

第116回モバイルコンピューティングと新社会システム 口頭発表[MBL116-8]

**VTuber2Vec:**

**VTuberタレントの定量的解釈に向けた文書ベクトルの応用**

---

徳山 儀亮, 吉井 健敏(株式会社D2C), 持橋 大地(統計数理研究所)

1. 研究背景
2. 提案手法
3. 実験
  - i. データセットの準備
  - ii. VTuber2Vecの可視化
  - iii. 階層型クラスタリングを使用したベクトル作成条件の評価
  - iv. タレント事務所別ファン特性の理解
4. まとめ
5. Appendix

- ✓ VTuberとはアバターを用いるオンライン活動者のことである<sup>[1]</sup>
- ✓ 所属・活動形式によって概ね3パターンに分類される

VTuberの類型 <sup>[2]</sup>	背景・例
個人型	<ul style="list-style-type: none"><li>セルフレプロデュースをしながら種々の配信・創作活動を行う 例：名取さな(なとり さな)さん</li></ul>
組織型	<ul style="list-style-type: none"><li>事務所内で他タレントとの交流を行いながら配信・タレント活動を行う 例：月ノ美兎(つきの みと)さん [にじさんじ]</li></ul>
	<ul style="list-style-type: none"><li>特定企業に所属し、アンバサダー活動を行う 例：ウェザーロイド Airiさん[ウェザーニューズ社]</li></ul>

本スコープ



[1] 武田太一, 濱崎雅弘, 後藤真孝. Vtuber におけるソーシャルネットワークに関する分析. Web インテリジェンスとインタラクション研究会 予稿集 第 15 回研究会, pp. 9-12. 2019.

[2] 発表者分類による.

- ✓ タレントのファン情報は通常非公開であり, 宣伝起用文脈での横比較は困難
- ✓ タレントの所属組織を横断した, ファン層の定量評価アプローチを確立したい

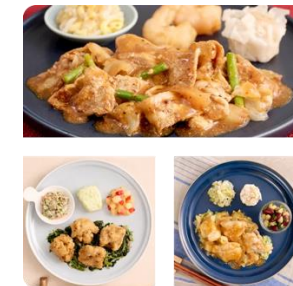
事例1: リポにじ応援祭キャンペーン<sup>[1]</sup>事例2: nosh宣伝活動<sup>[2]</sup>

商品

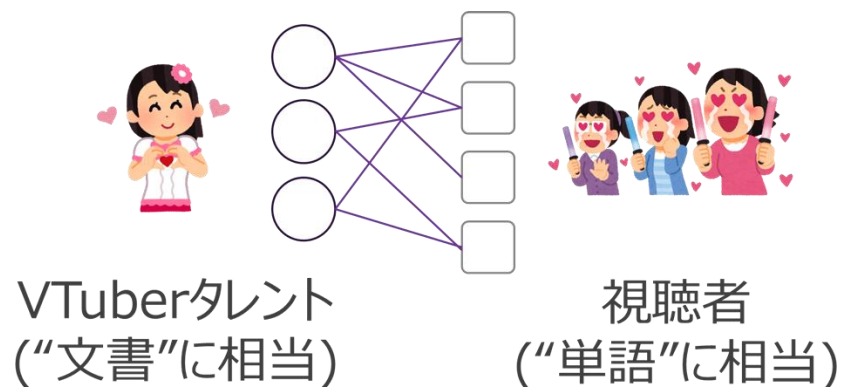
リポビタン(栄養ドリンク)

VTuber  
タレント(タレント: 三枝明那さん、不破湊さん、  
甲斐田晴さん)

nosh(冷凍宅配弁当)

(タレント: 宝鐘マリンさん、天音かなたさん、  
他多数)[1] <https://www.taisho.co.jp/company/news/2025/20250825001971/> より引用[2] nosh公式webサイト, <https://www.youtube.com/watch?v=h0uTcFe6T4c>, [https://www.youtube.com/watch?v=YjIGk\\_xgwp0](https://www.youtube.com/watch?v=YjIGk_xgwp0) より引用

- ✓ 動画チャットデータを用い, VTuberと視聴者を2部グラフとして表現する
- ✓ Word2Vecと等価な特異値分解を通じて<sup>[1]</sup> VTuberベクトル<sup>[2]</sup>を獲得し, 考察



タレント(c)	視聴者(w)	チャット数
A	a	3
A	b	10
...	...	...
B	b	4

表形式データによる  
2部グラフの表現

Shifted Positive PMI( $c, w$ )

$$= \max \left( \log \frac{p(c, w)}{p(c)p(w)} - \log k, 0 \right)$$

$$M = U \Sigma V^T$$

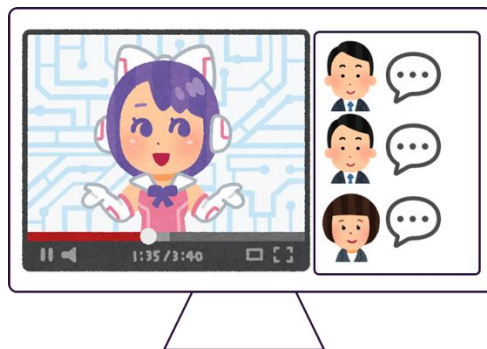
タレントとチャット者の  
組合せの珍しさを評価

$U \Sigma^{1/2}$ が単語ベクトル、 $V^T \Sigma^{1/2}$ が文書ベクトル  
を示し, 後者をVTuberベクトルと呼称

[1] Omer Levy and Yoav Goldberg. Neural word embeddings as implicit matrix factorization. Advances in neural information processing systems, Vol. 27, 2014.

[2] 徳山 儀亮, 吉井 健敏, 持橋 大地. メタ情報を考慮した VTuberのファン層の異なりの評価. 第4回計算社会科学大会(CSSJ 2025), P2-11, 2025.

✓ 25年1-6月におけるYouTubeのライブ配信データからチャット情報を取得した



YouTube Data API<sup>[1]</sup>  
を用いたデータ取得

主な抽出条件：

- [タレント条件]  
にじさんじ, hololive, ぶいすぽっ！,  
ななしいんくの4団体いずれかに所属
- [ライブ配信条件]  
配信時間が30分以上である
- [チャット者条件]  
3人以上のタレントにチャットを送っている

タレント	視聴者	チャット回数
A	a	3
A	b	10
...	...	...
B	b	4

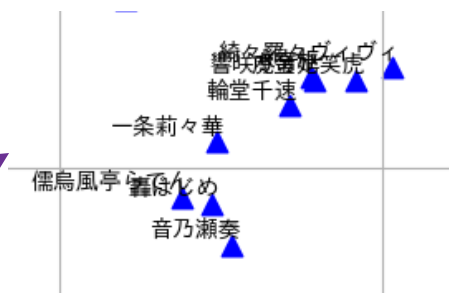
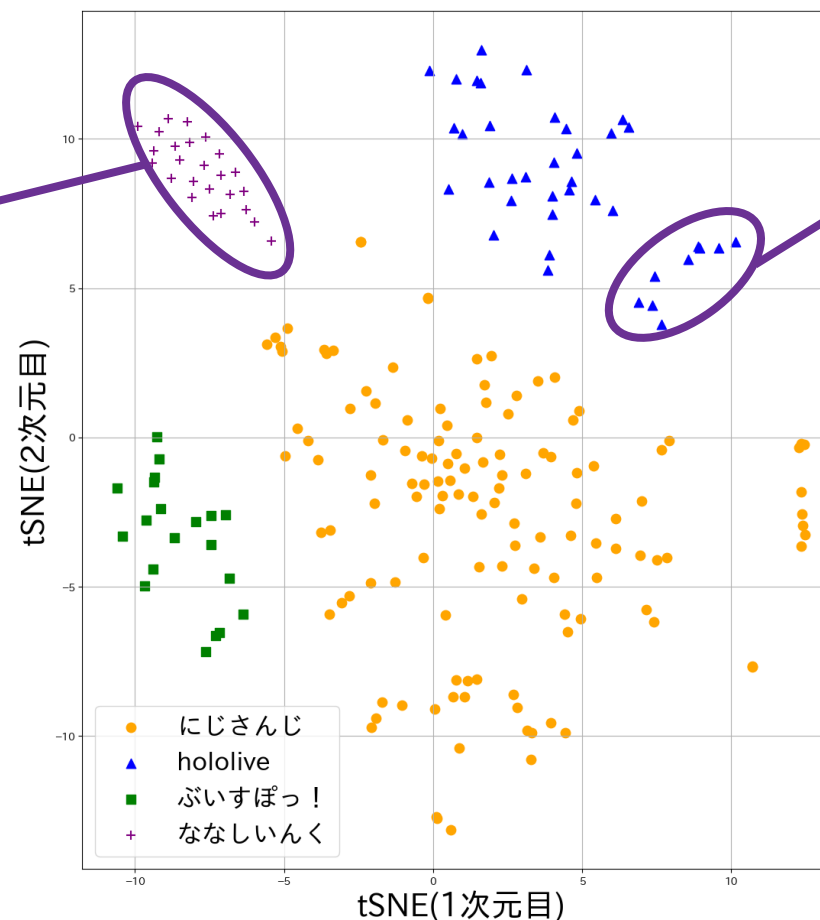
- タレント数: 213名
- 視聴者(チャット者): 約11万人
- チャット件数: 約31億件

[1] <https://developers.google.com/youtube/v3/getting-started?hl=ja>

- ✓ タレントの所属事務所がファン層を規定していることが確認された
- ✓ 局所的には所属事務所内のコラボレーションユニットが分布に影響した

図: t-SNEによるVTuberベクトルの可視化

点群が事務所毎に集中  
= 事務所がファン層を規定している



局所的には事務所内のサブユニットが影響

例: hololiveによる  
タレントグループ  
"hololive DEV\_IS"  
(デバイス)<sup>[1]</sup>(上拡大図)



- ✓ 独立成分分析(ICA)を用いたVTuberベクトルの可視化を行った
- ✓ 事務所によってファンの多様性を確保している軸が異なることが確認された

### 独立成分分析(ICA)の概要

- 各軸の独立性が最大になるような線形射影を求める方法
- 軸に対して意味解釈ができる

図: ICAの概念図<sup>[1]</sup>

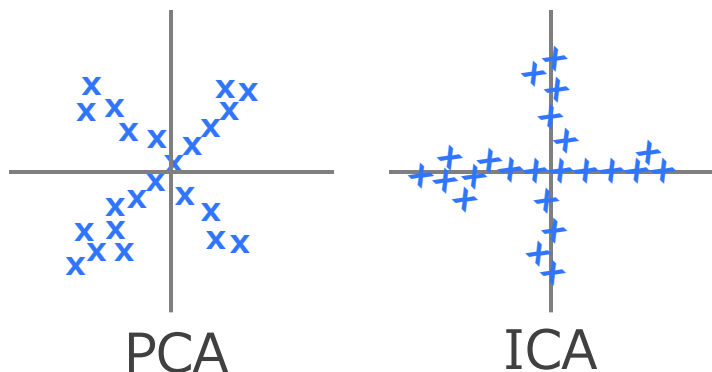
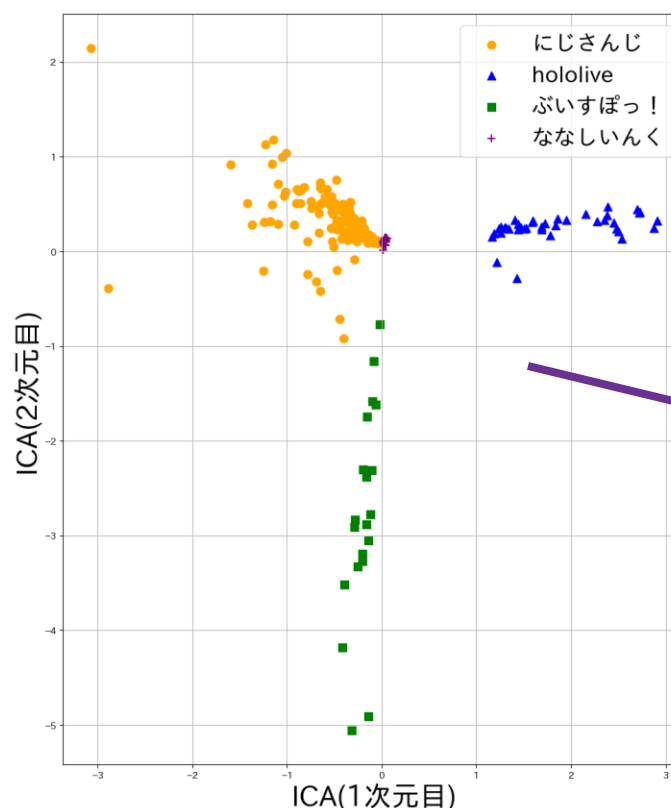


図: 独立成分分析によるVTuberベクトルの可視化



事務所毎に軸の  
取り方が異なる  
= 各事務所に  
よる異なった  
ファン獲得戦略  
が示唆される

[1] 『持橋大地. 統計的テキストモデル：言語へのベイズ的アプローチ. 確率と情報の科学. 甘利俊一, 麻生英樹, 伊庭幸人編. 岩波書店, 2025』を参考に発表者作成.



- ✓ 階層型クラスタリングによる樹形図上で、ファンが類似するタレントどうしが木構造上近くにいるほど良い分類とみなすための評価関数を考案する

図: タレント-視聴者間の2部グラフの例

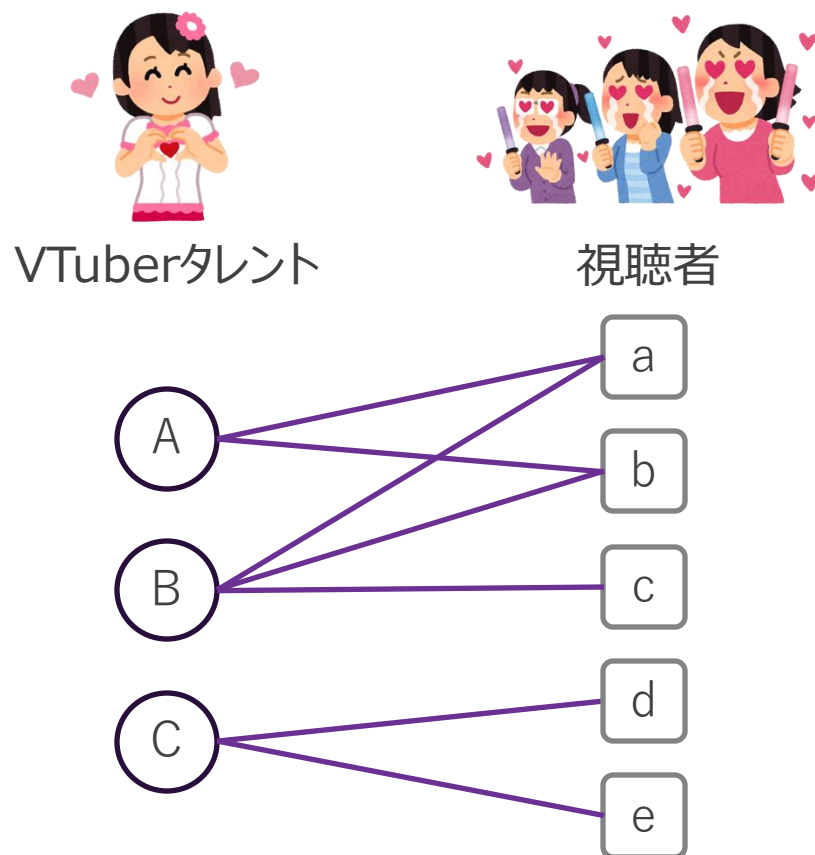
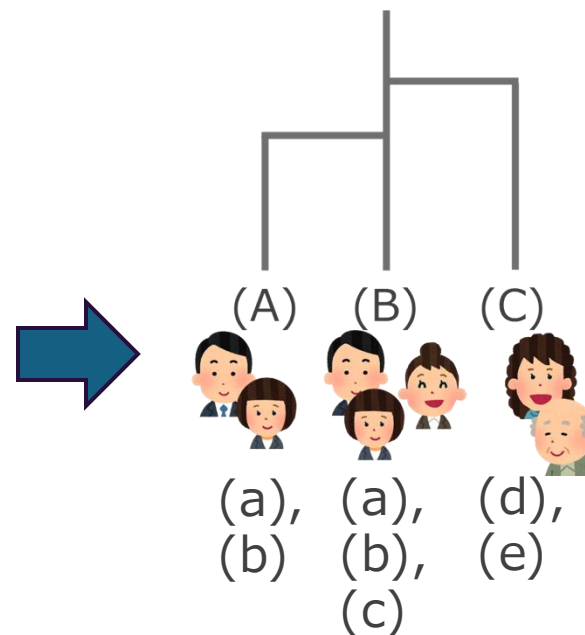


図: 階層型クラスタリングモデルによる樹形図の例



- タレントAとB:  
共通のファンが多い  
= 木構造上近くに位置することが望ましい
- タレントAとC:  
共通のファンがない  
= 木構造上は離れていることが望ましい

✓ 前頁の議論を数式的に表現し, 評価関数を作成した

**Algorithm 1** 階層型クラスタリングモデルのコストを求める関数

```

1: function EVALUATE_COST( $X, Z$ )
2:    $cost\_dict \leftarrow \{\}$ 
3:   for all  $i \leftarrow array[3, 4, 5, \dots, N]$  do
4:      $X' \leftarrow sample((X \text{ for } num\_sendee\_talents = i), 100)$ 
5:      $d \leftarrow 0$ 
6:     for all  $u \leftarrow X'$  do
7:       for all  $v_1, v_2 \leftarrow u$  do
8:          $d \leftarrow d + distance(Z, v_1, v_2)$ 
9:       end for
10:    end for
11:     $cost\_dict[i] \leftarrow d$ 
12:  end for
13:  return  $cost\_dict$ 
14: end function

```

Input: VTuberベクトル  $X$ ,  
階層型クラスタリングモデル  $Z$

(1) チャット者に着目し,  
送信先のタレント人数  $i$  ごとに層化抽出

(2) 各チャット者による送信先タレントの  
組み合わせ  $i \mathbf{C} 2$  通りで木構造上の距離を  
計算

(3) 距離を足し合わせる  
(評価時には適宜ペア数で正規化する)

- ✓ チャット回数によるチャット者の足切りと, コラボレーション動画の除外が, VTuberベクトルの作成において有用であることが確認された

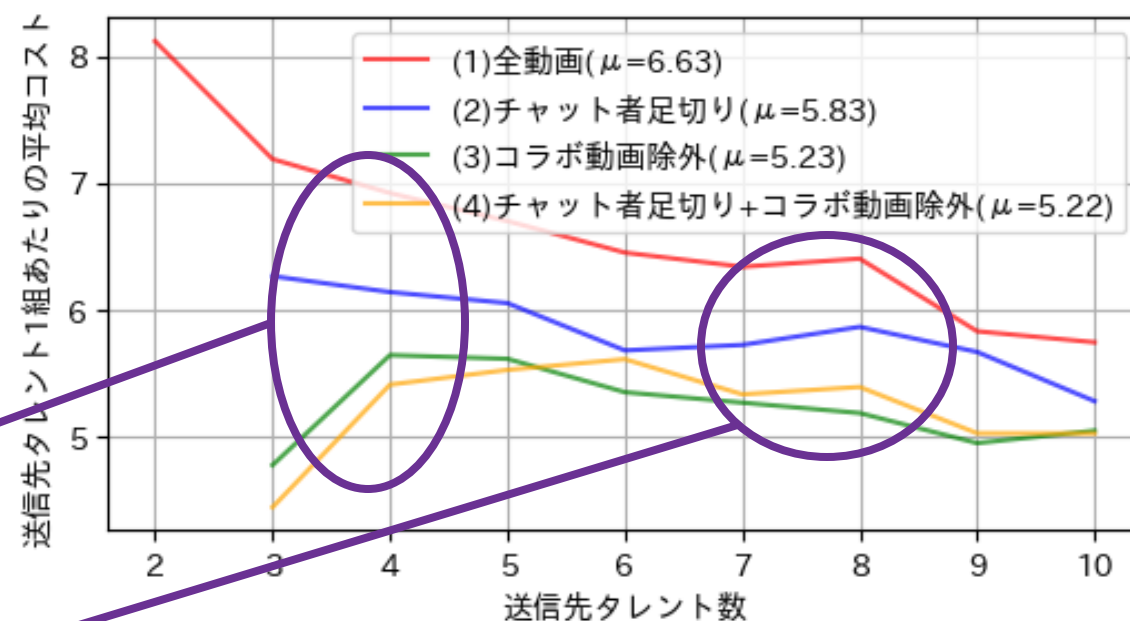
VTuberベクトルの評価対象作成条件:

- (1) 全動画[ベンチマーク]
- (2) チャット者をチャット回数30回以上に限定
- (3) コラボレーション動画<sup>[1]</sup>の除外
- (4) (2), (3)の同時適用

- 作成条件に応じて出力値が明確に変動

- 送信先タレント数の増加に従ってチャット回数の足切りは関数の出力値に影響しなくなるため, 条件の改良が必要

図: VTuberベクトル作成条件の評価結果<sup>[1]</sup>



- ✓ 内在次元数(ID)を評価し, 独立成分分析(ICA)を用いて各軸の解釈を図った
- ✓ 事務所ごとのファン獲得戦略の差や, 各軸の解釈可能な意味が示唆された

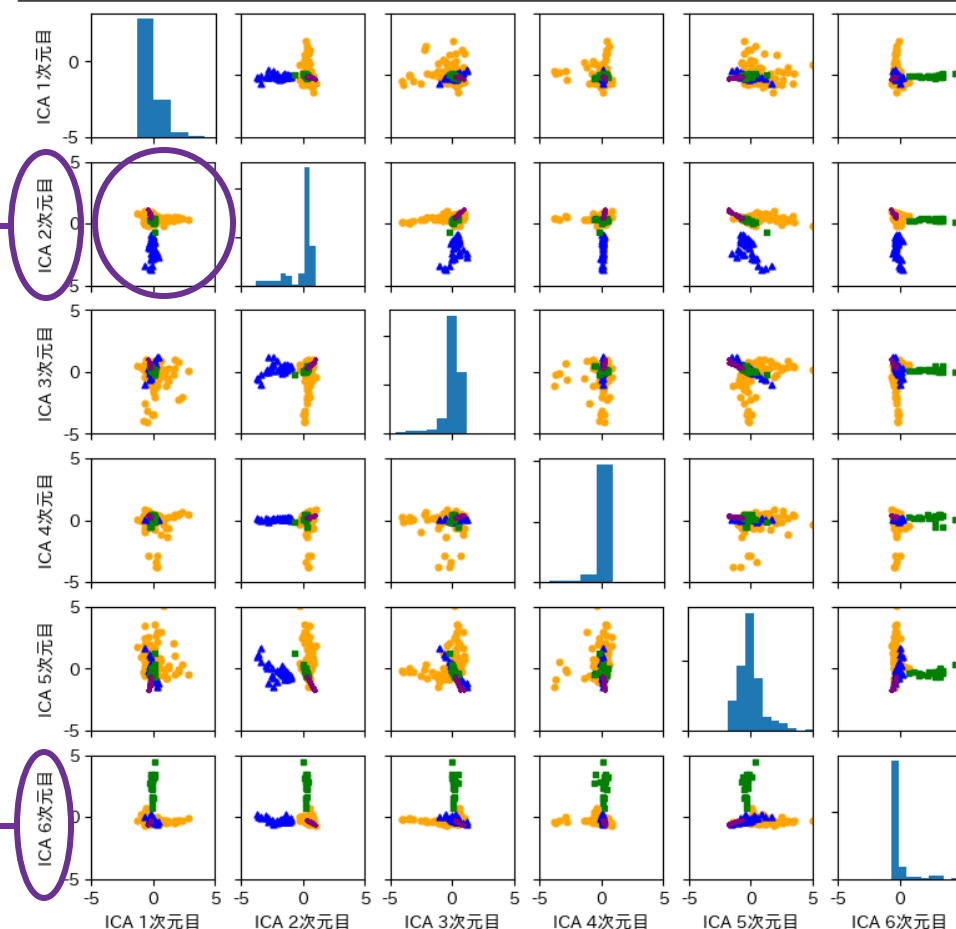
表: 内在次元数の評価

事務所(右図色)	ID数
にじさんじ(橙)	5.9
hololive(青)	6.1
ぶいすぽっ!(緑)	2.9
ななしいんく(紫)	2.3

各軸への想定解釈

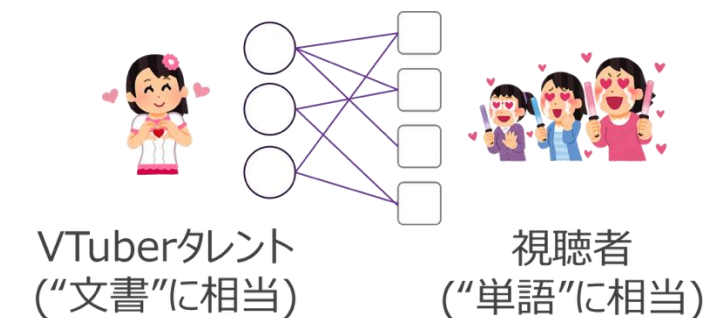
男性人気軸?  
[hololive 分散(大)]

ゲーム競技性軸?  
[ぶいすぽっ! 分散(大)]

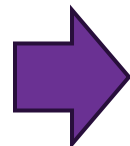
図: 独立成分分析のペアプロット<sup>[1]</sup>

[1] 当研究の予稿集より引用.

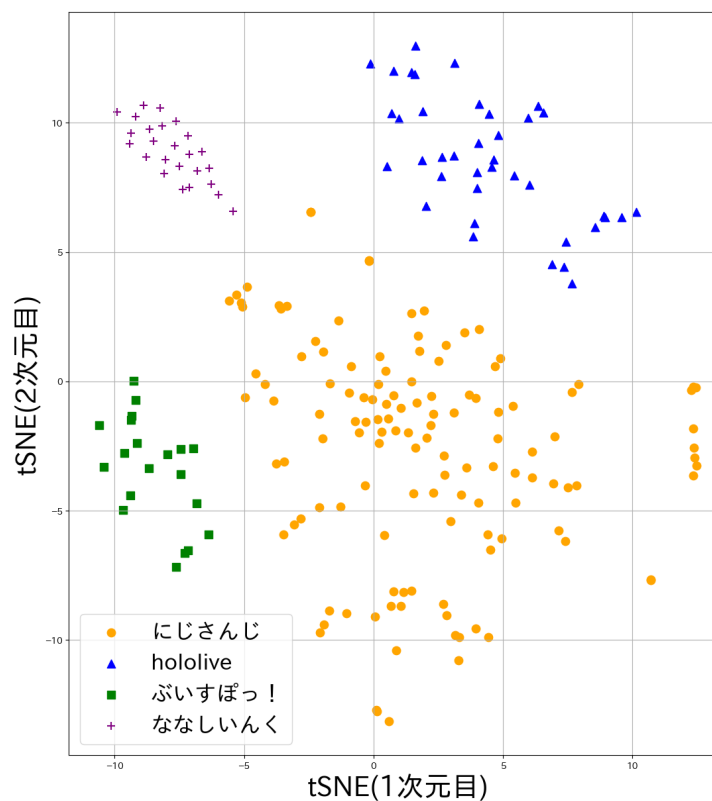
- ✓ VTuberと動画チャット者の2部グラフに着目し, VTuberベクトルを獲得した
- ✓ 自作評価関数や内在次元数を用いて, チャットデータのみでVTuberタレントのファン層把握やタレント事務所特性を明らかにできることが示された



タレント	チャット者	チャット数
A	a	3
A	b	10
...	...	...
B	b	4



### VTuberベクトルの獲得



### 階層型クラスタリングによる精度評価

→ チャット回数の足切りと  
コラボ動画の除外が有用

### 内在次元(ID)数による事務所別ファン特性の理解

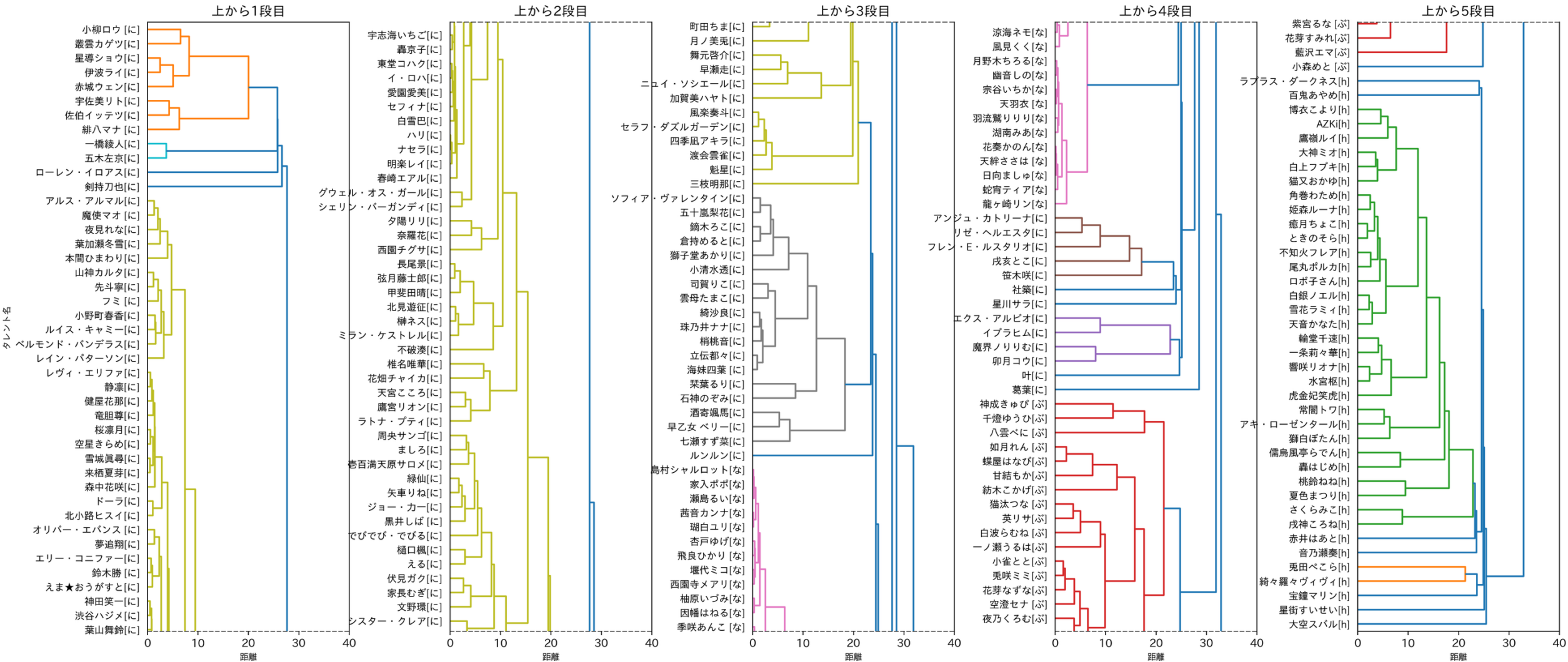
→ 男性人気軸など,  
意味づけが可能な軸が抽出  
された

### オンラインインフルエンサー全般への実験の拡大

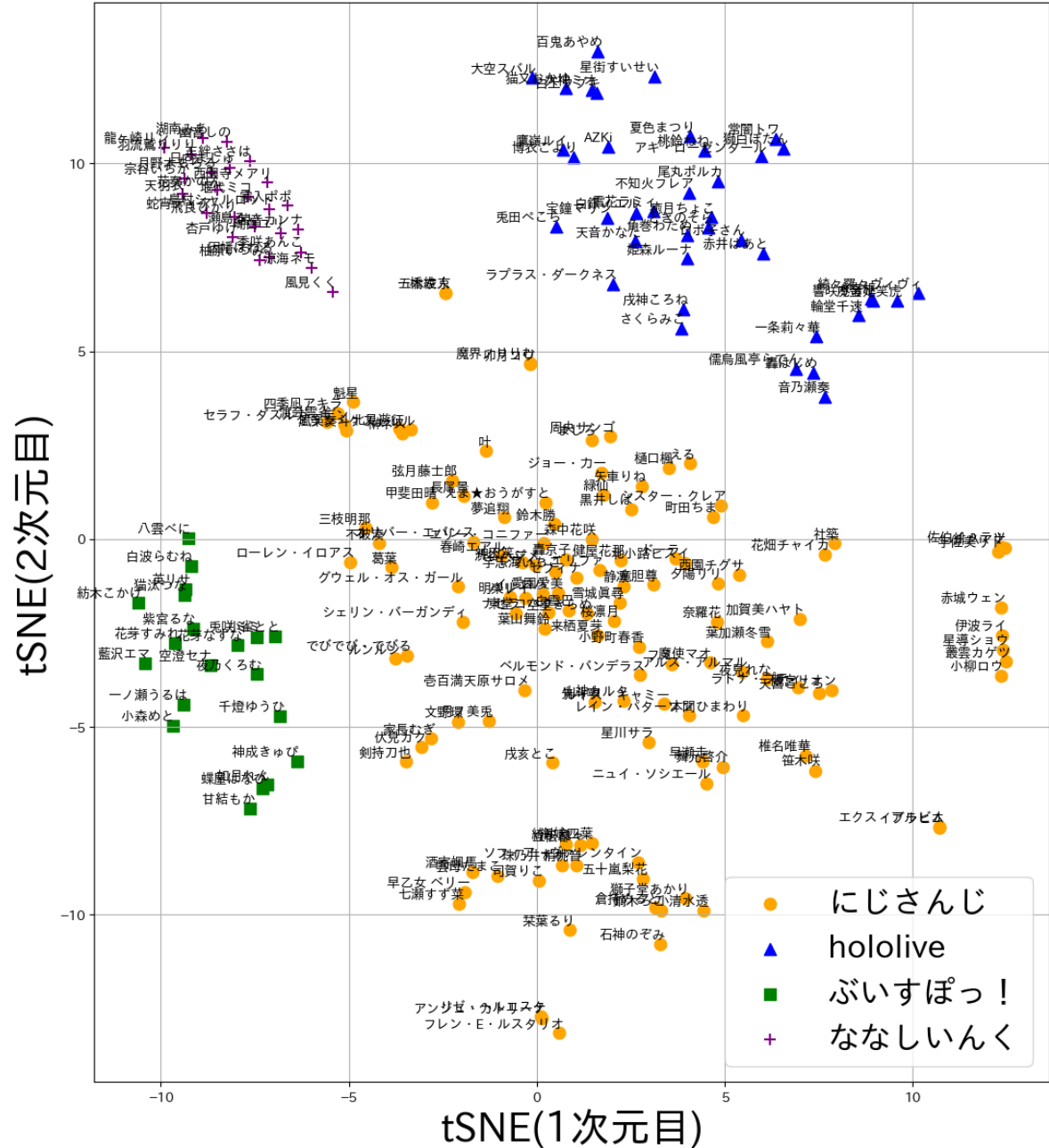
→ アバターを用いずに配信活動を行うYouTuberタレントはどのようなファンを持つのか？ VTuber市場という市場・分析の枠組みは適切なのか？

### 特定トピック(商材) に対するVTuberタレントの組合せシミュレーション

→ 「VTuberタレント 対 視聴者」の2部グラフを「動画 対 視聴者」にも応用し、動画タイトルの自然言語的埋め込み表現と組み合わせれば、タレント人気性とコンテンツ人気性を分離して分析できるのではないか？







**EoF**