非機能要件の詳細な自動分類手法の構築に向けて：セキュリティ要件の分類

宮崎 智己[1,a)]　大平 雅雄[2,b)]

概要：非機能要件の自動分類手法が提案されている．しかしながら，既存手法は類義語に対応できない，また分類の粒度が大きく，分類結果を更に目視して詳細に分類する必要がある問題点がある．本研究では，類義語対応可能かつ，詳細な分類を可能とした非機能要件自動分類手法を提案する．

1. はじめに

システムが満たすべき要件には，システムの機能に関する機能要件（FR）と Performance, Usability, Reliability, Security といったのシステムの性能や品質に関する非機能要件（NFR）がある．開発者は発注者と要件定義を行い，システムに必要な FR と NFR を定義し要件定義書にまとめ，要件定義書をもとに設計・実装を行う．しかしながら，FR の記述の中に NFR の内容が混在している，開発者によって同じ NFR を指す記述でも異なる表現になる [1] といった理由から，NFR が見落とされてしまうことがある．NFR の見落としはシステム開発の手戻りを引き起こしてしまう．

手作業による NFR の見落としを防止するために，教師あり学習を用いた NFR 自動分類手法が提案されている [2]．しかしながら，教師あり学習を用いた手法は大量の学習データが必要であり，学習データの作成にコストがかかる．そこで，学習データの作成コストを削減するために，Mahmoud は単語の関連度を算出する Normalized Google Distance (NGD)[3] と教師なし学習である階層クラスタリングを用いた手法を提案している [4]．しかしながら，Normalized Google Distance はドキュメント内の単語の共起関係を用いて単語間の関連度を算出するため，同じ意味をもつ単語の関連度を低く算出してしまうという問題点がある．また，既存手法が対象としている NFR は非常に幅広い分類となっており，既存手法の利用者は分類結果を更に目視して詳細に分類する必要があるという問題点がある．例えば，ISO25010[*1] では Security を Confidentiality, Integrity, Non-repudiation, Accountability, Authenticity の 5 つに分類している．既存手法では Security までを分類対象としているため，既存手法の利用者は Security と分類された結果を更に目視して分類する必要がある．そこで，本研究では，類義語に対応可能，かつ，詳細な NFR 自動分類手法を構築することを目的とする．特にセキュリティ要件の詳細な分類を対象とする．詳細な分類を可能にすることで，開発者の NFR の見落としや分類のコストを削減することを目指す．

2. 提案手法

本研究では，類義語に対応可能かつ，詳細な分類を可能とした NFR 自動分類手法を提案する．提案手法の概要を図 1 に示す．提案手法は以下の 4 つのステップで構成される．

（1）前処理

名詞や動詞といった単語は使われる文脈によって語形が変化するため，そのまま扱うと同じ単語を別の単語とみなしてしまう．そこで，対象とする文に対してステミング処理を行う．また，“is” や “the” といったストップワードは除去する．

（2）Word2Vec を用いた類義語の特定

NFR の記述文は自然言語で記述されるため，同じ NFR を指す記述文でも異なる単語が用いられる場合がある．例えば，“system” と “product” は同じ意味で用いられることがある．こういった類義語による表記ゆれは，分類精度に悪影響を及ぼすため，Word2Vec を用いて NFR 記述文に含まれる単語の類義関係を特定する．

1 和歌山大学大学院システム工学研究科
2 和歌山大学システム工学部
a) miyazaki.tomoki@g.wakayama-u.jp
b) masao@sys.wakayama-u.ac.jp

*1 http://iso25000.com/index.php/en/iso-25000-standards/iso-25010

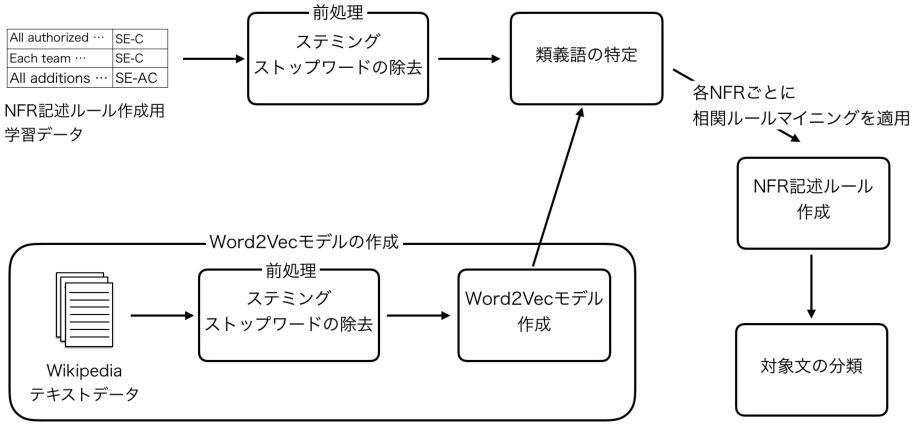

図 1: 提案手法の概要

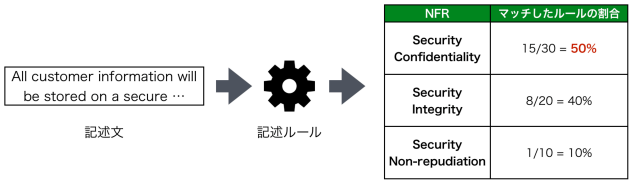

図 2: 分類方法

Word2Vec とは，ニューラルネットワークを用いた単語分散表現手法であり，同じ意味を持った単語の類似度を高く計算することができる．Word2Vec の学習には Skip-gram モデルを適用し，学習データには Wikipedia のテキストデータを用いる．Wikipedia のテキストデータを学習データに用いる理由は，Wikipedia の学習には大量の学習データが必要となるため，また，容易に取得が可能であるためである．学習した Word2Vec モデルを用いて，NFR 記述文中に含まれる全ての単語間の類似度を計算する．類似度の閾値を予め設定し，閾値以上の類似度を持つ単語を類義語とみなして，類義語を抽出する．

(3) 相関ルールマイニングを用いた NFR 記述ルールの作成

同じ NFR を指す記述文であっても表現方法が異なるため，様々な表現方法に対しても分類可能な手法を作成する必要がある．そのため，各 NFR の記述文に対して Apriori アルゴリズムを用いた相関ルールマイニングを適用して NFR 記述ルールを作成する．NFR 記述ルールとは，NFR の記述として満たすべき単語の共起パターンのことである．Apriori アルゴリズムでは，支持度と確信度の 2 つの指標を用いる．支持度と確信度のそれぞれに閾値を設定し，支持度と確信度がどちらも閾値以上の単語の共起パターンを NFR 記述ルールとして抽出する．

(4) NFR 記述ルールを用いた文の特定

最後に，作成した NFR 記述ルールを用いて記述文がどの NFR に該当するかどうか分類する．まず，分類対象の文が各 NFR 記述ルールにマッチするか調べる．分類対象の文に NFR 記述ルールに含まれる単語が全て含まれている場合はマッチしている，1 つでも記述ルールの単語が含まれていない場合はマッチしていないと判断する．次に，各 NFR ごとに分類対象の文にマッチした記述ルールの割合を計算し，マッチした記述ルールの割合が最も高い NFR を分類対象の文の NFR であると割り当てる．図 2 に分類判定の方法を示す．図 2 の例の場合，Security Confidentiality がマッチした割合が 50 %と最も高いため，Security Confidentiality の文として分類する．

3. おわりに

本研究では，類義語対応可能，かつ，詳細な分類を可能とした NFR 自動分類手法を提案した．今後は，ISO25010 に基づいて作成したデータセットを用いて，提案手法が実際に NFR の詳細な分類が可能であるか評価実験を行う．

謝辞 本研究の一部は，JSPS 科研費 JP17H00731，JP18K11243 による助成を受けた．

参考文献

[1] Abad, Z. S. H., Shymka, A., Pant, S., Currie, A. and Ruhe, G.: What are Practitioners Asking about Requirements Engineering? An Exploratory Analysis of Social Q A Sites, *Proc. of IEEE 24th International Requirements Engineering Conference Workshops, (REW '16)*, pp. 334–343 (2016).

[2] Zhang, W., Yang, Y., Wang, Q. and Shu, F.: An Empirical Study on Classification of Non-Functional Requirements, *Proc. of the 23rd International Conference on Software Engineering & Knowledge Engineering (SEKE '11)*, pp. 444–449 (2011).

[3] Cilibrasi, R. L. and Vitanyi, P. M. B.: The Google Similarity Distance, *IEEE Trans. on Knowl. and Data Eng.*, Vol. 19, No. 3, pp. 370–383 (2007).

[4] Mahmoud, A.: An information theoretic approach for extracting and tracing non-functional requirements, *Proc. of 23rd International Requirements Engineering Conference (RE '15)*, pp. 36–45 (2015).