

Convolutional Neural Network とドメインオントロジーを用いた 非機能要求の要約手法の提案

齋川 祐太[†] 長岡 武志[‡] 北川 貴之[‡] 位野木 万里[†]

工学院大学[†]

東芝デジタルソリューションズ株式会社[‡]

1. はじめに

Digital Transformation の社会実装の加速が求められている[1]. 新規のデジタルビジネスを提案するために、既存の情報システムの要求仕様書の分析は重要である。複雑化した情報化社会では、関連する情報システムは複数存在し、各システムの要求仕様書の規模は膨大になる傾向がある。それらの開発には関わっていない技術者が複雑かつ膨大な要求仕様書を理解し必要な情報を得ることは容易ではないと考えられる。

自然言語処理に基づく要約技術を用いて、要求仕様書を要約する取り組みが研究されている[2]. 自然言語処理技術による要約は、文書中に表れる単語の出現頻度を用いることが多い[3].

要求仕様書には、機能要求と非機能要求が含まれる[4]. 非機能要求の中には、「本システムは 24 時間 365 日稼働しなければならない」等の、1 行の記述であっても、開発において見逃されると、システム運用に多大な影響を与える要求が存在する。非機能要求の記述が、単語の出現頻度に基づいて要約されると、重要な事項が取りこぼされるリスクがある。また、要求仕様書に明示的に定義されていないにもかかわらず、ユーザが実運用する上で暗黙的に求める使い勝手や性能等の非機能要求も存在する。非機能要求の要約では、どの非機能要求がどの程度定義されているのか、またどの非機能要求が未定義なのか、という状況を、定量的かつ簡潔に示した情報が有効であると考えられる。そこで、本稿では、非機能要求の定量要約手法を提案する。

以下、本稿は次のように構成する。2 章では非機能要求の自動要約の課題と解決策へのアプローチを述べる。3 章では、非機能要求の定量要約手法を提案する。4 章では適用評価、5 章では考察、6 章では本稿のまとめを示す。

2. 非機能要求の要約の課題と解決策への アプローチ

要求仕様書中に、非機能要求のタイプ別に、各タイプの非機能要求がどの程度定義されているのかを把握するには、非機能要求の自動分類技術が有効と考えられる。Shreda らは、Word2Vec と Convolutional Neural Network (以下、CNN と呼ぶ) を非機能要求の分類ごとに自動分類を行う手法を提案した[5].

Younas らは大量の学習データを準備する手間を軽減するために、オントロジーを用いて非機能要求に含まれるキーワードの類似度から非機能要求の分類を行う手法を提案した[6].

前者の手法は、一般的な学習モデルを用いるため、非機能要求に特化した場合、分類精度が向上しないリスクがある。後者の Younas らの手法では、定義されたオントロジーの用語は限定的であり、オントロジーを充実させることによりさらに分類精度の向上が期待できる。

ところで、これらの手法は、文書を分類することにとどまり、要求仕様書全体を要約することには対応していない。そこで、本研究では、CNN モデルの構築、オントロジーの融合、オントロジーの拡張を行い、要求仕様書に記述されている非機能要求の自動分類を実現し、その自動分類結果を用いて要求仕様書の非機能要求の記述部分を定量的観点で可視化することで、要約情報として提供することを提案する。

3. 非機能要求分類を活用した定量要約手法

図 1 に提案する非機能要求の自動要約手法の概要を示す。提案する手法では、要求仕様書を与えると、その記述内容に応じて、可用性、運用性、使用性、セキュリティ等の 11 個のタイプに自動分類し、自動分類された結果を定量要約として出力する。要求の記述を非機能要求のタイプ別に分類するには、(a) CNN モデルに基づく分類と、(b) オントロジーに基づく分類の 2 つの方法で分類し、それぞれの結果の合議アルゴリズムを経てタイプを決定する。

3.1 CNN モデルに基づく分類

Zhang らは文書分類に有効な CNN モデルを提案している[7]. 本研究では、あらかじめラベリングされている公開済み学習データを利用して CNN モデルを構築し、そのモデルを用いて、入力された非機能要求を分類する。

3.2 ドメインオントロジーに基づく分類

オントロジーとは、単語を表現する概念、概念間の関係性を定義することで単語間の意味的類似度を得ることができる方法である。オントロジーに定義されたキーワードの同義語は、公開済み学習済み Word2Vec モデルを用いる。

本研究では、Younas ら[6]が定義したキーワードに加え、非機能要求グレード[8]に含まれるキーワードを有識者の意見を元に追加した。

3.3 非機能要求自動分類と定量要約手法

本研究では、非機能要求に対して自動分類精度を向上させるため、ラベリング済非機能要求を学習した CNN 手法とオントロジーによる分類手法を融合する。提案する手法による合議アルゴリズムでは、CNN とオントロジーの双方による分類結果から得られる、非機能要求の各タイプへの確率を用いて、(1) 各タイプへの分類確率を 11 タイプ中の順位に変換し、(2) その順位を用いて、各要求文の各タイプに対するスコアを設定し、(3) 最も高いスコアが設定されたタイプを、対象要求文の分類とする。また、分類結果をグラフ化し、定義数が少ない、または要求が定義されていない非機能要求を把握できるようにする。

Proposal of a Non-functional Requirements Summarization Method using Convolutional Neural Network and Domain Ontology

[†]Yuta Saikawa, Mari Inoki, Kogakuin University

[‡]Takeshi Nagaoka, Takayuki Kitagawa, Toshiba Digital Solutions Corporation

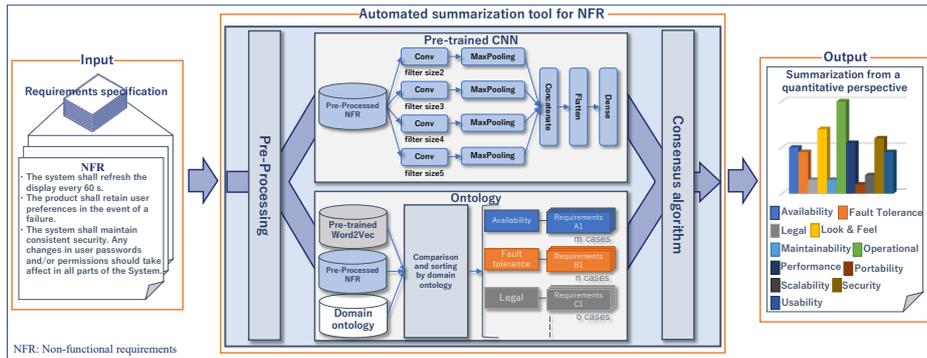


図1 非機能要求の分類手法

4. 適用評価

4.1 評価目的と方法

提案する手法である, CNN とオントロジーの合議による定量要約手法の有効性, 妥当性の確認のため, 事例を用いた評価を行った. 評価に用いたデータは, 「PROMISE Software Engineering Repository Dataset」 [9] (以下, PROMISE と呼ぶ) に含まれるラベリング済み要求文である.

本分類手法を用いて PROMISE に含まれる各要求文を, ケース 1: CNN のみ, ケース 2: オントロジーのみ, ケース 3: CNN とオントロジーの合議の方法により分類し, それぞれの適合率, 再現率, F 値を算出する. 合議手法による分類結果に基づいて非機能要求の定量要約情報を可視化し, 非機能要求の要約としての妥当性を評価する.

4.2 評価結果

本手法を適用した実験結果を表 1 に示す. 表 1 に示す通り, ケース 1 に比べケース 2 では再現率, 適合率が向上した. ケース 3 では CNN だけでは特定の難しい要求をオントロジーで特定することにより, 上位 3 件の正解率が向上した. また本手法を適用し 99 件の要求に対して分類を行った結果各要求の分類数を示した定量要約結果 (図 2) が得られた.

表 1: 実験結果

	ケース 1	ケース 2	ケース 3
再現率	0.343	0.490	0.495
適合率	0.345	0.395	0.389
F 値	0.324	0.332	0.377
正解率	0.464	0.313	0.464
上位 3 件正解率	0.696	0.606	0.747

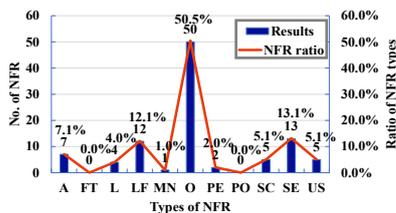


図 2: 非機能要求の定量要約

5. 考察

CNN を用いた場合, 一部のタイプの非機能要求において定義が不均衡になる等の学習データの偏りに起因する分類精度の低下がみられた. CNN とオントロジーとの融合による分類により, 分類精度の向上と, 少ない学習データでも, 定量要約が可能になることが確認された.

図 2 の非機能要求の定量要約結果によれば, 要求仕様書中の Fault Tolerance (FT), Portability (PO)に関する要求

が他のタイプと比較して少ないことがわかった. 実際に有識者により PROMISE の定義状況を確認したところ, 他のタイプの要求と比較し, とくに PO についての定義が不十分であるとの指摘が得られた. 図 2 のように非機能要求の定義状況が可視化されることで, 定義状況が不十分な非機能要求のタイプが指摘しやすくなり, 要求仕様書に内在する非機能要求の定義漏れによるリスク予測に有効と考えられる.

6. まとめ

本稿では, 要求仕様書に定義された非機能要求を CNN とオントロジーを用いて自動分類し, 非機能要求のタイプ別に定義状況を可視化することにより, 定量化による要約情報を自動生成する手法を提案した. 非機能要求の定量化要約は, 定義状況が明確化されるため, 要求仕様書に内在する非機能要求の定義漏れリスクを指摘可能と考えられる. 今後は, 要求仕様書の学習データと, オントロジーで使用しているキーワードの拡張により, 非機能要求の分類精度を高め, 高品質な定量要約情報の自動生成に取り組む.

謝辞

本研究は JSPS 科研費 JP19K11907 の助成を受けた.

参考文献

- [1] The Ministry of Economy, Trade and Industry (METI), DX Report 2 (Interim Report), Compiled as Interim Report by Study Group for Acceleration of Digital Transformation, <https://www.meti.go.jp/press/2020/12/20201228004/20201228004-1.pdf> (参照 2021-10-6).
- [2] A. Ferrari, Natural Language Requirements Processing: from Research to Practice, 2018 ACM/IEEE 40th International Conference on Software Engineering: Companion Proceedings (2018).
- [3] M. Day and C. Chen, Artificial Intelligence for Automatic Text Summarization, 2018 IEEE International Conference on Information Reuse and Integration (IRI), pp.478-484 (2018).
- [4] ISO/IEC/IEEE 29148:2018, Systems and software engineering — Life cycle processes — Requirements engineering (2018).
- [5] Q. Shreda, A. Hanani, Identifying Non-functional Requirements from Unconstrained Documents using Natural Language Processing and Machine Learning Approaches, IEEE Access, vol. 4, pp.1-22 (2016).
- [6] M. Younas, K. Wakil, D. Jawawi, M. Shah, and A. Mustafa, An Automated Approach for Identification of Non-Functional Requirements using Word2Vec Model, International Journal of Advanced Computer Science and Applications, vol. 10, no. 8, pp.539-547 (2019).
- [7] Y. Zhang, B. Wallace, A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentences Classification (2015).
- [8] 独立行政法人情報処理推進機構 社会基盤センター, 非機能要求グレード, <https://www.ipa.go.jp/sec/softwareengineering/reports/20130311.html> (参照 2021-10-03)
- [9] PROMISE Software Engineering Repository Dataset <http://promise.site.uottawa.ca/SERepository/> (参照 2021-04-23)