

# データマイニングビジネスにおける 最適化技術の活用

草野 隆史, 矢島 安敏

インターネットや計算機などIT技術の進歩により、顧客の購買履歴やWebの閲覧履歴といった大量のデータを扱うことのできる環境が整ってきた。同時に、このような大規模データに対応できるデータマイニング技術も登場してきた。その結果、企業はさまざまなマーケティング施策において、データに基づき予測モデルを構築し、より科学的な手法で施策の効率化を目指すことが可能になってきた。本稿では、弊社が行っている取り組みの中から「顧客コンタクト最適化」と「リスティング広告最適化」の二つを取り上げ、特に最適化部分を中心にその概要を紹介させていただく。

キーワード：データマイニング、最適化、区分線形関数、SOS 制約

## 1. はじめに

株式会社ブレインパッドは、データマイニング技術や最適化技術を活用して、企業内に蓄積された大量データに基づくマーケティングの意思決定支援を行うために2004年に創業された。現在、企業からのデータの分析依頼に対応する受託分析事業を中心に、自社開発のアルゴリズムを実装したASPサービスの提供や、分析関連のソフトウェア販売などを展開している。

その創業の背景には、インターネットやIT技術の普及・発展による企業の経営環境、特にマーケティング環境の変化がある。かつて、生活者が接触するメディアがマス中心であった際は、多くの企業では、マス広告を活用した製品やサービスの告知がマーケティングコミュニケーションにおいて大部分の比重を占めていたため、効率のよいマス広告を実現することが、効率のよいマーケティングを意味していた。

しかし、ネットやモバイルの普及とともに、生活者の接触メディアは非常に多様化し、従来のマス広告はその影響力が低下するとともに、正確に成果の測定ができるネット広告との対比で、その効果の不透明性が指摘されるようになっていく。さらに、期を同じくした、手数料の自由化とネット取引の容易化で増加した個人投資家からの要望を受け、企業における投資について厳しい説明責任が求められるようになっていく。

このような流れから、企業は、経営の効率化と説明性の向上を図るため、コミュニケーションチャンネルの変化と技術の進歩で蓄積が可能になった、生活者に関する大量なデータを活用したマーケティング上の合理的な意思決定を実施する必要が生まれている。

対して、これまでの日本企業における意思決定へのデータの活用は、製造現場を除くと非常に限定的であり、特に感性が問われたマーケティング領域では、人材や知見の蓄積に乏しい状況にあるため、これを支援することが弊社の事業となっている。

弊社の目指すものは、これらの事業を通して、データマイニングや最適化という技術を実ビジネスの現場で活用できる人材を育成するとともに、サービスの現場にハードサイエンスのアプローチを取り入れるサービス・サイエンス (Services Sciences) をビジネスのジャンルとして確立/認知拡大を行っていくことである。サービス・サイエンスは、今後のイノベーション創造に不可欠なものとして海外での積極的な取り組みが行われているが、日本での関心が相対的に低い。しかし、限られた経営資源や激化する国際競争の中で、競争力を確保していくためには不可欠な技術であり、その発展の一助として寄与したいと考えている。本稿では、これまでの弊社の実績から、最適化技術を実際のマーケティングに応用した事例を紹介する。

## 2. 顧客コンタクト最適化

弊社では、カード会社や総合通販会社などに向け、ダイレクトメールやカタログ送付といった施策の効果的な実施を支援してきた。過去の顧客の購買行動や、

くさの たかふみ, やじま やすとし  
(株)ブレインパッド  
〒141-0022 品川区東五反田 5-2-5

性別、年齢、職業といった顧客属性を独立変数にした商品への反応予測モデルを構築し高い反応が見込める顧客を抽出することで、投下可能な予算の中でより多くの反応を得ることが可能である。しかし、実際には複数の商品が同時に扱われたり、あるいは、顧客への推薦方法も郵送、電子メール、電話など多岐にわたる。さらに、商品ごとの可能販売数の上限や利用可能なチャネルの容量の制限もあり、これらの制約条件も考慮した上で最も効果的に顧客にコンタクトを行わなくてはならない。そのためには、最適化手法を用い、制約条件を満たした上で顧客とコンタクト方法の最適な組み合わせを求め、顧客コンタクト最適化が不可欠である。

我々の考える顧客コンタクト最適化とは、顧客一人ひとりに対して

- 商品
- チャネル
- タイミング

の各コンタクト要素を、実務上の制約条件を満たし、かつ売上げや利益といった目的を最大（あるいはコストを最小）にするように決定することである。ここで、チャネルとは、郵便、電話、電子メールといった顧客にコンタクトする手段であり、また、タイミングと呼んでいるのは、例えば四半期ごとに4回カタログを送付する施策などにおける「いつ」コンタクトするかといった時間的な要素である。

まず前提として、過去の顧客の購買履歴やダイレクトメールなどへの反応データを基に予測モデルが構築され、顧客と商品、チャネル、タイミングの各要素との組み合わせで、顧客が反応する確率や期待売上金額などが推定されていると仮定する。その上で、施策全体の売上や利益の最大化、あるいは全体コスト最小化などが最適化の目的である。

制約条件には施策全体に関係するものとして、大きく分けて「予算」、「コンタクト数」および「売上げや利益」に関係するものがある。予算に関する制約条件としては、まず、チャネルコストの総予算制約が考えられる。それに加え、投下コストの平準化といった目的でチャネル、商品、タイミングの各要素ごとにコスト制約が考えられる。コンタクト数に関する制約も予算制約の場合と同様で、商品ごとやタイミングごとにコンタクト数の上限制約が考えられる。さらに、最適化の目的をコストの最小化とした場合には、期待売上や期待利益に対する下限制約条件が通常設定される。

このような制約条件はいずれも施策全体に課されたものであることから、以降では、グローバルな制約条件と呼ぶ。

一方、顧客一人ひとりに対する制約条件も考えられる。例えば、一人にコンタクトする回数に上限を設定し、多くのコンタクトが一人の顧客に集中することがないようにしたり、あるいは、どの顧客にも必ずコンタクトしたい場合にはコンタクト回数の下限が設定される。他にも、推薦する商品数の上下限や、商品、チャネル、タイミングの組み合わせに関係した実務上のコンタクトルールに対応した制約条件が考えられる。以降、こういった制約をローカルな制約条件と呼ぶことにする。

次節では、以上の制約条件を考慮した最適化モデルとその効率的な最適化について述べる。

## 2.1 最適化モデル

まず、最も単純な最適化モデルの例として、ある総チャネルコストの中で、各顧客へのコンタクトを高々  $n$  回までとした場合、期待利益が最も大きくなる顧客へのコンタクト方法（すなわち、どの商品を、どのチャネルで、どのタイミングで）を求める問題を定式化する。

そこで、顧客、商品、チャネルおよび期間に対応する添え字集合をそれぞれ  $I, J, K, L$  とし、顧客  $i \in I$  が商品  $j \in J$  をチャネル  $k \in K$  を使ってタイミング  $l \in L$  でコンタクトした場合、得られる期待利益を  $v_{ijkl}$  とする。さらに、チャネル  $k$  を使い1回コンタクトするごとに発生するコストを  $c_k$ 、また総チャネルコストの上限を  $TC$  と表記する。

最適化の変数を  $x_{ijkl}$  とし、顧客  $i$  に商品  $j$  をチャネル  $k$  を使ってタイミング  $l$  にコンタクトする場合  $x_{ijkl}=1$ 、コンタクトしない場合0となる0-1変数と約束すれば、最適化問題は次のようにモデル化される。

$$\begin{array}{ll}
 \text{最大化} & \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} \sum_{l \in L} v_{ijkl} x_{ijkl} \\
 \text{制約} & \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} \sum_{l \in L} c_k x_{ijkl} \leq TC, \\
 & \sum_{j \in J} \sum_{k \in K} \sum_{l \in L} x_{ijkl} \leq n, i \in I, \\
 & x_{ijkl} \in \{0, 1\}, i \in I, j \in J, k \in K, l \in L.
 \end{array} \tag{1}$$

このモデルで、総チャネルコストの上限制約は前節で述べたグローバルな制約条件であり、一方、顧客ごとのコンタクト回数の上限制約はローカルな制約条件に相当する。

一般的には、次のようにモデル化される。 $s \in S$  を制約式に関する添え字、また、 $s$  に対応した添え字集

合を  $J^s \subseteq J, K^s \subseteq K, L^s \subseteq L$  とすれば、グローバルな制約条件は一般的に

$$\sum_{i \in I} \sum_{j \in J^s} \sum_{k \in K^s} \sum_{l \in L^s} a_{ijkl}^s x_{ijkl} \leq A^s$$

と表現される。また、ローカルな制約条件は、 $t \in T$  を制約式に関する添え字、また、 $J^t \subseteq J, K^t \subseteq K, L^t \subseteq L$  として、

$$\sum_{j \in J^t} \sum_{k \in K^t} \sum_{l \in L^t} b_{ijkl}^t x_{ijkl} \leq B^t, \quad i \in I$$

と表現されるものである。

例えば、チャンネルごとにコストの上限を設定する場合、チャンネル  $k \in K$  のコストの上限を  $C_k$  と記せば

$$\sum_{i \in I} \sum_{j \in J} \sum_{l \in L} C_k x_{ijkl} \leq C_k, \quad k \in K$$

と記述される。これは、グローバルな制約条件となる。

一方で、顧客へのコンタクトが複数回可能な場合、すなわち問題(1)で  $n$  が 2 以上の場合には、通常、多くのコンタクト制約が必要となる。例えば、

- 1 期間内でのコンタクトは高々 1 回、
- 同じ商品のコンタクトは高々 1 回、
- ある期間にコンタクトを行った場合には次の期間にはコンタクトを行わない、あるいは連続してコンタクトを行う、

などで、これらはいずれも 0-1 変数と等式や不等式系で表現が可能である。

以上述べてきたことをまとめると、顧客コンタクト最適化は一般的に次のようにモデル化される。

$$\begin{array}{l} \text{最大化} \quad \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} \sum_{l \in L} v_{ijkl} x_{ijkl} \\ \text{制約} \quad \sum_{i \in I} \sum_{j \in J^s} \sum_{k \in K^s} \sum_{l \in L^s} a_{ijkl}^s x_{ijkl} \leq A^s, \quad s \in S, \\ \quad \sum_{j \in J^t} \sum_{k \in K^t} \sum_{l \in L^t} b_{ijkl}^t x_{ijkl} \leq B^t, \quad i \in I, t \in T, \\ \quad x_{ijkl} \in \{0, 1\}, \quad i \in I, j \in J, k \in K, l \in L. \end{array} \quad (2)$$

## 2.2 大規模な最適化

上で示した問題(2)は通常、商品数やチャンネル数と比べ顧客数  $|I|$  が非常に大きく、顧客数が百万を超える場合もある。このような場合には、変数の個数は優に数千万から 1 億以上となってしまふ。また、ローカルな制約条件は顧客に対応しているため個数が膨大となり、問題(2)は、変数、制約とも超大規模な最適化問題となってしまふ。したがって、このままのモデルを通常の計算機環境で扱うことは不可能である。

そこで、グローバルな制約条件をラグランジュ緩和することで、問題を規模の小さな部分問題に分解し最適化を行っている。すなわち、 $\lambda^s, s \in S$  をグローバルな制約条件に対するラグランジュ係数として緩和問題を作ると、問題(2)は顧客  $i$  ごとに分解された子

問題：

$$\begin{array}{l} \text{最大化} \quad \sum_{j \in J} \sum_{k \in K} \sum_{l \in L} \sum_{s \in S} (v_{ijkl} - \lambda^s a_{ijkl}^s) x_{ijkl} \\ \text{制約} \quad \sum_{j \in J^t} \sum_{k \in K^t} \sum_{l \in L^t} b_{ijkl}^t x_{ijkl} \leq B^t, \quad t \in T \\ \quad x_{ijkl} \in \{0, 1\}, \quad j \in J, k \in K, l \in L \end{array} \quad (3)$$

となる。これは、0-1 整数計画問題となっているが、多くの場合、変数が数十個の小規模の問題となり、制約条件によってはグリーディーに最適化が可能な場合もある。劣勾配法によりラグランジュ係数を更新しながら、上の子問題(3)を高速に最適化することで実用的な計算時間での最適化を実現している。

## 3. リスティング広告の最適化

### 3.1 ネット広告

インターネットにおける広告は、従来からのマス広告 4 媒体に並ぶ主要な広告のひとつに成長している。表 1 に示すように、すでにネット広告は広告費全体の 1 割以上を占めており、これは新聞広告に匹敵する [1]。バナー広告や電子メールによる広告はネット広告の代表例であるが、特に近年は、リスティング広告が、他のネット広告と比べて効果的にユーザに接触可能であること、また費用対効果が計測しやすいことなどの理由で年率で 20% 程度の成長を遂げている。しかし、広告の掲載位置が入札金額に左右されることや、課金方法がクリック課金となっているため、コスト管理が難しい側面がある。

ここで、簡単にリスティング広告の仕組みを述べる。リスティング広告とは、Yahoo! や Google などの検索サイトで、ユーザがキーワードを入力し検索を行った際に、検索結果とともにページの上部や右のスペースに表示される広告である。ユーザは、広告をクリックすることで広告主のサイトへと遷移し、もしそこで目的の商品やサービスが見つければ購買などが行われる。このように、購買などが行われ Web ページから成果が得られることをコンバージョンと呼び、広告

表 1 2008 年度の広告費 (電通 (2008) 「2008 年 (平成 20 年) 日本の広告費」より)

媒体	広告費(億円)	対前年比 (%)	構成比 (%)
TV	19,092	95.6	28.5
新聞	8,276	87.5	12.4
ネット広告 (リスティング広告)	6,983 (1,575)	116.3 (122.9)	10.4 (2.4)
雑誌	4,078	88.9	6.1
ラジオ	1,549	92.7	2.3
総広告費	66,926	95.3	100.0

の効果を計る指標としてよく用いられている。広告を出稿しようとする広告主は、あらかじめ、自社の商品やサービスと関連の深いキーワードに対して入札を行う。その後、入札したキーワードで検索が行われた場合、そのキーワードに入札を行った広告主の広告が表示されるという仕組みである。複数の広告主が同一のキーワードに入札を行っていた場合には、それぞれの広告主の広告が表示されるが、その際、入札金額の多寡（および広告の質）により検索結果のページ内での広告の掲載位置が決められる。基本的には入札金額が高いほど、ページ上部の目立つ位置に広告が表示される仕組みになっている。

この際、広告がクリックされるごとに、入札金額に応じた費用が広告主に発生する。したがって、高額で入札すればより目立つ位置に広告が表示されるため、結果として多くのユーザを自社のサイトへと誘導でき、コンバージョンの増加が期待できる一方で、クリックが増加したことで広告費用も増加するため、広告の効果（コンバージョン）と費用とのトレードオフを考慮し、適切に入札金額を決定することが求められている。すなわち、

- ある一定の広告費用の中でコンバージョン数を最大とする入札
- ある一定のコンバージョン数を確保した上で広告費用が最小となる入札
- 1コンバージョン当たりの広告費用の上限を満たした中で、コンバージョン数を最大にする入札といったことを考慮し、膨大な数のキーワードに対する適切な入札価格を決める必要がある。

さらに、入札のルールとして、キーワードによっては最低の入札金額が設定されている場合もあり、適切な入札金額を決めることは簡単ではない。特に、入札するキーワード数が数十万を超えるような企業の場合、非常に多くのキーワードの入札金額を手で決定することは不可能であり、以下で述べるようなデータマイニングと最適化技術による問題解決が不可避である。

### 3.2 最適化モデル

まず、過去の入札履歴を分析することで、キーワードごとにコンバージョン数やコストの予測モデルが構築されていることを仮定する。すなわち、キーワード  $i$  に対して1日に期待できるコンバージョン数およびコストが、入札金額  $b_i$  の関数として推定できており、それぞれ  $v_i(b_i)$  および  $c_i(b_i)$  と表記する。これらの関数は入札金額に関して非減少関数ではあるが、必ずし

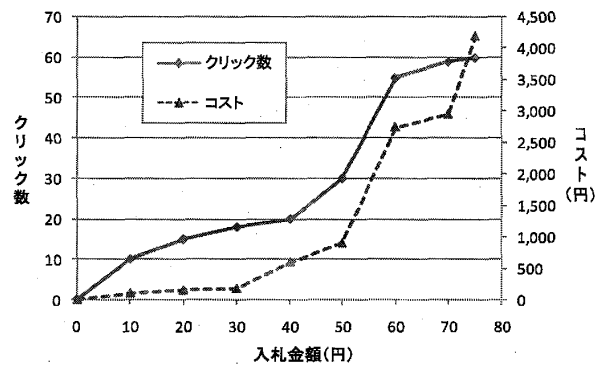


図1 クリック数、コストの予測モデル

も単純な線形関数にはならないことが過去の分析より明らかになっている。図1に、これらの関数の例を示した。

その上で、例えば1日の広告の予算を  $B$  とした場合、最もコンバージョンが獲得できるよう各キーワード  $i$  の入札価格  $b_i$  を求める最適化問題は

$$\begin{cases} \text{最大化} & \sum_{i \in I} v_i(b_i) \\ \text{制約} & \sum_{i \in I} c_i(b_i) \leq B, \\ & b_i \geq 0, i \in I \end{cases} \quad (4)$$

と記述される。ただし、 $I$  は全キーワードに対する添え字の集合である。

あるいは、予算制約の代わりに、1クリック当たりのコスト (CPC: Cost Per Click) の上限や、1コンバージョン当たりのコスト (CPA: Cost Per Acquisition) の上限も制約としてよく用いられる。例えば、CPA が  $P$  以下となる条件は

$$\frac{\sum_{i \in I} c_i(b_i)}{\sum_{i \in I} v_i(b_i)} \leq P,$$

と書き表せることから、この制約条件でコンバージョンを最大化する問題は

$$\begin{cases} \text{最大化} & \sum_{i \in I} v_i(b_i) \\ \text{制約} & \sum_{i \in I} c_i(b_i) - P \sum_{i \in I} v_i(b_i) \leq 0, \\ & b_i \geq 0, i \in I \end{cases} \quad (5)$$

となる。

以降では、最も単純なモデルとして問題(4)を取り上げ説明するが、実際には、予算制約や CPA 制約は、上で示したようなキーワード全体に対するものだけでなく、キーワードの部分集合に対する制約が複数組み合わせられることもある。

### 3.3 区分線形近似による最適化

前節で図示したように、コンバージョン数やコストは入札金額に対して非線形な関数であることから、最

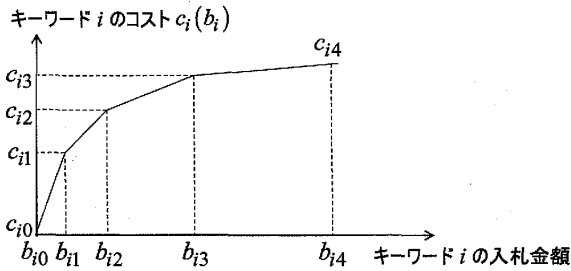


図2 区分線形関数

適化問題(4)や(5)は一般には非凸な最適化問題となってしまう、このままではキーワード数が数十万を超えるような大規模な問題を扱うことは困難である。そこで、各予測モデルを線形区分関数として推定することで、最適化問題を整数計画問題に帰着させ最適化を行っている。

ここでは、最適化問題(4)を例にして考える。まず、図2に示したように、キーワード  $i \in I$  ごとに入札金額を

$$b = 0, b_{i1}, b_{i2}, \dots, b_{ini}$$

に区分し、コストとコンバージョン数の予測関数  $c_i(\cdot)$ ,  $v_i(\cdot)$  を  $n_i + 1$  個の節点

$$(0, c_i(0)), (b_{i1}, c_i(b_{i1})), (b_{i2}, c_i(b_{i2})), \dots, (b_{ini}, c_i(b_{ini})), \\ (0, v_i(0)), (b_{i1}, v_i(b_{i1})), (b_{i2}, v_i(b_{i2})), \dots, (b_{ini}, v_i(b_{ini}))$$

を結んだ区分線形関数として扱うことにする。

その上で、各節点に対応した非負変数

$$x_{i0}, x_{i1}, \dots, x_{ini}, i \in I$$

を導入すれば、総予算  $B$  以下の範囲でコンバージョン数が最大となる入札金額を決定する問題は

$$\begin{cases} \text{最大化} & \sum_{i \in I} \sum_{j=0}^{n_i} v_i(b_{ij}) x_{ij} \\ \text{制約} & \sum_{i \in I} \sum_{j=0}^{n_i} c_i(b_{ij}) x_{ij} \leq B, \\ & \sum_{j=0}^{n_i} x_{ij} = 1, i \in I, \\ & x_{ij} \geq 0, i \in I, j = 0, 1, \dots, n_i, \\ & \{x_{i0} \leq x_{i1} \leq x_{i2} \leq \dots \leq x_{ini}\}_2, i \in I, \end{cases} \quad (6)$$

と表現できる。なお、 $x_{ij}^*$  をこの問題の最適解とすれば、キーワード  $i \in I$  に対する最適な入札金額は

$$b_i^* = \sum_{j=0}^{n_i} b_{ij} x_{ij}^*$$

と定めることができる。

また、問題(6)の制約条件

$$\{x_{i0} \leq x_{i1} \leq x_{i2} \leq \dots \leq x_{ini}\}_2$$

は2次の SOS (Special Ordered Set) 制約を表す。これは、非ゼロとなる要素がこの中で高々2つまで、

かつ  $\leq$  で指定されている順序に従って隣接することを要求する制約条件である。

例えば、次のような解

$$x_{i2}^* = 1/2, x_{i3}^* = 0, x_{i4}^* = 1/2$$

は非ゼロ要素が隣接しておらず、SOS 制約を満たしていない。これは、例えばコスト関数  $c_i(\cdot)$  であれば、隣接しない節点  $(b_{i2}, c_{i2})$  と  $(b_{i4}, c_{i4})$  を結んだ直線が用いられた場合に相当しており、このような解は SOS 制約によって排除されることになる。なお、2次の SOS 制約は分枝操作を行うことで考慮され、ほとんどの整数計画法のパッケージであれば、標準的にサポートしている機能の1つである。

入札方式の制限のひとつに、最低入札価格の制限がある。すなわち、もし入札を行うのであれば、ある一定の金額以上でなくてはならない制限である。例えば、キーワード  $i$  の最低入札価格を  $b_{i1}$  とすれば、変数  $x_{i0}$  に 0-1 整数制約を付ければよく、

$$\sum_{j=0}^{n_i} x_{ij} = 1, \{x_{i0} \leq x_{i1} \leq \dots \leq x_{ini}\}_2,$$

$$x_{i0} \in \{0, 1\}, x_{i1}, x_{i2}, \dots, x_{ini} \geq 0$$

と変更するだけである。

#### 4. おわりに

本稿では、弊社の最適化技術を用いた取り組みの中から、顧客コンタクト最適化とリスティング広告最適化について述べた。顧客コンタクト最適化では、顧客と商品・チャネル・タイミングの最も効果的な組み合わせを算出するため、極めて大規模な最適化が必要となることを述べた。最適化モデルは、顧客数の増加とともに、変数と制約条件ともに大規模なものになってしまうが、ラグランジュ緩和を用いて顧客ごとの子問題に分解することで、実用的な計算時間で最適化が可能である。また、リスティング広告の最適化では、一般には非凸なクリック数やコストの予測モデルを区分的に線形な関数として扱うことで、キーワードごとの最適な入札価格が SOS 制約のある最適化モデルに帰着できることを述べた。

本稿の中では割愛したが、いずれの例も、最適化を行うためには統計やデータマイニング技術を使い顧客の購買履歴や Web アクセスログなどの大量データを分析し、商品の購買確率や広告のクリック数の予測モデル構築プロセスが必要である。実際、マーケティングサイエンスなどの分野では、予測モデル構築などのデータ分析に多くの努力が注がれてきたのではないだ

ろうか。しかし、そこで得られたモデルや予測結果を合理的な意思決定に結びつけ、より効率的な施策を実現するためには、OR の分野で生み出された技術が必要で、最適化モデルはその一つであると考え。今後は、最適化技術とデータマイニングを融合した統合的なマーケティングソリューションの一つとして発展を

させたいと考えている。

#### 参考文献

- [1] 電通 (2008) 「2008 年 (平成 20 年) 日本の広告費」  
[http://www.dentsu.co.jp/marketing/adex/adex2008/\\_outline.html](http://www.dentsu.co.jp/marketing/adex/adex2008/_outline.html)