

日本語非機能要件の自動化における 学習データサイズの影響

オープンソースソフトウェア工学研究室

60216013
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する作業

＊機能要件

＊実装する機能の詳細

＊非機能要件

＊システムの性能や信頼性，セキュリティに関する要件

非機能要件の問題

- ◆機能要件と比べて曖昧な表現になりやすい
 - *判断が難しい
- ◆機能要件の記述と混在することがある
 - *目視で探す場合手間がかかる

先行研究・目的

◆先行研究[1]

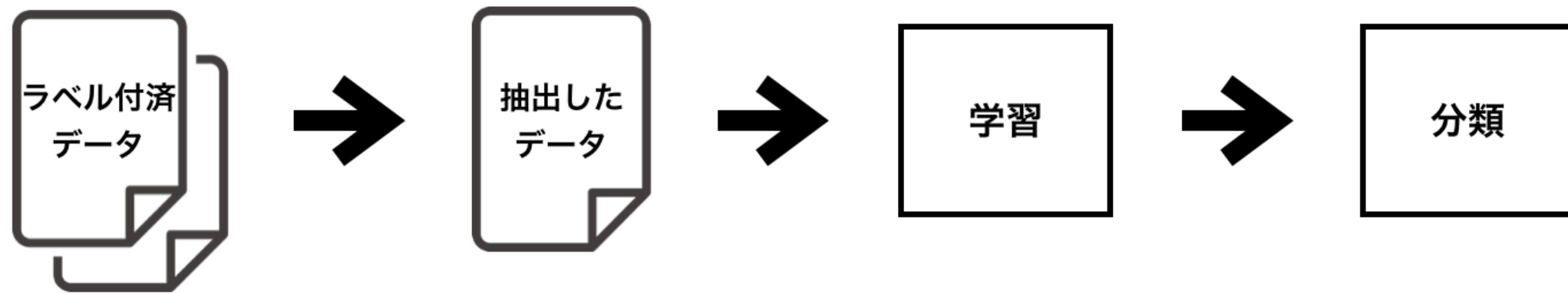
- *日本語非機能要件の分類を行っている
- *データサイズが分類器に与える影響が示されていない

◆目的

- *データサイズが分類器に与える影響の調査

[1]:大東 誠弥, 福井 克法, 宮崎 智己, 大平 雅雄, "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," ソフトウェア・シンポジウム 2019 in 熊本, pages 36--44, 2019年6月.

実験概要

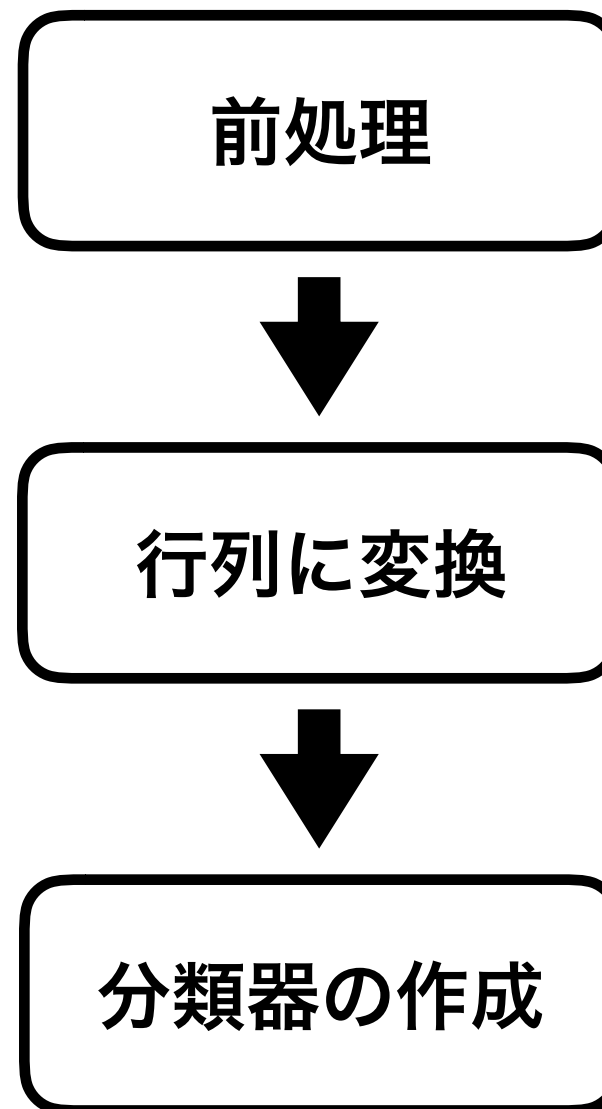


実験の流れ

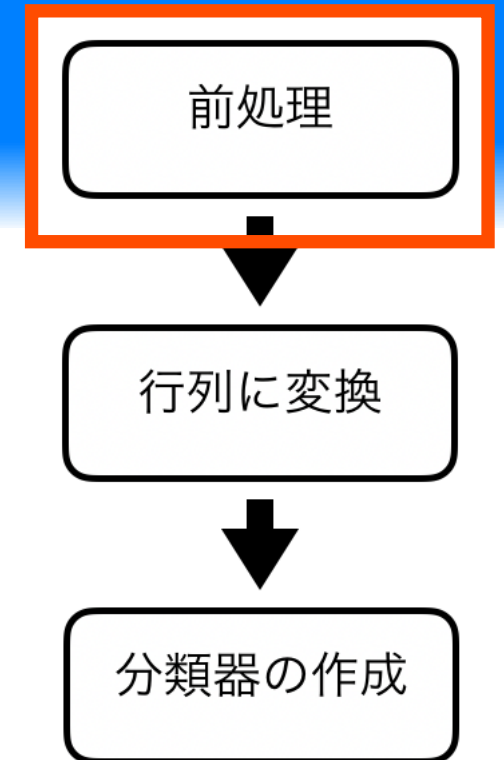
- ◆5分割交差検証を行う
- ◆図の操作を20回繰り返し、計100回の平均で評価
- ◆Precision, Recall, F値を用いて分類精度を評価

分類の流れ

◆以下の流れで非機能要件の分類を行う



前処理



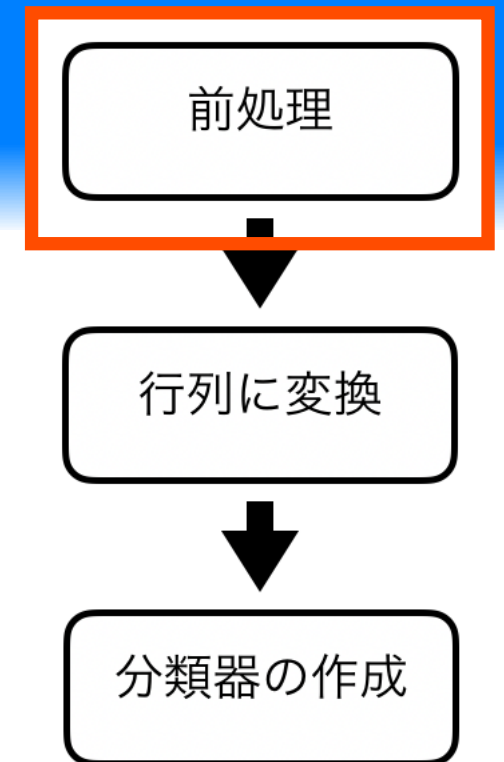
- ◆ アルファベットの小文字化

- ◆ 数字を全て0に置換

 - *例：2020年1月1日 → 0年0月0日

- ◆ 箇条書きの削除

前処理



◆形態素解析

＊単語が意味を持つ最小の単位に分割

＊形態素解析にはMeCabを使用

＊辞書にはmecab-ipadic-NEologdを使用

脆弱性の対策を行う	
脆弱性	名詞,固有名詞,一般,*,*,*,脆弱性,ゼイジャクセイ,ゼイジャクセイ
の	助詞,連体化,*,*,*,*,の,ノ,ノ
対策	名詞,サ変接続,*,*,*,*,対策,タイサク,タイサク
を	助詞,格助詞,一般,*,*,*,を,ヲ,ヲ
行う	動詞,自立,*,*,五段・ワ行促音便,基本形,行う,オコナウ,オコナウ

形態素解析の例

行列変換

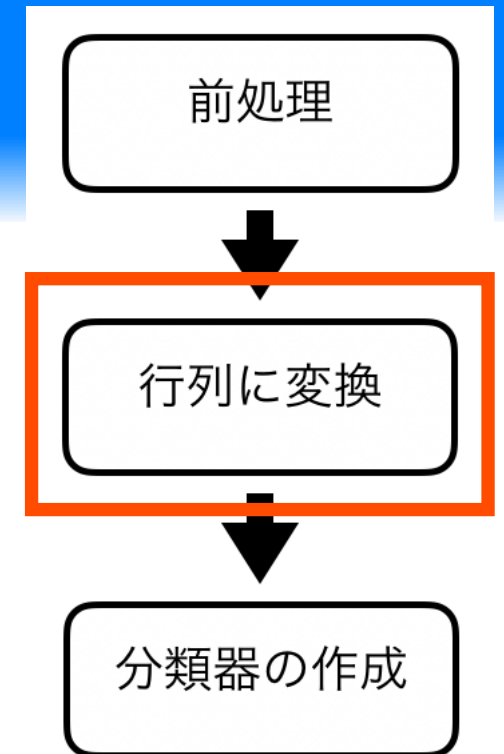
◆Bag of Words を用いて行列に変換

*単語の出現回数を特徴とする

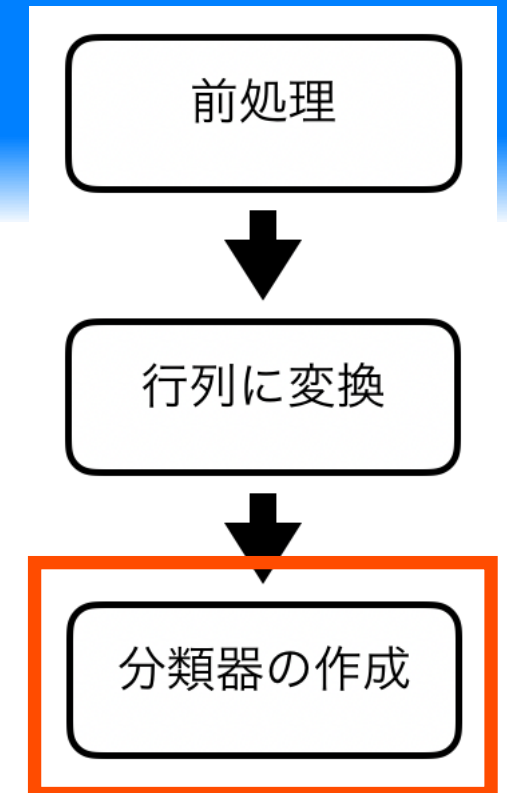
A.私はラーメンが好きです

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



分類器の作成



◆ナイーブベイズ

*ベイズの定理に基づいた確率的分類手法

◆k-近傍($k=1$)

*特徴空間における最も近いデータに基づいた分類手法

◆SMO

*2クラス分類を行うための識別境界を求める分類手法

データセット

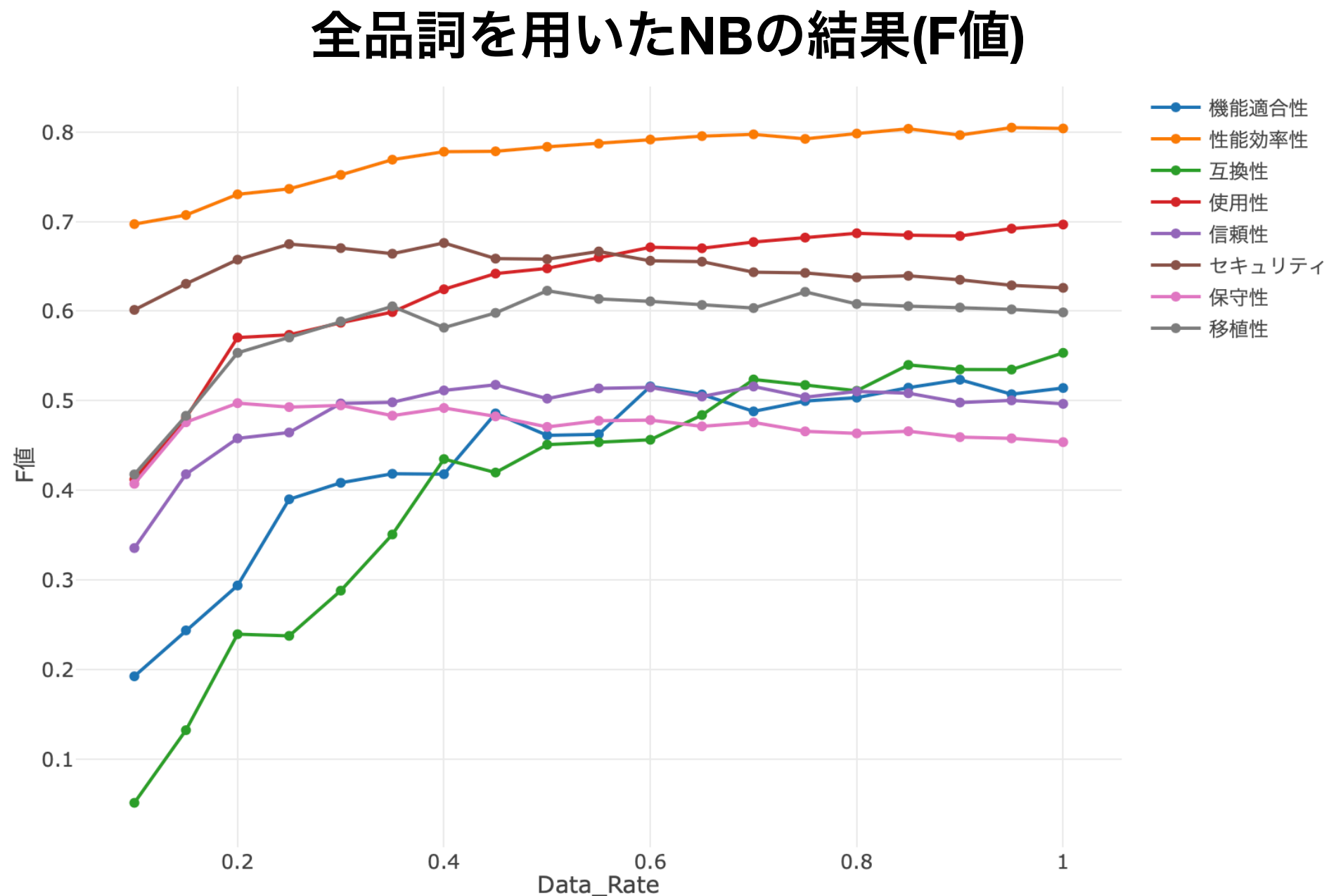
◆厚生労働省が公開

品質特性ごとの要件数

品質特性	元のラベル数	追加ラベル数	合計
機能適合性	22	68	90
性能効率性	126	641	767
互換性	13	60	73
使用性	36	240	276
信頼性	75	231	306
セキュリティ	107	456	563
保守性	110	391	501
移植性	46	168	214
その他	2497	6972	9469

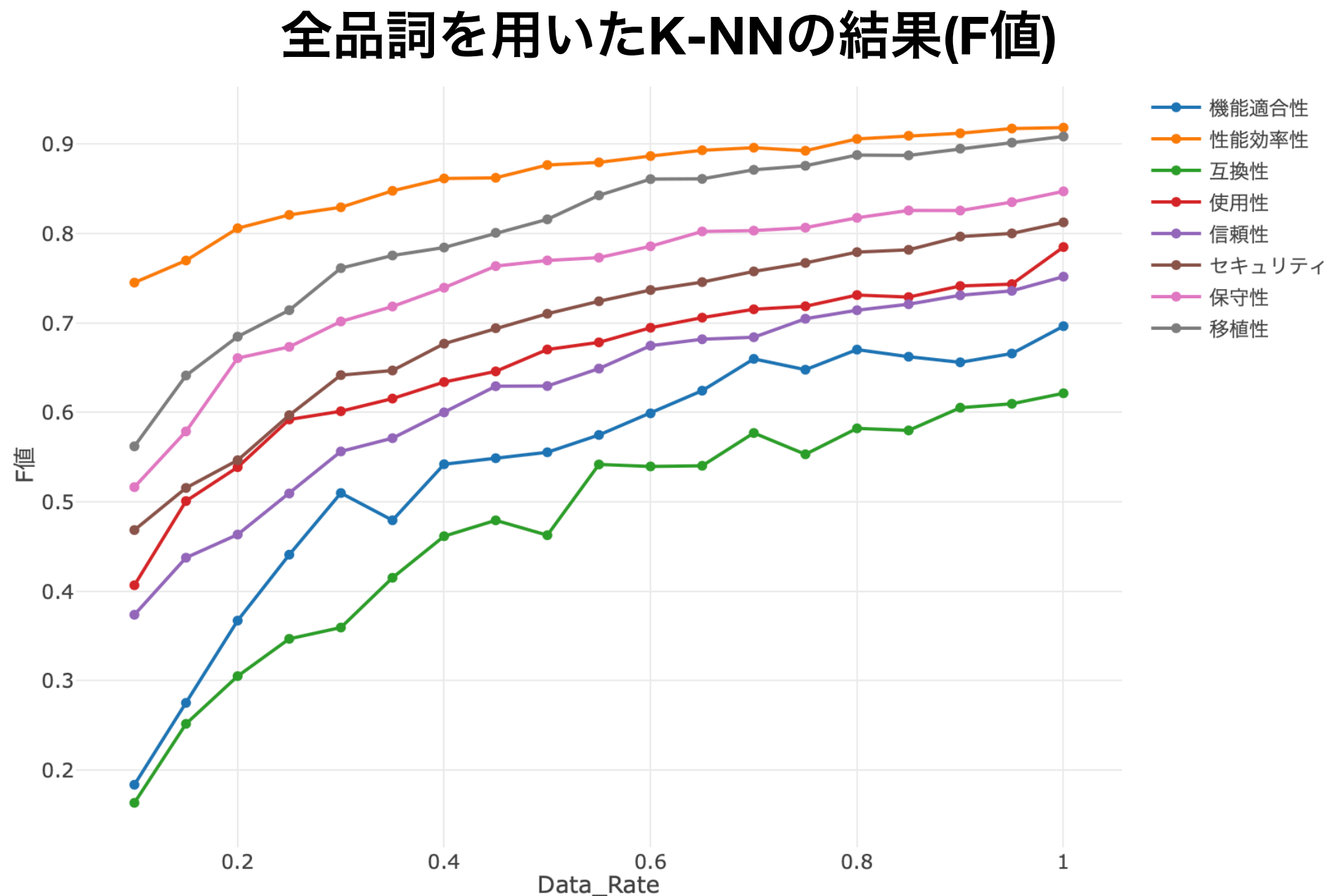
結果

◆データを増やせば精度が下がる場合もある



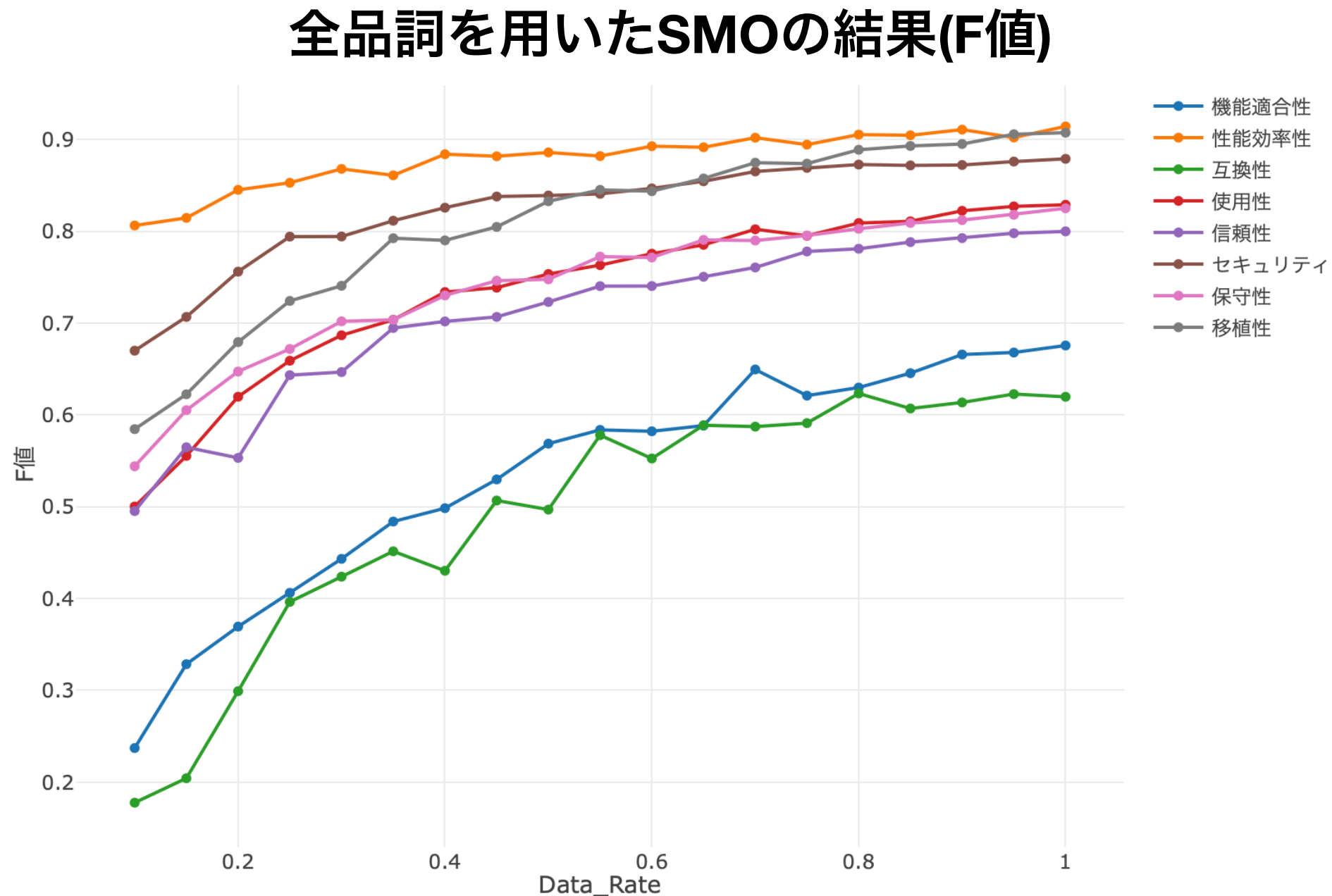
結果

◆データを増やせば精度が上がる



結果

◆データを増やせば精度が上がる



今後の予定

- ◆結果の考察
- ◆関連研究を探す