MWS Cup 2019 ポストミーティング
課題3 解説

株式会社ＦＦＲＩ
https://www.ffri.jp

2019/12/09

作問担当

問題作成・採点

押場 博光 (株式会社FFRI)

中川 恒 (株式会社FFRI)

茂木 裕貴 (株式会社FFRI)

末吉 大輝 (株式会社FFRI)

アジェンダ

テーマ

課題解説

分析

テーマ

課題3について

今回の課題3のテーマは**表層解析**

昨年までは動的解析がテーマ

- Cuckoo Sandboxの動的解析ログを用いた問題が中心

データセットの活用につなげたい

- FFRI Dataset 2018からは表層情報になっている

課題3について

今回の課題3のテーマは**表層解析**

昨年までは動的解析がテーマ

- Cuckoo Sandboxの動的解析ログを用いた問題が中心

データセットの活用につなげたい

- **FFRI Dataset 2018からは表層情報になっている**

表層情報について

表層解析データは非常に便利

検体を動かさなくても入手可能

- 解析コストが比較的低い場合が多く大量の検体の解析が可能
- 実行時のコンテキストを意識しなくてよい
  - データセットの拡張・分析がしやすく非常に使いやすい
- マルウェアを実行前に検知し得る

但し，**情報量は劣る**

表層情報について

表層解析データは非常に便利

検体を動かさなくても入手可能

- 解析コストが比較的低い場合が多く大量の検体の解析が可能
- 実行時のコンテキストを意識しなくてよい
  - データセットの拡張・分析がしやすく非常に使いやすい
- マルウェアを実行前に検知し得る

但し，**情報量は劣る**

最大限活用してもらいたい

課題解説

課題構成

集計・分析問題

マルウェア・良性ファイル分類問題

表層情報によるマルウェア検知の検討をイメージ

配点

設問	問題	配点
1	主張の正誤判定	3
2	impfuzzy衝突数の集計	2
3	原因の考察	2
4	データの分析	3
5	マルウェア・良性ファイル判定	15

課題 3-1

1. 以下の記述を読み、正しいものにすべてチェックをつけなさい。

- [ ] デジタル署名が付与されているマルウェアの占める割合は、マルウェア全体の 5% 以上も存在する。
- [ ] デジタル署名が付与されている良性ファイルの占める割合は、良性ファイル全体の 35% 以上である。
- [ ] ファイルごとに"strings"に含まれる文字列の平均長を計算する。文字列の平均長が 7 以下のマルウェアのサンプル数は、同じ条件の良性ファイルのサンプル数の 2 倍以上である。

課題 3-1 意図

何が特徴になり得るかのイメージを掴んでもらう

ここで挙げている情報は分類の特徴として利用可能

- ただし，これだけで勝てる特徴にはなっていない
- 3-5でこれを活用できれば，5点ほど獲得できる

スクリプト等で簡単に検証，活用してもらう想定

課題 3-1 解答

1. 以下の記述を読み、正しいものにすべてチェックをつけなさい。

- [ ] デジタル署名が付与されているマルウェアの占める割合は、マルウェア全体の 5% 以上も存在する。
- [x] デジタル署名が付与されている良性ファイルの占める割合は、良性ファイル全体の 35% 以上である。
- [x] ファイルごとに"strings"に含まれる文字列の平均長を計算する。文字列の平均長が 7 以下のマルウェアのサンプル数は、同じ条件の良性ファイルのサンプル数の 2 倍以上である。

課題 3-1 解答

1. 以下の記述を読み、正しいものにすべてチェックをつけなさい。

- [ ] デジタル署名が付与されているマルウェアの占める割合は、マルウェア全体の 5% 以上も存在する。
- [x] デジタル署名が付与されている良性ファイルの占める割合は、良性ファイル全体の 35% 以上である。
- [x] ファイルごとに含まれる文字列の平均長を計算する。文字列の平均長が…ファイルのサンプル…

マルウェアでも署名付与は0ではない (これは非常に重要)
一方で，明確に付与されている数には差がある

課題 3-1 解答

1. 以下の記述を読み、正しいものにすべてチェックをつけなさい。

- [ ] デジタル署名が付与されているマルウェアの占める割合は、マルウェア全体の 5%
- [x] デジタ… 良性ファイル全体…

パッカーなどの影響で差が生じる

- [x] ファイルごとに"strings"に含まれる文字列の平均長を計算する。文字列の平均長が 7 以下のマルウェアのサンプル数は、同じ条件の良性ファイルのサンプル数の 2 倍以上である。

課題 3-2, 3-3, 3-4

2. マルウェアと良性ファイルでそれぞれ impfuzzy の集合を考えたとき、その積集合のハッシュ値に含まれているサンプル数を答えなさい。

回答を入力

3. マルウェアと良性ファイルで impfuzzy が衝突する理由について簡潔に説明しなさい。

回答を入力

4. PEiD に含まれるシグニチャの内容から、impfuzzy の衝突が起きる原因を一つ取り上げ説明しなさい。

回答を入力

課題 3-2, 3-3, 3-4 意図

特性を理解する必要があることを体得してもらう

今回は題材として，impfuzzy※を取り上げた

- impfuzzyだけでマルウェアと良性ファイルに分類するのは難しい
- この理由の分析を通じて感覚を身に着けてもらう

※ https://blogs.jpcert.or.jp/ja/2016/05/impfuzzy.html

課題 3-3 impfuzzy

impfuzzyとは？

JPCERT/CCがリリースしたファジーハッシュ

- 簡単に言えばインポートテーブルを対象にssdeepを計算する
- マルウェアファミリの分類において効果が示されている

https://blogs.jpcert.or.jp/ja/2016/05/impfuzzy.html

課題 3-3 解答

impfuzzyとは？

JPCERT/CCがリリースしたファジーハッシュ

- 簡単に言えば**インポートテーブルを対象にssdeepを計算**する
- マルウェアファミリの分類[illegible]を効果が示されている

インポートテーブルが一致すればimpfuzzyも一致する

https://blogs.jpcert.or.jp/ja/2016/05/impfuzzy.html

課題 3-4 意図

impfuzzyとは？

JPCERT/CCがリリースしたファジーハッシュ

- 簡単に言えば**インポートテーブルを対象にssdeepを計算**する
- マルウェアファミリの分類[illegible]を効果が示されている

インポートテーブルが一致すればimpfuzzyも一致する

これが今回のデータで起きているのはなぜ？（課題 3-4）

https://blogs.jpcert.or.jp/ja/2016/05/impfuzzy.html

課題 3-4 想定解答

インポートテーブルが一致する可能性がある状況

- パッカー
- .NET Framework
- NSIS Installerなどのインストーラー形式
- …

様々な要因で発生し得る
(妥当性がありそうなら〇)

課題 3-3, 3-4

補足

impfuzzyの評価

次に、提案手法とimphash、ssdeepの三者を用いてマルウエアの類似性を評価する比較実験の結果を示します。

比較実験では、20種類のマルウエアについて、それぞれ異なる10検体ずつ（合計200検体）を用意しました。200検体から2検体を選ぶすべての組合せについて、3つの方法で類似度を計算して、類似度が30以上なった場合に同じマルウエアと判定しているものとみなしました。
なお、各検体はパックされている場合、アンパックしてから各手法を適用しています。
図1に比較実験の結果を示しています。
なお、今回の比較実験では、異[illegible]マルウエアだとする誤判定は、いずれの方式でも皆無でした。

なお、各検体はパックされている場合、アンパックしてから各手法を適用しています。

https://blogs.jpcert.or.jp/ja/2016/05/impfuzzy.html

JPCERT/CCの評価においてもこういった特性は考慮されている
- マルウェアファミリの分類とマルウェア・良性ファイル分類は違うタスクであることにも注意が必要

impfuzzyの有効性に関する話ではなく，タスク・特性を理解して利用しましょうというお話

課題 3-5

分類問題

マルウェア・良性ファイル分類問題

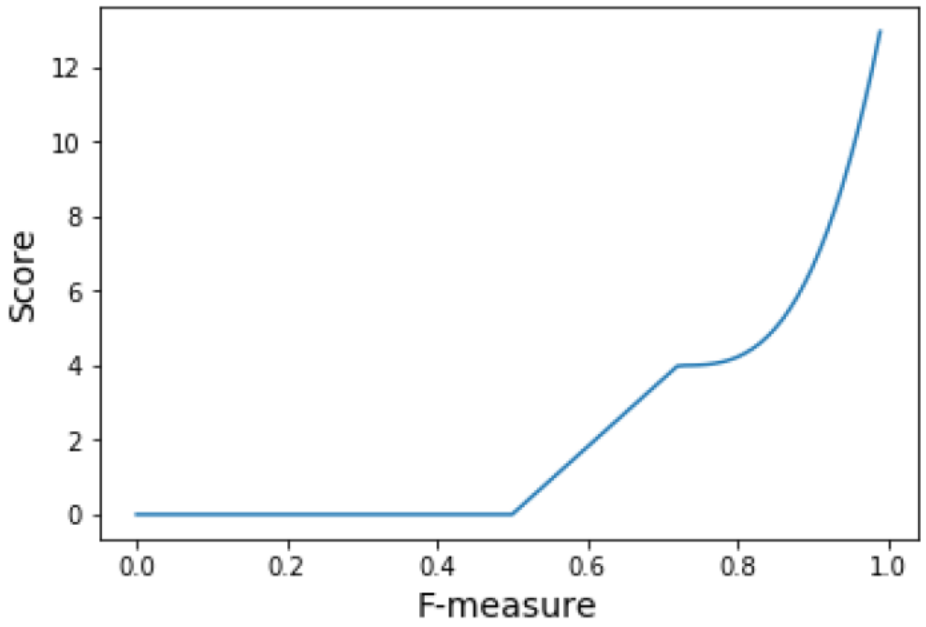
- 前回に引き続き出題

提出してもらった分類結果の精度ベースで評価

課題 3-5 分類結果の評価 意図

Weighted F-Measureを基に評価

Weighted F-Measureの値を下記のように点数化



課題 3-5 分類結果の評価 (参考)

Weighted F-Measure

$$F_\beta = (1 + \beta^2)\frac{\mathrm{precision} \cdot \mathrm{recall}}{\beta^2 \cdot \mathrm{precision} + \mathrm{recall}}$$

Weighted F-Measure の特徴

$\beta < 1$ Precision重視のスコアに

$\beta \to 0$ Precisionで評価することと等価に

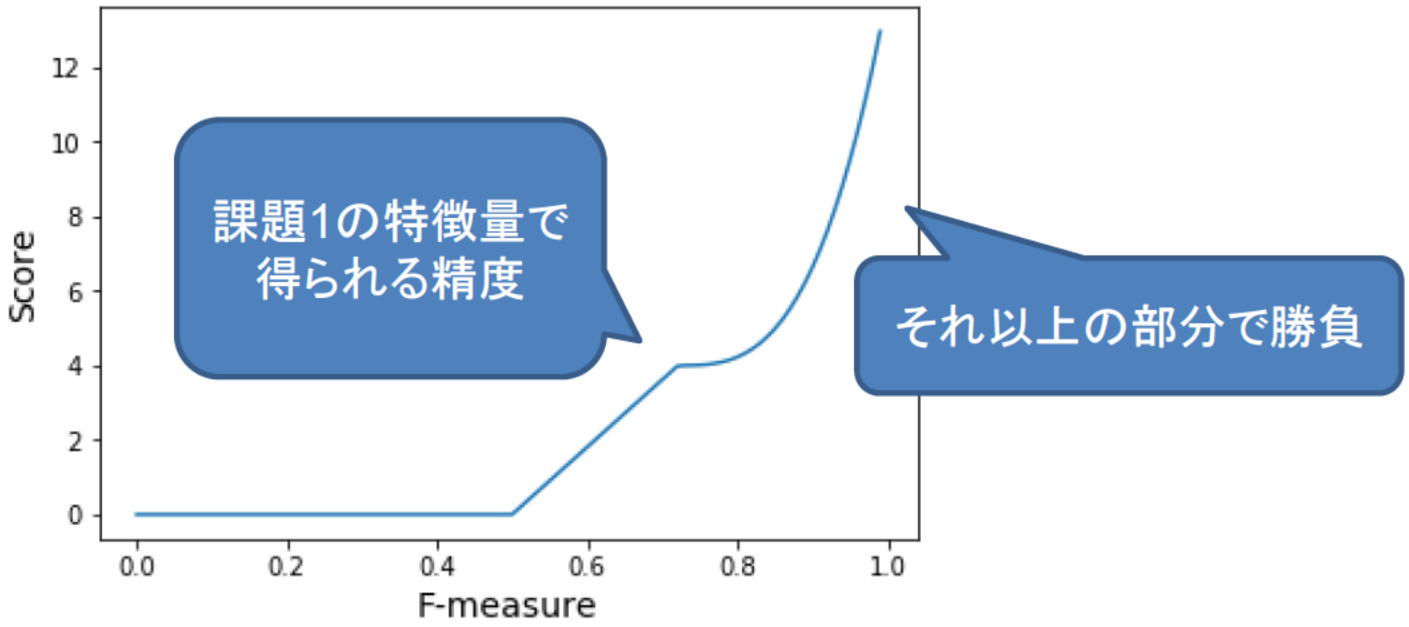
$\beta \to \infty$ Recallで評価することと等価に

今回は
β=0.7で評価

課題 3-5 分類結果の評価 意図

Weighted F-Measureを基に評価

Weighted F-Measureの値を下記のように点数化

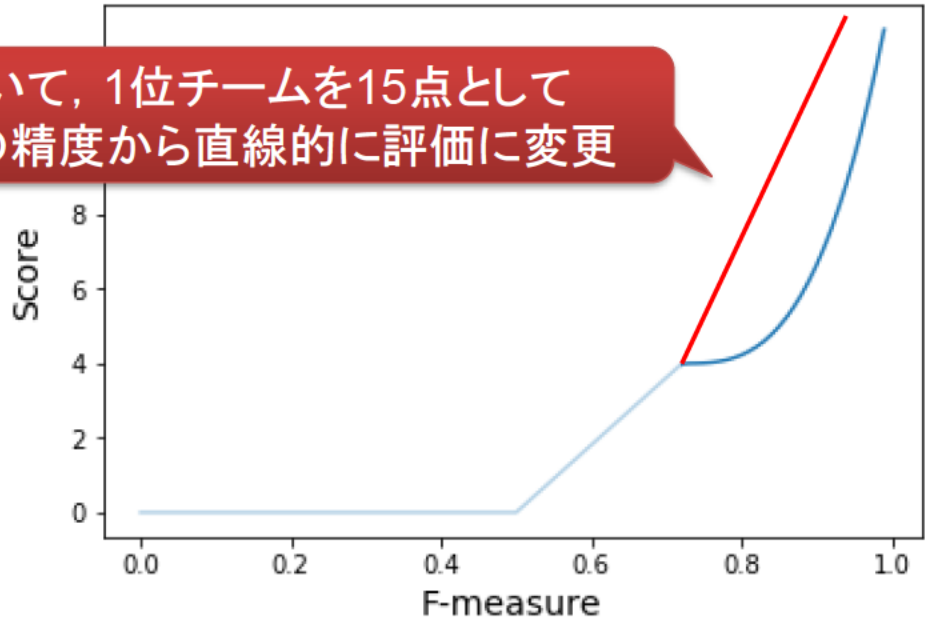


課題 3-5 分類結果の評価 調整

Weighted F-Measureを基に評価

Weighted F-Measureの値を下記のように点数化

この部分について，1位チームを15点として
課題1ベースでの精度から直線的に評価に変更

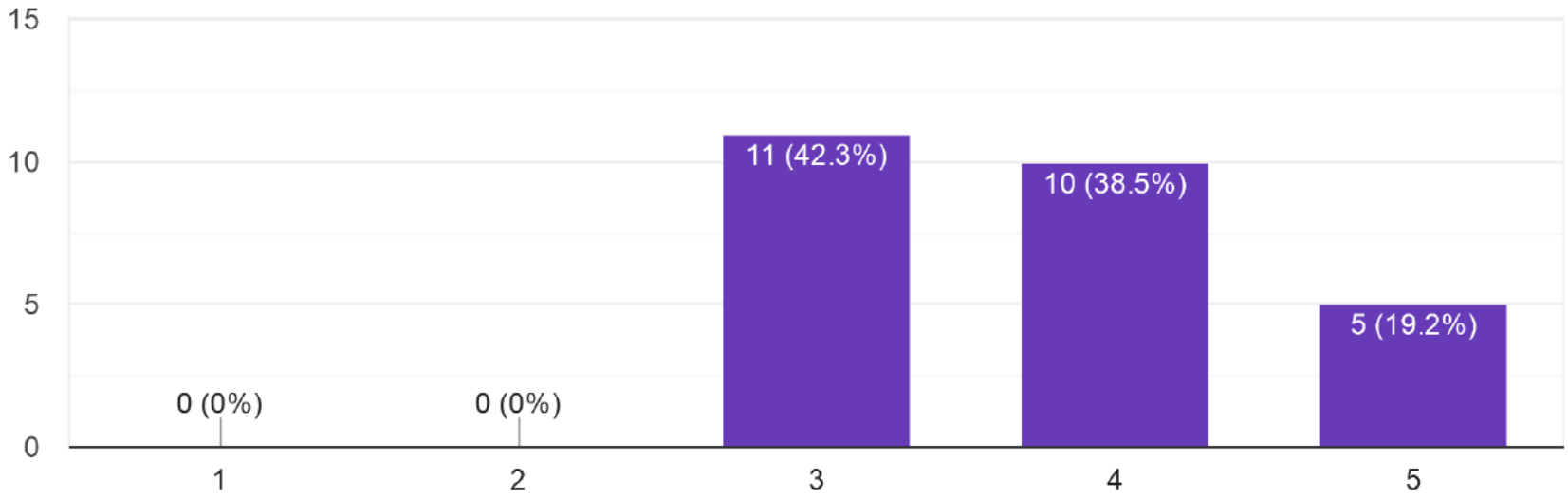


分析

課題3 分量

課題3の分量はどうでしたか？

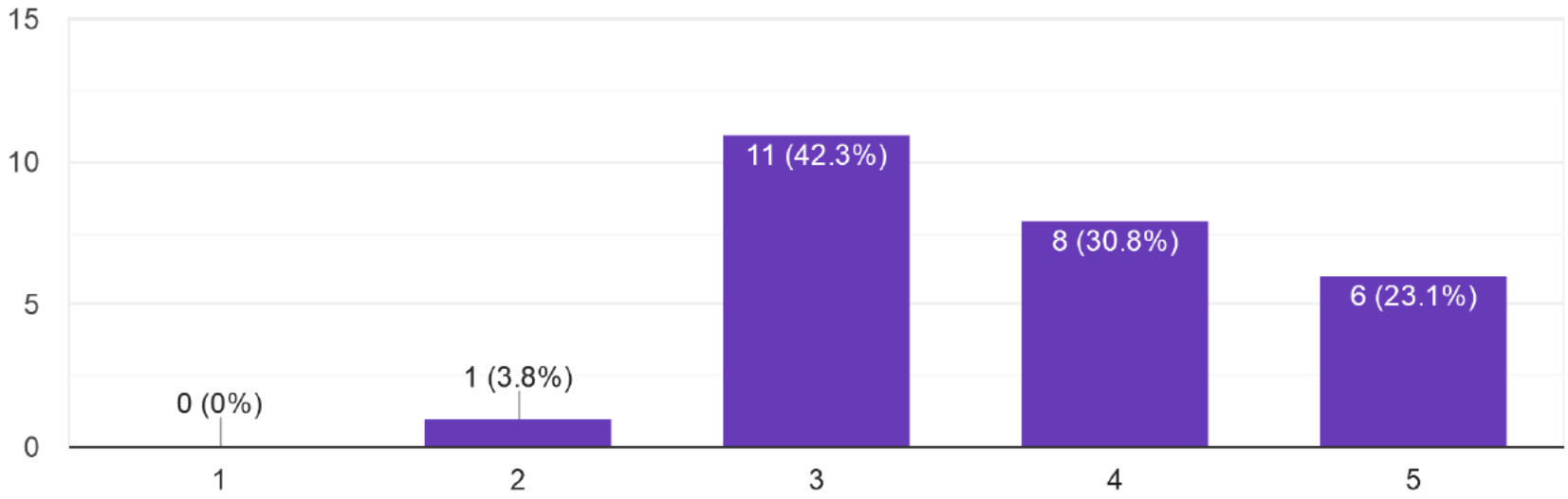
26 件の回答



課題3 難易度

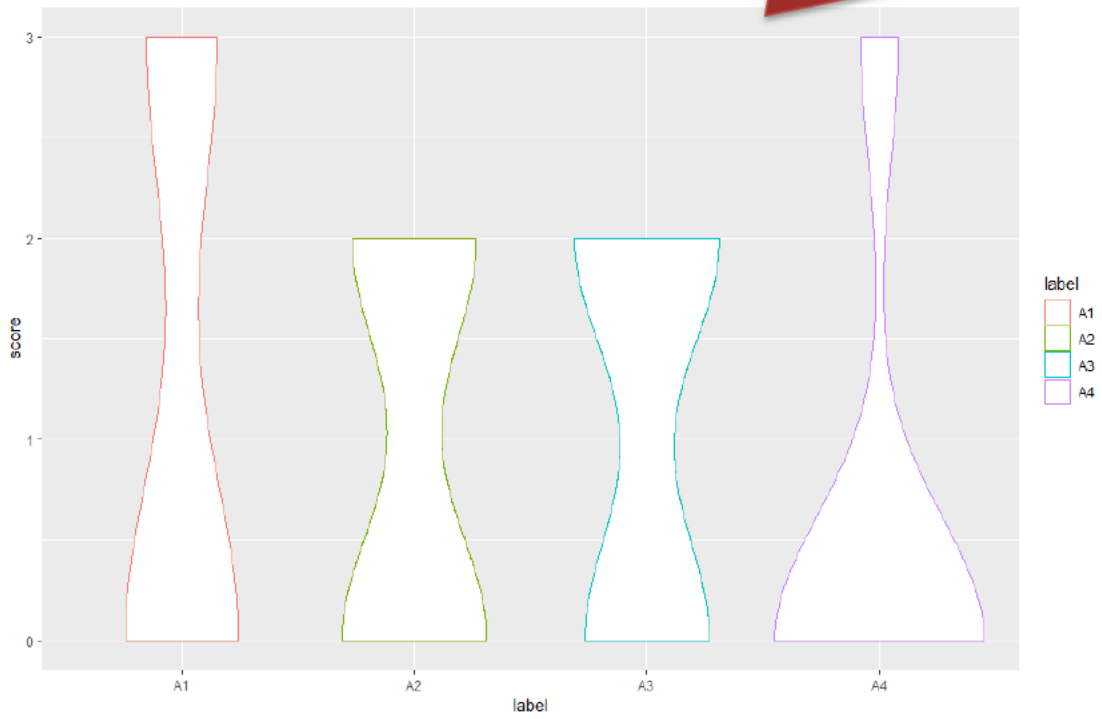
課題3の難易度はどうでしたか？

26 件の回答



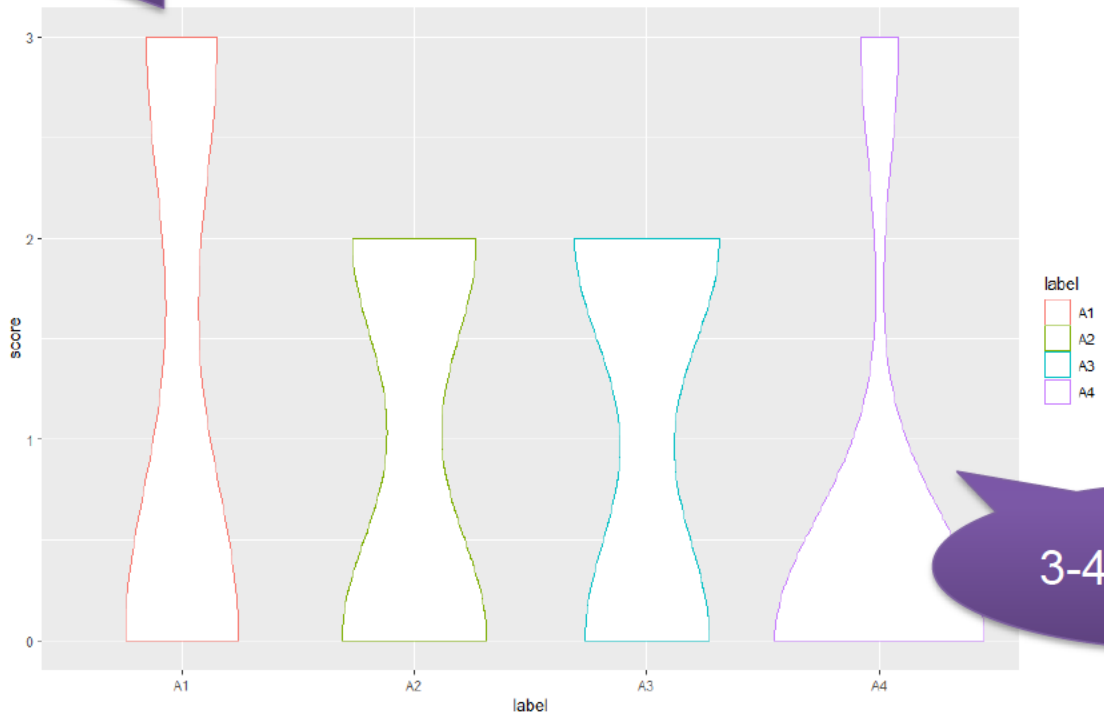
課題 集計･分析問題 分布

全体的にはっきり分かれた



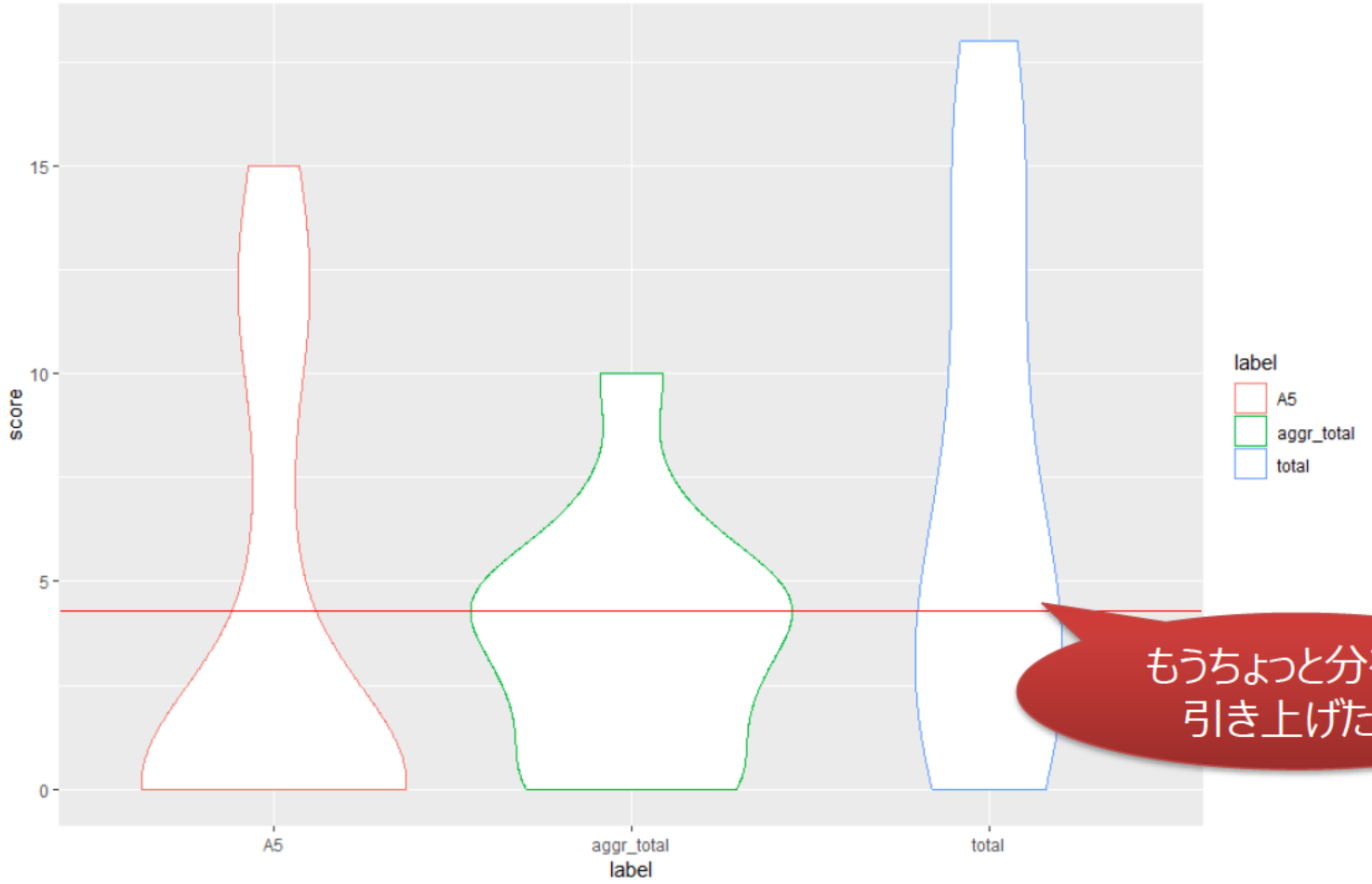
課題3 集計・分析問題 分布

思ったより3-1, 3-2で
得点してもらえなかった



3-4は3-3に依存

課題3 全体の分布


15
10
5
0
score
A5
aggr_total
total
label
label
A5
aggr_total
total
もうちょっと分布を
引き上げたい

全体を通して

全体を通して獲得得点は去年より上がった認識

一方で分類問題を完全に捨てたチームが増えた

(たしか)去年は0チーム，今年は3チーム

なにかしらとっかかりが必要？

反省

得点がもう少しばらけるようにしたい

- 問題数・問題構成・配点の見直し
- とっかかりやすさ

出題方法

- 問題文の改良・サンプルファイルの用意

システムテストの徹底

- ご迷惑をお掛けいたしました

Thank you !

FFRI,Inc.
https://www.ffri.jp

Hiromitsu Oshiba
research-feedback@ffri.jp