

# テキストマイニングに基づくガス専焼ラインの地震リスク被害の分析

22T0216Z 米山和輝

指導教員：丸山喜久

## 1. 研究背景・目的

日本では地震をはじめとする自然災害が多発しており、災害時に非常用電源設備として、ガスの火力を利用して自家発電を行うガス専焼自家発電設備の導入事例が増加している。この設備を導入するにあたって、専門家らはガス専焼自家発電設備用ガス供給系統評価委員会(ガス評価委員会)を開いて地盤の液状化リスクや導管の耐震性を議論してきた。しかし、これまでの評価委員会等の協議記録(議事録)を体系的に整理・活用する仕組みは十分に整っていない。

このような背景を踏まえて本研究では、議事録に対してテキストマイニングを適用しその内容を可視化し、「地震被害が疑われ詳細な検討を要する」(被害あり)文章と「地震被害の懸念がない」(被害なし)文章を判別するモデルの構築を目的とする。

## 2. 使用したデータとテキストマイニング

本研究ではガス評価委員会の議事録175回分を使用した(1998年10月25日~2025年3月7日)。

テキストマイニングとは、テキストデータを文字や単語、フレーズ等の単位に分解し、言葉の出現頻度や言葉同士の関連を数量化し客観的な分析結果を示す手法である。本研究では、テキストマイニングにおける情報抽出の手法として、単語間の共起関係を明らかにする共起ネットワーク図を作成した。作成した共起ネットワーク図を図1に示す。

また、議事録内での単語ごとの重要度を測るために、PageRank<sup>1)</sup>を採用した。PageRankとはWebページの相対的な重要度を測定する手法である。Webページの重要度は他のページからの被リンク数によって評価されるという考え方にに基づき、Webページごとに値が算出される。PageRankを算出する式を示す。

$$PR(A) = (1 - d) + d \left( \frac{PR(T_1)}{c(T_1)} + \dots + \frac{PR(T_N)}{c(T_N)} \right) \quad (1)$$

ここで、PR(A)はページAのPageRankの値を示している。ページT<sub>1</sub>からページT<sub>N</sub>はページAにリンクを貼っているページを示している。パラメータdはダンピングファクターと呼ばれるものであり、0から1で表される。ここではダンピングファクターd

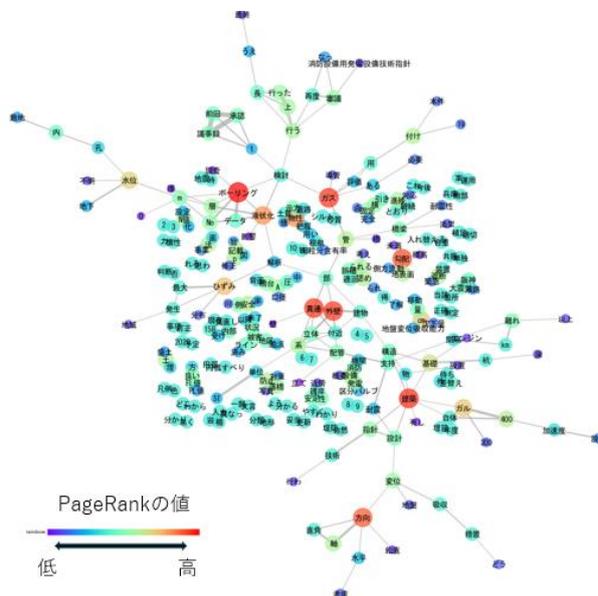


図1 共起ネットワーク図

表1 PageRankの値上位15語(特徴語)

| 順位 | 単語           | PageRank値 |
|----|--------------|-----------|
| 1  | <u>ボーリング</u> | 0.009533  |
| 2  | <u>建築</u>    | 0.008548  |
| 3  | <u>貫通</u>    | 0.008372  |
| 4  | <u>外壁</u>    | 0.008329  |
| 5  | <u>勾配</u>    | 0.008211  |
| 6  | <u>液状化</u>   | 0.008036  |
| 7  | <u>方向</u>    | 0.008027  |
| 8  | <u>物性</u>    | 0.007430  |
| 9  | <u>遮断</u>    | 0.006944  |
| 10 | <u>ひずみ</u>   | 0.006852  |
| 11 | <u>管</u>     | 0.006729  |
| 12 | <u>ガル</u>    | 0.006704  |
| 13 | <u>cm</u>    | 0.006594  |
| 14 | <u>水位</u>    | 0.006566  |
| 15 | <u>ガス</u>    | 0.006461  |

は0.85としている。またC(T<sub>N</sub>)はページT<sub>N</sub>から出ているリンクの数である。

本研究では、Webページの部分を単語に置き換えてこの考え方を用い、PageRank値上位15語(特徴語)を抽出した。また特徴語に共起している単語(共起語)

も抽出した。PageRank 値上位 15 語(特徴語)を表 1 に示す。

### 3. 被害判別モデルの構築と精度

地震による被害有無の判別モデルを構築するために、OpenAI が開発した大規模言語モデルに基づく対話型生成 AI のサービスである ChatGPT<sup>2)</sup>を用いて議事録の被害あり文章に頻出する表現(懸念語)の抽出と擬似被害文章の生成を行った。ChatGPT を利用した擬似被害文章の生成は、テキストマイニングで得られた特徴語と共起語に加え懸念語の 3 つの表現を使用して 2400 文(被害あり 1200 文、被害なし 1200 文)を生成した。どちらの文章にも、特徴語と共起語を利用する。また、被害あり文章には懸念語を含む文章と含まない文章の両方を生成した。被害なし文章には懸念語を含まずに生成した。

作成した擬似被害文章 2400 文を訓練データとしてロジスティック回帰分析(以下、ロジットモデル)を行い、実際の議事録の文章を被害ありと被害なしに分類するモデルを構築した。テストデータとして(a)すべての議事録の中から選んだ文章と(b)特徴語が含まれているものの中から選んだ文章の 2 つを用意した。

精度の検証には ROC 曲線を用いた。被害あり文章を被害ありと判定する率を感度 TPR、被害なし文章を被害なしと判定する率を特異度 TNR とする。TPR と TNR は表 2 の混同行列に基づき、式(2)(3)で表される。

$$TPR = \frac{TP}{TP+FN} \quad (2)$$

$$TNR = \frac{TN}{TN+FP} \quad (3)$$

表 2 被害予測モデル検証のための混同行列

|    |                  | 予想                              |                                 |
|----|------------------|---------------------------------|---------------------------------|
|    |                  | positive<br>被害あり                | negative<br>被害なし                |
| 実際 | positive<br>被害あり | True<br>positive(TP)<br>被害あり的中  | False<br>negative(FN)<br>被害あり外れ |
|    | negative<br>被害なし | False<br>positive(FP)<br>被害なし外れ | True<br>negative(TN)<br>被害なし的中  |

ROC 曲線は縦軸 TPR と横軸(1-TNR)で表される。判別能力が高いほど、ROC 曲線はグラフの左隅に近づく。テストデータ(b)の ROC 曲線を図 2 に示す。

精度は AUC(Area Under the Curve)で評価し 1 に近づくほどモデルの精度が高い。また、テストデータ

ごとの AUC、しきい値、TPR、TNR を表 3 に示す。本研究で AUC はそれぞれ(a)0.648、(b)0.666 となり特徴語を含む文章を対象とした場合の方が精度が高い結果となった。また、テストデータ(b)の TNR は約 0.85 と高く、被害なしデータの抽出に有効と考えられる。

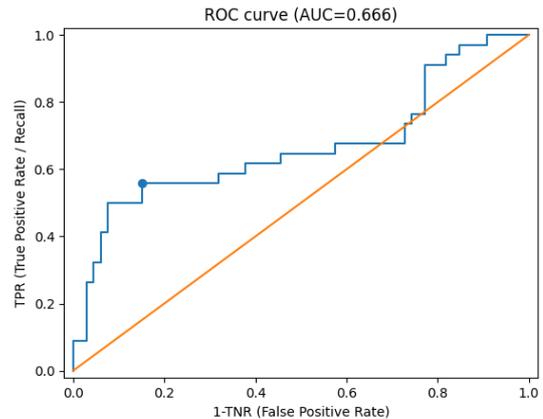


図 2 テストデータ(b)の ROC 曲線

表 3 ロジットモデルの判定結果の精度

| テストデータ | AUC   | しきい値  | TPR/TNR     |
|--------|-------|-------|-------------|
| a      | 0.648 | 0.713 | 0.617/0.713 |
| b      | 0.666 | 0.749 | 0.558/0.849 |

### 4. まとめ

本研究では、ガス評価委員会の議事録をテキストマイニングで可視化し、その内容を基に擬似被害文章を作成し、実際の議事録の文章を被害ありまたは被害なしに分類する被害判別モデルを構築した。2 種類のテストデータを用いた結果、特徴語を含む文章を対象とした場合の方が精度が高く、特に被害なしデータの抽出に有効と考えられる。しかし全体として十分な精度は得られなかった。

今後の展望としては特徴語の追加導入により、擬似被害文章の表現多様性を高めること、議事録のさらなる内容分析のためのテキストマイニングの手法を検討することで、生成する文章の内容を充実させることが挙げられる。また、機械学習の面では、本研究ではロジスティック回帰分析を用いたが、他のモデルの検討も行うことで、より被害の判別精度の高いモデルの構築を目指す。

#### 参考文献

1) Sergey Brin, Lawrence Page (1998) 「 The anatomy of a large-scale hypertextual Web search engine」『Computer Networks and ISDN Systems 30』 pp107~117

2)OpenAI: “Introducing ChatGPT”

<https://openai.com/blog/chatgpt> 閲覧日 1 月 22 日