

February 7, 2020

# Web 内容マイニングによる 複数キーワードに対する 3D 有向グラフを用いた発想支援

1. はじめに
2. 発想支援システムとは
3. インターネット上のテキストデータの分析

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. 発想支援システムの概要と課題
3. インターネット上のテキストデータの分析
4. 提案手法
5. おわりに

# 1. はじめに

2/10

## 背景

1. はじめに
2. 発想支援システムとは
3. インターネット上のテキストデータの分析

人間の創造活動において精神的な労力が必要である。この精神的な労力の負担を減らすためにい発想支援についての研究がされている。現在、情報処理技術の発達に伴って発想支援についての研究も進んでいる。これから時代はよりアイデア発想が重要になってくると考える。既存の発想支援には一つのキーワードに対してそのキーワードに関連するテキストデータから共起ネットワークにするものがある。

## 目的

- 1 複数キーワードに対して関連語を 3D 有向グラフにして描画するシステムの作成
- 2 単語の順序関係による有向な関係を生かした発想支援

2/10

## 2. 発想支援システムとは

3/10

### 発想のプロセス

1. はじめに
2. 発想支援システムとは
3. インターネット上のテキストデータの分析

- 1 Step1:情報の収集
- 2 Step2:発散的思考 - 関係のあると思われるところを洗い出す
- 3 Step3:収束的思考 - 発散的思考で出した関係の整理・本質の追求
- 4 Step4:アイデア結晶化 - アイデアの評価・決断

本研究の発想支援システムではこのプロセスに従ってシステムを構築する。

発想のプロセスをすべて行う発想システムは実現が難しいので、情報の収集から発散的思考と収束的思考までを行う発想支援システムが研究されている。

### 3. テキストマイニング

4/10

#### テキストマイニング

- はじめに
- 発想支援システムとは
- インターネット上のテキストデータの分析

情報を収集する行程では、多量のテキストデータがあるインターネットのテキストを収集・分析する。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。

#### インターネット上からのテキストデータの分析

任意の2つのキーワードについて、クローリングによりウェブサイトのURLを集めてくる。そのURLのテキストデータをクリーニング・単語分割・正規化・ストップワード除去し、名詞・形容詞・動詞に絞った単語にまとめる。

クローリング



自然言語処理



隣接行列



3D有向グラフ

## 4. アルゴリズム

### アルゴリズム

1. はじめに
2. 発想支援システムとは
3. インターネット上のテキストデータの分析

- ① 複数のキーワードから Google の検索結果の 1 ページ分の URL を保存
- 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データを取得
- ③ そのテキストデータに対して自然言語処理の前処理を行った後、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順序に応じた隣接行列を作る
- ⑤ 隣接行列の規模が大きすぎるので PageRank を使って関連性の高い単語に絞る
- ⑥ 隣接行列を 3D 有向グラフに適応する

隣接行列を 3D 有向グラフに使える形にしたデータ

- はじめに
  - 発想支援システムとは
  - インターネット上のテキストデータの分析

```
var text3D = {"nodes": [{"id": "リコピン", "color": 2}, {"id": "にくい", "color": 2}, {"id": "サラダ", "color": 2}, {"id": "入れる", "color": 2}, {"id": "ナス科", "color": 2}, {"id": "できる", "color": 2}, {"id": "サイト", "color": 2}, {"id": "ポリ塩化ビニル", "color": 2}, {"id": "レシピ", "color": 2}, {"id": "ポリマー", "color": 2}, {"id": "熱硬化性樹脂", "color": 2}, {"id": "用いる", "color": 2}, {"id": "大きい", "color": 2}, {"id": "耐熱性", "color": 2}, {"id": "ページ", "color": 2}, {"id": "合成樹脂", "color": 1}, {"id": "ジュース", "color": 2}], "edges": [{"source": "リコピン", "target": "にくい", "color": 2}, {"source": "にくい", "target": "サラダ", "color": 2}, {"source": "サラダ", "target": "入れる", "color": 2}, {"source": "入れる", "target": "ナス科", "color": 2}, {"source": "ナス科", "target": "できる", "color": 2}, {"source": "できる", "target": "サイト", "color": 2}, {"source": "サイト", "target": "ポリ塩化ビニル", "color": 2}, {"source": "ポリ塩化ビニル", "target": "レシピ", "color": 2}, {"source": "レシピ", "target": "ポリマー", "color": 2}, {"source": "ポリマー", "target": "熱硬化性樹脂", "color": 2}, {"source": "熱硬化性樹脂", "target": "用いる", "color": 2}, {"source": "用いる", "target": "大きい", "color": 2}, {"source": "大きい", "target": "耐熱性", "color": 2}, {"source": "耐熱性", "target": "ページ", "color": 2}, {"source": "ページ", "target": "合成樹脂", "color": 1}, {"source": "合成樹脂", "target": "ジュース", "color": 2}]};
```

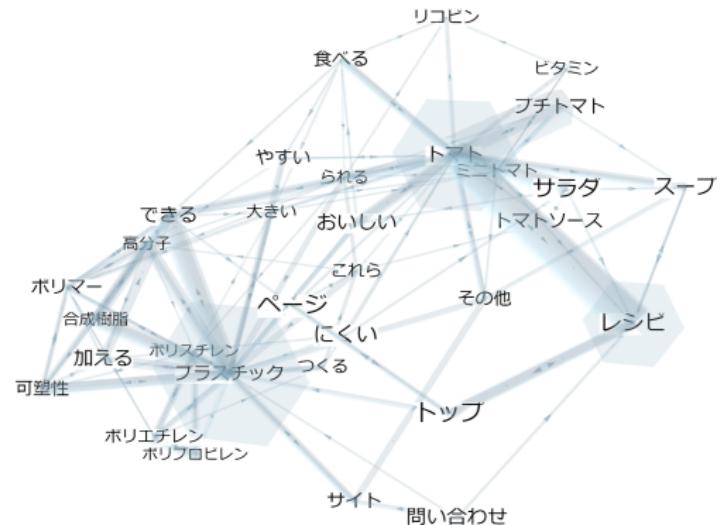
### 3D 有向グラフのノードとなる

### 3D 有向グラフのエッジとなる

1. はじめに
  2. 発想支援システムとは
  3. インターネット上のテキストデータの分析

「トマト」と「プラスチック」を入力したときの出力結果

出力結果から「トマト」→「食べる」→「やすい」→「プラスチック」の繋がりが分かる。その結果、「トマトを食べやすくするプラスチック」という発想支援が実現した。



## 4. 比較

8/10

### Google で「トマト プラスチック」と検索した際の一ページ目の URL タイトル

【プラスチック製品のケチャップ染みの落とし方】簡単!!食器や容器 ...  
プラスチック食器の皿にケチャップ汚れが残ってしまった時の落とし ...  
プラスチック容器汚れが落ちない時の解決法!電子レンジ ... - レシピ  
プラスチック保存容器にしみついた汚れは漂白剤と電子レンジで ...  
子供用食器(プラスチック製)の汚れが取れません・・・スパゲッティ ... - Yahoo!知 ...  
【楽天市場】トマト(コンテナ | プラスチック容器) : 農業資材 ...  
第8章 トマトケチャップ用チューブ容器の開発 - 一般社団法人 日本 ...  
タッパーの汚れ - マイレビ  
使い捨てプラスチックを減らす・使わずに生活する30の方法 | プラ ...  
Jacment プラスチックトマトストレージキーパーポッド 2 ... - Amazon

### Google で「トマト プラスチック」に関連する検索キーワード

#### トマト プラスチックに関連する検索キーワード

プラスチック 食器 色素沈着	プラスチック容器 白い汚れ
プラスチック 色素沈着	トマト ソース 汚れ
人参 プラスチック 色移り	プラスチック容器 洗い方
タッパー ケチャップ 取れない	トマト 色 落ちない
プラスチック カレー 黄ばみ 重曹	タッパー 油 落とし方

## まとめ

1. はじめに
2. 発想支援システムとは
3. インターネット上のテキストデータの分析

- ① 複数のキーワードに対する関連語を 3D 有向グラフにして発想支援システムとして提案した

## 今後の課題

- 1 キーワード間の単語の経路を重み付けすることで、より関連性が高い単語による発想支援の実現が考えられる。
- 2 隣接行列の作成に単語の順序関係のみでなく、係り受け解析も考慮することにより文章の依存構造を反映した発想支援の実現が考えられる。