

Revisiting the Impact Classification Techniques on the Performance of Defect Prediction Models

Baljinder Ghotra, Shane McIntosh, Ahmed E. Hassan
Software Analysis and Intelligence Lab (SAIL) School of Computing,
Queen's University, Canada
{ghotra, mcintosh, ahmed}@cs.queensu.ca

60171033 Ayaka Tsukiyama

背景

- ❖ 欠陥予測モデルには様々な分類手法がある
- ❖ **Lessmannの研究(2008年)**
 - 22種類の分類手法のうち、17種類の分類手法でそれぞれお互いにその性能について統計的な違いが見つからなかった
 - **NASAのコーパスを使用**
- ❖ **他の研究でも欠陥予測モデルはどの分類手法を使ってもその性能にあまり差がないと示している**

問題点

NASAコーパスについて

- ❖ ノイズ(誤り)が含まれている: **Shepperd**の研究(2013)
 - ノイズによって誤った結果が示されているのではないか
- ❖ 一つの組織内で独自に開発されたもの
 - 違う設定で開発されたもの(例:OSS)では違う結果になるのではないか

実験に使ったデータそのものが正しくない可能性

目的

- ❖ 分類手法によって生成される欠陥予測モデルの性能に差があるか正しく評価する
- ❖ RQ1
 - いろいろな分類手法の欠陥予測モデルはクリーンなNASAコーパスを使用した時でも同様な結果になるか
 - もとのノイズが含まれているNASAコーパスと対比して、整形されたものをクリーンと表現(2013年、Shepperdが提供)
- ❖ RQ2
 - いろいろな分類手法の欠陥予測モデルはOSSを使用した時でも同様な結果になるか

使用する欠陥予測モデルの分類手法

1. 統計的手法
2. クラスタリング手法
3. ルールベース手法
4. ニューラルネットワーク
5. 近傍法
6. SVM
7. 決定木
8. アンサンブルメソッド

この分類手法を
順位付けする

TABLE I
OVERVIEW OF THE STUDIED CLASSIFICATION TECHNIQUES.

Family	Technique	Abbreviation
Statistical Techniques	Naive Bayes	NB
	Simple Logistic	SL
Clustering Techniques	K-means	K-means
	Expectation Maximization	EM
Rule-Based Techniques	Repeated Incremental Pruning to Produce Error Reduction	Ripper
	Ripple Down Rules	Ridor
Neural Networks	Radial Basis Functions	RBFs
Nearest Neighbour	K-Nearest Neighbour	KNN
Support Vector Machines	Sequential Minimal Optimization	SMO

Decision Trees	J48	J48
	Logistic Model Tree using Logistic Regression	LMT
Ensemble Methods using LMT, NB, SL, SMO, and J48	Bagging	Bag+LMT, Bag+NB, Bag+SL, Bag+SMO and Bag+J48
	Adaboost	Ad+LMT, Ad+NB, Ad+SL, Ad+SMO and Ad+J48
	Rotation Forest	RF+LMT, RF+NB, RF+SL, RF+SMO, and RF+J48
	Random Subspace	Rsub+LMT, Rsub+NB, Rsub+SL, Rsub+SMO, and Rsub+J48

アプローチ

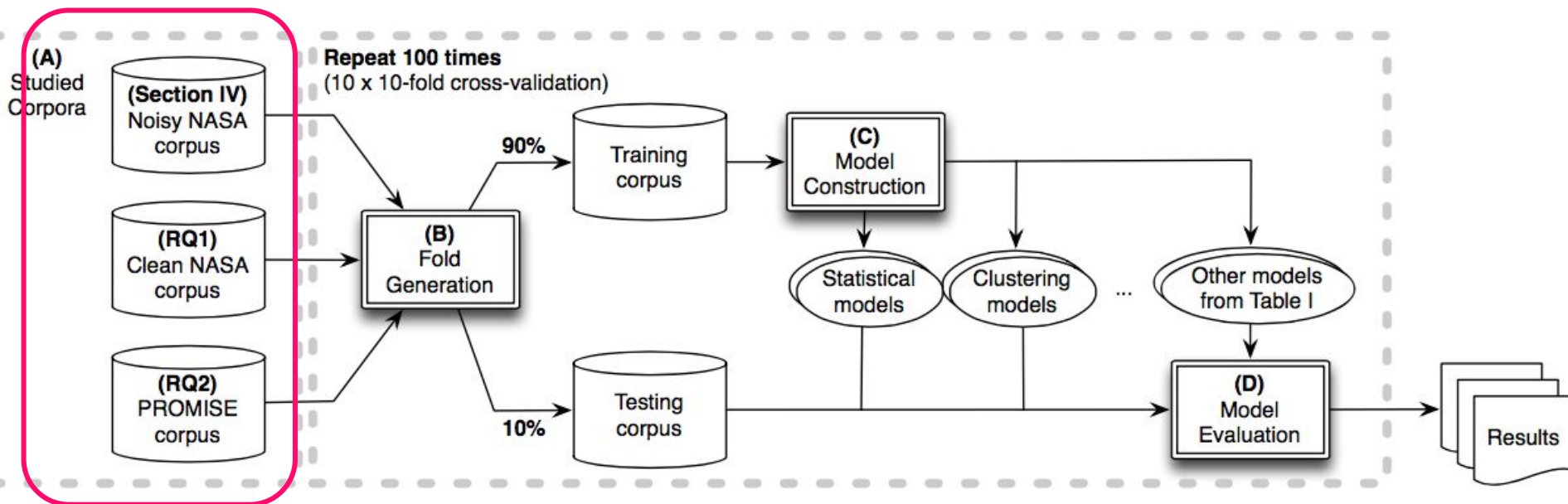


Fig. 1. An overview of our model construction and evaluation approach.

- ❖ A: データセット
 - ノイズを含むNASAコーパス
 - 整形されたクリーンなNASAコーパス
 - OSSなど整形された上で公開されているデータセット

TABLE II
AN OVERVIEW OF THE STUDIED CORPORA.

Corpus	Dataset	No. of Modules	Defective	Defective (%)
Noisy NASA (Section IV)	CM1	505	48	9.5
	JM1	10,878	2,102	19.3
	KC1	2,107	325	15.4
	KC3	458	43	9.3
	KC4	125	61	48.8
	MW1	403	31	7.6
	PC1	1,107	76	6.8
	PC2	5,589	23	0.4
	PC3	1,563	160	10.2
	PC4	1,458	178	12.2
Clean NASA (RQ1)	CM1	327	42	12.8
	JM1	7,720	1,612	20.8
	KC1	1,162	294	25.3
	KC3	194	36	18.5
	MW1	250	25	10
	PC1	679	55	8.1
	PC2	722	16	2.2
	PC3	1,053	130	12.3
	PC4	1,270	176	13.8
PROMISE (RQ2)	Ant 1.7	745	166	22.3
	Camel 1.6	965	188	19.5
	Ivy 1.4	241	16	6.6
	Jedit 4	306	75	24.5
	Log4j 1	135	34	25.2
	Lucene 2.4	340	203	59.7
	Poi 3	442	281	63.6
	Tomcat 6	858	77	8.9
	Xalan 2.6	885	441	46.4
	Xerces 1.3	453	69	15.2

アプローチ

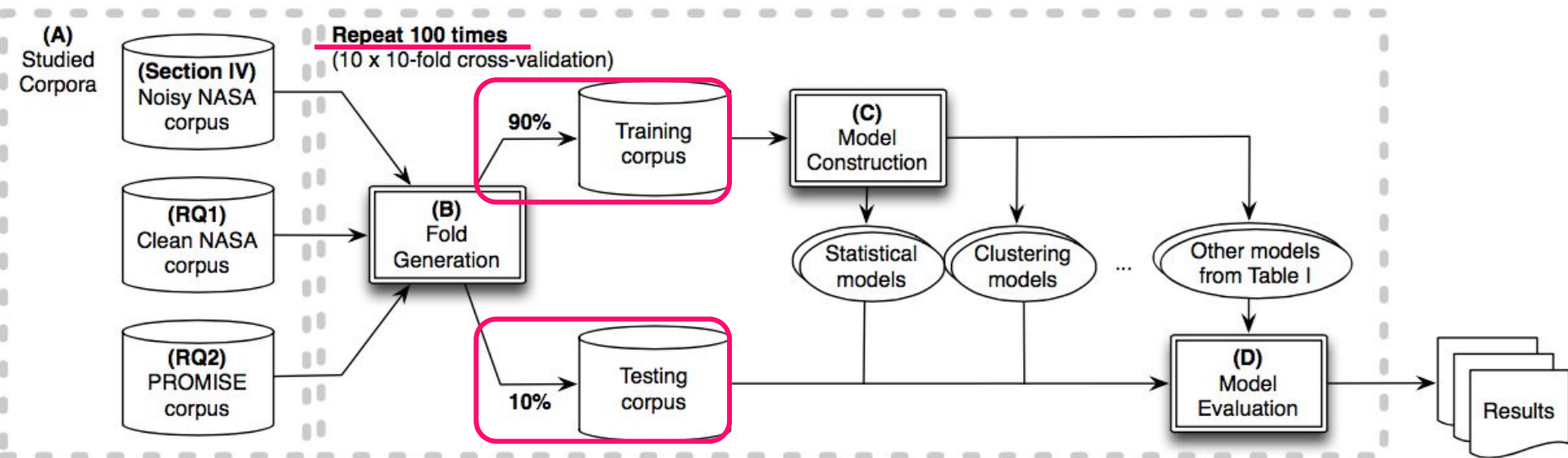


Fig. 1. An overview of our model construction and evaluation approach.

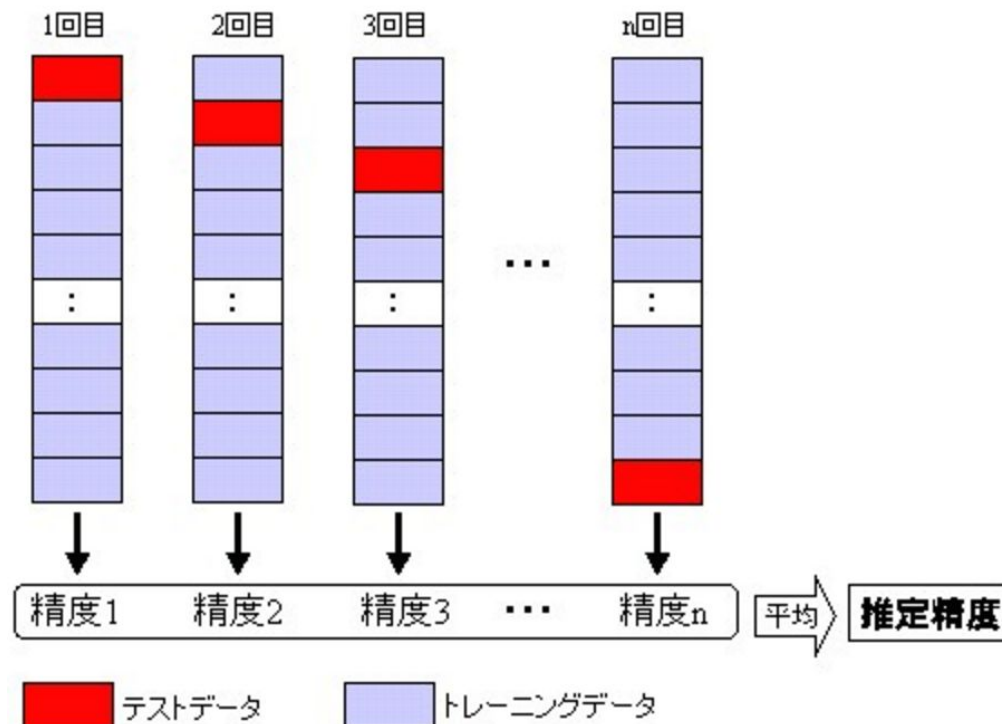
❖ B: 交差検証

- 標本データを分割し、その一部をまず解析して、残る部分でその解析のテストを行い、解析自身の妥当性の検証・確認に当てる手法

交差検証

❖ 機械学習

- データをランダムに10分割
- 9つをトレーニングデータ、1つをテストデータ



アプローチ

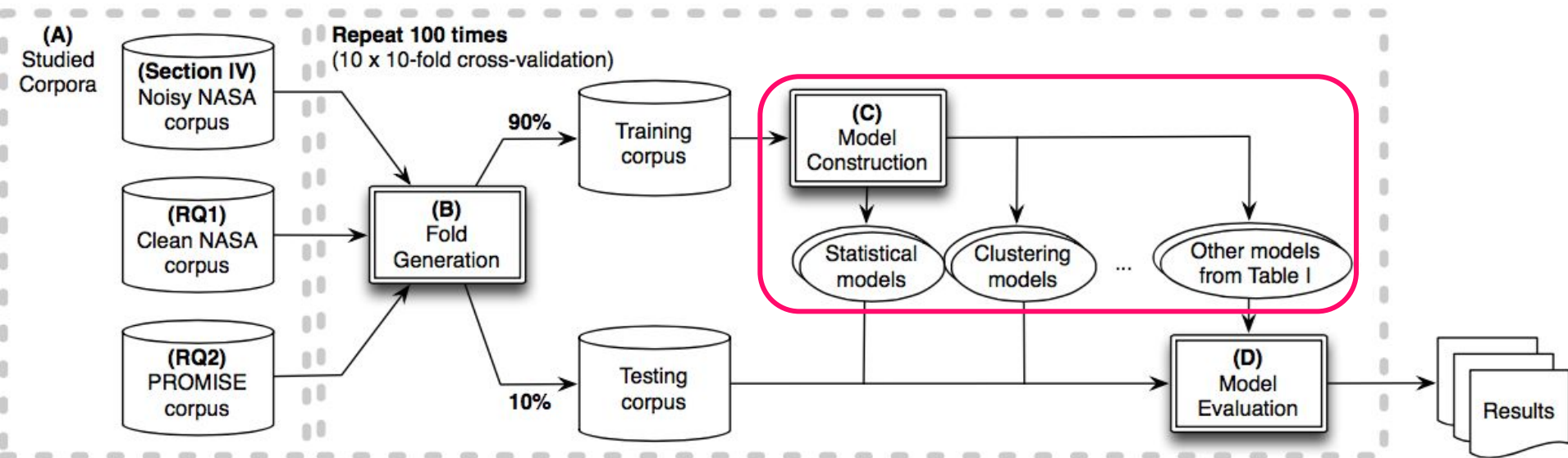


Fig. 1. An overview of our model construction and evaluation approach.

❖ C: モデル構築

➤ それぞれの分類手法でモデルを構築

アプローチ

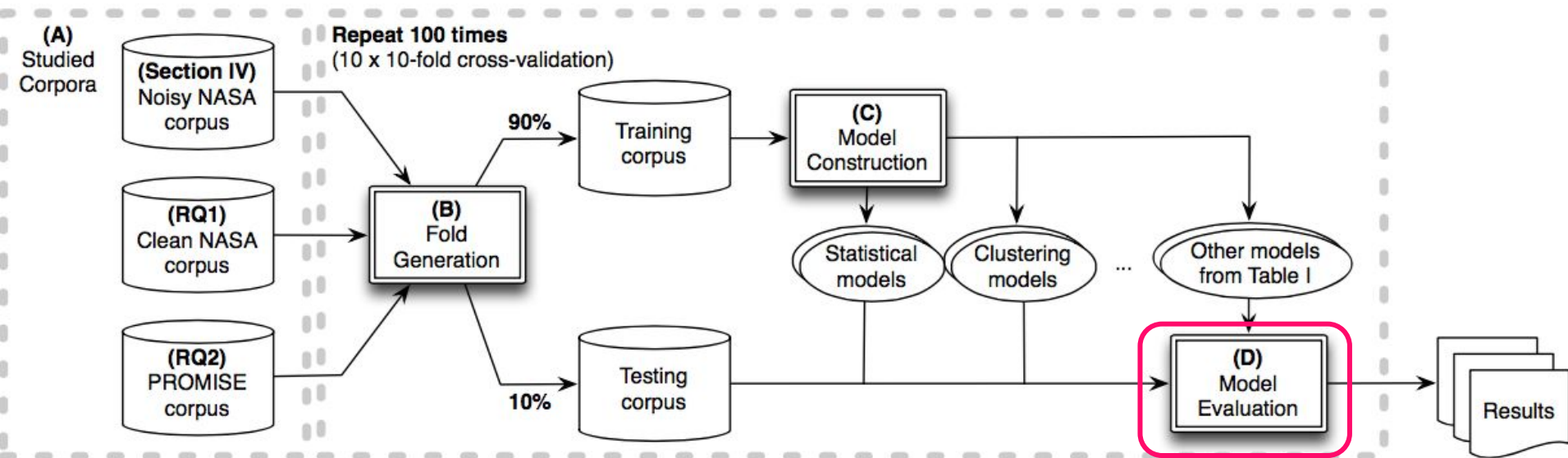


Fig. 1. An overview of our model construction and evaluation approach.

❖ D: モデル評価

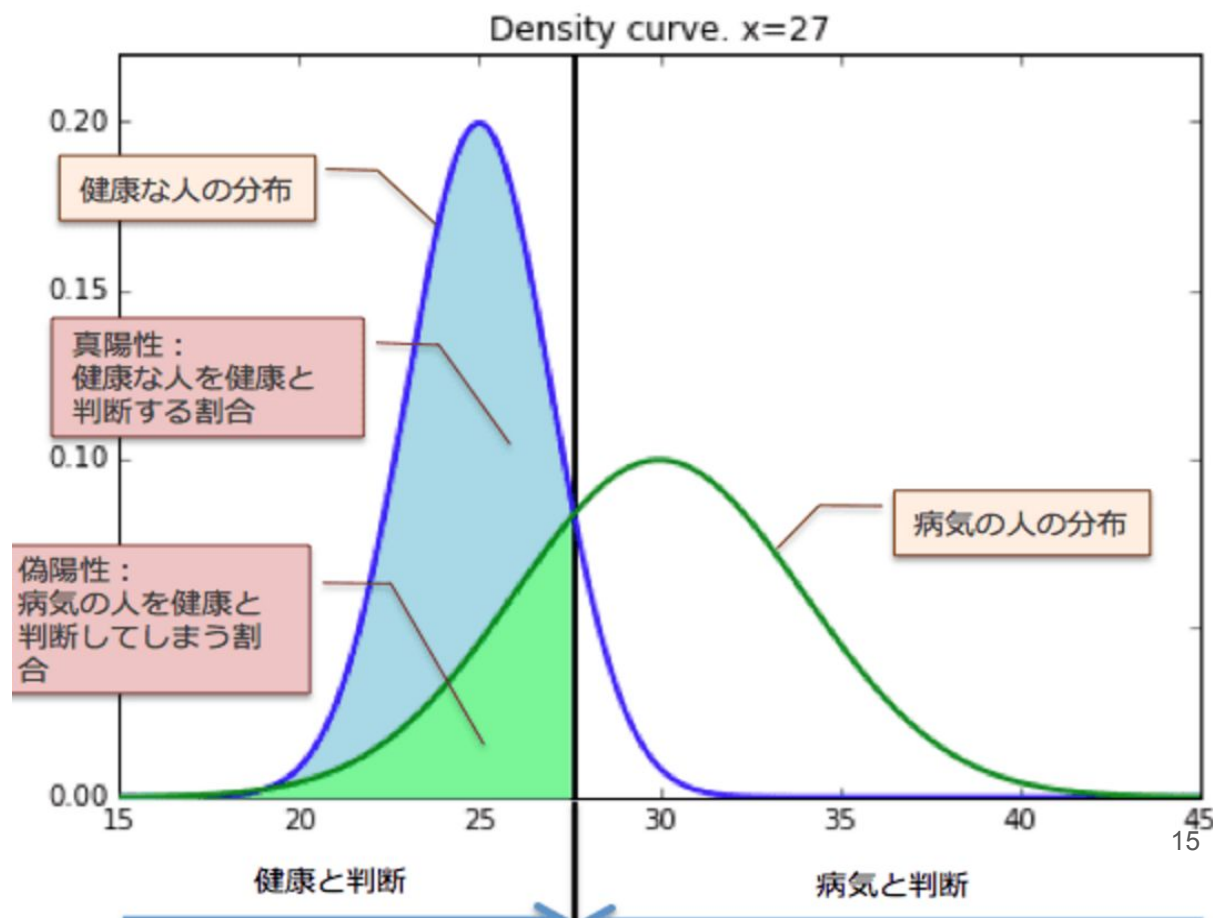
- AUC(値が0.5以上でランダム予測より優れている)
- Scott-Knott test.

Scott-Knott test

- ❖ 統計的にグループをランク分けする
 - ❖ AUCの値をもとにまず2つに分ける
 - ❖ それを再帰的に繰り返す
-
- ❖ 今回は4つにランク分けする

AUCとは

- ❖ 分類のアルゴリズムの性能の良さ
- ❖ 0から1まで
- ❖ ランダムな分類の場合は0.5



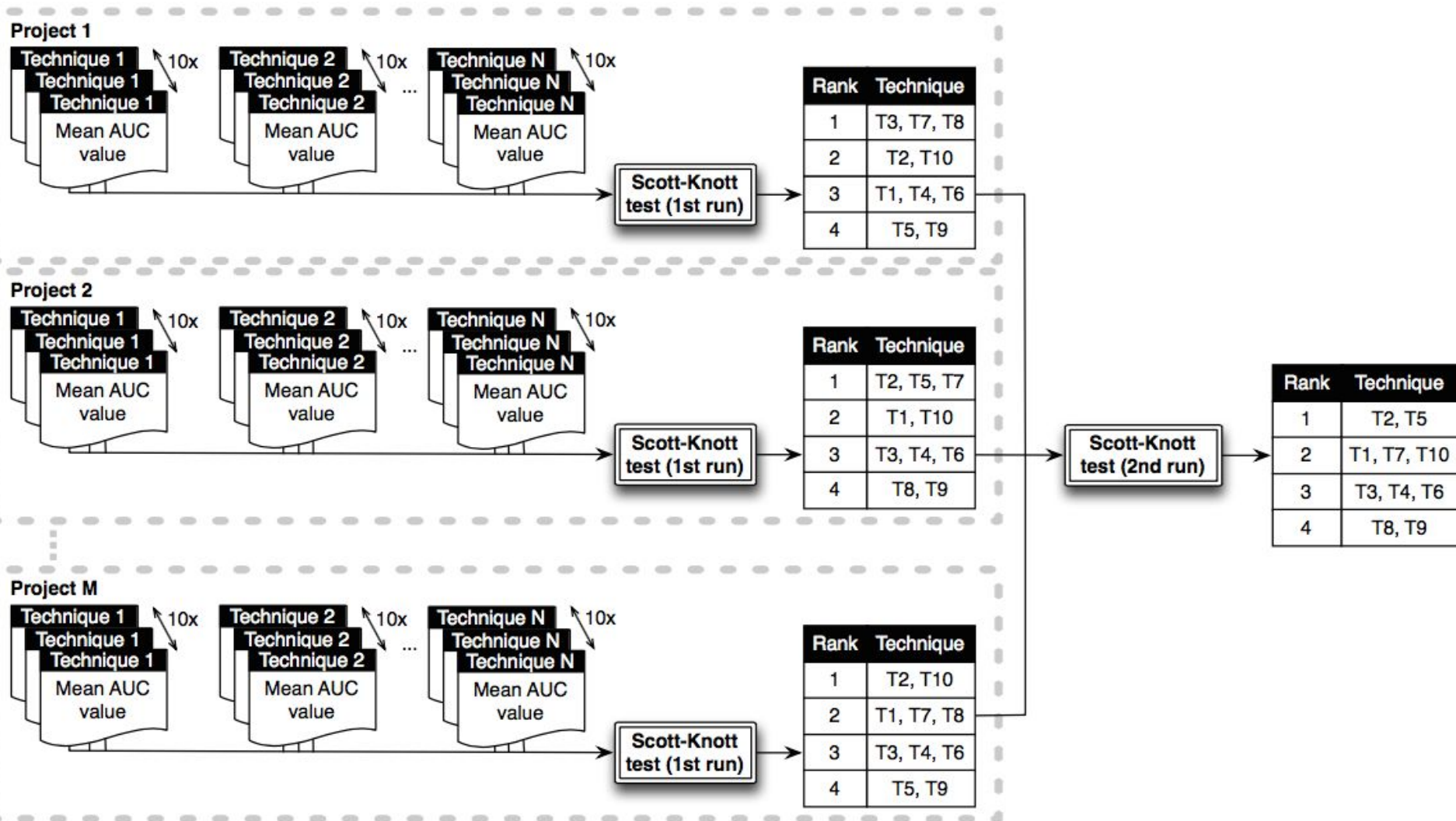
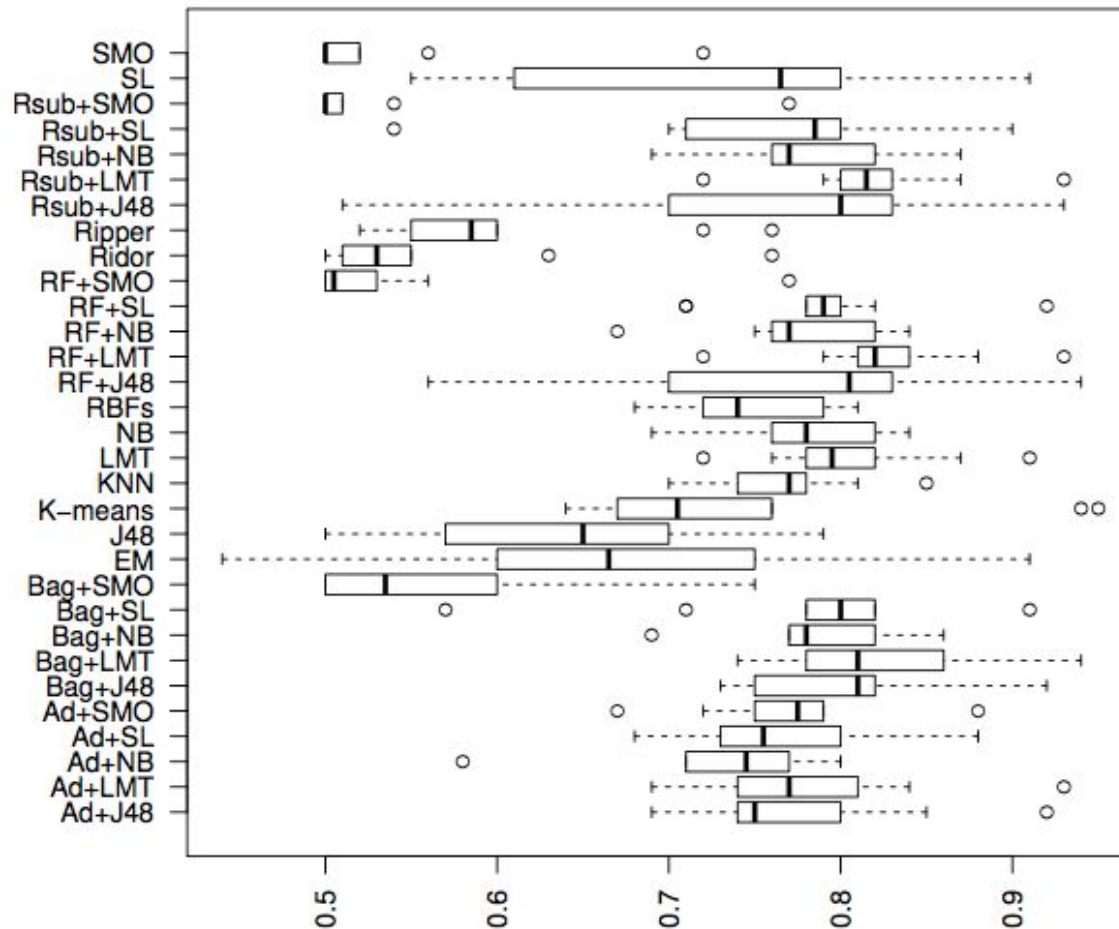


Fig. 2. Our approach for ranking classification techniques using a double Scott-Knott test.

結果(1/3)

- ❖ ノイズが含まれているNASAコーパス
 - 31種類中、24種類で統計的に区別がつけられなかった



結果(2/3)

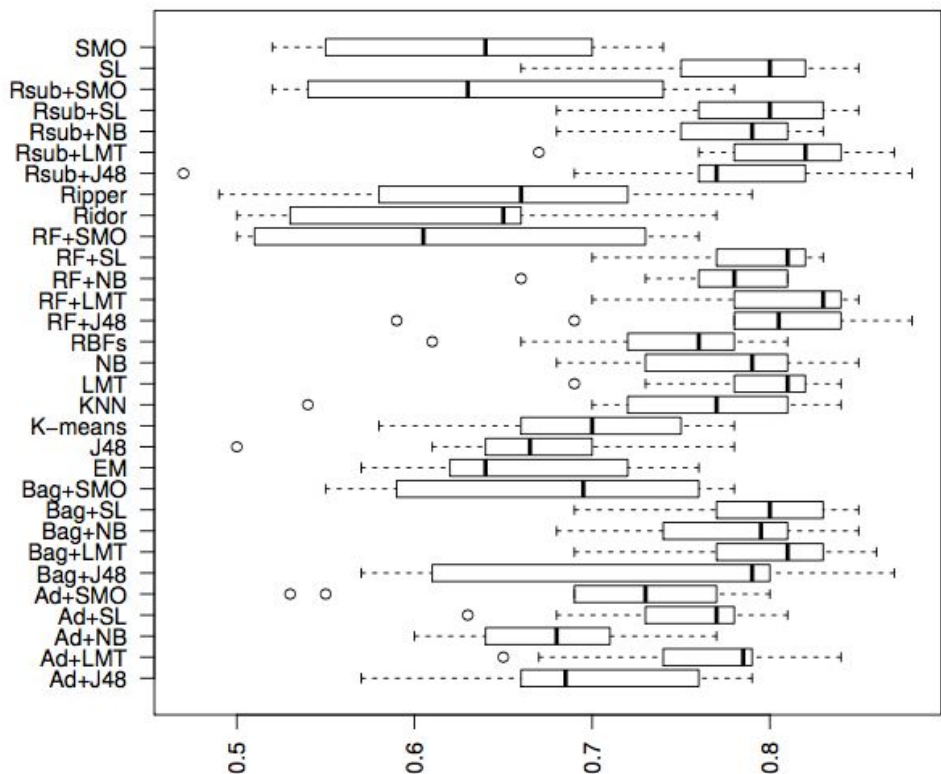
❖ 整形されたNASAコーパス

- アンサンブルメソッド、決定木、統計的手法 >
- 近傍法、ニューラルネットワーク >
- ルールベース法、クラスタリング法、SVM
- LMT(分類木)や(simple logistic:統計的手法)とアンサンブルメソッドを組み合わせたものが一番精度が高い

結果(3/3)

❖ PROMISEデータセット

- 決定木、統計的手法、ニューラルネットワーク、k近傍法
- ルールベース、クラスタリング法、SVM
- LMT(分類木)や(simple logistic:統計的手法)と他の分類手法を組み合わせたものが一番精度が高い



まとめ

- ❖ NASAコーパスとPROMISEコーパスで共通のパターンが観測できた
- ❖ Scott-Knott testについて
 - LMT(決定木)とアンサンブルメソッドを組み合わせた時がどちらも性能が高い
 - クラスタリング法、ルールベース法、SVMは性能が低い

所感

- ❖ オリジナルのNASAコーパスにノイズが多いのはなぜなのか疑問に思った
- ❖ 分類手法など専門用語が多かった(TABLE 1の表は勉強になりそう)