

# 通信キャリアの検索クエリログに基づく 飲食チェーン店舗の関心層分析

井上大地<sup>1</sup> 森勇人<sup>2</sup> 吉井健敏<sup>1</sup> 木下定知<sup>1</sup> 持橋大地<sup>3</sup>

<sup>1</sup> 株式会社 D2C <sup>2</sup> 株式会社 NTT ドコモ

<sup>3</sup> 統計数理研究所

daichi.inoue@d2c.co.jp

## 概要

検索クエリログは、ユーザの自発的な関心を反映する行動データである一方、店舗やブランドごとの主要なターゲット層に関する情報は一般に公開されていない。本研究では、国内大手通信事業者が運営するポータルサイトの検索クエリログと、同一事業者が管理する会員基盤情報に基づく性年代情報のみを用い、チェーン業態内における店舗ごとの関心層を推定する手法を提案する。正規化自己相互情報量 (NPMI) を用いた分析により、検索クエリと性年代情報のみから擬似的な競合分析が可能であることを示す。

## 1 はじめに

店舗やブランドが想定する主要なターゲット層は、競争戦略やマーケティング施策の中核を成す情報である。しかし、こうした情報は企業戦略上の理由から外部に公開されないことが多く、第三者が客観的に把握することは困難である。

一方で、検索クエリはユーザが自発的に入力するテキストであり、特定の店舗や商品に対する関心、検討段階の情報要求、さらには購買前後の行動を反映すると考えられている。これまでにも、検索クエリログを用いたユーザ行動分析や関心推定に関する研究が数多く行われてきた [1][2]。

本研究では、検索クエリログと性年代情報のみを用いて、店舗ごとの関心層を性別・年代条件付きで推定することを目的とする。特に、同一チェーン業態内における店舗間の相対比較に着目し、正規化自己相互情報量 (NPMI) を用いて「他のチェーン店舗と比較して、特定の性年代層でどの程度検索されやすいか」を定量化する点に特徴がある。このような分析は、公開情報に依存しない擬似的な競合分析と

して、マーケターにとって有用な知見を提供すると期待される。

## 2 関連研究

検索クエリログを用いた分析は、情報検索および自然言語処理分野において長年研究されてきた。Jain ら [1] は、検索クエリログからエンティティを抽出する手法を提案し、クエリが持つ意味的情報の有用性を示している。

また、Qiao ら [3] は、検索クエリログを用いて競合キーワードを抽出し、検索広告の最適化に応用する手法を提案している。この研究は、クエリログが競合関係の把握に有効であることを示しており、本研究の問題設定と密接に関連する。

一方、ユーザ属性を考慮した検索行動分析としては、医療分野における PubMed の検索ログ解析が知られている [2][4]。これらの研究では、検索行動がユーザ属性や目的によって異なることが示されているが、商業店舗や競合分析への応用は十分に検討されていない。本研究は、これらの先行研究を踏まえつつ、検索クエリと性年代情報のみから店舗間の相対的な関心構造を推定する点に新規性を有する。

## 3 使用データ

### 3.1 検索クエリログ

検索クエリログは、国内大手通信事業者が運営するポータルサイトにおける検索履歴から取得した。対象期間は 2024 年度 (2024 年 4 月~2025 年 3 月) である。プライバシー保護の観点から、分析対象ユーザは約 100 万人を無作為抽出している。

検索クエリは空白で分割し、単語単位で処理する。例えば、「〇〇バーガー 期間限定 持ち帰り」は「〇〇バーガー」「期間限定」「持ち帰り」という語列

として扱われる。

### 3.2 業態およびチェーン店の定義

本研究では、外食産業における代表的な業態として、とんかつチェーン、牛丼チェーン、ハンバーガーチェーン、ファミリーレストランチェーン、喫茶店チェーンの5業態を分析対象として定義する。

各業態は複数のチェーン店から構成される。ここでのチェーン店とは、実店舗(例：〇〇バーガー 汐留店)を指すものではなく、検索クエリ中でユーザーが区別して言及する店名(〇〇バーガー 汐留店の場合、〇〇バーガーがチェーン店名に該当する)を指す。

同一業態に属する複数のチェーン店は、それぞれ独立した検索対象として扱われ、業態内・業態横断の比較が可能となる。具体的なチェーン店名は、機密情報保護の観点から論文中ではマスキングし、「とんかつチェーンA」「牛丼チェーンB」のように表記する。本研究で分析対象とした業態およびチェーン店数を、Appendixの表1に示す。

検索クエリには、同一チェーン店に対してもひらがな・カタカナ・アルファベット表記、略称や俗称など、多様な表記揺れが存在する。本研究では、チェーン店ごとに表記揺れ辞書を事前に定義し、これらを統合して1つのチェーン店として扱う。この処理により、ユーザーの入力表現の違いによる検索回数の過小評価を防ぎ、チェーン店単位での検索行動を安定して分析することが可能となる。また、検索回数の集計にあたっては、同一ユーザーによる完全に同一の検索ワードについては、同日では1回の検索として集計し、過度な繰り返し検索による影響を抑制している。

### 3.3 性年代情報

ユーザーの性年代情報は、同一事業者が管理する会員基盤情報に基づく。性別は男性・女性の2区分とし、年齢については10代、20代のような単純な10年刻みではなく、ライフスタイルの違いを考慮した区分を採用した。

具体的には、就学段階や就業状況の違いが検索行動に影響を与える可能性を考慮し、10代を中学生以下と高校生に分割し、大学生層を独立した区分として定義した。また、60代についても、就労継続者と引退後層の違いを考慮し、65歳以下と66歳以上に分割している。本研究で用いた年齢区分の一覧は

Appendixの表2に示す。

## 4 実験設計

本研究では、検索クエリログと性年代情報に基づき、チェーン店と性年代層との関連度を正規化自己相互情報量(NPMI)として算出し、得られた性年代別NPMIベクトルに主成分分析(PCA)を適用することで、チェーン店間の客層構造を比較・可視化する。

本研究における比較の特徴は、同一業態内での比較ではなく、定義したすべての業態を比較対象とする点にある。具体的には、ある分析対象店舗 $s$ (例：とんかつチェーンに属する特定店舗)について、ハンバーガーチェーンや牛丼チェーンなど、全業態に属するチェーン店を含む検索クエリ全体と比較を行う。

すなわち、「特定店舗 $s$ が検索されたか否か」と「特定の性年代 $d$ に属するか否か」の共起関係を、業態を限定せずに算出することで、当該店舗が事前に定義した飲食チェーン全体の中でどの性年代層において相対的に検索されやすいかを評価する。NPMIにより共起性に注目することで、性年代分布の偏りや、業種ごとの検索数の差による影響を抑えつつ、チェーン店と性年代層との相対的な関係性を業態横断的に捉えることができる。

### 4.1 NPMIの定義と解釈

店舗 $s$ と性年代属性 $d$ (性別と年齢区分の組)との関連度を測るため、正規化自己相互情報量(NPMI)を用いる[5]。

$P(s, d)$ を、属性 $d$ を持つユーザーが店舗 $s$ を含む検索クエリを行う割合、 $P(s)$ を店舗 $s$ を含む検索クエリが行われる割合、 $P(d)$ を属性 $d$ を持つユーザーが検索を行う割合とすると、PMIは次式で定義される。

$$\text{PMI}(s, d) = \log \frac{P(s, d)}{P(s)P(d)} \quad (1)$$

なお、ここでの割合は検索回数ではなく、検索クエリログ中に集計されたユーザー数に基づいて算出する。また、これらの割合は日単位や月単位での集計ではなく、対象期間全体を通じた累積値に基づいて計算している。

PMIは値のスケールが一定でないため、本研究では次式で定義されるNPMIを用いる。

$$\text{NPMI}(s, d) = \frac{\text{PMI}(s, d)}{-\log P(s, d)} \quad (2)$$

本研究における NPMI は、店舗の絶対的な人気度を表すものではない。ある性年代  $d$  において店舗  $s$  の NPMI が高い場合、「事前に定義した飲食チェーン全体と比較して、店舗  $s$  が当該性年代層において相対的に検索されやすい」ことを意味する。

このように、全業態を比較対象とすることで、業態の違いを含めた横断的な視点から、各店舗の関心層の偏りを捉えることが可能となる。

## 4.2 性年代別 NPMI ベクトルと PCA による次元削減

各店舗  $s$  について、性別（男性・女性）と年齢区分（Appendix の表 2 に示す区分）の組合せからなる 20 個の性年代条件  $d_1, d_2, \dots, d_{20}$  を定義する。各条件に対し、式 (2) により NPMI を算出し、店舗  $s$  は次式で表される 20 次元の特徴ベクトルとして表現される。

$$\mathbf{v}_s = (\text{NPMI}(s, d_1), \dots, \text{NPMI}(s, d_{20})) \quad (3)$$

このように定義される NPMI ベクトルは、各店舗がどの性年代層に相対的に検索されやすいかを多角的に表現している。しかし、20 次元空間のままでは店舗間の関係性を直感的に把握することが難しい。

そこで本研究では、客層構成が類似したチェーン店同士を低次元空間上で可視化し、業態を横断した関心構造の類似性を把握するため、主成分分析 (PCA) を適用する。PCA により、性年代別の検索傾向に基づく主要な変動軸を抽出することで、「どのような客層に支持されている店舗同士が近いか」を視覚的に捉えることが可能となる。本研究では、第 1 主成分および第 2 主成分を用いて店舗を 2 次元空間上に配置し、その分布を分析対象とする。

## 5 実験結果

本章では、性別・年代ごとに算出した NPMI に基づく 20 次元特徴ベクトル (式 3) を用い、店舗間の関心構造を分析する。

PCA の結果、第 1 主成分の寄与率は 0.545、第 2 主成分の寄与率は 0.236 であり、この 2 軸で全体の約 78.1% を説明できることが確認された。このことから、性年代別 NPMI に基づく店舗間の差異は、主に第 1 および第 2 主成分によって特徴づけられてい

るといえる。各店舗について算出した NPMI ベクトルに対し、主成分分析を適用した。図 1 は、第 1 主成分および第 2 主成分による 2 次元空間上での店舗分布を示す。

### 5.1 主成分の解釈

第 1 主成分および第 2 主成分の寄与度の詳細は、Appendix(表 3) に示す。

第 1 主成分では、正の寄与が男性 30–60 代を中心とした男性層に集中しており、一方で負の寄与は女性の若年層から中年層に集中している。寄与度の符号が性別によって明確に分かれていることから、第 1 主成分は「女性寄り～男性寄り」という性別による客層の違いを主に表す軸であると解釈できる。すなわち、第 1 主成分の値が大きいほど男性志向が強く、小さいほど女性志向が強いチェーン店であることを示している。

一方、第 2 主成分では、正の寄与が 60 代以上の高年齢層に集中しており、負の寄与は学生層や 20 代以下の若年層に多く見られる。このことから、第 2 主成分は「若年層志向～高年齢層志向」という年代による客層構造を表す軸であると解釈できる。第 2 主成分の値が大きいほど高年齢層に、小さいほど若年層に相対的に検索されやすい傾向を持つといえる。

以上より、本研究で得られた PCA の第 1 主成分は主に性別を、第 2 主成分は主に年代を反映していると解釈できる。この低次元空間上では、業態が異なるチェーン店であっても、性別および年代の両面で客層構造が類似した店舗同士が近接して配置される。この結果は、検索クエリと性年代情報のみを用いて、チェーン店の客層構造に基づく擬似的な競合関係を把握できる可能性を示唆している。

### 5.2 チェーン店別 NPMI の解釈

次に、図 1 に示した PCA の結果と、各チェーン店における性年代別 NPMI の分布を対応付けて考察する。本研究では、PCA による業態横断的な全体構造の把握と、チェーン店ごとの NPMI 分布の詳細分析を組み合わせることで、客層構造の違いを多角的に捉えることを目的としている。

前節の寄与度分析より、第 1 主成分は主に性別の違いを表す軸であり、値が大きいほど男性寄り、小さいほど女性寄りの傾向を示す。また、第 2 主成分は主に年代の違いを表す軸であり、値が大きいほど

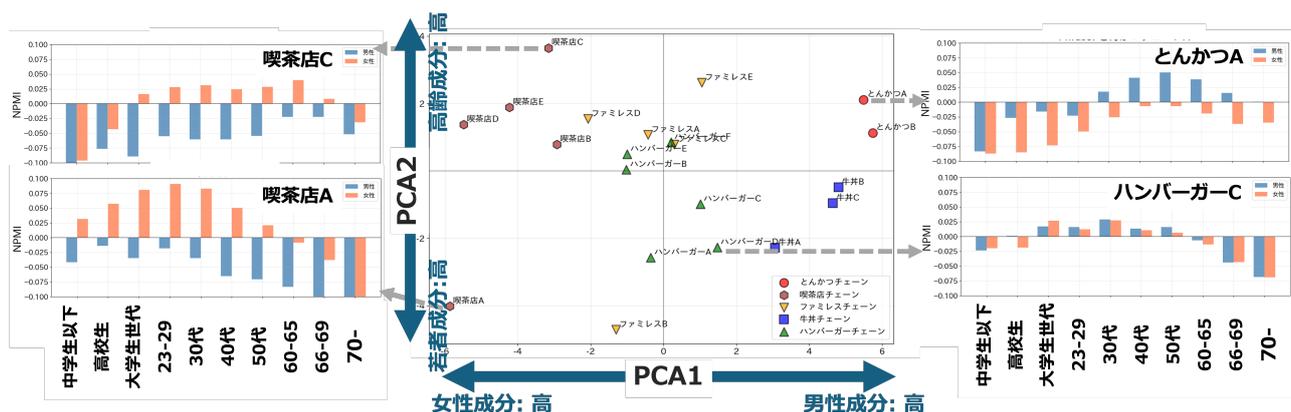


図 1: 性別・年代別 NPMI ベクトルに対する PCA とチェーン店ごとの NPMI

高年齢層寄り、小さいほど若年層寄りの傾向を表している。この解釈に基づき、各チェーン店の配置と NPMI 分布について考察する。

まず、とんかつ業態に属する店舗（とんかつチェーン A）は、PCA 空間上で第 1 主成分の正方向に位置しており、男性寄りの特徴を示す。性年代別 NPMI を見ると、男性 30 代から 50 代で相対的に高い値を示しており、当該層において検索されやすい傾向が確認できる。この結果は、該当企業へのインタビューで述べられたとんかつチェーン A の主要利用層（30～50 代男性）と整合的であり、NPMI が実際の利用年代を適切に捉えていることを示している。すなわち、検索クエリに基づく NPMI の結果は、実際の利用年代との乖離が小さいことを示している。同様に、牛井やハンバーガーといった他の外食業態に属するチェーンについても、PCA 空間上では相対的に男性寄り、かつ若年層から中年層寄りの領域に配置される傾向が確認できる。これは、これらの業態が比較的短時間・低価格での利用が想定されやすく、若年層や働き盛りの男性層から検索・関心を集めやすいという一般的な認識と整合的である。このように、本手法により、異なる業態間での客層構造の違いを同一の尺度上で比較できている点は重要である。

次に、喫茶店業態に属する店舗群に着目する。喫茶店業態の各チェーンは、PCA 空間上では女性寄りの領域に分布する傾向が見られるが、同一業態内であっても、性年代別 NPMI の分布には顕著な違いが存在する。例えば、あるチェーンでは女性の若年層から中年層で高い NPMI を示す一方、別のチェーンでは高年齢層で相対的に高い値を示すなど、関心を集める層が大きく異なっている。なお、喫茶店

業態に属する複数チェーン（喫茶店チェーン A-E）について、性年代別 NPMI を直接比較した結果は Appendix の表 3 に示す。これにより、同一業態内における客層構造の差異をより詳細に確認できる。

以上より、PCA による低次元可視化は、業態間の客層構造の違いを俯瞰するだけでなく、同一業態内におけるチェーン店ごとの関心層の差異を把握する上でも有効である。また、検索クエリと性年代情報のみを用いた分析であっても、一般に認知されているブランドイメージや利用者像と整合的な結果が得られる点は、本手法の妥当性と実用性を示している。

## 6 おわりに

本研究では、ポータルサイトの検索クエリと性年代情報を用い、飲食業態内における店舗ごとの関心層を分析した。

分析の結果、第 1 主成分は性別、第 2 主成分は年代構成を主に反映した軸として解釈できることが確認された。PCA 空間上では、業態を超えて検索傾向が類似したチェーン店が近接して配置されるとともに、同一業態内であっても関心層が異なる可能性が示された。また、性年代別 NPMI の分布は、企業インタビューから明らかとなった利用者層と概ね整合的であり、検索クエリと性年代情報のみから客層構造を一定程度捉えられる可能性が示唆された。

本研究の意義は、購買履歴や来店データを用いずに、検索クエリからチェーン店の客層構造や擬似的な競合関係を把握できる点にある。今後の課題として、検索クエリの時間的変化を考慮した分析や、他の属性情報を組み合わせることで、より多面的な客層分析への拡張などが考えられる。

## 参考文献

- [1] Alpa Jain and Marco Pennacchiotti. Open entity extraction from web search query logs. In **Proceedings of Coling 2010**, pp. 510–518, 2010.
- [2] Illhoi Yoo and Abu Saleh Mohammad Mosa. Analysis of pubmed user sessions using a full-day pubmed query log. **JMIR Medical Informatics**, Vol. 3, No. 3, p. e25, 2015.
- [3] Dandan Qiao, Jin Zhang, Qiang Wei, and Guoqing Chen. Finding competitive keywords from query logs to enhance search engine advertising. **Information & Management**, Vol. 54, No. 4, pp. 531–543, 2017.
- [4] Rezarta Islamaj Dogan, G Craig Murray, Aurélie Névéol, and Zhiyong Lu. Understanding pubmed® user search behavior through log analysis. **Database**, Vol. 2009, p. bap018, 2009.
- [5] Gerlof Bouma. Normalized (pointwise) mutual information in collocation extraction. In **Proceedings of the Biennial GSCL Conference**, pp. 31–40, 2009.

## A 補足資料

表 1: 分析対象とした業態とチェーン店数

業態	チェーン店数
とんかつ	2
牛丼	3
ハンバーガー	6
ファミレス	5
喫茶店	5
合計	21

表 2: 年齢区分

年齢区分
10代 (中学生以下)
10代 (高校生)
10-20代 (大学生)
20代 (23歳以上)
30代
40代
50代
60代 (65歳以下)
60代 (66歳以上)
70代以上

表 3: 第 1 主成分および第 2 主成分に対する性年代別寄与度 (上位・下位 5 件)

性年代区分	第 1 主成分	第 2 主成分
男性_50代	0.295	—
男性_60代 (65歳以下)	0.294	—
男性_40代	0.290	—
男性_60代 (66歳以上)	0.275	—
男性_30代	0.264	—
女性_40代	-0.291	—
女性_20代 (23歳以上)	-0.282	—
女性_30代	-0.281	—
女性_10-20代 (大学生)	-0.261	—
女性_50代	-0.225	—
女性_60代 (66歳以上)	—	0.348
女性_70代以上	—	0.342
女性_60代 (65歳以下)	—	0.286
男性_70代以上	—	0.214
女性_50代	—	0.179
男性_10-20代 (大学生)	—	-0.306
男性_10代 (中学生以下)	—	-0.302
男性_20代 (23歳以上)	—	-0.302
女性_10代 (中学生以下)	—	-0.292
男性_10代 (高校生)	—	-0.270

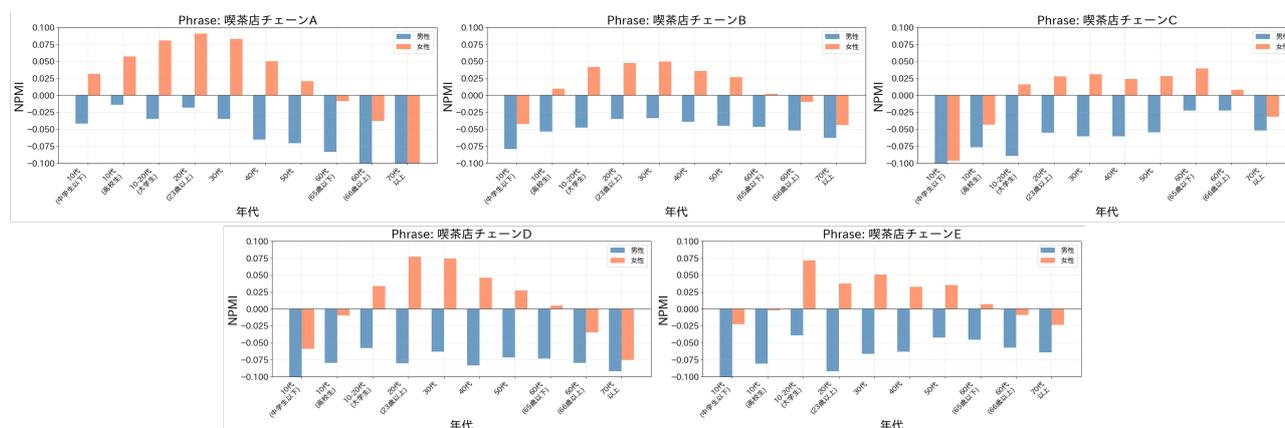


図 2: 喫茶店チェーン A-E における性年代別 NPMI の比較

同一業態であっても女性若年層に強く支持されるチェーンや高齢層で相対的に高い値を示すチェーンが存在し、関心を集める性年代層には大きな差異が見られる。この結果は、業態分類のみでは捉えられないチェーン店ごとの客層構造の違いを検索クエリから把握できることを示している。