転職サイト会員アンケートからのアウトカムを高めるトピック抽出
ー構造トピックモデルとグラフィカルモデルー

窪野 哲光[†]　　日吉 のぞみ[‡]　　明石 大樹[‡]

†パーソルキャリア コーポレート本部 〒100-0005　東京都千代田区丸の内 2-5-2 三菱ビル 8F

‡パーソルキャリア 転職メディア事業部 〒100-0004　東京都千代田区大手町 1-6-1 大手町ビル 8F

E-mail:　†‡{norimitsu.kubono, nozomi.hiyoshi, daiju.akashi}@persol.co.jp

あらまし　人材サービス会社パーソルキャリアが運営する転職サイト DODA の会員アンケートテキスト分析に，「構造トピックモデル」と「グラフィカルモデル」を併用して，アクションにつなげることのできる，アウトカム NPS を高めるためのトピック抽出方式について検討したので報告する。表現力豊かなトピックモデルである「構造トピックモデル」により得られた，トピック相関，トピック～メタデータの共変量の関連性を，「グラフィカルモデル」により特徴選択して可視化分析することが効果的であることを示す。更に，「構造トピックモデル」により得られた「文書ートピックの配分」行列は，「文書ー語」行列の潜在トピックに基づく次元圧縮であることから，文書自動分類に適した文書ベクトルであることも示す。

キーワード　構造トピックモデル，トピック相関，共変量，スパース性，グラフィカルモデル，特徴抽出，構造学習，グラフィカルラッソ，ベイジアンネットワーク，自己組織化マップ

Useful Topic Extraction Method to Increase Outcome from Career Site VOC
- Feature selection by Structural Topic Model and Graphical Model-

Norimitsu KUBONO[†]　　Nozomi HIYOSHI[‡]　and　Daiju AKASHI[‡]

† Corporate Headquarters, PERSOL CAREER　2-5-2 Marunouchi, Chiyoda-ku, Tokyo, 100-0005 Japan

‡ Career Change Media Division, PERSOL CAREER　1-6-1 Otemachi, Chiyoda-ku, Tokyo, 100-0004 Japan

E-mail:　†‡{norimitsu.kubono, nozomi.hiyoshi, daiju.akashi}@persol.co.jp

Abstract　We describe the result of examination applying Structural Topic Model and Graphical Model Structure Learning complementarily to text analysis of "user / withdrawal questionnaire" of job change site DODA operated by PERSOL CAREER. We have studied a useful topic extraction method for improving outcome NPS that can lead to action by using a Structural Topic Model and a Graphical Model. It is effective to visualize and analyze the relevance of topic correlation, topic ~ metadata covariates obtained by Structural Topic Model which is rich expressive topic by feature selection by Graphical Model Structure Learning. Furthermore, the topic allocation "document - topic" matrix of the document obtained by the Structural Topic Model is also a document vector suitable for document clustering because it is information dimension compression of the "document - word" matrix.

Keywords　Structural Topic Model, topic correlation, covariant, sparsity, Graphical Model, feature selection, structure learning,　Graphical Lasso, Baysian Network, Self-Organized Maps

1. はじめに

マーケティング分野における消費者インサイトの重要性，その抽出のためのトピックモデル適用を示した後，課題と問題設定について述べる。

1.1. 背景（消費者インサイト，見えない属性）

User-Generated Content（ユーザー生成コンテンツ，以下，UGC）には，「消費者インサイト」，つまり消費者が企業に理解して欲しいと潜在的に願うメッセージが含まれていることから，UGC のビジネスへの活用が，マーケティング活動を支援できると言われている。UGC は消費者インサイトを抽出するために価値ある情報源と見なされるが，文書などの非構造化データであることから，効率的に抽出する手法として<u>トピックモデルを適用することの役割，利点</u>について論じている[1]。UGC から抽出されるトピックのことを製品やサービスに関する「見えない属性(un-seen-attribute)」と表現している。この「見えない属性」は，消費者が重要視する製品やサービスの潜在的な特徴であり，消費者による自社の製品・サービスについての指摘，重要な事柄を発見することができる効果，企業間の競争力を促す効果があると言われている。

1.2. 課題，問題設定

人材サービス会社パーソルキャリア（2017 年 7 月にインテリジェンスから社名変更）は，転職サイト DODA（http://DODA.jp/）を運営している。会員に実施している「利用者・退会者アンケート」は特に重要な VOC である。業務ではテキストマイニング製品を利用して，頻出の単語ランキングやボリューム，単語同士のつながり，特定の属性でどんな単語が出やすいか，ということまでは把握できている。しかし，テキストに潜んでいる「話題」の種類と大きさ，「話題」と「話題」のつながり，「話題」と「属性（メータデータ）」との関連性については見えづらい，という課題がある。

「利用者」の顧客親密度の向上と利用促進，「退会者」の離反防止が強く求められていることから，施策の手がかりとなるような有益な「消費者インサイト」をアンケートから得るために，以下の問題設定をする。

【問題設定】

DODA 会員アンケートデータ（2.1 で後述）から「利用者インサイト」を意味する良質なトピックを的確に抽出し，次に，アウトカム NPS の点数を高めるために寄与する有益なトピックを特徴選択する。

取り組みの見取図を図 1 に示す。

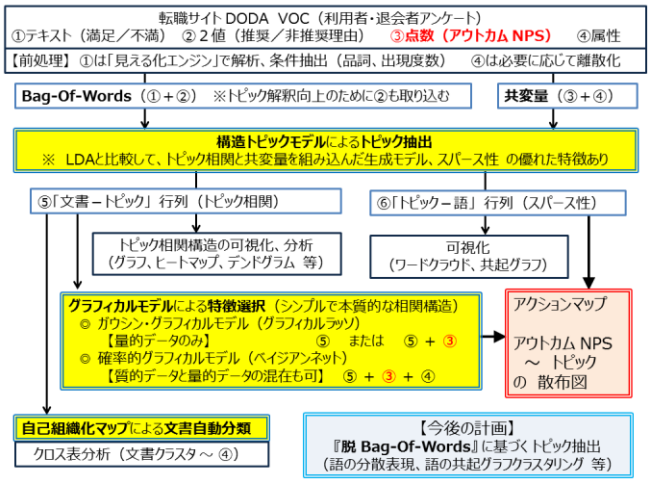


図 1. 取り組みの見取図

2. 分析に使用したデータセット

DODA 会員アンケートデータ～構造トピックモデル入力データの関連性について述べる。

2.1. DODA 会員アンケートデータ

4 種類のデータファイルから構成される。

①テキスト

会員アンケート（1996 年 7 月～12 月の 121,289 人）から，自由記述【DODA への満足・不満】を「見える化エンジン」で解析，条件抽出（品詞が名詞・形容詞・動詞，度数が 5 回以上）して得た 9,167 人（内訳は，利用者が 6,711 人，退会者が 2,456 人）。

②アセスメントデータ

ブランドイメージ，求人，メール，コンテンツ，イベント等 10 項目について，【DODA 推奨／非推奨理由】の双対で得た回答（0/1 の２値，複数回答）。

③アウトカム NPS

NPS（Net Promoter Score）は【顧客親密度】を 0～10 の点数で得た回答。点数の 0～5 は「批判者」，6～8 は「中立者」，9～10 は「推奨者」の，順序付き 3 クラスで統計集計。

④属性

会員属性のうち，会員区分（利用者／退会者），コア区分（コア／非コア），年齢，転職回数，回答年月。

2.2. 構造トピックモデルの入力データ

BOW(Bag-Of-Words)は，前出の①テキストと②アセスメントデータを結合して作成する（9167 文書＊1568 異なり語）。共変量は，③と④を結合して作成する。

3. 構造トピックモデルの適用

構造トピックモデルの特徴と位置づけを説明した後，構造トピックモデルの計算結果について述べる。

3.1. 構造トピックモデルとは

構造トピックモデル（Structural Topic Model 以下 STM と略す）[2]は，ハーバード大学の社会学分野で 2013 年に開発された比較的新しいトピックモデルである。日本では社会学分野の研究報告[3][4]がある。

STM はトピック相関，共変量情報を組み込むことができる生成モデルであるという点が，通常の LDA[5]との最大の違いであり，その特徴を図 2 に示す。

共変量により，例えば，文書の出版年や著者，その他のメタデータを組み込んだ形でトピックの比率および個別トピックのもつ単語の確率分布を推定できる。こうした共変量によってトピックがどう異なるか（文書～トピック配分の出現確率にメタデータの影響を与えるタイプの共変量であり，prevalence と呼ばれる），トピックの内容はどうか（トピック～語の出現確率にメタデータの影響を与えるタイプの共変量であり，contents と呼ばれる），といった問題を考える場合には有益なトピックモデルである。

STM は，LDA[5]以降の 3 種類の新しいトピックモデルの特徴を結合させたベイズ文書モデルである[2]。

- Correlated Topic Model (2007)
  LDA をベースにトピック相関の拡張
- Dirichlet-Multinomial Regression (2008)
  ディレクトリ分布を用いたベイズ的多項式回帰
- Sparse additive generative models (2011)
  トピックからの語の生成確率のスパース性
  潜在意味情報解析ではスパース性が重要であり，トピック相関と相性がよい

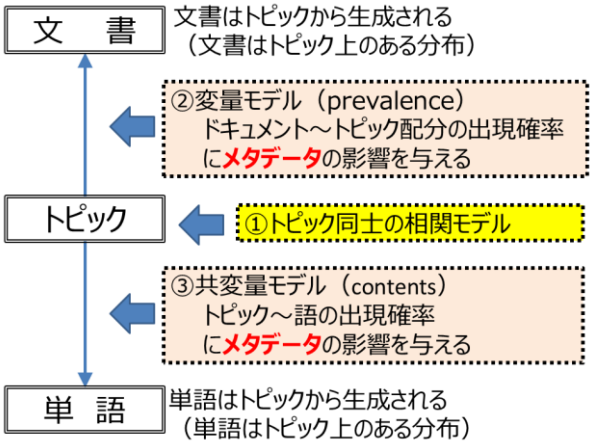


図 2. 構造トピックモデルの特徴

3.2. 最適トピック数の推定

最適トピック数の推定結果を図 3 に示す。スクリープロットから最適トピック数は 27 であると判断した。

トピックモデルは「文書ー語」行列の潜在トピックに基づく次元圧縮であることから,「残渣」は分かりやすい指標である。特性が異なる以下の 2 指標の相補的利用はヒューリスティックではあるが効果的である。

・値が大きいほど好ましい指標「確からしさ」
・値が小さいほど好ましい指標「残渣」

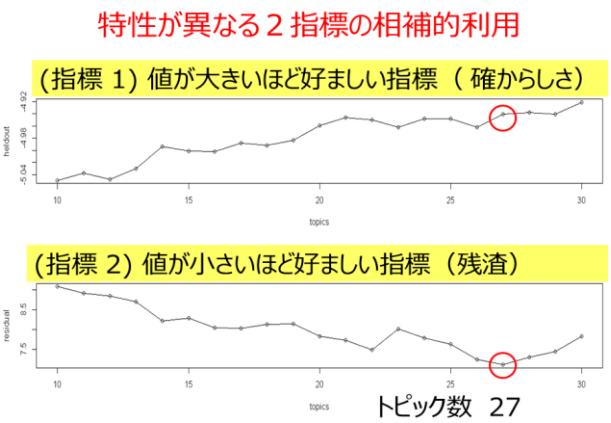


図 3. 最適トピック数の推定

3.3. トピックの大きさ，特徴語，可視化

抽出された 27 トピックの一覧を図 4 に示す。横軸はトピックの大きさを示す（文書数基準の割合）。トピックからの出現確率が高い語を特徴語として付与することで，トピックのポジネガの判別ができる。これにより，トピック（以下，T_通し番号 で表記）が大きい上位 3 の T_7，T_16，T_8 は以下のように解釈できる。

・T_7【ポジ系】DODA サイト・求人票の見やすさや検索しやすさに満足　→図 5 左図
・T_16【ネガ系】メールが大量であることに不満
・T_8【ポジ系】キャリアアドバイザーのサポートに対する高評価・感謝　→図 5 右図

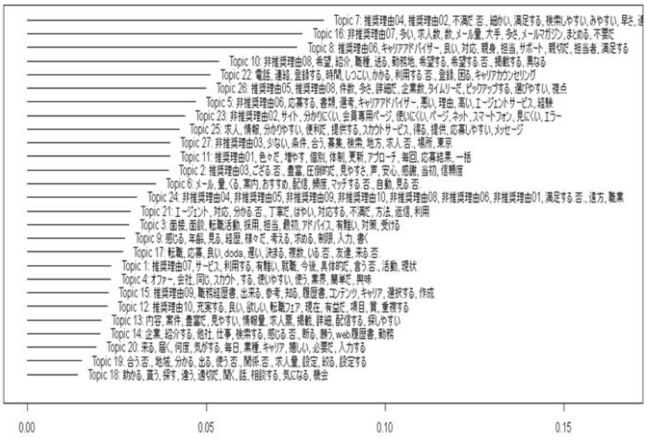
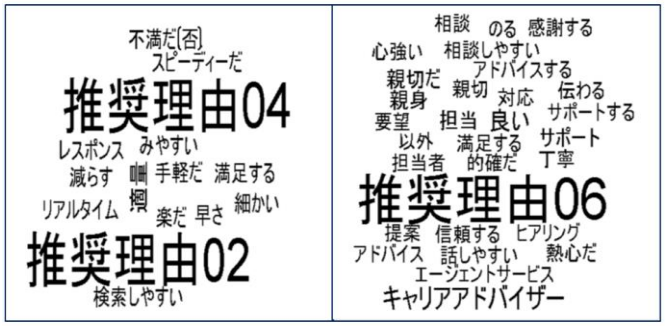


図 4. 抽出トピックの一覧（大きさ，特徴語）

出力結果表の 1 つである「トピックー語の出現確率」行列をトピック毎に可視化したものがワードクラウドであり，例としてポジ系の T_7，T_8 を図 5 に示す。

表記「推奨理由 xx」はアセスメントデータである（前出 2.1②）。「推奨理由 02」は「DODA サイト・マイページが見やすい，探しやすい」,「推奨理由 04」は「求人票が見やすい」,「推奨理由 06」は「キャリアアドバイザー（エージェントサービス）の対応が良い」の意味を表すラベルである。

アセスメントデータと語を同時に表示することで，通常の語だけの表示と比較して，トピックの特徴語の直感的な解釈をより支援できる。なお，図 5 左図の上部の語「不満だ(否)」は，テキストマイニング製品「見える化エンジン」の特徴的な出力形式であり，語「不満だ」の後に否定語が続くことを意味し，結果として肯定的な意味（「不満ではない」）となる。



T_7（第 1 位）　　T_8（第 3 位）

図 5. トピックのワードクラウド

3.4. 共起グラフ(トピック～語・係り受け)

トピック～語・係り受け（「見える化エンジン」解析結果）の共起グラフを図 6 に示す。矩形ノードはトピック，円ノードは語を表す。ノードの大きさは出現頻度に比例するように設定する。語と係り受けを同時に表示することで，通常の語だけの表示と比較して，トピックの特徴語の直観的な解釈をより支援できる。

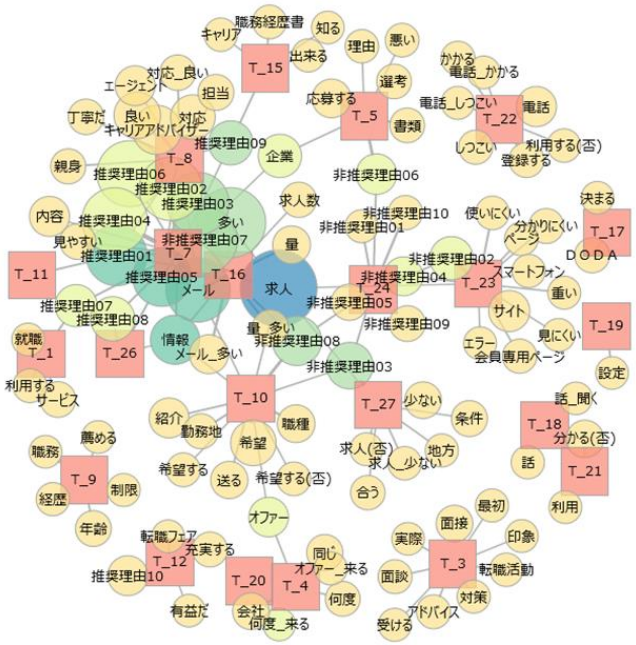


図 6. 共起グラフ（トピック～語・係り受け）

4. トピック相関の可視化分析

出力結果表の１つである「文書－トピックの配分」行列から，トピック相関構造の多面的・多角的な可視化分析を行う。トピックベクトル(行列の列ベクトル)を用いて，トピック同士の相関係数を計算する。

4.1. 相関係数によるヒートマップ

トピック同士の相関係数によるヒートマップを図 7 に示す。トピックはポジ系～ネガ系の強さの順序付けで対角線上に整列されていることが見られた。

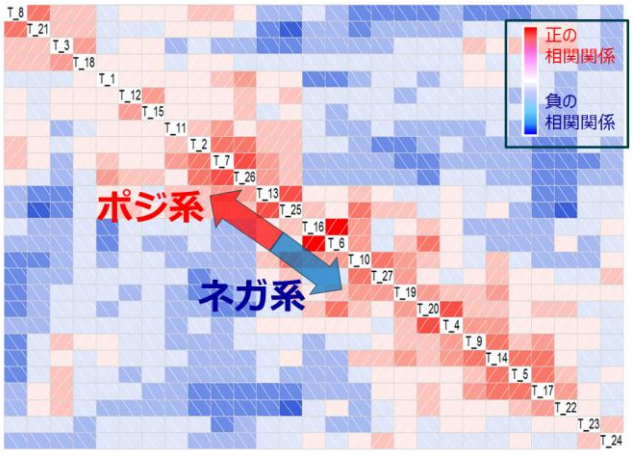


図 7. 相関係数によるヒートマップ

4.2. 相関係数による階層型クラスタリング

ブートストラップ法に基づく信頼度の高い，相関係数による階層型クラスタリング[6]の結果を図 8 に示す。最終的にはポジ系～ネガ系の 2 クラスタまで併合できることが分かった。

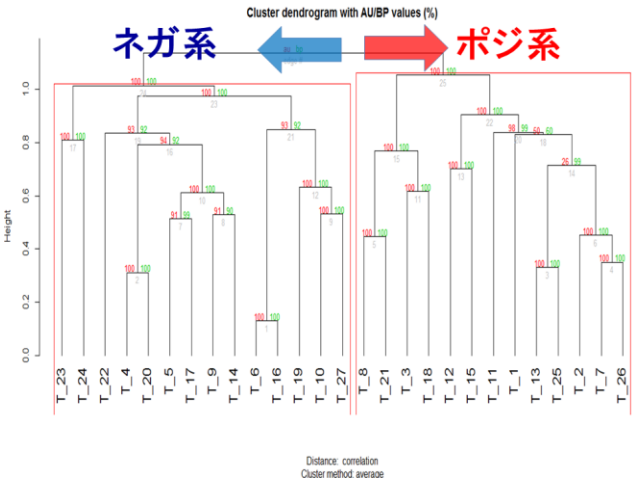


図 8. 相関係数による階層型クラスリング

4.3. 相関係数による共起グラフ

トピック同士の相関係数をエッジ重みとするトピック共起グラフを，観点が異なる 3 種類（①②③）で比較した結果を図 9，10 に示す。エッジ重み閾値は，孤立ノードが出現しない下限値の相関係数 0.18 とした。①すべてのエッジをつなぐ，②ネットワーク分析[7]の最小スパン木，③ネットワーク分析[7]のコミュニティ抽出（walktrap 法でグラフクラスタリング）

②③では，トピックはポジ系～ネガ系の 2 グループに大別され，前出の図 7，8 とよい対応が見られた。

STM はトピック相関性を組み込む生成モデルであることから，図 9,10 のトピック共起グラフを作成できる。一方，トピック相関性を組み込むことのできない通常の LDA ではトピック共起グラフは意味をもたない。

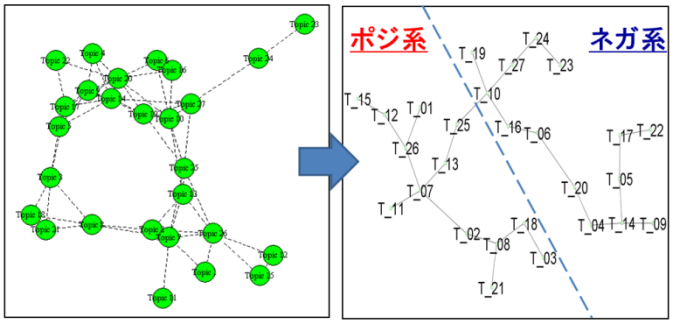


図 9. 相関係数による共起グラフ（① vs ②）

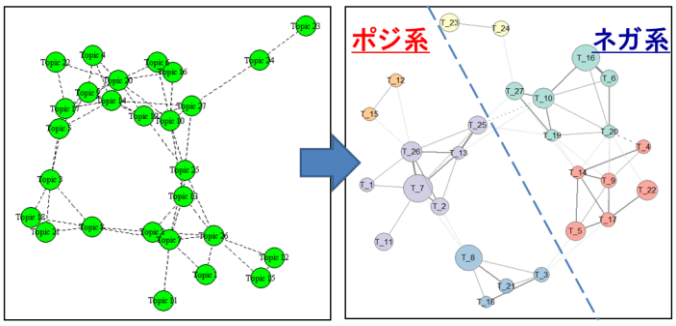


図 10. 相関係数による共起グラフ（① vs ③）

共起グラフの問題として 2 点ある。1 つ目は，最適トピック数とエッジ重み閾値の 2 つの設定により，グラフ（エッジ有無）が変わることである。2 つ目は，相関係数は 2 つの関係を見る際には有効であるが，3 つ以上のトピック同士の絡みが見えにくいことである。

4.4. グラフィカルモデル適用の意義

シンプルで本質的な相関構造を得るためには，トピック同士の関係を統計的な「条件付き独立性」に対応できるグラフィカルモデル[8][9][10]の適用が有用である。グラフィカルモデルは，確率変数間の「条件付き独立性」に基づいて，変数同士の依存構造を表現する確率モデルであり，複雑な多変量データの構造を分かりやすく表現できる手法である。「条件付き独立性」を適用することで，相関関係を「直接相関（真の相関関係）」と「間接相関」とに分けることが可能となり，シンプルで本質的な相関構造を得ることができる。

グラフィカルモデルは変数の尺度水準により，量的変数のグラフィカルモデル(ガウシアン･グラフィカルモデル[8]，以下 GGM と略す)，質的変数のグラフィカルモデル（確率的グラフィカルモデル[11]，以下 PGM と略す）の 2 種類に大別されるので，分析データ種類に応じて使い分ける。

5. ガウシアン･グラフィカルモデル

相関関係の強いデータセットに適用する場合の問題を説明した後で，その問題をスパース推定により解決するグラフィカルラッソについて述べる。そして，グラフィカルラッソを｢文書－トピックの配分｣行列，アウトカム NPS～「文書－トピックの配分」行列へ適用する提案方式について述べる。

5.1.「直接相関」と「間接相関」の定義

「直接相関」と「間接相関」を以下のように定義する。

- 相関係数=0　→　独立
- 偏相関係数=0　→　条件付き独立
- 相関係数は，「直接相関」と「間接相関」の和
- 偏相関係数は，「直接相関」の量を表す

5.2. 精度行列の計算に係る問題

「直接相関」構造を得るために伝統的には，共分散選択（偏相関係数のいくつかを 0 と置いた相関構造モデルを採用するアプローチ）がとられてきた[8]。

共分散行列（平均を 0，分散を 1 に標準化すると，相関係数行列）から偏相関係数行列を求める手順を R 言語で示すと以下のようになる。

```
R                    # 共分散行列（相関係数行列）
X                    # 精度行列
P                    # 偏相関係数行列
X <- solve(R)        # 精度行列の計算
d <- sqrt(diag(X))
P <- -X / (d %*% t(d))
```

変数同士に強い相関性がある場合には（多重共線性），共分散行列の逆行列である精度行列の計算不能という問題が指摘されている（変数の個数が多くなるビックデータでは避けられない問題）。これを解決するために，L1 正則化を導入したスパース性に基づいて精度行列を正確に計算できる「グラフィカルラッソ」[12]が提案されており，高速性，頑強性という優れた特徴が報告されている[13]。また，スパース性を導入した機械学習が注目されている[14] [15]。「文書－トピックの配分」行列ではトピック同士に強い相関関係が見られることから，精度行列の計算不能の問題に遭遇した。

5.3. グラフィカルラッソによる直接相関構造

本稿では，シンプルで本質的な相関構造を得ることが目的であることから，正則化パラメータ ρ を 0 から 1 までの間で逐次変化させながら，逸脱度 SRMR(Standardized Root Mean square Residual)[8]の差分，精度行列の差分（尤度比に比例する量）[9]の変化が大きいステップの ρ を採用する。もし大きいステップが特に見られない場合には，孤立ノードが出現する直前の値を採用した。

④相関係数でつなぐ場合，⑤グラフィカルラッソによる偏相関係数でつなぐ場合の比較を，図 11 に示す。（エッジ重みが 0.05 以上を表示）

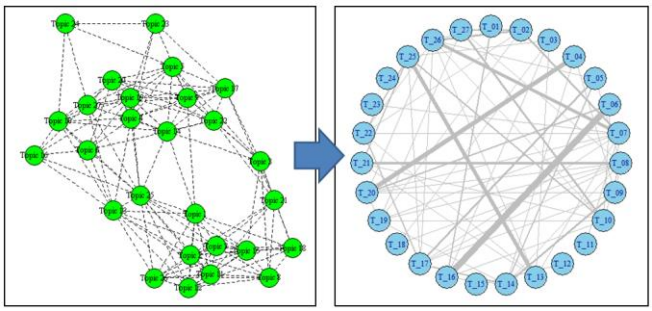


図 11. 共起グラフ（④相関係数 vs ⑤偏相関係数）

⑤から得られた，直接相関が強いトピック同士の代表的な組の例を以下に示す。([P]はポジ，[N]はネガ)

- T_06（[N]メールが多すぎることによる弊害に不満）～T_16（[N]メールが大量であることに不満）
- T_04（[N]オファーに対する不満）～T_20（[N]メールの頻度・内容に対する評価）
- T_07（[P]DODA サイト・求人票の見やすさや検索しやすさに満足）～T_26（[P]求人や求人票に対する評価）
- T_08（[P]キャリアアドバイザーのサポートに対する高評価・感謝）～T_21（[P]DODA 担当者の対応に満足）
- T_13（[P]求人・求人票の内容・情報量に満足）～T_25（[P]求人情報に対する評価）

次に，2 つの結合ファイル（アウトカム NPS～「文書ートピックの配分」行列）に，グラフィカルラッソを適用した「直接相関」構造を図 12 に示す。

アウトカム NPS～トピックとの偏相関係数の降順（左図）ではスパース推定により 0 要素が多くなっていることから，シンプルで本質的な相関構造を得ることができた。トピック T_22 は，マイナス方向で最も大きいことから，アウトカム NPS を高めるためには「最優先で改善」すべきトピックであることが分かる。

トピック	偏相関係数（NPS～トピック）
T_08	0.194
T_02	0.175
T_07	0.153
T_26	0.121
T_11	0.024
T_01	0
T_03	0
T_04	0
T_06	0
T_09	0
T_10	0
T_12	0
T_13	0
T_14	0
T_15	0
T_16	0
T_18	0
T_20	0
T_21	0
T_23	0
T_25	0
T_24	-0.001
T_27	-0.039
T_19	-0.044
T_17	-0.079
T_05	-0.096
T_22	-0.122

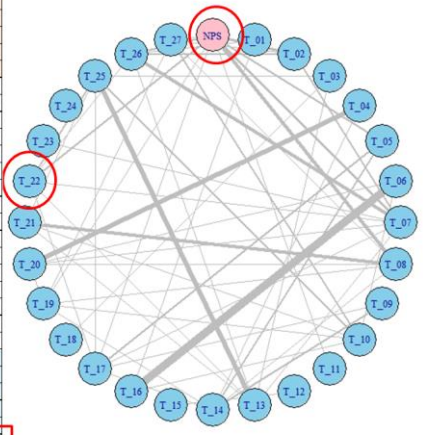


図 12. グラフィカルラッソ（アウトカム～トピック）

つながるエッジ重み（相互情報量）の総和とした。属性「年月」と T_12 のつながりが見られることから，T_12 に注目することで定点観測の可能性があることが分かった。また，相互情報量はエントロピー基準に基づく依存性という性質から，相関係数（偏相関係数）では得られない，ポジ系とネガ系のつながり（T_13～T_22，T_8～T_19）が見られ興味深い。

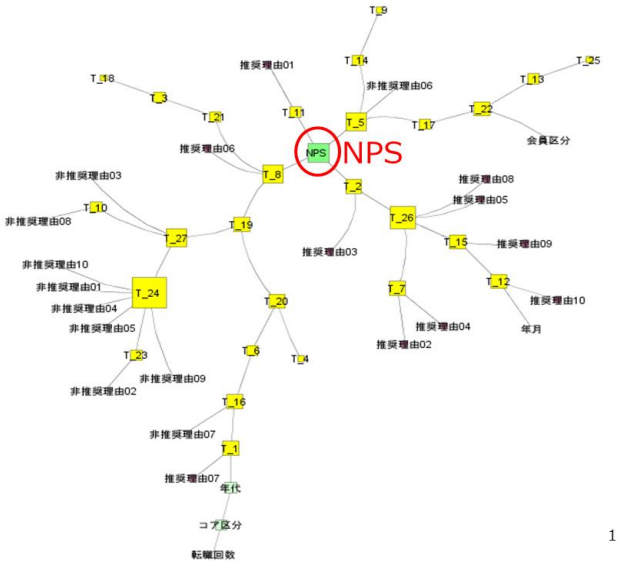


図 13.ベイジアンネット無向森グラフ

6. 確率的グラフィカルモデル

確率的グラフィカルモデルの一種であるベイジアンネット[11]は，確率変数同士の同時分布を条件付き独立性に基づいて逐次的因数分解した確率モデルである。質的データと量的データが混在するデータセットにも適用可能なベイジアンネット構造学習[16]を利用した結果について述べる。

6.1.「直接相関」と「間接相関」の定義

機械学習では，情報理論のエントロピー基準に基づく確率変数間の依存性の尺度である相互情報量を用いた特徴選択が注目されている[17]。ベイジアンネット構造学習[16]では相互情報量を用いて，「直接相関」と「間接相関」を以下のように定義する。

- 相互情報量=0　→ 独立
- 条件付き相互情報量=0　→ 条件付き独立
- 相互情報量は，「直接相関」と「間接相関」の和
- 条件付き相互情報量は「直接相関」の量を表す

6.2. ベイジアンネット構造学習による直接相関構造

3 つの結合ファイル（アウトカム NPS～「文書ートピックの配分」行列～属性）は，量的データと質的データとが混在するデータである。得られた無向森グラフを図 13 に示す。中心のアウトカム NPS の周辺にはトピックがつながり，その先の外辺には属性データがつながる構造が得られた。ノードの大きさはノードに

7. アクションマップの作成

図 4（抽出トピックの種類と大きさ），図 12（グラフィカルラッソ）の 2 つの結果を組み合わせた，アクションマップを図 14 に示す。トピック T_22 は，アウトカム NPS に対する偏相関係数がマイナス方向で最も大きく，かつ，トピックのボリュームが大きいことから，アウトカム NPS を高めるためには「最優先で改善」すべきトピックであることが分かる。

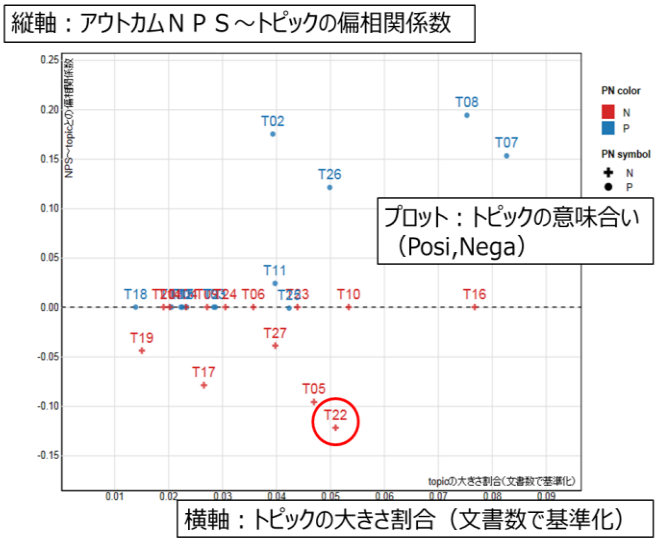


図 14.アクションマップ（アウトカム NPS～トピック）

8. 文書自動分類への応用

文書自動分類の問題を概観した後で，「STM－SOM（自己組織化マップ）法」の提案方式について述べる。

8.1. 文書自動分類の問題

文書自動分類の問題は，「文書－語」行列の問題（高次元，データスパース），クラスタリング手法の問題（k-means 法の初期値依存性，クラスタ数の所与）に整理できる。前者への対応としては SVD（主成分分析，LSA 等）やトピックモデル（pLSA，LDA）による情報次元圧縮が提案されている。後者への対応としては k-means 法の問題を緩和，回避できる手法として，例えば，自己組織化マップ[18]などが提案されている。

8.2. STM－SOM 法の比較実験

「LDA－SOM 法」に基づく文書自動分類[19]にヒントを得て，3 種類の SOM 入力データの比較実験をした。

・語：「文書－語」行列（1568 語）　　（）は次元
・LDA：「文書－トピックの配分」行列（27 トピック）
・STM：「文書－トピックの配分」行列（27 トピック）

【比較実験 1】　語 vs STM トピック

語と STM トピックで比較した結果を図 15 に示す。

左図（語）の場合には，非常に大きいサイズのクラスタ（C_1）があり分離度が悪い。それに対して右図（STM トピック）の場合には分離度が向上して，ほぼ均等なサイズのクラスタが形成されている。また，計算所要時間も数分の一に短縮されていた。

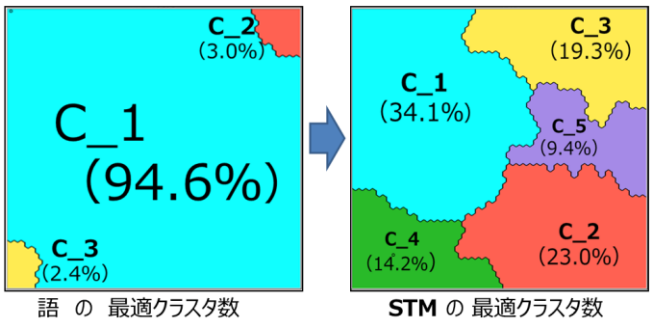


図 15. SOM クラスタの比較（情報次元圧縮の有無）

【比較実験 2】　LDA トピック vs STM トピック

「文書－トピックの配分」行列を，LDA で生成した場合と STM で生成した場合で比較した結果を図 16 に示す。左図（LDA トピック）と比較して，右図（STM トピック）では分離度が向上していることが見られた。

STM の各 SOM クラスタを特徴づけるトピック（プロファイルと呼ぶ）を標準化得点 Z 値により特徴選択した。前出の図 7,8,9,10 と良い対応関係があることが見られた。その一例として，図 10③共起グラフクラスタリングに SOM クラスタのプロファイルを追記したものを図 17 に示す。STM の 27 トピック～SOM の 5 クラスタには良い対応関係があることが見られた。

これらのことから，STM の「文書－トピックの配分」行列の推定精度は高く，文書自動分類に適した文書ベクトルと言える。

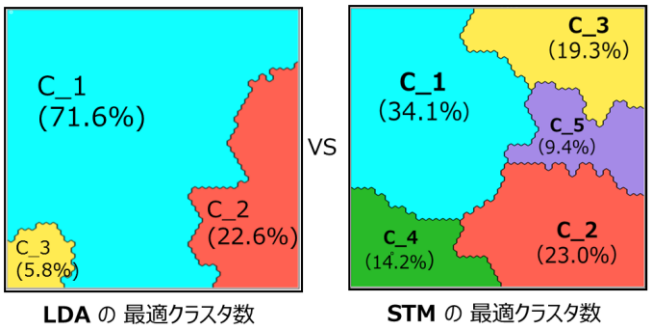


図 16. SOM クラスタの比較（トピックモデルの違い）

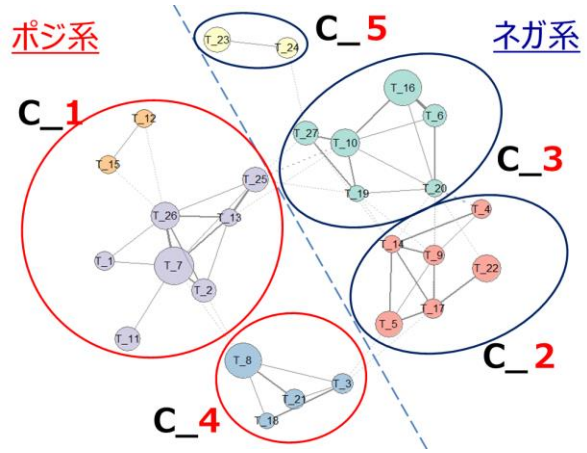


図 17. SOM クラスタのプロファイル

8.3. クロス表分析（アウトカム NPS～SOM クラスタ）

アウトカム NPS（3 カテゴリ）～SOM クラスタのクロス表分析リフト値を，図 18 ヒートマップに示す。項目間には強い関連性が見られた。特に「1:推奨者」では C_4 が最も特徴的であり，「3:批判者」では C_2 が最も特徴的であることが見られた。

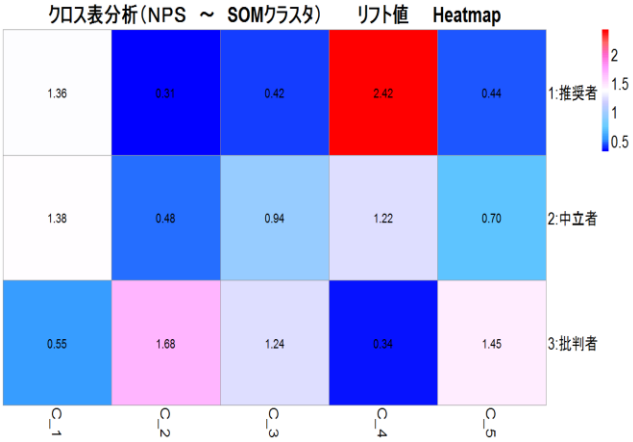


図 18. クロス表分析(アウトカム NPS ～SOM クラスタ)

9. おわりに（まとめ，課題・今後の計画）

構造トピックモデルとグラフィカルモデルの適用成果をまとめ，課題と今後の計画について述べる。

9.1. まとめ

DODA 会員アンケートはテキスト～メタデータの文書構造をもち，STM を適用できる有用な実データである。STM により「消費者インサイト」を意味する良質なトピックを適確に抽出することができた。

STM により得られた，トピック相関，アウトカム NPS～トピックの関連性を，グラフィカルラッソによる直接相関構造で特徴選択，可視化分析することが効果的であることを示した。

さらに，STM により得られた「文書ートピックの配分」行列は，「文書ー語」行列の潜在トピックに基づく次元圧縮であることから，文書自動分類を改善する文書ベクトルであることも示した。

9.2. 課題・今後の計画

BOW(Bag-Of-Words)ファイルでは，語の出現順などの意味的な関連度が欠落する課題が残る。

消費者インサイトをより的確に抽出するために，語の意味的な関連度を取り込んだ検討を進める必要性がある。一つのアイディアとしては，語の分散表現 word2vec と LDA の研究報告[20]を参考にし，新たに，語の共起グラフクラスタリング(図 19)も取り込んで，STM と分散表現との相補的利用法の検討を進める。

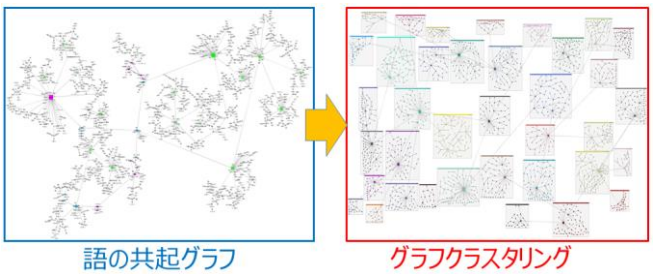


図 19. 語の共起グラフに基づくクラスタリング

謝　辞

東北大学国際高等研究教育機構学際科学フロンティア研究所の瀧川裕貴助教授から，構造トピックモデルについてご教示を受けました。

文　献

[1]佐藤,″マーケティング研究におけるトピックモデルの適用に関する一考察″,経済研究,第 68 巻第 3 号,大阪市立大学経済学部,2017.

[2]Roberts, Stewart, Tingley, and Airoldi. "The Structural Topic Model and Applied Social Science." Advances in Neural Information Processing Systems Workshop on Topic Models, 2013.

[3]大林,瀧川,″『理論と方法』におけるテーマの 30 年,方法の 30 年″,『理論と方法』,Vol. 31, 99-108, 2016.

[4]瀧川,″戦後日本社会学史への計算社会科学的アプローチー『社会学評論』1954-2015 の構造トピックモデルによる分析″,第 89 回日本社会学会大会,2016.

[5]David M Blei, Andrew Y Ng, and Michael I Jordan. "Latent dirichlet allocation". *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993-1022, 2003.

[6]R.Suzuki, H.Shimodaira, "Pvclust: an R package for assessing the uncertainty in hierarchical clustering", Bioinformatics, 22 (12): 1540-1542,2006.

[7] 鈴木,″ネットワーク分析 第 2 版 (R で学ぶデータサイエンス 8)″,共立出版,2017.

[8]宮川,″グラフィカルモデリング (統計ライブラリー) ″,朝倉書店,1997.

[9]C.M. ビショップ (原著),元田他(監訳),″パターン認識と機械学習 下 (ベイズ理論による統計的予測)″,第 8 章,丸善,2012.

[10]渡辺,″グラフィカルモデル (機械学習プロフェッショナルシリーズ) ″,講談社,2016.

[11]鈴木,植野(著,編集) ,″確率的グラフィカルモデル″,共立出版,2016.

[12]J. Friedman et al."Sparse inverse covariance estimation with the graphical lasso". Biostatistics, 9:432 - 441, 2008.

[13]井手,″依存関係にスパース性を入れる″,岩波データサイエンス,vol5, pp48-63, 2017.

[14]冨岡,″スパース性に基づく機械学習 (機械学習プロフェッショナルシリーズ) ″,講談社,2015.

[15]Trevor Hastie,Robert Tibshirani,Martin Wainwright,"Statistical Learning with Sparsity The Lasso and Generalizations",Chapman & Hall/CRC,2015.

[16] Suzuki, J., "A novel Chow-Liu algorithm and its application to gene differential analysis", International Journal of Approximate Reasoning, 2017.

[17]杉山,入江,友納,″相互情報量を用いた機械学習とそのロボティックスへの応用″,日本ロボット学会誌,vol33,No.2, pp86-91,2015.

[18] Kohonen, T. "Self-Organizing Maps" Springer Series in Information Sciences. New York: Springer, 3rd edition.2001.

[19]Jeremy R. Millar and Gilbert L. Peterson and Michael J. Mendenhall,"Document Clustering and Visualization with Latent Dirichlet Allocation and Self-Organizing Maps",the Twenty-Second International FLAIRS Conference,2009.

[20]江原,″生コーパスからの単語難易度関連指標の予測″,言語処理学会 第 23 回年次大会,2017.