

P14: Multivariate Hawkes process analysis of posts on a web service about uncomfortable gender experiences

郷田昌稔 (東京大学), 矢野良輔 (東京海上日動リスクコンサルティング), 水野貴之 (国立情報学研究所)

研究の背景

- 男女格差が社会問題となっている
- Web上で不満が伝搬している

研究の目的

ジェンダー問題に関する不満が投稿されるメカニズムをHawkes過程で解析



日本のジェンダー問題の構造を定量的に捉える

データの概要

- 不満調査データ*を使用 (時系列データ)
- 不満内容に「男/女」が含まれるものを抽出
- 性別年齢でグループ分けして分析

図1. グループ一覧

性別/年齢	0-19	20-29	30-39	40-49	50-59	60-
女性	G 1	G 2	G 3	G 4	G 5	G 6
男性	G 7	G 8	G 9	G 10	G 11	G 12

図2. 不満内容の例

不満内容	投稿時刻	グループ
自分には甘いのが相手のことになると相手を貶めるためだけに男尊女卑とか差別とか軽々しく使うのやめてほしい。コメンテーターレベルの議員は特に税金の無駄なんでやめてほしい。	2016/05/16 16:18:35	9
テレビでのオネエキャラへのいじりや流行り物のように男装女装を取り上げる限り日本のセクシャルマイノリティに対する理解は広まるどころか間違った方向に行くのではないかなと不安。服装、体、心、性対象を男性か女性かの二種類のどちらかに統一するものだという考えは間違えている。まずはメディアがきちんとした知識をつけて欲しい。	2016/09/28 12:09:26	2
...

*株式会社Insight Tech 不満買取センター <https://fumankaitori.com>

Hawkes過程

カーネルが指数型の12次元Hawkes過程

以下の強度過程を持つ点過程 $N_t = (N_t^1, \dots, N_t^{12})$:

$$\lambda_t^i = \mu_i + \sum_{j=1}^{12} \int_0^{t-} \alpha_{ij} e^{-\beta_{ij}(t-s)} dN_s^j, \quad i = 1, \dots, 12.$$

※ $\mu_i, \alpha_{ij}, \beta_{ij}$: パラメータ

※ $\rho_{ij} := \alpha_{ij}/\beta_{ij}$: カーネルの積分値

記号の解釈

- N_t^i : 時刻 t までのグループ i の累積投稿数
- λ_t^i : 時刻 t でのグループ i の投稿の起きやすさ
- μ_i : グループ i のPoisson的な投稿の起きやすさ
- ρ_{ij} : グループ j → i のトレンドの伝搬の大きさ

性質

- 各グループのトレンドの伝搬を記述可能
→ 地震, ニューロン, 感染症, etc.

推定量

□ QMLE :

- $\mu_i = 0, \alpha_{ij} = 0$
⇒ 適当な確率で $\hat{\mu}_i = 0, \hat{\alpha}_{ij} = 0$ と推定
- 非零のパラメータには漸近正規性が成立

□ LASSO :

- $\mu_i = 0, \alpha_{ij} = 0$
⇒ QMLEより正確に $\hat{\mu}_i = 0, \hat{\alpha}_{ij} = 0$ と推定
- 推定値にバイアスが発生.

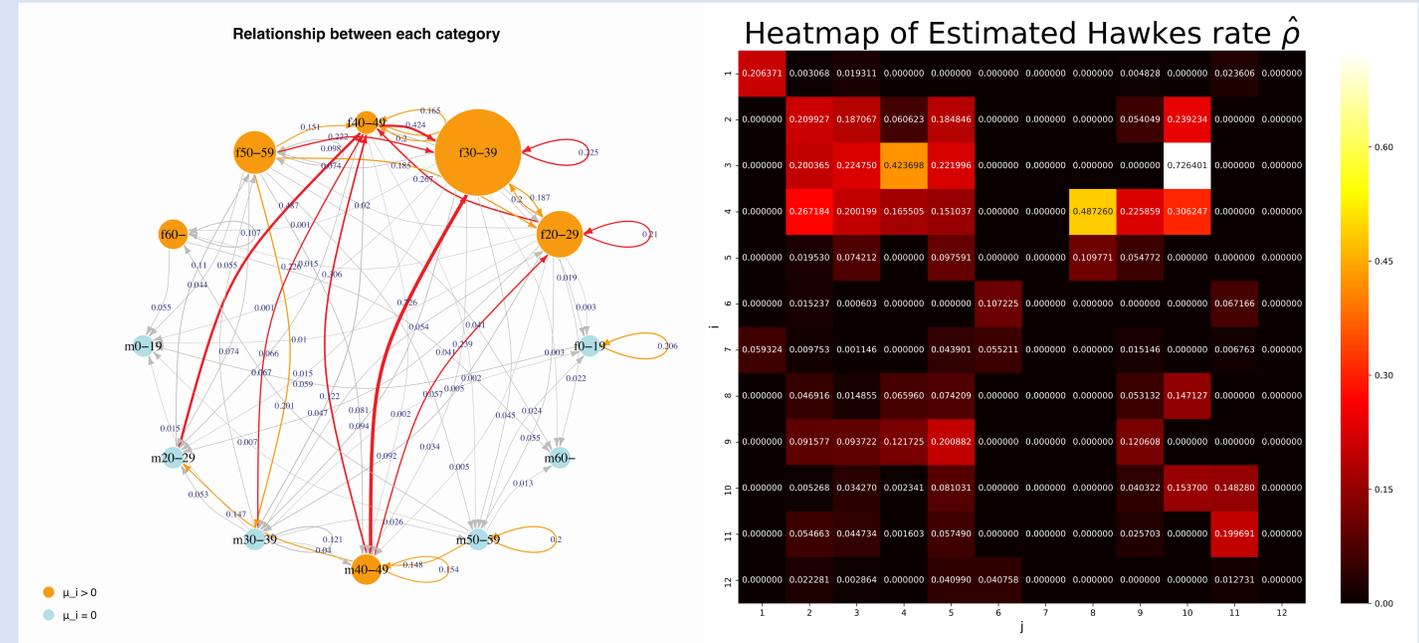


Step1 LASSO で 0 のパラメータを推定

Step2 正のパラメータを QMLE で再推定

解析結果

図4. 推定値 $\hat{\mu}_i, \hat{\rho}_{ij}$ によるHawkesグラフ・Hawkesヒートマップ



考察

- 20~30代女性のPoisson的な不満の発生が大
- 20~40代男性から30~40代女性への影響が大
- 40~50代男性は自身以外から受ける影響が小
- 男女共に10代の投稿数は少ない

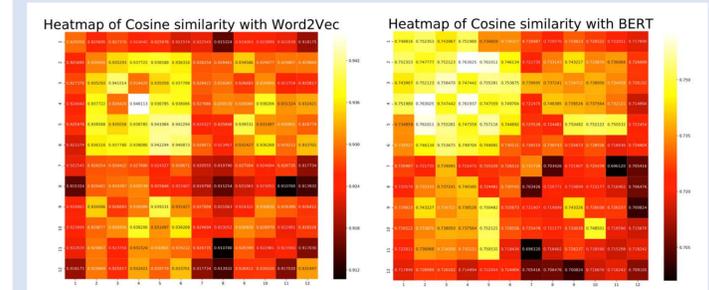


- 社会人女性の不満が顕著に多い
- 社会人男性→社会人女性の不満の伝搬がある
- 中高年男性は女性の不満への関心が低い
- 10代(学生)の不満は少ない

モデルの妥当性

- 相互作用があるグループ ⇒ 投稿内容に類似性
- Word2Vec, BERTにより投稿内容をベクトル化し, グループ間のコサイン類似度を計算

図5. コサイン類似度のヒートマップ



※左: Word2Vec, 右: BERT

図6. $\hat{\rho}_{ij}$ (図4) とコサイン類似度 (図5) の順位相関

	スピアマン順位相関係数	p 値
Word2Vec	0.3750	3.632×10^{-6}
BERT	0.4277	8.903×10^{-8}

→ 相関は有意 (トレンドが記述されている)