

# グルメサービスにおける投稿データと獲得リアクション数の 関係分析のための潜在クラスモデルに関する研究

情報数理応用研究

5216C017-9 坂元哲平  
指導教員 後藤正幸

## A Study on Latent Class Model for Relational Analysis of Posted Text Data and Acquired Reactions on Social Gourmet Service

SAKAMOTO Teppei

### 1 研究背景・目的

近年、ブログや SNS を始めとして、インターネット上で一般のユーザが情報を発信することが容易になった。そして、様々なユーザから発信された情報を解析することで、新たな知見を得ようとする試みが盛んになっている。一方で、一般ユーザから投稿された情報は大規模かつ多様な嗜好が混在するような複雑な構造のデータであるため、対象問題に合わせた適切な処理が必要である。このような問題に対し、PLSA[1] や LDA[2] を代表とする潜在クラスモデルの有用性が広く示されている。潜在クラスモデルは、異質なデータが混在している母集団の分析を可能とする手法であり、観測されたデータの背後に観測できない潜在的な離散ラベルの存在を仮定している。言い換えれば、潜在クラスモデルはデータ全体が特徴の異なるグループが混在した集合体であるとして仮定した統計モデルである。潜在クラスの仮定により、特にテキストデータやマーケティングデータの解析に対して有用性が示されている。

本研究では、ソーシャルグルメサービス Retty を対象とした解析を行う。Retty 上では、ユーザはレストランの推薦記事を投稿すること、他のユーザの推薦記事を閲覧しリアクション (e.g., 「いいね」) をすること、気に入ったユーザをフォローすること等ができる。このようなサービスにおいて、ユーザから発信される情報は重要な資産であり、サービス上に良質な投稿が増加することが望まれる。また、推薦記事を投稿するユーザにとって、どのような投稿が他のユーザから共感される (リアクションされる) かということに関心が高いと言える。以上のことから、良い投稿とは何かを明らかにすることができれば、その結果を開示して投稿ユーザの興味を集めることや、サービス内での記事の表示順序やまとめ記事作成等に活用することができる。そこで本研究では、投稿に対する獲得リアクション数を投稿の良さの指標として捉え、投稿記事データと獲得リアクション数の関係性を表現する分析モデルを構築することを目指す。特に、ユーザが様々な観点から自由に記述できるテキスト情報に注目する。

一方、獲得リアクション数に対しては、投稿したユーザのフォロワー数等が大きな影響を与えることがわかっている。そのため、投稿データと獲得リアクション数のモデル化を階層的に行うことを考える。まず、第一段階で投稿ユーザのフォロワー数等を説明変数とし、各投稿のリアクション数を予測する回帰モデルを構築する。そして、実測値と予測値の差である残差に焦点を当てる。この残差は、テキスト情報とレストランの効果によるベースライン (期待される獲得リアクション数) からのずれと解釈できる。この残差とテキスト情報、レストラン情報の関係モデルが構築できれば、リアクション数の増加に結びつく要因を明らかにすることができる。しかし、テキスト情報やレストラン情報は多様であり、リアクション数に与える作用はどの投稿に対しても均一ではないと考えられる。そこで、第二段階では、残差とテキスト情報、

さらにレストランを潜在クラスモデルによってモデル化することで、これらの異質性を表現する。

以上により本研究では、回帰と潜在クラスモデルを階層的に用いることで、ユーザの投稿データと獲得リアクション数の関係分析モデルを構築する手法を提案する。最後に、提案モデル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し、提案モデルの有効性を検証するとともに、分析により得られた新たな知見を示す。

### 2 準備

#### 2.1 グルメサービスと関連研究

グルメサービスとは、主にレストランの情報を検索、閲覧できるインターネット上のサービスであり、近年では利用するレストランを決める際にグルメサービスを活用する人が増加している。本研究では、事例として Web とアプリでグルメサービスを展開する Retty を対象とし、分析には主にアプリのサービスから得られたデータを用いる。Retty に投稿される推薦記事には、レストラン情報、3段階のおすすめ度、テキスト、画像などが含まれる。また、ある種の SNS のような機能を持ち、信頼できるユーザからレストランを探すために、気に入ったユーザをフォローすることや、他の推薦記事に対してリアクション (e.g., 「いいね」) ができる。

次に、グルメサービスを対象とした従来研究について述べる。例えば、Pantelidis[3] 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を行い、レストランの改善すべき点について指摘している。Kang ら [4] は、レストランのクチコミの感情語に着目し、ラベル付与を手で行った上で機械学習手法を用いてネガポジ分類を行っている。このように、様々な観点からグルメサービス上のデータを対象とした研究が行われている一方で、本研究のように推薦記事投稿そのものに注目し、他のユーザからのリアクション数との関係を明らかにすることを目的とした研究は行われていない。

#### 2.2 潜在クラスモデル

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変数の存在を仮定したモデルである。潜在変数の仮定は、異質なデータが混ざったような複雑で現実的な問題の分析を可能とする。言い換えれば、潜在クラスモデルは、データ全体が複数の特徴の異なるグループによる集合体であるとして仮定した統計モデルである。文書データにおける文書と単語や、購買履歴データにおけるユーザとアイテムへの適用については盛んに研究が行われ、その有用性が示されている。特に、文書データを対象とするときにはトピックモデルとも呼ばれる。

最も基本的なトピックモデルは、文書がある潜在クラスのもとで生起すると仮定したシングルトピックの混合ユニグラムモデル [5] である。これに対し、同一文書中の単語が複数の潜在クラスから確率的に生起することを許容したマルチトピックモデルの PLSA[1] が提案されている。さらに、PLSA に対してベイズの枠組みを導入した LDA[2] がある。他にも、対象とする問題設定に合わせて、

仮定する確率分布や考慮する事象の特徴をモデルに組み込むことで様々な拡張が行われている。その中には、単語や文書といった離散変数に加えて、評価値などの連続変数をモデルに取り入れ、その関係を表現したモデル [6],[7] もあるが、これらの研究で対象となる事象では、投稿者と評価者が同一である。一方、本研究で扱う対象は投稿者と評価者が異なるため、これらのモデルをそのまま適用することができない。しかしながら、対象とするデータはテキストデータであり、また多くのレストランとの関係性を表現するという観点から、潜在クラスモデルを援用することは有用であると考えられる。

### 3 提案分析手法

#### 3.1 概要

本研究の目的は、推薦記事の投稿とリアクション数の関係分析モデルの構築である。特に、テキスト情報がリアクション数に与える影響について注目する。

まず推薦記事投稿とリアクション数を分析するためのモデルが対象とする事象の特徴とそれに対するアプローチについて述べる。リアクション数を目的変数とした予測を行う場合に、テキスト情報の影響に比べて、投稿したユーザのフォロワー数や画像枚数などの投稿の基本情報の影響が大きいことが確認されている。そのため、それらのリアクション数に影響を与えやすい基本情報とテキスト情報を同列に扱い、同時に説明変数に組み込んだ場合には、テキスト情報の効果について十分な分析が行えない可能性がある。よって本研究では、階層的にモデルを構築する。具体的には、ステップ1として、まずフォロワー数等の基本情報を入力変数としたリアクション数の予測モデルを構築する。そして、リアクション数の実測値と予測値の差である残差に着目する。ここで、この残差は基本情報の影響を排除した値であり、テキスト情報とレストランの効果によるベースラインからのズレと解釈することができる。すなわち、同じ基本情報を持つ投稿に対して、残差が正であれば良い投稿であり、残差が負であればベースラインよりもリアクションが獲得できていない投稿であるといえる。次にステップ2として残差とテキスト情報、そしてレストランの関係をモデル化する。ここで、テキスト(単語)とレストランは種類が多く、単一の関係で表すことが難しい。そこで、それらを自動的にうまくまとめながら学習することを目的とし、潜在クラスを仮定したモデルを構築する。以上のアプローチのイメージを図1に示す。

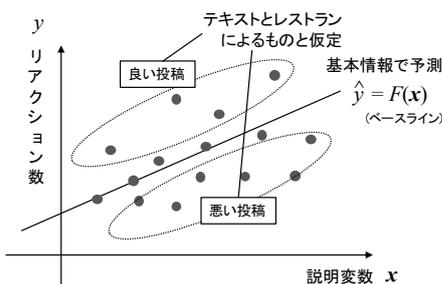


図1: 分析アプローチのイメージ

ここで、ステップ1において全学習データに対して回帰分析を行う場合、過学習を起し、ステップ2で注目する残差が過小評価されるおそれがある。すなわち、ステップ2の分析に必要な残差の変動が取り出せない。そこで、学習データを二つに分割し、それぞれステップ1とステップ2で別々に用いることを考える。これにより、過小評価されていない残差に対して潜在クラスモデルによる分析が可能となる。なお、過学習を防ぐための方法として正則化などもあるが、それに比べてデータを分割するアプローチはシンプルで有効な方法となっている。

具体的な手順として、まず全学習データをステップ1用の  $M$  件と、ステップ2用の  $N$  件に分割する。次に  $M$  件のデータを用いて予測関数  $F$  を学習する。そして、 $N$  件のデータについて、予測関数  $F$  を用いて予測値および残差を算出し、その残差とテキスト、レストランの関係を潜在クラスモデルで表現する。

#### 3.2 基本情報による回帰 (ステップ1)

ステップ1では、フォロワー数などの推薦記事の基本情報とリアクション数の関係を表現するモデルを構築する。具体的には、リアクション数を目的変数、基本情報を説明変数として回帰モデルを求める。いま、推薦記事投稿の  $D$  個の基本情報を表現するベクトル  $\mathbf{x} = (x_1, x_2, \dots, x_D)^T$  を定義し、その投稿に対するリアクション数を  $y$  と定義する。適当な回帰の予測関数  $F$  を用いれば、式(1)-(2)により投稿のリアクション数の予測値  $\hat{y}$  と残差  $r$  が得られる。

$$\hat{y} = F(\mathbf{x}) \quad (1)$$

$$r = y - \hat{y} \quad (2)$$

ここで  $\hat{y}$  は、基本情報により決まるリアクション数のベースラインと解釈することができる。残差  $r$  は基本情報では説明ができない変動部分であり、すなわちテキストとレストランの効果が含まれると仮定し、次のステップへ進む。

#### 3.3 残差と投稿データのモデル化 (ステップ2)

ステップ2では、ステップ1で得られた残差とテキスト情報、レストラン情報の関係をモデル化する。ここで、テキストやレストランは多様で表現が難しいため、潜在クラスモデルを用いる。

##### 3.3.1 定式化

まず用いる変数を定義する。単語のボキャブラリー  $V$  を  $V = \{w_i | 1 \leq i \leq I\}$ 、レストラン集合  $S$  を  $S = \{s_j | 1 \leq j \leq J\}$  と定義する。さらに、投稿の文書ベクトルを  $\mathbf{d} = (d^{w_1}, d^{w_2}, \dots, d^{w_I})^T$  と定義する。ただし、 $d^{w_i}$  は単語  $w_i$  が出現しているか否かを表す二値変数 ( $d^{w_i} \in \{0, 1\}$ ) である。また、 $K$  個の潜在クラス集合  $Z$  を  $Z = \{z_k | 1 \leq k \leq K\}$  と定義する。

このとき、ある投稿のテキスト情報、レストラン情報、残差の関係をそれらの共起として捉え、潜在クラスによって表現する確率モデルを式(3)で表す。

$$P(r, s_j, \mathbf{d}) = \sum_{k=1}^K P(z_k) P(r|z_k) P(s_j|z_k) P(\mathbf{d}|z_k) \quad (3)$$

$$P(r|z_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{(r - \mu_k)^2}{2\sigma_k^2}\right\} \quad (4)$$

$$P(\mathbf{d}|z_k) = \prod_{i=1}^I P(w_i|z_k)^{d^{w_i}} P(\bar{w}_i|z_k)^{1-d^{w_i}} \quad (5)$$

ただし、残差  $r$  については正規分布を仮定し、 $\mu_k, \sigma_k^2$  はそれぞれ潜在クラス  $z_k$  における正規分布のパラメータの平均と分散を表す。また、 $P(s_j|z_k)$  は潜在クラス  $z_k$  のもとでレストラン  $s_j$  について投稿される事象の確率を表し、多項分布で表現する。文書ベクトル  $\mathbf{d}$  の生起確率は、単語の条件付き二項分布の積として表現する。なお、 $P(w_i|z_k)$  は潜在クラス  $z_k$  のもとで単語  $w_i$  が出現する事象の確率を表し、 $P(\bar{w}_i|z_k)$  は出現しない事象の確率を表す。すなわち、 $P(w_i|z_k) + P(\bar{w}_i|z_k) = 1$  が成り立つ。

##### 3.3.2 パラメータの学習アルゴリズム

提案モデルにおけるパラメータ  $P(z_k), P(s_j|z_k), P(w_i|z_k), \mu_k, \sigma_k^2$  は、EM アルゴリズムによって推定する。

EM アルゴリズムは、尤度を極大化するように、対数尤度の潜在変数に対する期待値をとる操作 (E-step) と、それを最適化するパラメータを算出する操作 (M-step) を繰り返すことでパラメータを探索的に推定する。学習は対数尤度関数の値が収束するまで計算を行う。

ここで、全  $N$  件の推薦記事投稿のうち、 $n$  番目の投稿の基本情報ベクトルを  $\mathbf{x}_n = (x_{n1}, x_{n2}, \dots, x_{nD})^T$ 、投稿の対象となったレストランを  $a_n \in S$ 、文書ベクトルを  $\mathbf{d}_n = (d_n^{w_1}, d_n^{w_2}, \dots, d_n^{w_I})^T$  と表す。また、Step1 で予め学習済みの予測関数  $F$  と基本情報ベクトル  $\mathbf{x}_n$ 、獲得リアクション数の実測値  $y_n$  から計算される  $n$  番目の投稿の獲得リアクション数の残差を  $r_n$  と表す。このとき、全投稿の対数尤度関数  $LL$  は以下の式 (6) で表される。

$$LL = \sum_{n=1}^N \log P(r_n, a_n, \mathbf{d}_n) \quad (6)$$

提案モデルにおける EM アルゴリズムの各ステップの更新式は以下の式 (7)-(12) で計算される。

【 E-step 】

$$P(z_k | r_n, a_n, \mathbf{d}_n) = \frac{P(z_k, r_n, a_n, \mathbf{d}_n)}{\sum_{k=1}^K P(z_k, r_n, a_n, \mathbf{d}_n)} = \gamma_n^k \quad (7)$$

【 M-step 】

$$P(z_k) = \frac{1}{N} \sum_{n=1}^N \gamma_n^k \quad (8)$$

$$P(s_j | z_k) = \frac{1}{NP(z_k)} \sum_{n=1}^N \delta(a_n = s_j) \gamma_n^k \quad (9)$$

$$P(w_i | z_k) = \frac{1}{NP(z_k)} \sum_{n=1}^N d_n^{w_i} \gamma_n^k \quad (10)$$

$$\mu_k = \frac{1}{NP(z_k)} \sum_{n=1}^N r_n \gamma_n^k \quad (11)$$

$$\sigma_k^2 = \frac{1}{NP(z_k)} \sum_{n=1}^N (r_n - \mu_k)^2 \gamma_n^k \quad (12)$$

ただし、 $\delta(\cdot)$  は引数が真のときに 1、それ以外は 0 を返すインジケータ関数である。

## 4 実データ分析

本節では、グルメサービス Retty に蓄積された実データに対して提案手法を適用し、その結果を示す。

### 4.1 データセットと分析条件

本分析においては、Retty における 2016 年 3,4 月の推薦記事データ、レストランデータ、ユーザーデータを用いる。投稿データは適切な前処理を行い、各月で投稿数万件を抽出して分析対象とした。

まずステップ 1 では、3 月のデータを用いて回帰関数  $F$  を学習する。ここで、基本情報として「おすすめ度」、「画像枚数」、「フォロワー数」の 3 変数を用いた。また、予測関数  $F$  には、予測性能の良さとモデルのシンプルさの面から、ランダムフォレスト回帰 [8] を用いた。

つぎに、予測関数  $F$  を 4 月のデータに対して適用し、リアクション数の予測値を求め、予測値と実測値との残差を計算する。ステップ 2 では、この残差とレストランと文章に対して潜在クラスモデルに基づくクラスタリングを行う。このとき、対象とする単語は 4 月の投稿に頻出した名詞・動詞・形容詞に限定し、6,513 個のボキャブラリーを抽出した ( $I = 6,513$ )。なお、形態素解析には形態素解析エンジン MeCab を用い、MeCab のシステム辞書として、新語や口語に強いとされる mecab-ipadic-neologd を

用いた。また、レストランはお店のカテゴリでまとめ上げ、213 のレストランカテゴリを抽出した ( $J = 213$ )。潜在クラス数  $K$  は事前実験の結果より、AIC が最小となった 14 を用いた。

### 4.2 ステップ 1 の結果

表 1 にステップ 1 の推定結果を示す。評価は RMSE(Root Mean Squared Error) で行う。また、ランダムフォレストによって得られる変数の重要度を合計 1 に基準化した値を表 2 に示す。

表 1: 推定精度の確認

テストデータ	RMSE
3 月	19.31
4 月	40.75

表 2: 各変数の重要度

おすすめ度	画像枚数	フォロワー数
0.02	0.10	0.88

表 2 より、リアクション数に影響を与える度合いが大きい変数はフォロワー数であることがわかる。これは、獲得リアクション数は投稿が他のユーザに見られたか否かに依存するためである。また、おすすめ度の変化はリアクション数への影響が小さい。これは対象サービスの特徴として、「基本的に投稿ユーザがおすすめしたいレストランを紹介する」といったポリシーがあるため、おすすめ度があまり影響しなかったと考えられる。

### 4.3 ステップ 2 の結果

#### 4.3.1 学習結果と解釈

表 3 に学習されたパラメータを示す。単語とレストランについては、潜在クラスへの所属確率  $P(z_k | s_j)$  と  $P(z_k | w_i)$  から解釈を与えた。例えば、所属確率が高い単語がピザ、チーズ等の料理・食材そのものを表す単語であれば、潜在クラスのタイプを「料理・食材」とした。なお、表 3 は残差の平均値  $\mu_k$  で降順にソートし、クラスの番号を振り直している。すなわち、番号の小さいクラスが相対的に良い投稿のクラスという解釈を与えられる。

表 3: パラメータの学習結果

$z_k$	$P(z_k)$	$\mu_k$	$\sigma_k^2$	レストラン タイプ	単語 タイプ
$z_1$	0.02	6.36	3442.0	喫茶店	状況 他
$z_2$	0.04	6.25	4063.6	多国籍 (亜)	状況 他
$z_3$	0.05	2.07	2002.3	和菓子・パン	料理・食材
$z_4$	0.11	1.38	2709.6	食堂・和食	料理・食材
$z_5$	0.07	0.65	1873.4	ラーメン	詳細
$z_6$	0.09	-0.43	648.4	麵・中華	料理・食材
$z_7$	0.04	-2.17	3808.1	高級	ブログ
$z_8$	0.09	-2.29	10.7	多国籍	料理・食材
$z_9$	0.10	-2.76	26.0	高級 (和)	利用場面
$z_{10}$	0.03	-3.39	1628.8	スイーツ	ブログ
$z_{11}$	0.07	-3.45	1552.1	カレー	料理・食材
$z_{12}$	0.16	-3.62	1337.0	丼もの	料理・食材
$z_{13}$	0.05	-4.90	794.8	多国籍 (洋)	料理・食材
$z_{14}$	0.09	-12.1	2875.6	酒	ブログ

表 3 より、潜在クラス  $z_k$  の生起確率に偏りが少なく、特定のクラスに多くのデータが固まっているということはないことがわかる。残差の平均値  $\mu_k$  を見ると、各クラスで値が異なることがわかり、その値の変化とレストラン、単語の関係性を見ることでいくつかの傾向を把握することができる。

まず、残差の平均値が高いクラス 1 と 2 では、文書中の単語は料理・食材の名前ではなく、それ以外の要素 (店内の様子や、状況など) を表す単語の所属確率が高い。一方でクラス 11 ~ 13 を始めとして、残差の平均値がマイナスのクラスの多くでは、単語について見ると、料理・食材の名前そのものを表す単語の所属確率が高かった。す

なわち、推薦記事中には料理・食材についての言及だけでなく、状況や空間などについても言及することが良い投稿に繋がることが示唆される。

レストランについては、全体的な傾向としては低価格で庶民派なレストランほどリアクションが付きやすく、高級なレストランやなじみの薄いジャンルのレストランにはリアクションが付きにくい傾向にあると指摘できる。

最後に、分散パラメータ  $\sigma_k^2$  について考察する。いくつかの潜在クラスを除いて、分散の値が大きい。これは、提案モデルが残差の正規分布とレストランの多項分布、単語の二項分布を同等に扱っていることによるものと考えられる。すなわち、残差に対して重点的にフィットするようなモデルではないため、正規分布の分散を大きくすることで、全体として尤度が上がるようにパラメータの学習が行われるものと考えられる。

#### 4.3.2 パラメータ値を用いた分析

次に、得られたパラメータをもとに、単語に注目した分析を行う。ここでは、レストランジャンルによらず推薦記事内で一般に使われる単語を対象とすることを目的とし、出現回数が上位5%の単語のみに注目し、それぞれの単語について、単語の出現確率  $P(w_i|z_k)$  と残差の平均値  $\mu_k$  について順位相関係数を算出する。相関係数が高ければ、その単語が出現する投稿はベースラインよりもリアクションを獲得しやすいと解釈できる。表4に相関係数の値の上位と下位10単語をそれぞれ示す。

表4:  $P(w_i|z_k)$  と  $\mu_k$  の相関係数

上位		下位	
単語	相関係数	単語	相関係数
一番	0.75	リーズナブル	-0.67
違う	0.71	すごい	-0.66
前	0.69	得	-0.64
今	0.67	おすすめ	-0.61
頃	0.67	腹一杯	-0.61
店舗	0.66	絶品	-0.61
味わい	0.65	写真	-0.60
他	0.65	オススメ	-0.60
口	0.64	前菜	-0.60
強い	0.64	たくさん	-0.59

表4より、相関係数の上位の単語にいくつかの共通点がある。「一番」「違う」「他」という単語は、何かと比較するときに用いる単語である。次に、「店舗」については場所を、「前」「今」「頃」については時間を表す単語である。これらの時間や場所を表す単語と比較を表す単語を推薦記事に書くことが、投稿を見るユーザがリアクションしたくなることに繋がると考えられる。また、「味わい」「口」「強い」というのは味に関係する単語である。グルメサービスという特性もあり、味を伝えることが重要な要素であることが考えられる。一方、下位の単語に注目したときにも、いくつかの共通点がある。「リーズナブル」「得」「腹一杯」「たくさん」はお得感を表す単語である。すなわち、お得感についてはあまりリアクションにつながらない傾向にあると言える。また、推薦記事内では「おすすめ」という直接的な表記をせず、他の要素によってレストランを推薦することが好まれると示唆される。

次に、クラスごとの文書の長さ  $length_k$  と残差の平均値  $\mu_k$  の関係性を確認する。ここでは文書の長さの期待値は  $length_k$  は次式で計算できる。

$$length_k = \sum_{i=1}^I len(w_i)P(w_i|z_k) \quad (13)$$

ただし、 $len(w_i)$  は単語  $w_i$  の文字数を表す。なお、本研究では単語が出現するか否かの2値で扱うこと、品詞や頻度を理由に扱わない単語があること等により、厳密に  $length_k$  を算出することはできない。

表5にクラスごとの文書の長さ  $length_k$  を示す。また、クラスごとの文書の長さ  $length_k$  と残差の平均値  $\mu_k$  の相関係数は0.46であった。

表5: クラスごとの文書の長さ  $length_k$

$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$	$z_7$
157.6	229.7	45.4	100.3	87.4	41.6	205.8
$z_8$	$z_9$	$z_{10}$	$z_{11}$	$z_{12}$	$z_{13}$	$z_{14}$
42.4	45.9	105.0	109.7	45.5	48.4	78.1

表5から、クラスごとに文書の長さに違いがあることがわかる。また、相関係数から文書の長さや残差の平均値にはある程度の相関関係があり、文字数が多く、情報量が十分にある推薦記事であるほどリアクションの獲得につながる傾向があると言える。

以上より、提案モデルによって得られた学習結果から、様々な知見が得られることが示された。

#### 5 まとめと今後の課題

本研究では、グルメサービスを対象事例とし、投稿記事データと獲得リアクション数の関係分析モデルを構築した。その際、投稿の基本情報とテキスト情報やレストラン情報を残差に注目することで階層的に扱い、さらに潜在クラスモデルを用いてモデル化した。

実データへの適用では、ステップ1ではフォロワー数がリアクション数に与える影響が強いことが明らかになった。さらにステップ2では、単語やレストランによってリアクション数のベースラインからのズレに違いがあることが発見され、その傾向をとらえることができた。具体的には、推薦記事には、料理・食材そのもの以外について書くことや、文章中に比較の要素を含めることが良い投稿につながることである。以上のことから、提案手法の有効性が示されたといえる。

今後の課題として、リアクション数の予測精度の向上が挙げられる。また、現状ではパラメータ推定における残差の分散  $\sigma_k^2$  が大きいのが、分散を小さくするようにできれば、分析結果の信頼性が増すと考えられる。さらに、本研究では扱わなかったユーザの特性や画像の質についても、詳細に分析を行いモデルに組み込むことが望まれる。

#### 参考文献

- [1] Hofmann, T., "Probabilistic Latent Semantic Analysis," *Proc. the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 289-296, 1999.
- [2] Blei, D. M., Ng, A. Y., and Jordan, M. I., "Latent dirichlet allocation," *Journal of machine Learning re-search*, Vol. 3, pp. 993-1022, 2003.
- [3] Pantelidis, I. S., "Electronic meal experience: A content analysis of online restaurant comments," *Cornell Hospitality Quarterly*, Vol. 51, pp. 483-491, 2010.
- [4] Kang, H., Yoo, S. J., and Han, D., "Sentilexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews," *Expert Systems with Applications*, Vol. 39, pp. 6000-6010, 2012.
- [5] Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T., "Text classification from labeled and unlabeled documents using EM," *Machine learning*, Vol. 39, pp. 103-134, 2000.
- [6] Hofmann, T., "Collaborative filtering via gaussian probabilistic latent semantic analysis," *Proc. 26th Ann. international ACM SIGIR Conf*, 2003.
- [7] Mcauliffe, J. D., & Blei, D. M., "Supervised topic models," *Advances in neural information processing systems*, pp.121-128, 2008.
- [8] Breiman, L., "Random Forests," *Machine Learning*, Vol. 45, pp. 5-32, 2001.