

不具合情報からの因果モデル抽出技術の開発

清水 勇喜

(株)日立製作所 研究開発グループ 機械イノベーションセンタ

yuki.shimizu.kf@hitachi.com

1 はじめに

製品の品質・信頼性向上のための手段の一つとして、過去の不具合情報活用による不具合の再発防止がある。製造業では、不具合情報データベース（以下、DB）を構築し不具合情報の共有化を図っているが、十分に活用しきれていないと言われている[1]。その理由として、設計対象の自製品に関する不具合情報を的確に検索できていないことが挙げられる。

関係する不具合情報を探し出すためには、自製品の部品名、及び、不具合の発生要因となるストレス（例：温度、湿度、圧力、応力）を、不具合情報と比較する必要がある。不具合情報には、比較対象の部品名やストレス以外の情報も含まれるため、単純に、不具合情報の文章全体に対してキーワード検索すると、ノイズが多くなってしまう。部品、ストレスを的確に比較できるようにするための手法の一つとして、不具合情報の構造化、すなわち部品 - ストレス - 故障モードからなる因果モデルの抽出がある。故障モードとは、断線、短絡、折損、摩耗のような故障状態の形式による分類である。

因果モデル抽出の従来技術として、手掛かり語を用いる手法がある[2]。例えば「歯車において繰り返し応力のために亀裂に至った」という文章があった場合、手掛かり語の「において」の前に記述されている「歯車」を部品として、「のため」の前に記述されている「繰り返し応力」をストレスとして、「に至った」の前に記述されている「亀裂」を故障モードとして抽出する。この手法では、手掛かり語を用いて記載されている場合は、うまく抽出できるが、「歯車に亀裂が発生した。原因は、繰り返し応力だと考えられる。」のように記載された場合は抽出できず、不具合情報の表現の揺れに影響されやすいという課題がある。

そこで、本研究では、製品の品質・信頼性向上のための不具合情報の活用支援を目的として、表現の揺れに影響されることなく、不具合情報から因果モデルを抽出することを目指す。

2 因果モデル抽出

2.1 概要

本技術では、不具合情報に記載された項目間の関係を利用することで、因果モデルを抽出する。不具合情報は、「不具合名」「不具合発生迄の経過」「現象」「原因」「対策」の項目で記載されることが多い。例えば、科学技術振興機構が公開している失敗知識DB[3]にも同様の項目が含まれる。

不具合情報の各項目に書かれた文章には、下記のような傾向があり、本技術では、これらを用いることで因果モデルを抽出する。

- ・「不具合名」には、それを見て不具合の内容がわかるように、例えば、「電子システムの信号不良」のように「故障部品」「故障モード」が記載されることが多い。
- ・「現象」には、「故障部品」「故障モード」が記載される。「不具合発生迄の経過」には、「現象」が発生するまでの経過が記載されており、「現象」と「不具合発生迄の経過」には、共通して「故障部品」「故障モード」が記載されることが多い。
- ・「原因」には、「故障発生要因部品」「ストレス」が記載される。「対策」には、「原因」に記載された「故障発生要因部品」の「ストレス」への対応が記載されていることから、「原因」と「対策」には、共通して「故障発生要因部品」と「ストレス」が記載されていることが多い。

ここで、「故障部品」とは、故障が発生した部品を指し、「故障発生要因部品」とは「ストレス」によって故障の発生要因となった部品のことを指す。

2.2 因果モデル抽出処理

因果モデル抽出処理の流れを図1に示す。まず、データクリーニングにて、改行・スペース除去や全角・半角統一などを行い、表記を揃える。これは、後で行う単語のマッチングを適切に行うためである。

続いて、形態素解析を行い、単語に分解する。形態素解析ツールは、Mecab[4]を用いる。各項目の単語について、先に述べた傾向を用いてマッチ

ングすることで「故障発生要因部品」「ストレス」「故障部品」「故障モード」の候補となる単語を抽出する。この候補の中から「故障発生要因部品」「故障部品」「ストレス」「故障モード」を選定し因果モデルを抽出する。マッチングによる候補抽出処理を 2.3 節にて、因果モデル選定を 2.4 節にて、その詳細を説明する。

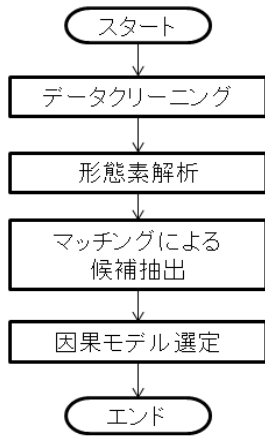


図 1 因果モデル抽出処理の流れ

2.3 マッチングによる候補抽出

マッチングによる候補抽出は、大きく (1)「ストレス」「故障発生要因部品」の候補抽出と、(2)「故障モード」「故障部品候補抽出」の候補抽出で構成される。以下、それぞれについて説明する。

(1)「ストレス」「故障発生要因部品」候補抽出
形態素解析を行うと、文章が単語に分解され、品詞が得られる。「原因」「対策」に書かれた文章を形態素解析して得られた単語に対して、マッチングを行い、両方に共通する語で、かつ品詞が名詞もしくは未知語のものを抽出する。

例えば、表 1 のような原因、対策では、「基板」「ウイスカ」が抽出される。

(2)「故障モード」「故障部品」候補抽出

(1)と同様の手法で、「不具合発生迄の経過」と「現象」に書かれた文章から共通語を抽出する。さらに、「不具合名」の形態素解析結果から、名詞もしくは未知語を抽出する。これらを「故障モード」「故障部品」の候補とする。

表 2 のような「不具合名」「不具合発生迄の経過」「現象」では、「電子システム」「基板」「ウイスカ」「信号不良」が抽出される。ここで、「電子システム」は、形態素解析によって「電子」「システム」に分解されるが、元の文章で単語が連続しており、複数単語の組合せによる複合語であるため、「電子システム」のように結合して 1 つの単語として扱うようにした。

表 1 「ストレス」「故障発生要因部品」の候補抽出例

項目	内容
原因	基板で、ウイスカが発生し信号ピンと短絡した。
対策	ウイスカが発生しないめっき仕様の基板に変更した。
共通語	基板, ウイスカ
ストレスと故障発生要因部品の候補	基板, ウイスカ

表 2 「故障モード」「故障部品」の候補抽出例

項目	内容
不具合名	ウイスカによる、電子システムの信号不良
不具合発生迄の経過	20XX 年：電子システムを設置 20XX 年：電子システムサービス開始 20XX 年：不具合発生 20XX 年：基板を代替品へ交換し、復旧
現象	電子システムに搭載している基板で、ウイスカが発生し、短絡した。このため、電子システムにて信号の不良が発生した。
故障モードと故障部品の候補	電子システム, 基板, ウイスカ, 信号不良

2.4 因果モデル選定

マッチングによって得られた候補には、部品と現象が混在している。ここから因果モデルを抽出するためには、単語がストレス・故障モードなどの現象か、部品かを判別する必要がある。そこで、現象を抽出するための指標 IEPH(Index for Extracting PHenomenon)と、部品を抽出するための指標 IEPA(Index for Extracting PArt)を考案した。

IEPH は、新たに考案した現象らしさを表す指標 PHL(PHENOMENON LIKENESS)に、従来から単語の重要性評価に用いられる TF(Term Frequency)・IDF(Inverse Document Frequency) [5]を掛け合わせた値で、式(1)で求める。

$$IEPH = PHL \cdot TF \cdot IDF = phl \times tf \times idf \dots \dots \dots (1)$$

$$phl = phdf / tdf$$

phdf: 末尾の文字列が「～が発生」「～が生じた」などの表現で記載されている文書数
 tdf: 末尾の文字列を含む全文書数
 tf: 単語の出現回数
 $idf = \log k/df$
 k: DB に蓄積されている文書総数
 df: 単語を含む文書数

例えば、「残留応力」や「高温劣化」などの現象を考えると、末尾の文字列「応力」「劣化」は、「～が発生」「～が生じた」などの表現で記載されることが多い。そこで、PHL は、末尾の文字列がこのような表現で記載されている文書数が多いほど評価値が大きくなるようにした。さらに、多くの文書に記載される一般的な現象も、記載があまりない特殊な現象も同様に評価できるように、その末尾の文字列を含む文書数で割るようにした。

IEPA は、部品名らしさを表す指標 PAL (P Art Likeness) と TF・IDF を掛け合わせた値で、式(2)で求める。考え方は IEPH と同様であるが、部品が「～の故障」「～の購入」などの表現で記載されることが多いことから、末尾の文字列がこのような表現で記載されている文書が多いほど評価値が大きくなるようにした。TF・IDF については式(1)と同様である。

$$IEPA = PAL \cdot TF \cdot IDF = pal \times TF \cdot IDF \dots \dots \dots (2)$$

$$pal = paldf/tdf$$

paldf: 末尾の文字列が「～の故障」「～の購入」などの表現で記載されている文書数
 tdf: 末尾の文字列を含む全文書数

「ストレス」「故障発生要因部品」の候補から、IEPH が最も高いものを「ストレス」として、IEPA が最も高いものを「故障発生要因部品」として選定する。表 1 の例では、「ウイスカ」が、IEPH が高くなり、「ストレス」として「基板」が、IEPA が高くなり、「故障発生要因部品」として抽出される。

同様に、「故障モード」「故障部品」の候補から、IEPH が最も高いものを「故障モード」として、IEPA が最も高いものを「故障部品」として選定する。このとき、既に「ストレス」「故障発生要因部品」として抽出されているものは除外する。

表 2 の例では、「電子システム」が「故障部品」として、「信号不良」が「故障モード」として抽出される。最終的には、因果モデルとして「基板」「ウイスカ」「電子システム」「信号不良」が抽出される。これは、「基板」に「ウイスカ」が発生し、それによって「電子システム」に「信号不良」が

生じたという因果関係を表している。

3 検証

3.1 評価方法

下記にて抽出精度の評価を行う。

(1) 評価対象

不具合情報には、物理的な直接原因と、人間の心理的作用や活動を対象とした動機的原因の 2 種類の原因が記載されることが多い。本評価では、物理的な因果関係を対象とするため、前者の直接原因を記載した不具合情報 90 件を対象とする。

(2) 抽出手法

提案手法と、手掛かり語を用いた従来手法について評価する。従来手法では、表 3 に示す手掛かり語を用いる。

表 3 従来手法で用いる手掛かり語

因果モデル	手掛かり語
部品	において、にて、に対して
ストレス	ために、環境で、環境下で、により、した結果、によって、したことで、することで、に伴って、による、が原因
故障モード	に至り、に至った

(3) 評価方法

抽出された「故障発生要因部品」「ストレス」「故障部品」「故障モード」に対し、ノイズの少なさ P と抽出漏れの少なさ R について評価する。それぞれ式(3) (4) によって求める。

$$P = \text{適切に抽出された件数} / \text{抽出された総件数} \times 100 \dots \dots \dots (3)$$

$$R = \text{適切に抽出された件数} / \text{抽出元の不具合情報総件数} \times 100 \dots \dots \dots (4)$$

3.2 評価結果

(1) 提案手法評価結果

抽出した因果モデルの一例を表 4 に示す。1 つ目の例は、ねじ部にめっきが付着したことで車両にて異音が発生したことを表している。2 つ目の例では、軸受潤滑オイル穴に巣が生じ、エンドブラケットから軸受潤滑オイルが漏れ出したことを表している。

表 4 因果モデルの抽出結果例

故障発生要因部品	ストレス	故障部品	故障モード
ねじ部	めっき 付着	車両	異音 不具合
軸受潤滑 オイル穴	巣	エンド ブラケット	軸受潤滑 オイル漏れ

評価結果を、図 2、図 3 に示す。提案手法 A が 2 章にて述べた手法の評価結果である。従来手法では、故障発生要因部品と故障部品の判別はできないため、故障発生要因部品として一纏めとしている。

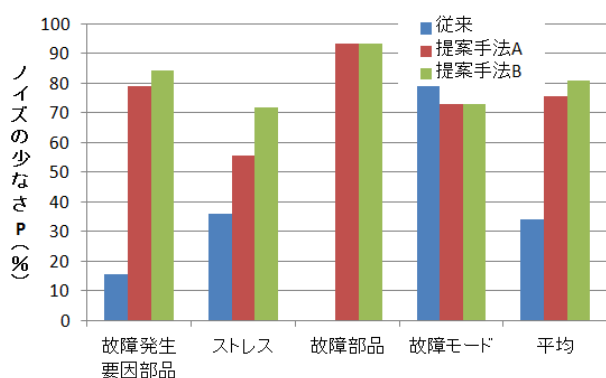


図 2 評価結果 (ノイズの少なさ P)

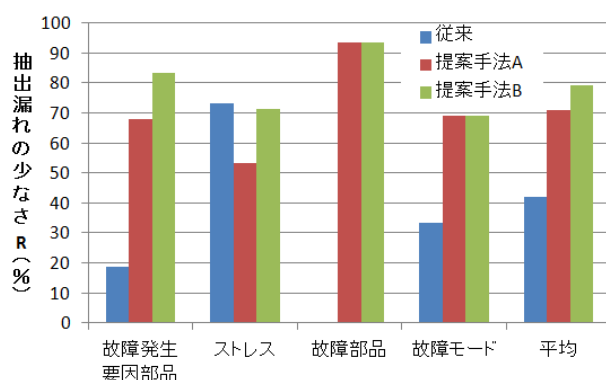


図 3 評価結果 (抽出漏れの少なさ R)

従来手法では、ノイズの少なさ P の平均が 34%、抽出漏れの少なさ R の平均が 42% となった。これに対して、提案手法 A では、それぞれ 75%、71% となった。従来手法は、手掛かり語を増やしたり、別の表現としたり調整することで、抽出漏れを少なくできる可能性もあるが、不具合情報の表現によっては、抽出漏れが減らなかったり、逆にノイズが増えたりする可能性がある。これに対して、提案手法 A では、このような調整は必要なく、表現の揺れに強い手法といえると考えられる。

また、ストレスの抽出漏れの少なさ R をみると、従来手法 73% に対し、提案手法 A では 56% となった。提案手法 A の方が低くなったのは、ストレスの発生要因となった部品を、不具合のない別の部品に取り換える、もしくは無くすという対策がとられるケースがあるためである。この場合は、対策にストレスの記載がないことが多く、提案手法 A では抽出できない場合がある。

そこで、ストレスの抽出精度を上げるために、従来手法の考え方を組合せ、ストレスの手掛かり表現を含む文章を原因より検索し、文頭から手掛かり表現までの名詞及び未知語を抽出し、「ストレス」「故障発生要因部品」の候補として追加するようにした。これが提案手法 B である。検証の結果、72% となり、当初に比べて 16% 精度を上げることができた。平均値でもノイズの少なさ P が 81%、抽出漏れの少なさ R が 80% と精度を向上できた。

以上により、提案手法によって、表現の揺れに影響されることなく、ノイズ・抽出漏れ双方について精度良く因果モデルを抽出できることを確認できた。

4. まとめ

製品の品質・信頼性向上のための不具合情報の活用支援を目的として、不具合情報の項目間の関係を利用することで、部品 - ストレス - 故障モードからなる因果モデルを不具合情報から自動抽出する技術を開発した。90 件の不具合情報を用いた検証の結果では、ノイズの少なさ、抽出漏れの少なさ双方について、80% 程度の精度で抽出でき、手掛かり語による抽出に比べ抽出精度を向上できることを確認した。

参考文献

- [1] 畑村, 中尾, 飯野, 失敗知識データベース構築の試み, 言語処理, Vol14, no. 7, pp. 733-739, July 2003
- [2] 坂地, 竹内, 関根, 増山, 構文パターンを用いた因果関係の抽出, 言語処理学会年次大会発表論文集, Vol14, pp. 1144-1147, 2008
- [3] <http://www.sozogaku.com/fkd/index.html>
- [4] <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>
- [5] Salton, G. and McGill, M. J, Introduction to Modern Information Retrieval, McGraw-Hill (1983)