

# 飲食業界の不満要因分析における情報エントロピーを用いた 新たな分析手法の提案

A Novel Analytical Method Using Information Entropy for Analyzing Dissatisfaction Factors  
in the Food Service Industry

張 宇帝 白井康之

Zhang, Yudi Shirai, Yasuyuki

大東文化大学 経営学部

Faculty of Business Administration, Daito Bunka University

概要：本研究は、飲食業界における不満要因を分析し、業態間の比較を可能とするため、情報エントロピーを用いた新たな分析手法を提案する。NIIの大規模不満データから飲食関連26万件を対象に、不満語の分布をカテゴリ別に分析し、語の広がりや特異性を定量的に評価した。その結果、共通する不満と業態特有の不満を識別することが可能となり、企業におけるサービス改善や商品戦略への実用的示唆が得られた。

Abstract: This study proposes a novel analytical method using information entropy to analyze factors of customer dissatisfaction in the food service industry and enable cross-category comparisons. Focusing on 260,000 food-related entries from NII's large-scale dissatisfaction dataset, we analyzed the distribution of dissatisfaction-related terms by category and quantitatively evaluated the breadth and specificity of these terms. As a result, we were able to identify both common and category-specific sources of dissatisfaction, providing practical insights for service improvement and product strategy development in the industry.

## 1. はじめに

近年、飲食業界では顧客満足度の向上と共に、インターネット上に投稿される不満情報の重要性が増している。特に、顧客の「不満」に着目したデータ分析は、サービス改善や商品開発に直接的な示唆を与えるため、企業にとって極めて重要である。しかし、膨大な不満データから効率的かつ体系的に不満要因を抽出し、さらにそれらを業態や場面ごとに比較・可視化することは、従来の手法では十分に達成されていない。

本研究では、NII 国立情報学研究所が公開した大規模不満データセットのうち、飲食業界に関連する約 26 万件のデータを対象とし、不満要因の抽出およびカテゴリごとの特徴分析を行う。特に、各カテゴリ（ファーストフード、レストラン、カフェ・喫茶、居酒屋、ファミリーレストラン）ごとに出現する不満語の「情報エントロピー」を計算することで、不満要因の共通性や独自性を新たな視点から明らかにする。

本研究の特徴は、情報エントロピーを用いて、多数の不満語間の関連や広がりやを定量的に評価する新しい分析手法を提案した点にある。本手法により、飲食業界の各業態が抱える典型的な不満要因や、その場面固有の問題点を効率的に抽出・可視化することが可能となり、今後のサービス改善に資する具体的な知見を提供することを目指す。

## 2. 関連研究

### 2.1 商品やサービスに対する不満情報を用いた既存研究

不満情報の活用に関する研究では、澤田ら [1] が肯定的特性抽出、吉川ら [2] が不満解決型推薦を提案している。テキスト分析では、末廣ら [3] がBoWによる素性化、村本ら [4] が語義曖昧性の解消を試みた。長谷川ら [5] は不満調査データのクラスタリングと可視化を行っている。

本研究は、これらを踏まえつつ、飲食業界の不満情報に対し、TF-IDF とエントロピーを用いた定量的な可視化を行う点で新規性がある。

### 2.2 不満情報分析の一般的な手法

従来、インターネット上の不満情報の分析には、テキストマイニングや自然言語処理 (NLP) を用いた手法が広く利用されてきた。具体的には、形態素解析を通じて重要なキーワードを抽出し、それらの頻度や共起関係を分析することで、顧客の不満要因を特定する方法が一般的である。また、近年では LDA (潜在的ディリクレ配分法) などのトピックモデルによる分析手法も用いられている。LDA は、不満情報を複数の「トピック」に分類することで、顧客の意見を効率的に要約し、重要な問題点を把握しやすくする手法として注目されている。

しかし、こうした頻度や共起に基づく分析は、不満語自体の出現傾向や一般的な特性を捉えることには有益である一方、特定の場面や業態間の違

いを数量的・客観的に評価することが難しい場合が多い。

## 2.3 情報エントロピーの応用と不満分析への展開可能性

情報エントロピー (entropy) は情報理論の重要な概念であり、テキスト分析分野においても、語彙の多様性や情報の不確実性を計測するために利用されることがある。

情報エントロピーを用いて単語の分布やカテゴリ間の違いを評価する研究も存在し、語彙分布の情報エントロピーを測定することで、特定の状況における単語の偏りや特徴を明らかにした事例が報告されている。本研究では、実際に飲食業界に特化して不満語のエントロピーを計算し、その結果を業態ごとの特徴抽出や不満語同士の関連性評価に応用し、業態別の不満語の特徴を明らかにする。

## 2.4 本研究の位置付け

以上の先行研究を踏まえ、本研究の新規性・独自性は以下の2点に集約される。

- 大規模な飲食業界不満データ (約 26 万件) を用いて、業態ごとに出現する不満語の情報エントロピーを測定し、不満語間の関連性を新たな視点から数的に分析・評価する。
- エントロピー分析結果をもとに、業態間の比較や特徴的な不満要因を特定し、実務的なサービス改善に役立つ具体的な知見を提供する。

## 3. データセットと分析対象

### 3.1 使用したデータセット

本研究では、国立情報学研究所 (NII) が株式会社 Insight Tech から提供を受けて研究者に提供している『不満調査データセット』 (2015 年~2017 年) を使用した。このデータには、日本全国の消費者が実際に投稿した商品・サービスに関する不満情報が集約されており、総件数は約 600 万件以上に及ぶ大規模なものである。各データは短文形式で提供されており、投稿内容には「不満テキスト」、投稿された業界・業態を示す「カテゴリ情報」などが含まれている。

### 3.2 分析対象カテゴリの選定

本研究の主な分析対象は、NII データセットの中でも飲食業界に関連する以下の5つの業態 (カテゴリ) に限定した: ファーストフード (fast food), レストラン (restaurant), カフェ・喫茶 (cafe), 居酒屋 (izakaya), ファミリーレストラン (family restaurant)。これら5カテゴリに該当するデータを抽出した結果、分析対象となったデータ件数は合計約 26 万件となった。これらのカテゴリを選定した理由は、各業態で顧客が抱える不満要因に差異

があり、比較分析に適していると考えられたためである。

## 4. 分析手法

本研究では、飲食業界における不満要因の分析において、「情報エントロピー」および「TF-IDF」を用いた分析アプローチを提案する。本章では、その具体的な分析手順について述べる。

### 4.1 不満語抽出と辞書の作成

まず、不満要因の分析の前段階として、McCab を用いて、前章で述べた不満テキストデータの形態素解析を実施した。解析結果のうち、不満を示す可能性が高い「名詞」と「形容詞」を中心に抽出し、各カテゴリの頻度集計を行った。

抽出した単語のうち、一定以上の頻度で出現する語彙を選定し、これらを「不満語」と定義して、分析に用いる辞書 (不満語辞書) を作成した。

### 4.2 TF-IDF スコアによる語の重要度算出

抽出した不満語について、各カテゴリにおける特徴的な語彙を浮かび上がらせるために、TF-IDF スコアを計算した。TF-IDF は、語の出現頻度 (TF) と、データ全体に対する逆文書頻度 (IDF) を組み合わせた指標であり、以下の式で定義される:

$$TF-IDF(w, c) = TF(w, c) \times \log \frac{N}{n_w}$$

ここで、

- $TF(w, c)$ : カテゴリ  $c$  における語  $w$  の頻度
- $N$ : 全カテゴリの数
- $n_w$ : 語  $w$  が出現するカテゴリの数

このスコアを用いることで、各カテゴリにおいて相対的に特徴的な不満語を抽出することが可能となり、後述するエントロピーとの対比にも有用である。

### 4.3 情報エントロピー計算方法

本研究では、以下の数式を用いて各不満語のエントロピーを計算した。カテゴリ集合を  $C = \{c_1, c_2, \dots, c_n\}$ 、不満語を  $w$ 、カテゴリ  $c_i$  における不満語  $w$  の出現確率を  $p(w|c_i)$  とするとき、不満語  $w$  のエントロピー  $H(w)$  は次のように定義される。

$$H(w) = - \sum_{i=1}^n p(w|c_i) \log p(w|c_i)$$

この式により、ある単語が特定の業態に偏って出現するのか、それとも幅広い業態で一般的に見られる不満語なのかを定量的に評価できる。

### 4.4 情報エントロピー分析による不満要因の特定と比較方法

本研究では、上記エントロピー計算の結果を利用して、不満要因を以下のように分析した:

- (1) 低エントロピー語：特定カテゴリに集中して出現し、業態特有の不満要因であると判断。
- (2) 高エントロピー語：多くのカテゴリで幅広く出現するため、業態を越えて共通する不満要因として判断。

この分類により、飲食業界における業態ごとの固有の不満要因と、業態を超えて普遍的な不満要因を明確に区別して分析することが可能となる。

## 5. 実験・結果

本節では、前章で述べた分析手法に基づいて実施した実験結果を報告する。主に (1) 不満語の出現傾向、(2) TF-IDF スコアによるカテゴリ特徴語の抽出、(3) エントロピー分布による語彙の汎用性・特異性の分析、(4) 業態間比較の考察の 4 点について述べる。

### 5.1 不満語の出現傾向

まず、全体の不満テキストから頻出語を集計した結果、共通して多く出現した語には「店員」「値段」「時間」など、サービスや店舗運営に関する語が多く含まれていた。

一方で、カテゴリごとに異なる語が上位に現れる傾向もあり、以下のような特徴が見られた：

- ファーストフード：「注文」「サイズ」「違う」
- レストラン：「禁煙」「店舗」「悪い」
- 居酒屋：「ビール」「ドリンク」「遅い」
- カフェ・喫茶：「アイス」「サイズ」「小さい」
- ファミリーレストラン：「メニュー」「子連れ」「待ち時間」

これらの語は、それぞれの業態の利用目的や利用者層の違いを反映していると考えられる。

### 5.2 TF-IDF スコアによるカテゴリ特徴語の抽出

次に、TF-IDF スコアを用いて、各カテゴリに特有の不満語を抽出した。TF-IDF スコアが高い語は、当該カテゴリで相対的に目立つ不満である。

代表的な例を表 1 に示す。

表 1：カテゴリ別高 TF-IDF 不満語の例（上位 5 語）

カテゴリ	高TF-IDFスコア語例
ファーストフード	ポテト、メニュー、マック、マクドナルド、セット
レストラン	メニュー、注文、店員、寿司、子供
カフェ・喫茶	コーヒー、カフェ、メニュー、高い、店員
居酒屋	居酒屋、注文、料理、店員、放題
ファミリーレストラン	メニュー、バー、ドリンク、もっと、少し

例えば、ファーストフードでは「ポテト」や「マクドナルド」など固有ブランドや商品名が強

く現れ、居酒屋では「放題」など提供形式に関する語が特徴的に出現している。これらの語は当該カテゴリに特有の不満傾向を示しており、エントロピーと併せて分析することで、より深い業態別特徴の把握が可能となる。

### 5.3 情報エントロピーによる不満語の分布評価

不満語ごとにカテゴリ別の出現割合を用いて情報エントロピーを算出した結果、語彙の分布に応じて以下のような傾向が確認された。

表 2 は、エントロピーの高低に基づいて抽出した代表的な不満語である。

表 2：代表的な不満語

分類	不満語	エントロピー値
高エントロピー語 (共通不満)	メニュー、店員、注文、少し、もっと	1.436 ~ 1.117
低エントロピー語 (業態特有)	ハッピー、おもちゃ、チョコ、フィギュア、キッズ	0.135 ~ 0.200

たとえば「メニュー」「店員」「注文」などは、どの業態でも頻繁に登場する不満語であり、飲食業界全体に共通する課題に関連していると考えられる。一方で「ハッピー」や「おもちゃ」などはファミリーレストランに特有の語であり、子ども向けサービスや景品に対する不満が集中していることが示唆される。

このように、エントロピー分析を通じて、不満語のカテゴリ依存性を定量的に把握することが可能であり、固有性・汎用性を兼ねた不満要因の把握に有効であると確認された。

### 5.4 業態間比較による傾向分析

また、単語単位だけでなく、不満語の組み合わせ（共起語ペア）に着目することで、より文脈的な不満傾向の把握を試みた。具体的には、各単語を「人物」「食べ物」「店名」、「概念」、「飲み物」、「サービス」、「道具」、「行事」の概念カテゴリに分割し、不満傾向における概念間の関連を把握した。

このうち、「人物」に関連する語のみを含む組み合わせを抽出し、カテゴリ別出現割合およびエントロピー値を分析したものを表 3 に示す。

表 3 に示すように、「大人-子供」ペアはファーストフード、居酒屋、ファミリーレストランなど幅広い業態で多く見られ、エントロピー値も 2.6147 と高いことから、業態を越えて共通する不満傾向であると考えられる。一方、「大人-子連れ」や「大人-女性」などのペアは、レストランやファミリーレストランに集中しており、特定の業態に固有の不満を反映していると解釈できる。

また表 4 では、「チョコ」「セット」「天ぷら」などの食べ物を表す語と、「注文」「予約」「会

計」などのサービスを表す語の共起ペアを抽出し、業態別出現傾向とエントロピーを分析した。

「セット-注文」「セット-テイクアウト」などはファーストフードやレストラン、居酒屋で広く見られ、エントロピーも高め（例：1.97, 0.18）で、共通的な不満要因と判断される。一方、「チョコ割引」や「天ぷら-テイクアウト」は特定業態でのみ出現し、エントロピーが非常に低い（0.0185, 0.0348）ことから、業態固有の不満を示していると考えられる。

また、「天ぷら-注文」や「天ぷら-予約」は居酒屋やレストランで多く見られ、料理ジャンルと提供形式の関係性が不満に影響することが示唆された。

このように、共起語エントロピーを用いることで、商品とサービス場面の組み合わせに対する不満構造をより具体的に把握できる。

表3 人物-人物ペアのカテゴリ別エントロピー値

word1	Type1	word2	Type2	ファースト フード	レストラン	カフェ・ 喫茶	居酒屋	ファミリー レストラン	Entropy
大人	人物	子供	人物	39.3850	46.1250	36.6410	39.5830	47.4360	2.6147
大人	人物	1人	人物	0.9230	4.2500	2.2900	10.4170	4.4870	0.9217
大人	人物	子ども	人物	9.2310	15.3750	5.3440	8.3330	10.2560	1.5942
大人	人物	女性	人物	0.6150	1.8750	1.5270	2.0830	0.6410	0.4079
大人	人物	2人	人物	0.9230	4.3750	5.3440	8.3330	6.4100	1.0385
大人	人物	子連れ	人物	0.9230	3.2500	6.1070	6.2500	1.9230	0.8290
大人	人物	男性	人物	1.2310	1.1250	2.2900	2.0830	1.2820	0.4726
大人	人物	スタッフ	人物	0.3080	0.3750	2.2900	3.1250	0.6410	0.3836
大人	人物	友達	人物	0.6150	0.5000	0.7630	2.0830	0.6410	0.3002

表4 食べ物-サービスのカテゴリ別エントロピー値

word1	Type1	word2	Type2	ファースト フード	レストラン	カフェ・ 喫茶	居酒屋	ファミリー レストラン	Entropy
チョコ	食べ物	注文	サービス	3.6310	8.5270	6.5730	0.0000	14.2860	1.1357
チョコ	食べ物	テイクアウト	サービス	0.5190	0.0000	1.1740	0.0000	0.0000	0.1146
チョコ	食べ物	割引	サービス	0.2070	0.0000	0.0000	0.0000	0.0000	0.0185
チョコ	食べ物	会計	サービス	0.2070	0.0000	0.7040	0.0000	0.0000	0.0688
チョコ	食べ物	追加	サービス	0.2070	0.0000	1.1740	0.0000	0.0000	0.0938
チョコ	食べ物	予約	サービス	0.0000	0.7750	0.4690	0.0000	0.0000	0.0907
セット	食べ物	割引	サービス	1.4200	0.6670	2.1300	1.0870	0.2730	0.3478
セット	食べ物	注文	サービス	11.3580	15.8880	15.5390	18.4780	9.8360	1.9747
セット	食べ物	テイクアウト	サービス	1.8490	0.3640	0.3760	0.0000	0.2730	0.1895
セット	食べ物	追加	サービス	0.7660	1.2730	1.0030	2.1740	1.0930	0.3919
セット	食べ物	会計	サービス	0.9340	1.2730	1.7540	4.3480	0.8200	0.4990
セット	食べ物	予約	サービス	0.2990	0.5460	0.5010	5.4350	0.5460	0.3738
セット	食べ物	デリバリー	サービス	0.0560	0.0000	0.0000	0.0000	0.2730	0.0293
天ぷら	食べ物	予約	サービス	0.0000	0.4450	0.0000	2.3260	0.0000	0.1610
天ぷら	食べ物	注文	サービス	6.1320	8.6860	0.0000	9.3020	2.7780	1.0155
天ぷら	食べ物	割引	サービス	0.9430	0.2230	0.0000	0.0000	0.0000	0.0831
天ぷら	食べ物	追加	サービス	0.4720	0.8910	0.0000	0.0000	0.0000	0.0971
天ぷら	食べ物	会計	サービス	1.4150	1.5590	0.0000	0.0000	0.0000	0.1805
天ぷら	食べ物	テイクアウト	サービス	0.0000	0.4450	0.0000	0.0000	0.0000	0.0348

## 6. 考察と今後の課題

本研究では、飲食業界における不満情報の比較分析を目的として、TF-IDF および情報エントロピーを用いた定量的アプローチを提案した。その結果、以下のような知見が得られた。

まず、TF-IDF に基づく分析により、各業態に特有の不満語を抽出することが可能となった。これらの特徴的な語は業態の改善ポイントを直接示すものであり、実務的な価値が高い。

一方、情報エントロピーを用いた不満語の出現分布分析では、飲食業界全体に共通する基礎的な不

満要因と、いくつかの業態に固有の不満を識別することができることが確認された。さらに、共起語ペアに着目した分析では、特定業態に集中して用いられる語の組み合わせが明らかとなり、単語単位では把握できない文脈的な不満構造を捉える上での有効性が示された。

これらの結果から、エントロピーを軸とした不満要因の分析手法は、単なる頻度分析やトピックモデルに比べて、より構造的かつ実用的な洞察を提供できることが確認された。

今後の拡張の可能性としては、LDA などのトピックモデル手法の導入、深層学習ベースの文脈モデリング、時系列・季節性分析、分析結果を実務者が直観的に理解できる可視化ツールとの連携が考えられる。

## 謝辞

本研究では、国立情報学研究所の IDR データセット提供サービスにより株式会社 Insight Tech から提供を受けた「不満調査データセット」を利用した。

## 参考文献

- [1] 澤田悠治, 北山大輔, 角谷和俊. ユーザの不満情報を用いたアイテムの肯定的特性の抽出. DEIM Forum 2018.
- [2] 吉川耀敬, 王元元. 商品レビューの不満・満足情報抽出に基づく不満解決商品推薦手法の提案. DEIM Forum 2021.
- [3] 末廣駿, 齋藤博昭. 不満調査データセットの素性ベクトル化. 言語処理学会 第 23 回年次大会, 2017
- [4] 村本英明, 鍛冶伸裕, 吉永直樹, 喜連川優. 意味カテゴリに基づく語義曖昧性解消における Web 資源の活用について. 情報処理学会論文誌, Vol.51, No. 10, 2010.
- [5] 長谷川徹, 北山大輔. 不満調査データセットを用いた不満グループの可視化. DEIM Forum 2017.
- [6] K. Mitsuzawa, M. Tauchi, M. Domoulin, M. Nakashima and T. Mizumoto. FKCC Corpus: a Japanese Corpus from New Opinion Survey Service. In proc. of the Novel Incentives for Collecting Data and Annotation from People: types, implementation, tasking requirements, workflow and results, 2016.