

日本語の非機能要件の 自動分類手法の提案(11/30)

大東 誠弥

目次

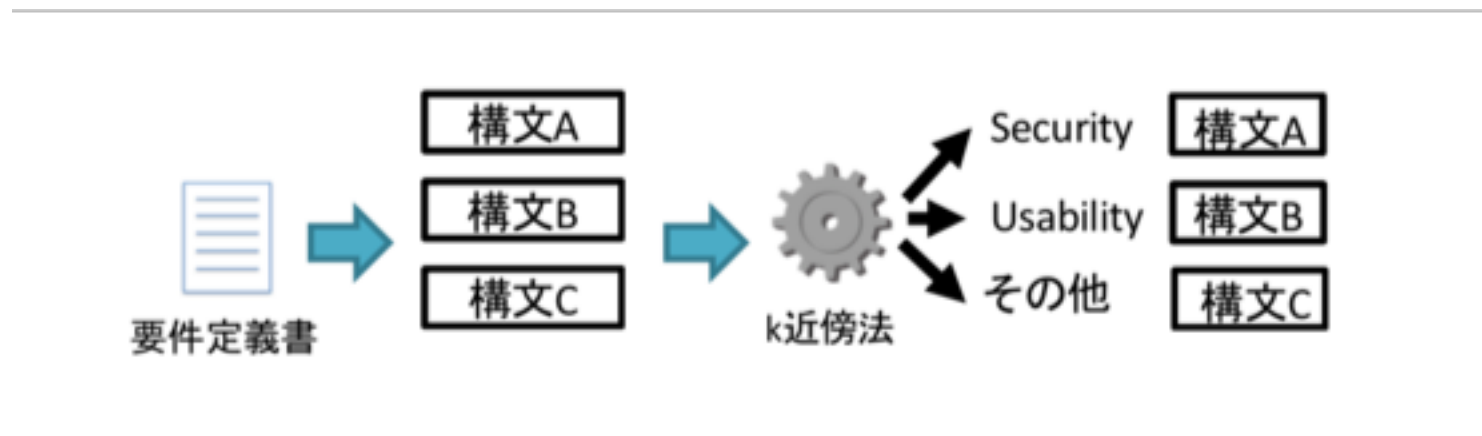
- 背景
- 目的
- 既存研究
- アプローチ
- 進捗状況
- 今後の予定

背景

- 要件定義のうち，非機能要件は書き手によって記述が異なったり，曖昧になったりすることがある.
- 非機能要件を人手で探すのにはコストを要する.

既存研究

- k近傍アルゴリズムを用いて、要件定義書(英語)内の非機能要件を分類している. [1]



- [1] John Slankas and Laurie Williams, Automated extraction of non-functional requirements in available documentation, NaturaliSE '13, page 9-16

目的

- 日本語の要件定義書から，非機能要件を抽出するのに適した，テキストマイニング手法を提案することを目的とする.

アプローチ

- 既存研究の分類手法を用いて，日本語の非機能要件の分類精度を調査する.
- 日本語の非機能要件での分類の精度が最も高い分類アルゴリズムを調査する.
- 企業ごとに分類の精度が異なるのかを調査する.

品質モデル(非機能要件の分類)

Functional Suitability (機能適合性)

Performance Efficiency (性能効率性)

Compatibility (互換性)

Usability (使用性)

Reliability(信頼性)

Security (セキュリティ)

Maintainability(保守性)

Portability(移植性)

* ISO25010に基づく分類

分類手法

1. 要件定義書より，非機能要件文を抽出し品質モデルごとにラベルをつける.
2. 抽出した要件文を品詞ごとに分割し，特定の品詞のみのリストを作る.
3. 特定の品詞のリストをBag Of Words(BOW)の行列に変換.
4. 変換したリストを訓練データとテストデータに分ける.
5. 訓練データとテストデータを元に最近傍法($K=1$)で分類.

分類手法

①要件定義書より，非機能要件文を抽出し品質モデルごとにラベルをつける．

電子データの送受信を行う際には電子署名や暗号化を用いることで偽造等から保護することが可能であること	セキュリティ
システム稼働後のデータの整合性・完全性を担保するためのマスタデータメンテナンス手法を構築すること	セキュリティ
個人情報へのアクセス状況の記録及び一定期間の保存並びにアクセス記録の定期的かつ迅速な分析を可能にするとともに、アクセス記録の改ざん、窃取又は不正な消去の防止等に必要な対応を講ずること。	セキュリティ
また、個人情報への不適切なアクセス監視のため、個人情報を含む又は含むおそれがある一定量以上のデータダウンロード時の警告表示機能や、当該警告表示を定期的に確認する機能等、個人情報の漏えいを防止するための必要な対応策について講ずること	セキュリティ
次期ハローワークシステムの機密性を確保するために、以下の要件を満たすこと。	セキュリティ
データが第三者や権限のないユーザに参照されないよう、データベースのアクセス権限を正しく設定し、必要に応じて暗号化を実施すること	セキュリティ
受託者は、以下の要件を満たし、本番稼働後の将来的な業務環境の変化に柔軟に対応できる拡張性の高いシステムを構築すること。	移植性
次期ハローワークシステムに必要な性能を維持するための処理能力の拡張性を確保するために、以下の要件を満たすこと。	性能効率性
次期ハローワークシステムにおいては、4(2)「システム方式に関する事項」に示した疎結合アーキテクチャの導入を前提とした上で性能の拡張性を確保すること	性能効率性

分類手法

②抽出した要件文を品詞ごとに分割し，特定の品詞のみのリストを作る．

(例)応答時間はX秒とする



Mecabで品詞ごとに分割

応答 / 時間 / は / X / 秒 / 以内 / と / する

名詞/名詞/助詞/名詞/名詞/名詞/助詞/動詞



名詞のみを抽出

応答 / 時間 / X / 秒 / 以内 /

分類手法

③特定の品詞のリストをBug Of Words(BOW)の行列に変換.

あらかじめ, 出現する単語で辞書を作成しておく.
[応答, セキュリティ, 時間, システム, . . .]

辞書と照らし合わせてその文に出現する単語を数字の列で表現する.

[1, 0, 1, 0, . . .]

分類手法

④変換したリストを訓練データとテストデータに分ける.

正解となるラベルがきちんと割り振られるように、訓練データ 7 割, テストデータ 3 割に分割する.

分類手法

⑤訓練データとテストデータを元に最近傍法($K=1$)で分類.

BOW行列をもとに，出現する単語の並びが最も近いもののラベルが正解であると考ええる.

分類に使用するデータ(富士通)

- 要件定義書および調達仕様書 8 件から，非機能要件と思われるものを**554**文抽出.

機能適合性	2 1 件
性能効率性	8 8 件
互換性	3 1 件
使用性	5 8 件
信頼性	3 2 件
セキュリティ	1 4 5 件
保守性	1 4 7 件
移植性	3 2 件

分類結果(富士通)

訓練データ 7 割, テストデータ 3 割で10回試行した平均を算出.

	Precision	Recall
機能適合性	0.408	0.150
性能効率性	0.491	0.641
互換性	0.257	0.144
使用性	0.540	0.723
信頼性	0.546	0.230
セキュリティ	0.663	0.757
保守性	0.683	0.580
移植性	0.561	0.430

分類結果(富士通)

訓練データ 7 割, テストデータ 3 割で10回試行した平均を算出.

	Precision	Recall
機能適合性	0.408	0.150
性能効率性	0.491	0.641
互換性	0.257	0.144
使用性	0.540	0.723
信頼性	0.546	0.230
セキュリティ	0.663	0.757
保守性	0.683	0.580
移植性	0.561	0.430

現時点での問題点

- 現段階では、非機能要件の中で各品質モデルの分類を行っていたが、要件定義書全体での精度を研究の目的にそぐわない。

今後の予定

- 残り 2 社(東芝, 日立)の分類データを用意し, それぞれの企業ごとに分類結果を出す.
- 分類にはどの品詞を用いるか, ストップワードはどう設定するか等を決める.
- 分類器を変えて比較するため, 分類器ごとの特徴を調べる.
- 参考文献を増やす.