

複数視点からのテキストマイニングによる設計品質の向上

Design Quality Improvement by Multi-View-Point Text Mining

テキストマイニングとは、大量のテキスト情報を分析して有用な情報を得る技術である。本論文では、製品開発における設計品質の向上を目的とし、不具合事例データベースを分析した事例について述べる。一般的なテキストマイニングでは、単語どうしの関連性の可視化や属性情報とのクロス分析を通じて分析結果を提供するが、設計品質向上を主目的とした不具合要因分析に適用した結果についての研究例は少ない。そこで、テキストマイニング技術を活用した不具合分析における分析モデルを構築し、当社のテキストマイニングシステム **CoreExplorer** への実装と評価を行った。その結果、現象－原因の関係性分析などにおいて一定の効果が得られることを確認した。

塚原 朋哉	Tsukahara Tomoya
高梨 勝敏	Takanashi Katsutoshi
渡邊 まり子	Watanabe Mariko
佐藤 俊也	Sato Shunya
井上 悠	Inoue Haruka
船木 裕	Funaki Yutaka

1. はじめに

近年、製品のリコールや基準不適合などが社会の重要な問題として認識されている。製品ライフサイクルの短縮や多品種少量生産、製品の高機能化が進む現在、品質の確保と作業の効率化の両立は企業にとって重要な課題であり、設計者への負担がますます増大する傾向にある。熟練技術者が不足する 2007 年問題への対処も社会的課題であり、過去のノウハウ・事例の再活用が重要視されている¹⁾。

設計段階でのチェック漏れや考慮漏れを防止するために、設計基準の確立や不具合事例検索システムを構築している企業は多いが、不具合対応による設計変更の影響範囲調査の負担、検索スキル不足による類似不良の見逃しといった問題はまだ解決されていない。

このような中、(株)日立東日本ソリューションズでは、各企業の品質保証部殿での不具合情報に関するテキストマイニング事例を元に、設計プロセスへフィードバックできる有用な知識を得る技術の研究を進めている。

これまでのテキストマイニング技術^{2) 3)}では、アンケート分析による顧客の声の抽出やコールセンターでのクレームの分類による FAQ 作成など、文書の傾向抽出を中心に研究開発が進められてきた。このため、設計・開発、製造、品質保証のプロセスを対象とする場合に従来

の手法では足りない場合がある。例えば、不具合の現象・原因・対策といった複数の項目を対象として分析する場合である。これまでは文書全体など、単一の項目から特徴的な単語を抽出するのみであるほか、2,3 の項目から、文書中の出現数に基づく小数の属性（顧客名や部品名など）を抽出してその間の同時出現関係を関連線で結ぶのみであり、項目間の因果関係に基づく関連を詳細に分析（原因から現象の絞込みなど）するのには向いていなかった。また、製品名・部品名・不具合現象などの複数の項目の関連を相互に調べたり、質問文と回答文など複数のテキスト情報をそれぞれ連動させて分析することも難しかった。

そこで、不具合情報から有効な知見を得るテキストマイニングソリューションを開発するため、項目間の関連を詳細に分析できる手法の研究を行った。研究が目標とする到達点は、製造業の企業が持つ不具合情報の分析から不具合・クレーム対策の迅速化と設計品質向上に有効な情報を抽出することである。特に、複数の項目の比較を容易にすることと、項目の変更や項目内の注目情報の変更を容易にすることで、ユーザが自由に分析の観点を変えながら不具合情報を分析できることを目的とした。各企業からヒアリングしたニーズを元に不具合情報分析のモデルを導出し、**CoreExplorer** に適用したテキストマイニングシステムの構築を行った。

2. 企業側のニーズ

各企業がどのような不具合情報を保持し、品質向上についてどのような情報を抽出することにニーズがあるかをヒアリングした。

2.1 不具合情報のデータ項目

各企業の保有している不具合情報は一般的に以下のような構造として定義できる。

(1) 製品構成の情報

- ・製品名、機種名
- ・モジュール名、パーツ名
- ・部品名

(2) 不具合の情報

- ・特定の製品や機種で発生した不具合の現象
- ・原因特定作業の経緯
- ・不具合の発生原因であるモジュール、パーツ、部品

2.2 不具合情報の活用目的

同じくヒアリングの結果、蓄積した不具合情報の活用目的は、大きく下記の二点に要約された。

(1) 不具合・クレーム対策の迅速化

日々収集・蓄積される不具合情報を分類・体系化することにより、以下の知見を得て、不具合やクレームの原因追求・対策を迅速化する。

- ・ある製品に特有の不具合傾向の発見
- ・未知の不具合原因の発見

(2) 設計品質の向上

既存製品の不具合の分析により、類似の新製品での不具合を未然に防止する。

- ・特定の部品を共通に使う複数の製品での不具合の予測
- ・既存機種での不具合の分析結果を用いた、新製品開発時のチェックリストの作成

上記ニーズから、製品、パーツ、部品および不具合事象間の因果関係を分析するモデルを定義し、モデルの実装を行った。

3. 分析モデルの検討

3.1 不具合情報分析モデル

3.1.1 入力情報

システムへの入力とは 2.1 節の不具合情報である。出力は 2.2 節のニーズを満たす情報となる。

入力情報は、製品名や不具合情報などが項目に分かれて 1 件ごとと格納されているデータベースのようなデータである (図 1)。

項目名	内容
製品名	製品A
タイトル	製品Aの液晶ディスプレイの不良
状況	液晶ディスプレイが定時的にちらつき、ちらつきが頻入るときもあれば、赤みがかったまましばらく画面が点滅し、ときがある。
原因	液晶ディスプレイのコネクタ部を見直し、コネクタ部を挿しなおした結果正常回復した。コネクタ部の接触不良と推定。
対策	コネクタを新品に交換した。補償範囲内。
発生日	2004年12月29日
発生日データ	20041229

図 1 検索対象データ (ノート PC の不良データ) 例

3.1.2 モデル

3.1.1 節のようなデータを入力として、2.2 節の不具合情報を活用できる不具合分析モデルを定義した (図 2)。

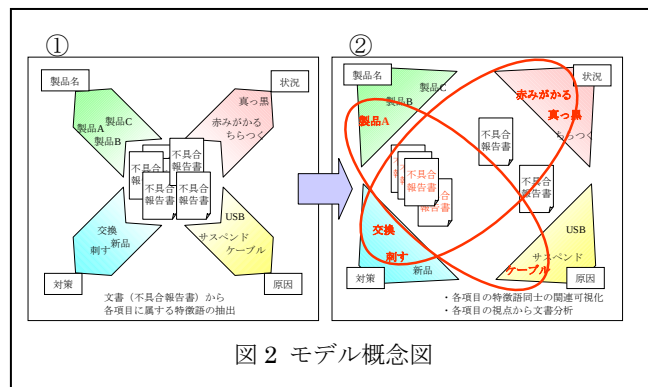


図 2 モデル概念図

- ① 入力情報となるデータベース中の不具合情報の項目ごとに属性 (図の「製品名」項目) や単語 (図の「状況」・「原因」・「対策」項目) を「特徴語」として抽出する。
- ② 抽出した特徴語どうしの関連性を求めるとともに、それぞれの項目の観点から不具合情報を分類する。

3.1.3 出力情報

出力では以下の(1),(2)の特徴を持たせることを目的とした。

(1) 項目間の比較・関連把握の容易性

出力情報は、データベースの複数の項目を比較したり、各項目内の情報の関連がわかるように、相互に関連し合う複数の特徴語を同時に一覧表示できることを目的とした。例えば、ある「製品名」と関連する「原因」の特徴語を見ることで製品特有の不具合を見つけたり、「状況」の特徴語に対応する「原因」「対策」の特徴語を知ること

で不具合の特定を容易にすることを狙う。

(2) ユーザ視点による情報分類の容易性

出力情報の別の表示形式として、ユーザが動的に「製品名」や「原因」などの各項目の上下関係（大分類・小分類など）を指定することで、任意の観点で不具合情報を分類（階層化）できることも目的とした。例えば、「製品名」で大きく分類し、それぞれの製品の「原因」でさらに不具合情報を分類したり、「原因」の観点で大きく分類した後「製品名」や「対策」の観点で分類することも可能とする。これにより不具合情報データベースを任意の集合に分けて分析でき、FAQ の作成や設計時のチェックリスト作りの支援を狙う。

この不具合情報分析モデルを、テキストマイニングシステム CoreExplorer をベースとして実装した。以下 3.2 節で CoreExplorer について簡単に説明し、3.3 節でモデルの実装について述べる。

3.2 CoreExplorer の特徴

CoreExplorer では RDB や XML など複数の項目を持つ情報を入力し、項目ごとに分析を行うことができるシステムである。

特徴は、検索実行時、検索された文書の一覧（文書群）と、検索された文書群に特徴的に出現する単語（特徴語）をあわせて出力することである。出力する特徴語は、特徴語スコアの上位のものから抽出している。特徴語スコアは、tf-idf法^{*1}に基づいた文書中の単語の重要度と、検索結果である文書のスコアの積を、検索文書群に渡って和を求めることで算出している。出力された特徴語はマップ画面に表示され、共起する特徴語どうしは近傍に配置される。出力された特徴語を見ることで文書群にどのような情報が含まれているかを俯瞰することができる（図 3）。

特徴語は検索された文書群中のどの文書と紐づいているのか（どの文書に含まれるのか）の情報を持っているため、特徴語をユーザが選択すると選択された特徴語すべてを含む文書を抽出することができる。

^{*1} ある文書において、単語の重要さを計算する（文書の内容を表すと考える）指標。単語が、注目している文書中で他の語よりも相対的に多く出現するほど(tf)、また、その他の文書にはあまり出現しないほど(idf)、重要と考える。

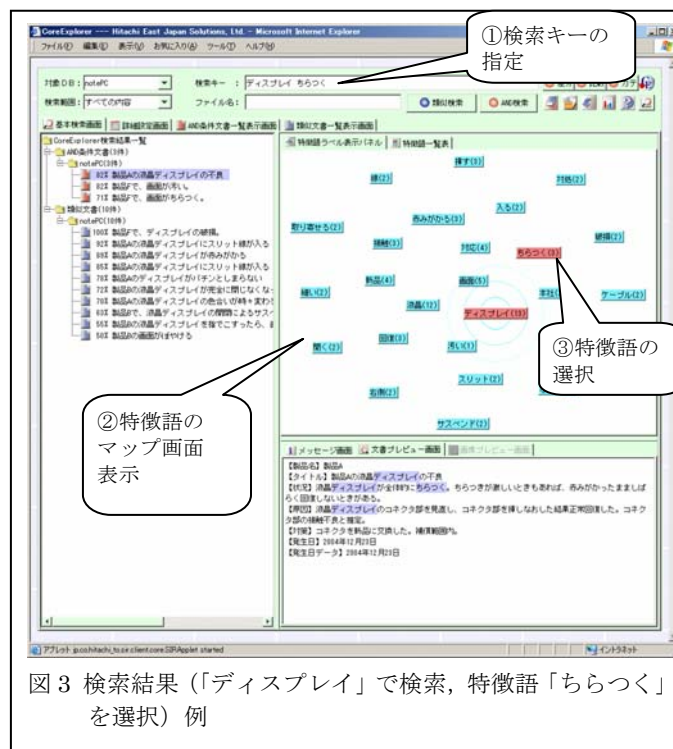


図 3 検索結果（「ディスプレイ」で検索、特徴語「ちらつく」を選択）例

3.3 CoreExplorer への適用

3.1 節不具合情報分析モデルを 3.2 節の CoreExplorer に適用し、不具合情報分析システムを開発した。

3.3.1 分析アルゴリズムの実装

分析モデルの出力情報の実装のため、複数の項目を比較できるように CoreExplorer を拡張した。図 3 に示すように、既存の CoreExplorer では検索対象項目（全文や、状況・原因といったデータベース中の各項目）である 1 つの項目のみから特徴語を抽出している。この処理を変更し、検索対象項目によらずすべての項目から特徴語を抽出するように処理を変更した。CoreExplorer では複数の項目それぞれに対してインデックスを作成（特徴語を集計）しているため、高速にすべての項目から特徴語を抽出することが可能となっている。また、特徴語の概念を拡張し、テキスト中の特徴的な単語のみならず、属性（製品名やパーツ名など）も特徴語として抽出した。

3.3.2 分析用画面の拡張（1）

特徴語の表示画面においては複数の特徴語を比較検討できるように特徴語表示画面（以下特徴語パネル）を複数配置した。1 つの特徴語パネルで特徴語を選択したとき、その特徴語に関連する文書が絞り込まれるだけでなく、選択された特徴語に関連する特徴語がわかるよう

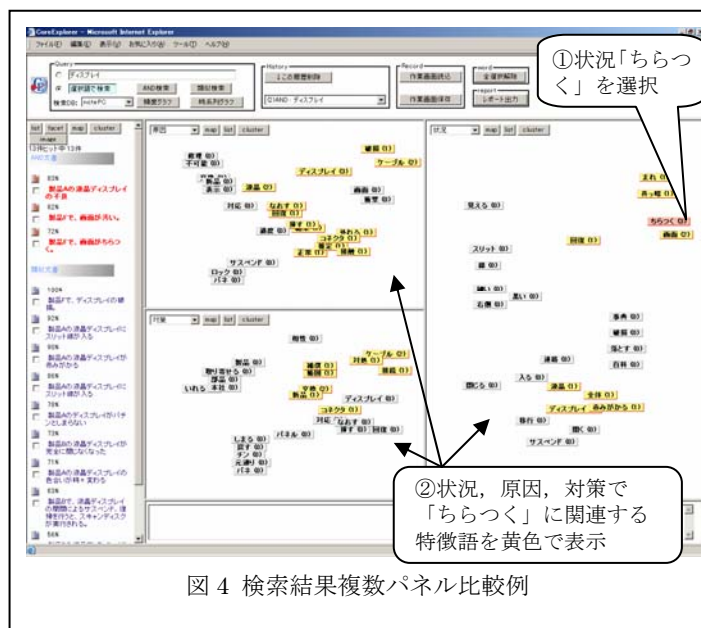


図 4 検索結果複数パネル比較例

に各特徴語パネルの特徴語の色を動的に変更した(図 4)。

特徴語どうしの関連は、絞り込まれた文書から求めている。ある特徴語を選択したときに絞り込まれる文書群があるとき、その文書群に紐づく特徴語は関連があると考えられる。例えば、「状況」項目の特徴語「ちらつく」を選択すると、ディスプレイ関連の文書が絞り込まれるため、「原因」項目の特徴語「ディスプレイ」が関連を表す色で表現される(図 4)。

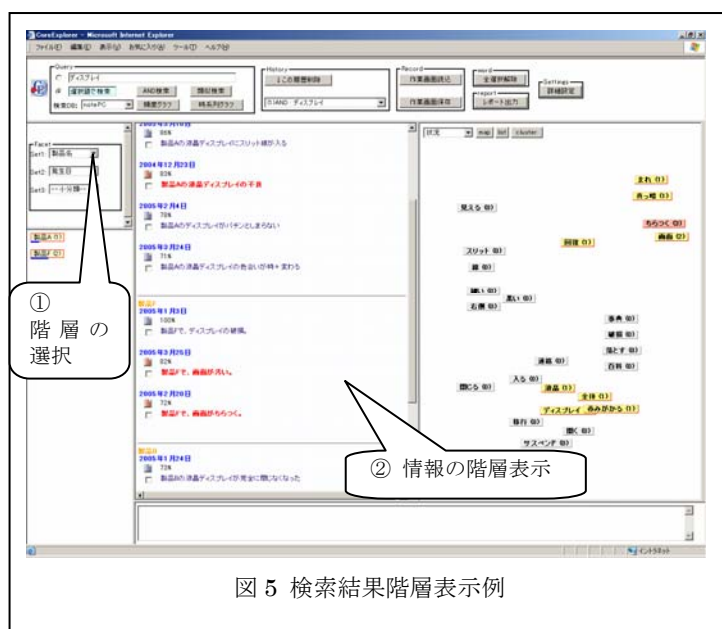


図 5 検索結果階層表示例

3.3.3 分析画面の拡張(2)

不具合分析モデルの階層表示では、複数パネルの比較の GUI と同じ情報を受け取り、項目名を階層に使用し、項目中の特徴語を階層内の情報の振り分けに使用した。

ユーザが動的に大分類・小分類などの階層を選択できるようにして、分析したい視点で情報を分類することを可能とした(図 5)。

大分類として選択した項目があるとき、その項目中の特徴語それぞれに対して、検索結果の文書群のうち特徴語に紐づく文書を分類する。このとき、複数の特徴語に紐づく文書は複数の特徴語それぞれに分類される。さらに中分類があるときは、大分類の項目の 1 つの特徴語に紐づく文書群を、大分類と同様に中分類の項目中の特徴語それぞれに分類する。小分類についても同様である。

4. 評価

4.1 分析モデルと実装システムの機能評価

本報告の不具合分析モデルは 3.1 節の CoreExplorer を用いた不具合・顧客の声などの分析コンサルティングサービスを通して得られた分析要望を元に作成している。以下、各要望を実現するための機能の評価結果について述べる。

(1) 不具合・クレーム対策の迅速化

実装機能を利用して分析を行った例を図 4 に示す。3.2 節であげたノート PC の不良についてのデータに対して検索した結果である。「ディスプレイ」に関する全文を分析したとき、状況・原因・対策の各項目の特徴語を、3 つの特徴語パネルにそれぞれ配置し、状況項目の特徴語「ちらつく」を絞り込み情報に選択した例である。選択した特徴語を赤で表し、選択された特徴語(複数の場合はすべて)に関連する特徴語を黄色で表し、選択された特徴語(複数の場合はいずれか)に関連しない特徴語は灰色で表している。

図から「ちらつく」には同じ状況項目で「真っ黒」や「赤みがかかる」が関連していることや、原因項目の「ケーブル」や「コネクタ」、対策項目の「接続」や「交換」が関連していることがわかる。

上記情報から今後同様の不具合がある場合、現象項目の検索により、原因の推定とそれに対応する対策をすばやく見つけることが期待できる。

(2) 設計品質の向上

分析対象の項目としてパーツ名や部品名の属性を設けることで、ある部品に不具合があるとき、同じ部品を使う他の製品見つけることが可能であり、不具合報告から同じ部品を使う他の製品での不具合予測と事前対策に利用することが期待できる。また、製品を選択した場合、

不具合の起こる部品やその頻度（特徴語右のカッコ内の数字）を知ることができ、予防保守や次の設計に生かす情報として利用することが期待される。

例えば、図 5 検索結果階層表示の例からは、「製品名」項目で分類し、さらに「不具合発生日」項目で分類された検索結果文書群（画面中央のパネル）に対して、「状況」項目の特徴語「ちらつく」を選択した例で、紐づく文書が赤字で表示されている。ユーザの考える分類規則で文書を分類できるため、全体を俯瞰することが容易になる。

以上の結果から、今回検討した分析モデル、及び実装したシステムが一定の有効性を持つことが確認できた。

5. 今後の課題

現在 CoreExplorer 上に表示されている単語は単純な名詞・形容詞・動詞のみであり、複合語（「スイッチの破損」など）の表示には対応していない。そのため、現状の分析では「スイッチ」と「破損」の関係性が分かったとしても、「スイッチの破損」が現象を表すのか、原因を現すのかといった点についてまでは分析できていない。

今後は単純な単語だけでなく「複合語」を扱えるようにするほか、文構造の前後関係から複合語どうしの因果関係を抽出するなどの工夫が必要だと考えている。

尚、設計品質を確保する試みは、信頼性工学では FMEA^{*2}や FTA^{*3}といった技法で既に行われており、本技術も同技法での利用が期待できる。今後更に研究を進め、不具合情報データベースから FMEA/FTA を支援するシステムの検討を進めていく。

6. おわりに

不具合分析についての顧客ニーズに着目し、製品、部品や現象などのデータ項目間に多くの関連がある場合に効率的に関連を把握できる技術を開発した。CoreExplorer はイントラネットでの情報共有システムとしての利用はもとより、PC1 台で手軽な分析ツールとしての利用が可能である。しかし、実際に不具合情報・

顧客の声などの分析を行うまでには、元データの整備・辞書の作成・ツールを利用した分析など手間が必要である。そこで、分析コンサルティングを行う専門スタッフが導入をサポートし、またデータを預かり、分析&レポート作成などの分析サービスを提供している。不具合情報がたまっているが活用しきれていない、顧客の声が集まっているがクレームや要望が把握できない、など分析に困っているお客様を強力にサポートするための、ツールとコンサルティングを連携させたソリューションを今後も展開する所存である。

参考文献

- 1) 失敗学会, <http://shippai.org/shippai/html/>
- 2) 塚原朋哉, CoreExplorer を活用したテキストマイニングソリューションの開発, 日立 TO 技報 11 号, pp.60-66, 2005.
- 3) 山西 健司, CGM マイニングと知識化, 情報処理学会誌 Vol.48, No.8, pp.830-836, 2007.



[塚原 朋哉] [1997]年入社
[ナレッジソリューショングループ]
[CoreExplorer・テキストマイニングツールの研究開発]
[tomo@hitachi-to.co.jp]



[高梨 勝敏] [1995]年入社
[ナレッジソリューショングループ]
[知識交流システム・CoreExplorerの開発, 知識処理ツールの研究開発]
[takana@hitachi-to.co.jp]



[渡邊 まり子] [2004]年入社
[ナレッジソリューショングループ]
[CoreExplorer を利用したテキスト分析・コンサルティング]
[w_mariko@hitachi-to.co.jp]



[佐藤 俊也] [1993]年入社
[ナレッジソリューショングループ]
[知識交流システム・CoreExplorer・テキストマイニングツールの拡販, コンサルティング]
[shu_sato@hitachi-to.co.jp]

^{*2} FMEA: "Failure Mode and Effects Analysis" (故障モード影響解析). システムを構成する部品やモジュールに故障が発生した場合、システムにどの程度影響を与えるかを解析する手法。

^{*3} FTA: "Fault Tree Analysis". 望ましくない事象を定義し、その事象を発生させる要因を抽出する手法。FMEA が、大きな問題を発生させる要因がどこに潜んでいるかを抽出する手法（ボトムアップ解析手法）なのに対し、FTA はトップダウン解析手法とも呼ばれている。



[井上 悠] [2002]年入社
[ナレッジソリューショングループ]
[CoreExplorer の開発・販売, 新規
GUI 開発, 開発リーダー]
[haruka-i@hitachi-to.co.jp]



[船木 裕] [1991]年入社
[ナレッジソリューショングループ]
[開発リーダー, 知識処理ツールの研究
開発]
[funaki@hitachi-to.co.jp]