

日本語非機能要件の
自動分類における
教師あり学習アルゴリズムの評価

発表者 池内未来

要件定義

- 実装すべき機能や満たすべき性能などを明確にしていく作業
- 機能要件
 - 実装する機能の詳細
- 非機能要件
 - 要求されるシステムの性能

背景

- 非機能要件は機能要件に比べて、曖昧なので見逃されやすい
 - システムの質の低下

先行研究

- 3つの教師あり学習アルゴリズムを用いて、非機能要件を分類
[1]
- Naive Bayes, k近傍, SVM
- 英語の非機能要件に関する分類
- 日本語の非機能要件においても同様の結果になるとは限らない

[1] J. Slankas and L. Williams, “Automated extraction of non-functional requirements in available documentation,” *NaturaLiSE '13*, pp.9–16, 2013.

目的

- データセットの数を増やし，日本語の非機能要件に適した教師あり学習アルゴリズムの評価

分類の流れ

1. 形態素解析
2. Bag of Wordsに変換
3. 分類

形態素解析

- MeCabを用いて文書を品詞ごとに分割
 - 辞書はmecab-ipadic-NEologdを使用

```
私はラーメンが好きです
私      名詞,代名詞,一般,* ,* ,* ,私,ワタシ,ワタシ
は      助詞,係助詞,* ,* ,* ,*,は,ハ,ワ
ラーメン 名詞,一般,* ,* ,* ,*,ラーメン,ラーメン,ラーメン
が      助詞,格助詞,一般,* ,* ,* ,*,が,ガ,ガ
好き    名詞,形容動詞語幹,* ,* ,* ,*,好き,スキ,スキ
です    助動詞,* ,* ,*,特殊・デス,基本形,です,デス,デス
EOS
```

形態素解析の出力例

Bag of Words

- 文書ごとの単語の出現回数を行列で表す

A. 私はラーメンが好きです

B. 私は餃子が好きです

	私	は	ラーメン	が	好き	です	餃子
A	1	1	1	1	1	1	0
B	1	1	0	1	1	1	1

分類

- 教師あり学習アルゴリズムごと
 - Naive Bayes
 - k近傍($k=1$)
 - SVM

データセット

品質特性ごとの要件数

	厚生労働省(13件)	独立行政法人(1件)	合計
機能適合性	23	0	23
性能効率性	130	5	135
互換性	15	1	16
使用性	49	14	63
信頼性	80	5	85
セキュリティ	116	13	129
保守性	119	8	127
移植性	57	5	62

実験

- 訓練データ7割, テストデータ3割
- Precision, Recall, F1値で精度を評価
- 試行回数は100回

結果

各アルゴリズムの結果(FI値)

	Naive Bayes	k近傍	SVM
機能適合性	0.135	0.307	0.377
性能効率性	0.686	0.569	0.717
互換性	0.038	0.187	0.143
使用性	0.342	0.352	0.595
信頼性	0.314	0.470	0.510
セキュリティ	0.432	0.609	0.687
保守性	0.309	0.535	0.570
移植性	0.317	0.416	0.523

今後の予定

- データ件数が少ないので、データを増やす