

# 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分析

1X13C130-0 劉 佩潔  
指導教員 後藤 正幸

## 1 研究背景・目的

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ていいる。グルメサービス上で、ユーザはレストラン検索に加えて、レストランに対する評価や推薦投稿、他ユーザが投稿した推薦記事に対してリアクション (e.g., 「いいね」) などをすることができる。実際にユーザがグルメサービス上でレストランを探す場合、他ユーザが投稿した推薦記事を参考にする事が多い。その際、投稿した推薦記事に対するリアクション数が多いほど、レストランを探すユーザにとっての判断材料になり得るものと思われる。また、レストラン推薦記事を投稿したユーザにとっても、多くのユーザに共感されることにより、次の記事を投稿するモチベーションに繋がる。このように、リアクション数はサービス全体の活性度を表す指標であると考えられ、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選ぶユーザを受信ユーザと定義する。発信ユーザは自身で評価したいレストランを選び、そのレストランへの満足度から評価得点を決定し、その理由を推薦記事の内容に反映する。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評価得点との関係」、並びに「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」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つくと考えられる。

そこで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。その際に潜在クラスモデル [1] を用いることにより、ユーザやレストランの異質性のモデル化が可能となる。さらに、提案手法を用いて、グルメサービス A に蓄積された実データを用いて分析を行い、その結果を示す。

## 2 問題設定

グルメサービスにおいて、ユーザは主に次の2つの行動をとる。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。

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好みのレストランへ行き、レストランでの経験を基に評価得点を与え、その経験に基づいて推薦記事を投稿する。その際、リアクションの獲得がモチベーションの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、共感する推薦記事へリアクションし、利用するレストランを決定するものと考えられる。

このようにリアクションは、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、リアクション数を増加するためのモデルを提案する。

## 3 提案手法

### 3.1 概要

本研究では、2節で述べた観点から2つのモデルを構築し、発信ユーザとレストランおよび評価、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。モデル化の際に潜在クラスモデルを用い、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。提案モデルは以下の2つである。

- (1) 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評価の関係性を表現する潜在クラスモデル (モデル1)。
- (2) 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 (モデル2)。

さらに、以上の2つの潜在クラスモデルの結果を用いて、統合的な分析を行う。モデル1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、モデル2では受信ユーザが投稿に対して行なったリアクション傾向を明らかにする。さらに、モデル1とモデル2の結果を統合的に分析することで、似た発信ユーザから好まれるレストラン、似た受信ユーザから共感するレストラン推薦投稿を明らかにすることができる。

### 3.2 モデル1 (発信ユーザ, レストラン, 評価得点)

モデル1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及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  $\mathcal{R} = \{r_t : 1 \leq t \leq T\}$ 、評価得点集合を  $\mathcal{S} = \{s_q : 1 \leq q \leq Q\}$  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  $r_t$  に評価得点  $s_q$  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。

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

$N$  件のレストラン推薦記事が与えられた時の対数尤度関数  $LL_1$  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、 $N = \sum_{i,q,t} n(x_i, s_q, r_t)$  である。また、モデル1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。

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

提案モデルではEMアルゴリズム [2] によるパラメータの推定を行う。このアルゴリズムでは、対数尤度関数が収束するまでE-step, M-stepを繰り返し計算し、パラメータを推定する。更新式は以下の式 (3), 式 (4)-(7) となる。

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

M-step )

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

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

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

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

### 3.3 モデル2 (受信ユーザ, 発信ユーザ, レストラン)

モデル2では, 受信ユーザが投稿に対して行な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及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 (8) で表現する.

$$P(x_i, y_j, r_t) = \prod_{l=1}^L P(x_i | v_l) P(y_j | v_l) P(r_t | v_l) P(v_l) \quad (8)$$

$M$  回のリアクションが与えられた時の対数尤度関数  $LL_2$  は式 (9) で与えられる. ただし,  $M = \sum_{i,j,t} n(x_i, y_j, r_t)$  である. また, モデル2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.

$$LL_2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^T n(x_i, y_j, r_t) \log P(x_i, y_j, r_t) \quad (9)$$

各パラメータはモデル1と同様に EM アルゴリズムにより推定する.

### 3.4 統合分析

本研究では, 提案モデル1と提案モデル2の結果を一つの表にまとめることで, 統合的な分析を行う. 表の行はモデル1の潜在クラス, 列はモデル2の潜在クラスを表すものとする. 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まとめられた集合, 各列は受信ユーザの好む投稿の特徴によってまとめられた集合を表す. また表の各セルに対して, 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, モデル1とモデル2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 分析

### 4.1 分析概要

提案手法を用いて, グルメサービス A 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. また, 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. 2011年4月30日から2016年3月1日までにリアクション数が1以上のレストラン推薦投稿データとすると, レストラン推薦投稿数およびリアクション数は  $N = 1,122,270$ ,  $M = 18,747,857$  となる. 分析対象となるレストラン推薦投稿の発信ユーザ数, レストラン数および受信ユーザ数は  $I = 56,931$ ,  $T = 68,695$ ,  $J = 60,778$  である. また, 評価得点は三段階  $Q = 3$  であり, ユーザはレストランに対し1,2,3の評価得点をつけることができる. 本分析では, 事前分析により, モデル1とモデル2の潜在クラス数は  $K = 10$ ,  $L = 10$  とした.

### 4.2 分析結果と考察

モデル1における潜在クラス (行), 及びモデル2における潜在クラス (列) での1投稿あたりの平均リアクション数を表1に示す. また, 提案手法の適用により得られたモデル1とモデル2の潜在クラスの割合と特徴をそれぞれ表2, 表3に示す.

表1: 1投稿あたりの平均リアクション数

	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_9$	$v_{10}$
$z_1$	41	27	38	21	17	33	16	25	17	33
$z_2$	24	17	35	11	11	21	11	18	10	19
$z_3$	20	17	29	13	10	22	13	16	13	22
$z_4$	30	16	34	22	15	22	10	14	12	26
$z_5$	41	25	39	19	17	34	15	22	15	34
$z_6$	35	22	30	24	14	26	15	18	14	26
$z_7$	19	16	35	10	11	21	13	14	12	17
$z_8$	13	11	12	11	7	15	7	15	9	15
$z_9$	23	15	35	13	11	20	11	16	12	22
$z_{10}$	19	13	21	11	9	15	11	12	10	15

表2: モデル1の各潜在クラスの生起確率と結果分析

モデル1	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$	$z_7$	$z_8$	$z_9$	$z_{10}$
$P(z_k)$	33%	3%	5%	2%	9%	4%	2%	1%	13%	28%
カテゴリ	肉	欧/珈	和食	鍋/酒	和食	亜	珈	酒	珈	和食
評価得点	3.0	2.0	1.9	2.1	2.9	2.8	2.0	1.9	2.1	1.9

\*ここでのカテゴリはレストランのカテゴリとなる. \*肉: 肉料理, 欧: ヨーロッパ料理, 珈: カフェ, 酒: 居酒屋, 亜: アジア料理.

表3: モデル2の各潜在クラスの生起確率と結果分析

モデル2	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_9$	$v_{10}$
$P(v_l)$	2%	14%	3%	8%	12%	16%	20%	1%	10%	14%
カテゴリ	和食	欧	和食	和/珈	珈	肉	和食	欧	亜/肉	酒

表1の各行よりモデル1の潜在クラス間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表2より発信ユーザが記事を投稿するレストランとその評価の違いによるものと考えられる. そのため, 発信ユーザにリアクション数が多いセルに属しているレストランを推薦することで, 縦方向のセルに発信ユーザを移動させることができれば, リアクション数の増加が期待できる. さらに, 表2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つつ, レストランを推薦することが可能である.

一方, 表1の各列ではモデル2の潜在クラス間のリアクション数の差異が分かる. このリアクション数の差異は表3より受信ユーザの好むレストランと記事を投稿した発信ユーザの違いによるものと考えられる. よって, 発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推薦記事を推薦し, 横方向のセルに発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. また, レストランのカテゴリを考慮することで, 発信ユーザの好むレストランを考慮しつつ, 推薦記事を推薦することが可能である.

### 5 まとめと今後の課題

本研究では, グルメサービスにおいて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評価の関係性を明らかにする潜在クラスモデルと発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各々構築し, 結果を統合的に分析する手法を提案した.

今後の課題として, 潜在クラス数の調整が考えられる. また, 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繋がる施策の立案が課題である.

#### 参考文献

- [1] T. Hofmann, "Probabilistic Latent Semantic Analysis," *Proc. of UAI' 99*, pp.289-296, 1999.
- [2] 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.