

Silver Egg Bulletin Report

2009年3月3日

レコメンデーションとは

ベネフィットと技術について

EXECUTIVE OVERVIEW

『レコメンデーション』は、人工知能技術の応用分野におけるひとつの成功事例として、すでに長い歴史をもっています。アマゾン・ドット・コムといったEコマースサイトは、多くのプロダクトとサービスへのユーザーアクセスを改善するというニーズをもち、そのニーズに答えるレコメンデーション技術が人気を博すことにつながっています。ここでは、レコメンデーションのもたらすベネフィットと、技術について簡単に説明します。

BENEFITS OF RECOMMENDATION

レコメンデーションを、単におすすめしてより多くの商品をユーザーに見せて買わせるという目的からだけで捉えては、本来のレコメンデーションのもつ戦略的意義は損なわれてしまいます。レコメンデーションとは、より良質のユーザー・エクスペリエンスを提供することでオンラインサイトの差別化を実現し、その結果、戦略的優位性を確保するためのもの。正しく運用されたレコメンデーションは、下記のようなベネフィット(利益)をサイトビジネスにもたらします。これらのそれぞれが累乗的に発揮されていき、サイト全体のトラフィックや売上増加をもたらします。

- 1) ユーザビリティ(ナビゲーション)の向上
- 2) ショッピング体験の質的向上(見つけやすさ、発見の驚き)
- 3) ページビュー毎の購買機会の増加
- 4) 顧客のリピーター化
- 5) 長期にわたる顧客関係の構築
- 6) 顧客インティマシー(親密な関係)、知識の蓄積
- 7) 顧客満足の向上(口コミ)

RECOMMENDATION TECHNOLOGIES

Eコマースサイトなどで提供される商品の数や種類が多岐にわたるようになり、ユーザーはますます欲しい商品を探すのが難しくなっています。このため、ユーザーがベストマッチの商品を素早く探せるようにと、様々なレコメンデーション技術が開発されてきました。現在、食品、アパレル、DVD・CD、書籍、雑貨など多岐にわたる業種において活用されています。



Phone: 03-5401-2562 (Tokyo) / 06-6386-1931 (Osaka)

■協調フィルタリング(Collaborative Filtering:CF)

レコメンデーション技術として一番普及しているのが、協調フィルタリング(CF)とよばれる技術です。簡単にいうと、あるユーザーと似た嗜好をもった多くの人のグループの嗜好をベースに、そのユーザーが好むであろうものを予測(或いはフィルタリング)するための技術ということです。

現在では、数え切れないほどのCFが存在しているのが事実です。シンプルなものから高度なものまで様々な味付けのものがあるといえます。なので、一口にCFといっても注意が必要です。

CFの最初のシステムとして有名なのが、Xerox社の有名なInformation Tapestryというプロジェクトで使われた技術です。もともと、文献検索から端を発しているようですが、その後この技術がプロダクト・レコメンデーションのための技術として使われるようになったのです。

アマゾン・ドット・コムもCFを使っています。彼らはアイテムベースのCFを使っています。ご存知のとおり、シルバーエッグ・テクノロジー社のアイジェントASPのエンジンに使われるCFもアイテムベースですが、これがユーザーベースのCFと比べて何が勝っているのでしょうか。

CFはもともとあるユーザーと嗜好があっているユーザーグループを探し出して、そのグループがほかにも購買したアイテムをおすすめするというものです。ユーザーが購買したアイテムにレーティング(評価)されるパターンのCFを表であらわすと、下のような感じです。このようなマトリックスを作って似ているユーザーを探していきます。

	アイテム1	アイテム2	アイテム3	アイテム4
ユーザー1	2	5	5	5
ユーザー2	5	1		3
ユーザー3	5	5		1
ユーザー4	2	5	5	4

でも、嗜好のパターンが重なるユーザーを探し出すなんて考えただけでもそんなに簡単なことではありません。同じ種類の映画に興味がある人たちでも、映画以外の興味となると、一人は海外旅行が趣味であったり、別の人は音楽が趣味であったりするからです。

つまり、嗜好の似た人をグルーピングするためには、大変な数のデータが必要となるわけです。また嗜好の相関も膨大なデータ量が集まらなければ薄い相関になってしまいます。これにより、レーティングしたことのない人(サイトに初めてくる人)にはおすすめ商品が算出できないという、CFの問題として有名な「コールドスタート」問題や、大規模サイトでなければうまく機能しないといった問題が存在

するわけです。

アイテムベースの CF では、人ではなくアイテム間の類似性を見ていくためユーザーベースよりも比較的少ないデータ量で相関が見出せます。マトリックスのイメージとしては、下記の通りになります。ボックス内の数字は併買の数となります。(説明のためにもものすごく単純なイメージにしています。)

	アイテム1	アイテム2	アイテム3	アイテム4
アイテム1		6	1	4
アイテム2	3		8	0
アイテム3	5	2		3
アイテム4	0	4	6	

このようにユーザーベースに比べて、アイテム間の相関をより少ないデータで見出せるため、コールドスタートといった問題が軽減されることと、ユーザーベース CF のもうひとつの問題である、ユーザー数やアイテム数が多いサイトでは計算能力のキャパを超えてしまうという問題もクリアされるため、効果的な手法として注目を集めました。

アイテムベースも、いろいろと相関の取り方が異なるものがあるのでこれも一概には同じではありませんが、ユーザーベースとの大きな違いはおわかりいただけたのではないのでしょうか。

■ルールベース (Rule-based recommendation)

ルールベースシステムは、もともと Expert system とよばれる、プロ(卓越したスキルをもつ人)のやり方をプロセス化しようとする人工知能の分野から派生したものです。字のとおり、ルールを作っていくというやり方です。1990年代には、米国でも大型規模のサイトに多く導入されましたが、ルールをメンテナンスする手間が大変かかることや、ダイナミックなおすすめができないなどで、現在では、その座を CF に明け渡している感があります。近年では、ナレッジベース・レコメンデーション(ここでは説明を割愛)のひとつとして分類されることもあるようです。

■コンテンツベースド・フィルタリング(Content-based Filtering:CBF)

コンテンツベースド・フィルタリングは、好まれる商品カテゴリーの中からおすすめを返すというレコメンデーション技術です。カテゴリーなどの商品情報がない場合、情報取得を自動的に行うタイプの CBF もあります。これは、商品説明テキストから、キーワードを抽出して、そのキーワードに対するユーザーの嗜好を計算するということをベースにしています。

下記の図でいうと、プリティウーマンを見ている人に、同じジャンルと主役のノッティングヒルをおすす

めするのが CBF です。

製品	コンテンツベース		
	ジャンル	主役	価格
プリティウーマン	ロマンス	ジュリア・ロバーツ	3500
ノッティングヒル	ロマンス	ジュリア・ロバーツ	3800
スパイダーマン	アクション	トビー・マクガイア	3500
エイリアン	SF	シガニー・ウィーバ	3300

当社のアイジェント ASP では、カテゴリフィルタリング機能を使うことができ、CBF と同様の効果を得ることができます。

■ハイブリッド・レコメンデーション (Hybrid recommendation)

上記やその他の技術の組み合わせによるレコメンデーション技術の総称をいいます。

シルバーエッグ・テクノロジー社のアイジェント ASP に使われている技術は、アイテムベースの協調フィルタリングをベースに、ベイジアン・ネットワークを応用した技術です。相関計算をリアルタイムで実現できる技術となっており、通常の協調フィルタリング技術のもつ欠点を克服し、新しいユーザーや新商品に対応できる設計となっています。

■レコメンデーションの技術を考えるときの重要なポイント

現在では CF や CBF 以外にも非常に多くの種類のレコメンデーション技術が開発されていますが、レコメンデーション技術を考えるときの重要なポイントは、何の技術を使っているかというよりも、「なぜ」その技術を採用したかという開発思想が非常に大切であるといえるのではないのでしょうか。

つまり、技術自体より、なぜその技術によって、精度の高いレコメンデーションが可能となるのかという点を問うべきだということです。

さらに、レコメンデーションは、オンライン上のプロセスであるため、大量のデータを効率よく計算をして、レコメンド結果を算出できるかというパフォーマンスとスケーラビリティも重要なポイントです。米国では、「リアルタイム」というのも大変重要な技術要素だと考えられています。

サイトにあったレコメンデーションを選ぶ際に、サイトのユーザー行動の特徴についても考慮する必要があります。ユーザーの嗜好性が高く、嗜好をレコメンデーションのベースにする必要がある場合は、

CF がベストでしょうし、医療アドバイスといった、嗜好が入りこむ余地がなく算出結果もほぼ変動ない場合は、ルールベースが適しているでしょう。また、ユーザーの嗜好がかなり頻繁に変わる場合や、新商品が頻繁にでる(商品の回転率が高い)などといった場合には、そのスピードに追いついていく技術を選ぶ必要があります。

DATA CONSISTENCY IS IMPORTANT!

最後に、レコメンデーション技術を利用する際に忘れてはならない重要なポイントがあります。それは、データそのものです。

レコメンデーション技術が使用するデータには、

- 暗示的データ
 - ユーザー行動(閲覧行為など)
 - ユーザートランザクション(購買や予約といった決済行動)
- 明示的データ
 - レーティング
 - レビュー
 - タグ
 - デモグラフィックデータ(地域、年齢、性別など)
 - その他制約条件(問い合わせ、優先順位など)

などがあり、技術によって使用するデータタイプは異なります(上記すべてを同一の技術で使えるのではありません)が、現在よく使われているのが、暗示的データの使用です。

とくに EC サイトの場合に、ユーザーをきちんとトラッキングできるように、一貫したユーザーID が使われていること、またプロダクトID やカテゴリID が整備されていることが、精度の高いレコメンデーションの相関ベースづくりに大変重要になります。新商品がでた場合に、以前の廃番商品のID を使いまわしをしたとすると、相関ベースの中には、以前の商品との相関が残っているため、相関自体が狂ってしまうということが起こり得ます。データの一貫性については、あまり目立たない部分ですが、レコメンデーション導入にとっては大切なポイントであるという認識が必要です。

シルバーエッグ・テクノロジー社では、今後とも、レコメンデーションの成果アップのためはもちろんのこと、EC サイトビジネスに役立つ様々な概念や手法などをご紹介してまいります。