Citizen science for plant identification: insights from Pl@ntNet

Joseph Salmon
Université de Montpellier
Institut Universitaire de France (IUF)

Mainly joint work with:

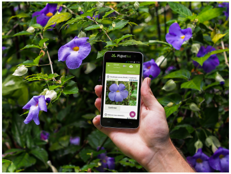
Tanguy Lefort (Univ. Montpellier, IMAG)
Benjamin Charlier (Univ. Montpellier, IMAG)
Camille Garcin (Univ. Montpellier, IMAG)
Maximilien Servajean (Univ. Paul-Valéry-Montpellier, LIRMM, Univ. Montpellier)
Alexis Joly (Inria, LIRMM, Univ. Montpellier)

and: Pl@ntNet

Pierre Bonnet (CIRAD, AMAP)
Antoine Affouard, **J-C. Lombardo**, **Titouan Lorieul**, **Mathias Chouet** (Inria, LIRMM, Univ. Montpellier)

Pl@ntNet

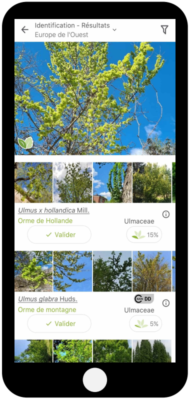
A **citizen science** platform using machine learning to help people identify plants with their mobile phones



Available on the App Store

GET IT ON Google Play

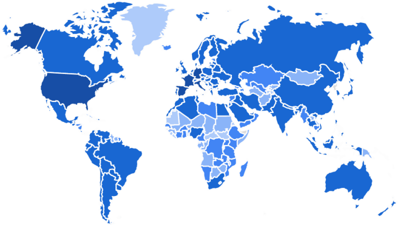
- Website: `https://plantnet.org/`
- Note: no mushroom identification!



PL@NTNET
USAGE AND POPULARITY (GROWING EVERY DAY!)

Pl@ntNet

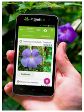
- Start in 2011, now **25M+ users**
- **200+** countries
- Up to **2M** image uploaded/day
- **50K** species
- **1B+** total images
- **10M+** labeled / validated



`https://identify.plantnet.org/stats`

Personal Usage

Nature, walks



Gardening

Phytotherapy

Professional Usage

Agro-ecology

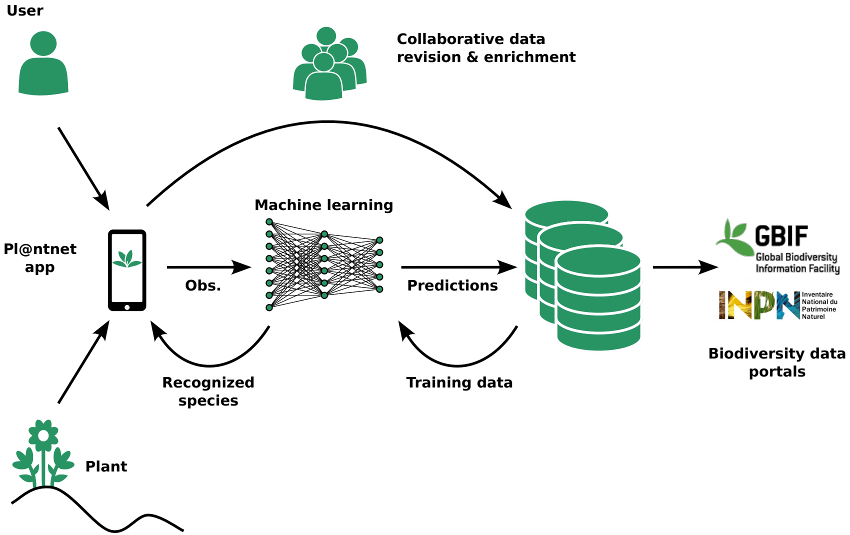
Natural Areas Management

Education, animation

Tourism

Trade

Key concept of Pl@ntNet
Cooperative Learning



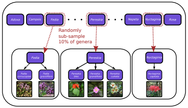
Outline

Motivation: an excellent ... but not a perfect app; **How to improve?**

- Community effort: machine learning, ecology, engineering, amateurs
- Many open problems (theoretical/practical)
- Need for methodological/computational breakthrough

Personal associated contributions
(Mostly methodological)

- Pl@ntNet-300K[1]:
  Creation and **release** of a large-scale dataset sharing the same property as Pl@ntNet; available for the community to improve learning systems



---

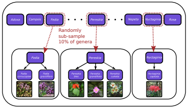
(1) C. Garcin, A. Joly, et al. (2021). “Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution”. *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*.

(2) T. Lefort et al. (2022). *Identify ambiguous tasks combining crowdsourced labels by weighting Areas Under the Margin*. Tech. rep., arXiv:2209.15380.

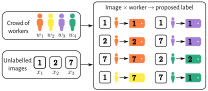
(3) C. Garcin, M. Servajean, et al. (2022). “Stochastic smoothing of the top-K calibrated hinge loss for deep imbalanced classification”. *ICML*.

Personal associated contributions (Mostly methodological)

- Pl@ntNet-300K[(1)]:
  Creation and **release** of a large-scale dataset sharing the same property as Pl@ntNet; available for the community to improve learning systems



- Learning & crowd-sourced data[(2)]:
  How to leverage multiple labels per image to improve the model? Need to **assert quality**: the workers, the images/labels, the model, etc.



---

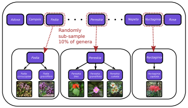
(1) C. Garcin, A. Joly, et al. (2021). "Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution". *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*.

(2) T. Lefort et al. (2022). *Identify ambiguous tasks combining crowdsourced labels by weighting Areas Under the Margin*. Tech. rep., arXiv:2209.15380.

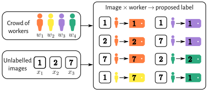
(3) C. Garcin, M. Servajean, et al. (2022). "Stochastic smoothing of the top-K calibrated hinge loss for deep imbalanced classification". *ICML*.

Personal associated contributions
(Mostly methodological)

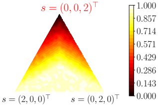
- Pl@ntNet-300K[(1)]:
  Creation and **release** of a large-scale dataset sharing the same property as Pl@ntNet; available for the community to improve learning systems



- Learning & crowd-sourced data[(2)]:
  How to leverage multiple labels per image to improve the model? Need to **assert quality**: the workers, the images/labels, the model, etc.



- Top-K learning[(3)]:
  Driven by theory, introduce new losses to cope with Pl@ntNet constraints to **output multiple labels** (such as user experience, Deep Learning framework, etc.)



[(1)] C. Garcin, A. Joly, et al. (2021). "Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution". *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*.

[(2)] T. Lefort et al. (2022). *Identify ambiguous tasks combining crowdsourced labels by weighting Areas Under the Margin*. Tech. rep., arXiv:2209.15380.

[(3)] C. Garcin, M. Servajean, et al. (2022). "Stochastic smoothing of the top-K calibrated hinge loss for deep imbalanced classification". *ICML*.

A need for new benchmarks

Popular datasets limitations:

- structure of label too simplistic (CIFAR-10, CIFAR-100)
- might be too clean (tasks easy to discriminate)
- might be too well-balanced (same number of images per class)

Motivation:
release a large-scale dataset **sharing similar features** as the Pl@ntNet dataset to foster research in plant identification $\Longrightarrow$ Pl@ntNet-300K[4]

[4] C. Garcin, A. Joly, et al. (2021). "Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution". *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*.

Asymetry of errors in Pl@ntNet

Asymetry of errors in Pl@ntNet

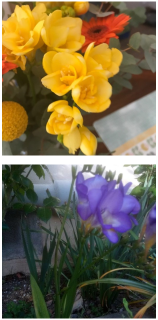
Intra-class variability: same label/species but very diverse images

Guizotia abyssinica

Diascia rigescens

Lapageria rosea

Casuarina cunninghamiana



Freesia alba

Based on pictures only, plant species are challenging to discriminate!

Asymetry of errors in Pl@ntNet

Inter-class ambiguity: different species but similar images

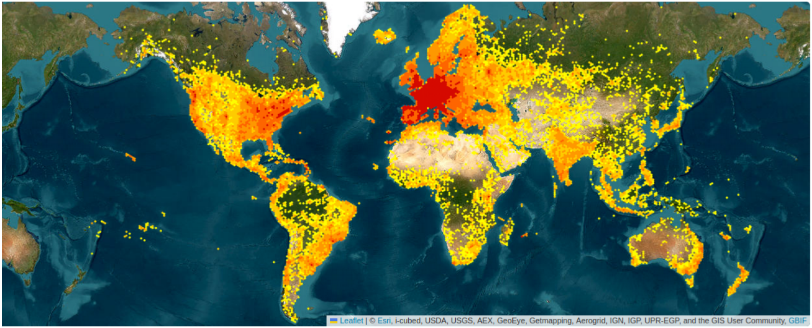
Cirsium *rivulare*	***Chaerophyllum*** *aromaticum*	***Conostomium*** *kenyense*	***Adenostyles*** *leucophylla*	***Sedum*** *montanum*

Cirsium *tuberosum*	***Chaerophyllum*** *temulum*	***Conostomium*** *quadrangulare*	***Adenostyles*** *alliariae*	***Sedum*** *rupestre*

Some species are visually similar (especially within genus)

Sampling bias (13/04/2024)

Spatial density of images collected by Pl@ntNet:



Top-5 most observed plant species in Pl@ntNet:

(a) *Echium vulgare L.*
25 134 observations

(b) *Ranunculus ficaria L.*
24 720 *observations*

(c) *Prunus spinosa L.*
24 103 observations

(d) *Zea mays L.*
23 288 observations

(e) *Alliaria petiolata*
23 075 observations

10 753 observations

Centaurea jacea

VS.

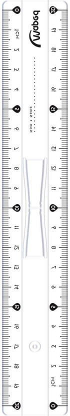
6 observations

Cenchrus agrimonioides

Sampling bias (13/04/2024)
Size

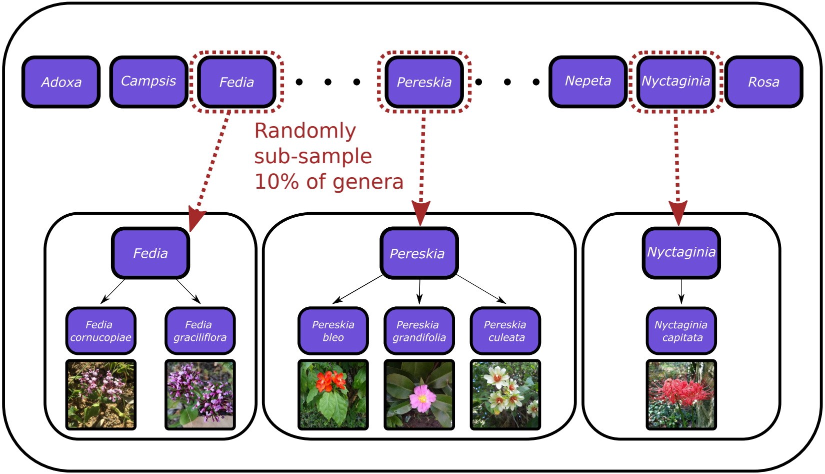
8 376 observations

Magnolia grandiflora



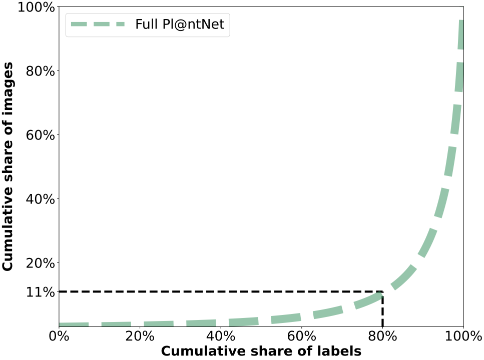
413 observations

Moehringia trinervia



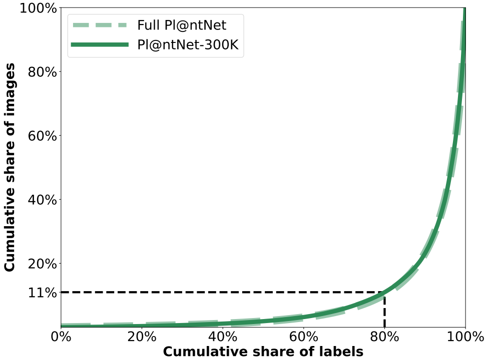
Sample at genus level to preserve intra-genus ambiguity (use hierarchical structure)

Long tailed distribution
Preserved with Subsampling of genera



80% of species | 11% of images $\iff$ 20% of species | 89% of images

Long tailed distribution
Preserved with Subsampling of genera

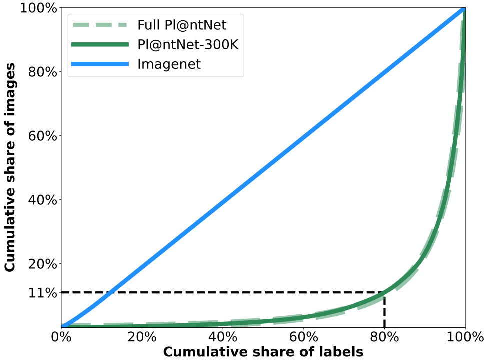


80% of species | 11% of images $\Longleftrightarrow$ 20% of species | 89% of images

Reminder:

- Pl@ntNet-300K: 1K+ species
- Pl@ntNet: 50K+ species
- Earth: 300K+ species

Long tailed distribution
Preserved with Subsampling of genera

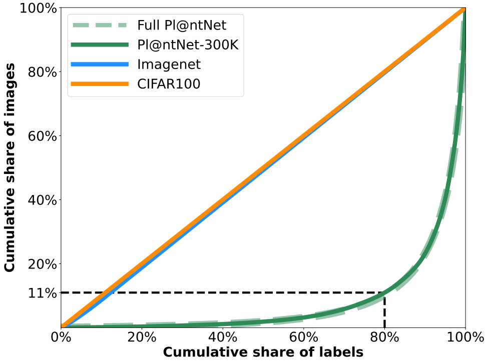


80% of species | 11% of images $\Longleftrightarrow$ 20% of species | 89% of images

Reminder:

- Pl@ntNet-300K: 1K+ species
- Pl@ntNet: 50K+ species
- Earth: 300K+ species

Long tailed distribution
Preserved with Subsampling of genera



80% of species | 11% of images $\Longleftrightarrow$ 20% of species | 89% of images

Reminder:

- Pl@ntNet-300K: 1K+ species
- Pl@ntNet: 50K+ species
- Earth: 300K+ species

Details on Pl@ntNet-300K
Size and links

- $306\,146$ color images
- 32 GB
- Labels: $K = 1\,081$ species
- $2\,079\,003$ volunteers "workers"

Zenodo, 1 click download

`https://zenodo.org/record/5645731`

Code to train models:

`https://github.com/plantnet/PlantNet-300K`

Outline

Pl@ntNet online votes

https://identify.plantnet.org/weurope/observations/1012500059

PL@NTNET ONLINE VOTES

https://identify.plantnet.org/weurope/observations/
1012500059

× *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

What about the labels?

- Images taken by users . . . so are the labels!
- But users can be wrong, or not experts
- Several labels can be available!

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

Espèce (Flore mondiale) Voter

Observation mal déterminée ? Votez pour Espèce indéterminée

Voter pour un organe

- 2 - - - -

Corrected initial submission

But sometimes users can't be trusted

https://identify.plantnet.org/weurope/observations/
1012500059

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

Espèce (Flore mondiale) Voter

Corrected ?

Voter pour un organe

5 - - - - -

Voter pour la qualité

But sometimes users can't be trusted

https://identify.plantnet.org/weurope/observations/1012500059

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.

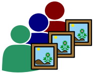
- The good: Fast, easy, cheap data collection

General.

- The good: Fast, easy, cheap data collection
- The bad: Noisy labels with different levels of expertise

General.

- The good: Fast, easy, cheap data collection
- The bad: Noisy labels with different levels of expertise
- The ugly: (partly) missing theory, ad-hoc methods for noisy labels

- Classes/labels/species:  k $\in$ $\{$  $\} \triangleq [K]$
- Images collected:  x_i $\in$  $\triangleq \mathcal{X}_{\text{train}}$
- Users/labelers:  u $\in$  $\triangleq [n_{\text{user}}]$
- Labels given by u to x_i : $y_i^u \triangleq$  $\in$ $\{$  $\}$
- Users labeling x_i :  $\triangleq \mathcal{U}(x_i)$

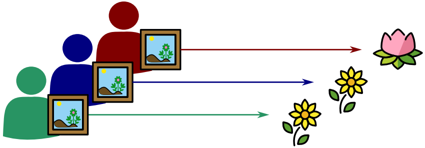
Provide for all images x_i an agregated label $\hat{y}_i$ based on the votes y_i^u of the workers $u \in \mathcal{U}$.

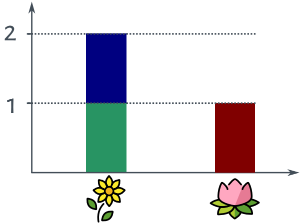
Majority Vote

Naive idea:
make users vote and take the most voted label for each image

Majority Vote

Naive idea:
make users vote and take the most voted label for each image





Result : $\hat{y}^{MV}$  = 

Definition: Majority Voting (MV)

Majority Voting outputs the most answered label:

$$\forall x_i \in \mathcal{X}_{\texttt{train}}, \quad \hat{y}_i^{\mathrm{MV}} = \underset{k \in [K]}{\arg\max} \left(\sum_{u \in \mathcal{U}(x_i)} \mathbb{1}_{\{y_i^u = k\}} \right)$$

Properties:

- ✓ simple
- ✓ adapted for any number of users
- ✓ usually efficient, often few labelers sufficient (say[5] <5)
- ✗ ineffective for borderline cases
- ✗ suffer from spammers / adversarial users

[5] R. Snow et al. (2008). "Cheap and Fast - But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks". *Conference on Empirical Methods in Natural Language Processing*. EMNLP 2008. Association for Computational Linguistics, pp. 254–263.

Outline

Weights, confidence and accuracy

Constraints: wide range of skills, different levels of expertise

Modeling aspect: add a user weight to balance votes

Weights, confidence and accuracy

Constraints: wide range of skills, different levels of expertise

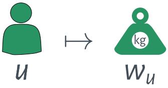
Modeling aspect: add a user weight to balance votes

$$u \mapsto w_u$$

Weights, confidence and accuracy

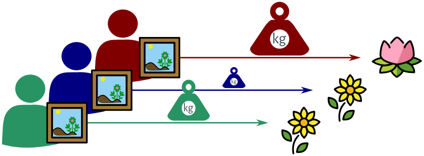
Constraints: wide range of skills, different levels of expertise

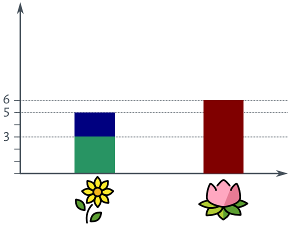
Modeling aspect: add a user weight to balance votes



Let us assume $(w_u)_{u \in \mathcal{U}}$ given for now

Weighted Majority Vote (WMV)
Example





Result : $\hat{y}^{\mathrm{WMV}}$  = 

Definition: label confidence

The label confidence $\mathrm{conf}_i(k)$ of label k for image x_i is the sum of the weights of the workers who voted for k:

$$\forall k \in [K], \quad \mathrm{conf}_i(k) = \sum_{u \in \mathcal{U}(x_i)} w_u \mathbb{1}_{\{y_i^u = k\}}$$

Confidence

Definition: label confidence

The label confidence $\mathrm{conf}_i(k)$ of label k for image x_i is the sum of the weights of the workers who voted for k:

$$\forall k \in [K], \quad \mathrm{conf}_i(k) = \sum_{u \in \mathcal{U}(x_i)} w_u \mathbb{1}_{\{y_i^u = k\}}$$

Size effect:

- more votes $\Rightarrow$ more confidence
- more expertise $\Rightarrow$ more confidence

Definition: label accuracy

The label accuracy $\mathrm{acc}_i(k)$ of label k for image x_i is the normalized sum of weights of the workers who voted for k:

$$\forall k \in [K], \quad \mathrm{acc}_i(k) = \mathrm{conf}_i(k) / \sum_{k' \in [K]} \mathrm{conf}_i(k')$$

Accuracy

Definition: label accuracy

The label accuracy $\mathrm{acc}_i(k)$ of label k for image x_i is the normalized sum of weights of the workers who voted for k:

$$\forall k \in [K], \quad \mathrm{acc}_i(k) = \mathrm{conf}_i(k) / \sum_{k' \in [K]} \mathrm{conf}_i(k')$$

Interpretation: only the proportion of the weights matters

Weighted Majority Vote (WMV)

Definition: Weighted Majority Voting (WMV)

Majority voting but weighted by a confidence score per user u:

$$\forall x_i \in \mathcal{X}_{\texttt{train}}, \quad \hat{y}_i^{\mathrm{WMV}} = \operatorname*{arg\,max}_{k \in [K]} \left(\sum_{u \in \mathcal{U}(x_i)} w_u \mathbb{1}_{\{y_i^u = k\}} \right)$$

Weighted Majority Vote (WMV)

Definition: Weighted Majority Voting (WMV)

Majority voting but weighted by a confidence score per user u:

$$\forall x_i \in \mathcal{X}_{\texttt{train}}, \quad \hat{y}_i^{\mathrm{WMV}} = \operatorname*{arg\,max}_{k \in [K]} \left(\sum_{u \in \mathcal{U}(x_i)} w_u \mathbb{1}_{\{y_i^u = k\}} \right)$$

Note: the weights w_u can be computed from confidence or accuracy

$$\hat{y}_i^{\mathrm{WMV}} = \operatorname*{arg\,max}_{k \in [K]} \left(\mathrm{conf}_i(k) \right) = \operatorname*{arg\,max}_{k \in [K]} \left(\mathrm{acc}_i(k) \right)$$

Suppose that you have a label estimate $\hat{y}_i$ for x_i:

Label validation

Suppose that you have a label estimate $\hat{y}_i$ for x_i:

Labels quality check: need for **expertise**

keep images with label confidence above a threshold θ_{conf}, validate $\hat{y}_i$ when $\mathrm{conf}_i(\hat{y}_i) > \theta_{\text{conf}}$

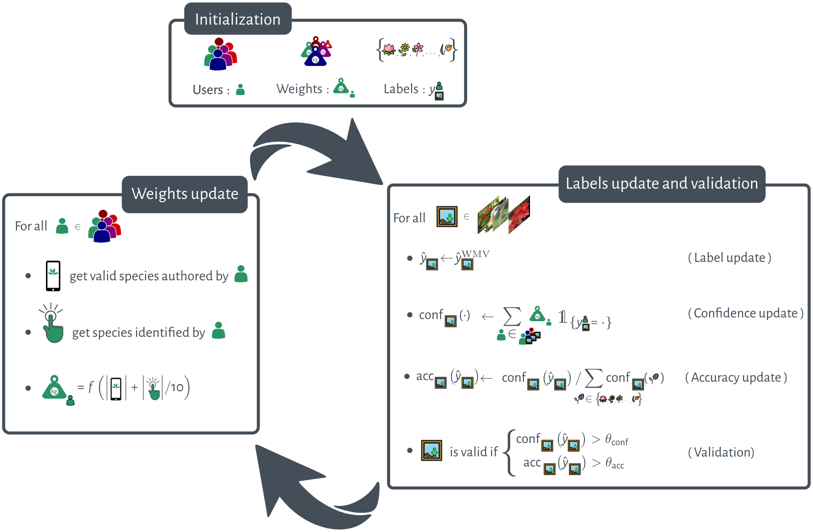
Suppose that you have a label estimate $\hat{y}_i$ for x_i:

Labels quality check: need for **expertise**
keep images with label confidence above a threshold θ_{conf}, validate $\hat{y}_i$ when $\mathrm{conf}_i(\hat{y}_i) > \theta_{\text{conf}}$

Agreement check: need for **consensus**
keep images with label accuracy above a threshold θ_{acc}, validate $\hat{y}_i$ when $\mathrm{acc}_i(\hat{y}_i) > \theta_{\text{acc}}$

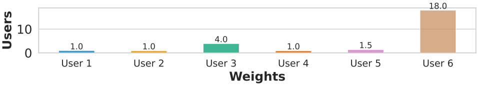
Pl@ntNet label aggregation (EM algorithm)

Weight user vote by number of identifications



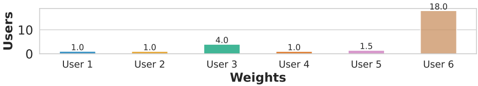
Weights example

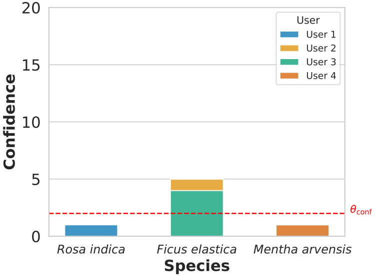
- $n_{\text{user}} = 6, K = 3$: *Rosa indica, Ficus elastica, Mentha arvensis*
- $\theta_{\text{conf}} = 2$ and $\theta_{\text{acc}} = 0.7$
- Users weights as follows:

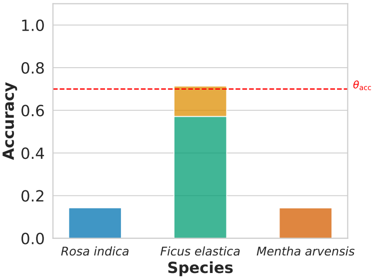


Weights example

- $n_{\text{user}} = 6, K = 3$: *Rosa indica, Ficus elastica, Mentha arvensis*
- $\theta_{\text{conf}} = 2$ and $\theta_{\text{acc}} = 0.7$
- Users weights as follows:

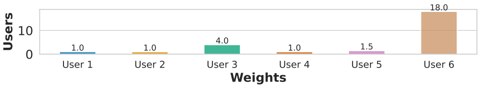


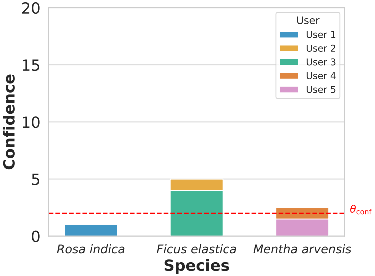


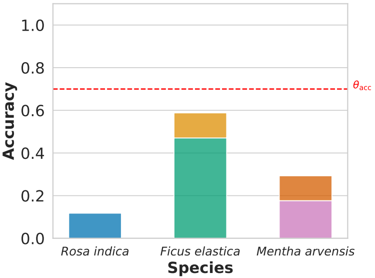


Take into account 4 users out of 6

- $n_{\text{user}} = 6$, $K = 3$: *Rosa indica*, *Ficus elastica*, *Mentha arvensis*
- $\theta_{\text{conf}} = 2$ and $\theta_{\text{acc}} = 0.7$
- Users weights as follows:



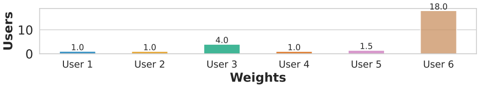


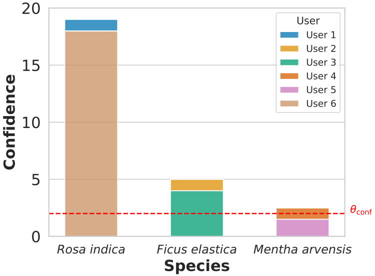


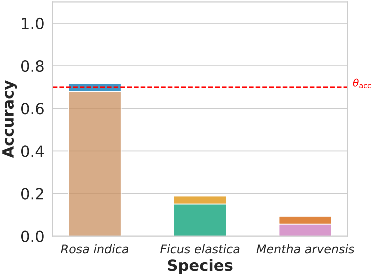
Invalidated label: Adding User 5 reduces accuracy

Weights example

- $n_{\text{user}} = 6, K = 3$: *Rosa indica, Ficus elastica, Mentha arvensis*
- $\theta_{\text{conf}} = 2$ and $\theta_{\text{acc}} = 0.7$
- Users weights as follows:



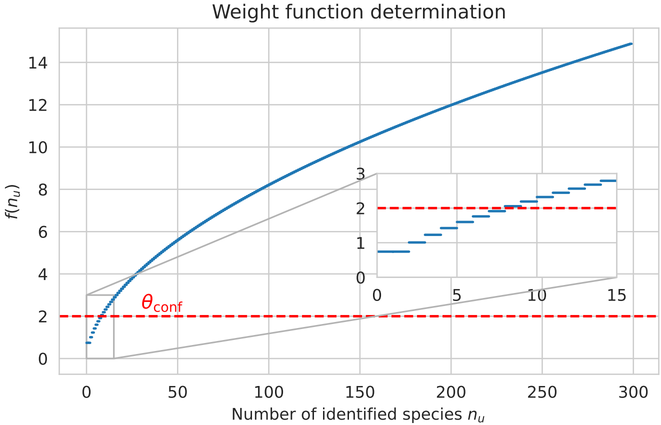




Label switched: User 6 is an expert

Choice of weight function

$$f(n_u) = n_u^{\alpha} - n_u^{\beta} + \gamma \text{ with } \begin{cases} \alpha = 0.5 \\ \beta = 0.2 \\ \gamma = \log(1.7) \simeq 0.74 \end{cases}$$



Other existing strategies

- **Majority Vote** (MV)

Other existing strategies

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

Outline

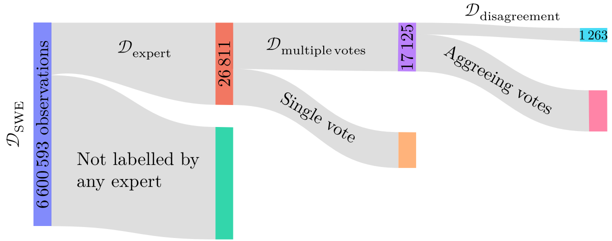
- South Western European flora obs since 2017
- $800K$ users answered more than $11K+$ species
- $9M+$ votes casted
- **Imbalance**: 80% of observations are represented by 10% of total votes

- South Western European flora obs since 2017
- $800K$ users answered more than $11K+$ species
- $9M+$ votes casted
- **Imbalance**: 80% of observations are represented by 10% of total votes

No ground truth available to evaluate the strategies

- Extract 98 experts : Tela Botanica + prior knowledge (P. Bonnet)

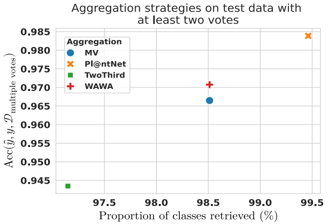
Pl@ntnet South-Western Europe flora dataset

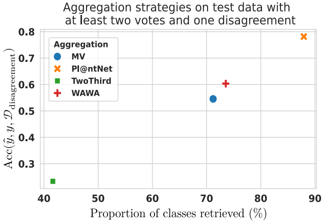


https://zenodo.org/records/10782465

Performance

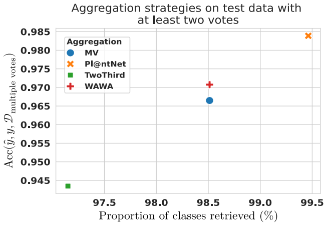
Accuracy and volume of classes kept

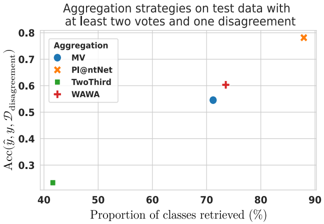




Performance

Accuracy and volume of classes kept



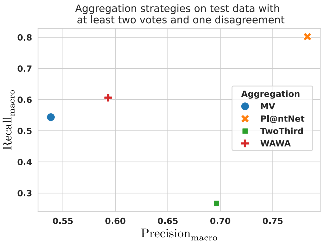


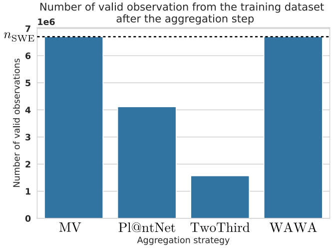
In short

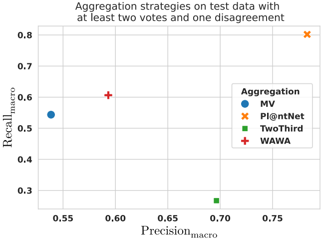
- Pl@ntNet aggregation performs better overall
- iNaturalist is highly impacted by their reject threshold
- In ambiguous settings (right), strategies weighting users are better

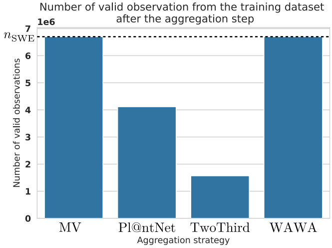
Performance

Precision, recall and validity







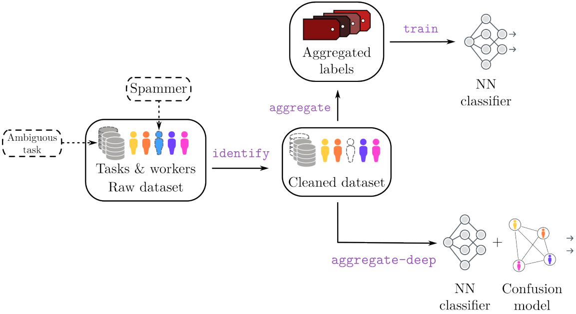


In short

- Pl@ntNet aggregation performs better overall
- TwoThird has good precision but bad recall
- We indeed remove some data but less than TwoThird

Aggregating labels: a new open source tools

Peerannot: Python library to handle crowdsourced data



Take home message(s)

- Citizen science challenges: many and varied (need more attention)

Take home message(s)

- Citizen science challenges: many and varied (need more attention)
- Crowdsourcing / Label uncertainty: helpful for **data curation**

Conclusion

Take home message(s)

- Citizen science challenges: many and varied (need more attention)
- Crowdsourcing / Label uncertainty: helpful for **data curation**
- Improved **data quality** $\Rightarrow$ **improved learning** performance

Conclusion

Take home message(s)

- Citizen science challenges: many and varied (need more attention)
- Crowdsourcing / Label uncertainty: helpful for **data curation**
- Improved **data quality** $\Rightarrow$ **improved learning** performance

Dataset release:

- Pl@ntNet-300K: `https://zenodo.org/record/5645731`
- Pl@ntNet SWE flora: `https://zenodo.org/records/10782465`

Conclusion

Take home message(s)

- Citizen science challenges: many and varied (need more attention)
- Crowdsourcing / Label uncertainty: helpful for **data curation**
- Improved **data quality** $\Rightarrow$ **improved learning** performance

Dataset release:

- Pl@ntNet-300K: `https://zenodo.org/record/5645731`
- Pl@ntNet SWE flora: `https://zenodo.org/records/10782465`

Code release:

- Toolbox: **`https://peerannot.github.io/`**
- Some benchmarks: **`https://benchopt.github.io/`**

Conclusion

Take home message(s)

- Citizen science challenges: many and varied (need more attention)
- Crowdsourcing / Label uncertainty: helpful for **data curation**
- Improved **data quality** $\Rightarrow$ **improved learning** performance

Dataset release:

- Pl@ntNet-300K: `https://zenodo.org/record/5645731`
- Pl@ntNet SWE flora: `https://zenodo.org/records/10782465`

Code release:

- Toolbox: **`https://peerannot.github.io/`**
- Some benchmarks: **`https://benchopt.github.io/`**

Future work

- Uncertainty quantification
- Improve robustness to adversarial users
- Leverage gamification for more quality labels **`theplantgame.com`**

Contact:

Joseph Salmon

joseph.salmon@umontpellier.fr

https://josephsalmon.eu

Github: @josephsalmon

Mastodon: @josephsalmon@sigmoid.social

References I

Garcin, C., A. Joly, et al. (2021). "Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution". *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*.

Garcin, C., M. Servajean, et al. (2022). "Stochastic smoothing of the top-K calibrated hinge loss for deep imbalanced classification". *ICML*.

Lefort, T. et al. (2022). *Identify ambiguous tasks combining crowdsourced labels by weighting Areas Under the Margin*. Tech. rep., arXiv:2209.15380.

Snow, R. et al. (2008). "Cheap and Fast - But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks". *Conference on Empirical Methods in Natural Language Processing*. EMNLP 2008. Association for Computational Linguistics, pp. 254–263.