

マーケティング分野における 協調フィルタリング手法に関する研究

B9EM1013 酒井洋輔

Contents

1. Motivation
2. Previous research
3. 提案モデルに関して
4. Reference

1. Motivation

- 2000年以降電子商取引数が急増すると共に、「推薦システム」が注目されることとなった。
 - 推薦システム...利用者にとって有用と思われる対象、情報、商品などを選び出し、それを利用者の目的に合わせた形で提示するシステム (神畠 2007)

- 推薦の手法は大きく2種類に分けられる

- 協調フィルタリング

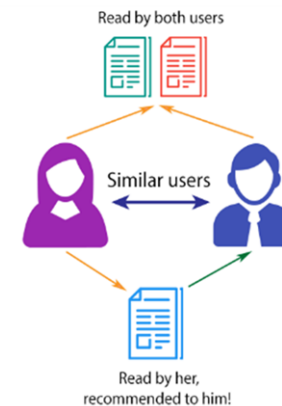
...推薦対象のユーザーaに近いユーザーbを選び出し、ユーザーbの評価値を用いて推薦を行う。

例) $\hat{r}_{aj} = \sum_b w_{ab} r_{bj}$

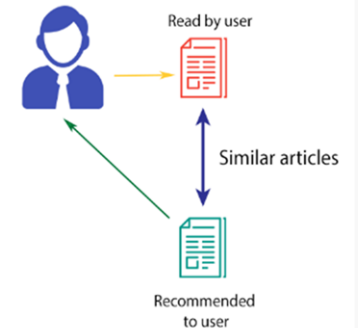
- 内容ベースフィルタリング

...商品の特徴ベクトルから、ユーザーaが高く評価した商品に近い商品を選び出し、推薦を行う。

Collaborative filtering



Contents based filtering



1. Motivation

- 「推薦」は情報処理の分野で盛んなトピック。
- Deep learning のような、解釈性には乏しいが、精度は高い手法を用いた研究が多い。
⇒「消費者の購買を促進する」というマーケティングの目的に適合しているのにも関わらず、
価格反応やブランド選択といったメジャーなトピックに比べて研究数が少ない。
- 本発表の目的...
解釈性を伴った推薦に関する研究を紹介、問題点や課題を明らかにするとともに、提案手法
のアイデアを紹介する。

2. Previous research

研究	手法	使用したデータ	概要
Ansari et al.(2000)	ベイズ線形回帰	評価行列、消費者の属性情報、商品のカテゴリー情報	属性情報、商品のカテゴリー情報毎の全体的な嗜好に加え、消費者毎、商品毎のrandom effectをモデルに取り入れた
Ying et al.(2006)	ベイズ線形回帰	評価行列、消費者の属性情報、商品のカテゴリー情報	評価行動を"select"と"evaluation"の二段階に分け、属性情報、商品のカテゴリー情報毎の、2段階それぞれへの影響を明らかにした
Jacobs et al.(2016)	Topic modeling	購買履歴データ、消費者の属性情報	購買履歴データから、潜在的な購買パターンを抽出し、推薦精度の向上を示した。

2. Previous research

1. Ansari et al.(2000) “Internet Recommendation Systems” Journal of Marketing Research

➤ ユーザー*i*による、商品*j*の評価（星の数） r_{ij} は、以下のモデルから生成されるとする。

➤ この時、 μ がfixed effect、 λ_i 、 γ_j がユーザー、商品ごとのrandom effectを表す。

➤ モデル： $r_{ij} = \text{constant} + \text{Genre}_j + \text{Demogra}_i + \text{Customer Hetero}_{ij} + \text{Movie Hetero}_{ij} + e_{ij}$

$$\text{Genre}_j = \sum_p \mu_p \text{Genre}_{jp}, \quad \text{Demogra}_i = \sum_{p'} \mu_{p'} \text{Demogra}_{ip},$$

$$\text{Customer Hetero}_{ij} = \lambda_{i1} + \sum_p \lambda_{ip} \text{Genre}_{jp}, \quad \text{Movie Hetero}_{ij} = \gamma_{j1} + \sum_{p'} \gamma_{jp'} \text{Demogra}_{ip},$$

➤ 事前分布： $\mu \sim \mathcal{N}(\eta, C^{-1})$, $\lambda_i \sim \mathcal{N}(0, \Lambda^{-1})$, $\gamma_j \sim \mathcal{N}(0, \Gamma^{-1})$

➤ 消費者毎、商品（ここでは映画）毎の異質性を考慮したことが新しい

2. Previous research

1. Ansari et al.(2000) “Internet Recommendation Systems” Journal of Marketing Research

パラメータの推定結果

<https://drive.google.com/file/d/1OrAL0c9nMEbcilZ1yMHkByBKUhGYT4fo/view?usp=sharing>

【解釈】

Fixed effect μ

- Movie genreに関するパラメータは、全て有意とはならなかった。
- Demographic に関しては、性別に関するパラメータは有意とはならなかったものの、年齢に関しては正に有意となった。→年齢が高いほど、高く評価する傾向がある。

Random effect Λ, Γ

- Movie genreに関する各消費者の嗜好のばらつきを表すパラメータ Λ は大きい値をとっている。
→消費者によって嗜好の違いが大きい
- Demographicに関する各映画の評価のばらつきを表すパラメータ Γ も大きい値をとった。
→映画の観測できない属性によって、属性毎の消費者による評価のされ方が異なる。

2. Previous research

2. Ying et al.(2006) “Leveraging missing ratings to improve online recommendation systems”
Journal of Marketing Research

➤ ユーザーによる商品の評価行動を、評価行動を”select”と”evaluation”の二段階に分け、それぞれ背後にある”潜在的な効用”によるものと仮定した。

➤ モデル： $U_{ijs} = \beta_{is}^G \text{Genre}_{js} + \beta_{js}^D \text{Demogra}_{is} + e_{ijs} \rightarrow$ ”選択”の効用
 $U_{ijp} = \beta_{ip}^G \text{Genre}_{jp} + \beta_{jp}^D \text{Demogra}_{ip} + e_{ijp} \rightarrow$ ”評価”の効用
 $Pr(Y_s = 1) = Pr(0 < U_s)$
 $Pr(Y_p = k) = Pr(K_{k-1} < U_p < K_k)$

➤ 事前分布： $\beta_{is}^G \sim \mathcal{MVN}(\mu_s^G, \Delta_s^G)$, $\beta_{js}^D \sim \mathcal{MVN}(\mu_s^D, \Delta_s^D)$, $\beta_{ip}^G \sim \mathcal{MVN}(\mu_p^G, \Delta_p^G)$, $\beta_{jp}^D \sim \mathcal{MVN}(\mu_p^D, \Delta_p^D)$
 $\log(\Delta K_i) \sim \mathcal{MVN}(\mu_c, \Delta_c)$

➤ 閾値 K_i も一緒に推定できる。 $\{\mu_s^G, \mu_s^D, \mu_p^G, \mu_p^D\}$ がfixed effect、 $\{\Delta_s^G, \Delta_s^D, \Delta_p^G, \Delta_p^D\}$ がrandom effectを表す。

2. Previous research

2. Ying et al.(2006) “Leveraging missing ratings to improve online recommendation systems”
Journal of Marketing Research

パラメータの推定結果

<https://drive.google.com/file/d/1HLvjvaY7IOIAg-F1D7km3ylb1cwWqQQI/view?usp=sharing>

【解釈】

Fixed effect μ

- Movie genreに関するパラメータは、“selection”、“evaluation”ともに有意な結果が多かった。
例) classic → “selection”は負に有意 “evaluation”は正に有意 → コアな古典映画ファンの存在の可能性
- Gender、Ageに関するパラメータは、“selection”が負、“evaluation”が正に有意だった。男性の方が映画を選びにくく、高く評価しやすい？年齢が高いほど映画を選びにくく、高く評価しやすい？

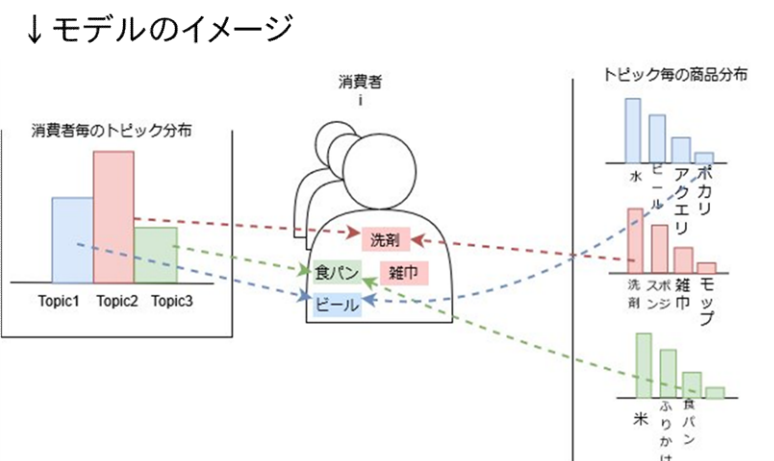
Random effect Δ

- Movie genreに関する各消費者の嗜好のばらつきを表すパラメータ Δ は、“horror”, “art/foreign”, “drama”など、なんだか好みに分かれそうなgenreが大きく、“comedy”, “animation”は小さくなっている

2. Previous research

3. Jacobs et al.(2016) “Model-based Purchase Predictions for Large Assortments” Marketing Science

➤ 購買履歴データにトピックモデルを用いて潜在的な購買パターンを抽出し、推薦精度の向上を示した。



↓結果

LDA-X M=15	Motivation 1 (Probability 0.21)		Motivation 2 (Probability 0.13)	
	Product	%	Product	%
1	Diapers – Pampers	20.11	Cleaning – Glorix	5.79
2	Baby/toddler nutrition – Nutrilon	19.26	Paper towels – Page	5.37
3	Baby/toddler nutrition – Olvarit	16.04	Dishwashing – Dreft	3.78
4	Baby/toddler nutrition – Bambix	10.13	Laundry – Robijn	3.54
5	Baby care – Zwitsal	7.94	Cleaning – Ajax	3.50
6	Baby care – Pampers	4.10	Laundry – Ariel	3.27
7	Pacifiers – Bibi	2.13	Disposables – Komo	3.08
8	Bottle appliances – Philips AVENT	2.05	Paper towels – Edet	3.03
9	Diapers – Huggies	1.70	Cleaning – Sorbo	2.97
10	Bottle appliances – Nuby	1.25	Cleaning – Cif	2.29

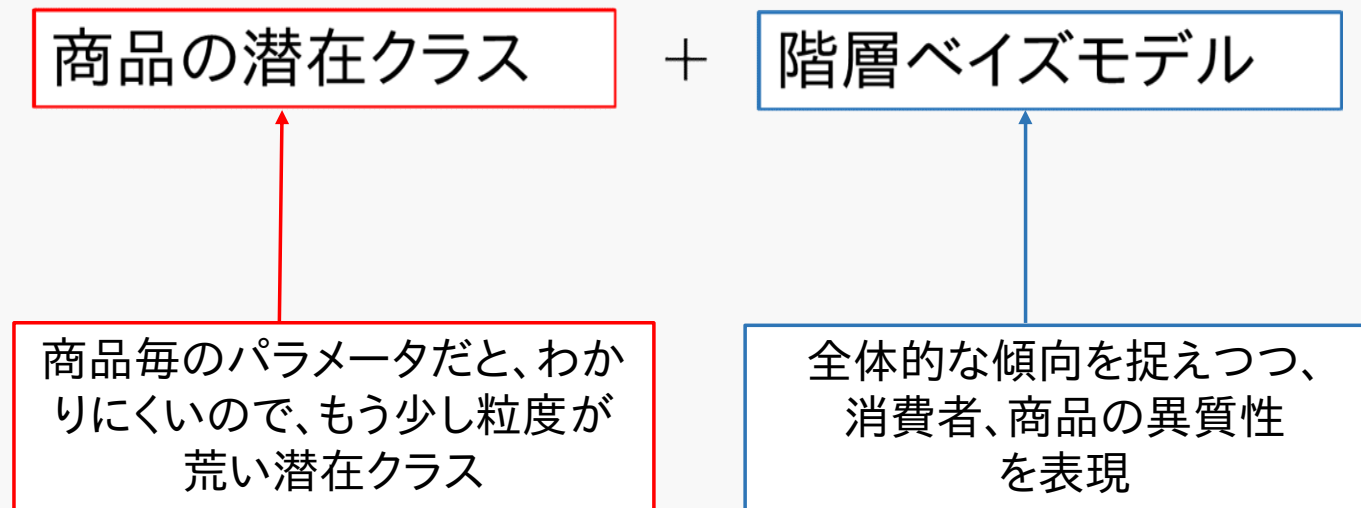
3. 提案モデルに関して

自身の研究に取り入れたい、あるいは明らかにしたい点

研究	取り入れたい点	改善したい点
Ansari et al.(2000)	消費者毎、商品毎の異質性をモデルに取り入れた ⇒ <u>階層ベイズモデル</u>	映画の観測できない属性によって、属性毎の消費者による評価のされ方が異なる ⇒ <u>どのような映画が、どの属性の消費者にウケるのか</u>
Ying et al.(2006)	・“selection”を評価行動の中に考慮する必要性 ・ <u>階層ベイズモデル</u>	<u>どのような映画が、どの属性の消費者にウケるのか</u>
Jacobs et al.(2016)	トピックモデルを用いて、商品の潜在的な選択パターンを明らかにする。	潜在的な選択パターンの抽出を、Ying et al.(2006)の“selection”のパートに適応できないだろうか？

3. 提案モデルに関して

➤ 提案モデルのイメージ



- 潜在クラス+階層ベイズモデルで検索をかけると、石垣先生の研究がヒットする
http://www.ifs.tohoku.ac.jp/cmd/reference/LS2_PDF/DI2-Ishigaki.pdf

3. 提案モデルに関して

石垣先生のマーケティング特論資料より http://www.ifs.tohoku.ac.jp/cmd/reference/LS2_PDF/DI2-Ishigaki.pdf

マーケティングと時代の要請

・消費価値観の変化に準じてマーケティングも進化

市場とマーケティングの進化

画一的価値観の市場

- ・ 十人一色の時代. メーカー主導の市場形成
⇒ マスマーケティング

$$\text{効用: } U_{itj} = f(\underline{a}, x_{itj}) + \varepsilon_{itj}$$

価値観・ライフスタイルの分化した市場

- ・ 少数セグメント毎へのアプローチ
⇒ セグメンテーション・マーケティング

$$\text{効用: } U_{itj} = f(\underline{a}_s, x_{itj}) + \varepsilon_{itj}$$

価値観・ライフスタイルの多様化した市場

- ・ 十人十色の時代. 個人毎へのアプローチ
⇒ one to oneマーケティング

$$\text{効用: } U_{itj} = f(\underline{a}_i, x_{itj}) + \varepsilon_{itj}$$

個人の時間異質性、文脈異質性

- ・ 一人十色の時代のマーケティング

$$\text{効用: } U_{itj} = f(\underline{a}_{it}, x_{itj}) + \varepsilon_{itj}$$

14

時代の流れと逆行してた・・・
個人毎、商品毎のパラメータが
推定できる以上、潜在クラスを
仮定する必要性は無い

3. まとめ

- 階層ベイズモデルをrating dataに用いることによって、個人毎、商品毎のパラメータを推定しつつ、全体的な傾向をモデル化することができた。
- 潜在クラスは、データがスパースかつ、推定すべきパラメータ数が多い時にこそ真価を発揮する。
⇒ ・Ansari et al.(2000)とYing et al.(2006)の研究は、個人毎、商品毎のパラメータが推定できていると推測されるため、潜在クラスを導入する意義がない。

4. Reference

- ・神嶌敏弘 (2007), “推薦システムのアルゴリズム” *人工知能学会誌 Vol.22 No.6*
- ・ Asim Ansari, Skander Essegaier, Rajeev Kohli (2000), “Internet Recommendation Systems”
Journal of marketing research Vol.37 pp. 363-375
- ・Bruno Jacobs, Bas Donkers, Dennis Fok (2016), “Model-based Purchase Predictions for Large Assortments.” *Marketing Science Vol.35, pp.389-404*
- ・YUANPINGYING, FRED FEINBERG, and MICHEL WEDEL(2006),“Leveraging Missing Ratings to Improve Online Recommendation Systems” *Marketing Science*