

レストラン検索サイトの メニュー名の統計的分析による 食の地域性の考察

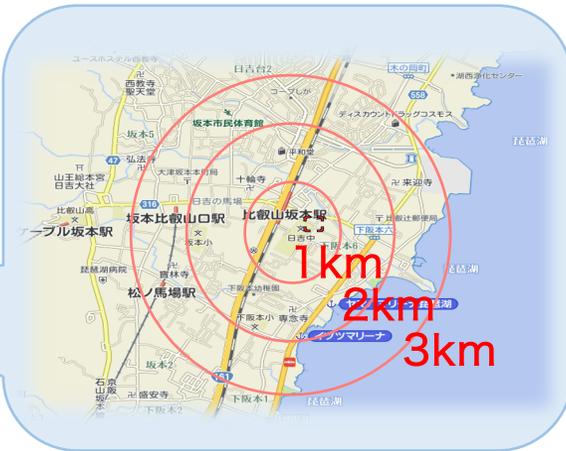
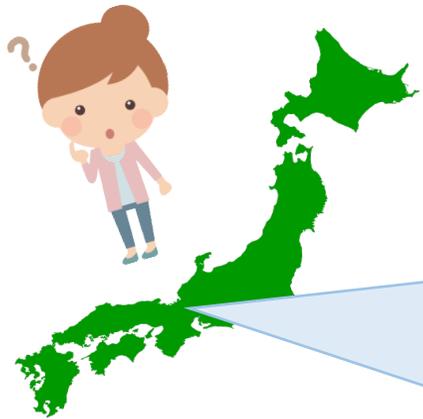
立命館大学情報理工学部

田中優里 山西良典 西原陽子

概要

<レストラン検索サイトを利用するとき>

- ・旅行や出張で普段行かない地域に行くとき
- ・誰かと食事に行くとき

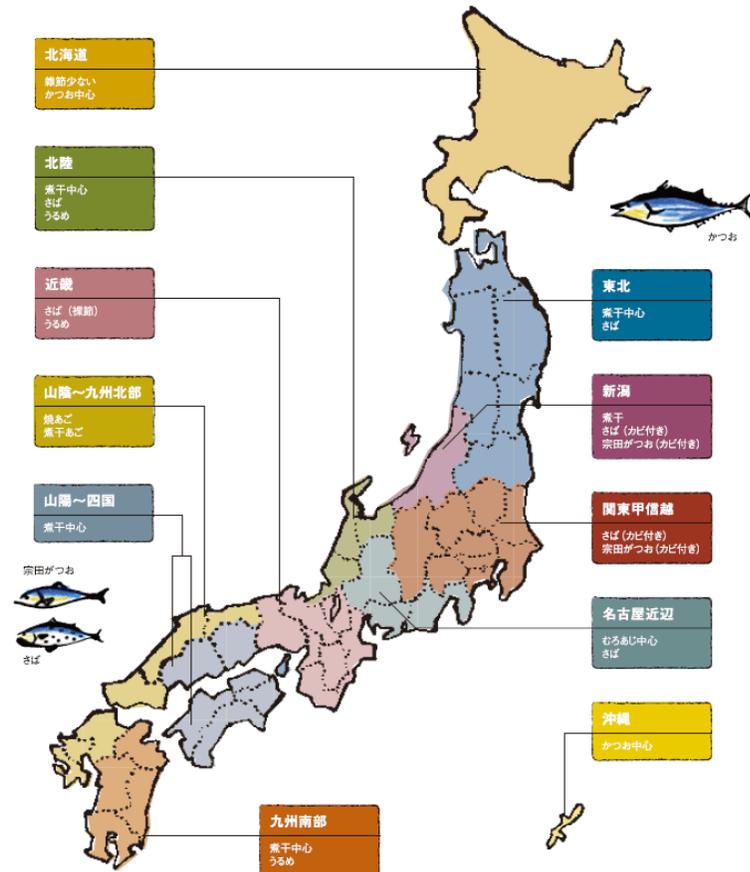


料理, 調理方法, 食材

地域で特色のある食の地域性がわかる！
⇒各地域の特色食材・料理の推薦が可能

背景と目的

<全国の出汁マップ>



<日清食品での活用例>

<東日本>
出汁：かつお
つゆの色：濃い



<西日本>
出汁：昆布
つゆの色：薄い

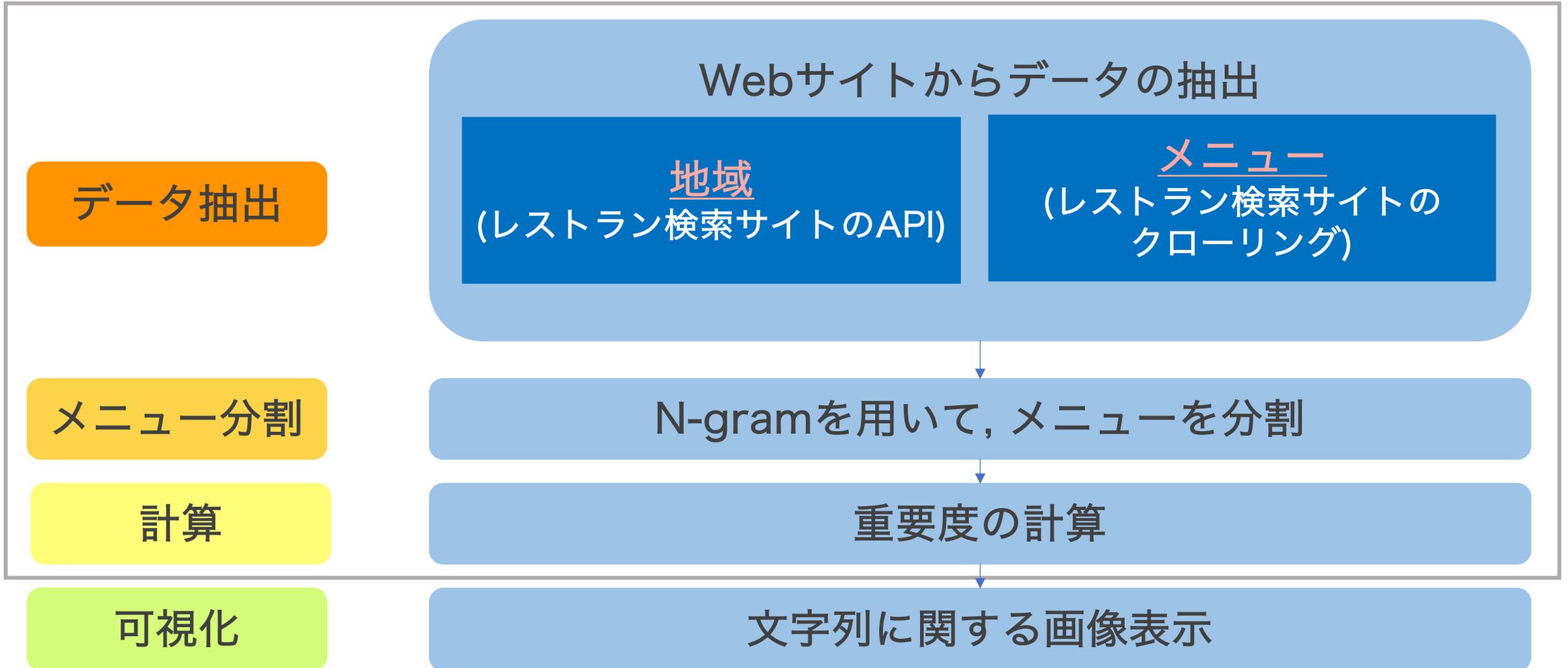
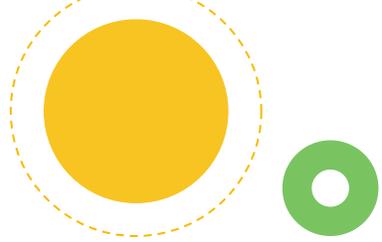


- ・ 「食」の選択
- ・ 外食産業の出店時のナビゲーション
- ・ 外国人観光客へのおもてなし
- ・ 食文化の継承

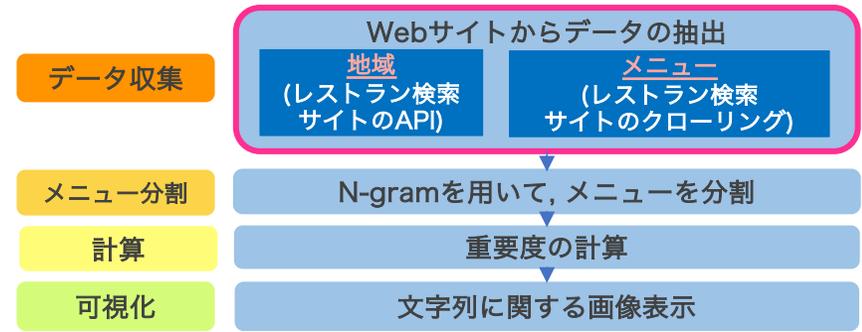


提案手法

提案手法



提案手法 - データ抽出 -



地域



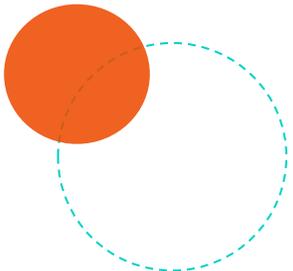
レストラン検索サイトのAPI

メニュー



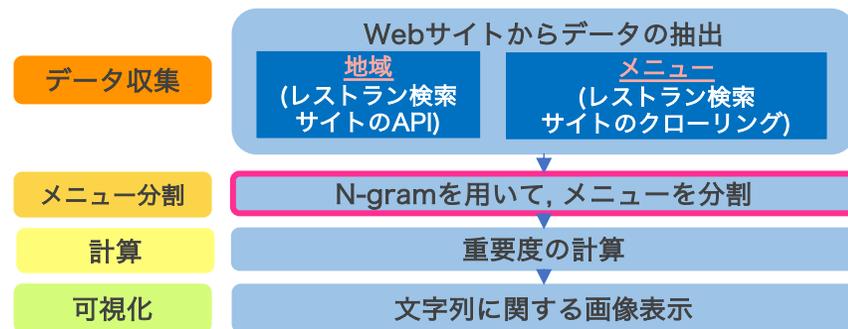
レストラン検索サイトからクローリング

地域情報とメニュー情報を抽出し、結びつける



提案手法

-メニュー分割-

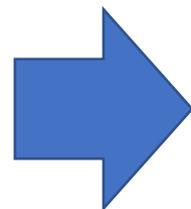


<N-gram：テキスト分割方法>

例) オムライス

- ・ N=2 の場合 (bi-gram)
 - オム
 - ムラ
 - ライ
 - イス
- ・ N=3 の場合 (tri-gram)
 - オムラ
 - ムライ
 - ライス
- ・ N=4 の場合
 - オムライ
 - ムライス
- ・ N=5 の場合
 - オムライス

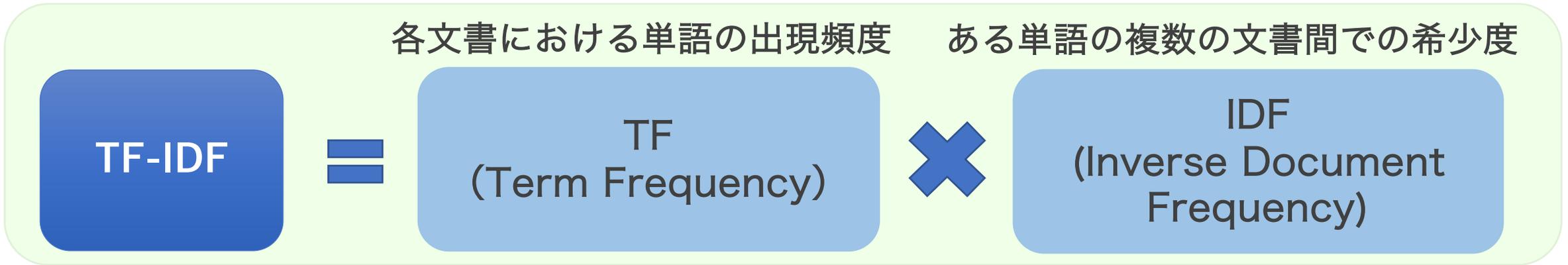
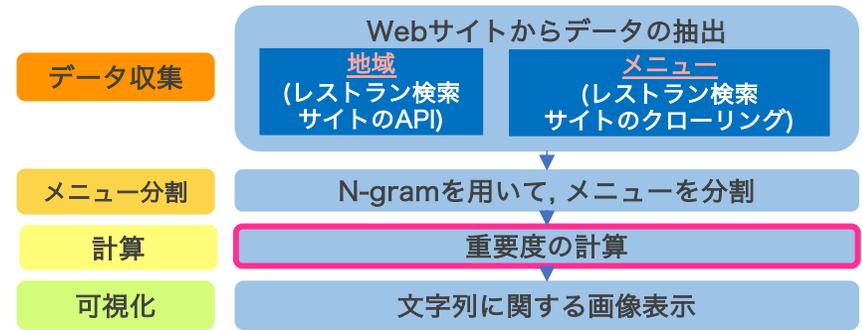
⋮



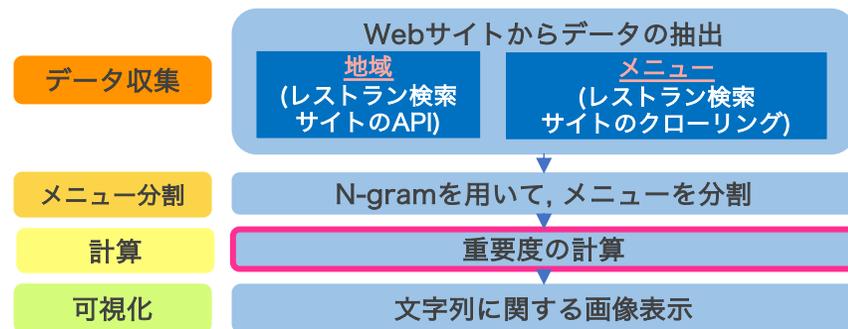
- ・ とろとろ卵のオムライス
- ・ 石焼きタコライス
- ・ マカオ風チキンライス
- ・ 自家製カレーライス
- ・ 月見オニオンスライス
- ・ ライスコロケ
- ・ サフランライス
- ・ Thai風ガパオライス

N-gramを全メニューに対して行うことで、
様々なメニューの統一を実現

提案手法 -重要度計算-



提案手法 -rfとilfの算出方法-



地域 a における文字列 t の出現レストラン RF :

$$rf_a(t) = \frac{n_a(t)}{\sum_{i \in T_a} n_a(i)}$$

$n_a(t)$: 地域 a において
文字列が出現したレストラン数

T_a : 地域 a に出現する文字列の集合

文字列 t の出現レストラン数 LF の逆数 :

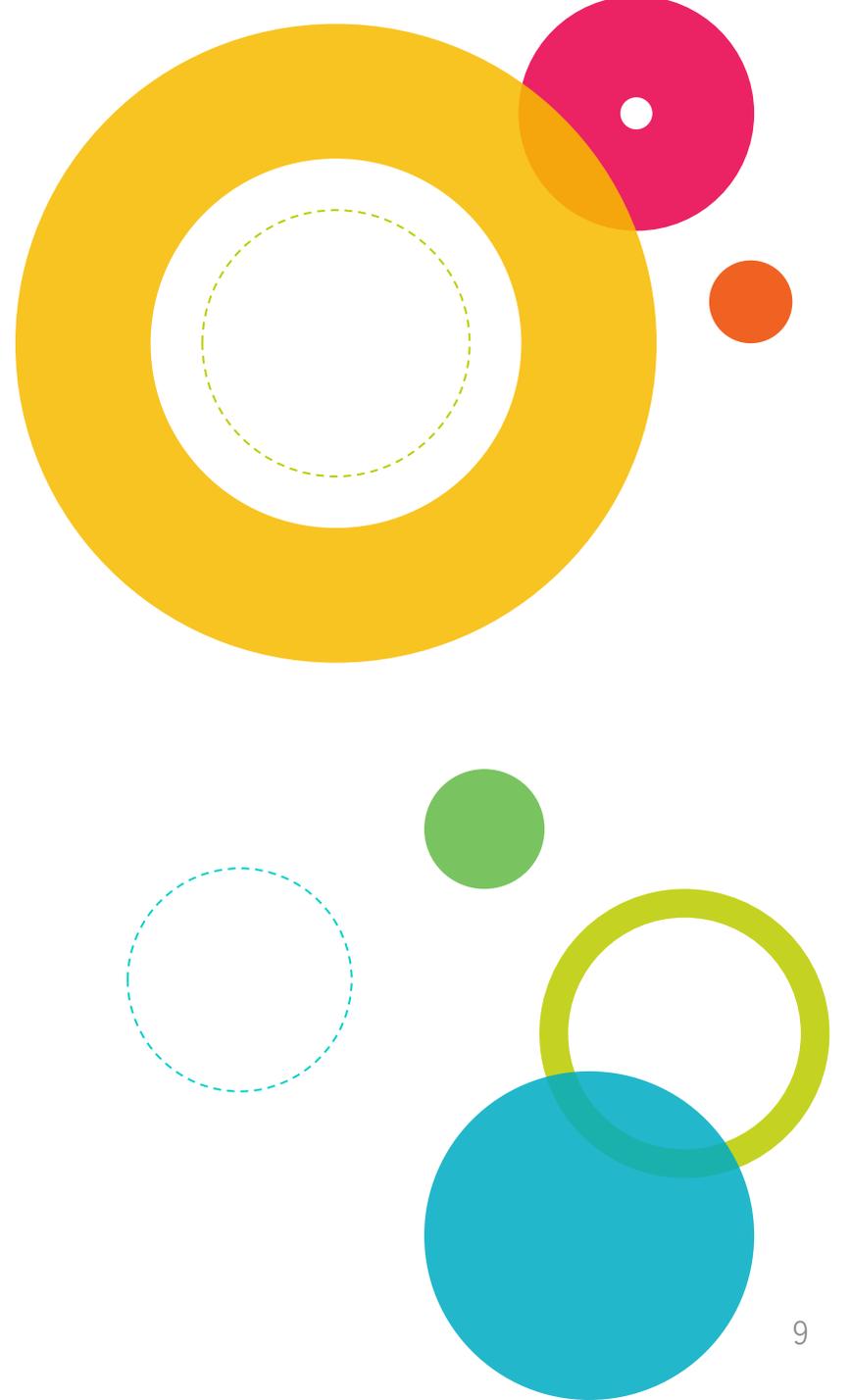
$$ilf(t) = \log \frac{N}{lf(t)} + 1$$

$lf(t)$: 文字列 t が出現した地域数

N : 全地域数

地域性を表す指標
(重要度) :

$$rf \cdot ilf_a(t) = rf_a(t) \times ilf(t)$$



実験と考察

実験の概要

<Webサイト>



Hotpepperグルメ

<分析対象>

- 地域：過去に，情報処理学会と人工知能学会の全国大会が開催された**63地域**
- メニュー：各地域で**上位100件**のレストランの全メニュー

<内容>

- 算出した $rf \cdot ilf$ 値をもとに考察を行う

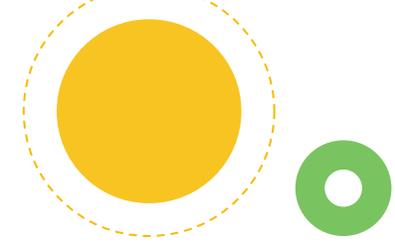
「 $rf \cdot ilf$ 値を用いて、地域で特色のある**食の地域性を抽出できるか**」

考察対象 (地域)

| 都道府県 | 地域 | 都道府県 | 地域 | 都道府県 | 地域 |
|------|-------------------|------|------------------|------|------------|
| 東京 | 御茶ノ水 | 徳島 | <u>徳島駅</u> | 東京 | 東小金井 |
| 東京 | 神保町 | 富山 | 富山駅 | 東京 | 八王子みなみ野 |
| 島根 | <u>松江駅</u> | 福岡 | <u>博多駅 (博多口)</u> | 東京 | 八王子市その他 |
| 山口 | 山口駅 | 福岡 | 福岡市東区 | 東京 | 永田町 |
| 東京 | 表参道 | 福岡 | 小倉・平和通駅・魚町銀天街 | 東京 | 赤羽橋 |
| 東京 | 田園調布 | 長崎 | 浦上 | 東京 | 半蔵門 |
| 東京 | 新宿西口 | 大分 | 府内町・大手町・金池 | 神奈川 | 川崎市多摩区その他 |
| 東京 | 北千住 | 宮崎 | 宮崎駅 | 神奈川 | 横浜市港北区その他 |
| 東京 | <u>大岡山</u> | 鹿児島 | <u>鹿児島中央</u> | 埼玉 | 高坂 |
| 千葉 | 津田沼 | 新潟 | <u>新潟駅前</u> | 千葉 | 野田市その他 |
| 神奈川 | 湘南台 | 石川 | <u>金沢駅</u> | 滋賀 | <u>草津市</u> |
| 愛知 | <u>名古屋駅</u> | 京都 | 北野天満宮 | 京都 | 出町柳 |
| 鳥取 | <u>鳥取駅</u> | 北海道 | 函館駅 | 京都 | 宝ヶ池 |
| 愛知 | 千種 | 北海道 | 五稜郭 | 京都 | <u>京田辺</u> |
| 愛知 | 鶴舞 | 福島 | <u>会津若松駅</u> | 大阪 | 樟葉 |
| 北海道 | 札幌駅 | 岩手 | <u>盛岡駅</u> | 愛知 | 豊田市その他 |
| 北海道 | 新札幌 | 東京 | 早稲田 | 宮城 | 太白区その他 |
| 北海道 | <u>旭川市中心</u> | 東京 | 西早稲田 | 広島 | 五日市 |
| 宮城 | 青葉区 | 東京 | 目白 | 山口 | 湯田温泉 |
| 茨城 | <u>つくば駅北・筑波大学</u> | 東京 | 後楽園 | 愛媛 | <u>松山駅</u> |
| 香川 | 高松駅・北浜 | 東京 | 船堀 | 福岡 | 戸畑 |

➤ 日本の各地区（北海道，東北，関東，中部，近畿，中国，四国，九州・沖縄）に分散するように選択した都市の合計16地域

➤ 16地域は，表中の赤文字下線

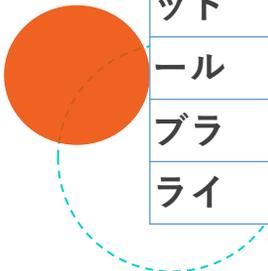


考察対象 (文字列)

| メニュー名の N-gram | メニュー名の N-gram | メニュー名の N-gram | メニュー名の N-gram |
|---------------|---------------|---------------|---------------|
| 焼き | イス | くら | カレー |
| サラ | ージ | 炒め | サラカ |
| ラダ | リー | チャ | マサラカ |
| サラダ | ポテ | ソー | サラカレ |
| チー | テト | ック | マサラカレー |
| ーズ | トマ | バー | 海老 |
| 揚げ | モン | ーキ | フラ |
| チーズ | 鳴門 | 野菜 | ステ |
| ース | カル | うどん | テー |
| 盛り | ルビ | うどん | ーキ |
| レー | カルビ | そば | ステーキ |
| ット | ムチ | チキ | 和牛 |
| ール | キム | チキン | コース |
| ブラ | キムチ | カレ | |
| ライ | いか | | |

➤ 全63地域の中から、
 $rf \cdot ilf$ 値が上位10件のN-gram,

16地域の中から、
 $rf \cdot ilf$ 値が上位5件ずつの
N-gramから得られた文字列



結果と考察① (文字列ごと)

文字列の重要度

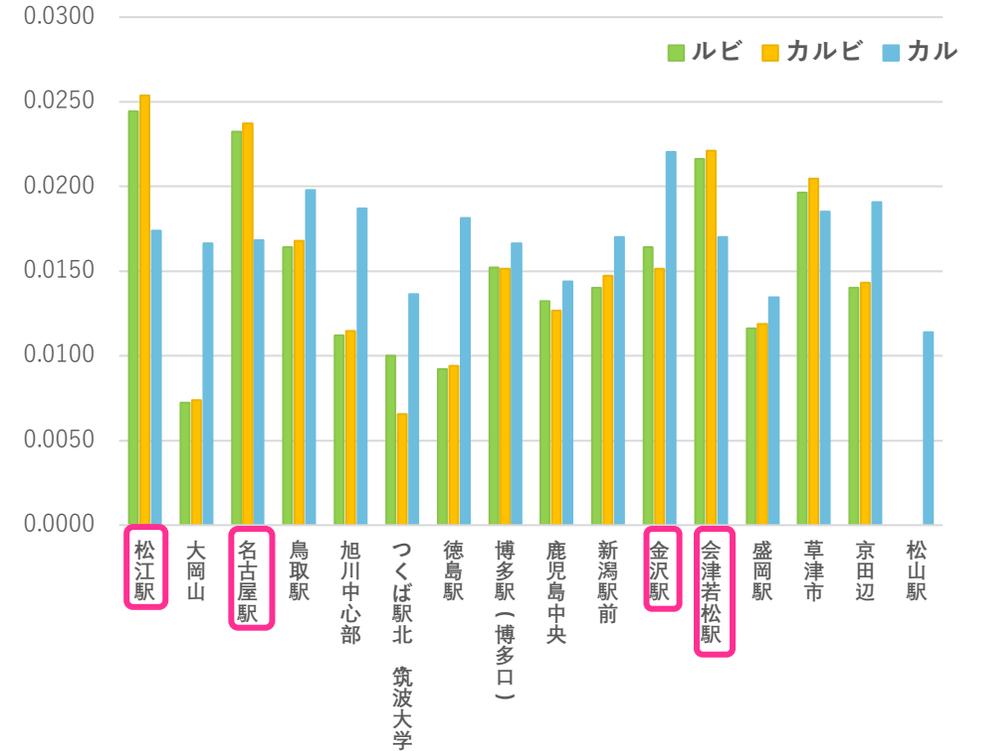
| | 島根県 | 愛知県 | 石川県 | 福島県 |
|------------------|--------|--------|--------|--------|
| メニュー名の N-gram | 松江駅 | 名古屋駅 | 金沢駅 | 会津若松駅 |
| ルビ | 0.0244 | 0.0232 | 0.0164 | 0.0216 |
| カルビ | 0.0254 | 0.0237 | 0.0151 | 0.0221 |
| カル | 0.0174 | 0.0168 | 0.0220 | 0.0170 |

“ルビ”, “カルビ”

- ・ 松江駅, 名古屋駅, 会津若松駅の重要度が同程度
 - “ルビ” ⇒ “カルビ”

“カル”

- ・ 金沢駅
 - 「カルパッチョ」, 「エスカルゴ」, 「マスカルポーネ」



文字列の共起確率について
考慮する必要あり

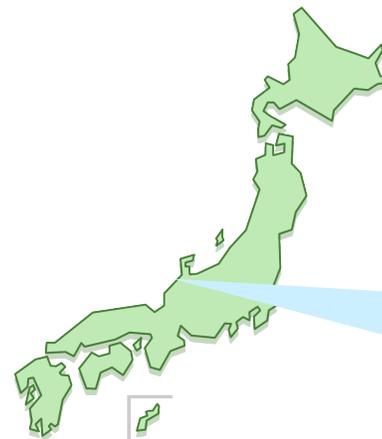
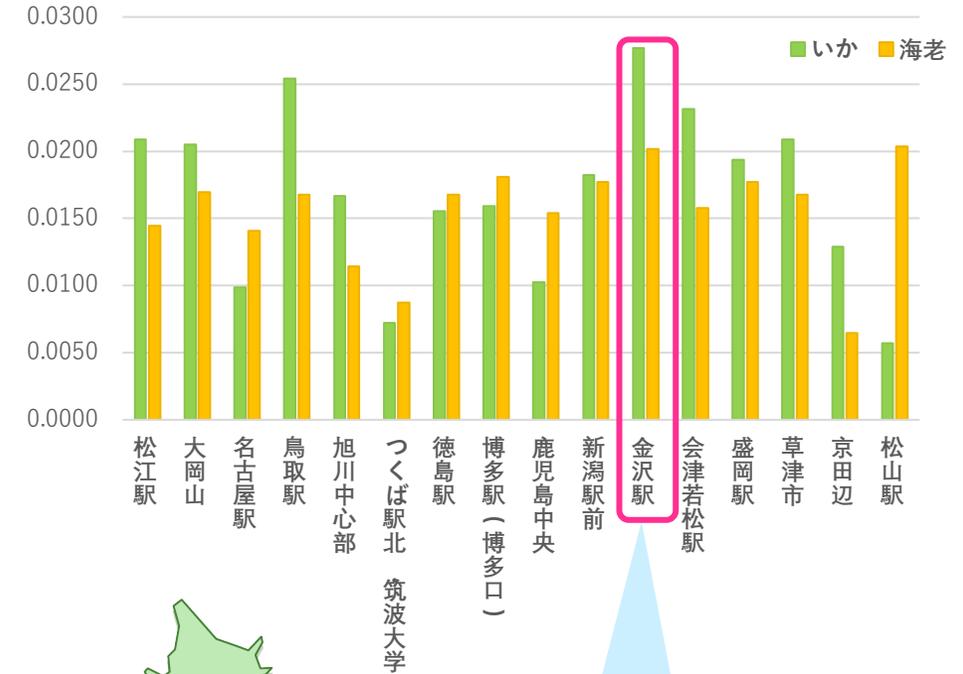
結果と考察②（食材）

文字列の重要度

| | 島根県 | 愛知県 | 石川県 | 福島県 |
|------------------|--------|--------|--------|--------|
| メニュー名の N-gram | 松江駅 | 名古屋駅 | 金沢駅 | 会津若松駅 |
| いか | 0.0209 | 0.0099 | 0.0277 | 0.0231 |
| 海老 | 0.0145 | 0.0141 | 0.0202 | 0.0158 |

“いか”，“海老”

- ・石川県は、日本海側に位置している
 - 「カルパッチョ」などの海鮮を使ったメニューも存在
 - 他にも、「ふぐ」「のどぐろ」などの海鮮類がよく使われている

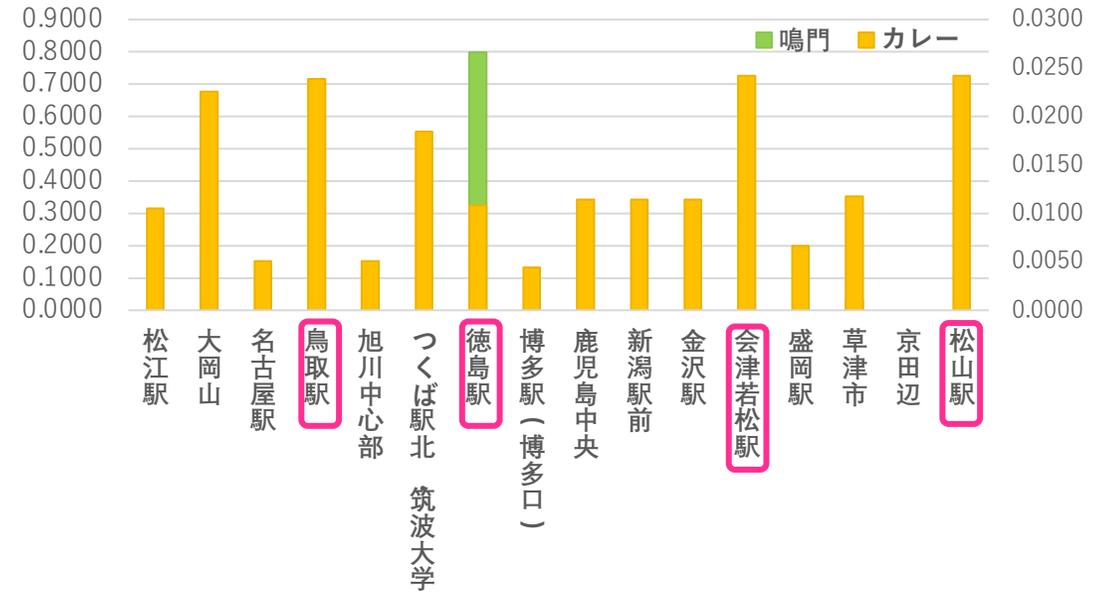


海鮮類が多い

結果と考察③ (地名・料理名)

単語の重要度

| | 鳥取県 | 徳島県 | 福島県 | 愛媛県 |
|------------------|--------|--------|--------|--------|
| メニュー名の N-gram | 鳥取駅 | 徳島駅 | 会津若松駅 | 松山駅 |
| 鳴門 | 0.0000 | 0.7989 | 0.0000 | 0.0000 |
| カレー | 0.0239 | 0.0108 | 0.0242 | 0.0242 |



“鳴門”

- 徳島県には地名の「鳴門」がある
- 「鳴門金時」
「鳴門わかめ」
「鳴門蓮根」
「鳴門鯛」などの地場産品が多い

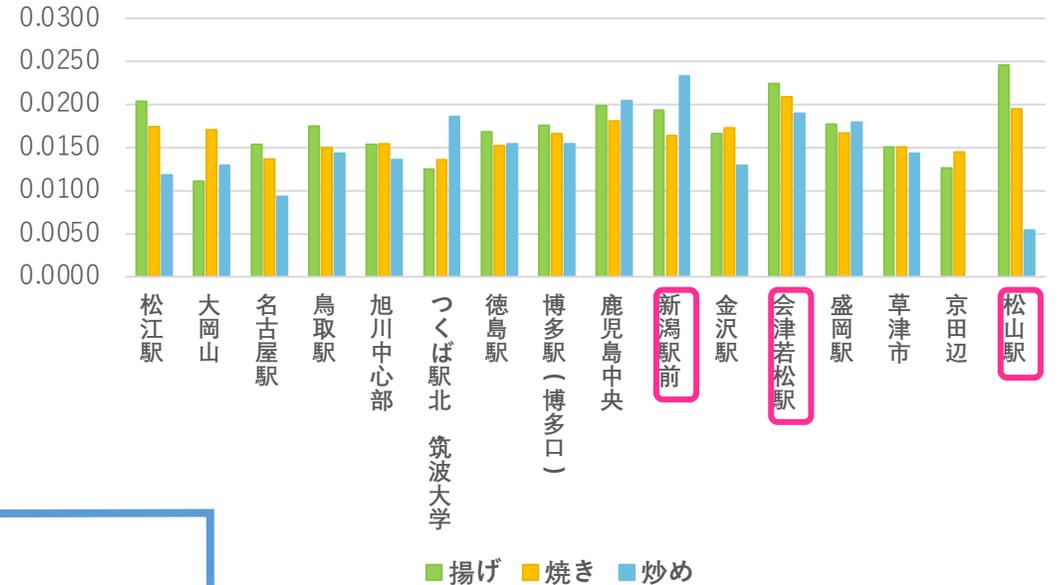
“カレー”

- 鳥取市
 - 1世帯あたりのカレー年間支出金額および購入数量で日本1位※
 - カレーを食べる機会が多い
 - 会津若松駅
 - 「会津カレー」のようなご当地のカレーメニューが存在
 - 松山駅
 - 「カレー」との深い関連は見られなかった
- ※2014年～2016年平均

結果と考察④ (調理方法)

単語の重要度

| メニュー名の N-gram | 新潟県 | 福島県 | 愛媛県 |
|------------------|---------------|---------------|---------------|
| | 新潟駅前 | 会津若松駅 | 松山駅 |
| 揚げ | 0.0193 | 0.0225 | 0.0246 |
| 焼き | 0.0164 | 0.0209 | 0.0195 |
| 炒め | 0.0233 | 0.0190 | 0.0054 |



“揚げ”

- 松山駅
- 「揚げ鶏」「竜田揚げ」「揚げごぼう」「揚げバナナ」などの様々な揚げ物メニューが存在

“焼き”

- 会津若松駅
- 「会津地鶏 レバー串焼き」「会津地鶏の玉子焼き」「【会津B級グルメ】カレー焼きそば」など、地鶏やカレーを使ったメニューが存在

“炒め”

- 新潟駅前
- 「海老と野菜のあっさり塩炒め」「ベビーホタテのガーリックバター炒め」など海鮮類を使用したものが多い
- 様々な味付けの中でも、ガーリックやバターを使用したものが多い

まとめと展望

<結論>

地域ごとに特色のある食の地域性（料理，調理方法，食材）を抽出できた。
⇒各地域の特色食材・料理の推薦が可能

<今後の展望>

- 63 地域をさらに広範囲へ広げた様々な地域（国内・海外）での分析
- N-gram ではなく形態素解析によってメニュー分割した分析
- 文字列の共起確率を用いた，メニューの統一
- 客観的評価の方法の検討