

意思決定のためのイベント・データ解析

— 個人別購買・販促記録からの最適ターゲティング・ ルールのマイニングを例として —

多種類、多量のイベントをその発生時刻と共に記録した時系列のデータを、イベント・データと呼びます。イベント・データは高度な解析技術によって、価格戦略や販促戦略など意思決定に直接役立てることができます。本稿では個別消費者へのマーケティング課題を例として、イベント・データを加工して最適な意思決定ルールを抽出するまでの課題解決プロセスをご紹介します。

1. 意思決定に役立つビッグデータ

Web上の行動データ、購買データ、あるいはセンサー・データなど、ビッグデータと呼ばれる大量のデータからビジネスに有用な知見を引き出す活動が盛んです。このトレンドはBigTable [1] に代表される莫大なデータ保存技術の発展と、隠れた知見を高精度で発見する機械学習やデータ・マイニング技術の飛躍的発展の双方によって引き起こされてきました。ビッグデータに対するデータ・マイニングは興味深い知見をもたらしてくれます。一方で知見が面白いだけでは、経営課題や販促、マーケティング課題などを解決することはできません。実際の課題解決においては、意思決定に役立つデータから優先的に記録、分析していく必要があります。

ビジネス上の意思決定に役立つビッグデータには共通の特徴があります。それは、i) 時系列データであり因果関係の統計的推測がある程度可能なこと、ii) 各日時に発生する個々のイベントがその属性と共に記録されていること、iii) 性質の異なる個体ごとにイベントが記録されていることです。これら i) ii) iii) の特性は日別の個人ID別 Point-Of-Sales (ID-POS) データを想像すると理解しやすいでしょう。ID-POS データでは各消費者の購買イベントが時系列に記録されています。それぞれの購買に現れる商品の種類は多様であり、例えば「商品、購買点数」の組を購買イベントの属性と見なすことができます。多くの消費者に対するデータがあることで、性質の似た消費者を同一グループにまとめてセグメンテーションして、セグメントごとに共通する購買行動パターンを見つけやすくなります。

本稿では属性付きのイベントを多く含む時系列ビッグデータをイベント・データと呼ぶこととし、イベント・データの意思決定への活用方法をご紹介します。図1はイベント・データを活用してより良い意思決定を行うまでのデータ加工と分析に関するプロセスを示したものです。2章から4章ではこのプロセスを具体

化して解説します。

分かりやすさのために全章を通して購買、販促イベントを題材にしますが、イベント・データの解析が役に立つのは消費者行動の分析と販売戦略、マーケティング戦略の策定ではありません。最後の5章ではイベント・データの活用における今後の展望を述べています。

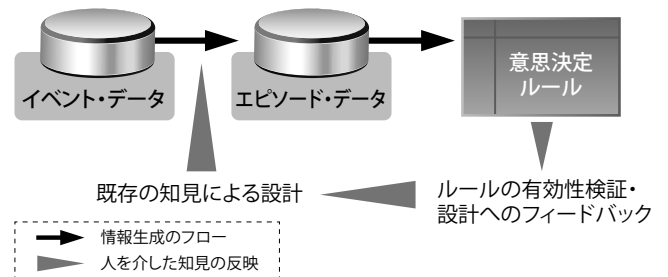


図1. イベント・データからの意思決定プロセス

2. イベント・データと予測

1 単位のイベントは個体の識別子 (ID)、発生時刻、属性という大きく分けて3種類の情報により定義されます。また、イベント・データは通常多量のイベントを並べたもので、消費行動分析においては表1のような形式をとります。

表1ではIDは顧客番号を表す文字列で、発生時刻には日付が記載されていますが、時刻に時間、分、秒を含めた方が役立つ

表1. 購買、販促イベント・データの例

ID	発生時刻	イベント属性		
		種類	チャネル	詳細
AA0001	2013/4/1	C	ダイレクト・メール	商品Xを紹介
AA0001	2013/4/3	R	コールセンター	商品Xについて問い合わせ
AA0001	2013/4/17	R	店舗	商品Xを購入;売上 ¥39,800
AA0002	2013/4/1	C	ダイレクト・メール	商品Yを紹介
AA0002	2013/4/23	R	Webサイト	販売ページを訪問;購買はなし
AA0002	2013/4/30	C	eメール	時限値下げ;1週間以内に買えば商品Yを25%割引!
AA0002	2013/5/5	R	Webサイト	商品Yを購入;売上 ¥16,800

つ場合もあります。形式が定まっている ID や発生時刻と違い、イベントの属性は厳密なルールによって記述されるとは限りません。明確に次元の分かっている構造化数値データであれば、その分のフィールドを準備してリレーショナル・データベースとして保存するのが良いでしょう。しかし非構造化情報も残す方が後の分析で役立つ場合があります。この場合、自然言語による自由文形式を利用することも考えられます。自由文を用いた場合には、分析の段階で自由形式をどのように加工して定型フォーマットに変換し、利用するかが担当者の腕の見せどころと言えます。

イベント属性をどう記述するかが、後述する最適意思決定の鍵となります。表 1 において「種類」という名称のフィールドがあり、ここでは C か R の文字が記載されています。C は Contact の頭文字で、販促活動やマーケティング・キャンペーンなどの企業側がお客様に「コンタクト」するイベントを表しています。一方、R は Response の頭文字で、お客様からの何らかの反応イベントを表します。個々の反応は必ずしも購買である必要はなく、購買に至る前に起きる顧客行動イベントも記録した方が有益です。購買はいくつかの中間的な反応の積み重ねによって起きることがあるためです。例えば 2013 年 4 月 3 日に ID='AA0001' のお客様が商品 X に関してコールセンターに問い合わせています。問い合わせは商品 X に関心がある証拠ですから、その記録は重要な情報としてその後の販促活動や店舗での対応に役立つべきでしょう。

表 1 ではチャンネルという定型文字列フィールドを用意していますが、チャンネルは最も重要なイベント属性です。Contact イベントのチャンネルとは、e メールやダイレクト・メールといった販促手法の区分です。Response イベントでは、チャンネルは購買や問い合わせの発生した店舗や組織の区分です。詳細なイベント属性を扱う前に、まずは各チャンネルの定義と、チャンネル別のデータ取得が重要です。例えば IBM Campaign という製品では Contact イベントと Response イベントをチャンネル別に記録し、どのキャンペーンが効果的だったかを事後分析できます。

分析の枠組みに進む前に、表 1 から消費者行動の背後にあるストーリーを想像してみましょう。4 月 1 日に ID='AA0001' のお客様にダイレクト・メールで商品 X を紹介すると、すぐに 4 月 3 日には該当商品に対するコールセンターへの問い合わせがありました。その後 2 週間して 4 月 17 日に商品 X が売れています。一方、ID='AA0002' のお客様には 4 月 1 日に同じくダイレクト・メールで、別の商品 Y を紹介しました。このお客様は 3 週間後の 4 月 23 日にオンライン・ショッピング・サイトを訪れましたが、特に何も買われませんでした。その 1 週間後、今度はダイレクト・メールよりも販促コストの低い e メールで 1 週間以内限定の値下げキャンペーンをお知らせしました。e メール送付後である 5 月 5 日には商品 Y が実際に売れたようです。このようなストーリーを想像すると、次に並べたような疑問が湧き上がってきます。

- ID='AA0001' のお客様が商品 X を買った決め手は、ダイレクト・メールだろうか？ それともコールセンターでの弊社の良い対応だろうか？ そもそもダイレクト・メールは開封していないかもしれない。前からもう商品 X を買うと決めていて、最終確認のためだけに問い合わせしてきた、とは考えられないか？
- ID='AA0002' のお客様にダイレクト・メールを送ってもしばらく反応が無かったが、効果は本当になかったのだろうか？ オンライン・ショッピング・サイトを閲覧してそのときは何も買わなかったが、サイトの閲覧によって商品 Y に対する期待が高まっていたかもしれない。
- ID='AA0002' のお客様に 4 月 30 日に 25% も値引きする必要があっただろうか？ しばらく待てば必然的に売上が計上された可能性はないか？ 値引き幅が 25% でなく 10% だったら結果は変わっただろうか？
- ダイレクト・メールは e メールよりも高い販促コストがかかる。どのお客様にはダイレクト・メールが必要で、どのお客様には e メールでも十分なのだろうか？

これらの疑問を解決するには、自己回帰モデル [2] と呼ばれる統計的時系列予測モデルを利用します。その際、少ないサンプル数の下で結果を都合よく解釈するリスクを防ぐには、大量のデータが必要です。少し単純化して、すべてのお客様からの合計購買点数が図 2 のような変遷をたどったとしましょう。

図 2 のシナリオでは、4 月 1 日の前週にダイレクト・メールを送った結果、2 週間後の 4 月 8 日の週には購買点数が上がっています。試しに送付後 2 週間の平均購買点数を計算しましょう。次に比較対照として、キャンペーンを行わなかった他の週を起点とした平均購買点数も計算します。その結果、キャンペーン実行後の 2 週間の方が平均購買点数がわずかに高いと分かりましたが、この一例だけではサンプル数が足りません。もっと長い期間、多くの消費者に対して分析することで、統計的に有意で効果の大きなキャンペーンを特定することができます。

販売促進の観点で重要なのは、個々のお客様からの購買確率を予測し、どのキャンペーンが購買確率を高めるかを理解することです。図 2 ではすべてのお客様を合計した購買点数に着目したので、目で見ただけでもおおよその推測ができました。しかし

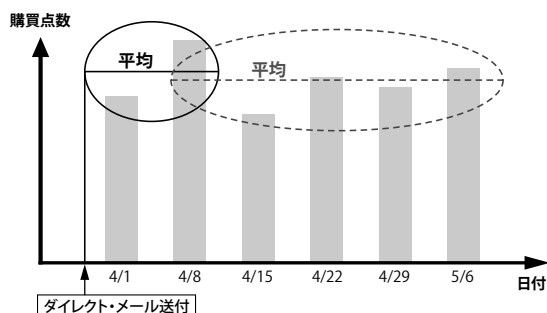


図 2. キャンペーンと合計購買点数の時系列

当然ながら、お客様の中にはキャンペーンに反応しやすい方とそうでない方がいます。一人一人のお客様に対して分析する場合、購買点数として記録されるのは0と1が大半であり、しかも1以上の購買点数はたまにしか発生しません。その結果、購買確率が上がるのか下がるのか、**図2**のように単純に集計することが難しくなります。

イベント・データに対するマイニングの一つの目標は、グレンジャーの因果 [3] と呼ばれる因果律の候補を発見することです。グレンジャーの因果は、どのような施策が購買を引き起こすかを示唆してくれます。Aタイプのイベントが起きるとその1週間後にはBタイプのイベントが通常よりも発生しやすい、という統計的法則を見つけたとします。このとき、イベントAからBへはグレンジャーの因果があるといい、イベントAがイベントBを引き起こしている可能性が高いという仮説が立てられます。グレンジャーの因果は真の因果とは必ずしも限りませんが、因果律の可能性が高いパターンを事前に知っておくことは、効果的なキャンペーンを練る上で非常に重要です。なお、価格の変化と購買との関係のように因果関係が明らかな変数については、グレンジャーの因果はすぐに意思決定に利用できます。

個人別にどのキャンペーンが購買確率を高めるかを特定するには、洗練されたデータ・マイニング・アルゴリズムが必要です。個人別の分析では単純集計が効かないという点に加えて、実データではイベントの種類が多岐に渡るという点にも着目してみましょう。実際の商取引ではチャネル、商品、販促方法のいずれも多数であり、これらの組み合わせによって消費行動は異なってきます。**図2**では2週間と仮定した持続時間も、実際にはイベントの種類ごとに異なっています。値引きのように短期的な影響が大きいイベントと、カスタマー・サポートのようにロイヤリティーの醸成によって長期に影響が持続するイベントとがあるため、実際の分析では持続時間の選定まで行う必要があります。優れたデータ・マイニング・アルゴリズムは、多様な消費者と多様なイベント種別、持続時間との間の複雑な依存性を見つけ出してくれます。

購買パターンを利用する意思決定にも優れたアルゴリズムが

表2. 表1を加工したエピソード・データの例

顧客ID	日時	報酬	状態変数				アクション変数		
			S1	S2	...	S11	S12	A1	A2
AA0001	2013/4/1	-70	0	0	...	0	0	0	1
AA0001	2013/4/2	0	0	0	...	1	1	0	0
AA0001	2013/4/3	-200	0	0	...	1	1	0	0
AA0001	2013/4/8	0	0	0	...	1	1	0	0
AA0001	2013/4/9	0	0	0	...	1	1	0	0
AA0001	2013/4/15	0	0	0	...	0	1	0	0
AA0001	2013/4/17	39,800	0	0	...	0	1	0	0
AA0001	2013/4/18	0	1	1	...	0	1	0	0
AA0001	2013/4/24	0	0	1	...	0	1	0	0
AA0001	2013/4/29	0	0	1	...	0	0	0	0
AA0001	2013/5/1	0	0	0	...	0	0	0	0
AA0002	2013/4/1	-70	0	0	...	0	0	0	1
AA0002	2013/4/2	0	0	0	...	1	1	0	0
...
AA0002	2013/4/15	0	0	0	...	0	1	0	0
AA0002	2013/4/23	0	0	0	...	0	1	0	0
AA0002	2013/4/29	0	0	0	...	0	1	0	0
AA0002	2013/4/30	0	0	0	...	0	0	1	0
AA0002	2013/5/5	16,800	0	0	...	0	0	0	0
AA0002	2013/5/6	0	1	1	...	0	0	0	0

必要です。日々変化する市場では、その変化にうまく追従できる意思決定ルールが必要です。そのようなルールでは、イベントごとの影響の強さや持続時間の違いが適切に考慮されなければなりません。値引きによって得られる短期の利益は、値引きをやめた途端に競合他社にお客様を取られることで消えてしまいます。一方で最大化したい長期の利益は、ロイヤリティーを醸成し、無理な値引きをせずともお客様に満足していただける状態を作ることによって初めて達成されます。

長期利益を最大化する戦略を見つけるにあたっては、イベント・データをさらに加工する必要があります。イベント間の複雑な依存性を利用した、意思決定に適する形式のデータをエピソード・データと呼んでいます。変化する消費者心理のように、特性が時間と共に変化する対象に対して適切に意思決定するアルゴリズムを強化学習 [4] と呼びます。エピソードは強化学習に適したデータ形式です。エピソード・データを利用した強化学習のビジネス・アプリケーションとしては IBM Customer Equity and Lifetime Management [5] や IBM Next-Best-Action Optimizer [6] があり、意思決定理論とデータ・マイニング・アルゴリズムの強力な組み合わせによって実務をサポートしています。

次章では、イベント・データをエピソード・データに加工する手順についてご紹介します。

3. イベント・データを加工してエピソード・データへ

1単位のエピソードとは、個体の識別子 (ID)、発生時刻、報酬、状態変数、アクション変数という主に5種類の情報により定義されます。多くのエピソードを集めたデータがエピソード・データです。IDと発生時刻はイベントと同様ですが、報酬、状態変数、アクション変数はイベント属性を加工して計算します。**表2**および**表3**には、**表1**の購買、販促データから加工したエピソード・データの例が紹介されています。

イベント・データをエピソード・データへ加工するには最大化

表3. 状態変数とアクション変数の一覧

変数名	意味
S1	過去1週間以内に商品を購入
S2	過去2週間以内に商品を購入
S3	過去4週間以内に商品を購入
S4	過去1週間以内にコールセンターに問い合わせ有
S5	過去2週間以内にコールセンターに問い合わせ有
S6	過去4週間以内にコールセンターに問い合わせ有
S1	過去1週間以内にショッピング・サイト閲覧
S2	過去2週間以内にショッピング・サイト閲覧
S3	過去4週間以内にショッピング・サイト閲覧
S7	過去1週間以内にeメール送付
S8	過去2週間以内にeメール送付
S9	過去4週間以内にeメール送付
S10	過去1週間以内にダイレクト・メール送付
S11	過去2週間以内にダイレクト・メール送付
S12	過去4週間以内にダイレクト・メール送付
A1	eメール (EM) の送付
A2	ダイレクト・メール (DM) の送付

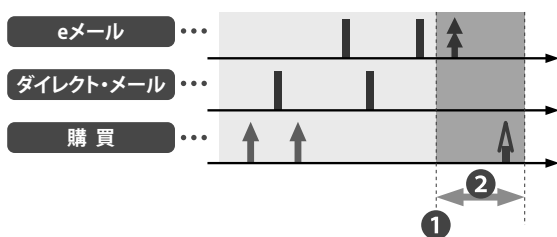
したい目的変数（例・長期の利益）を何にするか、既存の知見を状態変数、アクション変数にどう反映するかが極めて重要です。報酬、状態変数、アクション変数のそれぞれについて具体的な計算、定義方法を追っていきましょう。

報酬は各エピソードの発生時刻における売上値または販促コストです。表 2 では 4 月 1 日のダイレクト・メールの送付コストとして 70 円を、4 月 3 日に発生した問い合わせに人件費 200 円を計上しています。25% の値引きは購買が実際に発生した場合のみ計上されるので、5 月 5 日の購買時において売上から差し引いて計上されます。一連の報酬データは、売上からマーケティング・コストを引いた値を長期的に最大化するために利用します。利益の計算には売上からさらに原価を引くべきですが、原価はマーケティング活動により変化しないので、売上からマーケティング・コストを引いた値を最大化することが利益の最大化につながります。

状態変数には、各時点から過去を振り返ったときの各イベントの発生量が含まれています。例えば 4 月 30 日から過去 1 週間を見たとき、4 月 23 日に購買イベントが発生したため、4 月 30 日時点での状態変数 S1 に数値 1 が立っています。一方でアクション変数は瞬間的に記録されるものであり、販促活動が発生した時刻にのみ 1 を計上します。

ターゲット顧客を抽出するための指標が状態変数で、抽出された顧客に対して取れる販促活動の候補がアクション変数です。あるキャンペーンが将来の購買確率をどれだけ押し上げるか推定することで、それぞれのキャンペーンのターゲット顧客が抽出されます。したがって、購買確率に影響する可能性があるイベントはできる限り状態変数に含める必要があります。購買を決めるお客様の心理状態は、過去の購買経験とキャンペーンのどちらにも影響されるので、過去の Response イベントと Contact イベントの両方を状態変数に含めます。状態変数とアクション変数の双方に Contact イベントを用いる点がややこしく見えますが、図 3 のように現在と過去との因果関係を想像すると分かりやすいでしょう。

同一種類のイベントであっても、状態変数の計算には複数の持続時間を候補として含めます。最適な持続時間はマイニング



① 現在時刻	② 意思決定期間
■ 状態変数に使われる Contact イベント	▲ アクション変数に使われる Contact イベント
↑ 状態変数に使われる Response イベント	▲ 将来期待している Response イベント

図 3. 状態変数とアクション変数の違い

の結果として判明するのです。表 3 では 1 週間、2 週間、4 週間の 3 種類の持続時間が使われていますが、実際には長期的なロイヤリティーを捉えるために半年から 1 年くらいの持続時間も含めた方が効果的です。マーケティング分析で使われる RFM (Recency, Frequency, Monetary) 指標を、イベントの種類ごと・複数の持続時間ごとに計算したサンプルがエピソードだと考えると分かりやすいかもしれません。そして、分析担当者がイベントごとに最適な持続時間を事前知っている必要はありません。持続時間の候補さえデータに入力しておけば、後はマイニング結果から最適な持続時間が自動的に分かります。これは Lasso 回帰 [7] に代表される洗練された特徴選択アルゴリズムのおかげです。

有用なエピソード・データを作成するポイントを以下 (A) および (B) にまとめました。イベント・データのエピソード・データへの加工にはある程度の時間を要します。しかし、いったんエピソード・データを用意できれば、4 章に示す有用な意思決定ルールを自動的に得る事ができます。

(A) どのアクション変数がどの報酬を引き起こしたかの対応情報はデータ作成時には必要ありません。データの作成は会計処理に似ており、単に日時だけに着目して、売上やコストが発生した時刻に報酬値を記録していきます。状態変数やアクション変数と報酬との関係はマイニングの結果として、後から分かります。

(B) 小売業のように多くの商品を販売するビジネスでは、他のデータ・マイニング・アルゴリズムを併用すると効果的です。例えば IBM SPSS に搭載されている、k-平均クラスタリング [8] [9] や非負値行列因子分解 [10] [11] に代表される協調フィルタリングを利用して商品区分を事前に作成します。説明を簡潔にするため、表 2 および表 3 では商品の区別を入れない状態変数、アクション変数を用いました。しかし多数の商品がある場合は、以下のどちらかが必要になります。

(ア) 商品区分ごとに多くの状態変数、アクション変数を定義する
 (イ) 顧客 ID とお勧め商品とのペアを事前に作成した上で、購買チャンネル、販促チャンネルの区別を状態変数、アクション変数に含める

(ア) のアプローチでは、いつ、誰に、どの区分の商品を、どのチャンネルで勧めるべきかが分かります。(イ) のアプローチでは、誰に、どの商品を勧めるべきかという顧客対商品マッチングルールと、いつ、どの販促チャンネルで勧めるべきかというチャンネル選択ルールとが得られるため、2 種類のルールを組み合わせることで販促を行うことになります。

4. エピソード・データが示す最適意思決定戦略

エピソード・データを購買パターン発見アルゴリズム、および

最適意思決定アルゴリズムに入力した場合のアウトプット例をご紹介します。購買パターン・アルゴリズムはその内部で以下の処理を行っています。

- **動学的マイクロ・セグメンテーション**：将来の反応特性が同じと見なせるような状態変数の値域を、自動的にマイクロ・セグメントにまとめる。
- **セグメント別最適アクションの選定**：各マイクロ・セグメントごとに異なる最適アクションを選定し、セグメント内のターゲット人数も決定する。

図4と図5に動学的マイクロ・セグメンテーションの構造、マイクロ・セグメントを利用した最適意思決定ルール、そしてターゲット顧客IDの抽出手順をリレーショナル・データベースに対するSQLコマンド風に例示しました。マイクロ・セグメンテーションは状態変数を入力とした決定木構造をしており、その葉ノードには固有のマイクロ・セグメントIDがふられています。意思決定ルールは「日時、マイクロ・セグメントID、販促チャンネル、ターゲット人数」を多く並べて構成されます。決定木の性質上、あるマイクロ・セグメントには複数の消費者が所属します。該当マイクロ・セグメントに所属する消費者の全員あるいは一部にeメールやダイレクト・メールを送付することにより販促を行います。マーケティング・コストまで考慮して最大何人まで送付すべきか、が最適化されます。

図5の例では4月1日にマイクロ・セグメントID='956'の消費者6,008人にeメールを送付すべきだと計算されています。ID='956'の定義構文を決定木から抽出し、この条件にあった顧客IDの時刻を4月1日の時点で抽出することで、最適なeメールの自動送付が可能となります。

図4および図5のルール形式は一つの本質的理由と一つの実務的理由からデザインされました。本質的理由は、時間的に特性の変化する消費者に対して**時間的に変化しない統一ルール**で対応すべきであるという点です。実務的理由は、そのような統一ルールのうち最適なもの複雑になるため、アルゴリズムを用いて自動的に計算するのが効率的であるという点です。動学的

マイクロ・セグメンテーションに関するポイントをいくつかまとめました。

- A) **同じ顧客であっても時間がたつと違うセグメントに所属します**。例えば2013年に平均購買点数が大きく上昇した顧客ID='AA0003'のお客様が所属するマイクロ・セグメントが、2012年から2013年にかけて変化するなどです。
- B) **消費行動の変化に自動的に追従します**。「平均的購買セグメントにはeメールを、高頻度購買セグメントにはダイレクト・メールを送る」という統一ルールを持つだけで、同一顧客IDのお客様をターゲットとするチャンネルが2012年と2013年とで自動的に変化します。
- C) **顧客の違い・イベントごとの影響の違いを考慮した詳細なターゲティングが可能となります**。現実的なセグメントの設定には、現代の多様な消費社会について考慮する必要があります。つまり平均購買頻度のような1次元情報だけではなく、キャンペーンごとの反応特性、商品ごとの購買頻度など多次元情報を組み入れます。セグメンテーションを自動化することで、イベントの種類が莫大な場合に必要となる、複雑な場合分けを人手で行わなくて済みます。
- D) **アクションの閾値や順序が最適化されます**。状態変数には過去のContactイベントが含まれるため、過去の販促回数によりマイクロ・セグメントIDが変わります。「販促回数が閾値未満の人にはeメールを送るが、閾値以上の人には送らない」「eメールを送ったことのある人のみにダイレクト・メールを送る」といった閾値や順序を考慮した戦略は、マイクロ・セグメントごとにアクションを使い分けることで達成されます。
- E) **最適アクションはお客様の反応に応じて変化します**。適応的ルールの例は「何も反応がなければ3回までeメールを送る。ただし1回反応が起きた時点で送付をやめる」といったものです。状態変数に過去のResponseイベント回数を含めて、それが閾値を超えるとマイクロ・セグメントIDが変わるといったルールにします。その結果、反応回数が閾値を超えるかどうかで最適アクションが変わるように設計されています。

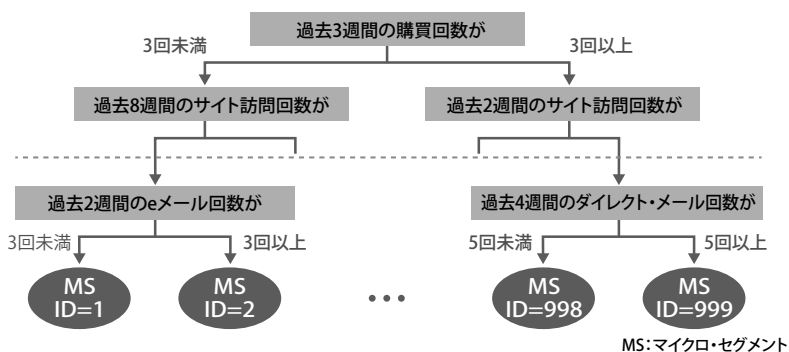


図4. 動学的マイクロ・セグメンテーションの構造

日時	マイクロ・セグメント ID	販促チャンネル	ターゲット人数
2013/4/1	956	EM	6008
2013/4/1	550	EM	1001
2013/4/1	1643	DM	5007

```
IF FreqOrder_32Weeks<3 AND
FreqWebvisit_8Weeks<3 AND ... AND
FreqEM_2Weeks>=2
THEN MSID=956

SELECT distinct CustomerID from DB
WHERE
FreqOrder_32Weeks<3AND
FreqWebvisit_8Weeks<3 AND ... AND
FreqEM_2Weeks>=2
```

図5. 最適アクション・ルールと実行時の顧客ID抽出

リアルタイムにデータを取得、分析するIT環境を整えると、A)からE)の利点を最大限に生かすことができます。E)のように適応的に販促を行う場合は、状態変数データベースを頻りにアップデートし、それに伴って各消費者のマイクロ・セグメントIDを再計算します。例えばIBM Campaignのように、アクションとお客様の反応をリアルタイムに記録し続けられれば、何らかの反応があった場合にすぐに追加販促ができます。複雑な販促ルールを実行する場合は、同時にレポート環境も整えた方が良いでしょう。その場合、少数の戦略セグメントを手動で定義した上で、それを詳細なマイクロ・セグメントに自動で分割する相補的なアプローチが有効です。戦略セグメントを示すカテゴリー変数を状態変数に含めた後にマイクロ・セグメンテーションを行うほか、戦略セグメントごとに収益やアクションの集計を行います。

5. イベント・データを利用した 意思決定技術の今後

販促課題以外にもイベント・データのアプリケーションは広く存在します。刻一刻とイベントが記録される中での意思決定は、アラート・システムでも要求される課題です。自動車を例にとると、車内に搭載されたセンサーからのデータに基づいて状態変数をリアルタイムに計算し、運転者に補助情報を与えるか各種機器設定を動的に変更することが考えられます。

しかし良かれと思って行ったドライバーへのガイダンスが、たとえその内容が有益であったとしても、メンタル・ワークロードなどの問題を引き起こして、かえって有害になることもあります。個々のガイダンスがもたらす収益や損失をデータとして記録し、状況別の最適なガイダンス・ルールを定める必要があるでしょう。別分野の例としては、証券取引所において高頻度取引 [12] によりたまっている莫大なトランザクション・ログを用いて、市場状態を監視することが考えられます。

イベント・データの価値をさらに生かすために、今後は複数の個人の間における相互作用モデリングと、シミュレーションによるデータ量の増幅が鍵を握ると予想されます。本稿の活用例では、個人消費者に対する販促戦略を、同一個人の経験したイベントに基づくルールとして導きました。一方でバイラル・マーケティングのようにクチコミ効果を狙う場合は、ある消費者の経験が別の消費者の購買にどう影響するか、相互作用を理解しなければなりません。相互作用を予測する数理モデルは、多数の消費者に対する同時ターゲットティングの成功を大きく左右します。加えて、多数の消費者に販促する上ではマス広告の取り扱いが欠かせません。その影響の大きさに反してマス広告への接触は現実的には観測の難しいイベントです。しかしメソリンガム・モデル [13] などによるシミュレーションを併用することで、必要なイベント・データを高精度に補間できるかもしれません。

データ・マイニングの精度はどれだけ多様な変数を入力するかにかかっており、シミュレーションは変数の多様性を大幅に増やしてくれます。気象予測などで確立されてきたデータ・マイニングとシミュレーションとを相補的に併用するアプローチが、マーケティングのような社会科学の領域でも今後普及していくでしょう。その普及を後押しする要因は、将来の見通しが立ちにくく不確実性の高い社会においても効果の大きな意思決定を行わなければならないという、多くの人の喫緊の課題にあるのです。

【参考文献】

- [1] F. Chang, et al., "Bigtable: A Distributed Storage System for Structured Data", Proc. OSDI'06: Seventh Symposium on Operating System Design and Implementation, 2006
- [2] G. Box, G. M. Jenkins, and G. C. Reinsel, "Time Series Analysis: Forecasting and Control, third edition", Prentice-Hall, 1994.
- [3] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", *Econometrica* 37 (3): pp.424-438, 1969
- [4] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction", MIT Press, 1998
- [5] N. Abe, P. Melville, C. Pendus, C. K. Reddy, D. L. Jensen, V. P. Thomas, J. J. Bennett, G. F. Anderson, B. R. Cooley, M. Kowalczyk, M. Domick, and T. Gardinier., "Optimizing debt collections using constrained reinforcement learning", In Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2010), pp. 75-84, 2010
- [6] A. Labbi and C. Berrospi, "Optimizing marketing planning and budgeting using Markov decision processes: An airline case study", *IBM Journal of Research and Development*, 51 (3): pp.421-432, 2007
- [7] R. Tibshirani, "Regression shrinkage and selection via the Lasso", *Journal of the Royal Statistical Society, Series B*, 58: pp. 267-288, 1994.
- [8] H. Steinhaus, "Sur la division des corps matériels en parties", *Bull. Acad. Polon. Sci. (in French)* 4 (12): pp.801-804, 1957
- [9] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations", *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability* 1, University of California Press, pp.281-297, 1967
- [10] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values". *Environmetrics* 5 (2): pp.111-126, 1994
- [11] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization", *Nature*, 401 (6755): pp. 788-791, 1999
- [12] IOSCO Technical Committee, "Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency", Consultation Report, 2011
- [13] R. Metheringham, "Measuring the Net Cumulative Coverage of a Print Campaign", *Journal of Advertising Research*, pp.23-28, 1964



IBM 東京基礎研究所 (IBM Research-Tokyo)
アナリティクス&インテリジェンス数理科学
スタッフ・リサーチャー

高橋 力矢

Rikiya Takahashi

【プロフィール】

2004年にIBM東京基礎研究所に入所。データ解析およびシミュレーションの研究に従事。消費者行動に関する時系列数理モデルやマーケティングのための意志決定サポート・ツールを開発している。