

## ユーザ嗜好の個人差と状況差を考慮した 映画推薦システムの実現と評価

小野 智 弘<sup>†1</sup> 黒川 茂 莉<sup>†1</sup>  
本村 陽 一<sup>†2</sup> 麻生 英 樹<sup>†2</sup>

筆者らは携帯電話などのユーザが持ち歩く端末での利用に適したコンテンツ推薦サービスを検討している。ユーザの嗜好はたとえば映画鑑賞の場合、家族と観るか恋人と観るか、元気であるか沈んでいるかなど、状況や気分に応じて変化するため、モバイル端末での利用を想定するとユーザごとの差異に加えて、同一ユーザ内でも状況に応じた差異を考慮してコンテンツを推薦する必要がある。ところが既存のコンテンツ推薦システムはユーザの個性を考慮できるが、状況を合わせて考慮して推薦することができない。本稿では、ユーザの特性に関する情報と気分や場所などの状況の複雑な依存関係を表現したベイジアンネットワークモデルに基づく映画推薦システムを提案する。具体的には、ユーザの属性、状況、映画評価を含むアンケートデータの収集手法、これらの変数間の依存関係を表現するベイジアンネットワークモデルの構築手法、および、同モデルに基づく映画推薦システムの設計と実装について述べる。また、2種類の評価実験により提案システムがユーザの嗜好の状況依存性をうまく反映できていることを示す。

### Implementation and Evaluation of a Movie Recommender System Considering both Users' Personality and Situation

CHIHIRO ONO,<sup>†1</sup> MORI KUROKAWA,<sup>†1</sup> YOICHI MOTOMURA<sup>†2</sup>  
and HIDEKI ASOH<sup>†2</sup>

We are developing a movie recommender system. Movie preferences change according to not only the users' personality, but also to the situation/context such as mood, location, accompanying person, and so forth, and the mobile user may access the recommender system under various situations. However, almost all existing researches on recommender systems have not yet dealt with users' situations. In this paper we propose a novel movie recommender system that provides context-aware personalized recommendations for mobile users. Here, we explain data acquisition process by a large-scale WWW questionnaire survey, then explain a novel Bayesian network model construction process using data acquired. Then we explain design and implementation of the context-aware personalized movie recommender system. The effectiveness of the proposed system is evaluated through two experiments.

#### 1. はじめに

近年、サービスの多様化とユーザの価値観の多様化とともに、ユーザがサービスを利用する状況も家の中や街中など多様化している。また、携帯電話経由のインターネットアクセスが増加していることによりこの傾向は顕著になっている。このような現状をふまえ、筆者らは、多様化するユーザの状況を考慮したコンテ

ンツ推薦サービスを検討している。

これまでにユーザが欲するであろうコンテンツを予測して推薦を行うシステムについては、研究と応用の両方の場面で、様々なアプローチが開発されてきた<sup>1),19),22)</sup>。ここでは、協調フィルタリングとコンテンツベース(属性ベース)の手法が、2つの主要な方法として広く用いられている。協調フィルタリングでは、あるユーザの嗜好を予測するためにそのユーザと類似した評価履歴を持つユーザの嗜好を利用する<sup>5),19)-21)</sup>。この手法は、多くのインターネットショッピングサイトでも用いられて、有効性が認められている。しかしながら、協調フィルタリングが機能するためには、多くの評価履歴データが必要である。このため、推薦シ

†1 株式会社 KDDI 研究所  
KDDI R&D Laboratories Inc.

†2 独立行政法人産業技術総合研究所  
National Institute of Advanced Industrial Science and  
Technology

システムが評価履歴をまったく持っていなかったり少数しか持っていなかったりするユーザに対して推薦を行うことは難しい。また、映画館で上映されている映画のように、推薦対象のコンテンツが頻繁に入れ替わる場合にも適用が難しい。

これに対して、コンテンツ、あるいはユーザ属性に基づく推薦では、推薦対象のコンテンツあるいはユーザをいくつかの属性値によって特徴づけ、その属性値と評価値の関係をモデル化する。したがって、新しいコンテンツやユーザに対しても適用することができる<sup>17),19)</sup>。しかしながら、コンテンツやユーザを表現するための適切な属性を選ぶことは難しい問題であり、多くの場合には、アドホックな方法が用いられている。

また、協調フィルタリング方式とコンテンツベース方式の両者を組み合わせた方式<sup>2),3),6),16),18)</sup>などが検討されている。

ところが、ユーザの嗜好はたとえば映画鑑賞の場合、家族と観るか恋人と観るか、元気であるか沈んでいるかなど、状況や気分に応じて変化する。このため、ユーザごとの差異に加えて同一ユーザ内でも状況に応じた差異を考慮してコンテンツを推薦する必要があり、携帯電話などのモバイル端末からインターネットにアクセスするユーザ（モバイルユーザ）にとっては特にこの要望が強い。ところが従来の推薦方式ではユーザの個性を考慮できるが、状況を合わせて考慮して推薦することが難しい。

一方、ユーザのいる場所に近いレストランを紹介するような状況に基づく情報サービスは行われているが、そこでは、個人の嗜好が考慮されていないことが多い。状況と個人の嗜好を考慮した例としては、Miller らによるモバイルデバイスにおける映画推薦システムの研究がある<sup>15)</sup>。そこでは、協調フィルタリング方式が用いられ、(1) ビデオを買う・借りるユーザ、(2) 映画館に映画を観に行こうとしているユーザ、の2つの状況に分けて推薦を行っている。ユーザの嗜好に影響を及ぼす状況には、ビデオなのか映画館なのかということ以外にも、鑑賞時の気分や、一緒にいる人など数多くあるが、Miller らの方式では考慮すべき状況の数が多い場合には対応が難しい。

本稿では、映画を対象とし、ユーザの個性と状況を同時に考慮した映画推薦システムを提案する。具体的には、対象とするユーザの個性と状況を表す変数および候補となる映画の特性を表す変数を入力とし、これらの変数とユーザがある状況である映画を評価した結果を表す変数間の関係性を表現するモデルを用いてユーザの映画評価を予測し、予測評価値の高い映画を

推薦するシステムを提案する。

ユーザの嗜好に影響を与える「状況」に関する情報は、ユーザの場所や位置、天気、時刻、気温などの環境情報、ビジネスやレジャーなどのシーン情報、1人、友だち、家族といった同伴者情報、忙しい、暇、気分（笑いたい・泣きたいなど）などの心理情報など様々あり、context, situation, occasion など様々な呼ばれ方がされている。利用するタスクごとに定義がなされているのが現状で、確立したものはないため、本稿では映画ユーザに対する調査結果に基づいて、3.3節の第6項に述べるように、映画鑑賞に関する43種類の状況概念を取り上げて検討した。

提案する推薦システムを実現するためには、変数間の複雑な関係を表現するモデル構築が重要な鍵となる。これらの関係は不確実性が高いため、確率的なモデル化が適していると考えられる。また、これらの関係は、非線形性や非正規性が高いと考えられるため、通常の線形予測モデルや、構造方程式モデルなどでは適切にモデル化することが難しい。変数間の関係を非線形なモデルでとらえる試みとして、たとえば Chiristakou ら<sup>8)</sup>は、ニューラルネットワークを用いて、協調フィルタリングとコンテンツベースの推薦を融合する映画推薦方式について検討している。また、Basilico ら<sup>2)</sup>は、サポートベクトルマシンを用いた方式を検討している。しかし、状況に関する変数を含めて検討している例はまだほとんどない。

近年、多数の確率変数間の複雑な依存関係を柔軟にモデル化できる手法としてベイジアンネットワーク<sup>12),26)</sup>が注目されており、プリンタの故障診断<sup>13)</sup>やユーザインタフェース<sup>11)</sup>などに応用されている。そこで、本研究では、上記の変数間の関係をモデル化するためにベイジアンネットワークを用いることとした。

実用規模のベイジアンネットワークモデルの構築は難しい課題であり、確立した構築方法は存在しない。またこれまでの研究事例で今回の問題にそのまま適用できる手法はないため、本稿で新たに検討する。

以下、まず2章で映画推薦タスクおよび推薦システム実現の課題を説明する。3章では提案するベイジアンネットワークに基づく映画嗜好モデル構築手順と構築したモデルについて、4章では映画推薦システムの設計と携帯電話上での実装について、5章では実装したシステムの評価実験について、それぞれ述べる。6章ではまとめと今後の課題を述べる。

## 2. 映画推薦と課題

### 2.1 映画推薦

本節では、筆者らが開発する映画推薦システムのタスクの性質について、Herlocker らの研究<sup>10)</sup> にならって、6つの観点から説明する。

#### (1) 推薦システムのタスク

ユーザがタイムリーに見たいと思われる映画を推薦する。本研究ではタイムリー性に対する要求が強い携帯電話上での利用も意識した映画推薦を対象とし、ユーザの嗜好の個人差と状況依存性を同時に考慮した推薦を行うことを目標とする。

#### (2) 推薦リスト

出力する推薦リスト数は表示デバイスや応用の種類により決定される。PC画面では10~20個程度表示可能であるが、画面サイズの小さい携帯電話は3~5個程度に限定される。

#### (3) 推薦対象となる集合

映画館で観る場合とレンタルやDVDで観る場合で対象が異なる。

(ア) 映画館：推薦対象となる上映中の映画作品数は、2007年3月15日時点で旧作上映や単館上映も含めて約250本である（全国規模の上映は約70本）。

(イ) レンタルやDVD販売、Video on demand：在庫数は店舗によるが数千~数万本規模となる。

本研究では、(ア)、(イ)の両者を対象としてシステムを構築している。対象によって、母集団となる映画作品や視聴環境は異なるが、嗜好モデルは両者を统一的に扱う共通のものを用いた。ただし、2006年度の公開作品の統計で、年間821本公開のうちの上位10本（本数比1.2%）が全興行収入の36%を占めていることから分かるように、特に、新作劇場公開映画については、ユーザは宣伝規模が大きなメジャー作品を非常に好んで観ることがよく知られている。評価実験の節（5.2節）でも述べるとおり、劇場公開映画を推薦する際には、対象ユーザが人気や宣伝をどの程度重視するかによって推薦リストを修正する、といった運用上の工夫が必要であると思われる。

推薦に利用する映画情報としては、映画に付随するジャンルや製作年などのメタデータ、および、映画の解説文やレビュー文を利用している。

#### (4) ユーザ集合

ユーザ集合が会員組織である場合とそうでない場合とで性質が異なる

(ア) 会員：ユーザ数はそれほど大きくないが、会員登録時などにユーザ属性を細かく取得できる。また特

定の映画情報へのアクセスや購買履歴も取得できる。

(イ) 非会員：ユーザ数は大きいユーザ属性は取得できない。ただし、履歴情報についてはCookie IDなどで対応づけることにより継続的に取得可能なことも多い。また、推薦要求を受ける際にユーザへの簡単な質問によりユーザ情報を取得して推薦することなども考えられる。

本研究では会員登録情報または推薦要求時の質問により、ユーザ属性や履歴を得られることを想定する。

#### (5) ユーザの嗜好評価の性質と粒度

明示的評価と暗示的評価の2種類がある。

(ア) 明示的評価：ユーザが映画の感想アンケートに答えたり推薦結果に対するフィードバックを行うことにより個々の映画に対する評価を入力する。通常は5段階で評価されることが多い。

(イ) 暗示的評価：映画情報サイトにおける映画紹介ページへのアクセスやDVDの購買履歴などを映画作品への評価と見なす。

本研究では明示的評価が得られることを想定する。

#### (6) 嗜好評価の密度

映画評価の密度を(評価値数)/(コンテンツ数 × ユーザ数)とすると、通常明示的評価では密度はあまり高くなく、ユーザや映画によりかなりばらつきがある。ミネソタ大学のMovieLensプロジェクトで公開しているデータセットはユーザ数が6,040、コンテンツ数が3,900、密度が約4.2%であった。本研究で収集したデータについては、3.4節で述べる。

### 2.2 課題と解決のアプローチ

映画推薦システム実現のために解決すべき課題は以下の3点となる。

[課題1] ユーザの状況と個性を反映する。

[課題2] 履歴が少なくても一定の推薦精度を実現する。

[課題3] 候補となる映画作品数が多くても推薦要求に対して実用的な応答時間を達成する。

上記の課題解決のため以下のアプローチをとる。

1) 課題1, 2に対しては、ユーザがある状況である映画を評価した結果に関するアンケートデータを収集し、ベイジアンネットによりアンケートの各変数間の関係性をモデル化し、このモデルを用いて映画推薦を行う。

2) 課題3に対しては、映画推薦システムの実装上の工夫として、候補となるすべての映画についての評価を予測するのではなく、近似的な手法により1回の予測計算で推薦映画を決定することとする。以降、各アプローチの詳細をそれぞれ3章、4章で述べる。

### 3. 映画嗜好ベイジアンネットワークモデルの構築手法の提案と構築したモデル

本章では、映画推薦システムで利用するための、ユーザ情報、映画情報、ユーザの評価情報、状況情報を利用したベイジアンネットワークによる映画嗜好モデルの具体的な構築方法および構築したモデルについて述べる。

#### 3.1 映画推薦のためのベイジアンネットワークモデル

ベイジアンネットワークは複数の確率変数の同時確率分布をモデル化する手段である。確率変数はネットワーク上のノードとして表現し、変数間の依存関係はノード間の有向リンクで表現する。変数間の条件付き独立性はネットワークの全体構造として表現され、効率的な確率推論に利用される。

映画嗜好モデルをベイズ情報処理の枠組みで定式化すると、観測可能なユーザ属性 ( $U$ )、観測可能な映画属性 ( $C$ )、観測可能なユーザの状況 ( $S$ )、ユーザの映画評価 ( $V$ ) 間の確率的依存関係を同時確率分布  $P(u, c, s, v)$  によってモデル化することとなる。

映画推薦の場合は、映画評価に関する条件付き確率  $P(v|u, c, s)$  を対象ユーザ  $U = u$ 、ユーザの状況  $S = s$  と候補映画  $C = c$  について計算し、高い評価を得られる確率の高い映画から順に推薦する。あるいは、条件付き確率  $P(c|u, s, v)$  を対象ユーザ  $U = u$  と評価  $V = \text{positive}$  について計算し、確率の高い映画作品を探すこともできる。

#### 3.2 映画嗜好モデルの構築方針

ベイジアンネットワークはネットワーク構造とネットワーク上の各ノードに付随する条件付き確率表 (CPT: Conditional Probability Tables) で規定される。ベイジアンネットワークにおけるモデル構築は大きく分けて下記の3通りのアプローチで行われる。第1のアプローチは対象とする分野や現象に関する知識を活用して、ネットワーク構造と条件付き確率表をすべて手で構築する方式である。第2のアプローチはデータからネットワーク構造と条件付き確率表を自動で推定する方式である。第2のアプローチには2通りの方法がある。第1の方法は、候補となる変数から AIC や MDL のような情報量基準を利用してネットワーク構造と変数を選択する方法である。典型的な例としては、Cooper らによる K2 アルゴリズムがある<sup>9)</sup>。第2の方法は、ノード間のリンクの有無を独立性の検定を利用して決定する方法である<sup>7)</sup>。ところがこれらの第2のアプローチはいずれも膨大な計算量を必要とするため小規模なネットワークの構築に限られる。第3のアプローチは大まかなネットワーク構造を手動で規定

し、個々のリンクの有無や条件付き確率表の値をデータから推定する方法である。本アプローチは対象とする分野に関する知識と学習データの双方が得られる場合に用いられる。また、本アプローチは実用的でありいくつかの実際の問題に適用されている<sup>4),14)</sup>。ところがこれらの事例で用いられているモデル構築の過程はヒューリスティックスを含んでおり標準的な手法は確立されていない。

本研究では、3.3 節で述べるように観測データに含まれる確率変数の数が膨大であるため、第3のアプローチを採用した。

#### 3.3 映画嗜好モデルのためのデータ収集

映画嗜好モデルはユーザがある状況である映画を評価した結果に関するアンケートデータから構築する。アンケートデータの収集は、1) 定性調査による質問候補の抽出と選定、2) 定量調査の2段階で実施した。

質問項目抽出のための定性調査はインタビュー形式<sup>24)</sup> および WEB 形式<sup>25)</sup> のグルーピング評価グリッド法で実施した。いずれも被験者に映画を提示して好きな映画グループと嫌いな映画グループに分類させ、好きな映画の好きな理由を収集する。被験者数はインタビュー形式が 17 人、WEB 形式が 1,408 人であった。収集方法の詳細は文献 24), 25) を参照されたい。定量調査は以下のとおり実施した。

1. 被験者数: 2,153 人
2. 実施時期: 2006 年 3 月
3. 採用条件: 映画鑑賞頻度が比較的高い被験者
4. 映画作品数: 197 作品
5. 映画評価値: 各被験者に視聴経験のある作品からランダムに 5~10 作品提示し評価させた。
6. 調査項目:
  - (ア) ユーザデモグラフィック・ライフスタイル属性 (30 属性): 年齢, 性別, 職業, 新商品には目がない, 対人関係に気を使う, など
  - (イ) ユーザの映画視聴に関する態度属性 (32 属性: 7 段階): 映画選択時の重視項目 (人気・話題性, ストーリーなど), ジャンル選好度合 (ホラー, コメディ, アクションなど), 映画を見る主な目的 (笑いたい, 泣きたい, 怖がりなど) など
  - (ウ) 映画鑑賞時の状況属性 (43 属性): 映画を観た場所や手段 (映画館で観た, 自宅で DVD をレンタルして観た, など), 日時 (休日, 平日, 午前中, など), だれと観たか (1 人で観た, 夫婦で観た, 親子で観た, 恋人と 2 人で観た, など 7 属性), 状態 (元気を出したい

- ときに観た, 疲れたときに観た, など) など
- (工) (各映画について)鑑賞した映画に対する印象属性 (358 属性: 7 段階): 映画の特徴に即した印象 (映像が美しい, 音楽が印象に残る, ストーリーが壮大, 脇役が名優, など), 映画を観たときに感じた気持ち (悲しい結末, テンポが軽快, 話が理解しやすい, など), 映画の全体的な印象 (感動した, 泣けた, 明るい, 面白おかしい, など) など
- (オ) (各映画について)鑑賞した映画に関する総合評価 (1 属性: 7 段階)

本アンケートの特徴は, ユーザの状況属性 (43 属性) やユーザの印象属性 (358 属性) など, MovieLens などの他のデータセットに比べて属性数が非常に多いことがあげられる. 嗜好評価の密度は約 2.2% であり, MovieLens の約 4.2% よりもさらに低い. 映画属性としては, ジャンルや製作年, 製作国などのメタデータと, 映画紹介文から抽出したキーワードをあわせて 26 属性を利用した.

ユーザが評価した映画作品数は 1 人あたり平均 4.26 作品, 中央値は 5, 標準偏差は 1.84 であった. 映画作品の被評価数については 1 作品あたり平均 44.6 人, 中央値は 30, 標準偏差は 43.56 であった.

3.4 映画嗜好モデル構築手順

まず, 定性調査の結果に基づき, ユーザの映画作品に対する評価を予測するための大まかなネットワーク構造の仮説を図 1 に示すように立てた. この仮説は, 主に評価グリッド法の結果に基づくもので, ユーザの総合評価は, “泣けた”, “笑った” などの印象属性グループとのみ直接的に関係し, ユーザ属性グループ, 状況属性グループ, 映画属性グループは印象属性グループを介してのみ総合評価グループと関係する, という階層的な構造を仮定するものである. リンクの向きについては条件付き確率テーブルが大きくなりすぎないように, 逆向きリンクで表現することにした.

次に, 上記のネットワーク構造の仮説に基づき, 観測データ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部分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き確率表を推定した. 詳細なステップを下記に示す. なお, 収集したアンケートデータから性能検証用データとして各ユーザからランダムに評価値を 1 つずつ抽出し, 残り学習用データとしてモデル構築に利用した.

1. データ前処理: 映画紹介文から映画属性を抽出する. 本モデル構築では以下を実施した.
  - (ア) 映画紹介文からキーワードを抽出した.

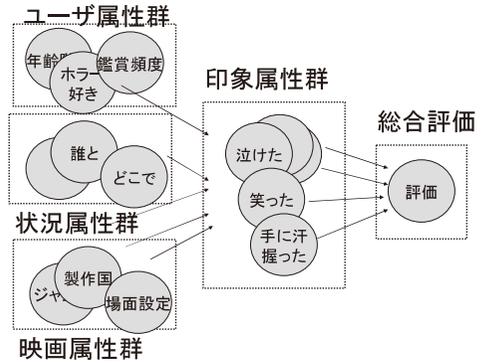


図 1 モデル構造に関する仮説 Fig.1 Model structure.

- (イ) 抽出したキーワードを「場面設定」などの 4 種類のカテゴリに大まかに分類した.
- (ウ) 各カテゴリ内のキーワードをさらに 3~5 種類に集約した.
2. 擬似映画属性の抽出: 映画属性を拡張するため, 印象属性のうちユーザごとの評価の差が少ないものを擬似映画属性とし, モデル構築時に映画属性として扱った. 具体的には学習データ中の各印象属性について,  $score = I(r, CID)/H(r|CID)$  を計算し,  $score$  の高い印象属性群を擬似映画属性とする.  $I(r, CID)$  は印象属性の値  $r$  とコンテンツ ID との相互情報量であり,  $H(r|CID)$  はコンテンツ ID が与えられた場合の  $r$  の条件付きエントロピーである. この理由は, コンテンツ ID との関連性が高く, ユーザごとの差が少ないものほど映画の属性を表現していると考えたためである. 本モデル構築では 30 属性を擬似映画属性として抽出した.
3. 変数群のグループ化: 変数群を仮説に基づくカテゴリにグループ化する. 本モデル構築では, ユーザ属性 ( $U$ ), ユーザ状況 ( $S$ ), 擬似属性を含む映画属性 ( $C$ ), 擬似属性を除いた印象属性 ( $I$ ), 総合評価 ( $V$ ) の 5 種類に分類した.
4. 各グループの属性のクラスタリング: 各グループについて, グループ内のすべての属性間の相互情報量を求め, 続いて各クラスタを代表する属性 (代表属性) を抽出する. 本モデル化では Ward 法<sup>22)</sup> により属性をクラスタリングした. 各グループのクラスタ数 (= 代表属性の数) を表 1 に示す.
5. 構造探索: グループ間のネットワーク構造を決定するため, 各グループの代表属性間の親子関係を探索した. ここでは, 候補となる構造を複数生成し, 情報量基準 (本モデル化では AIC を利用) を

表 1 グループごとのクラス数

Table 1 The number of clusters for each group.

グループ数	属性数	変換後のグループ数	クラス数
デモグラフィック・ライフスタイル属性	30	デモグラフィック・ライフスタイル属性	10
態度属性	32	態度属性	10
状況属性	43	状況属性	7
印象属性	358	印象属性	24
		映画属性	23
映画属性	26		

表 2 生成したネットワークのノード数とリンク数

Table 2 The number of nodes and links in the constructed networks.

	Number of nodes	Number of links
UCS-I-V	75	115
UC-I-V	68	107
UCS-V	50	49
UC-V	44	43

用いて最もデータに適合する構造を選択する。この過程では本村が開発したベイジアンネットワークツール BAYONET<sup>27)</sup> を利用した。候補となる構造の生成については、親ノードの最大数を 4 として生成した。後で述べる評価実験で用いるために以下の 4 種類のモデルを生成した。第 1 のモデルは 5 グループすべて (V, I, U, C, S) を利用した。第 2 のモデルは S を利用しなかった。第 3 のモデルは S を利用したが I を利用しなかった。第 4 のモデルは S も I も利用しなかった。構築したネットワークの情報を表 2 に示す。

#### 4. 映画推薦システムの設計と実装

映画推薦システム構成を図 2 に示す。Java Servlet 環境で実現した。ユーザに対する携帯 WEB ベースのインタフェースを提供するインタフェースモジュール、ユーザ嗜好モデルに基づいてベイジアンネット推論を行う推論エンジンを利用して対象となるユーザの推薦映画リストを出力する推薦モジュール、および、ユーザ情報、ユーザ評価履歴、映画情報をそれぞれ格納するデータベースから構成される。

ユーザに最適な映画を発見するためには、推薦要求

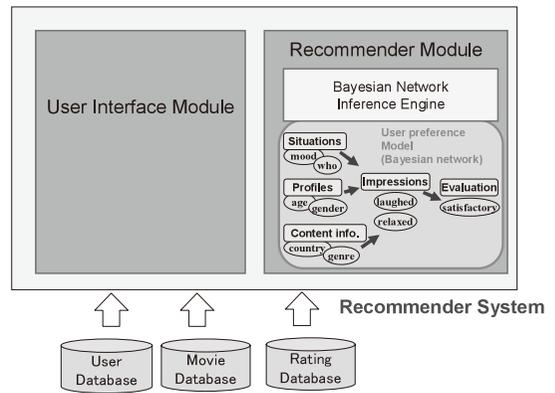


図 2 映画推薦システムの構成

Fig. 2 Overview of movie recommender system.

#### ③お勧めコンテンツとお勧め理由の提示



図 3 映画推薦システムの処理の流れ

Fig. 3 Flow of movie recommendation.

があったユーザに対する全映画の予測評価を求め、予測評価値の高い映画から順に薦めることが考えられる。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、以下の手法を提案・実装した。まず、各映画に対応する印象属性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は、ユーザ属性とユーザ状況を入力として印象属性の事後確率を求め、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。本方式に基づく推薦システムの処理の流れ、推薦画面をそれぞれ図 3、図 4 に示す。なお推薦映画の画像・タイトルは著作権上の理由から架空のものに置き換えてある。

1. 映画登録時に当該映画の特徴ベクトル (= 印象属性の値) を計算し、データベースに登録しておく。



図 4 映画推薦システムの画面例

Fig. 4 Screenshots of cellular phone display.

2. ユーザから携帯電話経由で状況に関する情報とともに映画推薦要求を受け付ける。
3. データベースから登録済みのユーザ属性情報を取得し、ユーザから入力された状況と合わせて特徴ベクトルを算出する。
4. ユーザの特徴ベクトル値とあらかじめ計算済みの映画作品群の特徴ベクトルとの類似性を判定し、類似度の高い映画を推薦する。
5. 推薦した映画に対するフィードバックをユーザから受け付け、予測精度の向上のための学習に用いる。

## 5. 映画推薦システムの評価

### 5.1 映画推薦モデルの精度評価

検証用データ内のユーザと映画の組について評価値を予測し、正解値との予測誤差をモデル間で比較した。予測精度の基準としては、平均絶対誤差 (MAE: Mean Absolute Error) を採用した。全評価値数を  $N$ 、評価の状態値数を  $r$ 、ユーザ  $i$  の映画  $j$  に対する状況  $k$  での正しい評価値を  $P_{ijk}$ 、予測評価値を  $v$  とすると、MAE は以下のように計算される。

$$\frac{1}{N} \sum_{ijk} \left| p_{ijk} - \sum_{v=1}^r v P(V=v|U=i, C=j, S=k) \right|$$

4 種類のベイジアンネットワークモデルを用いた予測、ベースライン予測、協調フィルタリング方式に基づく予測の 6 種類を比較した。ベースライン予測としては、各映画の平均評価値を予測値としたものを利用した。協調フィルタリング方式については 2 種類を利用した。1 つ目はユーザ間の評価値の類似性を利用したユーザベースの協調フィルタリング方式 (CF(user)) で、2 つ目は作品間の評価値の類似性を利用したアイ

表 3 モデルの精度評価結果

Table 3 Result of accuracy evaluation.

Model	MAE
Baseline	0.927
CF (user-based)	0.975
CF (item-based)	0.930
UC-V	0.887
UCS-V	0.869
UC-I-V	0.862
UCS-I-V	0.854

テムベースの協調フィルタリング方式 (CF(item)) である。類似性尺度にはピアソン相関値を利用した。

表 3 に評価結果を示す。実験ではユーザベースの協調フィルタリング方式は 4.1% のユーザ (72 人) に対してのみ予測結果を算出できた。またアイテムベースの協調フィルタリング方式では 12.6% のユーザ (221 人) のみ予測結果を算出できた。また、いずれの場合も MAE の値は悪かった。これは学習データの 1 人あたりの評価値数が少なかった (平均 3.26 個) が原因と考えられる。

4 種類のベイジアンネットワークモデルのすべてがベースライン予測と協調フィルタリング予測を上回った。印象属性を有するモデル (UC-I-V, UCS-I-V) が、印象属性を持たないナイーブベイズモデル (UC-V, UCS-V) よりも良い予測精度を示したことから、階層的なモデル構造と印象属性の利用は有効であったといえる。

また、状況変数を含むモデル (UCS-I-V, UCS-V) は状況変数を持たないモデル (UC-I-V, UC-V) よりもそれぞれ良い予測精度を示した。このことは、状況を考慮することが、嗜好の予測精度の向上のために有効であることを示している。

### 5.2 映画推薦システムの主観評価

被験者に映画推薦システムを実際に使用してもらい、アンケートによる主観評価を行った。評価の目的は、携帯電話による映画推薦システムへのニーズ、実装したシステムの使い勝手、そして推薦結果に対する満足度を調べることである。

1. 被験者数: 171 人 (10 代から 50 代までの男女)
2. 選定方法: ストリートキャッチ方式 (会場周辺の通行人であらかじめ設定された条件を満たす人を選ぶ) による会場テスト
3. 実施時期: 2007 年 1 月
4. 選定条件: 映画関係者でなく、映画鑑賞頻度が比較的高く、携帯電話によるメールやウェブ閲覧の経験がある人

## 5. 推薦対象：調査日時時点で公開中の映画

被験者は評価用の携帯電話を用いて推薦システムにログインし、映画推薦モデルで採用している年齢、性別、職業、ジャンル選好度合い、気分など19項目を入力して推薦結果を得る。各属性の値の入力は、5から7個程度の候補からプルダウンメニュー形式で選択することで行った。気分については、笑いたい、泣きたい、怖がりしたい、手に汗握りたい、スカッとしたい、癒されたい、感動したい、の7種類から選択した。この操作を数回行った後に、システムへのニーズと使い勝手に関する質問に回答した。

続いて、① 興行収入ランキング上位リスト、② ユーザ属性のみを入力した場合の推薦リスト、③ ユーザ属性と気分の両方を入力した場合の推薦リスト、④ 興行収入上位20作品（人気作品）の中でユーザ属性と気分の両方を入力した場合の推薦リスト、⑤ 興行収入21位以降（人気作品以外）でユーザ属性と気分の両方を入力した場合の推薦リスト、の5つのパターンの推薦リスト（いずれも上位5位まで）を紙上で提示し、現在の気分に適した推薦と思われる順に順位付けを行ってもらった。被験者にはそれぞれのリストの作成方法については説明していない。

結果の一部を下記にまとめる。

1) 映画推薦システムへのニーズに関する質問「あなたは、今後『おすすめ映画サービス』を携帯電話サービスの1つとして利用したいですか？」に対しては、81%が利用を希望した。

2) 「プロフィールと気分の両方を入力して自分専用に気分にあったおすすめをしてほしい」という問いに対しては、システムを利用したいと答えた被験者の約60%が、「そう思う」あるいは「ややそう思う」と答えており、その比率は、「プロフィールだけを入力して自分専用におすすめしてほしい」「気分だけを入力しておすすめしてほしい」と思う被験者よりも多かった。

以上から、携帯電話による、状況を考慮した映画推薦サービスへのニーズは高いといえる。

次にシステムの使い勝手に関しては、

1) 「操作の仕方や画面の見やすさなどを総合的に考えた場合、使いやすいと思いますか？」という質問に、76%が「使いやすい」あるいは「やや使いやすい」と答えている。

2) 「プロフィールの入力を手間と感じましたか？」には48%が「手間と感じない」「あまり手間と感じない」と答えている。

3) 「気分の入力を手間と感じましたか？」については、62%が「手間と感じない」「あまり手間と感じない」と

表4 5種類の推薦リストに対して、それぞれを最も今の気分に合わせているとしたユーザの割合

Table 4 Ratio of users who chose the recommendation list as most suitable to current mood.

推薦リストで考慮した情報	全ユーザ	人気ランキングを参考にしないユーザ
①興行収入	36 %	22 %
②ユーザ属性	11 %	14%
③ユーザ属性+気分	15 %	20%
④ユーザ属性+気分（人気作品）	24 %	20%
⑤ユーザ属性+気分（人気作品以外）	14 %	24%

答えている。

以上から、今回実装したシステムの使い勝手は良く、またプロフィールや状況の入力に対してあまり負担を感じない被験者が多いことが分かる。

最後に、5種類の推薦リストの評価に関しては、表4に示すように、そのリストが5種類中最も現在の気分に適している、と答えた被験者が多い順に、① = 36%、④ = 24%、③ = 15%、⑤ = 14%、② = 11%であった。ここで、興行収入ランキング上位リスト①を最適とした被験者が約4割を占めているのは、今回の実験での推薦対象が劇場公開中の映画であったためと考えられる。

一方、その4割を除いた残りの約6割の被験者は、なんらかの形で自分専用に推薦されたリストを好んでいることが分かる。そこで、①以外の推薦リストの優劣について詳細に検討する。まず、ユーザ属性と気分の両方を考慮した推薦リスト③が気分を考慮せずにユーザ属性のみを考慮した②に対して約1.8倍のユーザから支持を得ていることは、気分を考慮した推薦の有効性を示しているといえる。

また、気分に加えて人気も考慮した推薦リスト④が③の1.6倍のユーザから支持されていることから、公開中の映画を対象とする場合には、人気も合わせて考慮した推薦を行うことによりユーザ全体の平均的な満足度は高くなることが分かる。

さらに、映画を選択する際に「人気ランキングを見て決める」という質問に対して「あまりあてはまらない」「あてはまらない」と答えた被験者（59人）のみを抽出して集計したところ、⑤ = 24%、① = 22%、④ = 20%、③ = 20%、② = 14%となり、興行収入21位以降（人気作品以外）でユーザ属性と気分の両方

を考慮したリスト⑤を最適とする被験者数が、①を最適とする被験者数を上回り最大となった。このことは、ユーザ属性に加えて気分によって推薦リストを変えることで、より効果的な推薦ができる可能性を示している。

## 6. ま と め

本稿では、個性と状況の両方を考慮したモバイル端末のための映画推薦システムの実現と評価について述べた。まず、定性調査と大規模な定量調査の組合せによって、ユーザの嗜好を推定するペジアンネットワークを構築する手法を提案した。独自に映画ユーザに対するインタビューと大規模な Web アンケートを実施し、データからユーザあるいは映画を表現するための属性変数を抽出し、それらの情報と場所やだれなどの状況を表す変数、およびユーザの映画に対する印象や評価の間の複雑な依存関係を、ペジアンネットワークによってモデル化した。構築したモデルを用いて、携帯端末上で映画推薦システムのプロトタイプを構築した。提案するシステムの評価として、嗜好モデルの精度に関する評価実験および構築したシステムの被験者ユーザによる主観的な評価実験を行った。

それらの結果、状況を考慮することで、ユーザの映画への嗜好をより精度良く予測できること、携帯端末による状況に依存した映画推薦サービスへのニーズが高いこと、推薦リストの満足度を向上させることができること、などが明らかになった。以上を総合して、映画の推薦に状況を利用することの有効性および既存手法に対する優位性を示すことができたと考えられる。

今後の課題としては、まず、嗜好のモデルの洗練・高度化があげられる。さらに多くのデータから嗜好モデルの洗練を行うとともに、モデルの確率パラメータの円滑化などによって少数のデータから安定した嗜好モデルを構築する方法について検討することや、本研究の手法と協調フィルタリングとを融合することによって、より満足度の高い推薦を実現すること、などは興味深い課題である。また、本研究の手法は、携帯端末における Video on Demand 配信、さらには、映画以外の WEB ページ、本、写真、ブログ、SNS などのコンテンツ推薦にも有効であると考えられる。今後、携帯端末以外のデバイス上のサービスも含め、幅広いサービスに対して適用していきたい。

謝辞 日頃ご指導いただく KDDI 研究所秋葉重幸代表取締役所長に深く感謝いたします。

## 参 考 文 献

- 1) Adomavicius, G. and Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE Trans. Knowledge and Data Engineering*, Vol.17, No.6, pp.734-749 (2005).
- 2) Basilico, J. and Hofmann, T.: Unifying collaborative and content-based filtering, *Proc. 21st Int. Conf. on Machine Learning* (2004).
- 3) Basu, C., Hirsh, H. and Cohen, W.W.: Recommendation as classification: Using social and content-based information in recommendation, *Proc. 15th National Conference on Artificial Intelligence*, Madison, WI, pp.714-720 (July 1998).
- 4) Binder, J., Koller, D., Russell, S. and Kanazawa, K.: Adaptive probabilistic networks with hidden variables, *Machine Learning*, Vol.29, pp.213-244 (1997).
- 5) Breese, J.S., Heckerman, D. and Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering, *Proc. 14th Annual Conference on Uncertainty in Artificial Intelligence*, pp.43-52 (1998).
- 6) Burke, R.: Hybrid recommender systems: Survey and experiments, *User-Modeling and User-Adapted Interactions*, Vol.12, pp.331-370 (2002).
- 7) de Campos, L.M.: Independency relationships and learning algorithms for singly connected networks, *Journal of Experimental and Theoretical Artificial Intelligence*, Vol.10, pp.511-549 (1998).
- 8) Christakou, C. and Stafylopatis, A.: A hybrid movie recommender system based on neural network, *Proc. 5th Int. Conf. on Intelligent Systems Design and Application* (2005).
- 9) Cooper, G.F. and Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data, *Machine Learning*, Vol.9, pp.309-347 (2002).
- 10) Herlocker, J., et al.: Evaluating collaborative filtering recommender systems, *ACM Trans. Information Systems*, Vol.22, No.1, pp.5-53 (2004).
- 11) Horvitz, E.: Principles of mixed-initiative user interfaces, *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems* (1999).
- 12) Jensen, F.V.: *Bayesian Networks and Decision Graphs*, Springer-Verlag (2001).
- 13) Jensen, F.V., et al.: The SACS0 methodology for troubleshooting complex systems, *Artificial Intelligence for Engineering Design, Analysis*

- and Manufacturing (AIEDAM)*, Vol.15, pp.321-333 (2001).
- 14) Mani, S., McDermott, S. and Valtorta, M.: MENTOR: A Bayesian model of prediction of mental retardation in newborns, *Research in Developmental Disabilities*, Vol.8, No.5 (1997).
- 15) Miller, B.N., Albert, I., Lam, S.K., Konstan, J.A. and Riedel, J.: Movie-Lens unplugged: Experiences with an occasionally connected recommender system, *Proc. 8th Int. Conf. on Intelligent User Interfaces*, pp.263-266 (2003).
- 16) Mobasher, B., Jin, X. and Zhou, Y.: Semantically enhanced collaborative filtering on the Web, *Proc. European Web Mining Forum*, Berendt, B., et al. (Eds.), LNAI 3209, Springer (2004).
- 17) Mooney, R.J. and Roy, L.: Content-based book recommending using learning for text categorization, *Proc. 5th ACM Conference on Digital Libraries*, pp.195-204 (2000).
- 18) Ono, C., Motomura, Y. and Asoh, H.: A study of probabilistic models for integrating collaborative and content-based recommendation, *Working Notes of IJCAI-05 Workshop on Advances in Preference Handling* (2005).
- 19) Resnick, P. and Varian, H.R.: Recommender systems, *Comm. ACM*, Vol.40, No.3, pp.56-58 (1997).
- 20) Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedel, J.: GroupLens: An open architecture for collaborative filtering of netnews, *Proc. ACM Conference on Computer Supported Cooperative Work*, pp.175-186, ACM (1994).
- 21) Shardanand, U. and Maes, P.: Social information filtering: Algorithms for automating 'word of mouth', *Proc. CHI'95 Mosaic of Creativity*, pp.210-217 (1995).
- 22) Zekerman, I. and Alberecht, D.W.: Predictive statistical models for user modeling, *User Modeling and User-Adapted Interaction*, Vol.11, No.1-2, pp.5-18 (2001).
- 23) Ward, J.H.: Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.*, Vol.58, pp.236-244 (1963).
- 24) 芳賀, 小野, 本村: グルーピング評価グリッド法の開発と応用可能性の検討—映画コンテンツ評価構造の探索のために, 日本行動計量学会第33回大会発表論文抄録集, pp.130-131 (2005).
- 25) 芳賀, 小野: Webによるグルーピング評価グリッド法の開発と利用—また見たい映画はどんな映画?, 日本行動計量学会第34回大会発表論文抄録 (2006).
- 26) 本村, 岩崎: ベイジアンネットワーク技術, 東

京電機大学出版局 (2006).

- 27) 本村: ベイジアンネットソフトウェア BayoNet, 計測と制御, Vol.42, No.8, pp.693-694 (2003).

(平成 19 年 4 月 16 日受付)

(平成 19 年 10 月 2 日採録)



小野 智弘 (正会員)

1992 年慶應義塾大学理工学部電気工学科卒業。1994 年同大学大学院理工学研究科修士課程計算機科学専攻修了, 同年国際電信電話 (株) 入社。1999 年 9 月~2000 年 9 月スタンフォード大学電気工学科客員研究員。現在 (株) KDDI 研究所知能メディアグループ主任研究員。利用者嗜好抽出ユーザ嗜好抽出, ソフトウェアエージェント, データベース等の研究に従事。情報処理学会全国大会学術奨励賞受賞。電子情報通信学会会員。



黒川 茂莉 (正会員)

2005 年慶應義塾大学理工学部管理工学科卒業。2007 年同大学大学院理工学研究科修士課程開放環境科学専攻修了, 同年 KDDI 株式会社へ入社。現在 (株) KDDI 研究所 Web データコンピューティンググループに所属。ユーザ嗜好抽出, 確率モデルの構築・評価等の研究に従事。電子情報通信学会会員。



本村 陽一

1991 年電気通信大学通信学科卒業。1993 年同大学大学院電子情報科学専攻修士課程修了。1993 年通産省工技院電子技術総合研究所入所。1999 年アムステルダム大学招聘研究員。2001 年産業技術総合研究所情報技術研究部門主任研究員。2003 年同研究所デジタルヒューマン研究センター主任研究員。2005 年早稲田大学非常勤講師, モデライズ (株) 取締役を兼任。ベイジアンネットソフトウェアの開発, 人の認知・評価構造行動のモデル化と応用技術の研究に従事。人工知能学会研究奨励賞, 全国大会優秀賞, DCS 船井賞, ドコモモバイルサイエンス賞等受賞。電子情報通信学会, 人工知能学会, 神経回路学会, 行動計量学会, 認知科学会, マーケティングサイエンス学会, IEEE 各会員。



麻生 英樹

1981年東京大学工学部計数工学科卒業。1983年同大学大学院工学系研究科情報工学専攻修士課程修了。同年通商産業省工業技術院電子技術総合研究所入所。1993～1994年ドイツ国立情報処理研究センター客員研究員。現在、独立行政法人産業技術総合研究所情報技術研究部門主任研究員。学習能力を持つ知的情報処理システムの研究に従事。電子情報通信学会，人工知能学会，日本神経回路学会，行動計量学会各会員。

---