

グルメサービスにおけるレストラン推薦投稿への リアクション数増加を目的とした潜在クラスモデル分析

劉 佩潔¹ 山下 遥^{2,a)} 岩永 二郎³ 樽石 将人³ 後藤 正幸^{1,b)}

受付日 2017年4月19日, 採録日 2017年9月5日

概要: 近年, インターネット上でユーザの提供したレストラン情報を掲載するグルメサービスが広く利用されるようになった. 本論文で対象とするグルメサービス Retty では, ユーザはレストラン検索に加えて, レストランに対するおすすめ度や推薦記事の投稿, 他ユーザが投稿した推薦記事に対するリアクション (ex., 「いいね」) 等を行うことができる. このリアクションの数はサービス全体の活性度を表す指標として位置づけられ, サービス運営会社とユーザの双方にとって, リアクション数の増加が望まれる. そのため, ユーザの投稿やリアクションといった履歴情報を活用し, リアクション数の増加に結び付けられる施策を明らかにするための分析モデルは意義がある. 一方で, これらのユーザの投稿やリアクションは, 個々のユーザの判断に行われ, 多様な個人的嗜好を反映したものとなっている. これらのユーザの嗜好は多様であり, 全ユーザに対して単一的な統計的法則があるとは考えられないため, その異質性を表現した統計モデルを導入する必要がある. そこで本研究では, ユーザがレストランに対して情報を掲載する行動, ならびにその他のユーザがその情報にリアクションする行動を, 潜在クラスモデルによって表現する. 具体的には, ユーザがレストランに対しておすすめ度を含む推薦記事を投稿する事象を「投稿するユーザ」, 「レストラン」, 「おすすめ度」の3つの共起ととらえ, 「発信ユーザのおすすめ度傾向モデル」を提案する. さらに, 記事を読んでいるユーザが投稿にリアクションする事象を「読んでいるユーザ」と「レストラン」, 「記事を投稿したユーザ」の共起と考え, これらの事象に対して「受信ユーザのリアクション傾向モデル」を提案する. これらの評価傾向モデルとリアクション傾向モデルの両者を統合的に分析することで, リアクション数を増加させるための施策を検討する方法を示す.

キーワード: グルメサービス, 潜在クラスモデル, EM アルゴリズム, アクセスログデータ, ビジネスアナリティクス

An Analysis Model Based on Latent Class Models to Increase Reactions to Restaurant Recommendation on Social Gourmet Service

PEIJIE LIU¹ HARUKA YAMASHITA^{2,a)} JIRO IWANAGA³ MASATO TARUISHI³ MASAYUKI GOTO^{1,b)}

Received: April 19, 2017, Accepted: September 5, 2017

Abstract: In recent years, social gourmet services that post restaurant information provided by users on the Internet have been widely used. On a social gourmet service, users can make a restaurant search, recommendations to restaurants, reactions (ex., “like”) to the recommended articles posted by other users, and the like. The number of reactions is positioned as an indicator of the activity level of the service site, and an increase in the number of reactions is desired for both a service operating company and users. It is, therefore, meaningful to provide analytical models to exploit historical information such as posts and reactions by users and to clarify measures linked to an increase in the number of reactions. In this research, we propose the two statistical models of behaviors that users post information to restaurants and that other users react to that information introduced by latent class models. By analyzing both of the evaluation tendency model and the reaction tendency model in an integrated manner, we show a method of examining measures to increase the number of reactions.

Keywords: gourmet service, latent class model, EM algorithm, access log data, business analytics

¹ 早稲田大学
Waseda University, Shinjuku, Tokyo 169-0072, Japan

² 上智大学
Sophia University, Chiyoda, Tokyo 102-0081, Japan

³ Retty 株式会社
Retty Inc., Minato, Tokyo 108-0073, Japan

a) h-yamashita-1g8@sophia.ac.jp

b) masagoto@waseda.jp

1. はじめに

近年、インターネット上でユーザが提供したレストラン情報を掲載するグルメサービスが広く利用されるようになってきている。本研究が対象事例とするグルメサービス Retty では、2011年6月サービス開始以降、利用者数が増加し、2017年5月現在、月間利用者数は3,000万人（ユニークブラウザ数でユーザ数を集計した数値）を超えている。このように現在、グルメサービスは多くの人々の生活に浸透しているといえる。

このようなインターネット上で提供されるグルメサービスでは、レストランの情報のほか、ユーザの投稿履歴や閲覧履歴といったデータが蓄積されるため、これらのデータを活用した施策立案に結び付けることが期待されている。グルメサービスで獲得されるデータの分析については、これまでもいくつかの研究が行われている。たとえば、Pantelidis [1] は、グルメサービスにおけるレストランのおすすめ度を高めるための要因を分析している。また、Kangら [2] は、レストランの評価記事の感情語 (i.e., Negative/Positive) に着目した、記事の定量化に関する研究を展開している。一方、望月ら [3] は、レストランの投稿ユーザの経験によって作成された評価記事を受信ユーザにとって伝わりやすい表現に変えるための方法を提案している。また、Zhangら [4] はグルメサイト上での予約行動に着目し、予約につながる要因を分析している。

一方、本研究で対象とするグルメサービス Retty [5] 上では、ユーザはレストラン検索や情報閲覧に加えて、レストランに対しておすすめ度を含む推薦記事の投稿や、他ユーザが投稿した推薦記事に対するリアクション (ex., 「いいね」) をすることができる。推薦記事を投稿したユーザにとって、通常は多くのユーザにリアクションされることは多くの共感を獲得することができたととらえられるため、次の推薦記事を投稿するモチベーションの向上につながると思われる。このように、ユーザからのリアクション数はサービス全体の活性化につながると考えられ、サービス運営会社とユーザの双方の視点から、リアクション数の増加が望まれている。

本研究では、レストランの推薦記事を投稿するユーザを発信ユーザ、それらの推薦記事を参考にレストランを決定するユーザを受信ユーザと定義する。発信ユーザは実際に行ったレストランの中から、自身で周りにおすすめしたいレストランを選び、そのレストランの味、接客対応、価格等様々な観点からレストランの総合的なおすすめ度を決定し、その理由を推薦記事の内容に反映する。一方で、受信ユーザは推薦記事の内容への共感の程度から、発信ユーザの投稿する推薦記事に対してリアクションをするかどうかを決定する。すなわち、発信ユーザとレストランの特徴によって推薦記事を決定し、その推薦記事に対して受信ユー

ザがリアクションする。そのため、「発信ユーザが推薦記事を投稿するレストランと、おすすめ度との関係 (投稿モデル)」、ならびに「発信ユーザが投稿したレストランと、受信ユーザのリアクションの関係 (リアクションモデル)」を明らかにすることで、リアクション数増加のための施策検討に結び付けることができると考えられる。

グルメサービスにおけるリアクション数増加の施策を検討するためには、「発信ユーザが推薦記事を投稿するレストランと、評価得点との関係」、および「発信ユーザが投稿したレストランと、受信ユーザのリアクションの関係」を分析することが重要と考えられる。本研究ではリアクション数の増加を目的とした分析のため、上記2つの関係性を表現する統計モデルを構築するとともに、それらの関係性を統合的に考慮した分析手法を提案する。具体的にはまず、「発信ユーザが、ある1つのレストランを推薦する」という事象に対して、「発信ユーザ、個別のレストラン、おすすめ度」を確率変数とした確率モデル (投稿モデル) を提案することで、発信ユーザの投稿特性を分析可能とする。加えて、「受信ユーザが、ある1つのレストランの投稿に対してリアクションする」という事象に着目し、「発信ユーザ、個別のレストラン、受信ユーザ」を確率変数とした確率モデル (リアクションモデル) を構築し、受信ユーザのリアクション特性を分析可能とする。その際、本研究では潜在クラスモデル [6], [7], [8], [9] を導入することにより、ユーザやレストランの異質性を考慮したモデル化を可能とする。提案モデルにより、リアクション向上を目的として、発信ユーザに対して、リアクション数の期待値が高いと推測されるレストランのリストを提示することが可能となり、これにより、発信ユーザがこれらのレストランの中から投稿するレストランを選ぶことで、リアクション数の向上が期待される*1。本研究では、提案手法を用いてグルメサービス Retty に蓄積された実データを用いて分析を行い、その結果について考察を与える*2。その結果、提案したモデルを用いた統合的分析により、リアクション数増加のための具体的な施策について検討が可能であることを示す。

*1 提案モデルを用いて、各レストランに対して期待リアクション数を計算できるため、実際に発信ユーザに (リアクション数が見込める) レストランを推薦する際には、個々のレストランのリストを提示することができる。また、受信ユーザに対しても、リアクションがされやすいであろうレストランのおすすめ記事を提示することができる。このように、提案するモデル (投稿モデル、およびリアクションモデル) は、リアクションを向上させる施策を検討可能な分析結果を提示することを目的としている。

*2 本研究では、対象問題の性質上、個々のユーザの ID やレストランの具体的な名称を提示した分析結果や考察を論文で論じることができない。そこで、潜在クラスごとに所属するレストランの主たるカテゴリーを提示することで各潜在クラスの意味の解釈を行い、提案モデルの妥当性について考察する。

2. 準備

2.1 関連研究

グルメサービス上から獲得されるデータの分析については、これまでもいくつかの研究が行われている。たとえば、Pantelidis [1] は、グルメサービスにおけるレストランのおすすめ度を高めるための要因の分析、Kang ら [2] は、レストランのおすすめ記事の感情語 (i.e., Negative/Positive) に着目した、記事の定量化に関する研究を展開している。一方、望月ら [3] は、レストランの投稿ユーザの経験によって作成されたおすすめ記事を受信ユーザにとって伝わりやすい表現に変えるための方法を提案している。また、Zhang ら [4] はグルメサイト上での予約行動に着目し、予約につながる要因を分析している。このように、様々な観点からグルメサービスから得られるデータを対象とした分析が行われている。しかし、本研究のように、ユーザからのリアクション数の増加を目的とした研究は行われていない。

一方、本研究で提案する分析手法は、特にマーケティング分野で有効性が認められている潜在クラスモデル [6], [7], [8], [9] をベースとしている。潜在クラスモデルは、観測されたデータの背後に観測できない潜在的な変数の存在を仮定したモデルである。潜在的な変数の仮定は、異質のデータが混在しうる複雑な構造を持つデータに対して現実に即した分析を可能とするため、マーケティング等の分野でよく用いられるモデルとなっている [10]。たとえば、潜在クラスモデルの1つに、2つの確率変数が潜在クラスが与えられたもとで条件付独立に生起すると仮定する Aspect Model [6] がある。これは、Latent Semantic Indexing [11] の流れを汲んで提案されたため、Probabilistic Latent Semantic Indexing (PLSI) と呼ばれることも多いが、離散の潜在クラスを持つ確率モデルである。Aspect Model はもともとは言語データ解析のためのモデルとして提案されたが [6]、データに仮定する分布や、その構造を変化させることで自由度の高い分析を実現するため、実際に購買履歴分析 [12], [13] や推薦システム [14], [15], [16]、検索結果の要約 [17] 等に幅広く適用されている。また、顧客の購買履歴データの分析としては、ECサイトの閲覧履歴と購買履歴を同時に表現する潜在クラスモデル [18], [19] やトピック追跡モデル [20] 等、多方面へ拡張したモデルが提案されている。

以上の潜在クラスモデルのメリットは、嗜好の異なるユーザグループや特徴の異なるアイテムグループの混在という現象を適切に表現することができる点にあり、特に本研究で対象とする問題には有効であると考えられる。本研究では、潜在クラスモデルをベースとして「発信ユーザが推薦記事を投稿するレストランと、おすすめ度との関係」、および「発信ユーザが投稿したレストランと、受信ユーザ

のリアクションの関係」を明らかにするための分析モデルを提案する。

2.2 グルメサービス

グルメサービスとは、主にレストランの情報を検索できるインターネット上のサービスである。現代において最初のグルメサービスは1996年に誕生したとされている [21]。最初のグルメサービスが誕生して以来、今日に至るまで様々なグルメサービスが誕生、消失を繰り返してきた。グルメサービスはインターネットが普及し始めた頃に誕生した。当時はインターネットへの接続は従量課金とされていたため、この時期では自店のホームページを持っている飲食レストランはかなり少なかった。そこで、ホームページを持っていない飲食レストランは、グルメサービスに掲載された自店の情報を自店のホームページとして代用するというケースが多く存在した。このような背景から、グルメサービス誕生当初はレストランの情報を掲載するレストランガイド型 [22] のグルメサービスが主流であった。現在ではインターネットは広く普及し、レストランを決定する際にグルメサイトまたはグルメアプリを利用する人も増加してきている。

また近年、レストランガイド型のグルメサービスから、ユーザが提供するレストランの情報を掲載する口コミ型のグルメサービス [23] が主流になってきている。レストランガイド型のグルメサービスの場合、レストランの宣伝を書くことができるのは、レストラン側であり、ユーザではない。レストラン側は集客のため、自店の宣伝内容を記載できる。しかし、レストラン側が書いた自店の推薦内容は量に限りがあり、様々な角度、視点からの情報が得られないと考えられる。これに対して、ユーザが情報を掲載する口コミ型のグルメサービスでは、ユーザが様々な視点から自由に口コミを書くことができるため、情報量が多くなりレストランを選ぶ際に、より参考にすることができる。このようなことから、現在はユーザが提供するレストランの情報を掲載する口コミ型のグルメサービスが主流になってきている。

口コミ型のグルメサービスは主に2つの機能がある。1つ目はレストランを検索する機能である、ユーザはグルメサービスを用いてレストランを検索し、どのレストランに行くか、自分の好みに合わせて決めることができる。2つ目は行ったレストランに対して推薦記事を投稿し、おすすめ度を付ける機能である、ユーザはレストランに対して味やサービス、コンセプト等、様々な観点から自由に推薦記事を投稿し、おすすめ度を付けることができる。

このような口コミ型のグルメサービスを利用することにより、レストランを探すユーザが他ユーザの投稿したレストラン推薦記事を参考にし、普段行かない、もしくは知らないレストランの情報を手に入れることが可能となり、好

みのレストランを選択することが容易になる。また、レストランで食事した後は、ユーザが味、サービスやコンセプト等に対する満足度から、レストランの推薦記事として投稿する。推薦記事、他ユーザから共感された場合には、「いいね」等のリアクションを獲得することができる。

2.3 対象事例

2.3.1 問題設定

本研究で扱うグルメサービスにおける、2つの重要なユーザの行動を指摘しておく。1つ目は、おすすめしたいレストランの推薦記事を投稿する発信ユーザの行動、2つ目は、推薦記事を参考にし、レストランを探す受信ユーザの行動である。

まず、発信ユーザにとって、多くのリアクションを獲得することは好ましい。発信ユーザは自分の行ったレストランの中から、おすすめしたいレストランに対して経験に基づくおすすめ度を与え、推薦記事を投稿する。その際、リアクションの獲得が推薦記事の投稿に対するモチベーションの要因の1つになっていると思われる。一方、受信ユーザは共感するレストラン推薦記事に対してリアクションする。すなわち、受信ユーザはレストランページに表示される推薦記事を参考にし、利用するレストランを決定し、共感する推薦記事に対してリアクションするものと考えられる。

このようにリアクション数は、サービス内でのユーザの活動の程度を表しており、多い方が望ましいと考えられる。そこで、本研究ではこれらに着目し、リアクション数の増加を促す施策を考えるためのモデルを提案する。

2.3.2 対象グルメサービスの機能

本研究では、グルメサービス Retty [5] のデータを用いて分析を行う。グルメサービス Retty を利用するためには、ユーザが Facebook のアカウントでユーザ登録し、ログインする必要がある。すなわち、グルメサービス Retty は、「実名型」のグルメサービスであり、その他のグルメサービスとは大きく異なるといえる。匿名性の場合、口コミが誰によって書かれたものなのか、ユーザからは分からない。たとえば、受信ユーザが参考にした口コミを書いた発信ユーザが、受信ユーザとは性別や年齢が大きく異なっていた場合、参考にしないことが考えられる。しかし、実名性の場合、誰が投稿した口コミなのかユーザからも分かるため、自分と似た特徴を持つ発信ユーザや、友人等自分と近い関係にある人の口コミを取捨選択することができ、口コミが参考になりやすくなるという特徴がある。グルメサービス Retty では主に以下のような機能がある。

- おすすめしたいレストランに対して、推薦記事を投稿する（口コミ、写真）。
- 友達の推薦記事から行きたいレストランを見つける。
- 友達の推薦記事に対して、「いいね」や「行きたい」と

いったリアクションを付けることができる。

- 「行きたい」のリアクションがなされたレストランは、そのユーザのリストに自動的に追加される。
- エリアを指定し、地図でレストランを見る。
- 投稿ユーザの中でも特に専門のエリアとジャンルに詳しいトップユーザの投稿を参考に、行きたいレストランを探す。

また、ユーザの主な行為は以下である。

レストラン検索

ユーザはテキスト検索のほか、位置情報、ジャンル、営業時間、予算もしくはまとめ記事、友達のオススメから、レストランを探す。

行ったレストランに関する記事の投稿

ユーザはレストランに対しておすすめ度を投稿し（おすすめ度は3段階である。Excellent/Good/Normal）、推薦記事を投稿する。

行きたいレストランリストの作成

ユーザが好きな投稿もしくはレストランに対して「行きたい」を付けることで、自分だけの「行きたいレストランリスト」に自動的に追加される。

2.3.3 対象グルメサービスの基礎データ

グルメサービス Retty には多様なデータが蓄積されている。蓄積されているデータの例とその基礎集計を以下に示す。

ユーザ情報

ユーザプロフィール情報には、「ユーザ ID」を含むデモグラフィック情報等のデータが含まれている。

レストラン情報

レストラン情報としては、「レストラン ID」、「レストラン名」、「住所」、「エリア ID」等の情報が含まれる。これに加え、レストランカテゴリ情報として、「レストラン ID」、「詳細カテゴリ ID」、「作成日」等がある。また、レストランのカテゴリは、表 1 に一部を示すように「大カテゴリ」、「中カテゴリ」、「小カテゴリ」と階層分類されている。

レストラン推薦記事情報

レストラン推薦記事情報には、「レストラン推薦記事 ID」、「発信ユーザ ID」、「レストラン ID」、「おすすめ度」、「作成日」等のデータが含まれる。一方、レストラン推薦記事リアクション情報としては、「レストラン推薦記事 ID」、「発信ユーザ ID」、「受信ユーザ ID」、「リアクションタイプ（ex. 「いいね」、「行きたい」等）」、「作成日」等の情報が保存されている。

2.3.4 本研究に用いるデータの基礎分析

所在地で集計したレストラン数上位3県の全体数に対する割合を表 2 に示す*3。

表 2 から、研究対象となるグルメサービス Retty が蓄積

*3 実数については非公開指標データのため、本論文では割愛する。

表 1 レストランのカテゴリ情報

Table 1 Category information of restaurants.

大カテゴリ	中カテゴリ
和食	和食
	日本料理・郷土料理
	すし・魚料理
	ラーメン・麺類
	お好み焼・粉物
アジア料理	中華
	アジア・エスニック
ヨーロッパ料理	フレンチ
	洋食・西洋料理
	イタリアン
	アメリカ料理
	アフリカ料理
	珍しい各国料理
肉料理	焼肉・ステーキ
	焼鳥・串料理
	こだわりの肉料理
鍋料理	鍋
	しゃぶしゃぶ・すき焼き
営業形態	居酒屋・バー
	カフェ・スイーツ
	ファミリーレストラン・ファストフード
	ビュッフェ・バイキング
その他	その他

表 2 レストランの基礎集計

Table 2 Basic statistics of restaurants.

レストラン数上位 3 県	割合
1. 東京都	17.79%
2. 大阪府	9.43%
3. 愛知県	6.75%

しているレストラン情報データの中、東京都のレストラン店舗数が一番多いことが分かる。そこで、本研究の提案モデルでは東京都のレストランを対象として分析を行う。

提案手法を用いて、グルメサービス Retty に蓄積された実データを分析する。分析対象データは、2016 年 11 月現在東京エリア内で営業中のレストランとする。研究対象とするデータの基礎分析を行い、結果を示す。

ただし、本研究で分析に用いるカテゴリは営業形態以外は大カテゴリを表し、営業形態に関してはレストラン数が多いため、中カテゴリを使用する。表 3 は用いるカテゴリと各カテゴリに属するレストランの割合を示している*4。

以下で示す提案モデルを用いた実データの学習結果に対し、各潜在クラスの特徴を解釈するために、表 3 のレストランのカテゴリ情報を用いる。

表 3 カテゴリごと所属レストランの割合

Table 3 The number of restaurants in each category.

カテゴリ	レストランの割合
和食	27.84%
アジア料理	9.19%
ヨーロッパ料理	12.95%
肉料理	9.28%
鍋料理	1.31%
居酒屋・バー	20.56%
カフェ・スイーツ	15.46%
ファミレス・ファストフード	0.37%
ビュッフェ・バイキング	0.24%
その他	2.79%

3. 提案モデル

グルメサービス Retty には 2 つの重要なユーザの行動がある。1 つ目は、レストランを評価し、推薦記事を投稿する発信ユーザの行動、2 つ目は、推薦記事を参考にし、レストランを探す受信ユーザの行動である。上記の 2 つの行動は、リアクションの獲得につながるものと考えられるため、本研究ではこれらに着目し、リアクション数の増加を促す施策を考えるためのモデルを提案する。発信ユーザはレストランで食事をした経験を基におすすめ度を与え、受信ユーザは共感する推薦記事にリアクションするという 2 つの観点から 2 つのモデルを構築し、発信ユーザとレストランおよびおすすめ度、または受信ユーザと発信ユーザおよびレストランの関係性をそれぞれ分析する。さらに、それらを統合的に考慮した分析手法を提案する。モデル化では潜在クラスモデルを導入し、発信ユーザとレストランおよび受信ユーザ等が複数の異なる潜在クラスに属する度合いを所属確率で表すことで、混在した各ユーザの嗜好やレストランごとの特徴を表現する。構築するモデルは以下の 2 つである。

(1) 投稿モデル：発信ユーザとレストラン、および発信ユーザのレストランに対するおすすめ度の関係性を表現する潜在クラスモデル。

(2) リアクションモデル：受信ユーザのリアクションと発信ユーザ、および発信ユーザによって提供されるレストラン情報の関係性を表現する潜在クラスモデル。

具体的には、投稿モデルでは、「発信ユーザが、ある 1 つのレストランを推薦する」という事象に対して、「発信ユーザ、個別のレストラン、おすすめ度」を確率変数とした確率モデル化を行い、リアクションモデルでは「受信ユーザが、ある 1 つのレストランの投稿に対してリアクションする」という事象に着目し、「発信ユーザ、個別のレストラン、受信ユーザ」を確率変数とした確率モデル化を行う。

これにより、たとえば、発信ユーザに対して、リアクション数の期待値が高いと推測されるレストランのリストを提

*4 実数については非公開指標データのため、本論文では割愛する。

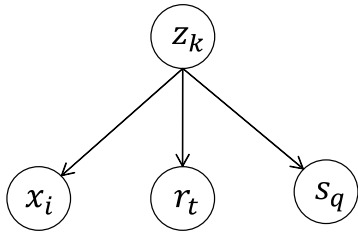


図 1 投稿モデルのグラフィカルモデル

Fig. 1 The graphical model of the posting model.

示することで、発信ユーザがこれらのレストランの中から投稿するレストランを選べば、リアクション数の向上が期待できる。その際、推定されたモデルを用いた期待リアクション数は、各レストランに対して計算可能であり、実際に発信ユーザに（多くのリアクション数が見込める）レストランを推薦する際には、個店レベルでのレストランリストを提示することができる。

3.1 投稿モデル

投稿モデルでは、リアクションを獲得している発信ユーザが推薦記事を投稿するレストランの特徴を明らかにするために、発信ユーザとレストランおよびそのレストランに対する評価の関係性をモデル化する。そこで、全 N 件のレストラン推薦記事データにおける、発信ユーザ集合を $\mathcal{X} = \{x_i : 1 \leq i \leq I\}$ 、レストラン集合を $\mathcal{R} = \{r_t : 1 \leq t \leq T\}$ 、おすすめ度の集合を $\mathcal{S} = \{s_q : 1 \leq q \leq Q\}$ と定義する。また、 K 個の潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ と定義する。発信ユーザ x_i がレストラン r_t におすすめ度 s_q を与えるという事象を (x_i, s_q, r_t) と記述し、各変数の分布にはそれぞれ多項分布を仮定する。このとき、レストラン推薦投稿の確率モデル $P(x_i, s_q, r_t)$ を式 (1) で表すことにする。

$$P(x_i, s_q, r_t) = \sum_{k=1}^K P(x_i|z_k)P(s_q|z_k)P(r_t|z_k)P(z_k) \quad (1)$$

また、投稿モデルのグラフィカルモデルは図 1 で表される。

ここで N 件のレストラン推薦記事が与えられたときの対数尤度関数 LL_1 は式 (2) で与えられる。ただし、事象 (x_i, s_q, r_t) の発生回数を $n(x_i, s_q, r_t)$ と記述し、 $N = \sum_{i,q,t} n(x_i, s_q, r_t)$ となる。また、投稿モデルでは $n(x_i, s_q, r_t)$ は多くの場合、0 または 1 であるが、同じ発信ユーザが 2 回以上同じレストランに同じ評価をつけることもあるため、 $n(x_i, s_q, r_t)$ が 2 以上になる場合もある。

$$LL_1 = \sum_{i=1}^I \sum_{q=1}^Q \sum_{t=1}^T n(x_i, s_q, r_t) \log P(x_i, s_q, r_t) \quad (2)$$

この対数尤度関数を極大化するパラメータは、次の EM アルゴリズムで探索することができる。導出については付録に示す。

パラメータ更新アルゴリズム：

Step 1：各パラメータの初期化

パラメータ $P(z_k)$, $P(x_i|z_k)$, $P(s_q|z_k)$, $P(r_t|z_k)$ の初期値は一様乱数を用いて生成する。

Step 2：パラメータ更新

E-step：

$$P(z_k|x_i, s_q, r_t) = \frac{P(z_k)P(x_i|z_k)P(s_q|z_k)P(r_t|z_k)}{\sum_k P(z_k)P(x_i|z_k)P(s_q|z_k)P(r_t|z_k)} \quad (3)$$

M-step：

$$P(z_k) = \frac{\sum_{i,q,t} n(x_i, s_q, r_t)P(z_k|x_i, s_q, r_t)}{\sum_{i,q,t,k} n(x_i, s_q, r_t)P(z_k|x_i, s_q, r_t)} \quad (4)$$

$$P(x_i|z_k) = \frac{\sum_{q,t} n(x_i, s_q, r_t)P(z_k|x_i, s_q, r_t)}{\sum_{i,q,t} n(x_i, s_q, r_t)P(z_k|x_i, s_q, r_t)} \quad (5)$$

$$P(s_q|z_k) = \frac{\sum_{i,t} n(x_i, s_q, r_t)P(z_k|x_i, s_q, r_t)}{\sum_{i,q,t} n(x_i, s_q, r_t)P(z_k|x_i, s_q, r_t)} \quad (6)$$

$$P(r_t|z_k) = \frac{\sum_{i,q} n(x_i, s_q, r_t)P(z_k|x_i, s_q, r_t)}{\sum_{i,q,t} n(x_i, s_q, r_t)P(z_k|x_i, s_q, r_t)} \quad (7)$$

Step 3：収束判定

$$LL_1 = \sum_{i=1}^I \sum_{q=1}^Q \sum_{t=1}^T n(x_i, s_q, r_t) \log P(x_i, s_q, r_t) \quad (8)$$

(1) 繰り返し数 $t = 1$ のとき、 $LL_{old} = LL_1$ とし、Step 2 へ戻る。

(2) 繰り返し数 $t \geq 2$ のとき、 $LL_{new} = LL_1$ とし、

a) $\frac{|LL_{new} - LL_{old}|}{|LL_{new}|} \leq \epsilon$ ならば終了

b) $\frac{|LL_{new} - LL_{old}|}{|LL_{new}|} > \epsilon$ ならば $LL_{old} = LL_{new}$ とし Step 2 へ戻る。

3.2 リアクションモデル

リアクションモデルでは、受信ユーザが投稿に対して行ったリアクション傾向を明らかにするために、受信ユーザと発信ユーザおよび発信ユーザによって提供されるレストラン情報との関係性をモデル化する。

N 件のレストラン推薦記事に対する全 M 回のリアクションにおける、受信ユーザ集合を $\mathcal{Y} = \{y_j : 1 \leq j \leq J\}$ 、 L 個の潜在クラス集合を $\mathcal{V} = \{v_l : 1 \leq l \leq L\}$ とする。また、受信ユーザ y_j が発信ユーザ x_i のレストラン r_t に対するレストラン推薦記事にリアクションするという事象を (x_i, y_j, r_t) と記述し、各変数の分布にはそれぞれ多項分布を仮定する。このとき、レストラン推薦記事へのリアクションの確率モデル $P(x_i, y_j, r_t)$ を式 (9) に、グラフィカルモデルを図 2 に示す。

$$P(x_i, y_j, r_t) = \sum_{l=1}^L P(x_i|v_l)P(y_j|v_l)P(r_t|v_l)P(v_l) \quad (9)$$

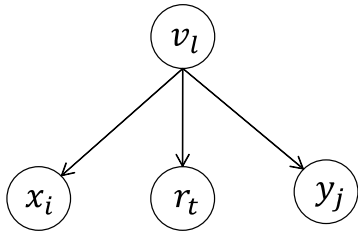


図 2 リアクションモデルのグラフィカルモデル
Fig. 2 The graphical model of the reaction model.

ここで、 M 回のリアクションが与えられたときの対数尤度関数 LL_2 は式 (10) で与えられる。ただし、 $M = \sum_{i,j,t} n(x_i, y_j, r_t)$ となる。また、リアクションモデルにおいても事象の発生回数 $n(x_i, s_q, r_t)$ はほぼ 0 または 1 であるが、2 以上になることもある。

$$LL_2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^T n(x_i, y_j, r_t) \log P(x_i, y_j, r_t) \quad (10)$$

提案リアクションモデルにおいても、観測できない $\mathcal{V} = \{v_l : 1 \leq l \leq L\}$ を含むため、EM アルゴリズム [24], [25] によるパラメータの推定を行う。投稿モデルの学習に対する EM アルゴリズムの導出と同様に、以下のパラメータ更新アルゴリズムを導くことができる。

パラメータ更新アルゴリズム

Step 1 : 各パラメータの初期化

パラメータ $P(v_l)$, $P(x_i|v_l)$, $P(y_j|v_l)$, $P(r_t|v_l)$ の初期値は一様乱数を用いて生成する。

Step 2 : パラメータ更新

E-step :

$$P(v_l|x_i, y_j, r_t) = \frac{P(v_l)P(x_i|v_l)P(y_j|v_l)P(r_t|v_l)}{\sum_l P(v_l)P(x_i|v_l)P(y_j|v_l)P(r_t|v_l)} \quad (11)$$

M-step :

$$P(v_l) = \frac{\sum_{i,j,t} n(x_i, y_j, r_t)P(v_l|x_i, y_j, r_t)}{\sum_{i,j,t,l} n(x_i, y_j, r_t)P(v_l|x_i, y_j, r_t)} \quad (12)$$

$$P(x_i|v_l) = \frac{\sum_{j,t} n(x_i, y_j, r_t)P(v_l|x_i, y_j, r_t)}{\sum_{i,j,t} n(x_i, y_j, r_t)P(v_l|x_i, y_j, r_t)} \quad (13)$$

$$P(y_j|v_l) = \frac{\sum_{i,t} n(x_i, y_j, r_t)P(v_l|x_i, y_j, r_t)}{\sum_{i,j,t} n(x_i, y_j, r_t)P(v_l|x_i, y_j, r_t)} \quad (14)$$

$$P(r_t|v_l) = \frac{\sum_{i,j} n(x_i, y_j, r_t)P(v_l|x_i, y_j, r_t)}{\sum_{i,j,t} n(x_i, y_j, r_t)P(v_l|x_i, y_j, r_t)} \quad (15)$$

Step 3 : 収束判定

$$LL_2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^T n(x_i, y_j, r_t) \log P(x_i, y_j, r_t) \quad (16)$$

- (1) 繰り返し数 $t = 1$ のとき、 $LL_{old} = LL_2$ とし、Step 2 へ戻る。
- (2) 繰り返し数 $t \geq 2$ のとき、 $LL_{new} = LL_2$ とし、

表 4 統合分析における集計表

Table 4 Summary table in the integrated analysis.

モデル1 モデル2	1	...	l	...	L
1	$n_{1,1}$...	$n_{1,l}$...	$n_{1,L}$
⋮	⋮	⋮	⋮	⋮	⋮
k	$n_{k,1}$...	$n_{k,l}$...	$n_{k,L}$
⋮	⋮	⋮	⋮	⋮	⋮
K	$n_{K,1}$...	$n_{K,l}$...	$n_{K,L}$

- a) $\frac{|LL_{new} - LL_{old}|}{|LL_{new}|} \leq \epsilon$ ならば終了
- b) $\frac{|LL_{new} - LL_{old}|}{|LL_{new}|} > \epsilon$ ならば $LL_{old} = LL_{new}$ とし Step 2 へ戻る。

3.3 統合分析

本研究では、提案投稿モデルと提案リアクションモデルの潜在クラスの特徴を 1 つの表にまとめることで、統合的な分析を行う方法を提案する。表の行は投稿モデルの潜在クラス $k(1 \leq k \leq K)$ 、列はリアクションモデルの潜在クラス $l(1 \leq l \leq L)$ を表すものとする。すなわち、各行は発信ユーザのレストランの嗜好によって分析された各クラスタの要素集合、各列は受信ユーザの好む投稿の特徴によって分析された各クラスタの要素集合を表す。また表の k 行 l 列のセルの要素 $n_{k,l}$ に対して、各セルに所属する発信ユーザ数、平均リアクション数、投稿数等の値を集計することで、投稿モデルとリアクションモデルで得られた情報を同時に用いてリアクション数増加のための様々な施策を考案することができる。

4. 実データによる分析

4.1 対象データおよび分析条件

提案手法を用いて、グルメサービス Retty に蓄積された実データを分析する。分析対象データは 2016 年 11 月現在東京エリア内で営業中のレストランとし、本分析での「リアクション」は、発信ユーザのレストラン推薦投稿に対して受信ユーザが「いいね」を押すこととした。また、2011 年 4 月 30 日から 2016 年 3 月 1 日までにリアクション回数が 1 以上のレストラン推薦投稿データを分析対象とする（データサイズは、レストラン推薦投稿数 N が数 100 万件、リアクション数 M が数 1,000 万件の規模である*5）。おすすめ度は三段階 $Q = 3$ であり、ユーザはレストランに対し 1, 2, 3 のおすすめ度の得点をつけることができる。また、事前分析により、投稿モデルとリアクションモデルの潜在クラス数は $K = 10$, $L = 10$ とした。各パラメータの初期値はランダムに決定し、収束条件を決定するパラメータ ϵ

*5 投稿数とリアクション数の正確な実数、ならびに分析対象となるレストラン推薦投稿の発信ユーザ数 I 、レストラン数 T および受信ユーザ数 J については、非公表データのため掲載できない。

を e^{-7} とした。ただし、 e は自然対数の底とする。

4.2 分析結果

4.2.1 パラメータの推定結果

投稿モデルのパラメータ $P(z_k)$ とリアクションモデルのパラメータ $P(v_l)$ の推定結果を表 5, 表 6 を用いて示す。

また、個々のユーザの ID やレストランの具体的な名称を公開することができないため、詳細を示すことができないが、投稿モデルから得られるパラメータ $P(x_i|z_k)$ から計算される $P(z_k|x_i)$, パラメータ $P(r_t|z_k)$ から計算される $P(x_i|z_k)$ の値, 実際のレストランごとのリアクション数を用いてリアクション数の期待値を計算し, 値の高いレストランを推薦することでリアクションを多く獲得することが期待される。一方, リアクションモデルでは, 得られるパラメータ $P(y_j|v_l)$ から計算される $P(v_l|y_j)$, パラメータ $P(r_t|v_l)$, $P(x_i|v_l)$, 実際の発信ユーザのレストランに関する記事に対するリアクション数を用いて発信ユーザの投稿記事に対するリアクションの期待値を計算し, 値が大きい記事を推薦することにより, リアクション数を増加させることが期待される。

4.2.2 推定パラメータの分析

ここでは, 推定されたパラメータを分析し, クラスの特徴を変数でよく表現していることを示していく。

まず, 投稿モデルに関しては, 潜在クラス z_k に所属する発信ユーザ数, 1 人当たりの投稿数, および発信ユーザのレストランにおけるおすすめ度の期待値, 各潜在クラスの特徴を表すことにする。計算式はそれぞれ式 (17)~式 (19) となる。

表 5 投稿モデルの潜在クラス

Table 5 Latent classes of the posting model.

潜在クラス	z_1	z_2	z_3	z_4	z_5
$P(z_k)$	0.33	0.03	0.05	0.02	0.09
潜在クラス	z_6	z_7	z_8	z_9	z_{10}
$P(z_k)$	0.04	0.02	0.01	0.13	0.28

表 6 リアクションモデルの潜在クラス

Table 6 Latent classes of the reaction model.

潜在クラス	v_1	v_2	v_3	v_4	v_5
$P(v_l)$	0.02	0.14	0.03	0.08	0.12
潜在クラス	v_6	v_7	v_8	v_9	v_{10}
$P(v_l)$	0.16	0.20	0.01	0.10	0.14

表 7 各潜在クラスにおけるおすすめ度の生起確率とおすすめ度の生起確率

Table 7 The analysis of latent classes of the posting model.

おすすめ度	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}	$P(s_q)$
3	1.00	0.00	0.04	0.15	0.93	0.91	0.00	0.31	0.17	0.00	0.48
2	0.00	0.99	0.85	0.85	0.05	0.02	0.96	0.30	0.78	0.95	0.48
1	0.00	0.00	0.11	0.01	0.02	0.07	0.03	0.39	0.05	0.05	0.04

$$\text{発信ユーザ} = \sum_i P(z_k|x_i) \quad (17)$$

$$1 \text{ 人当たりの投稿数} = \frac{NP(z_k)}{\sum_i P(z_k|x_i)} \quad (18)$$

$$\text{おすすめ度の期待値} = \sum_q P(s_q|z_k)s_q \quad (19)$$

ここで各潜在クラスにおけるおすすめ度の生起確率 $P(s_q|z_k)$, およびおすすめ度の生起確率 $P(s_q)$ をまとめると, 表 7 のようになった。表 7 から, おすすめ度の値はほとんど 2 か 3 となっていること, また, 潜在クラスごとのおすすめ度の生起確率 $P(s_q|z_k)$ に着目すると, おすすめ度の違いは値が 2 と 3 の部分のみに表れていることが分かる。このように, 各潜在クラスにおける「おすすめ度」の分布には確かに特徴が見られ, 提案モデルの妥当性が示唆される。

また, リアクションモデルに関しては発信ユーザ数と 1 人当たりの投稿数は投稿モデルと同様に計算し, 潜在クラス v_l に所属する受信ユーザ数を式 (20) で表し, 各潜在クラスの特徴を分析する。

$$\text{受信ユーザ数} = \sum_j P(v_l|y_j) \quad (20)$$

上記の計算方法を用いて, 投稿モデルとリアクションモデルの潜在クラスの結果分析を表 8, 表 9 にまとめる。

また, 投稿モデルとリアクションモデルのレストランのカテゴリを, それぞれの潜在クラスごとの生起確率 $P(r_t|z_k)$ と $P(r_t|v_l)$ における上位 10 位のレストランのカテゴリを用いて以下の表 10, 表 11 に示す。

次に潜在クラスごとに所属するレストランの主たるカテゴリを提示し各潜在クラスの意味の解釈を行い, 「分析結果から考えられること」や「分析結果に対する意味づけや理由の解釈」について述べていく。表 10, 表 11 のランキングの情報を用いて, 以下の 3 つの観点から潜在クラスごとレストランの特徴的なカテゴリを総合的に決定した。

- 一番多く出現したカテゴリ。ただし, 出現回数が同じカテゴリに対して, ランキング上位のカテゴリを優先
- 全潜在クラスを比較し, 潜在クラスでの出現回数が一番多いカテゴリ
- ランキング 1 位のカテゴリ, かつ潜在クラス内に 3 回以上出現したカテゴリ

たとえば, 潜在クラス z_8 において, 和食が 5 回出現し, 潜在クラス z_8 の中一番出現回数の多いカテゴリであるた

表 8 投稿モデルの潜在クラスの結果分析

Table 8 The analysis of latent classes of the posting model.

投稿モデル	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
発信ユーザ数	17,138	1,977	2,826	1,935	6,624	3,326	1,122	714	7,974	13,294
1人当たりの投稿数	21.4	17.65	18.83	15.23	15.32	14.49	16.29	15.17	17.83	23.84
おすすめ度	3.0	2.0	1.9	2.1	2.9	2.8	2.0	1.9	2.1	1.9

表 9 リアクションモデルの潜在クラスの結果分析

Table 9 The analysis of latent classes of the reaction model.

リアクションモデル	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
発信ユーザ数	1,054	6,849	1,379	5,789	9,911	5,453	11,523	695	7,923	6,354
1人当たりの投稿数	14.08	20.12	13.41	16.39	16.66	21.87	23.8	14.36	17.53	15.86
受信ユーザ数	1,582	9,137	1,725	4,706	7,853	9,278	10,330	741	6,751	8,674

表 10 投稿モデルの各潜在クラスにおける生起確率上位 10 位レストランのカテゴリ

Table 10 Categories of top ten restaurants with occurrence probability in each latent class of the posting model.

ランキング	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
1	欧	欧	和食	亜	和食	欧	欧	酒	珈琲	和食
2	肉	珈琲	和食	酒	酒	亜	和食	和食	欧	珈琲
3	和食	欧	和食	酒	和食	亜	和食	欧	珈琲	肉
4	和食	酒	亜	鍋	肉	和食	珈琲	珈琲	和食	和食
5	肉	珈琲	珈琲	和食	酒	亜	和食	酒	珈琲	欧
6	和食	和食	和食	欧	珈琲	珈琲	珈琲	和食	和食	肉
7	和食	欧	和食	欧	亜	和食	珈琲	和食	和食	酒
8	肉	欧	和食	欧	欧	亜	酒	酒	和食	亜
9	酒	亜	和食	和食	珈琲	肉	亜	珈琲	亜	和食
10	酒	珈琲	和食	酒	和食	亜	欧	和食	和食	欧

表 11 リアクションモデル潜在クラスごと生起確率上位 10 位レストランのカテゴリ

Table 11 Categories of top ten restaurants with occurrence probability in each latent class of the reaction model.

ランキング	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
1	和食	酒	和食	酒	珈琲	和食	珈琲	酒	肉	酒
2	和食	和食	和食	和食	和食	酒	珈琲	欧	欧	酒
3	欧	欧	和食	酒	珈琲	肉	和食	珈琲	和食	肉
4	和食	和食	和食	酒	珈琲	酒	和食	酒	酒	酒
5	和食	酒	和食	亜	肉	酒	肉	その他	亜	亜
6	珈琲	和食	酒	和食	珈琲	肉	欧	欧	肉	酒
7	珈琲	欧	和食	酒	欧	和食	和食	欧	亜	酒
8	亜	欧	和食	和食	和食	酒	和食	酒	和食	和食
9	酒	和食	珈琲	肉	和食	酒	和食	酒	肉	酒
10	酒	和食	和食	和食	欧	肉	和食	酒	欧	肉

め、和食を潜在クラス z_8 の特徴的なカテゴリとする。また、居酒屋・バーが潜在クラス z_8 に 3 回出現しており、他の潜在クラスと比べて一番多い。さらに、居酒屋・バーがランキング 1 位であり、かつ潜在クラス z_8 に 3 回以上出現したため、居酒屋・バーも潜在クラス z_8 の特徴的なカテゴリとして考える。

上記のカテゴリの決定方法等を合わせて、提案手法の適用により得られた投稿モデルとリアクションモデルの潜在

クラスごとレストランカテゴリの特徴をそれぞれ表 12、表 13 に示す。ただし、これらの表におけるカテゴリはレストランのカテゴリであり、次の略称を用いて表記している（肉：肉料理、和：和食、欧：ヨーロッパ料理、珈琲/珈：カフェ・スイーツ、鍋：鍋料理、酒：居酒屋・バー、亜：アジア料理）。この結果から、クラスごとにレストランのカテゴリが大きく異なっていることが分かり、提案モデルにより、レストランの特徴の違いをよくとらえられて

表 12 投稿モデルにおける各潜在クラスの主カテゴリ

Table 12 Main category of each latent class in the posting model.

投稿モデル	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
カテゴリ	肉/和	欧/珈	和食	鍋/酒	和食	亜	珈/和	酒/和	珈/和	和食

表 13 リアクションモデルにおける各潜在クラスの主カテゴリ

Table 13 Main category of each latent class in the posting model.

リアクションモデル	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
カテゴリ	和食	欧/和	和食	酒	珈琲	肉/酒	和食	欧/酒	亜/肉	酒

表 14 1人当たり投稿数のクロス分析

Table 14 Cross analysis of the number of postings per person.

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
z_1	14.10	20.05	12.36	14.42	15.21	25.01	22.50	10.70	17.18	16.80
z_2	12.81	15.21	15.42	12.74	13.28	15.03	15.55	10.75	12.13	11.83
z_3	10.93	17.54	15.99	14.25	11.92	14.12	16.88	12.51	13.66	14.78
z_4	14.55	12.34	13.32	15.10	12.14	10.40	12.07	10.30	10.52	12.32
z_5	12.89	15.38	11.20	12.25	11.20	16.61	13.96	13.82	12.05	14.89
z_6	9.80	15.40	8.18	14.00	11.15	13.90	13.15	12.64	10.62	10.25
z_7	7.99	16.69	11.14	10.50	11.09	15.01	15.59	7.99	12.41	9.07
z_8	7.54	12.15	4.45	11.57	10.00	11.24	8.99	10.67	9.16	9.89
z_9	12.32	15.04	11.50	13.01	13.33	14.33	16.99	17.56	13.96	13.04
z_{10}	10.78	17.52	9.21	14.59	16.14	15.96	24.32	11.06	19.55	12.85

いることが分かる。

4.3 統合的な分析

投稿モデルとリアクションモデルの分析結果を用いて統合的な分析を行う。その際に、以下のように (z_k, v_l) における発信ユーザ数 I_{kl} , (z_k, v_l) における投稿数 N_{kl} および (z_k, v_l) におけるリアクション数 M_{kl} をそれぞれ式 (21)~式 (23) に定義する。ここで発信ユーザ x_i がレストラン r_t に投稿した記事数を $N(x_i, r_t)$, 発信ユーザ x_i がレストラン r_t に投稿した全記事におけるリアクション数を $M(x_i, r_t)$ とする。

$$I_{kl} = \sum_i P(z_k|x_i)P(v_l|x_i) \quad (21)$$

$$N_{kl} = \sum_{i,t} P(z_k, v_l|x_i, r_t) \times N(x_i, r_t) \quad (22)$$

$$M_{kl} = \sum_{i,t} P(z_k, v_l|x_i, r_t) \times M(x_i, r_t) \quad (23)$$

投稿モデルにおける潜在クラス (行), およびリアクションモデルにおける潜在クラス (列) での 1人当たり投稿数を表 14 に示す。また, 表 14 の作成方法は式 (24) で表される。

$$1 \text{人当たり投稿数} = \frac{N_{kl}}{I_{kl}} \quad (24)$$

表 14 から, 1人当たり投稿数が一番多いのは 1行 6列のセルの値 $n_{1,6}$ であり, 一番少ないのは $n_{8,3}$ である。全体的にみると, z_1, z_{10}, v_2 と v_7 が 1人当たり投稿数が他

の潜在クラスと比べて多いため, z_1, z_{10}, v_2 と v_7 はよく投稿するアクティブな発信ユーザが集まる潜在クラスだと考えることができる。また, z_7, z_8, v_3 と v_8 が 1人当たり投稿数が他の潜在クラスと比べて少ないため, z_7, z_8, v_3 と v_8 は投稿の少ないアクティブでない発信ユーザが集まる潜在クラスだと考えることができる。

投稿モデルにおける潜在クラス (行), およびリアクションモデルにおける潜在クラス (列) での 1人当たりリアクション数をそれぞれ表 15 と表 16 に示す。

また, 表 15 の作成方法は式 (25) で表される。

$$1 \text{人当たりリアクションされた数} = \frac{M_{kl}}{I_{kl}} \quad (25)$$

表 15 の各セル間の 1人当たりリアクションされた数の差異が分かる。1人当たりリアクションされた数が一番多いのは 1行 6列のセルの値は n_{16} であり, 表 14 の 1人当たり投稿数と同じである。また, 1人当たりリアクションされた数が一番少ないのは 8行 3列のセル n_{83} であり, 表 14 の 1人当たり投稿数と同じである。よって, 1人当たり投稿数が多いほど, 1人当たりリアクションされた数が多い傾向があると考えられる。

表 16 の作成方法は式 (26) で表される。

$$1 \text{投稿当たりリアクション数のクロス結果} = \frac{M_{kl}}{N_{kl}} \quad (26)$$

表 16 の各行に注目すると, 投稿モデルの潜在クラス間

表 15 1人当たりの被りアクション数のクロス分析

Table 15 Cross analysis of the number of reactions per person.

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
z_1	674.98	596.02	577.49	350.49	302.77	982.30	424.93	332.63	331.22	610.82
z_2	375.08	284.89	679.94	164.31	172.97	364.51	192.69	240.71	145.39	257.02
z_3	246.87	338.81	528.14	229.21	139.08	318.67	253.96	246.78	204.50	363.02
z_4	554.74	223.41	534.62	361.86	206.57	254.52	135.69	190.00	144.55	343.45
z_5	605.71	437.81	557.90	269.82	216.25	669.26	240.42	407.62	210.55	566.81
z_6	407.12	392.02	319.94	391.38	175.46	434.29	225.60	300.89	181.48	304.20
z_7	189.40	304.71	418.80	118.68	139.29	342.07	240.65	113.86	173.81	167.73
z_8	139.91	164.92	63.68	153.53	93.72	215.69	71.70	192.96	97.80	184.20
z_9	344.64	259.85	482.06	192.05	171.65	335.48	221.60	367.00	186.00	307.84
z_{10}	266.47	268.44	230.16	201.55	183.02	282.12	336.64	193.74	238.51	219.24

表 16 1投稿あたりリアクション数のクロス分析

Table 16 Cross analysis of the number of reactions per posting.

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
z_1	47.86	29.73	46.74	24.30	19.91	39.27	18.88	31.10	19.28	36.36
z_2	29.29	18.72	44.09	12.90	13.02	24.26	12.39	22.40	11.99	21.73
z_3	22.59	19.32	33.03	16.08	11.66	22.56	15.05	19.73	14.97	24.56
z_4	38.13	18.10	40.14	23.96	17.02	24.47	11.25	18.45	13.74	27.88
z_5	46.98	28.47	49.83	22.03	19.31	40.28	17.22	29.50	17.47	38.07
z_6	41.56	25.46	39.09	27.96	15.74	31.23	17.15	23.80	17.09	29.67
z_7	23.70	18.26	37.59	11.30	12.56	22.79	15.43	14.25	14.00	18.50
z_8	18.55	13.57	14.31	13.28	9.37	19.18	7.97	18.08	10.68	18.62
z_9	27.97	17.28	41.91	14.77	12.87	23.41	13.05	20.90	13.33	23.60
z_{10}	24.72	15.32	24.98	13.81	11.34	17.68	13.84	17.51	12.20	17.07

のリアクション数の差異が分かる。これらのリアクション数の差異は、発信ユーザが記事を投稿するレストランのカテゴリとその評価の違いをよくとらえることができているものと考えられる。

このように、提案モデルで得られたパラメータを詳しく分析することにより、クラスの特徴を変数でよく表現しているモデルとなっていることが分かり、提案モデルの妥当性が示唆される。

5. 考察

本章では、提案投稿モデル、および提案リアクションモデルを用いた解析について様々な角度から考察していく。

5.1 提案モデルの共起要素について

本研究における提案投稿モデルでは発信ユーザとレストランのみの共起を考える従来のPLSAにかわり、おすすめ度の要素を加えた。この手法では発信ユーザがレストランで食事をした体験をもとにおすすめ度を与えたデータを分析し、発信ユーザの潜在的な嗜好を明らかにする。おすすめ度は、発信ユーザがレストランへ行った後、そのサービス、味またはコンセプト等に対する満足度と解釈することができ、発信ユーザが多く推薦記事をするレストランのタ

イプのみならず、そのカテゴリに対する満足度が分かるようになる。

さらに、提案リアクションモデルにおける受信ユーザが投稿にリアクションする事象では、発信ユーザ、レストランおよび受信ユーザのある関係性を元に同時生起していると考えている。発信ユーザは、推薦記事の内容に特徴があり、受信ユーザは、共感する推薦記事の内容に特徴があると考えられる。そのため、発信ユーザと受信ユーザのマッチングという観点からリアクション数の増加を促す施策を考えることが可能である。

提案手法では、投稿モデルの共起要素として、発信ユーザ、レストランとおすすめ度を設定しており、リアクションモデルの共起要素として、発信ユーザ、レストランと受信ユーザを設定している。2つのモデルとも発信ユーザとレストランがという要素が入っている。しかしながら、投稿モデルでは、発信ユーザがレストランへ行き、その満足度のよりおすすめ度を付けた事象に対するモデル化であり、いわゆる投稿データを扱うモデルである。一方でリアクションモデルでは、発信ユーザによるレストラン推薦記事の投稿に受信ユーザが閲覧したときのリアクション事象に対するモデル化であり、いわゆるリアクションデータを扱うモデルである。すなわち、類似したグラフィカルモデ

ルであってもどのような事象に対して、モデル化を行っているかによって得られる結果は大きく異なる。そのため、本提案手法における2つのモデルから得られる潜在クラス結果の解釈は大きく異なり、それを統合的に分析することによって、各々のモデルのみで得られる以上の情報を得ることができる。

5.2 潜在クラス数の決定方法について

潜在クラス数を決定する方法として、AIC（赤池情報量基準）とBIC（ベイズ情報量基準）等の情報量基準が知られている [26]。これらの情報量基準はパラメータの最尤推定量が漸近的に正規分布に従うモデルクラスに対して導出された基準であるが、潜在クラスモデルはこのような性質を満たすことが保証できないため、厳密にはAICやBICを適用してよい確率モデルではない。しかし、AICやBICを形式的に用いて潜在クラス数を選択することは可能である。

一方、今回の潜在クラスモデルでは、個々のユーザとレストランに多項分布を仮定するため、パラメータ数は（ユーザ数+レストラン数）×潜在クラス数程度となる。このパラメータ数は非常に大きいため、AICやBICではペナルティ項が相対的に重視され、データの統計的性質にかかわらず、パラメータ数の少ないシンプルなモデルが選ばれるようになる。一方で、このようなモデルを選んだ場合、潜在クラスによって多様なグループが混在しているという潜在クラスモデルの有用性が失われてしまうという問題が生じる。実際に、予備実験を行った際に、潜在クラス数 $K = 3, 4, 6, 8, 10$ のモデルに対して、AICとBICを用いてモデルの結果の評価を行ってみたところ、潜在クラス数が小さければ小さいほどAICとBICとも小さくなるという結果が得られ、うまく機能しなかった。

そこで本研究では「関係性分析」という目的を考慮し、実験的方法によって、結果の技術的解釈が可能な潜在クラス数を探索し、提案投稿モデルと提案リアクションモデルの潜在クラス数を決定することとした。予備実験で試した潜在クラス数の結果を分析し、その考察から分析者が判断した結果、 $K = 10$ 、 $L = 10$ の場合の分析結果を採用した。実際に、提案投稿モデルでは潜在クラス数がそれぞれ $K = 3, 4, 6, 8, 10, 15, 20$ のパターンを実験し、潜在クラスの解釈を行っている。潜在クラス数が $K = 10$ より少ない (ex., $K = 3, 4$ 等) のとき、レストラン推薦記事やユーザが1つの潜在クラスに集中し、潜在クラスごとに特徴の差異が明確とならなかった。また、潜在クラス数が $K = 10$ より多いとき、提案リアクションモデルと統合的に分析する際に、分析するセルが大量になってしまい、結果の解釈が困難となった。そのため、本研究では潜在クラスごとに特徴の違いがあり、なおかつ分析が容易なセル数であった潜在クラス数 $K = 10$ と $L = 10$ とした。

このように、提案モデルでは、潜在クラス間の差異と分析の容易さを考えて潜在クラス数を決定する必要がある。そのため、現状では分析者による判断や恣意性を完全に排除することは難しい。今後、広く提案モデルの利用を促すためには、クラスの数を決めるための新たな基準や方法が求められる。

5.3 おすすめ度の条件付確率のモデルについて

本研究で提案した投稿モデルのおすすめ度は順序数ではなくカテゴリ変数として定式化されている。すなわち、ユーザがつけたおすすめ度1, 2, 3の違いは取り扱える一方で、その大小関係をモデルに組み込むことができていない。これに対して、たとえば Gaussian pLSA [27] のように、おすすめ度に正規分布を仮定する等の方法によっておすすめ度の3段階の大小関係を取り扱うことが可能である。

おすすめ度が5段階や10段階のような場合、多項分布ではおすすめ度の大小を扱えないだけでなく、パラメータ数がそれぞれ4と9になる等、モデルの自由度が上がってしまう。そのため、パラメータ数2の正規分布を当てはめることによって、(おすすめ度は連続変数ではないものの) 精度の良い推定が可能になると考えられる。

一方、今回の対象問題では、おすすめ度の値が3段階と少なく、三項分布のパラメータ数も2で少ないことから、推定精度もそれなりに確保できると考え、多項分布モデルを採用した。また、離散値のカテゴリカル変数として扱うことは、定式化やモデルの推定が容易であるというメリットがあり、実際に Flexible Mixture Model [28] では評価値に多項分布を仮定した定式化が行われている。

以上のように、本研究では対象とする実問題のおすすめ度が3段階であることから、単純に多項分布を仮定したモデル化を行った。しかしながら、おすすめ度の等の得点に関する値の数が多いたデータに対し、本提案モデルを適用する場合には、得点の部分に対して順序数としてモデル化すべきであると考えられる。具体的には、正規分布やベータ分布等を仮定することで、少ないパラメータ数の確率分布で、おすすめ度の大小関係を取り扱うことができる。

5.4 施策の立案について

4章で得られた提案投稿モデルと提案リアクションモデルの結果を1つの表にまとめることで、統合的な分析を行った。その際に表の各列で「発信ユーザの好みを考慮しつつ、発信ユーザにリアクションされやすいレストランを推薦する」施策、および表の各行で「受信ユーザの好むレストランを考慮しつつ、受信ユーザに共感しそうな発信ユーザおよびレストランの推薦記事を推薦する」施策を考察した。

本研究で対象としているグルメサービス Retty では、多くの発信ユーザは、様々なレストランに対する有用な情報

を一般消費者に広く知らしめることに価値を見出ししており、「多くの他者からリアクションされた」ということが1つのモチベーションになっている。これは「他のユーザや一般閲覧者から見られること（自身の投稿情報が参考にされること）」が価値となっているためである。

そこで、たとえば次のようなサービスを提供することで、発信ユーザが次に訪れる（おすすめ記事を投稿する）レストランを決めるための有効な情報を提供することができる。

- 1) 発信ユーザが、自身の判断であるレストランのおすすめ記事を投稿する。
- 2) その投稿完了時に、「次に、これらのレストランのおすすめ記事を投稿してはどうでしょうか？」という推薦リストを、期待リアクション数の上位順で提示する。
- 3) 発信ユーザは、期待リアクション数の高いレストランのリストを参考に、次におすすめ記事を投稿するレストランを選ぶことができる。

以上のような期待リアクション数に基づくレストランの推薦により、発信ユーザは、新たにリアクション数を稼ぐことができそうなレストランを知ることができ、そのレストランのおすすめ記事を書くために実際にそのレストランに訪問することで、（発信ユーザによる）レストランへの訪問を促進することができるものと考えられる。

さらに、発信ユーザにレストランを推薦する際に表 12 のように集計したレストランのカテゴリを考慮することで、発信ユーザの好みを考慮した、個別のレストランを推薦することが可能となる。たとえば、モデル 2 において多くリアクションされている v_1 の中で、リアクション数 $n_{k,8}$ が最も少ない投稿モデルの潜在クラスは z_8 である。このセルに所属する発信ユーザがより多くのリアクションを得るためには、リアクション数の多いクラスに所属しているレストランの推薦記事を投稿すると効果的である。なぜなら様々なレストランを食べ歩いて有用な情報を一般消費者に広く知らしめることに価値を見出ししている発信ユーザは、リアクションされやすいレストランの記事を投稿したいと考えるからである。また、表 12 より z_8 のカテゴリ「居酒屋」に着目すると、 z_8 に所属する発信ユーザに対して、リアクション数が多い z_4 に属する居酒屋に対して推薦記事の投稿を促すことにより、リアクション数を増加させることができるものと思われる。上記のように受信ユーザにリアクションされそうなレストランを推薦することで、実際にそのレストランへの訪問を促進し、推薦記事を投稿させることができることが示唆される。

次に、受信ユーザに対する施策としては、受信ユーザがレストランを探す際、検索をかける前に受信ユーザが共感しそうな現在地周辺等のレストランに対する推薦記事を提示する、また、受信ユーザが推薦記事に対してリアクションを行った際に、個人ページに受信ユーザが共感する可能性がある推薦記事のコンテンツを表示するということが考

えられる。上記のような施策によって、受信ユーザが自分の知りたいレストランの情報を投稿する発信ユーザを見つけ、レストランを探すことが容易になる。このように、タイミングを考えた実際のグルメサービスにおける推薦が可能となる。また、表 16 の各列ではリアクションモデルの潜在クラス間のリアクション数の差異が分かる。よって、受信ユーザにリアクションの多いセルに属するレストランへの推薦記事を促し、横方向のセルに受信ユーザを移動させることにより、リアクション数の増加が期待できる。その際に、記事を推薦する際にレストランのカテゴリを考慮することで、受信ユーザの好むレストランを考慮することができる。

ここで z_1 において、1 投稿あたりリアクション数が最も少ない v_7 に着目する。このセルに所属する発信ユーザがより多くのリアクションをするためには、リアクション数の多いクラスに所属している発信ユーザの推薦記事を受信ユーザに推薦することが有効であると思われる。たとえば、表 13 より v_7 のカテゴリは和食である。そこで、 v_7 に所属する発信ユーザに対して、リアクション数が多い v_1 に属する和食レストランの推薦記事の閲覧を推薦することにより、リアクション数の増加が期待できる。

このように本研究で提案したモデルにより、当該グルメサービスにおいて様々な施策を講じることができる。ただし、当該サービスでは日々、新しい記事が投稿され、データが更新されていく。よって現実に則したモデルを得るためには、データが更新するたびにモデルを学習しなおす必要がある。サービス運営会社が得られる効果とそれにかかるコストを試算し、意思決定する必要がある。そこで、提案手法で推定可能な $P(x_i|r_t)$, $P(y_j|x_i, r_t)$ や、直近の閲覧履歴等のデータから期待リアクション数を算出し、得られる効果を定量化することによって意思決定が容易になる。

6. まとめと今後の課題

本研究では、グルメサービスにおけるリアクション数を増加させるための施策を考えることを目的とし、3つのステップからなる分析モデルを提案した。具体的には、リアクションを獲得する発信ユーザのレストランに対するおすすめ度の関係性を明らかにする潜在クラスモデルと受信ユーザが発信ユーザによって提供されるレストランとの関係性を明らかにする潜在クラスモデルをそれぞれ構築し、結果を統合的に分析するという手法である。また、グルメサービス Retty の実データに対して提案モデルを適用し、結果を検証した。その結果、発信ユーザ、受信ユーザの2つの関係性とリアクション数の関係が明らかになり、発信ユーザにリアクションされやすいレストランを推薦し、受信ユーザに共感しそうな発信ユーザおよびレストランの推薦記事を推薦する施策の立案を可能とした。このような施策を実際のグルメサービスにおいて実施することにより、

リアクション数が増加しサービスが活性化することで今までリアクションをしていなかったユーザにも波及効果を及ぼすことが可能であると考えられる。本研究では、異なる事象に着目して統合的に分析することで新たな知見を得ることができたが、これはグルメサービスに限ったことではなく様々なサービス、データにも応用できる汎用的な方法であるため、様々なデータに適用することで、新たな知見を得ることが可能になると考えられる。

適切な潜在クラス数の決定ルールの検討、ならびに本研究で得られた知見を活かした実際のグルメサービスの活性化につながる施策の立案等については今後の課題とする。

謝辞 本研究にあたり、熱心な議論とサポートをいただいた早稲田大学創造理工学部経営システム工学科後藤研究室の皆様、データ提供元である Retty 株式会社の皆様に深く感謝致します。本研究の一部は、科学研究費 (26282090, 26560167) の助成を受けたものである。

参考文献

- [1] Pantelidis, I.S.: Electronic Meal Experience: A Content Analysis of Online Restaurant Comments, *Cornell Hospitality Quarterly*, Vol.51, pp.483–491 (2010).
- [2] Kang, H., Yoo, S.J. and Han, D.: Senti-lexicon and Improved Naive Bayes Algorithms for Sentiment Analysis of Restaurant Reviews, *Expert Systems with Applications*, Vol.39, pp.6000–6010 (2012).
- [3] 望月理香, 渡部智樹, 並河大地, 田中 清, 山田智広: 個人に合わせて“たとえ”るコンシェルジュサービスの評価検討 (O分野: 情報システム, 査読付き論文), 情報科学技術フォーラム講演論文集 (FIT2013), Vol.12, pp.137–144 (2013).
- [4] Zhang, Z., Zhang, Z., Wang, F., Law, R. and Li, D.: Factors Influencing the Effectiveness of Online Group Buying in the Restaurant Industry, *International Journal of Hospitality Management*, Vol.35, pp.237–245 (2013).
- [5] Retty グルメ, 入手先 (<https://retty.me/>).
- [6] Hofmann, T.: Probabilistic Latent Semantic Analysis, *Proc. UAI '99*, pp.289–296 (1999).
- [7] Bishop, C.: *Pattern Recognition and Machine Learning* (Information Science and Statistics), 2nd printing, Springer (2010).
- [8] 三輪 哲: 潜在クラスモデル入門 理論と方法, Vol.24, No.2, pp.345–356 (2009).
- [9] 後藤正幸, 小林 学: 入門パターン認識と機械学習, コロナ社, pp.200–206 (2014).
- [10] Train, K.E.: *Discrete Choice Methods with Simulation*, Second edition, Cambridge University Press, Cambridge (2009).
- [11] Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R.: Indexing by Latent Semantic Analysis, *J. Society for Information Science*, Vol.41, pp.391–407 (1990).
- [12] 佐藤栄作, 廣松 毅, 椿 広計: 潜在クラスモデルを利用したデータセグメンテーション, 行動計量学, No.30-1, pp.121–133 (2003).
- [13] 石垣 司, 竹中 毅, 本村陽一: 日常購買行動に関する大規模データの融合による顧客行動予測システム, 人工知能学会論文誌, Vol.26-6-D, pp.670–681 (2011).
- [14] Hofmann, T. and Puzicha, J.: Latent Class Models for Collaborative Filtering, *Proc. 16th International Joint Conference on Artificial Intelligence*, pp.688–693 (1999).
- [15] Hofmann, T.: Latent Semantic Models for Collaborative Filtering, *ACM Trans. Inf. Sys.*, Vol.22, No.1, pp.89–115 (2004).
- [16] 大井貴裕, 三川健太, 後藤正幸: 評価と購買の両履歴データの学習による確率的潜在クラスモデルの推定精度向上に関する一考察, 日本経営工学会論文誌, Vol.6, No.4, pp.286–293 (2015).
- [17] 原島 純, 黒橋慎夫: PLSI を用いたウェブ検索結果の要約, 言語処理学会第 16 回年次大会発表論文集, pp.118–121 (2010).
- [18] Fujiwara, N., Mikawa, K. and Goto, M.: A New Estimation Method of Latent Class Model with High Accuracy by Using Both Browsing and Purchase Histories, *15th Asia Pacific Industrial Engineering and Management Systems Conference (APIEMS 2014)*, Jeju, Korea (2014).
- [19] Goto, M., Mikawa, K., Hirasawa, S., Kobayashi, M., Suko, T. and Horii, S.: A New Latent Class Model for Analysis of Purchasing and Browsing Histories on EC Sites, *Industrial Engineering & Management Science*, Vol.14, No.4, pp.335–346 (2015).
- [20] 岩田具治, 渡辺晋司, 山田武士, 上田修功: 購買行動解析のためのトピック追跡モデル, 電子情報通信学会 D, Vol.J93-D, No.6, pp.978–987 (2010).
- [21] 滝 久雄: ぐるなびー「No.1 サイト」への道, 日本経済新聞社 (2006).
- [22] 平尾勇司: Hot Pepper ミラクル・ストーリー——リクルート式「楽しい事業」のつくり方, 東洋経済新報社 (2008).
- [23] 堀江貴文: ばかウマ, 小学館 (2014).
- [24] Dempster, A.P., Laird, N.M. and Rubin, D.B.: Maximum Likelihood from Incomplete Data via The EM Algorithm, *J. Royal Statistical Society, Series B*, Vol.39, No.1, pp.1–38 (1997).
- [25] 宮川雅巳: EM アルゴリズムとその周辺, 応用統計学, Vol.16, No.1, pp.1–21 (1987).
- [26] 坂元慶行, 石黒真木夫, 北川源四郎: 情報量統計学, 共立出版 (1983).
- [27] Hoffman, T.: Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis, *SIGIR '03 Proc. 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pp.259–266 (2003).
- [28] Si, L. and Jin, R.: Flexible Mixture Model for Collaborative Filtering, *ICML '03 Proc. 20th International Conference on International Conference on Machine Learning*, pp.704–711 (2003).

付 録

A.1 EM アルゴリズムの導出 (投稿モデル)

提案した投稿モデルでは観測できない $Z = \{z_k : 1 \leq k \leq K\}$ を含むため, EM アルゴリズム [24], [25] によるパラメータの推定を行う. まず E-step では, 各パラメータ $P(z_k), P(x_i|z_k), P(s_q|z_k), P(r_t|z_k)$ が与えられたもとの潜在クラスの事後確率 $P(z_k|x_i, s_q, r_t)$ を計算する. そして M-step では, E-step で推定された事後確率 $P(z_k|x_i, s_q, r_t)$ を用いて期待対数尤度 LL_1 を最大化することにより各パ

ラメータを更新する. このように上記2つのステップを対数尤度関数が収束するまで繰り返して計算し, 投稿モデルのパラメータを推定する.

まず E-step では, 発信ユーザ, レストラン, おすすめ度に関するパラメータを固定したときの潜在変数の分布 $P(z_k|x_i, s_q, r_t)$ は以下の式 (A.1) により求める.

$$P(z_k|x_i, s_q, r_t) = \frac{P(z_k)P(x_i|z_k)P(s_q|z_k)P(r_t|z_k)}{\sum_k P(z_k)P(x_i|z_k)P(s_q|z_k)P(r_t|z_k)} \quad (\text{A.1})$$

次に, M-step では, 各パラメータを推定するために以下のように展開する.

$$\begin{aligned} LL_1 &= \sum_i \sum_q \sum_t n(x_i, s_q, r_t) \\ &\quad \cdot \log \sum_k P(z_k)P(x_i|z_k)P(s_q|z_k)P(r_t|z_k) \\ &= \sum_i \sum_q \sum_t n(x_i, s_q, r_t) \\ &\quad \cdot \log \sum_k P(z_k|x_i, s_q, r_t) \\ &\quad \cdot \frac{P(z_k)P(x_i|z_k)P(s_q|z_k)P(r_t|z_k)}{P(z_k|x_i, s_q, r_t)} \\ &\geq \sum_i \sum_q \sum_t n(x_i, s_q, r_t) \\ &\quad \cdot \sum_k P(z_k|x_i, s_q, r_t) \\ &\quad \cdot \log \frac{P(z_k)P(x_i|z_k)P(s_q|z_k)P(r_t|z_k)}{P(z_k|x_i, s_q, r_t)} \quad (\text{A.2}) \end{aligned}$$

式 (A.2) の右辺第1項から第2項への展開は Jensen の不等式による. 式 (A.2) の右辺をさらに変形し, 定数項を省略したものを式 (A.3) とし, これを最大化する.

$$\begin{aligned} h &= \sum_i \sum_q \sum_t n(x_i, s_q, r_t) \\ &\quad \cdot \sum_k P(x_i, s_q, r_t) \log P(z_k)P(x_i|z_k)P(s_q|z_k)P(r_t|z_k) \quad (\text{A.3}) \end{aligned}$$

ただし, 確率の和は1であるという制約条件があるため, ラグランジュの未定乗数法を用いる. 具体的には, 未定乗数を $\pi, \lambda_k, \eta_k, \xi_k$ とし, ラグランジュ関数を以下のように定義する.

$$\begin{aligned} g &= h - \pi \left(1 - \sum_k P(z_k) \right) - \sum_k \lambda_k \left(1 - \sum_i P(x_i|z_k) \right) \\ &\quad - \sum_k \eta_k \left(1 - \sum_q P(s_q|z_k) \right) - \sum_k \xi_k \left(1 - \sum_t P(r_t|z_k) \right) \quad (\text{A.4}) \end{aligned}$$

式 (A.4) を偏微分して0とおくことにより, 各パラメータの推定式は以下の式 (A.5)~(A.8) 表される.

$$P(z_k) = \frac{\sum_{i,q,t} n(x_i, s_q, r_t) P(z_k|x_i, s_q, r_t)}{\sum_{i,q,t,k} n(x_i, s_q, r_t) P(z_k|x_i, s_q, r_t)} \quad (\text{A.5})$$

$$P(x_i|z_k) = \frac{\sum_{q,t} n(x_i, s_q, r_t) P(z_k|x_i, s_q, r_t)}{\sum_{i,q,t} n(x_i, s_q, r_t) P(z_k|x_i, s_q, r_t)} \quad (\text{A.6})$$

$$P(s_q|z_k) = \frac{\sum_{i,t} n(x_i, s_q, r_t) P(z_k|x_i, s_q, r_t)}{\sum_{i,q,t} n(x_i, s_q, r_t) P(z_k|x_i, s_q, r_t)} \quad (\text{A.7})$$

$$P(r_t|z_k) = \frac{\sum_{i,q} n(x_i, s_q, r_t) P(z_k|x_i, s_q, r_t)}{\sum_{i,q,t} n(x_i, s_q, r_t) P(z_k|x_i, s_q, r_t)} \quad (\text{A.8})$$

以上により EM アルゴリズムの各更新式が導出された. 式 (A.2) で表される対数尤度関数が収束するまで各ステップを繰り返せばよい.



劉 佩潔

1993年生. 2017年早稲田大学創造理工学部経営システム工学科卒業. 現在, ヤフー株式会社. 機械学習を用いたデータ分析に関する研究に興味を持つ.



山下 遥

1987年生. 2010年東京理科大学理工学部経営工学科卒業. 2012年慶應義塾大学大学院修士課程修了. 2015年同大学大学院博士課程修了. 博士(工学). 2015~2017年早稲田大学創造理工学部助手. 現在, 上智大学理工学部助教. 品質管理, 統計学, 情報工学を融合させた新たなデータ解析方法に関する研究に従事. 応用統計学会, 日本経営工学会, 日本品質管理学会等, 各会員.



岩永 二郎

1982年生. 2006年早稲田大学理工学部数理科学科卒業. 2008年同大学大学院理工学研究科数理科学専攻修士課程修了. 2008年株式会社数理システム(現NTTデータ数理システム)入社. 研究員として数理最適化, 自然言語処理, 機械学習, 統計学をビジネスに応用するBtoBビジネスに従事. 2016年Retty株式会社入社. データサイエンティストとして数理科学全般を活かした事業課題の解決, およびBtoCビジネスに従事.



樽石 将人

1978年生。2000年電気通信大学電気通信学部情報工学科卒業。2002年同大学大学院電気通信学研究科情報工学専攻修士課程修了。工学修士(情報)。レッドハット日本法人に入社し、インターネット基盤の開発に従事。VA

Linux Systems Japan 株式会社, グーグル日本法人, 楽天株式会社を経て, 2014年 Retty 株式会社 CTO, 現在に至る。



後藤 正幸 (正会員)

1969年生。1994年武蔵工業大学大学院工学研究科修士課程修了。2000年早稲田大学大学院理工学研究科博士課程修了。博士(工学)。1997年早稲田大学理工学部助手。2000年東京大学大学院工学系研究科助手。2002年武

蔵工業大学環境情報学部助教授。2008年早稲田大学創造理工学部経営システム工学科准教授。2011年同大教授。情報数理応用とデータサイエンスの研究に従事。著書に、『入門パターン認識と機械学習』, コロナ社(2014), 『ビジネス統計 統計基礎とエクセル分析』, オデッセイコミュニケーションズ(2015)等。IEEE, 電子情報通信学会, 人工知能学会, 日本経営工学会, 日本オペレーションズ・リサーチ学会, 経営情報学会等, 各会員。