

レシピの時間的特徴を用いた レシピ検索法について

風間 一洋 (和歌山大学)

はじめに

- 桐本 宙輝（和歌山大学大学院→日立システムズ）との研究の概要を紹介
 - 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, IDR 2016. ※**優秀賞受賞**
 - 桐本宙輝, 風間一洋: レシピと調理動機の時間的特徴によるレシピ検索法の提案, IDR 2017.
 - その他3件の学会発表, 1件の論文, 2回学会受賞

研究の目的とアプローチ

- 目的

- ユーザの気分や置かれた状況を反映できるような情報検索・推薦を実現したい

- アプローチ

- 調理するレシピ選択には「時間」が影響

- 季節（春, 夏, 秋, 冬）
- 食材の旬
- 年中行事

- 正月, バレンタインデー, ひな祭り, クリスマスなど

- レシピの時間的特徴に着目

- cookpadはユーザ投稿型レシピサービス

- つくれぽ：cookpadのレシピを実際に調理し、写真付きで感想を投稿する仕組み

- つくれぽをユーザのレシピの調理履歴として利用

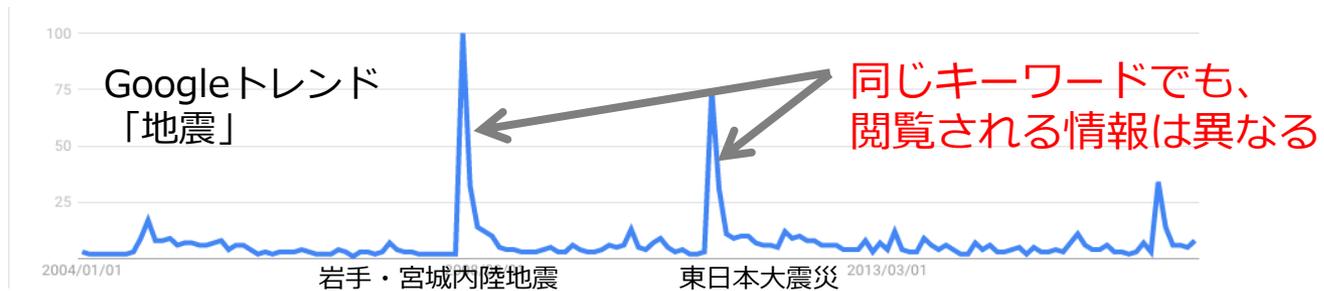
レシピの時間的な性質

Web検索の場合…検索する明確な動機・理由が存在

即時性：実世界のイベントをきっかけに検索される

変動性：情報の陳腐化が早く，内容が移り変わる

例：新製品の発表，突発的な災害・事件，etc...

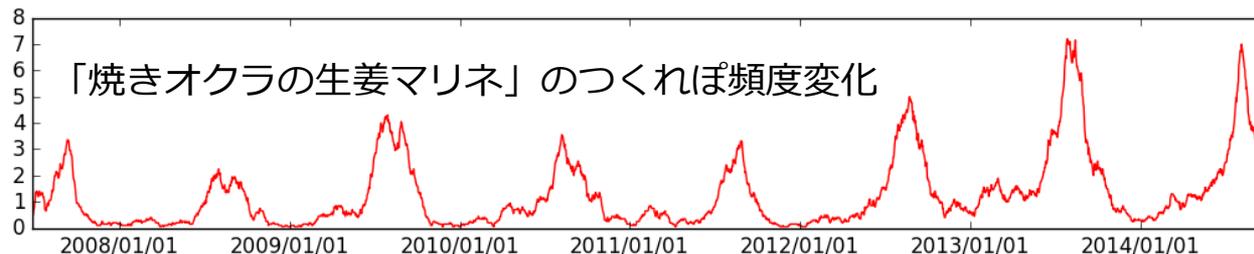


レシピの場合…人間は1年365日食事をしなければ生きられない

日常性：日常生活の季節や年中行事と結びつく

永続性：美味しいレシピは繰り返し参照される

→ **強い周期性が存在**



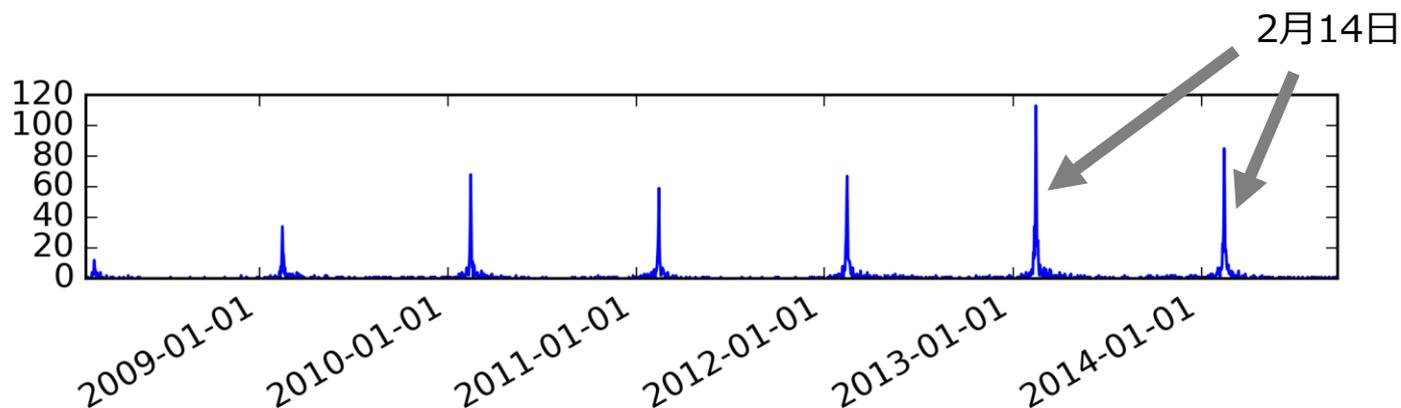
レシピの時間的特徴の定量化

- 調理・食事における周期性に着目して、レシピの時間的特徴を定量化
 1. レシピの1日のつくれば投稿数の頻度ベクトルを作成
 2. 頻度ベクトルを周期 T で折りたたんで特徴ベクトルを作成
 - 周期間の差異（期間・長さ）を吸収し、周期性を明確化
 - $T=1$ 週間, 1ヶ月, 1年

レシピの頻度ベクトル作成

- レシピ r_i のつくれぽの頻度変化をベクトルで表現
- つくれぽ数の**頻度ベクトル** f_{r_i} (次元数 T_{r_i})
 - T_{r_i} …つくれぽの期間
 - $f_{r_i}^{(t)}$ …時刻 t ($0 \leq t \leq T_{r_i} - 1$)のつくれぽ数

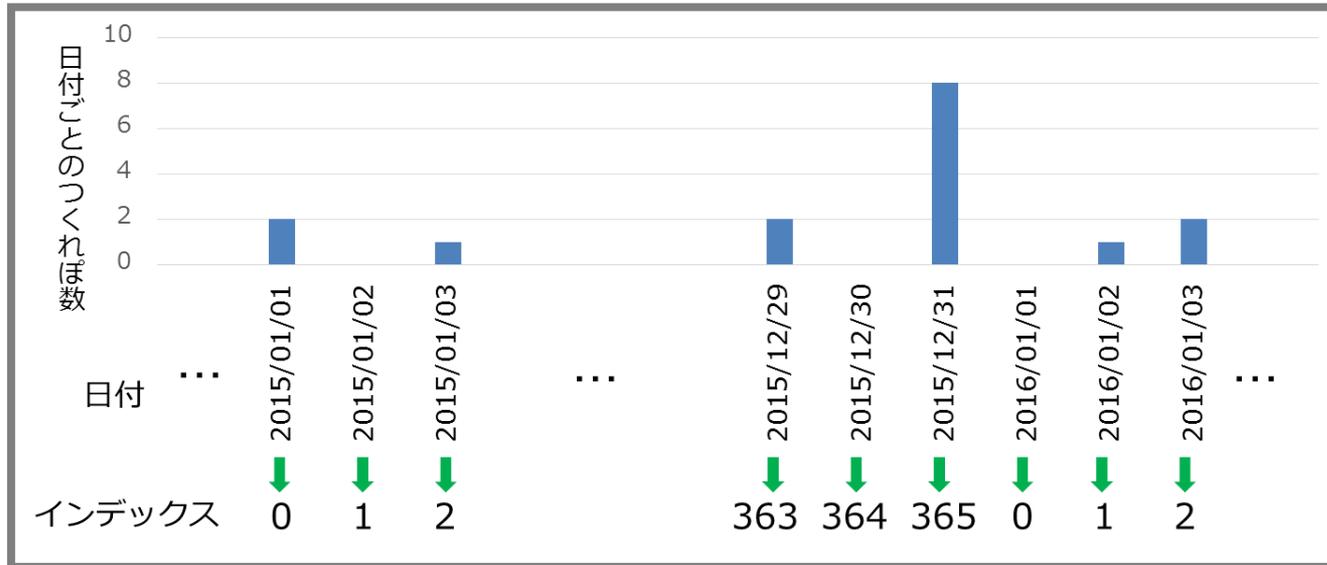
$$f_{r_i} = (f_{r_i}^{(0)}, f_{r_i}^{(1)}, \dots, f_{r_i}^{(T_{r_i}-1)})$$



レシピ「バレンタインにいかが？フォンダンショコラ」の頻度ベクトル

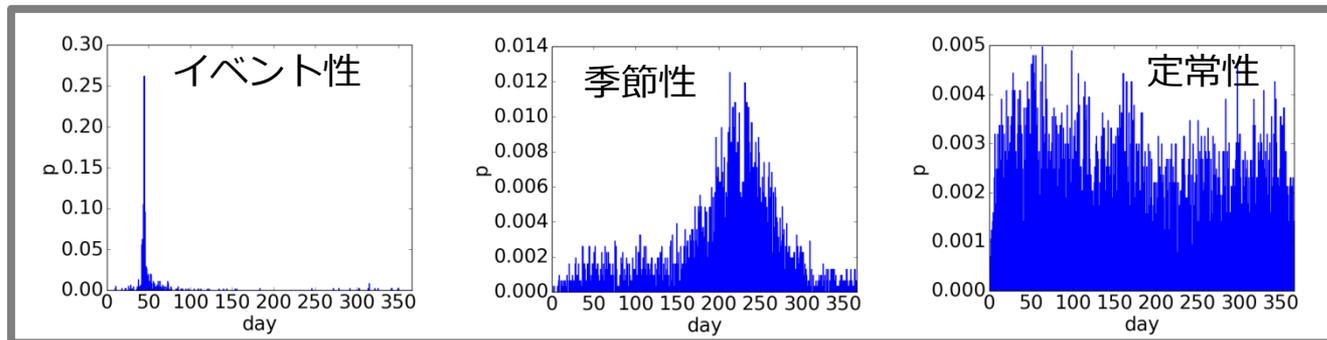
レシピの特徴ベクトル作成

頻度ベクトル



特徴ベクトル

周期 T で折りたたみ
総和が1になるように正規化



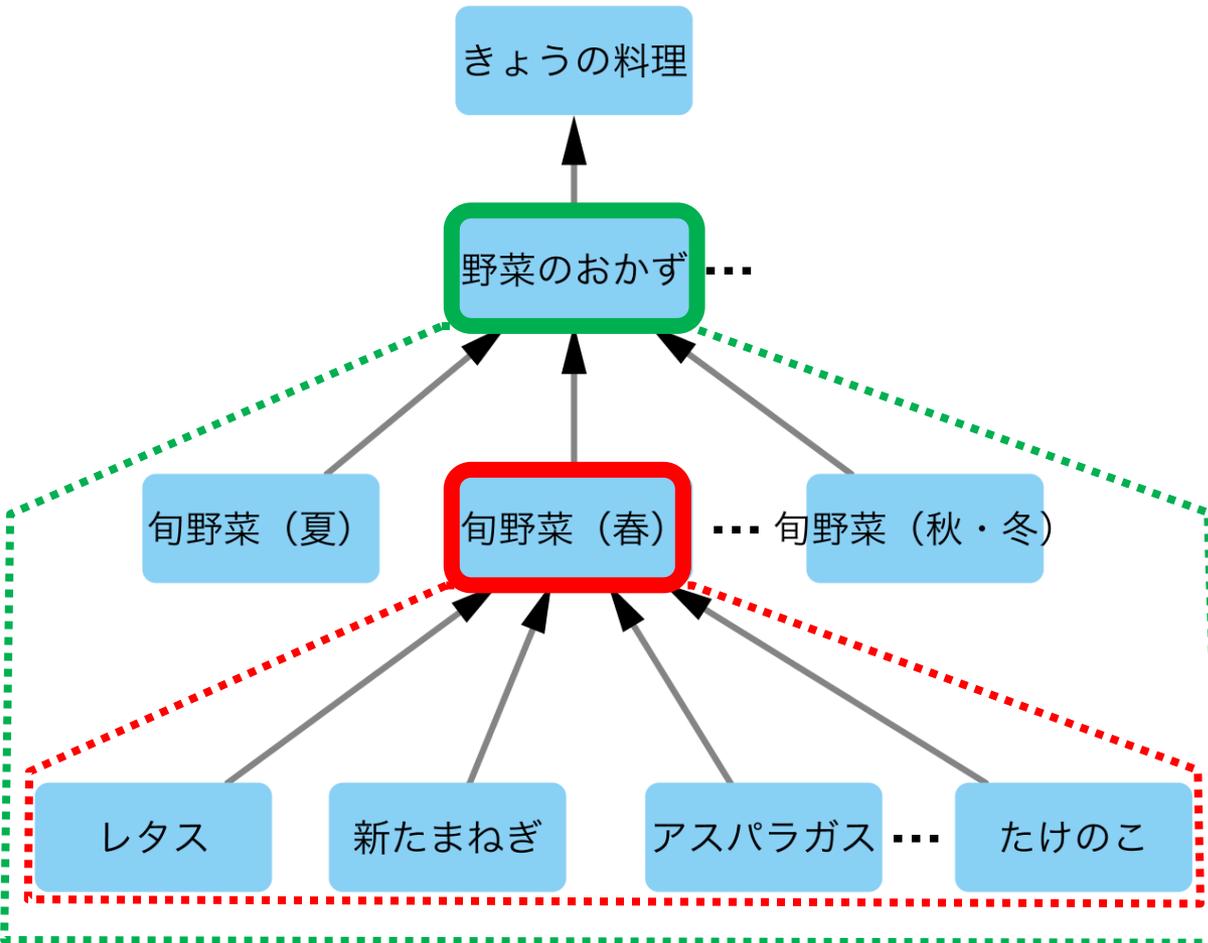
$T=1$ 年の場合

時間的特徴ベクトルによる レシピのクラスタリング

- 時間的特徴ベクトルの妥当性を確認するために, k-means法でレシピを15個にクラスタリング
- 各クラスタの内容を, クックパッドのカテゴリ名で要約
 - 全1099カテゴリ
 - 木構造
 - 人手で付与
 - ただし, 付与されていない, または複数付与されているレシピも存在

カテゴリ使用の問題点

例: 春に作られるレシピのクラスタの場合



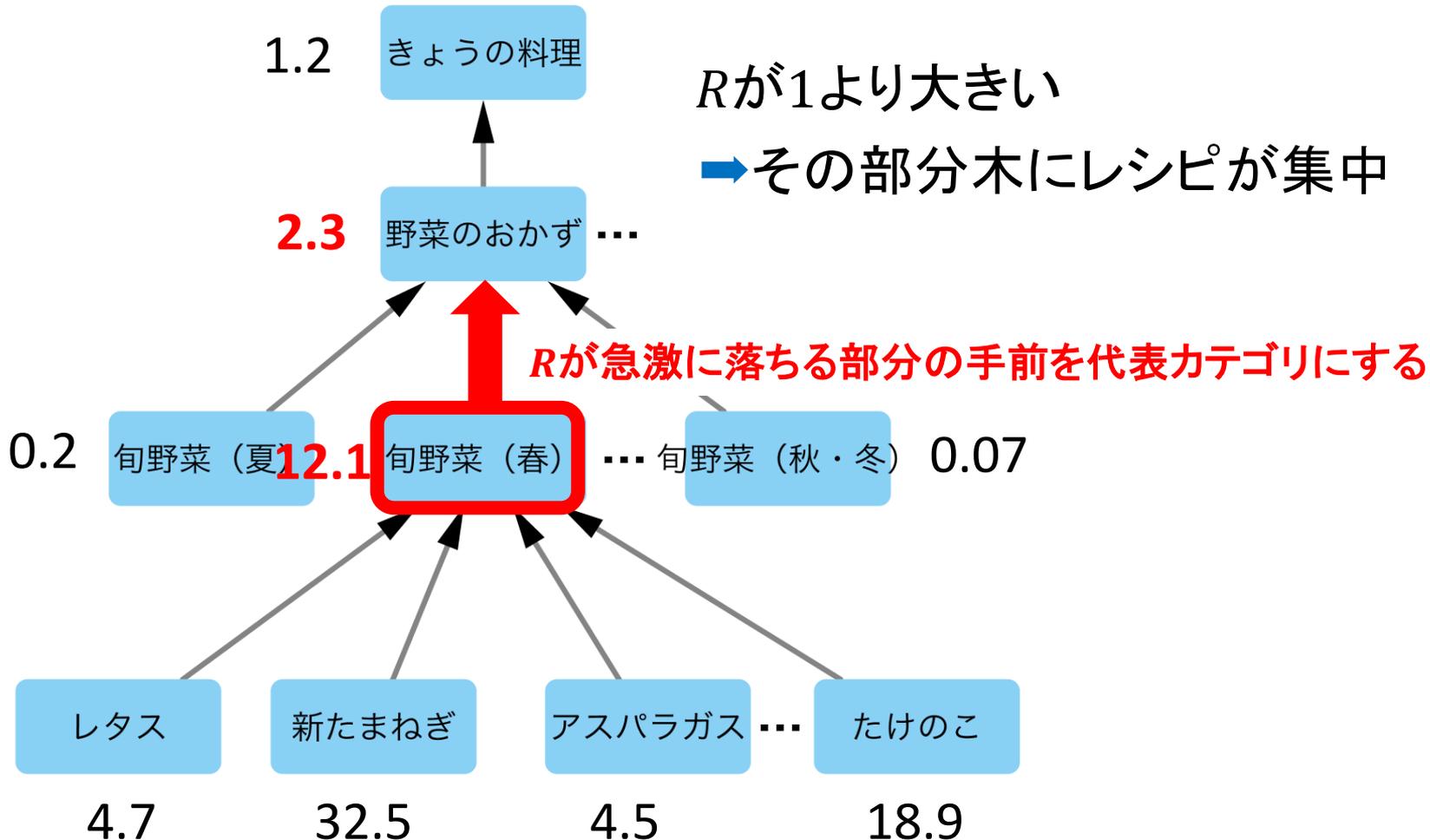
- 出現頻度が高いカテゴリの羅列では、内容がよくわからない

→階層関係を用いて上位層の**直感的なカテゴリ**（**代表カテゴリ**）に集約して表現

レ시피には下層の食材名・料理名が登録されている

クラスタ内容の要約方法

$$R = \frac{\text{あるクラスタにおけるカテゴリ部分木のレシピ存在確率(実確率)}}{\text{レシピ全体におけるカテゴリ部分木のレシピ存在確率(予測確率)}}$$

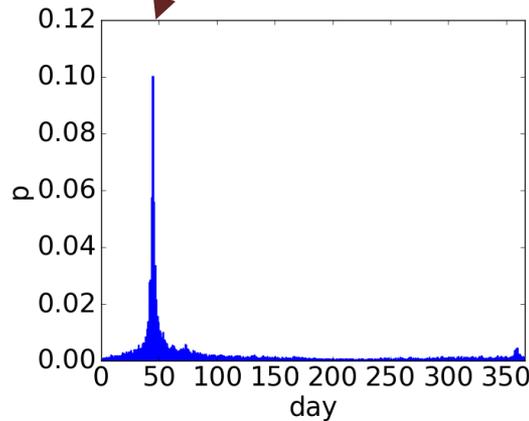


クラスタリング結果

イベント性

お正月、ひな祭り、ハロウィン、etc.

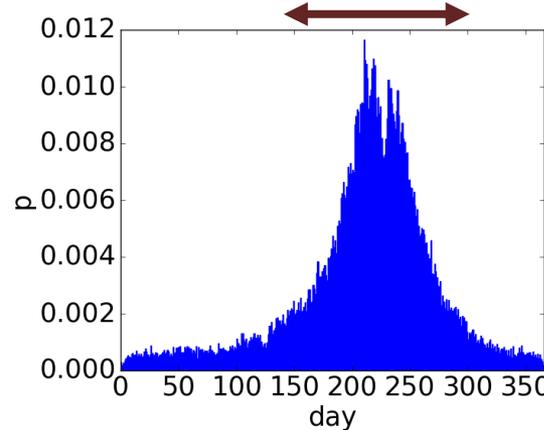
2月14日
(バレンタインデー)



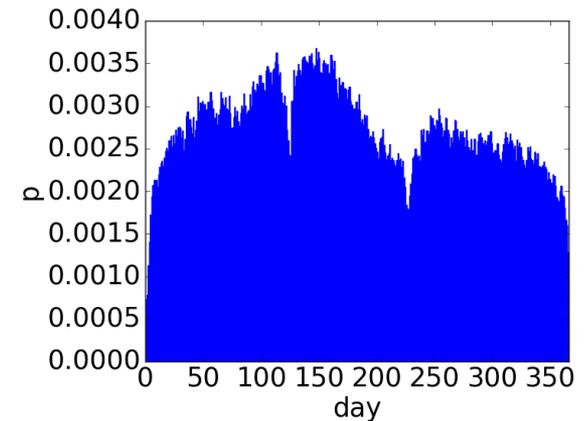
季節性

春夏秋冬や、その移り変わりの時期など

7月上旬～9月下旬(夏)



明確な周期性なし



代表カテゴリ(レシピ数が多い順に上位5件)

チョコレートのお菓子	103	旬野菜(夏)	149	人気のキャベツレシピ	151
もっと料理を楽しむ	36	夏に食べたい料理	21	お弁当のメインおかず	52
おもてなしデザート	9	そうめん	17	再現レシピ	38
持ち寄り・プレゼント	2	夏！おもてなし料理	15	子どもが喜ぶ♪かわいいおかず	37
シリコンスチーマーでお菓子	1	冷製・アイデアパスタ	10	運動会のお弁当	34

時間的特徴を用いたレシピ検索

1. 与えたレシピの時間的特徴で類似レシピ検索
 - コサイン類似度でランキング
2. 注目する事象の時間パラメータから推定した時間的特徴で類似レシピ検索
 - 検索語 \leftrightarrow 時間パラメータのDBを利用
 - ただし, 任意の検索語は使えない
3. レシピとつくればのテキストから求めた言語表現の時間的特徴で類似レシピ検索

時間パラメータによる 特徴ベクトル作成

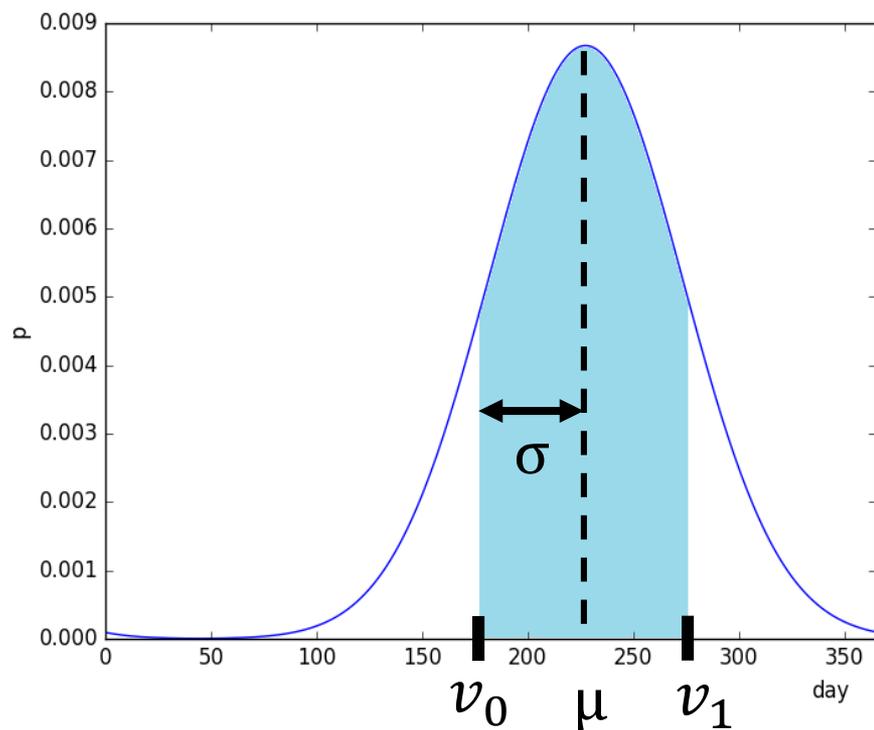
時間的状況を表す言語表現

例：夏, バレンタインデー



時間パラメータ + 確率密度関数

特徴ベクトル



固有の確率分布を持つと仮定

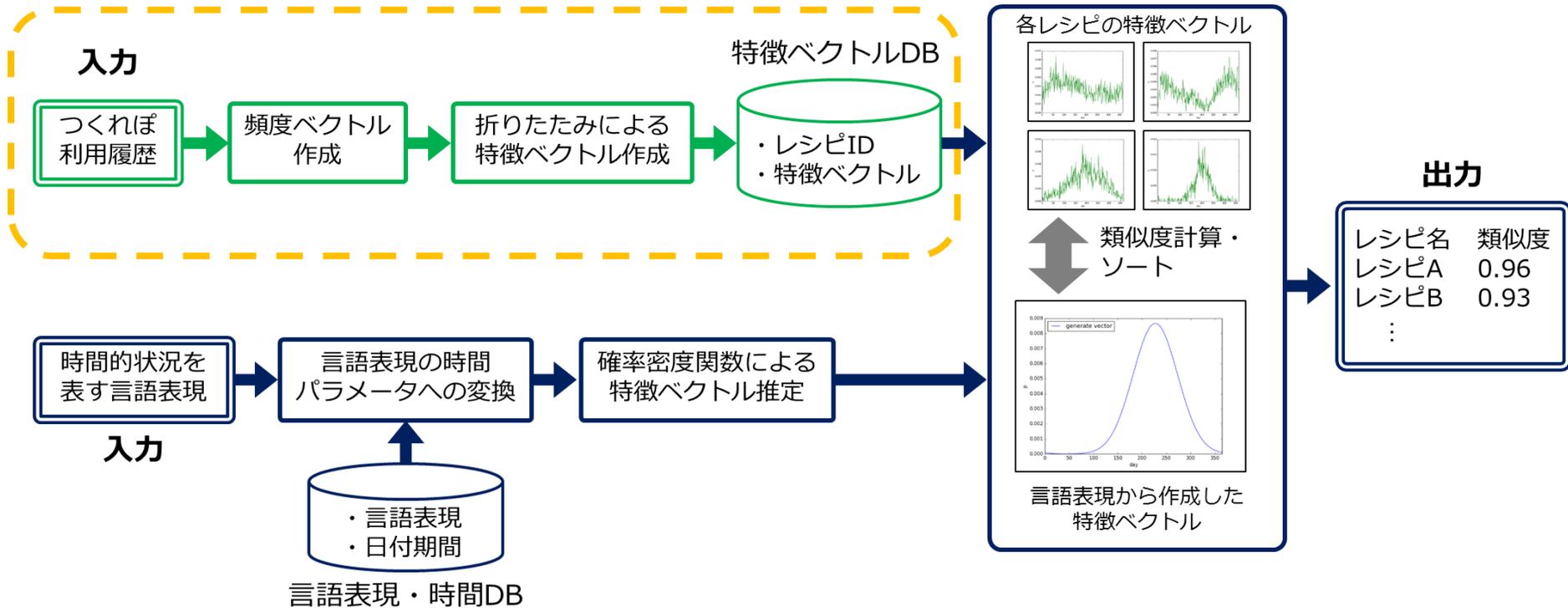
- イベント性と季節性：単峰性の正規分布
 - 実データもイベント日がピーク
- 時間的特性がない：一様分布

パラメータ設定

- 開始日： v_0
- 終了日： v_1
- 平均： $\mu = (v_0 + v_1)/2$
- 標準偏差： $\sigma = (|v_0 - v_1| + 1)/2$
- 期間が1日の場合は $v_0 = v_1$

時間パラメータによるレシピ検索

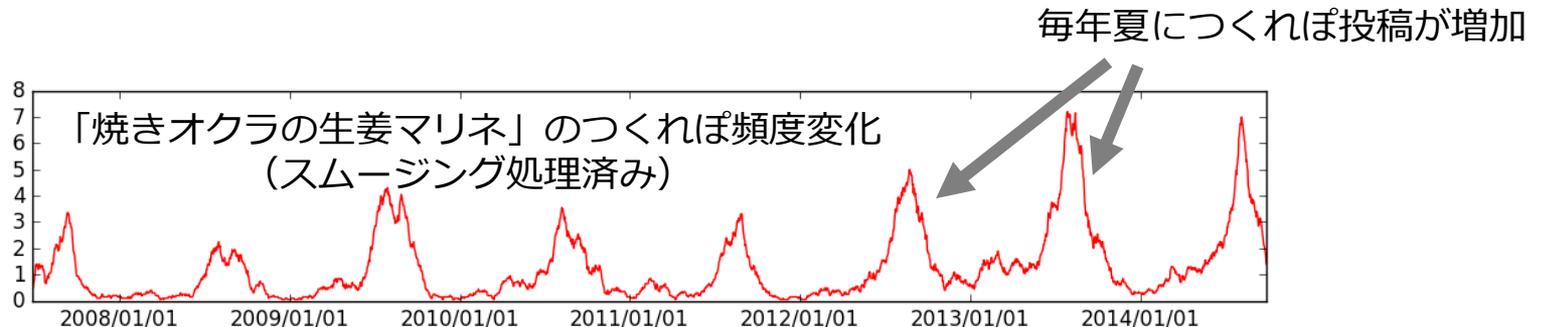
レシピの特徴ベクトルは事前に作成



レシピと言語表現の違い

- **レシピ**

- つくれぽの頻度変化から求められる



- **言語表現**

- レシピ中の全ての語がレシピと同じ時間的特徴を持たない
 - 「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」
 - 調理時期は初夏で「春」は比較対象



- **ノイズ除去**

- TF-IDFによる言語表現の選別
- スペクトルサブトラクション法 (SS法) による定常成分の除去

言語表現の抽出

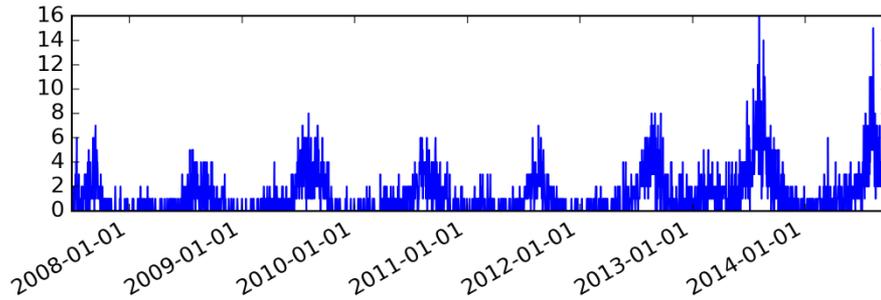
- レシピ本文に加えて、つくれぽの文章も用いる
 - 短く簡潔だが、複数のユーザが投稿
 - ➡調理動機などのレシピの特徴が繰り返し言及 **(集合知)**
- 1. レシピ・つくれぽのテキストを正規化し、形態素解析で単語に分割し、ストップワード・記号類を除去
- 2. TermExtractを用いて複合語化し、言語表現とする
 - 名詞が連続している部分を複合語として抽出
- 3. 各レシピ中のTF-IDF値が高い上位 M 件の言語表現を抽出し、言語表現と用いられているレシピ群の関係を表す言語表現・レシピ対応表を作成
 - $M = 48$ (テキストの語彙数の中央値240語の上位2割)

抽出された言語表現

- 多くが食材名
 - 「ピーマン嫌い」, 「大根消費」のような複合表現も
 - 特定の目的を示す言語表現
 - 「風邪気味」, 「プレゼント用」のような用途
 - 「花見」, 「ホワイトデー」などの行事・イベント名
 - 調理動機とは言えないような言語表現
 - 一般的な調味料など
- ➡ 時間的特性が顕著な言語表現をUIで利用できる可能性あり

言語表現の頻度ベクトル作成

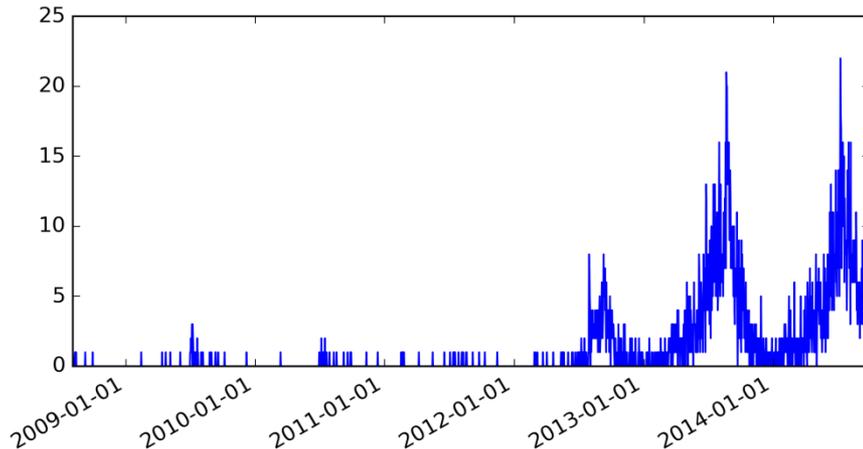
ある言語表現に関する
レシピの頻度ベクトル



+

⋮

+



- ある言語表現に関するレシピの頻度ベクトルの総和を計算

言語表現・レシピ対応表

言語表現	レシピ群
言語表現A	レシピa, レシピb...
言語表現B	レシピa, レシピc...
言語表現C	レシピd, レシピe...
⋮	

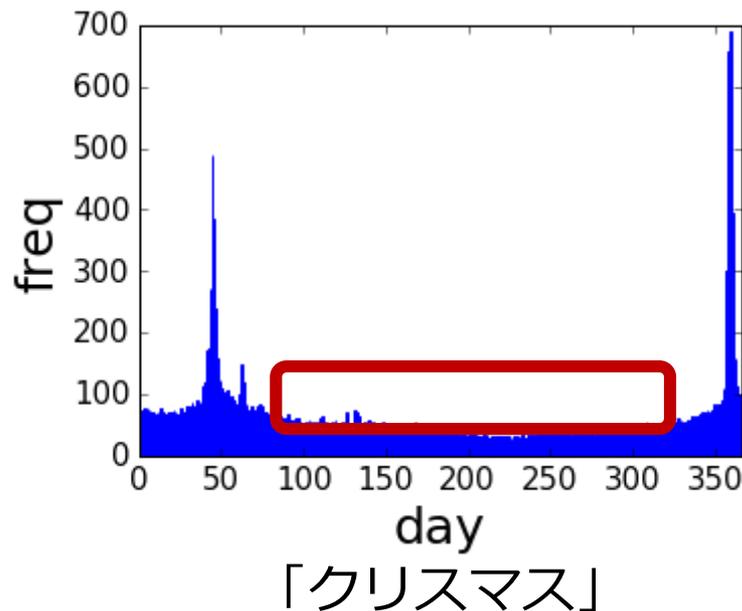
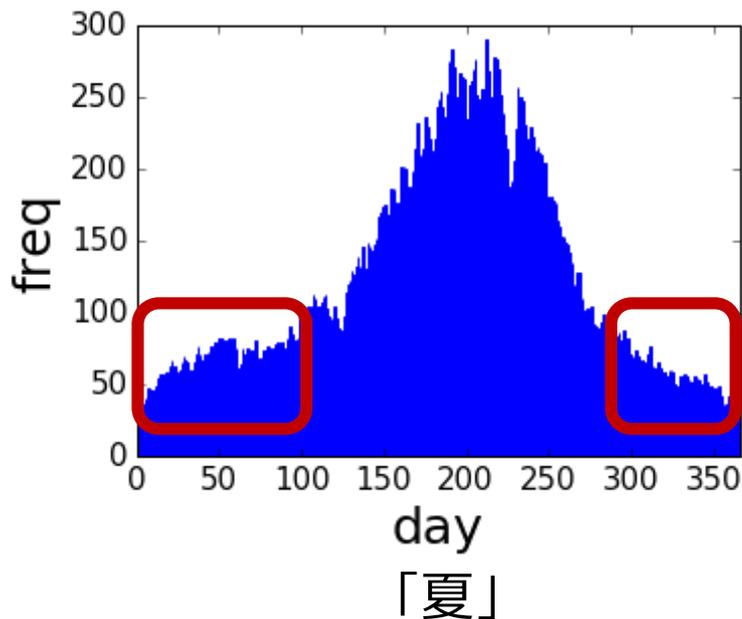
- レシピごとにつくれぽの投稿期間が異なる
 - 頻度ベクトルの次元数はデータの全期間数とする

特徴ベクトルからの定常成分の除去

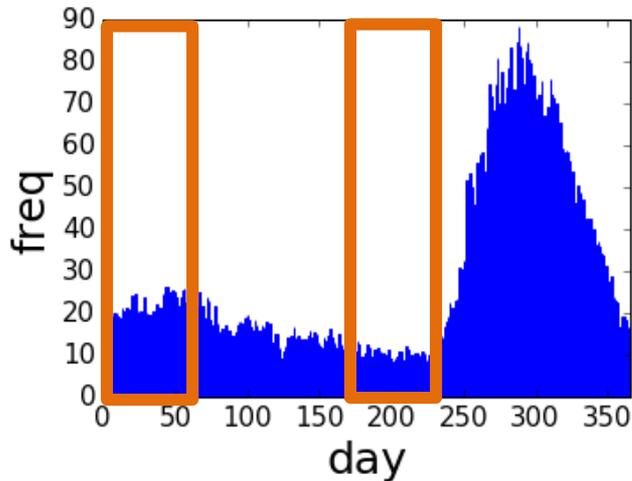
- 頻度ベクトルを折りたたんで特徴ベクトルを作成
- ただし、時間に伴って変動しない定常成分が存在
 - 類似度比較に大きく影響

→ スペクトルサブトラクション法で除去

- 音声強調に用いられる
- 入力信号のパワースペクトルからホワイトノイズ（定常成分）のパワースペクトルを減算して元信号を得る

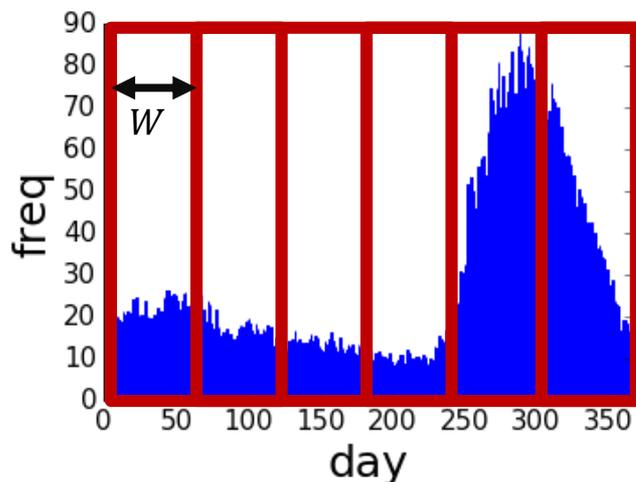


スペクトルサブトラクション法



定常成分の推定

- 変動の底の部分を定常成分とみなす
- 長さ W の窓を移動しながら平均値を計算
- 平均値が最も低い部分を用いる
- 366次元を等分割できる $W = 61$ とした

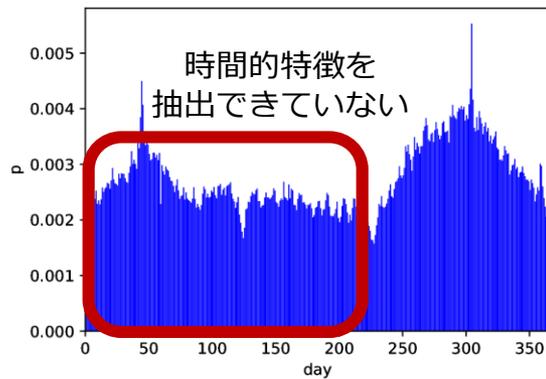


SS法の適用

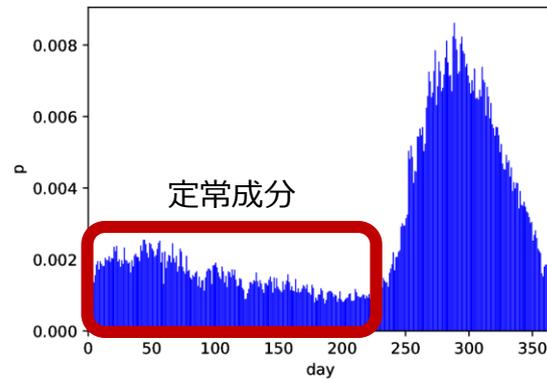
- 長さ W に分割した各区间ごとにSS法を適用
- 各要素の和が1になるように正規化

ノイズ除去の効果

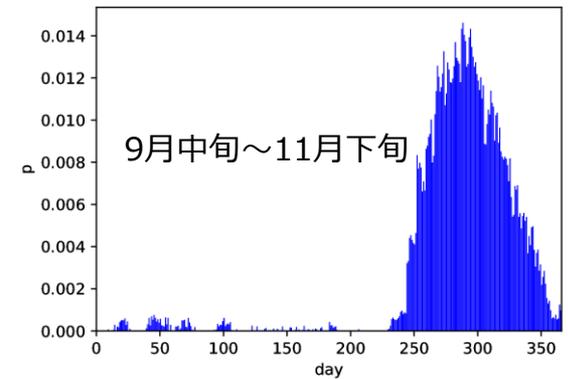
- 言語表現「秋」の特徴ベクトルで比較



言語表現をそのまま
全て用いた場合



TF-IDF値で言語表現の
選別を行った場合

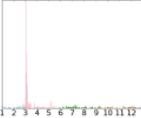


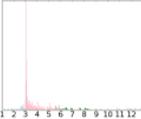
SS法で定常成分の除去を
行った場合

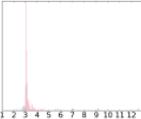
プロトタイプシステム

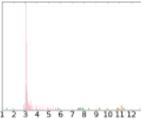
時間検索 検索

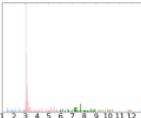
1ページ目(1-40件目表示中)

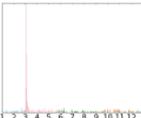
 **★ひな祭り★基本の五目ちらし寿司の素**
これさえ覚えておけば後は何をトッピングしてもおいしい。フライパン一つ。切って煮るだけ。友達にも大好評でした♪冷凍保存可能です！子供のためのひな祭り、ぜひ手作りです！
つくれば数：398件

 **ひなまつり＊簡単＊いちごのムースケーキ**
れぽ400感謝！固めるだけで、とっても簡単です★2011.4に話題入り★2012.4'ひな祭り'で1位！！
つくれば数：422件

 **ひな祭り♪三色ヨーグルトババロア**
ひな祭りに可愛いババロア、いかがですか(#^_^#)一口サイズでさっぱりとしたデザートどうぞ♪
つくれば数：201件

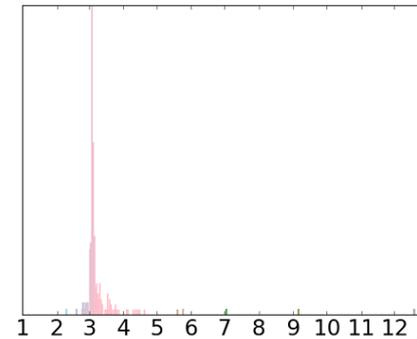
 **ひな祭り＊ちらし寿司ケーキ**
ひな祭りにぴったりの、ひし餅を意識した3層のかわいいちらし寿司です♡
つくれば数：162件

 **ちらし寿司に失敗無し薄焼き卵・錦糸卵コツ**
失敗しない薄焼き卵と錦糸卵のコツです。散らし寿司・冷やし中華などにどうぞ。白だしが無くても作れるレシピ記載しました。
つくれば数：311件

 **蛤のお吸い物＊**
昆布と蛤のお出汁が美味しいお吸い物です＊ 2012.12.17 3つ目の話題入り♡ 2013.02.26 カテゴリ登録♡
つくれば数：316件

検索モードと検索語を入力
(時間的特性3種類+全文検索)

特徴ベクトルを季節ごとに彩色
春, 夏, 秋, 冬
→季節性を直感的に表現可能



拡大図

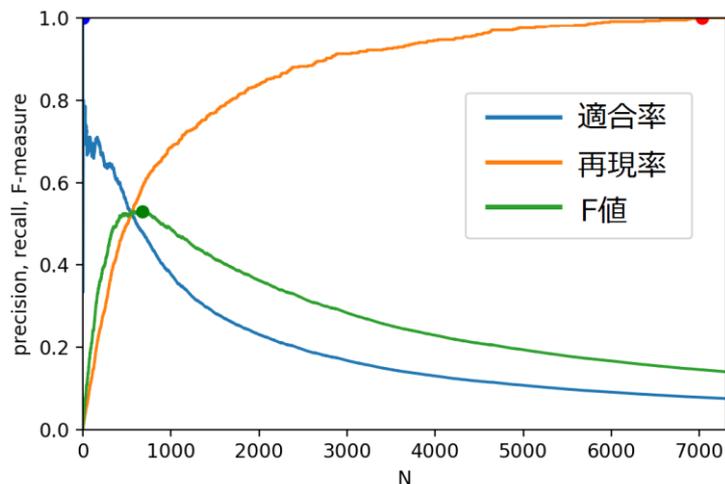
テキスト中に検索語の「ひな祭り」が含まれていなくても、関連メニュー（「ちらし寿司」など）を上位に提示できる

検索結果の妥当性の評価方法

- 調理動機の時間的特徴に合致したレシピが検索結果の上位に来ているか？
 - **カテゴリ情報**を正解データとして用いて評価
 - 名前に言語表現を含むカテゴリとその下位カテゴリのレシピを正解と見なす
 - **ただし、正例と負例が明確に区別できないことに注意**
 - 関連カテゴリが全て付与されるとは限らない
 - カテゴリが付与されていなくても正しい場合がある
 - cookpadのカテゴリ名は限られている
 - 提案手法で「遠足弁当」や「風邪気味」のような多彩な言語表現を抽出しても、対応する正解カテゴリがない
- ➡上位N件の適合率・再現率・F値の変化を評価

検索結果の妥当性評価 (1)

「夏」



正解カテゴリ数：41件

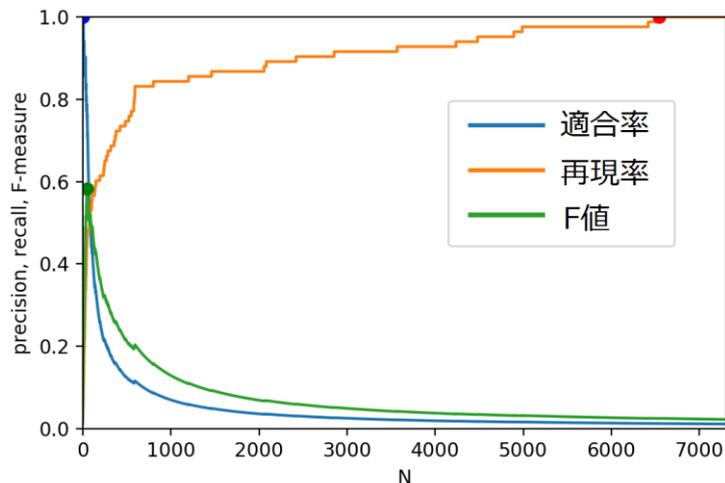
- 「夏！おもてなし料理」
- 「旬野菜（夏）」
- 「夏に食べたい料理」
- 上記3つの下位カテゴリ

正解レシピ数：552件

- 正解は検索結果の上位に集中
 - 「夏」のレシピを適切にランキング
- 上位でも適合率が0.7程度
 - 適切なレシピを提示できていたが、正解としたカテゴリ以外の夏らしいカテゴリが付与
 - 「冷やし中華」, 「そうめん」, 「アイスクリーム」など

検索結果の妥当性評価 (2)

「クリスマス」



正解カテゴリ数：20件

- 「クリスマス」 (歳時記の下位カテゴリ)
- 「クリスマス」 (その他の下位カテゴリ)
- 上記2つの下位カテゴリ

正解レシピ数：83件

- 適合率の低下が急
 - 正解レシピ数が少なく、早い段階でほとんど提示
 - 適切なレシピでも、「クリスマスケーキ」ではなく「チョコレートケーキ」「基本のスポンジ」などのカテゴリが付与されている場合も
- 再現率が1になかなか達しない
 - 下位カテゴリの「大皿料理」「プレゼントレシピ」「持ち寄りレシピ」などは、クリスマス以外にも調理される

まとめ

- レシピの時間的特徴を用いたレシピ検索法を紹介
- 現在後輩がクックパッドデータセットの献立データを用いて研究中

人工知能学会誌 2019年1月号

- 特集：「料理情報の知的処理」
 - エディタ
 - 原島 純（クックパッド）
 - 風間 一洋（和歌山大学）
 - 記事11件・執筆者17人
 - 実サービス, データセット, 自然言語処理, 画像認識, 検索, ユーザビリティ, 食事記録, スマート調理, スマートキッチン, 研究・学会の状況など広範囲
- もちろん, NII-IDRも紹介されています！
 - クックパッドデータセット, 楽天データセット