

# 予想材料に基づく為替の動向情報分析を目的とした ニュース記事分類手法

## Classifying News Articles with Factors for Foreign Exchange Forecast

岩崎有基<sup>1\*</sup> 松下光範<sup>1</sup>  
Yuki Iwasaki<sup>1</sup> Mitsunori Matsushita<sup>1</sup>

<sup>1</sup> 関西大学大学院  
<sup>1</sup> Kansai University

**Abstract:** The purpose of this study is to support reviewing on their own decision for foreign exchange trends. Foreign exchange trends can be predicted from economic and political events. This is a decision based on experience, and it enables improvement of accuracy by verifying the validity of own analysis. However, as the factors of foreign exchange fluctuations are diverse, it is difficult to verify the adequacy of foreign exchange trends. To solve this problem, this paper proposes a method to classify news articles based on investor analysis factor in order to grasp the relevance of news articles and foreign exchange trends.

### 1 はじめに

外国為替取引(以下、為替取引と記す)や株式売買は資産運用の一つとして広まっている。しかし、投資における資産運用には損失を被るリスクがあり、投資家はリスク軽減のために為替や株の動向を分析し、分析内容を根拠として通貨売買の意思決定を行う。

為替取引において、為替動向を分析する手法はテクニカル分析とファンダメンタル分析の2通りに大別される。テクニカル分析は為替レートの数値情報に基づいて、為替相場の変動を予測する手法であり、ファンダメンタル分析は経済や政治のニュース記事といったテキスト情報から為替相場に対する影響を見積もる手法である。これらの分析手法を統合的に用いることで投資家は為替動向の分析を行う。

ファンダメンタル分析を行う場合経済や政治に関する出来事から為替動向を予測するが、この行為は経験による判断が大きい。岩崎らの研究において熟達投資家は投資市場における外的要因を考慮しリスクテイク量(自身の資産のうちどの程度の投資を行うか)を調整していたが、非熟達投資家はどのような場合にも同じリスクテイク量をとっていた[6]。このように投資家は熟達度によって経済や政治に関する出来事といった外的要因を考慮する度合いが異なる。

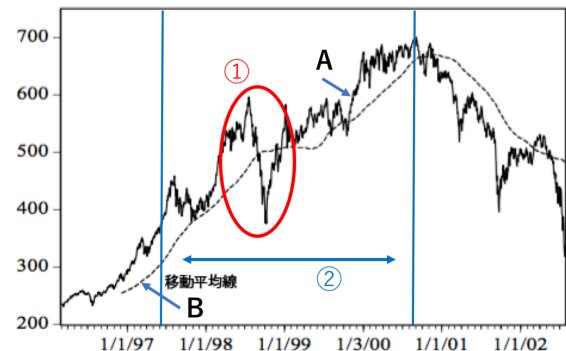


図 1: 移動平均線 (200) の例 [3] より

為替の動向分析において投資家は自身の分析精度を向上させるために分析から導き出した動向予測と実際に市場が動いた結果を照らし合わせ、自身の分析の妥当性を検証する必要がある。しかし、為替相場における要因は多様で非熟達投資家はどの要因が為替相場に影響を与えたのかを判断することが難しい。そこで、本研究は為替動向に対する自らの判断についての振り返りを支援することで、非熟達投資家の分析精度の向上を目的とする。熟達投資家と同じニュース記事の情報整理を非熟達投資家も行えるようにすることで分析精度の向上が期待される。そこで、本稿ではニュース記事と為替の関連性把握のために、ニュース記事を投資家の分類指標に基づいて分類する手法を提案する。

\*連絡先: 関西大学大学院総合情報学研究所知識情報学専攻  
〒 569-1095 大阪府高槻市霊仙寺町 2-1-1  
E-mail: k732146@kansai-u.ac.jp

## 2 為替における動向分析手法

本章ではテクニカル分析とファンダメンタル分析の詳しい説明と本研究における着眼点としてファンダメンタル分析を選択した理由を述べる。

### 2.1 テクニカル分析

テクニカル分析は過去の為替レートの値動きを用いることで未来の相場を分析する手法であり、主に数値情報を用いる。代表的な分析手法として移動平均線(図1-B)と呼ばれるものがある。移動平均線は分析困難な価格変動を平滑化し方向性(e.g., 円安への動きや円高への動き)の傾向を視覚化することができる。例えば、図1-Aを分析する際に図1-①の部分に着目すると下落方向の動きに見える。しかし、図1-②の期間を見ると下落ではなく上昇の一部である。このとき、図1-①内の点線(図1-B)に着目すると、右肩下がりでないことから下落ではなく、上昇の一部だとわかる。このように、テクニカル分析は次節で述べるファンダメンタル分析と比べ、価格変動のみに着目することで分析が容易である。

### 2.2 ファンダメンタル分析

ファンダメンタル分析はマクロ経済(e.g., インフレ, 原油価格, 失業率), 各業界特有の情報(e.g., コンペ, 需要と供給, 技術進化), 各企業特有の情報(e.g., 成長率, 業績, 訴訟事件)などの情報を用いて市場の動きを予想する分析方法である。ファンダメンタル分析では主にニュース記事のテキスト情報が用いられ、ニュース記事に記述されている経済や政治に関する出来事は為替の相場変化の要因となる。例えば、2009年のリーマン・ショックではリーマン・ブラザー・ホールディングスの倒産を契機に米ドル/円が円高に推移したように経済や政治に関する出来事は直接的に為替に影響を及ぼす。

ファンダメンタル分析はテクニカル分析に比べてニュース記事を俯瞰的に解釈する必要がある。そのため分析することが難しく、数値化することができず明確な指標が存在しない。また、ニュース記事の解釈は人により異なる。例えば、2016年の米大統領選挙ではトランプ氏の勝利により円高に推移すると予想する場合とトランプ氏の政策によりアメリカが成長するため円安に推移するといった2通りの考え方が存在した。また、ファンダメンタル分析ではマクロ経済学やミクロ経済学などの多方面の分野に精通していなければ理解が難しい。

表 1: 予想材料 (和泉ら [5] の研究より)

予想材料	元となる生データ
景気	GDP, NAPM, GNP 指数
物価	消費者物価指数, 卸売物価指数
金利	公定歩合, 長期金利
マネーサプライ	マネーサプライ
貿易収支	貿易収支
雇用	失業率, 非農業就労者数
個人消費	小売売上, 個人所得
介入	介入
要人発言	中央銀行総裁等の発言
マルク	ドルマルク, 円マルクのレート
石油	石油価格
政治	政情, 国際的な事件
株	株価
債券	債券価格
短期トレンド 1	先週の変動値
短期トレンド 2	変動値の変動値
長期トレンド	5 週間の変動

### 2.3 本研究の立ち位置

テクニカル分析やファンダメンタル分析は一方の分析方法のみ使用して為替動向の分析を行うことができる。しかし、テクニカル分析は統計的な手法のため過去にない動きなどは予想できない。また、ファンダメンタル分析は突発的な事象による為替動向の予測はできるが、テクニカル分析よりも大雑把にしか予測ができない。そのため、これらを統合的に用いることが為替動向の分析では重要である。

ファンダメンタル分析はテクニカル分析と比べ、明確な指標がなく、俯瞰的に解釈する必要があるため分析が難しい。そこで、本研究ではファンダメンタル分析に焦点を当てて支援を行うことで、非熟達投資家の分析の精度の向上を図る。

## 3 関連研究

外国為替取引市場において、投資家は大量のニュース記事などを用いて分析する。このような背景からニュース記事を為替動向の予測のために用いている研究は多い。Deshらは短い期間において為替の動向をリアルタイムのニュースヘッドラインから予測するテキストマイニング技術を提案した [4]。ニュースヘッドラインは最も重要なニュースの要約が記述されている。そのためニュースヘッドラインから予測を行うことが可能となっている。例えば、ニュースヘッドラインには“BOJ SEEN KEY CALL RATE STEADY ON THURSDAY” と

表 2: スクレイピングしたデータセット

head	date	body
米 F R B、2 回のレポで...	2008 年 01 月 1 日 10:11 JST	[ニューヨーク 3 1 日 ロイター] 米連邦準備理事会 (F R B) 傘下の...
前タイ首相派の国民の力党...	2008 年 01 月 1 日 10:20 JST	[バンコク 3 1 日 ロイター] 2 3 日のタイ下院選挙で第一党と...
北朝鮮、期限までに核計...	2008 年 01 月 1 日 10:25 JST	[ワシントン 3 1 日 ロイター] ケーシー米 국무省副報道官は 3 1 日、...

表 3: 整形したデータセット

head	date	body
米 F R B、2 回のレポで...	2008/01/1	米連邦準備理事会 (F R B) 傘下の...
前タイ首相派の国民の力党...	2008/01/1	2 3 日のタイ下院選挙で第一党と...
北朝鮮、期限までに核計...	2008/01/1	ケーシー米 국무省副報道官は 3 1 日、...

記述されているとする。「停滞している (STEADY)」、  
 「上昇した (UP)」などの情報からシステムは為替動向  
 を予測を行う。リアルタイムの取引に対してはこの技  
 術は有用だが、トレード後の検証時に振り返る際、ど  
 んな要因によって変動したかを詳しく知ることができ  
 ず、自身の経験として蓄積されない。そこで、本稿で  
 はニュース記事本文から為替変動の要因を抽出するこ  
 とでニュース記事を分類し、どの要因が為替変動に影  
 響を与えたのかをユーザが確認できるようにする。

株価の分析においてもニュース記事が主な判断材料  
 として使用されるため、ニュース記事を使用した予測や  
 分析といった研究も存在する。馬場らはトピックモデ  
 ルの一つである LDA (Latent Dirichlet Allocation) [1]  
 を拡張したヘテロトピックモデルを用いて株に関する  
 ニュースの分析を行う手法を提案している [7]。ニュー  
 ス記事は投資の判断材料となるような企業の情報を示  
 していることが多いことから、株価への影響を与えて  
 いる可能性が高い。そこで、馬場らは株価だけでなく、  
 ニュース記事も用いることで統合的な分析環境の構築  
 を行った。記事の分類の指標として用いているものは、  
 LDA によりニュース記事群から取得した潜在的なト  
 ピックである。ヘテロトピックモデルは LDA で取得  
 したトピックに対し株価情報を考慮してそのトピック  
 が企業に影響を与えるかの判定を行う。この研究では  
 潜在的なトピックを LDA でニュース記事から取得す  
 ることで分類を行っていたが、必ずしもそれが投資家  
 が分析する際に行う分類方法とは限らない。本稿では  
 熟達投資家の分類指標に基づいて分類を行うことで分  
 析を行うための基盤を構築する。

福元らは単語の共起関係を RF(Random Forest)[2]  
 や Jubatsu<sup>1</sup>を利用してニュース記事のカテゴリ分類を  
 行い、複数の学習アルゴリズムについて検証している

<sup>1</sup><http://jubat.us/ja/>

[8]。ニュース記事を分類する際にニュース記事をベク  
 トルデータとして扱うが、次元数が膨大になってしま  
 う問題がある。そこで福元らはシソーラスの単語の意  
 味属性を用いて共起頻度による共起行列を生成するこ  
 とで次元数の削減を図った。それらを RF や Jubatsu  
 の学習データとして用い、識別率の比較を行っている。  
 この研究では次元削除のために共起関係を用いている。  
 本研究では gensim<sup>2</sup>を使用し、ベクトル化に用いる単  
 語を出現回数から選定し、次元削除を行うことで RF  
 の学習データとして用いる。

## 4 デザイン指針

本章では、ニュース記事を分類するにあたり考慮し  
 た要因を述べる。

### 4.1 予想材料

本研究ではニュース記事を分類する指標として和泉  
 らの研究内で明らかとなっている予想材料を用いる [5]。  
 この予想材料は人工市場のマルチエージェントが考慮  
 する内容を決定するために、熟達投資家にインタビュー  
 を行い、為替相場が動く要因となるものを分類したも  
 のである。この予想材料を分類指標として用いること  
 によって熟達投資家と同じニュース記事の整理を行う  
 ことができ、ユーザの分析の効率を上げることができ  
 ると考えた。

和泉らの研究で用いられている予想材料は 17 種類  
 (表 1) がある。しかし、予想材料の一つであるマルクは  
 ドイツの旧通貨であり、1999 年に廃止され 2018 年 11  
 月現在用いられていない。また、短期トレンドや長期

<sup>2</sup><https://radimrehurek.com/gensim/> (2018/11 確認)

トレンドはテクニカル分析に用いるもので、為替レートといった価格が元となる生データである。そこで本研究では 17 種類のうち 13 種類(景気, 物価, 金利, マネーサプライ, 貿易収支, 雇用, 個人消費, 介入, 要人発言, 石油, 政治, 株, 債券)を用いてニュース記事分類を行う。

## 4.2 使用するデータ

2.2 節で述べたように、ファンダメンタル分析において経済や政治の出来事を為替動向変化の要因として使用している。個人投資家の経済や政治の出来事を知る情報源はニュース記事のため、本研究においてニュース記事を学習データとして用いている。ニュース記事分類モデル作成に用いる学習データはロイターのトップニュース 2008 年 1 月 1 日から 2009 年 12 月 31 日の 2 年分を用いる。ニュース記事は Web 上からスクレイピングを行い抽出している。取得したニュース記事の総数は 23,969 件である。表 2 にスクレイピングしてきたデータの一部を示す。1 列目にはニュース記事のタイトル, 2 列目にはニュース記事が公開された時間(以下, 公開日), 3 列目には実際に記述された記事の全文(以下, 本文)が入力されている。本文の文頭にはどの地域で最初に発行されたかを示す内容が記述されており, ニュース記事分類の要素として不必要なため, 前処理として除いている。また, 公開日の書式と何時何分に発行されたかという詳細な時間まで必要ないため除いた。これらの条件のもと, 整形したデータを用いたものが表 3 である。本研究では, このように収集し整形したニュース記事を RF の学習データとして用いる。

## 5 提案手法

和泉らの予想材料の手がかりではニュース記事内に存在するそれぞれの予想材料の特徴を全て包括できているとは言えない。本研究では予想材料ごとにそれぞれの特徴語を収録している辞書を作成し, それを元にニュース記事にラベル付けを行うことで学習データを作成した。

### 5.1 辞書の作成

ニュース記事を予想材料ごとに分類するために予想材料ごとの特徴語をまとめた辞書を作成した。辞書は予想材料の元となる生データ(表 1)を参考に, 日経シ

ソーラス<sup>3</sup>と Yahoo! ファイナンス<sup>4</sup>の経済指標カレンダーから用語を抽出した。日経シソーラスは金融や政治の用語が階層構造になっている。例えば, 金利の子には貸出金利, 金利据え置き, FF 金利などが含まれる。それらを予想材料(金利)の特徴語として辞書を作成し, 景気, 物価, マネーサプライなども同様に行い作成した。また, Yahoo!ファイナンスの経済指標カレンダーからは, 経済指標と呼ばれる政府が発表する GDP などの統計をそれぞれの特徴に入力した。しかし, この方法での辞書作成では, 政治政策の新しい用語(e.g., リーマンショック, オバマケア)や要人は任期によって変化することから, 政治と要人発言についての特徴語が日経シソーラスから抽出できなかった。そこで, 政治, 要人発言を除いた 11 種類の特徴語群を用いて記事を分類した。分類方法は, 辞書内にある予想材料の特徴語の単語の記事に含まれていれば, その予想材料の記事としている。例えば, ニュース記事中に「FF 金利」という言葉があれば, その記事は金利の予想材料に属すると考え分類を行う。11 種類のどの特徴にも属しなかった記事に政治や要人発言の予想材料が含まれると仮定し, 政治, 要人発言と思われるニュース記事を人手で分類した。人手で分類する際に, 政策に関する記述, 「~を述べた」や「~との見解を示した」といった発言と思われる記述があればそれらの条件を元にそれぞれ政治, 要人発言としている。

以上の 13 種類の記事群から, 頻出している単語の抽出を行い記事群の特徴語として辞書の拡張を行った。この拡張により「日経シソーラス」や「Yahoo!ファイナンス」から作成した用語だけでなく, 実際にニュース記事内で用いられている単語を特徴語とすることができる。異なる予想材料の特徴語が重複しないために, 特徴語が重複した場合は TF-IDF 値により, 値の大きいもののみを用いた。TF-IDF は TF(Term Frequency)と IDF(Inverse Document Frequency)を合わせたものである。TF はすべての文章にわたって頻出している単語を重要語とし, IDF はすべての文書に横断的に存在する単語の重要度は低いとしている。この TF と IDF を組み合わせることで, すべての文章に横断している単語の重要度は低く, 他の文章中に存在しない単語ほど重要度を高くできる。TF-IDF 値は scikit-learn<sup>5</sup>の tfidfVectorizer を用いて算出した。表 4 に以上の工程から作成された辞書の一部を示す。

<sup>3</sup>[urlhttp://t21.nikkei.co.jp/public/help/contract/price/20/thesaurus/](http://t21.nikkei.co.jp/public/help/contract/price/20/thesaurus/), (2018/11 確認)

<sup>4</sup>[urlhttps://info.finance.yahoo.co.jp/fx/marketcalendar/](https://info.finance.yahoo.co.jp/fx/marketcalendar/), (2018/11 確認)

<sup>5</sup><http://scikit-learn.org/>, (2018/11 確認)



表 4: 辞書のラベルと辞書データの例

ラベル	特徴語の例
景気	GDP, GNP
物価	消費者物価指数, 卸売物価指数
金利	公定歩合, 長期金利
マネーサプライ	資金移動, マネーサプライ
貿易収支	貿易収支, 貿易黒字
雇用	失業率, 非農業就労者数
個人消費	小売売上, 個人所得
介入	市場介入, ドル買い介入
要人発言	オバマ大統領, 日銀総裁
石油	原油在庫, 原油価格
政治	大統領選挙, 辞任表明
株	日経平均, ダウ
債券	社債, 個人向け国債

ニュース記事

米連邦公開市場委員会で**フェデラルファンド金利**の誘導目標を...FRB当局者は依然、経済の下振れリスクを懸念しているが、一方で原油や食料の値上がりを受けた**インフレ**圧力にも目配りしている

予想材料	景気	物価	金利	マネーサプライ	貿易収支	雇用	個人消費	介入	石油	株	債券	政治	要人発言
ラベル	1	0	1	0	0	0	0	0	0	0	0	0	0

図 2: ニュース記事へのラベル付与の例

## 5.2 辞書を用いたラベル付け

5.1 節で作成した辞書を用いて学習用のラベルを作成した。本研究では図 2 にニュース記事本文に対してラベル付けを行った例を示す。左から順番に、景気、物価、金利、マネーサプライ、貿易収支、雇用、個人消費、介入、石油、株、債券、政治、要人発言である。辞書内に存在する予想材料の特徴語がニュース記事本文に存在しているかを判定しラベル付けをしている。ニュース記事本文中に辞書内の単語が存在している場合は 1 を、存在していない場合は 0 を付与した。上記の方法でラベル付けしたところ、図 2 のようになった。この図からわかるようにニュース記事は予想材料単体のラベル(シングルラベル)だけでなく、複合してラベルが付与される(マルチラベル)ことがある。このことからニュース記事には複数の予想材料が含まれていることがわかる。そのため、それぞれの記事をマルチラベルからシングルラベルへの変換を行った。あるニュース記事が内包している予想材料のラベルの数だけニュース記事を複製し予想材料のラベルを単一に割り当てた。例えば、図 2 の場合だと景気と金利の予想材料を内包している記事なので 2 つの同じニュース記事が生成され、それぞれに景気や金利とラベル付けされる。このように 2008 年 1 月 1 日から 2009 年 12 月 31 日までの 23,968 件の記事に対してラベル付け処理を行った。

## 6 ニュース記事ごとの分類実験

本章では、5.2 節で作成したデータセットを元に scikit-learn の RF を学習させ、ニュース記事の分類を行い、正解率、適合率、再現率、F 値によるニュース記事分類モデルの評価を述べる。

### 6.1 実験の流れ

5.2 節で作成したデータセットを用い教師あり学習である RF を学習させ、学習後のモデルを使用しニュース記事の分類を行った。データセットの文は MeCab を用いて一般名詞や固有名詞を抽出しその他の動詞や形容詞、副詞可能、非自立、代名詞、助詞類接続、加えて 12 月などの数値情報などを除いて Bag-of-Words の形式にしている。また、「オバマ/大統領」といった連続して名詞が出現している場合「オバマ大統領」と名詞同士を結合している。その後これらの名詞情報からニュース記事をベクトル化することで、それぞれのニュース記事ごとの特徴量としている。その際、gensim を使用し、ベクトル作成に用いる単語はすべての記事にわたり 20 回以上出現し、5 割以上の記事に出現していないものとしている。これらを学習データ 36,819 件 (6 割)、テストデータ 24,546 件 (4 割) に分割して使用した。

RF はシード値を 0 に固定し、作成する決定木の数は 10 としている。この条件の元、上記の特徴量を RF に学習させ、テストデータの正解と予測データの比較を行った。比較を行った際の評価指標である正解率、適合率、再現率、F 値は以下としている。

$$\text{正解率} = \frac{\text{正解したデータ数}}{\text{予測した全体の記事数}} \quad (1)$$

$$\text{precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (2)$$

$$\text{recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (3)$$

$$\text{F-measure} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

上記の評価指標により算出された正解率、適合率<sup>6</sup>、再現率<sup>7</sup>、F 値<sup>8</sup>を表 5 に示す。学習データセットの分け方による正解率の違いを図るために 5 分割交差検定も行った。その平均値は 0.122 となった。これらの数値は小数点第四位を四捨五入を行い算出している。

<sup>6</sup>[http://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision\\_score.html](http://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_score.html)

<sup>7</sup>[http://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall\\_score.html](http://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html)

<sup>8</sup>[http://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1\\_score.html](http://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html)

表 5: 記事ごとの分類評価

正解率	適合率	再現率	F 値
0.113	0.425	0.113	0.179

表 7: 文ごとの分類評価

正解率	適合率	再現率	F 値
0.769	0.878	0.767	0.819

表 6: 辞書のラベルと辞書データの例

記事番号	文番号	本文	ラベル
0	0	米連邦準備理事...	00000000000000
0	1	内訳は 4 日物が...	00000000000000
0	2	受け入れ担保は...	00100000000000
0	3	2 日物は米国債...	00000000000000
0	4	応札額は 4 日物...	00010000000000

## 6.2 考察

ニュース記事ごとのラベル付けで RF に学習させたところ正解率、適合率、再現率、F 値はどれも低いものとなった。これは 5 分割交差検定の結果から学習に用いたデータセットの組合せの問題ではないことがわかる。データデータセット作成のために本文をベクトル化したのが、その際に他の特徴も含まれていたため、それらが影響し正解率などが低下したと考えられる。例えば、図 2 をラベルの数だけ複製し、それぞれに景気や金利といったラベルを付与していたが、この場合ニュース記事には景気と金利のベクトルが本文に存在している。そのため学習を行う際にノイズとなり正解率やその他の評価指標の低下につながったと考えられる。そこで、ニュース記事ごとのラベル付けではなく、ニュース記事を文ごとに分割しラベル付けを行うことで余分なベクトルデータを削減する方法での学習を行う。

## 7 文ごとの分類実験

本章では、6.2 節の考察を元にデータセットの再作成を行い、ニュース記事分類モデルを評価した。

### 7.1 実験の流れ

5.2 節ではニュース記事をラベルの数だけ複製し、それぞれに景気や金利といったラベルを付与していた。この場合それぞれの予想材料の特徴がベクトルデータに含まれてしまうため、ニュース記事を分割し、一文ごとにラベルを付与することで特徴ベクトルの削減を行った。あるニュース記事を分割しラベル付けを行った例を表 6 に示す。6.1 節のように MeCab を使用し、Bag-of-Words の形式でニュース記事文のベクトル化を行っている。その際 6.1 節と同じく、一般名詞や固有

名詞を残し、その他の動詞や形容詞、副詞可能、非自立、代名詞、助詞類接続、12 月などの数値情報などを除いて作成したデータセットを使用し RF の学習を行った。データの総数は 214,094 件あり、それを学習データ 128,454 件 (6 割)、テストデータ件 85,640 (4 割) に分割し用いている。正解率、適合率、再現率、F 値を表 7 に示す。また、5 分割交差検定を行ったところ平均値は 0.757 となった。

## 7.2 考察

7.1 節から予想材料を元にニュース記事を分類する場合、ニュース記事ごとに分類するよりも有用である。ニュース記事を文ごとに分割して分類した方がそれぞれの予想材料のラベルに適した特徴ベクトルをそれぞれの文が保持しているため、正解率、適合率、再現率、F 値が向上した。また、ニュース記事を文により分類することが可能になれば、ニュース記事をセグメンテーション化できるようになる。そうすることで、ニュース記事のどの部分が金利の予想材料なのか、景気の予想材料なのかをユーザに提示することが可能となり、ユーザはより詳しくニュース記事に記述されている為替動向の変化要因を知ることができる。以上の正解率、適合率、再現率、F 値とセグメンテーション化することができる有用性からニュース記事ごとに分類するよりも、ニュース記事の一文ごとに分類する方が有用である。

## 8 今後の展望

今後は本研究で作成したニュース記事分類モデルを一機能として実装したインタフェースの開発を行う。図 3 に理想のインタフェースを示す。このインタフェースはユーザの為替動向の分析を支援することを目的として作成する。

図 3-A には為替レートのグラフが表示され、図 3-B にはニュース記事のタイトルが表示されている。左図に表示されているグラフは 1 日ごとの高値、安値、始値、終値が表示されており、折れ線グラフより詳細な情報を確認することができる。右図にはニュース記事のタイトルが表示されているが、このタイトルをクリックすることでその記事のタイトルがハイライトされ、本文を表示することができる。また、その選択したニュース記事がグラフ上のどの位置の記事がを提示してくれる。記事本文が表示された際には本文下部に 3 種類のボタ

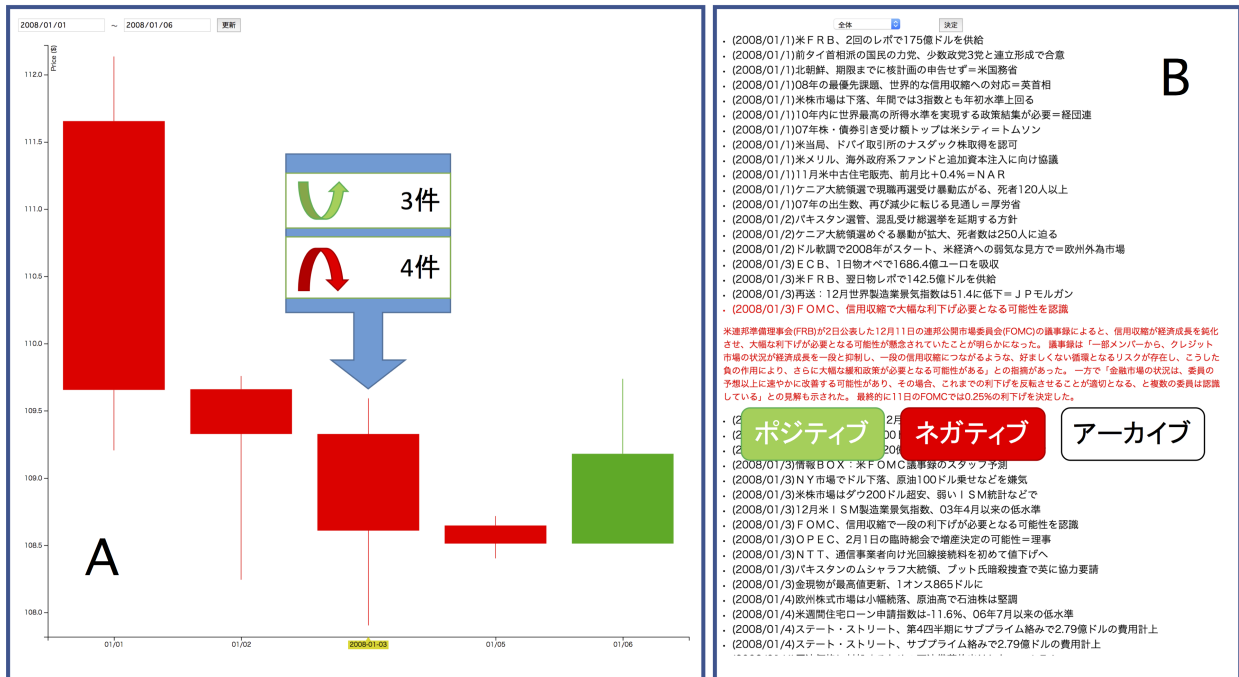


図 3: 理想のシステム図

ンが表示され、ユーザはそのニュース記事が為替動向の上昇要因か、下落要因か、それとも関係がないかをのタグを付けることができる。そうすることで、振り返りの際に図 3-A 上に対応するグラフの日付とどのようにユーザ自身がニュース記事を分類したかをシステム側から提示され、自身の思考の流れを思い返すことができる。図 3-A 上部にはプルダウン式で日付が選択でき、表示したい期間を指定することができる。この期間指定の機能はグラフの表示期間だけでなく、ニュース記事の期間も指定され、指定された期間外のニュース記事は表示されなくなる。図 3-B の上部に表示されているプルダウン部分は何の予想材料を表示したいかを選択する際に使用でき、例えば景気を選択した場合、景気予想材料を含んでいる記事が提示される。

このようなインタフェースを実装することで、実際にユーザは分析の効率を向上させられるかについての検証を行う予定である。

## 9 おわりに

本研究では為替動向に対する自らの判断についての振り返りの支援を目的に、本稿ではニュース記事と為替の関連性把握のために、ニュース記事を熟達投資家の分析指標に基づいて分類する手法を提案する。ニュース記事ごとのラベル付けを行ったデータセットとニュース記事 1 文ごとのラベル付けを行ったデータセットで

RF を学習させ、ニュース記事分類モデルを作成した。それぞれのニュース記事分類モデルを正解率、適合率、再現率、F 値で評価したところ、ニュース記事ごとのラベル付けよりもニュース記事を文ごとに分割してラベル付けしたデータセットで学習したニュース記事分類モデルの方が有用だと示唆された。

## 謝辞

本研究の遂行にあたり、文部科学省科学研究費（課題番号:15H02780）の助成を受けた。記して謝意を表す。

## 参考文献

- [1] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent dirichlet allocation, *Journal of machine Learning research*, Vol. 45, No. 1, pp. 993–1022 (2003)
- [2] Breiman, L.: Random forests, *Machine learning*, Vol. 45, No. 1, pp. 5–32 (2001)
- [3] Griffioen, G. A. W.: Technical analysis in financial markets, Doi: <http://dx.doi.org/10.2139/ssrn.566882>, University of Amsterdam (2003)

- [4] Peramunetilleke, D. and Wong, R. K.: Currency exchange rate forecasting from news headlines, *Australian Computer Science Communications*, Vol. 24, No. 2, pp. 131–139 (2002)
- [5] 和泉潔, 植田一博, 中西晶洋: 人工市場モデルによる外国為替ディーラーの学習行動の分析, 情報処理学会研究報告,1996-ICS-106, Vol. 106, No. 105, pp. 91–98 (1996)
- [6] 岩崎雄斗, 和泉潔, 伊藤祐輔, 植田一博: 投資家のリスク態度と熟達度ならびに市場動向が投資行動に与える影響, 日本認知科学学会論文誌, Vol. 22, No. 3, pp. 389–408 (2015)
- [7] 馬場慧, 馬強: 株価とニュースの統合分析のためのヘテロトピックモデル, データ工学と情報マネジメントに関するフォーラム (2017)
- [8] 福元伸也, 淵田孝康: 単語の共起関係に基づく機械学習による文書分類, 情報処理学会研究報告, Vol. 2014-DBS-160, No. 28, pp. 1–5 (2014)

# 深層学習における分類パターンの解釈支援

## Interpretation Support System for Classification Patterns in Deep learning

安藤 雅行<sup>1\*</sup> 河原 吉伸<sup>2,3</sup> 砂山 渡<sup>4</sup> 畑中 裕司<sup>4</sup>  
Masayuki ANDO<sup>1</sup> Yoshinobu KAWAHARA<sup>2,3</sup> Wataru SUNAYAMA<sup>4</sup> Yuji HATANAKA<sup>4</sup>

<sup>1</sup> 滋賀県立大学大学院工学研究科

<sup>1</sup> Graduate School of Engineering, The University of Shiga Prefecture

<sup>2</sup> 大阪大学産業科学研究所

<sup>2</sup> The Institute of Scientific and Industrial Research

<sup>3</sup> 理化学研究所革新知能統合研究センター

<sup>3</sup> RIKEN Center for Advanced Intelligence Project

<sup>4</sup> 滋賀県立大学工学部

<sup>4</sup> School of Engineering, The University of Shiga Prefecture

**Abstract:** This paper describes an interpretation support system for classification patterns based on the contents of learning results in deep learning with texts, and verified its effectiveness. It is well known that classification patterns by deep learning models are often difficult to interpret the reasons derived. The proposed system extracts the contents of learning results in deep learning with texts and provides seeds for interpretations of the patterns learned. Then, the system displays learned network structures so that anyone can easily understand learning results. In verification experiments to confirm the effectiveness of the system, based on the learning result of deep learning classifying sentences, test subjects were instructed to give meanings of classification patterns peculiar to each output. The results show that test subjects who represent novice data scientists could understand the meanings of the classification patterns of deep learning with texts.

## 1 はじめに

近年、深層学習を用いた AI システムが世の中を席卷している。画像認識を始めとし、車の自動運転や株の自動取引、医師を助ける診断アシストなどさまざまな分野において、活用が進められている。

人間が何らかの判断を行う際には根拠を問い、その判断が妥当かどうかを客観的に鑑みることが大切である。しかし現状の AI は、判断は行ってもその根拠は示さないという大きな問題点がある。この AI による判断の根拠を、人間が解釈可能な形で取り出し、「根拠」を条件部、「判断」を帰結部とする因果関係を知識として得ることができれば、人間がその知識を共有することで、関連する仕事の効率を上げるために役立てられる。そこで本研究では、テキストベースの文章分類を題材として、深層学習により学習された分類パターンを表

すネットワークの解釈を支援するシステムを提案する。

本研究の狙いは、ブラックボックスと言われている深層学習による学習結果の解釈支援にある。そのため、文章分類の精度向上を目指すアプローチではなく、人間による学習結果の解釈が可能なモデルを検討する。今回は人間の解釈は言葉によって行われるとし、単語をそのまま利用する BoW(Bag of Words) 形式の入力を扱うことで、学習結果への意味づけを目指す。また深層学習の構造としても、最も単純な DNN(Deep Neural Network) による構造を扱い、本構造における解釈を可能にすることを目指す。本研究による知見が、単語の分散表現を入力とする深層学習や、RNN(Recurrent Neural Network) など他の構造の深層学習の学習結果の解釈を支援するシステムの、ベースラインとして役立てられることを期待する。

以下本論文では、2 章で関連研究について述べる。3 章で深層学習による分類パターンの解釈支援システムの構成と詳細について述べる。4 章で提案システムの評価実験について述べ、5 章で本論文を締めくくる。

\*連絡先：滋賀県立大学大学院工学研究科 電子システム工学専攻  
安藤雅行

〒 522-8533 滋賀県彦根市八坂町 2500  
E-mail: oh23mandou@ec.usp.ac.jp



## 2 関連研究

インターネットの普及に伴い、また、SNS (Social Networking Service) の出現により、世の中の文章データが大規模になり、自分が欲しい文章や要らない文章を見分けるのが難しくなっている。そこで注目されているのが、深層学習を用いたテキストマイニングシステムである [ダヌシカ 14, Arisoy 12]。深層学習は、一般に多層から構成されるニューラルネットワークを用いた学習を指す。畳み込みニューラルネットワーク [LeCun98, Krizhevsky12] の出現により画像を用いた場合に限らず多くの場面で高い分類性能を実現できることが報告されている。

その一方で、深層学習は、その出力を導いた根拠についての解釈が困難であることも知られている。畳み込みニューラルネットワークを用いた画像認識においては、この問題に対する研究も最近進められており、いくつかの方法が提案されている。例えば、入力画像に対応する畳み込みニューラルネットワークにおける層間のスコアの勾配を計算することでネットワークの可視化を行う方法や [Zeiler 14, Mahendran 15]、学習済みのネットワーク中間層のノード情報を用いて、対応する画像中の画素への寄与度を計算することにより画像の分類に重要な部位を表示する方法などが提案されている [西銘 15]。また最近では、解釈性の高いリカレントニューラルネットワーク [Wu17] や Generative Adversarial Network [Chen 16] など、他のニューラルネットワークモデルにおいても議論が進んでいる。

しかし自然言語への深層学習の適用においては、上記のような画像認識における方法を直接適用できないため、その分類根拠を抽出する効果的な方法については議論が進んでいないのが現状である。そこで本研究ではこのような問題意識の下、文章 (テキスト) の分類問題を例として、深層ニューラルネットワークを用いた学習によってネットワークの各層に付けられた重みの値を抽出し、各中間層が学習した情報を単語の集合として表現する。そして、単語の組み合わせによって中間層ごとの学習された情報を解釈し、そこから分類のパターン、つまり出力を導くルールの意味付けをする手助けを行うシステムの開発を目指す。

## 3 テキストベースの深層学習における分類パターンの解釈支援システム

本章では、テキストを分類するための深層学習ネットワークにおいて、その分類パターンの解釈を支援するためのシステムについて述べる。

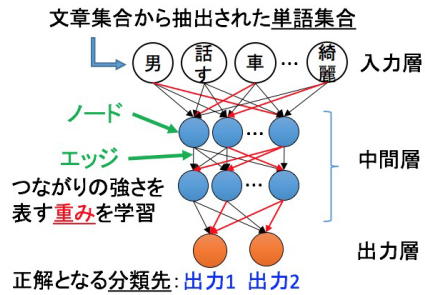


図 1: テキスト分類ネットワーク

### 3.1 深層学習によるテキスト分類

深層学習とはニューラルネットワークの中間層を 2 層以上に多段化したものを指す。本研究では、最も単純な DNN (Deep Neural Network) による深層学習を考え、テキスト集合を分類する問題に適用する場面を考える (図 1)。すなわち、正解ラベルが与えられたテキスト集合をもとに、各テキストの単語ベクトル (テキスト中の名詞、動詞、形容詞を対象としてその出現の有無を 0,1 で表す) を入力として、中間層が 3 層の全結合型の DNN におけるエッジの重みをその分類精度が高くなるように学習させる。学習された分類パターンを解釈するために、適切な学習 (学習精度が 90% 以上) が行われていることを前提としている。

本研究においては、ここで学習されたテキストの分類パターンを表す学習ネットワークをもとに、その分類パターンを人間が解釈することの支援を行う。

### 3.2 テキスト分類パターン解釈支援システムの構成

提案するテキスト分類パターン解釈支援システムの構成を図 2 に示す。まず、深層学習により学習された分類パターンを表すネットワークを入力とする。次に、入力されたネットワークの中から、特定の出力に結びつく部分ネットワークをパスとして抽出し、パス上のノードにテキスト中の単語を用いたラベルづけを行う。その後、これらのパスとラベルを可視化した分類パターンの解釈支援機能を、ユーザに提供する。ユーザは、可視化される重要パスとパス上のラベル、ラベルの単語が出現する原文、などを参照しながら、分類パターンに解釈を与える。

また本研究では、「ある単語を含むと、ある出力に分類される」という分類パターンを解釈に利用することを仮定する。これは、テキスト中で用いられている単語情報と、分類先となる出力との関係を表す最もシンプルな表現として設定する。



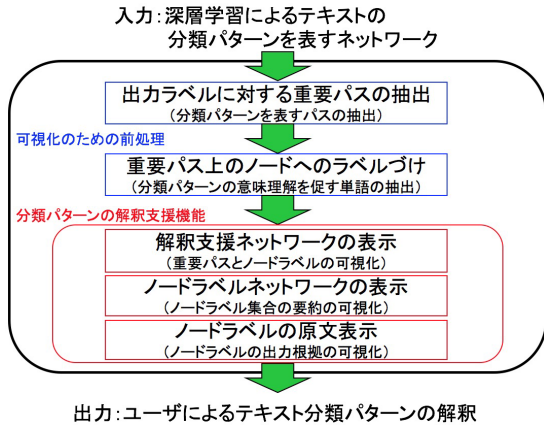


図 2: テキスト分類パターン解釈支援システムの構成

### 3.3 出力ラベルに対する重要パスの抽出

学習ネットワークにおいて、ある出力を導く部分ネットワークの中で、分類に大きく寄与するパスを抽出する。ここでパスとは、入力層のノードから指定の出力ノードまで各中間層のノードを1つずつ辿った一本道のネットワークを表す。入力層のノード数を  $n_0$ 、第一中間層、第二中間層、第三中間層のノード数をそれぞれ  $n_1, n_2, n_3$  としたとき、ある出力層へのパスの総数は、 $n_0 \times n_1 \times n_2 \times n_3$  本となる。

ノード  $i$  からノード  $j$  へのエッジ  $e(i, j)$  の重みを  $W_{i,j}$  で表すとす。このとき、ノード  $i$  からノード  $j$  を経由してノード  $k$  に至るパス  $path = \{e(i, j), e(j, k)\}$  の重要度を、パスに含まれるエッジの重みの積として、 $W_{i,j} \times W_{j,k}$  で定義する。

すなわち、パス  $path$  の重要度  $\text{Imp}(path)$  を、式 (1) で定義する。

$$\text{Imp}(path) = \prod_{e(i,j) \in path} W_{i,j} \quad (1)$$

エッジの重みには負の値を持つものもあるが、重要度は重みの積として計算されるため、最終的に重要度の正の値の大きい順に、重要なパスとして抽出する。

### 3.4 重要パス上のノードへのラベルづけ

本節では、前節の方法により抽出された重要パス上の中間層ノードの意味を表すラベルを付与する方法について述べる。

重要パスが含む中間層のノードについて、各ノードと強いつながりのある入力層のノードが表す単語を、つながりの強い順に抽出し、抽出された単語集合を、各中間層ノードのラベルとする。

各ノードと強いつながりのある入力層のノードの抽出には、前節で述べた重要パス抽出の方法を転用する。

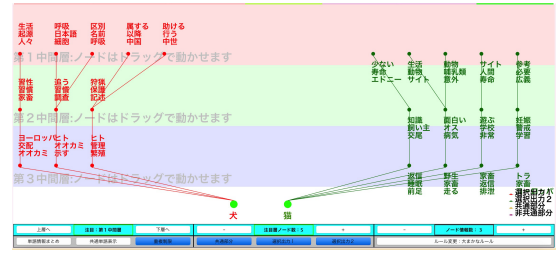


図 3: 解釈支援ネットワークの表示例

すなわち、中間層ノード  $m$  における入力層ノード  $w$  の重要度  $\text{Label}(w, m)$  を、 $w$  から  $m$  に至るすべての部分パス  $kpath$  の重要度の和として、式 (2) で定義する。ただし式中の  $\text{Allkpath}(w, m)$  は、入力層ノード  $w$  から中間層ノード  $m$  に至るすべての部分パスの集合を表す。

$$\text{Label}(w, m) = \sum_{kpath \in \text{Allkpath}(w, m)} \text{Imp}(kpath) \quad (2)$$

この重要度の値の高い順に、指定された数だけ入力層ノードの単語を取り出して、中間層ノードのラベルとする。

### 3.5 解釈支援ネットワークの表示

前節までに抽出した、ある出力を導く重要パスと、パス上のノードのラベルを可視化した解釈支援ネットワークを表示する。図 3 に解釈支援ネットワークの表示例を示す。3つの中間層と出力層の4つの領域ごとに色分けを行い、最下部を出力層としてネットワークを表示する。

ユーザは初めに、インターフェースの下部に並べられた出力ラベルから、1つまたは2つの出力ラベルを選択する。システムは選択された出力ラベルに対して、指定数の重要パスを表示する (図 3 の例では 5 本)。またパス上のノードに対して、ノードラベルを指定した数を最大として表示する (図 3 の例では 3 個)。ノードラベルは、一つのパス上で同じ単語が重複して現れる場合は、出力層に近いノードにのみ表示させる。ユーザは、出力されたネットワークを眺めることで、どのような単語の利用が出力を導いているかのパターンを見つけ、その解釈を行う。

### 3.6 ノードラベルの原文表示

分類パターンの解釈に向けて、単語情報だけでは、その単語が実際にどのような文脈で使われていたのかを把握することは難しい。そのため、ノードラベルとし

犬はもともと **オオカミ** と同じように、群れで生活する動物です。

は広くイヌ科に属する動物 (イエイヌ、**オオカミ**、コヨーテ、ジャッカル、キツネ、タヌキ) の方向転換で能として働いたが、**オオカミ** などと比べると細く短く、また、日中は **オオカミ** の2年に比べて早熟である。

動物であり、手に仔犬 (イヌか **オオカミ** かはつきりしない) を持たせて埋葬された、1万5000年前に **オオカミ** から分化したと推定されている。

図 4: ノードラベル (オオカミ) の原文表示例

て出力された単語が、学習に用いたテキスト内で実際にどのように使われているかを表示させる。

図 4 にノードラベルの原文表示例を示す。ユーザは、解釈支援ネットワーク上で単語をクリックすることで、その単語を含む原文を表示させて参照することができる。

### 3.7 分類パターンの比較支援機能

分類パターンの解釈において、特定の出力を導く分類パターンの解釈以外に、複数の出力の分類パターンを比較したい場合も想定される。そのため本システムにおいては、2つの出力に対するそれぞれの重要パスを比較するための機能を設ける。

解釈支援ネットワークにおいては、2つの出力ラベルに対する重要パスをノードラベルを、それぞれ赤と緑で可視化すると共に、2つの出力ラベルに対する共通の重要パスを茶色で表示する。

これらの機能によって、2つの出力を導く分類パターンについて、それぞれの出力に独自の分類パターンの解釈を支援する。

## 4 テキスト分類パターン解釈支援システムの有効性の検証実験

本章では、提案システムが、分類パターンの解釈に有効に用いられるかを確認するために行った実験について述べる。本実験では、理系の大学生、大学院生 14 名に、システムの利用方法について理解してもらった後、3種類のデータ分析課題を与え、分類パターンに興味づけを行ってもらった。

### 4.1 実験手順

#### 4.1.1 実験手順概要

実験の手順を以下に示す。

1. サンプルデータの分析: システムの操作を一通り試してもらう。

2. 課題 1 「動物」: 動物の生態を説明するテキストの分類: 2種類の動物の生態の違いをまとめてもらう。
3. 課題 2 「映画」: 映画「シン・ゴジラ」のレビュー記事の分類: 映画「シン・ゴジラ」の評価されている箇所と、評価されていない箇所を分析してもらう。
4. 課題 3 「ツイート」: 受験にまつわるツイートの分類: 受験にまつわるツイートの中で、どのようなツイートが、より「いいね」をもらいやすいかを分析してもらう。

#### 4.1.2 実験手順詳細

3つの課題のそれぞれについて、以下のステップで分類パターンに興味づけを行ってもらった。

- 1) 2種類の出力ラベルを選択する: 「動物」課題では2種類の動物を被験者に選んでもらった。「映画」課題では評価が5のレビューと、評価が1または2のレビュー、「ツイート」課題では、いいねが11以上のツイートと、いいねが0のツイートとした。
- 2) 3つの中間層のそれぞれに表示されている単語を元に、それぞれの出力につながるルールを推定して答えてもらう。
- 3) それぞれの出力につながる5つのパスとパス上の単語を元に、各出力につながるルールを推定して答えてもらう。
- 4) ステップ 2) とステップ 3) で答えた内容を整理して、2種類の出力の相違点についてまとめてもらう。

### 4.2 実験データ詳細

学習データは、課題「動物」においては、50種類の動物において、動物名と「生態」というキーワードで検索エンジンによる検索結果の上位の Web ページから、100件ずつ、合計 5000 件のテキストを利用した。課題「映画」においては、映画レビューサイト<sup>1</sup>において、映画「シン・ゴジラ」の映画レビューの項目から、レビュー記事と評価値を 1000 件 (評価 1 と 2 を評価 153 件、評価 3 を 119 件、評価 4 を 330 件、評価 5 を 398 件) 利用した。課題「ツイート」においては、キーワード「受験」で検索した Twitter<sup>2</sup>のツイートから、

<sup>1</sup>映画.com (URL) <http://eiga.com>

<sup>2</sup>Twitter: (URL) <http://twitter.com>

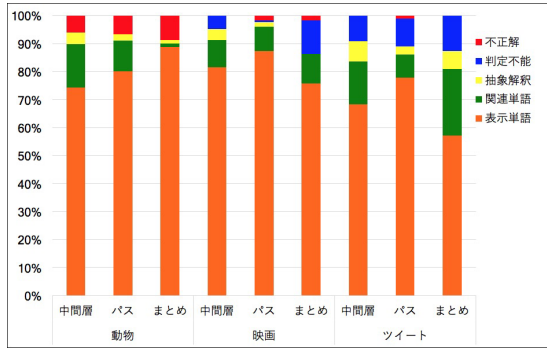


図 5: 被験者に記述された解釈の内訳 (被験者平均)

フォロワー数が 100 人以上のユーザーのツイートを 6000 件 (いいね 0 個, いいね 1 から 10 個, いいね 11 個以上をそれぞれ 2000 件) を利用した。学習は全結合型のディープニューラルネットワークによって行い, いずれの課題も中間層は 3 層とした。

### 4.3 実験結果と考察

被験者により記述された解釈の内訳を図 5 に示す。ただし, 表に示す解釈の内訳は以下に定義する内容をもとに, 著者の 1 名が分類を行った。また, 実験手順詳細のステップ 2) から 4) に対応する, 中間層, パス, まとめごとに被験者平均を算出している。

- 表示単語: 提案システムが表示する単語を使用しており, その内容の正しさが原文から確認できる。
- 関連単語: 表示単語は使われていないが, 表示単語が使われている原文に含まれる単語を使用しており, その正しさが原文から確認できる。
- 抽象解釈: 表示単語, および表示単語が使われている原文に含まれる単語が使われていないが, その内容の正しさが原文から確認できる。
- 判定不能: 解釈の正しさが判定できない。
- 不正解: 解釈の内容が確実に誤っている。

図 5 の結果から, 多くの解釈は「不正解」と「判定不能」以外に分類され, その正しさが確認できており, 実際に当てはまる事例を具体的な知識として捉えることができていた。このことから, 中間層とパスの解釈において, 本システムが出力する単語情報まとめと意味づけ支援ネットワークが, 分類パターンの解釈に役立てられたと考えられる。また表 5 において, 「表示単語」だけでなく「関連単語」や「抽象解釈」に分類される解釈も一定の割合で存在しており, システムが提示する単語を活用した解釈が行われていたことがわかる。

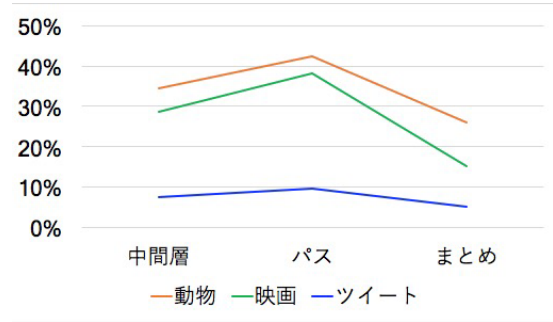


図 6: 被験者の解釈が当てはまる原文の割合 (被験者平均)

課題ごとの比較として, 「動物」の課題では不正解が多く, 「映画」と「ツイート」の課題では判定不能が多くなった。これは動物の生態はその正しさが確認しやすかったのに対して, レビューやいいねの数などの人間の感覚に依存する評価においては, 正しさの絶対的な基準がなかったことによると思われる。そのため, 即座にその正しさを確認できない解釈について, その判断を支援する情報を提供できれば, より効果的なシステムに改善できると考えられる。

また課題「映画」と「ツイート」においては, 中間層とパスの解釈に比べ, まとめの解釈において表示単語の割合が少なくなり, 判定不能に分類される解釈が多くなった。このことから, 個別の中間層やパスの解釈をまとめる際に, 一定の抽象化が行われ, 表示単語そのもの以外の言葉による解釈が行われていたことがわかる。そのため, 複数の解釈を集めてまとめることは, より汎用的な知識を得る上で一定の意味があると考えられるが, 先の考察と同様に, その正しさを検証できる仕組みが必要と考えられる。

被験者の解釈を著者らがまとめたところ, 課題「映画」においては, 『音楽や映像などの演出, 映画が扱っているテーマ, については高い評価が得られ, ストーリーやセリフ, これまでのシリーズとの違い, については低い評価が得られた』となる。また課題「ツイート」においては, 『近況報告, 人との関わり, 自分の心境, のツイートについては「いいね」をもらいやすく, 勉強方法や, 広告, のツイートは「いいね」がもらえない』となる。特に直感的には「いいね」がもらえらえると思われる勉強方法のツイートは, かえっていいねに結びつかないという興味深い結果も得られている。

実際の分析の場面で同様のまとめを得るためには, 一人の分析者が主要な複数の解釈を列挙してまとめる道筋を用意することや, 分析の初心者であっても複数人が分析を行うことで, 一人のメタ分析者が結果をまとめあげる形式をとることで, 有効な知見を得られる可能性もあると考えている。

被験者の解釈が当てはまる原文の割合を、図6に示す。ただし集計対象は、先の解釈の内訳で「表示単語」「関連単語」に分類されたもののみとし、それらの単語を含む原文の数を、各課題ごとに2種類の出力ラベルの学習に用いたテキスト数<sup>3</sup>で除した値を、解釈が当てはまる原文の割合としている<sup>4</sup>。ただし、図6の中間層の項目では、3つの中間層に対する解釈が当てはまる原文の合計を、パスの項目では、5つのパスに対する解釈が当てはまる原文の合計を、まとめの項目ではまとめとして書かれた1つの解釈が当てはまる原文の数を用いている。

まとめの解釈においては、課題「動物」「映画」「ツイート」の順に、26%(52件)、15%(84件)、5%(208件)の原文に当てはまっている。この数値には判定不能となった解釈の数値は含まれていないこと、また先述の複数人の解釈をまとめる方法をとることなどにより、この割合を増やしていける可能性があることから、本システムによって、学習データとなったテキスト集合に対して一定の解釈を与える支援を行うことができたと考えている。

課題「動物」「映画」「ツイート」の順に、まとめの解釈が原文に当てはまる割合が下がったのは、テキストが対象とする内容の範囲の広さによるものと考えられる。すなわち、課題「動物」に用いたテキストは特定の動物の生態についての話であり、内容には一定のまとまりがあると考えられるのに対して、「映画」は特定の映画に対するレビューであるものの、個人の主観に依存する部分があるため、内容に広がりが生じたと考えられる。「ツイート」は、受験をテーマにしている以外には制約がないため、個人の主観に加え、個人の置かれている状況や環境の違いがあるため、さらに内容の幅が広くなり、全体に当てはまる解釈を導くことは難しかったと考えられる。

## 5 おわりに

本論文では、テキストベースの文章分類を題材として、深層学習により学習された分類パターンを表すネットワークの解釈を支援するシステムを提案した。提案する環境が、深層学習に精通していない、データ分析の初心者でも分類パターンの解釈につなげられることを実験により確認した。

今後は、文章分類におけるリカレントニューラルネットワークやその発展手法における学習結果の解釈や、より大規模なネットワークの解釈に繋がられる環境に改

良していきたい。また、ユーザが与えた解釈の正しさを裏付けるデータを、学習データの内外から取得してユーザに提示することで、仮説として得られた知識の検証を行える枠組みを構築していきたいと考えている。

## 参考文献

- [Arisoy 12] Ebru Arisoy, Tare N. Sainath, Brian Kingsbury, Bhuvaba Ramabhadran: Deep Neural Network Language Models, In Proceedings of the NAACLHLT Workshop, Will We Ever Really Replace the N-gram Model?, pp.20–28, 2012.
- [Chen 16] X. Chen, Y. Duan, R. Houthoof, J. Schulman, I. Sutskever, and P. Abbeel, “Info-GAN: Interpretable representation learning by information maximizing generative adversarial nets,” In Proceedings of NIPS, 2016.
- [ダヌシカ 14] ボレガラ ダヌシカ：自然言語処理のための深層学習, 人工知能学会誌, Vol.29, No.2, pp.195–201, 2014.
- [Krizhevsky12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” In Proceedings of NIPS, 2012.
- [LeCun98] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” In Proceedings of the IEEE, 1998.
- [Mahendran 15] Aravindh Mahendran and Andrea Vedaldi, “Understanding deep image representations by inverting them,” In Proceedings of CVPR, 2015.
- [西銘 15] 西銘 大喜：ディープニューラルネットワークによる画像からの表情表現の学習, 第29回人工知能学会全国大会, 3L4-3, 2015.
- [巢籠 16] 巢籠 悠輔：Deep Learning Java プログラミング 深層学習の理論と実装, 株式会社インプレス, 2016.
- [Wu17] T. Wu, X. Li, X. Song, W. Sun, L. Dong, and B. Li, “Interpretable R-CNN,” In Proceedings of arXiv:1711.05226, 2017.
- [Zeiler 14] Matthew D. Zeiler and Rob Fergus, “Visualizing and understanding convolutional networks,” In Proceedings of ECCV’14, pp.818–833, 2014.

<sup>3</sup>課題「動物」では200件、「映画」では551件、「ツイート」では4000件。

<sup>4</sup>解釈に用いられた単語が使われている原文においても、解釈が当てはまらない可能性も考慮して、一通り原文の内容を確認し、解釈の当てはまらないうまくないことを確認している。



# 食材と時季に基づいた料理レシピの人気分析

## Popularity Analysis of Cooking Recipes Based on Ingredients and Seasons

池田真大 熊野雅仁 小堀聡 木村昌弘\*

Masahiro Ikeda, Masahito Kumano, Satoshi Kobori, Masahiro Kimura

龍谷大学理工学部電子情報学科

Department of Electronics and Informatics, Ryukoku University

**Abstract:** 近年、料理レシピ共有サイトが世界的に普及しており、かつては観測が困難であった一般家庭の料理を反映して大量に投稿されたレシピ群がビッグデータとして成長するに従い、人々が料理に対して行う行動の背後に潜む原理や未知の要因を理解することを目指す上で、各地域の料理レシピに含まれる特に利用食材などの観点から人々の日常生活における料理レシピを科学的に分析する研究が注目されている。我々は、日本における一般家庭の料理に対して人々が行う行動と料理レシピとの関係に注目しており、特に、人気を得る料理レシピの要因について食材だけでなくその利用時季および順序性に注目している。そこで本研究では、日本の料理レシピ共有サイトに投稿された大量のレシピデータに着目し、人気を得る料理レシピの要因として、食材そのもの、ある時季に頻出する食材、食材の利用順序性、ある時季に頻出する食材の利用順序性という4つの観点から料理レシピの人気分析を行う。そして、日本の料理レシピ共有サイト CookPad の実データを用いた実験で、人気分析の結果を報告する。

### 1 はじめに

ソーシャルメディアの発展は、人々の日常生活に大きな影響を与えているだけでなく、実世界の人々の日常の動向を捉えるための様々な分析を行う研究へも影響を与えつつある。近年、料理レシピ共有サイトが世界的に普及しており、これまで観測が困難であった一般家庭の料理を反映して大量に投稿されたレシピ群がビッグデータとして成長するに従い、一般的な人々が料理に対して行う行動の背後に潜む原理や未知の要因を理解することを目指す上で、利用食材などの観点から人々の日常生活における料理レシピを科学的に分析する研究 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] が活発化している。

Ahn ら [2] は、食材どうして共有される香料化合物に着目したネットワークを分析し、西洋料理では香料化合物を共有した食材ペアを利用するという食材ペアリング仮説の検証を世界規模で行った。Jain ら [5] は、Ahn らの研究の応用としてインド料理の香料化合物ネットワークを分析した。Teng ら [3] は、レシピ内における共起食材のネットワークとレシピ作成における代替食材のネットワークを用い、一部の食材を代替食材に置き換えたレシピが人気を得るかを調べ、レシピ推薦へ

応用した。Sajadmanesh ら [8] は、レシピ内の食材群に着目することで、世界各地の郷土料理を判別することができる可能性を示した。これらの研究は、人々が料理に対して行う行動の背後に潜む原理や未知の要因を理解する上でレシピに含まれる食材群に着目するアプローチの可能性を押し広げていると考えられる。

ところで、Ahn ら [2] は、西洋料理が多くの香料化合物を共有した食材ペアを利用する傾向を持つことを明らかにしたが、東アジア料理が香料化合物を共有した食材ペアの利用を避ける傾向を持つことも示した。東アジアに含まれる日本の和食の特徴は、食材そのもの持ち味や新鮮さを生かすことや、季節や伝統的な行事などの時季を重んじる傾向がある（詳しくは [11] を参照）。しかし、世界の様々な料理が取り込まれている日本において、その傾向が料理レシピ共有サイトに投稿される一般家庭料理の動向に反映されているかは不明である。また、我々は、人気を得る料理レシピの要因に注目しているが、人気の要因としては、一般的にレシピ投稿者に関する人気に左右される場合もあれば、見た目の良さ、食材そのものの人気、ある時季によく使われる食材を用いることでの人気など、これら以外にも様々な要因が存在し得ると考えられる。特に我々は、料理の味や風味、色合いなどに関連すると思われる調理における順序性にも注目している。本研究では、日本の料理レシピ共有サイトに投稿された大量

\*連絡先：龍谷大学理工学部電子情報学科  
〒520-2194 大津市瀬田大江町横谷 1-5  
E-mail: kimura@rins.ryukoku.ac.jp

表 1:  $2 \times 2$  分割表

	$D_1$	$\bar{D}_1$
$a$	$n_{a,1}$	$n_{a,\bar{1}}$
$\bar{a}$	$n_{\bar{a},1}$	$n_{\bar{a},\bar{1}}$

のレシピデータに着目し、人気を得る料理レシピの要因を調べる第一歩として、食材そのものに着目した食材ユニグラム、ある時季での利用状況が活発化する食材に着目したアクティブ食材ユニグラム、また、順序性として、ある食材とその次に使われる食材に着目した食材バイグラム、ある時季での出現状況が活発化する食材バイグラムに着目したアクティブ食材バイグラム、という4つの観点から料理レシピの人気分析を行う。そして、日本の料理レシピ共有サイト CookPad の実データを用いた実験で、人気分析の結果を報告する。

## 2 分析法

対象とする日本の料理レシピ共有サイトにおいて、ここ数年間に投稿されたレシピ全体の集合  $D$  を考える。レシピ総数を  $|D| = N$  とする。  $D$  内の各レシピを調査し、どのようなレシピが人気を得たかについて分析する。  $D$  に属するレシピのうち、人気を得たレシピ全体の集合を  $D_1$  とし、それ以外を  $\bar{D}_1$  とする。すなわち、  $D = D_1 \cup \bar{D}_1$ ,  $D_1 \cap \bar{D}_1 = \emptyset$  である。本研究では、レシピに関する属性  $a$  を導入し、レシピ  $d \in D$  に対して、  $d$  が属性  $a$  をもつことと  $d$  が人気レシピであること (i.e.,  $d \in D_1$ ) に関係性があるかどうかを、表1の様な  $2 \times 2$  分割表を作成しカイ二乗検定に基づいて調べる。ここに、  $n_{a,1}$  は属性  $a$  をもつ  $D_1$  の元の数、  $n_{a,\bar{1}}$  は属性  $a$  をもつ  $\bar{D}_1$  の元の数、  $n_{\bar{a},1}$  は属性  $a$  をもたない  $D_1$  の元の数、  $n_{\bar{a},\bar{1}}$  は属性  $a$  をもたない  $\bar{D}_1$  の元の数それぞれ表す。このとき、  $N = n_{a,1} + n_{a,\bar{1}} + n_{\bar{a},1} + n_{\bar{a},\bar{1}}$  である。本研究では、特に、利用食材や時季に関わるようなレシピ属性  $a$  について考える。食材全体の集合を  $V$  とする。

### 2.1 食材ユニグラムの分析

まず、その食材が利用されたことがレシピの人気獲得に大きく貢献したというような食材 (すなわち、食材ユニグラム) の抽出を目指し、レシピに関する属性  $a$  として、特定の食材がレシピで利用されている (すなわち、食材ユニグラムがレシピに出現する) という考えを調べる。各食材  $v \in V$  に対して、カイ二乗値  $\chi^2(v)$  と対応する  $p$  値  $p(v)$  を求め、レシピの人気獲得と関係

性があったと考えられるような食材の集合  $V_1$  を抽出し分析する。

## 2.2 アクティブ食材ユニグラムの分析

### 2.2.1 概要

日本には四季があり旬の食材を利用したレシピは人気を獲得する傾向があると推測されるので、次に、レシピの人気獲得と関係するような旬食材の抽出について考える。我々は、季節変化の周期性を考慮して、各食材  $v \in V$  のレシピ出現頻度の時系列データを数年にわたって調べ、  $v$  がその平年における平均頻度と比べて顕著に出現しているような時季を検出して、それを  $v$  のアクティブ時季  $\mathcal{I}_v$  と呼び、  $v$  の旬の時季とみなすことにする。そして、食材  $v \in V$  とそのアクティブ時季  $\mathcal{I}_v$  の組  $(v, \mathcal{I}_v)$  全体の集合  $\mathcal{B}$  を考える。任意の  $(v, \mathcal{I}_v) \in \mathcal{B}$  に対して、食材  $v$  がアクティブ時季  $\mathcal{I}_v$  内に利用されていることをレシピに関する属性  $a$  とし、カイ二乗値  $\chi^2(v, \mathcal{I}_v)$  と対応する  $p$  値  $p(v, \mathcal{I}_v)$  を求め、レシピの人気獲得と関係性があったと考えられるような食材とそのアクティブ時季の組の集合  $\mathcal{B}_1$  を抽出し分析する。

### 2.2.2 アクティブ時季の検出法

各食材  $v \in V$  に対して、そのアクティブ時季  $\mathcal{I}_v$  の検出を考える。本研究では、食材ネットワーク分析において我々 [9, 10] が与えた食材ペアのアクティブ期間抽出法を応用して、以下の様に  $\mathcal{I}_v$  を求める。

まず、時間単位を指定すると、1年は区間  $[0, T)$  上の整数点の集合で表すことができることに注意する。ここで正整数  $T$  は、その指定した時間単位における1年の長さである。本研究では、平年を  $\mathbb{R}/T\mathbb{Z}$  上の格子点の集合、

$$S = \{[t] \in \mathbb{R}/T\mathbb{Z} \mid t = 0, 1, \dots, T-1\}$$

と同一視する。ここに、  $\mathbb{R}$  は実数全体の集合で  $\mathbb{Z}$  は整数全体の集合である。平年  $S$  における時季という概念を次のように定義する。任意の整数  $t, t'$ , ( $0 \leq t, t' \leq T-1$ ,  $t \neq t'$ ) に対して、時季  $[[t], [t']]$  は、  $t < t'$  のとき  $[t, t')$  ( $\subset [0, T)$ ) であり、  $t > t'$  のとき  $[t, T) \cup [0, t')$  ( $\subset [0, T)$ ) である。我々は、食材  $v$  のアクティブティに関する観測時系列データ、

$$\mathcal{X}(v) = \{x_{[t]}(v) \mid [t] \in S\}$$

に基づいて  $\mathcal{I}_v$  を検出する。ここに、  $x_{[t]}(v)$  は時間ステップ  $[t]$  に投稿されたレシピのうち  $v$  を利用していたものの総数である。また、時間ステップ  $[t]$  に投稿されたレシピの総数を  $N_{[t]}$  とし、時間ステップ  $[t]$  に投稿さ



れるレシピに食材  $v$  が利用されている確率を  $P(v|t]$  とする。

各  $t = 0, 1, \dots, T-1$  に対して,  $x_{[t]}(v)$  はパラメータが  $N_{[t]}$  と  $P(v|t]$  である二項分布から生成されると仮定する. さらに,  $v$  のアクティビティは  $J_v$  個の時季に分割できると仮定する. ここに,  $J_v$  は正の整数である.  $J_v = 1$  の場合,  $v$  はアクティビティ変化点をもたないと言い,  $\mathcal{T}^1(v) = \emptyset$  とし,

$$P(v|t] = p^1(v), \quad \forall [t] \in \mathcal{S}$$

とする.  $J_v \geq 2$  の場合,  $v$  はアクティビティ変化点の集合,

$$\mathcal{T}^{J_v}(v) = \{t_1(v), \dots, t_{J_v}(v)\}$$

( $0 \leq t_1(v) < \dots < t_{J_v}(v) \leq T-1$ ) をもち, それに対応する  $J_v$  個の確率の組,

$$\mathbf{p}^{J_v}(v) = (p_1(v), \dots, p_{J_v}(v))$$

が存在すると仮定する. ただし,  $j = 1, \dots, J_v$  に対して,

$$P(v|t] = p_j(v), \quad \forall [t] \in \mathcal{S}_j(v) = [[t_j(v)], [t_{j+1}(v))]]$$

であり, また,  $p_j(v) \neq p_{j+1}(v)$ ,  $t_{j+1}(v) = t_1(v)$ ,  $p_{J_v+1}(v) = p_1(v)$  である. このとき,

$$\mathcal{S} = \bigcup_{j=1}^{J_v} \mathcal{S}_j(v) \quad (\text{disjoint union})$$

が成り立つことに注意する.  $(\mathcal{T}^{J_v}(v), \mathbf{p}^{J_v}(v))$  を  $v$  のアクティビティ変化パターンと呼び, 各  $\mathcal{S}_j(v)$  を  $v$  の第  $j$  時季と呼ぶ.

観測時系列データ  $\mathcal{X}(v)$  からアクティビティ変化パターン  $(\mathcal{T}^{J_v}(v), \mathbf{p}^{J_v}(v))$  を検出することを考える.  $J_v = 1$  の場合,  $\mathcal{X}(v)$  の対数尤度は,

$$\mathcal{L}^1(\mathcal{X}(v); p^1(v)) = \sum_{[t] \in \mathcal{S}} \{x_{[t]}(v) \log p^1(v) + (N_{[t]} - x_{[t]}(v)) \log(1 - p^1(v))\}$$

であるので,  $p^1(v)$  の最尤推定値  $\hat{p}^1(v)$  は,

$$\hat{p}^1(v) = \frac{\sum_{[t] \in \mathcal{S}} x_{[t]}(v)}{\sum_{[t] \in \mathcal{S}} N_{[t]}}$$

で与えられる. また,  $J_v \geq 2$  の場合,  $\mathcal{X}(v)$  の対数尤度は,

$$\begin{aligned} & \mathcal{L}^{J_v}(\mathcal{X}(v); \mathcal{T}^{J_v}(v), \mathbf{p}^{J_v}(v)) \\ &= \sum_{j=1}^{J_v} \sum_{[t] \in \mathcal{S}_j(v)} \{x_{[t]}(v) \log p_j(v) \\ & \quad + (N_{[t]} - x_{[t]}(v)) \log(1 - p_j(v))\} \end{aligned}$$

であるので,  $\mathcal{T}^{J_v}(v)$  が指定された時,  $\mathbf{p}^{J_v}(v)$  の最尤推定値  $\hat{\mathbf{p}}^{J_v}(v) = (\hat{p}_1(v), \dots, \hat{p}_{J_v}(v))$  は,

$$\hat{p}_j(v) = \frac{\sum_{[t] \in \mathcal{S}_j(v)} x_{[t]}(v)}{\sum_{[t] \in \mathcal{S}_j(v)} N_{[t]}} \quad (j = 1, \dots, J_v)$$

で与えられる. 我々は, 目的関数として,

$$\begin{aligned} & f(\mathcal{T}^{J_v}(v)) \\ &= \mathcal{L}^{J_v}(\mathcal{X}(v); \mathcal{T}^{J_v}(v), \hat{\mathbf{p}}^{J_v}(v)) - \mathcal{L}^1(\mathcal{X}(v); \hat{p}^1(v)) \end{aligned}$$

を定義し,  $f$  を最大化することにより最適なアクティビティ変化点の集合  $\mathcal{T}^{J_v}(v)$  を検出する (詳細は [9, 10] を参照).

さて, 観測時系列データ  $\mathcal{X}(v)$  から, アクティビティ変化パターン  $(\mathcal{T}^{J_v}(v), \hat{\mathbf{p}}^{J_v}(v))$  が検出されたとする. まず,  $\hat{p}^1(v)$  は平年において  $v$  がレシピで利用される平均頻度を表し, 各  $\hat{p}_j(v)$  は第  $j$  時季  $\mathcal{S}_j(v)$  において  $v$  がレシピで利用される平均頻度を表していることに注意する. よって, 各  $j = 1, \dots, J_v$  に対して,  $\hat{p}_j(v) > \hat{p}^1(v)$  ならば  $\mathcal{S}_j(v)$  はアクティブであると呼び,  $\hat{p}_j(v) \leq \hat{p}^1(v)$  ならば  $\mathcal{S}_j(v)$  は非アクティブであると呼ぶ. 我々は, アクティブな  $\mathcal{S}_j(v)$  の全体を  $v$  のアクティブ時季  $\mathcal{I}_v$  として検出する.

## 2.3 食材バイグラムの分析

料理レシピは食材リストを含んでいるだけではなく, 一般に, 調理における食材の利用の仕方を説明しており, その食材の利用の仕方には順序性があると考えられる. その食材の順序性がレシピの人気獲得に大きく貢献したというような食材の利用順序 (すなわち, 食材バイグラム) の抽出を目指し, レシピに関する属性  $a$  として, 特定の食材  $u \in V$  の次に  $v \in V$  が利用されている (すなわち, 食材バイグラムがレシピに出現する) ということを考える. ここに, 特定の食材  $u \in V$  の次に  $v \in V$  が利用されることを  $u \rightarrow v$ , ( $u \neq v$ ) と表す.

図 1 は, 食材リストと調理手順が記載された一般的な料理レシピから食材バイグラムを抽出する処理過程を示している. まず, 事前準備の処理について述べる. 料理レシピの食材リストには, 野菜などの主食材やそれ以外の食材 (調味料・風味調味料) だけでなく「キッチンペーパー」などの道具として使う材料が含まれる場合がある. 本研究では, 道具として使われる材料を食材リストから事前に除外した. 次に, 調理手順が記載された文章は順序立てて (番号順に) 説明が行われている. その順序に沿って説明された各文章を順序性を保ちながら 1 つの文章として連結し, 形態素解析を行って順序性を維持しながら名詞群を抽出する. 事前準備

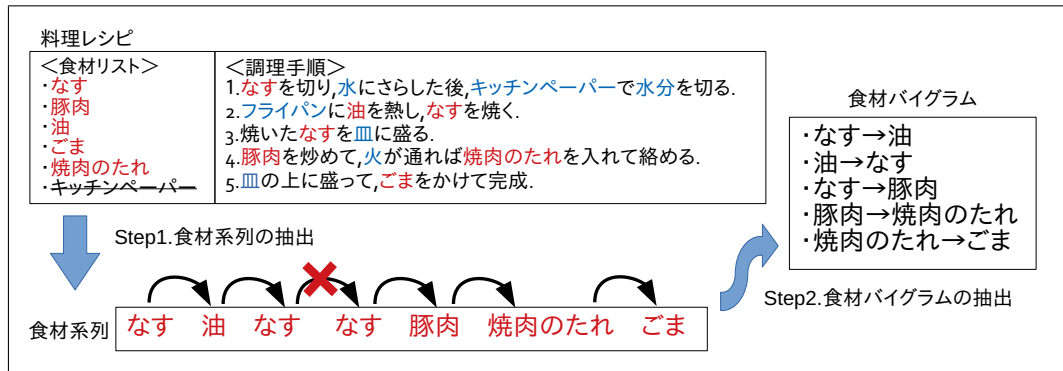


図 1: 食材バイグラムの抽出手順

備を終えると、Step 1として料理レシピそれぞれの説明文書から抽出された名詞群とその食材リストに現れる各食材とを照合し、食材リストに存在する食材(図1の赤字で表された食材)のみを順序性を保ちながら食材系列として生成する。なお、図1の調理手順を説明する文章では、主食材や調味料・風味調味料以外に、青い字で表した「フライパン」、「皿」、「火」などの名詞が存在しているが、これらの名詞は例示した料理レシピの食材リストには存在しないため、食材系列には加わらないことに注意しておく。ただし、文書中の「水」は、食材リストに含まれていれば食材系列に加わることに注意しておく。次に、Step 2では、食材系列に沿って先に現れた食材から隣り合う次の食材への有向リンクを結び、これを食材バイグラム  $u \rightarrow v$  とする。ただし、「なす」から「豚肉」や「なす」から「ごま」など、食材系列上で2つ以上離れた食材どうしの順序関係については本研究では対象外とした。また、「なす」から「なす」など、同じ食材が続く場合は食材バイグラムから除外している。以上の条件のもとに食材バイグラムの集合を抽出する。各食材バイグラムに対して、カイ二乗値  $\chi^2(u \rightarrow v)$  と対応する  $p$  値  $p(u \rightarrow v)$  を求め、レシピの人気獲得と関係性があったと考えられるような食材バイグラムの集合を抽出し分析する。

## 2.4 アクティブ食材バイグラムの分析

本研究において、調理における食材の利用の仕方に順序性があると考え、食材バイグラムを抽出しているが、アクティブ食材ユニグラムにおいて時季性を考えたように、食材バイグラムについても時季性が存在し得ることが予想される。そこで、レシピの人気獲得と関係するような旬の食材バイグラムとしてアクティブ食材バイグラムの抽出を行う。食材バイグラムのアクティブ時季の検出については、アクティブ食材ユニグラムにおけるアクティブ時季の検出法を適用し、食材バイグラム  $u \rightarrow v$  とアクティブ時季  $\mathcal{I}_{u \rightarrow v}$  の全体の集合

$\mathcal{C}$  を考える。任意の  $(u \rightarrow v, \mathcal{I}_{u \rightarrow v}) \in \mathcal{C}$  に対して、食材バイグラム  $u \rightarrow v$  がアクティブ時季  $\mathcal{I}_{u \rightarrow v}$  内に利用されていることをレシピに関する属性  $a$  とし、カイ二乗値  $\chi^2(u \rightarrow v, \mathcal{I}_{u \rightarrow v})$  と対応する  $p$  値  $p(u \rightarrow v, \mathcal{I}_{u \rightarrow v})$  を求め、レシピの人気獲得と関係性があったと考えられるような食材とそのアクティブ時季の組の集合を抽出し分析する。

## 3 実験

### 3.1 実験設定

日本の料理レシピ共有サイト「Cookpad」の2011年1月から2013年12月までの3年間に投稿された料理レシピに関する実データを用いた。Cookpadでは、レシピが投稿された時間および利用された食材、調理手順などの情報が得られる。Cookpadの料理レシピに含まれる食材の名前には「たまご」「卵」「玉子」のように漢字やひらがな、カタカナなどによる表記揺れが存在するため、食材用語に関する辞書を作成することで表記揺れを手動で解消し、調味料・風味調味料を含む489種類の食材を抽出した。また、料理レシピに用いられる食材や手順の傾向は「肉料理」や「野菜料理」、「お菓子」など、料理のジャンルによって異なり得ると考えられることから、実データの期間において料理ジャンルに関する情報を持つ料理レシピのみを抽出し、料理ジャンルごとに分析を行う。本研究では、Cookpadに存在する料理カテゴリーを料理ジャンルと見なし、レシピ数が多い11種類の料理カテゴリーに焦点を当てるが、本稿では、頁数の制限から、特に「野菜のおかず」カテゴリーについて分析を行った結果を報告する。「野菜のおかず」カテゴリーに含まれるレシピ数  $N=14,590$  であった。また、「野菜のおかず」カテゴリーに含まれる料理レシピ群について、料理レシピごとにその投稿日から1年間に付与されたつくレポ数の多さに基づいて人気レシピとそうでないレシピを仕分ける。本研究で

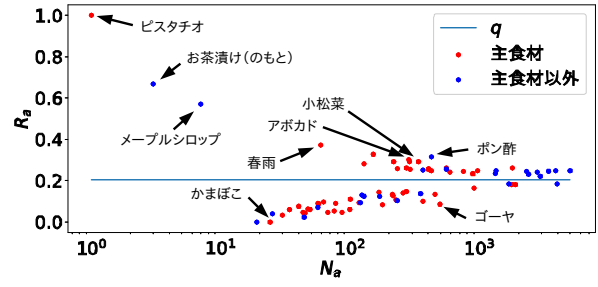
表 2: 食材ユニグラムの  $\chi^2$  値ランキング

(a) 主食材 Top 10.

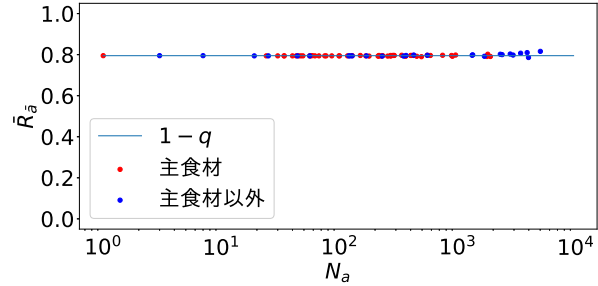
主食材	$\chi^2$	$R_a$	$\bar{R}_a$	$N_a$
ゴーヤ	45.448	0.085	0.790	494
ごま	38.322	0.261	0.802	1,799
昆布	25.770	0.099	0.792	363
菜の花	16.217	0.084	0.793	178
アボカド	16.211	0.301	0.796	282
豚ばら肉	15.647	0.292	0.797	336
しいたけ	14.961	0.133	0.792	451
パプリカ	13.861	0.106	0.793	226
小松菜	13.635	0.293	0.796	287
かにかま	13.633	0.327	0.796	150

(b) 主食材以外 Top 10.

主食材以外	$\chi^2$	$R_a$	$\bar{R}_a$	$N_a$
醤油	80.299	0.247	0.816	5,014
砂糖	59.133	0.248	0.810	3,875
油	45.059	0.246	0.807	3,413
ボン酢	32.332	0.315	0.798	425
酒	27.888	0.242	0.803	2,787
みりん	25.072	0.244	0.802	2,297
塩	17.381	0.183	0.786	3,979
片栗粉	16.027	0.248	0.799	1,341
ワイン	14.640	0.104	0.793	230
塩こしょう	12.186	0.232	0.800	2,380



(a)  $R_a$  と  $N_a$  の関係.



(b)  $\bar{R}_a$  と  $N_a$  の関係.

図 2: 食材ユニグラムの分析結果

は、各料理カテゴリーの料理レシピ総数に対するつくレポ数の平均値を上回るものを人気レシピ集合  $D_1$  と定めて分析を行った。料理レシピ総数  $|D| = N$  に対する人気レシピ数  $|D_1| = n_{a,1} + n_{\bar{a},1}$  の比を、

$$q = |D_1|/|D|$$

と表す。  $q = 0.205$  であった。

分析を行うにあたり、属性  $a$  をもつレシピの数を、

$$N_a = n_{a,1} + n_{a,\bar{1}}$$

と表し、そのようなレシピが人気レシピになる確率を、

$$R_a = n_{a,1}/N_a$$

と表す。属性  $a$  をもつレシピについては、  $R_a > q$  ならば  $R_a$  の値が大きいほど人気レシピになる傾向が強く、  $R_a < q$  ならば  $R_a$  の値が小さいほど人気レシピにならない傾向が強いと考えられる。また、属性  $a$  をもたないレシピが人気レシピにならない確率を、

$$\bar{R}_a = n_{\bar{a},1}/(N - N_a)$$

と表す。このとき、属性  $a$  のカイ二乗値は、

$$\chi^2(a) = q(1 - q)N_a(R_a - q)^2 + q(1 - q)(N - N_a)\{\bar{R}_a - (1 - q)\}^2 \quad (1)$$

と表されることに注意する。本分析では、カイ二乗検定においてレシピの人気獲得との関係が有意であると判定された属性のみを対象とする。

### 3.2 食材ユニグラムに基づいた分析

表 2 は、各食材ユニグラムに対応する  $\chi^2$  値でランキングを行った Top 10 を示しており、表 2a は主食材のみ、表 2b は主食材以外を抽出しており、それぞれの  $R_a$ 、 $\bar{R}_a$ 、 $N_a$  を併記している。また、 $N_a$  に対して  $R_a$  と  $\bar{R}_a$  に着目して分析した結果を図 2 に示す。図 2a および図 2b ともに赤い点は食材ユニグラムのうち主食材を、青い点は食材ユニグラムのうち調味料や風味調味料などの主食材以外を表している。ここで、図 2b の傾向から、 $\bar{R}_a \simeq 1 - q$  と見なせるため、式 (1) の第 2 項は無視できることから、料理レシピの人気獲得に貢献するものとして、 $|R_a - q|$  や  $N_a$  の大きさが  $\chi^2$  値に貢献することがわかる。ただし、図 2a の傾向から、 $N_a$  が大きくなるほど  $|R_a - q|$  が小さくなる傾向があるため、 $N_a$  が大きすぎても人気獲得には貢献しない傾向があることが見てとれる。図 2a において、 $R_a$  が最も高い食材ユニグラムは主食材の「ピスタチオ」であった。しかし、そのレシピ数  $N_a$  は非常に少ないことから、人気レシピの主要な要因とは言い難い。また、主食材で  $R_a$  が低い「ゴーヤ」や主食材以外の「ワイン」、さらに図 2a から、「かまぼこ」などについては利用するとむしろ人気に対して逆効果となる傾向がある。一方、主食材の「アボガド」や「春雨」(図 2a) などは  $N_a$  がある程度数に達しているだけでなく  $R_a$  が大きめであり、これらの食材ユニグラムは料理レシピが人気を得ることについてある程度貢献する可能性が期待できると考えられる。

表 3: アクティブ食材ユニグラムのカイニ乗ランキング

(a) 主食材.

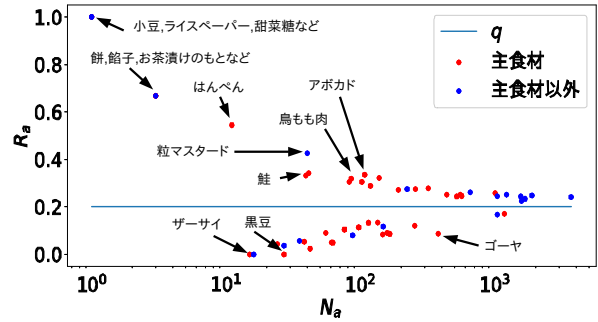
主食材	$\chi^2$	$R_a$	$\bar{R}_a$	$N_a$
ゴーヤ	33.198	0.087	0.791	378
ごま	17.087	0.257	0.798	998
菜の花	14.876	0.085	0.793	165
パプリカ	13.731	0.082	0.793	146
昆布	13.365	0.089	0.793	158
しいたけ	11.911	0.119	0.793	253
小松菜	11.337	0.321	0.796	137
かにかま	11.324	0.336	0.795	107
ほうれんそう	10.323	0.278	0.796	317
玉ねぎ	9.967	0.170	0.791	1,167

(b) 主食材以外.

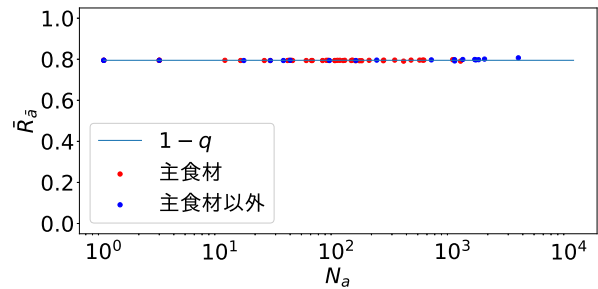
主食材以外	$\chi^2$	$R_a$	$\bar{R}_a$	$N_a$
醤油	40.149	0.242	0.807	3,664
砂糖	24.895	0.249	0.801	1,876
酒	16.761	0.251	0.799	1,220
水	14.562	0.243	0.799	1,546
片栗粉	12.979	0.261	0.797	655
粒マスタード	11.839	0.425	0.795	40
みりん	10.761	0.245	0.798	1,036
オリーブオイル	9.407	0.168	0.792	1,039
油	8.573	0.233	0.798	1,667
からし	8.380	0.080	0.794	87

### 3.3 アクティブ食材ユニグラムに基づいた分析

表 3 に、アクティブ食材ユニグラムにおける主食材 (表 3a) と主食材以外 (表 3b) のカイニ乗ランキングを示す。また、図 3 は、アクティブ食材ユニグラムに基づいた分析結果である。アクティブ食材ユニグラムについても  $|R_a - q|$  や  $N_a$  の大きさが  $\chi^2$  値に貢献することがわかる。主食材では「小松菜」、「かにかま」、「ほうれんそう」、「ごま」などが人気獲得にある程度貢献することがわかるが、図 3a から、「はんぺん」なども人気獲得への貢献度が高い可能性がある。しかし、「しいたけ」、図 3a から「ザーサイ」や「黒豆」、特に「ゴーヤ」や「玉ねぎ」などは  $N_a$  が大きい、すなわち多くの料理レシピで利用されているにもかかわらず、利用すると人気に対して逆効果であることがわかる。図 4 は表 3 の各食材  $v$  のアクティブ時季  $\mathcal{I}_v$  を黒い区間として表わしている。料理レシピの人気獲得への貢献が期待される「小松菜」、「かにかま」、「ほうれんそう」、「ごま」はそれぞれ異なる時季にアクティブになることがわかる。主食材以外では「粒マスタード」が人気獲得に貢献する度合いが高い傾向を持ち、そのアクティブ時季が 2 月から 4 月に限られることがわかる。また、「酒」、「みりん」も人気獲得への貢献が期待されるが、「ごま」

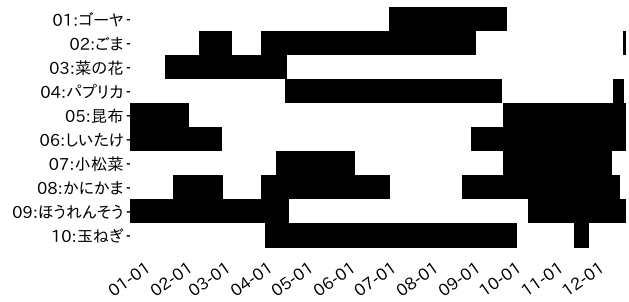


(a)  $R_a$  と  $N_a$  の関係.

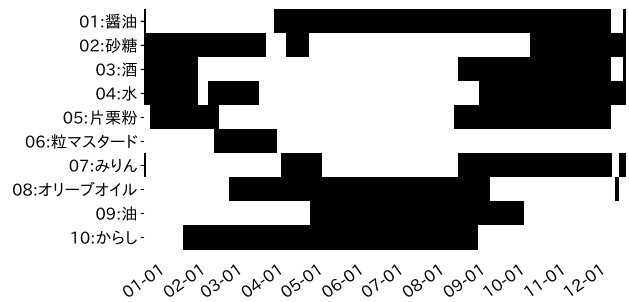


(b)  $\bar{R}_a$  と  $N_a$  の関係.

図 3: アクティブ食材ユニグラムの分析結果



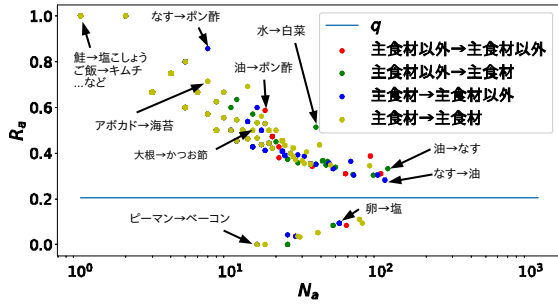
(a) 食材ユニグラムのアクティブ時季 (主食材).



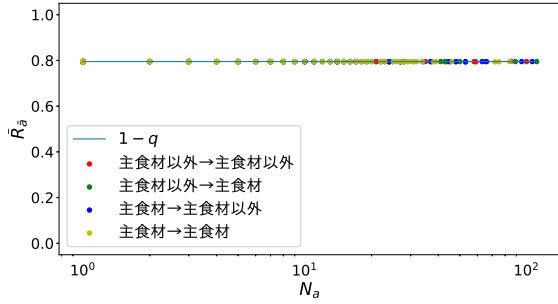
(b) 食材ユニグラムのアクティブ時季 (主食材以外).

図 4: アクティブ食材ユニグラムの時季

も含め、年末から年始にかけてアクティブになっており、「酒」、「みりん」、「ごま」は、日本の伝統文化である正月用の料理に関連した人気レシピの要因となっている可能性が示唆される。



(a)  $R_a$  と  $N_a$  の関係.



(b)  $\bar{R}_a$  と  $N_a$  の関係.

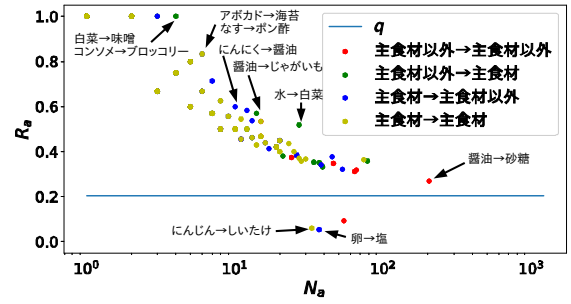
図 5: 食材バイグラムの分析結果

表 4: 食材バイグラムのカイ二乗ランキング

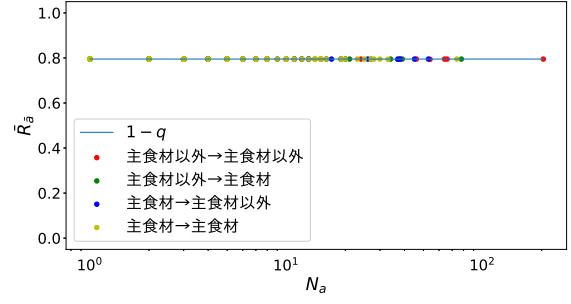
食材バイグラム	$\chi^2$	$R_a$	$\bar{R}_a$	$N_a$	種類
水 → 白菜	21.558	0.514	0.795	37	主食材以外 → 主食材
小松菜 → 卵	18.888	0.692	0.795	13	主食材 → 主食材
なす → ポン酢	18.217	0.857	0.795	7	主食材 → 主食材以外
水 → コンソメ	17.490	0.388	0.796	85	主食材以外 → 主食材以外
油 → ポン酢	15.273	0.588	0.795	17	主食材以外 → 主食材以外
えりんぎ → 油	14.315	0.600	0.795	15	主食材 → 主食材以外
バター → ベーコン	12.716	0.436	0.795	39	主食材 → 主食材
片栗粉 → じゃがいも	12.519	0.636	0.795	11	主食材以外 → 主食材
バター → コーン	12.505	0.562	0.795	16	主食材 → 主食材
海老 → 塩こしょう	11.732	0.667	0.795	9	主食材 → 主食材以外

### 3.4 食材バイグラムに基づいた分析

図 5 は、食材バイグラムに基づいた分析結果である。また、図 2a の食材ユニグラムと図 5a を比較すると、 $|R_a - q|$  が大きくなっており、人気の獲得への貢献やその逆効果となる食材バイグラムがより顕著に現れる傾向を持つことがわかる。表 4 に食材バイグラムのカイ二乗ランキングを示す。人気への貢献度が高いことが期待されるものとして、「主食材 → 主食材以外」パターンの「なす → ポン酢」や「海老 → 塩こしょう」などがあり、図 5a から、「主食材 → 主食材」パターンの「アボカド → 海苔」、「大根 → かつお節」、「主食材以外 → 主食材以外」パターンの「油 → ポン酢」などがあることもわかる。また、「油 → なす」はその逆の「なす → 油」より人気獲得への貢献度が高いことわかる。一方、「主食材 → 主食材」パターンの「ピーマン → ベーコン」は、利用すると人気に対して逆効果となる傾向をもつことがわかる。



(a)  $R_a$  と  $N_a$  の関係.



(b)  $\bar{R}_a$  と  $N_a$  の関係.

図 6: アクティブ食材バイグラムの分析結果

表 5: アクティブ食材バイグラムのカイ二乗ランキング

アクティブ食材バイグラム	$\chi^2$	$R_a$	$\bar{R}_a$	$N_a$	種類
水 → 白菜	16.236	0.519	0.795	27	主食材以外 → 主食材
サラダ油 → キャベツ	15.471	1.000	0.795	4	主食材以外 → 主食材
アボカド → 海苔	14.493	0.833	0.795	6	主食材 → 主食材
なす → ポン酢	14.493	0.833	0.795	6	主食材 → 主食材以外
コンソメ → ブロッコリー	11.602	1.000	0.795	3	主食材以外 → 主食材
白菜 → 味噌	11.602	1.000	0.795	3	主食材 → 主食材以外
なす → ピーマン	11.573	0.365	0.795	74	主食材 → 主食材
醤油 → じゃがいも	11.495	0.571	0.795	14	主食材以外 → 主食材
油 → なす	11.316	0.359	0.795	78	主食材以外 → 主食材
海老 → 塩こしょう	11.105	0.714	0.795	7	主食材 → 主食材以外

### 3.5 アクティブ食材バイグラムに基づいた分析

図 6 は、アクティブ食材バイグラムに基づいた分析結果である。また、表 5 に、アクティブ食材バイグラムのカイ二乗ランキングを示す。人気への貢献度が高いことが期待されるものとして、「主食材以外 → 主食材」パターンの「サラダ油 → キャベツ」、「コンソメ → ブロッコリー」、「主食材 → 主食材以外」パターンの「白菜 → 味噌」が非常に高い人気獲得率を持つことがわかる。それらは  $N_a$  については相対的に少ないが、利用されると人気レシピとなっていることから、人気レシピの要因として高い貢献度があると考えられる。しかし、「主食材 → 主食材」パターンの「にんじん → しいたけ」や「主食材 → 主食材以外」パターンの「卵 → 塩」などは人気に対して逆効果となる傾向があることがわかる。図 7 は、表 5 の各食材バイグラムのアクティブ時季を示したものである。「サラダ油 → キャベツ」は 4



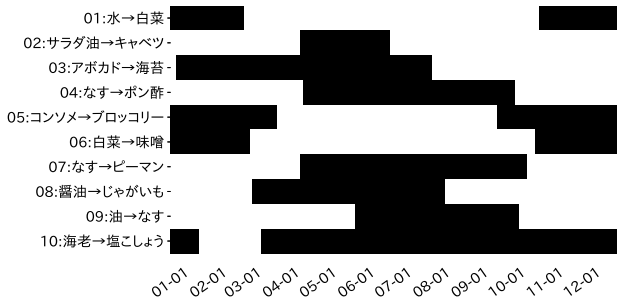


図 7: アクティブ食材バイグラムのアクティブ時季

月前から7月前の期間だけにアクティブ時季があることがわかる。「コンソメ → ブロッコリー」, 「白菜 → 味噌」はいずれも秋から春までの期間にアクティブ時季があることがわかる。すなわち, アクティブ食材バイグラムの人気への貢献度が高い時季はか限られる傾向があると考えられるため, 時季とともに順序性に注目することは, 人気分析を行ううえで意義深いことが示唆された。

これらの結果から, 食材バイグラムは食材ユニグラムよりも人気の要因を捉える点でより深い知見が得られる可能性があり, 時季性を捉えることも人気の要因を探求する上で意義深いことを示唆していると考えられる。すなわち, 食材単体の分析を行うだけでなく, 利用食材の調理上の順序性や時季を捉えるような本研究が着目した料理レシピの人気分析における新たな観点は, 人々が料理に対して行う行動の背後に潜む原理や未知の要因を理解する上でレシピに含まれる食材群に着目するアプローチの可能性をより押し広げている点で, 本研究の有効性を示唆していると思われる。

## 4 まとめ

本研究では, 日本の料理レシピ共有サイトに投稿された大量のレシピデータに着目し, 食材ユニグラム, アクティブ食材ユニグラム, 食材バイグラム, アクティブ食材バイグラムという4つの観点から人気を得る料理レシピとの関連性について分析を行った。そして, 料理レシピの人気分析において, 食材だけでなく, 食材利用の順序性を捉えることの有効性を示唆する結果を示した。今後は, 食材利用の順序性について, 新たな観点を導入し, 料理レシピの人気の要因についてさらなる探求を行う予定である。

## 謝辞

本研究では, クックパッド株式会社と国立情報学研究所が提供するクックパッドデータを利用した。

## 参考文献

- [1] Kinouchi, O., Diez-Garcia, R. W., Holanda, A. J., Zambianchi, P., and Roque, A. C.: “The nonequilibrium nature of culinary evolution,” *New Journal of Physics*, vol. 10, No. 7, pp. 073020:1–073020:8, 2008.
- [2] Ahn, Y. Y., Ahnert, S. E., Bagrow J., and Barabasi, A. L.: “Flavor Network and the Principles of Food Pairing,” *Scientific Reports*, vol. 1, pp. 196:1-196:7, 2011.
- [3] Teng, C.-Y., Lin, Y.-R., and Adamic, L.: “Recipe recommendation using ingredient networks,” in *Proceedings of WebSci’12*, pp. 298–307, 2012.
- [4] Nedovic, V.: “Learning recipe ingredient space using generative probabilistic models,” in *Proceedings of CwC’13*, pp. 13–18, 2013.
- [5] Jain, A., N K, R., and Bagler, G.: “Analysis of food pairing in regional cuisines of India,” *PLoS ONE*, vol. 10, no. 10, pp. 1–17, 2015.
- [6] Jermsurawong, J. and Habash, N.: “Predicting the structure of cooking recipes,” in *Proceedings of EMNLP’15*, pp. 781–786, 2015.
- [7] Jiang, Y., Skufca, J. D., and Sun, J.: BiFold visualization of bipartite datasets, *EPJ Data Science*, vol. 6, pp. 2:1–2:19, 2017.
- [8] Sajadmanesh, S., Jafarzadeh, S., Osia, S. A., Rabbiee, H. R., Haddadi, H., Mejova, Y., Musolesi, M., Cristofaro, E. D., and Stringhini, G.: “Kissing cuisines: Exploring worldwide culinary habits on the Web,” in *Proceedings of WWW’17 Companion*, pp. 1013–1021, 2017.
- [9] Kikuchi, Y., Kumano, M., and Kimura, M.: “Analyzing dynamical activities of co-occurrence patterns for cooking ingredients,” in *Proceedings of IEEE ICDMW’17*, pp. 17–24, 2017.
- [10] 菊地悠樹, 熊野雅仁, 木村昌弘, “料理レシピ共有サイトにおける食材のアクティブ共起パターンの抽出,” *情報処理学会論文誌「数理モデル化と応用」*, vol. 11, no. 2, pp. 30–40, 2018.
- [11] 熊倉功夫, “和食の神髄: ユネスコ無形文化遺産に登録された本当の理由,” *ヘルシスト*, 226, pp. 2–7, 2014.



# POI訪問ダイナミクスの可視化分析

## Visualization Analysis of POI Visit Dynamics

藤井輝 熊野雅仁 小堀聡 木村昌弘\*

Teru Fujii, Masahito Kumano, Satoshi Kobori, Masahiro Kimura

龍谷大学理工学部電子情報学科

Department of Electronics and Informatics, Ryukoku University

**Abstract:** 位置情報ベースのソーシャルネットワークサービス (LBSN) の普及にともない、都市観光での人々の足跡に関する大量データが蓄積されつつあり、このような大規模な時空間データを有効活用することに期待が高まっている。本研究では、観光プランニング支援や観光マーケティングへの活用基盤として、観光都市における POI(point-of-interest) 群への人々の訪問ダイナミクスの特徴を視覚的に分析できるシステムの構築を目指し、ソーシャルメディアのアイテムに対するアテンション到着過程の確率モデルである RPP モデルと DPM-RPP モデルを援用した、POI 訪問ダイナミクスの可視化分析法を提案する。Foursquare データを用いた実験において、東京およびニューヨークでの POI 群に対する訪問ダイナミクス分析のケーススタディを通し、提案法の有効性を検証する。

### 1 はじめに

近年、Foursquare Swarm や Facebook Places などの位置情報ベースのソーシャルネットワークサービス (LBSN) が普及し、都市観光での人々の足跡に関する大量データが蓄積されつつある。LBSN では、その都市の観光スポットや有名店舗などが POI(point-of-interest) として登録されており、それら POI 群の位置情報および、ユーザが POI にチェックインした時間情報が観測可能である。そして、このような大規模な時空間データに基づいて、観光都市における人々の行動の傾向や特徴を理解し、観光推薦システムや観光マーケティングに有効活用することが期待されている [1], [2], [3].

一方、Twitter や料理レシピ共有サイトなどのソーシャルメディアでは、投稿されたアイテムが他のユーザ達から共有されポピュラリティを獲得していくという現象がよく起こる。そして、このようなオンラインアイテムへのアテンション到着過程のモデリングが、ポピュラリティ予測、トレンド予測、マーケティングなどの観点から、近年、ソーシャルメディアマイニングの分野で注目されている [4]. 個々のアイテムのポピュラリティ予測については、既存の時系列解析法を用いる研究 [4], [5] や、特徴量を導入して既存の回帰問題や分類問題として扱う研究 [6], [7], [8] が、これまで主に行われてきた。最近 Shen ら [9] は、RPP モデル (Reinforced Poisson Process model) と呼ばれる個々のアイテムへのアテンション到着過程モデルとその学習

法を与え、既存のポピュラリティ予測法よりも精度が高いことを示した。ここに、RPP モデルは、アイテムの魅力度を表すフィットネス、アイテムの新たなアテンション獲得能力に対するエイジング特性を表す対数正規的な時間緩和関数 [10] および、“rich-get-richer” 現象を表す線形的強化関数 [11] という 3 つの構成要素からなる非一様 Poisson 過程である。我々 [12] は、対象とするソーシャルメディアのアイテム群全体に対するアテンション到着過程の確率モデルとして、Dirichlet 過程 [13] と RPP モデルを融合した DPM-RPP モデル (Dirichlet Process Mixture - RPP model) とその学習法を提案した。そして、料理レシピ共有サイトデータにおけるアイテムのポピュラリティ予測に対して、DPM-RPP モデルの精度が RPP モデルを含む従来法のものよりも平均的に見てかなり高いことを示した。さらに、DPM-RPP モデルを用いて、エイジングタイプの観点から料理レシピ群の構造を分析した。

本研究では、観光プランニング支援や観光マーケティングへの活用基盤として、観光都市における POI 群への人々の訪問ダイナミクスの特徴を視覚的に分析可能なシステムの構築を目指し、DPM-RPP モデルと RPP モデルを用いた POI 訪問ダイナミクスの可視化分析法を提案する。提案可視化法では、各 POI に対して、そのポピュラリティ予測の観点から DPM-RPP モデルと RPP モデルを比較し、予測精度が高い方のモデルを用いて訪問ダイナミクスに関するその魅力度とエイジング特性を可視化する。まず、DPM-RPP モデルによる POI のエイジング特性の可視化については、推定されたそのエイジングタイプを表示する。一方、RPP モデ

\*連絡先: 龍谷大学理工学部電子情報学科  
〒 520-2194 大津市瀬田大江町横谷 1-5  
E-mail: kimura@rins.ryukoku.ac.jp

ルによる POI のエイジング特性の可視化については、推定されたその対数正規的な時間緩和関数の特徴を効果的に表現する可視化法を新たに与え、それを表示する。そして、Foursquare データ [1] を用いた実験において、東京およびニューヨークでの POI 群に対する訪問ダイナミクス分析のケーススタディを通し、提案法の有効性を検証する。

## 2 分析法

対象とする観光都市における POI 全体の集合を  $D$  とし、期間  $[0, T)$  におけるそれら POI 群への人々の訪問ダイナミクスに対して、RPP モデル [9] と DPM-RPP モデル [12] を用いた可視化分析の枠組みを与える。ここに、 $[0, T)$  は分析対象とする期間を表し、 $T > 0$  はその期間の長さを表している。任意の POI  $d \in D$  に対して、 $[0, T)$  における  $d$  への訪問者の到着時刻の系列を、

$$\mathbf{t}^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$$

とする。ここに、 $n_d(T)$  は  $[0, T)$  における  $d$  への訪問者数であり、 $t_i^d$  は  $d$  への第  $i$  番目の訪問者の到着時刻を表し、 $0 \leq t_1^d \leq \dots \leq t_{n_d(T)}^d < T$  を満たしている。我々は、観測データ、

$$\mathcal{T}(T) = \{\mathbf{t}^d(T); d \in D\}$$

に基づいて、 $[0, T)$  における  $D$  への訪問ダイナミクスを分析する。

### 2.1 RPP モデル

まず、各  $d \in D$  に対して、 $\mathbf{t}^d(T)$  を RPP モデルによりモデル化することを考える。RPP モデルはレート関数  $\lambda_d(t)$  の非一様 Poisson 過程であり、 $\lambda_d(t)$  は、 $d$  の魅力度を表すフィットネス  $\eta_d$ 、 $d$  のエイジング特性を表す対数正規的な時間緩和関数  $g(t; \mu_d, \sigma_d)$  および、“rich-get-richer” メカニズムを表す線形的強化関数  $r_d(t)$  の積、

$$\lambda_d(t) = \eta_d g(t; \mu_d, \sigma_d^2) r_d(t)$$

で表される。ここに、 $\eta_d > 0$ 、 $\mu_d \in \mathbb{R}$ 、 $\sigma_d > 0$  はパラメータであり、

$$g(t; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{(\log t - \mu)^2}{2\sigma^2}\right\} \quad (1)$$

であり、 $r_d(t)$  は、 $t_{i-1}^d \leq t < t_i^d$  に対して、

$$r_d(t) = i - 1 + \nu \quad (2)$$

である。ただし、 $\nu > 0$  はハイパーパラメータである。

RPP モデルによる分析では、 $d \in D$  ごとに  $\mathbf{t}^d(T)$  が完全に独立にモデル化され、モデルパラメータ値が  $d \in D$  ごとに異なり得ることに注意しておく。

### 2.2 DPM-RPP モデル

次に、 $\mathcal{T}(T)$  を DPM-RPP モデルによりモデル化することを考える。DPM-RPP モデルでは、POI 群  $D$  は完全に独立ではなくなんらかの関係性を持つと考えている。そして、各 POI の魅力度を表すフィットネスについては、RPP モデルと同様に POI ごとに異なり得ると仮定するが、POI のエイジング特性を表す時間緩和関数については、POI ごとに完全に異なるのではなく、いくつかのエイジングタイプに分かれると仮定する。すなわち、RPP モデルに  $D$  のエイジングタイプ構造を組み込むことにより、DPM-RPP モデルは構築される。また、エイジングタイプの総数  $K$  も観測データから学習するために、DPM-RPP モデルには Dirichlet 過程 [13] が組み込まれている。DPM-RPP モデルでは、各  $d \in D$  に対して、 $\mathbf{t}^d(T)$  は以下の様に生成される。

まず、 $k = 1, 2, 3, \dots$  に対して、エイジングタイプ  $k$  の時間緩和関数  $g(t; \mu_k, \sigma_k^2)$  (式 (1) 参照) のパラメータ値  $(\mu_k, \sigma_k^2)$  を、次の Normal-Inverse-Gamma 分布から生成する。

$$p(\mu_k, \sigma_k^2 \mid \tau_1, \tau_2, \tau_3, \tau_4) = \frac{\sqrt{\tau_2 \tau_4^{\tau_3}}}{\sqrt{2\pi\sigma_k^2 \tau_3 + 3}\Gamma(\tau_3)} \exp\left\{-\frac{\tau_2(\mu_k - \tau_1)^2 + 2\tau_4}{2\sigma_k^2}\right\}$$

ここに、 $\tau_1 \in \mathbb{R}$ 、 $\tau_2, \tau_3, \tau_4 > 0$  はハイパーパラメータである。無限次元離散確率分布  $\boldsymbol{\pi} = (\pi_k)_{k=1}^\infty$  を Stick-Breaking 過程から  $k = 1, 2, 3, \dots$  に対して、

$$v_k \mid \gamma \sim \text{Beta}(1, \gamma), \quad \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$

と生成する。ここに、 $\text{Beta}(1, \gamma)$  はパラメータが 1 と  $\gamma > 0$  のベータ分布である。ランダム測度  $G$  を

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{(\mu_k, \sigma_k^2)}$$

と定義する。ここに、 $\delta_{(\mu_k, \sigma_k^2)}$  は位置  $(\mu_k, \sigma_k^2)$  に集中した確率測度を表す。このとき  $G$  は、密度関数  $p(\mu, \sigma^2 \mid \tau_1, \tau_2, \tau_3, \tau_4)$  が定める確率測度を基底測度とし、 $\gamma$  を集中度パラメータとする Dirichlet 過程  $\text{DP}(\gamma, p(\mu, \sigma^2 \mid \tau_1, \tau_2, \tau_3, \tau_4))$  にしたがって分布していることに注意しておく。

次に、 $d$  のエイジングタイプ  $z_d$  をランダム測度  $G$  から、

$$z_d \mid G \sim G$$

と生成する.  $d$  のフィットネス  $\eta_d$  をパラメータが  $\alpha > 0$  と  $\beta > 0$  のガンマ分布  $\text{Gamma}(\alpha, \beta)$  から,

$$\eta_d | \alpha, \beta \sim \text{Gamma}(\alpha, \beta)$$

と生成する. ただし,  $\alpha$  と  $\beta$  はハイパーパラメータである.  $d$  に対する強化関数  $r_d(t)$  を式 (2) で定義する. このとき,  $\lambda_d(t)$  を

$$\lambda_d(t) = \eta_d g(t; \mu_{z_d}, \sigma_{z_d}^2) r_d(t)$$

で定義し, レート関数  $\lambda_d(t)$  の非一様 Poisson 過程から  $t^d(T)$  を生成する.

## 2.3 学習法と予測法

まず, 各  $d \in D$  に対して, 観測データ  $t^d(T)$  から RPP モデルを学習する (詳細は [9] を参照). 時間緩和関数のパラメータの推定値  $\theta_d^* = (\mu_k^*, \sigma_k^{*2})$  および, フィットネスに対する事前分布のハイパーパラメータの推定値  $\alpha^*, \beta^*$  が得られる. RPP モデルを用いるとき, 任意の  $t > T$  に対して,  $d$  への期間  $[0, t)$  での訪問者総数の予測値  $\langle x_0^d(t) \rangle$  は,

$$\begin{aligned} \langle x_0^d(t) \rangle &= \int_0^\infty x_0^d(t) p(\eta_d | t^d(T), \theta_d^*, \alpha^*, \beta^*) d\eta_d \\ &= (n_d(T) + \nu) \\ &\quad \times \left( \frac{X_d(\theta_d^*) + \beta^*}{X_d(\theta_d^*) - Y_d(\theta_d^*) + \beta^*} \right)^{n_d(T) + \alpha^*} - \nu \end{aligned}$$

で与えられる. ここに,

$$x_0^d(t) = (n_d(T) + \nu) \exp \{ \eta_d Y_d(\theta_d^*) \} - \nu$$

であり, また,

$$\begin{aligned} X_d(\theta) &= (n_d(T) + \nu) \int_0^T g(t; \theta) dt \\ &\quad - \sum_{i=1}^{n_d(T)} \int_0^{t_i^d} g(t; \theta) dt \end{aligned}$$

であり,

$$Y_d(\theta) = \int_T^t g(s; \theta) ds$$

である.

次に, 観測データ  $\mathcal{T}(T)$  から DPM-RPP モデルを学習する (詳細は [12] を参照). エイジングタイプの総数  $K^*$ , 各エイジングタイプ  $k$  の時間緩和関数のパラメータの推定値  $\theta_k^* = (\mu_k^*, \sigma_k^{*2})$ , フィットネスに対する事前

分布のハイパーパラメータの推定値  $\alpha^*, \beta^*$  および, 各  $d \in D$  が各エイジングタイプ  $k$  に属する事後確率,

$$\phi_{d,k} = P(z_d = k | t^d(T), \Theta^*, \alpha^*, \beta^*, \gamma)$$

が得られる. ただし,  $\Theta^* = \{\theta_1^*, \dots, \theta_{K^*}^*\}$  である. DPM-RPP モデルを用いるとき, 任意の  $t > T$  に対して, 各  $d \in D$  への期間  $[0, t)$  での訪問者総数の予測値  $\langle x_1^d(t) \rangle$  は,

$$\begin{aligned} \langle x_1^d(t) \rangle &= \sum_{k=1}^{K^*} \phi_{d,k} \int_0^\infty x_1^d(t) p(\eta_d | t^d(T), z_d = k, \theta_k^*, \alpha^*, \beta^*) d\eta_d \\ &= (n_d(T) + \nu) \\ &\quad \times \sum_{k=1}^{K^*} \phi_{d,k} \left( \frac{X_d(\theta_k^*) + \beta^*}{X_d(\theta_k^*) - Y_d(\theta_k^*) + \beta^*} \right)^{n_d(T) + \alpha^*} - \nu \end{aligned}$$

で与えられる. ただし,

$$x_1^d(t) = (n_d(T) + \nu) \exp \{ \eta_d Y_d(\theta_{z_d}^*) \} - \nu$$

である. また, DPM-RPP モデルでは, 各  $d \in D$  に対して,

$$z_d^* = \operatorname{argmax}_{1 \leq k \leq K^*} \phi_{d,k}$$

により,  $d$  のエイジングタイプ  $z_d^*$  を決定し,  $D$  のエイジングタイプ構造を抽出する.

## 2.4 可視化法

まず,  $T' > T$  を固定し, 各  $d \in D$  に対して, 期間  $[0, T')$  での  $d$  への訪問者総数の予測を考える. そして, RPP モデルによる予測値  $\langle x_0^d(T') \rangle$  と DPM-RPP モデルによる予測値  $\langle x_1^d(T') \rangle$  を比較し, 予測精度が高い方のモデルに基づいて  $d$  の魅力度とエイジング特性を可視化することにより, 期間  $[0, T)$  における POI 群  $D$  への訪問ダイナミクスの特徴を分析する. ここに, DPM-RPP モデルに従っている POI の訪問ダイナミクスは, エイジングタイプ構造の観点から  $D$  内の他の POI 群の訪問ダイナミクスと関係していると考えられ, 一方, RPP モデルに従っている POI は,  $D$  内の他の POI 群との関係があまりない独自の訪問ダイナミクスをもっていると考えられる.

以下において我々は, POI 訪問ダイナミクスの主な特徴である魅力度とエイジング特性について, POI ごとにそれらを効果的に可視化する手法を与える. まず, DPM-RPP モデルに従っている POI に対して可視化する手法を与え, 次に, RPP モデルに従っている POI に対して可視化する手法を与える.

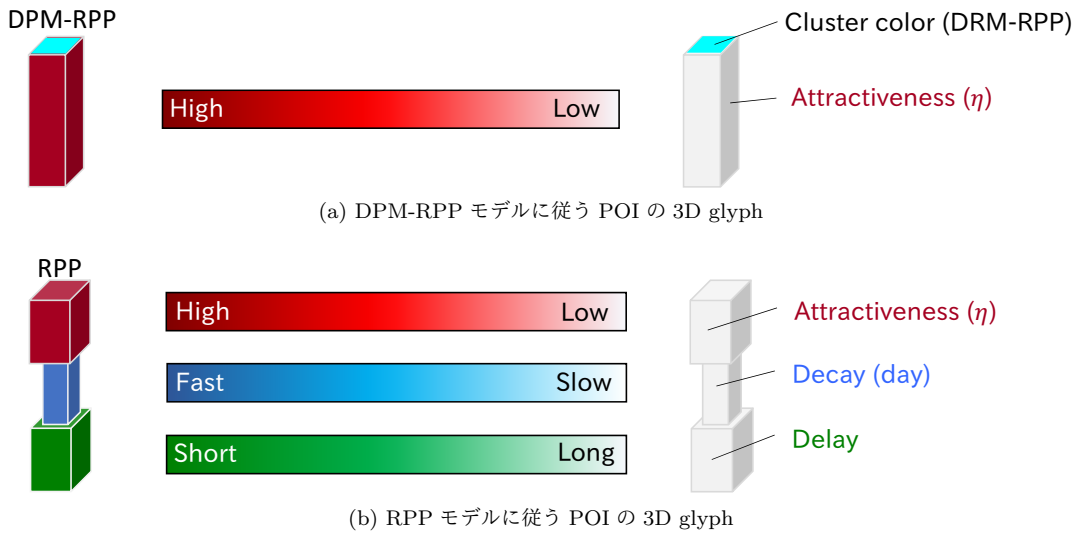


図 1: POI のフィットネス（魅力度）とエイジング特性の可視化法

本研究では、各 POI が位置情報を持つことを有効活用し、地図システム上で相互の POI に付随する情報を比較可能とするため、POI に付随する情報を四角柱の組み合わせで表した立体的な 3D glyph として可視化する手法を考える。DPM-RPP と RPP モデルに従う POI を 3D glyph として表現する方法を図 1 に示す。

#### 2.4.1 DPM-RPP モデルによる可視化

DPM-RPP モデルに従う POI 群は、同じ時間緩和関数を共有するいくつかのクラスタに分かれるが、地図上では相互に離れた位置に存在し得るため、同じクラスタに所属する POI であることを視認できるように、図 1a のような天井面の色が同じ四角柱を同一クラスタの POI として表す。また、DPM-RPP モデルに従う各 POI の魅力度を表すフィットネス  $\eta$  の高さを図 1a の濃い赤から白までのグラデーションで表す。フィットネスが高いほど四角柱の側面は濃い赤となり、フィットネスが低いほど白となる。これにより、同一クラスタに所属する DPM-RPP モデルに従う POI であっても、

それぞれ異なるフィットネスを持つ POI として容易に識別することができる。なお、DPM-RPP モデルに従う POI では、クラスタごとに時間緩和関数が異なるが、クラスタ数は多くはないため、各クラスタの時間緩和関数は補助グラフにより参照可能にする。

#### 2.4.2 RPP モデルによる可視化

RPP モデルに従う各 POI は、それぞれ異なる時間緩和関数を持つが、地図上に配置される POI の数が多いことから、各 POI の時間緩和関数の形状を地図上で容易に視認可能にする方法として図 1b のように、四角柱を三段縦に積み重ねる方法を用いる。図 1c は時間緩和関数の例を示している。緑色の時間緩和関数は遅延日数が短く急激に減衰し、青色の時間緩和関数は緑色のものと同じく遅延日数が短い減衰は緩やかとなっている。また、赤色の時間緩和関数は遅延日数が長く急激に減衰し、黄色の時間緩和関数は赤色のものと同じく遅延日数が長い減衰は緩やかとなっている。

それらの時間緩和関数の違いを可視化で視認可能とする上で、まず、時間緩和関数が極大となる日、

$$t_{DL} = \exp(\mu - \sigma^2)$$

を基準日  $t=0$  からの遅延 (Delay) と定義し、3D glyph の下から 1 段目にある四角柱で表す。基準日からの日数が短いほど基準日から近い頃に訪問者が急増する傾向を示すことを意味し、基準日からの日数が長いほど日数が経過してから POI への訪問者が急増する傾向を示すことになる。言い換えれば、遅延日数が短いほど人気は早期に得られる傾向を示し、遅延日数が長いほど人気は晩期に得られる傾向を示している。ここでは、遅延日数が短いほど濃い緑となり、遅延日数が長いほど白となるように配色した。次に、人々が POI を訪問する行動がピークを迎えてから減衰する傾向を表すうえで、時間緩和関数が極大となる日から急激に減衰したり、ゆっくり減衰することを特徴として視認可能とするため、3D glyph の下から 2 段目の四角柱で減衰傾向を表す。時間緩和関数において、それが極大となる日から右側の変曲点までの日数、

$$t_{DC} = \exp\left(\mu + \frac{-3\sigma^2 + \sigma\sqrt{\sigma^2 + 4}}{2}\right) - t_{DL}$$

をその減衰度 (Decay) と定義する。  $t_{DC}$  が小さいということはピークからの速い減衰を表しており、それが大きいということはピークからゆっくりとした減衰を表している。ここでは、減衰が速いほど濃い青となり、ゆっくりであるほど白くなるように配色した。最後に、各 POI の魅力度を表すフィットネスは、最上段にある四角柱の側面および天井面を用いて図 1a と同様に、濃い赤から白からまでの色を配色して表す。

### 3 実験

#### 3.1 実験データ

Foursquare の check-in データに含まれる、東京とニューヨークにおける 2012 年 4 月 3 日から 2013 年 2 月 16 日までの実データを用いた。活発な POI に焦点を当てて動的分析を行うため、check-in 数が 10 件以上の POI を対象とし、データセットを構築した。そのデータセットをここでは Foursquare データセットと呼ぶ。図 2 は Foursquare データセットにおけるエイジング特性の観点から POI 訪問行動に関する動的分析を行った結果である。エイジング特性の検証においては初日の設定が必要となる。Foursquare の POI 訪問行動においては、史跡や海岸、洞窟の他、最初の日を決定し難い分析対象が多い。そこで、本研究では基準日を定めるうえで、Foursquare データセットの期間における初

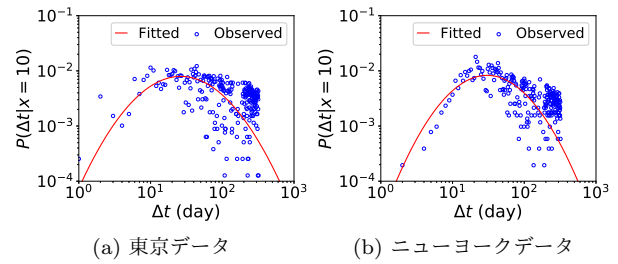


図 2: Foursquare データにおけるエイジング特性

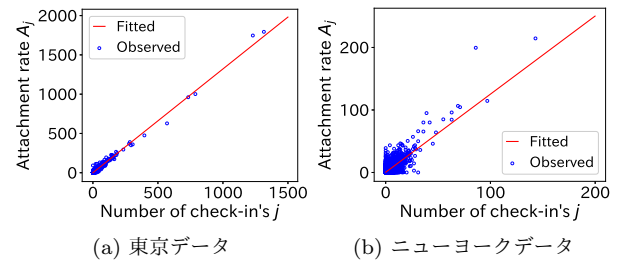


図 3: Foursquare データにおける”rich-get-richer”性

日をすべての POI 共通の基準日とした。先行研究 [10] と同様に、エイジング特性の観点から東京 (図 2a)、ニューヨーク (図 2b) の結果を見ると、ともに対数正規分布に従う傾向があることが見てとれる。本研究では時間緩和関数として対数正規分布を用いることが妥当であると考えられる。また、図 3 は”rich-get-richer”性の検証において、先行研究 [11] の手法を Foursquare データに適用した結果である。Attachment 比は、2012 年 4 月から 6 月の POI 訪問行動を調べた場合の check-in 数に対するその直後の一ヶ月間 (7 月) の check-in 数との関係を示している。check-in 数に対する Attachment 比が明瞭な線形性を示していることから、Foursquare データの東京 (図 3a) およびニューヨーク (図 3b) においても”rich-get-richer”性を確認することができる。

本研究では、動的分析における季節性の有無を調べるため、東京とニューヨークのデータに対して、4 月から 6 月までの春と 10 月から 12 月までの秋についての 2 つのデータセットを構築した。なお、東京データの春データセットにおける POI 数は 3,339、check-in 総数は 162,861、また、東京データの秋データセットにおける POI 数は 2,165、check-in 総数は 97,740 となった。次に、ニューヨークデータの春データセットにおける POI 数は 2,386、check-in 総数は 62,329、また、ニューヨークデータの秋データセットにおける POI 数は 801、check-in 総数は 16,829 となった。また、基準日については、春と秋それぞれのデータ上の初日とした。なお、先行研究 [12] に従ってパラメータ  $\gamma$  を 1 に設定し、 $\nu$  を 10 に設定した。

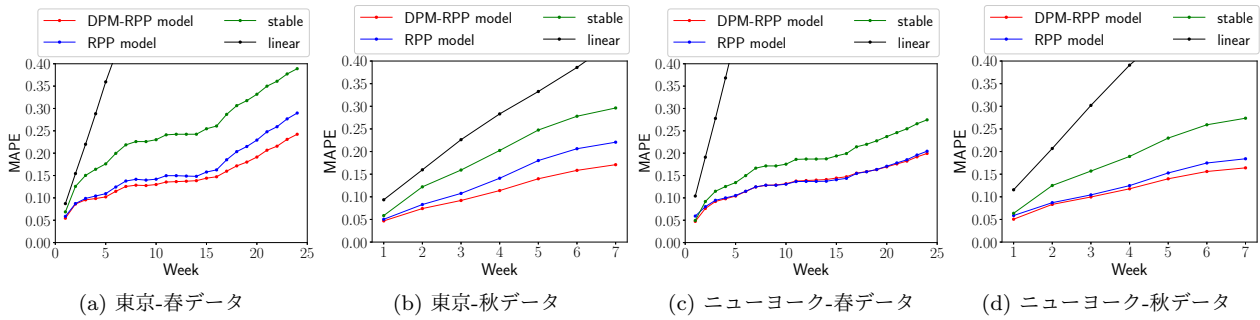


図 4: ポピュラリティ予測性能の比較

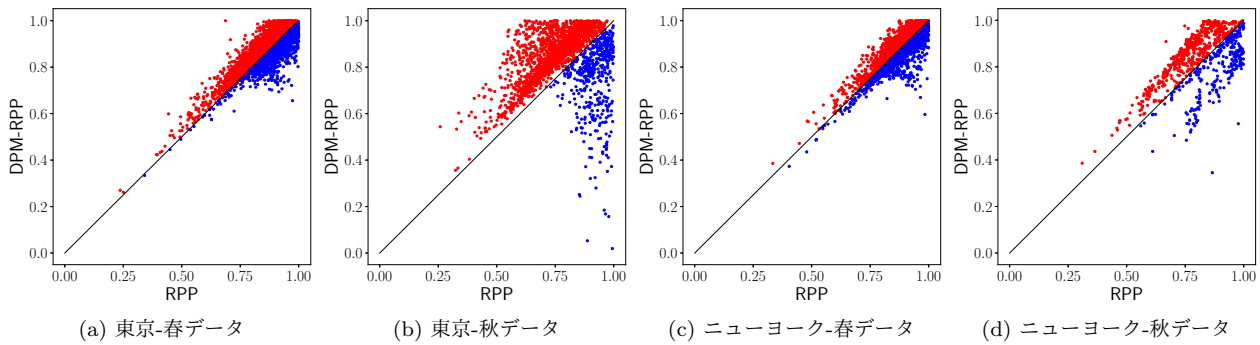


図 5: DPM-RPP と RPP による POI のポピュラリティ予測精度

表 1: DPM-RPP に従う POI 数と RPP に従う POI 数.

	東京		ニューヨーク	
	春	秋	春	秋
DPM-RPP (赤)	2,151	1,678	1,236	527
RPP (青)	1,188	481	1,150	274

### 3.2 ポピュラリティ予測の結果

全ての POI に対して、期間  $[0, T')$  内の訪問ダイナミクスを用いて、将来の  $t (\geq T)$  時間後までに獲得する check-in 総数  $x^d(t)$  の予測を行い、その性能を評価した。DPM-RPP および RPP モデル以外に、単純法として学習期間の最終時点から過去 1 週間の変化量に基づいて線形予測を行うモデル (linear) と、学習期間の最終時点から POI 訪問は変化しないと見なす定常予測モデル (stable) との比較を行う。我々は、先行研究 [12] と同様にハイパーパラメータを設定し、ポピュラリティ予測において標準的に用いられる Mean Absolute Percentage Error (MAPE) を指標として予測性能を測定した。図 4 に、4 つのモデルに基づく MAPE を指標として用いたポピュラリティ予測の結果を示す。ニューヨークの春データについては DPM-RPP と RPP の予測性能に顕著な差は現れなかったものの、東京の春と秋のデータおよびニューヨークの秋データの結果においては  $T' - T$  が大きくなるほど DPM-RPP は RPP よ

りも予測性能が高くなる傾向が見てとれる。

### 3.3 可視化分析の結果

可視化分析においては、各 POI ごとに DPM-RPP モデルと RPP モデルを適用したポピュラリティ予測において、より高い精度を示した方のモデルに対応する 3D glyph を使って可視化を行うことに注意しておく。なお、予測期間の長さが短いニューヨーク-秋データに合わせ、 $T' - T$  を 6 週間とした。図 5 は、各 POI に対して、横軸に RPP モデルによる予測精度、縦軸に DPM-RPP モデルによる予測精度をとったポピュラリティ予測精度の POI 分布であり、図 5a と図 5b はそれぞれ東京の春と秋のデータに関する結果、図 5c と図 5d はそれぞれニューヨークの春と秋のデータに関する結果である。それぞれの図において、赤い点は DPM-RPP の精度が高かった POI、青い点は RPP の精度が高かった POI を表している。表 1 は、図 5 において DPM-RPP の予測精度が RPP に対して上回った赤い POI の数と RPP の予測精度が DPM-RPP に対して上回った青い POI の数を東京およびニューヨークの春と秋のデータで比較したものである。いずれも、DPM-RPP に従う POI の数が多いことから、RPP に従う POI よりも DPM-RPP に従う POI が多いことが



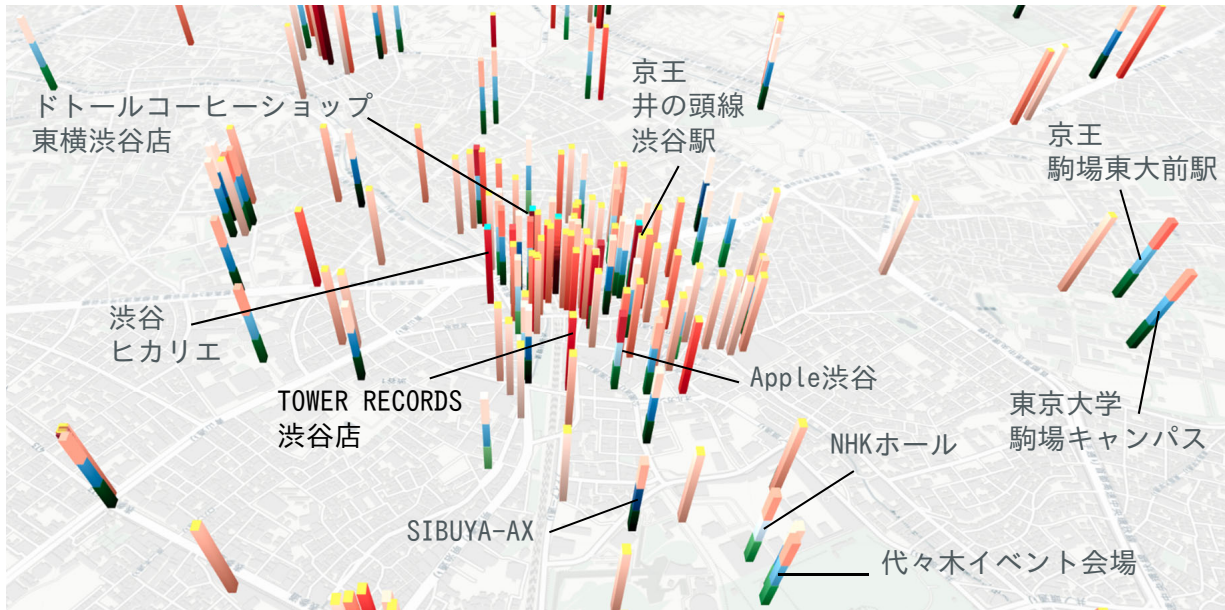


図 8: 東京-春データ (渋谷周辺) の可視化結果

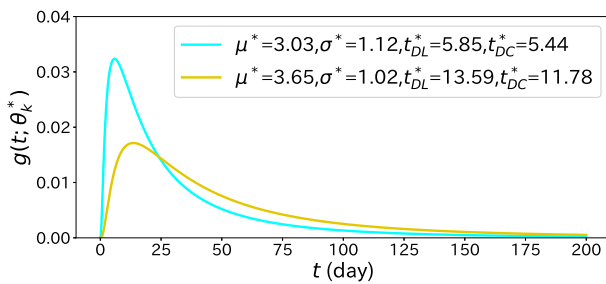


図 6: 主要クラスタの時間緩和関数 (東京-春データ)

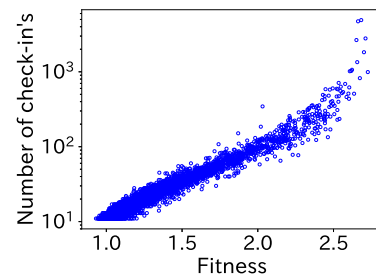


図 7: check-in 数とフィットネスとの関係 (東京-春データ)

わかる。ここでは頁数の制限から、東京-春データに着目した可視化結果を用いて地図上での POI の地域分析を試みる。なお、3D glyph を描画する地図システムとして、OSM Buildings を用いた。

### 3.3.1 東京-春データの結果

図 6 は、DPM-RPP モデルに基づいて推定された異なるエイジングタイプに対応する時間緩和関数である。東京-春データでは、2つの主要なエイジングタイプが得られた。黄色のエイジングタイプに従う POI 数は 2,990、水色のエイジングタイプに従う POI 数は 348 であった。また、黄色のエイジングタイプの遅延  $t_{DL}$  が 13.59 日であるのに対して、水色のエイジングタイプでは遅延  $t_{DL}$  が 5.85 日であり、時間緩和関数が極大となるまでの日が半分程度と遅延が短い傾向があることがわかる。また、DPM-RPP モデルに従う 2つのエイジングタイプ構造の減衰傾向としては、水色の方が黄

色の方よりも急峻に減衰することがわかる。すなわち、水色のエイジングタイプに従う POI 群は、もう 1つのエイジングタイプに対して高いアテンションを獲得するがすぐに忘れ去られる傾向があることを示唆している。図 7 は、東京-春データに含まれる各 POI ごとに check-in 数とフィットネスの関係を示した散布図である。各 POI への訪問 (check-in) 数 (人気度) はフィットネスと関係性があることがわかる。すなわち、フィットネスを魅力度と見なすのは妥当であると考えられる。

DPM-RPP モデルに従う 2つのエイジングタイプそれぞれに所属する POI 群や RPP モデルに従う POI 群が実世界上でどのように分布しているかを確かめるために、各 POI に対応する 3D glyph を地図システム上に描画する。東京-春データに基づいて、渋谷周辺を可視化した結果が図 8 である。渋谷の中心街周辺には DPM-RPP における黄色のエイジングタイプに従う POI 群が多く、それらの多くの魅力度が低めであることがわか

る。ただし、TOWER RECORDS 渋谷店は黄色のエイジングタイプであるが魅力度が非常に高いことがわかる。京王井之頭線渋谷駅や渋谷ヒカリエ、ドトールコーピーショップ東横渋谷店は DPM-RPP における水色のエイジングタイプに従い、魅力度が高く、エイジングタイプ構造の観点から  $D$  内の他の POI 群の訪問ダイナミクスと関係していると考えられる。一方、SIBUYA-AX というライブ会場は RPP に従い、3D glyph の 1 段目が濃い緑、すなわち遅延がかなり短く、2 段目が濃い青、すなわち減衰もかなり速い傾向があり、他の POI 群との関係があまりない独自の訪問ダイナミクスをもっていると考えられる。また、SIBUYA-AX に対して東京大学駒場キャンパスや京王駒場東大前駅、代々木イベント会場は 3D glyph の 1 段目が少し濃い緑、すなわち遅延が少し長めで、2 段目が水色、すなわち減衰も少し緩やかであるが、NHK ホールは 1 段目が緑、すなわち遅延がさらに長めで、2 段目が白、すなわち減衰は非常にゆっくりであることがわかる。

以上のように、提案法による可視化分析により、POI 群のそれぞれがどのような POI 訪問ダイナミクスを示すかという動的な観点から地域分析を可能にし得ることが示唆された。

## 4 まとめ

本研究では、観光プランニング支援や観光マーケティングへの活用基盤として、観光都市における POI 群への人々の訪問ダイナミクスの特徴を視覚的に分析できるシステムの構築を目指し、ソーシャルメディアのアイテムに対するアテンション到着過程の確率モデルである RPP モデルと DPM-RPP モデルを援用した、POI 訪問ダイナミクスの可視化分析法を提案した。Foursquare データを用いた実験において、東京およびニューヨークでの POI 群に対する訪問ダイナミクス分析のケーススタディを通し、動的なアテンション獲得過程について行動データ分析を行える可能性を示した。これらの結果は観光マーケティングにつながると考えている。

## 参考文献

- [1] Yang, D., Zhang, D., Zheng, V.W., and Yu, Z.: Modeling user activity preference by leveraging user temporal characteristics in LBSNs, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 45, No. 1, pp. 129–142 (2015).
- [2] Chen, S., Yuan, X., Wang, Z., Guo, C., Liang, J., Wang, Z., Zhang, X.L., and Zhang, J.: Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 22, No. 1, pp. 270–279 (2016).
- [3] 相原健郎: ビッグデータを用いた観光動態把握とその活用: 動体データで訪日外客の動きをとらえる, 情報管理, Vol. 59, No. 11, pp. 743–754 (2017).
- [4] Szabo, G. and Huberman, B.: Predicting the popularity of online content, *Communications of the ACM*, Vol. 53, No. 8, pp. 80–88 (2010).
- [5] Yang, J. and Leskovec, J.: Patterns of temporal variation in online media, *Proceedings of WSDM'11*, pp. 177–186 (2011).
- [6] Bandari, R., Asur, S. and Huberman, B.: The pulse of news in social media: Forecasting popularity, *Proceedings of ICWSM'12*, pp. 26–33 (2012).
- [7] Pinto, H., Almedia, J. and Goncalves, M.: Using early view patterns to predict the popularity of youtube videos, *Proceedings of WSDM'13*, pp. 365–374 (2013).
- [8] Cheng, J., Adamic, L., Dow, P., Kleinberg, J. and Leskovec, J.: Can cascades be predicted?, *Proceedings of WWW'14*, pp. 925–936 (2014).
- [9] Shen, H., Wang, D., Song, C. and Barabási, A.-L.: Modeling and predicting popularity dynamics via reinforced Poisson processes, *Proceedings of AAAI'14*, pp. 291–297 (2014).
- [10] Wang, D., Song, C. and Barabási, A.-L.: Quantifying long-term scientific impact, *Science*, Vol. 342, No. 6154, pp. 127–132 (2013).
- [11] Redner, S.: Citation statistics from 110 years of Physical Review, *Physics Today*, Vol. 58, pp. 49–54 (2005).
- [12] 松谷貫司, 木村昌弘: ソーシャルメディアのアイテム群に対するアテンションダイナミクスの学習, 情報処理学会論文誌「数理モデル化と応用」, Vol. 10, No. 2, pp. 14–24 (2017).
- [13] Neal, R. M.: Markov chain sampling methods for Dirichlet process mixture models, *Journal of Computational and Graphical Statistics*, Vol. 9, No. 2, pp. 249–265 (2000).

# 研究評価指標に関する考察

## Consideration on Scientific Indicators

清水 勝太<sup>1</sup> 高間 康史<sup>1</sup>

Shota Shimizu<sup>1</sup>, Yasufumi Takama<sup>1</sup>

<sup>1</sup> 首都大学東京システムデザイン研究科

<sup>1</sup>Graduate School of System Design, Tokyo Metropolitan University

**Abstract:** In recent years, many scientific indicators have been proposed for measuring the impact of research, and as the selection criteria of academic journals. Although these indicators give us quantitative evaluation based on the number of citations and downloads of papers, and the number of publications of authors, there is a problem that these indicators are not based on the contents of a paper. In order to support researchers to find appropriate journals for their paper submission and information gathering, a novel indicator that overcomes the above-mentioned problem is needed. This paper compares existing scientific indicators as a preliminary stage for the final objective of research evaluation based on contents.

## 1 はじめに

近年、インターネットの普及に伴い、学術論文の出版、古い論文の電子化、学会や講演における発表資料の公開など、Web を利用した学術情報の流通が盛んである。American Journal Expert 社の調査[1]によれば、2016 年までの 10 年程度で論文出版数は倍増している。また、最も論文出版数の多い学術雑誌 5 つの内 4 つがオープンアクセスジャーナルである。近年では、学術情報専門の媒体が Social Networking Service (SNS) などを通して、専門家でない人からもよくアクセスされるようになっている。このような Web 上での学術情報の流通は今後拡大していくと考えられる。

論文の出版数増大に伴い、様々な問題も発生している。これらの問題は査読と情報収集に関する問題に分類することができる。査読に関する問題の一つとして、投稿論文数増加による査読負担の増大が挙げられる。学術雑誌に投稿される論文が時々刻々と増加しているが、査読を担当する人間はほとんど増えておらず、大量の査読負担が一部の査読者に集中することが懸念されている[2]。情報収集に関する問題としては、流通する学術情報の増大と人間の処理能力のギャップが大きくなることによって、研究者にとって読むべき論文の選定が困難になっていることが挙げられる。

以上の問題に対して、Impact Factor (IF) [3]などの研究評価指標が用いられることがある。例えば、研究者は自身の論文を投稿する学術雑誌をその雑誌の

IF で判断することがある。同様に、研究評価指標に基づいて読むべき論文を決めることもある。このように研究評価指標は様々な面での選択基準の一つとして利用されている。

本稿では、増加した論文をその内容に基づき評価することを最終的な目標とし、その前段階として、被引用数に基づく既存研究評価指標と内容に基づく論文評価指標の比較に基づき分析を行い、研究評価指標が満たすべき要件と今後の課題について考察する。

## 2 研究評価

研究活動の質を管理する方法として、研究評価が長年行われている。査読は最も古く[4]、現在でも行われている研究評価方法の一つである。査読の目的は、学術雑誌の限られた誌面に掲載する論文を選定すること、投稿された研究が科学的に妥当であるかどうかを検討することである。近年では、競争的研究資金の獲得や、研究活動の説明責任の追及、また、大学や研究機関の評価など、様々な背景から研究評価の必要性が取り上げられている。

現代の研究評価は研究の結果のみならず、そのプロセスやその後の波及効果なども対象とすることがある。しかし、研究成果の主要な発表手段は論文であるので、本稿では、学術雑誌あるいは論文を対象とした研究評価を対象とする。現在主流の被引用数に基づく研究評価指標はいくつか存在するが、そのほとんどが典型的な問題点を抱えている。以下、主な研究評価指標について述べる。

## 2.1 Impact Factor

Impact Factor (IF) は最も有名な研究評価指標の一つであり、学術雑誌に対して与えられる定量的な指標である。これは対象とする学術雑誌に掲載された論文が直近二年間で、平均してどの程度引用されたかを示す指標で、一年ごとに算出される。例えば、ある学術雑誌のある $y$ 年の IF,  $I_y$ は以下のように定められる。

$$I_y = \frac{C_{y-2}^y + C_{y-1}^y}{P_{y-2} + P_{y-1}}, \quad (2.1)$$

ここで、 $P_y$ は該当する学術雑誌が $y$ 年に掲載した論文の総数、 $C_x^y$ は $x$ 年に掲載された論文が $y$ 年に引用された回数を示す。

IF は当初、収蔵スペースに限りのある図書館が購読する学術雑誌を決めるための参考として考えられた。そのため、同分野の学術雑誌を比較する際には有用である。一方で、IF は論文単体に与えられる評価ではなく、学術雑誌に与えられる評価であることや、被引用数しか考慮しないため、研究の内容評価に用いられることは疑問視されている[5]。例えば、2015年の Nature の IF は 38.1 だが、掲載された論文の 75.5% は 35 回以下の被引用数しか持たない。つまりほとんどの論文が過大評価されている状態である。こういった欠点があるにもかかわらず、IF を研究評価の指標として用いる場面は多数存在している。

また、IF を部分的に改良したものや、分野ごとの評価値を正規化したものなど、IF をベースラインとした被引用数に基づく評価指標が多数提案されている[6]-[9]。しかし、これらの指標は被引用数に基づくため 2.4 節で後述する問題を抱えている。

## 2.2 h-index

h-index[10]は論文の著者に対する評価指標である。これは研究者が発表した論文数とその論文がどの程度引用されているかを示す定量的指標であり、定義は「被引用回数が $h$ 回以上である論文が $h$ 本以上あることを満たす最大の数値 $h$ 」である。h-index を用いることで、ある著者の論文出版数と被引用数（論文の質とみなされている）を同時に扱うことができる。研究分野や研究慣習の異なりを超えて、著者の持つ研究への量的、質的な貢献度を測ろうというのが h-index による評価の意図である。

h-index の欠点として、10 回引用された論文を 100 本持つ研究者と、100 回引用された論文を 10 本もつ研究者への評価が同じになってしまう点や、研究歴

の短い研究者はそもそも論文出版数が少なく、研究歴の長い研究者に対して不利である点などが挙げられる。

## 2.3 Altmetrics

IF は学術雑誌、h-index は著者に与えられる指標であった。一方、Altmetrics は論文出版数が増えたことや、SNS の利用が増大したことに伴い、論文単位で評価を行うために提唱された[11]。Altmetrics とは、論文や研究成果の影響を、ソーシャルメディアを通じて定量的に測定する手法と、およびこれを用いた研究評価活動のことを指す。これにより、即時的かつ多面的に論文単位での評価が可能となる。Altmetrics の特徴として、専門家に限らず一般の人に対する研究の影響を測定できる点、被引用数による評価の補完・代替となる可能性、論文発表直後から評価を行える即時性などが挙げられる。Altmetrics 計測の要素として、Altmetrics 計測サービスの Impact Story は以下の 5 つを挙げている[12]。

1. Viewed
  - PDF ファイルなどのダウンロード数
2. Discussed
  - SNS での言及回数
3. Saved
  - Mendeley<sup>1</sup>などでのブックマーク数
4. Cited
  - 論文や Wikipedia での引用数
5. Recommended
  - プレス記事などでの推薦数

これらの項目に基づいて論文の評価を行う。例えば、新規論文が Twitter で言及された回数や、論文 PDF ファイルのダウンロード数で、研究の影響度を測定する。しかし、より新しい論文が有利に測定されやすい点や、SNS での言及回数の測定方法に一貫性がない点など、ソーシャルメディア利用による問題も存在する。

## 2.4 被引用数に基づく研究評価指標の課題

被引用数に基づく研究評価指標には、以下に挙げる点が問題として付随する。Altmetrics は評価項目として、被引用数も含んでいるので、被引用数に基づく研究評価指標として考える。

1. 研究発表から評価までに時間を要する
2. 分野による研究慣習の異なりに影響を受ける

<sup>1</sup> Mendeley <https://www.mendeley.com>

3. データベースに依存する
4. 内容に基づいていない
5. 論文の種類による引用のされやすさが異なる
6. 共著者の貢献度を評価できない

これらのうち、1, 2 はこれまでに述べた通りである。データベース依存性は被引用数に基づく評価方法の典型的な課題である。例えば、IF は Web of Science に収録された論文にしか適用されず、これに収録されていない論文からの引用も被引用も考慮されていない。

また、どの評価手法でも、論文の内容を考慮しているものはない。論文の内容を考慮せず、被引用数に基づき評価する場合の弊害として、評価が論文の種類に依存することが挙げられる。様々な研究内容について触れるレビュー論文は引用のベースラインとして利便性が高いので引用されやすく、レビュー論文を含む学術雑誌は被引用数に基づく評価が高くなりやすい。

さらに、被引用数に基づく評価では、共著の論文に関して、著者別の貢献度を評価できないことも、問題点として挙げられる。被引用数に基づく評価では、これらの問題が研究評価の妥当性に対して常に存在する。

### 3 内容に基づく 論文評価指標の検討

前節で挙げた被引用数に基づく研究評価指標の問題点のうち、本稿では論文内容を考慮していない点に着目し、これを考慮した評価可能性について検討する。

一般に、情報検索やテキストマイニングなどでは文書の内容としてそのトピックに着目し、トピックの類似性に基づいて文書の検索やクラスタリングを行う。一方、トピックとは異なる観点として、論文の書き方・表現も内容の一種と考えられる。採択率の低いトップジャーナルやトップカンファレンスでは、研究内容の新規性や信頼性だけでなく、論文の書き方についても査読により厳しく評価されている。その結果、十分な研究業績を持つ研究者は論文執筆スキルも高いといえる。従って、トップジャーナルなどに掲載された論文と、書き方に関する類似性の高い論文の質は高い事が期待できる。

この仮説について検証するために、本稿では文献[13]で提案されている、学術雑誌間類似度を用いた予備実験結果について示す。また、単一の雑誌のみを用いた内容に関する評価指標として、同文献[13]で提案されている自動要約指標を用いた手法の予備実験結果も示す。

### 3.1 学術雑誌間類似度

この手法では、論文中のテキストを文単位でベクトル化し、ベクトル間の類似度をコサイン類似度で定義し、論文間類似度を定義する。さらに、論文間類似度に基づいて学術雑誌間類似度を定義し、学術雑誌の評価を行う。それぞれの定義を以下に示す。

$$S_{\text{Journal}}(X, Y) = \frac{1}{|X_N||Y_N|} \sum_{p_i \in X_N} \sum_{p_j \in Y_N} S_{\text{paper}}(p_i, p_j), \quad (3.1)$$

$$S_{\text{paper}}(p_i, p_j) = \frac{1}{|p_i||p_j|} \sum_{s_k \in p_i} \sum_{s_l \in p_j} S_{\text{sentence}}(s_k, s_l), \quad (3.2)$$

$$S_{\text{sentence}}(s_k, s_l) = \frac{\langle s_k, s_l \rangle}{|s_k||s_l|}, \quad (3.3)$$

ここで、 $S_{\text{sentence}}(s_k, s_l)$  は文をもとに生成されたベクトル  $s_k, s_l$  のコサイン類似度を示す。 $S_{\text{paper}}(p_i, p_j)$  は論文  $p_i, p_j$  の論文間類似度を示す。2 件の論文について、文間類似度 (式(3.3)) の平均が論文間類似度である。 $S_{\text{Journal}}(X, Y)$  は学術雑誌  $X, Y$  の学術雑誌間類似度を示す。 $p_i, p_j$  はそれぞれ  $X_N, Y_N$  に含まれる論文を示す。 $S_{\text{Journal}}(X, Y)$  を算出する際、取得可能な論文数の差、計算量を考慮し、一定数 ( $N$ ) の論文を各学術雑誌からサンプリングして論文間類似度の算出に用いる。

本稿では、基準となる既存評価指標として、Scimago Journal Rank (SJR) [6] と h-index [10] を用いた。図 3.1, 3.2, 3.3 に論文間類似度を用いた手法による学術雑誌の評価を示す。

図 3.1 は分子生物学分野の学術雑誌について、トップジャーナルである Cell との学術雑誌間類似度を示したものであり、横軸に SJR の値、縦軸に基準雑誌 (Cell) との類似度を示している。SJR の値が高いほど基準雑誌との類似度が高い傾向が確認できる。

図 3.2, 3.3 は h-index を基準指標とした学術雑誌間類似度を示している。基準となる著者の論文と対象となる学術雑誌の分野は計算機科学分野とした。

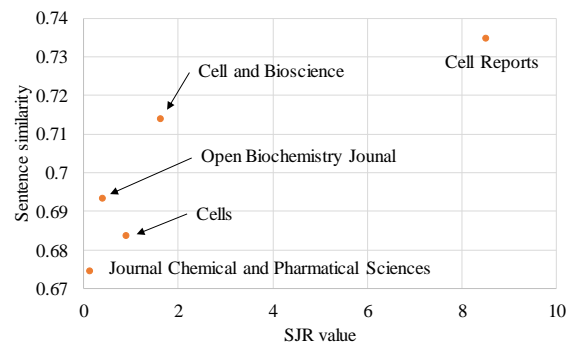


図3.1 分子生物学分野雑誌の学術雑誌間類似度



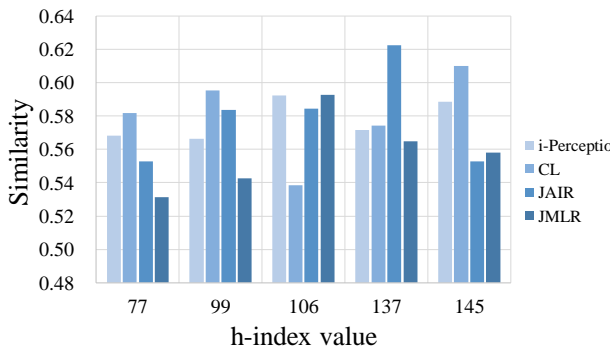


図3.2 著者別の学術雑誌間類似度

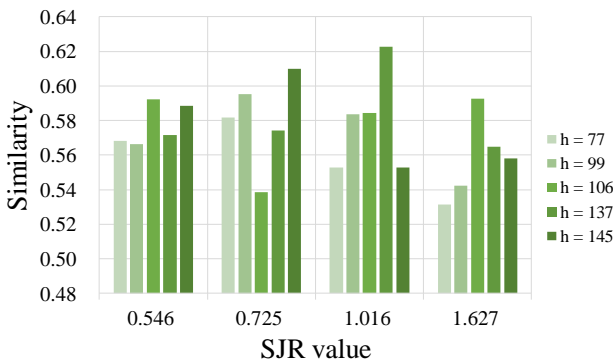


図3.3 雑誌別の学術雑誌間類似度

凡例に雑誌名と著者の h-index 値を示している。両図から h-index と SJR, 学術雑誌間類似度の間に一貫した傾向がないことがわかる。ここから, h-index を基準にしても既存評価指標と相関のある評価は行えないと考えられる。

### 3.2 要約評価指標に基づく評価

自動要約システムの評価は, システムによる要約 (評価対象) と人手の要約 (正解データ) との一致度合いに基づいて行われる。Abstract, Summary といった論文の要約セクションを正解データ, 研究内容を示す本論セクションを評価対象として ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [14][15] を適用することで, 要約セクションと本論セクションの一貫性が評価可能と考える。ROUGE による正解データ  $R$  と評価対象  $S$  に関する評価値  $S_{ROUGE}(S, R)$  は式(3.4)で定義される。

$$S_{ROUGE}(S, R) = \frac{\sum_{e \in n_{gram}(S)} C_{match}(e)}{\sum_{e \in n_{gram}(R)} C(e)}, \quad (3.4)$$

ここで,  $n_{gram}(\cdot)$  はテキストに含まれる単語 n-gram を示す。  $C(e)$  は文書中における  $e$  の出現頻度,  $C_{match}(e)$  は正解データ  $R$  と評価対象  $S$  に含まれる  $e$  の共起回数を示す。式(3.4)は Recall と同じ定義であるが, 同様

表3.1 Cell の要約評価

	Precision	Recall	F-value
Introduction	0.106	0.506	0.175
Results	0.016	<b>0.710</b>	0.031

表3.2 Cell Reports の要約評価

	Precision	Recall	F-value
Introduction	0.113	0.515	0.185
Results	0.018	<b>0.713</b>	0.035

に Precision を算出することもでき, そこから F-value も求めることができる。

表 3.1, 3.2 に Cell, Cell Reports の要約評価を示す。Summary を正解データ, 研究内容を示す Introduction, Results を評価対象としている。両表からどちらの雑誌でも, Results の方が Recall の値が高いことがわかる。これは, Introduction には研究背景として, 当該論文以外の研究内容が相対的に多く含まれるのに対し, Results に示される研究成果の主要な部分は Summary にも含まれるためと考える。

## 4 研究評価指標の比較と考察

3 節の結果を踏まえて, 表 4.1 に研究評価指標の比較を示す。2.4 節で述べた各問題点の解消を要件とみなし, それぞれに対する各評価指標 (手法) の相対的, 定性的な性能比較を ○, △, × で示す。

評価までの時間に対しては, h-index が最も性能が悪く, 論文が引用されるまで評価が与えられない。IF は 2 年ごとに評価値が与えられ, Altmetrics は発表された瞬間からダウンロード数などが与えられ,

表4.1 研究評価指標の比較  
 “-” は評価なしを示す。

	Impact Factor	h-index	Altmetrics	$S_{Journal}$ / ROUGE
評価までの時間	△	×	○	○
異分野間対応	×	-	-	△
DB 非依存性	×	×	○	△
著者の評価	×	△	×	×
内容に基づく評価	×	×	×	○
論文単位の評価	×	-	○	△

3 節の手法 ( $S_{\text{Journal}}$  / ROUGE) も評価対象が発表されてすぐに評価が可能である。

分野の異なりに対して IF は前述のとおり、影響を受けるので最も評価が悪い。h-index, Altmetrics は分野間の異なりを対象としないので評価なしとした。3 節の手法は分野の異なりに部分的に対応している [13]。

IF, h-index はその算出がデータベースに依存するので、データベース非依存性の評価が最も悪く、Altmetrics では論文が多様なメディアにより流通されてから評価が行われるので非依存性が高いと判断される。一方、3 節の手法は、データベースには依存しないが、評価対象のジャーナルに収録された論文を取得・分析する必要があるため中程度の評価とした。

著者の評価に対しては、全ての評価指標で十分ではないといえる。例として、複数の共著者に対して、貢献度を差別化できない点が挙げられる。h-index は著者個人に評価を与えるので中程度の評価としたが、各論文への貢献度を判断できない点は他の指標と同様である。

内容に基づく評価に対しては、既存評価指標は全て十分ではないといえる。3 節の手法は現時点で十分な性能を持つとは言えないが、相対的に高い評価としている。

論文単位の評価に対しては、IF は論文単体を対象としていないので低い評価となる。h-index は論文出版数に基づき算出されるものの、論文に対して評価を与えないので、評価なしとした。3 節の手法は、ROUGE を用いる方法が論文単体に対応しているといえる。

以上より、IF, あるいはそれに類する被引用数に基づく研究評価指標を研究や研究者の評価として用いることの妥当性は低く、内容に基づく評価指標の導入が必要と考える。

## 5 おわりに

本稿では、被引用数に基づく評価指標と内容に基づく評価指標による学術雑誌・論文の評価について比較考察を行った。

考察の結果、被引用数に基づく評価指標は共通の課題を抱えていることを明らかにした。これら課題の解消を要件として、各研究評価指標を比較した結果、代表的な研究評価指標として頻繁に用いられる IF は、ほとんどの要件に対して評価が悪いことを示した。h-index も同様に、ほとんどの要件に対して、不十分であるか、評価なしであった。Altmetrics はそれらに比べ、対応している評価要件が多い、論文の内容に基づいて評価が行えない。一方、3 節の手法は

内容に基づいて評価が行える利点があり、他の要件に対しても、相対的に高い評価となることを示した。

本稿で示した比較結果は相対的かつ定性的なものであり、3 節で示した手法により信頼性の高い評価が可能かどうかは検証されていない。定量的な評価実験などを通じ、評価指標としての確立を目指すことが今後の課題である。

## 参考文献

- [1] American Journal Expert.: AJE Scholarly Publishing Reports:2016, American Journal Expert, (2016)
- [2] M. Kovanis, R. Porcher, P. Ravaud, and L. Trinquart.: The global burden of journal peer review in the biomedical literature: Strong imbalance in the collective enterprise, PLoS ONE, Vol. 11, No. 11, (2016)
- [3] E. Garfield.: The History and Meaning of the Journal Impact Factor. Journal of the American Medical Association, Vol. 295, No. 1, pp. 90-93. (2006)
- [4] R. Spier.: The history of the peer-review process, Trends in biotechnology, Vol. 20, No. 8, pp. 357-358, (2002)
- [5] E. Callaway.: Publishing elite turns against impact factor, Nature, Vol. 535, No. 14, pp. 210-211, (2016)
- [6] V. P. Guerrero-Bote and F. Moya-Anegón.: A further step forward in measuring journals' scientific prestige: The SJR2 indicator, Journal of Informetrics, Vol. 6, No. 4, pp. 674-688, (2012)
- [7] R. M. Alguliyev and R. M. Alguliyev.: Modified Impact Factors, Journal of Scientmetric Research, Vol. 5, No. 3, pp. 197-208, (2017)
- [8] J. D. West, T. C. Bergstrom, and C. T. Bergstrom.: The Eigenfactor Metrics™: A Network Approach to Assessing Scholarly Journals, College & Research Libraries, Vol. 71, No. 3, pp. 236-244, (2010)
- [9] H. F. Moed.: Measuring contextual citation impact of scientific journals, Journal of Informetrics Vol. 4, No. 3, pp. 265-277, (2010)
- [10] J. E. Hirsch.: An index to quantify an individual's scientific research output, Proceedings of the National Academy of Sciences of the United States of America, Vol. 102, No. 46, (2005)
- [11] J. Priem, D. Taraborelli, P. Groth, and C. Neylon.: altmetrics: a manifesto, <http://altmetrics.org/manifesto/>, (2010), 最終アクセス 2018 年 11 月 8 日.
- [12] Impact Story.: A new framework for altmetrics, <http://blog.impactstory.org/31524247207/>, (2012), 最終アクセス 2018 年 11 月 8 日.
- [13] Shota Shimizu and Yasufumi Takama.: Preliminary Investigation on Quantitative Evaluation Method of Scientific Papers based on Text Analysis, In Proceedings of ACM MEDES'18, (2018)
- [14] C. Y. Lin and E. Hovy.: Automatic Evaluation of summaries using N-gram co-occurrence statistics, In Proceedings of NAACL'03, Vol. 1, pp.71-78, (2003)
- [15] C. Y. Lin.: ROUGE: A package for automatic evaluation of summaries, In Proceedings of WAS2004, pp. 74-81, (2004)

# SIGIR2018 参加報告

## Report on SIGIR2018

野本 昌子<sup>1\*</sup> 富樫 陸<sup>1</sup> 鈴木 翔吾<sup>1</sup> 椎野 弘章<sup>1</sup>  
Masako Nomoto<sup>1</sup> Riku Togashi<sup>1</sup> Shogo Suzuki<sup>1</sup> Hiroaki Shiino<sup>1</sup>

<sup>1</sup> ヤフー株式会社

<sup>1</sup> Yahoo Japan Corporation

**Abstract:** The 41st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018) took place in Ann Arbor, USA in July 2018. A total of 409 papers for the full paper track were submitted and reviewed by six tracks, for an acceptance rate of 21%. About 46.0% of the accepted papers are related to neural network, distributed over about 73.1% of technical sessions.

## 1 はじめに

情報検索 (Information Retrieval, 以下 IR) のトップカンファレンス SIGIR2018 が 2018 年 7 月に米国で開催され、約 740 名が参加した<sup>1</sup>。

本稿では SIGIR2018 の参加報告として、SIGIR の概要、SIGIR2018 の概要と動向、論文の一部、次回の概要と国内のコミュニティの動向を紹介する。

## 2 SIGIR

SIGIR (SIGIR Conference on Research and Development in Information Retrieval) は IR 分野のトップカンファレンスで、主催組織は検索・情報アクセス分野の国際学会 ACM の IR の分科会 SIGIR (Special Interest Group on Information Retrieval)<sup>2</sup>である。SIGIR は 1978 年から毎年開催されており、ACM SIGIR が後援している国際会議の中ではもっとも歴史が古い。

IR の周辺分野には自然言語処理、人工知能、データマイニング、Human Computer Interaction などがある ([1])。

Microsoft Academic の報告 [2] によると、SIGIR の論文は CIKM<sup>3</sup>、WWW<sup>4</sup>、WSDM<sup>5</sup>等、周辺分野の国際会議と相互に論文の参照関係がある。2015 年から 2017

年の間に SIGIR の論文を参照した venue、SIGIR の論文が参照した venue を各々、図 1<sup>6</sup>、図 2<sup>7</sup>に示す。SIGIR の論文は上に述べた 3 会議の他に、理論寄りの ICTIR<sup>8</sup>、情報推薦の RecSys<sup>9</sup>、データマイニング系の KDD<sup>10</sup>等から参照されている。逆に SIGIR の論文からも上述の 3 会議に加えて、KDD、機械学習系の NIPS<sup>11</sup>や ICML<sup>12</sup>、自然言語処理の ACL<sup>13</sup>や EMNLP<sup>14</sup>、人とコンピュータのインタラクション (HCI) に関する CHI<sup>15</sup>など、多様な会議の論文を参照している。

図 3 に SIGIR の大陸別開催回数を示す<sup>16</sup>。従来 SIGIR は欧米で開催されていたが、近年はアジアとオセアニア、南北アメリカ、ヨーロッパを順に回っている。

## 3 SIGIR2018

### 3.1 開催概要

SIGIR2018 は米国ミシガン州アナーバーのミシガン大学で、7 月 8 日から 12 日まで 5 日間に渡り開催され

<sup>6</sup>[2] のグラフ Top Venue Citation Over Time の 2015 年から 2017 年のデータを元に作成。

<sup>7</sup>[2] のグラフ Top Venue Reference Over Time の 2015 年から 2017 年のデータを元に作成。

<sup>8</sup>ACM International Conference on the Theory of Information Retrieval

<sup>9</sup>ACM Conference on Recommender Systems

<sup>10</sup>ACM SIGKDD Conference on Knowledge Discovery and Data Mining

<sup>11</sup>Conference on Neural Information Processing Systems

<sup>12</sup>International Conference on Machine Learning

<sup>13</sup>Annual Meetings of the Association for Computational Linguistics

<sup>14</sup>Empirical Methods in Natural Language Processing

<sup>15</sup>ACM CHI Conference on Human Factors in Computing Systems

<sup>16</sup>[3] を元に作成。

\*連絡先：ヤフー株式会社

千代田区紀尾井町 1-3

E-mail: mnomoto@yahoo-corp.jp

<sup>1</sup>SIGIR2018 のオープニングの紹介によると、登録者数 740 名、うち約 40% は企業と非アカデミアの組織から参加。

<sup>2</sup>本稿では区別のため、以降、'ACM SIGIR' と記す。

<sup>3</sup>ACM International Conference on Information and Knowledge Management

<sup>4</sup>ACM International World Wide Web Conference

<sup>5</sup>ACM International Conference on Web Search and Web Data Mining

表 1: SIGIR2018-2016 の概要

会議(回)	SIGIR2018(41)	SIGIR2017(40)	SIGIR2016(39)
会期	July 8-12	Aug 7-11	July 17-21
開催地	Ann Arbor Michigan, USA	Tokyo, Japan	Pisa, Italy
参加者数	740	911	566
フルペーパー	86/409 (21%)	78	62
ショートペーパー	98/327 (30%)	121	104
デモ	18/36 (50%)	17	21
チュートリアル	11/13 (85%)	8	12
ワークショップ	10/15 (67%)	8	7
SIRIP(Industry Track)	9/16 (56%)	12	12

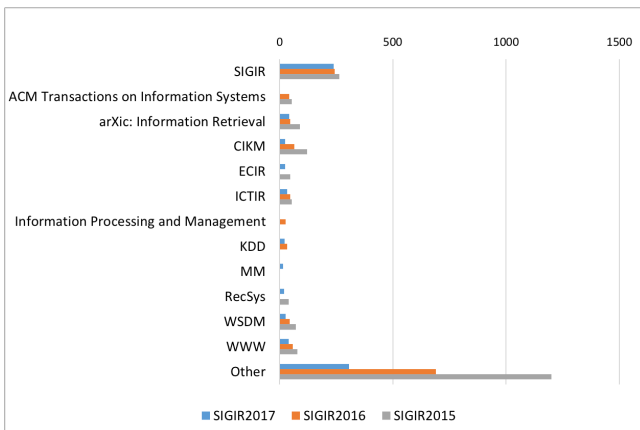


図 1: SIGIR の論文を参照した論文の venue (SIGIR2015-SIGIR2017)

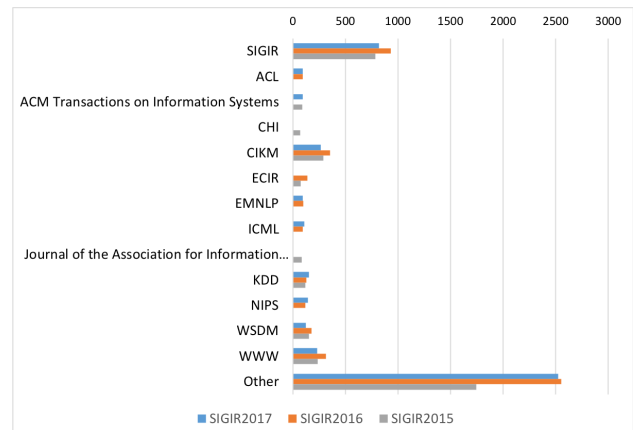


図 2: SIGIR の論文が参照した論文の venue (SIGIR2015-SIGIR2017)

た. 表 1<sup>17</sup>に過去 3 回の SIGIR の開催概要を示す. 今回のプログラムも例年通りチュートリアルとワークショップが各 1 日, 本会議が 3 日間であった. 大きな変化は, 企業の IR の応用に関する SIRIP<sup>18</sup>が 2 日間に拡大されたことである.

過去最大の 911 名の参加者を集めた昨年につき, 今回も約 740 名(登録者数)と多数の参加があった. なお, 2001 年以降で SIGIR の参加者数が 700 名を超えたのは 2006 年以来である ([4],p.9).

日本からは, アカデミアから少なくとも 13 名, 企業からはヤフー, リクルート住まいカンパニー, リクルートジョブズ, メルカリ, NTT コムウェア, Honda R&D 等から少なくとも 17 名が参加した.

表 2 に今回のスポンサーを示し, 前回, 前々回のスポンサーに該当する場合は'\*'を付した<sup>19</sup>. スポンサー全体としては米国と中国の企業が多い. 前回は開催国日本の企業が目立っていたが, 今回は見られない. 情報

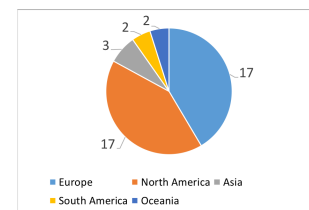


図 3: 大陸別開催回数

通信の企業が多いのは例年通りで, 米国企業の顔ぶれは変わらず, アジア企業が増加を続けている. 今回スポンサーに加わったのは中国の車のライドシェアサービスの DiDi, 音楽ストリーミングサービスの Spotify, 米国の小売業や e コマースの企業である.

### 3.2 投稿・採択状況

図 4 に SIGIR のフルペーパーの投稿数, 採択数, 採択率の推移 (1999 年から 2018 年まで) を示す<sup>20</sup>. 今回は昨年を上回る 409 本のフルペーパーが投稿され, うち 86 本が採択され, 採択率は 21.0%であった. SIGIR

<sup>17</sup>フルペーパー以降の行は SIGIR2018 については採択数/投稿数(採択率), 他は採択数を示す.

<sup>18</sup>The SIGIR Symposium on IR in Practice. Industry Track に相当する.

<sup>19</sup>[5]pp.xxxviii-xxxix, [6]pp.xxxv-xxxvi, [7]pp.xvii-xviii を元に作成.

<sup>20</sup>[5] の Publication, Acceptance Rate の表より作成.

表 2: SIGIR2018 のスポンサー

国	分類	企業	2017	2016
アメリカ	情報通信	Microsoft	*	*
		IBM	*	*
		Bloomberg		*
		Google	*	*
			*	*
	小売/e コマース	ebay	*	*
		amazon	*	*
		wayfair		
		Home Depot		
中国	情報通信	Alibaba	*	
		Baidu	*	*
		Sogou	*	
		Byte Dance		
		kika		
	旅行 (ライドシェア)	DiDi		
韓国	情報通信	NAVER	*	
スウェーデン	音楽	Spotify		

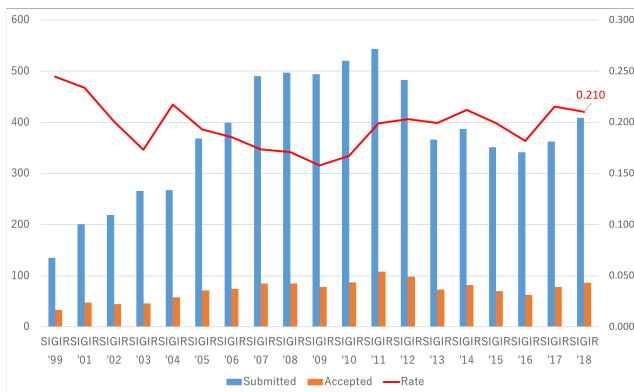


図 4: フルペーパーの投稿数, 採択数, 採択率の推移 (1999 年から 2018 年まで)

全体 (1999 年から 2018 年) の平均採択率 19.1% に比べると今回はやや易しく, 昨年の 21.5% と同程度である。

全投稿論文の国別著者数 (2016 年から 2018 年) を図 5<sup>21</sup> に示す。2016 年時点の投稿数は 1 位が米国, 2 位が中国であったが, 2017 年に 1 位と 2 位が入れ替わった。

次に, フルペーパーの採択論文の国別著者数の割合 (2016 年, 2018 年) を図 6<sup>22</sup> に示す。2016 年の時点では米国が 34%, 中国が 23% であり, 昨年も順位の変動はなかったが<sup>23</sup>, 今回はついに中国が 34%, 米国が 30% と順位が入れ替わった。前回のフルペーパーは日本から少なくとも 3 本が採択されたが, 今回は日本からの採択はなく, 本会議の Interface のセッションでは梅本ら [9] の TOIS (ACM Transactions on Information Systems) のジャーナルの論文の口頭発表<sup>24</sup>があった。

著者の所属の分布を見るため, SIGIR2018 の全著者の所属をアカデミア, 企業, 両方, それ以外に分類し

<sup>21</sup>[5]p.iii, [6]p.iii, [7]p.iii を元に作成。

<sup>22</sup>[5]p.iii, [7]p.v を元に作成。

<sup>23</sup>[8], SUBMISSIONS BY COUNTRY(p.6).

<sup>24</sup>SIGIR2017 から TOIS の論文の著者は CHIIR, SIGIR, IC-TIR のいずれかの国際会議で口頭発表を行えることになった [10].

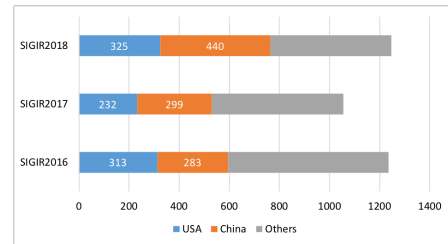


図 5: 全投稿論文の国別著者数 (2016 年-2018 年)

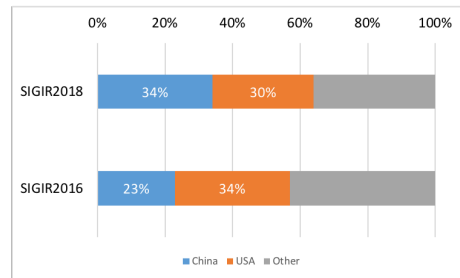


図 6: フルペーパー採択論文の国別著者数の割合 (2016 年, 2018 年)

たところ, SIGIR2018 の全著者の約 78% がアカデミア, 約 20% が企業の著者であった<sup>25</sup>.

次に, SIGIR2016 から SIGIR2018 のフルペーパーの採択論文の著者の所属を同様に分類した結果を図 7 に示す<sup>26</sup>。アカデミアの著者が年々増えているが, 企業の著者数は年により異なる。SIGIR2018 ではアカデミアが 81.6%, 企業が 15.6% であった。

さらに, 2016 年から 2018 年のフルペーパーの企業別著者数を図 8 に示す<sup>27</sup>。2016 年は Yahoo と Microsoft が突出していたが, 2017 年以降は Microsoft が単独で他を引き離している。今回は特に Spotify 等, 新しいスポンサーが目立つ。

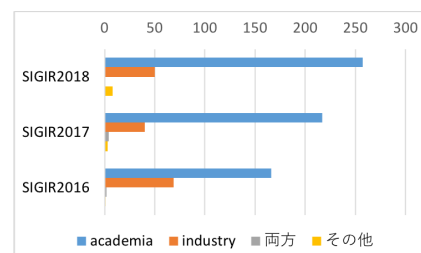


図 7: フルペーパーの著者の所属グループ別著者数 (2016 年から 2018 年)

<sup>25</sup>[11] の full papers, short papers, demo papers, SIRIP Industry Papers, Doctoral Consortium のいずれかの表に含まれる著者の所属を元に分類した。

<sup>26</sup>[11] の full paper の著者の所属, [12], [13] のプログラムの著者の所属を元に作成。

<sup>27</sup>[11] の full paper の著者の所属, [12], [13] のプログラムの著者の所属を元に作成。



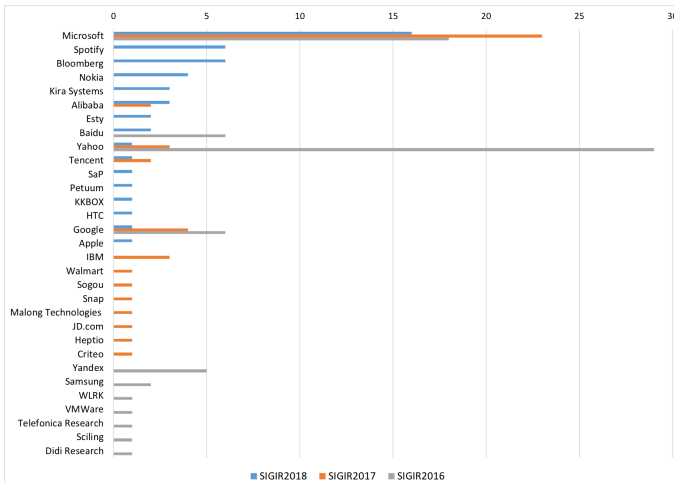


図 8: フルペーパーの企業の著者の所属 (2016 年から 2018 年)

表 3: SIGIR2018 のトラック

トラック
(a) Search and Ranking
(b) Foundations and Future Directions
(c) Domain-Specific Applications
(d) Content Recommendation, Analysis and Classification
(e) Artificial Intelligence, Semantics, and Dialog
(f) Human Factors and Interfaces

### 3.3 トラック

今回は例年と異なり、表 3 に示した 6 つのトラック別に投稿、採択が行われた。投稿の多いトラックは (d) のコンテンツ推薦で、124 件 (投稿論文の 30.0%) の投稿があった<sup>28</sup>。参考のため、前回の推薦に関する投稿を図 9 のトピック別投稿数で見ると、推薦システム (recommender system) が 50 件以上 (投稿論文の 13.8% 以上)、推薦のパーソナライズ (personalized recommendation) が 40 件以上 (同 11.0% 以上)、推薦アルゴリズム (recommendation algorithms) が 40 件弱 (同 11.0% 弱) であった。推薦に関する投稿が多い傾向自体は前回から続いている。

一方、今回採択数の多いトラックは (c) の分野に特化した応用と (f) Human Factors で各々 20 件 (採択論文の 23.2%) である ([5]p.iii)。参考のため、前回の採択論文の中で多かったトピック<sup>29</sup>を見ると、行動分析 (15%) やユーザスタディ (14%) が含まれている。採択論文にユーザに関する研究が目立つのも昨年から続く傾向と考えられる。

<sup>28</sup>[5]p.iii.

<sup>29</sup>The most popular call-for-papers topics for accepted papers were ranking algorithms (19%), text representation (17%), behavioral analysis (15%), recommender systems (14%), and user studies (14%).' ([6]p.iii)

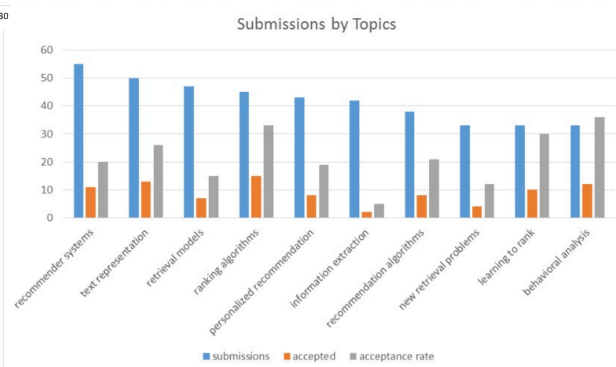


図 9: トピック別投稿数 (SIGIR2017, 投稿数順)

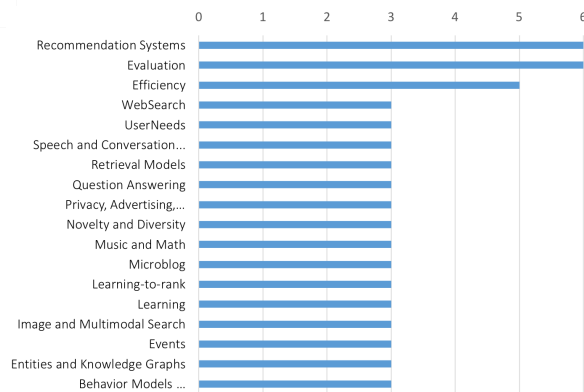


図 10: セッション別論文数 (SIGIR2016)

### 3.4 セッション

SIGIR2016, SIGIR2017, SIGIR2018 のセッション別の論文数を図 10, 図 11, 図 12 の青線で示す。

2016 年は推薦、評価、効率に関するものが多く、新しいセッションでは音声と対話やセマンティック・ウェブのエンティティと knowledge graph があった。

2017 年はフィルタリングと推薦、検索モデルとランキング、検索のインタラクションなどが多かった。また、2016 年に続き、エンティティや、対話と質問応答のセッションもあった。

今回はランキング学習の論文数が突出しているように見えるが、ユーザーの行動の分析に関するもの、推薦に関するものも複数セッションにまたがっている。対話やセマンティック・ウェブ関連のセッションは定着した感がある。新しいセッションの一つにソーシャルグッド (Social Good) があり、これは基調講演のキーワードでもある。

### 3.5 ニューラルネットに関する動向

SIGIR2016 の基調講演で Christopher Manning 氏 (スタンフォード大学) は今後 2, 3 年は深層学習が SI-

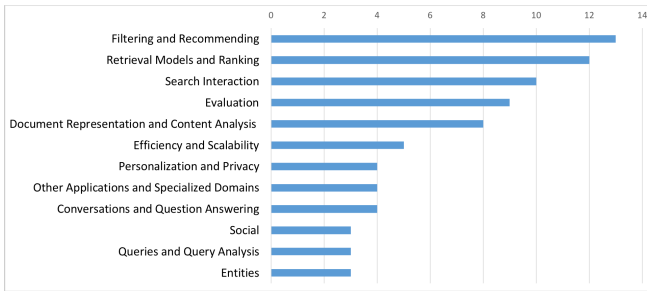


図 11: セッション別論文数 (SIGIR2017)

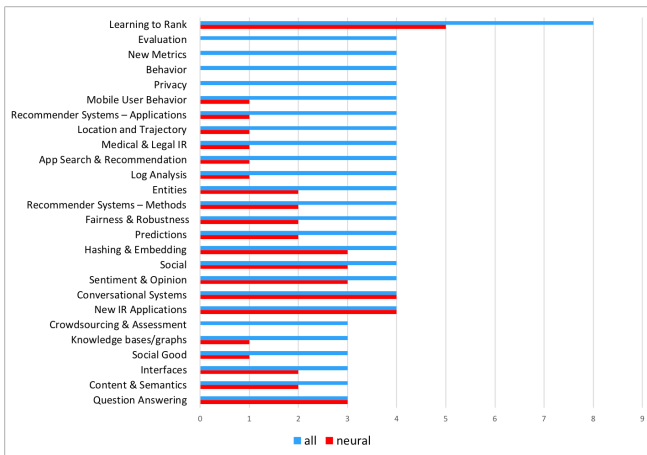


図 12: セッション別の論文数及びニューラルネット関連論文数 (SIGIR2018)

GIR を支配すると予測した [14]。また、前回 SIGIR2017 の最大の話題はニューラルネットで、Neural Information Retrieval のワークショップ (Neu-IR)<sup>30</sup> のオープニングでは約 4 分の 1 の論文が neural IR に関するものであり、「ニューラルネットの年 (The year of the Neural Nets)」と報告された。

そこで、SIGIR2018 の本会議のテクニカルセッションの論文を対象にニューラルネットに関する論文かどうかを調べたところ<sup>31</sup>、フルペーパーの 40/87 本 (約 46.0%)、TOIS の論文も含めると 45/102 本 (約 44.1%) が該当した。各セッションの発表論文のうち、ニューラルネット関連の論文数を図 12 の赤線で示す。さらにセッション単位の分布を見ると、フルペーパーを対象とするとテクニカルセッションの約 73.1% (19/26)<sup>32</sup>、TOIS の論文も含めると同約 80.8% (21/26) でニューラルネット関連の論文が見られる。

<sup>30</sup><https://neu-ir.weebly.com>

<sup>31</sup>調査は基本的にはキーワードベース ('neural net', 'deep learning') で、future work として挙げたもの等、非該当と容易に判別できるもののみ除いた。

<sup>32</sup>Session 1D: Learning to Rank I と 3D: Learning to Rank II はまとめて 1 セッションとカウントした。

### 3.6 基調講演

シカゴ大の研究助教<sup>33</sup>の Ghani 氏はかつてのオバマ大統領の選挙キャンペーン (2012 年) のデータサイエンスチームの中心人物で、シカゴ大ではデータサイエンティストが実社会の問題解決に取り組むようにするための活動<sup>34</sup>に関わっている。講演 [15] ではソーシャルグッド<sup>35</sup>と公共政策を取り上げ、政策や社会問題の解決に機械学習やデータサイエンスを役立てる取組み等を紹介した。

Gerald Salton 賞<sup>36</sup>を受賞したタンペレ大学教授の Jarvelin 氏は Information Seeking や IR の評価に関する研究で著名で、IR の適合性の評価指標 nDCG の最初のバージョンを提案した ([16], pp.32-33) 人物である。講演 [17] は情報のインタラクションに関するもので、これまでの研究を振り返り、情報科学の研究は、人間を仕事や生活の文脈で情報の環境を認識してそれらとインタラクションするアクターとして研究するものと考えたこと、IR の研究でも人に関わる部分が重要であることを強調した。

### 3.7 チュートリアル

今回は表 4 に示す 11 のチュートリアルがあった<sup>37</sup>。

(f) の深層学習のチュートリアル [18] では、検索におけるマッチングと推薦におけるマッチングを取り上げ、それらは各々 SIGIR, RecSys<sup>38</sup>という異なるコミュニティで研究されているが、統一的な視点で整理できることを示した上で、さらに、マッチングの伝統的な手法から最新の深層学習の手法までを概観した。

### 3.8 SIRIP(Industry days)

SIRIP<sup>39</sup>は、企業での IR の応用に関するもので、SIGIR2016, SIGIR2017 では本会議中の 1 日に、テクニカルセッションと並行して行われていたが、今回は 2 日間 (Industry Days) で、基調講演、招待講演と一般の講演があった。

アマゾンの Carmel ら [19] からはアレクサショッピングの商品に関する QA (質問応答) の研究とその課題等の紹介があった。研究では商品に関する主観的な質問にフォーカスしているとし、例えば、贈り物のアイデアへの意見や具体的な商品のアドバイス、商品の

<sup>33</sup>Research Assistant Professor.

<sup>34</sup>Data Science for Social Good Summer Fellowship

<sup>35</sup>社会に対して良いインパクトを与えようとする活動やサービス、製品などのこと。

<sup>36</sup>Gerald Salton 賞は IR の研究に継続的に顕著な貢献を行ってきた人に対して 3 年毎に贈られる賞。

<sup>37</sup><http://sigir.org/sigir2018/program/tutorials/>

<sup>38</sup>ACM Conference on Recommender Systems

<sup>39</sup><https://sigir.org/sigir2018/program/industry-days/>

表 4: Tutorial(SIGIR2018)

Tutorial	説明
(a) Knowledge Extraction and Inference from Text:	知識抽出・推論
(b) Probabilistic Topic Models for Text Data Retrieval and Analysis	テキストデータ検索・分析向け確率的トピックモデル
(c) SIGIR 2018 Tutorial on Health Search (HS2018)	医療・健康に関する検索
(d) Conducting Laboratory Experiments Properly with Statistical Tools	統計的ツールを用いる (実験室) 実験の方法
(e) Information Discovery in E-commerce	E コマースにおける情報発見
(f) Deep Learning for Matching in Search and Recommendation	深層学習の検索と推薦におけるマッチングへの応用
(g) Generative Adversarial Nets for Information Retrieval	敵対的生成ネットワーク (GAN) を用いた IR
(h) Tutorial on Utilizing Knowledge Graphs for Text-centric IR	Knowledge Graph を用いたテキスト情報検索
(i) Neural Approaches to Conversational AI	ニューラルネットを用いた対話型 AI
(j) Efficient Query Processing Infrastructures	効率的なクエリ処理の基盤
(k) Fusion in Information Retrieval	IR におけるフュージョン

使い方, 特定商品の比較や商品群の中でのおすすめへのコメント等を例に挙げた. 回答にはカタログや web データの他, レビューや QA などにも利用している. 商品検索の研究課題として音声インターフェイスのフレンドリーな回答生成, 分散した異種の情報からの回答生成, 個人適応等の他, 特に主観的な質問に関する課題としては複数の観点からの回答 (multi-aspect answer), 質問の評価を挙げた.

マイクロソフトの Agrawal ら [20] はインド市場向けのチャットボット, ルー (Ruuh) の開発経験を紹介した. インド市場は若年層が多く, ルーは全ての若者の BFF(永遠の大親友) となることを目指しており, 単に適合した応答を返すだけでは不十分である. 検索ベースのチャットボットで対話と応答のペアをインデクシングして会話に再利用しているが, twitter 等の生データのクレンジングでは攻撃的な発言や個人を特定できる情報を除くほか, 実際にはできないことを言う表現 (例: 電話するね, 等) も削除している. ユーザのシェアした画像や動画へのコメントを求められることも多く, 感情や意見の表出が期待されている. また, インドでは言語が多様で会話の 3-5% でコードミクシング (複数言語の混ざった発話) があり, ボットがコードミクシングで応答すると利用可能性は 30% にまで上がるという.

リンクトインの Geyik ら [21] からは優秀な人材を検索・推薦するシステム LinkedIn Recruiter の紹介があった.

### 3.9 ワークショップ

今回は表 5 に示す 10 のワークショップがあった<sup>40</sup>. 全日のワークショップは検索への対話的アプローチ, knowledge graph の利用, e コマースと前回からの継続が多い. 新しいワークショップとしては特定分野の専門家による複雑な検索の他, ITS に関するもの, 驚きの定量化, 理由を説明可能な推薦と検索等があった.

e コマースのワークショップでは商品タイトルから商品のカテゴリを予測する機械学習のデータチャレンジ

(Rakuten Data Challenge<sup>41</sup>) の結果が報告された. カテゴリは ID の木構造 (leaf は 3,000 強) である. 商品タイトルとカテゴリ ID のパスのペア (0.8M) で学習して, テスト用の商品タイトル (0.2M) のカテゴリを推定するタスクで, 末端まで推定できた場合のみ正解とする. 1 位の Skinner[22] の手法は character ベースの LSTM と F1 最適化を組み合わせたものであった.

### 3.10 SIGIR2018 のトレンド

今回のトレンドを以下にまとめる.

- ニューラルネットに関する論文が大幅に増加 (フルペーパーの約 46.0%, テクニカルセッションの 73.1% に分布)
- 投稿数, 採択数とも中国からの論文が増加
- トラック別投稿数はコンテンツの推薦が多く, 前回に続き, 推薦に関する論文の投稿が多い
- トラック別採択数は分野特有の応用と Human Factors が多く, 今回もユーザよりの論文が多い
- 新しいキーワードはソーシャルグッド (social good)
- 新しいワークショップは分野の専門家の複雑な検索, ITS 関連, 驚きの定量化, 理由を説明可能な推薦と検索等
- 対話やセマンティック・ウェブのセッションは定着

### 3.11 テクニカルセッション

テクニカルセッションのうち, 一部の論文を紹介する.

#### 3.11.1 受賞論文

Best paper はフェアネスとロバストネスのセッションの Canameres ら [23] で, 推薦システムにおける人気 (popularity) の扱いに関するものである. 近年, 推薦システムの評価やアルゴリズムでは人気重視され

<sup>40</sup><http://sigir.org/sigir2018/program/workshops/>

<sup>41</sup><https://sigir-ecom.github.io/data-task.html>

表 5: Workshop (SIGIR2018)

Workshop	回	区分	トピック
CAIR ' 18	2	全日	検索への対話的アプローチ. マルチターンの複雑な対話の支援, マルチモーダルインターフェイス.
ECOM ' 18	2	全日	e コマース (商品検索と検索).
ProfS2018	1	全日	特定分野の専門家による複雑な検索.
KG4IR ' 18	2	全日	検索での knowledge graph 利用.
BIRNDL ' 18	3	半日	自然言語処理, IR, サイエントメトリクス, テキストマイニング, 推薦による学術文献の理解/分析/検索.
CompS ' 18	1	半日	Computational surprise(驚きの定量化, どのようにして IR に取り込むか, 等).
DATA:SEARCH ' 18	1	半日	構造化/半構造化されたデータの検索.
EARS ' 18	1	半日	理由を説明可能な推薦と検索.
Intelligent Transportation Informatics	1	半日	道路交通に関する異種情報の獲得, 統合, 分析.
Learning from Limited/Noisy data for IR	1	半日	IR のための限られたデータまたはノイズの多いデータからの学習.

る一方, バイアスとして取り除くべきとする研究もあり, 答えが出ていない. そこで, 推薦に人気を用いる効果, オフラインの実験で効果を正確に測定できているかなどを調べ, 結論としては人気に基づく推薦が良いかどうかは一概には決められないことを示した.

### 3.11.2 ログ分析

Gathright ら [24] は音楽推薦システムの精度向上のため, ユーザの満足度に関する詳細な調査と分析を実施し, 満足度の予測を行った. Spotify の音楽推薦を 10 週間利用したユーザを対象に利用目的やユーザ体験に関するインタビューを行い, さらにユーザ約 18,500 人を対象に利用目的に加えて週毎と全体的な満足度のサーベイを実施した. これらの分析の結果, 音楽推薦システムの 4 つの利用目的を抽出し, 利用目的によってユーザ行動が異なること, ポジティブなユーザ体験が具体的にどのような行動に現れるか等を明らかにし, 満足度の予測モデルを構築した. さらにユーザの評価におけるバイアスやユーザが満足するための条件, 満足度の推定に役立つもの等の知見も得た.

### 3.11.3 行動

Lu ら [25] はモバイル環境でニュース記事を読む際のユーザの preference<sup>42</sup>の変化や, その変化に影響を与えるものをユーザスタディをふまえて分析した. ユーザの preference は記事を読む前後とタスクの後の 3 段階を区別し, タスクが終わった後の preference を真の preference としている. 得られた知見の例としては, クリックはユーザの preference の指標としては不十分であること, 記事への preference は記事を読む前後, タスク後の 3 段階で変化すること, preference の変化に影響を与えるのは記事の品質, ユーザインタラクションのコンテキストで, 例えば直前に読んだ記事の真の preference が低いと次の記事を過度に高く評価する傾向がある, 等である. さらにユーザの行動, 記事の品

<sup>42</sup>本論文では 'like' を 'preference' と呼んでいる.

質, ユーザのインタラクションのコンテキストも用いてユーザがクリックした記事の真の preference を予測するモデルを構築し, 従来のクリックベースの指標を予測した preference で置き換えることでユーザの満足度を改善できることを示した.

## 4 SIGIR2019

次回, SIGIR2019 は第 42 回で, 2019 年 7 月 21 日から 25 日までパリ (会場: シテ科学産業博物館) で ARIA<sup>43</sup>の後援を受けて開催される. すでにフルペーパーの CFP, 会場, オーガナイザの情報は公式ページ<sup>44</sup>に掲載されており, 今後の更新情報については, 同ページに加えて公式 Twitter<sup>45</sup>を参照されたい.

## 5 国内のコミュニティの動向

日本で開催した SIGIR2017 の成功を受けて, 昨年 12 月, 主催組織の ACM SIGIR の初めての支部である東京支部<sup>46</sup>が誕生した. SIGIR 関連のセミナー等のイベントや会員向けの情報発信を行い, 国内外の研究コミュニティの交流を促進し, IR 研究における日本のプレゼンスの向上を目指す. 2018 年 10 月には今回の SIGIR2018 を含む最近の主要な国際会議の論文 (約 27 本) の読み会 (IR Reading2018 秋)<sup>47</sup>を開催した.

## 6 まとめ

SIGIR は IR 分野の最高峰の国際会議で, 今回の SIGIR2018 は米国で開催され, 約 740 名が参加した. フルペーパーの採択率は約 21 % (86/409) で, 国別に見ると投稿数, 採択数とも中国の躍進が目立った. トレ

<sup>43</sup>French Association for Information Retrieval and Applications

<sup>44</sup><http://sigir.org/sigir2019/>

<sup>45</sup><https://twitter.com/sigir2019>

<sup>46</sup>Tokyo ACM SIGIR Chapter, <http://sigir.jp/>

<sup>47</sup>[http://sigir.jp/post/irreading\\_2018fall/](http://sigir.jp/post/irreading_2018fall/)

ンドはニューラルネット (フルペーパーの約 46.0%, テクニカルセッションの約 73.1%に分布), トラック別投稿数ではコンテンツ推薦が多く, 前回に続き推薦に関する投稿が多い. トラック別採択数では分野に特化した応用と Human Factors に関する論文が多く, ユーザよりの論文が多い傾向が続いている. その他のキーワードとしてはソーシャルグッドがあった.

次回 SIGIR2019 は 2019 年 7 月にパリで開催される. 国内のコミュニティの動向としては, 2017 年 12 月に ACM SIGIR 初の地方支部である東京支部が誕生した. 今後の SIGIR において, 日本のプレゼンスがさらに向上することを願っている.

## 参考文献

- [1] Toshihiro Kamishima: ML, DM, and AI Conference Map, <http://www.kamishima.net/archive/MLDMAImap.pdf>. (2018)
- [2] Microsoft Academic: SIGIR Conference Analytics, <https://www.microsoft.com/en-us/research/project/academic/articles/sigir-conference-analytics/>. (2018)
- [3] SIGIR: List of SIGIR Locations and History, <http://sigir.org/general-information/history/>.
- [4] 櫻 惇志: SIGIR2017 参加 (開催) 報告-会議概要とニューラルネットな情報検索-, 第 11 回テキストアナリティクス・シンポジウム, <http://www.ieice.org/~nlc/attachment/MLC20170908-SIGIR-2017-Keyaki.pdf>. (2017)
- [5] The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, <https://dl.acm.org/citation.cfm?id=3209978>. (2018)
- [6] Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, <https://dl.acm.org/citation.cfm?id=3077136>. (2017)
- [7] Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, <https://dl.acm.org/citation.cfm?id=2911451>. (2016)
- [8] Arjen de Vries, Hang Li, Ryen White: ACM SIGIR 2017 - Opening - PC Chairs, <https://www.slideshare.net/arjenpdevries/acm-sigir-2017-opening-pc-chairs>. (2017)
- [9] Kazutoshi Umemoto et al.: Search by Screenshots for Universal Article Clipping in Mobile Apps, ACM Transactions on Information Systems (TOIS) - Special issue: Search, Mining and their Applications on Mobile Devices, Vol.35, Issue.4, pp.34:1-2. (2017)
- [10] SIGIR: TOIS Presentation, <http://sigir.org/conferences/tois-presentation/>. (2017)
- [11] ACM SIGIR 2018 Conference Organization: ACM SIGIR 2018 Accepted Papers, <http://sigir.org/sigir2018/accepted-papers/>. (2018)
- [12] SIGIR2017, <http://sigir.org/sigir2017/>. (2017)
- [13] SIGIR2016, <http://sigir.org/sigir2016/>. (2016)
- [14] Christopher Manning: Natural Language Inference, Reading Comprehension and Deep Learning, <https://nlp.stanford.edu/manning/talks/SIGIR2016-Deep-Learning-NLI.pdf>, p.72. (2016)
- [15] Rayid Ghani: Data Science for Social Good and Public Policy: Examples, Opportunities, and Challenges, Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, p.3. (2018)
- [16] 酒井哲也: 情報アクセス評価方法論, コロナ社. (2015)
- [17] Kalervo P. Jarvelin: Salton Award Keynote: Information Interaction in Context, Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp.1-2. (2018)
- [18] Jun Xu, Xiangnan He, and Hang Li: Deep Learning for Matching in Search an Recommendation, <http://comp.nus.edu.sg/~xiangnan/sigir18-deep.pdf> (2018)
- [19] David Carmel, et al.: Product Question Answering Using Customer Generated Content - Research Challenges, Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 1349-1350. (2018)
- [20] Manoj Kumar Chinnakotla and Puneet Agrawal: Lessons from Building a Large-scale Commercial IR-based Chatbot for an Emerging Market, Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 1361-1362. (2018)
- [21] Sahin Cem Geyik, et al.: Talent Search and Recommendation Systems at LinkedIn: Practical Challenges and Lessons Learned, Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 1353-1354. (2018)
- [22] Michael Skinner: Product Categorization with LSTMs and Balanced Pooling Views, [https://sigir-ecom.github.io/ecom18DCPapers/ecom18DC\\_paper\\_9.pdf](https://sigir-ecom.github.io/ecom18DCPapers/ecom18DC_paper_9.pdf). (2018)
- [23] Rocio Canameres and Pablo Castells: Should I Follow the Crowd?: A Probabilistic Analysis of the Effectiveness of Popularity in Recommender Systems, Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 415-424. (2018)
- [24] Jean Garcia-Gathright, et al.: Understanding and Evaluating User Satisfaction with Music Discovery, Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp.55-64. (2018)
- [25] Hongyu Lu, Min Zhang, and Shaoping Ma: Between Clicks and Satisfaction: Study on Multi-Phase User Preferences and Satisfaction for Online News Reading, Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp.435-444. (2018)