

商品推薦システムの構築事例

株式会社 KDDI 研究所 知能メディアグループ 小野智弘

1. はじめに

最近“オススメ”、“レコメンド”サービスや技術紹介を新聞、ニュース、WEB などでご覧になった方も多いと思われる。これらは膨大な情報や商品からユーザの好みにあったものを自動で抽出し提示するサービスの総称である。商品推薦を支える技術自体は90年代から存在したが低迷期を経て最近ニーズとシーズが揃ってきたことを背景に、特にビジネス的側面から大きな注目を集めている。これに呼応して学術界でも論文や学会で商品推薦技術の特集が組まれるなど注目を集めている。本稿では、商品推薦システムの構築事例を映画推薦システムを例に解説する。

2. 商品推薦システムの概要

商品推薦システムとは、ユーザの興味に関わる情報を手がかりに膨大な情報や商品から個々のユーザの好みにあったものを自動で抽出し、「あなたへのオススメ」などの欄で提示するサービスの総称を指す(図1)。重要なのは、「個々のユーザの」という部分で、オススメと称していてもランキングなどにより全ユーザに画一的に提示するものは含まない。

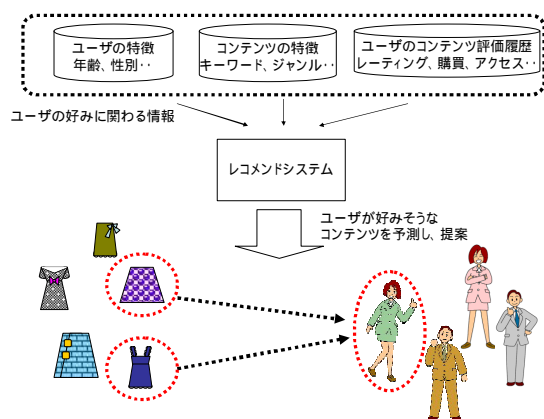


図1. 商品推薦システムイメージ

推薦対象となる商品はWEBサイトで販売している商品、ニュースやブログ記事、あるいは、サイトそのものであるなど様々である。商品推薦はユーザにとってもサイト運営者側にとってもメリットを生み出す。ユーザにとってのメリットは1) 好みのコン

テンツへ迅速にアクセスできる、2) 自分では気づかない好みの商品を発見できることなどが挙げられる。サイト運営者にとってのメリットとしては、1) サイトアクセスやコンテンツ購買率の向上が期待される、2) 人気上位でないいわゆる「ロングテール」コンテンツの売り上げ向上が期待されることなどが挙げられる。携帯サイトのように表示できる文字数が少ない環境では特に重宝される。

3. 商品推薦システムの構築事例

3.1. KDDI 研究所の商品推薦技術

商品推薦システムを支える技術は様々あり、行動が類似したユーザ同士は嗜好も類似しているという考えに基づきユーザの嗜好を過去の行動(=履歴)という形で保存し、対象ユーザと似たような行動をとっているユーザの履歴から対象ユーザの嗜好を推測する協調フィルタリング方式や、ユーザがある商品を好きであれば類似した特徴を持つ商品も好きであろうという考えに基づきユーザの好むコンテンツに類似した特徴を持つ商品を推薦する内容(コンテンツ)ベース方式などがある。

KDDI 研究所では、ユーザや商品の特徴をそれぞれ複数の属性で表現しそれらの属性間の依存関係を表現することにより商品を推薦する属性ベースの方式をとっている。本方式の特長としては、ユーザが商品を利用するシーンや気分などの状況を属性に含めることにより個人差と状況に応じて最適な商品を推薦するシステムを実現していること、および、これらの複雑な依存関係を確率モデル(ベイジアンネット)により表現していることある[1]。このベイジアンネットモデルを利用して商品を推薦する通常の手順は、ユーザ U_i から商品推薦要求を受け付けた際に、候補となる商品 C_j について、その属性群および U_i に関する属性情報を入力とし、ユーザ U_i が商品 C_j を V_{ij} と評価する事後確率 $P(V_{ij} | U_i, C_j)$ を求め、 V_{ij} =良いとなる確率の高い順に商品を推薦する。ところが候補となる映画作品数と同じ回数分の推論が必要となり、計算時間の観点から現実的ではない。そこで推薦要求時の推論回数を抑えるために、以下の手法

をとった。まず、各映画に対応する印象属性（後述）の事後確率を平均的なユーザを想定して（つまり、ユーザ属性値を設定せずに映画属性値のみを入力として）あらかじめ計算して映画の特徴ベクトルとして保存しておく。そして、ユーザからの推薦要求時にユーザの状況（同伴者や気分）を入力させ（図2）あらかじめ登録済みのユーザ属性情報と併せて印象属性の事後確率を求め（図2）、保存済みの映画の特徴ベクトルとのマッチングを行い、近い映画から順に推薦を行う（図2）。

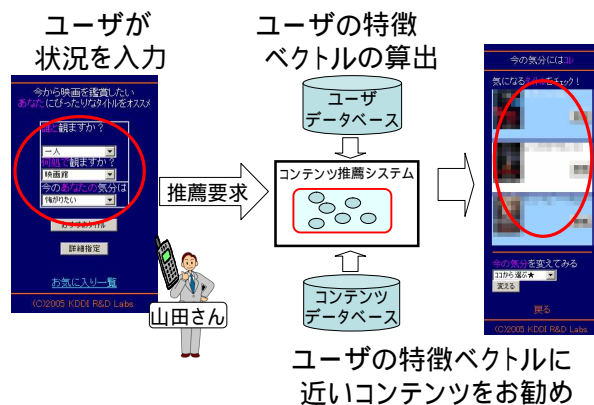


図2：推薦システムの流れ

3.2. ベイジアンネットモデルの構築事例

本節では推薦システムで用いるベイジアンネットモデルの構築ステップを述べる。ベイジアンネットのモデル構築手法としては対象とする分野や現象に関する知識を活用して手動で構築する方式、データから自動で探索する方式などがあるが、筆者らは、変数間の依存関係に関するおおまかな仮説に基づき、学習用データへの適合性を評価しながら依存関係を探索する半自動方式により構築した。モデル構築の大まかなステップは以下の通り。

- 1) 学習用データの収集
- 2) データ加工
- 3) モデル構造およびパラメータの決定
- 4) モデル評価

モデル構築では2)~4)のステップを繰り返しながらモデルを精練していくため、各ステップの処理が可視化できることを重視してステップ2),4)ではデータマイニングツール VMStudio(Visual Mining Studio)を利用した。ステップ3)はベイジアンネット構築ツールである Bayonet を利用した。

学習データはWEB アンケート調査により取得した。

ここでは2153人の映画鑑賞頻度が比較的高い被験者に対して視聴経験のある作品からランダムに5~10件の映画を提示し、ユーザ属性、映画評価を収集した。質問項目は1)年齢や性別などのデモグラフィック属性とライフスタイルに関する属性をあわせた30種類のユーザ属性、2)映画選択時の重視項目(俳優重視、映像重視)や主要目的(癒されたい、笑いたい)などの32種類の映画視聴態度属性、3)誰と観たかなどの43種類の状況属性、4)泣けた、怖かった、映画に入り込めた、など、鑑賞した映画に対する358種類の印象属性、5)映画の総合評価とした。本アンケートの特徴はユーザの状況属性や印象属性など、他のデータセットに比べて属性数が非常に多いことが挙げられる。映画属性としては、ジャンルや製作年、製作国、映画紹介文から抽出したキーワードなど26種類利用した。

2)のデータの加工ステップでは、ベイジアンネットモデルは離散変数を用いるため連続変数の離散化や属性値の圧縮(7値5値)を行った。またモデル構築ツールに投入する変数を絞るため、ユーザ属性、コンテンツ属性、印象属性といった各グループについてグループ内の全ての属性間の相互情報量を求め、続いて各クラスを代表する属性(代表属性)を抽出することにより重要変数を選定した。

3)のモデル構築ステップでは、投入した属性から効果的な属性を選択し、さらに、選択した属性間の依存関係を反映したネットワーク構造を決定し、続いてデータを利用して条件付確率表を推定した。構築したユーザモデルは、ユーザ属性群(20ノード)、コンテンツ属性群(23ノード)、状況属性群(7ノード)、印象属性群(24ノード)、総合評価の合計75ノードとなった。

4)のモデル評価ステップでは、学習に利用していない正解付き検証用データを用いて複数の推薦手法間の予測精度の比較を行った。

参考文献

- [1] 小野、本村、麻生：移動端末におけるユーザの状況を考慮した嗜好抽出技術, 情報処理学会誌, Vol.28, No.9, 2007