

Account-Based Marketing のためのターゲット企業推薦システムの構築

Development of A Target Company Recommendation System for Account-Based Marketing

早川 敦士^{*1}
Atsushi Hayakawa

北内 啓^{*1}
Akira Kitauchi

^{*1} 株式会社 FORCAS
FORCAS, Inc.

In this paper, we developed a system for the new B2B marketing method ABM (account-based marketing). The system recommends target companies as future potential customers by analyzing current customers. There are two requirements in recommending target companies. (1) Users can grasp the impact of future importance on prediction score. (2) Users can update the model by modifying the importance of features without affecting that of other features. We propose a model that extends Naive Bayes classifier with a modified smoothing method. The experimental results show that the classification accuracy of our proposed model achieves AUC equal to or better than logistic regression and GBDT.

1. はじめに

新しい B2B マーケティング手法として ABM (account-based marketing, アカウントベースマーケティング) の活用が広がっている。アメリカに本社をおくアドバイザーファーム ITSMA が 2003 年に ABM を初めて提唱した [Bev 17]。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誕生している [庭山 16]。シンフォニーマーケティング株式会社は「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り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 [シンフォニーマーケティング 19]。株式会社 FORCAS 代表の佐久間は ABM を「ターゲットアカウントを定義し、アカウント別に営業、マーケティング情報を集約・分析し、アカウント別に営業、マーケティング組織を再編成し、ターゲットアカウントの LTV 最大化を目指すマーケティング」と位置づけている [佐久間 17]。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のが大きな違いである [庭山 16]。

成約確度が高いと推測されるターゲット企業の特長には、自社の営業データ(特に成約済みの顧客データ)と豊富な属性データをもつ企業マスター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マグラフィックと呼ばれるデータがよく利用されるが、企業が利用するサービス、求人の募集職種といっ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の生産性を向上する企業が増えてきており、ABM の重要性が高まっている。

我々は、成約済みの企業データ(以下、既存顧客データとよぶ)を分析して既存顧客以外の企業(以下、潜在企業とよぶ)の中から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を構築した。ターゲット企業の推薦を分類タスクととらえ、ナイーブベイズを用いて特徴量の重要度を算出し、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指標とともに提示することが可能となる。また、ユーザーが特徴量の重要度を変更でき、他の特徴量に影響を

与えずに潜在企業の予測スコアを再計算することも可能となる。提案手法の分類精度を評価し、特徴量間の依存関係を考慮するロジスティック回帰、GBDT と同等以上の精度を達成することを確認する。

2. システムの概要

本システムの構成を図 1 に示す。システムは大きく 3 つのステップからなる。(1) ユーザーがアップロードした既存顧客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、ファーマグラフィックデータおよび行動解析的なデータの特徴量としてモデルを学習する。(2) モデルをもとに潜在企業の成約確度を企業スコアとして予測し、予測スコアの高い順に提示することでターゲット企業を推薦する。また各企業がもつ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ーが理解できる。(3) マーケターの経験や意思に応じてユーザーが特徴量の重要度を変更でき、その変更内容をもとにモデルを再学習して潜在企業の予測スコアを更新する。

本システムが上記の 3 つのステップを持つ理由を述べる。(1) 日本国内には同名の企業が存在したり、ユーザーが管理する既存顧客データにおける企業名と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顧客データの対応付けができていて、モデルの学習に必要なデータを作ることができる。(2) ターゲット企業に集中し、優先順位をつけるために必要な予測スコアである。(3) 既存顧客の中には、経営者どうしが親しいなどの外部要因による獲得や、現在とは異なるマーケティング戦略による獲得が含まれている場合がある。そのため、マーケターや営業の意思や経験をターゲット企業の推薦アルゴリズムに反映する仕組みが必要である。既存顧客データから意思や経験に反する企業を除外する方法では、企業に結びつく特徴量は複数存在するので戦略上重要な特徴量も同時に失う可能性がある。既存顧客データに存在しない特徴量を重視するような戦略をモデルに反映することもできない。したがって、特徴量の重要度を直接変更する方法を提案する。

連絡先: 早川敦士, 株式会社 FORCAS,
atsushi.hayakawa@uzabase.com

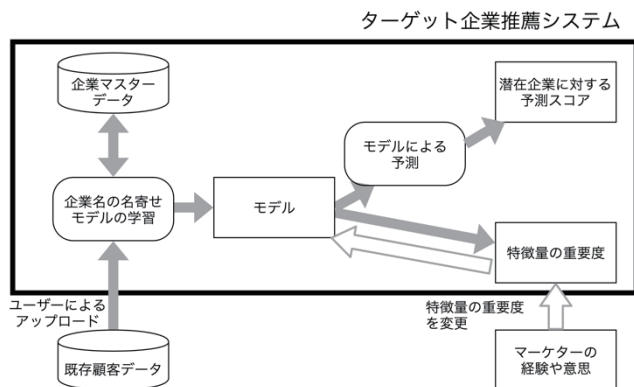


図 1: ターゲット企業推薦システムの構成

3. 問題設定

本システムにおいて、自社の既存顧客データを用いてターゲット企業を推薦する際に生じる以下の課題を解決する手法を提案する。なお、ターゲット企業の推薦には分類、回帰、協調フィルタリングなど様々な手法が存在するが、本研究では2値分類タスクととらえ、分類スコアをもとに潜在企業の予測スコアを算出する。また、企業の属性データのほとんどが質的データであるため、従業員数等の量的データを2値データに変換することで、すべての特微量を2値データとして扱う。

課題1. 特微量の重要度が予測スコアの増減に与える影響を解釈しやすいこと

課題2. 推定時に各特微量が相互に影響を与えないこと

前述の通り、本システムではターゲット企業がもつ特微量の重要度を予測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と営業の連携が不可欠となる。マーケティングから営業にターゲット企業のリストを提供したとき、その根拠が明確でないと、営業側が押し付けられたと感じるなど信頼を得にくい。明確な根拠があれば、適切な営業手段を選択できる。そのためには、重要度が予測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(課題1)。また、特微量どうしの依存関係がある場合でも、特微量単体でどの程度重要かが分かるような指標がよい。すなわち、推定時に他の特微量の影響を受けないモデルがよい(課題2)。

また、ある特微量の重要度を変更したときに、他の特微量の重要度も変更されてしまうと、マーケティングや営業の意思や経験をうまく反映することができない。したがって、他の特微量に影響を与えずに潜在企業の予測スコアを再計算できるモデルが必要となる(課題2)。

4. 関連研究

河村らは、勘と経験に頼っていた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[河村17]。この研究では、ベテランな営業への聞き取り調査を2度実施した。1度目の調査により作成した特微量に基づくモデルに対しての推定結果を提示し、フィードバック(2度目の調査)を得て特微量を作り、営業の経験を加味したモデルを構築した。特微量の追加は、現在のマーケティング戦略に基づくターゲット企業の予測精度向上には貢献するが、マーケティング戦略を変更した場合の予測スコア算出には対応できず、課題2を解決できない。

Takahamaらは、人の知識をモデルに反映する研究として、誤分類されやすい画像データを分類できるようにするためにクラウドワーカーに対して特微量定義に一致するかどうかを答えさせて特徴ベクトルを作成した[Takahama18]。この方法では、提示された企業の属性情報を確認しなければ、自社のマーケティング戦略に一致するかを判断できず、課題2に対応できない。つまり、属性情報を個別に判断しているのと同様である。意思や経験を組み込む必要があり現在のデータで解決できない問題に対しては、所与のデータを越えた一般化を実現するために、そのデータを越えた知識や仮説を学習者は盛り込む必要があるとしている[Domingos12]。

前述したように、本研究ではターゲット企業の推薦を2値分類タスクととらえる。代表的な分類手法であるロジスティック回帰は回帰係数を特微量の重要度とし、大きさを変えることによって該当する企業の予測スコアも変えることができ課題1の要件を満たすが、回帰係数の推定には他の特微量の影響を受けるため課題2の要件を満たさない。Gradient Boosting Decision Tree (GBDT)も特微量の重要度を算出できるが、重要度の大きさを意図的に変えることは木の構造を変えることになるので、変えた特微量の下位ノードに影響を与えてしまい、課題2の要件を満たさない。

5. 提案手法

課題1,2を解決するため、ナイーブベイズを拡張した手法を用いる。 x_i を特微量として、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。

$$P(y|\mathbf{x}) = P(y) \prod_i P(x_i|y)$$

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 $R(\mathbf{x})$ は次式ようになる。この値を潜在企業の予測スコアとする。

$$\begin{aligned} R(\mathbf{x}) &= \frac{P(y_1|\mathbf{x})}{P(y_2|\mathbf{x})} = \frac{P(y_1) \prod_i P(x_i|y_1)}{P(y_2) \prod_i P(x_i|y_2)} \\ &= \frac{P(y_1)}{P(y_2)} \prod_i \frac{P(x_i|y_1)}{P(x_i|y_2)} \end{aligned}$$

また $R(\mathbf{x})$ の中の下記の部分 $S(x_i)$ を特徴 x_i の重要度とする。

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

$S(x_i)$ は、既存企業における特徴 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と言える。 $S(x_i)$ は x_i のみに依存するため、課題2の要件を満たす。なお、 $S(x_i)$ を人為的に変更した場合、それに応じて $R(\mathbf{x})$ も変わり、 $S(x_i)$ 以外の特微量の重要度には影響を与えない。

$P(x_i|y_j)$ は次式によって算出する。

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

$N(x_i, y_j)$ はクラス y_j において特徴 x_i をもつ企業数、 $N(y_j)$ 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。 α ($0 < \alpha < 1$)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 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き

と同じ値となる。上記 2 点の性質から提案手法は課題 1 の要件を満たす。

6. 実験

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。

6.1 データセット

FORCAS¹ 上にある既存顧客データと企業マスターデータを用いる。企業マスターデータの特徴量は 1,677 種類あり、従業員数、外部利用サービス等が含まれる。従業員数のような数値データはビン分割により 2 値データに変換している。分析対象となる企業は約 12 万社ある。マーケティング戦略の異なる 4 種類の既存顧客データを用意した。それぞれに含まれる企業数は Data1 が 95 社、Data2 が 167 社、Data3 が 349 社、Data4 が 817 社である。

6.2 実験条件

scikit-learn に含まれる分類手法のうち、ラプラススムージングによるナイーブベイズ(NB)、ロジスティック回帰(LR)、および GBDT の 3 種類を提案手法と比較する。既存顧客を正例、それ以外を負例として 3 分割の交差検証を行った。評価尺度には AUC を用いた。

6.3 実験結果

実験結果を表 1 に示す。LR と GBDT は Data1, Data3 において AUC が 0.9 を下回ったのに対し、提案手法と NB はいずれのデータセットにおいても 0.9 以上であった。提案手法と NB の AUC は大きな差がなかった。ABM での営業データは既存顧客数が少なく不均衡データであるため、提案手法の精度が LR, GBDT を上回ったと考える。

表 1: 各手法による AUC の比較

手法	Data1	Data2	Data3	Data4
提案手法	0.929	0.986	0.929	0.920
NB	0.904	0.983	0.916	0.924
LR	0.877	0.985	0.881	0.925
GBDT	0.746	0.989	0.822	0.874

7. おわりに

既存顧客データを分析して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を構築した。本システムは予測したターゲット企業の根拠を解釈しやすく、マーケティングの経験や意思をターゲット企業の予測に反映できる特徴をもつ。それらの特徴を実現するため、ナイーブベイズを拡張することで企業の予測スコアを算出し、スムージングにより特徴量の重要度を補正する手法を提案した。提案手法の有効性を評価した結果、既存顧客データの数が少ない場合でも精度の高い予測ができることを確認した。

今後は、特徴量の重要度を変更したときの分類精度についても評価する予定である。たとえば、ある時点までの既存顧客データとそれ以降の既存顧客データの 2 種類を用意し、前者をもとに算出した特徴量の重要度を変更してターゲット企業を予測し、後者を用いて評価するといった方法を考えている。

参考文献

- [Bev 17] Bev Burgess, Dave Munn: A Practitioner's Guide to Account-Based Marketing: Accelerating Growth in Strategic Accounts (English Edition), Kogan Page (2017).
- [庭山 16] 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).
- [シンフォニーマーケティング 19] 株式会社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/> (2019-1-17 最終閲覧).
- [佐久間 17] 佐久間衡: Account Based Marketing を実践する, MarkeZine 第 21 号 特集「戦略的 BtoB マーケティング」, 翔泳社 (2017).
- [河村 17] 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, Web インテリジェンスとインタラクション 2017, 人工知能学会 (2017).
- [Takahama 18] Ryusuke Takahama, Yukino Baba, Nobuyuki Shimizu, Sumio Fujita, Hisashi Kashima: AdaFlock: Adaptive Feature Discovery for Human-in-the-Loop Predictive Modeling, The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), pp. 1620-1626 (2018).
- [Domingos 12] Domingos, P.: A Few Useful Things to Know About Machine Learning, Communications of the ACM, Vol. 55, No. 10, pp. 78-87 (2012).

¹ <https://www.forcas.com/>