

このアプローチはわかりやすいけれども、しかし事前にすべてのルールをきちんと決めておかなければならない。オンラインショッピングなどの狭い分野では実用的かもしれないが、しかし情報アクセス全般に適用するには無理がある。

これに対して、より深い解析手法によって情報の相関関係を求めているこうというアプローチが、この十年ほどの間にいくつも生まれてきている。

たとえばコンテンツベース・フィルタリングという手法。

これは、情報の属性と利用者の好みをうまく適合させて、お勧めを表示する方法だ。先端的な例としては、アメリカの音楽レコメンデーションサービス、パンドラがある。楽曲の中身を徹底的に解析し、利用者の好みに合わせて音楽を配信してくれるサービスだ。

パンドラに登録して「新しいラジオ局を作る」というボタンを押すと、好きなアーティストや曲の入力を求められる。たとえばポール・マッカートニーの「ダンス・トゥナイト」を入力すると、「ダンス・トゥナイト・ラジオ」という仮想的なラジオ局が作成され、「ダンス・トゥナイト」に似た曲が次々に再生される。

このサービスを支えているのは、「音楽ゲノムプロジェクト」という技術だ。それぞれの楽曲を、メロディーやリズム、傾向といったさまざまな要素によって構成されていると捉え、それらの要素を「遺伝子」と呼ぶ。そしてひとつひとつの楽曲を、専門家が手作業で約四百のゲ

ノムに基づいて点数化し、最初にユーザーが選んだ楽曲と似たゲノム構造の曲を選んで、再生していく——という仕組みになっている。さらにこのゲノムの類似性によって再生された曲に対し、ユーザーの側は「好き」「嫌い」というボタンで自分の好みを加えることもできる。これによって、楽曲と利用者の適合度をさらに高めることができるというわけだ。

このパンドラのアプローチは、意味を持つたウェブという考え方と非常に似通っている。

セマンティック・ウェブは、ウェブの考え方を発明したことで有名なイギリスのコンピュータ科学者、ティム・バーナーズ・リーが二〇〇一年から提唱しているプロジェクトの名称だ。HTMLで書かれている通常のウェブに対し、セマンティック・ウェブは情報の注釈を付け加えたXMLで記述することによって、標準的な形式でウェブに意味を与えることができる。

たとえばある人が、バリ島に旅行に行くために情報を収集するシーンを思い浮かべてみよう。どのようなツアーが設定されているのかを調べるため、検索エンジンで「旅行 バリ島」と検索すると、大量の情報が検索結果に表示されてしまう。ツアー情報は大量にあるが、それらのツアーの価格、日程、宿泊ホテル、オプションの有無などがどうなっているのかは、検索結果が渾然となってしまうため、すぐには拾い上げられない。検索結果にはツアー情報だけでなく、個人の旅行記や書籍の紹介なども含まれてしまっている。また検索キーワードを「旅行 バリ島」にしてしまうと、「旅」や「トラベル」などのキーワードからわかる情報を取

りこぼしてしまう可能性もある。

だがこうしたバリ島の旅行情報がセマンティック化されていると、検索はずっと簡単になる。「旅行のツアー情報である」「どこに向かうツアーか」「価格」「日程」「宿泊ホテル」「オプションの有無」などが、RDFと呼ばれるフォーマットによって注釈メタデータとして記述されているから、一覧表を作って簡単に比較できてしまうのだ。

### セマンティック・ウェブと協調フィルタリングの問題点

セマンティック・ウェブは非常に大きな可能性を秘めているが、ウェブをこの方向に進化させるのはそう簡単ではない。最大のハードルは、すでに存在しているウェブページをすべて作り替えなければならないことだ。

加えて情報アクセスの観点から言えば、もうひとつの問題がある。セマンティック・ウェブはウェブに意味を付け加えることによつて、ウェブ同士の関連性をすぐに調べることができるようになるが、しかし利用者とウェブの関連性を調べることはできない。

つまりセマンティック・ウェブは、情報と情報の適合性を高めることができるけれども、情報と人の適合性を高めるためには、別の手法が必要になるということだ。

そこで、この「情報と人との相関関係」に注目したテクノロジーとして最も有名なのは、プロログでも紹介した協調フィルタリングである。プロログでアマゾンのトイレ洗剤と自殺本の関係を紹介したように、協調フィルタリングは、インフリメント暗黙ウェブでもある。だから利用者の好みを自動収集することができる。

また、パンドラやセマンティック・ウェブのようなアプローチでは、情報の属性を最初に解析し、分類しておかなければならない。パンドラでは、まず膨大な数の楽曲をゲノムに基づいて点数化するという専門家の作業が必要なのだ。

しかし協調フィルタリングでは、他の利用者との好みの類似性だけを見ているため、その情報がどのような内容を持っているかは、考慮しなくてよい。つまり情報を事前に分類しておくという作業がいつさい不要になるわけだ。

この性質には、一見ユーザーの好みとは合わない予想外のお勧めが行われることもあるという副効用を伴う。この予想外のお勧めによつて、ユーザーは「あ、こんな良いものがあつたのか」と驚き、「自分にもこんな好みがあつたんだ」と新たな気づきを得る可能性がある。つまりは幸運を偶然見つける力だ。

とはいえ、協調フィルタリングにもいくつかの大きな欠陥がある。それはいま挙げた利点の裏返しにもなっている。

まず第一に、協調フィルタリングは情報の中身を見る必要がない代わりに、当初から膨大な数の利用者データを必要とする。利用者データが集まっていなければ、満足のいく解析結果を得ることができないからだ。

第二に、協調フィルタリングは、情報がどのような内容を持っているのかを知らないから、トンチンカンな情報をお勧めしてしまう可能性を否定できない。

たとえばアマゾンが、利用者の購入した商品データに基づいて他の顧客のデータと関連づけ、「あなたにはこんなお勧めの映画があります」と勧めてくれるようなケース。だがこの場合には、自分がどんな監督や俳優、ジャンルの映画が好きかという属性はまったく考慮されていない。だからたとえば、理系のある分野の技術書をひんばんに購入していると、それらの技術書の読者たちが比較的好むアニメをお勧めされる、ということが起きてしまう。技術書は読むがアニメが好きではない人にとって、この提案は有効ではない。

第三に、協調フィルタリングは、利用者同士の行動の類似性を見ているだけで、利用者の属性を見ていない。たとえば、アマゾンで妻にプレゼントするために夫が女性用化粧品を購入すると、その後しばらくは女性用化粧品をさかんに勧められるような現象が起きてしまう。

このように見てみると、協調フィルタリングも決して完璧な情報アクセス手段ではない。パンドラのようなコンテンツベース・フィルタリングにもいくつかの欠陥があり、協調フィルタ

リングにもいくつかの欠陥がある。

ではこれらを昇華させ、すべての問題をクリアできるような手法はないのだろうか？

## クチコミ五百万件が掲載されるサイト

前章でも書いたように、情報へのアクセスを高度化させるためには、以下の三つの要素を盛り込む必要がある。

①情報の属性（その情報がどのような分野の情報であり、どのような名前を持ち、どのような内容なのか）

②人々の属性（人々の性別、住居地、年齢、趣味志向など）

③情報と人々の結びつきの履歴（過去にどの人がどの情報にアクセスしたのかという履歴）

この難題に対しては、さまざまな企業がさまざまなアプローチで取り組んでいる。たとえば国内最大の化粧品情報サイトとして知られる「アットコスメ」は、情報にアクセスする人の属性だけでなく、その人の一時的な気分を盛り込んだ商品お勧めシステムを開発している。

情報を選ぶという行為には、自分自身がどのような背景コンテキストを持つているのかという要因だけ

でなく、自分がその時どんな気分なのか、何をしているところなのかという、その時々<sup>アドホック</sup>の要因も影響する。その日の気分やいろいろな外的要因に合わせて、人が好む方向性はくるくると移り変わる。つまりはTPOだ。人と人の関係にTPOがあるように、人とモノの関係性にもTPOがあるというわけだ。

アットコスメを運営するアイスタイル社長の吉松徹郎は、私の取材に対して次のように話している。

「化粧品は、女性がただきれいになるための道具ではない。たとえば食事をする店でも、『今日は恋人と行くからちょっと高級なイタリアンレストラン』『今日はひとりだからささっと吉野家で』とTPOに分けて店を替えるように、化粧品もその時々で替える。あるいは安い商品だつて『こんなに安くて良いものを見つけた』という喜びを感じることもあるし、高い化粧品に『持つ喜び』を感じることもある。同じ化粧品であっても、それに対して女性が抱く気持ちはずさまさまざまだ」

加えて化粧品には、もちろんユーザーの属性も非常に重要な意味を持っている。そのユーザーが「乾燥肌か、敏感肌か」「年齢は何歳代か」といった属性だ。

このアットコスメには、五百万件以上のクチコミが掲載されている。たとえば、無印良品の化粧水。「さっぱりタイプ」と「しっとりタイプ」が発売されているこの五百八十円の商品に対しては、二千万件以上のクチコミが書き込まれ、この結果、アットコスメが震源地となつて

大ヒットすることになった。

「値段も安いし、そんなに期待してなかったけど、想像以上によかったです」

「とってもシンプルですが、しつかり潤うのに変な刺激もなくお気に入りです。安いし、たっぷり使えます。刺激も特にありませんでした♪」

「ほどよく潤うし、ティーンの肌には充分じゃないかなと思います」

ただし、クチコミ数がきわめて多いというのは、ひとりのユーザーから見るとさほど重要なことではない。

ユーザーにとっては、自分に不要な選択肢が多いことよりも、自分が必要としている商品のクチコミ、それに加えて自分と志向の近い人のクチコミがどれだけ充実しているかどうかの方が、ずっと大切だからだ。

化粧品というのはビッグビジネスであり、巨大な消費マーケットを擁しているけれども、しかし大量生産・大量販売の商品ではない。女性ひとりひとりで年齢や肌質は異なっている。先ほどの吉松社長のコメントにあるように、TPOによっても使う化粧品は替わってくる。

同社の代表取締役兼アットコスメ主宰の山田メユミはこう話している。

「昔だったら、デパートの化粧品売り場で洗顔料からクリームまでまとめてそろえている人が

多かつた。得られる情報も少なかつたから、売り場でお勧めされるものを買うしかなかつた。でも今はインターネットで情報を得やすくなり、この結果、「私に合った化粧水はこれとこれ」とベストの商品をTPO別に細かく選ぶことができるようになった」

アットコスメは、こうした属性や行動の方向性に合わせ、クチコミから情報をピックアップするお勧めシステムを作り上げている。

たとえば「ピックアップ」という機能では、クチコミ全体から、

- ①効果（潤い、毛穴・角質ケア、アクネケア、美白・くすみなど）
- ②集計期間
- ③年齢
- ④肌質（普通肌、乾燥肌、脂性肌、混合肌、敏感肌、アトピー）
- ⑤髪量
- ⑥髪質
- ⑦タイプ・好み（メイク大好き！、スキンケアの鬼、ボディケア命、ネイル通、フレグランス大好きなど）
- ⑧購入した場所
- ⑨その他（現品使用、リピート）

などの要素によってクチコミを並べ替え、自分に最適なクチコミを選び出すことができるようになった。

また商品情報のページには、「相性チェック」というボタンもある。「年齢」と「肌質」を選択し、これをクリックすると、その商品について「あなたの総合相性は〇〇%です」と表示される。

たとえばある女性が「三十代」で「乾燥肌」だったとすると、三十代の人がある商品はどう評価し、乾燥肌の人がある商品はどう評価しているのかが数値で表示される仕組みだ。さら

に、この「年代相性」と「肌質相性」の数値の下にある「クチコミを見る」ボタンをクリックすれば、三十代の人のクチコミや乾燥肌の人のクチコミが、まとめて表示される。

とはいえ、これらはどちらかといえば能動的なインターフェイスである。「あなたにお勧めの商品はこの商品です」と受動型のお勧めが行われているわけではない。

なぜなら女性は多くの場合、TPOによって使う化粧品を分けており、このTPOという方向性を要素として数値化できない以上、その時々<sup>アドホック</sup>に情報共有圏<sup>インフォコムズ</sup>を逐次生成していくことは非常に難しいからだ。

「デートで使う」「仕事で使う」といった目的を要素として数値化するのは簡単だが、しかしTPOの要素は用途だけではない。季節や気候の違いによって肌の質は微妙に変わってくるし、同じ仕事の用途であっても、会う相手によって化粧品の使い方は違う。取引先のプレゼンテーションがある日と、単なる部内会議しかない日では、化粧の気合も変わってくる。

そして女性の多くは、こうしたTPOを細かく分類し、超絶的な皮膚感覚によって化粧品を使い分けている。この皮膚感覚を数値化するのは、まだ先のことになるだろう。

## ベイズ理論を使ったサービス

より技術に特化した先端的なケースとしては、日本のゼロスタートコミュニケーションズが発表している「zero zone」という技術がある。同社は山崎徳之ら、ライブドアの役員や執行役員だった人たちが設立したインターネットベンチャーだ。

生粋の技術者である山崎の採ったアプローチは、ベイズ理論である。迷惑メールをふるい落とす迷惑メールフィルタを使っていて、「こうしたフィルタで使われているベイズ理論をレコメンドに応用できないだろうか」と思いついたのが開発のきっかけだった。

ベイズ理論は、十八世紀のイギリスのアマチュア数学者トーマス・ベイズが考え出した確率論の定理のひとつだ。わかりやすく言えば、新たなできごとを予測する際に、すでに起きている事象を予測の材料として次々と加えていくことによつて、予測の精度が高められる——という考え方である。

たとえば、ある企業の株価が上がるかどうかは、過去のその企業の四半期ごとの業績や、マスメディアにどの程度取り上げられたのかという露出度、過去の株価の変動などの要素を集計していくことによつて、ある程度は予測できる。株に興味を持っている人ならたいはいはこうしたことを無意識のうちに頭の中で計算し、「よし、この会社は上がるぞ」などと直感的に判

断しているのだが、この判断を数値化していこうというのがベイズ理論の考え方だ。

I T分野では、ベイズ理論は迷惑メールフィルタに使われている。メールソフトなどに組み込まれている迷惑メールフィルタは最初に、あらかじめ分類されている迷惑メールと普通のメールを比較して、「それぞれにどのような単語が含まれ、どのような出現頻度になっているのか」「単語同士の組み合わせはどうか」「メールのヘッダやHTMLにはどのような情報が含まれているのか」などを調べる。

そのように比較した結果をデータベースとして構築し、次に新しいメールAを受け取ると、そのデータベースに基づいてAが迷惑メールかどうかを判断する。その判断が正しければ、データベースにはAのメールの解析内容が加えられるし、もし判断が誤っていて利用者が「これは迷惑メールではない」と指摘すると、フィルタのシステムはそれも学習して分類の精度を上げていく。つまりベイズ理論は、予測した結果も再びフィードバックして要素のひとつとして取り込み、さらに新たな予測を計算していくことができる自己進化的な仕組みを持っているのだ。

山崎はこのベイズ理論を、情報アクセスに応用することを考えた。  
どのような応用なのか。

協調フィルタリングが、人と人の相関関係を相関係数で表すのに対し、ベイズ理論では、ある人がある情報を選択する可能性を「確率」によって表すことができる。具体的には、情報の

名称や中身などの注釈メタデータを取り出し、この注釈を使って、人がその情報を選択する確率を計算している。

山崎の作ったベイズ理論レコメンデーションには、次の三種類がある。

- ①人と情報のページアン
- ②人と人のページアン
- ③情報と情報のページアン

ソーシャルブックマークを例にとつて考えてみよう。

たとえば百あるウェブサイトの中から、ある人がサイトを十個ブックマークしていたとする。この時、ブックマークされた十のサイトに含まれている注釈メタデータと、残りの九十のサイトに含まれている注釈を比較して、十のサイトを選んだユーザーが、残りの九十のサイトからどれを選び出すかを推測する。これが①の「人と情報のページアン」だ。

②の「人と人のページアン」は、たとえばAさんが十五のサイト、Bさんが十八のサイト、Cさんが十三のサイトをそれぞれブックマークしたとする。これらのサイトの注釈メタデータを比較すれば、AさんとBさん、Cさんの好みがどれだけ近いのかを推測することができる。

また③の「情報と情報のページアン」は、それぞれの商品が持っている注釈メタデータを比べて、そ

|     | 商品1 | 商品2 | 商品3 | 商品4 | 商品5 | 商品6 |
|-----|-----|-----|-----|-----|-----|-----|
| Aさん | ○   | ○   | ×   | ○   | ×   | ○   |
| Bさん | ×   | ×   | ○   | ○   | ×   | ×   |
| Cさん | ○   | ○   | ×   | ○   | ○   | ○   |
| Dさん | ○   | ×   | ×   | ○   | ×   | ○   |
| Eさん | ○   | ○   | ○   | ×   | ×   | ○   |

### ユーザーの好みと協調フィルタリング

これらの情報同士の近さを推測するというものだ。

またゼロゾンではこの三つのページアンに加えて、二種類の協調フィルタリングも内包させている。

通常の協調フィルタリングが人と人の関連性だけを調べているのに対し、ゼロゾンの協調フィルタリングは商品と商品の関連性も調べている。上の表を見てほしい。

人と人の協調フィルタリングで見ると、AさんとCさんはかなり好みが似通っているので、同じ情報をお勧めできる。

一方、情報の側から見ると、商品1と商品6は好む人が似通っている。そこで商品1を選択した人に対しては、商品6もお勧めできるということになる。これが情報と情報の協調フィルタリングだ。

このゼロゾンは、USENの音楽配信サイト「**On Gen**」などに採用されている。

過去、ベイズ理論を使ったサービスとしては、マイク

ソフトがオフィス・スイートに搭載したユーザー・アシスタントがあった。ベイズ理論に基づいて、必要な場面で必要な内容のヘルプを的確にユーザーに提供できるというふれこみだったが、しかし「何の脈絡もなく現れて、的外れなヘルプを言ってくる」とさんざんな評価で、その後姿を消している。ベイズ理論の応用がまだ未開拓の分野であるのは間違いないが、しかしこうしたアプローチが採用されていくことによって、アルゴリズムの高度化が今後進むのは間違いないだろう。

### ビッグブラザーに監視される不安

しかし——ここでひとつ重要な問題が浮上してくる。

それは、アマゾンやゼロゾンなどのアルゴリズムによるアプローチは、利用者側から見えて可視化されていないという問題だ。どのような情報の注釈ノットや人の属性をもとにして、どのような計算が行われ、どのような相関関係の可能性が見出されたのかという過程プロセスは、利用者には情報開示されない。

自分の情報がシステムに転写コピーされているのに、その転写された自己がどのようなになっているのかわからない——というのは、存在論的な不安を引き起こす。自分のアイデンティティが強制的に細分化され、システムの側と自分の側に分断させられているような感覚を呼び起こすか

らだ。

この結果、人々はこう感じるようになる。

「自己の一部をシステムに収奪され、プライバシーを侵害されている」

このプライバシーの問題は、本書で取り上げられているようなソーシャルサービスでは、避けて通れない問題と化している。ITのシステムとプライバシー懸念が結びつくと、多くの人はイギリスの作家ジョージ・オーウェルが小説『1984年』で描いた全体主義国家を思い起す。ビッグブラザーが国民の生活と行動のすべてを監視するという恐怖の未来像だ。こうした不安の社会的背景があつて、先に紹介したフェースブック・ビーコンもプライバシー懸念の反発を引き起こし、サービス変更を余儀なくされたわけだ。

ではどうすれば、このようなプライバシー懸念を防ぐことができるのか。

ITのシステムにおけるプライバシー懸念は、「何らかの要因がそこに存在し、その結果生じる実害」という三段論法で語ることでできる問題ではない。何か特定の要因を削除すれば、プライバシーの侵害が収まるという話ではないのだ。

それは先ほども言及したように、ある種の存在論的不安であつて、何らかの実害を伴うものではないからだ。もちろんプライバシーが漏洩したことによってなりすましなどの犯罪の対象となり、実害が生じる場合もないわけではないが、それらはどちらかと言えばレアケースだ。プライバシー懸念の多くは、ロジカルに定義することは不可能なもので、単に「プライバ

シーを侵害されて気持ち悪い」「なんだか不安を感じる」という情念でしかない。ではそのような情念的な不安を、どうすれば解消することができるのだろうか。

この困難な問題を解消できるアプローチは、たった一つしかない。それを次章で説明しよう。

# 第 4 章

---

ウェブ3・0は  
「信頼」と「友情」を両立させる

## 「ウェブ3・0の定義」のコンテスト

「プライバシーをシステムに収奪されている」という不安は、情報共有圏に立ちはだかる最大の障壁となる。情報共有圏が利用者の背景や属性、行動履歴などをベースとして生成されるものである以上、自分自身の情報がシステム側に渡るといのは避けられない事態だからだ。この問題をどうすれば解決できるのだろうか？

その解決に向かうための要素を、ここで提示しよう。それは二点ある。

まず第一に、情報が利用者のもとに再集約されること。

第二に、同じ情報共有圏をとる他人に対し、信頼だけでなく友情を導き出せるようになること。

これら二点の確立は、情報共有圏の捉え方を決定的に変容させる。

システムによって生成された情報共有圏は、「システムからの収奪」ではなく、人が自分自身の主導権によって、他の人たちとの友情によって作り出されつつある中間共同体として認識されるのだ。前章でも書いたように、フェースブックのような友情システムは、信頼に基づいて情報アクセスを高度化させた情報共有圏を形作ることはできない。しかし信頼だけが存在し、友情の欠如した情報共有圏は、深刻なプライバシー不安を引き起こしてしまう。