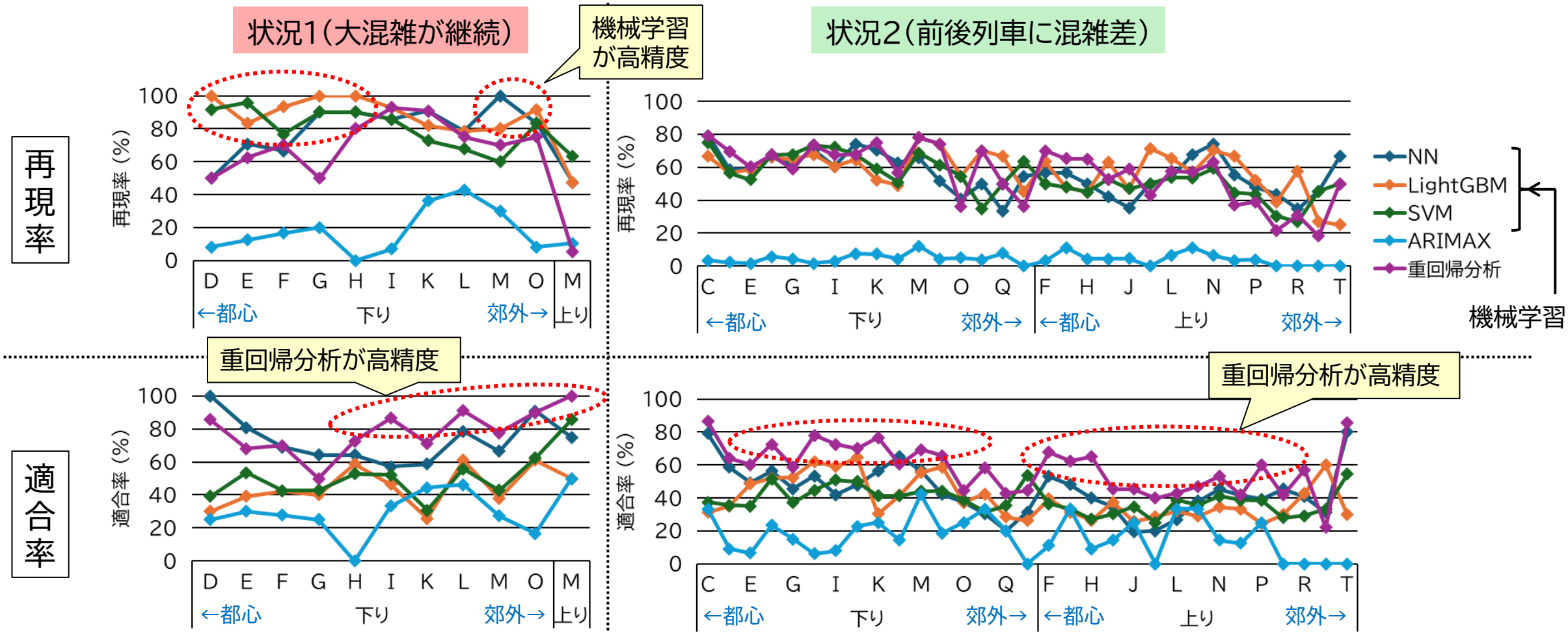
再現率を重視→多少、大混雑が継続と誤予測しても、大混雑の継続を見逃さない。
適合率を重視→多少、大混雑の継続を見逃しても、大混雑が継続と誤予測しない。

走行時間帯毎の精度検証結果



再現率は機械学習、適合率は重回帰分析の精度が高い傾向を確認。

予測対象駅毎の精度検証結果



（走行時間帯毎と同様、）

再現率は機械学習、適合率は重回帰分析の精度が高い傾向を確認。

本日の発表

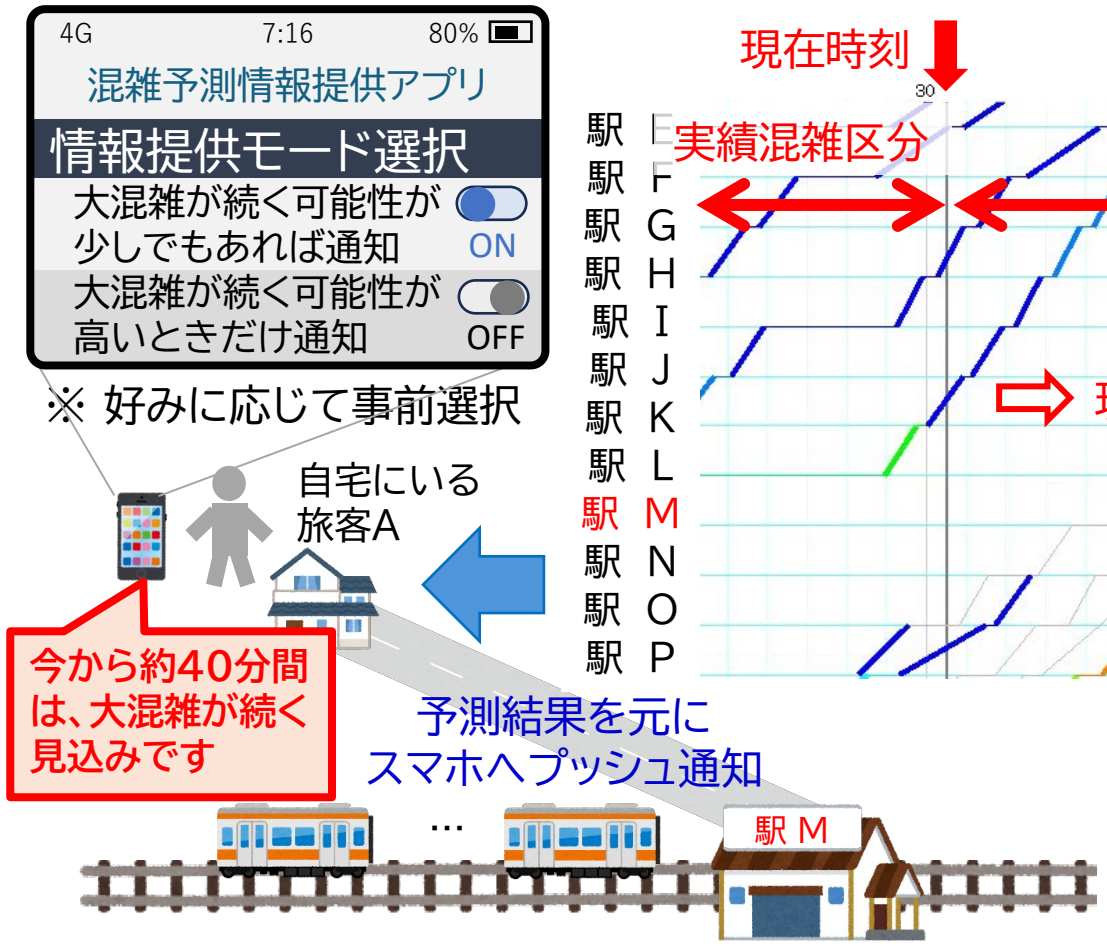
- 混雑予測モデルの構築
  - –混雑予測に用いる手法の選定
  - –各手法を用いたモデルの入出力データの検討
- 精度特性分析
- まとめ

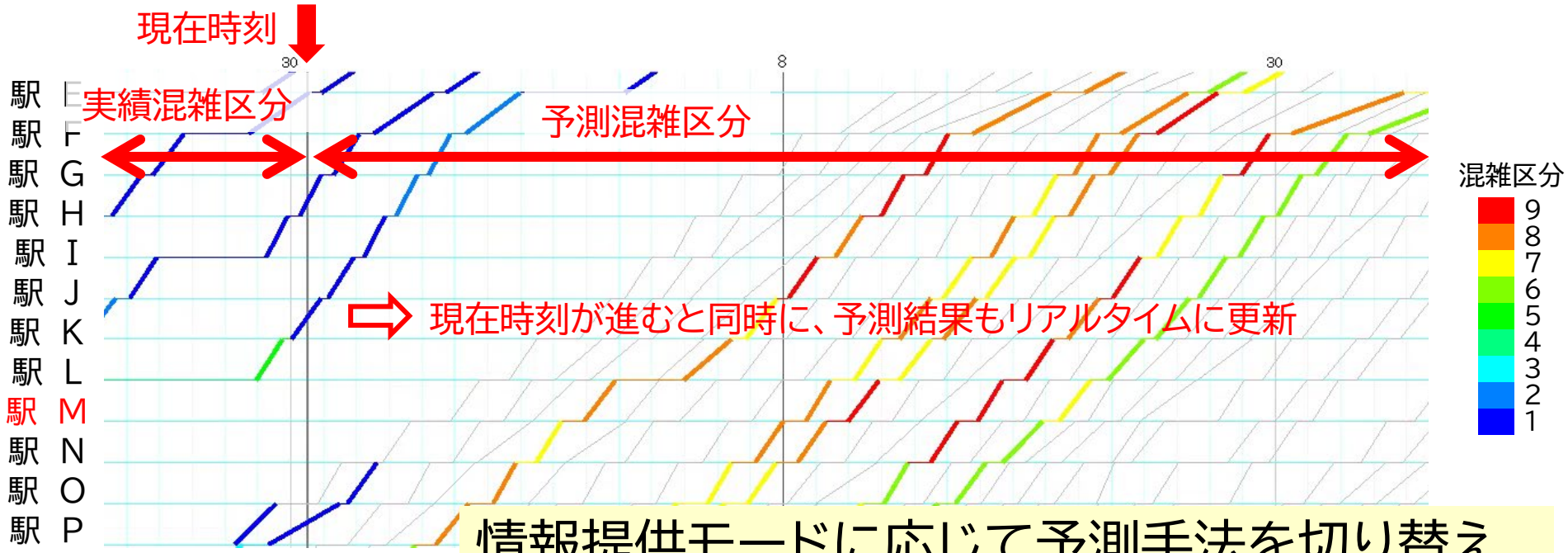
まとめ

- ダイヤ乱れ時の混雑を予測する5種類のモデルを構築。
- ダイヤ乱れ時に特有の2状況を対象に、各モデルの精度特性を把握。
  →機械学習は再現率、重回帰分析は適合率が高い傾向を確認。
- 得られた精度特性を踏まえ、実用性のある混雑予測手法の構築を検討。
  - 例）混雑が続く可能性があることを知りたい個人→機械学習の予測結果
    誤案内を許容しない個人や集団の旅客→重回帰分析の予測結果

成果の活用

- 開発成果は、旅客への混雑予測情報を提供するWebアプリ等の開発のための要素技術として活用できる可能性あり。





情報提供モードに応じて予測手法を切り替え

- 混雑が続く可能性を知りたい→機械学習
- 誤案内を許容しない→重回帰分析