

## 不具合事例からの因果関係抽出に向けた実体性のある名詞の抽出手法の検討

大森信行<sup>†</sup>森辰則<sup>‡</sup><sup>†</sup>横浜国立大学大学院環境情報学府  
長野県工業技術総合センター<sup>‡</sup>横浜国立大学大学院環境情報研究院

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

## 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、自動車のリコール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、不具合事例から、不具合に関与する製品や部品の記述を特定・抽出する手法について述べる。製品や部品は実体性を持つことに着目し、ある名詞が抽出対象の部品であるかどうかを実体性の有無により判断することを検討する。文書に現れる名詞が実体物を表すかどうかの判断のために、名詞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。

## 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

不具合事例は、ある製品や部品等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した文書である。実体物としては不具合の発生した実体物の他に、原因事象を引き起こした実体物や結果として影響を与えた実体物等も含まれる。これらの各実体物について実体単体および複数の実体物群に要求される機能や役割等の属性についての記述、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が記述される。不具合事例文は、これらを組み合わせて構成されている。

図1は、本研究で抽出しようするのは不具合の原因事象の実体物である「パイプ」ならびに原因事象の結果として不具合の生じた実体物である「ボルト」である。

## 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, 5]。

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、不具合事例の活用するためには、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、記載された因果関係を中心に参照することになる。不具合における因果関係の例として不具合発生前後の時間的な遷移過程を表すものと、不具合発生後の状態で下位機能の不具合が上位機能の不具合を起こした状態を表すもの(機能階層関係)の等がある[6]。また、因果関係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[7, 8]。乾ら[7]によれば、上記の2種類の因果関係は、事態間の原因を示す cause 関係であると考えられる。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。不具合事例文の因果関係抽出にもこれらの手法が有効と考えられる。

不具合事例から開発中の製品に関連する情報を取り出すために、不具合事例における原因や結果の記述から、製品や部品を表す語を抽出する。この抽出において中心的手続きとなる実体物の抽出について、次節以降で述べる。本稿における実体物の定義は2.5節で述べる。

## 2.3 不具合事例に纏わる「モノ」を示す語の抽出に関連する研究

2.1節で述べた不具合の原因や結果は、製品、部品に関する事態の記述である。不具合事例に纏わる事態の記述には「モノ(物体)」と「コト(事態)」を示す語が登場するので、開発中等の製品に関する不具合事例を検索するには、このうちのモノを抽出した上で、それが検索したい製品と一致するかどうかを照合する必要がある。

固有表現抽出では IREX[9]をはじめとして、固有物名 (ARRIFACT) に属する表現の抽出を行っている。IREX の定義によれば、固有物名は、「人間の活動によって作られた具体物、抽象物を含む物の固有物の名前」であるので、本研究で目指す実体物を含む。しかし、曲名等の作品名、法律名といった抽象物の名前も同時に含むため、固有物名に属する表現ということのみでは、製品や部品を示す語かどうかの判断としては不十分である。関根らの拡張固有表現[10]においても同様に、カテゴリに実体性を持つ語と、持たない語の両方が含まれることがあり、あるカテゴリに属するかどうかで、実体性を判断できるとは限らない。

不具合事例文から、不具合に纏わるモノとしての抽出を目指した研究として、稗方ら[5]は船舶に関する不具合事

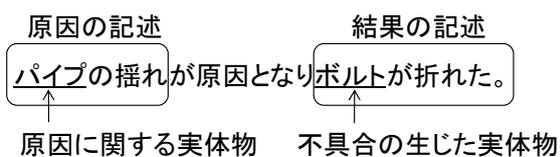


図 1 不具合事例を構成する因果関係と実体物の例

例から注目語の文節内の統語パターンの利用により不具合の生じた部品や不具合状態等を抽出しているが、不具合の生じた部品の抽出の適合率が 20%程度であり、精度向上が必要と報告している。

## 2.4 不具合事例からの製品や部品を示す語の抽出

本節では、不具合事例に纏わる事態に登場するモノと、そのうち本研究で抽出しようとするものについて述べる。

本研究において想定する電気製品や機械製品等の製品に関しては、不具合の起こった箇所や部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られる。

一方、不具合事例においては、不具合の生じた製品や部品の他に、オペレータ(操作者)や、床面(設置箇所)、荷物(積荷)等も登場する。これらの語は不具合の原因や結果、影響に纏わるモノ(具体物)を示すが、製品や部品には該当しない。開発中の特定の製品についての事例を検索する上では、これらの情報はノイズとなる可能性があることから、抽出しないことが望ましいものである。

したがって不具合事例文において、不具合に纏わる事態に関連するモノから、製品、部品を示す語のみを取り出すことが本研究の目的である。本研究では、抽出の手法として、後述する語の実体性に着目する。文脈の情報から語の実体性を判断する研究や、不具合事例に纏わる事態に登場するモノの抽出を語の実体性に基づいて行う研究は我々の知る限り行われていない。

## 2.5 対象とする製品と実体物の定義

これらの製品や部品の記述を抽出するため、本稿では実体物を「抽象的な概念でなく実在するもので、工業製品やその部品として利用されうるモノ」と定義する。これは、溝口らによる具体物[11]のうち、存在に空間が必要なものに包含される概念である。また、ある語が実体物を示す文脈で利用されるときに、その語は実体性を持つと定義する。名詞のうち、実体物を持ち、製品や部品となり得るものの抽出が本研究の目標である。

考察を行う不具合情報の分野として、電気製品・機械製品を構成する製品・部品を対象とした。これらの分野の製品は、収集した不具合事例文に多く含まれていること、部品や部材を加工したり組み立てるといった製造方法が共通しており、故障の状況の記述に一定の規則性が期待できるためである。

化学製品等についての不具合事例は本稿では対象とし

## ・注目語の前後1文節の形態素列

### 事態性判別の素性

- ・注目語の分類語彙表中[13]での分類項目の上位4 桁
- ・注目語が複合名詞であったとき、各名詞についての分類語彙表中の分類項目の上位4 桁
- ・同一文節中にサ変名詞+サ変名詞がある
- ・後ろにサ変名詞がある
- ・後ろに助詞+サ変名詞が続く
- ・前にサ変名詞がある
- ・後ろに名詞の接尾辞がある
- ・前にサ変名詞、後ろに助詞がある

### 係り受け関係に関する素性

- ・注目語の文節の末尾の格助詞
- ・注目語の係り先文節の動詞
- ・注目語の係り先文節と係り受け関係にある文節の格助詞と格要素の名詞

### 隣接文節に関する素性

- ・注目語の前に出現する文節の助詞
- ・注目語の前に出現する文節の助詞の前の名詞等
- ・注目語の後に出現する文節の助詞
- ・注目語の後に出現する文節の助詞の前の名詞等

表 1 実体性の判別に利用した素性

ない。これは、加工・組立により製造する電気・機械製品に対して、化学製品等は、原材料の化学反応が主な製造工程であることから、不具合の内容も異なることと、液体や気体等の製品の表現は、本稿で想定した実体物とは異なると考えられ、化学製品等を示す語の抽出には実体物の抽出とは異なる手法が適することが予想されるためである。

## 3 実体物の抽出手法

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

本研究における不具合事例文中の名詞の実体性の判断に関するタスクを以下のように定義する。

- ・実体性を判断する語が、その出現する文および文中の出現位置情報と共に与えられる。
- ・与えられた語が実体性を持つ語であるか、持たない語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、抽出する素性等の実装について述べる。

### 3.2 事態性判別に関する素性の検討

名詞の実体性の判断は、名詞が属するクラスを判断する問題といえる。本研究に関連する語のクラス判別の研究として、小町らは名詞の事態性の判断を行っている[12]。事態性とは、名詞が事態を示す用法で使われている状態のことである。コト(事態)を指すかモノ(物体)を指すかという意味的な違いに対応する。本研究で抽出しようとする語

は、不具合に纏わる事態に登場する製品や部品を表す語であり、モノを表す名詞であるので、事態性を持たない名詞である。従って、小町らの事態性判別のための素性を、事態性のある名詞を取り除くために利用することを検討する。図 1 の例では「揺れ」と「原因」という 2 つの名詞のうち、「揺れ」は事態性判別のための素性により、実体物を表さないと判断されることを期待する。

### 3.3 係り受け関係に関する素性

事態性の無い名詞であっても、以下に示す例のように全て実体性があるとは限らない。そこで、実体性のある語に特徴的な統語パターンを素性として利用する。

不具合事例は、製品等の構成要素間の関係や、不具合の原因や結果の状況等、不具合事例に纏わる事態におけるモノの関係を記述したものである。このうち、実体物を表す語は、これらの関係における格に対応した後置詞句の要素として現れると考え、以下を素性として抽出する。ただし、(2)の動詞が「する」の場合はその前の語を抽出する。

- (1) 注目している語を含む文節の格助詞
- (2) 注目している語を含む文節の係り先文節の動詞
- (3) 上記の係り先文節を係り先に持つ文節の格助詞と、格助詞直前の名詞

以下は事態性を持たず、かつ、実体物を示さない名詞の例である。

- ・時間(1 時間)
- ・組織名(A 社)
- ・属性名、性質名(酸化物、危険物、剛性、脆性)

### 3.4 隣接文節に関する素性

並列関係を表わす助詞によって接続されている二つの名詞は同じ性質をもつことが暗示されることから、次を仮定する。

「いま注目している語が別の語と並列関係を表わす助詞により等位接続されて現れており、その別の語が実体物を指し示すのであれば、いま注目している語も実体物である」

この仮定に従い、前後の文節の助詞および助詞の前の名詞を素性として抽出する。この助詞には格助詞のみでなく他の助詞も含む。例えば部品間の接続状態や取り付け方法等を表現する「ケースとパイプをねじで固定した。」において、「ケース」の実体性の判断に後続文節の「パイプ」と「を」という語を利用する。この素性は、係り受け関係は利用せず、注目している語の文節に対して、出現順で前または後に出現する文節の素性を利用する。

小町ら[12]は「A の」という注目語の前の表現の情報を事態性判別の素性としているが、小町らが特定の助詞が出現する前の文節の素性に着目するのに対し、本研究では対象を前後の文節と、全ての種類の助詞に拡張したものに相当する。

選択した素性	精度	再現率	F値
(1) 形態素列	61.8%	72.0%	66.4%
(2) (1) + 事態性判別に基づく素性	72.5%	67.2%	69.7%
(3) (1) + 係り受けの素性	63.3%	79.8%	70.5%
(4) (1) + 隣接文節の素性	62.1%	74.0%	67.4%
(5) (2) + 係り受けの素性	70.8%	78.8%	74.4%
(6) (2) + 隣接文節の素性	73.1%	73.4%	73.0%
(7) 係り受け + 隣接文節の素性	63.4%	78.0%	69.8%
(8) (2) + 係り受け + 隣接文節の素性	70.0%	76.8%	73.1%

表 2 実体性判別の実験結果

利用した隣接文節の素性	精度	再現率	F値
(9) 助詞のみ	70.3%	76.4%	73.0%
(10) 名詞のみ	70.1%	78.6%	74.0%

表 3 隣接文節の素性ごとの判別性能

## 4 実験

### 4.1 実験設定

提案手法に基づく実体性の判別には SVM<sup>light</sup> を用い、3 節の素性を利用して語の文脈における実体性を学習した。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 1000 語ならびに語の出現する文(1 文あたり平均 46.8 語)である。このうち正例および負例は、ともに 500 語ずつであり、異なり数は正例 418 語、負例 414 語である。本研究で抽出しようとしている製品や部品を示す語は正例、それ以外の語は負例とした。この 1000 語について、以下に示す条件で、10 分割交差検定を実施した。

- (1) 形態素列 (ベースライン)
- (2) (1) + 事態性判別に基づく素性
- (3) (1) + 係り受けの素性
- (4) (1) + 隣接文節の素性
- (5) (2) + 係り受けの素性
- (6) (2) + 隣接文節の素性
- (7) 係り受け + 隣接文節の素性
- (8) (2) + 係り受け + 隣接文節の素性

実体性判別の実験に用いた素性は表 1 に示す。(1)形態素列は、注目語の前後 1 文節の形態素列の素性である。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 語の複合語として実体性の判別を行った。

### 4.2 実験結果

実験結果について表 2 に再現率と精度、F 値を示す。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、精度の最大は「(6) (2) + 隣接文節の素性」、再現率の最大は、「(3) (1) + 係り受けの素性」、F 値の最大は「(5) (2) + 係り受けの素性」を利用する場合であった。

ベースラインとの比較では、新たに提案した係り受けと隣接文節の素性を追加した(3)と(4)は、いずれも精度、再

現率、F 値が向上した。事態性判別の素性を利用すると、再現率が低下したが、精度と F 値は向上した。

## 5 考察

事態性に基づく素性を利用した(2)、(5)、(6)、(8)は、ベースラインと比較して、いずれも精度は大きく向上する。一方で(2)の事態性に基づく素性のみでは、再現率が低下することから、今回の事態性に基づく素性を利用することで、実体性を有する語に対して、実体性が無いという誤った判断をしやすいく傾向がある。例として、「フィルター」や「機体」は、今回の不具合事例文では実体物を示す語として利用されていたが、事態性判別の素性を使った(2)では実体物でないと判断され、係り受けの素性と隣接文節の素性も使った(5)、(6)、(8)では正しく実体物と判断されていた。したがって、事態性に基づく素性は、今回利用したような再現率の向上に有効な素性と組み合わせることで、精度・再現率をバランスよく向上できると考えられる。特に係り受けの素性を用いた(5)は、(2)よりも再現率が大きく向上しており、事態性判別の素性よりも広範囲の文脈情報を利用することで、多くの実体物を表す語が抽出できていると考えられる。

今回新たに導入した隣接文節の素性について、名詞と助詞のそれぞれが判別性能に与える影響を調べるために、隣接文節の素性のうち、(9)助詞のみ、(10)名詞のみを利用した際の判別性能を調査した。結果を表 3 に示す。助詞と名詞両方を使用した(8)と比べて、助詞のみの(9)は顕著な性能の変化は認められなかった。名詞のみの(10)については、精度を維持しつつ、再現率と F 値が向上し、係り受け関係の素性を利用した(5)と近い性能を示した。したがって、隣接文節の素性のうち、名詞はノイズを増やすことなく、実体性のある名詞の抽出するために有効な素性であると考えられる。また、実体性判別において隣接文節という狭い範囲の局所的な情報により、係り受け関係の広域な情報と同程度の性能が達成できることが分かった。

今回の実験結果では、不具合事例に纏わる「コト」を示し、かつ、製品や部品を示さないが、誤って実体性があると判別された語があった。全素性を利用した(8)において、「人」、「地盤」、「A 社」等の語である。つまり、現状の判断手法では、これらの語も製品や部品を示す語と判断してしまう。このような、不具合事例に纏わる「コト」を示す語であるが、製品や部品を示さない語を取り除く方法は今後の課題である。

## 6 まとめ

不具合事例文からの製品や部品の抽出のために、実体性の判別に基づく手法を提案し、実験により係り受けや隣接文節の情報を用いて性能向上ができることを確認した。

今後は、より性能向上のために適した素性の検討や、不具合事例に纏わる「コト」を示すが、製品や部品ではない語を取り除く方法等が考えられる。

## 参考文献

- [1] Nobuyuki Ohmori and Tatsunori Mori. Novel Approach for Test Methods Automatic Selection in Product Reliability. International Conference on Machine Learning and Applications, 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 (<http://shippai.jst.go.jp/>).
- [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] 來村徳信, 西原稔人, 植田正彦, 池田満, 小堀聡, 角所収, 溝口理一郎. 故障オントロジー. 人工知能学会誌, 14(5), pp.828-837, 1999
- [7] 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
- [8] 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
- [9] Satoshi Sekine and Hitoshi Isahara. IREX: IR and IE Evaluation project in Japanese. Second International Conference on Language Resources and Evaluation, 2000
- [10] Satoshi Sekine and Chikashi Nobata. Definition, Dictionary and Tagger for Extended Named Entities. Forth International Conference on Language Resources and Evaluation, 2004
- [11] 溝口理一郎. オントロジー工学. オーム社, pp. 156-158, 2005
- [12] 小町守, 飯田龍, 乾健太郎, 松本裕治. 名詞句の語彙統語パターンを用いた事態性名詞の項構造解析. 自然言語処理, 17(1), 2006
- [13] 国立国語研究所: 分類語彙表, 大日本図書株式会社, 2004