

はじめに
準備
提案分析方法
実データ分析
まとめ

トピックモデルを用いたテレビ視聴における トレンド分析方法の提案

長瀬 永遠

富山県立大学 情報基盤工学講座

May 7, 2021

1.1 はじめに

2/16

はじめに
準備
提案分析方法
実データ分析
まとめ

本研究の背景

消費者の嗜好や余暇時間の使い方の多様化に伴い、テレビ局は視聴者の嗜好に寄り添った番組を編成する必要性が増した。視聴者の嗜好に寄り添った視聴傾向を把握するためには、全体の視聴傾向を観察するだけでなく、視聴者や番組を視聴傾向の類似性という観点でクラスタリングし、クラスタの特徴を分析する必要がある。

⇒しかし、視聴傾向から番組の関係性をまとめながら視聴者の嗜好を把握することを目的とした分析事例は乏しい。

本研究の目的

視聴者と番組を紐づけながら同時にクラスタリングすることが可能なPLSAに対して時間的変化の概念を導入することで、視聴者の嗜好の変化をとらえられる分析方法とその可視化手法を提案する。

2.1 準備

3/16

対象データ

- VR CUBIC データ（株式会社ビデオリサーチ）

視聴者が分単位の各時点でのチャンネルを視聴していたかが記録されている。本研究では、2017年4月からの1年間に放送されたドラマとその視聴者情報を使用。

クラスタリング手法

- PLSA (Probabilistic Latent Semantic Analysis)

トピックモデルの一種。共起関係にある両者を同時にクラスタリングする手法。

可視化手法

- サンキーダイアグラム

時間変化の中でのクラスタの成長や衰退、結合や分離などを視覚的に表現。

3.1 概要

4/16

提案手法における 3 ステップ

- ① 全期間の視聴履歴データに対して視聴傾向を PLSA で学習し、期間によらない視聴傾向全体をモデル化.
- ② 視聴履歴データを一定期間で分割し、ステップ 1 の結果を活用しながら、再度各期間の PLSA を学習。視聴者を期間ごとのクラスタに割り当てる.
- ③ 期間の間の視聴者の遷移数をインプットにサンキーダイアグラムによる可視化を行う.

2 種類の分析方法

- ① 各期間で視聴者をクラスタに割り当てる際に最も嗜好度の高いクラスタに所属させる（分析法 1）.
- ② 視聴者が複数のクラスタに対して嗜好度を持つことを考慮して潜在クラスタを新たに構成し、視聴者をそのクラスタに割り当てる（分析法 2）.

3.2.1 分析法 1

はじめに

準備

提案分析方法

実データ分析

まとめ

ある視聴者がある番組を視聴する事象の確率モデル

$$P(x_i, y_j) = \sum_{k=1}^K P(z_k) P(x_i|z_k) P(y_j|z_k) \quad (1)$$

I 個の番組 : $x_i (i = 1, \dots, I)$

J 人の視聴者 : $y_j (j = 1, \dots, J)$

K この潜在クラスタ : $z_k (k = 1, \dots, K)$

最尤度関数

$$LL = \sum_{i,j} n_{i,j} \log P(x_i, y_j) \quad (2)$$

$n_{i,j}$: 視聴者 y_j が番組 x_i を見た回数

3.2.2 分析法1

6/16

ステップ1

EM アルゴリズムの更新式 【E ステップ】

$$P(z_k|x_i, y_j) = \frac{P(x_i, y_j, z_k)}{\sum_k P(x_i, y_j, z_k)} \quad (3)$$

【M ステップ】

$$P(z_k) = \frac{\sum_{i,j} n_{i,j} P(z_k|x_i, y_j)}{\sum_k \sum_{i,j} n_{i,j} P(z_k|x_i, y_j)} \quad (4)$$

$$P(x_i|z_k) = \frac{\sum_j n_{i,j} P(z_k|x_i, y_j)}{\sum_{i,j} n_{i,j} P(z_k|x_i, y_j)} \quad (5)$$

$$P(y_j|z_k) = \frac{\sum_i n_{i,j} P(z_k|x_i, y_j)}{\sum_{i,j} n_{i,j} P(z_k|x_i, y_j)} \quad (6)$$

はじめに

準備

提案分析方法

実データ分析

まとめ

3.2.3 分析法1

はじめに

準備

提案分析方法

実データ分析

まとめ

ステップ2

ステップ1の結果を初期値とし、各期間 ($s_l (l = 1, \dots, L)$) に対してステップ1と同様の処理を行う。

期間 s_l において視聴者 y_j に割り当てるクラスタ $c_{j,l}$ を次式で決定する。

$$c_{j,l} = \begin{cases} z_{K+1} & \sum_i n_{i,j}^l = 0 \\ \arg \max_{z_k} P(z_k | y_j)^l & otherwise \end{cases} \quad (7)$$

ステップ3

ステップ2の結果をもとにサンキーダイアグラムを作成する。

3.3.1 分析法2

視聴者の嗜好度分布の出現確立

期間 s_l に視聴者 y_j がクラスタ z_k を選択した数を

$N_{j,k}^l = \sum_{i=0}^I P(z_k|x_i) n_{i,j}^l$ として計算し、全期間で視聴者 y_j がクラスタ z_k を選択した回数を $N_{j,k}^l = \sum_{l=0}^L N_{j,k}^l$ と定義する。

$N_j = (N_{j,1}, \dots, N_{j,k})$ とすれば、 N_j は全期間での視聴者 y_j の嗜好の分布となる。ここで、視聴者の嗜好度分布 N_j を新たな潜在クラス v_m ($m = 1, \dots, M$) でクラスタリングすることを考えると視聴者の嗜好度分布の出現確率 $P(N_j)$ は以下のように表せる。

$$\begin{aligned}
 P(N_j) &= \sum_{m=1}^M P(v_m) P(N_j|v_m) \\
 &= \sum_{m=1}^M P(v_m) \frac{\Gamma(\sum_k N_{j,k} + 1)}{\prod_k \Gamma(n_j, k + 1)} \prod_{k=1}^K P(z_k|v_m)^{N_{j,k}}
 \end{aligned} \tag{8}$$

3.3.2 分析法2

はじめに
準備
提案分析方法
実データ分析
まとめ

EM 法の更新式

【E ステップ】

$$P(v_m|N_j) = \frac{P(v_m)P(N_j|v_m)}{\sum_{m=1}^M P(v_m)P(N_j|v_m)} \quad (9)$$

【M ステップ】

$$P(v_m) = \frac{\sum_{j=1}^J P(v_m|N_j)}{\sum_{m=1}^M \sum_{j=1}^J P(v_m|N_j)} \quad (10)$$

$$P(z_k|v_m) = \frac{\sum_{j=1}^J P(v_m|N_j)N_{j,k}}{\sum_{m=1}^M \sum_{j=1}^J P(v_m|N_j)N_{j,k}} \quad (11)$$

4.1 分析条件

はじめに

準備

提案分析方法

実データ分析

まとめ

対象データと分析条件

- 全期間：2017 年 4 月 3 日～2018 年 3 月 31 日
- 視聴者：3247 人 ($= J$)
- 番組：197 番組 ($= I$)
- 分割期間：1 か月 ($L = 12$)
- 放送時間の 50 %以上を視聴した場合に「視聴」とカウント.

4.2 有効性検証

11/16

検証方法と結果

今回は、サンキーダイアグラムの視認性という観点から、エントロピーを用いて提案手法を定量的に評価する。エントロピーの式を以下に、結果を以下の表1にそれぞれ示す。

$$Entropy = \frac{1}{L-1} \sum_{l=1}^{L-1} \sum_{k=1}^{K+1} \frac{|Z_k^l|}{J} \sum_{k'=1}^{K+1} P(z_{k'}^{l+1} | z_k^l) \log P(z_{k'}^{l+1} | z_k^l) \quad (12)$$

Table 1: エントロピーを用いた評価

	entropy
従来法	2.15
分析法 1	1.83
分析法 2	1.72

4.3.1 分析法1の結果

はじめに
準備
提案分析方法
実データ分析
まとめ

Table 2: 番組特徴： $P(x_i|z_k)$ の上位 5 番組

z_1 大河ドラマ	z_2 一般（女性主演）	z_3 夜中	z_4 深夜・海外	z_5 一般（恋愛・ヒューマン）
おんな城主直虎 (Q1, Q2, Q3)	火曜ドラマ・あなたのこ とはそれほど (Q1)	木曜ドラマ・恋がヘタで も生きてます (Q1)	ドラマ 24・孤独のグルメ Season6 (Q1)	金曜ドラマ・リバース (Q1)
西郷どん (Q4)	週保護のカホコ (Q2)	木曜ドラマF・リピート・ 道を変える (Q4)	コードネームミラージュ (Q1, Q2)	日曜劇場・ごめん、愛し てる (Q2)
日曜劇場・陸王 (Q3)	火曜ドラマ・カンナさー ん！ (Q2)	木曜ドラマ・脳ススメホ が埋められた！ (Q2)	GRIMM/グリムシリーズ シ 1 (Q2, Q3)	愛してたって、秘密はあ る。 (Q2)
土曜ドラマ・植木等との ぼせもん (Q2)	異様は、取り扱い注意 (Q3)	木曜ドラマF・ブラック リベンジ (Q3)	ドラマ 24・下北沢ダイ ハード (Q2)	僕たちがやりました (Q2)
日曜劇場・小さな巨人 (Q1)	木曜劇場・人は見た目が 100 パーセント (Q1)	木曜ナイトドラマ・女因 セブン (Q1)	木ドラ 25・さばりーマン 甘太郎 (Q2)	anone (Q4)
z_6 テレビ朝日刑事	z_7 NHK	z_8 一般（男性主演）	z_9 一般（刑事）	z_{10} ヒット作
相棒 (Q3, Q4)	ドラマ 10・この声をきみ に (Q3)	ボク、運命の人です。 (Q1)	CRISIS 公安機動捜査隊 特捜班 (Q1)	コード・ブルー・ドクター ヘル緊急救命 (Q2)
木曜ミステリー・斜探研 の女 (Q3, Q4)	ドラマ 10・フバキ文具店・ 鎌倉代書屋物語 (Q1)	もみ消して冬・わが家の 問題なかったことに (Q4)	日曜劇場・小さな巨人 (Q1)	金曜ドラマ・コウノドリ (Q3)
木曜ドラマ・ドクターX・ 外科医 大門未知子 (Q3)	土曜ドラマ・4 号警備 (Q1)	先に生まれただけの僕 (Q3)	日曜劇場・99.9・刑事専 門弁護士 (Q4)	日曜劇場・陸王 (Q3)
刑事 7 人 (Q2)	大河ファンタジー・精霊 の守り人・最終章 (Q3)	ウチの夫は仕事ができな い (Q2)	木曜劇場・刑事ゆがみ (Q3)	木曜ドラマ・ドクターX・ 外科医 大門未知子 (Q3)
木曜ミステリー・遺留搜 査 (Q2)	ドラマ 10・ブランケット・ キャップ (Q2)	ブランケンシュタインの 恋 (Q1)	貴族探偵 (Q1)	木曜ドラマ・アンナチュ ル (Q4)

Table 3: 視聴者特徴

	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}	全データ
平均年齢	48.4	41.4	41.7	43.4	41.6	47.4	47.6	40.3	43.7	42.0	43.8
男性比率	0.58	0.40	0.53	0.65	0.44	0.52	0.56	0.47	0.50	0.48	0.51
未婚比率	0.28	0.25	0.35	0.39	0.26	0.27	0.35	0.31	0.26	0.24	0.29

4.3.2 分析法1の結果

13/16

はじめに
準備
提案分析方法
実データ分析
まとめ

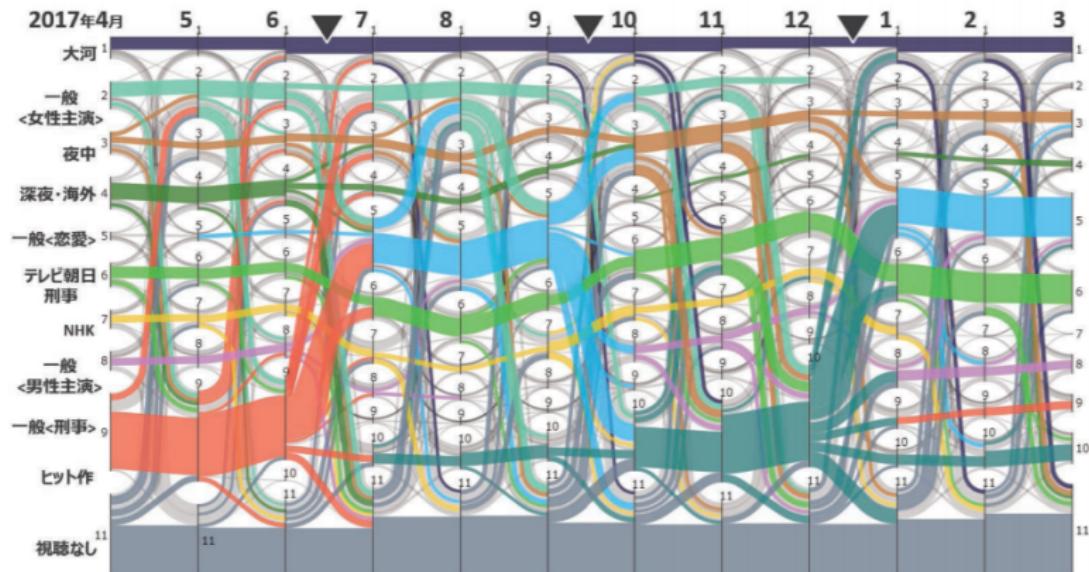


図 1: 分析法1によるトレンド可視化結果

4.4.1 分析法2の結果

はじめに
準備
提案分析方法
実データ分析
まとめ

$v_m \setminus Z_k$	大河	一般<女性主演>	夜中	深夜、海外	一般<恋愛>	テレビ朝日刑事	NHK	一般<男性主演>	一般<刑事>	ヒット作
クラスタ1	0.11	0.11	0.11	0.06	0.11	0.08	0.11	0.09	0.13	0.08
クラスタ2	0.03	0.10	0.08	0.02	0.12	0.22	0.03	0.09	0.19	0.12
クラスタ3	0.02	0.21	0.06	0.01	0.19	0.04	0.01	0.16	0.16	0.13
クラスタ4	0.24	0.12	0.03	0.02	0.12	0.07	0.05	0.09	0.14	0.12
クラスタ5	0.19	0.03	0.04	0.04	0.05	0.32	0.09	0.04	0.12	0.09
クラスタ6	0.03	0.07	0.23	0.21	0.08	0.10	0.05	0.07	0.10	0.06
クラスタ7	0.01	0.16	0.20	0.07	0.15	0.04	0.02	0.14	0.13	0.08
クラスタ8	0.02	0.02	0.06	0.04	0.05	0.51	0.03	0.03	0.15	0.10
クラスタ9	0.12	0.04	0.15	0.35	0.03	0.05	0.16	0.03	0.04	0.03
クラスタ10	0.54	0.02	0.02	0.03	0.03	0.04	0.22	0.02	0.04	0.04

⇒ 多様
 ⇒ 刑事もの
 ⇒ 一般系
 ⇒ 大河 & 一般系
 ⇒ テレビ朝日刑事 & 大河
 ⇒ 深夜系
 ⇒ 夜中 & 一般系
 ⇒ テレビ朝日刑事
 ⇒ 深夜系 & NHK系
 ⇒ 大河 & NHK

図 2: 嗜好度分布のクラスタリング結果

4.4.2 分析法2の結果

15/16

はじめに
準備
提案分析方法
実データ分析
まとめ

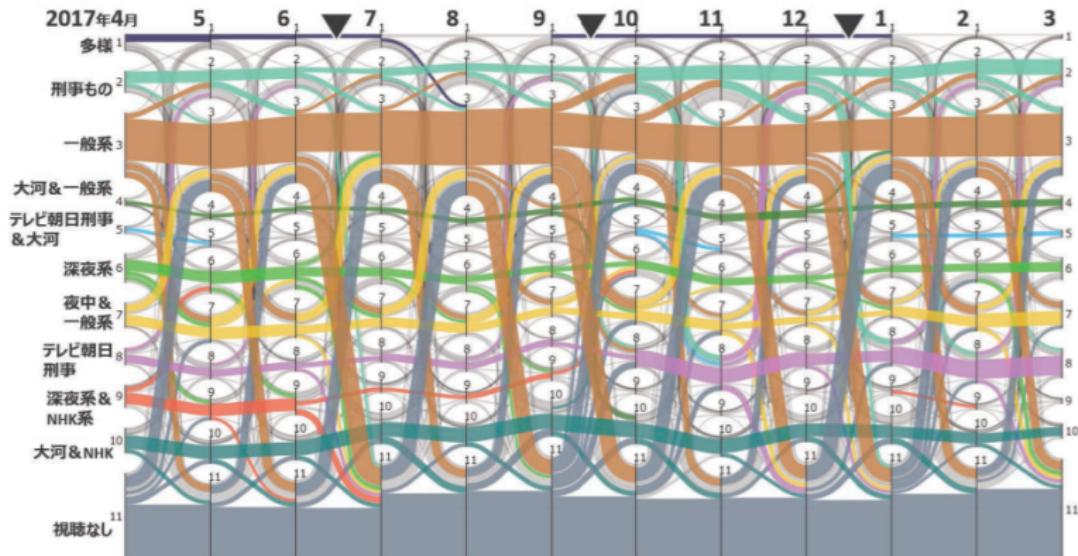


図 3: 分析法2によるトレンド可視化結果

はじめに
準備
提案分析方法
実データ分析
まとめ

実現したこと

- トピックモデルの一種である PLSA を用いることで番組と視聴者を同時にクラスタリングする手法の提案.
- 結果をサンキーダイアグラムにすることでテレビ番組におけるトレンドを可視化.

今後の課題

- 出演者等の分析を加える.
- サイコグラフィック情報等を加えた分析を行う.
- 提案手法に対する評価指標を定め、クラス多数の決定方法を考える.