Enhancing Crowdsourced Plant Identification: From Label Aggregation to Personalized Recommendations

Tanguy Lefort
INRIA Lille, Scool

Ongoing work with...

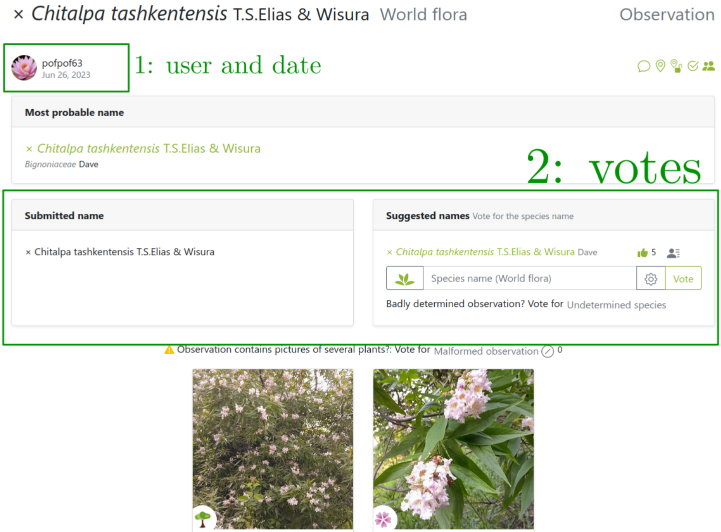
- Odalric Ambrym Maillard
- Alexis Joly
- Vanessa Hequet

- Benjamin Charlier
- Joseph Salmon
- Pierre Bonnet
- Antoine Affouard
- Jean-Christophe Lombardo

Publications

- Label aggregation: Methods in Ecology and Evolution 2024 (part of PhD)
- Recommender system: WIP (part of postdoc)

PL@NTNET ONLINE VOTES



× *Chitalpa tashkentensis* T.S.Elias & Wisura World flora

Observation

pofpof63
Jun 26, 2023

1: user and date

Most probable name

× *Chitalpa tashkentensis* T.S.Elias & Wisura
Bignoniaceae Dave

2: votes

Submitted name

× Chitalpa tashkentensis T.S.Elias & Wisura

Suggested names Vote for the species name

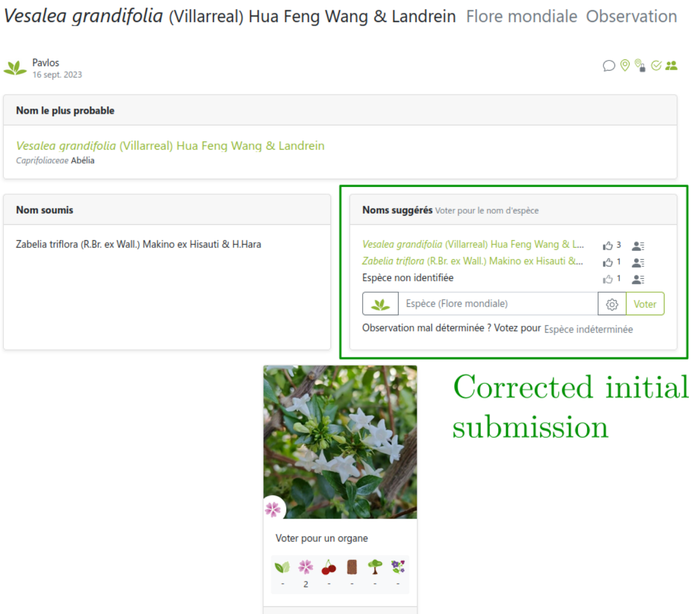
× *Chitalpa tashkentensis* T.S.Elias & Wisura Dave 👍 5

Species name (World flora) Vote

Badly determined observation? Vote for Undetermined species

⚠ Observation contains pictures of several plants?: Vote for Malformed observation ⊘ 0

Users can make corrections



Vesalea grandifolia (Villarreal) Hua Feng Wang & Landrein Flore mondiale Observation

Pavlos
16 sept. 2023

Nom le plus probable

Vesalea grandifolia (Villarreal) Hua Feng Wang & Landrein
Caprifoliaceae Abélia

Nom soumis

Zabelia triflora (R.Br. ex Wall.) Makino ex Hisauti & H.Hara

Noms suggérés Voter pour le nom d'espèce

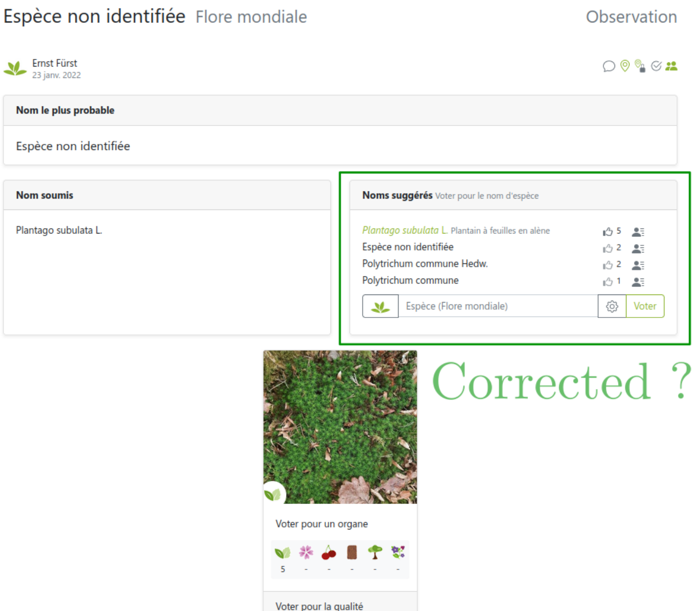
Vesalea grandifolia (Villarreal) Hua Feng Wang & L...	3
Zabelia triflora (R.Br. ex Wall.) Makino ex Hisauti &...	1
Espèce non identifiée	1

Observation mal déterminée ? Votez pour Espèce indéterminée

Voter pour un organe

Corrected initial submission

But sometimes users can't be trusted



Espèce non identifiée Flore mondiale

Observation

Ernst Fürst
23 janv. 2022

Nom le plus probable

Espèce non identifiée

Nom soumis

Plantago subulata L.

Noms suggérés Voter pour le nom d'espèce

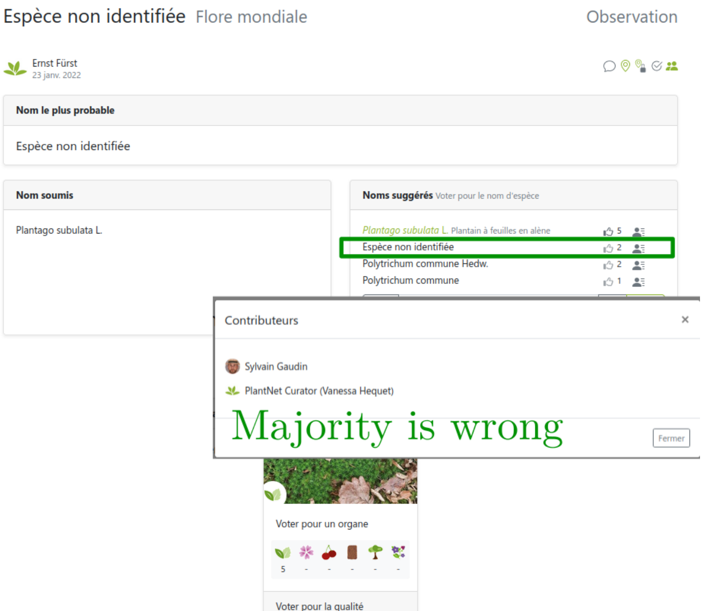
Plantago subulata L. Plantain à feuilles en alène	5
Espèce non identifiée	2
Polytrichum commune Hedw.	2
Polytrichum commune	1

Voter pour un organe

Voter pour la qualité

Corrected ?

But sometimes users can't be trusted



Espèce non identifiée Flore mondiale

Observation

Ernst Fürst
23 janv. 2022

Nom le plus probable

Espèce non identifiée

Nom soumis

Plantago subulata L.

Noms suggérés Voter pour le nom d'espèce

Plantago subulata L. Plantain à feuilles en alène	5	
Espèce non identifiée	2	
Polytrichum commune Hedw.	2	
Polytrichum commune	1	

Contributeurs ×

Sylvain Gaudin

PlantNet Curator (Vanessa Hequet)

Majority is wrong

Fermer

Voter pour un organe

5 - - - - -

Voter pour la qualité

General.

- The good: Fast, easy, cheap data collection

Crowdsourcing for classification

The good, the bad and the ugly

General.

- The good: Fast, easy, cheap data collection
- The bad: Noisy labels with different level skills

Crowdsourcing for classification

The good, the bad and the ugly

General.

- The good: Fast, easy, cheap data collection
- The bad: Noisy labels with different level skills
- The ugly: Very few theory, ad-hoc methods to handle noise from users

Crowdsourcing for classification

The good, the bad and the ugly

General.

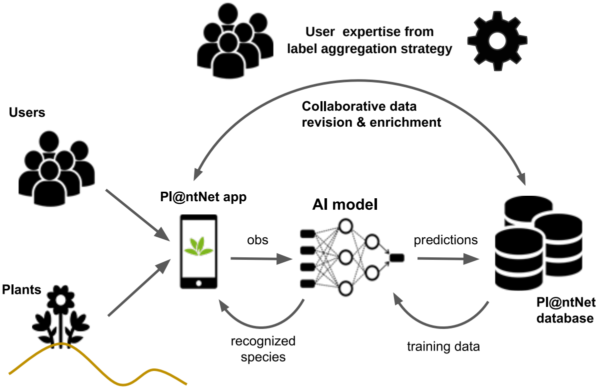
- The good: Fast, easy, cheap data collection
- The bad: Noisy labels with different level skills
- The ugly: Very few theory, ad-hoc methods to handle noise from users

Pl@ntNet.

- 20+ million observations from around the world
- 6+ million users
- 22+ million votes
- 49 720 species

Pl@ntNet general design

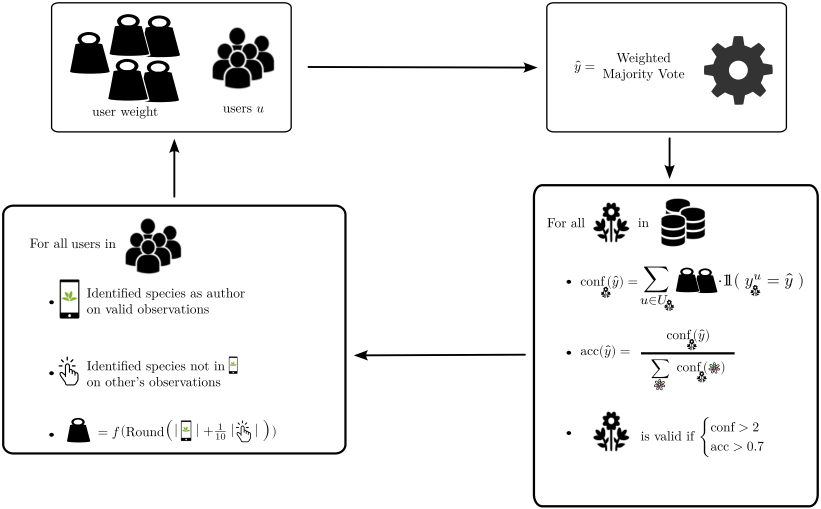
Key concept of Pl@ntNet: Collaborative AI



Pl@ntNet label aggregation

EM based algorithm

Weighting users vote by their estimated number of identified species



user weight

users u

$\hat{y} =$ Weighted Majority Vote

For all ✿ in ⛁

- $\mathrm{conf}_{✿}(\hat{y}) = \sum_{u \in U_{✿}}$ ⚖ $\cdot \mathbb{1}(y^{u}_{✿} = \hat{y})$
- $\mathrm{acc}(\hat{y}) = \dfrac{\mathrm{conf}_{✿}(\hat{y})}{\sum_{❀} \mathrm{conf}_{✿}(❀)}$
- ✿ is valid if $\begin{cases} \mathrm{conf} > 2 \\ \mathrm{acc} > 0.7 \end{cases}$

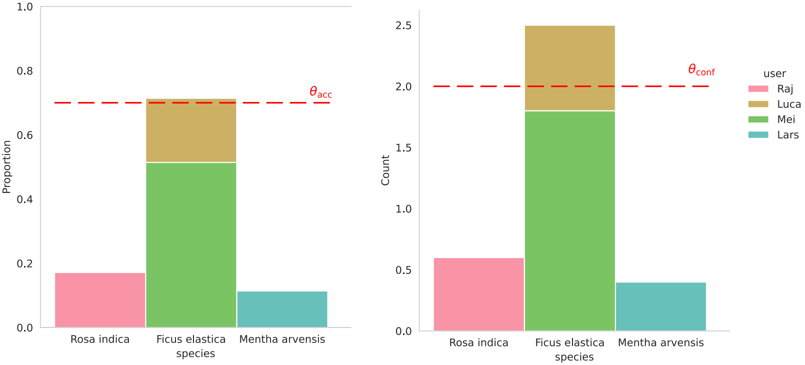
For all users in 👥

- 📱 Identified species as author on valid observations
- 👆 Identified species not in 📱 on other's observations
- ⚖ $= f(\mathrm{Round}(\,|$📱$| + \frac{1}{10}\,|$👆$|\,))$

Active dataset

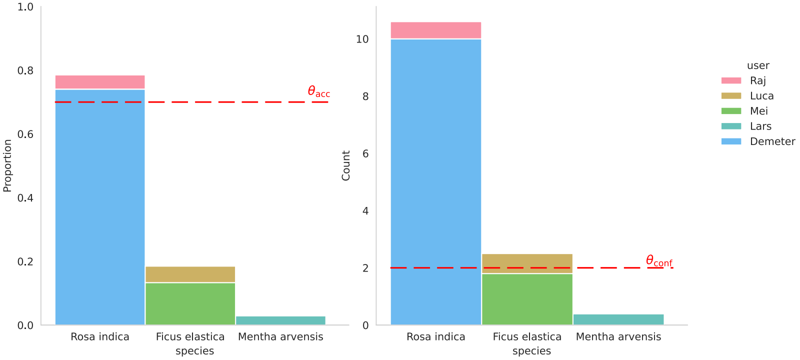
Any observation labeling is active

Initial setting



Active dataset
Any observation labeling is active

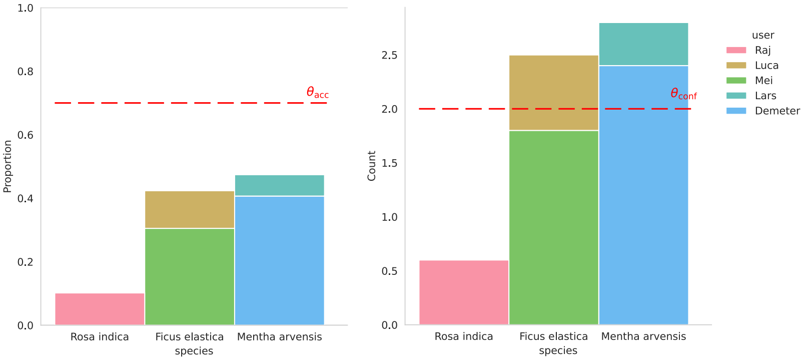
Label switch



Active dataset

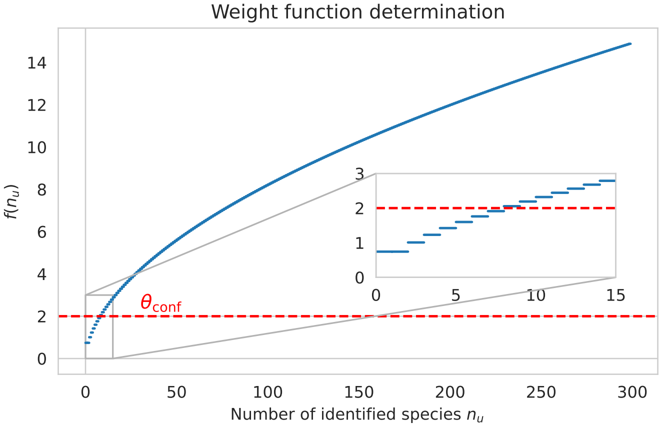
Any observation labeling is active

Invalidating label



Choice of weight function

$$f(n_u) = n_u^{\alpha} - n_u^{\beta} + \gamma \text{ with } \begin{cases} \alpha = 0.5 \\ \beta = 0.2 \\ \gamma = \log(2.1) \simeq 0.74 \end{cases}$$



Other existing strategies

- **Majority Vote** (MV)

Other existing strategies

- **Majority Vote** (MV)
- **Worker agreement with aggregate** (WAWA, Appen 2021)
  - Majority vote
  - Weight user by how much they agree with the majority
  - Weighted majority vote

Other existing strategies

- **Majority Vote** (MV)
- **Worker agreement with aggregate** (WAWA, Appen 2021)
  - Majority vote
  - Weight user by how much they agree with the majority
  - Weighted majority vote
- **TwoThird** (from iNaturalist)
  - Need at least 2 votes
  - 2/3 of agreements

Extracting a subset: Pl@ntNet-CrowdSWE

Design and some numbers

- South Western European flora obs since 2017
- 823 000 users answered more than 11000 species
- 6 700 000 observations
- 9 000 000 votes casted
- **Imbalance**: 80% of observations are represented by 10% of total votes
- zenodo: `https://zenodo.org/records/10782465`

Extracting a subset: Pl@ntNet-CrowdSWE

Design and some numbers

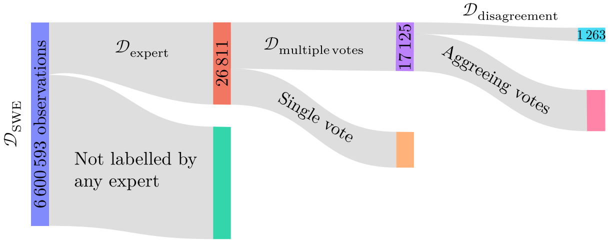
- South Western European flora obs since 2017
- 823 000 users answered more than 11000 species
- 6 700 000 observations
- 9 000 000 votes casted
- **Imbalance**: 80% of observations are represented by 10% of total votes
- zenodo: `https://zenodo.org/records/10782465`

No ground truth available to evaluate the strategies

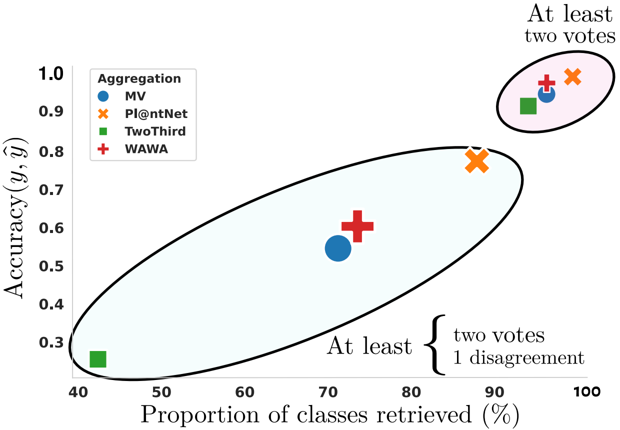
Extracting a subset of a Pl@ntNet
Creation of test sets

- Extraction of 98 experts (TelaBotanic + prior knowledge – thanks to Pierre Bonnet)

Pl@ntnet South-Western Europe flora dataset



Performance



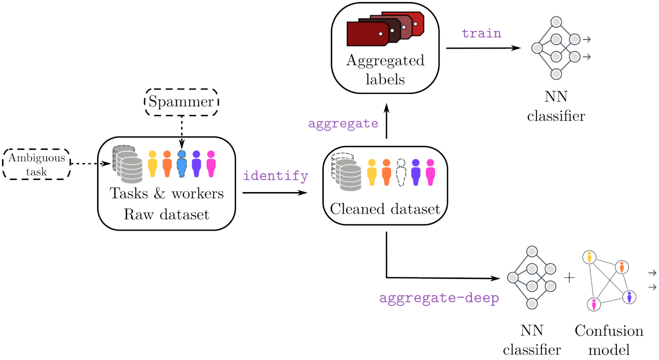
In short

- Pl@ntNet aggregation performs better overall
- We indeed remove some data but less than TwoThird

Aggregating labels: with what tools?

https://peerannot.github.io/

Peerannot: Python library to handle crowdsourced data



Ambiguous task

Spammer

Tasks & workers
Raw dataset

identify

Cleaned dataset

aggregate

Aggregated labels

train

NN classifier

aggregate-deep

NN classifier

+

Confusion model

Recommender system for botanical data

Why?

- "As an expert in XXX I only want to see observations related to XXX"
- Personalized flow of observations to annotate
- Have more valid observations in the long term

Recommender system for botanical data

Why?

- "As an expert in XXX I only want to see observations related to XXX"
- Personalized flow of observations to annotate
- Have more valid observations in the long term

How

- RL: Sequential flow of arriving observations to learn from
- Tool: Contextual Multi-armed bandits (the context is the user's expertise)
- Bonus 1: We can exploit the botanical taxonomy
- Bonus 2: We have a current estimate of the species using Pl@ntNet computer vision model
- Issue: Recommender systems are mostly based on popularity, and we don't want many votes on each observation

Mortal Multi-armed bandits

- Neurips 2008: **Mortal Multi-armed bandits** Chakrabarti et al.
- In our work: user=context and arm=observation to recommend

Mortal Multi-armed bandits

- Neurips 2008: **Mortal Multi-armed bandits** Chakrabarti et al.
- In our work: user=context and arm=observation to recommend

Mortal bandit algorithm in crowdsourcing

```
**Input:** Recommender system $f$, arms $\mathcal{A}$, constraint functions $\Gamma_{\text{agg}}$, user $u$, budget $T$, user weights $W$
**Output:** Set of valid observations
**for** t=1,...,T **do**
    $i \leftarrow f(u)$                                  {recommend a new observation}
    **if** $y_i^u \notin \emptyset$ **then**
        $r_{u,i} \leftarrow 1$
        **if** $\Gamma_{\text{agg}}(i, W, \{y_i^u\}_{i,u}) = 1$ **then**
            $\mathcal{A} \leftarrow \mathcal{A} \setminus \{i\}$                    {observation is valid}
    **else**
        $r_{u,i} \leftarrow 0$
    Update $f$ following its policy
```

Our recommender system: PhyloCrowdRec

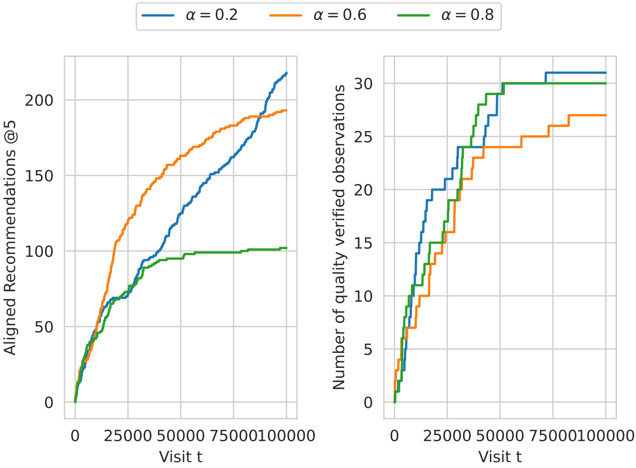
- Keypoint: recommend a genus and then select the observation

```
Input: Recommender system $f$, Constraint functions $\Gamma_{\text{agg}}$, Budget $T$, History of interactions with genera $(g, u, r_{g,u})_{g,u}$, User votes on observations $\{y_i^u\}_{i,u}$
Output: Set of valid observations $\mathcal{D}_{\text{valid}}$, User weights $W$
$\mathcal{D}_{\text{valid}} \leftarrow \emptyset$, $w_u^0 = 1$ for all users                         {Initialization}
for t=1,...,T do
    $g \leftarrow f(u)$                                                          {Recommend genus}
    if $r_{g,u} = 0$ then
        Update CMAB and go to next visit                          {Unaligned recommendation}
    else
        $\mathcal{D}_g \leftarrow \{i | \text{genus}(x_i) = g\}$
        $i_t \leftarrow$ First $(x_i | \text{genus}(x_i) = g,\ \Gamma_{\text{agg}}(i, W, \{y_i^u\}_{i,u}) = 0,\ w_u \geq \max_{u' \in \mathcal{U}_i} w_{u'})_i$,
        Observe $y_{i_t}^u$
        Aggregate $\{y_i^u\}_{i,u}$ and get new weights
        $W \leftarrow (w_u^t)_u$                                                  {Update weights}
        if $\Gamma_{\text{agg}}(i_t, W, \{y_{i_{t'}}^u\}_{i_{t'},u,t'<t} \cup \{y_{i_t}^u\}) = 1$ then
            $\mathcal{D}_{\text{valid}} \leftarrow \mathcal{D}_{\text{valid}} \cup \{i_t\}$                  {observation is valid}
        Update CMAB with $r_{g,u} = 1$
```

Experimental setting

- MovieLens-100K dataset with TwoThird aggregation
- A user likes a genre of movies if they liked over 5 movies of this genre (binary classification: good or bad movie)
- A user likes a movie if rating is 5 stars
- In total: 19 genres, 1682 movies, 100K ratings
- LinUCB bandits for online recommendation

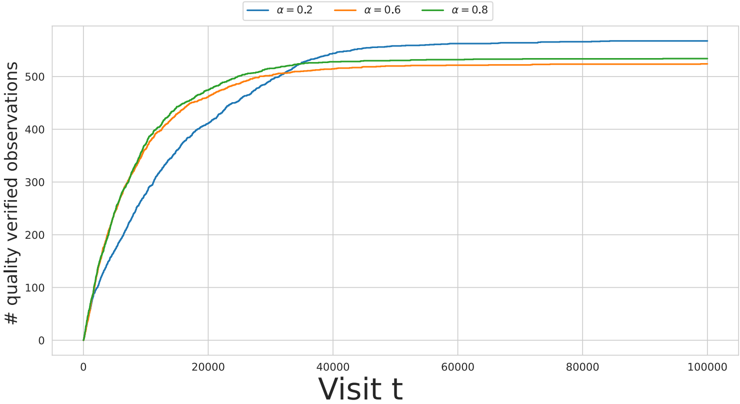
Results Online Mortal bandits



In short

- Too many arms, poor performance overall

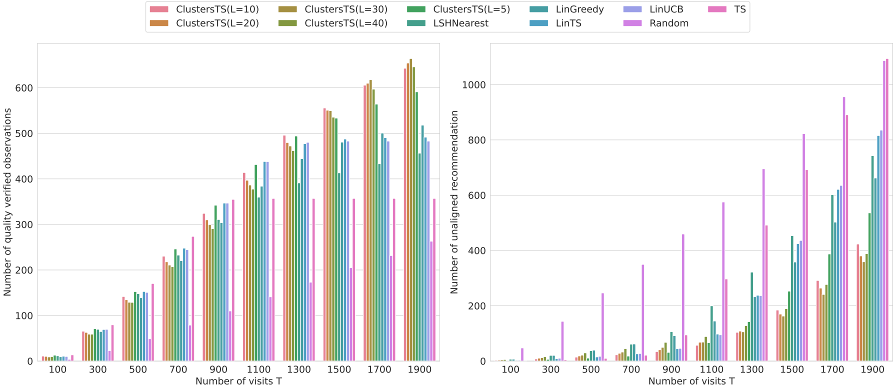
Results PhyloCrowdRec



In short

- More than 550 quality verified movies for the same budget

Other bandit types?
Offline experiment



In short

- Bandits that cluster contexts outperforms others
- Contextual bandits outperform non-contextual bandits

What about Pl@ntNet recommender system?

Work in progress

- What is the user profile?
- What happens when we add the weights?
- Lots of observation are seen by a very few users

Take home message

- Crowdsourcing in large scale classification settings can be handled by the Pl@ntNet aggregation strategy
- Bandit-based recommender systems can exploit the data phylogeny to improve user interactions and quality control
- Python library if you want to try it out: `https://peerannot.github.io/`
- Pl@ntNet-CrowdSWE available on zenodo `https://zenodo.org/records/10782465`

Thank you!