

2022-04-13

Treefitのパッケージング

須藤功平 株式会社クリアコード Bio"Pack"athon 2022 #4

自己紹介



✓ 須藤功平

- ✓ 株式会社クリアコード代表取締役
- ✓ フリーソフトウェア開発者
- ✓ TreefitのR版・Python版パッケージの 開発者・メンテナー

バックグラウンド



- ✓ バイオ関連のバックグラウンドはない
- ✓ PythonとRは日常使いしていない
 - ✓ Pythonのパッケージを作ったことはあった
 - ✓ Rのパッケージは作ったことがなかった
 - ✓ C・C++・Rubyをよく使うしパッケージも作っている
- ✓ データ関連ではApache Arrowの現PMC chair

今日の話



- ✓ Python・Rのパッケージングの実例紹介
 - ✓ どんなツールを使っているか
 - ✓ どうメンテナンスしているか
- ✓ よさそう!と思ったところは参考にしてね

Treefitの立ち位置



- ✓ 応用ではなく基盤となる機能を提供
 - ✓ 細胞分化の軌跡推定という応用を想定しているが 木っぽさの指標を与えるというのは汎用的
- ✓ 細胞分化の軌跡推定に限らず使って欲しい

Treefitのパッケージの方針

ClearCode

- ✓ できるだけ依存関係を少なく
 - ✓ 依存関係が多いほど他のパッケージと競合しやすい
 - ✓ 導入の障壁を減らしてユーザーを増やしたい
 - ✓ メンテナンスコストも上がる
- ✓ 依存するパッケージはデファクトのものを
 - ✓ 著名パッケージと統合されていると使いやすい 例:PythonでのNumPy

Treefitの研究プロセス



- 1. アイディアを実験する
- 2. 実験結果を元にディスカッション よい成果がでるまで1. 2. を繰り返す
- 3. よい成果がでたアイディアをリリース colfeで次のリリースをできるかも?

Treefitの研究チーム



- ✓ 早水先生・鈴木さん・石井さん
 - ✓生物・数学などに詳しい人たち
 - ✓ 担当:アルゴリズム考案(1. と2.)
- ✓ 須藤
 - ✓ プログラミングとかパッケージング一般に詳しい人
 - ✓ 担当:パッケージング(3.)

コードベース



- 1. 実験用コードベース
 - ✓ 非公開:Jupyter Notebookベース
 - √ 必要なら便利なものはバンバン使う
 - ✓ Haskellを使っていたこともある
- 2. リリース用コードベース
 - ✓ 公開:https://github.com/hayamizu-lab/
 - ✓ 前述の方針で1. を元に実装

Treefitパッケージの実装



- ✓ アルゴリズムが洗練されているので 実装は非常にコンパクト
- ✓ R版の依存パッケージ
 - ✓ 演算: igraph, pracma
 - ✓ 可視化:ggplot2, patchwork

Treefitのドキュメント

ClearCode

- ✓ ユーザー獲得のために重視!
- ✓ 全体のデザイン: 今どきっぽく
- √ チュートリアル:
 - ✓ 図とサンプルコードをたっぷり
- ✓ リファレンス:網羅

ドキュメントのデザイン:R版



pkqdown

- Treefit for R The first software for quantitative trajectory inference

 The soft of the software for the s
- ✓ なんかシュッとしていてキレイ
- ✓ いろんなRのパッケージで使われている
- ✓ vignetteやRoxygenも含めて サイト一式を生成できる

ドキュメントのデザイン:Python版

Sphinx

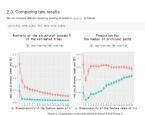
- ✓ Read the Docsテーマ
 - ✓ よく見るやつ
 - ✓ 今ならPyData Sphinxテーマを使う
- ✓ 追加文書やdocstringも含めて サイト一式を生成できる



チュートリアル

ClearCode

- ✓ 図をたっぷり
 - ✓ 研究者が論文を読むときは まずざっと図を眺める
- ✓ サンプルコードをたっぷり
 - ✓ リファレンスを見なくても使えるくらい



3 Interpreting the results 9

リファレンス

ClearCode

- ✓ 最低限:網羅する
 - ✓ どんな公開APIがあるかわかる
- ✓ 使用例と実行結果をつける
 - ✓ 文章での説明よりピンとくる



ドキュメントのデプロイ



- ✓ GitHub Pages
 - ✓ 各プロジェクトごとにWebページを管理できる
 - ✓ github.com/hayamizu-lab/treefit-r/ →
 hayamizu-lab.github.io/treefit-r/
 - ✓ github.com/hayamizu-lab/treefit-python/ →
 hayamizu-lab.github.io/treefit-python/
- ✓ タグを打ったらCIでデプロイ
 - ✓ 未設定 難しくないのでメンテナンスコストを減らすためにやったほうがよい

Treefitのテスト



- ✓ CIを整備する
 - ✓ GitHub Actionsを使用
 - ✓ Linux/macOS/Windowsでテストできるのがうれしい
 - ✓ CRANへのsubmitには複数環境でのテストが必須!
- ✓ Winbuilder/R-hubとの使い分け
 - ✓ CRANへのsubmit前のチェックだけ
 - ✓ 通常はGitHub ActionsのCIでチェック

Treefitのパッケージ登録



- ✓ Rは難しい
 - ✓ CRANへのsubmit後にレビューがあるから
- ✓ Pythonは難しくない
 - ✓ twine uploadですぐにPyPIに反映

CRANへのパッケージ登録・更新

- ✓ https://blog.atusy.net/2019/06/28/cran-submission/か https://cran.r-project.org/web/packages/policies.htmlを読む
 - ✓ 忘れるので毎回読んでいる
- ✓ devtools::release()とかdevtools::*()を 使ってsubmit
 - ✓ ↑も忘れるのでRakefileに書いてある Rで書くべきなんだろうけど。。。

Treefitのメンテナンス



- ✓ CRANでのパッケージの維持
- ✓ GitHubでのフィードバック対応

CRANでのパッケージの維持

ClearCode

- ✓ パッケージが削除されることがある!
 - ✓ 公開してから2年で2回あった
 - ✓ 原因:関連ライブラリーの非互換な更新
 - ✓ 原因:CRANのテスト環境の更新
- ✓ 削除される前にCRANから連絡がある
 - ✓ 少なくとも1回目はあった
 - ✓ 2回目は気づかなかった もしかしたら連絡がなかったのかも。。。

CRANでパッケージが削除されたら

- ✓ 問題を修正してバージョンアップ
- ✓ 1回目のsubmitよりは早く処理してもらえる
 - ✓ 2回目以降は差分を中心にレビューしているから?
- ✓ CIがあるとマジ助かる!
 - ✓数ヶ月単位で実装を触らない期間があるから

GitHubでのフィードバック対応

- https://github.com/hayamizu-lab/treefit-r/issues/1
 - ✓ 露崎さんが問題を報告してくれた!
 - ✓ pull requestへの誘導とレビュー・マージ
 - ✓ CRANへのsubmit
- ✓ みんなも使って気になることがあったら 報告してね!

今日の話



- ✓ Python・Rのパッケージングの実例紹介
 - ✓ どんなツールを使っているか
 - ✓ どうメンテナンスしているか
- ✓ よさそう!と思ったところは参考にしてね
 - ✓ 経験を積みたいならTreefitの研究・開発に 参加するといいかも!