

特集号招待論文

Account-Based Marketingのためのターゲット企業推薦モデルの改善

新井和弥¹ 北内 啓² 高柳慎一¹ 早川敦士¹ 林 樹永¹ 長田怜士¹

¹ (株) ユーザベース ² (株) ニューズピックス

B2B (Business To Business) 領域における企業情報活用が著しい飛躍を遂げており、企業情報を用いた新たなB2Bマーケティング手法としてABM (Account-Based Marketing, アカウ
ントベースドマーケティング) の活用が広がっている。ABMをソフトウェアによって実践する方
法の1つとして、既存企業のデータを教師データとし、受注確度が高いと考えられる既存企業と
の類似度の高いターゲット企業を推薦するモデルを構築する方法がある。最終的に人間がどの企
業に対してマーケティング施策を行うのかという意思決定を行う都合から、ターゲット企業の推
薦モデルにはモデルの解釈性、すなわち、特徴量が企業類似度に与える影響を解釈しやすいこと
が要件として望まれる。(株) ユーザベースが提供するB2B事業向け顧客戦略プラットフォーム
FORCASでは企業の所属業界、企業の推定利用サービス等の特徴量としたナイーブベイズを独自
拡張したモデルを従来使用していたが、ナイーブベイズにおいて仮定される特徴量の独立性は現
実には満たされていない。このため特徴量間の正相関の効果によりアカウントスコアの値が嵩上
げされるという問題が生じ、結果として推薦されるターゲット企業に偏りが出てしまっていた。
この問題を解決するため、本稿ではL2正則化項付きのロジスティック回帰モデルを用いて企業類
似度を算出し、この類似度に基づいた推薦アルゴリズムを提案する。FORCASが保有する1社あ
たり約1,500個の特徴量を持つ合計約11万社の企業データに対し、業種が異なる4種類の既存企
業のデータに対して検証した結果、シンプルな手法ながらもナイーブベイズ拡張モデル、また一
般的に高精度と評価されるGBDTと同等以上の精度を達成することを確認した。また、提案手法
は業種の違いに対して頑健であり、ABMを実践するためのモデルとして適切であることも確認し
た。

1. ターゲット企業推薦モデルとは

1.1 ターゲット企業推薦モデルとは

近年、Web上の情報に効率的にアクセスするためのAPIを代表とした技術の普及により、さま
ざまな企業において顧客データの収集と管理が進められている。また、人工知能によるデータ処
理技術の発展とその普及も日進月歩の勢いで進んでおり、結果として、多くの企業において顧客
データの収集、およびその活用が進んできている。特に、B2B (Business To Business) 領域

における企業情報活用が著しい飛躍を遂げており、企業情報を用いた新たなB2Bマーケティング手法としてABM（Account-Based Marketing, アカウトベースマーケティング）の活用が広がっている[1]。ABMはアメリカに本社をおくアドバイザリーファームITSMAが2003年に初めて提唱した概念である[2]。2010年頃よりアメリカで注目され始め、ABMに特化したソリューションが次々と誕生している[3]。シンフォニーマーケティング（株）は「全社の顧客情報を統合し、マーケティングと営業の連携によって、定義されたターゲットアカウントからの売上げ最大化を目指す戦略的マーケティング」とABMを定義している[4]。また、（株）ユーザベースFORCAS執行役員CEO田口慎吾は、ABMを「ターゲット企業（アカウント）を定義し、ターゲット企業（群）別に営業・マーケティング情報を集約し、ターゲット企業（群）別に営業・マーケティング組織を再編成し、ターゲット企業（群）からのLTV最大化を目指すマーケティング」と位置づけており、ABMは顧客戦略プラットフォームへとさらに進化していくと報告している[5]。従来のデマンドジェネレーション（マーケティング活動において営業部門への見込み顧客を渡す活動全般）は個人に集中して実施されていたのに対し、ABMではターゲットとなるアカウント（企業）に集中して実施する点が最も大きな違いである[3]。また、ABMを実行するためのABMプラットフォーム覇権争いも激化してきており、米国大手企業の参入も続いている。米オラクル社はABMを「マーケティングに関するカスタマイズされたアプローチであり、見込み客へのターゲティングを通じてブランドの認知や製品購入意欲を高めてもらいながら、情報提供を通じた関係構築を深めるのに役立つマーケティングである。特に、マーケティングの際に使用されるクリエイティブや送付メッセージは、顧客のある特定の問題点に関連したものである」と定義している[6]。

受注確度が高いと推測される企業（以下、ターゲット企業）の特定には、自社の営業データ、特に受注済みの顧客データと豊富な属性データを持つ企業マスタデータが必要となる。属性データとしては、従業員数、業種、上場/非上場など企業の規模や種別を表すファームグラフィックデータと呼ばれるデータがしばしば利用されるが、企業をより特徴付けているあたかもユーザの行動ログのように見做せるデータ、たとえば企業が利用しているアクセス解析ツールやチャットツールなどのサービス、特定の国・地域に進出しているか否か、企業が求人票において募集している職種、といった企業の現在の行動に基づいた行動解析的なデータも有用である。これらのデータを組み合わせたマーケティング活動を実施することで営業活動の生産性を飛躍的に向上させる企業が増えてきており、ABMの重要性が高まっている。特に、資金力に乏しく初期顧客となり得るターゲット企業に焦点を絞ったアカウントベースの戦略を通じて急速な収益成長を促進する必要のあるスタートアップ企業向けにABMのベストプラクティスに関する実践的な手引が存在するほどである[7]。

ABMをソフトウェアによって実践する方法の1つとして、受注済みの企業（以下、既存企業）のデータを教師データとし、受注確度が高いと考えられる、企業類似度の高いターゲット企業を推薦するモデル（以下、ターゲット企業推薦モデル）を構築する方法がある。最終的に人間がどの企業に対してマーケティング施策を行うのかという意思決定を行う都合から、受注確度の高い企業をターゲット企業推薦モデルはモデルの解釈、特に企業の特徴量が企業類似度に与える影響を解釈しやすいことが要件として望まれる。

我々は、既存企業のデータを分析して既存企業以外の企業（以下、潜在企業）の中からターゲット企業を特定できるシステム（以下、本システム）を2019年に構築した[8]。本システムにおいては、ターゲット企業の推薦問題を受注するかしないかの二値分類問題として捉え、ナイーブベイズを独自に拡張させ、特徴量の重要度を算出しつつさらにスムージングによって補正することで、推薦するターゲット企業の根拠を解釈しやすい特徴量の重要度とともに提示することが可能となった。また、本システムはユーザが特徴量の重要度を変更でき、ほかの特徴量に影響を与えずに潜在企業の企業類似度を再計算することも可能となるよう構築された。なお、企業類似度は分かりやすさの観点から本システムのユーザに対してはアカウントスコアと呼称されており、本稿においても適宜使い分ける。

以下に、典型的なシステム構成を図1に示す。システムは大きく2つのステップからなる。
 (1) ユーザがアップロードした既存企業のデータを名寄せ処理によりシステム内の企業マスターデータと紐づけ、ファームグラフィックデータおよび行動解析的なデータを特徴量としてターゲット企業推薦モデルを学習する。
 (2) ターゲット企業推薦モデルをもとに潜在企業の受注確度を企業類似度として推定し、出力された企業類似度の高い順に提示することでターゲット企業の推薦とする。また各企業のデータから作成した特徴量を重要度とともに提示することで、その企業が推薦された根拠をユーザが理解できる。本システムが上記の2つのステップを持つ理由を述べる。
 (1) については、同名の企業が存在する場合や、ユーザが管理する既存企業のデータにおける企業名と弊社システム内で管理する企業名の表記が異なる場合がある。アップロードされた各既存企業を一意に特定するために、企業名、所在地、電話番号、法人番号などの情報を用いて名寄せする必要がある。名寄せ後は企業マスターデータと既存企業のデータの対応付けができていたので、ターゲット企業推薦モデルの学習に必要なデータを作り、実際に機械学習モデルを学習させる。以上のような理由がある。
 (2) については、ターゲット企業にビジネスの活動を集中し、優先順位をつけるために必要な推薦アルゴリズムであり、また、特徴量の重要度も算出されるため、解釈性に優れているという理由がある。

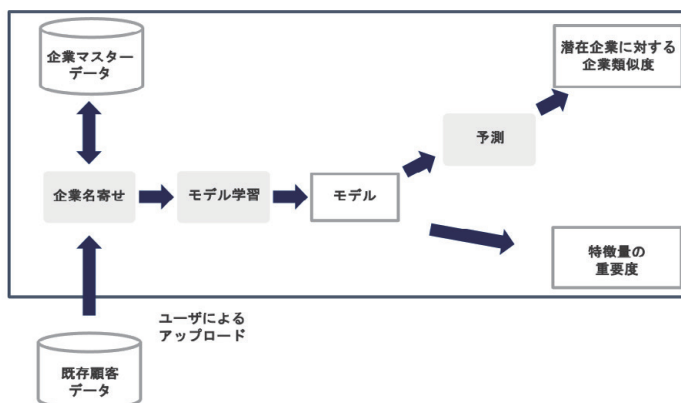


図1 ターゲット企業推薦システムの構成

1.2 ABMに関連する研究

公表されているABMに関する研究，特に機械学習モデルの理論・実践はきわめて少ない．[9]，[10]においては，B2B業界におけるABM自体の進化，およびそのメリットについて論じているものの，機械学習を用いたターゲット企業推薦モデルの構築法について論じたものではない．河村ら[11]は，勘と経験に頼っていた不動産営業を置き換えるために，機械学習を用いて申込み顧客リストを作成することによって営業の効率化を支援する推薦モデルを提案した．この研究では，ベテラン営業への聞きとり調査を2度実施している．1度目の調査により作成した特徴量に基づくモデルに対しての推定結果を掲示し，フィードバック（2度目の調査）を得て特徴量を作り，営業の経験を加味したモデルを構築している．内見時のアンケート結果，物件情報の基礎データのほかに熱意と地域ポテンシャルという独自の特徴量を加え，ランダムフォレストを用いて二値分類問題として解いたところ，53.8%のPrecision（適合率）で申込み顧客を推薦することができたと報告している．中山ら[12]は法人向けの営業活動における初期訪問から受注に至るまでの一連のプロセスの各ステップおよび各ステップで選択されるアクティビティに注目し，営業日報などの非構造化データを入力情報として，受注確率の高い営業プロセスの規則性を発見する学習モデルを構築，法人営業活動を営む企業において営業活動の意思決定支援システムを構築した．実際の営業担当者による試用を通して3カ月の営業活動期間を対象にして営業活動意思決定支援システムの適用前後を比較したところ，受注率において10%程度の改善効果を確認している．これら2つの研究は営業効率向上という意味ではABMと類似しているものの，ABMそのものに主眼を当てた研究ではない．また，（株）セールスフォース・ドットコムSalesforce Einsteinという人工知能製品においてもABM機能を搭載している旨が公表されているが，その詳細なアルゴリズムなどは不明である[13]．

2. 本研究の位置づけ

2.1 ターゲット企業推薦システムの抱える課題

本節では，本システムで使用している既存手法の紹介をするとともにその問題点を明らかにする．ターゲット企業の推薦問題は分類，回帰，協調フィルタリングなどさまざまな手法で定式化できる．本研究では二値分類問題と捉え，潜在企業の企業類似度を算出する．すなわち，企業類似度は0から1の間の値をとり，その値が1に近い値をとる企業ほど既存企業に近しく受注確率が高いと考えられるので優先的にマーケティング施策・営業攻勢をかけるべき企業として推薦される．また，企業の属性データのほとんどがあるサービスを利用しているか否かや，ある特定の条件（シナリオと呼称）に当てはまるか否かなど二値データとなるため，従業員数や売上げ等の量的データも二値データに変換することで，すべての特徴量を二値データとして扱う．すべての特徴量を二値データ化するため，特徴量の解釈方法を統一させることが容易であり，ユーザビリティの向上にもつながる．一般に，Webサイトからの問合せやオンラインセミナーの参加者情報，メールマガジンの配信などマーケティング施策によって生じる見込み客はリードと呼ばれる．マーケティング施策の結果，今後もフォローすべきであると判断されたリードは営業へと引き継がれ，そこから先の営業活動が営業担当者の仕事となる[14]．ABMにおいてはターゲット企業のリストがリードとなる．本システムではターゲット企業が持つ特徴量の重要度をアカウントスコアの根拠として提示する．推薦されたターゲット企業をもとに営業活動を行う際，その根拠が明確で理解しやすいことが重要である．ABMではマーケティング担当者と営業担当者の連携が不可欠であり，マーケティング担当者から営業担当者にターゲット企業のリストを提供したときそのリ

ストがなぜでき上がったのか？ の根拠が明確でないと、営業側が一方的にリストを押し付けられ、営業を命じられたと感じるなど信頼を得にくいからである。一方、明確な根拠があれば、営業側も適切な営業手段を選択でき、売上の増大に貢献する可能性が飛躍的に高まる。そのためには、重要度がアカウントスコアの増減に与える影響を解釈しやすいことが望ましい。また、特徴量同士の依存関係がある場合でも、特徴量単体でどの程度重要かが分かるような重要度が良い。すなわち、推定時における特徴量間の相互作用の影響がなるべく小さくなるようなモデルが良いと考えられる。

上述の要件を考慮し、既存手法として採用した手法はナイーブベイズを独自に拡張した手法（以下、NBEM）である。xを特徴量ベクトル、 x_i をxのi番目の要素とすると、ナイーブベイズにおけるクラスyの条件付き事後確率は以下となる：

$$P(y|x) = P(y) \prod_i P(y_i).$$

ただしここでiは特徴量ベクトルのすべての要素に対して走る。また、既存企業、潜在企業のクラスをそれぞれ y_1 , y_2 とすると、双方の条件付き事後確率の比 $R(x)$ は次式のようにになる。この値をNBEMにおける潜在企業のアカウントスコアと定義する：

$$R(x) = \frac{P(y_1|x)}{P(y_2|x)} = \frac{P(y_1) \prod_i P(x_i|y_1)}{P(y_2) \prod_i P(x_i|y_2)} = \frac{P(y_1)}{P(y_2)} \prod_i \frac{P(x_i|y_1)}{P(x_i|y_2)}.$$

また $S(x_i)$ を以下のように定義し、特徴量 x_i の重要度とする：

$$S(x_i) = \frac{P(x_i|y_1)}{P(x_i|y_2)}.$$

ここで $S(x_i)$ は、既存企業における特徴量 x_i の生起確率が、潜在企業におけるそれと比べてどの程度大きいかを表す値である。 $P(x_i|y_j); j=1,2$ は次式によって算出する：

$$P(x_i|y_j) = \frac{N(x_i, y_j)}{N(y_j)} + \alpha.$$

ここで $N(x_i, y_j)$ はクラス y_j において特徴量 x_i を持つ企業数、 $N(y_j)$ はクラス y_j の企業数である。右辺の第1項は多変数ベルヌーイモデルでの最尤推定によって導出される値である。 $\alpha(0 < \alpha < 1)$ はスムージングのための値であり、 $N(x_i, y_j)$ が小さいほど $S(x_i)$ は1に近づく。また、既存企業における特徴量 x_i の生起確率が潜在企業におけるそれと等しい、すなわち $N(x_i, y_1)/N(y_1) = N(x_i, y_2)/N(y_2)$ のとき $S(x_i) = 1$ となり、スムージングしないとときと同じ値となる。本システムにおいて、この特徴量の重要度は星マークによる9段階表記にてユーザに表示される。

このように、本システムでは企業の所属業界、企業の利用サービス等を特徴量としたナイーブベイズを独自拡張したモデルを使用していた。一方、この計算方式では、アカウントスコアの計算において特徴量間の正相関の効果によりアカウントスコアの値が嵩上げされるという問題があり、推薦されるターゲット企業に偏りが出てしまう場合があった。この問題はあるサービスを利用しているか否かを表現する特徴量のうち、ある特定のジャンルのサービス群に属するサービス間において強く出る傾向にある。たとえばソフトウェア開発会社においては複数のプログラミン

グ言語や開発フレームワークを用いて開発することは至ってあたり前のことであり、その結果として当該利用サービス（あるプログラミング言語やライブラリ、または開発フレームワーク等）における特徴量間の正の相関が強くなることに起因している。たとえばプログラミング言語Pythonを利用している企業は機械学習ライブラリであるPyTorchやscikit-learnも使用している傾向にあることは直感的にも理解できよう。

本問題を解決するため、アカウントスコア算出口ジックにおいて、L2正則化項付きロジスティック回帰モデルを用いた方法を提案する。

2.2 L2正則化項付きのロジスティック回帰モデルを用いた解決方法の提案

2.1節で説明した問題を解決するため、本稿ではL2正則化項付きのロジスティック回帰モデル（以下、L2LR）を用いた企業類似度を新たに定義し、この類似度に基づいた推薦アルゴリズムを提案する。正則化に関しては、スパースモデリングの文脈においてL1正則化法の一つであるLASSO（Least Absolute Shrinkage and Selection Operator）[15]が非常に有名である。モデルのパラメータを最小二乗法で推定する際、サンプルサイズがモデルのパラメータ数に比べて過小である場合や、あるいは特徴量間の相関が高いパラメータが複数存在する場合において、パラメータの最小二乗推定量が存在せず推定自体が不安定化してしまう。このとき、LASSOを用いることで、パラメータのL1ノルムがペナルティとして学習時の損失関数に加算され、パラメータの推定が安定的に行われると同時に変数選択も行うことができるという特徴がある。

一方、ターゲット企業推薦モデルの改善においてはLASSOを用いたモデルは適切ではないと判断しその採用を見送った。その理由は、LASSOを用いた場合、重要ではあるがほかの特徴量との相関が高い特徴量が選ばれずに、人が見たときに直感的に分かりにくい結果となってしまう場合があるからである。たとえば、売上高や従業員数という特徴量は、東証一部に上場しているという特徴量と強い相関を持っているため、モデル推定としてはそのどちらかが削除されてしまっても問題ない場合が多い。一方、人が直接的に各企業の特徴量をフィルタリングの条件として使用する際には売上高X円以上、従業員数Y人以上、東証一部上場企業という形でフィルタリングを行うことは一般的である。したがって、たとえ特徴量間の相関が強かったとしても、モデル推定の結果として当該特徴量に対するパラメータが有限の値を持ち、また、その特徴量の重要度が算出されることが強く望まれる。

そこで本稿においてはL2LRを用いたモデル構築を行うこととした。L2正則化のメリットとしては、ある特徴量のパラメータの値だけを極端に大きくするというのをしないために過学習を防ぐことができ、たとえ相関の強い特徴量があっても、ある一方の特徴量が完全に削除されることはなく、それぞれの特徴量に対応したパラメータに対しある程度の大きさの値を割り振ってくれるよう作用する点である。このようなL2正則化の効果により、推薦されるターゲット企業に存在する偏りを解消することが期待される。またL2LRはロジスティック回帰モデル同様、標準化されたデータに対し、パラメータの大きさそのものが予測モデルへの影響度合いを表しており、解釈性が非常に高い点もABMのモデルとして好ましい。L2LRのパラメータ推定は以下のように定式化される：

$$\min_{w,c} \frac{1}{2} w^T w + C \sum_{i=1}^n \log(\exp(-y^{(i)}(x^{(i)T} w + c)) + 1).$$

ただしここで、 $y^{(i)}, x^{(i)}$ はそれぞれ*i*番目のサンプルのクラスと特徴量ベクトル、 n はサンプルサイズ、 w は各特徴量に対して推定されるパラメータ、 c は切片項、また C は正則化の強度の逆数を表す。アカウントスコアは、L2LRの出力する受注確度の高い企業である確率に対し、独自の確率分布の補正などを行い最終的な値を算出している。

3. 実証実験

本章では、提案手法がどの程度ターゲット企業を正しく推薦できるかを確認するため、分類性能を評価した結果を紹介する。また、その結果としてABMとして有用なモデルであるかについての評価について述べる。

3.1 データセット

FORCAS上にある既存企業データと企業マスタデータを用いる。本実証実験においては、FORCASが保有する1社あたり約1,500個の特徴量を持つ合計約150万社の企業データに対し、業種が異なる4種類の既存企業のデータを、当該業種が対象とするであろう企業群を想定して作成し検証を行う（以下、検証データ）。異なる業種に所属するデータA～Dに対して検証を行うことで、提案手法の業種に対する頑健性を確認し、ABMツールとしての本システムの妥当性を検証する。なお、本稿における業種は筆者らが独自に命名したものであり、既存ユーザの企業の業界に合うよう設定したものである。

企業マスタデータの特徴量には従業員数、業界区分、各企業が利用していると推定される利用サービス（例：広告サービス、チャットサービス）、また独自に作成しているシナリオ（例：北米進出企業、増収企業）等が含まれる。利用サービスやシナリオは二値のデータであり、従業員数のような数値データもビン分割により二値データに変換し、すべての特徴量を二値のデータとして扱う。また、ある程度の規模の企業を分析したいというFORCASユーザの要望とモデル推定および予測の計算コストを勘案し、実際に推薦対象となる企業は約11万社となっている。検証データに含まれる既存企業数を表1に示す。

表1 検証データの概要

ID	業種	既存企業数
A	インターネットサービス	約 200
B	電気機器	約 900
C	情報通信	約 2000
D	工業製品	約 100

3.2 実験条件

NBEM, ロジスティック回帰 (以下, LR), および勾配ブースティング決定木 (以下, GBDT) の3種類を提案手法であるL2LRと比較する. 既存企業を正例, それ以外の企業を負例として3分割の交差検証を実施する. 評価指標としてはAUC (Area Under the Curve) と Precision@Nの2種類を採用した. ここで, Precision@Nとは推薦した上位N個の企業における既存企業の割合で定義される評価指標である. 本項において, この2種類の評価指標を採用した理由は,

- 業界で標準的に使われているAUCにより大域的な精度を検証
- Precision@Nで推薦結果ランキングの上位のみという条件付きの局所的な精度を検証

という2点についての検証を行いたいためである. Precision@Nを採用した理由は, FORCASのターゲット企業推薦結果画面の最初に現れる推薦結果が正しいものであることをできるだけ高精度にするためである. これは実際に推薦された企業に対しマーケティング・営業活動を行いたいというユーザーニーズの特に強い推薦結果ランキングの上位をより局所的かつ重点的に評価したいがためである. 既存企業のデータにおける教師ラベルの割合はおおよそ正例:負例=1:1000と不均衡になっている. また, L2LRにおける正則化項のパラメータはグリッドサーチを用い, 最も精度が高くなる値を採用した.

3.2 実験結果

提案手法をNBEM, LR, GBDTと比較した. その結果, 全体的な傾向として

- AUCについてはL2LRが最も良い
- Precision@20, 100についてもL2LRが良い傾向にある

ことが確認された (表2). これらの精度改善の結果を受け, 我々はL2LRをシステムとして運用していくモデルとして採用することを決定した.

表2 各手法のパフォーマンス比較

ID	モデル	AUC	Precision@20	Precision@100	Precision@500
A	NBEM	0.96	0.20	0.11	0.08
	GBDT	0.93	0.40	0.31	0.19
	LR	0.81	0.25	0.28	0.14
	L2LR	0.98	0.55	0.35	0.22
B	NBEM	0.88	0.60	0.35	0.21
	GBDT	0.86	0.60	0.58	0.37
	LR	0.85	0.55	0.47	0.32
	L2LR	0.92	0.55	0.60	0.39
C	NBEM	0.90	0.10	0.24	0.27
	GBDT	0.87	1.00	0.75	0.50
	LR	0.94	0.75	0.69	0.52
	L2LR	0.96	0.70	0.73	0.61
D	NBEM	0.93	0.15	0.08	0.05
	GBDT	0.84	0.25	0.18	0.08
	LR	0.77	0.20	0.12	0.04
	L2LR	0.97	0.30	0.21	0.09

3.2 実証実験から得られた知見と考察, およびAccount-Based Marketingとしての評価

本節では実証実験から得られた知見および考察をまとめる。特に, 独自にモデルを開発するL2LRがなぜ最も良い結果となったかについての考察を行う。

提案手法であるL2LRのAUCが全体的にLRと比べて高い。これはL2正則化による過学習の抑制が効果を発揮し, 精度を向上させたためであると考えられる。同様にL2LRはGBDTやNBEMに比べてもAUCが高い傾向にある。推薦結果ランキングの上位のみという条件付きの局所的な精度を表すPrecision@Nにおいても同様の傾向が出ており, これにより, 実際に推薦された企業に対しマーケティングや営業活動を行いたいというユーザーニーズの特に強い推薦結果ランキングの上位に関しても, 十分に実務に耐え得る精度であると考ええる。

一般に高いパフォーマンスを発揮する手法として知られているGBDTが、本実験において高いパフォーマンスを発揮しなかった原因としては特徴量間の相互作用を捉えられるGBDTの良さが発揮できなかったことに起因すると考えられる。たとえば、すべての特徴量はすべて0か1の二値で表現されており、そのため最適な分岐点を探索できず高精度の分類器として学習しづらいのではないかと考えられる。あるいは特徴量の性質から深い木構造になりやすく過学習を起こしやすいのではないとも考えられる。詳しい原因を知るためにはさらなる研究が必要である。また、検証データCについてはLRとLRL2の評価指標の差が大きく変わらない結果となった。Cは既存企業数が約2,000とほかのデータに比べて多く、既存企業数とパフォーマンスの関係について関連性があると考えられ、この点に関するより詳細な研究が必要である。

次にABMとしての評価について述べる。検証データA～Dまでの既存企業データは異なる業種に所属するデータであった。ABMを実践するためのモデルとして提案手法であるL2LRは業種の違いに対して頑健であり有効に動作すると考える。したがって、業種が異なる企業に属するユーザであっても、提案手法を使用することに問題はなく高いパフォーマンスを発揮することが期待される。一方、今後、本システムが想定するユーザの業種が拡大するにつれ、本研究において用いた検証データもアップデートしていかなければならない。また、アカウントスコアを算出するためのモデルとして、提案手法であるL2LRはNBEMやGBDT、そしてLRと比べても解釈性が同程度、あるいはそれ以上に高い。NBEMにおいてみられた特徴量間の正相関の効果によりアカウントスコアの値が嵩上げされる問題も表面化することはなく、業種によらずに安定的なパフォーマンスを発揮している。したがって、ABMのモデルとして本提案手法を採用することは妥当であると考えられる。

4. まとめと今後の展望、および課題

本稿では、既存企業データを分析し、成約確度の高い企業をターゲット企業として推薦するシステムの概要について説明した。既存手法として、ナイーブベイズを拡張した手法（NBEM）を用いて企業のアカウントスコアを算出し、スムージングにより特徴量の重要度を補正する手法を実装していたが、特徴量間の正相関の効果によりアカウントスコアの値が嵩上げされるという問題が生じ、結果として推薦されるターゲット企業に偏りが出てしまう場合があり、この問題を解決するため、L2正則化項付きのロジスティック回帰モデル（L2LR）を用いて企業類似度を算出し、この類似度に基づいた推薦アルゴリズムを提案した。提案手法の有効性を評価した結果、提案手法がNBEMや一般的に高精度と評価されるGBDTと同等以上のAUC、およびPrecision@Nを達成することを確認した。

今後は、解釈性ある、より高い予測精度のモデルを構築していく手法を検討したい。最終的なビジネス施策の実行を担うセールスやマーケティング担当者が納得感をもってビジネスに邁進できるようにするためには、機械学習モデルの出力した結果が機械学習の素人であっても理解できることが肝要である。機械学習、および人工知能における解釈性の研究は日進月歩の状況ではあるが、現状その問題が解決されているとは言いがたい[16]。そのため、我々自身が自分たち自身のニーズに応えられるような解釈性の高いモデルあるいはそのフレームワークを構築する価値があると考えられる。

また、現状、特段の特徴量選択のアルゴリズムを本システムに搭載していないが、今後の機能として重要な特徴量の自動抽出を検討したい。本システムには約1,500個の特徴量が搭載されている。数多くの特徴量があるおかげで各企業を特徴量空間においてより正確に表現し、その類似度をうまく計算することができている一方、L2正則化項付きモデルで高い精度が出たことからアカウントスコアや重要度に寄与しない不要な特徴量が存在することも事実である。これらの特徴量を自動的に削除するアルゴリズムを組み込み、ユーザに対し見る必要のない情報は出さないというユーザエクスペリエンスを提供していきたいと考えている。

参考文献

- 1) Bacon, A. : Strategic Account-Based Marketing : How to Tame This Beast, Management for Professionals, in : Uwe G. Seebacher (ed.), B2B Marketing, chapter 17, pp.419-435, Springer (2021).
- 2) Burgess, B. and Munn, D. : A Practitioner's Guide to Account-Based Marketing : Accelerating Growth in Strategic Accounts, Kogan Page (2017).
- 3) 庭山一郎：究極のBtoBマーケティング ABM（アカウントベースドマーケティング），日経BP社（2016）。
- 4) （株）シンフォニーマーケティング：アカウント・ベースド・マーケティング（ABM）とは、<https://www.symphony-marketing.co.jp/abm/about/>（2021年8月29日現在）
- 5) 田口慎吾：データとテクノロジーの力で顧客を再定義するーB2Bマーケティングの成果を最大化する『ABM』の基本概念とその実践事例ー，MarkeZine Day 2021 Spring（2021）。
- 6) Oracle Corporation : Account-Based Marketing Handbook, <https://go.oracle.com/LP=79765>（2021年8月29日現在）
- 7) Day, D. and Shi, S. : Automated and Scalable : Account-Based B2B Marketing for Startup Companies, Journal of Business Theory and Practice 8, p16, 10.22158/jbtp.v8n2p16（2020）。
- 8) 早川敦士，北内 啓：Account-Based Marketingのためのターゲット企業推薦システムの構築，人工知能学会（2019）。
- 9) Kumar, G. P., and Rajasekhar, K. : Account-based Marketing in B2B industry, Journal of Interdisciplinary Cycle Research, Volume XII, Issue II, February/2020, ISSN NO : 0022-1945（2020）。
- 10) Paavola, A. : Designing an Account-based Marketing Program, Master's Thesis, Lappeenranta University of Technology（2017）。
- 11) 河村一輝，諏訪博彦，小川祐樹，荒川 豊，安本慶一，太田敏澄：飲食店向け不動産営業を支援する申込み顧客推薦モデルの提案，人工知能学会（2017）。
- 12) 中山義人，森 雅広，斎藤 忍，成末義哲，森川博之：営業活動における意思決定システムの適用と評価，In IEICE Conferences Archives. The Institute of Electronics, Information and Communication Engineers（2019）。
- 13) Saini, N. K. and Sharma, H. : Salesforce Einstein : Artificial Intelligence for Customer Success Platform, International Journal of Scientific Research & Engineering Trends, Vol.6, Issue 3, May-June-2020, ISSN (Online) : 2395-566X（2020）。
- 14) 福田康隆：THE MODEL，翔泳社（2019）。
- 15) Tibshirani, R. : Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society : Series B (Methodological) 58.1, pp.267-288（1996）。
- 16) Molnar, C. : Interpretable Machine Learning A Guide for Making Black Box Models Explainable, Lulu.com（2020）。

新井和弥（非会員）kazuya.arai@uzabase.com

2020年名古屋工業大学大学院工学研究科博士前期課程修了。修士（工学）。2020年（株）FORCAS入社。2021年統合により（株）ユーザベースへ転籍。B2B事業向け顧客戦略プラットフォームFORCASの開発に従事。

北内 啓（非会員）akira.kitauchi@uzabase.com

1998年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。同年（株）NTTデータ入社。2014年（株）ユーザベース入社。以来、自社サービスのアルゴリズム開発に従事。2017年（株）FORCASに転籍。2021年（株）ニューズピックスに転籍。推薦システム開発に従事。

高柳慎一（正会員）shinichi.takayanagi@uzabase.com

2020年総合研究大学大学院複合科学研究科統計科学専攻博士課程修了。博士（統計科学）。2020年（株）FORCAS入社。2021年統合により（株）ユーザベースへ転籍。B2B事業向け顧客戦略プラットフォームFORCASの開発に従事。徳島大学客員准教授。情報処理学会ビッグデータ解析のビジネス実務利活用研究グループ幹事を兼任。

早川敦士（非会員）atsushi.hayakawa@uzabase.com

2015年電気通信大学大学院情報理工学研究科博士前期課程修了。修士（工学）。2018年（株）FORCAS入社。2021年統合により（株）ユーザベースへ転籍。B2B事業向け顧客戦略プラットフォームFORCASの開発に従事。徳島大学客員准教授を兼任。

林 樹永（非会員）tatsunaga.hayashi@uzabase.com

2017年慶應義塾大学大学院理工学研究科博士前期課程修了。修士（理学）。2019年（株）FORCAS入社。2021年統合により（株）ユーザベースへ転籍。B2B事業向け顧客戦略プラットフォームFORCASの開発に従事。

長田怜士（非会員）ryoji.nagata@uzabase.com

2015年大阪電気通信大学情報通信工学部情報工学科卒業。2020年（株）FORCAS入社。2021年統合により（株）ユーザベースへ転籍。B2Bに特化した顧客戦略プラットフォームFORCASの開発に従事。

受付日：2021年8月30日

採録日：2021年10月15日

編集担当：戸田貴久（電気通信大学）