

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

桐本 宙輝^[1]

風間 一洋^[1]

[1] 和歌山大学大学院システム工学研究科

はじめに

背景

- 多くのユーザに共通する代表的な調理動機として、季節や年中行事などの時間的狀況が存在する
 - 夏にはそうめん、冬には鍋料理、お正月にはお節料理、節分には恵方巻きなど

目的

- 時間的な調理動機を表す言語表現とその時間的特徴を、レシピやつくれぽのテキストから発見的に抽出する
 - 言語表現：「バレンタイン用」「遠足弁当」など
 - 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
- シチュエーションや季節性などの時間的特徴に基づいたレシピ検索に応用

レシピの時間的な特徴

つくれぽの頻度変化から容易に時間的特徴を求められる



言語表現の時間的な特徴

比較や否定のために用いられる場合もあり、時間的特徴を推定するのは難しい

- 例：「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」
- このレシピが作られる時期は初夏で、「春」は比較のために用いられている

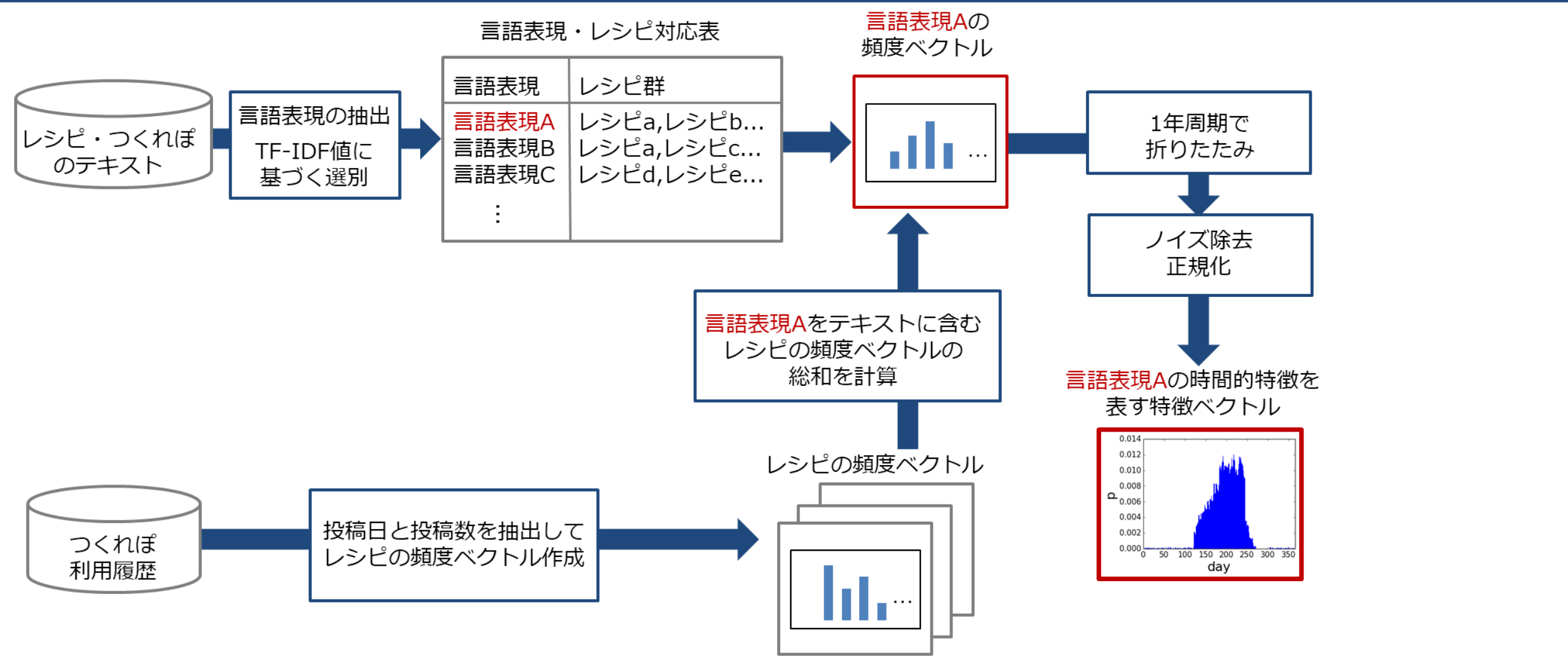
この問題を解決するには…

- レシピ本文とあわせてつくれぽのテキストも用いる
 - 短く簡潔なため、レシピの主な特徴について言及される
 - 複数のユーザによって投稿されるため、調理動機などのレシピの特徴を的確に示す言語表現が繰り返し使われ強調される



- TF-IDF値に基づいた言語表現の選別が有効
- つくれぽ数の頻度変化を統合すれば、言語表現の時間的特徴を推定できる

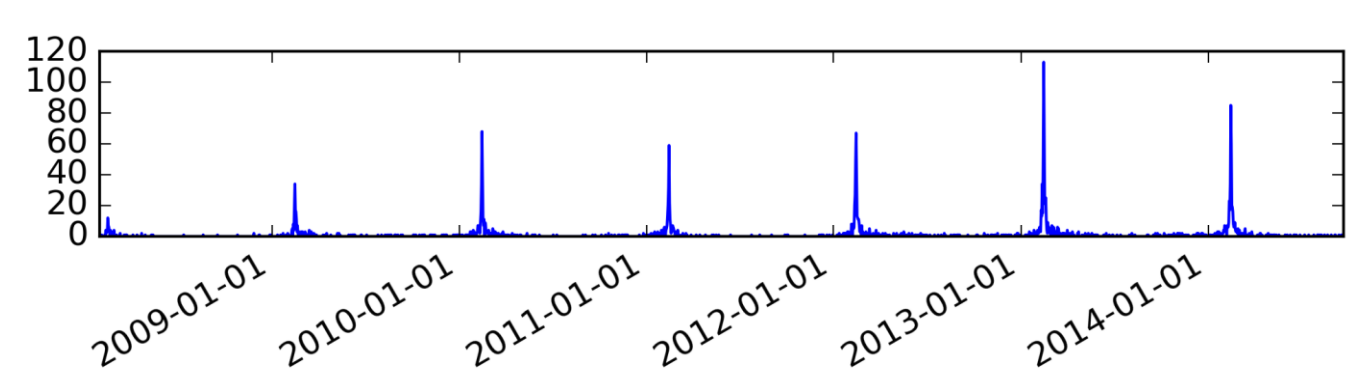
提案手法の概要



言語表現の抽出

- レシピおよびつくれぽのテキストを正規化し、単語に分割して、ストップワードや記号類を除去
- TermExtractを用いて複合語化し、言語表現とする
- 各レシピ中のTF-IDFの上位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示す言語表現・レシピ対応表を作成
 - テキストの語彙数の中央値が240語であったため、2割のM=48とした

レシピの頻度ベクトル作成

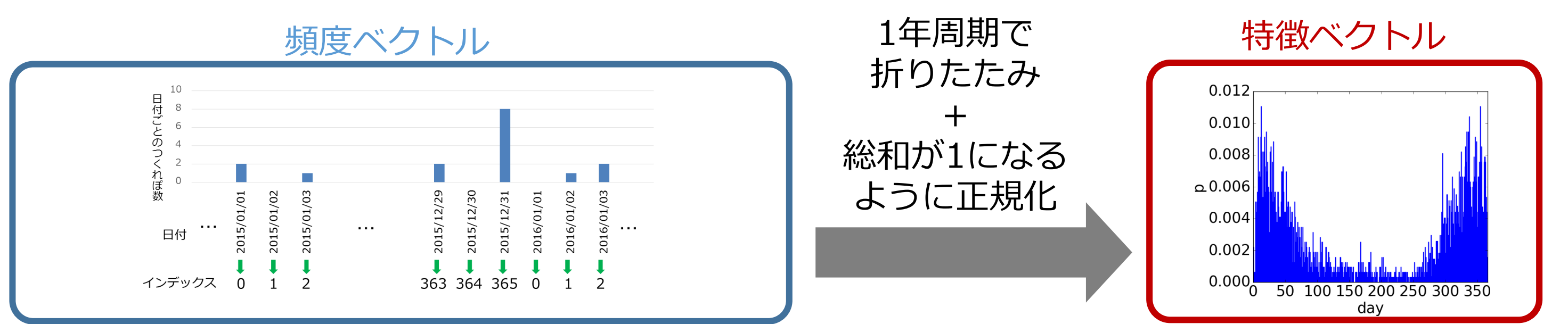


レシピ「バレンタインにいかが？フォンダンショコラ」の頻度ベクトル

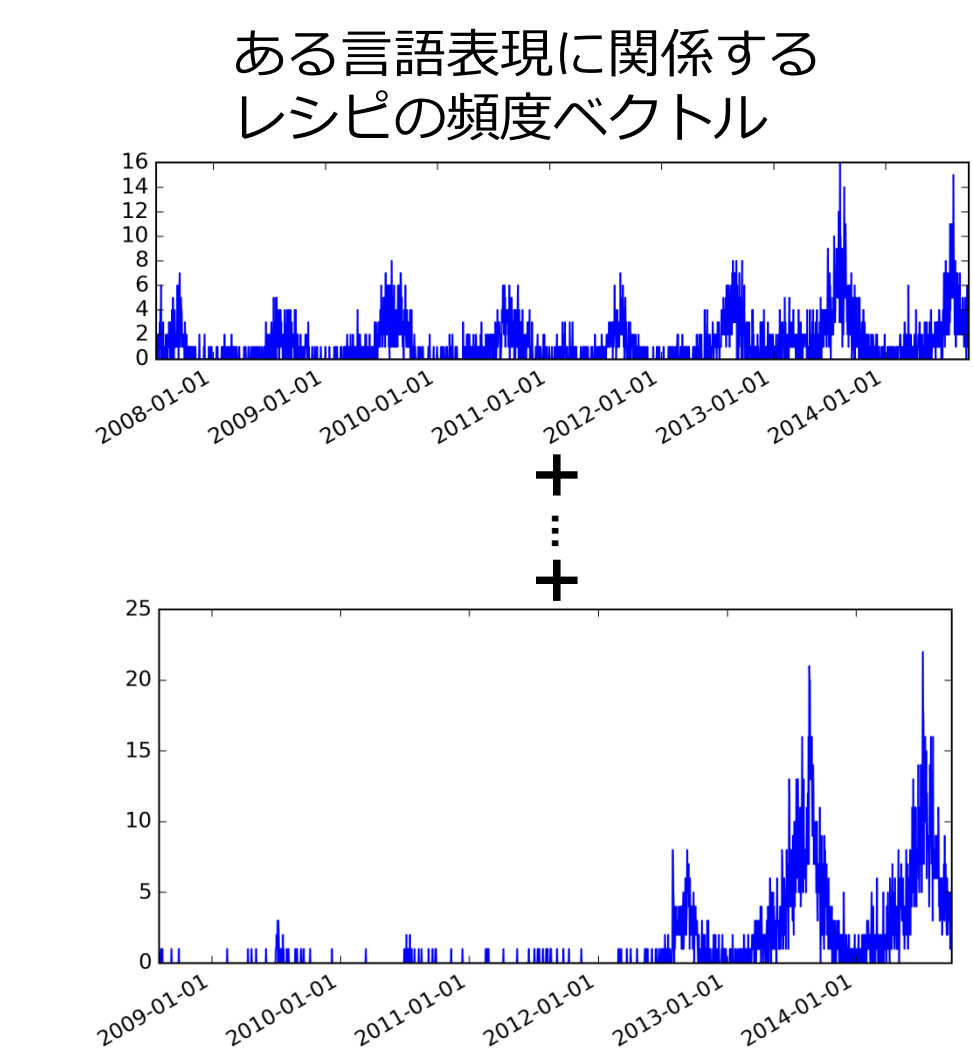
- レシピ r_i が期間 T_{r_i} のつくれぽの情報を持つ
- 時刻 $t(0 \leq t \leq T_{r_i} - 1)$ のつくれぽ数は $f_{r_i}^{(t)}$
- 次元数 T_{r_i} のつくれぽ数の頻度ベクトル f_{r_i}

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

レシピの特徴ベクトル作成



言語表現の頻度ベクトル作成



- 言語表現・レシピ対応表に基づいて、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
- レシピごとにつくれぽの投稿期間が異なるので、頻度ベクトルの次元数はデータの全期間数にする

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...

言語表現の特徴ベクトル作成

- 1年周期での折りたたみに関してはレシピの特徴ベクトルと同様
- 言語表現の場合、季節によって変動しない定常成分が含まれている



- スペクトルサブトラクション法で定常成分を除去
 - あらかじめノイズを何らかの方法で求めておく必要がある
- 言語表現の頻度変化は変動成分と定常成分の合成で、時間的な言語表現の場合は、その変動の底の部分では定常成分だけになる点に着目
- 366次元に折りたたんだ頻度ベクトルを、長さWのウィンドウを移動しながら平均値を算出し、平均値が最も低い区間をノイズとして用いる
- 頻度ベクトルを長さWに分割した各区間ごとにスペクトルサブトラクション法を適用し、総和が1になるように正規化することで特徴ベクトルを作成する

調理動機によるレシピの検索

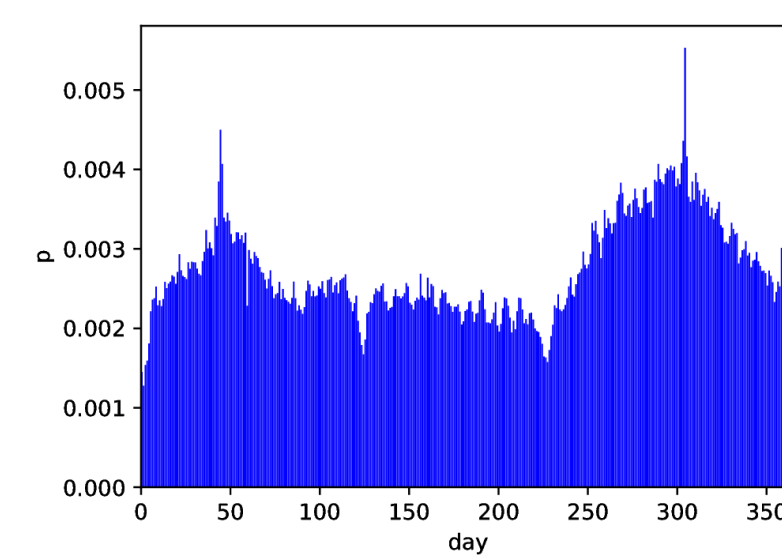
- レシピの特徴ベクトルと、言語表現の特徴ベクトルのコサイン類似度を計算
- 類似度が高いレシピほど、そのレシピが調理された時間的狀況が言語表現の時間的特徴に近いとして、検索用GUIでは類似度上位から順に表示する

評価

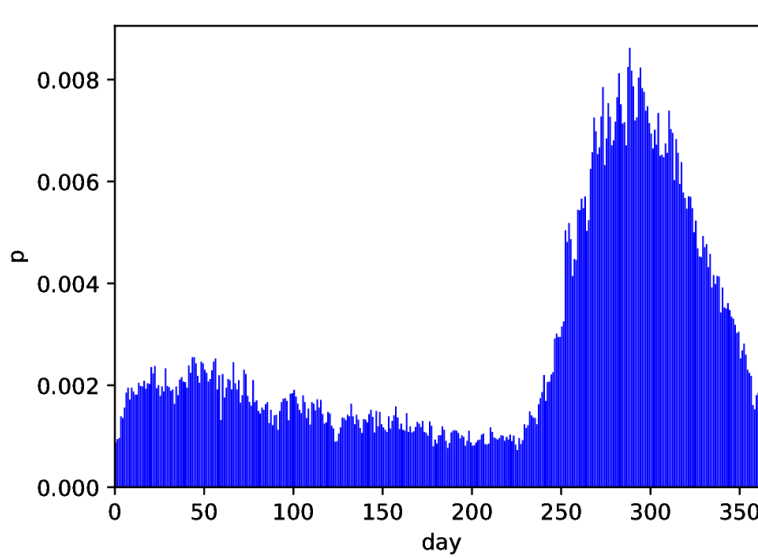
- クックパッド株式会社と国立情報学研究所が提供する「クックパッドデータ」を利用
- つくれぽ数100件以上、つくれぽ投稿期間366日以上レシピ11,507件が対象
- 評価にはその中からカテゴリが付与されている7,304件のレシピを使用

ノイズ除去の効果

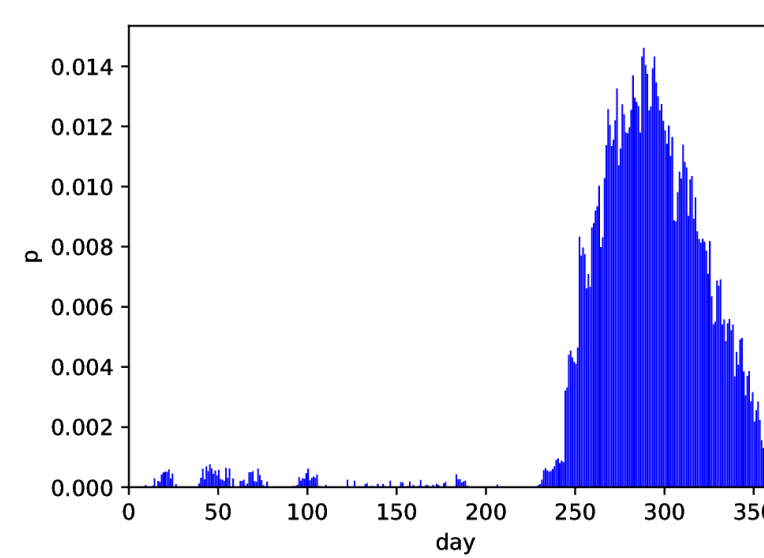
- 言語表現「秋」の特徴ベクトルでそれぞれの場合を比較



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



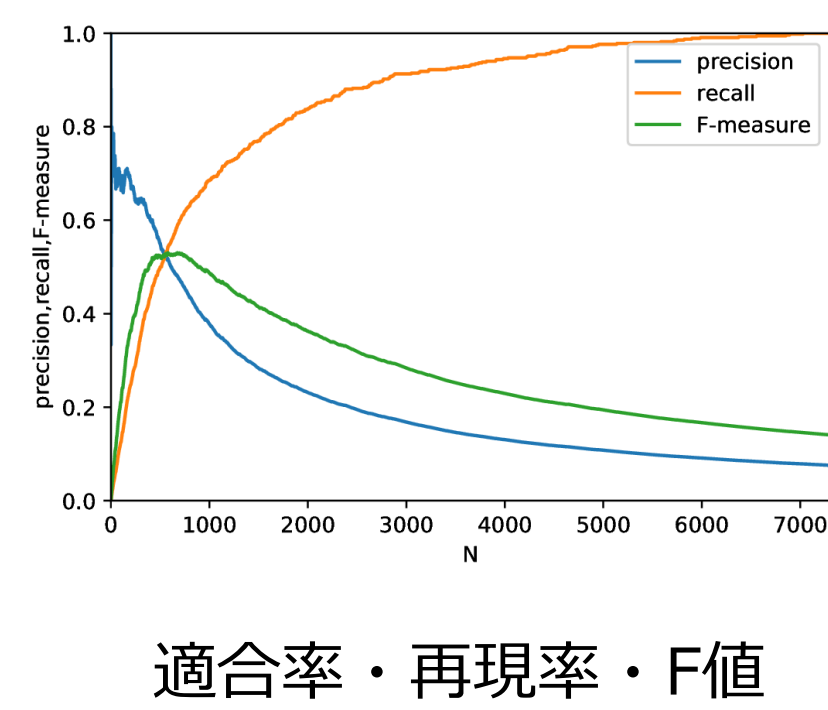
スペクトルサブトラクション法で定常成分の除去も行った場合

検索結果の妥当性（評価方法）

- cookpadのレシピに付与された1099個の木構造のカテゴリを用いて、検索結果の妥当性を定量的に評価
 - 但し、正例と負例が明確に区別できる正解データとしては使用できない
- カテゴリ名に言語表現と同じ文字列を含むカテゴリとその下位カテゴリが付与されているレシピは少なくとも正例であると見なす
- 言語表現の検索結果から、カテゴリが付与されていないレシピを除いた上位N件の検索結果の適合率・再現率・F値を求め、その分布を調べる
- 提案手法では「遠足弁当」や「風邪気味」「夏バテ」「大根消費」のような多彩な言語表現が抽出できるが、cookpadのカテゴリ名は比較的限られていることから、季節性を示す「夏」と年中行事を示す「クリスマス」2つの言語表現を評価に用いる

検索結果の妥当性（評価結果）

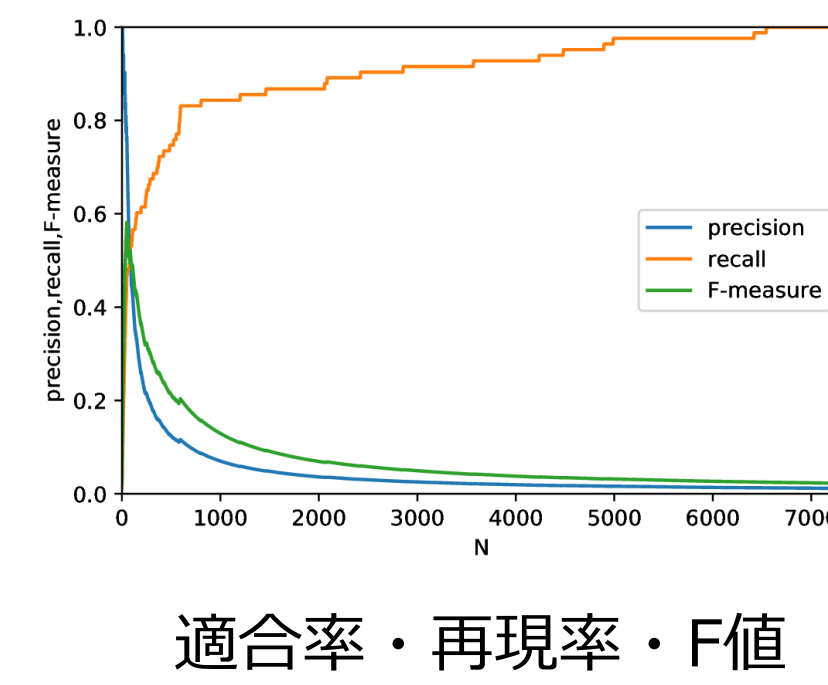
「夏」



適合率・再現率・F値

- 正解カテゴリは「夏！おもてなし料理」「旬野菜（夏）」「夏に食べたい料理」とその下位カテゴリの合計41カテゴリ、正解レシピ数は552件
- 正例は検索結果の上位に集中し、逆に下位になるほど含まれていないことが分かる
- 適合率が1にならないのは、「冷やし中華」や「そうめん」「アイスクリーム」などの、カテゴリに名に使用した言語表現を含まなくても、いかにも夏らしいカテゴリが存在するため

「クリスマス」



適合率・再現率・F値

- 正解カテゴリは「歳時記」と「その他」の下位カテゴリである2つの「クリスマス」と、その下位カテゴリの合計20カテゴリ、正解レシピ数は83件
- 「夏」より適合率の低下が急で、Nが小さい時にF値のピークが現れているが、これは正解としたレシピ数が少なく、早い段階でほとんど正解を提示できているため
- 再現率はN=589で0.8を超えるが、1になかなか達しないのでなだらかに推移しているのは、下位カテゴリの「大皿料理」や「プレゼントレシピ」「持ち寄りレシピ」が、実際にはクリスマス以外にも調理されているため

- 実際にクリスマスの時期に調理されているレシピでも、「チョコレートケーキ」や「基本のスポンジ」のカテゴリが付与されるが、「クリスマス」やその下位カテゴリが付与されていなかったケースが多く見られた
- 人手による作業では、網羅性の高いカテゴリ付与を行うのは難しいことが分かる

今後の課題

提案手法で抽出した言語表現から、食材名を含む複合表現や用途、行事・イベント名などの、調理動機や調理条件を表す、検索の手がかりとして有用なものを抽出する