

大規模インタラクションデータによる Crowd Computing

Crowd computing driven by large scale interactions

吉井 伸一郎^{*1}
Shinichiro Yoshii

河内 佑美^{*2}
Yumi Kawachi

^{*1} サイジニア株式会社
Scigineer Inc.

^{*2} サイジニア株式会社
Scigineer Inc.

This paper describes how to deal with large scale data obtained from online advertisement business. It is an adaptive solution that is suitable for a competitive environment such as online business. We developed an ad-serving platform capable of self-learning through enormous interactions and discuss its performance.

1. はじめに

サイジニア株式会社では、EC やメディアサイトへのレコメンドーションサービスや、ネット広告配信におけるクリエイティブ最適化及びターゲティング事業を行っている。ネットの世界では、絶えず新サービスの登場、競合するプレーヤーの戦略の進化により、常に有効な最適解は存在しない。一方、ネットの世界における膨大なビヘイビアは、ユーザの限定合理的な判断に基づき、自身の効用を最大化する方向へ向かう。本稿では、ネット広告事業領域の最新動向を概説し、当社が開発した、膨大なデータを解析する手法と、インタラクションを通じてオンラインで最適化する仕組みの効果について述べる。

2. 背景

ネット広告業界は、昨年から、大きな構造変化が進んでいる。これまで、媒体の広告枠をメディアレップが販売する、あるいは、アドネットワークと呼ばれる事業者が広告枠在庫の一部を提供し、アドネットワーク事業者がこれら広告枠在庫を束ねて巨大な広告ネットワークを構成し、ネット広告が販売されてきた。

しかし、昨年あたりから、アドエクスチェンジと呼ばれる新しい形態の広告配信プラットフォームが登場した。これは、広告枠をリアルタイムに、ユーザが目にする 1 インプレッションを取引する仕組みである。RTB (Real Time Bidding) と呼ばれる接続方式によって、ユーザが Web サイトにアクセスした際、誰がどのサイトにアクセスしているかという情報が、瞬時に、接続する広告配信事業者に開示され、各広告配信事業者はそのインプレッションにいくらで入札するかを応答する。最終的には、セカンドプライスオークション等のメカニズムによって、高値を付けた入札者が落札し、広告を掲出する権利を得るといったオンラインのオークションシステムである。通常、これらの取引は 10~100ms 以内に完了するため、ユーザは以前と変わらぬユーザ体験で Web ページを閲覧することになる。

また、第三者配信 (3rd Party Ad Serving: 3PAS) という仕組みにより、広告配信時のクリエイティブを、第三者のアドサーバーから配信することができる。結果、第三者配信事業者が、ターゲットとなるユーザや配信条件に応じて、1 インプレッションごとに広告内容を最適化することが可能になった。このような機能は、DSP (Demand Side Platform) と呼ばれるシステムから利用する。

現在、国内には、数社の DSP 事業者が存在し、その取引量は

日に日に増加する傾向にある。

当社では、EC サイトに情報収集タグを設置、ブラウザのクッキー単位でいつ誰が何を閲覧・購入したかを判別し、表示する商品情報をパーソナライズするレコメンドーションサービスを提供している。さらに、解析した情報を広告として配信するターゲティング広告事業を展開している。第三者配信の仕組みによって接続するネットワークには、Google 社の AdWords や AdX、Yahoo! 社の RightMedia、および OpenX などのアドエクスチェンジ、Advertising.com 社などのアドネットワーク、さらには広告代理店が運営する DSP がある。取り扱うデータ量としては、月間で 8000 万ユニークブラウザと、インプレッション数は 100 億程度に上る。クッキーごとにお薦めの商材を解析し、ネットワーク上に該当するユーザが出現した場合、リターゲティングや RTB によりインプレッションを落札、第三者配信の仕組みを使って広告配信を行っている。

EC 事業者などの広告主は、1 クリック当たりのコスト (CPC)、1 コンバージョンにかかるコスト (CPA) や、投じた広告費用に対して得られる収益 (ROAS) などの広告効果測定指標を厳密に用いて、各ソリューションを使い分けている。広告配信事業者は、自身の収益 (= 広告主にとってのコスト) も含めて、既存手法の効率を上回らなくてはならない。また、各ソリューションを利用するに当たっては個別に運用コストが発生するため、局所的にパフォーマンスが良いだけでは事業としては成立しない。したがって、膨大なトランザクションを効率よく最適化するアプローチが必要となる。

3. データ解析手法

近年、IT 化が進んだ結果、発生・収集される大量データを、どう活用するかが注目され、ビッグデータというキーワードと共に語られるようになった。しかし、その多くは、Hadoop などのツールに焦点が当てられがちである。まず、取り扱うデータの本性、構造特性について議論すべきではないだろうか。

当社では、何百もの EC サイト、ポータルサイト、その他のユーザ行動履歴を解析してきた。爆発的にヒットする売れ筋商品からロングテールの商品まで、ベキ分布に従うのは、サイトの大小・商材の種別によらず、成熟したネット事業のデータに共通する特性である。このようなデータセットに対して、商品の関連性を分析する際、商品同士の共起確率などを計算するアプローチでは、アウトプットに大変な偏りが生じてしまう。それを防ぐために、様々なデータ前処理、アドホックなフィルタリングルールを適用することも可能だが、それでは、スケラビリティを追求できない。そこで、当社では、ユーザ行動履歴を、ユーザとアイテムを

ノード、閲覧・購買行動をリンクとするネットワークを構成して、データ解析するシステムを構築した。

スケールフリーネットワークのサブグラフが、フラクタル様の構造特性を備えていることを鑑みたロジックにより、各ノード基点で local modularity が最大化するようにグラフを分割する(図1)。コミュニティ分割の計算ロジックは、複数ユーザによる購買行動や、閲覧/購買など異種アクションを統合して解析できるように、多重辺ネットワークに拡張されている。各ノードとは、ブラウザのクッキーID であったり、商品アイテムを指すので、結果として得られるコミュニティは、互いに嗜好の類似するユーザ群や商品アイテムグループとなる。これにより、特定オーディエンスのセグメントに対して広告を配信したり、過去に閲覧・購入した商品に類似したお薦め商材を広告としてリコメンデーションする。

4. 膨大なインタラクションを通じての機械学習

ネット事業は、参加するプレーヤーの戦略が共進化していくゲーム環境である。現時点での最適解は瞬間に陳腐化し、常に改良をしていかなくては、市場からの退場を余儀なくされる。

一方で、ネットの世界では、各制御変数は可観測である。ユーザは全ての選択肢を比較することはできないが、現状知り得る選択肢の中から限定合理的に最善の1クリックを決定する。それは最小のコストで最大の効用を得ようとする行為である。この行動原理に基づけば、膨大なインタラクションを通じて、システムのパフォーマンスを改善することが可能になる。たとえば、ある商品に何をリコメンデーションすべきかは、次式のようなロジックによって品質を改善できる。

今、アイテム集合を $I = \{i_k \mid 1 \leq k \leq n\}$ 、アイテム i_x のリコメンデーション集合を $R(i_x) = \{i_r \mid 1 \leq r \leq s, i_x \neq i_r, i_r \in I\}$ (s はアイテム i_x に対するリコメンデーションアイテム数)とすると、アイテムとの関連性 w_{xx_r} は以下の学習式によって更新される。

$$w_{xx_r}(t+1) \leftarrow w_{xx_r}(t) + \alpha \times n_{x_r} / \sum_{m=1}^s n_{x_m} \quad (\alpha > 0)$$

($C(i_x, i_{x_r}) = n_{x_r} \in \mathbf{N}$ はユーザによる選択回数)

当社では、この考えを拡張し、広告配信条件の最適化している。主要な配信条件としては、入札価格(CPM)、配信ドメイン数、フリークエンシー、キャンペーン数、単位時間当たりのIMP数などがある。配信結果に対する評価は、ユーザによるクリック率(CTR)とする。図2は、あるアドエクスチェンジにおいて、広告主のキャンペーンがどのように改善されているかを示している。ここでは、多次元の配信条件を多次元尺度法によって、xy平面にプロットし、その条件におけるCTRをz軸にプロットしている。数字は、キャンペーン開始からの時間経過を表している。キャンペーン開始当初は低いCTRが、時系列に沿って最適化されていく様子がうかがえる。

5. おわりに

近年、IT化の進展に伴い、あらゆる行動がデータ化され、分析して応答することが可能になった。ネット広告業界は、その最たる例であり、多数のプレーヤーが参加して効率化を図っている。本稿では、グラフ理論を活用した行動履歴解析手法と、オンラインで自動学習させる広告配信システムについて述べた。

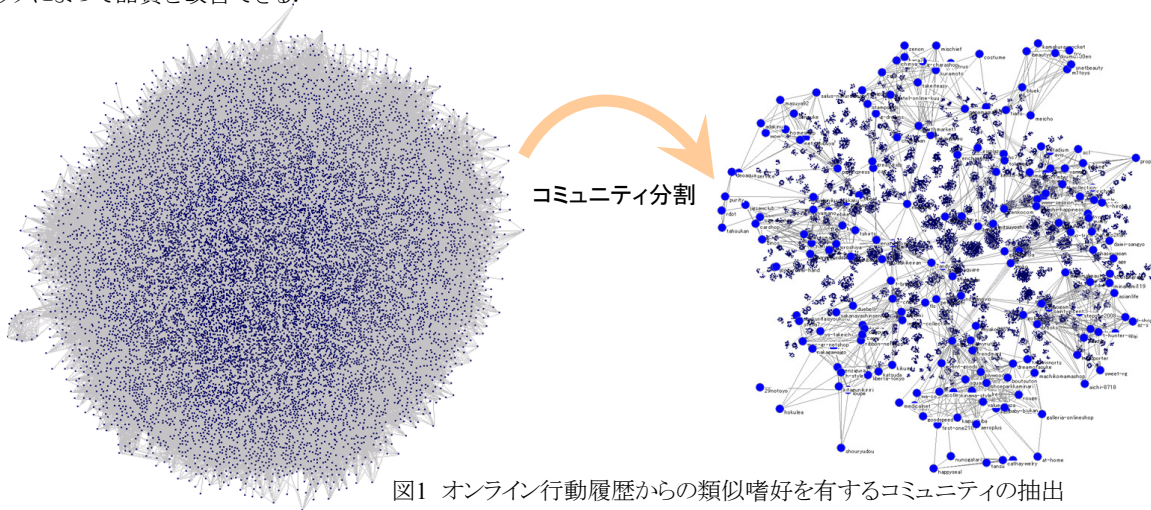


図1 オンライン行動履歴からの類似嗜好を有するコミュニティの抽出

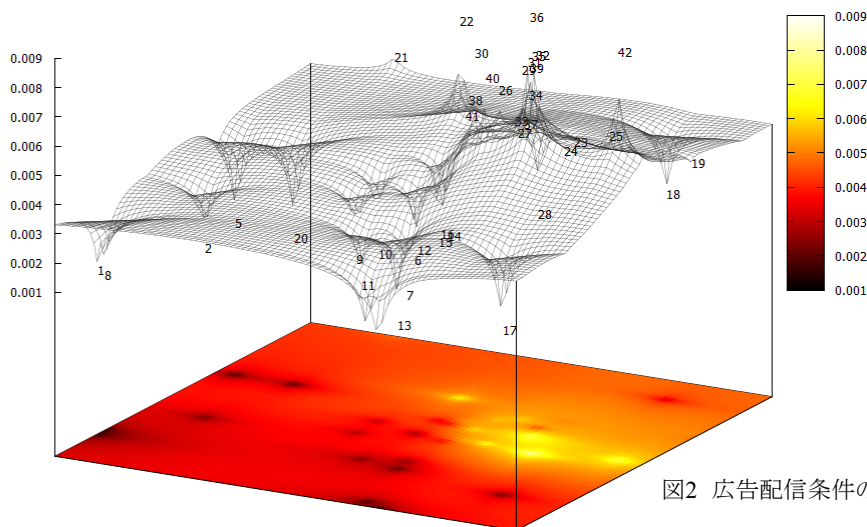


図2 広告配信条件の自動学習によるCTRの向上