

経営情報学会 2018 年春季全国研究発表大会 企業事例セッション

経営情報学会 2018 年春季全国研究発表大会では、2017 年秋季大会に続いて企業事例セッションを開催しました。産業界の方に（1）経営情報学に関わる企業での実践例や（2）経営情報学に関わる技術・サービスをご紹介いただくことで、他企業や学識者との情報交換・交流の促進を意図するものです。企業同士となると日常ではなかなか情報交換の機会はなく、学会が交流の場を設定することには実務家の方にとって価値提供になりえ、また産学の交流という視点でも共同研究（事業）への発展の可能性もあることから、本セッションにより大会がより一層魅力的な場となることを目指しました。

今回は以下にご紹介する 9 件の発表をしていただきました。当日、セッション会場では活発な議論が交わされました。聴講者の評価をふまえ、大会委員長から「経営情報学会 2018 年春季全国研究発表大会企業事例セッション優秀報告賞」として株式会社アナリティクスデザインラボ 野守耕爾氏による「人工知能技術を応用した特許文書分析が生み出す新たな技術戦略の検討」、JFE スチール株式会社 茂森弘靖氏による「データサイエンスによる鉄鋼製品の品質管理の革新」を表彰しました。前号と今号にわたり、ご寄稿を承諾いただいた参加者による内容紹介を掲載いたします。

本誌をご覧の読者の方にも参考になる部分が多いと思いますので、今後の事業・研究にご活用いただければ幸いです。なお、本セッションは次回大会でも開催予定です。企業の皆様はぜひ発表をご検討ください。

- 1) ○山田広明（株式会社富士通研究所）、大堀耕太郎（株式会社富士通研究所）、岩尾忠重（株式会社富士通研究所）、吉良知文（九州大学マス・フォア・インダストリ研究所）、神山直之（九州大学マス・フォア・インダストリ研究所）、吉田宏章（株式会社富士通研究所）、穴井宏和（株式会社富士通研究所）
「不確実な状況下での空港旅客流のモデリングとマネジメントー福岡空港の事例から」
- 2) ○野守耕爾（株式会社アナリティクスデザインラボ）
「人工知能技術を応用した特許文書分析が生み出す新たな技術戦略の検討」
- 3) ○羽鳥冬星（株式会社リクルートテクノロジーズ）
「目的変数の異なるモデル間での stacking モデルについて」
- 4) ○齋木 啓（新日鉄住金ソリューションズ）、向 正道（新日鉄住金ソリューションズ）
「巧妙化するサイバー攻撃の手法とその対策—当社セキュリティサービス NSSEINT の提供を通じて—」
- 5) ○真鍋友則（Sansan 株式会社）
「名刺交換から見えてくるビジネスにおける『出会い』の分析」
- 6) ○加藤惇雄（株式会社大和総研）
「アウトバウンド型ビジネスにおける営業員特徴抽出」
- 7) ○森 正人（ヤフー株式会社）
「広告主のポテンシャル推定と分類」
- 8) ○茂森弘靖（JFE スチール株式会社）
「データサイエンスによる鉄鋼製品の品質管理の革新」
- 9) ○瀬之口潤輔（カムイ・キャピタル株式会社）
「実数値 GA による変数選択を用いた株価予想モデル」

大会委員長

優秀報告賞

データサイエンスによる鉄鋼製品の品質管理の革新

茂森弘靖（しげもり ひろやす）

JFE スチール株式会社

この度、経営情報学会 2018 年春季全国研究発表大会企業事例セッション優秀報告賞をいただきました。大変光栄に思います。これを励みに今後も本研究を深化させていきたいと思ひます。また、本技術分野の発展に微力ながら貢献していきたいと思ひます。誠にありがとうございました。

今回は、普段仕事でお世話になっている実行委員の方からお声をかけていただいたのがきっかけで、発表させていただきました。

本報告では、製鉄所に蓄積される大量データを有効活用したデータサイエンスによる鉄鋼製品の品質管理について解説しました。JFE スチールは、Just-In-Time モデリングの一種の局所回帰モデルを用いた品質設計および品質制御システム、ならびに、多変量統計的プロセス管理技術を用いた品質操業管理システムを実用化しています。これらは、製品品質の高精度化、従来の品質管理では発見できない異常の検出などを可能とし、品質管理の向上に大きく貢献しています。また、JFE スチールのさまざまなプロセス、さまざまな事業所で展開されています。技術教育およびツール開発など社内実施展開と継続的なレベルアップのための工夫について述べました。

質疑応答の時間および懇親会では、異業種の方から、また、大学の先生方から多くの質問や意見を頂くことができました。普段の仕事の中では聞くことができないものであり、今後の新しい研究の方向性を決める上で参考になりました。

本研究を進めるためには、製鉄所にある多くの工程のさまざまな製造設備のデータをリアルタイムかつ大量に収集する必要があります。鉄鋼業界では古くから計算機システムによる自動化が進められてき

ており、多くのデータを集めるための設備投資が行われてきました。それにより、データを集めるには比較的恵まれた環境にあるように思ひます。しかしながら、工程毎に分散してセンサーデータを蓄積しているため、多くの工程を跨ったデータ解析は難しい状況にあります。また、代表の 1 点に情報を縮約して品質管理を行っていました。すなわち、当時の計算機能力の制約のため有用な情報を捨てるをえず縮約したデータだけがデータベースに格納されていました。そのため、本研究に必要なより精密なデータを集めるためには、全国の事業所に赴き、製造現場までデータを取りに行く必要がありました。

Society 5.0 に向けた技術開発が全産業で活発に行われており、それを実現できる計算機システムの再構築が現在行われています。本報告ではその一端を紹介できたと考えています。しかしながら、さらに高度な品質管理ができるように研究開発をしていくことが必要にならうかと思ひます。従来は代表 1 点の 0 次元の情報に縮約していましたが、センサーデータで集めている 1 次元、2 次元、3 次元の情報をそのまま扱うようにする必要があらうかと思ひます。IoT、AI 技術の進歩を見据えた研究開発を今後も続けていくつもりです。

略歴

茂森弘靖（しげもり ひろやす）

1992 年入社以来、鉄鋼プロセスの計測・制御・システムの研究開発、建設、保全業務に従事。近年は主にデータサイエンスの研究、データサイエンティスト育成に従事。京都大学博士（工学）。

目的変数の異なるモデル間での stacking モデルについて

羽鳥冬星 (はとり とうせい)
株式会社リクルートテクノロジーズ

1. はじめに

本稿では筑波大学東京キャンパスで行われた2018年春季全国研究発表大会での発表の概要と今後の展開について述べます。

2. CV 予測問題の背景と課題

データを利活用する企業において、日々運用される予測モデルには様々なものがあります。広告コストの最適配分をするための応募数予測モデル、3カ年計画策定のための売上予測モデル、効率的な支店配置のための地域ポテンシャル予測モデルなど、多くの予測モデルが日々の意思決定に寄与しています。

一方で、現実のビジネス要件に沿って目的変数を設定していくと、しばしばその構造が複雑になっていくことがあります。例えばユーザーに広告を配信し、そのユーザーがコンバージョン（以下 CV とします）したかどうかに興味がある場合を考えます。最もシンプルな場合は、表1のような場合です。このとき、ユーザーは広告配信後、現在に至るまでに CV したかどうかだけを対象に目的変数 y が構成されています。

一方、現実的には CV までにかかる時間にも興味がある場合もあります。例えば実際に広告配信からユーザーの CV を予測するモデルを構築したとしても、それがいつ起こるのかわからなければ、実際の施策に結びつけるのは難しくなってしまいます。そのような場合には表2のように目的変数 y を構成します。ここでは CV したユーザーのうち、当日に CV したものを正例として扱い、目的変数 y を構成しています。このような状況下では、モデルは各ユーザーが CV するかどうかに加えて、それが当日に起こるのかどうかも併せて予測する必要があります。そのため、ただ単に CV 有無を予測する場合に比べて難しい問題となっています。本稿では上記

表1 時間制約がない CV

	当日	翌日以降
CVあり	$y=1$	$y=1$
CVなし	$y=0$	$y=0$

表2 時間制約がある CV

	当日	翌日以降
CVあり	$y=1$	$y=0$
CVなし	$y=0$	$y=0$

のような時間制約がある CV の予測問題を対象に、アンサンブル学習の一手法である stacking (Wolpert, 1992) を用いて、高精度なモデルが構築可能であるかを検討します。

3. 提案手法

本研究では表2のような、時間制約がある CV を予測する問題における初段と次段の目的変数が異なる stacking モデルを提案します。

モデルの全体像は表3の通りです。本手法と通常の stacking モデルとの差分は手順4にあります。ここで分割したデータに対する初段のモデルにおける学習において、表2のような時間制約のない CV を目的変数と学習を行います。

通常の stacking モデルにおいては初段と次段の目的変数は同じものを使用します。このときに、初段の精度が良ければもちろんその分だけ次段の精度も向上します。しかし時間制約がある CV を予測する問題においては、初段モデルの予測値の精度が低い場合が起こり得えます。一方、提案する stacking モデルは初段で単純な目的変数を持つモデルを学習し、次段では複雑な目的変数を持つモデルを学習します。そのようにして得られた予測値は、CV の有無を当てるものとしては信頼性が高く、次段のモデルの予測に十分寄与すると考えられます。

表3 提案手法のアルゴリズム

1.	Dを学習に使用する教師データの全体とする
2.	データを D_1, \dots, D_n に分割する
3.	$i=1$ とする
4.	$T_i = \{D - D_i\}$ を目的変数に時間制約のないCVの有無を設定した初段モデルで学習する
5.	D_i を予測し予測値を P_i として保存する
6.	$i=i+1$ とする
7.	$i=n$ でなければ4に戻る
8.	P_1, \dots, P_n を用いて、次段のモデルの学習を行う

4. 数値実験

今回は、時間制約があるCVを予測する問題について、実データを用いて数値実験を行いました。いくつかのアルゴリズムについて、このデータを既存手法で予測した場合と、提案手法を用いて予測した場合のROC-AUCを比較したのが表4です。Naive Bayes Classifier, Logistic Regression, RandomForest, SVMでは既存手法に比べ、精度向上が見られました。これらのアルゴリズムにおいては初段の予測値が汎化性能の向上に寄与していると考えられます。

これらとは対照的に、XGBoostにおいては精度低下が起こっています。この原因の一つは、初段の予測値が特徴量として有効に機能していなかったことだと考えられます。

5. 今後の展開

本研究では、時間制約があるCV予測問題において、初段と次段の目的変数が異なるstacking手法を提案しました。数値実験ではいくつかのアルゴリズムにおいては精度向上が見られましたが、実務適用の際にはいくつか課題があると考えられます。

表4 アルゴリズムごとの精度比較

アルゴリズム	holdoutスコア	手法
Naive Bayes	0.6088	既存手法
Naive Bayes	0.6417	提案手法
Logistic Regression	0.6769	既存手法
Logistic Regression	0.6798	提案手法
SVM	0.68	既存手法
SVM	0.6853	提案手法
RandomForest	0.6789	既存手法
RandomForest	0.6827	提案手法
XGBoost	0.6884	既存手法
XGBoost	0.6862	提案手法

stackingモデルの宿命として、層を重ねるごとに必要な計算機資源のコスト、学習の際にかかる時間が増大します。また、予測の際にも複数のモデルを通す必要があるためレスポンスタイムが長めにかかることとなります。これらのコストが精度向上によるビジネスインパクトと釣り合うかは慎重に検討すべき事項です。

今後は実務実装にあたってのこれらの課題を検討したいと考えています。

参考文献

[1] Wolpert, D. H., "Stacked generalization," *Neural networks*, Vol. 5, Issue. 2, 1992, pp. 241-259.

略歴

羽鳥冬星 (はとり とうせい)

2015年7月より株式会社リクルートテクノロジーズにてデータ分析業務に従事。