レシピに関する時間的特徴を持つ言語表現の抽出法の検討

桐本　宙輝†　　風間　一洋†

† 和歌山大学 システム工学部　〒 640–8510 和歌山県和歌山市栄谷 930
E-mail: †{s171063,kazama}@sys.wakayama-u.ac.jp

あらまし　食事という日常的な行動は人間の生活と密接な関係にあり，季節や年中行事の影響を大きく受けるので，料理の時間的特徴に着目すれば，ユーザが想定している時間的状況下で作られる料理を和食，洋食，菓子などの垣根を越えて探すことが可能になる．しかし，そのような時間的特徴を表すために実際にどのような言語表現が用いられているのか，また時間的特徴を表す言語表現の利用が時期によってどのように変化するかは，必ずしも明らかではない．本稿では，cookpad で公開されている料理のレシピに対して，まずレシピとそのつくれぽのテキストから特徴的な言語表現を抽出し，さらにそれぞれの言語表現が使われているレシピ群のつくれぽ数の頻度変化を統合することで，レシピに関する時間的特徴を示す言語表現とその時間的特徴の抽出を試みる．

キーワード　cookpad, レシピ, つくれぽ, 時系列解析, 周期性

1. はじめに

食事という日常的な行動は人間の生活と密接な関係にあり，季節や年中行事の影響を大きく受けるので，人間がどのようなレシピを調理するかは，時期や目的によって大きく変化する．たとえば，夏には素麺，冬には鍋料理のような季節特有のレシピを，正月にはおせち料理，節分には恵方巻きのような年中行事にちなんだレシピを調理することは，日常的に行なわれていると言える．このようなレシピをレシピ検索サイトで探す場合には，探したい特徴を持つ料理に関する何らかの時間的特徴を手掛かりにすることが多いと考えられる．

たとえば，cookpad の検索履歴分析サービス「たべみる」(注1) を用いれば，ユーザがレシピを探す時に想定している時間的状況をある程度知ることができる．ただし，クエリは複数のレシピを探索するためのきっかけにすぎず，通常はクエリに適合する多数のレシピの中からユーザ自身の好みに合うレシピを探索的に見つけることが多いことから，クエリは必ずしも具体的であるとは限らず，また最終的に調理したレシピの内容を的確に示すとも限らない．

そこで，レシピとそれを実際に調理した感想であるつくれぽのテキスト中に現れる，時間的特徴を示す言語表現に着目する．この言語表現とは単一の意味のまとまりであり，単一の単語に加えて，複合語やフレーズを意図している．すでに我々はレシピに対するつくれぽの投稿履歴からレシピ固有の時間的特徴を抽出する手法を提案した [1], [2] が，さらにどのような時間的特徴を示すレシピで各言語表現が用いられているかを分析することで，その言語表現が持つ時間的特徴を求めることができる．このような言語表現の持つ時間的特徴が明らかになれば，ユーザのクエリから想定される時間的状況に合致したレシピの提示や，類似した時間的特徴を示す言語表現に基づいたレシピ群の分類のような，レシピの使用食材や調理手順からは分からないような時間的特徴や時間的共起性に基づいた処理が可能となる．

そこで，本稿では，言語表現が持つ時間的特徴を，検索ログではなく実際に調理を行ったレシピとそのつくれぽから求める手法を提案する．具体的には，cookpad で公開されている料理のレシピに対して，まずレシピとそのつくれぽのテキストから特徴的な表現を抽出し，さらにそれぞれの表現が使われているレシピ群のつくれぽ数の頻度変化を統合することで，レシピに関する時間的特徴を示す表現と，その時間的特徴の抽出を試みる．さらに，抽出した言語表現を時間的特徴に基づいてクラスタリングすることで，年中行事や季節と関連のあるような時間的に特有な性質を示す言語表現を抽出する．また，言語表現が持つ時間的特徴を妥当に抽出できているのかを分析することで，手法の有用性を評価する．

2. 関連研究

2.1 レシピの特徴語抽出

金内らは，投稿型レシピサイトにおけるレビュー情報から，料理タイトルを自動生成する研究を行った [3]．料理が苦手であったり投稿型レシピサイトに不慣れなユーザは，レシピを投稿する際にそのレシピの特徴を的確に表現したタイトルを付けられるとは限らない．そこで，レシピのレビュー情報として cookpad のつくれぽを用いて，レシピの「対象者」,「作った意図」,「欲求」,「好み」,「イベント性」の 5 つの特徴・目的を示す単語を抽出し，それをもとに料理タイトルを自動生成するための手法を提案した．

本研究はレシピに現れる単語が持つ時間的な特徴を推定することが目的であるため，金内らの手法とは抽出した単語の利用方法が異なる．

2.2 レシピの推薦・検索

cookpad のようなレシピ投稿サイトでは日々新たなレシピが投稿されるため，収録レシピ数が膨大で，類似レシピも大量に存在する．そのため，ユーザが自身のニーズに合致したレシピを探すには，気に入るレシピが見つかるまでレシピを探索する

(注1)：https://info.tabemiru.com

必要があり，手間がかかる．この問題を解決するために，新たなレシピの検索・推薦方法が提案されている．

上田らは，食材の利用履歴から個人の嗜好を抽出することで，ユーザの好みを反映したレシピを推薦する手法を提案した [4]．また，上田らは，レシピの閲覧・摂食履歴から個人の嗜好を抽出し，レシピの推薦に用いる手法を提案した [5]．

平川らは，レシピに添付された料理画像の色情報に基づいて，目的に応じたレシピの選択支援を行う手法を提案した [6]．平川らの手法では，料理の完成写真はユーザが調理するレシピを決定する重要な要因であると考え，レシピテキストの修飾語と料理画像の色情報を用いてレシピを推薦する．

門脇らは，ブログ型レシピの文章からレシピの誕生・使用事由を，食材や調理器具，時間の有無や，調理時の季節や気候，調理者の気分や体調などのカテゴリに分けて抽出することで，レシピの使用事由に基づいた検索を行う手法を提案した [7]．

上田らと平川らによる研究は，新しいレシピの検索・推薦手法によって，ユーザの目的に合致したレシピを簡単に検索できるようにするという目的は本研究と同じであるが，アプローチの仕方がまったく異なる．門脇らの手法は，調理時の季節による検索を考慮している点で本研究と類似しているが，門脇らがレシピテキストから季節性を抽出するのに対して，本研究ではつくれぽの投稿履歴を用いており，実際に多くの人々がそのレシピを作ったという事実から時間的特徴を抽出している点が異なる．

3. 時間的な特徴を持つ言語表現の抽出

3.1 本研究のアプローチ

文字列の時間的特徴の分析には，通常は単語や文章の生成時刻が特定できる新聞記事や Twitter のツイートのような時系列テキストや，検索システムのクエリログなどが用いられることが多い．ただし，一度作られた文章が，変更されずに長い期間参照され続けるような利用形態を持つレシピの場合には，単語や文章の生成時間に着目した分析はできない．

そこで，本稿ではテキストの利用時刻に着目する．cookpad では，公開されているレシピに対して調理した感想を投稿することでレシピ投稿者への感謝を示す「つくれぽ」と呼ばれる機能を提供している．そこで，つくれぽが書かれたレシピの文章がその時刻に利用されたと仮定すれば，ある言語表現が出現するレシピと，そのレシピに対して書かれたつくれぽの関係から，その言語表現の利用に関する時間的特徴を推定できると考えられる．

また，通常つくれぽの文章は短く簡潔なことから，単体では言語表現の時間的特徴の分析には不十分だが，元のレシピの主な特徴について言及したり，内容を短く要約する効果があると推測される．さらに，複数のユーザによってつくれぽが書かれることを考慮すると，元のレシピに関する特徴を的確に示す言語表現だけが，複数のユーザによって繰り返し使われる強調効果もあることから，文字列のつくれぽにおける出現時刻も合わせて考慮することで，言語表現の時間的特徴をより正確に推定できるようになると思われる．

ただし，ここで，あるレシピに対してつくれぽが書かれたとしても，そのレシピに出現するすべての言語表現が利用されたと単純に考えてはいけないことに注意しなければならない．実際に，同じ単語でもあっても，様々な文脈で用いられることがある．たとえば,「春が終わって暑くなってきた時期にぴったりのレシピです」という文章では，このレシピが作られる時期は初夏であるが,「春」は比較のために用いられている．つまり，文脈によっては，言語表現は比較や否定のために用いられることがある．これらの違いを正確に判定することは困難であることから，レシピのテキストにおいて言語表現が繰り返し使われるなどの重要性が高い場合に限りつくれぽとの関連性が高いと見なして，時間的特徴の推定に使用する．

3.2 提案手法の概要

本稿では，レシピに関する時間的特徴を持つ言語表現の抽出法を提案する．その概要を図 1 に，処理手順の概略を以下に示す．

（1） レシピとつくれぽのテキストから，TF-IDF 値に基づいて重要と判断される言語表現を抽出し，それらの言語表現と出現するレシピ群の関係を示す言語表現・レシピ対応表を作成する．

（2） つくれぽ利用履歴から，各レシピの時間的な性質を表す頻度ベクトルを作成する．

（3） 言語表現・レシピ対応表に基づいて，ある言語表現に関係するレシピの頻度ベクトルの総和を計算し，その言語表現の頻度ベクトルを作成する．

（4） 言語表現の頻度ベクトルを 1 年周期で折りたたみ，正規化して，言語表現の持つ時間的な性質を表す特徴ベクトルを作成する．

3.3 レシピとつくれぽからの言語表現と時間的特徴の抽出

単語を抽出する対象は，レシピのタイトル，概要，生い立ち，コツ・ポイント，そのレシピに対するつくれぽのテキストとする．なお，抽出する単語は名詞に限定する．

単語の抽出は以下の手順で行う．

（1） 対象テキストを MeCab(注2) で日本語形態素解析して，単語に分割する．

（2） ストップワードを除去する．たとえば，1 文字の記号類などが相当する．

（3） 単語を複合語化し，一つの意味のまとまりである言語表現を作成する．現時点では，専門用語自動抽出システムである TermExtract(注3) を用いて，単語を複合語化した結果を言語表現として扱っている．

（4） TF-IDF 値が高い上位 N 語を，レシピの特徴的な言語表現として抽出する．

（5） 言語表現と，それが出現するレシピ群の関係を示す言語表現・レシピ対応表を作成する．

なお，レシピとそのつくれぽのテキストに出現する語彙数の中央値が 270 語であったため，TF-IDF 値が高い上位 1 割の

(注2)：http://taku910.github.io/mecab/
(注3)：http://gensen.dl.itc.u-tokyo.ac.jp

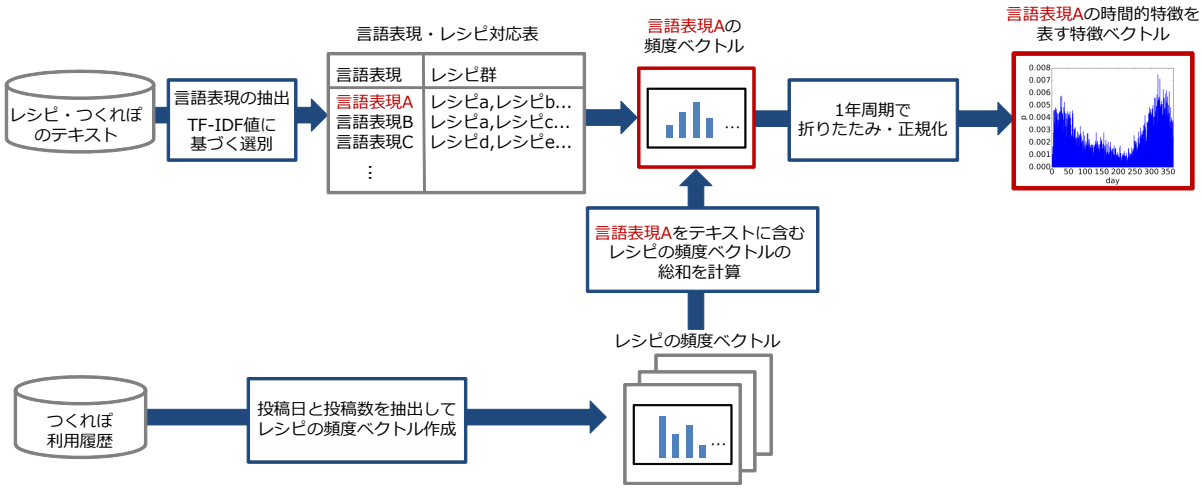


図 1　提案手法の概要

$N = 27$ とした．

3.4　レシピの頻度ベクトルの作成

つくれぽの利用履歴から，各レシピの時間的な性質を表す頻度ベクトルを作成する．

まず，レシピ r_i の最初のつくれぽの投稿から最後のつくれぽの投稿までの期間が T_{r_i} の場合に，時刻 $t(0 \leqq t \leqq T_{r_i} - 1)$ のつくれぽ数を $f_{r_i}^{(t)}$ として，次元数 T_{r_i} のつくれぽ数の頻度ベクトル $\boldsymbol{f}_{r_i}$ を求める．

$$\boldsymbol{f}_{r_i} = (f_{r_i}^{(0)}, f_{r_i}^{(1)}, \ldots, f_{r_i}^{(T_{r_i}-1)}) \tag{1}$$

頻度ベクトルの各要素は，その日におけるつくれぽの投稿数を表す．つくれぽの投稿は，年中行事に関連するレシピであれば行事開催日の前後数日に，特定の季節性を持つレシピであればその季節に活発になる．頻度ベクトルはそのようなレシピの時間的な性質を反映していると言える．

3.5　言語表現の特徴ベクトル作成

抽出した言語表現の集合を W とした時，W に含まれる言語表現 w_i の頻度ベクトル $\boldsymbol{f}_{w_i}$ を以下の式で求める．

$$\boldsymbol{f}_{w_i} = \sum_{r_j \in R} correct(\boldsymbol{f}_{r_j}) \tag{2}$$

ここで，集合 R は言語表現 w_i をテキストに含んだレシピの集合である．また，レシピごとにつくれぽの投稿期間が異なるため，頻度ベクトル $\boldsymbol{f}_{r_j}$ の次元数も不揃いで，そのままでは総和が計算できない．そのため，異なる投稿期間を補正する $correct(\boldsymbol{f}_r)$ 関数を導入して計算する．

次に，$\boldsymbol{f}_{w_i}$ を次元数 C に折りたたんだベクトル $\boldsymbol{v}_{w_i,C}$ を，以下のように作成する．

$$\begin{cases} \boldsymbol{v}_{w_i,C} & = & (v_{w_i,C}^{(0)}, v_{w_i,C}^{(1)}, \ldots, v_{w_i,C}^{(C-1)}) \\ v_{w_i,C}^{(j)} & = & \frac{sum(\{f_{w_i}^{(t)} | fold(t,C)=j\})}{|\{f_{w_i}^{(t)} | fold(t,C)=j\}|} \end{cases} \tag{3}$$

ここで，$fold(t, C)$ は，$\boldsymbol{f}_{w_i}$ の時刻 t を $\boldsymbol{v}_{w_i}$ の $0 \sim C-1$ の範囲のインデックス値に折りたたむ関数である．本稿では閏日も含めた 1 年周期（$C = 366$）とし，その年で何日目かを示す値 $1 \sim 366$ をそれぞれ $0 \sim 365$ に変換する．$sum(S)$ は集合 S の総和を求める関数である．

最後に，$\boldsymbol{v}_{w_i,C}$ を総和が 1 になるように正規化し，言語表現 w_i の周期 C の特徴ベクトル $\boldsymbol{p}_{w_i,C}$ を求める．$\boldsymbol{p}_{w_i,C}$ は，つくれぽの周期 C の各要素の時点の生起確率である．

$$\begin{cases} \boldsymbol{p}_{w_i,C} & = & (p_{w_i,C}^{(0)}, p_{w_i,C}^{(1)}, \ldots, p_{w_i,C}^{(C-1)}) \\ p_{w_i,C}^{(j)} & = & \frac{v_{w_i,C}^{(j)}}{\sum_{k=0}^{C-1} v_{w_i,C}^{(k)}} \end{cases} \tag{4}$$

4.　時間的特徴による言語表現のクラスタリング

提案手法によりレシピやつくれぽのテキストから抽出した言語表現は，年中行事に関連するものや季節との関連があるもの，逆に時期による変化がない定常的なものなど，様々な種類の言語表現が混在していると考えられる．そこで，言語表現の時間的な性質を表す特徴ベクトルを用いてクラスタリングすることで，言語表現を種類別に分類する．

また，提案手法で抽出された言語表現の特徴ベクトルに，言語表現が使用される文脈の違いによりノイズが混入する可能性については既に述べた．そこで，分類結果から，抽出された特徴ベクトルの妥当性も同時に検証する．

クラスタリング手法としては階層型クラスタリングの一種であるウォード法を，言語表現間の距離としてはユークリッド距離を用いる．なお，年中行事や季節と関連する言語表現の場合には，特徴ベクトルの分布形状が同じでも，ピークとなる時点は言語表現により大きく異なることから，特徴ベクトルの各要素を降順にソートしてからクラスタリングすることで，そのようなピーク時期によらない，分布の偏りに基づく分類を可能にする．

5.　評　　価

5.1　クックパッドデータセット

Cookpad から提供された 1998 年 4 月 23 日から 2014 年 9 月 30 日までの 1,715,595 件のレシピ [8] から，つくれぽ数が 100 件以上あり，つくれぽ投稿期間が 366 日以上の 11,507 件のレシピを抽出して，評価に用いた．なお，レシピの ID，タイトル，概要，生い立ち，コツ・ポイント，つくれぽ投稿日，つく

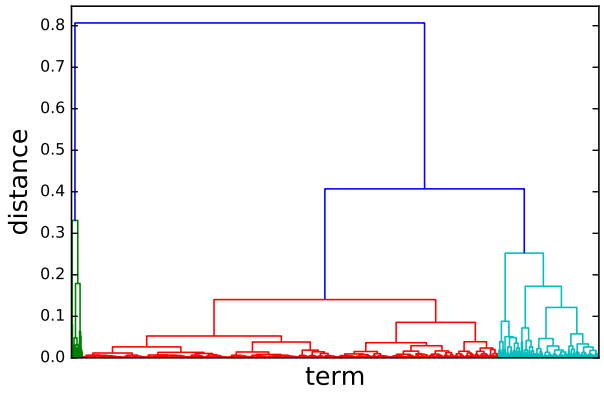


図 2　クラスタリング結果（デンドログラム）

れぽ投稿数，つくれぽのテキストを使用する．

5.2　言語表現のクラスタリング結果

データセットとして用いた 11,507 件のレシピとそのつくれぽのテキストから言語表現を抽出した結果，518,476 語が得られた．その中から各レシピの TF-IDF 値が高い上位 27 語をそのレシピの特徴付ける言語表現として，30 件以上のレシピで出現する 1,246 語を抽出した．

これらの語をクラスタリングした結果，図 2 に示すデンドログラムが得られた．ここで，横軸の各要素は言語表現，縦軸はクラスタ間の距離を表す．デンドログラムの形状から，クラスタ数が 3 個より少なくなると，クラスタ間の距離が大きく離れたクラスタ同士を統合してしまうことが分かる．そのため，クラスタ数が 3 個になるようにデンドログラムを切断し，各クラスタを図 2 でそれぞれ緑色，赤色，水色に着色した．

緑色で示したクラスタに属する言語表現は 26 語で，そのほとんどが年中行事に関する語であった．たとえば「バレンタイン」や「お正月」,「ひな祭り」などの言語表現がこのクラスタに属していた．

水色で示したクラスタに属する言語表現は 237 語で，季節に関連する語と，年中行事に関連する語で構成されていた.「春」,「夏」,「秋」,「冬」のように直接的に季節性を示すものや,「苺ジャム」や「冷やし中華」のような旬の食材・料理などが確認できた．また,「冷製」や「花見」のような季節特有の語もみられた．一方で,「クリスマス」という言語表現がこのクラスタに属しており，年中行事に関連のある言語表現の一部は，季節性のある言語表現と綺麗に分離できているとは言えない結果となった．

赤色で示したクラスタに属する言語表現は 983 語で，特定の年中行事や季節との関連がない定常的な語で構成されていた.「醤油」や「塩」のような常用する調味料の他に,「ごはん」や「定番」などが確認できた．

以上の結果から，おおまかに (1) 年中行事に関連した言語表現，(2) 季節に関連した言語表現，(3) 年中行事や季節と関連のない定常的な言語表現の 3 種類にクラスタリングできたと言える．しかし，年中行事に関連する言語表現の一部は，季節に関連した言語表現との分離性が悪かった．

5.3　言語表現の時間的な性質を表す特徴ベクトル

抽出された言語表現の特徴ベクトルの形状を見て，言語表現が持つ時間的特徴を妥当に表現できてるかを評価する．

図 2 において緑色のクラスタに属する，年中行事に関連のある言語表現の特徴ベクトルを図 3 に示す．なお，特徴ベクトルのグラフの横軸は 1 月 1 日からの経過日数，縦軸はその日におけるつくれぽの生起確率を表す．特徴ベクトルの形状から,「友チョコ」はバレンタインデーに,「ハロウィン」はハロウィンの日に鋭いピークが発生していることが分かる．また,「初節句」はひな祭りの日と，こどもの日の両方に鋭いピークが確認でき，年中行事の時間的な特徴を適切に表現できていると言える．

次に，水色のクラスタに属する，季節性のある言語表現の特徴ベクトルを図 4 に示す．特徴ベクトルの形状から,「夏バテ」は夏に,「秋刀魚」は秋に盛り上がりがあることが確認できる．また,「遠足弁当」は幼稚園や小学校などで遠足が行われることが多い春・秋に盛り上がりがあり，季節特有の時間的特徴を表現できていると言える．

次に，赤色のクラスタに属する，季節性を示さない定常的な言語表現の特徴ベクトルを図 5 に示す．特徴ベクトルの形状から,「冷凍」や「常備菜」,「弁当」のような特定の季節との関連がない定常的な言語表現はピークがなく，平らな形状になっていることが分かる．これらは，冷凍食品や常備菜のような作りおきをする料理や，お弁当のような 1 年中需要がある時間的特徴を表現できていると言える．

以上の結果から，年中行事に関連する言語表現，季節性のある言語表現，定常的な言語表現の 3 種類において，時間的な特徴を抽出することができたと言える．

5.4　TF-IDF 値による言語表現の選別の有用性

今回は 1 つのレシピあたり TF-IDF 値が高い 27 語をそのレシピを特徴付ける言語表現として採用した．言語表現の選別がどの程度特徴ベクトルの形状に影響するかを分析する．

「夏」という言語表現について，TF-IDF 値による言語表現の選別を行ってから作成した特徴ベクトルと，選別を行わずに作成した特徴ベクトルを図 6 に示す．

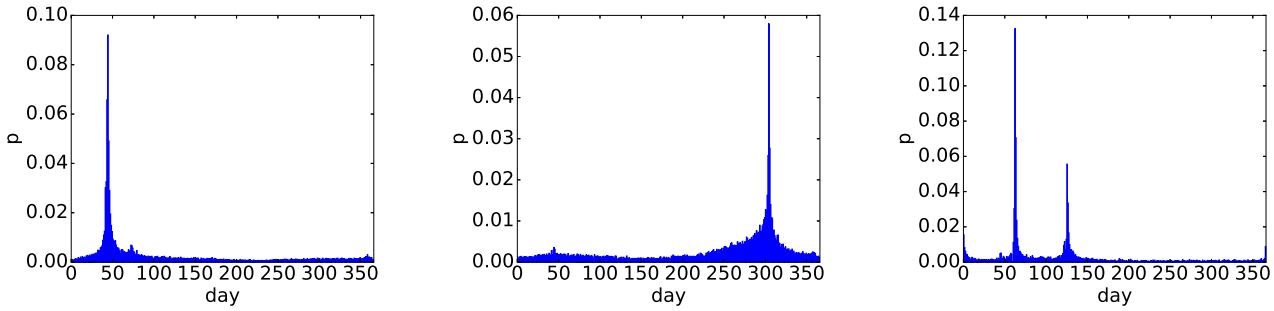
図 6(a) は一般的に夏と言われる 6 月上旬〜8 月下旬あたりにピークが現れているのに対して，図 6(b) では夏にピークが現れていない上に，バレンタインデーの日に局所的なピークが現れている．これは,「夏」という言語表現は多くのレシピやつくれぽのテキストに含まれるため，TF-IDF 値による選別を行わなかった場合，言語表現の頻度ベクトルを作成する際に実際には夏にほとんど作られないノイズとなるレシピの頻度ベクトルも足し合わせてしまうことが原因である．

レシピの持つ季節性に着目すると，以下の 3 つの集合が考えられる．

- U：全レシピの集合
- A：季節性を持つレシピの集合
- B：テキスト中に季節性が明文化されているレシピの集合

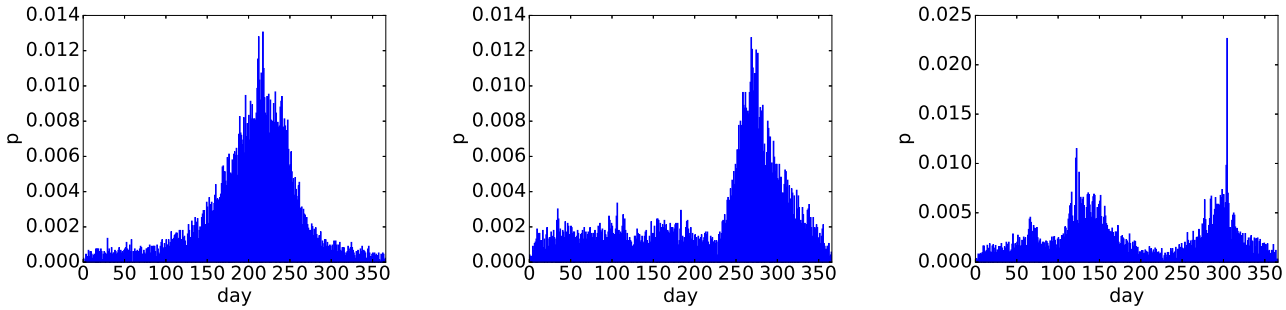
これらの集合の関係を，ベン図として図 7 に示す．

図 7 に示したベン図は，4 つの領域に分けることができる．



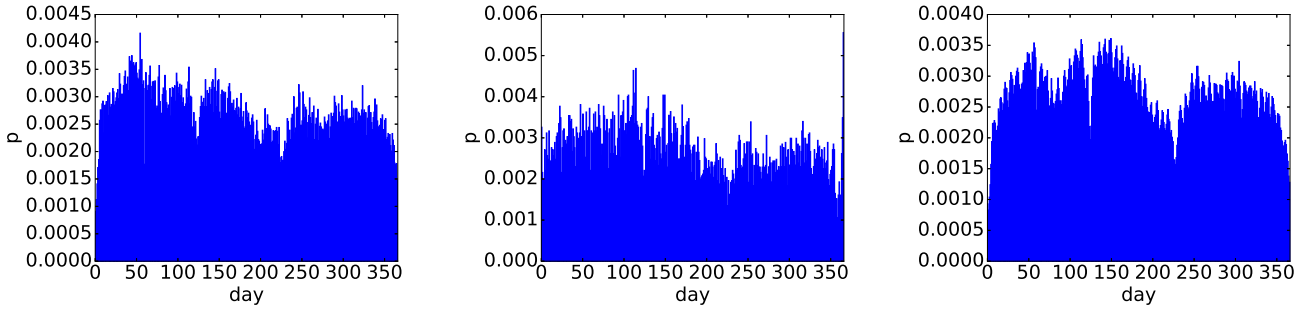
(a) 友チョコ　(b) ハロウィン　(c) 初節句

図 3　年中行事に関連する言語表現の特徴ベクトル



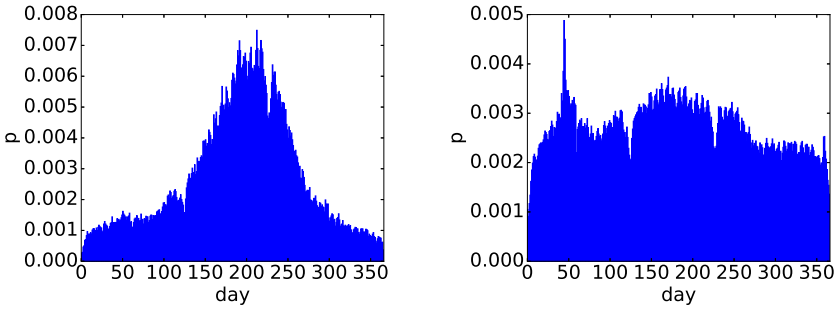
(a) 夏バテ　(b) 秋刀魚　(c) 遠足弁当

図 4　季節性のある言語表現の特徴ベクトル



(a) 冷凍　(b) 常備菜　(c) 弁当

図 5　定常的な言語表現の特徴ベクトル



(a) 選別を行った場合　(b) 選別を行わなかった場合

図 6　選別の有無による「夏」の特徴ベクトルの比較

- $A \wedge \overline{B}$：レシピが季節性を持っているが，その季節性は明文化されていないため，言語表現の特徴ベクトル作成に利用できない．
- $A \wedge B$：レシピが季節性を持っており，その季節性は明文

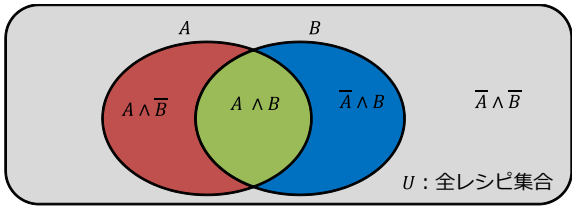


図 7　レシピの季節性とテキストの関係

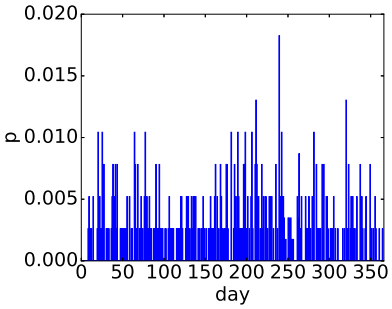


図 8　レシピ「夏にスタミナ満点 !! レバニラ炒め」の特徴ベクトル

化されているため，言語表現の特徴ベクトル作成に利用できる．

- $\overline{A} \wedge B$：レシピは季節性を持っていないが，季節性が明文化されているため，言語表現の特徴ベクトル作成の際にノイズになる．
- $\overline{A} \wedge \overline{B}$：季節性のある言語表現の特徴ベクトルの作成に影響しない．

ここで問題となるのは，$\overline{A} \wedge B$ の領域に属するレシピである．この領域に属するレシピの例として，「夏にスタミナ満点 !! レバニラ炒め」がある．このレシピの特徴ベクトルを図 8 に示す．なお，レシピの特徴ベクトルは，言語表現の頻度ベクトルから特徴ベクトルを作成するのと同様に，レシピの頻度ベクトルに対して 1 年周期の折りたたみと正規化を行うことで作成する．

このレシピは，タイトルに「夏」という言語表現が含まれているにも関わらず，図 8 の形状から分かるように夏に顕著なピークが現れておらず，つくれぽは 1 年中投稿されている．これは，レシピの投稿者と調理者（つくれぽの投稿者）の認識にギャップが存在することが原因である．このレシピのつくれぽ数は 393 件だが，つくれぽのテキストに「夏」が含まれているのは全体の約 2%にあたる 8 件のみであった．$\overline{A} \wedge B$ の領域に属するレシピはこのように，その言語表現の出現頻度が低い傾向にある．そのため，TF-IDF 値による選別はノイズを減らすのに効果的であると言える．

また，言語表現の選別を行った場合，語彙数は 11,068 語から 1,246 語に大幅に減少した．除去されたものの中には，バレンタインの中でも「本命チョコ」や「練習用」,「会社用」のような用途を示す言語表現や,「餅消費」,「夏バテ知らず」のような季節特有かつ直感的に理解しやすい言語表現も含まれていた．これらの言語表現は検索クエリ推薦やカテゴリ分類，内容要約に非常に有用であると考えられる．そのため，有用な言語表現を漏らすことなく，なおかつ特徴ベクトルにノイズが含まれないような手法に改良することが今後の課題として挙げられる．

6. 考　　察

年中行事に関連する言語表現の一部は，季節性のある言語表現の特徴ベクトルとの分離性が悪く，不適切なクラスタリング結果になっていた．たとえば「クリスマス」や「ホワイトデー」などが該当する．

これらの言語表現の特徴ベクトルを，図 9 に示す．図 9(a) から,「クリスマス」の特徴ベクトルはクリスマスの日だけではなくバレンタインデーにもピークが現れていることが分かる．また，値は小さいがひな祭りの日にもピークが出現している．これは，クリスマスに作られるようなお菓子のレシピは，バレンタインデーやひな祭りなどのイベントや祝い事の時にも需要があることから生まれる時間的共起性を示唆している．

さらに，図 3 に示した特徴ベクトルと比較すると，ピーク以外の定常的な成分が多い．提案手法である特定の日にピークを示すような特徴ベクトルの類似度を計算する場合には，ピーク部分よりも期間が長い定常成分の類似性の影響が大きくなることが原因で，季節性のある言語表現の特徴ベクトルとの分離性が悪くなったと考えられる．

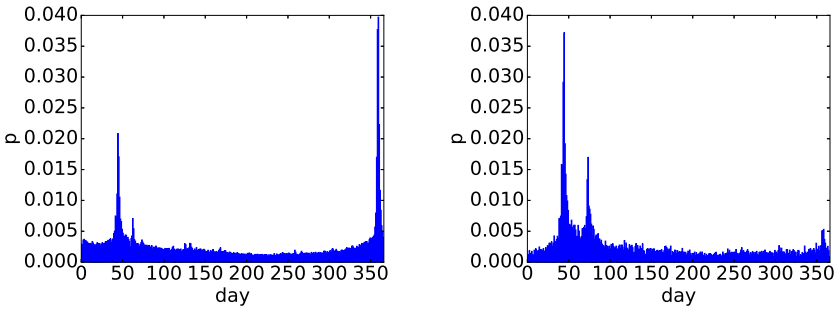
図 9(b) に示した「ホワイトデー」の特徴ベクトルで最も大きいピークが現れているのはバレンタインデーの日で，次にホワイトデー，クリスマスとなっている．この場合も定常的な成分が多く，年中行事の日に限らず日常的に調理されるようなお菓子のレシピの時間的特徴が影響していると考えられる．

ピークが存在するが定常的な成分も薄く持つ言語表現は，クラスタ数が 5 個に分割される部分までデンドログラムの切断箇所を下げれば季節性のある言語表現と別のクラスタに分離された．定常的な成分は TF-IDF による言語表現の選別では完全には除去できないが，ハイパスフィルタでノイズを取り除く処理を加えれば，クラスタ数を増やすことなく高い分離性で「年中行事に関連する言語表現のクラスタ」,「季節性のある言語表現のクラスタ」,「定常的な言語表現のクラスタ」の 3 種類に分けることができると考えられる．

7. おわりに

本稿では，cookpad で公開されているレシピに対してテキストから言語表現を抽出し，各レシピのつくれぽ数の頻度変化を統合することで，言語表現の持つ時間的な特徴を抽出した．さらに，抽出した言語表現を時間的特徴でクラスタリングすることで，年中行事や季節に関連するような，特有な時間的特徴を示す言語表現の抽出を行った．また，それらの特徴ベクトルの形状が，言語表現が持つ時間的な性質を反映していることを検証した．

今後の課題として，語彙数を多く保ちつつ，言語表現の特徴ベクトル作成の際にノイズとなるようなレシピの頻度ベクトルを足し合わせないような手法への改良が挙げられる．また，クラスタの分離性を向上させるために，ハイパスフィルタを用いて作成後の特徴ベクトルから定常成分を除去する処理も取り入れる必要がある．



(a) 「クリスマス」の特徴ベクトル　　(b) 「ホワイトデー」の特徴ベクトル

図 9　不適切なクラスタリング結果になった言語表現の特徴ベクトル

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．

文　　献

[1] 桐本宙輝, 風間一洋. Cookpad のつくれぽ数の時間変動に基づく類似レシピ抽出法の提案. 第 8 回データ工学と情報マネジメントに関するフォーラム, 2016.

[2] 桐本宙輝, 風間一洋. レシピ利用履歴の時間特性に基づいた時間表現によるレシピ検索法の提案. 情報処理学会論文誌データベース（TOD）, Vol. 9, No. 4, pp. 11–16, dec 2016.

[3] 金内萌, 難波英嗣, 角谷和俊. 投稿型レシピサイトにおけるレビュー情報に基づく料理タイトル自動生成. 第 8 回データ工学と情報マネジメントに関するフォーラム, 2016.

[4] 上田真由美, 石原和幸, 平野靖, 梶田将司, 間瀬健二. 食材利用履歴に基づき個人の嗜好を反映するレシピ推薦手法. 日本データベース学会 Letters, Vol. 6, No. 4, pp. 29–32, 2008.

[5] 上田真由美, 高畑麻理, 中島伸介. レシピ閲覧・摂食履歴を用いた嗜好の抽出. Web とデータベースに関するフォーラム (WebDB Forum 2011), 情報処理学会シンポジウムシリーズ, 3G-1-2, 2011.

[6] 平川芽依, 牛尼剛聡, 角谷和俊. 料理画像の色情報に基づく目的に応じたレシピ選択支援. 第 8 回データ工学と情報マネジメントに関するフォーラム, 2016.

[7] 門脇拓也, 山肩洋子, 森信介. 誕生・使用事由によるレシピ検索: 生い立ちレシピサーチ. 日本データベース学会和文論文誌, Vol. 13, No. 1, pp. 78–85, 2014.

[8] Jun Harashima, Michiaki Ariga, Kenta Murata, and Masayuki Ioki. A large-scale recipe and meal data collection as infrastructure for food research. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, pp. 2455–2459, 2016.