

# 火災状況の自動理解のためのオントロジー構築に向けた 出火原因の自動検出

松場 建都<sup>†</sup> 中村 匡秀<sup>†,††</sup> 佐伯 幸郎<sup>†††</sup>

<sup>†</sup> 神戸大学 〒657-8501 神戸市灘区六甲台町 1-1

<sup>††</sup> 理化学研究所・革新知能統合研究センター 〒103-0027 東京都中央区日本橋 1-4-1

<sup>†††</sup> 高知工科大学 〒782-8502 高知県香美市土佐山田町宮ノ口 185

E-mail: <sup>†</sup>matsuken@ws.cs.kobe-u.ac.jp, <sup>††</sup>masa-n@cs.kobe-u.ac.jp, <sup>†††</sup>saiki.sachio@kochi-tech.ac.jp

**あらまし** 神戸市消防局の作成している火災調査報告書には、出火原因や、火災に至った問題点などの情報が含まれているが、これらの情報を集約し、活用するという取り組みは進んでいない。火災調査報告書は多くの項目が文章として記録されているため、特徴抽出を機械的に行うことは困難である。そこで、日本語自然言語処理の技術を活用することで、神戸市消防局の持つ各種データから、火災の出火原因である発火源、経過、着火物の機械的な抽出を行った。その結果、発火源を 93.1%、着火物を 77.4% の精度で抽出でき、経過の抽出に対しては、42.4% のケースでおおむね妥当な結果が得られた。

**キーワード** 火災、出火原因、テキストマイニング、火災調査報告書、自然言語処理、形態素解析、構文解析

## Automatic Detection of Fire Causes for Ontology Development for Automatic Understanding of Fire Situation

Kento MATSUBA<sup>†</sup>, Masahide NAKAMURA<sup>†,††</sup>, and Sachio SAIKI<sup>†††</sup>

<sup>†</sup> Kobe University Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan

<sup>††</sup> Riken AIP, 1-4-1 Nihon-bashi, Chuo-ku, Tokyo 103-0027

<sup>†††</sup> Kochi University of Technology 185 Miyanokuchi, Tosayamada, Kami City, Kochi, 782-8502, Japan

E-mail: <sup>†</sup>matsuken@ws.cs.kobe-u.ac.jp, <sup>††</sup>masa-n@cs.kobe-u.ac.jp, <sup>†††</sup>saiki.sachio@kochi-tech.ac.jp

**Abstract** The fire investigation report prepared by the Kobe City Fire Department contains information on the causes of the fire and the problems that led to the fire, but efforts to consolidate and utilize this information have not progressed. Since many items in fire investigation reports are recorded as text, it is difficult to extract features mechanically. By using Japanese natural language processing technology, we were able to extract the ignition source, the course of the fire, and the ignition materials from various data provided by the Kobe City Fire Department. As a result, we were able to extract the ignition source with an accuracy of 93.1% and the ignition material with 77.4%. For the extraction of the process, the results were generally reasonable in 42.4% of the cases.

**Key words** fire, cause of fire, text mining, fire research report, NLP, morphological analysis, syntax analysis

### 1. はじめに

兵庫県神戸市における火災の発生件数は減少傾向にあり、令和 2 年における火災件数 385 件は、平成 23 年の 640 件と比較すると、その件数は約 60% にまで減少している [1]。火災件数の減少には、地域特性や社会情勢の変化などの要因が複雑に関係しており、その理由を解明することで、火災減少のための具体的な防火施策の構築や、広報活動に役立てることができる。

神戸市消防局では、すべての火災の原因及び損害並びに関係

者の行動等を明らかにし、火災予防施策及び警防対策に必要な基礎資料等消防行政を推進するため、発生した全火災に対して原因、対応等を調査し、火災調査報告書を作成している [2]。火災調査報告書には、各火災の全体の概要、り災程度、予防行政への反映事項などが記載されている。これらのうち、出火日時や、出火原因として重要な 3 つの要素である発火源、経過、着火物といった情報は、報告書とは別のデータベース（以下、このデータベースに登録されるデータのことを火災統計データと呼称する）に記録され、火災の統計報告に活用されている。

火災調査報告書には、火災統計データには登録されていない、地域性の高い情報や、火災統計データに登録する際に抽象化された情報、省略された情報等を含んでいる。しかし、この火災調査報告書に焦点を当てた分析は積極的には行われておらず、消防署内で保管されたままになっており、詳細な火災の状況や教訓などの貴重な情報が未来の防災のために十分に活用されているとはいえない。そこで、火災調査報告書のデータを活用し、火災の事案の様々な角度から分析・調査する研究に取り組んでいる。

本稿ではその第一歩として、火災調査報告書の火災概要の文章を機械的に分析し、火災の出火原因として重要な要素である発火源、経過、着火物を自動的に抽出することを目指す。研究のキーアイデアとして、形態素解析、構文解析などの**日本語自然言語処理**の技術を活用する。

課題解決へのアプローチとして、火災調査報告書に対し以下のA1～A3の操作を順に実施する。

A1: 専門辞書に基づく発火源の抽出

A2: キーワードと構文木に基づく着火物の抽出

A3: 構文的意味に基づく経過の抽出

A1では、COTOHA API [3] を用いて火災調査報告書の火災概要の文章を文節単位に分解・構文木を作成し、各文節と専門辞書に登録されている発火源一覧を照合することで、発火源の抽出を行う。A2では、A1で作成した火災概要の構文木の中で、着火物を含んだ文節とリンクする可能性の高い単語を「着火物キーワード」として設定し、着火物キーワードを含む文節とリンクする文節を探索し、着火物として適切な単語を抽出する。A3では、A1、A2で抽出された発火源、着火物を含むそれぞれの文節から、構文木の根に至るまでのすべての文節から成り立つ文章を抽出し、発火源、着火物を含む火災概要を要約した文章を作成する。

上記のアプローチを用いて、神戸市消防局の作成した2020年1月出動分の火災調査報告書33件から発火源、経過、着火物の抽出を試みた結果、発火源は33件中27件において、正しい発火源が抽出された。また着火物は、33件中24件において、抽出した複数の文節の中に着火物を含む文節が含まれていた。また33件中17件において、着火物を含む文節のみを抽出した。最後に経過の抽出を行い、消防局員4名による評価を行った結果、42.4%のケースでおおむね妥当な結果が得られたと評価された。

## 2. 準備

### 2.1 発火源、経過、着火物

消防において、**発火源・経過・着火物**の3つの組み合わせが火災の原因とされており、発火源、経過、着火物の定義は、それぞれ以下になっている [4]。

**発火源:** 出火に直接関係し、またはそれ自体から出火したもので、原則として「火」または「高温体」になる。

**経過:** 火災に至るまでの経過のことで、「現象」・「状態」・「行為」に分類できる。

**着火物:** 発火源によって最初に着火したもの。

この3つの要素に着目して火災に至る経緯を整理、分類し、それに対して適切な予防施策を実施することで、火災の発生を効果的に抑制することができる。

### 2.2 神戸市消防局の持つデータ

神戸市消防局では、通報を受けて出動したすべての火災について、火災調査規定 [2] に基づき、火災の出火原因、火災初期の対応、損害規模などの調査を行っている。調査の結果は、**火災調査報告書**や**火災統計データ**といった形式で記録されている。

#### 2.2.1 火災調査報告書

消防局員によって調査された火災に関する情報は、**火災調査報告書**という形式でまとめられ、消防署内に保管されている。火災調査報告書は、主に以下の内容から構成されている。

- 1: 火災の概要および火災原因概要
- 2: 火災時の状況（発見、通報、初期消化、延焼拡大等）
- 3: 避難の状況（使用器具等、避難経路）
- 4: 死傷者発生状況（場所、経過等）
- 5: 防火管理の状況（防火管理者、自衛消防隊の対応等）
- 6: 防災設備の状況（奏功、不奏功、防火区画等）
- 7: 過去の状況（用途変更、査察、違反等）
- 8: 問題点または教訓

また、火災調査報告書に記載された情報のうち、文章で記載されている状況や教訓を除く、出火日時や出火原因といった情報は、統計報告にまとめやすい形に変換し、火災統計データとして登録される。

火災調査報告書の項目の一つである**火災概要文**には、火災統計データに登録する際に抽象化されたり、省略された情報を含んでいる。

#### 2.2.2 火災統計データ

火災調査報告書に記載した項目のうち、火災の出火日時や出火原因（発火源、経過、着火物）などの情報は、火災調査報告書とは別に電子データとして保存され、統計報告などに活用されている。このデータを**火災統計データ**と呼ぶ。火災統計データのサンプルを表1に示す。火災統計データ1件につき、132個の項目が登録されており、本稿ではその中の「発火源」「経過」「着火物」の項目に着目する。

#### 2.2.3 専門辞書

火災調査報告書の内容を火災統計データに登録する際、統計報告の際に火災傾向などをとらえやすくするため、発火源、経過、着火物はそのまま登録されるのではなく、統計報告に活用しやすい形に抽象化されて登録される。例として、着火物が「タオル」であると火災調査報告書に記載されていた場合、火災統計データには「繊維製品」として登録される。このように、火災調査報告書に記載された発火源、経過、着火物を、どのように変換して統計データに登録すべきかが記載された一覧を、それぞれの**専門辞書**と呼ぶ。

### 2.3 現状の問題

火災調査報告書は、消防局内に閉じた形で保管されており、電子化への取り組みや、報告書のテキストマイニングによって防火施策に活用するといった取り組みは行われていない。そのため、火災統計データに登録する際に失われた情報や、火災統

表 1 火災統計データのサンプル

| No | 出火日時（年月） | 火災種別 | 火元焼損程度 | 発火源   | 経過       | 着火物   | 統計上の原因   | ... |
|----|----------|------|--------|-------|----------|-------|----------|-----|
| 1  | 2020/01  | 建物   | ぼや     | ガスこんろ | 不適の用に用いる | 繊維製品  | ガスコンロ    |     |
| 2  | 2020/01  | 車両   | 部分焼    | たばこ   | 引火する     | 不明    | 不明       |     |
| 3  | 2020/01  | 建物   | 全焼     | ライター  | 放火       | 第一石油類 | 放火、放火の推定 |     |

計データに登録されない情報は、活用されないままになっている。また、火災概要文はテキストデータであり、火災概要文の書式・構成は書き手によって異なる。そのため、火災統計データのような分類や分析が困難であり、機械的な分析には日本語自然言語処理のような技術が必要である。

#### 2.4 日本語自然言語処理

自然言語処理とは、人間が日常的に利用している自然言語を機械的に処理する技術の総称である[5]。自然言語処理は扱う言語によって大きく処理が異なり、日本語を処理する自然言語処理の基礎技術には形態素解析や構文解析がある。

##### 2.4.1 形態素解析

日本語にかかわる処理を機械的に行うためには、文章中の単語を認識し、分かち書きを行う必要がある。日本語では、単語の識別、活用語処理、品詞の同定が密接に関連するため、これらの処理が同時に行われることが普通であり、これらの全体的な処理を形態素解析と呼ぶ[6]。

##### 2.4.2 構文解析

日本語自然言語処理における形態素解析の次のステップとして、構文解析（係り受け解析とも呼ぶ）が行われる[7]。これは、文節・単語単位に分解済みの文章を対象に、文法や統語論を元に単語間の関係を決定する作業である。構文解析により得られた文章内の文節間の修飾関係は、木構造の形で可視化することができ、これを構文木と呼ぶ。

##### 2.4.3 COTOHA API

本研究では、日本語文章の形態素解析、および構文解析に、COTOHA API[3]を用いた。COTOHA APIは、NTTコミュニケーションズが提供する自然言語処理 API プラットフォームであり、形態素解析、構文解析といったサービスをインストール不要で利用することができる。

実際に、ある火災概要文を参考に筆者が作成した「何者かがライター等で木製ベンチ上の物品に火を着け、放火した。」という文章に対し、COTOHA API の構文解析機能を用いた例を示す。COTOHA API の構文解析 API のパラメータに文章を入力して送信すると、構文解析の結果が JSON 形式で得られる。得られた構文解析結果を構文木として可視化した結果を図 1 に示す。なお、JSON ファイルから構文木への可視化には、Python のオープンソースグラフ描画ライブラリである Python-Graphviz を用いた。

図 1 内の丸は、その文章における文節の一つを表す。そのため、「何者かがライター等で木製ベンチ上の物品に火を着け、放火した。」という文章は、7つの文節から構成されることがわかる。また、文節から文節に伸びる矢印は、2つの文節が何らかの関係で係り受けを持っていることを表している。矢印に書かれた単語は意味ラベルであり、2つの文節がどのような修飾関

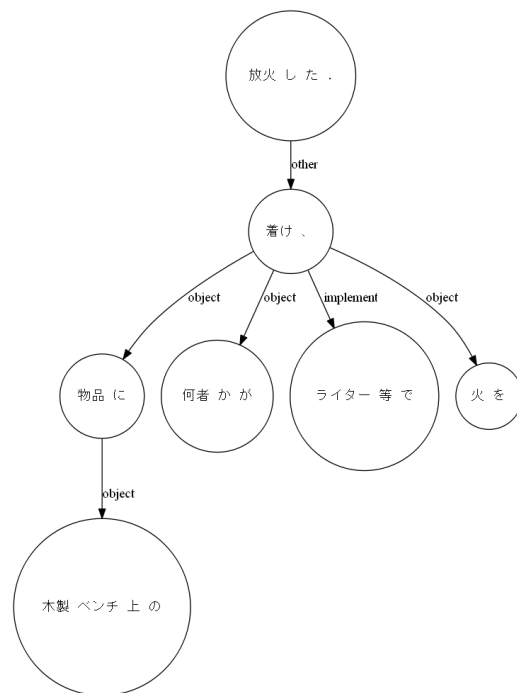


図 1 COTOHA API の構文解析により得られた構文木

係にあるかを表している。COTOHA API の構文解析 API で出力される文節間の意味ラベルは全 24 種あり、その一部を表 2 に示す。

### 3. 提案手法

#### 3.1 研究目的とキーアイデア

本稿の目的は、火災調査報告書に記載されている火災概要の文章から、火災の出火原因として重要な 3 つの要素である、発火源、着火物、経過を抽出することである。キーアイデアとして、神戸市消防局の持つ火災統計データ、および各種専門辞書と、日本語自然言語処理の技術を活用し、発火源、着火物、経過を抽出する手法を提案する。

#### 3.2 全体のアプローチ

研究目的達成のために、以下の A1～A3 のアプローチを、Python を用いて順に実施する。

- A1: 専門辞書に基づく発火源の抽出
- A2: キーワードと構文木に基づく着火物の抽出
- A3: 構文的意味に基づく経過の抽出

#### 3.3 A1: 専門辞書に基づく発火源の抽出

A1 では、火災統計データの登録された発火源を利用し、火災調査報告書の火災概要文から発火源の抽出を行う。火災概要文中における発火源は、火災統計データに登録された発火源と類似した文言で出現することが多い。そこで、COTOHA API を

表2 COTOHA API が出力する意味ラベルの一例

| 意味ラベル名称           | 説明                      | 例              |
|-------------------|-------------------------|----------------|
| object            | 動作・変化の影響を受ける対象          | 電話帳に ← 着け      |
| goal              | 事象の主体または対象の最後の位置        | 雑品に ← 延焼拡大した   |
| agent             | 有意動作を引き起こす主体            | 左裾付近が ← 触れた    |
| agentnonvoluntary | 動作を引き起こす主体（主体に意思性がない場合） | タオルが ← 燻焼して    |
| place             | 事象の成立する場所               | 日除けネットに ← 放火した |

用いて文節単位に分解した火災概要の文中から、火災統計データに登録された発火源と最も近い文字列を含む文節を抜き出すことで、発火源の抽出を行う。文字列の近さの判定方法として、火災概要文の全文節に対し、火災統計データに登録された発火源の文字列に含まれる文字数を調べ、重複する文字数が最も多い文節を選択する。

### 3.4 A2: キーワードと構文木に基づく着火物の抽出

着火物は発火源と違い、物体が「火」「高温体」に限られないため、火災統計データに登録する際にの抽象化度合が発火源より大きい。そのため、発火源抽出のような文字列比較による抽出は現実的ではなく、着火物を含む一文節を的確に抽出することは困難である。

そこで A2 では、火災概要文の構文木において着火物とリンクすることが多い着火物キーワードを利用し、着火物を含むであろう文節を複数個抽出する。火災概要文において、「～に着火」「～が接触し」「～に延焼し」のように、着火物と修飾関係を持つ場合が多い動詞が存在する。本稿では、そのような動詞を「着火物キーワード」と定義し、着火物キーワードを含む文節とリンクを持つ文節内の名詞を、着火物の候補として抽出する。着火物キーワードの選定は、2020 年 1 月の火災概要文の構文解析結果を考慮し、「着火」「着け」「つけ」「接触」「触れ」「延焼」「拡大」「放火」「焼損」「発火」の順に 10 種類とした。また COTOHA API では、文節間の係り受け関係を意味ラベルとして出力するが、着火物キーワードと着火物の間の意味ラベルは、object あるいは goal であることが多い。そのため、着火物キーワードを含む文節とリンクする文節との間の意味ラベルを調べ、それが object, あるいは goal であれば、その文節が着火物を含む可能性が高い。

これらのことから、火災概要本文から着火物を抽出するために、以下の A2-1～A2-5 の操作を実施する。

A2-1: 火災概要本文を COTOHA API の構文解析 API を用いて文節単位に分解し、構文木を出力する。

A2-2: 着火物キーワードを「着火」から順に参照し、着火物キーワードが含まれる文節があれば、その文節を  $c_1, c_2, \dots, c_n$  として探索キュー queue に追加し、次のステップ A2-3 に進む。着火物キーワードが含まれる文節が存在しない場合、着火物抽出失敗として処理する。

A2-3: queue から文節を一つ取り出し、これを文節 C とする。構文木において、C を根とする部分木の中で、C とリンクする文節を  $d_1, d_2, \dots, d_m$  とする。

A2-4: C とリンクする m 個の文節に対し、C と  $d_j$  の文節間の意味ラベルを調べる。このとき、 $d_j$  が「火を」「何者か」

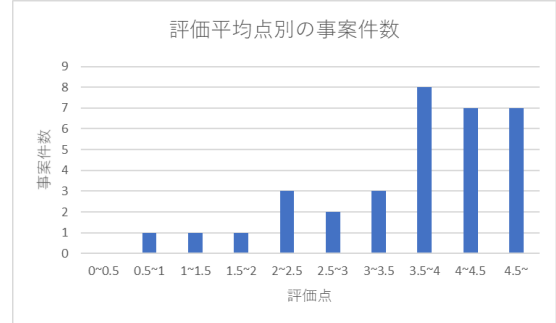


図2 評価平均点別の事案件数

といった、火災概要本文に頻出し、かつ着火物抽出とは関係のないフレーズを含む文節はスキップする。意味ラベルが object, あるいは goal であった場合、 $d_j$  を着火物を含む文節として抽出する。C と  $d_j$  間の意味ラベルが manner であった場合、 $d_j$  の次の文節を調べ、意味ラベルが object である文節を着火物を含む文節として抽出する。

A2-5: 着火物を含む文節が A2-4 で見つかった場合、着火物の抽出を終了する。見つかっていなかった場合、m 個の文節  $d_1, d_2, \dots, d_m$  を queue に追加し、A2-3 に戻る。

### 3.5 A3: 構文の意味に基づく経過の抽出

A3 では、A1, A2 で抽出された発火源、着火物を利用し、経過の抽出を行う。火災概要文の構文木において、3.3 節、3.4 節で明らかにした発火源、着火物を含む文節を基準に、火災概要文全体の要約を行うことで、火災における経過の抽出を行う。火災概要文の構文木において、「(発火源)を～」「(着火物)に～」のように、着火物、あるいは発火源を含む文節からつながる文節には、発火源からどのように着火物に燃え移ったかが記載されることが多い。そこで、発火源を含む文節から構文木の根に至る文章と、着火物を含む枝から構文木の根に至る文章を組み合わせ、発火源と着火物を含む文章を抽出する。

具体的なアルゴリズムは以下の A3-1～A3-3 の通りである。

A3-1: 発火源、着火物のいずれかが抽出できていない場合、経過の抽出を失敗として処理する。

A3-2: 発火源を含む文節を  $c_s$ 、着火物を含む枝の末端の文節を  $c_f$  とする。

A3-3:  $c_s$  から構文木の根に至るすべての文節と、 $c_f$  をから構文木の根に至るすべての文節を含む文章を作成する。

## 4. 評価実験とその結果

本章では、3. 章で提案した手法の評価のため、実際の火災調査報告書の火災概要文から、発火源、経過、着火物の抽出を行う。

表3 発火源、着火物、経過の抽出評価

| 出動番号         | 3      | 116    | 171    | 177  | 187    | 223  | 238    | 310    | 333    | 382    | 404    |
|--------------|--------|--------|--------|------|--------|------|--------|--------|--------|--------|--------|
| 発火源の抽出評価     | ○      | ○      | ○      | ○    | ○      | ○    | ○      | ○      | ○      | ○      | ○      |
| 着火物の抽出評価     | 1      | 4      | 2      | 4    | 4      | 4    | 3      | 2      | 4      | 4      | 4      |
| 3要素の抽出評価の平均値 | 2.25   | 4.25   | 3.75   | 3.5  | 4.5    | 5    | 4.25   | 3.25   | 4.5    | 3.75   | 4.5    |
| 3要素の抽出評価の分散  | 0.1875 | 0.1875 | 0.6875 | 0.25 | 0.25   | 0    | 0.6875 | 3.6875 | 0.25   | 0.6875 | 0.25   |
| 出動番号         | 434    | 437    | 438    | 448  | 474    | 491  | 504    | 535    | 545    | 565    | 568    |
| 発火源の抽出評価     | ○      | ○      | ○      | ○    | ○      | ○    | ○      | ×      | ○      | ×      | ○      |
| 着火物の抽出評価     | 4      | 1      | 3      | 3    | 3      | 4    | 2      | 2      | 4      | 4      | 4      |
| 3要素の抽出評価の平均値 | 2.5    | 3      | 3.75   | 4.5  | 3.75   | 4    | 2.75   | 0.75   | 3.75   | 2      | 4.25   |
| 3要素の抽出評価の分散  | 1.25   | 2.5    | 0.6875 | 0.25 | 0.6875 | 0.5  | 3.6875 | 0.1875 | 0.6875 | 1      | 0.6875 |
| 出動番号         | 620    | 628    | 668    | 688  | 689    | 715  | 726    | 738    | 783    | 821    | 838    |
| 発火源の抽出評価     | ○      | ○      | ○      | ×    | ○      | ×    | ○      | ×      | ○      | ○      | ×      |
| 着火物の抽出評価     | 4      | 3      | 1      | 1    | 4      | 4    | 1      | 3      | 4      | 3      | 4      |
| 3要素の抽出評価の平均値 | 4.25   | 4.25   | 1.75   | 1    | 4.5    | 3.5  | 3.25   | 2.25   | 4.5    | 4      | 3.75   |
| 3要素の抽出評価の分散  | 0.1875 | 0.1875 | 0.6875 | 0    | 0.25   | 0.25 | 0.1875 | 0.1875 | 0.25   | 0.5    | 0.1875 |

表4 着火物抽出の主な失敗要因

| 出動番号 | 主な失敗要因                  |
|------|-------------------------|
| 3    | 着火物キーワードと着火物の間の意味ラベルが特殊 |
| 171  | 火災概要文に着火物の記載なし          |
| 310  | 火災概要文に着火物の記載なし          |
| 437  | 着火物キーワードと着火物の間の意味ラベルが特殊 |
| 504  | 火災概要文に着火物キーワードなし        |
| 535  | 火災概要文に着火物キーワードなし        |
| 668  | 着火物キーワードの設定が不適切         |
| 688  | 着火物キーワードと着火物の間の意味ラベルが特殊 |
| 726  | 着火物キーワードと着火物の間の意味ラベルが特殊 |

表5 評価点が2点未満の火災概要要約文

| 出動番号 | 抽出された文章                                     |
|------|---------------------------------------------|
| 3    | 電気ストーブの上部を焼損させた。                            |
| 535  | (抽出されず)                                     |
| 565  | ドラム缶付近に置かれた廃材、雑品等に着火した。                     |
| 668  | 「仏壇のろうそく立てにろうそくを着けていたところ燃え移った。」と供述している。     |
| 688  | ところ、置いていた火をつけたままのカセットコンロの炎に触れたことにより着火し、至った。 |
| 738  | 何らかの火源を用いて放火したことにより、トイレトペーパーが焼きた。           |

評価に使用するデータとして、2020年1月における兵庫県神戸市消防局の火災調査報告書全33件を用いた。発火源の抽出においては、火災統計データに登録された発火源と比較し、「正しく抽出できた」「誤った文節を抽出した」の2段階の評価を行った。着火物の抽出においては、火災統計データに登録された着火物と比較し、次の4段階で評価を行った。

- 1: 誤った文節を抽出した
- 2: 文節が抽出されなかった
- 3: 正しい1文節を含む複数の文節を抽出した
- 4: 正しい1文節を抽出した

経過の抽出においては、それぞれの火災概要文から抽出された文章を神戸市消防局の職員4名が確認し、抽出した発火源、

経過、着火物の3要素、特に経過においてそれが妥当なのかどうかを、以下の5段階で評価した。

- 1: まったく見当はずれである
- 2: やや間違っている
- 3: あっているところと間違っているところがある
- 4: おおむね妥当である
- 5: 妥当である

発火源、経過、着火物それぞれの抽出評価の結果を、表3に示す。また、着火物抽出において正しい着火物を抽出できなかった主な要因を表4に示す。さらに、抽出された経過に対する評価平均点別の事案件数を図2、抽出された経過に対する評価平均点が2.5点未満のケースにおいて、抽出された文章を表5に示す。

第一に、発火源の抽出結果に対する評価について論じる。使用した火災調査報告書33件のうち、発火源を含む文節が正しく抽出できたケースは27件、誤った文節を抽出したケースは6件であった。誤った文節を抽出した6件のうち、火災概要文に発火源が記載されていなかったケースは、出動番号565、715、738、838の4件あり、これらのケースでは火災概要文から発火源を抽出することは不可能である。出動番号535のケースでは、概要文中の発火源が「落ち葉」、火災統計データに登録された発火源が「枯れ草焼き」、抽出された発火源が「火災」であった。出動番号688のケースでは、概要文中の発火源が「カセットコンロ」、火災統計データに登録された発火源が「ガスコンロ」、抽出された発火源が「ところ」であった。

第二に、着火物の抽出結果に対する評価について論じる。使用した火災調査報告書33件のうち、着火物を含んだ枝を適切に抽出できたケースは24件であった。そのうち、枝を一本のみに絞り込めていたケースは17件、複数の枝が出力され、その中に正解の枝があるケースは7件であった。

最後に経過の抽出結果に対する評価についてだが、火災調査報告書33件に対する神戸市消防局4名による評価点の平均は3.5であった。また、評価点が「おおむね妥当である」を表す4点以上であるケースは、33件中14件、評価点が3.5点以上で

あるケースは 33 件中 22 件であった。

## 5. 考 察

神戸市消防局の作成した 2020 年 1 月出動分の火災調査報告書全 33 件に対し、火災の出火原因である 3 要素を抽出した結果に対し、考察を行う。

### 5.1 発火源抽出における考察

表 3 より、火災調査報告書全 33 件に対する発火源抽出では、27 件のケースで正しい発火源を抽出できた。また、失敗した 6 件のケースのうち、火災統計データに登録された発火源が火災概要文内に記載されておらず、発火源の抽出が不可能なケースが 4 件あることから、93.1% (29 件中 27 件) の精度で発火源を抽出できたといえる。このことから、火災概要文から発火源を抽出することにおいて、提案手法は大きな効果を示したと言える。

発火源抽出における問題点として、「カセットコンロ」→「ガスこんろ」のような、仮名・カナ表記や漢字表記などの表記ゆれに対応していなかった点が挙げられる。これは、文字コードによる仮名・カナ変換や、COTOHA API による形態素のカナ読み出力機能などを用いることで、対応が可能である。

また、「落ち葉」→「枯れ草焼き」のように、実際の発火源と発火源統計辞書におけるラベルが大きく異なり、文字列比較による抽出に失敗する問題が発生した。このようなケースを解消するためには、文字列比較時に重複する文字数に閾値を設定し、抽出失敗の判定をすることや、着火限抽出のような文節間の意味ラベルを考慮した手法を構築する必要がある。

### 5.2 着火物抽出における考察

着火物の抽出では、火災調査報告書 33 件中、火災統計データに登録された着火物を抽出できたケースは 24 件である。表 4 より、火災統計データに登録された着火物が火災概要文に記載されていないケースが 2 件あったため、着火物を含む枝を抽出する精度は 77.4% (31 件中 24 件) であった。

出動番号 504, 535 においては、火災概要文中に今回設定した着火物キーワードが含まれていなかった。着火物の抽出を着火物キーワードに依存していること、着火物キーワードの設定を手動で行うことは、提案手法の欠点の一つであり、本研究における最も解決困難な課題である。

また、出動番号 3, 437, 688, 726 においては、着火物キーワードと着火物の文節間の意味ラベルが、goal, object ではなかったため、着火物として抽出されなかった。この 4 件における着火物キーワードと着火物の文節間の意味ラベルはそれぞれ、agentnonvoluntary (タオルが ← 燻焼して), place (敷布団に ← 撒き), agent (左裾付近が ← 触れた), place (日除けネットに ← 放火した。) であった。また、出動番号 668 のケースでは、着火物キーワードが文章中に複数存在し、予期せぬ挙動を示した。これらのことから、分岐条件の追加・見直しなどによるアルゴリズムの改良が求められる。

### 5.3 経過抽出及び火災概要の要約における考察

経過の抽出および火災概要の要約において、評価点が「おおむね妥当である」を表す 4 点以上であるケースは 42.4% (33 件

中 14 件)、評価点が 3.5 点以上であるケースは 66.7% (33 件中 22 件) という成果が得られた。

表 5 にある評価点が 2.5 点未満であるケースにおいて、出動番号 565 では発火源、出動番号 3, 668, 688 では着火物が抽出されていなかった。このように、提案手法では発火源および着火物を含む文節に基づいて経過を抽出するため、いずれかが欠損していれば適切な経過を出力することができない。よって前工程である発火源、着火物の抽出精度を高めることで、経過の抽出精度も高めることができると考えられる。

また、出動番号 668 における口述文のような、地の文以外の特殊な文から火災概要文が構成されている場合も精度が低く、提案手法全体の前処理として、形態素解析・構文解析のしやすい形に文章を編集する操作を行うなどの対応が考えられる。

## 6. ま と め

本稿では、神戸市消防局の持つデータと日本語自然言語処理の技術を用いて、火災調査報告書の火災概要文から火災の出火原因である発火源、経過、着火物を機械的に抽出する手法を提案した。

加えて、33 件の火災調査報告書と火災統計データを用いて提案手法の評価実験を行い、火災概要文から発火源、着火物が欠損している場合を除き、発火源を 93.1%, 着火物を 77.4% の精度で抽出できた。また、抽出した経過に対して消防局員 4 名が評価を行い、66.7% のケースで 5 点中 3.5 点以上の評価点を得た。

今後の展望として、より多くの火災調査報告書を用いて評価実験を行い、より高い精度で出火原因を抽出できるよう、提案手法の改良を行う。また、提案手法によって得られたデータを用いた、より詳細な火災統計の分析を行う。

**謝辞** 本研究の一部は JSPS 科研費 JP19H01138, JP18H03242, JP18H03342, JP19H04154, JP19K02973, JP20K11059, JP20H04014, JP20H05706 の研究助成を受けて行われている。

本研究は神戸市消防局との共同研究の一環で行われている。

## 文 献

- [1] “神戸市: 令和 2 年の消防局の災害・救急出動状況 (速報),” <https://www.city.kobe.lg.jp/a21572/032139488072.html>. (Accessed on 02/09/2022).
- [2] “火災調査規定,” [https://www1.g-reiki.net/city.kobe/reiki\\_honbun/k302RG00001396.html](https://www1.g-reiki.net/city.kobe/reiki_honbun/k302RG00001396.html). (Accessed on 02/09/2022).
- [3] “Cotoha api,” <https://api.ce-cotoha.com/contents/index.html>. (Accessed on 02/13/2022).
- [4] “大阪市消防局・大阪市立大学共同研究「大阪市における火災分析」,” [https://www.city.osaka.lg.jp/contents/wdu010/digitalbook/syobo/kasaibunseki\\_r2/#target/page\\_no=1](https://www.city.osaka.lg.jp/contents/wdu010/digitalbook/syobo/kasaibunseki_r2/#target/page_no=1). (Accessed on 02/14/2022).
- [5] P. McCann, “fugashi, a tool for tokenizing Japanese in python,” Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS), pp.44–51, Association for Computational Linguistics, Online, Nov. 2020. <https://aclanthology.org/2020.nlpss-1.7>
- [6] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸, “形態素解析システム「茶釜」,” 情報処理, vol.41, no.11, pp.1208–1214, 2000.
- [7] 工藤拓, 松本裕治他, “チャンキングの段階適用による日本語係り受け解析,” 情報処理学会論文誌, vol.43, no.6, pp.1834–1842, 2002.