

POSデータを用いた注文傾向による店舗間の関係分析

安井 彰悟^{1,a)} 武藤 敦子^{1,b)} 島 孔介¹ 森山 甲一¹ 松井 藤五郎² 犬塚 信博¹

概要：消費者の購買行動を分析する技術として POS データ分析がある。商品の売上向上のために小売店の売上全体におけるジャンル別の商品売上割合を元に店舗をクラスタリングし、クラスタ内の店舗で扱う商品や立地等傾向を明らかにする手法がある。しかし、飲食店においては売上だけではなく顧客個人の注文方法に傾向がみられる。そこで本研究では、レシート毎の注文から非負値行列因子分解を用いて注文傾向を抽出し店舗の分類を行う手法を提案する。この手法では頻出する注文傾向で店舗を分類するため、店舗のニーズと立地・環境との関連性の特定に活用が可能になる。実験では、実店舗である飲食店の POS データに対して本提案手法を適用し有効性を確認した。

1. はじめに

POS (Point Of Sales) データとは、小売店や飲食店等の売り上げを記録したデータであり、購買時刻や商品、購入数などの購買に関連した情報が日々、収集されている。このデータによって、経営者は販売実績や売れ行きの動向などを容易に把握することができる [1]。このようなデータを用いた、消費者の購買行動についての分析 [2][3][4][5] などによって、自店舗における現状や課題などを特定することができ、経営における意思決定の支援に繋がると考えられる。POS データを用いた研究の 1 つとして、小売店等の複数店舗の POS データを分析し、複数店舗でみられる購買傾向や類似した売り上げ傾向にある店舗を特定する研究が存在する [6]。また、この研究を活用し、購買傾向が類似した店舗に共通している特徴を特定し、購買とその特徴の関係性についての分析もなされている [7]。[7] は、商品ジャンル別の商品売上割合から店舗分類することで、特定の商品ジャンルの売上と立地等の関係性を特定する手法を提案した。店舗の分類を行うことで、各店舗ごとに効果的な品揃えや販売促進等についての提案が可能になると考えられる。しかし、小売業以外の商品数が少なく、商品ジャンル間に大きな違いが存在しない業種において、従来手法の適用では、各店舗のジャンル単位での購買傾向の特徴から顧客のニーズを特定することができないという課題が存在する。そこで、本

研究では POS データから顧客の注文傾向を抽出し、店舗の分類を行う手法を提案する。

2. 関連研究

本節では、提案手法で利用している非負値行列因子分解について説明し、その後、店舗分類に関する関連研究について述べる。

2.1 非負値行列因子分解

非負値行列因子分解 (Non negative Matrix Factorization: NMF) とは、非負値で表された $I \times J$ サイズの入力行列 X を因子分解し、下記のように $I \times R$, $R \times J$ サイズの因子行列 A , B の積の形へ分解する手法である [8]。 R は因子数である。

$$X \approx AB^T \quad (1)$$

因子分解では因子行列 A , B の積と入力行列 X の距離を最小化することで分解結果を得ることができ、得られた分解結果がデータ中のパターンに対応する。例として、図 1 のような入力行列 X (レシート × 商品) に NMF を適用した場合、潜在的な購買のパターンを因子行列 B の因子 r ($1 \leq r \leq R$) の列を見ることで把握することができ、各レシートはどの購買パターンにどの程度基づいて購買されたのかを因子行列 A の各ユーザ i ($1 \leq i \leq I$) の行を見ることで把握することができる。

2.2 関連研究

POS データの分析として、複数の店舗を売り上げ傾向が類似している店舗に分類する方法がある。この分析では、チェーン店など、同一企業が経営する複数の店舗において

¹ 名古屋工業大学
Nagoya Institute of Technology, Nagoya, Aichi 466-8555, Japan

² 中部大学
Chubu University, Kasugai, Aichi 487-8501, Japan
a) clf14125@nitech.jp
b) mutoh.atsuko@nitech.ac.jp

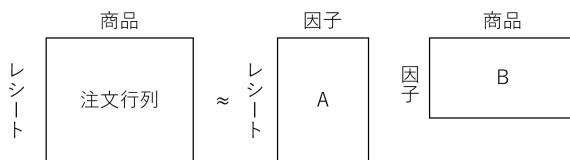


図 1 NMF
Fig. 1 NMF

収集された POS データから、店舗を販売傾向が類似しているものでグループ分けすることで、チェーン店に存在する様々なニーズを特定することができる。例として,[6] は小売店のチェーンストアにおいて、扱う品目や重点とする品目が異なるいくつかの店舗スタイルの妥当性を確認するために店舗の分類を行った。この研究では、各店舗の商品の各中分類売上の割合をもとにクラスター分析を用い、店舗の分類が行われている。また、店舗分類の結果と店舗特徴といった POS データ以外の情報を組み合わせ、両者の関係性を特定するといった研究が存在する。実際に、小売店の複数店舗において収集された POS データから、購買の傾向によって店舗を分類し、店舗間の関係性やその他店舗特徴との関連性について分析がされている。例として,[7] は、平日・週末のデータごとに店舗分類を行い、販売特性の変化と店舗の地理的特徴・商圏環境の特徴の関係性について分析した。この研究では、5 チェーンのコンビニエンスストア 287 店舗における商品カテゴリの販売実績データに対して、因子分析とクラスター分析を用いることで、店舗は販売特性の異なる 7 つのグループに分類されている。これらの研究では、商品ジャンル別売り上げ数を店舗ごとにデータから算出し、それらの値をもとに店舗の分類が行われている。しかし、小売店以外の商品数が少ない POS データの場合、この手法の適用は必ずしも適切であるとは限らない。理由として、飲食チェーン店の場合、取り扱う商品が全て食品であり、商品ジャンル間にも大きな違いが存在しないため、店舗分類後の各店舗クラスタにおける、ジャンル単位の注文傾向に違いが少ないことが挙げられる。これにより、チェーン店内に存在するニーズなどの情報が分析によって十分に得られない。また、このような店では、一部の看板商品を中心とした注文が多い傾向があると考えられるため、顧客のニーズを特定するためには商品単位かつ、どの商品とどの商品が同時に注文されやすいかという顧客の注文傾向について詳しく調べる必要がある。

3. 提案手法

本研究では、飲食店などの商品数が少ないチェーン店において、顧客の注文傾向と店舗特徴との関係性を特定するため、POS データから顧客の注文傾向を抽出し、店舗の分類を行う手法を考える。2.2 節で説明したように、飲食店等では従来手法のような店舗単位・ジャンル単位のデータに

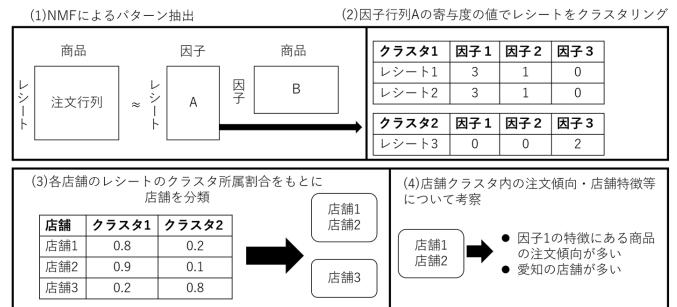


図 2 提案手法概要
Fig. 2 Proposed method

よる店舗分類は適切でないため、提案手法では伝票(レシート)単位・商品単位のデータを用いる。提案手法では、NMF とクラスター分析を組み合わせ、POS データから注文傾向を抽出し、店舗の分類を行う。また、店舗分類の結果より、店舗における注文傾向の特徴と立地・環境といった特徴の関係性についても分析する。

3.1 提案手法の流れ

提案手法の流れを図 2 に示し、説明する。

- (1) **NMF によるパターン抽出**: 商品単位での、顧客の注文傾向を抽出するために全店舗の POS データからレシート × 商品の入力行列を作成し、NMF によって 2 つの因子行列 A, B に分解する。因子行列 B は各因子の注文の特徴を示し、因子行列 A はレシートがどの因子の特徴にどの程度基づいて注文をしているかを示す。
- (2) **因子行列 A の寄与度でレシートをクラスタリング**: 因子行列 A の各行はレシートの各因子への寄与度を示しており、これらの値が類似しているレシートは似た注文となる。そのため、因子行列 A の各行を各レシートの特徴ベクトルとし、レシートをクラスタリングすることで、注文内容によるレシートの分類を行う。
- (3) **各店舗のレシートの所属割合をもとに店舗を分類**: クラスタリング結果より各店舗のレシートがどのクラスタにどの割合所属しているかを算出し、その値をもとに店舗を分類する。図 2 の例の場合、店舗 1, 2 ではクラスタ 1 の値が大きいため、クラスタ 1 の注文傾向(因子 1 と因子 2 の注文特徴)が強いと解釈することができる。
- (4) **店舗クラスタ内の注文傾向と店舗特徴の関係性について考察**: 各店舗クラスタは注文傾向とその割合が類似した店舗群となっているため、顧客層などに共通点が存在することが考えられる。顧客層等と注文傾向との関係性を特定するために、顧客層に関して間接的に関係のある店舗の立地・環境に共通点がないかを考察する。

4. 実験

本実験では、提案手法と従来手法を実際の飲食店の POS データに対して適用する。本節では、まず提案手法において NMF の適用結果を述べ、次に店舗分類の結果と立地環境との関係性について述べる。その後、従来手法でも同様に NMF の結果、店舗分類の結果について述べる。それぞれの店舗分類結果から分析手法を定性的に評価し、提案手法の有効性を確認する。

4.1 提案手法による実験

4.1.1 実験環境

協力企業の同一チェーンレストランであるアロハテーブル 18 店舗における 2021 年 7 月 5 日から 2021 年 12 月 5 日の POS データを用いる。この飲食店ではランチタイム（11 時から 16 時）とディナータイム（16 時以降）によってメニューの取り扱いが異なるため、来店数の多いランチタイムのデータを利用する。POS データでは、どのメニューが 1 人分の注文であるかが不明であるため、人数が異なるデータや人数の多いデータにおいて注文傾向の抽出が複雑になってしまう。また、友人やカップルといった顧客の関係性は注文内容と関連があることが予想されるため、顧客同士の関係性が予想しやすく、データが多くなった 2 人組の注文データを対象とする。

4.1.2 データの加工

分析に適した入力行列を算出するために、以下のようなデータの加工を行った。

- (1) 各店舗からレシートを 500 枚ずつランダムに抽出
- (2) テイクアウトメニューの注文があるレシートを削除
- (3) 総注文数が少ない商品とそれを注文しているレシートを削除

このデータの加工によって、レシート枚数は 9000 枚から 8506 枚に減少した。

4.1.3 NMF の定式化

分析に用いる NMF は [9] を参考にした。NMF の近似 \approx の尺度には、式 (2) に示すユークリッド距離 d_{EU} を利用した。 $x_{ij} \in X$ が入力行列の要素、 $\hat{x}_{ij} \in \hat{X}$ が近似行列の要素である。

$$d_{EU}(x_{ij} \parallel \hat{x}_{ij}) = \frac{1}{2}(x_{ij} - \hat{x}_{ij})^2 \quad (2)$$

d は行列の要素同士の距離であり、行列同士の距離 \mathcal{D}_{EU} は式 (3) のように表せる。

$$\mathcal{D}_{EU}(X \parallel \hat{X}) = \sum_{i,j=1}^{I,J} d_{EU}(x_{ij} \parallel \hat{x}_{ij}) \quad (3)$$

ここで、入力行列 X と因子行列の積 \hat{X} の距離を $\mathcal{D}_{EU}(X \parallel \hat{X})$ とし、NMF は式 (4) に示す最適化問題を解くことで因子行列を出力する。因子行列を更新していく上で入力行列と因

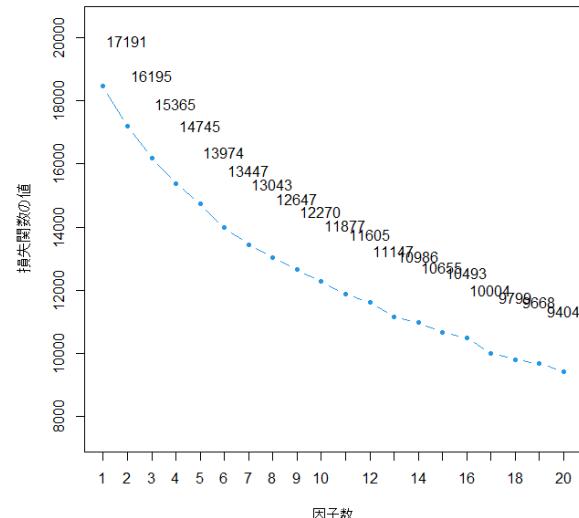


図 3 損失関数

Fig. 3 Cost function

子行列の積の誤差（距離）を損失関数の値とする。

$$\begin{aligned} & \arg \min_{A,B} \mathcal{D}_{EU}(X \parallel \hat{X}) \\ & s.t. A, B \geq 0 \end{aligned} \quad (4)$$

この最適化問題を解くために式 (5)–(6) のアルゴリズムを利用した。

$$b_{jr} \leftarrow b_{jr} \frac{\sum_{i=1}^I x_{ij} a_{ir}}{\sum_{i=1}^I \hat{x}_{ij} a_{ir}} \quad (5)$$

$$a_{ir} \leftarrow a_{ir} \frac{\sum_{j=1}^J x_{ij} b_{jr}}{\sum_{j=1}^J \hat{x}_{ij} b_{jr}} \quad (6)$$

a_{ir} に更新式にある b_{jr} は式 (5) にて更新した後の値を利用する。

4.1.4 因子数の決定方法

因子数は [10] と同様に、損失関数の値の減少幅が明らかに小さくなった時の因子数を最適な因子数として設定する。

4.1.5 NMF 結果

因子数を 1 から 20 で設定し、それぞれ入力行列を因子分解した際の、損失関数の値を 3 に示す。因子数の増加に対し、損失関数の値の減少幅はほぼ一定であった。そこで、因子数は分析で設定した最大因子数 20 と設定した。

NMF によって抽出された因子の特徴として「メインメニューとドリンク」、「メインメニュー」、「セットメニューとドリンク」が多くみられた。

4.1.6 レシートのクラスタリング結果

レシートのクラスタリングには k-means 法を用い、クラスタ数は 5 とした。本研究では、分類したレシートのクラスタ内に頻出している注文の特徴を注文傾向と解釈するため（以降、各レシートクラスタの注文特徴を注文傾向と呼

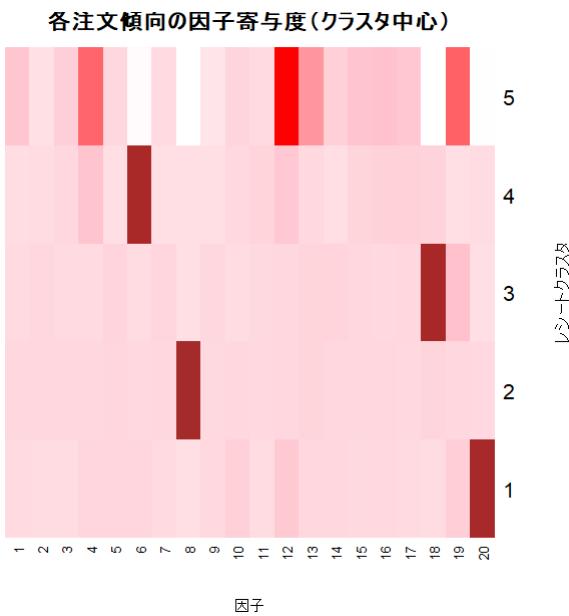


図 4 クラスタ中心
Fig. 4 Cluster centers

表 1 店舗の分類結果
Table 1 Result of clustering stores

店舗クラスタ 1	店舗クラスタ 2	店舗クラスタ 3	店舗クラスタ 4
テラスモール湘南	大崎	栄ミナミアロハ	コレットマーレ
横浜		京橋	星が丘
海老名		金山	
千葉		湘南	
		赤坂	
		仙台	
		代官山	
		大宮	
		中目黒	
		飯田橋	
		豊洲	

ぶ), 本分析ではレシートのクラスタリングによってレシートデータから 5 つの注文傾向を抽出する。各クラスタ内レシートの注文傾向を示すクラスタ中心のヒートマップを図 4 に示す。

なお、レシートのクラスタリングでは、クラスタ 5 に分類されたものがデータの半数を占めていた。

4.1.7 店舗分類結果

店舗の分類には k-means 法を利用し、クラスタ数は 4 とした。表 1 に店舗分類の結果、表 2 に各店舗クラスタにおけるレシートのクラスタ所属割合示す。

表 2 より、各店舗においてレシートクラスタ 5 に所属するレシートが最も多かった。このレシートクラスタでは人気のメインメニューを中心とした注文の特徴が強く、どの店舗においても同様のニーズがあることが特定できた。次に各店舗クラスタごとに表 2 と図 4、因子行列 B の数値行列から注文傾向を特定し、表 3 に示す。表 3 では、各注文傾向において特徴が強かった商品を太字で示している。

表 2 各店舗クラスタ内レシートのクラスタ所属割合

Table 2 Percentage of receipts in each store cluster belonging to each cluster

レシートクラスタ	1	2	3	4	5
店舗クラスタ 1	0.08	0.09	0.15	0.13	0.52
店舗クラスタ 2	0.10	0.01	0.06	0.25	0.56
店舗クラスタ 3	0.08	0.04	0.12	0.00	0.74
店舗クラスタ 4	0.11	0.03	0.14	0.07	0.63

表 3 各店舗クラスタの注文傾向

Table 3 Ordering and store characteristics for each cluster

店舗クラスタ	特徴的な注文傾向
1	カウカウセット カウカウレモネード アイスコーヒー アイスカフェオレ
2	ロコモコ アロハ MIX プレート ガーリックシュリンプレート モチコチキンプレート サーモンアボカドライス レモネード
3	ロコモコ モチコチキンプレート サーモンアボカドライス フライドポテト レモネード
4	特徴なし

4.2 従来手法による実験

提案手法での手法の評価を定性的に行うため、同じデータに対して従来手法を適用し、店舗分類を行い、結果について比較する。

4.2.1 従来手法の流れ

従来手法として、[7] における店舗分類の流れを以下に示す。

- (1) POS データから店舗ごとの商品カテゴリ別売上を算出し、店舗×商品カテゴリの入力行列を作成
- (2) 因子分析を用いて、データから相関のある商品カテゴリを因子の特徴として抽出
- (3) 各店舗の寄与率が高い上位因子のみを用いて、店舗を分類

4.2.2 実験環境

本分析では、飲食店の商品を 6 つのジャンル（メインメニュー・単品メニュー・デザート・コース・ドリンク・アルコール）に分類した。今回の分析では店舗分類によって得られる各店舗クラスタの注文傾向がジャンル単位となり、店舗間の注文傾向の特徴に大きな違いが存在しないため、注文傾向の直接的な比較は行わない。また、従来手法では共通因子の抽出に因子分析を用いているが、本分析では提案手法の分析と同様に NMF を利用して行う。

4.2.3 NMF 結果

因子数を 1 から 20 で設定し、入力行列を因子分解した結

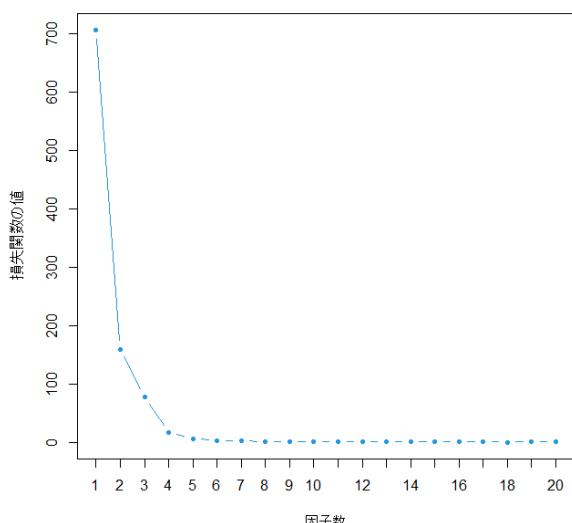


図 5 損失関数
Fig. 5 Cost function

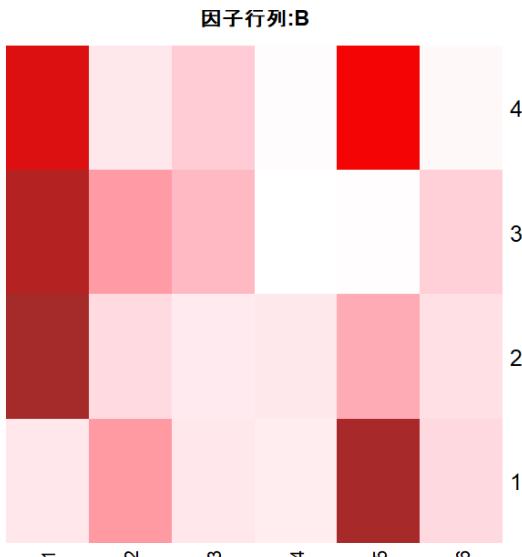


図 6 因子行列 B
Fig. 6 Factor matrix B

果を図 5 に示す。今回の分析では因子数が 4 以降、損失関数の減少幅が大幅に減少したため、因子数は 4 と設定した。次に因子数が 4 の時の、因子行列 B のヒートマップを以下の図 6 に示す。因子行列 B は各行の要素が因子、各列の要素が商品ジャンルを示しており、左側からメインメニュー・単品・デザート・コース・ドリンク・アルコールの順に対応している。

4.2.4 店舗分類結果

因子行列 A の値をもとにクラスタリングし、得られた店舗分類の結果を表 4、各店舗クラスタの因子の寄与度(クラスタ中心)を図 7 に示す。クラスタリングには提案手法と

店舗クラスタ 1	店舗クラスタ 2	店舗クラスタ 3	店舗クラスタ 4
湘南	栄ミナミアロハ	コレットマーレ	海老名
仙台	赤坂	テラスモール湘南	京橋
代官山		横浜	金山
		大崎	星が丘
		中目黒	千葉
		飯田橋	大宮
		豊洲	

各店舗クラスタの因子寄与度(クラスタ中心)

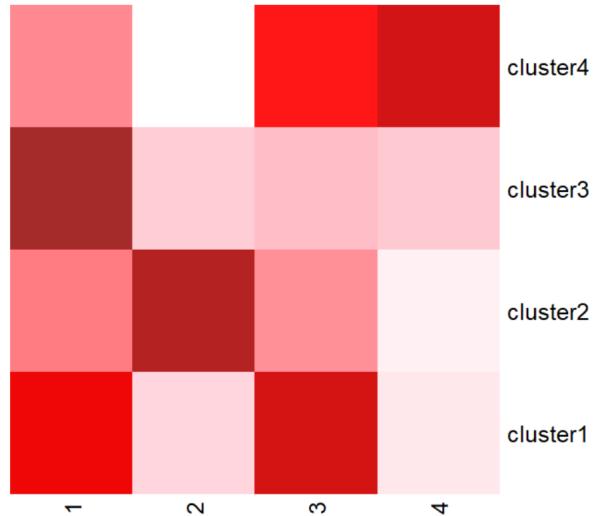


図 7 各店舗クラスタの因子の寄与度(クラスタ中心)
Fig. 7 Cluster centers

表 5 各店舗クラスタの注文傾向
Table 5 Ordering and store characteristics for each cluster

店舗クラスタ	特徴的な注文傾向(ジャンル)
1	メイン、ドリンク、単品
2	メイン、単品
3	ドリンク、単品
4	メイン、ドリンク、デザート

同様に k-means 法を用い、クラスタ数は 4 とした。

図 7 より、各店舗クラスタにおいて因子 1 の注文の傾向が最も高かった。注文傾向 1 ではドリンクの特徴が強く、どの店舗にもニーズがあることが特定できた。次に各店舗クラスタの注文傾向を図 6 と図 7 の数値行列を比較することで特定し、店舗特徴とまとめたものを表 5 に示す。

分析結果より、各店舗クラスタで人気のある商品ジャンルを特定することができた一方で、各店舗クラスター内に共通した店舗の特徴は店舗クラスタ 2 のみでしか見られなかった。

4.3 分析手法の比較

店舗分類で得られた各店舗クラスタ内の店舗において、立地や環境等に何らかの傾向が存在するかを確認し、手法の比較を行う。表 6 に提案手法と従来手法における各店舗

表 6 各店舗クラスタの店舗特徴

Table 6 Store characteristics for each cluster

手法	クラスタ	店舗立地や環境の特徴
提案手法	1	駅前商業施設
	2	ビジネス街と住宅地の狭間に立地
	3	ランチ需要が高い
	4	ビジネス客が多い繁華街に立地
従来手法	1	不明
	2	ビジネス街に立地
	3	不明
	4	不明

クラスタ内の店舗における立地や環境の傾向を示す。本研究では、POS データ以外の店舗に関するデータが収集できていないため、各店舗クラスタ内の店舗特徴については、データの提供元企業が考察した。

従来手法での店舗分類では、各店舗でニーズの多い商品ジャンルを特定することができたが、各店舗クラスタにおける店舗特徴に関連性は見られなかった。一方で、提案手法では各店舗クラスタ内において店舗の立地・環境に共通点がみられた。店舗の立地や環境といった店舗の特徴は、顧客層との間に関係があると考えられるため、各店舗クラスタにおいて店舗特徴に共通点がみられた提案手法の方が、顧客層ごとの異なるニーズを捉えた店舗分類ができていると考えられる。これは分析の対象とした飲食店ではメインメニュー やドリンクにおける看板商品を中心とした注文が多く、複数の人気商品を同じジャンルに分類していた点が、提案手法と結果が大きな異なった要因の 1 つであると考えられる。よって、商品ジャンルの間に大きな違いが存在しない飲食店において、顧客の注文傾向をもとに店舗分類を行う提案手法の適用が適していると考えられる。

5.まとめと今後の課題

本研究では POS データから NMF とクラスタリングによって顧客の注文傾向をもとに店舗分類する手法を提案した。実験では、実際の飲食店の POS データに対して提案手法と従来手法の両方を適用し、有効性を確認した。課題としては、提案手法では各店舗クラスタの店舗数に大きな差があり、店舗クラスタ 3 では分析対象のおおよそ半分の店舗が所属していたため、より適切な店舗クラスタの決定方法が挙げられる。他にも、評価実験としてより細かい商品のジャンル分けや他の飲食店での実験等が必要である。

謝辞 本研究のデータ提供元である株式会社ゼットンに感謝する。本研究は JSPS 科研費 22K12273 の助成を受けたものです。

参考文献

- [1] 小沢佳奈: 近年の POS データ研究に関するレビュー, 繊維機械学会誌, Vol. 61, No. 5, pp. 357–363, (2008)
- [2] 森本康彦, 八木一光: POS データからの売上変動パターン発掘, 日本データベース学会 Letters, Vol. 4, No. 2, pp.

97–100, (2005)

- [3] 宇佐美俊, 大竹恒平, 生田目崇: スーパーマーケットチェーンにおける店舗コーナーを用いた顧客特徴と購買行動の評価, 情報処理学会第 79 回全国大会, Vol. 5K-07, No. 1, pp. 505–506, (2017)
- [4] 若井拓哉, 中平勝子, 北島宗雄: 顧客満足度向上のための購買行動と POS 販売履歴の比較, 情報処理学会第 76 回全国大会, pp. 1-551–552 (2014)
- [5] 田中孝昌, 濱口智大, 西郷拓海, 津田和彦: スーパーマーケットの店舗別販売傾向と RFM 分析を利用した優良顧客分類, SIG-KSN, Vol. 20, (2017)
- [6] 山田実俊, 舟山貴光, 河野克哉, 服部敦, 山本義郎: 売り上げ傾向による店舗の分類と購買傾向の分析と可視化, 計算機統計学, Vol. 29, No. 2, pp. 159–167, (2016)
- [7] 箸本健二, 駒木伸比古: コンビニエンスストアの店舗類型とその平日・週末間での差異 –首都圏 287 店舗の POS データ分析を通して–, 都市地理学, Vol. 4, pp. 1–19, (2009)
- [8] D.D. Lee and H.S. Seung: *Learning the parts of objects with nonnegative matrix factorization*, Nature, Vol. 401, No. 6755, pp. 788–791, (1999)
- [9] 亀岡弘和: 非負値行列因子分解, 計測と制御, vol. 51, pp. 835–844, (2012)
- [10] 安川武彦: 非負値行列因子分解を用いたテキストデータ解析, 計算機統計学, Vol.28, No.1, pp.41–55, (2015)