

推薦システムの現状と未来



経営戦略研究科准教授(経営戦略専攻) 羽室 行信

20世紀末に始まったWWW (World Wide Web) の爆発的な発展は、現在に至るまで様々なビジネスを生み出してきた。その中でも書籍販売を中心としたネット通販であるamazon.comは、草分け的な存在として位置づけられるであろう。amazonの強みの一つは、強力な推薦システム (recommendation system) を有している事にある。ある本を買い物カゴに入れると、その本に関連の深い他の本を推薦してくれるシステムである。この推薦システムのおかげ(せい?) で、本の注文量が増えてしまった経験を持つのは私だけではないであろう。推薦システムの性能向上に対する関心は高く、例えば、米国のオンラインレンタルビデオ店の大手Netflixは、推薦システムの精度を現行のシステムより10%向上させれば100万ドルの賞金を与えるコンテストを開いている (<http://www.netflixprize.com/> : 2006年末に始まり現在トップチームは9.56%の改善を達成している)。

推薦システムとは

推薦システムは非常に単純な考え方に基づいて構築されている。推薦システムには大きく分けて二つあり、一つは商品の内容によって推薦する方式で、書籍の例で言えば、タイトル、著者、分野などの情報によって消費者が興味ある商品に類似した商品の推薦を行う方式である。もう一つは、「強調フィルタリング」と呼ばれる方式で、前もって多くの人から商品に対する評価 (5段階評価得点など) データを収集しておき、そのデータに基づいて商品を推薦する。ある人に商品を推薦する際には、その人と類似した評価を行っている人を検索し、それらの人の評価情報に基づいて商品を推薦する。しかしながら、この方法では、全く商品を評価していない新規顧客に商品を推薦することが困難であるという問題がある。amazonはこの問題点を克服する為に、異なるアプローチを採用している。似た人を探すのではなく、似た商品を探し、顧客が閲覧もしくは購入した商品に類似した商品を推薦するという仕組みを採用している。

頻出パターンマイニング

以上に紹介したamazonの推薦システムの仕組みでは、2商品間の類似度をあらかじめ計算しておけばよい。類似度の定義にもよるが、大規模データにおいては、その計算効率が大きな問題となる。この問題を扱っているのが「頻出パターンマイニング」と呼ばれる研究領域である。これはデータマイニングという言葉が出てきた1990年代初頭から活発に研究が進められてきており、今現在でもデータマイニング研究の中心的な領域である。

頻出パターンマイニングの基本はパターンの列挙および数え上げにある。パターンの代表例は、amazonで見られるような商品の組み合わせである。数百万タイトル×数百万人のデー

タから、共起頻度の高い商品の組み合わせをパターンとして列挙する。Amazonの例では2商品の組み合わせに限定して列挙すればよいことになる。そしてパターンはさらに複雑なものへと拡張できる。例えば、「崖上のポニョ」と「となりのトトロ」の2作品のレンタルDVDでは、どちらを先に借りるかによって意味は異なるであろう。ポニョが先であれば、初めて宮崎作品を見て気に入ったために過去の作品を次の機会に借りたのかもしれない。このような時系列を考慮したパターンの列挙なども盛んに研究されている。また顧客の性別や年齢といったデモグラフィック属性をも加えることができれば、さらに興味深いパターンが発見できるかもしれない。今後の推薦システムは、単なる2商品の共起というパターンだけでなく、上記に示したようなより複雑なパターンに基づいたものに発展していくことになるであろう。

多様化する推薦システム

推薦システムは何もamazonのようなネット上で商品を販売するビジネスだけで利用されるものではない。例えば、Webページの検索と広告の連動により大きな成功をおさめているGoogleが提案する検索システムも、単なる検索として捉えるのではなく、ユーザが与えた単語に基づいてwebページを推薦するシステムとして考えるべきであろう。そしてgoogleは既に、gmailを初めとする様々なアプリケーションを無料で提供することでユーザのプロファイルや行動ログを獲得している。それらの情報を融合する事で、ユーザの特性に応じたよりの確な検索結果を推薦するシステムへと発展していくであろう。

またネットの外のビジネスに目を向けてみると、例えば小売業界での商品の割引クーポン券の発行は、推薦システムの一つとして考えることができる。これまでは、メーカー主導で自社商品のシェア獲得のために実施される事が多かったが、それをユーザの購買履歴に基づいた推薦システムへと転換することも十分に考えられる。近年のポイントカードの普及は、それらの企業に顧客の購買履歴が蓄積されていることを意味する。そしてRFIDの普及は顧客の店舗での導線をリアルタイムに追跡する事を可能にする。これらの情報に基づいた実店舗での推薦システムが普及していく事は十分に考えられることである。

おわりに

現在、推薦システムは黎明期にある。多くの企業が大量データを蓄積してはいるものの、それをビジネスに生かしている企業はまだほんの一部である。しかしながら、この分野における技術発展のスピードは非常に速く、Webの内外関係なく、より優れた推薦システムの構築が今より容易に実現できるようになるのも決して遠い未来の話ではないであろう。

CRM（顧客関係管理）の重要性が言われて久しいが、顧客と企業との関係構築においては、パターンという抽象化されたモデルに基づいた推薦システムのような仕組みが今後ますます重要視されていくであろう。しかしながら、そういった動きに一抹の不安を覚えるのは著者だけではないはずである。