

# ベイジアンネットワークによるセグメント説明モデルと映画推薦への応用

吉田 真<sup>†, a</sup> 本村 陽一<sup>†, b</sup>

<sup>†</sup>東京工業大学 <sup>††</sup>産業技術総合研究所

a) *yoshida.m.ba@m.titech.ac.jp* b) *y.motomura@aist.go.jp*

**概要** 協調フィルタリングの特徴は、年齢や性別に関係なく嗜好が似ている人を参考にして、ユーザが最も欲するであろうアイテムを推薦できることである。しかし、嗜好が似ている人どうしで推薦しあうので、推薦できるアイテムに新規性や意外性がなくなる問題がある。人それぞれ好みは異なるが、ある一定の年齢層や性別に商品が流行するといった現象もあり、年齢や性別といったユーザの属性を推薦に応用することで推薦できるアイテムの幅が広がる。本研究ではベイジアンネットワークによる pLSA のセグメント説明モデルを構築し、映画の推薦へ応用する方法を提案する。ベイジアンネットワークを介すことで、ユーザの属性を推薦結果に反映することができ、従来では推薦できなかったアイテムを推薦することができるようになる。

**キーワード** 推薦システム, 確率的潜在意味分析, ベイジアンネットワーク

## 1 はじめに

Amazon.com など、推薦システムの実現手段として協調フィルタリングが多く利用されている。協調フィルタリングの実装方法としては、最近傍推薦 [1]、アイテムベース最近傍推薦 [2] などが提案されている。また、次元圧縮の技術を応用した推薦 [3][5] も提案されている。これら協調フィルタリングの特徴は、年齢や性別などといったユーザが持つ属性を利用することなく、過去のユーザの履歴さえあれば推薦が可能である点である。そのため、その人の年齢や性別に問わず、嗜好の似ている人を参考にして、ユーザが最も欲するであろうアイテムを選び出すことが可能である。その一方で、嗜好が似ている人どうしで推薦しあうことになるので、推薦できるアイテムに新規性や意外性がない。時間の経過とともに、推薦システムの出力に価値がなくなってしまう。

一般的に、同じ年齢や性別、同じ趣味を持っていたとしても、人それぞれに好みは異なっている。そのため、年齢や性別が同じだからといって同じアイテムを推薦することは望ましいとはいえない。しかし、女子中高生を中心に特定の商品やサービスが大流行することがニュースにとりあげられることがあるように、同じ属性を持つ人たちどうしの間で一部のアイテムを求めることもある。そのために、年齢や性別、趣味といったユーザの属性を利用することは、推薦できるアイテムの幅を広げることにつながる。

本研究では、pLSA による協調フィルタリングをもとに、ベイジアンネットワークを使って、pLSA によって作られるユーザのセグメントを彼らの属性により説明するモデルを構築する。ユーザの属性をベイジアンネットワークに入力することで、pLSA のどのセグメントにユーザが属しそうかを予め計算し、その結果に基づき映画の推薦を行うシステムを提案する。このシステムでは

従来の単純な協調フィルタリングでは得にくい推薦結果を得ることができる。加えて、新規ユーザであっても、ユーザの属性さえ入力すれば、推薦結果を得ることができるため、コールドスタート問題も回避することができる。

## 2 pLSA による推薦

Probabilistic Latent Semantic Analysis (pLSA)[4] は文書と単語の関係をいくつかの潜在変数を使って確率的に表現する分析手法である。pLSA では文書と単語を Aspect モデルで表現して、それぞれのパラメータを EM アルゴリズムを使って学習させる。潜在変数を 1 つの話題として考えることができ、文書集合をセグメント化するために利用することができる。

Hoffman は、pLSA を推薦技術へ応用している [4]。ユーザ  $u$  がアイテム  $y$  を取得する確率を式 (1)。

$$p(u, y) = \sum_z p(y|z)p(z|u)p(u) \quad (1)$$

あるユーザ  $u$  を定めたとき、確率  $p(y|u)$  が高いアイテムをユーザ  $u$  に推薦する。

$$p(y|u) = \sum_z p(y|z)p(z|u) \quad (2)$$

## 3 ベイジアンネットワークによる潜在クラスの説明モデル

pLSA による推薦を拡張して、各潜在変数に対してベイジアンネットワークによるユーザの属性を使った説明モデルを構築する方法を提案する。

ベイジアンネットワークは、確率変数間の確率的依存関係をグラフ構造によって表現したものである。また、それぞれの確率変数に対して、観測された値 (エビデン

ス)を与えることで、未観測の変数の確率分布を計算することができる。

ベイジアンネットワークを使って pLSA による推薦を拡張するために、まず、pLSA によって得られた結果をもとにユーザのセグメントを作成する。次に、そのセグメントに所属するユーザが回答したアンケート結果を使い、ユーザセグメントを説明するベイジアンネットワークを作成する。最後に、作成したベイジアンネットワークを使って、映画を推薦する手法を示す。

### 3.1 ユーザのセグメント化

はじめに、潜在クラスにより分類されるユーザセグメントの説明モデルを構築するために、ユーザ  $u$  に対して最も生起する確率が高い潜在クラス  $z_m$  を割り当てる。

$$z_m = \arg \max_z p(u, z) = \arg \max_z p(u|z)p(z) \quad (3)$$

同じ潜在クラスが割り当てられたユーザどうして1つのセグメントを形成する。ユーザのセグメントは潜在クラスの数  $K$  だけ生成される。それぞれのセグメントについて、ユーザがそのセグメントに所属するかを推定するためのベイジアンネットワークを作成する。

### 3.2 ユーザセグメントを説明するベイジアンネットワークの構築

生成したユーザのセグメントを説明するためのベイジアンネットワークを構築する。構築するベイジアンネットワークは、ユーザセグメントへの所属の有無、映画に関する嗜好、年齢、性別から構築される。ベイジアンネットワークを利用することで、ユーザセグメントをその構造を用いて分かりやすく説明できる。さらに、エビデンスを与えることで、新規のユーザがどのセグメントに所属しそうかを予測することができる。

ベイジアンネットワークの構造を学習するために、BayoNet [6] を使用する。BayoNet はベイジアンネットワークの構造を学習し、推論を行うためのアプリケーションである。構造の学習には Greedy アルゴリズムを使用し、情報量基準 AIC を最小にするようにベイジアンネットワークを構築する。

### 3.3 推薦方法

新規ユーザへ映画を推薦するための手順を示す。

はじめに、新規ユーザ  $u_{new}$  に対して、データセットを生成したときと同様のアンケートをとる。次に、各セグメントについて、アンケートの回答をエビデンスとしてベイジアンネットワークに与え、ユーザセグメントへの所属確率を推定する。得られた  $K$  個の所属確率を正規化し、 $p(z|u_{new})$  と見なす。最後に、 $p(y|u_{new})$  を算出し、確率の高い映画を推薦する。

ベイジアンネットワークの全てのノードに対してエビデンスを与える必要はないため、必ずしも全てのアンケート項目に答える必要がない。そのため、ユーザは答えられる範囲でアンケートに答えればよいいため、映画を推薦するための負担は大きくない。

## 4 実験について

本研究では、産業技術総合研究所が2009年に実施した「映画コンテンツの視聴態度に関するインターネット調査」のデータを使用する。このアンケートは、6,483名のユーザの659本の映画に対する視聴の有無に関する視聴履歴と、ユーザの性別や年齢、映画鑑賞に関する嗜好などといったユーザの属性に関するデータから構成されている。映画鑑賞に関する嗜好では、「アクション映画であること」、「実話に基づいていること」、「子供と一緒にでも安心して見られること」などについて、どのくらい重視するかを「全く重視しない」から「非常に重視する」までの7段階で尋ねた38項目の質問や、「ショッピング」、「ゲーム」、「旅行、ドライブ」などの趣味があるかを尋ねた20項目の質問(ダミー変数)が含まれる。

実験ではこのデータセットを使って pLSA の結果の意味付けを行う。

## 5 まとめ

本研究ではベイジアンネットワークによる pLSA のセグメント説明モデルを構築し、映画の推薦へ応用する方法を提案した。従来の協調フィルタリングでは時間の経過とともに推薦できるアイテムに新規性や意外性がなくなるという問題があった。本推薦システムではユーザの年齢や性別、嗜好を利用することで、協調フィルタリングでは推薦できなかったアイテムを推薦することができる。加えて、新規のユーザに対しても、その年齢や性別などといった属性を入力することで推薦するアイテムを選び出すことができ、コールドスタート問題を回避することができる。

実験では、産業技術総合研究所が実施した映画に関するアンケートを用いて、本推薦システムの性能を評価する予定である。

### 参考文献

- [1] Herlocker, J. L. and Konstan, J. A., et al.: An Algorithmic Framework for Performing Collaborative Filtering, Proc. of ACM SIGIR Conference, pp. 230-237, 1999.
- [2] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J.: Item-based collaborative filtering recommendation algorithms, Proc. of the 10th International Conference on World Wide Web, pp. 285-295, 2001.
- [3] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J.: Application of dimensionality reduction in recommender systems - a case study, Proc. of the ACM WebKDD Workshop, p, 2000.
- [4] Hofmann, T.: Probabilistic Latent Semantic Analysis, Proc. of the 15th conference on Uncertainty in artificial intelligence, pp. 289-296, 1999
- [5] Hofmann, T. and Puzicha, J.: Latent semantic models for collaborative filtering, Proc. of the 16th International Joint Conference on Artificial Intelligence, pp. 688-693, 2004.
- [6] 本村陽一:ベイジアンネットワークソフトウェア BayoNet, 計測と制御, vol. 42, no. 8, pp. 693-694, 2003