

January 10, 2020

Web 内容マイニングによる個別知識の 共通単語を介した相関ルールからの発想支援

1615052 山元 悠貴

富山県立大学 情報基盤工学講座

1. はじめに
2. 発想支援システムとは
3. データドリブンによる最適化
4. おわりに

January 10, 2020

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

1. はじめに

2/18

- 1. はじめに
- 2. 発想支援システムとは
- 3. テキストマイニング

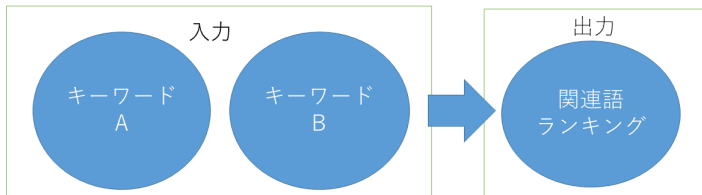
背景

現在，**AI** の発達に伴って発想支援についての研究も進んでいる．これからの時代はよりアイデア発想が重要になってくると考える．

目的

Google の検索結果を用いて複数のキーワードからの発想支援に着目した．

本研究では複数のキーワードを入力した際にそれらのキーワードに関連したアイデア発想につながるヒントを出力したい．



2. 発想支援システムとは

3/18

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

サイバー空間からのテキストデータ収集

発想支援において重要なことはキーワードからより関連度の高い単語をより多く表示させることである。

そこで、より良いデータを多く収集するためにサイバー空間からテキストデータを収集することとする。

自然言語処理

自然言語処理とは私たちが使う言語をコンピュータに処理させる技術のこと。

- コーパス
- 辞書
- 形態素解析
- 構造解析



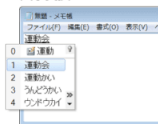
2. 自然言語処理の例

4/18

身近な例

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

■文字変換



■検索エンジン



■機械翻訳



■対話システム



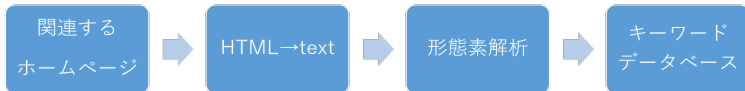
3. テキストマイニング

5/18

テキストマイニングとは

テキストマイニングとは、大量かつ多量なデータを様々な観点から分析し、役に立つ情報を取り出そうとする技術である..

今回、インターネット上のテキストを用いることで大量のデータを活用することができる。



1. はじめに
2. 発想支援システムとは
3. テキストマイニング

3. データドリブンによる最適化

6/18

単語の分散表現ベクトル

単語をベクトル表現する技術がある。(word2vec)

複数のキーワードの関連語がなかった場合にも発想支援できるように単語をベクトル表現することで近いベクトルの単語同士を関連語として見れるのではないかと考える。

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

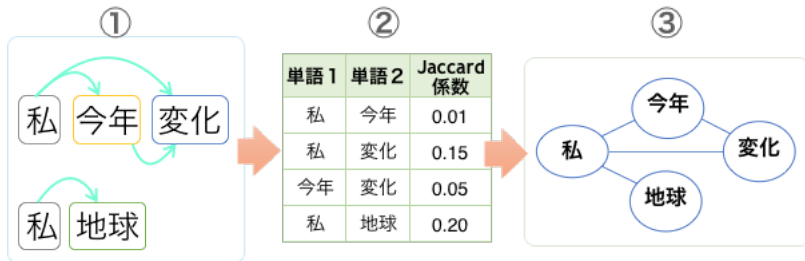
3. 複数のキーワード間の共起ネットワーク

7/18

複数のキーワード間の共起ネットワーク

Jaccard 係数を使って共起ネットワークを作成する

1. はじめに
2. 発想支援システムとは
3. テキストマイニング



4. 現状と課題

8/18

現状

- ① テキストマイニングでキーワードから Google の検索結果の X ページ分の URL を取得
- ② URL からそのページの HTML を取得して HTML タグを消去したテキストを取得
- ③ そのテキストに対して形態素解析し, 品詞を絞ってリストにする
- ④ そのリストから単語の遷移に応じた隣接行列を作る
- ⑤ 隣接行列からベイジアンネットワークを作る

4. 現状と課題

9/18

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

```
[ 'https://retty.me/area/PRE16/LCAT5/CAT290/city/16211/', 'https://tabelog.com/toyama/C16211/rstLst/ramen/',
'https://tabelog.com/toyama/C16211/rstLst/MC/', 'https://www.jalan.net/gourmet/cit_162110000/g2_3g130/', 'h
ttps://loco.yahoo.co.jp/search/?areacd=16211&genrecd=0106', 'https://ja.wikipedia.org/wiki/%E3%83%A9%E3%83%
BC%E3%83%A1%E3%83%B3', 'https://r.gnavi.co.jp/hg8enmn90000/', 'https://r.gnavi.co.jp/hg8enmn90000/', 'http
s://ramendb.supleks.jp/', 'https://www.marutakaya.co.jp/', 'https://www.syodai-marugen.jp/', 'https://www.m
enya-iroha.com/', 'https://www.hachiban.jp/shop/toyama/', 'https://www.rairaitei.co.jp/menu/', 'https://ww
w.itmedia.co.jp/business/articles/1905/28/news037.html', 'https://ichiran.com/np/ramen/', 'https://misoya.n
et/menu/' ]
```

お品書き | 味噌らーめん専門店 麵場 田所商店

お品書き - 味噌らーめん専門店 麵場 田所商店

ホーム

田所商店のこだわり

お品書き

店舗案内

会社案内

FC加盟

お問い合わせ

お品書き

北海道味噌・信州味噌・九州味噌・伊勢味噌・江戸前味噌など、各店舗(地域)に応じて3種類のお味噌を基本とし、それぞれの味噌に対して「味噌漬け炙りチャーシュー麺」「肉ねぎらーめん」「野菜らーめん」など数種のメニューをご用意致しております。

また、らーめんには味噌の土地にちなんだ特産品をトッピング。

北海道は、新鮮なじゃがいもをお店で揚げた「フライドポテト」。

信州は、らーめんには珍しい、彩り豊かな「山菜」。

九州には、味噌とも良く合う、野菜たっぷりの「さつま揚げ」。

伊勢には、味噌との相性抜群、風味豊かな「あおさのり」。

江戸前は大きな大きな「のり」と「あさり」をトッピング。

4. 現状と課題

10/18

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

[品書き、|、味噌、ら、ー、めん、専門、店、麵、場、田所、商店、品書き、味噌、ら、ー、めん、専門、店、麵、場、田所、商店、ホーム、田所、商店、こだわり、品書き、店舗、案内、会社、案内、FC、加盟、問い合わせ、品書き、北海道、味噌、信州、味噌、九州、味噌、伊勢、味噌、江戸前、味噌、店舗、(、地域、)、応じ、3、種類、味噌、基本、し、それぞれ、味噌、味噌、漬け、炙り、チャーシュー、麵、肉、ねぎ、ら、ー、めん、野菜、ら、ー、めん、数、種、メニュー、用意、致し、おり、ら、ー、めん、味噌、土地、ちゃん、特産、品、トッピング、北海道、新鮮、じゃがいも、店、揚げ、フライドポテト、信州、ら、ー、めん、珍しい、彩り、豊かな、山菜、九州、味噌、良く、合、う、野菜、たっぷり、さつま揚げ、伊勢、味噌、相性、抜群、風味、豊かな、あおさ、のり、江戸前、のり、あさり、トッピング、サイド、メニュー、味噌、チャーハン、味噌、唐、揚げ、味噌、だれ、食べる、餃子、味噌、こだわ、メニュー、こ、用意、し、おり、メニュー、ー、例、店舗、取り扱い、無い、場合、ござ、い、お手数、店舗、問い合わせ、頂け、確、実、定番、味噌、北海道、味噌、ら、ー、めん、濃口、北海道、味噌、赤色、系、濃口、味噌、大地、恵み、育ま、れ、力強く、パンチ、効い、味、塩分、控え目、コク、あり、深み、ある、味わい、信州、味噌、ら、ー、めん、濃口、全国、味噌、生産、量、3、分の、1、以上、占める、代表、的、淡色、系、味噌、酸味、ある、芳醇、香り、特長、九州、麦、味噌、ら、ー、めん、甘口、九州、地方、麦、味噌、主、生産、地、温暖、気候、ため、熟成、期間、短く、色、淡色、系、淡、赤色、麦、香ばし、さ、感じら、

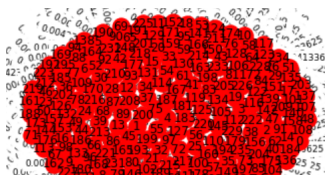
[illegible]

4. 現状と課題

11/18

- はじめに
- 発想支援システムとは
- テキストマイニング

	数	Inc	のり	あさり	1	情報	表示	基本	的	土地	...	彩り	(辛味	例	香り		入り
数	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	...	0.004	0.004	0.004	0.004	0.004	0.004	0.004
Inc	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	...	0.004	0.004	0.004	0.004	0.004	0.004	0.004
のり	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	...	0.004	0.004	0.004	0.004	0.004	0.004	0.004
あ																		
さ	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	...	0.004	0.004	0.004	0.004	0.004	0.004	0.004
り																		
1	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	...	0.004	0.004	0.004	0.004	0.004	0.004	0.004
情報	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	...	0.004	0.004	0.004	0.004	0.004	0.004	0.004



4. 現状と課題

12/18

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

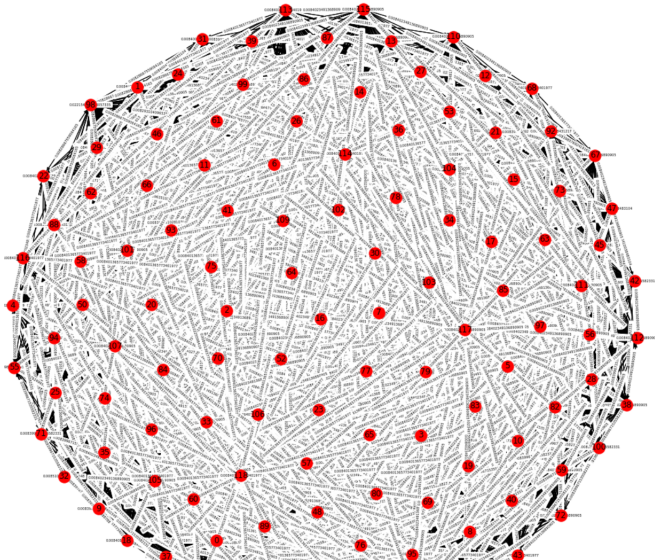
['多い', 'やすい', '細い', 'なし', 'なし', 'なし', 'なし', 'やすい', 'なし', '詳しい', '楽しく', 'よろしい', '多い', '高い', '美味しい', '細い', '無い', '速い', '美味しい', '忙しい', '暑い', 'なし', 'なし', 'なし', 'なし', 'やすい', 'なし', '詳しい', '楽しく', 'よろしい', 'よかつ', 'よかつ', '旨い', '美味しく', '旨い', '高い', '美味しかつ', '小さい', '甘', '美味しい', '甘い', '生臭', '美味しい', '暑く', '美味しい', 'なく', '少なく', 'やすく', '美味しゅう', '美味しい', 'にくい', '遅く', '遅い', 'いい', '旨い', '美味しく', '旨い', '美味しい', '高い', '美味しい', '高い', '美味しかつ', '小さい', '甘', '美味しい', '遅い', 'いい', '美味しい', '安い', 'ない', '詳しく', '旨い', '酸っぱい', '酸し', '古い', 'なかつ', '多く', '多く', '柔らかい', 'な', '短い', 'なく', '詳しく', 'やすく', '良く', '多く', '大きかつ', '小さく', '近い', '厳しい', '遅く', '長い', '多く', 'ない', '黒い', '悪', '珍しく', '面白', '正しい', '甘', '大きく', '多い', 'なく', '大きく', '多い', '正しい', '新しい', '著しい', 'なかつ', '広く', '高く', '安く', '長い', 'ほしく', '高く', '多く', '多い', '多い', 'やすい', '厳しい', '望ましい', '薄い', '多い', '近い', '多い', '浅く', '多い', '多い', '遠い', '細長く', '太', 'やすい', '白い', '多く', '多い', '細かく', '深い', '細長い', '多い', '多い', 'にくい', '古い', '強く', '多い', '浅', 'ない', '新しい', 'うまい', '細い', '細長い', '細', '多い', '早', '多い', 'ぬく', '温かい', 'ぬくい', '温かい', '少ない', '多い', '多い', 'いい', 'やばし', '長い', '高く', '甘', 'おもしろ', '甘く', '美味しい', '近き', '佳', 'やばし', '美味しい', '事新しい', '旨い', '数少ない', '甘', '多い', '旨い', '旨い', '暖かい', '冷たい', '近い', '美味しい', '旨い', '新しい', '美味しい', '甘', '美味しい', '甘', '良い', 'ぎこちなかつ', '美味しい', '重', 'ない', 'おいしい', '白い', '美しい', '深い', '珍しい', '重', '甘', '重', 'なし', '安い', 'なし', 'なし', 'なし', 'なし', '新しい', '心地よい', '美味しい', 'うまい', '有り難う', '固く', 'とろ', 'とろ', 'とろい', 'とろい', '甘', 'とろ', 'とろい', 'とろい', '甘', 'とろ', '甘', '甘', 'うまい', '美味しい', '愉し', '美味しい', '愉し', '美味しい', '愉し', '良い', '愉し', '愉し', '多く', '愉し', 'なし', '甘', 'おいしかつ', 'にくく', '良し', '美味しい', 'いい', '正しい', '良い', '珍しく', '高い', '安く', 'なく', 'なく', 'なく', 'なく', '少なく', '美味しい', '美味しく', 'やすく', '多く', '広い', '多い', '多く', '嬉しい', '遅く', '近い', 'いい', 'おいし', 'おいしい', '美味しい', 'ほど近く', '古き', '良き', '白い', '甘く', '程よく', '強い', 'いい', '近い', '細い', 'にくい', 'ほど近く', 'おいしい', 'ほど近い', 'おいしい', '甘', '甘', '欲しい', '美味しかつ', 'おいしい', '明るい', 'やすい', 'やすく', '安く', '美味しい', '甘', 'うい', '甘', '甘', 'うまい', 'うまい', 'うま', 'うま', 'ない']

[[0.01679752 0.00839876 0.00839876 ... 0.00839876 0.00839876 0.00839876]
[0.00840235 0.0168047 0.00840235 ... 0.00840235 0.00840235 0.00840235]
[0.00840235 0.00840235 0.0168047 ... 0.00840235 0.00840235 0.00840235]

4. 現状と課題

13/18

1. はじめに
2. 発想支援システムとは
3. テキストマイニング



4. おわりに

14/18

まとめ

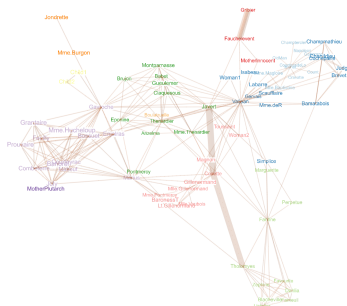
- ① アイデア発想支援の現状とこれからについて述べた.

今後の課題

- ① 複数のキーワードごとにネットワークを作り同じ単語が出てきたときにつなげる
- ② ベイジアンネットワークにした時の表示を見やすくする

これまでの課題

- ① ベイジアンネットワークにした時の表示を見やすくする
- ② そのために **3D** のネットワーク図にするために今の隣接行列のデータを適用する



4. 現状と課題

16/18

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

```
[{"id": "美味し": {"美味し": 0.0, "甘い": 0.0, "安い": 0.0, "高い": 0.0, "懐かし": 0.0, "正しい": 0.0340986003, "かわい": 0.0, "強く": 0.0, "多かつ": 0.0, "寂しい": 0.0, "詳しく": 0.0, "痛": 0.0, "古く": 0.0, "詳しい": 0.0, "やすい": 0.0, "難しい": 0.0, "": 0.0, "馬鹿らしく": 0.0, "薄く": 0.0, "美しい": 0.0, "なし": 0.0, "白い": 0.0, "かしい": 0.0, "よかつ": 0.0, "もの寂し": 0.0, "細長く": 0.0, "深かつ": 0.0, "っ若い": 0.0, "なつかし": 0.0, "ほしい": 0.0, "安く": 0.0, "悪い": 0.0, "かわいしい": 0.0, "不味い": 0.0, "寒": 0.0, "長く": 0.0, "無い": 0.0, "高く": 0.0, "幅
```

実際の隣接行列のデータ

```
{
  "nodes": [
    {"id": "さとし", "group": 1},
    {"id": "Napoleon", "group": 1},
    {"id": "Mlle.Baptistine", "group": 1},
    {"id": "Mme.Magloire", "group": 1},
    {"id": "CountessdeLo", "group": 1},
    {"id": "Geborand", "group": 1},
    {"id": "Champtercier", "group": 1},
    {"id": "Cravatte", "group": 1},
  ]
}
```

```
"links": [
  {"source": "Napoleon", "target": "さとし", "value": 1},
  {"source": "Mlle.Baptistine", "target": "さとし", "value": 8},
  {"source": "Mme.Magloire", "target": "さとし", "value": 10},
  {"source": "Mme.Magloire", "target": "Mlle.Baptistine", "value": 1},
  {"source": "CountessdeLo", "target": "さとし", "value": 1},
  {"source": "Geborand", "target": "さとし", "value": 1},
  {"source": "Champtercier", "target": "さとし", "value": 1},
  {"source": "Cravatte", "target": "さとし", "value": 1},
]
```

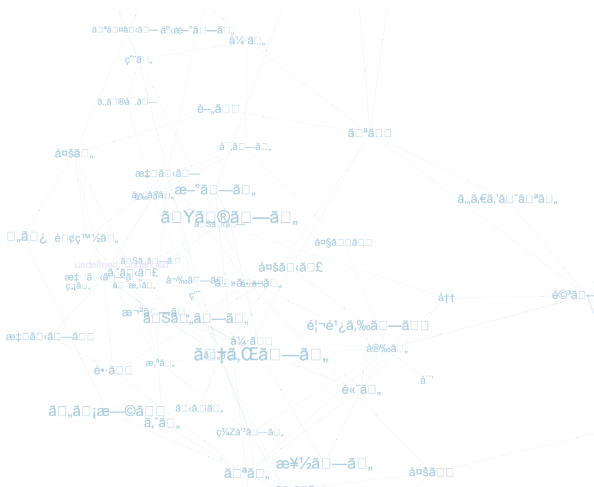
使える形のデータ

使える形にしたデータ

4. 現状と課題

18/18

1. はじめに
2. 発想支援システムとは
3. テキストマイニング



使える形にしたデータを適用した結果

課題

- ① ベイジアンネットワークにした時の表示の文字化けを直す
- ② jupyter のプログラムを.py で使えるようにする
- ③ キーワード

