

## II-107

## 津波防災アンケートを対象としたテキストマイニング手法の適用

東北大学大学院 正 員 ○渡邊まり子

東北大学大学院 正 員 今村文彦

## 1. はじめに

現在、アンケート調査は世論調査をはじめ、社会調査、意識調査などで幅広く使用されている。防災分野でも様々なアンケート調査が行われており、調査結果は防災事業を行う上での重要な判断材料として役立てられている。

一般的にアンケートの質問項目は、すでに用意されている選択肢から回答を選択する「選択式」と「自由記述」に大別される。選択式の質問が統計的な手法で分析される一方、これまで自由記述回答の分析は殆ど行われてこなかった。<sup>a)</sup> また、自由記述回答を人手で分析している例もあるが、定量的な分析までは行われておらず、分析者の主観に影響を受けるのは避けられない。

自由記述文は調査対象者の真意を知るための貴重な情報源であり<sup>1)</sup>、テキストマイニングによって定量的に分析可能なことが報告されている。そこで、本研究では防災分野のアンケートの自由記述回答に対するテキストマイニング手法の有効性を検証し、その効果と課題を示す。

## 2. テキスト分析手法

「テキストマイニング」は、テキストを定量的に分析する手法として近年注目を集めている手法である。テキストマイニングにも様々な手法があるが、本研究では「テキスト分類」を適用した分析結果を報告する。

テキスト分類とは「テキストを予め定められたカテゴリに分類すること」<sup>2)</sup>をいう。テキストとカテゴリ間の類似度を測り、最も類似しているカテゴリにテキストを分類する。類似度を測る方法は様々であるが、分類規則を単語または単語の組み合わせで表し、テキストと単語単位での比較を行う方法が一般的である。

テキスト分類は自由記述回答の分析に応用されることが多く、分類を自動化する研究も進められている<sup>3)</sup>。しかし、防災分野のアンケートに関しては、これまで適用した事例がなく、カテゴリも分類規則も未知である。そこで、本研究では防災分野の自由記述回答に対し、テキスト分類の手法を用いた分析と評価を行った。

## 3. 本研究でのテキスト分類

本研究では「平成13年岩手県津波避難対策検討委員会報告書<sup>4)</sup>」内の住民意識調査の自由記述回答を対象に分類を行った。設問は「津波防災対策に関して、特に要望すること、気づいたこと等があればお書きください。」であり、有効回答数は226件であった。

類似度を測る方法として、2で示した単語単位での比較を行った。防災分野での分類事例がないことから、アンケート中の他の設問や防災事典<sup>5)</sup>等の知識体系を参考にしてカテゴリと分類規則を定めた(表1)。自由記述回答中に表1で示す単語のいずれかが含まれていれば、カテゴリへ分類する。

表1：使用するカテゴリと分類規則

| 大分類   | カテゴリ名   | 分類規則                                |
|-------|---------|-------------------------------------|
| ハード対策 | 防災設備    | 防波堤 or 防潮堤 or 水門 or 堤防              |
|       | 避難施設    | 避難+所 or 避難 or 場所 or 避難+ビル           |
|       | 避難経路    | 経路 or 道路 or 歩道                      |
|       | 防災無線    | 防災+放送 or 防災+無線 or スピーカー or 拡声器      |
|       | 避難標識    | 標識                                  |
| ソフト対策 | 防災訓練    | 訓練                                  |
|       | 情報伝達    | 情報 or テレビ or ラジオ or 注意報 or 警報 or 勧告 |
|       | 災害時要援護者 | 災害+援護者 or 高齢者 or 障害者 or 弱者          |
|       | 防災教育    | 教育 or 教える or 指導 or 学校 or 学習 or 知識   |
|       | 景観・自然   | 景観 or 景色 or 環境 or 自然 or 生態          |
|       | 防災マップ   | 浸水予測+図 or マップ                       |
|       | 備蓄      | 備蓄 or 準備 or 常備 or 非常持ち出し            |
|       | 観光客対策   | 観光客 or 約り客 or 外来者 or 訪問者            |
|       | 自主防災組織  | 自主+防災+組織                            |

一つの自由記述回答中に、複数の観点からの回答が含まれている可能性があることを考慮し、下記に示すルールを定めた上で分類作業を行った。

## 【比較単位】

- 分類する最小単位は1文とし、内容に応じて複数文の塊として分類する。この塊を「1意見」とする。

## 【比較時の規則】

- 文字列の部分一致で比較する。

a) 2002年地域安全学会論文集に掲載されている論文の内、アンケート調査を行っている論文は7件あり、その内、自由記述回答を分析しているものは0件であった。また、Webで入手できる自治体による防災に関する意識調査8件のうち、自由記述回答を設けているものが3件あったが、分析が行われているものは0件だった。

b) 平成16年「和歌山県 防災(地震)に関する県民意識調査」では、自由記述回答に対し独自の分類が行われていた。

- 異表記の単語は同義語として扱う。  
(ex:釣り客=つり客)
- 1 意見中に複数の規則が含まれる場合は、複数のグループに分類する。

#### 4. 分類結果

自由記述回答を分類した結果を表2に示す。

表2：分類規則に基づいた分類結果

| 大分類       | カテゴリ    | 規則で分類できた意見 |      |   |       | ④  |
|-----------|---------|------------|------|---|-------|----|
|           |         | 件数         | ①    | ② | ③     |    |
| ハード<br>対策 | 防災設備    | 48         | 44   | 4 | 8.3%  | 1  |
|           | 避難施設    | 36         | 35   | 1 | 2.8%  | 3  |
|           | 避難経路    | 18         | 16   | 2 | 11.1% | 2  |
|           | 防災無線    | 14         | 14   | 0 | 0.0%  | 5  |
|           | 避難標識    | 8          | 8    | 0 | 0.0%  | 0  |
|           | ハード対策合計 | 135 (件数+④) |      |   |       |    |
| ソフト<br>対策 | 防災訓練    | 38         | 37   | 1 | 2.6%  | 2  |
|           | 情報伝達    | 25         | 24   | 1 | 4.0%  | 5  |
|           | 災害時要援護者 | 17         | 17   | 0 | 0.0%  | 3  |
|           | 防災教育    | 14         | 14   | 0 | 0.0%  | 4  |
|           | 景観・自然   | 8          | 8    | 0 | 0.0%  | 3  |
|           | 防災マップ   | 8          | 7    | 1 | 12.5% | 0  |
|           | 備蓄      | 4          | 4    | 0 | 0.0%  | 3  |
|           | 観光客対策   | 3          | 3    | 0 | 0.0%  | 1  |
|           | 自主防災組織  | 1          | 1    | 0 | 0.0%  | 0  |
|           | ソフト対策合計 | 139 (件数+④) |      |   |       |    |
| 分類済の意見合計  |         | 242        | ④の合計 |   |       | 32 |
| その他未分類    |         | 49         |      |   |       |    |
| 全意見の合計    |         | 323        |      |   |       |    |

- ①回答中に分類規則を含み、かつカテゴリにふさわしい内容の意見の数  
 ②分類規則を含むが、内容がふさわしくない意見の数  
 ③規則を含む全意見中の②の割合  
 ④規則を含まないが、カテゴリに含まれるべき内容の意見の数

まず、適用した分類手法の有効性について検証した。カテゴリに分類できたのは全意見323件のうち242件で、全体の約75%であった。7割以上の意見が機械的に分類できたことから、防災分野の自由記述回答は単語の規則を用いて分類できているといえる。

ただし、表2の②と③のような誤分類や未分類、そして④のように分類されなかった意見が存在することから、カテゴリと分類規則については、評価を繰り返す必要がある。

分類後の各カテゴリを集計した結果、ハード対策への意見は135件で全体の約42%、ソフト対策への意見は139件で約43%となった。この結果から、ハード・ソフト対策とも意見数はほぼ等しく、意見の偏りは見られなかったことが分かる。最も意見が多かったのは「防災設備」で、全体の約15%を占めていた。

これらの集計データを元にすれば、自由記述欄全体の特徴や傾向等を定量的に調べることができる。テキストマイニングの技術により、自由記述回答からも有意なデータを得られる可能性が示唆されたといえる。

カテゴリの情報を用い、更に詳細な分析を進めることが可能である。次章で具体的な事例を示す。

#### 5. 分析事例

4の分類結果と回答者の属性を組み合わせ、さらに詳細な分析を行った。自由記述回答を性別で集計した結果、男性111名、女性115名であった。表2の分類結果から最も意見数が多かった「防災設備」について、正しく分類された①44件のうち、重複を除いた回答人数の性別比を求めると、図1の結果が得られた。



図1：性別の割合(防災設備)

図1より、回答人数36人のうち男性の回答数が約7割を占めていた。25人という値は男性111人の約23%を占めることから、防災設備カテゴリは特に男性の意見が多い傾向にあることが判明した。

#### 6. おわりに

テキストマイニングの技術を活用し、分類規則を定めて防災分野アンケートの自由記述回答を分類・分析できることを示した。今後は本研究の成果を多数のアンケートに適用し、防災分野の自由記述回答に最適な分類手法とその評価方法、そして分類の自動化について研究を進める予定である。同時に今回の分類結果や過去の研究を元にした、防災知識の体系化についても検討を重ね、分類結果の評価基準となるような防災分野の知識体系の構築にも尽力する。

#### 参考文献

- [1]鎌原雅彦,宮下一博,大野木裕明,中澤潤.心理学マニュアル質問紙法.北大路書房.1998.
- [2]徳永健伸.情報検索と言語処理.京大出版会.1999.
- [3]金明哲,村上勝,永田昌明,大津起夫,山西健司.言語と心理の統計.岩波書店.2003.
- [4]岩手県津波避難対策検討委員会報告書  
[http://www.pref.iwate.jp/~hp010801/tsunami/houkokusyo\\_index.htm](http://www.pref.iwate.jp/~hp010801/tsunami/houkokusyo_index.htm)
- [5]日本自然災害学会.防災事典.築地書館.2002.