

# テキストデータを対象とした 文章における特有表現に着目した 意見抽出方式

池上 藍羽†      石井 雄大†      北 椋太†      中西 崇文†

† 武蔵野大学データサイエンス学部データサイエンス学科

E-mail: † s2022003@stu.musashino-u.ac.jp, takafumi.nakanishi@ds.musashino-u.ac.jp

# 目次

1

## 研究概要

研究背景 文章の構成要素 研究目的

2

## 提案手法

データセット 意見文抽出機能 意見文分類機能 意見文可視化機能

3

## 実験

4

## まとめ

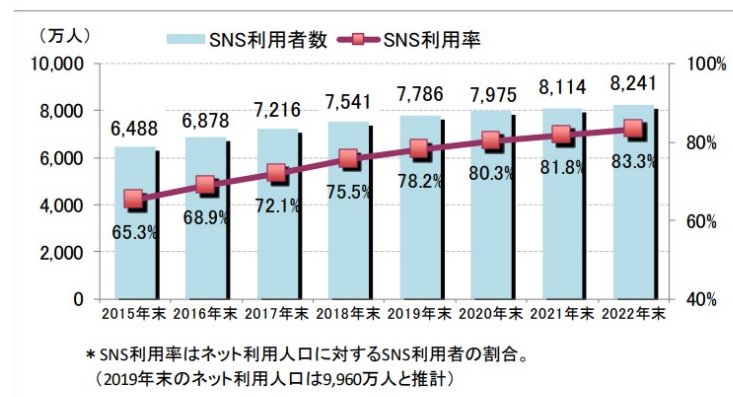
# 研究背景

## 新聞の電子化・SNSの発達

Web上において、多くのテキストが散在している

→ データを取得する機会の増加

■表1. 日本におけるSNS利用者数



ICT 総研調べ

## 意見と事実は混同して書かれることがある

→ ユーザが大量のテキストデータから意見と事実を判断し、抽出することが困難

テキストの形式ごとに意見を抽出することが重要である

# 文章の構成要素

## 文章の構成における主要素は「意見」「事実」の2つと仮定

意見：筆者の考えや判断に基づくもの

事実：調査や実験で確認が可能なもの

意見と事実の割合・内容は文章の形式により異なる

事実が多い例：ニュース・辞典

意見が多い例：レビュー・随筆

→意見が多い文章は発信者側の意向に沿った内容となる傾向にある

# 研究目的

## 文章における特有表現に着目した意見抽出方式の実現

様々なテキストデータから意見を抽出

テキストの形式ごとに意見の割合および、意見の主なトピックについての可視化をする

文章内から意見を抽出

→ その文章内にどのような考えが込められているのか客観的な把握が可能

本研究が対象とするテキストデータ：ニュースやレビューに関する記事

# 目次

## 1 研究概要

研究背景 文章の構成要素 研究目的

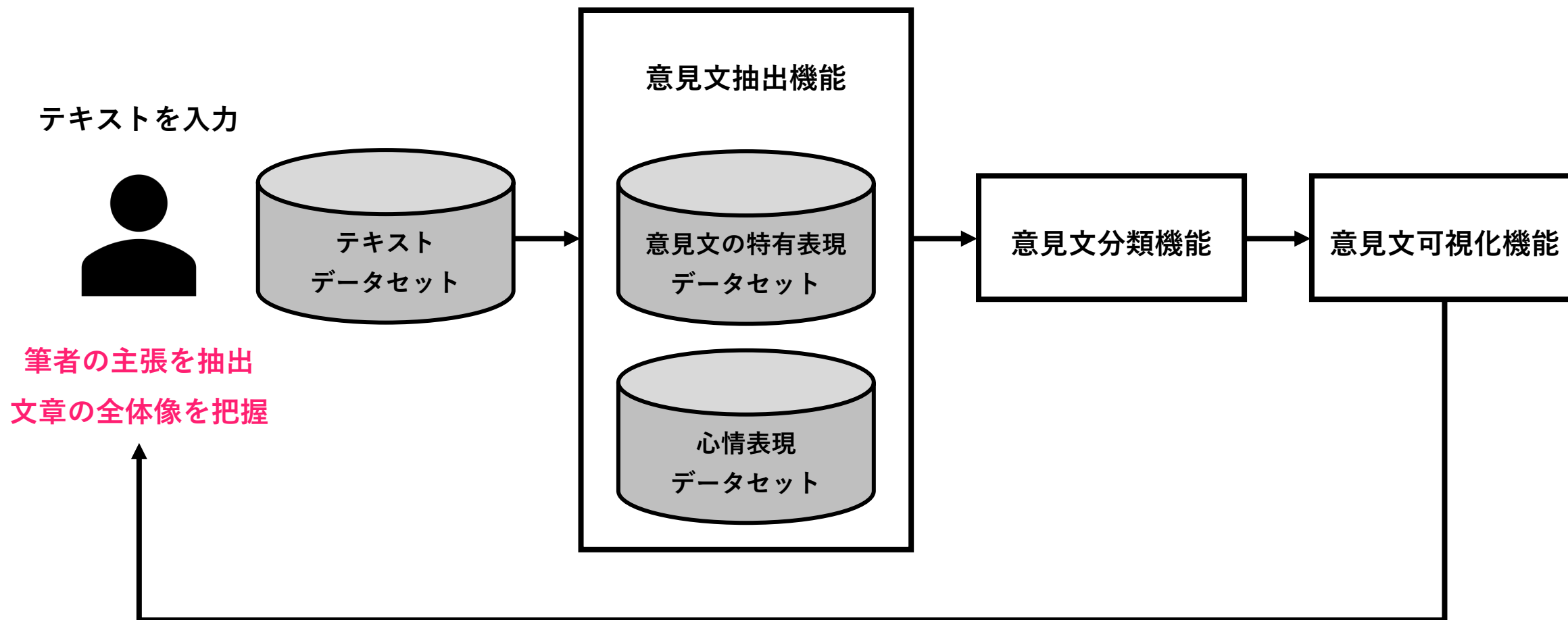
## 2 提案手法

データセット 意見文抽出機能 意見文分類機能 意見文可視化機能

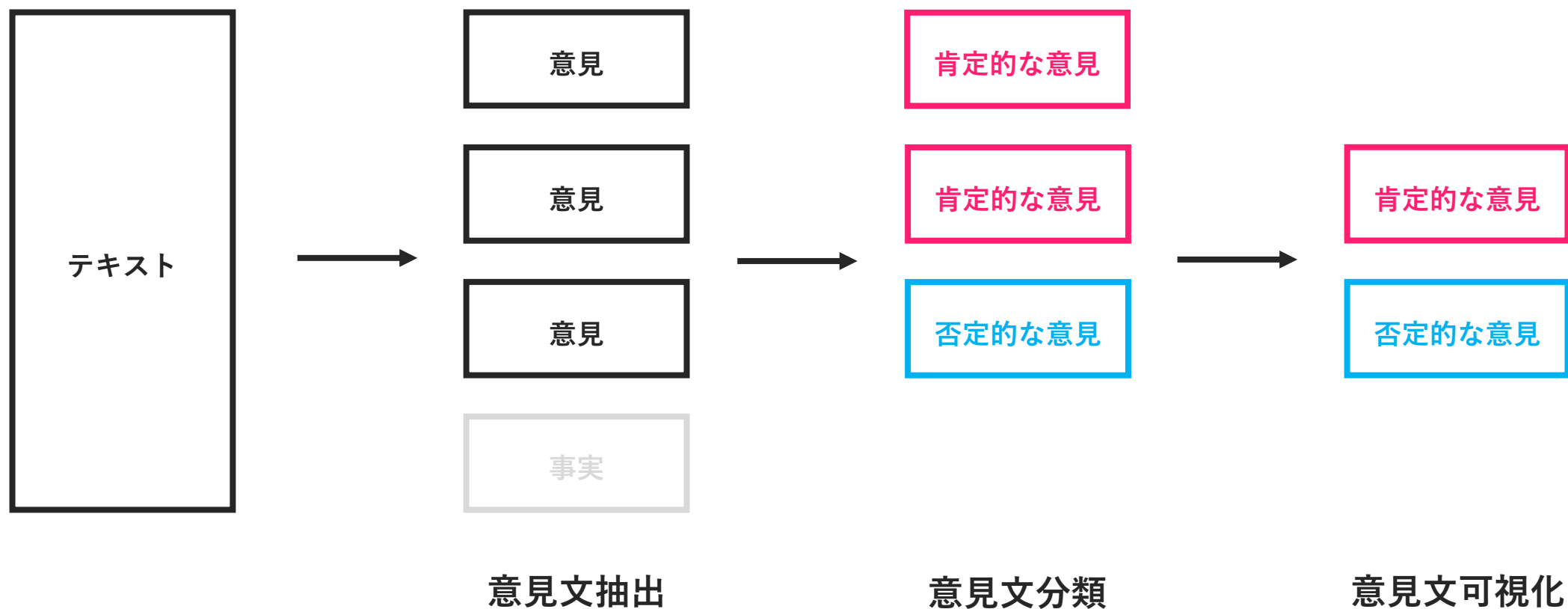
## 3 実験

## 4 まとめ

# 提案手法の全体像



# 提案手法の全体像





# 提案手法における各機能

1

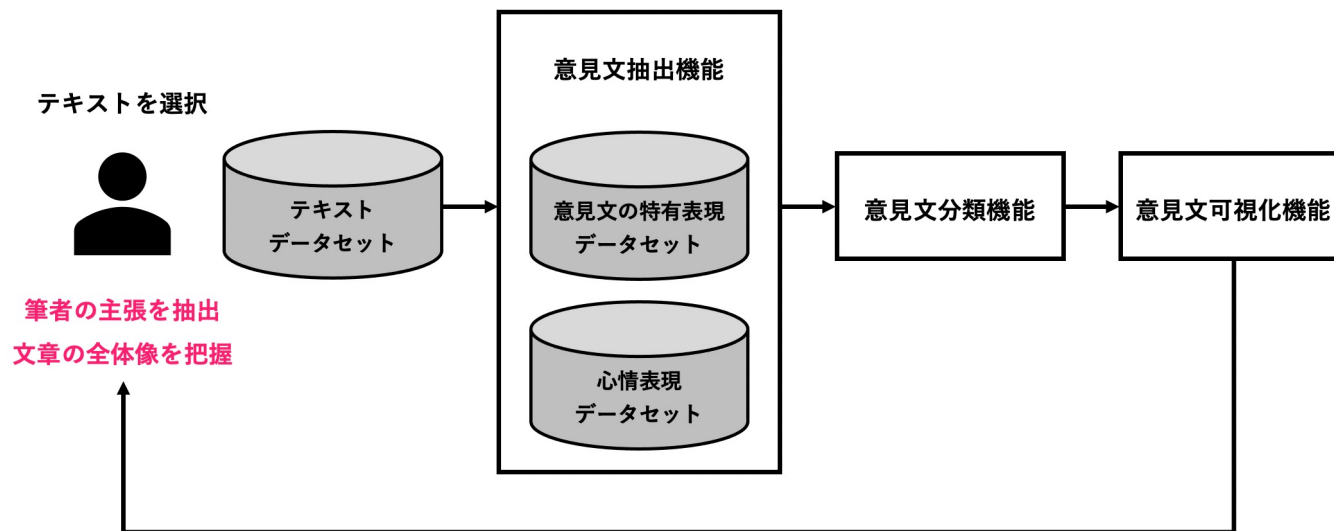
意見文抽出機能

2

意見文分類機能

3

意見文可視化機能



# 意見文抽出機能

## テキストから意見文の抽出

テキストデータと意見文の特有表現・心情表現データセットをマッチング

→ 筆者の意見が述べられている文を抽出する

この漫画は大人気だ。

今私は幸福だ。

この商品はおしゃれだと思う。

外に出ることが面倒に感じる。

# 文章における特有表現・心情表現

## 文章における特有表現に着目した意見抽出方式の実現

**意見文における特有表現：**

「思う」「感じる」など考えを述べる際に用いられる表現と、それらの類語

**心情表現：**

長岡技術科学大学の自然言語処理研究室が公開している日本語感情表現辞書

# 意見文における特有表現データセット

主観的な考えを述べる際に用いられる表現と、それらの類語を格納するデータベースを作成(242語)

## 特有表現の例：

考える・思う・感じる・評価する

## 類語の例：「考える」

考案・考察・想う・見なす・仮定・推定

# 心情表現データセット

長岡技術科学大学の自然言語処理研究室が公開している日本語感情表現辞書より抽出  
(1978語)

例：哀愁・愛・忌み嫌う・虚無・幸福・嘆く・不快・憂鬱

# 意見文分類機能

## 抽出された意見を肯定的な意見と否定的な意見に分類

日本語評価極性辞書を用いて、肯定的な意見文と否定的な意見文进行分类する

肯定的な意見文 1.0

必要以上に**恐れず**に、これまでの対策の徹底を求めています。

否定的な意見文 -0.5

「ワクチン**効果**の**低下**が**懸念**される」と指摘。

# 意見文可視化機能

## WordCloudを用いて肯定的な意見と否定的な意見を可視化する

文章内においてどのようなトピックが多く見られたか、視覚的に表現

→肯定的な表現を含む文と、否定的な表現を含む文において多く見られた単語を比較

抽出された意見文を形態素解析

→形容動詞語幹を用いて肯定的な意見と否定的な意見に分類

形容動詞語幹の例：高級だ → 高級



# 目次

## 1 研究概要

研究背景 文章の構成要素 研究目的

## 2 提案手法

データセット 意見文抽出機能 意見文分類機能 意見文可視化機能

## 3 実験

## 4 まとめ



# 実験

1

テキスト間における抽出された意見の割合の  
差異を明らかにする

2

肯定的な文と否定的な文に分類し、それぞれ  
可視化する手法の有効性を検証する

# データセット

## テキストデータセット

新型コロナウイルスに関するニュース記事(100件)

レビューに関する記事(100件)

ニュース記事：客観性が求められる

レビュー記事：個人的な感想が述べられる

# テキスト間における抽出された意見の割合の差異

## 抽出された意見の割合

対象：テキスト全体

文章内容	意見文の割合
ニュース	17.0262%
レビュー	28.3083%

ニュース：客観性が求められる

レビュー：個人的な文章

→対象物に対しての評価や感想を述べることが多い

# テキスト間における抽出された意見の割合の差異

## 抽出された意見の割合

対象：新型コロナウイルスに関する特定の記事A・B

文章内容	意見文の割合
記事A	18.1818%
記事B	43.4782%

### 記事A：緊急事態宣言の延長に関する報道

→ 政府の発表をありのまま報じている

### 記事B：変異株に関する医師の見解

→ 医師の考えが反映された記事

# 肯定的な意見文・否定的な意見文の可視化

## レビュー記事における肯定的な意見と否定的な意見

対象：レビュー記事



肯定的な意見(左)



否定的な意見(右)

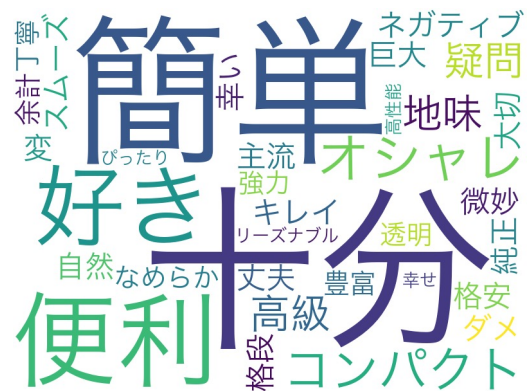
## 肯定的な意見：高級・安全

否定的な意見：可能・必要

→対象が不足している要素や  
要望が多く述べられている

# 肯定的な意見文・否定的な意見文の可視化

「iPhone SE (第2世代)」に  
関するレビュー記事A



肯定的な意見

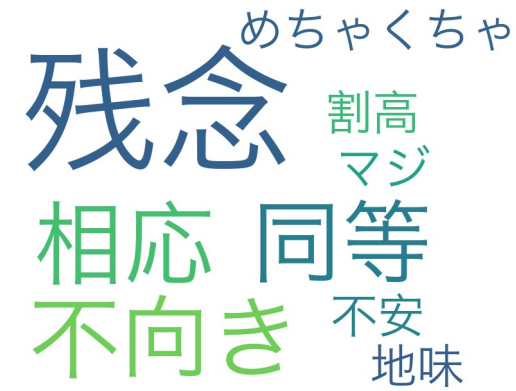


否定的な意見

「iPhone SE (第2世代)」に  
関するレビュー記事B



肯定的な意見



否定的な意見

# 目次

## 1 研究概要

研究背景 文章の構成要素 研究目的

## 2 提案手法

データセット 意見文抽出機能 意見文分類機能 意見文可視化機能

## 3 実験

## 4 今後の展望

# まとめ

## 意見文における特有表現に着目した意見抽出方式

意見を抽出することで文章の核となる要素（筆者の主張）を抽出可能とする

意見における主なトピックについての可視化

→ その文章内にどのような考え方が込められているのか客観的な把握が可能



# 今後の展望

事実に偽装した意見を抽出

→ フェイクニュースの発見

小説における登場人物の心情描写を抽出

→ 人物像を明らかにする

# 参考文献

- [1]小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, 自然言語処理, Vol.12, No.3, pp.203-222(2005).
- [2]長岡技術科学大学 自然言語処理研究室: 日本語感情辞書,  
〈<http://www.jnlp.org/SNOW/D18>〉 (参照2020-11-23).
- [3]oseti, 〈<https://pypi.org/project/oseti/>〉 (参照2021-1-3).

# 参考文献

[4]東山昌彦， 乾健太郎， 松本裕治， 述語の選択選好性に着目した名詞評価極性の獲得， 言語処理学会第14回年次大会論文集， pp.584-587， 2008.

[5]日本語評価極性辞書，

〈<http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary>〉（参照2021-1-3）.