

# セキュリティ要件分析支援ツール TACT

半田剣一<sup>†1</sup> 大崎人士<sup>†1</sup> 竹内泉<sup>†1</sup>

**概要:** 本研究は、階層的クラスタリング技術をセキュリティ分野の要件分析に応用するものである。本研究では、数百～数千件程度のセキュリティ要件から半自動的に階層的構造を抽出することにより、要件群の体系化を支援するソフトウェアツールを開発した。このツールにより、要求分析プロセスにおいて大幅なコスト削減が期待される。

**キーワード:** 要件分析, 階層的クラスタリング

## Security Requirements Analysis Supporting Tool: TACT

KENICHI HANDA<sup>†1</sup> HITOSHI OHSAKI<sup>†1</sup> IZUMI TAKEUTI<sup>†1</sup>

**Abstract:** This research is to apply the hierarchical clustering method to the field of security requirements analysis. We have developed a software tool that supports to systematize hundreds to thousands of security requirements by extracting the hierarchical structure semi-automatically, which leads to big cost down of requirements analysis process.

**Keywords:** Requirements Analysis, Hierarchical Clustering

### 1. はじめに

新しい技術が社会導入されて、社会環境が変化すると、新たなセキュリティ脅威が顕在化する。機器のセキュリティレベルを一定に保つためには、機器のセキュリティ要件や管理策を体系化し、国際基準や関係法令等のセキュリティ規範と対応づけをした上で継続的に見直し更新していく必要がある。しかし数百～千件にのぼる要件や管理策を人手で構造化し体系づけるには多大なコストがかかる。

本研究の目的は、システム開発の要求分析プロセスにおいて、大量の要件一覧を半自動的に階層的にグループ化することで、セキュリティ要件の構造化及び体系化にかかるコストを大幅に減らし、さらに思い込みや見落としによる分析ミスを減らすことである。

近年のビッグデータの広がりとともに機械学習やデータマイニング等の研究は急速に進歩し、テキストを対象とした分野でも実用的な分析手法が提案され、テキストの趣旨や傾向を抽出できるようになった ([1]他)。

筆者らは、従来ビッグデータの傾向分析に適用されていたクラスタリング技術を利用して、数百～千件の文書データを対象としたより精緻な階層化を可能にするための技術開発を行い、要件分析支援ツール TACT を作成した。

### 2. TACT の概要

TACT の処理の流れをエラー! 参照元が見つかりません。に示す。まず自然言語文の形態素解析を利用し各要件からキーワードを抽出し、それを元に各要件の特徴ベクトルを決定する。次に特徴ベクトルを元に類似度を計算して階層的クラスタリング技術により要件群を階層的に分類する。

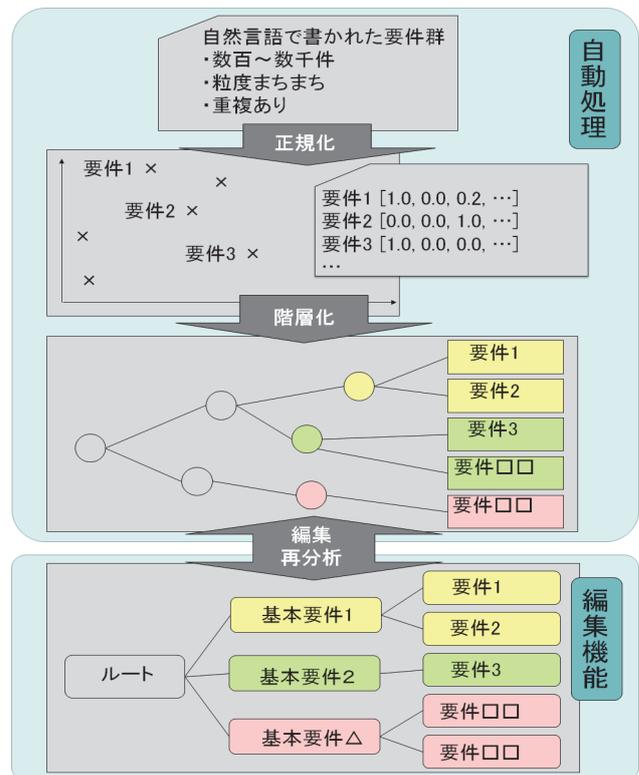


図1 TACT の処理の流れ

ここまでは自動処理で行う。最後に人手で階層木を編集する機能を用い、一部の自動再分析を繰り返しながら、要件全体を整合性のとれた体系へと改良していく。TACT の動作画面を図2に示す。

### 3. TACT の要素技術

#### 3.1 要件の特徴量抽出

要件の文章は名詞句によって特徴付けられる傾向が強いため、形態素解析により名詞を抽出し、要件群全体につ

<sup>†1</sup> 国立研究開発法人 産業技術総合研究所

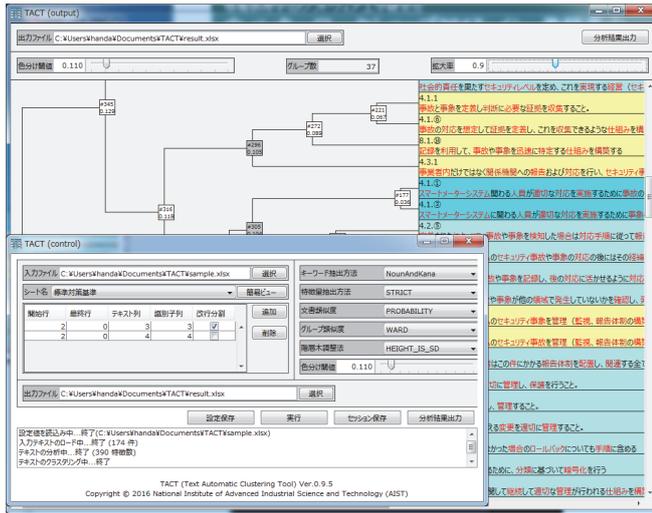


図2 TACTの動作画面

いての連続した名詞の出現パターンを考慮してキーワードを決定する。さらに、領域固有の用語辞書によるキーワードの重要度係数と、下記に述べるトピックモデル解析によって得られる隠れキーワードの出現確率から、各要件の特徴ベクトルを算出する。

### 3.2 階層木生成

階層木の生成には凝集型の階層的クラスタリング[2]技術を用いる。文書間の距離関数は、これまで提案された主な計算法（ユークリッド距離，コサイン距離，マハラノビス距離等）の他に，特徴ベクトルの要素を確率値として扱う確率距離を新たに実装した。これは，文書  $x, y$  が特徴  $f$  を持つ確率が  $P_{xf}$  と  $P_{yf}$  である場合，特徴  $f$  についての距離を  $P_{xf} * (1 - P_{yf}) + (1 - P_{xf}) * P_{yf}$ ，すなわち一方のみが  $f$  を持つ確率とするものである。例えば両文書とも 0.5 の確率で  $f$  を持つ場合， $f$  についての距離は 0.5 となる。

クラスター間の距離関数もこれまで提案された主なもの（最短距離法，群平均法，ワード法等）を実装した。ユーザは対象文書群に適した関数の組を選択できる。

### 3.3 トピックモデル解析

人手で書かれる要件の文章にはその要件に関するキーワードがすべて出現するとは限らないが，要件に隠されたキーワードの出現確率を推定することで，より精緻な階層木が得られる。TACT は統計学的処理を行うトピックモデル[3]によりこの推定を行う。

トピックモデルは，基本的には，特徴ベクトルで表現された要件群全体を行列として捉え，それを積によって近似するような，より低次元の2つの行列（文書-話題行列と話題-キーワード行列）を算出することで，各文書に内在する話題および各話題に各キーワードがどう影響しているかを発見するものである。

しかし，この2つの行列の積による近似行列に注目すると，それは各文書が各キーワードを持つ確率を表現しているものと考えることができる。TACT はこの情報を元に隠れキーワードを推定する。

## 4. 考察

### 4.1 階層木生成のパラメータ

前節で示した複数の関数を事例研究で試したところ，要件群の体系化という目的からは，安定してバランスの取れた階層木を生成するワード法が良好な結果を示した。また，短い文章（つまり特徴ベクトルの殆どの要素が 0）同士はたとえ共通キーワードが無くとも同じグループに分類されがちであるが，隠れキーワードの推定とともに確率距離を使用することで，この傾向を削減することができた。

### 4.2 生成された階層木の調整

階層的クラスタリングは2分木を生成するため，要素数が増えると木の深さも増えていく。しかし，体系化という目的のためには，クラスター要素の標準偏差がさほど変化しない分類はあまり意味を持たない。そこで，標準偏差が大きく変化する分類階層のみに階層木を縮退させる機能を開発した。これにより編集機能を使つての人手での調整の手間が軽減する。

### 4.3 事例研究

あるIoT機器システムの脅威分析を元に挙げられた200件弱のセキュリティ対策要件の体系化にTACTを利用し，多数の類似要件の検出や国際規格と抽象度を合わせた体系化を効率良く行えることを確認した。

また，人手で精査され構造化された文章のレビューにもTACTが有効であることを確認した。具体的にはスマートメータシステムの管理策及び対策例[a]300件弱から，精査モレ（非常に似た対策例の不統合）や管理策から対策例への具体化の不手際を検出することができた。

## 5. おわりに

大量のセキュリティ要件の体系化を支援するツールTACTについて述べた。本研究の事例から，特徴ベクトルの抽出に，自然言語処理や統計学的処理を組み合わせることで適用することにより，対象文書データに対して精度の高い特徴づけができることが分かった。本研究の要素技術は，あらゆる産業分野や製品に適用可能である。クラスタリング技術をコアとする要件分析の応用事例は，セキュリティ対策だけでなく，機能安全や保安等の目的にも利用することができる。本研究は，今後，自動車を含む我が国の重要なづくり産業のための品質特性に対しても応用の範囲を広げていく予定である。

## 参考文献

- [1] 橋本 泰一,他. 文書クラスタリングによるトピック抽出および課題発見. 社会技術研究論文集, 2008, vol. 5, p.216-226.
- [2] 元田,他. データマイニングの基礎. オーム社, 2006.
- [3] 佐藤一誠. トピックモデルによる統計的潜在意味解析. コロナ社, 2015.

[a] 平成27年7月スマートメータ制度検討会セキュリティ検討WG報告書の付属書「統一的なガイドラインの標準対策要件に盛り込むべき事項」より抽出