

進捗報告(10/9)

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

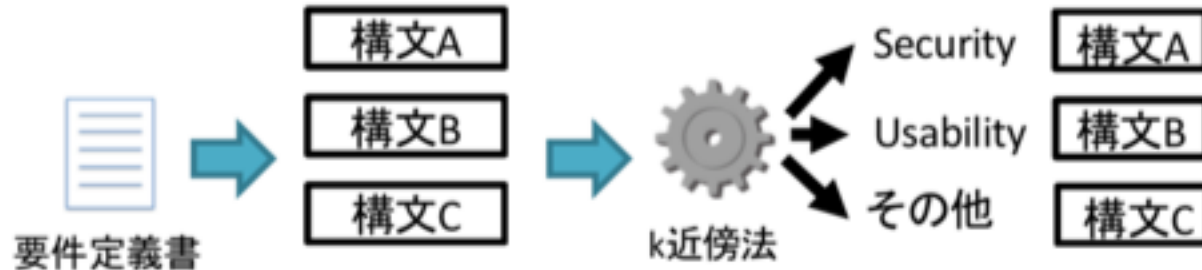
- 要件定義のうち，非機能要件は書き手によって記述が異なったり，曖昧になったりすることがある.
- 非機能要件を人手で探すのにはコストを要する.

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.

既存研究

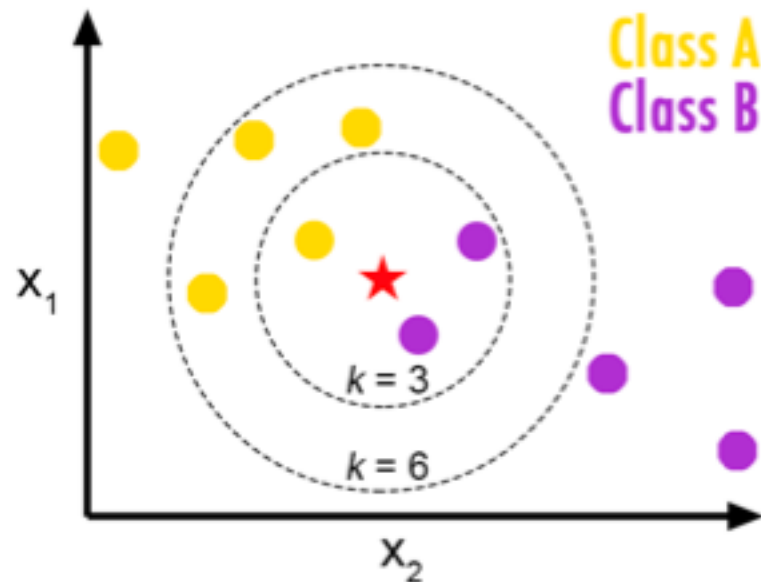
- k近傍アルゴリズムを用いて、要件定義書(英語)内の非機能要件を分類している. [1]



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

k近傍アルゴリズム

- 調べたいデータが、どのクラスに属するかを分類するアルゴリズム.
- 距離が近いデータをk個取得し、多数決によって属するクラスを分類する.



レーベンシュタイン距離

- 文字を置換，挿入，削除した回数を距離とする。
(例) **l**ight → **r**ight (置換1回 距離は 1)

(応用例) **I** have an apple **yesterday**.

→ **You** have an apple.

(置換1回，削除1回 距離は 2)

アプローチ

- 既存研究の手法を用いて，日本語版での分類の精度を調査する.
- 日本語版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.

進捗状況

日本語の要件定義書10件を， 8つの品質特性に分類済み。

70	受託者以外の者が市場で調達することが可能であるサービスであること。	使用性		
71	受託者以外の者が市場で調達することが可能である製品であること。	使用性		
72	アプリケーションプログラムの動作、性能等に支障を来たさない範囲において、オープンソースソフトウェア(OSS)製品(ソースコードが無償で公開され、改良や再配布を行うことが誰に対しても許可されているソフトウェア製品)の活用を図ること。	不明		
73	受託者は、利用するソフトウェアはサポート期間を考慮して選定し、ソフトウェアベンダーによるサポート又は他の事業者によるサポートサービスを必ず受けること。	不明		
74	本システムの特性及び基本方針書を踏まえ、効率的に設計・開発及び保守可能な手法を採用すること。	保守性		
75	本システムの利用者は、本省職員及び地方厚生局等職員合わせて約 910 名である。	機能適合性		
76	本システムを利用する拠点数は、本省 1 か所、地方厚生局等 47 か所及び地方厚生局 分室 2 か所である。	機能適合性		
77	検索、参照、登録、更新及び削除に係る処理については、業務の繁忙期においても 平均処理応答時間(レスポンスタイム):3 秒以内を実現可能とすること。	性能効率性		
78	導入するハードウェアの選定に当たっては、過剰な設備投資により必要以上の性能を確保することなく、適切なハードウェアを導入すること。	性能効率性		
79	平均処理応答時間(レスポンスタイム):3 秒以内は、通常運用時に実現可能とし、縮退運転時はこの限りではない。	性能効率性		
80	可用性に係る指標は「稼働率」として目標値を 99.5%とする。	信頼性		
81	障害発生時等の可用性を確保するため、本システムを構成する各サーバ等を適切な方式で冗長化し、迅速かつ円滑な切り替えによる業務遂行を可能とすること。	信頼性		
82	データの紛失や改ざんからデータを保護し、データの正確性及びデータの一貫性を保証すること。	セキュリティ		

今後の予定

- 既存手法を用いて，日本語の非機能要件を分類するための前処理をどのようにすれば良いかを考える．(日本語にレーベンシュタイン距離を使うのには無理がある)