コールセンターのログデータを用いた製品等の不具合の早期発見

宅間 大介　野美山 浩

Early Detection of Product Defects by Analyzing Call Logs

Daisuke Takuma　Hiroshi Nomiyama

階層的な製品カテゴリ, 不具合カテゴリ上の相関分析によってコールログから半自動的に製品の不具合を発見する手法を提案する. 提案手法は, 従来の解析で困難であった製品リリース直後の早期問題発見を可能にし, 実験では,“ ACアダプタから騒音がする ”等の不具合を, 不具合に関するコールが非常に少ない段階で検知することができた. 特定のカテゴリの用語で記述できない問題の検知に関しては, コールログに特化した時系列モデルを構築し, 既存技術を上回る精度を実現した.

We have developed a technique to detect semi-automatically the defects of products from call logs by correlation analysis on hierarchical product categories and defect categories. The proposed method enables early problem finding immediately after the releases of new products, which has been difficult by conventional analysis. In experimentation, defects such as “noise from AC adapters” could be detected at the stage when there were very few calls on defects. For the problems which cannot be expressed in terminologies of a specific category, a time-series model specialized to call logs was built to realize accuracy exceeding the accuracy obtained by existing technologies.

Key Words & Phrases：テキストマイニング, 変化点検出, 時系列解析, コールセンター
Product defect, text mining, change point detection, time series analysis, call center

1. はじめに

近年多くの企業において, 製品の品質管理への取り組みが重要視されている. 特に, 不具合への対応の遅れは, 顧客からの信頼を失墜させ, 致命的な損害をもたらす場合があるため, 企業側は不具合を早期に検知する必要に迫られている. そこで本稿では, コールセンターのコールログを例に, テキスト・データを製品の不具合の早期発見に用いる手法を提案する.

人手で扱えないほどの大量なテキスト・データから単語やフレーズを抽出して解析を行い, 製品の評判や消費者のニーズ等, 有用な情報を得る技術は, テキスト・マイニングと呼ばれている［1］, ［2］. その中でも, 本研究で対象とするテキスト・マイニングは, 予め蓄積されたテキストからの情報抽出という従来の解析とは異なり, 常時追加されていくテキストから変化を適時に検知することを目的としている. この種の技術は, コールログからの不具合や要望の検知の他, WEBデータや新聞記事から新トピックを発見する目的でも研究されている［3］.

定型化されたデータとは異なり, テキスト・データ上では, 表記の揺れ, 略記, 間接的な表現等により, 解析したい事象の頻度さえも確実な値を求められるとは限らない. しかし, そのような状況でも, 同じ抽出基準でカウントした単語頻度の時間的推移（時系列）や, 単語同士の相関等, 本来同等である値を比較することで得られる指標は有効になり得る. 提案手法における不具合発見の手掛かりも, こうした「一様性からのずれ」の統計的な評価に基づいている.

早期発見へのアプローチとして, 我々は, 不具合を「ある＜製品＞のある＜部品＞のある＜問題＞」といった, カテゴリの組み合わせで表現可能なクラスと, 「○×▲というエラー・メッセージが出る」等, トリガーとなるキーワードが予測できず, カテゴリの組み合わせでは表現不可能なクラスに分類した. そして, 前者に対しては, 特定の製品に偏って起こる不具合を早期に発見できる相関分析を, 後者に対してはコールログに特化した時系列解析を考案した. ここでの提案の新規性は, 以下の2点であると考える.

提出日：2004年8月30日　再提出日：2005年5月11日

(1)階層型カテゴリ上で相関分析を繰り返し行うことで, 特定の製品種別固有の不具合等, 時系列解析では検知できない不具合を発見する.
(2)コールセンターの電話件数の特性を考慮して, 時系列モデルを構築する.

実験では, PCコールセンターのコールログを用いた. 相関分析では, 製品の不具合に関する警告の上位24件のカテゴリの組のうち, 実際に不具合と判断できるものとして, {機種▲▲▲, ACアダプタ, 騒音}(機種▲▲▲でACアダプタから騒音がする)等, 4件が見つかった. この中には, 単語共起としては初出の時間区間で警告されているものもあり, そういった不具合を時系列解析によって検出するのは困難である. 時系列解析については, 提案手法と既知のモデル3種について, 警告を発する閾値を変化させながら, 同じ警告数における警告の的中率を比較し, 提案手法の優位性を示した.

2. 関連研究

トピック抽出やネットワーク侵入検知等, 時間情報を含むデータ上の変化検出に関しては, 対象分野, 解析手法によらない問題定義として“Activity Monitoring”がFawcett [4] らにより与えられており, その観点として,
(1)Granularity:変化の内容を適切な粒度で報告すること.
(2)Multiple alarms:同じ現象に対して複数の警告を出さないこと.
(3)Benefit of timely alarms:できるだけ早期に報告すること.
が挙げられている. 本研究はこれらの要件を満たすべく, (1), (2)に関しては,“マウスの不具合”,“ポインティング・デバイスの不具合”といった冗長な警告を統計的に意味のある粒度のものに集約する方法を, (3)に関しては, 繰り返し行う相関分析を製品リリース直後のデータが少ない状況に対応させる方法を提案している.

テキスト・データに特化して, 単語, 係り受け等の頻度時系列を利用した変化検出としては, Kleinberg [5] が, トピック/サブトピックの構造を含めた頻度の急上昇の検出を行っているが, これは過去のデータ上の変化点検出であり, 本研究が対象とする, 将来が未知の状態での動的な変化検出とは異なる. [5] のモデルは, 藤木ら [6] が考察しているとおり, 一定期間の文書数がポアソン分布に従うという仮定に相当する. 我々もポアソン分布を用いてモデル化しているが, 提案手法ではポアソン分布で期待値と分散が等しくなる性質を利用して、定常状態での変動スケールを効率良く推定しており, その効果は実験でも示されている.

その他の変化検出研究のトピックとしては, 計算の高速化と, 異なるタイム・スケールの時系列への対応が挙げられる([7], [8]). 本研究では, 前者への要求レベルは低いが, 後者は重要である. この点に関して, 本稿で提案する相関分析は, 累積的な頻度に基づいているため, 短期間で急激に起こる不具合も長期間で緩やかに起こる不具合も区別せずに扱える.

3. 不具合の早期発見

この章では, カテゴリを用いた不具合の表現, カテゴリと文書との対応付け, 相関分析, 時系列解析について述べる. 相関分析と時系列解析は, **表1**に示すような特性の違いがあり, 相補的な効果が期待される.

表1. 相関分析と時系列解析の比較

手法	相関分析(3.3節)	時系列(3.4節)
対象とする不具合	表現可能:カテゴリの組み合わせで表現可能な不具合	表現不可能:カテゴリの組み合わせでは表現不可能な不具合
判定基準	単語, カテゴリの共起の偏り	単語等の頻度の非定常な増加
手法の特性	緩やかな増加でも検知可能. 不具合に関するデータが少なくても検知可能.	急激な頻度変化のみに対応. 数期間分のデータが必要.

3.1 シソーラスによる領域知識の表現

この節では, シソーラスによって領域知識を表現する方法について述べる. 例として, PCコールセンターにおける“ノートPCのシリーズAの製品がHD(ハードディスク)を認識しない”という不具合を想定する. この場合, キーワードとして“シリーズA”,“HD”が, 係り受けでは“HD…認識しない”が内容を特徴付けており,“シリーズA”は<製品>,“HD”は<不具合に関係する部品>,“HD…認識しない”は<不具合の

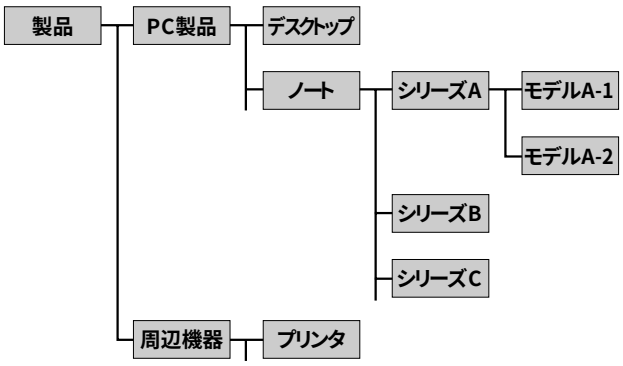

図1. 製品シソーラスの例

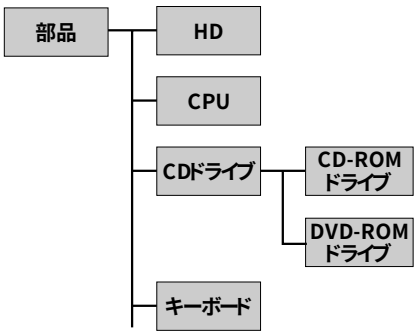

図2.　部品シソーラスの例

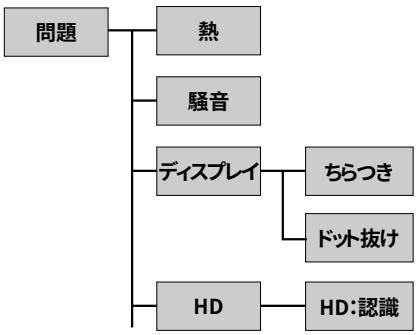

図3.　問題シソーラスの例

内容＞という観点での情報となっている. そこで, これらの観点をツリー構造のカテゴリ（シソーラス）に反映する. **図1**, **図2**, **図3**はそれぞれ製品, 部品, 問題のシソーラスの例である.

以上の例を一般化して, 本稿では,“ シソーラス ”とは以下の(1), (2), (3)を満たすツリー構造を意味することとする.

(1)各ノードはカテゴリに相当し, 人間に認識可能なラベルが振られている.
(2)親ノードと子ノードは不具合の表現として一般化／詳細化の関係[1]にある.
(3)シソーラス内で分類の観点は一貫している.

3.2　不具合の表現と文書へのマッピング

前節のシソーラス上のカテゴリを組み合わせると,“ ○○の製品で, ○○の部品に, ○○の問題が生じる ”というパターンの不具合を表現できる. 以下ではこれらの不具合を, 表現可能クラスと呼び, この節では表現可能クラスの不具合を文書と対応付ける方法について述べる.

まず, シソーラスのリーフに位置するカテゴリに対し, 関連する単語, 係り受けを登録する. 問題シソーラスの“ 騒音 ”カテゴリに登録する単語の例を以下に挙げる.

“ 騒音 ”,“ 異音 ”,“ うるさ（い）”,“ 音…鳴（る）”,
“ 音…大き（い）” （“ … ”は係り受け）

単語と係り受けの登録により, 任意のリーフ・カテゴリに対して, そのカテゴリに属する単語, 係り受けを含む文書集合が定まる. リーフでないカテゴリについては, そのカテゴリの下位のリーフ・カテゴリがマップされる全文書集合の和集合を対応させる.

以上により, カテゴリの組み合わせによる不具合表現と, 各カテゴリから文書集合へのマッピングが定義された. 以後, 不具合の内容を複数のシソーラスX_1, X_2, …, X_nに属するカテゴリの組み合わせ{C_1, C_2, …, C_n}で表したものを“ カテゴリセット ”と呼び, C_1, C_2, …, C_nに対応する文書集合の共通部分の文書の総数を#{C_1, C_2, …, C_n}で表してカテゴリセットの“ 頻度 ”と呼ぶ.

3.3　表現可能クラスの不具合の相関分析

ここでは, 表現可能クラスの不具合を早期に検知し, 冗長性を除いて警告する方法について述べる. 一般的に, 製品の不具合への反応の多くは製品のリリース直後から, コールログ中に現れることが多い. しかし, 従来の単語やカテゴリの頻度上昇によって不具合を検出する手法では, 製品リリースに伴う頻度上昇と不具合に起因する頻度上昇を区別できない. これに対し提案手法では, 各製品カテゴリにおける不具合の偏りに注目することで, 時系列としての特徴を持たない不具合の検出も可能にしている.

以後は, 再び3.1節のシソーラスを例に用いることとする. 相関分析では製品ごとの不具合の傾向の偏りに注目する. まず, 製品シソーラスのカテゴリで, それより下位の製品カテゴリで不具合の傾向が似ていると考えられるカテゴリC_{base}を定める（**図4**）. 製品の不具合は, 部品カテゴリC_1と問題カテゴリC_2の組{C_1, C_2}で表せたので, 製品の平均的な不具合の傾向を表す指標として, 製品に関するコールの中の不具合{C_1, C_2}に関するコールの割合：

$$\#\{C_{base}, C_1, C_2\} / \#\{C_{base}\}$$

を計算できる. 同様にC_{base}より下位の各製品$C_{product}$（図4）についても, 不具合の傾向

$$\#\{C_{product}, C_1, C_2\} / \#\{C_{product}\}$$

を計算できる. これらを用いて, 製品$C_{product}$における不具合{C_1, C_2}の偏りの度合いを

[1] カテゴリ同士が“ is-a ”の関係にある必要はなく, 例えば,“ キーボード ”カテゴリの下位カテゴリとして“ part-of ”の関係の“ テンキー ”があっても良い. これらは不具合を表現する内容としては, 包含関係にある.

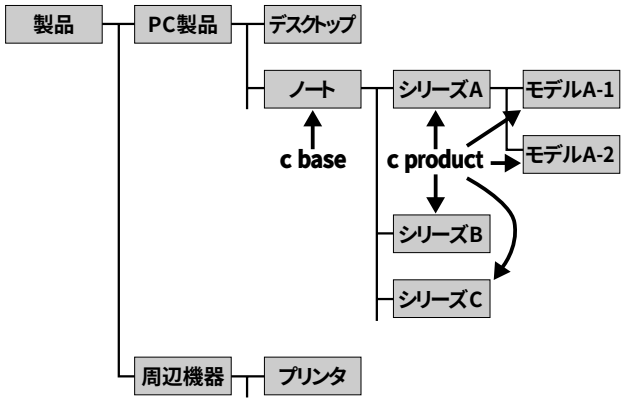

図4. 製品シソーラスにおけるc_{base}, $c_{product}$の例

$$R(\{C_1, C_2\}, C_{product}) = \frac{\#\{C_{product}, C_1, C_2\} / \#\{C_{product}\}}{\#\{C_{base}, C_1, C_2\} / \#\{C_{base}\}}$$

と定義し, 相対頻度と呼ぶことにする.

以後は$C_{product}$を固定し, 相対頻度を不具合のカテゴリセット$\{C_1, C_2\}$のみの関数として$R(\{C_1, C_2\})$と略記する. 実際のモニタリングでは, 製品がリリースされた時点からデータが追加されるごとに相対頻度を繰り返し計算し, 予め定めた閾値を越えた時点で警告する. ただしここで, 製品リリース直後では, まだ相対頻度の値を十分な信頼性を持って計算できるだけのデータが揃っていないという可能性を考慮する必要がある. そのため, 少ないデータによる偶然の偏りを検知することのないように, サンプル数を考慮した区間推定を行い, 真の相対頻度の値が一定確率(80%,90%,...)以上で含まれる区間を算出し, その左端の値$R^*(\{C_1, C_2\})$を補正済みの相対頻度として不具合検出の指標に用いる.

このようにして, 各製品について偏って起こる不具合のカテゴリセットを検出できるが, 警告には以下のような冗長なものが含まれる.

(1)シソーラス上でカテゴリが上位下位の関係にある.
カテゴリセットA:{部品／ハードウェア／ポインティングデバイス}
カテゴリセットB:{部品／ハードウェア／ポインティングデバイス／マウス}
(2)一方に別のシソーラスのカテゴリが含まれている.
カテゴリセットC:{部品／ハードウェア／ファン}
カテゴリセットD:{部品／ハードウェア／ファン, 問題／ハードウェア／騒音}
(3) (1),(2)の組み合わせ

(1)では, 警告をどこまで詳細にするかが問題となる. この場合, 詳細なカテゴリほど対応する文書数が少なく, 偶然の偏りが起きやすいことのみを考慮すれば良いので, $R^*(A)$と$R^*(B)$を比較し, 大きい方を優先する.

(2)の場合は,“ ファン ”の不具合と“ 騒音 ”の不具合が独立に起こっているのか, それとも“ ファンからの騒音 ”の不具合が起こっているのか, を判断する必要がある. そこで{ファン}の相対頻度と{騒音}の相対頻度を用いて, ファンと騒音の不具合が独立に起こっている場合の{ファン, 騒音}の相対頻度を計算し, 実際の{ファン, 騒音}の相対頻度がそれより有意に大きければ{ファン, 騒音}を{ファン}, {騒音}より優先するようにする. 独立性を仮定した場合の相対頻度の計算は, カテゴリセットEを{問題／ハードウェア／騒音}とすると, $R^*(C) \times R^*(E)$となる.

最後に(3)の場合を, (1), (2)の比較の組み合わせに帰着させることで, 任意の冗長な2つのカテゴリセットに対し警告の優先度を決定できる.

以上の手法を用いることにより, 特定の製品に特異的な不具合のカテゴリセットを検出し, 冗長性を排除して警告を出すことができる.

3.4 表現不可能クラスのための時系列解析

不具合の早期発見タスクにおいては, 3.1節で定義した表現可能クラスの不具合に合致しない未知の不具合(表現不可能クラス)も検出できることが望ましい. 表現不可能クラスについては, 従来の手法と同様に, 単語や係り受けの頻度の増加に注目する. 以下では, テキスト・データをコールログに限って, 既存手法以上の精度を持ち, 短い期間の観測で変化を検出できる時系列解析法を提案する.

まず, 実データの観察に基づいて, 一定時間区間ごとの特定の単語を含むコールの頻度時系列の性質として, 以下の2点を仮定する.
(1)時系列は定常状態と非定常状態とを遷移する. 定常状態 → 非定常状態で頻度が増加することが多い.
(2)定常状態では, 各時間区間での頻度はパラメータが等しい独立なポアソン分布に従う.

(1)は経験則であり, (2)はコールセンターでは電話をかける人に過去の頻度のフィードバックがないこと(独立性)と, 定常状態でのコール頻度がポアソン分布に従う性質を反映している. 不具合の検出には定常状態のノイズとして標準偏差を推定し, 相対的に大きな外れ値を取る瞬間をとらえる必要がある. 通常, 標準偏差や分散の推定には長い時間の観測を必要とするが, 提案手法では, ポアソン分布の“ 期待値＝分散 ”という性質を用いて, 分散を期待値として推定することによりこれを回避する. つまり, (2)の利用の効果は, 少ないサンプル数しかない段階でも, ノ

イズのスケールの正確な計算ができることと言える. これらに基づいて, 分布の平均の推定値には過去の頻度の重み付き平均を用いて, 非定常性の計測には定常状態の分布からの外れの度合いである確率的情報量を採用することで, 時系列からの非定常事象を発見することにする.

4. 実験

4.1 表現可能クラスの不具合の相関分析

3.3節で述べた相関分析を検証するために, PCコールセンターのコールログ約35万文書を月ごとに解析した. 相対頻度の区間推定における信頼係数は90%を用いた. その結果, R^*を閾値4で切った上位24件のカテゴリセットうち, 警告として価値があると判断できるものが4件あった. **表2**にその内容を示す. 有用と判断しなかった警告には, 警告後に相対頻度が下がったものの他, 特定の機種にのみ取り付けられているコンポーネント名に関する警告等があった.

表2. 実験で検出された不具合

不具合内容	不具合カテゴリセットのラベル
ACアダプタから異音がする	機種/○○○ 部品/ハードウェア/電源/バッテリ・ACアダプタ 問題/ハードウェア/騒音
ディスプレイにドット抜けがある	機種/××× 部品/ハードウェア/ディスプレイ 問題/ハードウェア/ディスプレイ/ドット
ハードディスクを認識しない	機種/■■■ 問題/ハードウェア/ハードディスク
ネジ穴に関する質問が多い	機種/▼▼▼ 部品/ハードウェア/ネジ

警告時期については, ACアダプタの騒音, ドット抜け, ネジに関する問い合わせは, 該当カテゴリが初めて共起した月に検知できている.

冗長な不具合警告の排除処理については, {電源}, {バッテリー／ACアダプタ}, {電源, 騒音}, {騒音}, {バッテリー／ACアダプタ, 騒音}から{バッテリー／ACアダプタ, 騒音}が選択されるといった効果が見られた.

4.2 表現不可能クラスの不具合の時系列解析

3.4節で提案した時系列解析の精度については, 4.1節と同じコールログの“一般名詞(その他)”,“固有名詞”に分類される400語の週ごとの頻度時系列を用いて比較検証した. これらの単語は専門用語が多く, 一語で不具合を特徴付けている可能性が高い. **表3**に, 提案手法と比較実験に用いた3つの手法について示す.

表3. 非核実験に用いた時系列解析手法

手法	増加判定の指標	時系列に対する仮定
提案手法	過去の平均値をパラメータとするポアソン分布における最新の頻度の情報量	各週の頻度は独立同分布なポアソン分布に従う
差分	今週の頻度 − 先週の頻度	時系列は連続的に変化する
正規化	(今週の頻度 − 平均)／分散	各週の頻度は独立同分布に従う
畳み込み	加重平均の増加分	大きなスケールで見て連続的に変化する

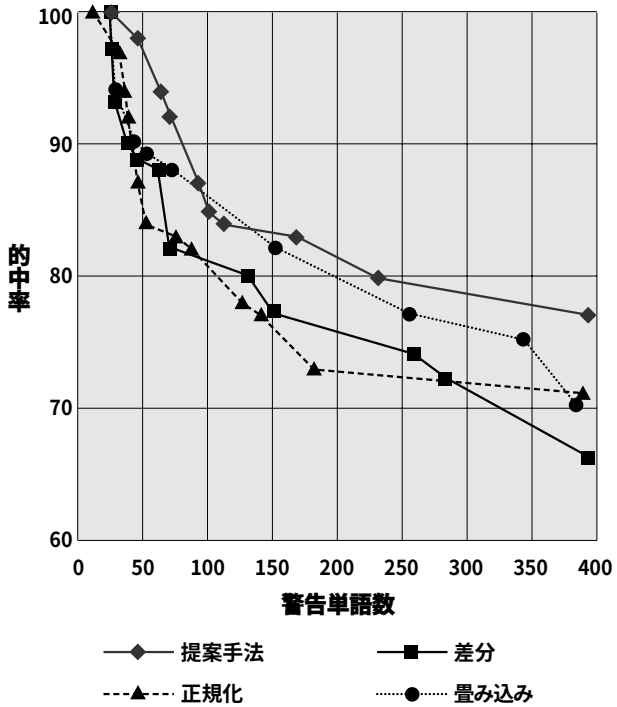

図5. 400単語の頻度時系列における各解析手法の警告数と的中率の関係

警告は増加判定の指標と閾値との大小評価に基づくため, 比較実験では各手法について閾値を変化させながら, 警告された単語数と的中率の関係を調べた. 的中率については, 警告週の前後5週間の平均頻度を比べ, 後5週間の方が大きければ“的中”とした. また, 多重警告回避のため, 各単語に対する警告は初めの一回のみとした. **図5**に比較実験の結果を示す. この図から提案手法では, 差分手法, 正規化手法と同じ警告の的中率を, 約倍の警告数で実現でき, 畳み込み手法と比べても優れていることが確認できる. また, 指標としてはほぼ同等の値を用いた正規化手法との比較により,“期待値=分散”という性質を用いる効率化の効果も示された.

実際に警告された単語は, ウィルス名など, 有用なものもあったが, 一方で増加判定は成功していても,

警告としては価値の低い, 新しいコール担当者の名前等も検知された.

5. 考察

5.1 表現可能クラスの不具合の相関分析

表2の結果から, 警告の指標が高い値を示しているカテゴリセットを人手で検証するというプロセスで, 不具合を発見できることが期待される. また, 警告はカテゴリセットの頻度が初めて正値となった月から出ており, 時系列解析では不可能な段階での検出に成功している. 精度に関しては, 製品に特異的な不具合ほど警告指標が顕著な反応を示している傾向が見られた.

5.2 表現不可能クラスのための時系列解析

時系列解析の非定常性の検知精度については, 図5の結果から, 提案手法の優位性は明らかである. 実用面では, 有用な警告を出すこともできたが, 不要な単語の増加警告が大量に出される問題は, 製品カテゴリなど自明な増加が予測されるカテゴリを除いたにもかかわらず, 依然として解決されていない. 不要警告の多くは, 増加予測という意味の時系列解析では成功しているため, 改善には, 単語を選別する等, 言語処理面での更なる工夫が必要と考えられる.

6. 結論

本研究ではテキスト・データを用いた不具合の早期発見というタスクに対して, 不具合をカテゴリの組み合わせで表現し, テキスト・データと対応付けた上で, 相関と時系列の二種の情報に基づくアプローチを試みた. 相関情報は, これまでテキストからの変化検出では利用されていなかったが, 階層的カテゴリの利用, サンプル数の考慮により実用的な結果が得られた. 時系列解析では, コールログの特性を再考することで, 解析精度を向上させた. 両アプローチの併用による相補的な効果については, それぞれが性質の異なる不具合を発見できた点では意味があったが, カテゴリを分けることで, 自明な頻度増加が頻繁に起こる単語群を時系列解析の対象から除くことは依然として困難であった.

既存のテキスト・マイニング技術([9])と比較した場合, 提案手法には, 不具合の内容を単語や係り受けだけでなく, カテゴリの組として様々な粒度で分かりやすく表現できるという利点がある. また, 予め分析するカテゴリを定めておくことで, 分析を自動化できるため, 分析するカテゴリを人手で設定していた従来の作業よりはるかに効率的である. 自動処理によって警告された不具合を人手によるインタラクティブな操作で検証することで, 網羅率も信頼性も高い不具合検出が可能であると考える. **図6**に, 提案手法を用いた不具合検出プロセスを示す.

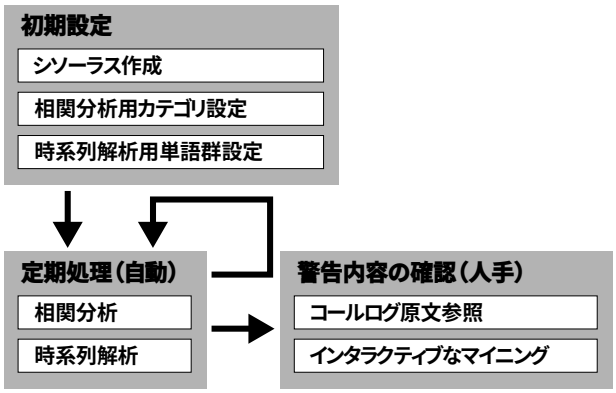

図6. 不具合検出プロセス

シソーラスの構築に関しては, 通常, 大きなコストがかかると考えられているが, 実験で用いたシソーラスの構築コストは比較的小さかった. これは, シソーラスで表現する領域知識が意味的に狭い範囲に限られていたこと, 製品分類など既に作成された情報を有効利用できたことによる.

不具合の早期発見は, 今後も多くの企業が取り組むべき問題だが, 提案手法は, コールログというユーザからの直接のフィードバックを利用できる手法として有効なアプローチと言える.

謝辞

本論文は, 情報処理学会研究報告[10]をベースに加筆・修正したものである. 本論文を本誌に転載することを許可していただいた情報処理学会に感謝致します.

参考文献

[1] 林 俊克, Excelで学ぶテキストマイニング入門, オーム社, 2002

[2] M.A.Hearst, Untagling text data mining, in Proc.og the 37the Annual Meeting of the Association for Computational Linguistics, 1999

[3] James Allan, Jaime Carbonell, George Doddington, Jonathan Yamron and Yiming Yang, Topic Detection and Tracking Pilot Study Final Report, Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval, 1998

[4] Tom Fawcett and Foster Provost, Activity Monitoring: Noticing interesting changes in

behavior, In Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 53-62, 1999
[5] Jon Kleinberg, Bursty and hierarchical structure in streams, In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002
[6] 藤木 稔明, 南野 朋之, 鈴木 泰裕, 奥村 学, document streamにおけるburstの発見, 情報処理学会研究報告, 2004-NL-160, pages85-92
[7] Yunyue Zhu and Dennis Shasha, Efficient Elastic Burst Detection in Data Streams, Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 336-345, 2003
[8] Yunyue Zhu, Dennis Shasha, Statstream: Statistical Monitoring of Thousands of Data Streams in Real Time, In Proceedings of 28th International Conference on Very Large Data Bases, pages 358-369, 2002
[9] Tetsuya Nasukawa, Tohru Nagano, Text Analysis and Knowledge Mining System, IBM SYSTEMS JOURNAL, VOL 40, NO 4, 2001
[10] 宅間 大介, 野美山 浩, テキストデータを用いた問題の早期発見手法, 情報処理学会研究報告 2004-NL-162, pp.19-26, 2004

日本アイ・ビー・エム株式会社
東京基礎研究所
副主任研究員

宅間 大介 Daisuke Takuma

[プロフィール]
2003年，日本IBM入社．テキストマイニング・システム IBM TAKMI Text Analysis and Knowledge Mining の設計・開発を担当．異常検知のためのデータ解析，テキスト・データ解析を高速に行うためのインデクシング技術の研究に従事している．

ta9ma@jp.ibm.com

日本アイ・ビー・エム株式会社
東京基礎研究所
主任研究員

野美山 浩 Hiroshi Nomiyama

[プロフィール]
1985年，日本IBM入社．東京基礎研究所にて機械翻訳、情報検索、テキストマイニング等の研究に従事．現在，アイ・ビー・エム ビジネスコンサルティングサービスに出向中．

nomiyama@jp.ibm.com