

---

2020年  
NTTデータ数理システム  
学生奨励賞応募資料

# 食卓日記データを用いた 季節ごとの特徴語に関する分析

中央大学大学院経営システム工学専攻  
北嶋弓月



# 目次

---

- 背景
- 目的
- 使用データ
- 分析
  - 特徴語抽出
  - ことばネットワーク
- 結果
  - 各季節の特徴語
  - 各季節の単語ネットワーク
- 考察
  - 季節に適したレコメンド単語の提案
- まとめ

# 背景

- 個人の感性
  - 個人の価値観
- ➔ 購買行動に影響をもたらす

## ➤ 個人の感性・感情に着目

### ジョブ理論

顧客(ユーザー)が**行動**するに至った**目的(ジョブ)**を明らかにするという理論



# ジョブについて

ジョブ理論

言語処理

言語処理を用いて、文書内の単語から**ジョブ**を明らかにする

ジョブ = 日常生活上のイベント

- ・ 運動会
- ・ 冠婚葬祭
- ・ 年末
- ・ クリスマス

周期的なジョブ

- ・ 風邪
- ・ 体調
- ・ 予定

突発的なジョブ

- ・ 気分
- ・ 感情

潜在的なジョブ

## 目的

- 商品の購買行動は、消費者のジョブ（イベント）の影響を受けている
- ジョブ（イベント）には、周期的/突発的/潜在的 なイベントが存在
- テキストデータはユーザ自身が文章で記録を行っている



季節による購買行動の変化に着目

### 目的

消費者に対して、季節ごとにどのようなレコメンドを行うことが効果的なのか

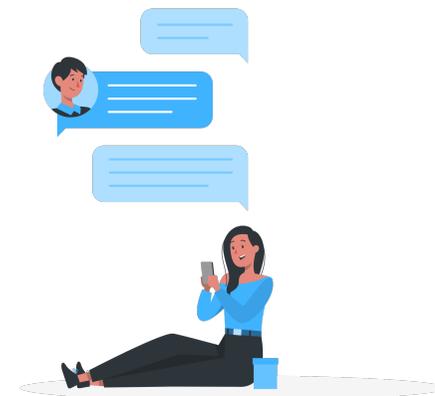
# 言語処理について

## 言語処理

人間が日常的に使用している言葉を、コンピュータを用いて処理

(今回扱うデータで実際に記されている文章の1例)

「先日スーパーで買ってきた高野豆腐を利用して、…」



先日	スーパー	で	買っ	て	き	た	…
名詞	名詞	助詞	動詞	助詞	動詞	助動詞	…

文章における特徴的な単語や、単語間の関係性を把握することが可能

## 使用データに関して

- 本分析内では、企業から提供いただいた「食卓日記データ」を用いて分析を行う。

使用データ	生活行動日誌データ
データ取得期間	1990年～2014年
使用カラム	背景データ, 材料データ

- また、本分析では、NTTデータ数理システム提供の「Visual Mining Studio」と「Text Mining Studio」を用いて分析を行なった。

## 分析の流れ

本分析の流れは以下の通りである。

```
graph TD; A[日記データの形態素解析] --> B[季節ごとにおける特徴語抽出]; B --> C[季節ごとにおける特徴語の共起ネットワーク作成]; C --> D[考察];
```

日記データの形態素解析

季節ごとにおける特徴語抽出

季節ごとにおける特徴語の共起ネットワーク作成

考察

## 季節ごとの特徴語抽出

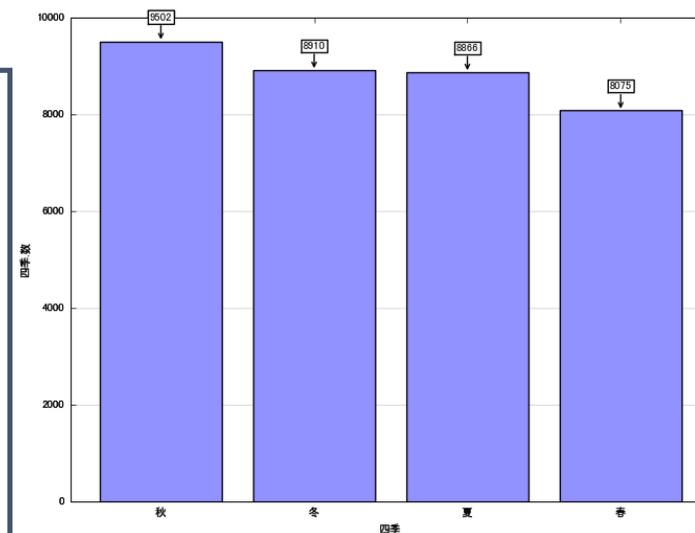
- 季節ごとのレコメンドを行えるように、春夏秋冬それぞれで、特徴語の抽出を行った。
- 形態素解析は分かち書き + 係り受けを用いた。

### TFIDF

全体の文書中に含まれる単語の重要度を測る指標

$$tf * idf = (\text{単語の出現頻度}) * (\text{各単語の希少度})$$

で表される。



TMS内の「特徴語抽出」で属性を選択し、各季節に対する特徴語を抽出

## 季節ごとの特徴語抽出

各季節に対する特徴単語は以下のように算出することができた。

春 特徴語	指標値 春	夏 特徴語	指標値 夏	秋 特徴語	指標値 秋	冬 特徴語	指標値 冬
竹の子	561.583	暑い	1418.999	買う	899.98	白菜	1116.41
喜	260.202	ナス	918.581	さつまい	633.191	大根	942.042
イチゴ	249.168	トマト	882.835	大根	513.955	餅	660.9
春休み	242.431	キュウリ	754.909	サンマ	409.991	入れる	542
ご飯	234.129	食欲	546.429	スーパー	371.617	葱	505.107
キャベツ	222.866	ソーマン	540.504	行く	355.207	寒い	498.552
お弁当	217.948	哀	529.932	おでん	333.975	作る	408.597
海苔	203.813	ピーマン	493.214	夕食	331.204	鍋	392.748
香り	180.214	夏休み	454.589	白菜	329.546	人参	388.127
ちらし寿司	172.183	食べる	433.413	運動会	311.015	お正月	361.052
米	171.743	暑さ	394.567	人参	308.174	鶏肉	356.712
新玉葱	156.396	怖	361.035	栗	286.014	焼く	348.308
旬	147.314	きゅうり	353.874	お弁当	232.828	椎茸	317.268
混ぜる	144.792	冷やす	353.514	りんご	226.503	風邪	311.655
飲む	143.56	喜	338.019	里芋	223.645	煮る	302.306
アスパラ	141.722	ハム	329.111	じゃが芋	212.904	お餅	287.207
炒める	136.387	食べる+したい	271.435	体	209.722	砂糖	286.267
筍	130.847	酢	267.499	寒い	204.409	豆腐	283.796
ビール	124.834	栄養	253.054	温まる	191.188	鶏卵	280.089
マヨネーズ	123.04	昼食	248.85	こんにゃく	190.619	残る	276.571

# 季節ごとの特徴語抽出

各季節に対する特徴単語は以下のように算出することができた。

春 特徴語	指標値 春	夏 特徴語	指標値 夏	秋 特徴語	指標値 秋	冬 特徴語	指標値 冬
竹の子	561.583	暑い	1418.999	買う	899.98	白菜	1116.41
喜	260.202	ナス	918.581	さつまい	633.191	大根	942.042
イチゴ	249.168	トマト	882.835	大根	513.955	餅	660.9
春休み	242.431	キュウリ	754.909	サンマ	409.991	入れる	542
ご飯	234.129	食欲	546.429	スーパー	371.617	葱	505.107
キャベツ	222.866	ソウメン	540.504	行く	355.207	寒い	498.552
お弁当	217.948			おでん	333.975	作る	408.597
海苔	203.813			夕食	331.204	鍋	392.748
香り	180.214			白菜	329.546	人参	388.127
ちらし寿司	172.183			運動会	311.015	お正月	361.052
米	171.743			人参	308.174	鶏肉	356.712
新玉葱	156.396	仰	361.035	栗	286.014	焼く	348.308
旬	147.314	きゅうり	353.874	お弁当	232.828	椎茸	217.268
混ぜる	144.792	冷やす	353.514	りんご	226.503		
飲む	143.56	喜	338.019	里芋	223.645		
アスパラ	141.722	ハム	329.111	じゃが芋	212.904		
炒める	136.387	食べる+したい	271.435	体	209.722		
筍	130.847	酢	267.499	寒い	204.409		
ビール	124.834	栄養	253.054	温まる	191.188	鶏卵	280.089
マヨネーズ	123.04	昼食	248.85	こんにゃく	190.619	残る	276.571

春休みやお弁当  
→子供のいる家庭が  
強く出ている

秋から寒さを感じ、  
温かいものが人気

# 季節ごとの特徴語抽出

各季節に対する特徴単語は以下のように算出することができた。

春 特徴語	指標値 春	夏 特徴語	指標値 夏	秋 特徴語	指標値 秋	冬 特徴語	指標値 冬
竹の子	561.583	暑い	1418.999	買う	899.98	白菜	1116.41
喜	260.202	ナス	918.581	さつまい	633.191	大根	942.042
イチゴ	249.168	トマト	882.835	大根	513.955	餅	660.9
春休み	242.431	キュウリ	754.909	サンマ	409.991	入れる	542
ご飯	234.129	食欲	546.429	スーパー	371.617	葱	505.107
キャベツ	222.866	ソーマン	540.504	行く	355.207	寒い	498.552
		哀	529.932	おでん	333.975	作る	408.597
		ピーマン	493.214	夕食	331.204	鍋	392.748
		夏休み	454.589	白菜	329.546	人参	388.127
		食べる	433.413			お正月	361.052
		暑さ	394.567			鶏肉	356.712
新玉葱	156.396	怖	361.035			焼く	348.308
旬	147.314	きゅうり	353.874			椎茸	317.268
混ぜる	144.792	冷やす	353.514			風邪	311.655
飲む	143.56	喜	338.019	芋子	223.645	煮る	302.306
アスパラ	141.722	ハム	329.111	じゃが芋	212.904	お餅	287.207
炒める	136.387	食べる+したい	271.435	体	209.722	砂糖	286.267
筍	130.847	酢	267.499	寒い	204.409	豆腐	283.796
ビール	124.834	栄養	253.054	温まる	191.188	鶏卵	280.089
マヨネーズ	123.04	昼食	248.85	こんにゃく	190.619	残る	276.571

暑いという単語や栄養という単語が目立つ

お正月やお餅など反対に寒いなどの単語はない

## 季節ごとの特徴語抽出

各季節に対する特徴単語は以下のように算出することができた。

春 特徴語	指標値 春	夏 特徴語	指標値 夏	秋 特徴語	指標値 秋	冬 特徴語	指標値 冬
竹の子	561.583	暑い	1418.999	買う	899.98	白菜	1116.41
喜	260.202	ナス	918.581	さつまい	633.191	大根	942.042
イチゴ	249.168	トマト	882.835	大根	513.955	餅	660.9
春休み	242.431	キュウリ	754.909	サンマ	409.991	入れる	542
ご飯	234.129	食欲	546.429	スーパー	371.617	葱	505.107
キャベツ	222.866	ソーマン	540.504	行く	355.207	寒い	498.552
お弁当	217.948	哀	529.932	おでん	333.975	作る	408.597
海苔	203.813	ピーマン	493.214	夕食	331.204	鍋	392.748
香り	180.214	夏休み	454.589	白菜	329.546	人参	388.127
ちらし寿司	172.183	食べる	433.413	運動会	311.015	お正月	361.052
米	171.743	暑さ	394.567	人参	308.174	鶏肉	356.712
新玉葱	156.396	怖	361.035	栗	286.014	焼く	348.308
旬	147.314	きゅうり	353.874	お弁当	232.828	椎茸	317.268
混ぜる	144.792	冷やす	353.514	りんご	226.503	風邪	311.655
飲む	143.56	喜	338.019	里芋	223.645	煮る	302.306
<div style="border: 2px solid blue; border-radius: 15px; padding: 10px; display: inline-block;"> <p>季節に関する単語が特徴語として多く現れている</p> </div>							
マヨネーズ	123.04	昼食	248.85	こんにゃく	190.619	残る	276.571

## 共起語の特定

形態素解析を係り受けありの状態で行い、特徴語分析で抽出された単語を用いて、「ことばネットワーク」を作成した。



季節ごとの特徴語と一緒に使われやすい単語の把握

### 条件

- 単語出現頻度2回以上
- 補完類似度を使用



# 共起語の特定

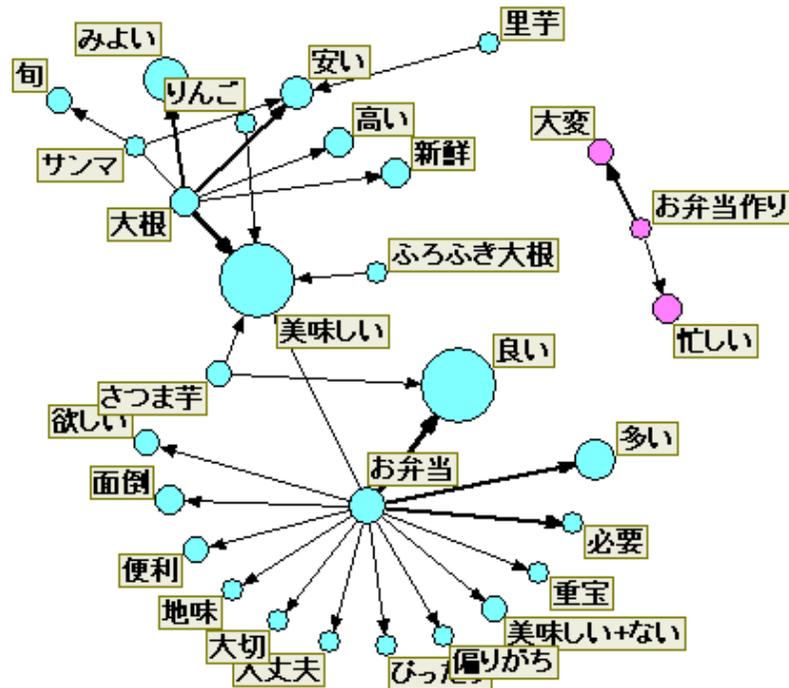


図3 秋の特徴語に対するネットワーク

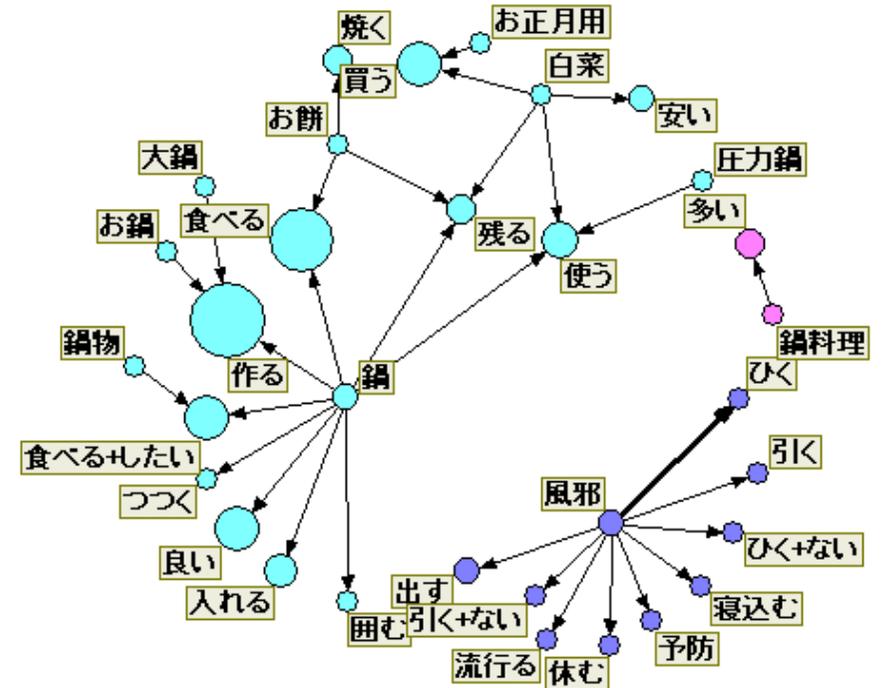


図4 冬の特徴語に対するネットワーク

本分析ではこの結果から、季節ごとの食品販売に関するレコメンド文を考察した。

# 考察—春—

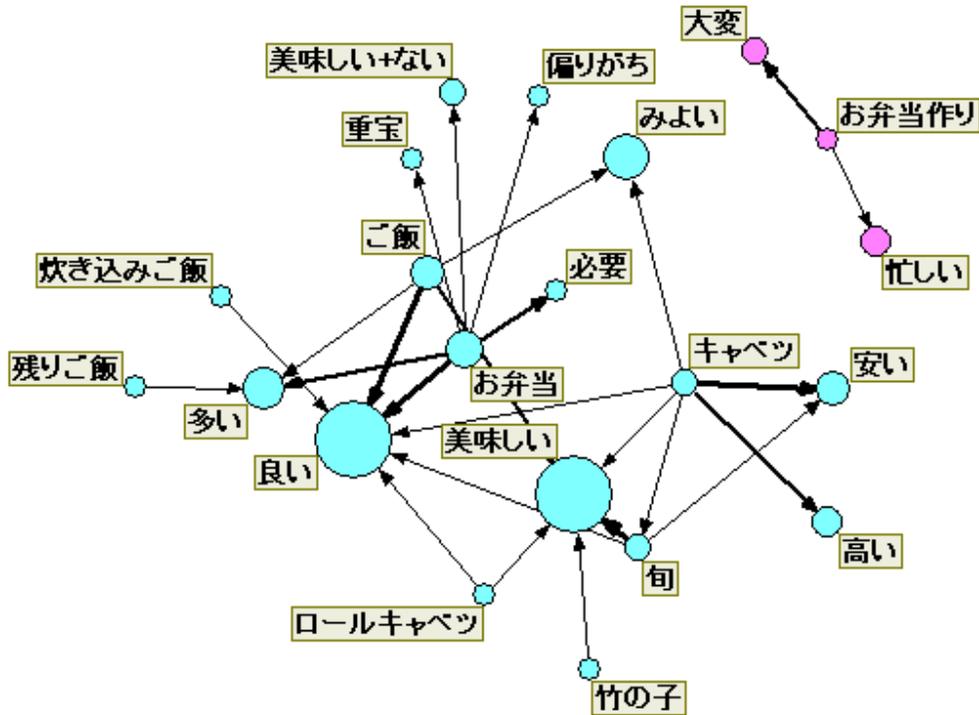


図1 春の特徴語に対するネットワーク

## 春の特徴語とネットワークから 提案するレコメンド文

### お弁当材料に対するレコメンド文

→ 「忙しい人に」  
「簡単調理」  
「偏りがちな」 などの単語を含む

また、食材としては  
「竹の子」「ちらし寿司」などが旬



# 考察—秋—

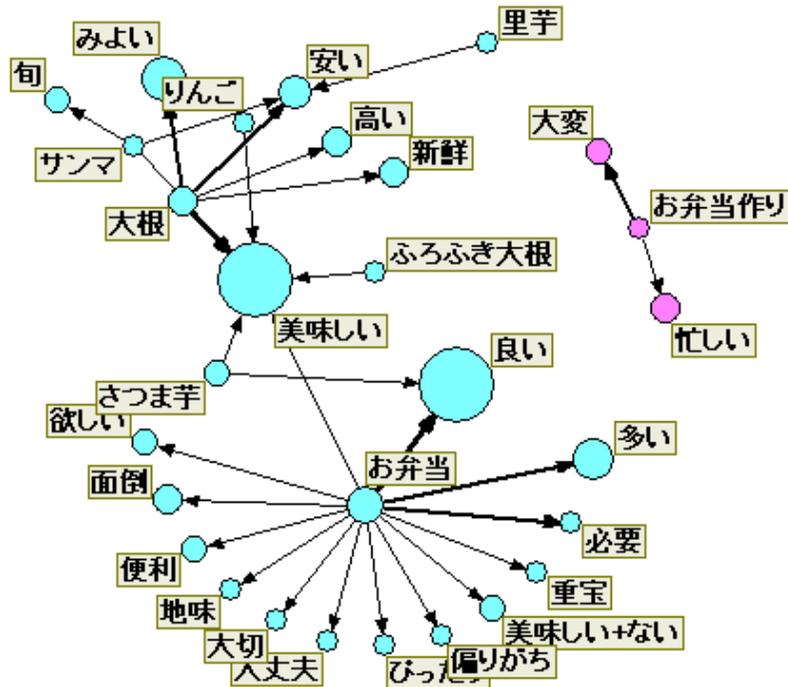


図3 秋の特徴語に対するネットワーク

## 秋の特徴語とネットワークから 提案するレコメンド文

→ 「温かい」  
「重宝」  
「楽ちん」などの単語を含む

また、食材としては  
「おでん」「大根」などが旬  
→冬よりも秋の方がおでんが旬

# 考察—冬—

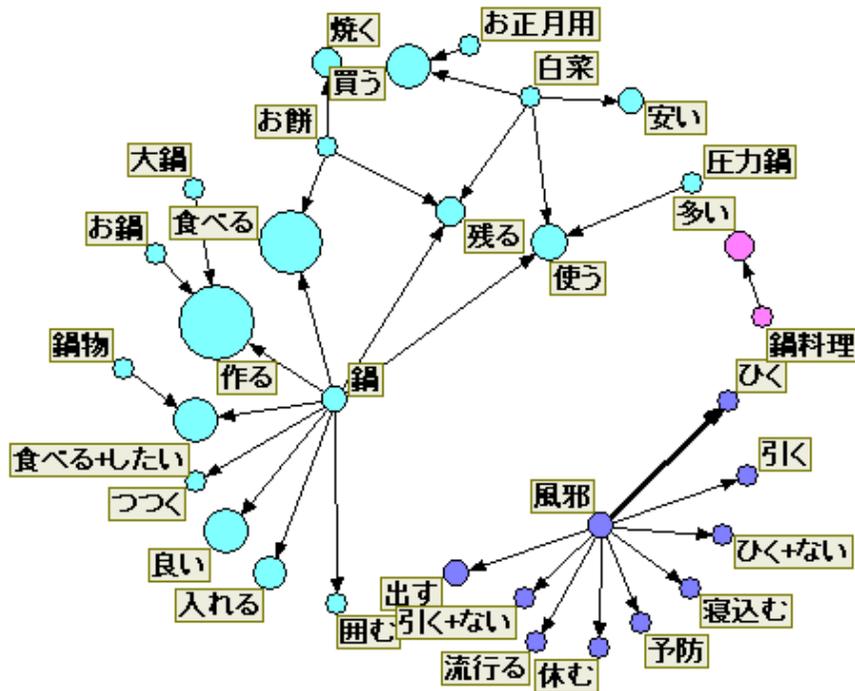


図4 冬の特徴語に対するネットワーク

## 冬の特徴語とネットワークから 提案するレコメンド文

→ 「つつく」  
「風邪に効く」  
「鍋にぴったり」などの単語を含む

また、食材としては  
「白菜」「お餅」などが旬  
→ 雑煮や鍋料理に入れる

# まとめ

---

## まとめ

- 本分析では、季節ごとの重要単語の特定を目的に、特徴語分析やことばネットワークを用いた分析を行なった。
- 分析の結果、季節ごとではやはり特徴的な単語は異なり、それぞれの季節に適した単語が特徴語として抽出できた。
- 同時に使われている単語から、その商品にあったレコメンド文を提案できた。

## 今後の課題

- 本分析では季節というところにしか着目していないが、ユーザの感情という部分にも着目した分析を行うことで、よりユーザに適したレコメンドを行うことができるのではないかと考えた。