

## 共通要因としての単語に注目した事象の進展の統合化・可視化 Visualization and integration for event progress

箕輪 弘嗣<sup>†</sup>  
Hirotsugu Minowa

### 1. 緒言

原子力、化学関連の事業所では些細な事が事故に繋がり兼ねない。ひとたび、事故が起きれば、人命だけでなく経済、環境、社会にも多大な損失が発生するため、事故は最も避けるべき事柄である。近年、コスト削減による教育不足、熟練者の退職による事業所の保安力低下から、事故の発生が懸念されていた。それを裏付けるように熟練技術者の大量退職が始まった 2007 年以後、大事故がたて続けに起きており、事故の未然防止の基盤確立は急務である。回避方法として事故事例の活用方法が模索されている。その手段として、事例データを収集して未来に活かそうという取り組みだが、有効な活用方法が見当たらないままである。

本研究では、大局的にデータを捉えるために統計的手立てを模索している。これを実現する方法として、筆者は提案した統合化進展法[1][2]における理論に基づき自動的に解析できるようアルゴリズムを確立する事を目指している。本手法は、事象の説明の核となる事象の進展の主格-述格という部分を共通要因とみなして抽出、統計的に解析する事で複数の事例を大局的に捉えようと試みである。

本論文では、その過程において電算機による機械的な可視化の実現、活用方法の提案について報告する。

### 2. 研究内容

事故事例の解析において、自然言語処理における形態素解析、係り受け解析は高い精度を誇るが、文脈解析や意味解析は研究途上である。では、専門家の解析を活用できるかであるが、こういった解析が適切かは明らかでない。多くは、起回事象や原因事象など特定の要因において人手で分類した項目を事例データに設けていた。例えば PEC-SAFER 事故事例集[3]において、専門家が定義した起回事象の区分を表 1 に示す。表 1 から、起回事象は、特定の機器から起きている事が分かる。しかし、この情報から原因や進展の詳細を直接的に明らかにするのは難しい。

事故の発生に起回事象が最終的にどういった事象（以下、最終事象）に繋がるかを明らかにするためには、事故へ至る経路上に頻発する要因の連なりがその手掛かりとならなにか検討している。従来の進展解析においても、その途中はブラックボックスであり、どういった経緯で起こりうるか俯瞰的に知る事はできなかった。もし、本研究で進展の経緯詳細が可視化され、共通した誘発要因(トリガー)が明らかとなれば、より事故へ至る進展の可視化の実現が期待できると考える。

<sup>†</sup>岡山商科大学, Okayama Shoka University

表 1 起回事象カテゴリ一覧

#	カテゴリ名
1	静止機器の腐食・劣化・破損
2	プロセス状態の変動・異常
3	静止機器の故障、機能喪失・低下
4	計装機器の故障・機能低下・破損・劣化
5	動機器の停止・機能低下・破損
6	その他
7	電気機器の故障・機能低下・破損・劣化
8	動機器の誤起動、予定外の起動

### 2.1 提案内容

事例データ内の進展は、図 1 の様なドミノ現象に代表されるようなイベント駆動型に進展する。その事象を説明文には、ほぼ必ずと言えるほど、それぞれ主部と述部が含まれている。この主部、述部は説明文の核となる部分であるため、図 2 の通り、主部、述部もしくは主語・述語という事象の核を意味する要素で結ぶ事で共通要因としてグループ化し、事例の大局的な傾向を抽出しようという試みである。

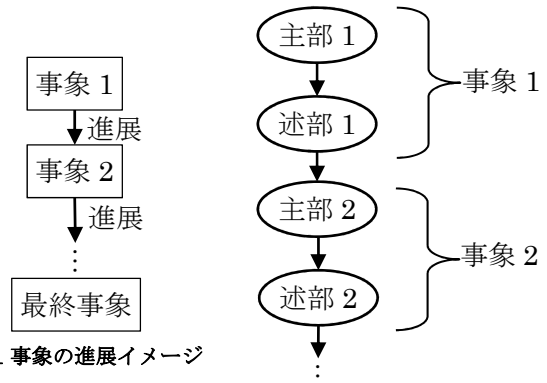


図 1 事象の進展イメージ

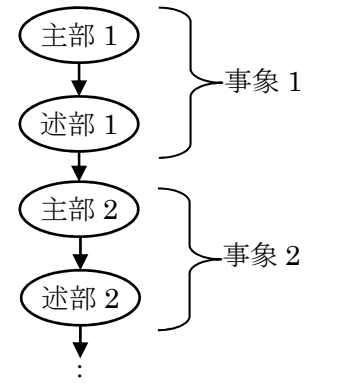


図 2 提案する事象の進展のモデル

### 2.2 相違点

共通要因の抽出に使われる指標として TF-IDF, Key Graph など出現頻度に応じた指標を用いられる事が多い。本研究は、頻度を指標として考慮するが、事象の進展の解析を目的としており、文の骨格となる主部、述部の抽出した要素を共通要因とみなして集約しているところが異なっている。

### 2.3 手順

グラフ可視化については次の通り定義した。

## Step 1. 前処理

対象の事例集の中から、事象の進展に関する説明の抽出。句点揃え。箇条書き、連番の除去。文の整形などを実施する。

## Step 2. 共通要因の頻度解析を実施

説明文の中における共通要因を抽出する。主部・述部のペアを抽出できれば良いが、検出精度の課題から形態素解析でもよいとする。

## Step 3. フィルタリング

解析結果の内、原因となりえない、汎用的な事象といった要因は除去する。

本時点では形態素解析によって得られた名詞の形態素(最小単位の単語)の頻度解析結果を利用する。これには形態素を潜在的なカテゴリとして利用する意図がある[4]。

## Step 4. 頻度上位の名詞を含む単語の解析

事例内に該当単語が含まれる場合は、可視化対象に数え、重複はカウントしない。

## Step 5. 事象の結果との関連付け

Step 2 で求めた単語に対して最終結果(事故)を関連付ける。最終結果の算出手段として、自然言語処理に基づき、最終結果を求める方法、または、事例データに記載されている最終事象という項目を用いる方法が考えられる。

## Step 6. グラフ描画

定めたトップ事象(グラフの根)と最終結果を結び形でグラフを生成する。直観的な把握を助けるため、頻度に応じて辺の太さを変える。また、正確にその進展度合いを取得できるように辺付近に頻度値を記述する。

## 2.4 実装

Python で開発。形態素解析には MeCab, グラフ可視化には Graphviz を使用した。Python には NetworkX や igraph などライブラリの使用を試みたが、NetworkX では有向グラフの矢印の形状が三角ではなく四角形と一見認識しにくく、Python 版 igraph では日本語の表示が困難な問題があった。

## 2.5 使用データ

本評価解析には、PEC-SAFER 事故事例[3]の 423 件を使用した。PEC-SAFER 事故事例集は、石油各社あるいは

石油連盟が所有している資料、事例、データのほかに、業界内のみでは不足する情報を補うため、専門家により分析、解析あるいは加工されたデータも含まれている。本事例における事故への経緯を説明した文章を基に、事故事象といった項目などを加味して解析を実施した。

## 2.6 結果・考察

タンクという単語を含む事例の進展結果を図 3 に示す。主に「漏洩・噴出」、次に「火災・爆発」へ進展している事が視覚的に理解できる。

タンクという存在自体から、漏洩や火災爆発に繋がる事は容易に想像が付く、しかし、「漏洩・噴出」、「火災・爆発」という最終事象の発生確率が半々の割合である事は、必ずしも想像できるわけではない。「火災・爆発」に至るためには漏洩という事象の後であるため、漏洩の半分(結果的に 1/4)は「火災・爆発」に繋がっている可能性がある。今回の解析の要素はタンクだったが、事例に含まれているか否かだけしか考慮していないため、原因事象がタンクである事を抽出する必要があるのは今後の課題である。

## 2.7 結言

本研究は自由記述で書かれた説明文から自然言語解析を用い、共通要因から事象の進展を解析し、大局的に事象の傾向を探ろうと試みた。形態素解析における頻度が高く、事故に関係が強い形態素に対して事象進展の可視化を実現した。今後は、単語の文法上の接続関係を考慮したモデルの提案に注力していく。

## 謝辞

本研究は JSPS 科研費 16K16367 の助成を受けたものです。本支援に感謝致します。課題名「複数事例の進展統合化グラフによる事業所内の潜在的リスク可視化システムの開発」

## 参考文献

- [1] 箕輪 弘嗣, 宗澤 良臣, 事象の主体と振舞に注目した進展事象の統合解析法, 安全工学, 53(5), 317-324 (2014).
- [2] H. Minowa, "A integrated analyzing method for the progress event based on subjects and predicates in events", *Proc. IEEE/SICE International Symposium on System Integration*, Tokyo, Japan, pp. 678-683 (2014).
- [3] 中野義之, 西尾芳男, PEC-SAFER ヒヤリハット事例データベースの概要, 安全工学, vol. 46, no. 5, pp. 303-310 (2007).
- [4] 箕輪弘嗣, 宗澤良臣 他, 自然言語処理による事故事例の可変カテゴリ分類の有効性, 安全工学, 安全工学, 51(5), pp.319-326 (2012).

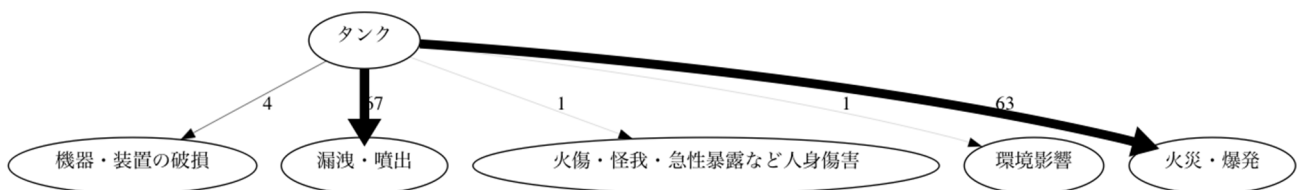


図 3 「タンク」を含む事例の最終事象