

# 不具合事例文書からの製品・部品に関する因果関係抽出手法の検討

大森信行<sup>†</sup>

<sup>†</sup>横浜国立大学大学院環境情報学院  
長野県工業技術総合センター

E-mail: {ohmori, mori}@forest.eis.ynu.ac.jp

森辰則<sup>‡</sup>

<sup>‡</sup>横浜国立大学大学院環境情報研究院

## 1 はじめに

製造業における製品開発においては、信頼性確保は常に重要な課題であるが、製品の不具合やリコールの状況に代表されるように、十分な信頼性が達成できた状況とは言えない。

我々は、蓄積されている不具合事例や事故事例等の故障情報に着目し、これらの不具合事例等を効率的に活用することで、信頼性の高い製品開発を支援する手法を検討してきた[1]。

製品開発に不具合事例等を活用するには、不具合事例等から開発中の製品に関連する情報を抽出して提示する必要がある[2]。このための一つの手法として、本稿では、不具合事例文に対して、因果関係抽出を適用し、不具合事例文から製品や部品に生じた不具合の原因に関する記述を取り出すことを検討する。ある文が製品や部品のトラブルに関する因果関係を表すかどうかの判断のために、文中の語の出現パターンに基づき学習した 2 値分類器を用いる手法を提案し、実験により抽出性能を評価する。また、抽出性能向上のために、不具合の表現においては製品や部品の挙動の記述が中心的内容を表すことから、不具合に関与する部品や製品を示す語の出現パターンに基づく判断手法を検討する。

## 2 背景とアプローチ

### 2.1 不具合事例文の記述内容と構成要素

不具合事例文書は、情報共有を主な目的に不具合に関する情報を記載して作成された文書である。事例名称、不具合の事象内容、原因、影響、再発防止策といった項目について、項目名とともに内容が記述されている。このような不具合事例文書に現れる文を本項では不具合事例文という。不具合事例文は、不具合事例文書を構成する文であり、ある製品や部品(以下、製品等という)の実体物やそれらの実体物間の関係が、原因事象により変化する状況を表現する。不具合の発生した製品等の他に、原因となる事象を引き起こした製品等や、不具合が影響を与えた製品等も含まれる。これらの製品等について要求される機能や役割等の属性の記述、および原因となる事象によってこれらの機能や役割が失われる状況についての記述が含まれる。

図 1 は、本研究で抽出しようとする製品等に関する不具合の因果関係の例である。製品である「パイプ」および「ボルト」について発生した不具合に関する因果関係を表している。

### 2.2 製品開発における不具合事例の利用

信頼性等の向上のため開発中の製品に関連する不具合情報を活用する試みは広く行われている [2, 3, 4]。

製造業における製品開発では、設計者が必要としているのは、不具合事例の全般的な情報ではなく、開発中の製品に関連する不具合情報である[2]ので、不具合事例の活用のためには、不具合事例から必要な情報を取り出す必要がある。

開発する製品の信頼性向上のために不具合事例を利用する際には、製品に関する不具合の因果関係を中心に参照することになる。つまり、以下の 3 つの記述を含む文を抽出する必要がある。

- ・製品等に関する記述
- ・不具合に関する記述
- ・因果関係に関する記述

なお、不具合事例文に含まれるが抽出対象とならない文の例としては、因果関係を記述しているが製品等に関与しない記述、例えば「A 社は業績が悪化し、B 事業を廃止した」、製品に関する記述であるが不具合でない記述、例えば「部品 C の動作が正常に停止した」である。

次節以降で、これらの記述を抽出に関する研究について述べる。

### 2.3 製品に関する不具合の因果関係抽出に関する研究

#### (1) 因果関係抽出に関する研究

不具合に限らず因果関係一般の記述を抽出する研究はこれまでも行われており、因果関係を構成する原因、結果を精度よく抽出する手法が報告されている[6, 7]。乾らは、「ため」を含む文を対象に cause 関係等、因果知識の分類、抽出を行っている。

坂地ら[7]は、新聞記事を対象として、文中の手がかり表現に着目し、文中の助詞のペアや、オントロジーに基づく語の意味等に基づいて因果関係を抽出している。

本研究では、他種類の手がかり表現に対応できること、抽出対象が名詞句や動詞句ではなく、文全体を対象としていることから、不具合事例文の因果関係抽出に坂地らの手法を採用する。その際に、本節で述べる実体語の抽出の観点から改良を試みる。なお、本研究においては、手がかり表現はその表現の出現する文が因果関係を示す可能性が高いと考えられるものであり、人手によりあらかじめ与えるものである。

#### (2) 製品や部品を示す語の抽出

不具合事例文では、不具合の起こった製品等は不具合事例に纏わる事態に登場するモノとして記述されると考え

抽出する文：  
パイプが揺れたために、ボルトが折れた。

図 1 不具合事例を構成する因果関係と実体物の例

られる。そこで、製品に関する記述を抽出するために、モノを示す語に関する素性を利用することを検討する。

新聞記事やWebから製品情報を抽出する研究は従来より行われているが、我々は不具合事例文等から製品やその部品に関する情報、すなわち実体語を抽出する研究[8]を行っている。

実体語抽出では、不具合事例文における不具合事例に纏わる事態の記述に登場する「モノ(物体)」と「コト(事態)」を示す語のうち、モノを示す実体語を抽出する。

本研究においては、製品等に関する不具合の因果関係を抽出するために、実体語抽出に関する手法を利用し、実体語や、実体語が出現する構文的なパターンに関する素性等を利用する。

### (3)不具合の記述の抽出

不具合の抽出に関して、丹治ら[9]により、トラブルを表す文を抽出する研究が行われている。丹治らの研究におけるトラブルは、人の行動に関する表現も広く扱っており、「パソコンの設定方法が分からなかった」といった製品等は正常に機能している状態を表す場合であってもトラブルと判断される。不具合として製品が正常に機能しない状況を中心とする本研究とは目的が異なっており、抽出しようとする対象が異なるため、丹治らの方法を本研究に適用することはできない。

## 3 因果関係の抽出手法

本章では、本研究におけるタスクを定義し、機械学習による実現方法、および、そこで利用する素性について述べる。

### 3.1 タスクの定義と実現方法

本研究における製品等に関する不具合の因果関係抽出に関するタスクを以下のように定義する。

- ・判断対象の文が、その出現する文書番号および文書内の項目名と共に与えられる。
- ・与えられた文が製品等に関する不具合の因果関係を表現する文であるか、表現しない文であるかを判断する2値分類を行う。

本研究では、このタスクの2値分類を機械学習により実現する。具体的には、文脈から素性を抽出してサポートベクターマシンにより分類を実施する。次節以降で、抽出する素性等について述べる。

### 3.2 因果関係抽出に関する素性の検討

因果関係抽出のために2.3節で述べた坂地ら[7]が利用した素性を利用する。素性の抽出例を図2に示す。

(1)助詞のペア：構文的な素性として助詞ペアを以下の手順で抽出する。手がかり表現を含む文節(核文節)が係り先となる文節の助詞を前部助詞リストに追加する。手がかり表現の係り先文節(基点文節)が係り先となる文節のうち、

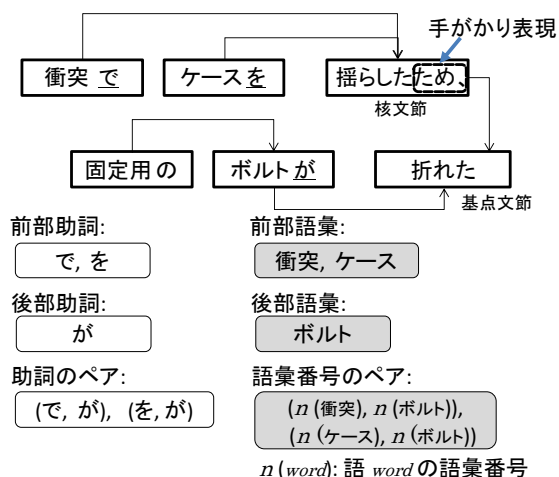


図 2 素性の抽出例

助詞を後部助詞リストに追加する。核文節および基点文節が係り先となる文節に助詞が存在しない場合は、対象となる助詞が存在しないことを示す記号 null を、それぞれ前部助詞リスト、後部助詞リストに追加する。前部助詞リストと、後部助詞リストの助詞の全ての組み合わせ(助詞のペア)を素性とする。今回利用した手がかり表現は4.1節で述べる。

(2)語の語彙番号のペア：前部助詞リストおよび後部助詞リストの助詞を取り出した文節にオントロジーに含まれる語があれば、オントロジー中の語番号を、それぞれ前部語彙、後部語彙として取り出し、前部語彙、後部語彙の各組み合わせを素性とする。前部語彙、後部語彙が、それぞれ取り出せなかったときは null を追加する。オントロジーとしては、坂地らは、小林らの拡張言語オントロジー[10]の6階層目の意味カテゴリを用いたが、本研究では分類語彙表[11]の語と一致した見出しの分類番号の上位4桁を利用した。

また、本研究では、概念の階層構造を素性として利用する研究[8]の手法を利用している。すなわち、分類語彙表の分類番号が上位の桁ほど、より上位の概念に対応している関係を表現するため、1つの語の分類番号を先頭からi桁分の数字列(i=1~4)で表現される4つの数値に分割して表現する。

語彙番号のペアとしては、同一の桁数の語彙番号のペアのみを取り出す。例えば、1234, 5678 という4桁の語彙番号からは、(1,5), (12, 56), (123, 567), (1234, 5678) の4つの語彙番号ペアを素性とする。

### (3)手がかり表現の直前形態素の品詞：

手がかり表現の直前の語により、その文が因果関係を表す確率が異なることから、手がかり表現の直前形態素の品詞を素性とする。図2の例では、「ため、」の直前の「た」の品詞である助動詞となる。

(4)ユニグラム・バイグラム：文に出現する形態素のユニグラムとバイグラムを抽出する。

### 3.3 製品等の抽出に関する素性の検討

与えられた文書に含まれる語がモノを示す語であるかどうか、つまり実体語であるかどうかを判別する手法[8]に基

を背景に ために により、	を背景に、 に伴い、 によって	を受け を反映して を受けて	ため、 で、 に伴う
---------------------	-----------------------	----------------------	------------------

図 3 利用した手がかり表現の例

づいて、文中のモノの状態や挙動の表現に関する素性を抽出する。本研究では、実体語を判別する手法により、対象文書に現れる実体語をあらかじめ抽出する。素性を抽出する際に、文に出現した語のうちあらかじめ抽出した実体語に一致する語について、当該語を実体語であると判断する。これらの素性は、前節の坂地ら[7]の因果関係抽出手法を実体語に関する記述に拡張した素性である。

(1) 実体語と共起する助詞のペア：前節の助詞のペアと同様に素性を抽出する。ただし、助詞と同一文節に実体語が存在しない場合には、当該助詞を含むペアは取り出さない。

(2) 実体語の語彙番号のペア：前節(2)の語の語彙番号のペアと同様に素性を抽出する。ただし、語が実体語でない場合には、当該語の語番号は前部語彙、後部語彙へ追加しない。

(3) 実体語のペア：(1)において、前部助詞リストおよび後部助詞リストの助詞と同一文節に出現する実体語を、それぞれ前部実体語リスト、後部実体語リストに追加する。前部および後部実体語リストの全ての組み合わせ(実体語のペア)を素性として取り出す。

(4) 実体語のユニグラム：文中に出現する実体語のユニグラムを素性とする。

(5) 実体語と助詞の組：同一文節内に出現する実体語と助詞の組を素性として取り出す。図 2 の例では(ケース、を)、(ボルト、が)となる。

(6) 実体語の語彙番号と助詞の組：上記(5)において、実体語を語彙番号に置き換え、語彙番号と助詞の組を素性とする。前節(2)のとおり、語彙番号は、1 語について 4 つの素性で表現するため、ここでも上記(5)の 1 組から 4 つの語彙番号と助詞の組を素性として取り出す。

### 3.4 不具合の抽出に関する素性

不具合事例文から抽出しようとする不具合は、過去に発生した事象であるため、文中には過去形で記述される。これに対して製品等の正常な動作や機能は、日本語においては非過去形で記述される。このことから、文の時制(非過去形または過去形)により不具合を表しているかどうかを判断できると考え、文の時制を表す文末の語とその品詞を素性として用いる。

## 4 実験

### 4.1 実験設定

提案手法に基づく因果関係の抽出実験には SVM<sup>light</sup> を用い、3 節の素性を利用して製品の不具合に関する因果関係を表す文・表さない文をそれぞれ正例・負例として学習した。SVM<sup>light</sup> は二次の多項式カーネルを用い、その他のパラメータはデフォルトとした。学習事例ならびに訓練

表 1 利用する素性と実験結果

	素性	再現率	適合率	F値
ベースライン1	ユニグラム・バイグラム	<b>76.50%</b>	76.36%	76.31%
ベースライン2	先行研究の方法	73.50%	87.72%	79.78%
ベースライン1 に素性を追加	(1) 実体語と助詞の組	69.00%	86.86%	76.71%
	(2) 実体語の意味番号と助詞の組	69.50%	86.62%	76.89%
	(3) 実体語ペア	71.50%	86.00%	77.85%
	(4) 実体語の分類語彙表番号のペア	72.50%	84.82%	77.91%
	(5) 実体語と共起する助詞のペア	71.00%	87.06%	77.97%
	(6) 実体語ユニグラム	68.50%	87.32%	76.57%
	(7) 文末の語	70.50%	87.22%	77.65%
ベースライン2 に素性を追加	(1) 実体語と助詞の組	74.00%	87.80%	<b>80.11%</b>
	(2) 実体語の意味番号と助詞の組	74.00%	87.32%	79.89%
	(3) 実体語ペア	71.50%	87.04%	78.25%
	(4) 実体語の分類語彙表番号のペア	74.00%	84.92%	78.82%
	(5) 実体語と共起する助詞のペア	72.50%	87.06%	78.98%
	(6) 実体語ユニグラム	69.50%	<b>88.86%</b>	77.72%
	(7) 文末の語	71.00%	<b>88.96%</b>	78.43%

事例は WWW 等で公開されている不具合事例文[3]のうち、2.5 節の定義に一致する 400 文(異なり数は 321 文)ならびに語の出現する文書情報であり、正例および負例とも 200 語ずつである。本研究で抽出しようとしている製品等に関する不具合の因果関係を表す文は正例、それ以外の語は負例とした。この 400 文について、以下に示す素性を利用する条件で、5 分割交差検定を実施した。交差検定の各検定において、評価事例の文と同一文書から取り出した文が訓練事例に含まれる場合は訓練事例から当該文をとりぞき、さらに、評価事例と同一の文が訓練事例に含まれた場合も、訓練事例から当該文を除いた上で、評価を行った。

実験に用いた素性は表 1 の素性欄に示す。(1)から(7)が、今回新たに提案した素性であり、(1)から(6)はそれぞれ 3.3 節の(1)から(6)の素性、(7)は 3.4 節の文末の語に関する素性である。ユニグラムとバイグラムに関する素性を利用した条件をベースライン 1 とし、坂地らの方法である先行研究の方法をベースライン 2 とした。係り受け解析には、Cabocha 0.53 を利用した。形態素解析結果で、名詞および未知語が連続したものは 1 語の複合語として扱った。手がかり表現は、図 3 の例を含む合計 34 個を利用した。ある語が実体語であるかどうかの判断は、[8]の方法に従い以下の手順で実施した。

- (1) 不具合事例文[3]中の語について、正例負例各 500 語、合計 1000 語を訓練事例として作成する。異なり数は 700 語であり、このうち 213 語が 400 文中に出現する。
- (2) 不具合事例文から取り出した語(合計約 14 万語)について実体語であるかどうかの判断を行う。
- (3) 14 万語のうち複数回出現する語については、多数決により最終的な実体語であるかどうかを判断する。
- (4) この結果、異なり数 12643 語を実体語として抽出したこの 12643 語に一致した語は、本実験において実体語であると判断した。なお、実体語の抽出を完全に正しく実行できれば、実体語を含む文の集合は本研究で抽出しようとする文を完全に包含することになる。よって、実体語を含む文をあらかじめ抽出したうえで、その文が不具合に関する因果関係を含むかどうかを 2 値分類すればよくなり、問題設定が本研究と異なる。しかし、完全に正しい実体語抽出は実現できていないため、本研究においては素性の

一つとして扱うこととした。

## 4. 2 実験結果

実験結果について表1に再現率と適合率、F値を示す。これは交差検定における各検定の平均値である。性能指標毎に、最良の性能を示す素性の組み合わせを調べたところ、適合率の最大は「ベースライン 2(坂地らの方法)+文末の語」、再現率の最大は、「ベースライン 1(ユニグラム・バイグラム)」、F値の最大は「ベースライン 2 + (1)実体語と助詞の組」を利用する場合であった。性能指標毎に1位の数値を太字で、2位の数値には下線を追加して表現した。

ベースライン 1 との比較では、新たに提案した実体語に関する素性を追加することで適合率、F値が向上した。

ベースライン 2 との比較では、「ベースライン 2 + (1)実体語と助詞の組」では適合率・再現率・F値が向上し、F値は全条件の中で最大となった。「ベースライン 2 + (7)文末の語」では、適合率が向上し全条件の中で最大となった。

## 5 考察

ベースライン 1 に今回提案した素性(1)から(6)を追加すると、適合率とF値が向上するものの、再現率が低下した。ベースライン 2 に今回提案した素性を追加すると、(1)および(2)ではF値が改善したものの、F値が低下する素性もあった。F値の低下の原因調査のため(3)の素性について調査した。(3)の素性では、対象とした400文のうち半数以上で、3.3節の前部および後部実体語リストともに追加する語がないことを示す(null, null)という素性が出現した。この素性の正例と負例の出現文数の比は7:3程度だった。3.2節の(1)の助詞ペアに関する素性では、同様の(null, null)という素性の出現する文は、20文程度であり、ほとんどの文で前部および後部助詞リストに追加する助詞が抽出できているので、実体語という制約が(null, null)という素性が多くなった原因と考えられる。

また、(3)の素性のうち(null, null)以外では、9割以上の素性で出現した文の数が1文のみであり、2文から3文のものが1割程度であるので、複数文に共通する統語的パターンを表す素性はほとんどない。以上から、取り出す実体語がないことを示す素性(null, null)が、多くの正例および負例として出現したことが性能低下の原因と考えられる。今回は、素性として扱う語の範囲を核文節と基点文節を係り先とする文節としたが、坂地ら[7]が提案するより広い範囲を対象とすることで、範囲内に実体語が見つかり、この影響を削減できる可能性がある。

再現率が低い理由について調査するため、ベースライン 2 に(1)実体語と助詞の組を追加した場合で、抽出しようとしている製品の不具合に関する因果関係が含まれている文について、判別が正しかった文と誤った文に分けて、ある素性がいくつの文に出現したかを調査した。上位にはユニグラム・バイグラムの他に、手がかり表現の前の語の品詞に関する素性が多く出現しており、名詞一般と名詞-サ変接続がいずれも上位20位以内に出現していた。手

がかり表現周辺の情報としてはこの素性だけしか抽出していないために、高い適合率で判別するためには情報が不足していた可能性があるため、詳細な調査と新たな素性の検討が課題である。

## 6 まとめ

不具合事例文からの製品開発に利用できる因果関係を抽出するために、製品等に関する不具合の因果関係を取り出す手法を提案し、実験により実体語に関する情報を用いると性能が向上することを確認した。今後の課題として、今回利用した手がかり表現を含まない文における判別性能を向上させるための新たな素性の検討や、文が不具合・製品等・因果関係を含むかどうかにより、判別性能へどのような影響があるかの調査等が考えられる。

## 参考文献

- [1] N.Ohmori and T.Mori. Novel Approach for Test Methods Automatic Selection in Product Reliability. International Conference on Machine Learning and Applications, pp.834-839, 2010.
- [2] 畑村洋太郎, 中尾政之, 飯野謙次. 失敗知識データベース構築の試み. 情報処理学会誌, 44(7), pp.733-739, 2003
- [3] 独立行政法人 科学技術振興機構. 失敗知識データベース, 畑村創造工学研究所 (<http://www.sozogaku.com/hatamura/> 2012.1 閲覧).
- [4] 独立行政法人 製品評価技術基盤機構. 平成19年度事故情報収集制度報告書. 2008.
- [5] 稗方和夫, 大和裕幸, 辻本翔. オントロジーを用いた製造現場の不具合情報検索手法に関する研究. 人工知能学会 知識・技術・技能の伝承支援研究会 SIG-KST-2008-09, 2008
- [6] 乾 孝司, 乾 健太郎, 松本 裕治. 接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得. 情報処理学会論文誌 45(3), pp.919-933, 2004
- [7] 坂地泰紀, 関根聡, 増山繁, 酒井浩之. 聞記事中の文が因果関係を含むか否かの判定. 電子情報通信学会技術研究報告 110(142), pp.47-50, 2010
- [8] 大森 信行, 森 辰則. 不具合事例文からの製品・部品を示す語の抽出. 電子情報通信学会論文誌 D J95-D(3), 2012
- [9] 丹治 広樹, 村田 真樹, 柿澤 康範ら. トラブルを表す文の Web からの抽出. 言語処理学会第15回年次大会 発表論文集, 2009
- [10] 小林 暁雄, 増山 繁, 関根 聡. Wikipedia と汎用シソーラスを用いた汎用オントロジー構築手法, 電子情報通信学会論文誌 D, J93-D(12), pp.2597-2609, 2010.
- [11] 国立国語研究所. 分類語彙表, 大日本図書株式会社, 2004