

テキストマイニングによる食レポ口コミからのユーザー嗜好モデリング

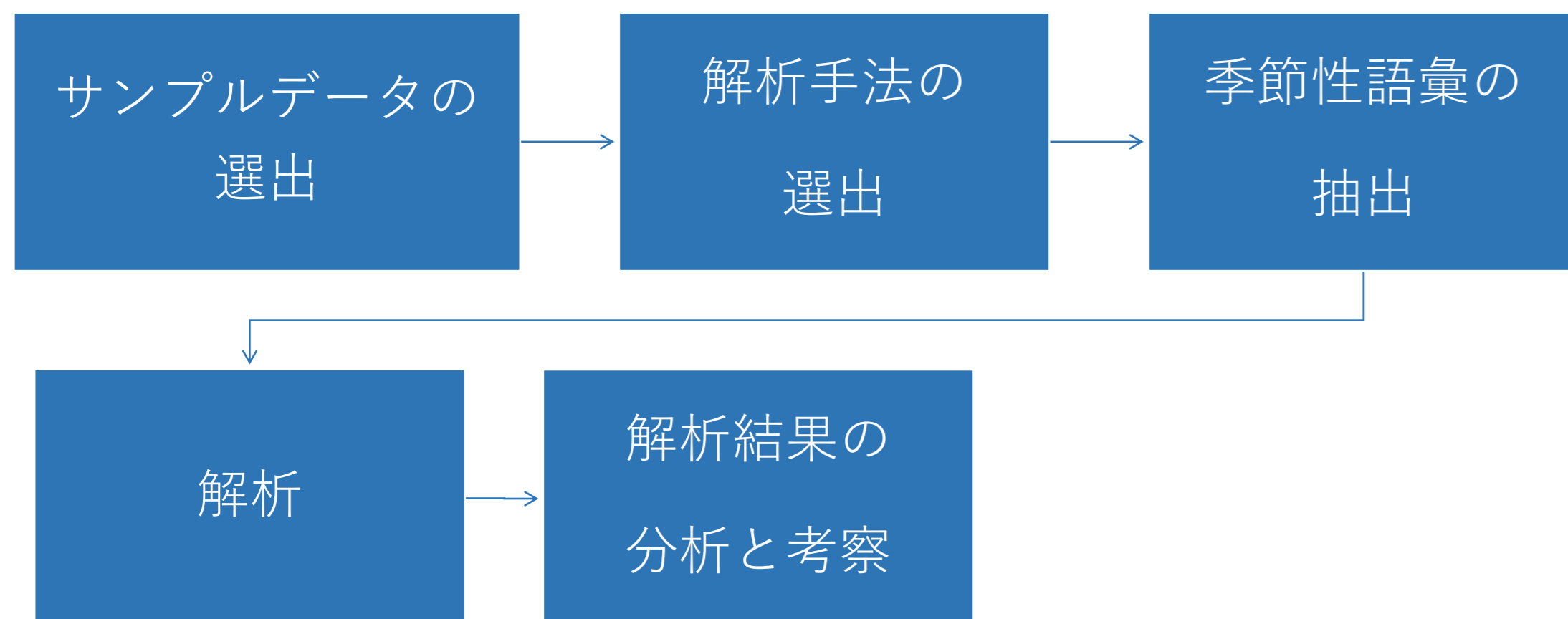
経営システム工学科 2年 吉田妃那 石黒董 小島幸奈 酒井あかり
指導教員 上原宏 教授

Introduction

新しい視点からの検索機能！

全国の飲食店の情報と共にユーザーの口コミが掲載されているサイト「食べログ」では口コミ件数順などのソート機能や地域を絞る情報フィルタリング機能が実装されており、ユーザーが求めている店舗を比較的見つけやすくなっている。しかし、食べログのサイトには季節ごとに店舗を検索する機能がない。本研究では既存の視点とは異なる四季という視点から旬にこだわっている飲食店や季節性の高い食べ物（鍋、おでんなど）が楽しめる店舗を提案する方法の有用性を示すことを目的としている。また、洋食と和食で季節を特徴づける単語に差があるのかという検証も本研究の目的である。

研究の流れ



Method

「食べログ」からの口コミ収集

食べログから洋食和食それぞれ6店舗、多ジャンル・様々な地域の店舗から収集。

洋食				和食			
店舗	ジャンル	地域	件数	店舗	ジャンル	地域	件数
ア・ニエルトウルヴェ・ヴー	フレンチ	東京都	436	茶茶白雨	京料理	東京都	284
レフェルヴェソンス	フレンチ	東京都	491	草喰 なかひがし	京料理	京都府	336
クッチーナイタリアーナガ	イタリアン	愛知県	237	銀座 しまだ	割烹料理	東京都	349
カーザ ヴィニタリア	イタリアン	東京都	354	島之内 一陽	割烹料理	大阪府	201
SALONE 2007	イタリアン・フレンチ	神奈川県	415	虎白	懐石料理	東京都	242
カセント	スペイン料理	兵庫県	299	稚加榮 本店	懐石料理	福岡県	481
計			2232	計			1893

Pythonを使ったプログラミング解析

口コミの本文を分析するにあたり、以下の手順を行った。

クローラーによる口コミデータの収集

MeCabによる口コミ本文の形態素解析

Word2VecやJuman意味フィルターを用いて
季節性のある重要な言葉を抽出

K-meansによるクラスタリングと主成分分析

※Jumanとは、日本語形態素解析システムであり、文章を一つ一つの単語列に区切った際に意味カテゴリ名やドメイン名が付与される。これらを用いてフィルターをかけることによって、重要な語句のみを抽出することができる。

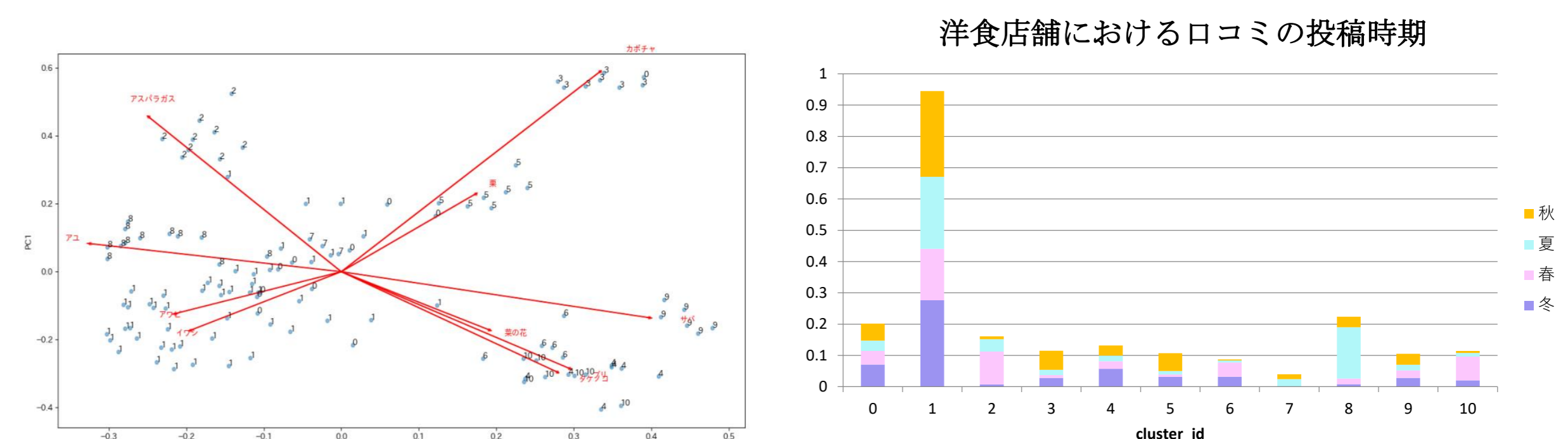
※word2vecとは、各単語はその周辺の語彙と何らかの関係性を持つという仮定に基づいて単語の特徴をベクトル化することのできるモデル。ベクトルの成分は、文書の特徴づける単語を抽出するために単語の重要度を測る値であるtf-idfを用いている。

Result

k-meansによるクラスタリングはクラスタ数を11個、主成分平面は第3主成分、第4主成分を軸として分析を行った。また、口コミが投稿された日時をもとにして口コミを春夏秋冬に分類している。ここでは12~2月を冬、3~5月を春、6~8月を夏、9~11月を秋として以下のグラフに表示している。

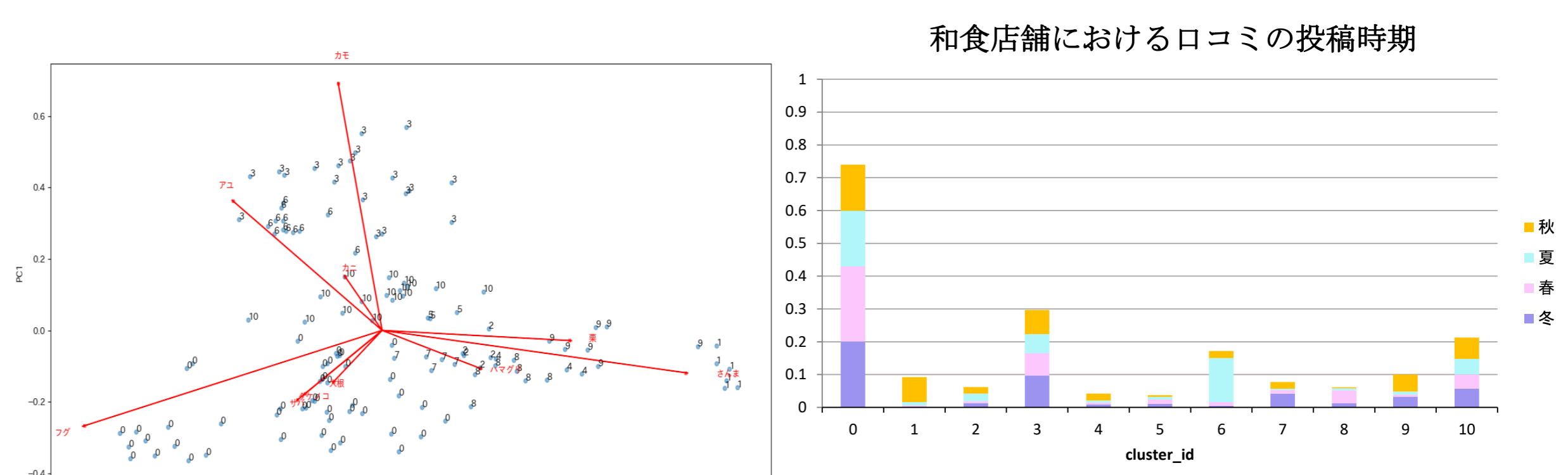
洋食

右下のグラフより、クラスタ2、クラスタ10は春、クラスタ8は夏と秋、クラスタ3、クラスタ4、クラスタ5は秋と冬、クラスタ6は春と冬に投稿された口コミが多いことが分かる。実際に、クラスタのtf-idfの値の高い単語を見ると、クラスタ2はアスパラガスや空豆や桜海老、クラスタ10はブリや桜や春菊、クラスタ8はアユやスイカやキュウリ、クラスタ3はセロリ、4はブリや大根、5は栗、6は菜の花やカリフラワーなど、口コミが投稿された時期と一致している旬の食材の名前が多く挙がっていた。



和食

右下のグラフを見ると、クラスタ7、クラスタ8は春と冬、クラスタ6は夏、クラスタ1は秋、クラスタ9は秋と冬に投稿された口コミが多いことが分かる。実際に、クラスタのtf-idfの値の高い単語を見ると、クラスタ7は菜の花やミカン、クラスタ8はハマグリや菜の花や桜、クラスタ6はトウモロコシやアユ、クラスタ1はさんま、クラスタ9は栗やハウレンソウなど洋食同様、口コミが投稿された時期と一致している旬の食材の名前が多く挙がっていた。



Consideration

季節と口コミの語彙、洋食と和食の関係

クラスタを特徴づける語彙の季節性と口コミが投稿された季節が一致しているとき、そのクラスタに多く属している店舗は旬の食材にこだわり、季節感の強い料理を提供する店舗であるといえるだろう。その一方で、季節という基準では上手く分類されなかったクラスタがある。これは近年、栽培環境や生産技術の向上、海外からの輸入などによって通年食べられる食べ物が増えていることが影響しているのではないかと考えられる。また、和食と洋食では分析結果を比較すると、季節感を感じさせる単語にはそれほど大きな差はなかった。しかし、主成分分析のグラフについては両方の概形の差が大きいため、一致したとは言い切れない。今回の分析結果から利用者の季節感に近い料理を提供する店舗の提案や、洋食と和食の互換性をもった提案を第三者に提供するまでには至らなかったが、さらなる研究を重ねることでこれらは可能になると考えられる。