

# レビューの要約を目的とした単語の意味領域に基づく上位語選択

渡辺 一生<sup>†</sup> 楠 和馬<sup>††</sup> 吉丸 直希<sup>††</sup> 波多野賢治<sup>†</sup>

<sup>†</sup> 同志社大学文化情報学部 〒610-0394 京都府京田辺市多々羅都谷 1-3

<sup>††</sup> 同志社大学大学院文化情報学研究科 〒610-0394 京都府京田辺市多々羅都谷 1-3

E-mail: <sup>†</sup>cgjd0306@mail4.doshisha.ac.jp, <sup>††</sup>{kusu,yoshimaru}@mil.doshisha.ac.jp,

<sup>†</sup>khatano@mail.doshisha.ac.jp

**あらまし** Web上に存在する文書から情報を取得する機会が増加しており、膨大な数の文書全てに目を通すことは困難であるため、それらを要約する技術が求められている。既存研究ではレビュー文書に出現する重要語に対してクラスタリングを行い、クラスタ全体を表す代表語を手で選択してラベルとして付与することで要約を実現している。しかし、レビュー文書によって適切な観点は異なるため人手による解釈は現実的ではない。そこで本研究では、単語の意味的広さを領域で表現可能な Word2Box を用いて単語集合に対して適切な観点を選択する手法を提案する。各単語の領域の重なりに基づき単語集合に対して上位語を選択する。

**キーワード** 情報抽出, 文書要約, 観点推定

## 1 はじめに

インターネット技術の発達に伴い、Web上に存在する膨大な数の文書には有用な情報が多く存在しており、それら文書の中から情報要求に合う文書を探し出す要求が高まっている。例えば、オンラインショッピングサイトや飲食店の口コミサイトなどに投稿されているレビューには、基本的には実際に商品を購入して使用した経験に基づいて記述されるため、購入検討者にとって有用な情報を含んでいる。しかし、膨大な数のレビューが投稿されている場合、全てのレビューに目を通すことは困難であり、レビューを閲覧するユーザが求めている情報を発見するまでに時間や労力が必要である。Yangらは、ユーザが意思決定を行う際、複数のレビューを閲覧することを前提として研究を行っている[1]。つまり、ユーザは意思決定を行う際に一つのレビューではなく複数のレビューを参照するため、閲覧負担が発生する可能性が高い。そのため、膨大な数のレビューから商品の購入を検討しているユーザにとって有用な情報を整理して提示することで、レビューを閲覧する際の負担を軽減することが求められている。レビューから商品の情報を効率的に把握するために有効な手法として、レビューを評価観点に基づいて分類する手法がある[2]。評価観点とは、レビューで述べられている評価対象への評価内容をカテゴリ化したものである。評価対象がレストランだった場合、「料理」、「サービス」、「立地」などの評価観点が考えられる。例えば、「この店の接客は良かった。」というレビュー文があった場合、「接客」という単語はレストランの「サービス」という観点に関係するため、このレビュー文は「サービス」の観点に分類される。このようにレビューを評価観点ごとに分類すれば、ユーザは欲しい情報を効率良く取得することが可能となる。

レビューから評価観点を抽出し、レビューを分類する研究が数多く行われている。評価観点の抽出方法には、ルールベース

や教師あり学習を用いる方法が採用されている[3]。しかし、評価対象によって適切な評価観点は異なるため、人手による観点の抽出ルールを作成することや教師あり学習のための正解データを準備することは人的コストを要することが問題である。

この問題を解決するために、教師なし学習を用いて評価観点の抽出を行い、抽出した評価観点に基づいてレビュー文の分類を行う手法が提案された[4]。しかし、Heらの手法は教師なし学習を用いて評価観点の抽出に要する人的コストを削減できるとしているが、作成された単語集合に対して人手で意味解釈を行っているため完全な自動化には至っていない。また、作成された単語集合の数が増加するにつれて人手による評価観点名の付与が困難になるため、レビューから得られた単語集合に対して適切な評価観点名を機械的に付与可能にする必要がある。

単語集合の意味解釈を行う際の人手による負担を軽減するための手法として、単語間の上位語・下位語の概念構造に基づく上位語を用いる手法が存在する[5]。この手法では各単語集合に対して複数の代表語を抽出して、WordNet<sup>1</sup>などの外部リソースから代表語の上位語を取得して単語集合に対してラベルとして付与する。その結果、付与された上位語によって単語集合の意味解釈が容易になったことが報告されている。しかし、WordNetのような人手によって作成された辞書ベースの外部リソースは、辞書を更新するコストが高く、現在ではその更新が行われていないため、把握できない概念関係が増加していく。

そこで本研究では、Word2Box[6]を用いて単語集合に対して適切なラベルとなる上位語を選択する手法を提案する。Word2Boxとは、教師なし学習によって単語の意味的広さを領域で表現する手法である。そのため、単語の概念関係や多義性、概念的な広がりを見視的に把握することが可能になる。本研究は、Word2Boxの特質を活用することで、人手による単語の意味解釈の処理を不要としつつ、文書の要約に必要な観点を文単

1: WordNet, <https://wordnet.princeton.edu/>, (2023/2/23 閲覧)

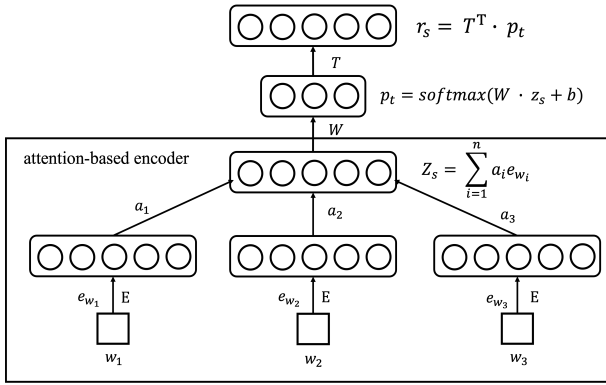


図1 ABAEの構造

位で付与することを可能にしている。

本稿は、一つのレビュー文に対して一つの観点ラベルを予測することを目的としている。実際のレビュー文書は複数の観点で構成されている場合も多いため、本研究はレビュー文書の要約自動化を行うための前段階的位置づけである。

## 2 関連研究

本節では、本研究と関連する評価観点推定の既存研究と、既存研究の問題点について説明する。また、既存研究の問題点を踏まえて、本研究の提案手法に関連する研究について説明する。

まず、2.1で教師なし学習で評価観点の抽出を行った研究の説明および問題点について述べる。次に、2.2で本研究と目的が一致している単語集合の機械的な意味解釈に取り組んだ研究について説明を行う。さらに、2.3で単語の分散表現を領域で表現することを可能にした Word2Box について述べる。これまでの単語埋め込み手法では、単語間の概念構造までは把握できなかったが、単語の概念を領域で捉えることを実現できたため、単語間の概念構造を把握できるようになっている。

### 2.1 評価観点抽出に関する研究

レビューを閲覧する際にかかる負担の問題を解決することを目的として評価観点抽出の研究が行われていた。評価観点抽出の既存手法ではルールベースや教師あり学習を用いているため、評価観点の抽出ルールの作成や学習のためのデータ整備にコストがかかる。Heらは既存手法で生じる問題を解決するために、教師なし手法による評価観点抽出を行い、レビュー文書から抽出された評価観点に基づいてレビューを分類する手法を提案している[4]。この手法を Attention-based Aspect Extraction (ABAE) という。

ABAEは Word2Vec モデル[7]による単語の分散表現を用いて、意味的に類似している単語同士を埋め込み空間内で近くに配置する。また、学習過程でアテンション機構によりレビュー文における評価観点と無関係な単語のフィルタリングを行うことで評価観点を表している重要な語の情報を保持することができる。

図1に ABAE のモデル構造の概要を示す。まず、レビューを構成する各単語  $w_i$  に対して、Word2Vec モデルを用いて分

散表現  $e_{w_i}$  を作成する。図中の  $E$  は各単語の分散表現からなる単語埋め込み行列である。次に、図中の  $a_i$  は、文の評価観点を捉えるために単語  $w_i$  が注目すべき正しい単語である確率と解釈することができる重みである。重み  $a_i$  は、単語ベクトル  $e_{w_i}$  と各単語ごとの埋め込み平均を用いて計算される。そして、重み  $a_i$  によって重み付けされた  $e_{w_i}$  の加重平均によって、文のベクトル  $Z_s$  を作成する。 $Z_s$  は評価観点を捉えるうえで重要な単語とそうではない単語をフィルタリングすることができる。文 (sentence) のベクトル  $Z_s$  を  $K$  次元に圧縮を行い、 $K$  個の評価観点 (topic) ベクトルに対する重みベクトルである  $p_t$  を求める。 $K$  は評価観点の数であり、事前に定義をする必要がある。 $p_t$  は入力されたレビュー文がどの評価観点到に属するのかを表した確率である。最後に、 $K$  次元に圧縮を行った  $Z_s$  を再構成したベクトルである  $r_s$  を作成する。 $r_s$  は  $p_t$  によって求めた評価観点ベクトル行列  $T$  を線形結合したものである。

このように、ABAE はフィルタリングされた文の埋め込み  $Z_s$  をその再構成ベクトル  $r_s$  にできるだけ歪みを少なく変換することを目的とした次元削減と再構成のプロセスにより、評価観点を表す上で重要な語の情報を保持することができる。

ABAE は評価観点ごとにベクトルを学習して、学習から得られたベクトルに基づき、各評価観点到に属する重要語を抽出し、単語集合を作成する。この単語集合に対して人手で意味解釈を行い、評価観点名のラベルを付与する。ここでの評価観点名は事前にデータセットによって定義されている評価観点名から選択する。しかし、人手による意味解釈は作成される単語集合の数が増加するにつれて人手による時間的・金銭的・労働的コストが大きくなるため困難になるが、Heらの手法では評価観点抽出の完全な自動化には至っていない。

### 2.2 単語集合の意味解釈に関する研究

Tsengらは、単語集合に対して意味解釈を容易にするために、与えられた単語集合に対して機械的に意味を捉えることを可能にする手法として、単語間の概念構造に基づく上位語を付与することを提案している[5]。この手法では、まず単語集合を作成する際に用いたベクトル表現から各単語の重みを計算し、重みに基づいて複数の代表語を抽出する。次に、各代表語に対して WordNet を用いて単語間の概念構造における上位語を求め、単語集合全体を表現しているラベルとして付与を行っている。その結果、単語集合の意味解釈が容易になったことを報告している。

ここで、単語間の概念構造における上位語・下位語の関係について説明する。単語間では上位下位の概念関係を持っているものがある。例えば、「家具」と「椅子」という単語の間には上位下位の概念関係が存在する。「家具」という単語は「椅子」という単語を概念的に包含しているため、「椅子」としての概念構造における上位語である。上位語は下位語をより抽象的に表現した単語であり、下位語は上位語よりも具体的に表現した単語である。

この手法で用いられている WordNet は単語間の概念構造に基づいた上位下位の関係を記載した辞書データベースである。

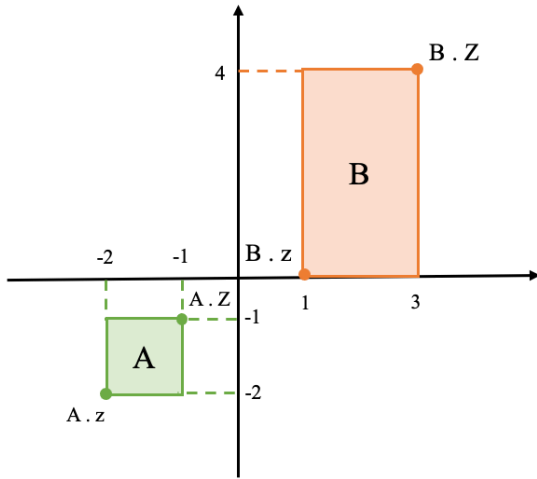


図2 Box Embeddings の実装イメージ

しかし、WordNet を用いる手法では辞書に含まれていない単語に対して概念関係を取得できないという問題がある。また、WordNet における登録単語数の問題から、一つの単語で複数の意味を持つような多義語に対して考慮することができず誤って特定される場合があるという問題も報告されている [8]。このことから、WordNet のような辞書型のデータベースを用いずに、単語間の概念構造に基づく上位下位の関係を取得する必要がある。

### 2.3 単語間の概念構造に基づいた領域表現に関する研究

本研究で単語間の概念構造を把握する際に用いた Word2Box という手法について紹介する [6]。

単語の分散表現はベクトルで表現されることが一般的とされており、Word2Vec などの単語ベクトルを生成するモデルでは単語をユークリッド空間上の点として捉えていた。しかし、単語をベクトルで表現すると概念構造における単語間の包含関係などが表現できないことが問題点とされている。この問題点を解決するために単語を長方形の領域で表現したものが Box Embeddings である [9]。

Box Embeddings は領域の広さや領域同士の重なりから意味的な広がりや単語間の包含関係を捉えることができる。概念構造における上位・下位関係がある単語ペアを Box Embeddings で表現した場合、上位語の領域表現が下位語の領域表現を含むように表示される。

Box Embeddings は  $n$  次元空間における領域表現を  $n$  個の始点と終点のペア中の最小、最大座標で表現できる。Box Embeddings によって獲得できる、ある集合  $X$  の要素  $x$  の領域表現は、以下の式のように直積で定義することができる。

$$\text{Box}(x) := \prod_{i=1}^n [x_i^-, x_i^+] = [x_1^-, x_1^+] \times \cdots \times [x_n^-, x_n^+] \quad (1)$$

図2は、Box Embeddings から得られる領域表現のイメージである。

この図では最小の座標である  $A.z$  と  $B.z$ 、最大の座標である  $A.Z$  と  $B.Z$  から、 $A$  と  $B$  の領域表現を獲得している。つまり、長方形の左下の角と右上の角の座標から領域を求め、面積などを計算することが可能である。

Word2Box は Box Embeddings を教師なし学習によって取得する手法である。Box Embeddings を取得するための教師なし学習にはテキストデータが多数収録されたデータセットの単語の分布に基づいて学習される。本研究の提案手法では2.2項の主張に基づいて上位語を文のラベルとして付与するため、Word2Box が適していると考えた。

ここで、Word2Box の学習方法について説明する。まず、テキストデータセットの一文から、ある単語を中心語として設定する。次に、同じ文中で中心語の周辺に出現する単語を文脈語として設定する。そして、Box Embeddings に変換した際に中心語と文脈語の領域表現における重なり部分が大きくなるように学習を行う。また、適当な単語をサンプリングして中心語と置き換えを行い、負例を作成する。置き換えを行った負例の単語と文脈語は Box Embeddings の領域表現における重なり部分が小さくなるように学習を行う。Word2Box は単語の多義性や概念的な広がりを表現することが可能な Box Embeddings を、特殊な教師データを必要とせず取得することができるため、本研究ではこの手法に基づいて提案を行った。

## 3 提案手法

本研究では、一つのレビュー文で述べられている評価観点を機械的に抽出することを目的としている。評価観点を機械的に抽出するためには、2.2の主張から上位語を取得する必要がある。したがって、提案手法では教師なし学習によって単語間の概念構造を把握可能な Word2Box を活用した。本研究では文に出現する単語の概念構造を把握したうえで、上位語を選択する手法を提案する。そして、選択された上位語を文の観点ラベルとして付与することで機械的な評価観点の抽出を行う。

### 3.1 文の主題を捉えるための方法

ここでは、Word2Box を用いて文で述べられている評価観点を推定する際に、文で中心の話題となっている主題をどのように捉えるかということについて説明する。

まず、Word2Box は学習データにおける単語の出現する分布から Box Embeddings を作成している。教師なし学習を行った Word2Box は学習元データセット内に出現した全単語に対して Box Embeddings を取得することが可能である。ここで出力される Box Embeddings は長方形で表現した際の左下の角の座標と右上の角の座標で出力される。例として Word2Box で “Apple” という単語の Box Embedding を取得した際に得られる出力を式 (2) に示す。

$$\text{Box}(\text{“Apple”}) = [[0.1345, 0.7966], [0.3894, 1.1778]] \quad (2)$$

式 (2) のように Word2Box によって長方形で表現する際の左下の角と右上の角にあたる点の座標を取得することができ、単語を領域で表現することが可能である。単語ごとに出力され

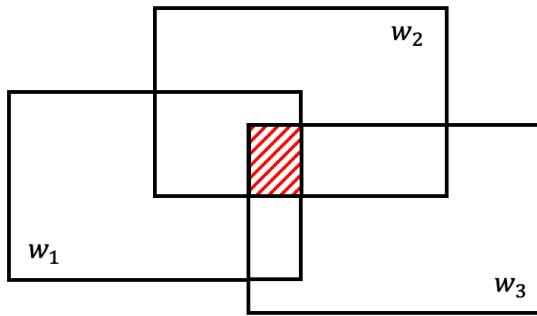


図 3 出現単語の重なり領域

る領域の広さはデータセット内の単語の出現回数に基づいており、単語の多義性を表現している。概念構造における上位語は下位語よりも多義性があり、出現回数が多くなるため出力される長方形の領域は広くなる。

また、意味が類似している単語ほど出現する文脈語が似ていることから、Box Embeddings 上で得られる座標が近くなるため領域同士の重なりが生まれる。このことから、文の出現単語間で領域同士の重なり数が多くなっている部分が文で中心的に述べられている主題を表現していると考えた。そのため、提案手法では出現単語間で生成される重なり領域が集中している部分に注目する。

提案手法を説明するにあたって評価観点を推定したい文を  $S$  とおき、その中で出現する単語を  $w = w_1, \dots, w_n$  とする。図 3 は出現単語を Word2Box によって領域で表した例で、斜線部分が最も重なり数が多い領域である。今回は簡易的に説明を行うため  $n = 3$  の場合を想定する。

まず、出現単語間で重なり領域が存在するかを確認する。Word2Box では、式 (3) のように単語ごとに領域表現における左下の角と右上の角の座標が与えられる。

$$w_i = [[lx_i, ly_i], [rx_i, ry_i]] \quad (3)$$

この座標から二単語間の重なり領域を求める。そして、全ての重なり領域を求めた後、重なり数が最も多い領域を求める。重なり数が最も多い領域の座標は、出現単語間の重なり領域における各座標  $lx, ly, rx, ry$  の最頻値から求めることができる。

重なり数を考慮する際に注意する点は、重なり数の最大値が複数存在する場合である。図 4 のように三つの単語が重なっている領域が二つ存在する時、文で述べられている主題が二つに分かれているため、どちらの領域を考慮するのかという問題が生じる。そのため、上位語を推定するためにどの領域に基づくのかを選択する必要がある。本研究では、意味が類似している単語同士の重なりが大きくなるという Word2Box の特徴に基づいて、最大重なり数の領域が複数ある場合、領域の面積が大きい方を考慮するとした。

### 3.2 上位語の推定

ここでは、2.2 項で説明した単語集合の意味解釈に上位語が

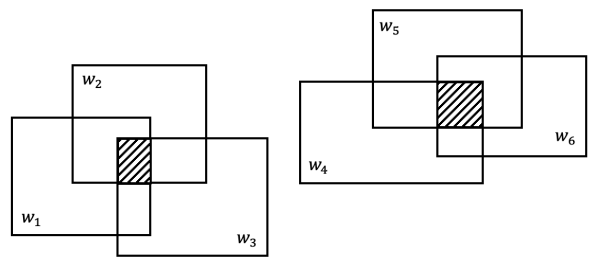


図 4 出現単語の重なり領域

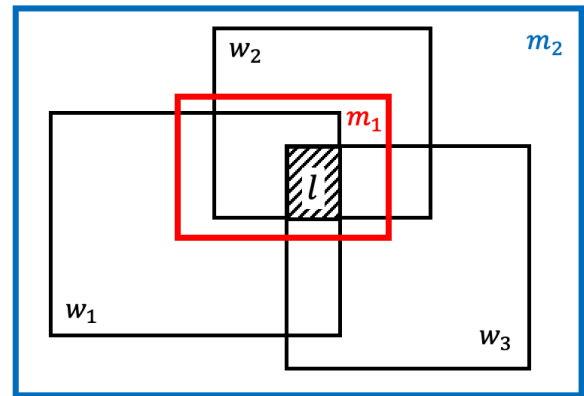


図 5 重なり領域を含んでいる語 A,B

有効であるという主張に基づいた上で、文中に出現した単語から Word2Box における領域表現を用いて、適切な上位語を選択する手法について説明する。

出現単語間で生成される重なり数が最大の領域部分が文の主題であることは前述したとおりである。Word2Box では、多義性が高い単語であるほど、領域表現が広くなるように学習が行われているため、文の主題を表している領域部分を含んでいる領域表現の単語が、文  $S$  における相応しい上位語であると考えた。上位語を決める際、文の主題を表す領域に対して上位語を表す領域の面積が大きすぎると、階層関係が離れているため適切な上位語とは見せせない。そのため、領域面積同士の差を見ることで適切な階層の上位語を推定する必要がある。

これを踏まえて、上位語を取得するための方法を説明する。上位語を決定するまでの全体の流れは、Algorithm 1 に示す。

提案手法の詳細を図 5 を用いて説明する。まず、文  $S$  の出現単語  $w_i$  に対して Word2Box を用いて領域表現を取得する。次に、3.1 で述べたように文で述べられている主題を捉えるために重なり数が最大になる領域を取得する。この領域を  $l$  とする。図 5 では、 $w_1, w_2, w_3$  の三単語が重なっている斜線部分が、重なり数が最大となる領域  $l$  である。

次に、Word2Box で取得可能な全ての領域表現から  $l$  と重なりがある単語を取得して上位語候補とする。上位語候補の領域面積を  $m_j$  とする。図 5 では、 $m_1$  と  $m_2$  の領域表現が  $l$  と重なっているため、この二単語が上位語候補となる。本稿では、上位語候補の中から適切な上位語を一語選択することを目的としている。上位語候補の中から、適切な上位語を選択するため

## Algorithm 1 提案手法の流れ

Input : 文  $s$

Output : 上位語  $w_{\text{sup}}$

- 1:  $B = \text{Word2Box}$  で取得可能な領域表現集合を全て取得
  - 2:  $B_s = \text{文 } s \text{ に出現する単語の領域表現集合}$
  - 3:  $W = \text{文から出現単語を抽出}$
- foreach: 文の出現単語  $w$  in  $W$ :
- 4: 文の出現単語  $w$  の領域表現  $b_w$  を作成
  - 5: 領域表現の集合  $B_s$  に  $b_w$  を追加
  - 6:  $l = \text{Box 集合から最多な重なり領域を導出}$
  - 7:  $B_c = B_s$  集合の要素を除く  $l$  と重なっている  $B$  の部分集合を取得
- foreach:  $bc$  in  $B_c$ :
- 8:  $bc.\text{score} = \text{式 (4)}$
  - 9: 式 (4) によって求められる  $bc.\text{score}$  が最小の  $bc \in B_c$  を適切な上位語として決定

に、上位語候補の面積に対してどれだけ  $l$  の領域を含んでいるかを求める。  $l$  と上位語候補が重なっている部分の面積を  $lm_j$  として、式 4 で差を求める。

$$m_j - lm_j \quad (4)$$

図 5 の上位語候補である  $m_1$  と  $m_2$  は  $l$  の領域を全て含んでいるため、  $lm_{m_1}$  と  $lm_{m_2}$  は  $l$  の面積と同じである。そして、式 (4) の計算を行い、差が最も少ないものを上位語として文  $S$  に付与する。図 5 の例では、重なり領域と面積の差が少ない  $m_1$  が上位語として選択され、文  $S$  にラベルとして付与される。

## 4 評価実験

本節では、提案手法である文の観点ラベルを付与する手法の有効性を確認するために行った、文単位で観点的分類精度を測る評価実験について説明する。

まず、4.1 で提案手法で用いた Word2Box の学習について説明する。4.2 で評価実験で使ったデータセットについて説明を行い、4.3 で評価実験の方法について述べる。そして、4.4 で結果と考察について述べる。

### 4.1 Word2Box の学習

2.3 で説明した Box Embeddings を取得するための教師なし学習に用いるデータについて説明する。今回、学習に用いたデータは WikiText-103 である [10]。WikiText-103 は、Wikipedia の記事から構成されており、豊富な語彙が収録されたコーパスである。本データを学習に用いる際に、まず WikiText-103 コーパス中の 1,158,870 文に対して前処理を行った。前処理では句読点の削除、全ての単語を小文字に変換、出現回数が 100 回未満であるトークンの削除、数字の削除を行った。上記の処理を施したデータセットを学習データとして学習を行う。

### 4.2 使用データ

本研究で行う観点抽出は文単位で行うため、評価実験では一文ごとにその文が何の観点を持っている文なのか識別できるラベルが付与されたデータセットを用いる必要がある。そ

表 1 評価実験で使った観点

大まかな観点	細かな観点
Art	broadcast, film, music, painting, written art
Building	hospital, hotel, library, restaurant, sports facility, theater
Event	war, disaster, election, protest, sports event
Location	bodies of water, GPE, island, mountain, park, road
Organization	company, education, government, newspaper, political party, religion, show organization, sports league, sports team
Person	medical, actor, artist, athlete, director, politician, scholar, soldier
Product	airplane, car, food, game, ship, software, train, weapon

こで、今回評価実験を行う際に用いたデータセットは A Few-shot Named Entity Recognition Dataset (Few-NERD) である [11]。Few-NERD は主に固有表現抽出タスクで用いることを目的に作成されたデータセットである。文で述べられている話題を認識して手動でアノテーションされたデータであるため、文の主題を捉えるという点で本研究の評価実験で用いることが可能と考えた。大まかな観点と細かな観点的ラベルが文単位で付与されたデータであり、Wikipedia から 188,238 文を取得し、4,601,160 語から構成されている。

今回、評価実験で使ったデータは Few-NERD の中で大まかな観点として付与されている 7 種類の観点と細かな観点として付与されている 47 種類の観点的ラベルを用いた。実際に用いた観点的ラベルの詳細は表 1 に示す。Few-NERD を用いた理由は、観点的ラベルによる推定精度の違いを確認するためである。

また、評価実験を行う際に、一文に対して複数の観点的ラベルが付与されているものは含まず、一文に対して単一のラベルが付与されているデータのみをテストデータとして使用する。

### 4.3 評価方法

評価実験では、Few-NERD で定義されている観点的ラベルを候補となる語として、それぞれで Box Embeddings を求め、文に出現する単語との重なり領域に基づいて、文の観点的ラベルを表現しているラベルとして候補から一語選択している。

提案手法の有効性を示すために、ベースラインとして 2.1 で紹介した ABAE を用いる。比較手法で用いる ABAE は、観点的ラベルのベクトルを作成する必要があるため、Few-NERD データセットから 65,000 件のデータを学習用データとして、七つの観点的ラベルのベクトルを作成した。そして、学習によって作成された七つの観点的ラベルに基づいて作成された重要語の単語集合を見て、意味解釈を行いどのベクトルがどの観点的ラベルを表しているのかを手で意味解釈を行った。意味解釈を行った観点的ラベルに基づいて、文単位でどの観点的ラベルに属するかを判定し、ラベルを付与する。

表 2 比較実験の結果

	Macro-F1	Micro-F1
ベースライン	0.150	0.168
提案手法	<b>0.153</b>	<b>0.177</b>

表 3 比較実験の結果

	Macro-F1	Micro-F1
ベースライン	<b>0.027</b>	0.034
提案手法	0.021	<b>0.039</b>

Few-NERD から 4,600 件をテストデータとして評価を行った。提案手法と比較手法それぞれで付与された観点が Few-NERD によって付与されている正解ラベルと一致しているかを評価する。評価指標には適合率、再現率、各ラベルの F1 値から計算される Macro-F1 値、Micro-F1 値を用いて、評価を行う。正解ラベルを正しく予測できたものが True Positive (TP)、正解ラベルではない文を誤って正解ラベルと予測したものが False Positive (FP)、正解ラベルの文を正解ラベルではないと予測してしまったものが False Negative (FN) である。適合率 (Precision) は、予測したラベルの中で実際に正解ラベルを予測できた割合で、式 (5) のように計算できる。再現率 (Recall) は実際に正解ラベルが正解となるものの中で、正解を予測できた割合で、式 (6) のように計算できる。F1 値 (F1-score) は適合率と再現率の調和平均であるため、式 (7) のように求めることができる。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Macro-F1 値は、分類モデル全体を評価するための指標である。観点集合を  $T$  とした時に各観点  $t$  の F1 値を  $F_t$  とする。Macro-F1 値は各観点の F1 値を平均したものであるため、式 (8) で求めることができる。観点集合の要素数  $|T|$  は表 1 の通り、大まかな観点の場合 7 であり、細かな観点は 47 である。

$$\text{Macro-F1} = \frac{1}{|T|} \sum_{t \in T} F_t \quad (8)$$

Micro-F1 値は、観点間のデータ偏りを考慮した評価指標である。Micro-F1 値は各観点の結果をまとめて計算するため、適合率の平均と再現率の平均を用いて、式 (7) の計算を行うことで求めることができる。

#### 4.4 結果と考察

表 2 に 7 種類の観点を分類した際の Macro-F1 値と Micro-F1 値を示す。次に、表 3 に 47 種類の観点を分類した際の Macro-F1 値と Micro-F1 値を示す。

提案手法で大まかな観点を推定した際の Macro-F1 値と Micro-F1 値は、ベースラインよりもわずかに高いが、ほと

んど同程度の精度である。また、細かな観点を推定した際の Micro-F1 値はベースラインを上回っているが、Macro-F1 値は下回る結果になった。提案手法とベースラインはほとんど同程度の精度であるが、どちらも精度としては極めて低い結果になった。

提案手法で精度が低くなっている要因として考えられることは、Few-NERD で付与されているラベルの単語で上位語として適切なものではないものが存在しているということが挙げられる。Few-NERD によって付与されている観点ラベルを上位語候補として扱っているが、観点によって出力される領域表現の面積が極端に小さい語が存在する。Word2Box では領域面積の広さが多義性を表現しているため、面積が小さい観点ラベルの語は上位語として適切ではない。これらのことから、Few-NERD で付与されている観点ラベルは文で主題として述べられている話題を表しているが、必ずしも出現単語の上位語として機能するラベルではないということである。

これは、Few-NERD が上位語を用いて文の主題を推定する本研究のタスクと合っていないことから生じた問題である。そのため、データセットを変更して追加で実験を行い、結果に応じて新たに提案手法を改良するための考察を行う必要がある。

## 5 おわりに

本研究では、文書を閲覧する際にかかる負担を軽減するために行われる文書要約において重要である、文書の観点抽出手法に関する提案を行った。具体的には、教師なし学習によって Box Embeddings と呼ばれる領域表現を取得することができる Word2Box という手法を用いて、文に出現する単語間の領域表現の重なりから概念構造における上位語を選択し、観点のラベルとして付与を行った。評価実験では、提案手法が人手による意味解釈処理を行うことなしに既存研究をわずかに上回る分類精度が得られることを確認した。

今後の課題は、4.4 で述べた、評価実験で用いた Few-NERD データセットのラベルが上位語として機能していないことが問題だと考える。したがって、評価実験は比較対象として用いた ABAE で使用されていた Citysearch コーパス [12] が妥当であると考えられる。Citysearch コーパスはレストランレビューに対して文単位で評価観点ラベルが付与されたデータであり、評価観点を推定する今回のタスクに適しているからである。

また、今回の評価実験では提案手法の分類精度を既存手法と比較するため、ラベル候補からラベルを選択する形式とした。しかし、文書要約を行う際は候補からの選択ではなく Word2Box から取得可能な全単語からラベルを付与することが求められる。そのため、候補がない状態で選択された上位語が観点として適切なものか評価を行うことも今後の課題とする。

## 謝 辞

本研究の一部は JSPS 科研費 JP19H04218 の助成を受けたものである。ここに記して謝意を表す。

## 文 献

- [1] Min Yang, Qiang Qu, Ying Shen, Qiao Liu, Wei Zhao, and Jia Zhu. Aspect and Sentiment Aware Abstractive Review Summarization. In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 1110–1120, 2018.
- [2] Ziyu Guan, Long Chen, Wei Zhao, Yi Zheng, Shulong Tan, and Deng Cai. Weakly-Supervised Deep Learning for Customer Review Sentiment Classification. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 3719–3725, 2016.
- [3] Qian Liu, Zhiqiang Gao, Bing Liu, and Yuanlin Zhang. Automated Rule Selection for Aspect Extraction in Opinion Mining. In *Proceedings of the 24th International Conference on Artificial Intelligence*, pp. 1291–1297, 2015.
- [4] Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. An Unsupervised Neural Attention Model for Aspect Extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, pp. 388–397, 2017.
- [5] Yuen-Hsien Tseng, Chi-Jen Lin, Hsiu-Han Chen, and Yu-I Lin. Toward Generic Title Generation for Clustered Documents. In *Information Retrieval Technology*, pp. 145–157, 2006.
- [6] Shib Dasgupta, Michael Boratko, Siddhartha Mishra, Shriya Atmakuri, Dhruvesh Patel, Xiang Li, and Andrew McCallum. Word2Box: Capturing Set-Theoretic Semantics of Words using Box Embeddings. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, pp. 2263–2276, 2022.
- [7] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Vol. 2, pp. 3111–3119, 2013.
- [8] 山田一郎, 呉鍾動, 鳥澤健太郎, 黒田航, 風間淳一, 村田真樹. Wikipedia を利用した日本語 WordNet への用語追加の検討. 言語処理学会年次大会発表論文集, Vol. 16, pp. 948–951, 2010.
- [9] Tejas Chheda, Purujit Goyal, Trang Tran, Dhruvesh Patel, Michael Boratko, Shib Sankar Dasgupta, and Andrew McCallum. Box Embeddings: An open-source library for representation learning using geometric structures. In *Conference on Empirical Methods in Natural Language Processing*, pp. 203–211, 2021.
- [10] Stephen Merity, Caiming Xiong, James Bradbury, and Richard Socher. Pointer Sentinel Mixture Models. In *International Conference on Learning Representations*, pp. 1–15, 2016.
- [11] Ning Ding, Guangwei Xu, Yulin Chen, Xiaobin Wang, Xu Han, Pengjun Xie, Haitao Zheng, and Zhiyuan Liu. FewNERD: A Few-shot Named Entity Recognition Dataset. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, Vol. 1, pp. 3198–3213, 2021.
- [12] Gayatree Ganu, Noémie Elhadad, and Amélie Marian. Beyond the Stars: Improving Rating Predictions using Review Text Content. In *International Workshop on the Web and Databases*, Vol. 9, pp. 1–6, 2009.