

メール型広告におけるタイトルが開封に与える影響

吉井 健敏^{1,a)} 城田 晃希¹ 市川 匠¹ 佐野 雄一¹ 持橋 大地²

概要: メール型広告において開封はクリック、コンバージョンにつながる最初のアクションであり、開封率の改善は収益や宣伝効果に重要な役割を果たす。特にメールのタイトルは、ユーザが広告内容を把握して開封するか判断するための数少ない手段である。本研究は株式会社 D2C が運用するメール広告配信サービス「メッセージ S(スペシャル)®」の大量の配信ログを用いて、広告クリエイティブの一部であるメールタイトルが開封に与える統計的な影響を明らかにする。配信ログ内のメールタイトルに形態素解析を行ってから統計的なフレーズ化を行い、(1) フレーズの有無とデモグラフィック情報から L_1 ロジスティック回帰モデルを用いて開封への影響を評価する手法 (2) 開封された広告フレーズと開封したユーザの特徴との自己相互情報量を計算し、性別年代ごとに開封に貢献したフレーズを明らかにする手法を提案した。

キーワード: L_1 ロジスティック回帰, 自己相互情報量, 電子メール

1. はじめに

「メッセージ S (スペシャル)®^{*1}<以下、メッセージ S>」は株式会社 D2C が広告販売・運用を行う、株式会社 NTT ドコモのスマートフォン向けメールサービスに配信されるメール型広告配信サービスである。約 3300 万人 [1] の登録者を持つキャリアが運営する、スマートフォン向けメール広告においては日本最大級のメディアである。メール型広告において開封はクリック、コンバージョンにつながる最初のアクションであり、開封率の改善は収益や宣伝効果に重要な役割を果たす。特にメールタイトルは、ユーザが開封するか判断するまでに広告内容を把握する数少ない手段である。本研究はメッセージ S の配信ログを用いて、広告クリエイティブの一部であるメールタイトルが開封に与える影響を明らかにするものである。デモグラフィック情報を使ったクリック予測の研究が盛んに行われているなかで、a) 開封に注目している、b) クリエイティブが与える影響を調査することで、デモグラフィック情報に依存せずに効果的なクリエイティブの制作に応用できる、c) 他に類を見ない大規模なデータセットを使用しているという三点において非常にユニークな研究となっている。

メールタイトルは広告主名や商品名などの固有名詞が多

く含まれるほか、同一ジャンルの商品にまたがって登場するユニークなフレーズが存在するため、Wikipedia や Web クローリング結果を用いて作成された Word2Vec はタイトルの持つ情報を表現するベクトル化にはそぐわない。そこで、本研究ではメールタイトル文を形態素解析し、形態素の出現頻度からフレーズ化を行った後、各フレーズを含むか含まないかの 2 値情報を元にベクトル化する手法を採用した。本研究では (1) デモグラフィック情報とフレーズの有無から L_1 ロジスティック回帰モデルを用いて開封予測モデルを作成し、メールタイトルが開封へ与える影響を評価する手法 (2) 開封された広告タイトルが含むフレーズと開封したユーザの特徴の共起性を自己相互情報量を用いて定義し、性別年代ごとに開封に貢献したフレーズを明らかにする手法の 2 つを提案する。

本稿ではまず、§2 で L_1 ロジスティック回帰、NPMI、NPMI を用いた教師なしフレーズ化について予備知識を導入する。§3 でメールタイトルのベクトル化、タイトルのスコアリング、NPMI を使ったフレーズの評価などの方法について提案。次に §4 で実験に使用したメッセージ S 配信ログについて、前処理の方法、学習時のパラメータ等について記載。最後に §5 でタイトルのスコアリング、フレーズの評価について実験結果を述べる。この結果よりメールタイトルが開封率に対して影響を与えていること、フレーズごとの性別・年代への開封への影響を可視化できることを示した。

¹ 株式会社 D2C
D2C Inc.

² 統計数理研究所
The Institute of Statistical Mathematics

a) taketoshi.yoshii@d2c.co.jp

*1 メッセージ S (スペシャル) は、株式会社 NTT ドコモの登録商標です。

2. 予備知識

2.1 L_1 ロジスティック回帰

ロジスティック回帰モデルは、ベルヌーイ分布に従う目的変数 $(y_i) = Y \in \{0, 1\}^N$ について説明変数 $(x_{ij}) = X \in \mathbb{R}^{N \times M}$ を用いて表現する統計モデルの一種であり、以下の式のように表すことができる。

$$\hat{y}_i = \sigma\left(\sum_j^M w_j \cdot x_{ij} + \beta\right) \quad (1)$$

ここで、 $\hat{y}_i (0 \leq \hat{y}_i \leq 1)$ は予測値で $(w_j) \equiv \mathbf{w} \in \mathbb{R}^M$ は説明変数ごとの重みを表すベクトル、 β は定数を表し、 $\sigma(x)$ はシグモイド関数で以下の式で定義される。

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \quad (2)$$

このとき、 \mathbf{w}, β は学習可能な変数で、以下の式から最適化される。

$$\mathbf{w}, \beta = \arg \min_{\mathbf{w} \in \mathbb{R}^M, \beta \in \mathbb{R}} \sum_i^N (y_i - \hat{y}_i)^2 \quad (3)$$

本研究ではメールタイトルを、特定のフレーズを含むか含まないかのベクトル形式で与えて説明変数として使用する。このとき考慮するフレーズの種類が膨大になることで特徴量の数 M が大きくなってしまい、過学習に陥ってしまう可能性がある。これを避けるため最適化の式 (3) に L_1 正則化項を導入する。

$$\mathbf{w}, \beta = \arg \min_{\mathbf{w} \in \mathbb{R}^M, \beta \in \mathbb{R}} \left(\sum_i^N (y_i - \hat{y}_i)^2 + \lambda \sum_j^M |w_j| \right) \quad (4)$$

ただし λ は正の実数。 L_1 正則化により寄与の小さい特徴量の重みを零化することができ、過学習を避けながら解釈性の向上も期待できる。

2.2 NPMI (Normalized Pointwise Mutual Information)

周辺確率 $p(x), p(y)$ で定義される離散確率変数 X, Y について自己相互情報量 (PMI; Pointwise Mutual Information) は以下の式で定義される。

$$\text{PMI}(x, y) \equiv \log \frac{p(x, y)}{p(x)p(y)} \quad (5)$$

ただし、 $p(x, y)$ は X, Y の同時確率。 $\text{PMI}(x, y)$ は x, y の共起性を表しており、2つが完全に独立なとき 0、共起しやすければ正、しづらければ負の値を取る。 $\text{PMI}(x, y)$ には x, y の出現頻度が非常に少ないとき値が大きく、同時確率 $p(x, y)$ が小さくなると無限に小さくなってしまふなどの問題点がある。それらの問題点を規格化により解決したものが Normalized PMI [3] であり、次の式で定義される。

$$\text{NPMI}(x, y) \equiv \log \frac{p(x, y)}{p(x)p(y)} \Big/ (-\log p(x, y)) \quad (6)$$

規格化により $\text{NPMI}(x, y)$ は -1 から 1 の間の値をとり、 x, y が独立なら 0、完全に共起しない場合は -1 、完全に共起する場合は 1 となる。

2.3 NPMI による教師なしフレーズ認識

Word2Vec [4] では”New York”や”Toronto Maple Leafs” (アイスホッケーチーム) のような、要素である ”new” や ”maple”, ”leaves” の組み合わせからは意味内容が異なるワードを ”new_york”, ”toronto_maple_leafs” などとフレーズと認識して Word2Vec を適用する手法が取られている。 Word2Vec では、単語 v と単語 w がフレーズとなるスコアを以下のように定義している。

$$\text{score}(v, w) = \frac{n(v, w) - \delta}{n(v)n(w)} \quad (7)$$

ただし $n(v, w), n(v), n(w)$ はバイグラムおよびユニグラムがコーパス中に出現する回数で、 δ は頻度の低いペアが上位に来るのを抑えるためのディスカウント係数。このスコアが閾値以上になると単語バイグラムを 1 単語としてまとめてフレーズとして認識することができる。このときスコアを確率で定義して考えると、コーパス中のユニグラムの総数を N として $p(v, w) = n(v, w)/N$, $p(v) = n(v)/N$, $p(w) = n(w)/N$ なので、 δ 部分を除くと、

$$\begin{aligned} \text{score}(v, w) &= \frac{n(v, w)}{n(v)n(w)} \\ &= \frac{p(v, w)}{p(v)p(w) \cdot N} \end{aligned} \quad (8)$$

となり、自己相互情報量 $\text{PMI}(v, w)$ (式 (5)) の形で書かれていることがわかる。^{*2} ただし、式 (8) は $1/N$ のファクターがあるため、コーパスに依存する量になってしまっている。そこで §2.2 で導入した Normalized PMI を導入して新たに以下のスコアを定義する。

$$\text{score}(v, w) \equiv \log \frac{p(v, w)}{p(v) \cdot p(w)} \Big/ (-\log p(v, w)) \quad (9)$$

この値がしきい値 t ($-1 \leq t \leq 1$) を超えたとき単語バイグラムを新たな一単語としてまとめる。得られた単語群を新たにフレーズ認識することで繰り返し回数 N_{iter} につき最大 $2^{N_{\text{iter}}}$ 単語からなるフレーズを取得することができる。

3. 提案手法

3.1 メールタイトルのベクトル化

メール型広告のタイトルは広告製品である性質上、広告主名や商品名などの固有名詞を多分に含むほか、同一ジャンルの商品にまたがって登場するようなユニークなフレー

^{*2} 以下の内容は、<http://chasen.org/~daiti-m/diary/?20210414> で議論されているものを基にしている。

ズが頻繁に出現する。本実験ではメールタイトル独特のニュアンスを損なわないようにするためタイトルに出現するフレーズを抽出し、そのフレーズを含むか含まないかの2値情報を使ってメールタイトルのベクトル化を行った。この操作によってメールタイトル s は以下のようにベクトル化できる。

$$\mathbf{x}^{\text{sub}} = (\delta(s, p|P)) \quad (10)$$

このとき $\{p\} = P$ はメールタイトル群から抽出したフレーズ集合で、 $\delta(s, p|P)$ はタイトル s にフレーズ p が与えられたとき 0 か 1 を返す関数である。1 となるフレーズ p はタイトル s を最大の被覆率かつ最小の分割数になるように選択する。具体例を上げて説明するとタイトル $s = \text{“美容液たっぷり”}$ 、フレーズの集合 $P = \{\text{“美容液”}, \text{“美容”}, \text{“容液”}, \text{“美”}, \text{“たっぷり”}\}$ が与えられたとき、 s は P のなかの“美容液”と“たっぷり”というフレーズで最大の被覆率で最小の分割数になる。タイトル s のベクトルは $\mathbf{x}^{\text{sub}} = {}^t(1, 0, 0, 0, 1)$ となる。本操作はタイトルベクトルの集合 (\mathbf{x}^{sub}) から多重共線性を排除するために行う。

3.2 タイトルのスコアリング

タイトルが開封率に与える影響を定量化する手法を提案する。フレーズの開封率への影響は開封の有無 ($y_i = Y \in \{0, 1\}^N$ (N はデータの行数) を目的変数とする L_1 ロジスティック回帰モデルの重みを用いて決定する。説明変数には、ユーザのデモグラフィック情報 ($x_{id}^{\text{demo}} \in \mathbb{R}^{N \times D}$ (D はデモグラフィック情報の数) とタイトルベクトル ($x_{iv}^{\text{sub}} \in \mathbb{R}^{N \times V}$ (V はフレーズの数) を使用すると、 L_1 ロジスティック回帰モデルは以下のように表される。

$$\hat{y}_i = \sigma \left(\sum_d w_d^{\text{demo}} \cdot x_{id}^{\text{demo}} + \sum_v w_v^{\text{sub}} \cdot x_{iv}^{\text{sub}} + \beta \right) \quad (11)$$

タイトルのスコアは、以下のように定義する。

$$(\text{score})_i = \sum_v w_v^{\text{sub}} \cdot x_{iv}^{\text{sub}} \quad (12)$$

複数のモデルを作成した場合には重みの平均 \bar{w}_v^{sub} と標準偏差 σ_v^{sub} を取ることができるので、スコアの標準偏差は誤差伝搬から以下のように計算できる。

$$\sigma_{\text{score}, i} = \sqrt{\sum_v (\sigma_v^{\text{sub}})^2 \cdot x_{iv}^{\text{sub}}} \quad (13)$$

スコアは開封に肯定的な影響を与えるフレーズを含むほど高く、否定的な影響を与えるフレーズを含むほど低くなる。

3.3 NPMI を使ったフレーズの評価

開封されたメールについてフレーズ p とユーザ特徴 f の

関係を定量的に評価するために以下のような指標を定義する。

$$\begin{aligned} \text{NPMI}(p, f) &= \log \frac{p(p, f)}{p(p) \cdot p(f)} / (-\log p(p, f)) \\ &= \log \frac{p(p|f)}{p(f)} / (-\log p(p, f)) \end{aligned} \quad (14)$$

ここで $p(f)$ は開封されたメールにユーザ特徴 f が出現する割合、 $p(f|p)$ はフレーズ p を含むタイトルを開封した際にユーザ特徴 f が出現する割合、 $p(p, f)$ は開封されたメールにユーザ特徴 f とフレーズ p が同時に出現する割合。NPMI(p, f) の値の正負によりフレーズ p がユーザ特徴 f に対して開封に肯定的あるいは否定的な影響を与えることを示すことができる。

4. 実験

4.1 データ準備

実験においてメッセージ S の配信ログを使用する。配信ログには配信日、ユーザのデモグラフィック情報 (性別、年齢、都道府県) と広告情報 (広告主名、タイトル文)、当日中に開封されたかどうかを表す符号が記載されている。また、本研究ではキャリアによるユーザインターフェースの違いを考慮し、Android スマートフォンユーザに絞って調査を行った。集計期間は 2021/9/1~2022/5/31 までとし、2021/9/1~2022/1/31 の 5 ヶ月間分を学習データとし、2022/2/1~2022/5/31 の 4 ヶ月分をテストデータとした (表 1)。

極端な開封率を示すユーザの除外 メッセージ S の配信においてユーザの中には開封率が著しく高いまたは低い者がいる。これらユーザは受信したメールの内容にかかわらず開封の意思を決定してしまいうため本実験の学習データ、テストデータからは除外する。除外するユーザは学習データの集計期間内で開封率が 80% 以上のユーザと 20% 以下のユーザを対象とした。

ダウンサンプリング メッセージ S の配信通数は 1 日あたりおよそ 1000 万件と膨大である。そこで学習にかかる時間、メモリ削減のため学習、テスト集計期間のログから 500 万件ずつランダムにサンプリングして小さくしたものを学習、テストデータとした。

4.1.1 前処理

デモグラフィック情報 年齢は標準偏差で標準化を行い、性別、都道府県は One-Hot Encoding を行い L_1 ロジスティック回帰モデルの説明変数とした。

タイトル メールタイトルは記号を除き数字を # で置換した後、IPA 辞書を使用して MeCab[2] で形態素への分解を行った。形態素の集合から助詞を除いたあとで、§2.3 で定義した $t = 0.5$, $N_{\text{iter}} = 1$ のパラメータでフレーズ化を行った。

表 1 実験で使ったデータの概要

	学習データ	テストデータ
集計期間	2021/9/1~ 2022/1/31	2022/2/1~ 2022/5/31
データサイズ (500 万件)	419MB	467MB
クライアント数	589 社	804 社 (488 社はテストデータのみ)
タイトル数	2176 件	1540 件 (860 件はテストデータのみ)

4.1.2 フレーズの選定

学習データ内において 1 つの広告主に対してのみ使用されるフレーズは過学習の原因となるため削除した。フレーズの選定により学習時のフレーズは 3023 個中 762 個となった。

4.2 学習

前処理を行ったデータに対して L_1 ロジスティック回帰モデル ($\lambda = 1$) の学習を行う。評価は 5 分割交差検証法により実施した。テストデータの予測は交差検証で得られたモデルの予測値の平均値を用いる。

4.3 評価指標

AUC L_1 ロジスティック回帰モデルの開封予測モデルの精度は AUC (Area Under ROC) で評価する。

相関係数 タイトルのスコアは高ければ高いほど開封率を向上することが期待されるので、スコアの値とタイトルごとの開封率の相関係数で評価する。

5. 結果

5.1 タイトルのスコアリング

学習データについてデモグラフィック情報、タイトル情報を使って 5 分割交差検証法で得られた精度は $AUC=0.5388$ であった。これはデモグラフィック情報のみを使って予測した場合に得られた $AUC=0.5143$ より高く、精度の向上が確認できた。このことからタイトル情報が開封に影響を与えることを示すことができた。

また、5 分割検証法で得られた 5 つのモデルから §3.2 で定義したタイトルのスコアを計算することができる。学習データとテストデータに登場するタイトルのうち、期間内に 100 通以上の配信されたものについてタイトルスコアと実際の開封率を計算しそれらの関係を表したものが (図 1) である。タイトルスコアが平均 $(score)_i$, 標準偏差 $\sigma_{score,i}$

表 2 §5.1 タイトルのスコアリング結果概要

	AUC	相関係数
学習データ	0.5388	0.431±0.005
テストデータ (全体)	0.5027	0.065±0.007
テストデータ (学習データに未出現タイトル)	0.5023	0.054±0.006

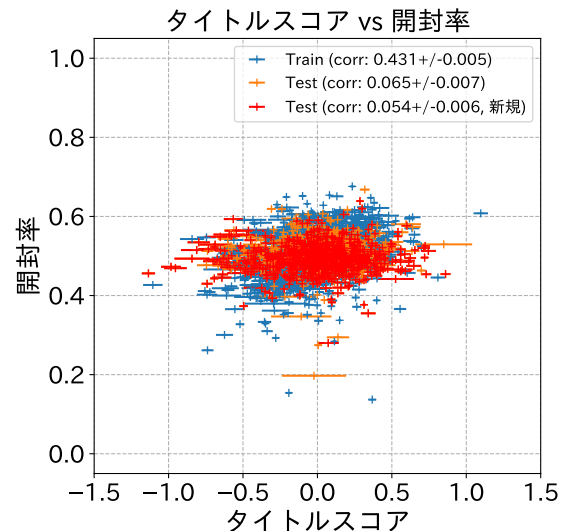


図 1 タイトルスコア (横軸) と開封率 (縦軸) の関係を表した散布図。青色が学習データ、橙色がテストデータ、赤色がテストデータのなかから学習データにない新規の広告についてプロットしたもの。学習データ、テストデータともに期間内に 100 通以上配信されたタイトルを選択。相関係数はサンプル数 20,000 回のモンテカルロシミュレーションで得られたものを表示している。

の正規分布に従うと仮定し、それぞれの学習、テストデータについて疑似タイトルスコアデータを 20,000 セット生成、疑似データごとに開封率との相関係数を計算し、それらの平均と標準偏差から母集団の分布を推定するモンテカルロ・シミュレーションを行った。これによると学習データ内で 0.431 ± 0.005 という高い正の相関を得ることができた。同モデルでテストデータについて評価を行うと 0.065 ± 0.007 という正の相関が得られた。特に、テストデータ内には学習データには現れないタイトルがあり、これらタイトルについても 0.054 ± 0.006 という正の相関があることがわかった。

これらの結果からタイトルが開封に影響を与えていて、タイトルスコアの値から開封率の当たりを事前につけることができることがわかった。

5.2 相互情報量を使ったタイトルの評価

§3.3 で定義した $NPMI(p, f)$ を学習データ内で出現するフレーズ p と性別・年代の組み合わせを特徴 f とすることで計算を行った。(表 3),(図 2) はフレーズごとの性別・年代への開封に与える影響を $NPMI$ を用いて表したものである。(表 3) は性別・年代ごとに開封に肯定的、否定的なフレーズの中から絶対値が大きい順に最大 8 個列挙したものである。また、(図 2) はフレーズがどの性別・年代の開封に影響を与えているかを棒グラフで表したもので横軸に年齢、縦軸に $NPMI$ をとり赤が女性、青が男性を示している。棒グラフは年齢の変化に対してなめらかに変化して、フレーズがどの性別、年代の開封に影響を与えているかを

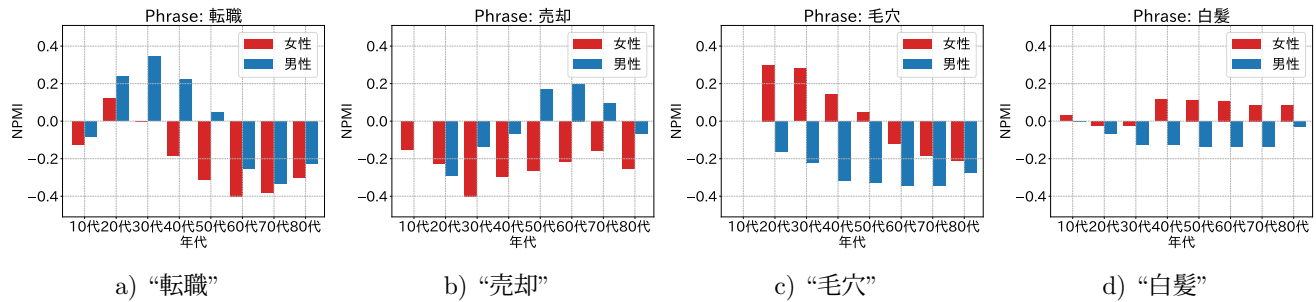


図 2 フレーズを含むメールタイトルがどの性別・年代に強く影響を与えているかを NPMI の棒グラフで表したもの。左から“白髪”、“毛穴”、“転職”、“売却”のフレーズを含むメールタイトルの配信ログからえられた結果を示している。それぞれ横軸に年代、縦軸に NPMI をとり赤が女性、青が男性を表している。正の値が大きいほど開封に肯定的な影響を、負の値が大きいほど否定的な影響を与えると解釈できる。

視覚的に理解することができる。§Appendix に (表 3 に登場するフレーズについて棒グラフでまとめたものの一覧を掲載している。

(表 3) より、若い男性には“仕事”や“転職”、“年収####”などの求人・転職関係のフレーズが、高齢の男性には“売る方法”や“価格”などの資産に関するフレーズが開封に肯定的な影響を与えていることがわかる。また、“毛穴”や“シミ対策”、“美容”など美容関連のフレーズは女性の若年層から高齢層まで幅広い層で開封に肯定的な影響を与えていることがわかる。次に、(図 2) の年代別影響度を確認する。(図 2 a) より“転職”というフレーズにおいて、20~40 代の男性では開封に肯定的な影響を、高齢層の男女では否定的な影響を与えていることがわかる。これは高齢世代は若年層より転職活動を行いにくいという一般的な見識と一致する。同様に“売却”というフレーズの開封への影響を表したのが(図 2 b)である。“売却”は不動産や車の査定の広告で使われているフレーズで、50~70 代男性の開封に貢献しているが、女性や若年層には否定的な影響を与えている。(図 2 c,d) は美容関連のフレーズである“毛穴”と“白髪”の開封への影響である。どちらも女性には肯定的な、男性には否定的な影響を与えている。興味深いのは“毛穴”が若年層により肯定的な影響を与えているのに対して、“白髪”は高齢層に肯定的な影響を与えていることである。これは女性が世代ごとに気になる美容のカテゴリが異なることを示しており、一般的な理解とも整合している。

6. まとめと今後の展望

本研究では株式会社 D2C の運営するメッセージ S の大量の配信ログを使って、メールタイトルが開封率に与える影響について異なる 2 つの手法を用いて明らかにした。

一つ目は、配信されたユーザーのデモグラフィック情報とメールタイトル情報を説明変数とする広告の開封を予測する L_1 ロジスティック回帰モデルを作成し、フレーズごとの重みからタイトルのスコアを定義した。そのスコアと実際の開封率の相関から影響を定量的に評価。結果よりメー

ルタイトルが開封に影響を与えることを明らかにした。また、タイトルのスコアを用いて事前に開封率の当たりを付けることができることを示した。タイトルスコアと開封率について正の相関を示すことができたが、因果関係は示せていない。今後、配信実験を通して因果関係の有無を明らかにしていく。

二つ目は、開封された配信ログに含まれるフレーズ p とユーザーの性別・年代の特徴 f の共起性を NPMI を用いて定義し、フレーズ p がどの性別・年代へ開封に影響を与えているかを定量的に明らかにした。結果は社会通念上の一般的な見解と合致しており、フレーズ単位でメールタイトルの広告効果を測定する新たなマーケティング指標としての活用が期待できる。

謝辞 本研究にあたって、ご討論頂いた統数研持橋先生、実験協力・論文校正を手伝っていただいた城田氏、市川氏、研究の機会を与えてくださった佐野氏、阿部氏、畔柳氏に感謝の意を表します。

参考文献

- [1] docomo business : 「メッセージ S」とは (online), 入手先 (<https://www.ntt.com/business/services/message-s>) (2022).
- [2] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, (<https://taku910.github.io/mecab/>)
- [3] Gerlof Bouma : Normalized (Pointwise) Mutual Information in Collocation Extraction, Proceedings of GSCL (2009).
- [4] Tomas Mikolov et al. : Distributed Representations of Words and Phrases and their Compositionality, CoRR (2013).

表 3 性別年代ごとに開封に肯定的 (NPMI > 0.2)、否定的 (NPMI < -0.2) な影響を与えたフレーズから一部抜粋してきたもの。
 “仕事” や “転職”、“年収###” などの求人・転職関係のフレーズが若い男性で、“売る方法” や “価格” などの資産に関するフレーズが高齢の男性で開封に貢献している事がわかる。一方で、“毛穴” や “シミ対策”、“美容” など美容関連のフレーズが女性の若年層から高齢層まで幅広い層で開封に肯定的な影響を与えていることがわかる。

	男性		女性	
	肯定的; NPMI>0.2	否定的; NPMI<-0.2	肯定的; NPMI>0.2	否定的; NPMI<-0.2
10代	“モデル”, “仕事”, “なり たい”, “間に合う電気”, “だった”	なし	“だった”, “手数料”, “楽 しもう”	なし
20代	“失敗し”, “失敗”, “させ”, “すべき”, “するべき”, “本 音”, “年収###”, “経験 者”	“コツ”, “韓国”, “ニュー ス”, “高”, “極上”, “ジェ ル”, “お届け”, “女性”	“黒ずみゴッソリ”, “ゴッ ソリ”, “洗顔”, “毛穴”, “失敗し”, “すべき”, “土 日祝”, “休み”	“感動”, “話題沸騰”, “ 糖質”, “クーポン付”, “ ニュースアプリ”, “お届 け”, “高く売る”, “価格”
30代	“年 収###”, “転 職”, “市”, “マイホーム”, “###万”, “土日祝”, “す べき”, “値”	“一変”, “さん絶賛”, “売 れ”, “募集中”, “爪”, “洗 顔”, “ダウンロード”, “朗 報”	“ゴッソリ”, “黒ずみゴッ ソリ”, “洗顔”, “ご案内”, “歳以上”, “毛穴”, “ママ”, “朗報”	“たい方”, “研究成分”, “生 活”, “爪”, “一”, “価格”, “楽しいプレゼント”, “中 性脂肪”
40代	“たい人”, “好条件求人”, “好条件”, “転職”, “転職 サイト”, “探し方”	“看護師”, “スキンケア”, “ホワイトニング”, “名”, “空気”, “モニター”, “## ~##代”, “毛穴洗浄”	“化粧品”, “香り”, “体験”, “シミ対策”, “ご案内”, “高 級”, “ダイエット”, “魅了”	“凄い”, “平均年収”, “年 収アップ”, “ニュースアプ リ”, “ダウンロード”, “締 切迫る”, “ニュース”, “て ない”
50代	“男性”, “医療保険”, “年 収アップ”	“クーポン”, “残り”, “ハリ ツヤ”, “時代”, “敏感肌”, “選ぶ”, “CM”, “特集”	“再入荷”, “放送後”, “今 月”	“機能”, “問アンケート”, “ゴッソリ”, “売る方法”, “知らない”, “好条件”, “不 動産”, “本音”
60代	“売る方法”, “価格”, “方 法”, “お家”, “たい方”, “###年”, “最高”, “ 方”	“学資保険”, “転職サイ ト”, “品薄中”, “選”, “モ ニター”, “残り”, “実感”, “超特価”	“卵殻膜”, “美容”, “先着”, “研究成分”, “美”, “東大”, “成分”, “超特価”	“駅近”, “わかる”, “査定”, “上昇中”, “急募月収”, “医 療保険”, “高く売る”, “埼 玉”
70代	“ニュースアプリ”, “安 心”, “ニュース”, “アプ リ”, “チェック”	“年収アップ”, “噂”, “特 別価格”, “取れ”, “学資保 険”, “人気商品”, “カタロ グプレゼント”, “成分”	“大ヒット”	“月収”, “働き”, “月入社”, “育”, “ママ”, “保険料”, “ご案内”, “最新版”
80代	なし	“濃厚”, “可能”, “男性”, “施工管理”, “癒し”, “## 代”, “誕生日”, “残り”	なし	“わかる”, “不動産”, “平均 年収”, “###P”, “ご 案内”, “台~”, “月々”, “シワ”

Appendix

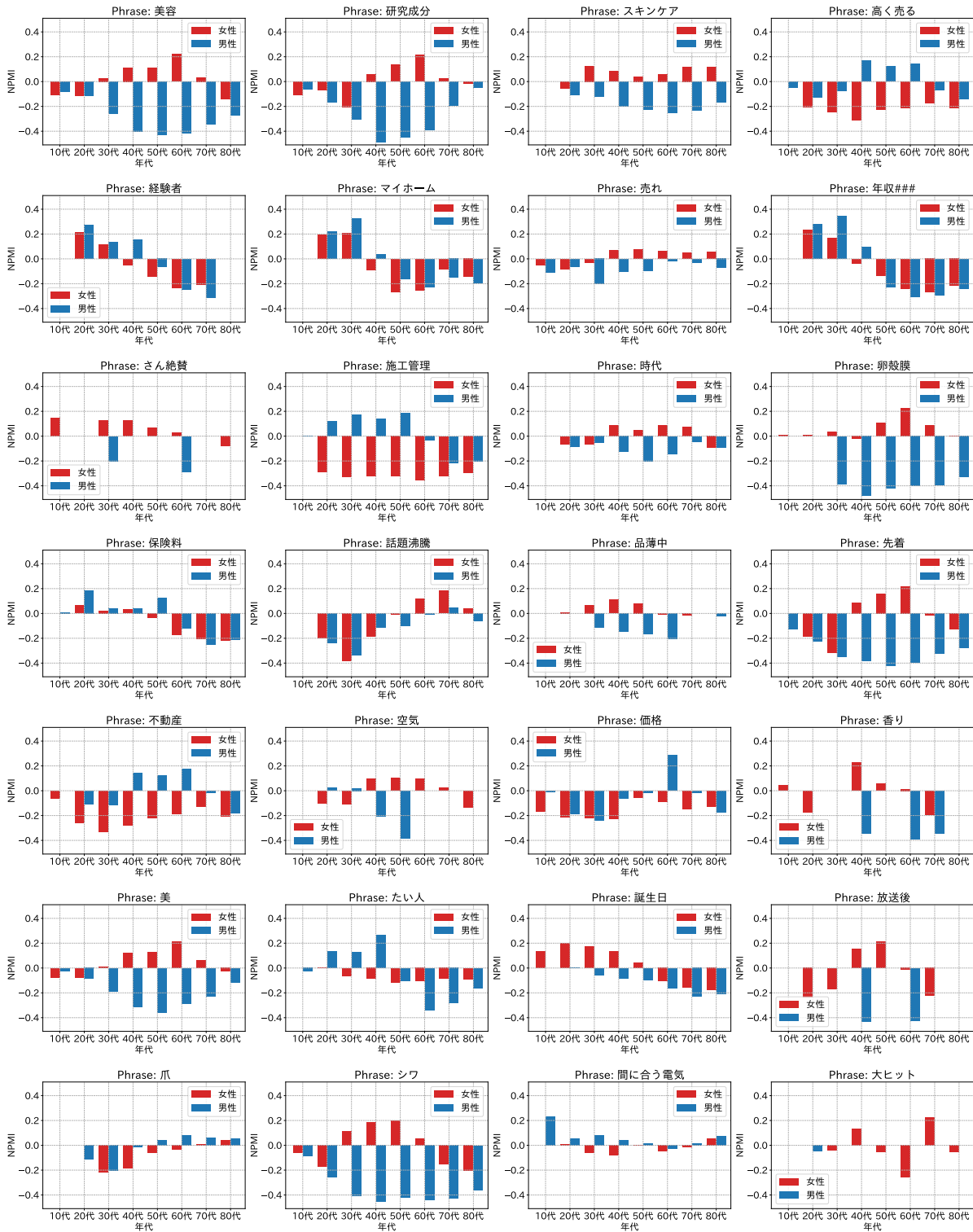


図 3 (表 3) に登場するフレーズについて、開封に与える影響を性別・年代ごとに棒グラフで表したもの (1/5)

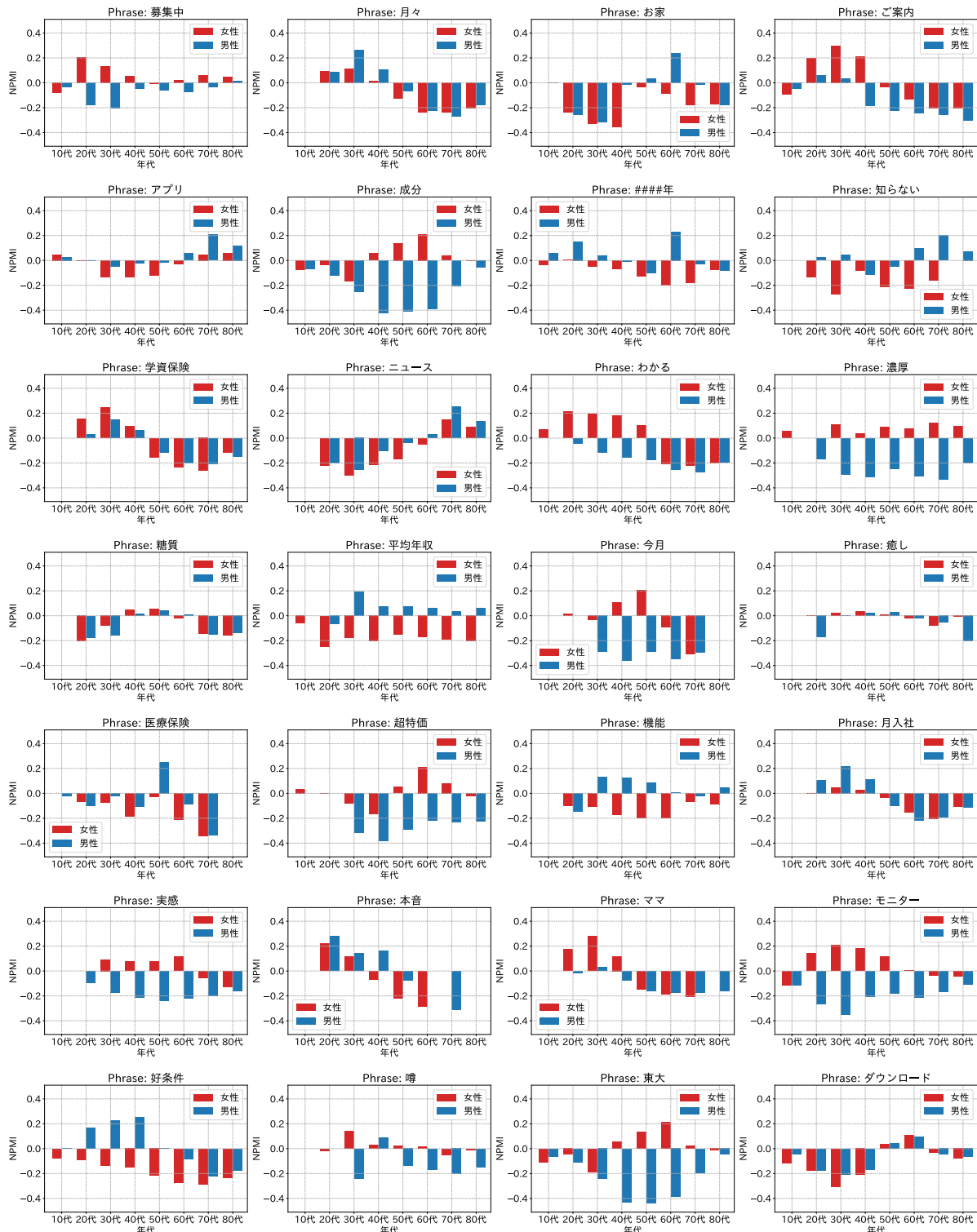


図 4 (表 3) に登場するフレーズについて、開封に与える影響を性別・年代ごとに棒グラフで表したもの (2/5)

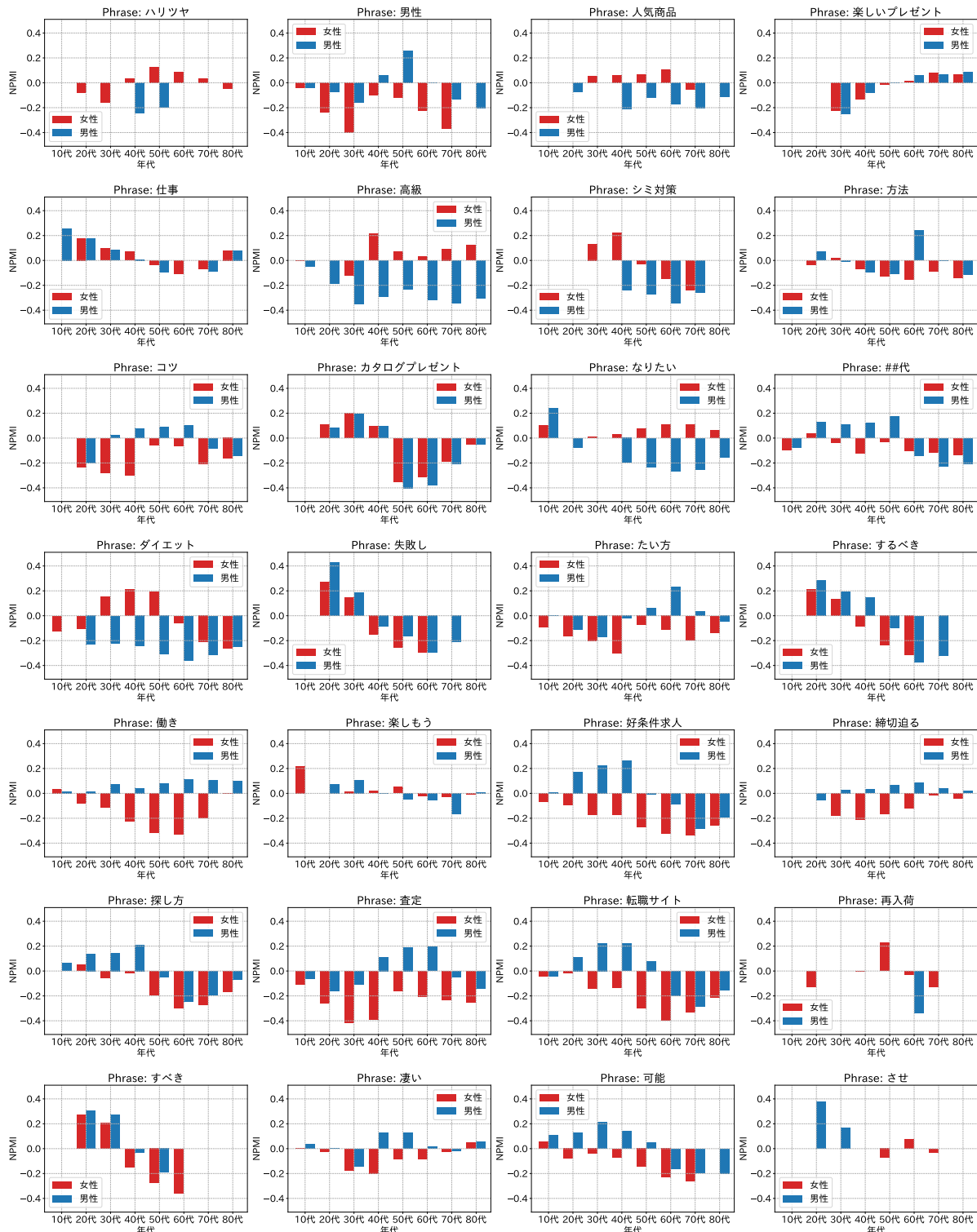


図 5 (表 3) に登場するフレーズについて、開封に与える影響を性別・年代ごとに棒グラフで表したもの (3/5)

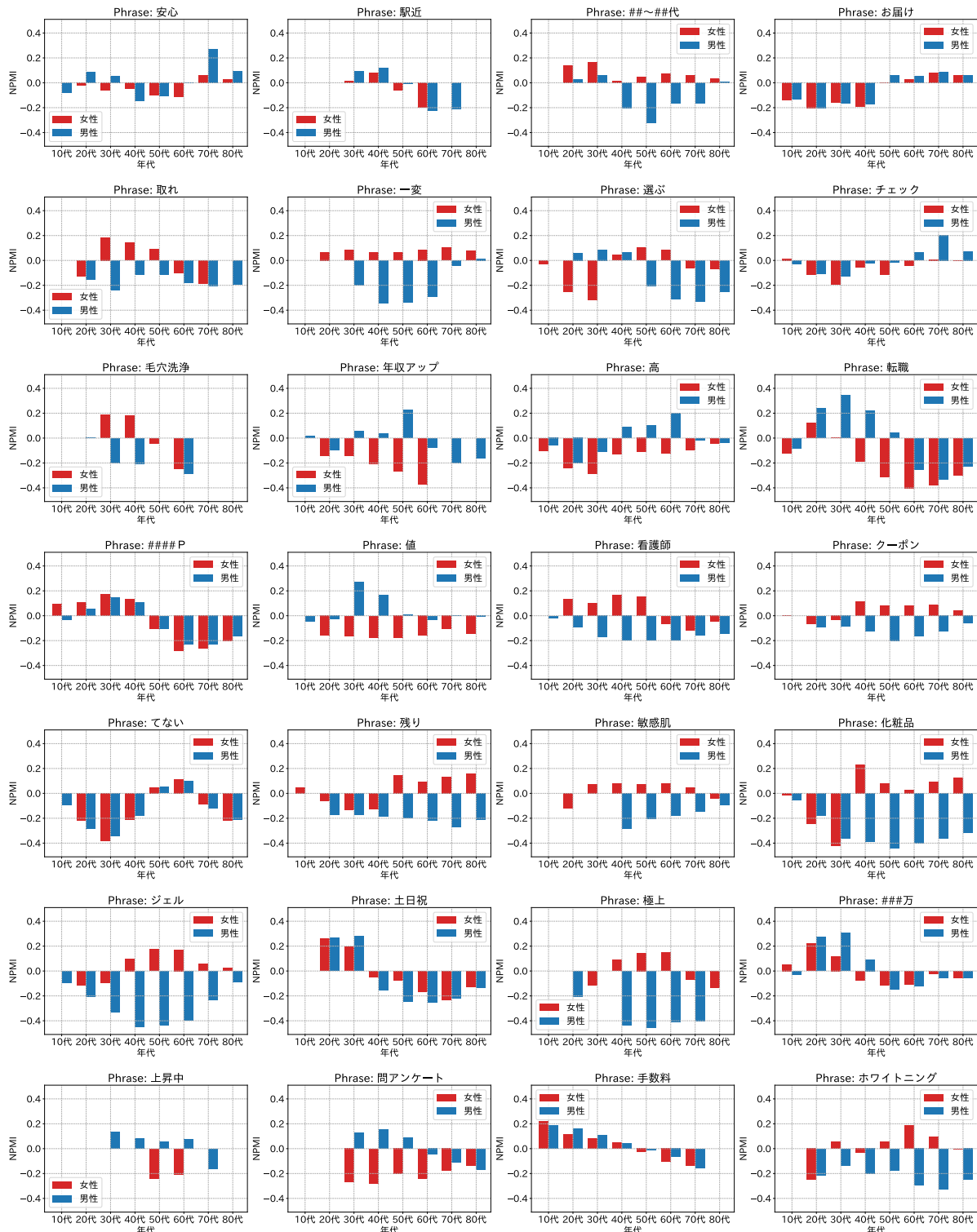


図 6 (表 3) に登場するフレーズについて、開封に与える影響を性別・年代ごとに棒グラフで表したもの (4/5)

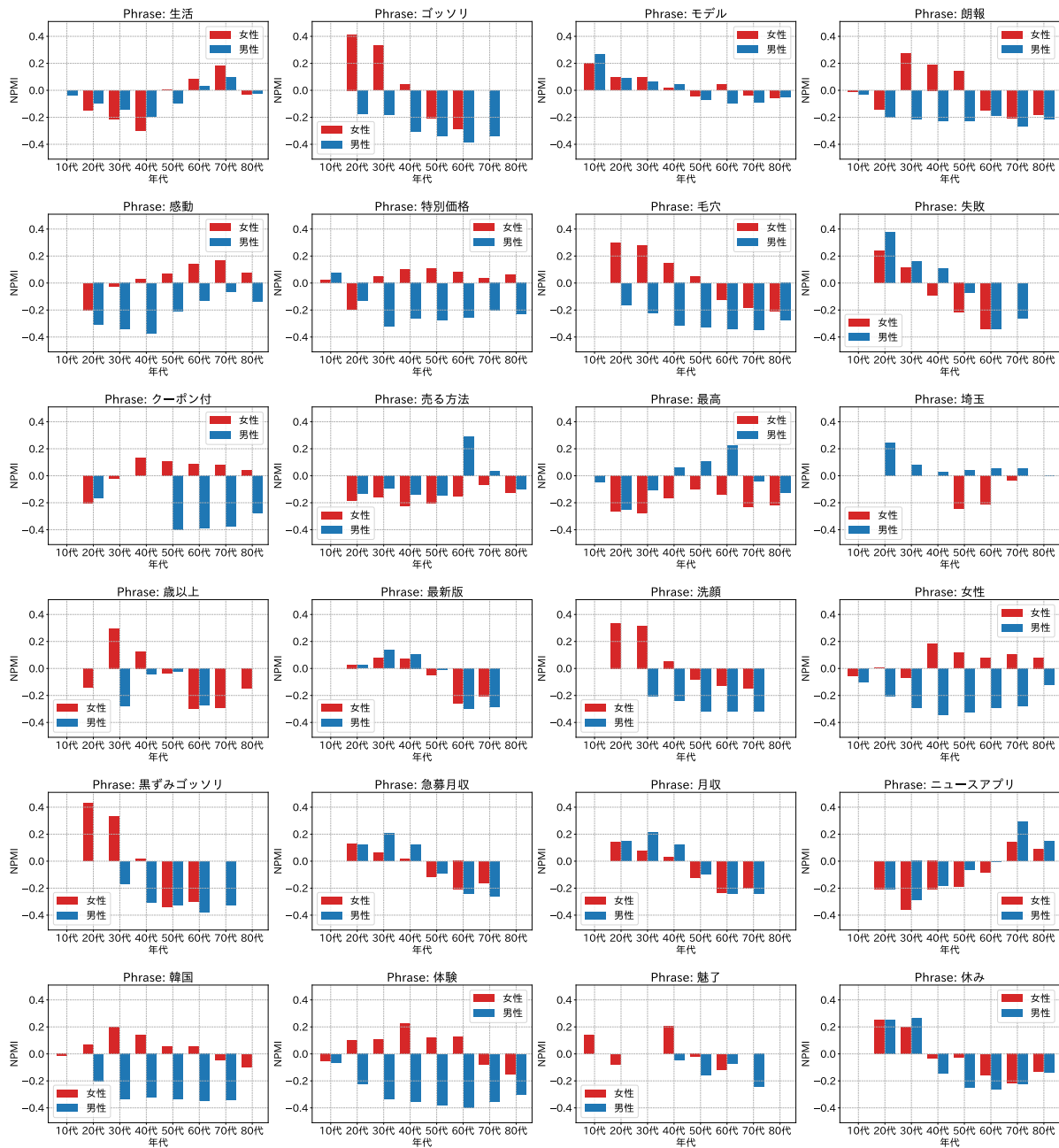


図 7 (表 3) に登場するフレーズについて、開封に与える影響を性別・年代ごとに棒グラフで表したもの (5/5)