

教育用データ解析コンペティション基盤の設計と実践

馬場 雪乃¹ 高瀬 朝海² 新 恭兵² 小山 聡² 鹿島 久嗣³

¹筑波大学 ²北海道大学 ³京都大学

データ解析技術の教育は、実社会でのデータ解析利活用の促進に向けてきわめて重要である。データ解析技術の習得には、統計・機械学習の理解はもちろんのこと、多様な実データと実課題を対象にした実践演習が必要不可欠である。Kaggleに代表されるデータ解析コンペティションが、そのような演習の場として注目されている。一方、現在のデータ解析コンペティションは、主にデータ解析技術者の腕自慢の場として設計されており、教育効果を高めるためのデータ解析コンペティションの設計については十分な検討がされていない。筆者らは、教育用途のデータ解析コンペティション基盤「ビッグデータ大学」を構築し、数年に渡り運用してきた。本稿では、ビッグデータ大学で導入されている、学習者支援のアプローチを紹介する。また、それらの効果検証の結果を示し、今後の展望を述べる。

1. はじめに

データ解析技術の教育は、実社会でのデータ解析の利活用を促進するためにきわめて重要である。データ解析技術を実社会で活用できる人材には、統計・機械学習の理解はもちろんのこと、問題の背景を理解した上で適切な機械学習手法を選択・実装し、運用する技能が求められる。このような技能の習得には、座学だけでは不十分であり、多様な実データ・実課題を対象にした実践演習が必要不可欠である。

データ解析技術者が自身の実践的スキルを示す場として、データ解析コンペティションが注目を集めている。データ解析コンペティションは、参加者が同じデータを用いて予測モデル構築に取り組み、その精度を競う大会である。機械学習コミュニティでは、KDD Cupに代表される学会主催のデータ解析コンペティションが開催されており[1]、機械学習の研究開発の促進に利用されてきた。また、Netflix社は2006年から映画推薦を対象にしたデータ解析コンペティションNetflix Prize[2]を開催し、コンペティションを通じて多数のデータ解析技術者を巻き込み、映画推薦技術の向上に取り組んだ[3]。Netflix社の取り組みを汎用化したKaggle [4]やCrowdANALYTIX [5]等のコンペティション基盤も登場している。これらのコンペティション基盤を用いることで、データを保有する企業は、容易にコンペティションを開催できるようになった。コンペティションを通じて、自社の問題を解決するための機械学習手法の開発に、世界中のデータ解析技術者を動員できるようになった[6]。

近年、データ解析技術の実践演習の場としても、データ解析コンペティションが活用され始めている。たとえばKaggleは、教育現場でのコンペティション開催を支援するツールとしてKaggle in Classを提供している。教師は、Kaggle in Classを用いて、予測結果の評価やリーダーボード、フォーラム等のKaggleが提供する機能をそのまま活用して、講義用のコンペティションを開催できる。Kaggle in Classは、世界中の数百の大学で演習に用いられている。これは、データ解析コンペティションの仕組みが、データ解析技術の演習に有効と認知されていることのあらわれだといえる。実データと実課題が提供され、成果物を提出すると速やかにフィードバックが得られるコンペティションの仕組みは、データ解析技術の自学自習を促す点で効果的と考えられる。また、参加者同士での競い合いは、学習の動機づけに効果的と考えられる。

一方、教育効果を高めるためのデータ解析コンペティションの在り方については、十分な検討がなされていない。Kaggle in Classは、熟練したデータ解析技術者向けのコンペティションの機能をデータ解析教育に用いているが、熟練者向けのコンペティションと学習者向けのコンペティションでは、求められる機能が異なる。たとえば、学習者向けコンペティションでは多くの初学者の参加が想定されるため、初学者を支援する仕組みづくりが重要である。また、熟練者向けコンペティションは、競い合いを主目的としているため、参加者同士の協力を促すような仕組みは導入されていない。一方、学習者向けのコンペティションでは、参加者同士が協調して学習できるような仕組みづくりが、教育効果を高めるために重要だろう。さらに、熟練者向けコンペティションでは主に金銭報酬を参加のインセンティブとして用いるが、教育用途では、学習意欲向上のための別の動機づけの仕組みが求められる。

筆者らは、教育効果を高めるためのデータ解析コンペティション基盤の設計を研究するために、独自のコンペティション基盤「ビッグデータ大学」[7]を開発・運用してきた。初学者の参加を支援するための機能や、参加の動機づけのための機能を、コンペティション基盤に導入してきた。本稿では、ビッグデータ大学の機能を紹介するとともに、教育効果向上に向けた実証実験の結果を報告する。実証実験では、チュートリアル提供が参加促進につながるという結果が得られた。また、投稿回数制限が不要なリーダーボードの導入が、参加意欲を向上させ、ひいては参加者のパフォーマンス向上につながるという結果が得られた。これらの結果を紹介するとともに、これまでの運用経験に基づき、教育用データ解析コンペティション基盤の今後の方向性についても議論する。

2. 教育用データ解析コンペティション基盤「ビッグデータ大学」の概要

ビッグデータ大学はWeb上のデータ解析コンペティション基盤であり、誰でも自由に参加できる。2014年の開設以降、登録者数は順調に増加しており、2018年5月現在では約700名が登録している。複数の大学で講義・演習に利用されるほか、2015年の機械学習サマースクールのデータ解析コンペティション等、学術イベントに併設のコンペティションでも利用されている。コンペティションの題材として、主に企業から提供される実データを利用している。これまでに、オンライン通販サイトでのユーザの購買データを利用した購買予測のコンペティションや、名刺画像データを利用した名刺記載項目の予測コンペティション等、実践的なデータ解析課題を扱ってきた。

参加者は、以下の手順でコンペティションに参加する。まずビッグデータ大学へのユーザ登録をした後に、コンペティション用データをダウンロードする。予測モデルを構築し、与えられたテストデータに対する予測結果をコンペティション基盤に投稿する。投稿すると直ちに中間スコ

アが表示される。中間スコアは、テストデータの一部を対象に算出される。参加者は、コンペティションの期間中、繰り返し投稿できる。ただし、多くのコンペティションでは、1日の投稿回数は3回に制限されている。コンペティション終了後、最終スコアにもとづいて順位が決まる。最終スコアは、中間スコアの算出に用いなかった残りのテストデータを対象に算出される。一般的なデータ解析コンペティションと異なり、ビッグデータ大学では、入賞者に賞金を提供しない。ただし、イベント併設のコンペティション等では、賞品や、論文の共著者になる権利を提供することがある。

3. 学習者支援のアプローチ

熟練者向けのコンペティションと異なり、教育用途のコンペティションでは、学習者を支援するために、初学者のサポートや参加動機づけのための工夫が求められる。ビッグデータ大学で導入されている、学習者支援のアプローチを紹介する。

3.1 チュートリアル

特にデータ解析の初心者にとっては、コンペティションで用いるデータをダウンロードした後、最初の結果投稿を行うまでの障壁が高い。初学者が最初の投稿をスムーズに完了できるように支援する仕組みが重要である。ビッグデータ大学では各コンペティションにおいて、コンペティションデータを用いて予測モデルを構築し、予測結果を出力するまでのチュートリアルを提供して、初学者を支援している。チュートリアルでは、データの入力・加工から、前処理や予測モデルの学習、テストデータに対する予測結果の出力といった一連の手続きのPythonコードを提供している。チュートリアルは、題材となるデータセットや問題設定に沿ってコンペティションごとに用意され、チュートリアルにしたがってコードを実行すれば、予測結果が出力できるようになっている。

3.2 リーダーボード

コンペティションの会期中、全参加者の中間スコアをリーダーボードに公開し、参加者が現在の順位を確認できるようにしている。これにより、参加者の競争意欲を高め、多くの参加が期待できる。図1に示す通り、時系列での順位変動も可視化しており、参加者が自身の順位の変化を確認できるようにしている。

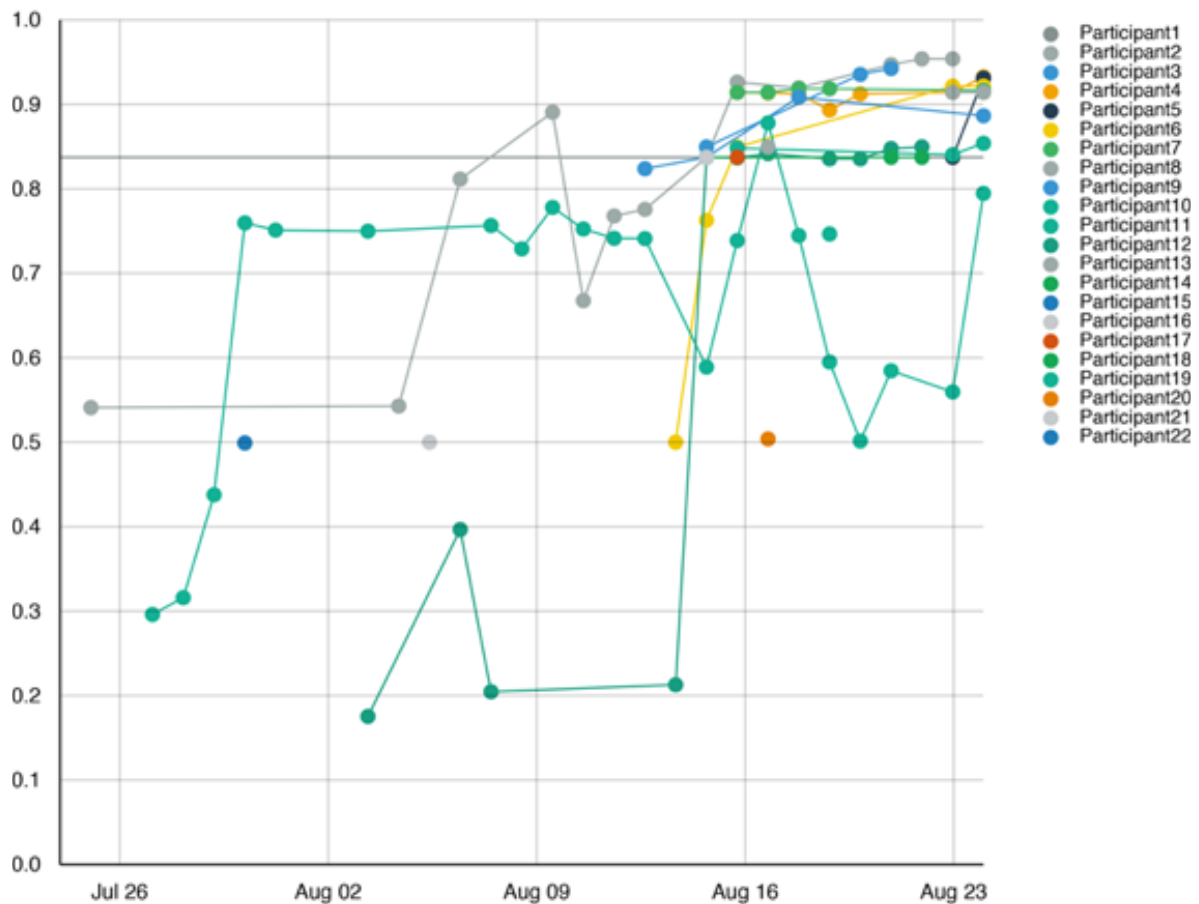


図1 リーダーボードの時系列可視化：コンペティション開催期間中の各日（横軸）における各参加者のスコア（縦軸）を可視化している。また、主催者提供のベースラインのスコア（この図では0.83）が示されている。

通常のリーダーボードでは、参加者が新しい結果を投稿するたびに中間スコアを更新する。このようなリーダーボードでは、参加者がリーダーボードから不当に多くの情報を引き出せないように投稿回数を制限する必要がある。しかし、投稿回数の制限は、意欲ある学習者にとっては試行錯誤の妨げになってしまう。そこでビッグデータ大学では、Ladderアルゴリズム[8]を用いたリーダーボードを用意している。Ladderアルゴリズムでは、新しい投稿のスコアが、自身の最良のスコアを統計的に有意に改善する場合のみ、中間スコアを更新する。これにより、リーダーボードからの情報漏えいを防げるため、Ladderアルゴリズムでは投稿回数の制限が不要となる。

3.3 レーティング

複数コンペティションへの参加を促すため、コンペティションを跨いだレーティングシステムも導入されている。このレーティングシステムは、TopCoder Marathon Matchで導入されているレーティング方法[9]を適用しており、あるコンペティションにおいて、自分よりレーティングが高い参加者を上回る成績を残すと、コンペティション終了後に自分のレーティングが大きく向上するようになっている。全員のレーティングとそのランキングが公開されており、参加者が自身の能力をアピールする場にもなっている。

3.4 入賞者レポート

レーティングと同様、上位の参加者が自身をアピールする場として、各コンペティションの入賞者によるレポートを掲載している。入賞者の要望があれば、本名や所属も併せて公開する。入賞者は、モデリングの基本方針や、前処理・特徴生成の方法、実装上の工夫等をレポートする。これらのレポートは、当該コンペティション参加者に、上位者がどのようなアプローチを採用していて自分はこういった点が不足していたのかを振り返る良い教材になる。また、以降のコンペティションの参加者にとっても、手法の方針を決める上での優れた資料となっている。

3.5 投稿類似度の可視化

大学等の講義でコンペティションを用いる場合、参加者が同じ講義に参加するため、オフラインでの交流が可能となる。参加者同士のオフラインでの議論の活性化を目的として、図2に示すような、各投稿の類似度を可視化したマップを提供している。この可視化では、各投稿を n 次元ベクトルと捉え、多次元尺度構成法を用いて2次元空間にマップしている。 n はテストデータ中のサンプル数に相当する。このような可視化により、自分と他の参加者の投稿の類似度を知ることができ、オフラインでの情報交換のきっかけになると期待される。

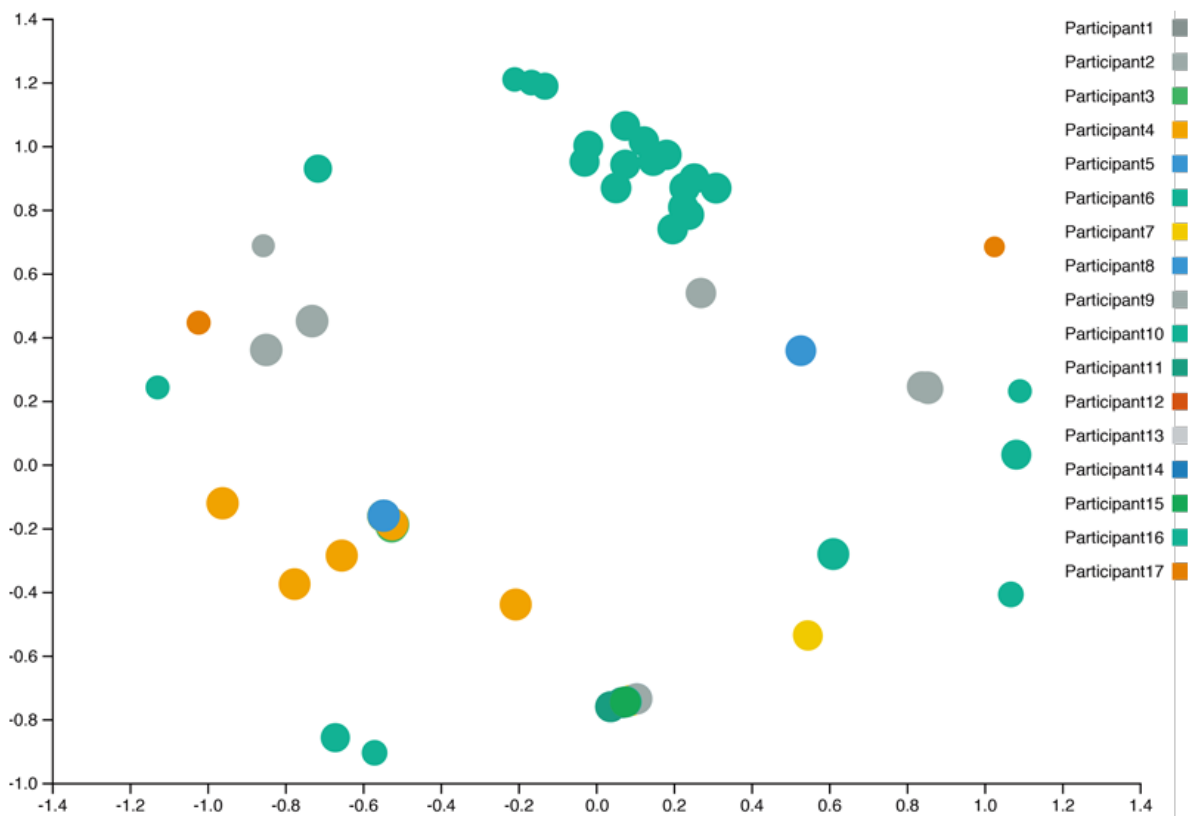


図2 投稿類似度の可視化：各投稿を2次元空間に射影することで投稿間の類似度を可視化している。

4. ケーススタディ

ビッグデータ大学で採用されている学習者支援のアプローチのうち、チュートリアルとLadderアルゴリズムに基づくリーダーボードについて、その効果測定のための実験を行った。実験内容と結果を紹介する。

4.1 チュートリアル

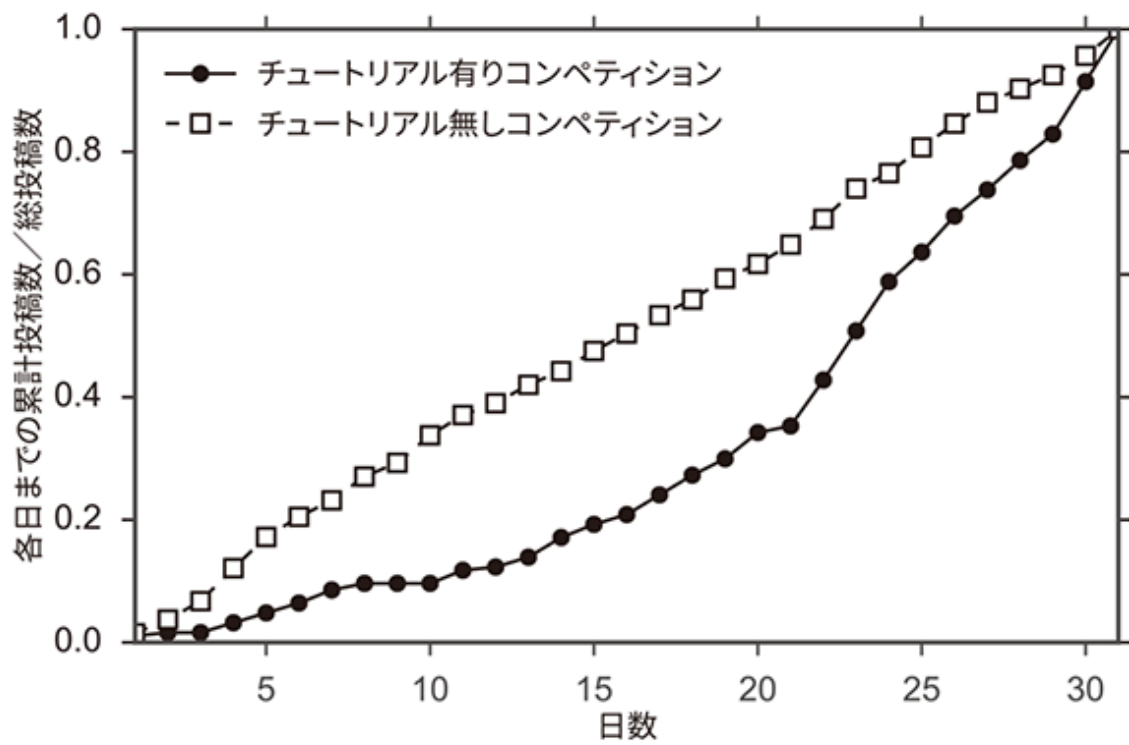
同時期に開催した2つのコンペティションを用いて、チュートリアルを提供が参加促進につながるかどうかを検証した。実験に用いたコンペティションは、オンライン通販サイトでの購買予測コンペティションと、ソーラーパネルの発電量予測のコンペティションである。このうち、購買予測コンペティションの方で、コンペティションの途中でチュートリアルを公開した。

購買予測コンペティションでは、1週間程度の期間内での各ユーザの行動情報（商品ページ閲覧、商品のカートへの追加・注文等）やユーザの属性情報を参加者に提供した。参加者はこれらのデータを用いて、特定の期間内にあるユーザがある商品を購入するかどうかを予測するモデルを構築する課題に取り組んだ。データ中のユーザ数は7,314名、商品数は10,099個であり、行動情報は78,107件、予測対象のユーザ・商品ペアは511,083件であった。発電量予測コンペティションでは、各時刻の発電量と気象情報（風速、気温、湿度等）を訓練データとして、気象情報からある時刻の発電量を予測する課題を扱った。訓練データは528件、テストデータは247件であった。

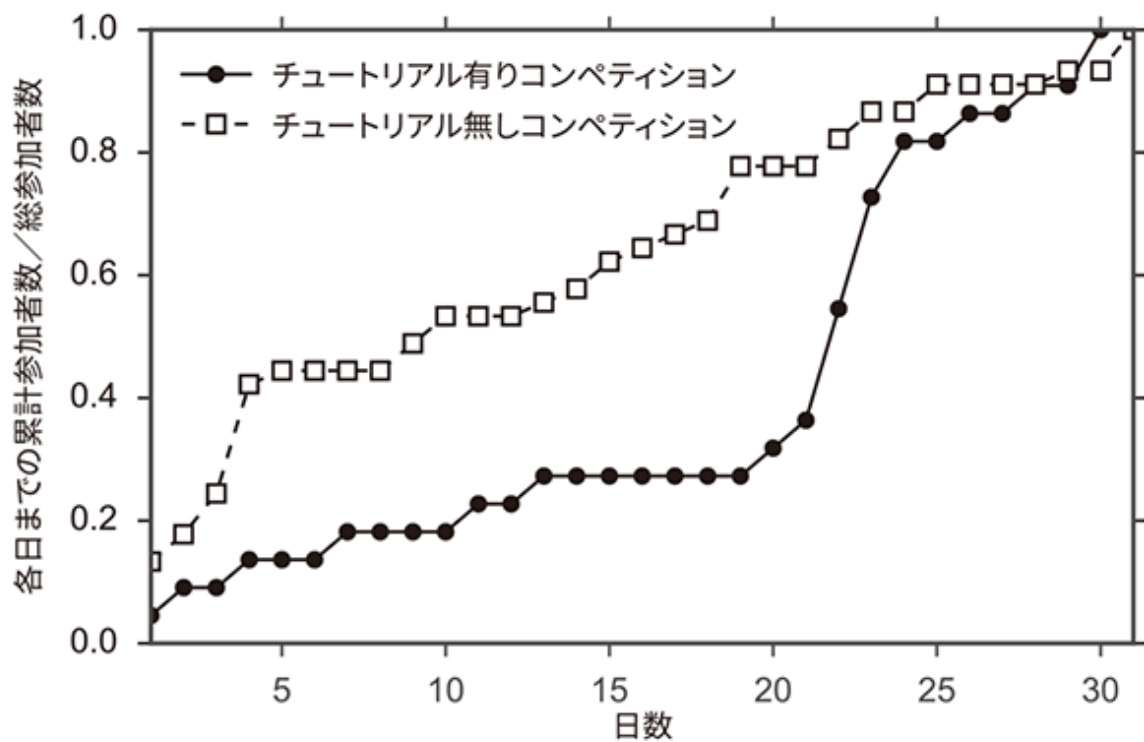
2つのコンペティションは、Machine Learning Summer School Kyoto 2015 (MLSS)の参加者を対象にしたコンペティションとして、まったく同じ期間に開催された。開催期間は2015年7月24日から8月23日である。購買予測コンペティションの参加者は22名、発電量予測コンペティションの参加者は45名であった。両方のコンペティションに参加した人は17名いた。MLSSは国内外の学生等に2週間の機械学習講義を提供するイベントであり、両コンペティションの参加者の多くは大学院生であった。総投稿数は、購買予測コンペティションでは187件、発電量予測コンペティションでは669件であった。発電量予測コンペティションの方が、データ形式が単純であり参加障壁が低かったために、参加者・投稿数が多かったと考えられる。購買予測コンペティションは、購買履歴とページ閲覧履歴からあるユーザがある商品を購入する確率を予測する問題であったが、複数のデータを統合して扱う必要があり、発電量予測コンペティションと比較して、データの取り扱いに工夫が求められた。

購買予測コンペティションでは、コンペティション開始から21日目にチュートリアルを公開した。両コンペティションでの、累積投稿数のグラフを図3 (a) に示す。チュートリアルを公開した購買予測コンペティションでは、公開後の22日目以降、投稿数が伸びていることが確認できる。公開前の21日間の総投稿数は66件で、1日あたりの平均投稿数が3.1件であったのに対し、公開後の10日間では121件の投稿があり、1日の平均投稿数が12.1件と大幅に向上している。なお、同時期に開催した、チュートリアルなしの発電量予測コンペティションでは、投稿数の伸びに大きな変化はないため、購買予測コンペティションでの活発度の変化に時期的な影響はなかったといえる。

図3 (b) に累積参加者数を示す。購買予測コンペティションでは、チュートリアル公開前の参加者は8人であったが、チュートリアル公開後に新たに14人の参加者を獲得している。これらの参加者も複数回投稿を行っており、チュートリアルの公開が、コンペティションへの積極的参加を促したと考えられる。



(a) 投稿数



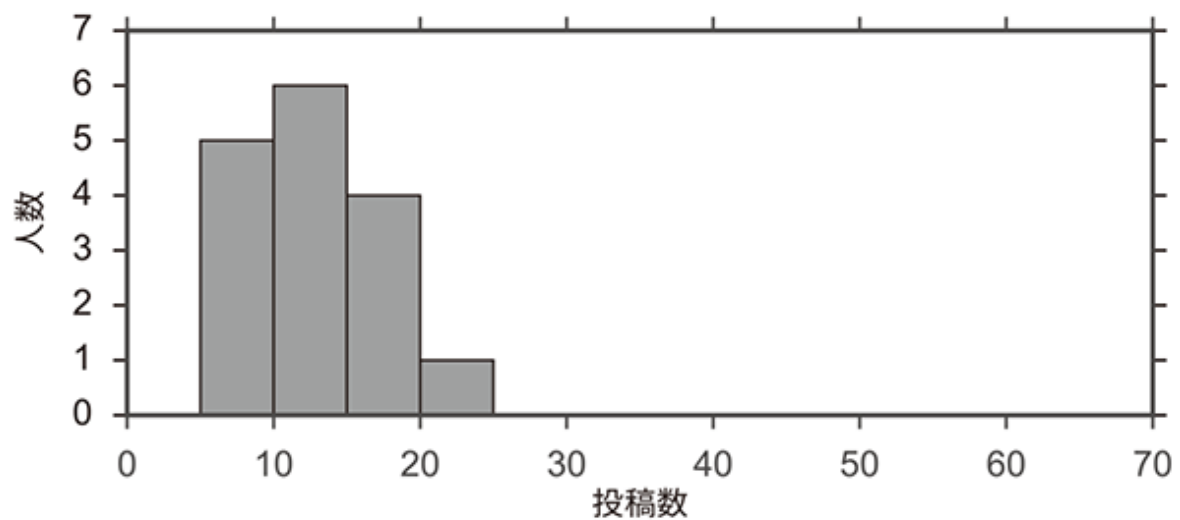
(b) 参加者数

図3 チュートリアルの有無による、投稿数と参加者数の比較

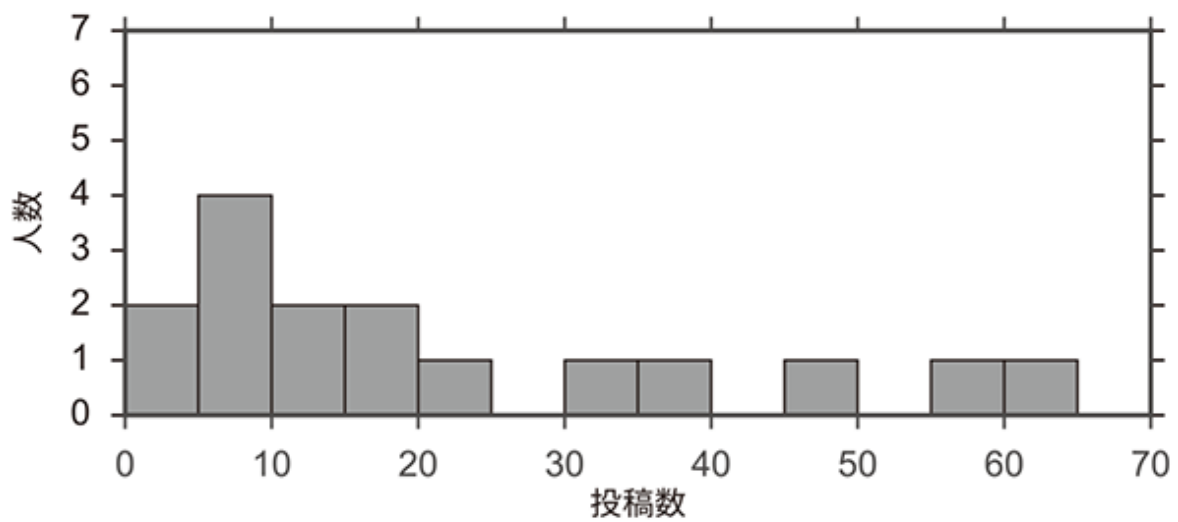
4.2 リーダーボード

同じく、同時期に開催した2つのコンペティションを用いて、Ladderアルゴリズムに基づくリーダーボードの利用による参加回数制限の撤廃が学習者の意欲向上につながるかを検証した。この実験で用いる2つのコンペティションは、データセット・予測問題・開催期間はすべて同じである。片方のコンペティションでは通常のリーダーボードを用い、もう一方のコンペティションではLadderアルゴリズムに基づくリーダーボード（Ladderリーダーボード）を用いた。前者のコンペティションでは1日の投稿回数は3回に制限し、後者では制限はなかった。ある授業の受講者を16人ずつの2グループにランダムに分け、各コンペティションに振り分けた。コンペティションの題材は単純な回帰問題であり、開催期間は2017年6月23日から7月27日であった。

図4に通常のリーダーボードを用いたコンペティションとLadderリーダーボードを用いたコンペティションでの、投稿回数の分布を示す。通常のコンペティションでは、投稿回数は最大22回であるのに対し、Ladderリーダーボードのコンペティションでは、投稿回数が20回以上の参加者が6人となり、意欲ある学習者が多数の投稿を行ったことが確認できた。通常のコンペティションでは総投稿数は191件であるのに対し、Ladderコンペティションでは352件となり、約1.8倍の活発度となった。



(a) 通常リーダーボードのコンペティション



(b) Ladderリーダーボードのコンペティション

図4 参加者の投稿数分布の比較

表1に、各コンペティションの上位3名の最終スコアと投稿回数を示す。最終スコアは同じテストサンプルを用いて算出されている。また、最終スコアは平均二乗誤差を用いているため、値が小さい方が高精度の予測結果となる。Ladderリーダーボードを用いたコンペティションでは、上位3名が、通常コンペティションの1位を上回る予測精度を達成した。特に、Ladderリーダーボードを用いたコンペティションの1位、2位は投稿回数が34回、61回であり、多くの試行が高い予測精度に結びついたと考えられる。図5に、通常リーダーボードのコンペティションとLadderリーダーボードのコンペティションでの最終スコアの分布を示す。Ladderリーダーボードのコンペティションの方が、全体として良いスコアを達成する傾向にある。また図6に、両コンペティションでの投稿回数と最終スコアの関係を示す。Ladderリーダーボードのコンペティションでは、投稿回数が多いほど最終スコアが優れているという結果が確認できた。以上の結果から、Ladderリーダーボードの導入による回数制限の撤廃には、参加者の試行錯誤を促し、高いパフォーマンスを引き出す効果があるといえる。

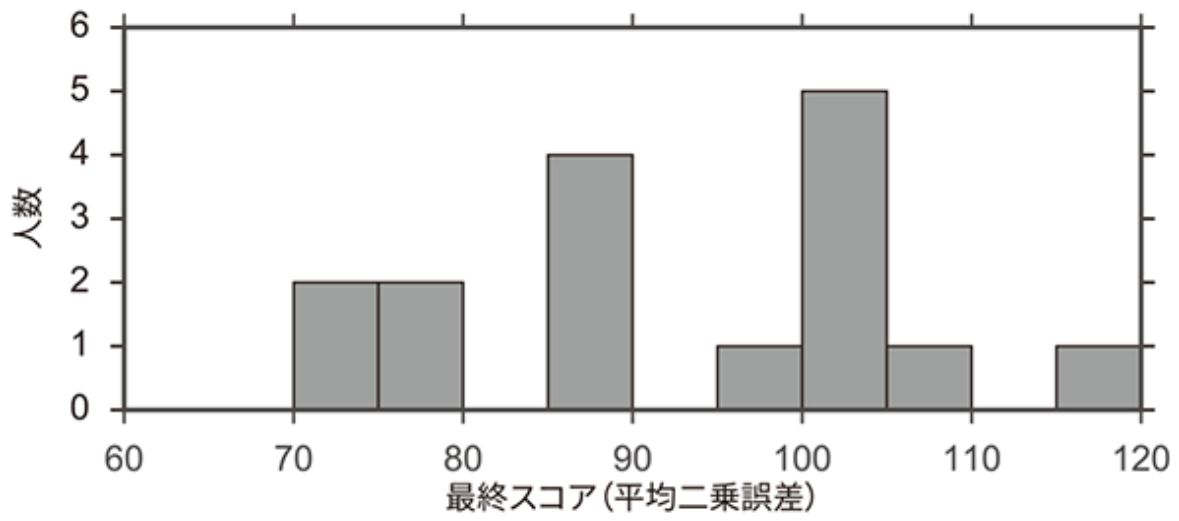
表1 上位3名のスコアと投稿数の比較

(a) 通常リーダーボード

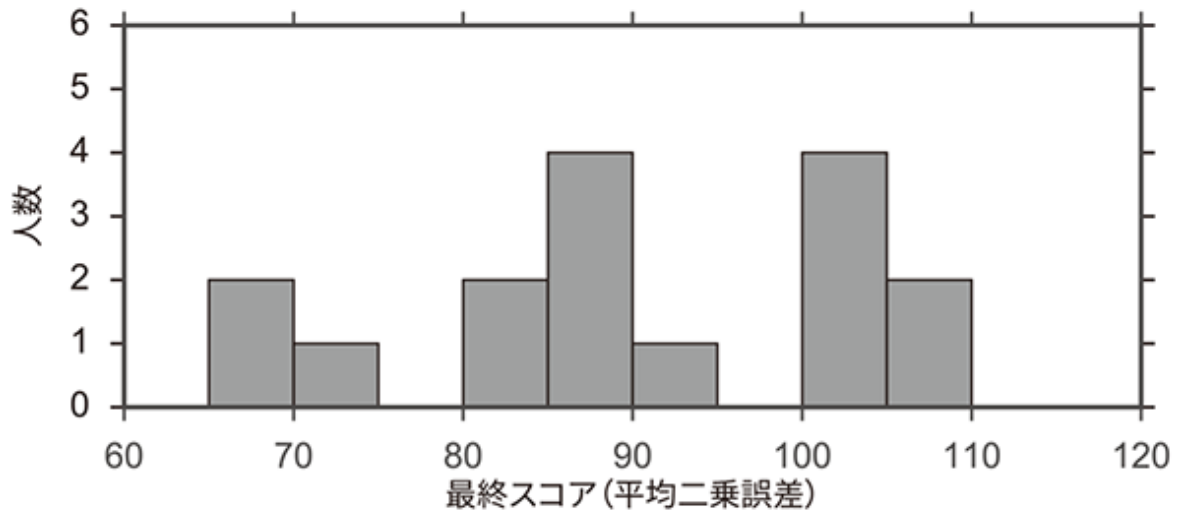
順位	最終スコア (平均二乗誤差)	投稿数
1	72.70	19
2	72.89	7
3	76.91	11

(b) Ladderリーダーボード

順位	最終スコア (平均二乗誤差)	投稿数
1	67.57	34
2	67.79	61
3	71.64	10

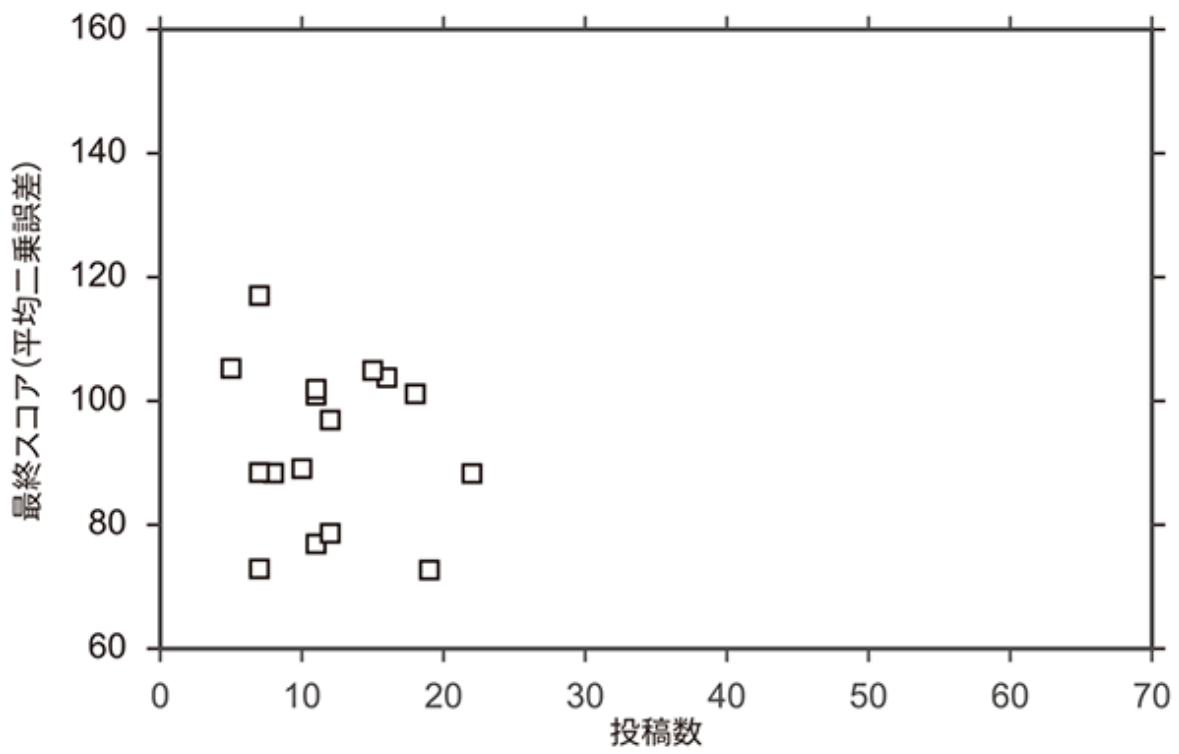


(a) 通常リーダーボードのコンペティション

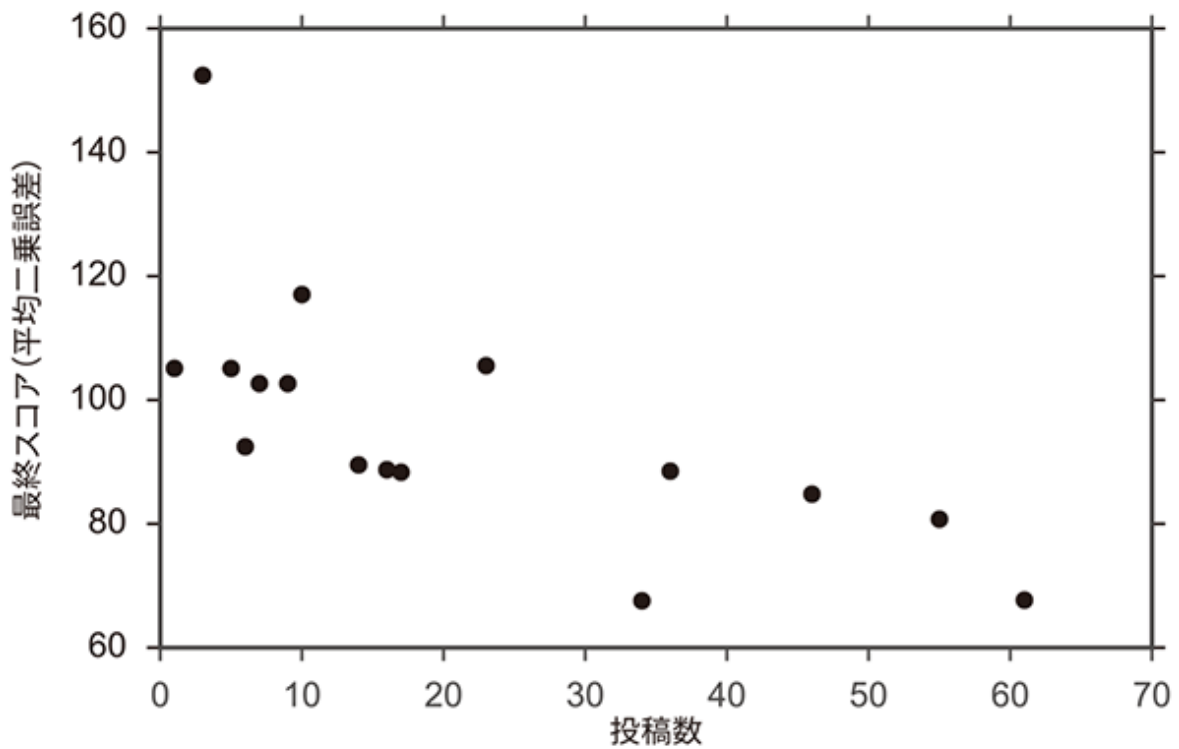


(b) Ladderリーダーボードのコンペティション

図5 参加者の最終スコア分布の比較



(a) 通常リーダーボードのコンペティション



(b) Ladderリーダーボードのコンペティション

図6 投稿数と最終スコアの関係の比較

5. 今後の展開

教育用データ解析コンペティション基盤「ビッグデータ大学」の運営経験に基づいて、教育効果を高めるためのコンペティションの方向性について議論する。

5.1 協調学習

従来のデータ解析コンペティションは、個人（あるいはチーム）の競争を主目的としていたが、教育用途の場合は、スキル向上のために参加者同士が協力し合い、協調的に学習する仕組みづくりが学習効果の向上のために有効と考えられる。実際に演習で用いた中で、参加者がある予測モデル（多層パーセプトロン）を試し、その有効性を演習に参加する友人数名にオフラインで伝えたところ、友人らもその利用により大きくスコアを改善することがあった。なかには、最初に多層パーセプトロンを試した参加者のスコアを追い抜く人すらいた。このような情報共有をオンラインで活発化できると、教室・組織の垣根を超えた協調学習が実現できる。活発に情報を提供して他の参加者の学習を支援する参加者に対する、アシストポイントのようなものを設計できると、協調学習の促進につながるだろう。

また、データ解析コンペティションにおいて、参加者同士の協力を促すための報酬メカニズムが提案されている[10]。このメカニズムでは、ある参加者が予測モデルを提出するとそれが他の参加者に公開され、別の参加者がそのモデルを改善できる。これを何度も繰り返し、最終的に各ステップにおけるモデル改善の度合いに応じて報酬が支払われる。このようなシステマティックな仕組みの導入は、協調学習促進の1つのアプローチと成り得る。

5.2 スキルの可視化

データ解析技術を実社会で活用する上では、データ解析技術者には多様なスキルが求められる。たとえば、データ解析のプロセスは前処理・特徴設計・予測モデルの構築といった段階にわかれており、各段階で異なるスキルが求められる。また、回帰・分類・推薦・異常検知といった問題設定ごとのスキルもあるし、自然言語処理・コンピュータビジョン等ドメインごとのスキルもある。このような細分化されたスキルにおける、学習者の長所・短所の可視化ができると、学習効果の向上につながるだろう。デザイン等のコンテストにおいて、コンテストでの勝敗情報とコンテストの特徴量を用いて、スキルの可視化を行う技術が開発されており[11]、このような可視化技術のデータ解析コンペティションでの活用が期待される。

5.3 探索型データ解析のコンペティション

ほとんどのデータ解析コンペティションでは、データから予測モデルを構築する予測モデリングを対象にしている。別の種類のデータ解析に、探索型データ解析がある[12]。探索型データ解析はデータの視覚化・集計等を通じてその中身を調べ、知見を獲得する作業である。データ解析のプロジェクトでは、探索型データ解析に多くの時間を割くこともしばしばあり、データ解析教育においては探索型データ解析のスキル育成も重要である。探索型データ解析を対象にしたコンペティションも開催され始めている。たとえば2000年のKDD Cupでは、「購買履歴データを使って、高額購入者の特徴を分析してください」という課題が出題された[1]。Kaggleでも、米国の人口統計データの分析コンペティション等が開催されている。

予測モデリングを対象にしたコンペティションでは、予測精度という定量的な評価指標を用いて順位を決定できる。一方、探索型データ解析では、定量的な評価指標を用意するのが難しい。2000年のKDD Cupでは、投稿結果それぞれについて専門家が、プレゼンテーション・内容の正

しさ・重要性等の観点で評価を行ったが、投稿数が増えると専門家による評価は労力が掛かり現実的ではない。

解決策として期待されるのが、MOOCs等でも導入されはじめている相互評価である[13]、[14]。相互評価は、教師の代わりに学習者にお互いの成果物を評価させ、評価の大規模化を実現する方法である。お互いに成果物を見せ合うことは、教育的にも良い効果をもたらすと考えられる。ただし、相互評価の結果によってコンペティションの順位が決まる場合、評価者間の利害関係が課題となり、公平な評価へのインセンティブ付与や結託攻撃への対応等が求められる。相互評価以外にも、インターネット上で不特定多数の人に仕事を依頼する仕組みであるクラウドソーシングで、評価を依頼する方法も考えられる[15]、[16]。クラウドソーシングでは評価の信頼性が課題となるが、クラウドソーシングでデータ解析結果の評価を依頼した場合、その信頼性は高いという実験結果もある[17]。相互評価やクラウドソーシングに基づく定量評価の効率的な仕組みの構築が、探索型データ解析のコンペティションの実現に向けて求められる。

6. おわりに

本稿では、筆者らが構築した教育用データ解析コンペティション基盤「ビッグデータ大学」を紹介した。ビッグデータ大学では、教育効果向上のためにさまざまな機能を導入している。ケーススタディを通じて、チュートリアル提供が参加促進に有効であることを示した。また、Ladderリーダーボードによる投稿回数制限の撤廃が、参加者の意欲を引き出すのに有効であることを示した。教育用データ解析基盤の今後の展開としては、協調学習を促す仕組みの導入、スキルの可視化、探索型データ解析のためのコンペティション設計が挙げられる。

参考文献

- 1) Kohavi, R., Brodley, C. E., Frasca, B., Mason, L. and Zheng, Z. : KDD-Cup 2000 Organizers' Report : Peeling The Onion, ACM SIGKDD Explorations Newsletter, Vol.2, Issue 2, pp.86-93 (2000) .
- 2) Netflix Prize : <https://www.netflixprize.com>
- 3) Töschler, A., Jahrer, M. and Bell, R. M. : The BigChaos solution to The Netflix Grand Prize.
https://www.netflixprize.com/assets/GrandPrize2009_BPC_BigChaos.pdf (2009) .
- 4) Kaggle : <https://www.kaggle.com>
- 5) CrowdANALYTIX : <https://www.crowdanalytix.com>
- 6) Baba, Y., Nori, N., Saito, S. and Kashima, H. : Crowdsourced Data Analytics: A Case Study of A Predictive Modeling Competition, In Proceedings of The 2014 International Conference on Data Science and Advanced Analytics (2014) .
- 7) ビッグデータ大学 : <http://universityofbigdata.net>
- 8) Blum, A. and Hardt, M. : The Ladder : A Reliable Leaderboard for Machine Learning Competitions. In Proceedings of The 32nd International Conference on Machine Learning, pp. 1006-1014 (2015) .
- 9) TopCoder Marathon Match Rating System :
<https://community.topcoder.com/longcontest/?module=Static&d1=support&d2=ratings>
- 10) Abernethy, J. D., and Frongillo, R. M. : A Collaborative Mechanism for Crowdsourcing Prediction Problems, In Advances in Neural Information Processing Systems 24, pp.2600-2608 (2011) .
- 11) Baba, Y., Kinoshita, K. and Kashima, H. : Participation Recommendation System for Crowdsourcing Contests, Expert Systems with Applications, Vol. 58, pp. 174-183

- (2016) .
- 12) Tukey, J. W. : Exploratory Data Analysis. Addison-Wesley (1977) .
- 13) Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., and Koller, D. : Tuned Models of Peer Assessment in MOOCs, In Proceedings of The 6th International Conference on Educational Data Mining (2013) .
- 14) Raman, K., and Joachims, T. : Methods for Ordinal Peer Grading, In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1037-1046 (2014) .
- 15) Baba, Y. and Kashima, H. : Statistical Quality Estimation for General Crowdsourcing Tasks, In Proceedings of The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 554-562 (2013) .
- 16) Sunahase, T., Baba, Y., and Kashima, H. : Pairwise HITS : Quality Estimation from Pairwise Comparisons in Creator-evaluator Crowdsourcing Process, In Proceedings of The 21st AAAI Conference on Artificial Intelligence, pp.977-983 (2017) .
- 17) Baba, Y. and Kashima, H. : Crowdsourcing Data Understanding: A Case Study Using Open Government Data, In An Adjunct to The Proceedings of The 3rd AAAI Conference on Human Computation and Crowdsourcing (2015) .

馬場 雪乃 (正会員) baba@cs.tsukuba.ac.jp

2012年東京大学大学院情報理工学系研究科にて博士(情報理工学)を取得。2012年～2014年東京大学大学院情報理工学系研究科特任研究員。2014年～2015年国立情報学研究所特任助教。2015年～2018年京都大学大学院情報学研究科助教。2018年より筑波大学システム情報系准教授。機械学習, データマイニング, ヒューマンコンピューテーションの研究に従事。

高瀬 朝海 (非会員) takase_t@complex.ist.hokudai.ac.jp

2012年東京大学工学部精密工学科卒業。2014年同大学大学院工学系研究科精密工学専攻修士課程修了。2018年北海道大学大学院情報科学研究科情報理工学専攻博士課程修了見込。ニューラルネットワークの探索と収束の制御に関する研究に取り組む。

新 恭兵 (学生会員) atarashi_k@complex.ist.hokudai.ac.jp

2016年北海道大学工学部情報エレクトロニクス学科卒業。2018年同大学大学院情報科学研究科修士課程修了。2018年同博士課程入学。特徴の組合せを用いる機械学習手法や、クラウドソーシングを用いた機械学習手法の研究に興味を持つ。

小山 聡 (正会員) oyama@ist.hokudai.ac.jp

2002年京都大学大学院情報学研究科博士後期課程修了, 博士(情報学)。2002年～2007年京都大学大学院情報学研究科助手, 2007年～2009年同助教。2009年より北海道大学大学院情報科学研究科准教授。2005年度人工知能学会論文賞, 2009年度日本データベース学会上林奨励賞, 2015年度同学会若手功績賞受賞。

鹿島 久嗣 (正会員) kashima@i.kyoto-u.ac.jp

IBM 東京基礎研究所勤務。2009～2014年 東京大学大学院情報理工学系研究科数理情報学専攻准教授。2014年より京都大学大学院情報学研究科知能情報学専攻教授。2016年より理化学研究所 革新知能統合研究センター ヒューマンコンピューテーションチーム チームリーダー(兼任)。人工知能, 特に機械学習, データマイニング等のデータ解析技術の研究開発とその応用に従事。

採録決定：2018年7月23日
編集担当：細野 繁（日本電気（株））