

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

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

- 近年、cookpadなどのユーザ投稿型レシピサイトが流行している
- 収録レシピ数が膨大で、類似レシピも大量に存在する
- ユーザの目的・嗜好に合ったレシピを探すのは困難という問題がある

目的

- ユーザの想定状況に合致したレシピ検索を実現する
 - 想定状況：季節・行事のような時間的状況
 - 時間的状況は言語表現を使用

既存研究との比較

既存研究

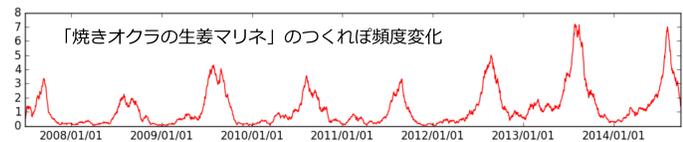
- 個人の嗜好を反映したレシピ推薦^[1]
 - 食材の利用履歴やレシピの閲覧・摂食履歴に基づく個人の嗜好を反映した推薦
- レシピの生い立ちによる検索^[2]
 - ブログ型レシピの文章から抽出したレシピの誕生・使用事由に基づいた検索

本研究では「つくれば」機能に着目

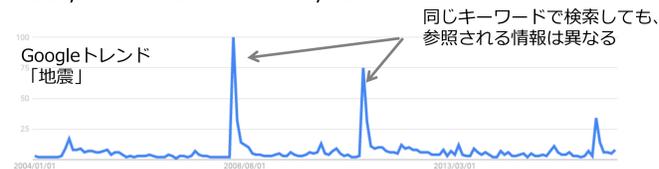
- レシピを実際に調理し、写真付きで感想を投稿する仕組み
- 投稿の累積によりレシピ固有の時間的特性が発生

レシピの時間的な特徴

日常性：普段の生活や季節・行事と結びつき、周期性を持つ
 永続性：美味しいレシピは繰り返し参照される

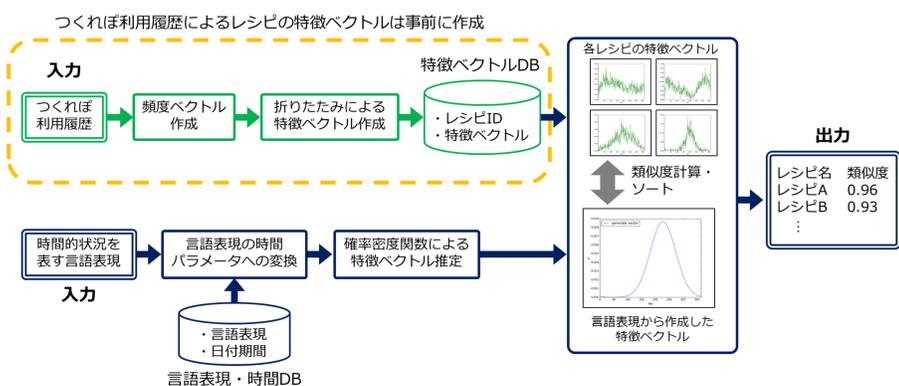


ニュース記事は非日常的で、情報の陳腐化が早い
 例：新製品の発表、突発的な災害・事件, etc...



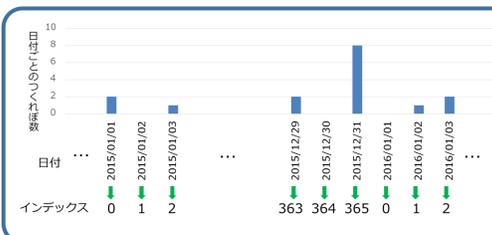
レシピ検索では、従来の検索システムのログの使用法と異なるアプローチがとれる

提案手法のシステム構成



レシピの特徴ベクトル作成

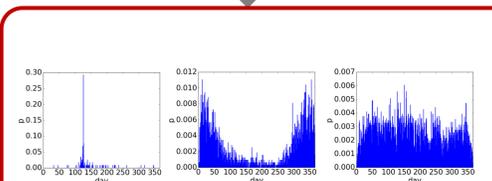
頻度ベクトル (日ごとのつくればの頻度をベクトルで表現)



- レシピ r_i が期間 T_{r_i} のつくればの情報を持つ
- 時刻 $t(0 \leq t \leq T_{r_i} - 1)$ のつくれば数を $f_{r_i}^{(t)}$ とする
- 次元数 T_{r_i} のつくれば数の頻度ベクトル f_{r_i}

$$f_{r_i} = (f_{r_i}^{(0)}, f_{r_i}^{(1)}, \dots, f_{r_i}^{(T_{r_i}-1)})$$
- このままではレシピによって次元数がバラバラ

特徴ベクトル 1年周期で折りたたみ + 正規化



頻度ベクトル f_{r_i} を次元数 C に折りたたむ
 1年周期の場合は $C=366$

$$v_{r_i,C} = (v_{r_i,C}^{(0)}, v_{r_i,C}^{(1)}, \dots, v_{r_i,C}^{(C-1)})$$

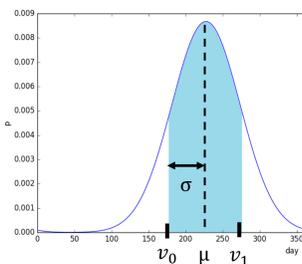
$$v_{r_i,C}^{(j)} = \frac{\sum \{f_{r_i}^{(t)} \mid \text{fold}(t,C) = j\}}{\sum \{f_{r_i}^{(t)} \mid \text{fold}(t,C) = j\}}$$

$\text{fold}(t,C)$: f_{r_i} の時刻 t を $v_{r_i,C}$ の $0 \sim C-1$ に折りたたむ関数
 1年: $0 \sim 365$, 1ヶ月: $0 \sim 30$, 1週間: $0 \sim 6$
 $\text{sum}(S)$: 集合の要素の値の総和を求める関数

時間的状況の特徴ベクトル作成

時間的状況を表す言語表現 (例: 夏, バレンタインデー) → 特徴ベクトル

時間パラメータ + 確率密度関数



固有の確率分布を持つと仮定

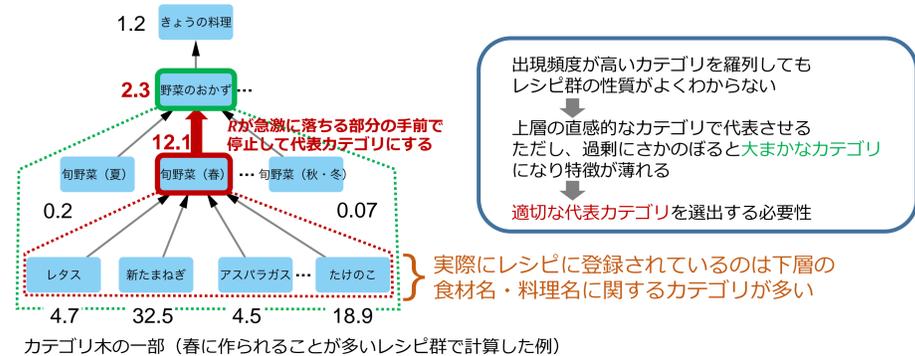
- イベント性と季節性：単峰性の正規分布
- 時間的特性がない：一様分布

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 $v_0 = v_1$

検索結果の要約

- レシピに登録されたカテゴリ情報を用いて、検索結果にどのようなレシピがあるのかひと目で理解できるように要約する
- カテゴリは木構造で、合計1100個に細分化されている

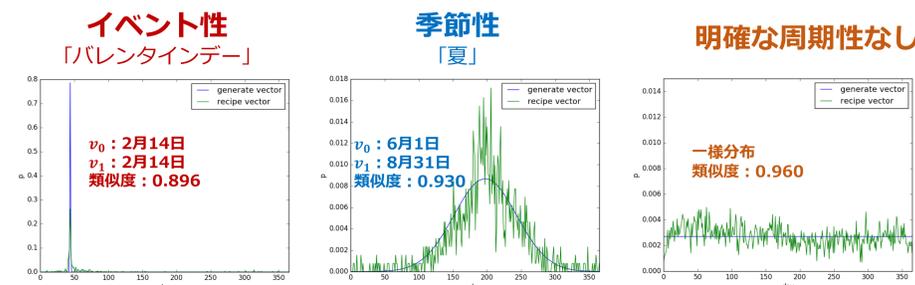


$$R = \frac{\text{あるレシピ群におけるカテゴリ部分木のレシピ存在確率(実確率)}}{\text{ランダムと仮定した場合のカテゴリ部分木のレシピ存在確率(予測確率)}}$$

R が1より大きい→そのカテゴリ部分木にレシピが集中している

検索結果の評価

- 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
- つくれば数100件以上、つくれば投稿期間366日以上以上のレシピ11,507件を抽出



類似度を降順にした場合の上位200件のレシピの代表カテゴリ (レシピ数が多い順に上位5件)

カテゴリ	数	カテゴリ	数	カテゴリ	数
チョコレートのお菓子	114	旬野菜 (夏)	92	おもてなし料理	25
もっと料理を楽しむ	48	夏に食べたい料理	16	人気のキャベツレシピ	13
おもてなしデザート	9	麺	12	キャベツ	11
持ち寄り・プレゼント	2	冷製・アイデアパスタ	7	脇役野菜・珍しい野菜	10
シリコンスチーマーでお菓子	1	夏! おもてなし料理	4	じゃがいも	10

キーワード検索との比較

代表カテゴリの比較 (レシピ数が多い順に上位5件)	提案手法	キーワード検索	
旬野菜 (夏)	92	旬野菜 (夏)	48
夏に食べたい料理	16	夏に食べたい料理	21
麺	12	麺	11
冷製・アイデアパスタ	7	夏! おもてなし料理	11
夏! おもてなし料理	4	豆腐	7

共通レシピは38件(19%)で大きく異なる

- レシピや食材の季節性は背景知識として共有されるためレシピ文章に必ずしも明文化されない
- 提案手法では調理者の視点、キーワード検索ではレシピ作者の視点が反映されている

参考文献

- [1] 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).
- [2] 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, DBSJ Japanese journal=日本データベース学会和文論文誌, Vol.13, No.1, pp.78-85(2014).

連絡先

桐本 宙輝 (和歌山大学大学院システム工学研究科)
 E-mail: s171063@center.wakayama-u.ac.jp