

キーワード抽出を用いた地域防災計画における 災害廃棄物管理に関する比較分析

劉 英楠^{†1} 畑山 満則^{†2}

概要：地域防災計画は、災害対策基本法（第40条）に基づき、各地方自治体の長がそれぞれの防災会議に諮り、防災のために処理すべき業務などを具体的に定めた計画である。災害が発生した時には、地域防災計画に記述されている内容に従って災害対応が行われるが、実際には自治体（特に市町村）が円滑に計画を実行できないことも多く、それゆえに計画の内容が問題視されることも多い。地域防災計画は人が作成したものであるため、不備や問題点は当然存在する。計画の定期的な見直しが必要である。しかし、地域防災計画の分量が非常に多く、自治体の数も多いゆえ、専門家が全ての地域防災計画を見定め・評価するのは非現実的な時間と労力がかかる。そのため、地域防災計画の大量なテキストデータを用いて自動的または半自動的に問題点や過去の被災経験から得られた知見を発見できる仕組みが求められている。本研究では、地域防災計画の実行時に発生しうる問題点を発見するアプローチを提案し、その有効性について考察・検証を行った。

キーワード：地域防災計画, キーワード抽出, 自然言語処理

Comparative Analysis on Disaster Debris Management in Local Disaster Management Plan with Keyword Extraction Methods

YINGNAN LIU^{†1} MICHINORI HATAYAMA^{†2}

Abstract: The local disaster management plan decides concretely the act of the local governments for disaster reduction. The head of municipality formulate the details of the plan. This plan is based on the disaster countermeasures basic act. However, when a disaster occurred, it does not always work so well even the municipality government act on the plan. This problem is gaining attention. The local disaster management plan is made by human so some problems and miss will always be there. Since that the local disaster management plan need to be reviewed regularly. For there are so many local governments and each of them have a very long local disaster management, it's not possible for experts to review all of them in time. Since that, an effective method to find problems and experiences automatically or semi-automatically in the parts of disaster debris management is required. In this research, we developed an approach to do a potential problem analysis, and to verify the validity.

Keywords: Local disaster management plan, NLP, Keyword extraction.

1. はじめに

1.1 研究背景

地域防災計画は、災害対策基本法（第40条）に基づき、各地方自治体の長がそれぞれの防災会議に諮り、防災のために処理すべき業務などを具体的に定めた計画である。災害が発生した時には、地域防災計画に記述されている内容に従って災害対応が行われるが、実際には自治体（特に市町村）が円滑に計画を実行できないことも多く、それゆえに計画の内容が問題視されることも多い。地域防災計画は人が作成したものであるため、不備や問題点は当然存在する。計画の定期的な見直しが必要であるが、自治体内の防災会議で議論し、自治体内部で完結するような見直しが行われているのが一般的である。計画の定期的な見直しが必要であるが、自治体内の防災会議で議論し、自治体内部で完結するような見直しが行われているのが一般的である。計画の定期的な見直しが必要であるが、自治体内の防災会議で議論し、自治体内部で完結するような見直しが行われているのが一般的である。そのため、被災していた自治体

は災害対応の経験を積むことができ、その経験や教訓を地域防災計画の改訂に反映している一方、被災していない自治体は災害対応の経験を積むことができず、被災した自治体が災害対応で得られた貴重な知見は、他の自治体で十分に活かされていない可能性がある。

1.2 研究目的

本研究では、キーワード抽出を用い、市町村レベルの自治体の地域防災計画の中の廃棄物処理に関連する章・節に適用する。各文書の特徴的な単語を抽出することで地域防災計画の内容を評価する時の労力の軽減と速や問題発見を図る。また、自治体計画文書の変遷から、自治体が過去の被災経験などからどのような見直しをしたかを明らかにし、その理由を究明することで得られた知見を他自治体に提供することを目指す。

2. 先行研究

本研究に関連する先行研究は以下が挙げられる。

(1) 地域防災計画における情報伝達の機能的障害の発見と見直しに関する研究

畑山ら[1]は、地域防災計画における情報伝達の機能的障害の発見手法の開発をしたが、自動的にかつ大量な文書からの問題発見はできなかった。本研究では大量な文書から問題発見の自動化を図る。

(2) 災害廃棄物処理に関連する研究

平山ら[2]は、「広域災害時における災害廃棄物処理の広域連携方策に関する研究」で、阪神・淡路大震災における廃棄物処理の事例と経験を踏まえ、広域災害時の災害廃棄物処理のフローに問題があると指摘した。そこで5つの広域連携方法を提案し、それに対して首都直下地震のシナリオに基づくシミュレーションを行った。その結果、広域災害に対応できる広域連携方策を示した。しかし、地域防災計画にそれらの内容が書かれているか否かを検証する方法については言及していない。本研究では、地域防災計画中の災害廃棄物処理に関連する章に焦点を当て、その内容を分析し、広域連携の問題を含めどのような問題が存在するかを探る。岡山ら[3]は、自治体の水害廃棄物処理計画の策定状況を把握するため、全国の1,742市町村および都道府県の災害廃棄物・災害瓦礫処理を担当している部署宛に対してアンケート調査を行った。その結果、36%の自治体が廃棄物処理計画を策定しておらず、さらに全体の49%の自治体がこれまでに災害廃棄物処理の経験がないと回答している。また、計画に対する実効性のなさや、仮置場の設置や廃棄物の分別に関する問題が指摘されている。

(3) テキストマイニングに関連する研究

那須川ら[4]が、テキストマイニングで用いられる技術や分析手法についてレビュー論文を発表した。テキストマイニングでよく用いられる基盤技術について紹介した。テキストマイニングは、自然言語処理やデータマイニングなど多様な技術を組み合わせた複合技術であり、テキスト自動分類、重要語句の抽出など、様々な課題を解決するために研究されてきた。テキストマイニングの技術でどのような問題を解決できるかを示したため、ここで記述されている知見を踏まえて、地域防災計画の分析への応用を検討する。

(4) キーワード抽出に関する研究

文書の中からキーワードを抽出する技術は、自然言語処理の分野の中でも重要な研究領域の一つとされており、今まで数多くのキーワードを抽出する方法が研究されてきた。例えば Tf-idf, BM25[5]といった古典的な統計情報を使用し、文書内の単語にスコアをつける手法がある。この手法ではコーパス全体において特徴的な単語に対し、高い点数をつける傾向にある。また、PageRankを自然言語処理のプロセスに適用し、グラフ理論を駆使した TextRank[6]や、辞書や形態素解析に頼らない N-gram を利用したキーワード抽出[7][8]などが検討されているが、どのようなデータに対しても精度が高いアルゴリズムは存在しない。その理由は、キーワード抽出の正解は人間が作成するものであるため、

人の経験の相違によって正解も変化するのである。評価は難しいとされている。扱うデータの性質やタスクや目的も様々なので、研究者各自が自分に需要によってアルゴリズムをアレンジする必要がある。

3. 関連技術

本研究では、以下の技術を使用している。

3.1 形態素解析

形態素解析とは、文法的な情報の注記のない自然言語のテキストデータから、対象言語の文法や辞書と呼ばれる単語の品詞等の情報に基づき、形態素の列に分割し、それぞれの形態素の品詞等を判断する作業である。本研究では、形態素解析のツールとして MeCab[9]を使用する。なお、形態素解析の精度の向上のため、拡張辞書として MeCab 新語辞書[10]を導入している。

3.2 Opaki BM25

本研究では、Tf-idf を改良した BM25 アルゴリズムをベースにキーワードの抽出を行う。BM25 とは、Tf-idf 考え方に基づき、同じ単語が同じ文に大量に出現しても、単語のスコアが極端に大きくならないように Tf の値に補正をかけたものである。BM25 を選んだ理由として、実装が相対的に容易であり、Tf-idf の改良版であるため、Tf-idf より高い精度が出せると言われているからである。また、TextRank では、キーワードの抽出の計算は一つの文書内で完結するため、コーパス全体の性質を反映しにくいという問題が存在する。今回の分析対象は地域防災計画の災害廃棄物関連章であり、TextRank を適用した場合、キーワードの抽出は1自治体の文書で完結し、自治体間の特徴が把握できない。よって本研究では、BM25 を採用する。

前述のように、BM25 は広く知られている特徴量の Tf-idf の改良版の一つである。算出された BM25 の値は、Tf-idf と同じ、文書に含まれる単語がどれほど特徴的であるかを示す値である。各自治体の文書の中で、ある単語の BM25 のスコアは以下のように算出される：

$$w(d, k) = \frac{(k_1 + 1) \cdot tf(d, k)}{k_1 \cdot \left((1 - b) + b \cdot \frac{dl}{avdl} \right) + tf(d, k)} \cdot \log \left(\frac{N - df(k) + 0.5}{df(k) + 0.5} \right)$$

($w(d, k)$: 単語 k が文書 d における BM25 値。 dl : 単語が含まれる文の長さ。 $avdl$: この自治体のすべての文の長さの平均値。 k_1, b : 定数のパラメータ。本研究では $k_1=2.0$, $b=0.75$ とする)

3.3 Tf-Ridf

Ridf[11]は、ポアソン分布の考え方を導入した単語への重み付け手法である。ポアソン分布を利用し単語の文書頻度の推定値を計算し、それを単語の文書頻度と差をとることと得られる。ポアソン分布は、文書の特徴を依存せずコ

ーパス中に普遍的に使用される単語（一般語）に対し良い近似になるため、文書の特徴、内容を表す内容語には近似しないことになっている。その理由は、内容語は普遍的に各文書に出現せず、限定的な文書で集中的に出現するためである。ある単語が集中的に限定的な文書に出現すると、単語の大域的頻度が上昇するが、文書頻度がそこまで上昇しないためである。実際の文書頻度とポアソン分布によって推定された文書頻度を比較することで、文書の特徴に依存しない一般語の重みを下げると同時に、内容語に大きい値をつけることができる。結果的に、内容語がより抽出されやすくなる。

ポアソン分布の計算式は以下のように：

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

ある単語を w_i とする。 F_i をある単語 w_i の大域的頻度（文書集合全体での出現頻度）とし、ポアソン分布を計算するための λ_i は以下の式で推定することができる：

$$\lambda_i = \frac{F_i}{n}$$

上記の式を用いて、単語 w_i が文書中に k 回出現する確率は $P(k; \lambda_i)$ で求めることができる。 $k=0$ の場合、 $p(0; \lambda_i)$ は w_i が文書に 1 回も出現しない確率となる。ある単語 i の $Ridf_i$ は下の式により計算される：

$Ridf_i = \text{実際のIDF} - \text{ポアソン分布により推定されたIDF}$

$$= \log \frac{n}{n_i} - \log \frac{n}{n(1-P(0;\lambda_i))}$$

$$= \log \frac{n}{n_i} + \log(1 - e^{-\frac{F_i}{n}})$$

$Ridf$ を $Tf-idf$ の idf と置き換えたものが $Tf-Ridf$ である。地域防災計画には、文書の特徴を表す固有名詞（地名など）や特徴的な表現は、ここで言う内容語に相当する。 $Ridf$ を適用することで、 idf を適用した場合よりも一部の文書のみで高頻度で使われる単語が高いスコアを獲得しやすくなり、抽出されやすくなる。以上の理由から、本研究では $Tf-Ridf$ を採用する。

4. 分析対象

4.1 地域防災計画の取得

日本全国には 1,718[12]自治体が存在する。各市町村の地域防災計画を分析するためには地域防災計画のテキストデータが必要であり、最終的に 1,113 自治体の計画のテキストデータを入手した。なお、各自治体が被災経験から学習した経験や教訓を探るために、近年土木学会が災害調査団を派遣した災害の被災地と平成30年7月豪雨の被災地の自治体をピックアップし、国立図書館インターネット資料収集保存事業からそれぞれの自治体の過去数年間の地域防災計画を収集した。最終的に旧計画を 16 自治体から 69 文書を収集し、最新版の 1,113 文書と合わせ、合計 1,182 文書を

収集でき、取得率が 64.8%である。

4.2 「災害廃棄物処理」関連章の取得

地域防災計画には様々な内容がそれぞれの章や節に書かれているが、今回は災害廃棄物処理に焦点を当て分析する。したがって、関連する章・節のみでコーパスを構築した。

- ・旧計画を取得できた 16 自治体 69 文書：手動で災害廃棄物処理の章を抽出した。そのため関連のない内容の混入はなく、純粋なテキストデータを取得できた。

- ・最新版のみの 1,113 自治体の 1,113 計画文書：自動抽出を行った。旧計画の手動抽出の結果の単語頻度を調べ、出現頻度が上位、かつ災害廃棄物処理の章にしか出ないと思われる単語を選択し、以下の 6 単語を検索キーワードとした。

[ゴミ、ごみ、廃棄物、廃棄物処理、災害廃棄物、し尿]

災害廃棄物処理の関連章・節では、上記の単語群が頻繁に出現することが考えられる。これらの単語を含む文の行番号を抽出し、これに対し階層型クラスタリングを行い（Ward 法、ユークリッド距離）、閾値を「上記単語を含む文の間の最大ユークリッド距離の 2%」に設定した。閾値で切られるクラスターは上記の単語群が密に出現する段落であり、クラスターに含まれる最小行番号と最大行番号の間の文を抽出した。

結果的に、合計 1,182 災害廃棄物処理に関連する文書を作り、この集合を「災害廃棄物処理コーパス」と呼ぶ。

4.3 自動抽出の精度検証

自動抽出のアルゴリズムを考案したが、アルゴリズムの正当性について検討していない。本節では、自動抽出の精度について定量的な評価を行う。

自動抽出結果を評価対象とし、手動抽出結果を正解と見なす。それぞれの文書に対し Kappa 係数 (kappa_sta)、 precision , recall , F-measure を利用して評価を行う。「旧計画セット」の 69 文書に対し計算を行って各項目の平均値を取り、その結果は以下の通り：

表 1 旧計画データセットの評価結果

kappa_sta	0.628
precision	0.596
Recall	0.758
F_measure	0.651

「旧計画セット」を使用し評価を行なった結果、 kappa 係数が 0.628 になり、「substantial agreement」[13]の範疇に入っている。また、 F-measure も 0.651 に上昇した。まだ精度が十分良いとは言えないが、適合率が 0.596 となり、再現率も 0.758 となった。比較的高い精度を取れたため、手動抽出では現実的ではない現段階では、自動抽出された文書でも十分地域防災計画・災害廃棄物処理に関連する章・節の性質を反映できると考えられるため、この自動抽出のア

ルゴリズムを「最新版セット」に適用し、「災害廃棄物処理コーパス」を構築することにした。

5. キーワード抽出手法の提案・適用

本研究では、4.2 節で構成された「災害廃棄物処理コーパス」を分析対象とし、それぞれの文書がどのような特徴を持つかを分析することを目的としている。BM25 と Tf-Ridf の特徴量を利用し、各文書中出现する単語にスコアを付け、そのスコアを利用してキーワードの抽出を行う。これらの特徴値の上位単語は各文書の特徴的な単語であり、特徴語と見なす。一方、BM25 と Tf-Ridf の下位単語はどの文書でも普遍性を持っており、ある文書ではそのような単語を使用していないことは、文書的内容的な欠陥や欠如の発見につながることを期待でき、欠損語と見なす。Ridf が一般語に対して小さな重みを付与するという特性を利用し、Tf-Ridf の特徴量を導入したことで、さらに抽出されたキーワードを絞り込む効果が期待できる。

特徴量の上位と下位の両方向からキーワードを抽出することで、文書の特徴を把握し、評価する。5.1 節では、この手法を利用した地域防災計画の見直し作業負担の軽減に貢献できるシステムを提案する。

5.1 提案システムの構成

本研究では主に災害廃棄物処理の章・節に焦点を当て、キーワード抽出結果を手動で分析しているが、長期的に一連のプロセスをシステムとして構築することを目標とする。なお、研究段階では本システムのユーザが研究者の方になるが、最終的に本システムをサービスとして提供し、自治体職員の方をユーザとして想定している。具体的な実現方法の論述は 6.3 節で行う。構想されたシステムを図 1 のように提案する。

- (1) データの収集及びクリーニング：各自治体の地域防災計画の PDF ファイルを収集した。テキストデータへ変換し、必要とする部分を抽出する。また、変換・抽出されたテキストデータを処理しやすいように整形する。
- (2) キーワード抽出：テキストデータをテキストマイニングの手法を適用し、各自治体の計画の文書の統計的な情報などを取り出す。
- (3) 結果分析：取り出した情報を手動で分析し、問題点になりうる箇所・知見になりうる箇所を見つける。
- (4) 問題識別：計画の実行の妨げになりうる箇所を発見した場合、該当自治体にヒアリング調査などを行い、実際に計画の実行の妨げになるか、あるいは想定した知見と一致するかを確認する。また、人がどのようにそのような箇所を発見したかを分析し、その理由や過程を他の自治体に適用することが可能かどうかを検討し、自治体に助言を行う。

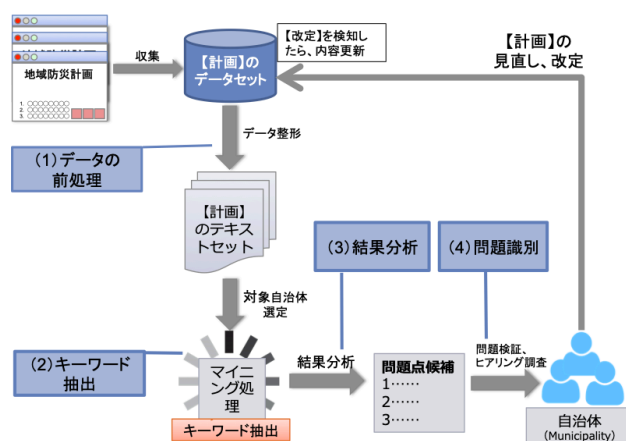


図 1 提案システムの構成

提案システムでは、大きく分けて 4 つのステップがある。その中 (1)、(2) に関しては、汎用性を持っており、どのようなテキストデータにも適用できるが、(3)、(4) は専門家の助言が必要とする。本研究の場合、分析対象が地域防災計画の災害廃棄物処理に関する章・節であるため、抽出結果を分析し解釈するには地域防災計画を熟知し、専門的な知識が必要である。よって防災あるいは廃棄物処理の専門家の助言が必要とする。一方、(1) と (2) プロセスより、専門家がすべてのテキストを見ることなく、抽出結果を分析することで問題発見および知見の抽出ができるため、労力と負担が大きく軽減される。

5.2 不要単語リストの作成

災害廃棄物処理に関連する章の中ではほぼ確実に出現し、かつ分析にあたって必要性を感じない単語が存在する。単語自体は普遍的に各自治体の計画に高い頻度で出現しているため、BM25 などの特徴量の計算式中の Tf 値が高くなり、抽出されやすい傾向にある。これらの文章の特性に依存する頻出単語を除外し、「分析対象外単語」とする。以下の 6 単語を選出し、「分析対象外単語」とした：

[ゴミ、ごみ、廃棄物、廃棄物処理、災害廃棄物、し尿]

5.3 特徴語の抽出手法

ある自治体の地域防災計画において、特徴的な単語が多数抽出されたことは、他の自治体ではあまり書かれていない単語が多数書かれていることを意味する。特徴的な単語として抽出された単語は、その自治体の独特な取り組みを表す可能性が高く、ポジティブな要素を示す単語（以後、Positive 要素単語と呼ぶ）として捉えることができる。

コーパスに対し、BM25 と Tf-Ridf のアルゴリズムを適用し、各自治体の中での全ての単語に BM25 と Tf-Ridf の値を付与する。各自治体において BM25 の値の上位 20 単語を取り出し、Positive 要素単語の候補とする。ただし、これらの特徴値から抽出された単語は、あくまで統計上特徴的な単語であるため、上位に来たとしても問題発見にあたり価値のある単語とは限らない。Positive 要素として高頻度で抽出された単語は、相対的な価値が低いと考えられ、コーパス

中の全ての文書において X 以上回抽出された単語を削除し、Positive 要素単語の数を抑え、価値の低い単語を濾過した。抽出方法は以下の式で表現する：

「Positive 要素単語」= (「BM25 上位 20 単語」&「Tf-Ridf 上位 20 単語」) - 「不要単語」- 「X 回以上抽出された単語」

X は閾値であり、今回は文書数の 1/8 に相当する 154 に設定した。すなわち、ある単語が他の自治体において 154 回以上抽出された単語を抽出結果から除去した。

5.4 欠損語の抽出手法

BM25 で文書内の全すべての単語にスコアを付与すると、BM25 の順位が下位にあるほど単語の tf と idf が低くなっていく傾向がある。よって下位にある単語は全コーパスにおいて一定程度の普遍性を持っており、どの自治体の計画の文書でも出現しやすい傾向があると考えられる。この「ほぼ全ての自治体において普遍的に使われた単語」がある自治体の計画で書かれていないことは、その自治体の計画に本来であれば書くべき内容が書かれていない可能性があることを示唆し、内容的な漏れの発見につながる可能性が高いと考えられる。つまり、このような単語は、文書のネガティブな要素を絞り込むための単語（以下、Negative 要素単語）として捉えることができる。

コーパスに対し、各文書の単語に BM25 順の下位 50% の単語を抽出し、各単語が何回抽出されたかをカウントし、1)抽出回数が Y 回以上、2)この自治体において一回も出現していない、3)分析対象外単語リストにない、4)その単語は部分文字列として抽出されなかった、この 4 つの条件を全て満たす単語を抽出する。抽出手法は以下の式で表現する：

「Negative 要素単語」= (「BM25 下位 50% の単語」&「Tf-Ridf 下位 50% の単語」) & 「Y 回以上抽出された単語」& 「tf(d, k)=0」- 「不要単語」- 「出現した単語の部分文字列である単語」

tf(d, k)=0 は、単語 k が文書 d で一回も出現しなかったことを意味する。Y は閾値であり、今回は 591 (災害廃棄物処理コーパスにある自治体数の半分に相当する値) に設定した。

5.5 実験結果

表 2 は 4.2 と 4.3 で行った実験結果の一例である。

表 2 Positive, Negative 要素単語の抽出結果 (一例)

都道府県	自治体	Positive 要素単語	Negative 要素単語
愛知県	あま市	産業廃棄物、発令、警戒宣言、保管、促進、附属、指導、連携、広域	応急、衛生、被災地、被害
福井県	あわら市	石綿、福井坂井、くみ取り	災害時
千葉県	いすみ市	策定、津波、指針、漁港、編成、協定、被害	応援、地域、応急

表 3 は、栃木県小山市を対象に 4.4 で述べた分析を行った例である。この例では、小山市の計画から抽出された Positive 要素単語の変化を表した。Positive 要素単語の増減

から計画の特徴の変化を読み取ることができ、Negative 要素単語の増減から、どのような単語が消され、書き足されたかが読み取れる。

表 3 旧計画の変遷の抽出結果 (一例)

栃木県 小山市 Positive:	消えた単語	増えた単語
H23→H26	20 節、市町村、決定	広域、除染、保健
H26→H28	(なし)	放射性物質、家屋
H28→H29	広域、家屋、保健、単独	特例、土壌、搬入、協定、汚染

6. 考察

6.1 キーワード分析から考えられる仮説

この節では、専門家の意見と自分の知識と経験を総合的に用いながら、抽出結果に対し考察を行い、問題点や過去の経験が含まれそうな文書を抽出する。この節で抽出された文書から立てた仮説は、後ほどの 6.2 節で検証する。

6.1.1 熊本県 熊本市

表 4 のように、熊本市では、H25 年版と H27 年版の Positive 要素単語に「宇城広域連合」があったが、H29 年版に消えている。

表 4 熊本市 Positive 要素単語の変遷

熊本県 熊本市	消えた単語	増えた単語
H21-H22	(なし)	地区
H22-H23	(なし)	(なし)
H23-H25	熊本、支部、(社)、熊本県、建設業協会	宇城広域連合、道路、山鹿植木広域行政事務組合
H25-H27	地区	熊本
H27-H29	宇城広域連合、道路、熊本、河川、センター、山鹿植木広域行政事務組合、移送	法、住民

宇城広域連合の資料によると、熊本市は平成 20 年 10 月に市町村合併により宇城広域連合に加入し[14]、平成 26 年 3 月に同連合を離脱した[15]。実際計画の本文を調べたところ、H21 年版には「宇城広域連合」に関する記述はなかった。「宇城広域連合」が初めて計画に出現したのは H22 年版であるが、全編を通して 1 回しか使われていないため、Positive 要素として抽出されなかった。H25 年版では初めて Positive 要素単語として抽出された。H27 年版の計画に「宇城広域連合」に関する記述が残っており、現状の反映に若干の時差があることがわかる。また、「宇城広域連合」が H29 年から消えた理由は、熊本市が連合を離脱したため消したのか、それとも熊本地震で何かが起きて消したのか、その理由を明らかにする必要があると感じた。

一方、岩手県盛岡市の H23、H25、H27 と H30 年の比較では、H23、H25、H27 年版では「稗貫」が Positive 要素単語として抽出されたが、H30 年版では「稗貫」が Positive 要素単

語の抽出結果から消えた。本文を調べた結果、「紫波、稗貫衛生処理組合」という表現が H30 年版の文書から消えている。その理由として、紫波、稗貫衛生処理組合が平成 30 年度に解散されたためであることが岩手県の資料[16]により分かり、盛岡市の計画が適時に見直されたことを示している。

6.1.2 北海道 帯広市

表 5 のように、北海道帯広市では、H25、H26、H27、H30、4 つの計画は抽出された単語の変化は少ないことから、4 版に渡って内容がさほど変更されてないことが推測できる。しかし、「埋却」が H26 年版およびその後の計画から Positive 要素単語として抽出されなくなっている。

表 5 帯広市 Positive 要素単語の変遷

北海道・帯広市	消えた単語	増えた単語
H25-H26	埋却	所有者
H26-H27	(なし)	(なし)
H27-H30	(なし)	くり

計画の本文の一部は下の黒い枠で示している。H25 年版では死亡獣畜の処理に関する部分で「移動し難い死亡獣畜について集中焼却または埋却するものとする」と明記されているが、H26 年版から「死亡獣畜取扱場において行う」という表現に変更され、移動し難い死亡獣畜の「埋却」に関する文章が消え、代わりに「十勝総合振興局保健環境部長の指導を受け臨機の措置を講ずる」と書かれていた。化製場等に関する法律（第 2 条の 2）によると、「死亡獣畜の解体、埋却又は焼却は、死亡獣畜取扱場以外の施設又は区域で、これを行なってはならない」と明記されている。

これにより、死亡獣畜の処理方法に関する記述が見直されたことが分かった。その理由については、帯広市が法律への抵触を懸念し「埋却」を明記することを避けたのではないかと推測する。もしこれが事実であれば、他の自治体の Positive 要素単語に「埋却」が出現した場合、その内容を確認した上、もし同じ内容が書かれているのであれば、帯広市と同じ理由で見直し時に「埋却」を再度検討する必要があるかと考える。

6.1.3 栃木県 小山市

表 6 のように、栃木県小山市の Positive 要素単語の抽出結果では、H26 年版「除染」が新たに抽出され、更に H28 版に「放射性物質」が抽出され、原子力事故対策に連想させる単語が増えている傾向にあることがわかった。

表 6 小山市 Positive 要素単語の変遷

栃木県 小山市	消えた単語	増えた単語
H23-H26	決定、20 節、市町村	除染、広域、保健
H26-H28	(なし)	家屋、放射性物質
H28-H29	家屋、単独、広域、保健	汚染、土壌、協定、特例、搬入

計画の本文を見ると、H26 年版の計画から「放射性物質に

より汚染された廃棄物の処理」の節が追加されており、旧計画を取得した 16 自治体の中では他に例を見ない小山市の独特な取り組みだと捉えることができる。また、最新版計画のみを調べた結果、「放射性物質」が Positive 要素単語として抽出された文書が 25 個あり、その中の 9 個が栃木県に集中している。原子力災害対策編が独立する自治体もあり、そこに原子力対策の詳細が書かれている場合もあるが、災害廃棄物処理の章・節に限定すれば、栃木県全体が放射性物質対策に非常に注力しているも推測できる。その理由について、東日本大震災における福島第一原発事故の影響が考えられる。また、小山市や栃木県の他の自治体と同じ理由で他県の自治体でも原子力対策について災害廃棄物処理の章の内容の強化を行うべきかどうかを検討するよう、助言する価値があると考えられる。

6.2 検証

6.1 節で行った考察で立てられた仮説を検証する。熊本市、帯広市、小山市にインタビューすることができ、6.1 節で述べた仮説が事実であるか否かを自治体に問い合わせし、問い合わせの内容とその結果をこの節にまとめる。

6.2.1 熊本県熊本市に関する事実

熊本市が H29 年版の地域防災計画に「宇城広域連合」に関する記述が残っている理由については明らかになっていないため、これを熊本市政策局、危機管理防災総室に電話インタビューを行った。以下のような回答を得た：

熊本市が平成 20 年に富合町と合併した。新熊本市が防災や廃棄物処理に関する業務の整合性を取るため、一時的に宇城広域連合に加入し、諸業務の調整期間を取ることにした。そして平成 26 年 3 月に、諸業務の調整が終わり、防災・廃棄物処理に関する業務において元富合町と整合性を取った新熊本市が宇城広域連合から脱退した。熊本市の H27 の地域防災計画は、H26 中に見直しながされ、見直しが行われた時点でまだ宇城広域連合から離脱していないため、その記述が残っていることが判明した。この事実から、熊本市が熊本地震ではなく、宇城広域連合から離脱したため「宇城広域連合」を削除したことがわかった。

6.2.2 北海道帯広市に関する事実

北海道帯広市は、「埋却」に関する記述を H26 年度版から明記しなくなったことがわかったが、その理由について帯広市に問い合わせた結果、以下の回答が得られた：

地域防災計画の中の死亡獣畜の処理に関する部分は、現在も見直している途中であり、今後も変わっていく可能性がある。おそらく当時、死亡獣畜がどれほど発生するかに関して試算を行った結果、焼却でも十分対応できる結果だったため、「埋却」を削除し、基本的には焼却で対応する方針にした。ただし、十勝総合振興局は帯広市だけでなく、周辺自治体も管轄しているため、死亡獣畜が大量に発生する時に焼却しきれない場合もありうる。それを懸念し、「埋却」を否定するものではなく、「指導の元で対応する」

という趣旨の文書に見直された。

こちらの回答から、帯広市は法律への抵触を懸念している訳ではなく、「焼却処理」を優先的に行うためであったことがわかった。

6.2.3 栃木県小山市に関する事実

小山市や栃木県は、他自治体よりも原子力事故の対応に注力していることがわかった。

背景として、2011年3月11日、東日本大震災が発生した。その一連の災害の中の一環として福島第一原子力発電所事故があった。放射性物質が大量に放出され、福島県を始め、福島および周辺の都道府県に放射性物質が飛来し、深刻な事故となった。そこで距離的に放射性物質が栃木県まで飛来しないと推測されたが、その後放射性物質が栃木まで飛んできたことがわかり、ホットスポットと呼ばれる放射線量が局所的に高い場所が見つかった[17]。また、栃木県の周辺に、福島第一原子力発電所、福島第二原子力発電所、東海第二発電所に囲まれている。福島第一原子力発電所の事故から、栃木県全体が原子力事故への緊張感が高まり、その対策に注力してきた。

小山市の取り組みは、栃木県の方針と関連があるかについて、小山市にインタビューを行い、以下の趣旨の回答が得られた：

H26版の地域防災計画に「除染・放射性物質により汚染された廃棄物の処理」という節が新たに追加された理由は、「除染関係ガイドライン」[18]がH25年5月に環境省より作成されて、除染に関する指針などが詳細に記述されており、栃木県がそれに合わせて栃木県の地域防災計画の見直しを行った。その後、栃木県内の市町村も順次、県の計画に合わせて計画を見直した。

H29年版の計画に「土壌」が新たに Positive 要素単語として抽出された理由は、「除染関係がガイドライン」が平成28年9月に追補があり、除染土壌を扱う時の注意事項に関する記述が追補された。栃木県がこれに合わせて県の地域防災計画を見直し、除染土壌に関してより詳しく記述することになり、これに呼応して栃木県と足並みを合わせて、小山市も変更したことがこのような結果になっているのではないと思う。

7. おわりに

7.1 まとめ

本研究では、地域防災系計画の中で災害廃棄物処理に関連する内容に焦点を当て、地域防災計画の実行時に発生しうる問題点を発見するアプローチを模索した。収集した全国の自治体の計画文書に対し、テキスト化を行い、収集してきた1,233文書を「調査団セット」、「西日本豪雨・広域連合セット」、「最新版セット」、「旧計画セット」の4セットに分類した。その中、「調査団セット」、「西日本豪雨・広域連合セット」、「旧計画セット」は手動で災害廃

棄物処理に関連する章・節を抽出したもので、純度の高いテキストデータである。

「最新版セット」は自動抽出により生成されたデータセットであり、近似的なものである。大量な文書から災害廃棄物処理に関連する章・節を抽出し、「災害廃棄物処理コーパス」を構築必要があった。しかし手動抽出では作業量が多量にも膨大であり、近似的に自動抽出を用いた階層型クラスタリングの手法でコーパスの構築を図った。抽出結果を「西日本豪雨・広域連合セット」と「旧計画セット」と比較し、抽出精度を評価した。抽出精度を評価した結果、精度が十分高いとはいえず、今後の重要課題の一つとして認識しているが、今回はこの抽出結果を用いてコーパスを構築した。

キーワード抽出の手法として、BM25とTf-Ridfの特徴量を利用し、BM25とTf-Ridfの上位単語である Positive 要素単語、およびBM25とTf-Ridfの下位単語である Negative 要素単語の抽出を行った。その結果を分析し、いくつかの自治体文書から懸念事項を発見することができた。更に「旧計画セット」を用いて前後の計画の抽出結果に対する比較分析を行った。その結果、予備実験段階で発見した懸念事項を再発見することができ、更にいくつかの仮説を立てることができた。3つの当該自治体にインタビューを行い、計画文書の変化の理由を明らかにした。

7.2 今後の課題

高質なテキストデータの獲得：

地域防災計画への適用では自動抽出を用いた近似的なデータセットからコーパスを構築した。しかし近似的なデータセットの精度が十分とは言えず、災害廃棄物処理に関連する章・節に属しない文の混入や、属しているのに抽出されていない文章が存在することが想像できる。コーパス全体の性質に対する影響は限定的であるが、個々の自治体の特徴や問題点を発見するに当たって抽出結果を分析する時に大きな影響を与えてしまう。この問題を解決するために自動抽出の精度を更に上げるか、膨大な時間と労力をかけて手動抽出を行うか、どちらかの手段を取る必要があるが、どちらでも容易ではない。今後の課題として解決しなければならない。今後はキーワード抽出の精度をさらに上げるために単語をベクトルで表現できる word2vec などの技術を使い、同義語や意味の近い単語を計算機に認識させ、抽出精度の向上や重複抽出の防止に注力することが必要だと考える。

キーワード抽出精度の向上：

本研究では、キーワードの抽出により地域防災計画の問題点発見を試みたが、表記の揺れや、単語の間の同義性の認識については考慮していない。

地域防災計画の時系列データベースの構築：

今回は「旧計画セット」で、計画の変遷を追跡することができた。その結果、いくつかの重要な事実や懸念事項を

発見することができた。計画の変遷を分析することが問題点の発見に繋げることが大きく期待できる。これからも災害が発生することや、その都度地域防災計画が改定・見直しされ、新たな経験や教訓が改定で反映されることを想定し、地域防災計画の時系列データベースを作成し、定期的に地域防災計画の最新版を収集する必要があると感じた。

問題点の自動発見：

本研究では、提案システムの一連のプロセスで地域防災計画の変遷から問題点と知見を抽出した。しかし提案システムでは文書の獲得とキーワード抽出機能を自動化することができたが、問題発見はまだ人力に頼っている。研究の段階では専門家の見解で問題点をもっているが、今後は応用段階に入ると、問題発見も自動化することが望ましい。それを実現するために、抽出されたキーワードがどのような問題点に繋げられるかの知見を備蓄し、キーワードと問題点の間に関連づけることで、問題点の自動発見を図る必要がある。

参考文献

- [1] 畑山満則, 安藤恵. 地域防災計画における情報伝達の機能的障害の発見手法の開発. 情報処理学会論文誌, 2013, Vol.54, no.1 p202-p212.
- [2] 平山修久, 河田恵昭. 広域災害に行ける災害廃棄物処理の広域連携方策に関する研究. 土木学会論文集, 2007, vol. 63, no. 2, p. 112-119.
- [3] 岡山朋子, 伊藤秀行. 自治体の水害廃棄物処理計画の策定状況に関する調査研究. 第24回廃棄物資源循環学会研究発表会公園論文集. 2013. A12
- [4] 那須川哲哉, 河野浩之, 有村博紀. テキストマイニング基盤技術. 人工知能学会誌, 2001, vol. 16, no. 2, p. 201-211.
- [5] S.E. Robertson and S.walker, "Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval," Proc. 17th Annual International ACM SIGIR Conference (SIGIR'94), 1994.
- [6] R. Mihalcea and P. Tarau. 2004. TextRank – bringing order into texts.
- [7] 森信介, 長尾真. n グラム統計によるコーパスからの未知語抽出. 情報処理学会論文誌, 1998, vol. 39, no. 2, p2093-2100.
- [8] 梅村恭司. 未踏テキスト情報中のキーワードの抽出システム開発. Technical report, <http://www.ipa.go.jp/archive/NBP/12nendo/12mito/mdata/10-36h/10-36h.pdf>, 2000.
- [9] MeCab: Yet Another Pair-of-Speech and Morphological Analyzer. <http://taku910.github.io/mecab/>
- [10] Mecab-ipadic-neologd. <https://github.com/neologd/mecab-ipadic-neologd>
- [11] 北研二, 津田和彦, 獅々堀正幹. 情報検索アルゴリズム. 共立出版, 2002年 p43-p45.
- [12] 総務省. <http://www.soumu.go.jp/gapei/gapei2.html>
- [13] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1):159-174.
- [14] 平成29年版消防年報くまもと. 熊本市. https://www.city.kumamoto.jp/common/UploadFileDsp.aspx?c_id=5&id=17095&sub_id=1&fid=118019
- [15] 消防年報 平成25年度版. 宇城広域連合消防本部. http://www.ukisyobo.or.jp/dl?q=5893_filelib_4025afcd572385ba0a1d7581e1a4911.pdf
- [16] 県央ブロックごみ・し尿処理広域化 県央ブロックごみ・し尿処理広域化基本構想(案)の概要. <http://town.iwate.iwate.jp/wp-content/uploads/2014/09/siryou2-gaiyouban.pdf>
- [17] 福島原発事故による土壌の放射能汚染対策. 関 勝寿. 経営論集 78号 p.13-26
- [18] 除染関係ガイドライン平成25年5月 第2版. 環境省. <http://josen.env.go.jp/material/>