

嗜好の個人差と状況依存性を考慮した映画推薦方式の検討

小野 智弘[†] 本村 陽一[‡] 麻生 英樹[‡]

株式会社KDDI研究所[†]

独立行政法人 産業技術総合研究所[‡]

近年の情報やコンテンツの爆発的な増大およびユーザニーズの多様化に伴い、ユーザが欲する情報の選択を支援する推薦システムへの要求が益々高まっている。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等、状況や気分に応じてユーザの嗜好は変化する。これまでの推薦方式はユーザの履歴情報やプロフィール、コンテンツの特徴に基づいて推薦を行う方式が主流でユーザの状況に関わらず同じ結果となってしまう。一方場所などの状況に基づいて推薦候補を絞りこむ方式も提案されているが同一の状況に対してはどのユーザにも同じコンテンツが推薦されてしまう。筆者らはこれらの問題を解決するためにユーザの履歴やプロフィール等の情報と気分や場所などの状況の依存関係をベイジアンネットにより表現し、個人差と状況に応じて最適なコンテンツを推薦する方式の研究を行っており、本発表ではこの経過を報告する。

Study of Movie Recommendation System considering both users' personality and situation.

Chihiro Ono[†], Youichi Motomura[‡] Hideki Asoh[†]

KDDI R&D Laboratories Inc.[†]

National Institute of Advanced Industrial Science and Technology[‡]

With the flood of various information and contents through the Internet, the need for recommendation systems that assist users in finding the information they desire is increasing. In case of movie recommendation, user preference may change according to his situation such as mood and location etc.. Existing approaches which use user profiles and/or histories are insufficient as they provide the same results regardless of users' situation. Context-aware methods such as location-based services are neither sufficient as they cannot provide personalized recommendation. In this paper we propose a recommendation system which achieves context-aware personalized recommendation based on Bayesian network by making use of various information related to user profiles, histories, situations, and content attributes.

1. はじめに

近年の情報やコンテンツの増大およびユーザニーズの多様化を受け、ユーザが欲する情報の選択を支援する推薦システムへの要求が益々高まっており、amazonの書籍の推薦やTutaya Online

のレンタル・販売DVDの推薦等効果を挙げている例も報告されている[1][2]。これまでにユーザ個別の嗜好に基づく推薦を行う枠組みとして、ユーザの購買やコンテンツ評価履歴からユーザの嗜好傾向を探り、嗜好傾向の類似した他人が好むコ

コンテンツを推薦する協調フィルタリング方式や、コンテンツそのものや属性群からユーザの嗜好構造を反映する特徴の抽出を行い、ユーザの好きなコンテンツ群の特徴と類似した特徴を持つコンテンツを推薦するコンテンツベース(属性ベース)の方式が提案されている[3]。これらは一定の効果があるものの、前者は大量の履歴データが必要なこと、後者は適切な特徴抽出が困難といった欠点がある。このため、両者を組み合わせた方式として、両者の結果を重み付けして結合した方式[4][5]や、アスペクトモデルと呼ばれる潜在クラスを用いた統計的混合分布モデルで結合した方式[6][7][8][9]が検討されている。Hofmannらはキーワードとコンテンツの共起データを潜在変数の集合と結びつけるためにアスペクトモデルを導入し[6]、さらにその拡張として、ユーザとコンテンツの共起データにアスペクトモデルを適用して協調フィルタリングに適用した[7]。Popesculらはさらに拡張し、ユーザ、コンテンツ、キーワード間の共起情報を元にベイジアンネットワークでユーザの嗜好を表現した[8]。小野らは複数のユーザ属性と複数のコンテンツ属性の活用を試みた[9]。ところが映画鑑賞では、家族と観るか恋人と観るか、元気であるか沈んでいるか等、状況や気分に応じてユーザの嗜好は変化するが、上記の推薦方式はユーザがどんな状況でも同じ結果となってしまうという問題点がある。

一方、Tsutaya Online[2]におけるユーザの気分やシチュエーション別検索やユーザの位置に応じた情報提供サービス等が提供されているが、これらは同一の状況に対してはどのユーザにも同じコンテンツが推薦されてしまう。またユーザの指定した項目に基づいて検索を行う場合、項目が増えていくと該当コンテンツが少なくなるという問題がある。以上をまとめると表1のようになる。

ユーザの個性と状況を同時に考慮した推薦を

行うためには、ユーザの履歴やプロフィール等の情報と気分や場所などの状況の情報をユーザの個性や状況に応じて重み付け等を変更しながら統合する必要があるが、既存の方式の単純な組み合わせでは実現は困難である。

表1：推薦方式の分類

	個人差 への対応	状況 への対応
協調フィルタ		×
コンテンツ(属性)ベース		×
ハイブリッド方式		×
状況ベース	×	

近年、確率事象間の複雑な依存関係をグラフ構造によって表現する手法として、ベイジアンネットワーク[10]が注目されており、ソフトウェアの機能推薦[11]やプリンタの故障診断[12]等で利用されている。

筆者らはこれまでに、ユーザのプロファイル、評価履歴、コンテンツの特徴情報間の複雑な依存関係をベイジアンネットにより表現し、コンテンツを推薦する方式を検討している[13]。本稿ではユーザの履歴やプロフィール等の情報に加えて気分や場所などの状況も含めて複雑な依存関係をベイジアンネットにより表現し、個人差と状況に応じて最適なコンテンツを推薦する方式の検討経過を報告する。以下、2章では本研究の問題設定、3章でベイジアンネットワークによる推薦方式の方針について述べる。4章ではモデルの作成概要、5章で推薦システムについて述べる。6章で考察を行い、最後に7章でまとめを述べる。

2. 映画推薦問題の記述

本研究で扱う推薦システムの目的と利用可能な情報を以下に記す。

1) 推薦システムの目的：ユーザの個人差と状況

依存性を同時に考慮してユーザのその時点の嗜好を推定し、最適な映画コンテンツを提示する。

- 2) ユーザ情報：以下が利用可能である
 - (ア) 基本属性、性別や年齢等のデモグラフィック属性や余暇の金額等のライフスタイル属性等
 - (イ) 映画視聴に関する態度属性：好きなジャンルや視聴目的等
 - (ウ) ユーザの状況：どこで、誰と、気分等
 - (エ) コンテンツ評価履歴：視聴した映画に対する評価履歴
- 3) 映画コンテンツの情報：映画コンテンツに付随するジャンルや製作年、映画の解説文等

3. ベイジアンネット

ベイジアンネットとは、変数間の複雑な依存関係を表すためにグラフ構造を利用し、依存関係のある変数間の条件付確率を局所的に定義していった非循環有向グラフで表す確率モデルである [10]。例えば20代の女性100人のうち30人があるコンテンツを購入したという統計情報があれば、20代の女性の条件付の購入確率は30%である。このような $P(\text{購入}|\text{年齢, 性別})$ という条件付確率を網羅的にモデル化できれば、目の前の新しいユーザについての購入見込みの予測や、購入する可能性の高いコンテンツの予測が可能となる。またこの確率の変化率を計算することで年齢や性別の他にも予測に強く影響する要因を発見できる。このような要因(変数)の間を結んで多段のネットワークにしたものがベイジアンネットになる。そして購買可能性のような確率値や可能性の高い商品を予測する計算を確率推論と呼ぶ。

図1に示すように、ベイジアンネットは確率変数をノードで表し、これらの間の依存関係を非循環の有向リンクで表しているが、有向リンクの元にあるノードを親ノード、先にあるノードを子ノ

ードと呼ぶ。有向リンクは親から子の向きに条件付きの依存関係があることを意味し、子ノード X にリンクを張る親ノードの集合を U とすると、この子ノードの確率変数は条件付確率分布 $P(X|U)$ に従う。たとえば、子ノードの確率変数 X が $X=x_1, \dots, x_m$ の m 通りの値を取り、親ノードの確率変数の集合 U が $U=u_1, \dots, u_n$ の n 通りの値を取る場合、条件付確率 $P(X|U)$ は $P(x_1|u_1), \dots, P(x_m|u_n)$ を並べた表(条件付確率表: Conditional Probability Table: CPT)によって与えられる。ベイジアンネットワークが表現する確率分布は、ノード集合とグラフ構造と各ノードに割り当てられる CPT とによって定義される。

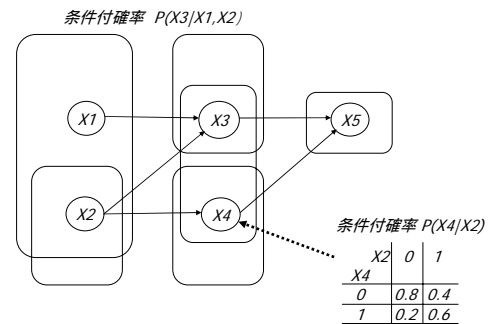


図1 ベイジアンネット概要

各ノードにおけるCPTは、完全データであればデータから学習が可能であり、データ数が十分大きい場合は最尤推定により求めることができる。一方ユーザのデータや状況のデータなど簡単には収集できないような問題の場合は不完全データとなるため、EMアルゴリズム [14] 等で確率分布を推定して求める必要がある。こうしたモデル構築を容易にするソフトウェアとしてBayonet [15] 等がある

4. ユーザ嗜好モデルの構築

4.1. モデルの構築方針

ベイジアンネットを用いてユーザの嗜好をモデル化する場合、ユーザ情報、コンテンツ情報、ユーザの評価情報、状況情報の間の依存関係をモデル化する必要がある。ベイジアンネットにおける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すべて手動で構築する方式、データから自動で探索する方式、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方法(半自動の方法)との3通りが考えられるが、学習用データが得られかつ構造に関する仮説がある場合には半自動方式が優れていることが多いため、今回は半自動方式を取ることとし、おおまかな仮説として図4のような評価構造を用いた。この評価構造は、ユーザのコンテンツ評価が、コンテンツに対する感じ方によって決まり、感じ方はユーザ属性、状況属性、コンテンツ属性によって決まるということを意味している。また、学習用データは、WEB アンケート調査によって収集した。

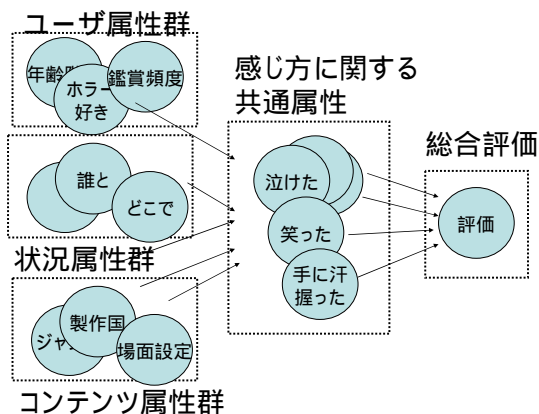


図2 評価構造の仮説

4.2. データ

2004年12月(第1回目)と2005年8月(第2回目)と2回に分けてWEBアンケート調査によりデータを取得した。

第1回目はユーザ属性、コンテンツ属性、コンテンツ評価履歴を取得した。概要は以下の通り。

- 1) 被験者数：1616人
- 2) リクルート条件：提示した100コンテンツ中20作品以上視聴経験のある被験者
- 3) 調査項目(各ユーザ):
 - (ア) デモグラフィック・ライフスタイル属性：年齢、性別、職業等27項目
 - (イ) 映画視聴に関する態度属性：映画コンテンツ(自宅での鑑賞も含む)の鑑賞頻度、映画選択時の重視項目、映画を見る主要目的(感動したい等7項目)
 - (ウ) ユーザのコンテンツ評価履歴：視聴経験のある20作品(ランダムに抽出)について、評価{良い・悪い}、その時の気分(感動した等7項目)
- 4) コンテンツ属性(100コンテンツ分):
 - (ア) メタ属性：ジャンル、収録時間等
 - (イ) 特徴属性：映画紹介文からテキストマイニングで抽出したキーワード群

第2回目は状況属性を取得した。概要は以下の通り。

- 1) 被験者数：1088人
- 2) リクルート条件：提示したコンテンツのうち15件以上が見たことの有る被験者
- 3) 調査項目(各ユーザ)：以下の項目を自由記述で取得
 - (ア) どこ(場所)
 - (イ) 誰と何人
 - (ウ) どんな映画どんな時

4.3. モデル構築手順

モデル構築ではベイジアンネット構築ツール BayoNet [15] を利用し、以下の手順で行った。

- 1) データの加工：コンテンツに付随するコンテンツの説明文からテキストマイニングによりキーワードを抽出し、抽出したキーワードを場面設定等の4種類に大まかに分類した。続いて各分類内のキーワードをそれぞれ3

- ～ 5 種類に集約した。状況データについては、自由記述された 3 種類のデータをそれぞれについて 3～5 種類に分類・集約した。
- 2) 属性のグループ分け：各種属性を仮説に基づきユーザ属性群，コンテンツ属性群，共通属性群，総合評価にグループ分けした。
 - 3) 部分モデルの構築：まず Bayonet の Greedy 探索により親ノード候補を探索し，続いて Tree Selection 機能で AIC/MDL[16]等の情報量基準を活用しながら依存関係を細かく分析し，親ノード群を決定した。
 - 4) 全体モデルの構築：部分モデルを仮説に基づき結合して全体モデルを構築した。なお，情報量基準や条件付確率票（CPT）の分布形を見ながらリンクの方向を調整した。
 - 5) モデルの検証：上記の手順で構築したモデルに対し，1つのノードの状態値を固定する（証拠情報を与える）ことにより条件を設定し，その他のノードの確率値をベイジアンネットワークの確率推論によって求め，これを実際のテストデータとの一致性をみることによって検証を行った

構築したモデルは，ユーザ属性群（22 ノード）とコンテンツ属性群（6 ノード），状況属性群（3 ノード），共通属性（7 ノード），総合評価の合計 36 ノードとなった。

5. 推薦システムの実装

構築したユーザ嗜好モデルを用いて映画推薦システムのプロトタイプを構築した。推薦システムの処理の流れを図 3 と以下に示す。

- 1) ユーザから携帯電話または PCWEB 経由で状況に関する情報とともにコンテンツ推薦要求を受け付ける。
- 2) データベースから登録済みのユーザ属性情報を取得し，状況を利用して特徴ベクトル（＝共通属性の値）を算出する。

- 3) ユーザの特徴ベクトル値とあらかじめ計算済みのコンテンツ群の特徴ベクトルとの類似性を判定し，類似度の高いコンテンツを推薦する。
- 4) 推薦したコンテンツに対するフィードバックをユーザから受付，予測精度の向上のための学習に用いる。

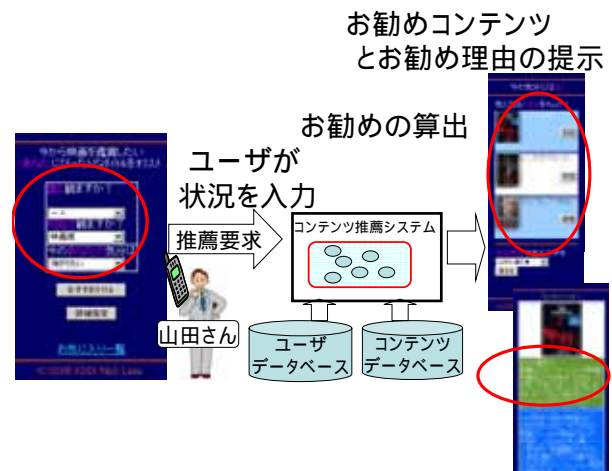


図3 コンテンツ推薦システムの概要

推薦画面のスクリーンショットを図4に示す。画面左側にユーザ毎の推薦結果およびユーザ個別の推薦理由が出力されている。ユーザは画面右側上部のボタン押下により気分や手段，誰と等を設定・変更することにより推薦結果をカスタマイズできる。また，同じく画面右側上部の「なったつもり」メニューにより家族や有名人等の個性を利用して気分等の状況に基づく推薦結果のカスタマイズができる。他人の個性の取得方法は，サーバからのダウンロードや相互交換等が考えられる。このように価値観の単位でシステム間でやりとりできることも本システムの大きな特徴となる。



図4 PCベースの推薦画面

6. 考察

- 1) 嗜好のモデル化では総合評価に強い影響を与える変数を抽出することが重要である。これまでの研究では既に取りあげたユーザ属性やコンテンツ属性について評価との相関などから選択することが多く、本研究で構築したベイジアンネットモデルでも既存のユーザ属性やコンテンツ属性を利用している。現在、より良い変数をユーザから抽出するため、評価グリッド法に基づく調査を行っており[17]、今後はここでユーザから抽出した重要変数を用いてモデル構築を実施する予定である。
- 2) 本研究では、状況に関するデータはユーザ情報等とは独立に取得し、それらを組み合わせてモデルを構築した。現在、状況属性付き学習データの収集実験を行っており、今後はここで得られた定量データを利用してモデルの精度を向上させる予定である。

7. おわりに

本稿では、ベイジアンネットによるユーザの個人差と状況を考慮したユーザモデルの構築経過、および、同モデルに基づく推薦システムのプロトタイプ構築結果を報告した。最後に日頃ご指導頂く KDDI 研究所浅見代表取締役所長に深く感謝致します。

参考文献

- [1] <http://www.amazon.co.jp>
- [2] <http://www.tsutayaonline.co.jp>
- [3] P. Resnic and H.R. Varian, "Recommender systems," Communications of the ACM, 40, No.3, 56-58, 1997
- [4] M. Balabanovic and Y. Shoham, "Fab: Content-Based, Collaborative Recommendation," Communications of the ACM, 40, No.3, pp. 66-72, 1997
- [5] C. Basu, H. Hirsh, and W. W. Cohen, "Recommendation as classification: Using social and content-based information in recommendation," Proc. 15th National Conference on Artificial Intelligence, pages 714--720, Madison, WI, July 1998
- [6] Hofmann, T., Probabilistic latent semantic analysis, In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, pp. 289-296, 1999
- [7] Hofmann, T. and Puzicha, J.: Latent class models for collaborative filtering, In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, pages .688-693, 1999
- [8] A. Popescul, L. H. Ungar, D. M. Pennock, and S. Lawrence, "Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments," Proc. 17th Conference on Uncertainty in Artificial Intelligence, 2001
- [9] 小野, 本村, 麻生, "確率モデルによる映画推薦方式の検討", 人工知能学会研究会資料, SIG-FPAI-A502-02, pp7-12, 2005
- [10] 本村陽一, "ベイジアンネットワーク", 学習システムの理論と実現(渡辺澄夫著), 4 章, 森北出版, 2005.
- [11] E. Horvitz, "Principles of Mixed-Initiative User Interfaces," Proc. ACM SIGCHI Conference on Human Factors in Computing Systems, 1999.
- [12] F.V. Jensen et.al., "The SACSO methodology for troubleshooting complex systems," Artificial Intelligence for Engineering Design, Analysis and Manufacturing(AIEDAM), Vol. 15, pp. 321-333, 2001.
- [13] 小野, 本村, 麻生, "ベイジアンネットによる映画コンテンツ推薦方式の検討," 信学技法, NC2004-66, pp.55-60, 2004
- [14] Dempster, A.P., Laird, N. M., and Rubin, D. B., Masimum likelihood from imcomplete data via the EM algorithm. Journal of Royal Statistic Society. B, 39: pages 1-38, 1977.
- [15] 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
- [16] J. Suzuki, "A construction of Bayesian networks from databased based on an MDL principle," Proc. the 9th Conference on Uncertainty in Artificial Intelligence, pp.266-273, 1993
- [17] 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, pp.131-132, 2005