

教育データ解析チャレンジコンテストにおける 教育データ解析事例の紹介

第35回 大学等におけるオンライン教育とデジタル変革に関する
サイバーシンポジウム「教育機関DXシンポ」

丸山 翔平

概要

教育データ解析技術を社会全体で向上させることを目的とした「教育データ解析チャレンジコンテスト」の、優勝者の解析方法と解析結果を共有する。

目的

「教育データ解析チャレンジコンテスト」で開発した教育データの解析方法と解析結果を共有する。

背景

“内閣府・戦略的イノベーション創造プログラム（SIP）第2期ビッグデータ・AIを活用したサイバー空間基盤技術（管理法人：NEDO）に採択された「エビデンスに基づくテラーメイド教育の研究開発」の研究では、教育・学習支援システムを利用したログデータの分析を実施しております。そこで今回、デジタル教材の閲覧履歴データを用いて成績を予測し、その精度を競う「教育データ解析チャレンジコンテスト」を開催することに致しました。これは、エビデンスに基づくテラーメイド教育の実現を目指して、教育データの分析技術を、コンペティションを通して社会全体で向上していくことを目的としています。”
(<https://sites.google.com/view/sipdatachallenge/> より抜粋して引用)

教育データ解析チャレンジコンテストの概要

デジタル教材の学習の途中経過から、学生の最終成績を予測する。
学習行動データには、学生がデジタル教材上で行った操作の履歴などが含まれている。

お題

- コース開始から5週経過時点、10週経過時点、15週経過時点のデジタル教材の学習行動データから、学生の最終成績を予測する。

背景

- 単位を落としそうな学生を早期に発見できれば、成績向上に向けた介入ができる。

成績予測に使えるデータ (学習行動データの中身)

- 操作日時
- 学生ID
- 教材ID、ページ番号
- 操作の種類 (教材を開いた、次のページに移動した、など)、マーカー (重要 or 難しい)、メモの長さ
- デバイスの種類

方法 | 説明変数の作成

成績を左右する要因として「学生に起因する要因」と「教材に起因する要因」の2種類の仮説を立て、それらの仮説に対応する変数を作成。

成績を左右する要因の仮説と、仮説に対応する変数

#	仮説	仮説に対応する変数
1	成績上位者に特有の操作パターンがある。 例) 下位者は一方向にページを送り続けるが、上位者は頻繁にページをジャンプする、など。	<ul style="list-style-type: none">• 操作の種類別のn-gram (1-gram、2-gram、3-gram) 別の操作数• マーカーの種類別の操作数• メモの長さ• 操作時間、操作別の操作時間• デバイスの種類別の操作数• 時間別の操作数、曜日別の操作数、週別の操作数
2	好成績を収めるうえで重要な教材やページがある。 例) 易しい教材がある、テストに出る内容が書かれているページがある、など。	<ul style="list-style-type: none">• 各教材の閲覧フラグ• 教材別の操作数• 教材別・ページ番号別の操作数• 教材別・ページ番号別の滞在時間

※1 受講した教材の数の影響を除くため、まず教材別に操作数をカウントし、その値を教材で集約した。

※2 集約は min、median、max、mean、std、sum を用いた。

方法 | その他の工夫

「途中経過から最終成績を予測する」というお題と整合させるため、与えられた全期間の学習行動データを打ち切り、途中経過のデータを擬似的に作成。

データ拡張

- 「途中経過から最終成績を予測する」というお題に対し、与えられたデータは開講中の全期間の学習行動データだった。
- お題とデータを整合させるため、学習行動データを打ち切り、途中経過のデータを擬似的に作成した。

特徴選択

- 大量の変数 (約17,000個) を作成したため、すべて使うと過学習を引き起こす。
- 過学習を緩和するため、予測に対する変数の重要度の一種である null importance を計算し、null importance 上位の約650個に変数を絞った。

モデル

- テーブルデータのコンテストでよく用いられる勾配ブースティング決定木を採用した。

結果 | 予測精度

学生の最終成績 (100点満点) を平均的に±11点程度の誤差で予測できた。
→ 介入が必要な学生をスクリーニングするには十分な精度。

順位	チーム名 (ニックネーム)	スコア
1	まるちゃん	10.89
2	NTTCS	14.71
3	mi-fujita	15.38
4	chsc	15.91
5	のぞもと	16.29
6	mh	17.84
7	khiroyuki1993	21.91
8	上智大学 データサイエンス特論チームA	23.08
9	熊本大学 喜多研究室	25.25
10	mprg	25.89
11	koudou-ai	30.12

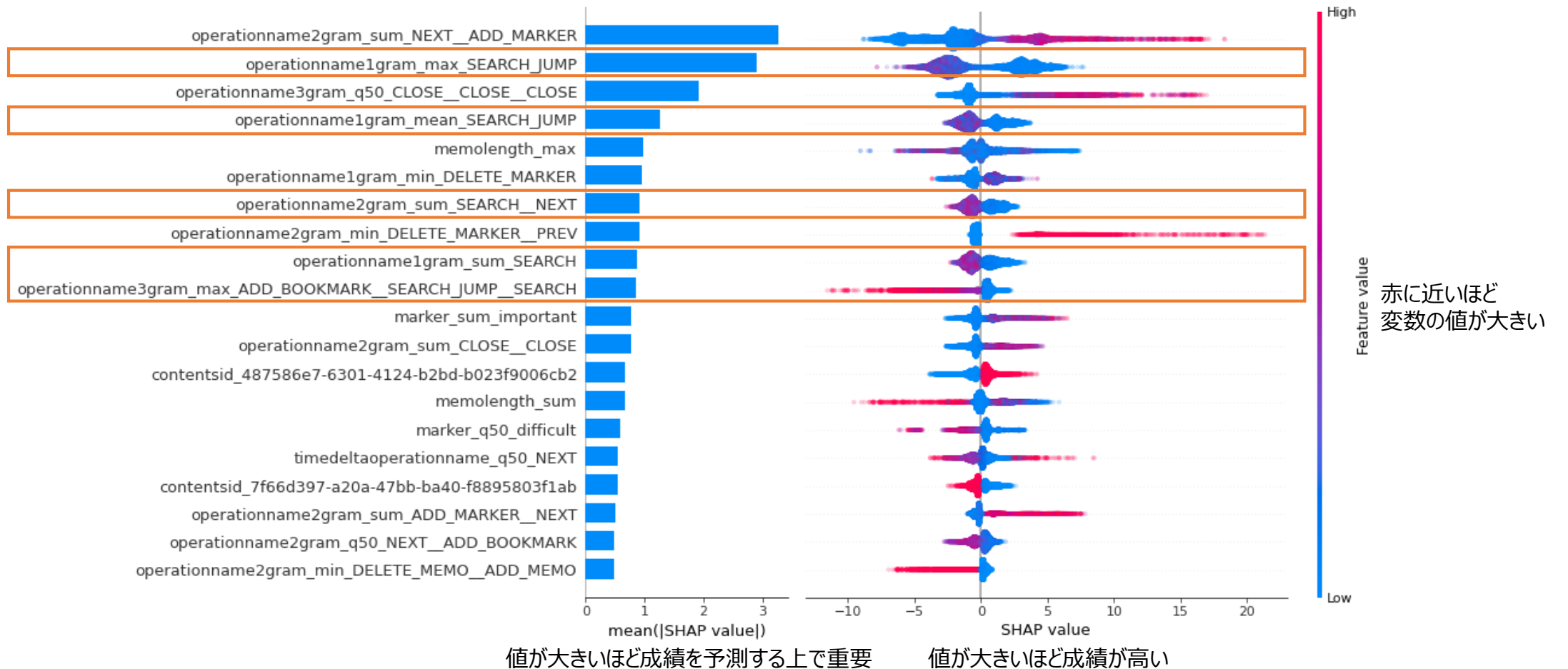
20210309 (最終成績) | [20210210](#) | [20210209](#) | [20210201](#) | [20210127](#) | [20210126](#) | [20210125](#) | [20210122](#) | [20210121](#) | [20210120](#) | [20210118](#)

(<https://sites.google.com/view/sipdatachallenge/result> より引用)

結果 | 成績予測の重要変数 - 成績が低い学生の特徴 (1/4)

「成績が低い」と予測される学生は、検索やジャンプを繰り返している。
→ 学習意欲があり、内容を理解しようと努力しているが、理解できていない。

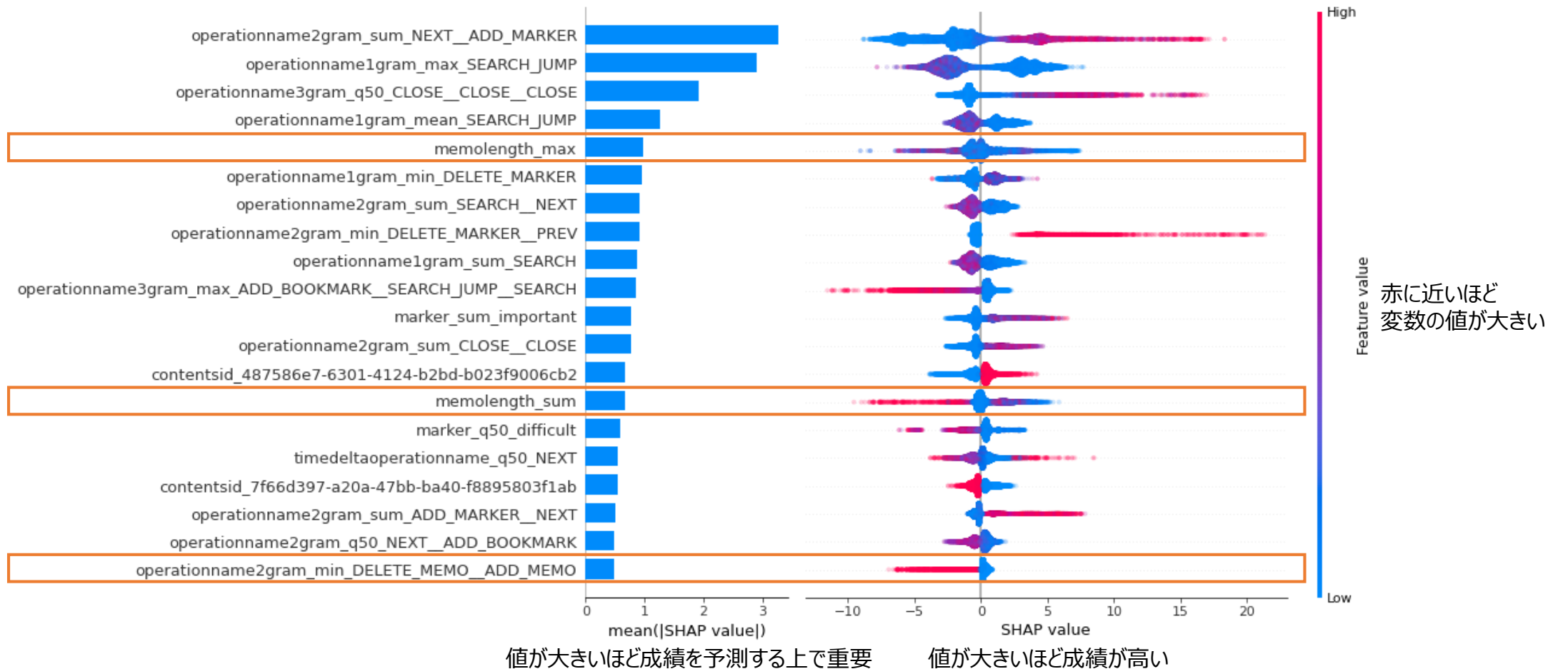
成績を予測する上で重要な変数 TOP20



結果 | 成績予測の重要変数 - 成績が低い学生の特徴 (2/4)

「成績が低い」と予測される学生は、一生懸命メモを取っている。
→ 学習意欲があり、内容を理解しようと努力しているが、理解できていない。

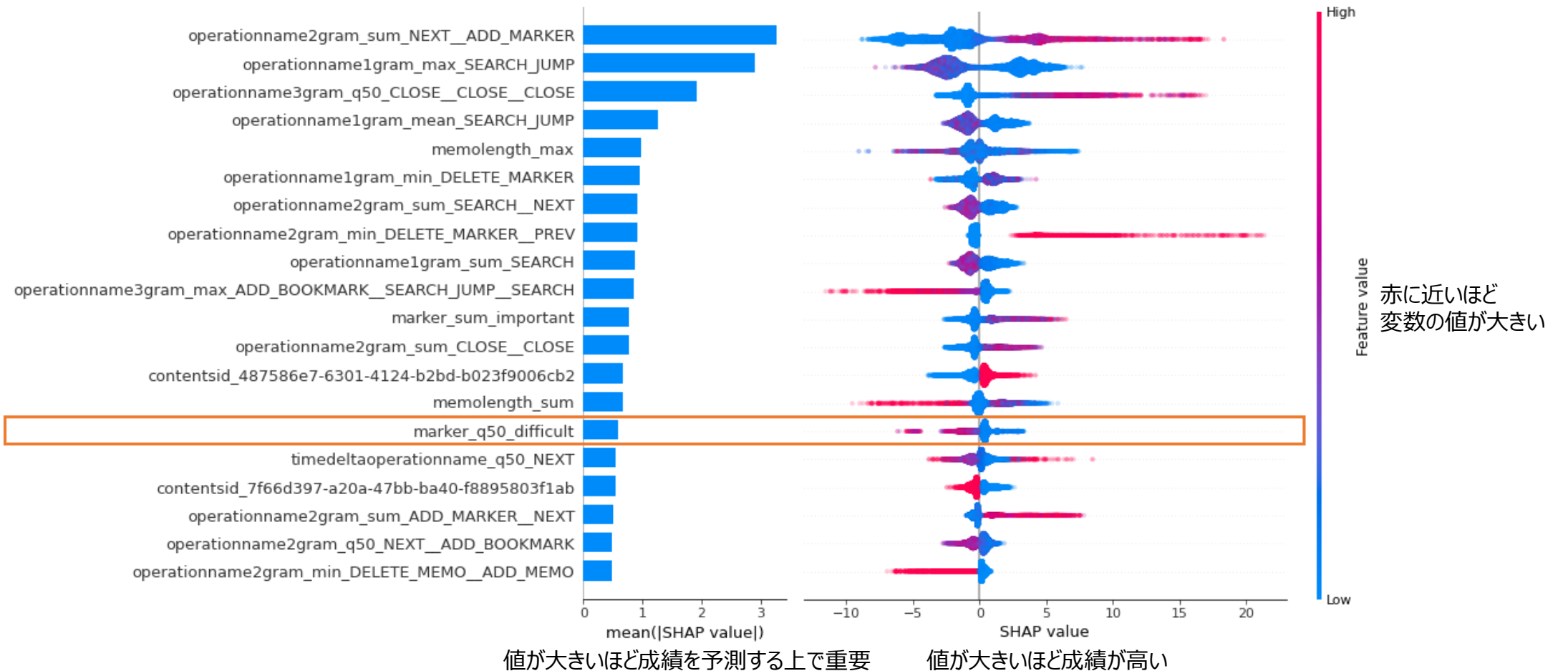
成績を予測する上で重要な変数 TOP20



結果 | 成績予測の重要変数 - 成績が低い学生の特徴 (3/4)

「成績が低い」と予測される学生は、難しいと思うページにマーカーを付けている。
→ 学習意欲はあるが、内容を理解できていない。

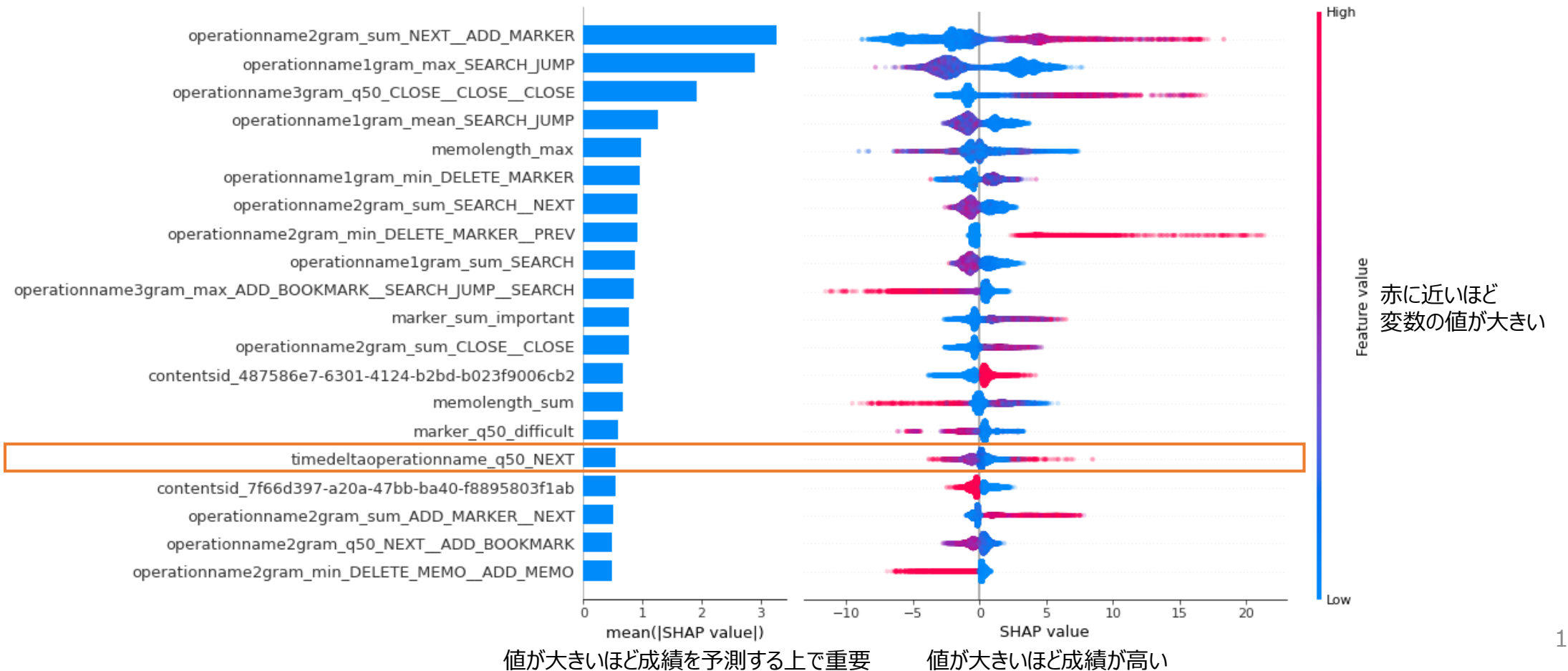
成績を予測する上で重要な変数 TOP20



結果 | 成績予測の重要変数 - 成績が低い学生の特徴 (4/4)

「成績が低い」と予測される学生は、ページを送る間隔が短い。
→ 学習意欲自体が低い。

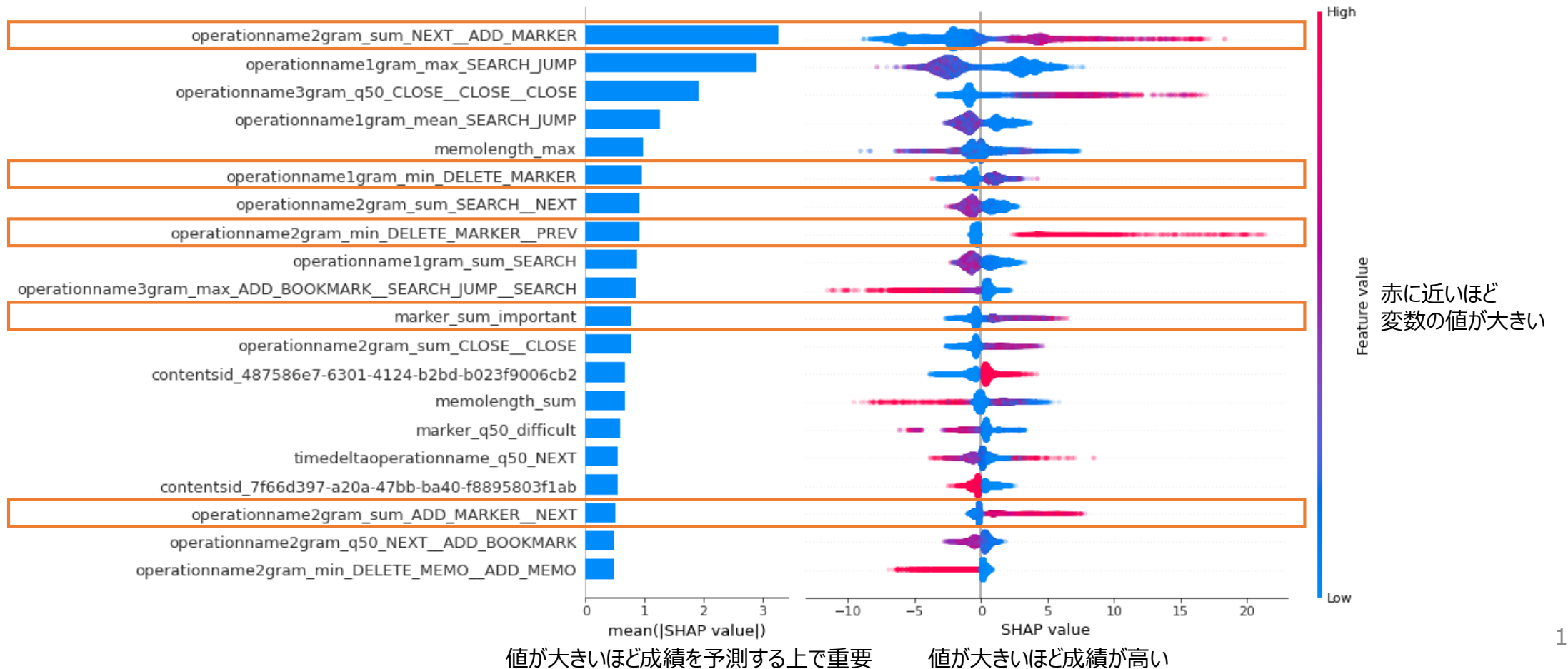
成績を予測する上で重要な変数 TOP20



結果 | 成績予測の重要変数 - 成績が高い学生の特徴 (1/2)

「成績が高い」と予測される学生は、「重要」マーカ-を活用している。
→ 内容を理解できている。

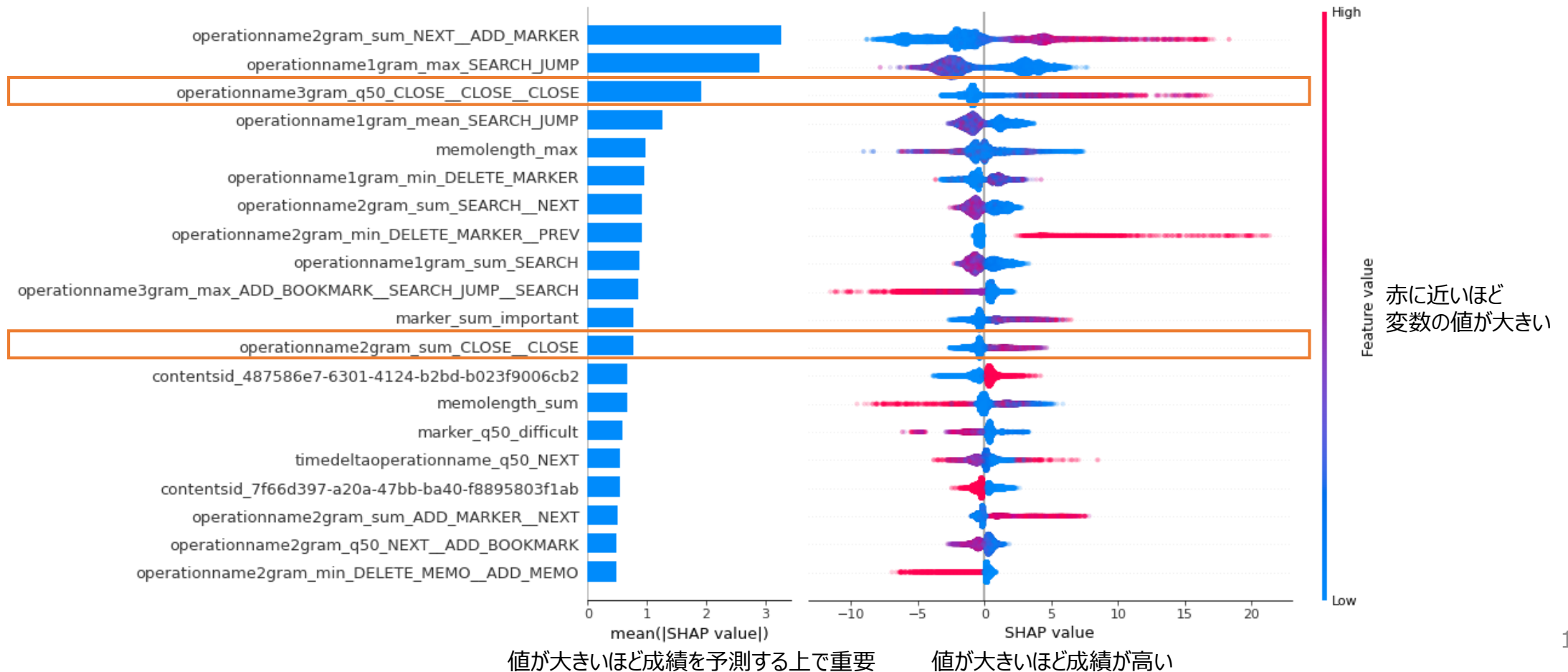
成績を予測する上で重要な変数 TOP20



結果 | 成績予測の重要変数 - 成績が高い学生の特徴 (2/2)

「成績が高い」と予測される学生は、教材をちゃんと閉じている。
→ 学習中に離脱せず、自らの意思で学習を中断・終了できている。

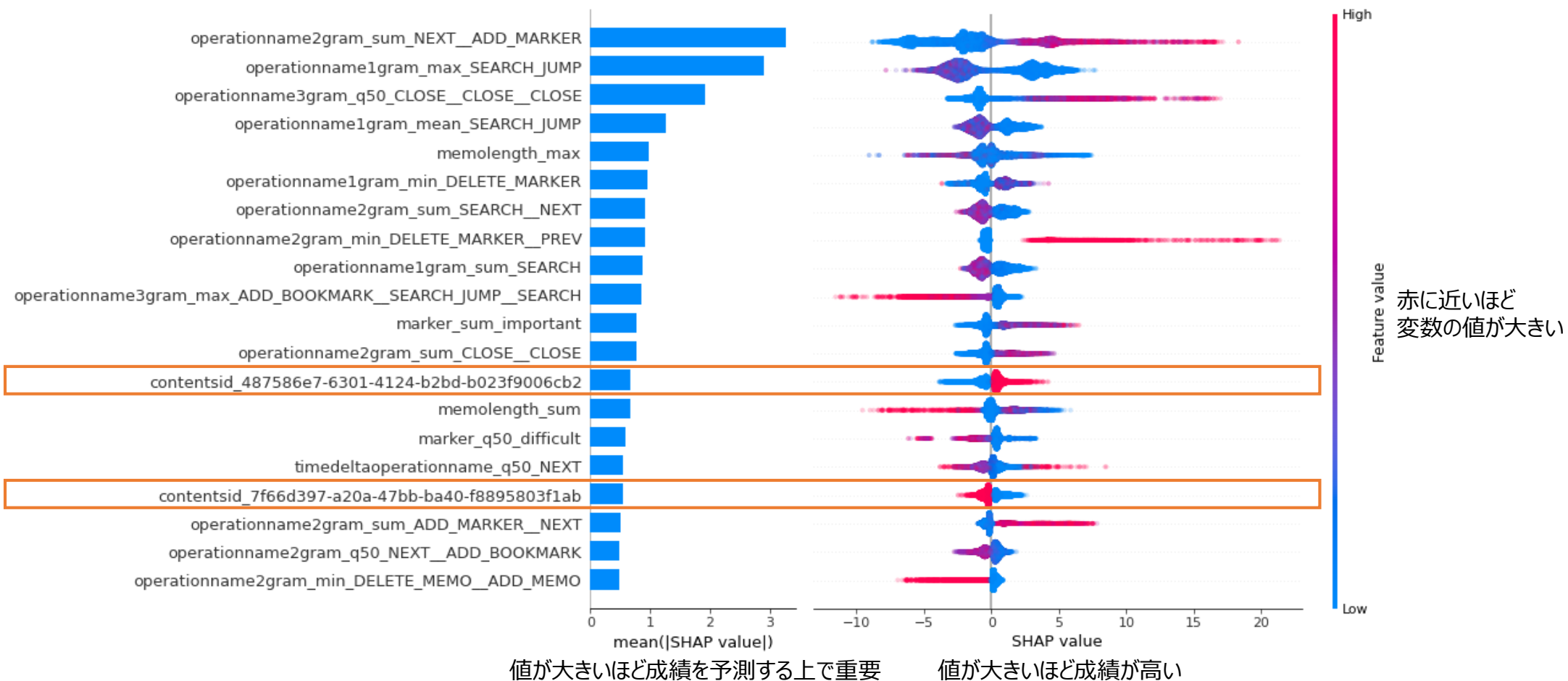
成績を予測する上で重要な変数 TOP20



結果 | 成績予測の重要変数 - 教材の特徴

成績が高くなりやすい教材、低くなりやすい教材がある。
→ 教材の難易度によっても成績が左右される。

成績を予測する上で重要な変数 TOP20



学習行動データの解析から見た課題

学習意欲と理解の有無によって学習中の行動が変化し、その結果が学習行動データに表れる。学習意欲と理解の有無に応じた適切な介入を行うことで、成績向上が望める。

学生の分類と、成績向上のための介入案

学生の分類			学習行動データの特徴	介入案
学習意欲	内容の理解	理解する努力		
ある	理解できた	-	・「重要」マーカーを付ける。	(介入の必要なし)
	理解できなかった	努力した	・検索やジャンプを繰り返す。 ・メモを残す。	・理解を助ける補足資料や 関連教材を提供する。
		あきらめた	・「難しい」マーカーを付ける。 ・途中離脱する。	・自己学習が難しいと 思われるため、個別指導を行う。
ない	-	-	・ページを送る間隔が短い。 ・途中離脱する。	・教材の内容の動機を補足する。 ・進路相談を受けてもらう。

まとめ

「教育データ解析チャレンジコンテスト」における 教育データ解析事例（学習行動データにもとづく成績予測）のご紹介

- 成績を左右する要因として「学生に起因する要因」と「教材に起因する要因」の2種類の仮説を立て、それらの仮説に対応する変数を作成することで高精度な予測を実現した。
- 成績の予測モデルを解析した結果、学習意欲と理解の有無によって学習中の行動が変化し、その結果が学習行動データに表れていることがわかった。

コンペティションを通じた社会全体での教育データ解析技術の向上

- コンペティションという形で教育データ解析を一般へ解放したことにより、機械学習の専門知識を持つ人材が参入し、教育データ解析技術の向上につながった。
- 今後も同様の取り組みを続けることで、学際的なアプローチで教育データ解析技術が発展することを期待する。