AI技術を活用した確率論的リスク評価手法の高度化研究 信頼性データベース構築のための自動故障判定手法の開発（2022年度）

森本　達也*　　氏田　博士**

Development of Probabilistic Risk Assessment Methodology Using Artificial Intelligence Technology Automatic Fault Detection Method for Building Reliability Database (FY2022)

Tatsuya Morimoto* and Hiroshi Ujita**

本稿では、文部科学省からの委託業務として、国立研究開発法人 日本原子力研究開発機構（以下、JAEA）と共に当社が実施中の、「AI 技術を活用した確率論的リスク評価手法の高度化研究」における「信頼性データベース構築のための自動故障判定手法の開発」について紹介する。

Keywords: PRA、信頼性データベース、AI、自然言語処理、テキストマイニング

1. はじめに

原子力発電所の確率論的リスク評価（以下、PRA）は、解析作業が膨大で事業者の負担となっているとともに、国際的に検討されているリスク情報活用アプローチを国内に導入するに当たっての懸案となる可能性がある。この要因は、膨大な設計資料等の読み込みと理解、信頼性データや評価モデルの構築を習熟した技能者が経験に基づいて手作業で入力しなければならない状況にある。この課題を解消するため、本研究では、文部科学省からの委託業務として、AI やデジタル化技術を活用し、手作業を自動化することによって原子力発電所 PRA の省力化・等質化を目指して、運転時の PRA におけるフォルトツリー（以下、FT）作成及び信頼性データベース構築に着目した AI ツールを 2022 年度からの 3 か年で開発し、PRA 手法の高度化を図る計画である[1]。

本研究のうち、JAEA と共に、当社が担当している「信頼性データベース構築のための自動故障判定手法の開発」では、信頼性データベース構築に資するために、テキストマイニングや深層学習等の AI 技術を活用し、各原子力プラントの故障及びトラブル情報から PRA に必要な故障原因を自動的に判定する手法を開発している。

2022 年度は、信頼性データベース構築に必要な情報として NUCIA[2]から故障機器、故障原因等を AI 技術等により抽出し、データベース化する AI ツールの方法論を構築し、試作した。2023 年度は、そのデータの事象や時間の相違を AI 技術により分析することにより、電力会社、プラント及び職種の相違判断、福島第一原子力発電所事故前後の相違判断等が可能な手法を試作する。2024 年度は、共通特性を AI 技術により分析することにより、共通機器、共通操作、共通組織特性等の共通要因を判断可能な手法を試作する[1]。

本稿では、2022 年度の試作内容と得られた成果・課題（専門家による評価）について紹介する。

2. 信頼性データベース構築に関する背景

信頼性データベース構築に関する背景としては、電力中央研究所や原子力発電推進協会の努力で軽水炉のためのデータベースとして NUCIA が、

*アドバンスソフト株式会社　第 5 事業部
5th Computational Science and Engineering Group, AdvanceSoft Corporation
**アドバンスソフト株式会社　リスク研究開発センター
Risk Research and Development center, AdvanceSoft Corporation

また JAEA の努力で高速炉のためのデータベースとして CORDS[3]が開発されている。これらは、個別事例も参照できるが、データベース探索により分析の支援も可能であり、故障率データベース作成も実施されてきている。

しかし、データ量が膨大であるため分析に多大な人的資源が費やされてきたこと、また多くの担当者が関わってきたために分析にバラツキや偏りの可能性が高いことが課題となっている（これを一次分析とする。本研究 2022 年度対象）。さらには、膨大なデータ量の全体を見通すことが困難なために、時間的・空間的な事象の特徴を抽出することが実質的に不可能であることが、最大の課題と言える（これを二次分析とする。本研究 2023 年度以降対象）。

そこで、AI 手法を活用することにより、下記のように課題を解決できるものと考える。最大の期待は、AI による分析により人間では見出すことができない新たな知見を見出す可能性にある。

- 課題 1（一次分析）: 信頼性データベース構築の効率向上
  - 迅速性、省力化の促進（ビックデータ処理）
  - 正確性、統一性の促進（分析者の個性差無し）
- 課題 2（二次分析）: 時間的・空間的特徴抽出の分析能力向上
  - 統一性の向上（テキストマイニング・データマイニング手法による総合的分析の促進）
  - 新知見の獲得（ビックデータ処理による時間的・空間的特徴の自動抽出による共通因子や経時変化等の故障特性の摘出）。

3. 試作した AI ツールの全体フロー

2022 年度は、信頼性データベース構築に必要な情報として、NUCIA から故障発生個所（系統・機器）及び故障原因を AI 技術等により抽出し、データベース化する AI ツールの方法論を構築し、試作した。その全体フローを図 1 に示す。

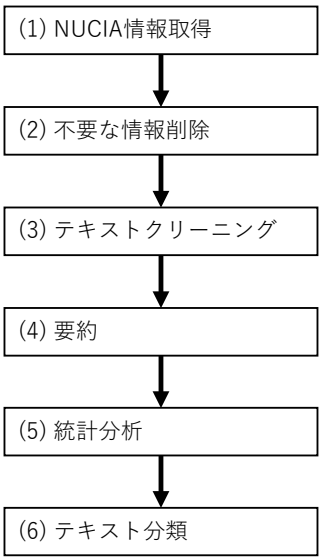

図 1　NUCIA から PRA の信頼性データベースを構築する AI ツールの全体フロー

(1)～(6)の概要について以下に示す。また、(1)～(6)の要点について 4 章に示す。

(1) NUCIA 情報取得

NUCIA サイト（トラブル情報等⇒国内原子力発電所⇒情報検索）に記載されている全情報（2023 年 1 月 27 日時点で 7377 件の事象が存在）を取得し、csv ファイルに保存する。

(2) 不要な情報削除

NUCIA 記載の全情報が信頼性データベースに必要となるわけではないため、専門家判断による削除方針を実装した独自プログラムによって、ルールベースで不要な事象を削除する（2023 年 1 月 27 日時点の 7377 件の場合⇒1941 件となる）。

(3) テキストクリーニング

NUCIA の「事象の原因」文章は、箇条書きであったりなかったり、経緯や背景まで含めた長文であったり原因のみの短文であったりと、事象によって非均質な文章構造である。

よって、後段の要約や統計分析の精度向上のために、NUCIA の「事象の原因」文章に対して、文章構造等を考慮した独自プログラム及び自然言語処理フレームワークの機能でテキストクリーニングを実行する。

(4) 要約

人間の目で見ても必要十分な情報量を見定めることが可能となるように、また、統一的に文章を比較する際の情報量の差を少なくするために、ある程度文章量を揃える必要がある。よって、NUCIA の「事象の原因」文章（テキストクリーニング済み）に対して要約処理を実行する。

(5) 統計分析

人間では見出すことができない新たな知見を見出す可能性を考慮した分析の一環として、NUCIA の「事象の原因」文章から抽出した単語を対象として、頻出グラフや共起ネットワークを作成する。

(6) テキスト分類

NUCIA では基本的に、1 つの事象に 1 つの原因分類コードが設定されている（ただし、1941 件中 63 件は 1 つの事象に複数の原因分類コード有り）。

今回、専門家によって確認したところ、本来複数の原因分類コードが設定されているべき事象が散見されたため、NUCIA で各事象に設定されている原因分類コードをそのまま教師データ（正解ラベル）としては利用できないと判断した。

よって、NUCIA の「事象の原因」文章に対して、教師無しのマルチラベル分類として Zero-shot 分類を実行する。マルチラベルの候補は、NUCIA で使用されている原因分類コードとする。これにより、故障原因の自動抽出を実現する。

4. 試作した AI ツールの各処理の要点

4.1. NUCIA 情報取得

Python で開発されたオープンソース（3 条項 BSD ライセンス）のクロールフレームワークである Scrapy を使用し、NUCIA サイトのトラブル情報等⇒国内原子力発電所⇒情報検索の 1 ページ目（通番昇順でソート、1 ページ 100 件表示）をスタート URL として、取得処理を実行した。

便宜上、NUICA サイトの事象一覧画面を「Search 画面」、各事象画面を「View 画面」と呼ぶ。Search 画面には全 10 個のデータ項目が、View 画面には全 38 個のデータ項目が存在する。全ページ（2023 年 1 月 27 日時点で 74 ページ）及び全事象（2023 年 1 月 27 日時点で 7377 件）におけるそれらのデータ項目を全て取得し、csv ファイルに一覧表として整理した。

4.2. 不要な情報削除

2022 年度は、専門家判断で削除方針を検討し、その明確なルールを Python で実装して削除処理を実行した（その結果、2023 年 1 月 27 日時点での 7377 件が 1941 件となった）。

その削除方針（4 点）を以下に示す。

- NUCIA の事象には「トラブル情報（T）」「保全品質情報（M）」「その他情報（S）」の 3 種類が存在するが、重要性と記載情報量が比較的少ない「その他情報（S）」は信頼性データベース構築には不要とする。
- 平成 15 年（2003 年）9 月以前のトラブル情報等は正式の報告義務前の情報につき、データ要件の相違があり信頼性データとして用いることは難しいと判断し不要とする。
- 「事象の原因」と「再発防止対策」の両方の記載内容が「検討中」「調査中」「記載無し」等の事象は、情報不足として削除する。
- 「事象の原因」や「再発防止対策」の記載内容が、「他の事象参照」「添付ファイル参照」等となっている事象は、重複情報として削除する。

また、NUCIA 記載のデータ項目の内、信頼性データベース構築に必要と考えられる以下の項目のみを、csv ファイルに出力するようにした。

NUCIA 記載の重要項目

- 通番
- 情報区分
- 事象発生日時
- 会社名
- 発電所
- 件名

- 事象発生箇所_設備
- 事象発生箇所_系統
- 事象発生箇所_機器(大分類)
- 事象発生箇所_機器(小分類)
- 事象発生箇所_部品
- 原因調査の概要
- 事象の原因
- 再発防止対策
- 原因分類(大分類)
- 原因分類(小分類)

4.3. テキストクリーニング

4.3.1. 独自プログラムによるクリーニング

NUCIA の「事象の原因」文章は上述した通り非均質な文章構造であるため、また、必要な数値情報を残す必要があったため、まずはそれらに対応可能な独自プログラムを Python で作成してクリーニングを実施した。これにより、必要な情報を残しつつ、後段の形態素解析結果精度向上を実現した。

具体的な処理内容は、正規化、読点の統一、箇条書き項目とセンテンス途中の改行の区別、箇条書き用の数字・記号の削除、不要な特殊文字の置換や削除、原子力関係用語の統一、NUCIA で設定されているコード（原因分類、設備、系統、機器）の単語を抽出しやすいように区切る、などである。

4.3.2. 自然言語処理フレームワークの機能によるクリーニング

日本語用の自然言語処理フレームワーク spaCy/GiNZA[4][5]（MIT ライセンス）と、国語研長単位モデル ja_gsdluw[6]（CC BY-SA 4.0 ライセンス）を使用して、NUCIA の「事象の原因」文章に対して形態素解析を実施した。

その際、可能又は必要な範囲で、GiNZA が使用しているトークナイザーにユーザー定義辞書を追加し、spaCy の機能で固有表現辞書及びストップワードを追加した。

これらにより、形態素解析によるトークン化の精度が向上し、後段の統計分析において必要となる単語のみを正確に抽出できるようにした。また、要約処理におけるシードキーワードによる誘導をしやすくした。

4.4. 要約

自然言語処理における要約方法には、大別して抽出型と抽象型の 2 種類がある。本業務では、原子力トラブル情報を扱う性格上、文章や単語の表現の正確性を重視し、既存の文章をそのまま抽出する抽出型要約を採用した。

まずは、Sentence-BERT の日本語モデル（バージョン 2）である sonoisa/sentence-bert-base-ja-mean-tokens-v2[7]（CC BY-SA 4.0 ライセンス）を使用して、NUCIA の「事象の原因」文章をセンテンス（句点区切り）毎に文脈を踏まえてベクトル化（埋め込み）した。

次に、NUCIA コード（原因分類、設備、系統、機器）の単語をシードキーワードとし、シードキーワードで重み付けしたセンテンスベクトルと元のセンテンスベクトルでコサイン類似度を計算し、文章を構成する各センテンスの関連性を表す類似度マトリックスを作成した。シードキーワードを考慮することにより、シードキーワードを含むセンテンスの重要度が増加し、要約文章として抽出されやすくなる（シードキーワードによる誘導）。

最後に、LexRank で重要なセンテンスをランク付けして抽出した。LexRank とは、文章を構成する各センテンスの関連性（ここでは、コサイン類似度による類似性マトリックス）をスコアにして、関連性の高いものが重要なセンテンスであるとして抽出する、グラフベース手法のアルゴリズムである。スコアの高い順にセンテンス配列をソートし、スコア上位 10 個を抽出することで要約文章とした（センテンスが 10 個以下の場合は全て抽出）。

4.5. 統計分析

NUCIA の「事象の原因」文章から抽出した単語（ここでは、品詞を名詞、動詞、数詞の 3 種類に限定）に対して、自然言語処理における基本的な可視化を簡略化した nlplot[8]（MIT ライセンス）

や、複雑なネットワーク構造等を作成・操作する NetworkX[9]（3 条項 BSD ライセンス）の Python パッケージを使用して頻出単語グラフと共起ネットワークを試作した。それらの試作例を図 2、図 3 に示す。

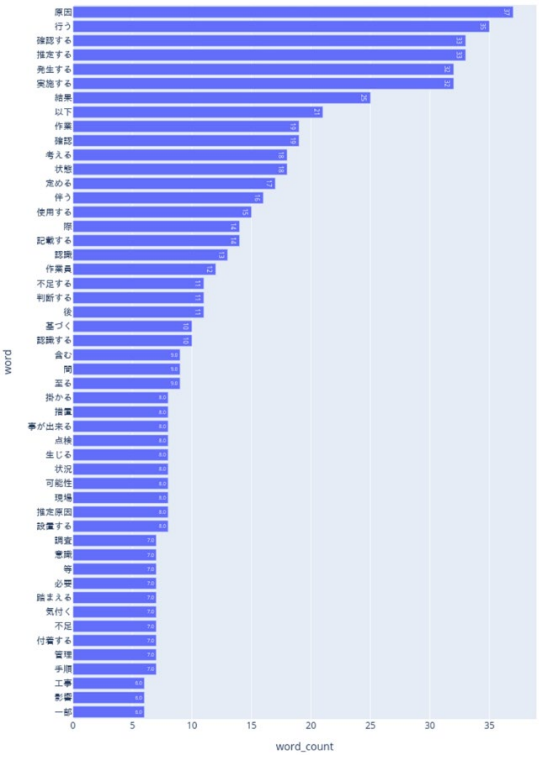

図 2　頻出単語グラフの一例（原因分類コードが「管理不良」の場合）

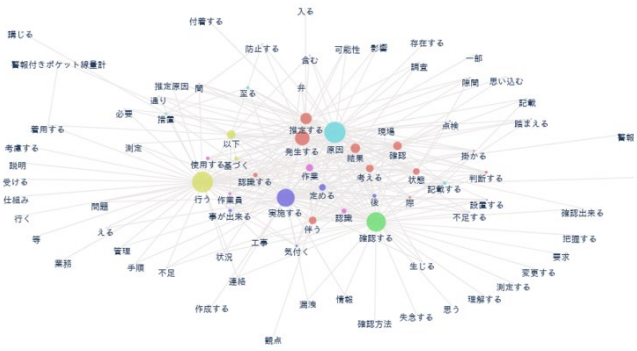

図 3　共起ネットワークの一例（原因分類コードが「管理不良」の場合）

4.6. テキスト分類

日本語にも対応した Zero-shot テキスト分類用の事前学習済み自然言語モデル MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7[10]（MIT ライセンス）を使用し、NUCIA の原因分類コード（大分類：9 種類、小分類 37 種類）をマルチラベル候補（ユーザ指定）として、NUCIA の「事象の原因」文章に対して Zero-shot テキスト分類を試行した。

先述の通り、NUCIA の原因分類コードは基本的に 1 事象に 1 つ設定されているが、本来は複数原因が存在すると考えられる事象が散見される。現状、複数原因で分類されたマルチ正解ラベルは一部（今回処理対象とした 1941 件中 63 件）しか存在しないため、試行結果の分類精度を評価することができない。よって本業務では、処理対象とした各事象（1941 件）に対して設定されている原因分類コードを、予測結果の中で最もスコアが高くなるべきラベルと仮定して分類精度を評価した。すなわち、少なくとも既存の分類を予測できているかどうかを、とりあえずの精度評価とした。

本業務での試行結果に対する分類精度を表 1 に示す。分類精度は、PyTorch で用意されている metrics の 1 つである MultilabelRankingLoss によって算出した。MultilabelRankingLoss は、誤って並べ替えられたラベルペアの平均数に対応するものであり、最高スコアはゼロとなる。本業務においては、予測結果のスコアが最も高いラベルと正解ラベルが全事象で一致すれば Loss=ゼロとなる（正解ラベルが N 個ある場合は、スコアのランキング上位 N 個が一致してればゼロ）。

表 1　Zero-shot テキスト分類試行結果の分類精度

要約	大分類コード	小分類コード
無し	0.30289	0.18632
有り	0.31712	0.19863

要約文章の方が分類精度が低い理由は、要約によって必要な単語やセンテンスが除去されてしまう場合があるためだと考えられる。また、小分類コードの方が分類精度が高い理由は、小分類コードの方がより文章に現れやすく、類似した表現を見つけやすい単語であるため、Zero-shot におけ

る仮説文章として正しいと判断されやすいのだと考えられる。

5. 専門家による結果評価

要約、統計分析、テキスト分類の結果について、一部の事象を抽出して専門家による評価を実施した。その内容を以下に示す。

5.1. 要約の評価

(1) 短文の事例：NUCIA の通番 8820 の事象

全文で 100 文字程度であるためほぼ要約処理はなく、適切な内容と判断できる。

(2) 長文の事例：NUCIA の通番 11455 の事象

要約処理により文章量が約 1/6 になっており、知見要約、原因要約、直接原因はほぼ記載されているが、背後要因（安全文化の評価）は記載がなく不十分である。

重要ではあるが、以下の a.及び b.のような文章を省く判定基準が必要である。なお、本事例のようにサイト間の比較と共通要因の抽出を実施した分析例は少なく、このような特徴のある分析内容を AI 手法で別途導き出す手順は重要である（2023 年度の課題）。

a. 異なる号機間の対応比較
b. それに基づく共通要因

逆に、背後要因（安全文化の評価は必須要件）は残す判定基準が必要である。また、ヒューマンファクタや安全文化の評価手法は、重要な研究テーマである（2024 年度の課題）。

5.2. 統計分析の評価

人間では見出すことができない新たな知見を見出す可能性を考慮した分析の一環として実施したが、特筆すべき発見は無かった。とりあえず、原因が記載された文章において使用されるケースが多い表現（原因、確認、推定、発生、結果、認識、不足、判断、可能性など）を把握した。これらの表現は、要約処理におけるシードキーワードとして活用できると考えられる。

5.3. テキスト分類の評価

複数原因が関連する事例として、NUCIA の通番 12504 の事象について評価した。

大分類と小分類を独立に見ると、それぞれ NUCIA 分類と Zero-shot 分類の対応性は高い。ちなみに NUCIA では、ルール表から大分類を決定しそこからさらに小分類を選択している。

NUCIA の原因分析の経緯を見ると、8 個の直接要因、9 個の組織要因（7 個のリスク管理と 2 個の安全文化の課題）を摘出している。最終的には、小分類で原因を特定することになるので、小分類の分析精度が決め手であり、今回の例で見る限りほぼ同じ原因を特定できている。

ただし、専門家分析では、原因分類コードに無い分類を採用する可能性があり、自由度の高い分類方法の検討が必要であろう（2023 年度の研究課題）。長期的課題としては、複数原因の因果関係を明らかにし、原因ネットワークを作図できるようにすることである（2023 年度の研究課題）。もう一つの長期的課題としては、組織要因を摘出する機能が必要となる（2024 年度の研究課題）。

6. まとめ

NUCIA から信頼性データベースを構築する AI ツールの方法論を、自然言語処理フレームワークや事前学習済み大規模自然言語処理モデル等を活用して構築し、試作した。これにより、信頼性データベース構築に必要な情報である故障発生個所（系統・機器）や故障原因等を自動抽出可能となった。

試作したツールでの要約処理や特徴抽出結果を専門家の分析と比較し、課題はあるものの実用化可能な方法であると判断した。さらに、抽出した課題に基づいて今後の分析方針を策定し、次年度以降の方針を確立することができた。

本研究は、文部科学省原子力システム研究開発事業 JPMXD0222682583 の助成を受けたものである。

参考文献

[1] 二神　敏，山野　秀将，栗坂 健一，氏田博士, AI 技術を活用した確率論的リスク評価手法の高度化研究,（1）AI ツールの開発計画，日本原子力学会　2023 年春の年会,2C10(2023.3.14)

[2] 「ニューシア　原子力施設情報公開ライブラリー」、一般社団法人　原子力安全推進協会（http://www.nucia.jp/）

[3] 栗坂健一「高速炉機器信頼性データベースの開発」、動力炉・核燃料開発事業団、動燃技報　No.98 P.18-P.31、1996 年 6 月.

[4] https://spacy.io/

[5] https://megagonlabs.github.io/ginza/

[6] https://github.com/megagonlabs/UD_Japanese-GSD/

[7] https://huggingface.co/sonoisa/sentence-bert-base-ja-mean-tokens-v2

[8] https://github.com/takapy0210/nlplot

[9] https://networkx.org/

[10] https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7

※ 技術情報誌アドバンスシミュレーションは、アドバンスソフト株式会社 ホームページのシミュレーション図書館から、PDF ファイル（カラー版）がダウンロードできます。（ダウンロードしていただくには、アドバンス/シミュレーションフォーラム会員登録が必要です。）