Improving Multiclass Classification in Crowdsourcing by Using Hierarchical Schemes
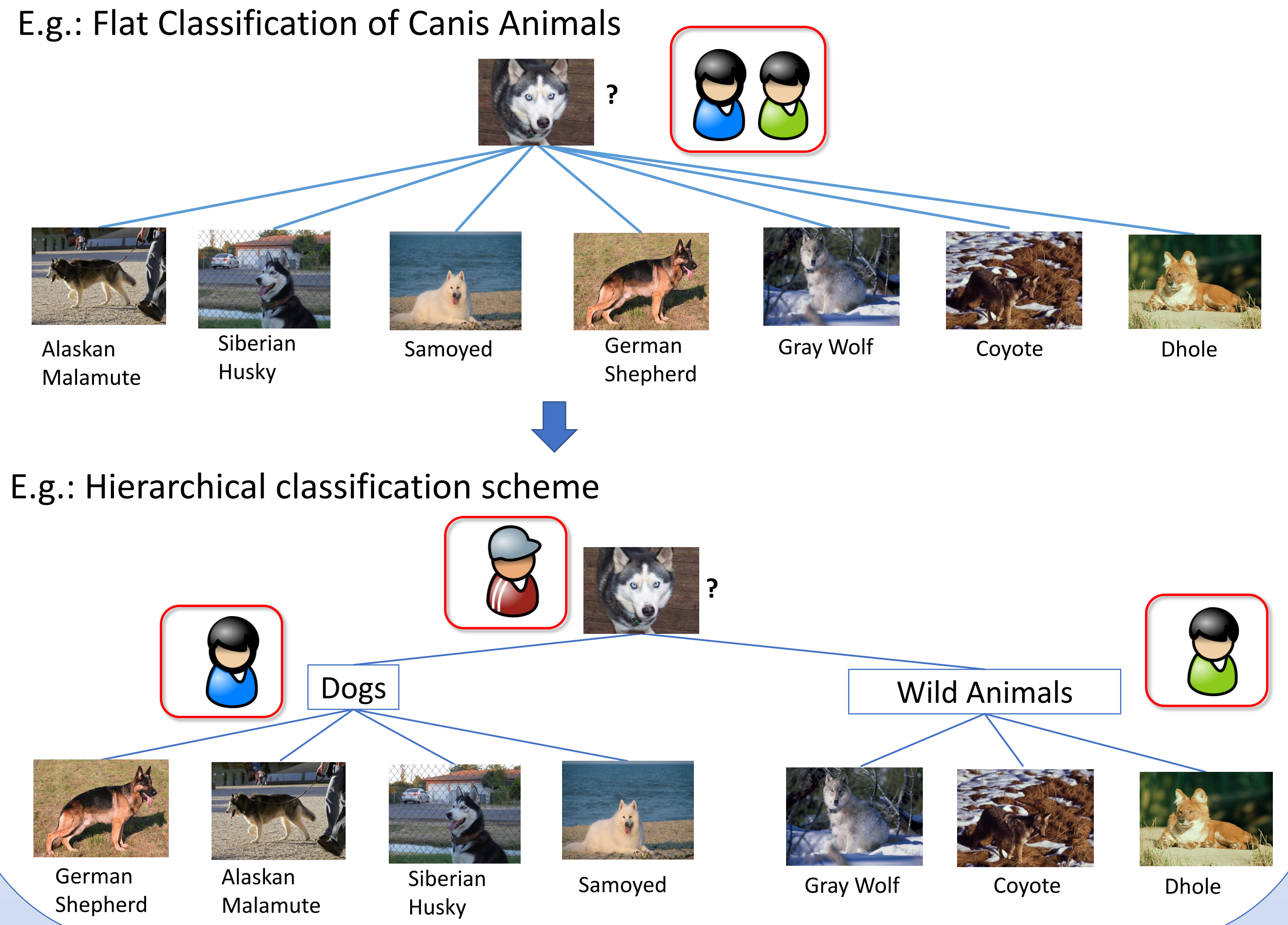
Xiaoni Duan, Keishi Tajima
School of Informatics, Kyoto University

京都大学
KYOTO UNIVERSITY

Different workers may be good at distinguishing different items

Worker A

Siberian Husky

Alaskan Malamute

Worker B

Gray Wolf

Coyote

Hierarchical Scheme Selection :

1. Generate many candidate hierarchical schemes.
2. For each scheme, simulate the worker allocation process, and estimate the expected accuracy.
3. Choose the scheme with the best accuracy.

In hierarchical classification, we can assign different workers to different sub-tasks

E.g.: Flat Classification of Canis Animals

?

Alaskan Malamute | Siberian Husky | Samoyed | German Shepherd | Gray Wolf | Coyote | Dhole

E.g.: Hierarchical classification scheme

?

Dogs

German Shepherd | Alaskan Malamute | Siberian Husky | Samoyed

Wild Animals

Gray Wolf | Coyote | Dhole

Worker Allocation Algorithm
### Greedy algorithm focusing on variance of worker ability

Overview:
1. We publish a flat classification task as a qualification task.
2. We calculate accuracy of each worker for each subtask by using the ground truth.
3. We assign workers to subtasks by a greedy algorithm giving priority to workers whose accuracy largely changes depending on tasks.

Example of worker allocation:

Create three worker lists sorted by accuracy for subtask AB, A, and B.

l_{AB}

Worker	Accuracy AB	Jobs
Worker 1	1	60
Worker 2	0.98	20
Worker 3	0.95	80
...	...	...

l_A

Worker	Accuracy A	Jobs
Worker 2	1	20
Worker 1	0.95	60
Worker 3	0.90	80
...	...	...

l_B

| Worker	Accuracy B	Jobs
Worker 3	0.98	80
Worker 2	0.95	20
Worker 1	0.93	60
...	...	...

AB   A   B

● : task instances
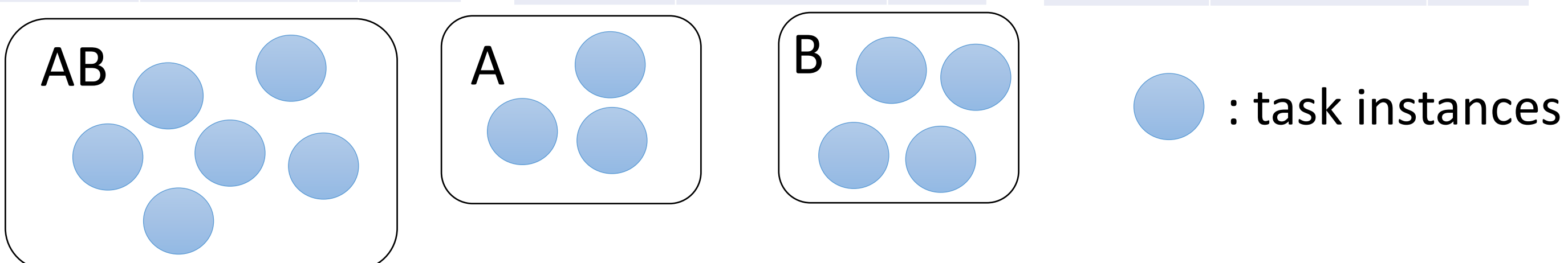
Calculate standard scores of workers' accuracy in each subtask.
Sort workers by the variance of the standard scores in subtasks.

| Worker	Variance	Standard Score AB	Standard Score A	Standard Score B
Worker 3	1.48454	-1.314	-1.225	1.314
Worker 2	0.34768	0.146	1.225	-0.146
Worker 1	0.311	1.119	0	-1.119

Worker 3 has the largest variance.

Assign worker 3 to the subtask that he can do best – subtask B.
Remove task instances assigned to worker 3 from subtask B.

AB   A   B
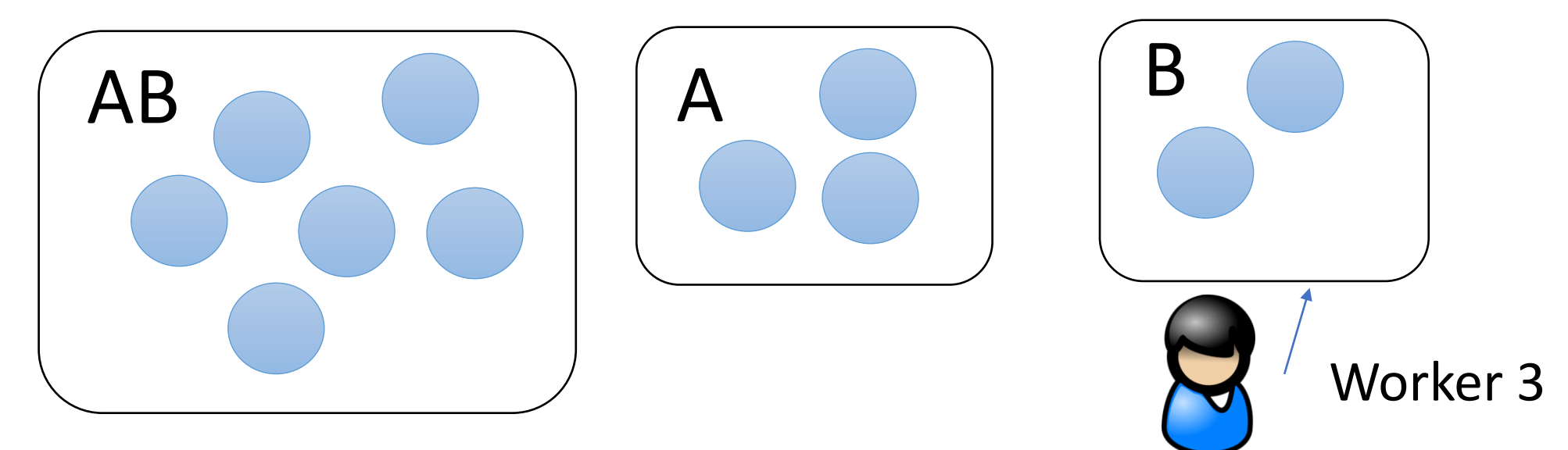
Worker 3

Assign Worker 2, then Worker 1, in their priority order.

AB   A   B

Worker 1   Worker 2   Worker 3
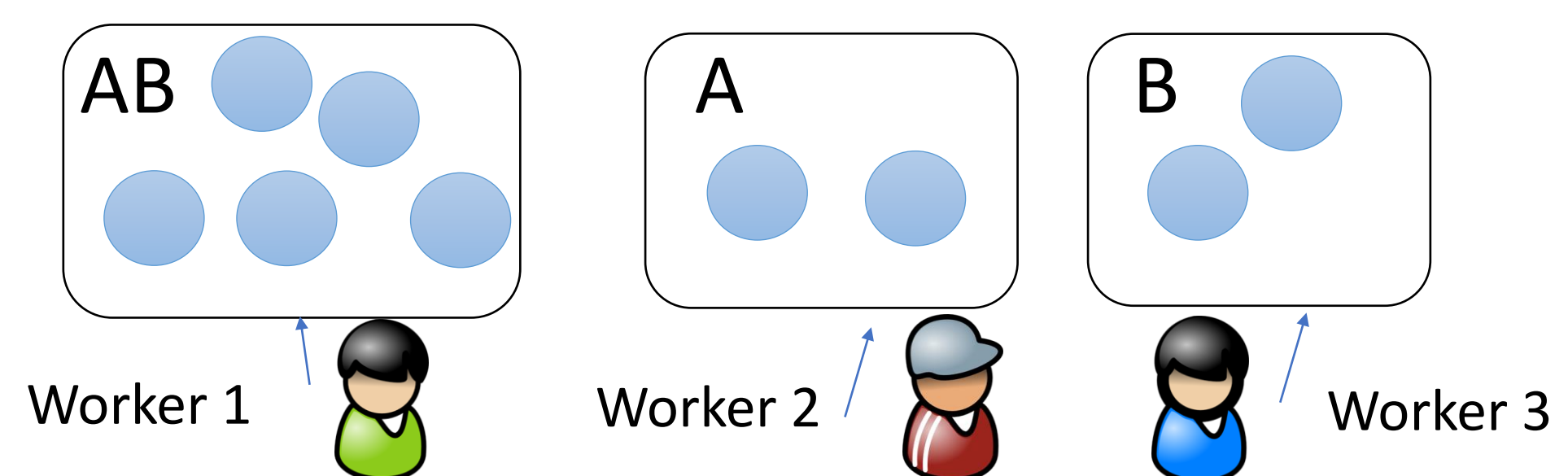
## Experiment 1 — Canis Animals

- Data: 800+ photos of 7 categories
- Qualification Task: a flat classification with 200 photos
- Collect 6420 answers from 152 workers
- Easily mistaken pair:
  - Alaskan Malamute & Husky

Scheme	Total Accuracy
A: Samoyed, Coyote, Wolf B: others	0.875
A: Alaskan malamute, Coyote B: others	0.590
A: wolf B: others	0.765
Flat classification (majority vote)	0.833
Flat classification (EM)	0.767

Result of Experiment 1 on AMT

Experiment 2 — Reptile and Amphibian Animals

- Data: 1000+ photos of 10 categories of animals
- Qualification Task: 200 photos & 307 workers
- Generate hierarchical schemes with 3-6 sub-tasks
- We double the number of workers in flat classification in order to compare the accuracy at the same cost.

Majority Voting by 3 (flat 6)

0.86 0.855 0.845 0.82 0.755
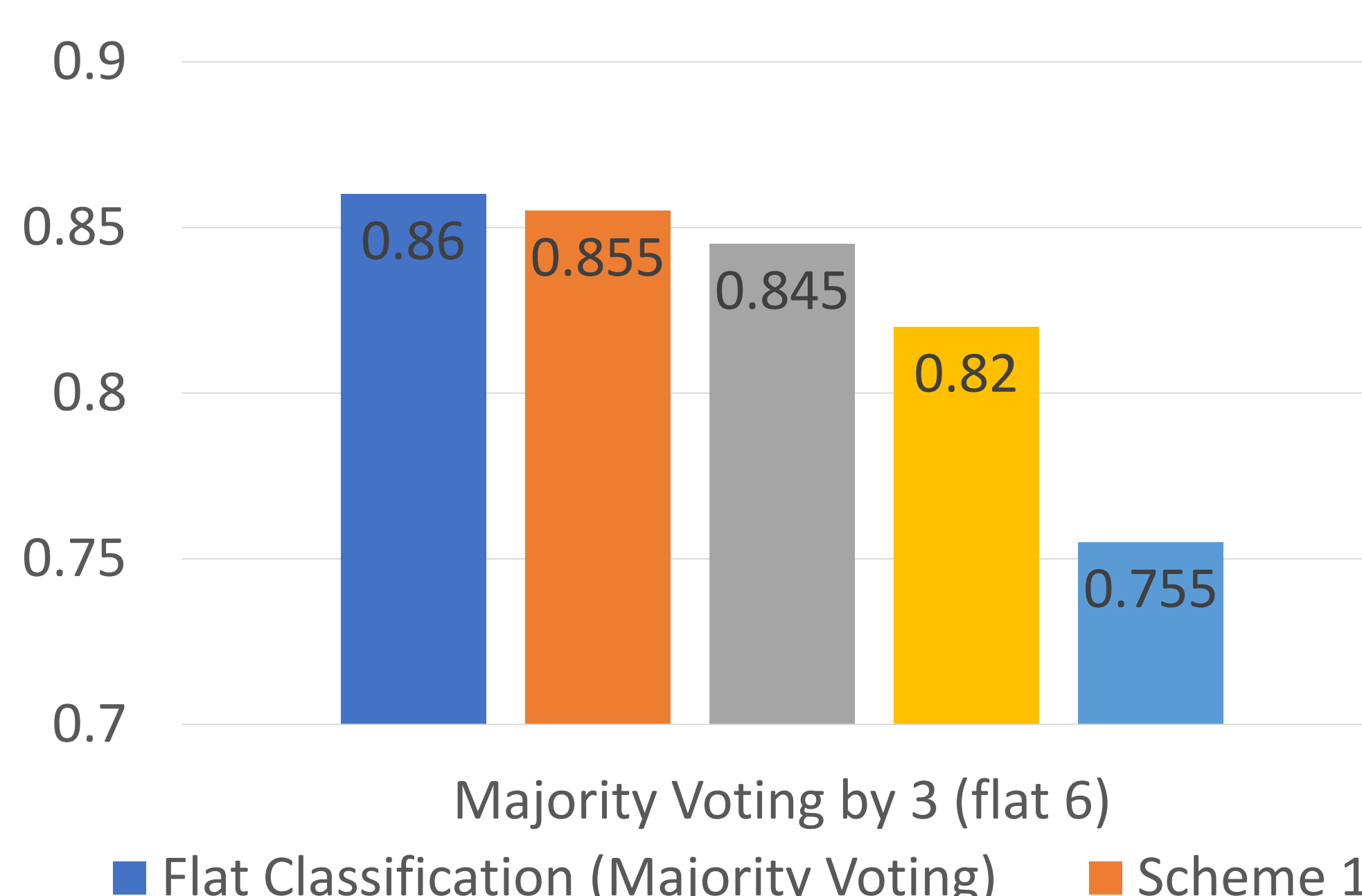
Result of Experiment 2 on AMT

■ Flat Classification (Majority Voting) ■ Scheme 1 ■ Scheme 202 ■ Scheme 6422 ■ Bottom Scheme

Summary

- Published 2 experiments on Amazon Mechanical Turk.
- Compare the accuracy of flat and hierarchical schemes with majority voting and EM-based weighted voting.
- Conclusion: Hierarchical schemes improve the accuracy if we choose an appropriate hierarchy by our algorithm.

Majority Voting by 5 (flat 10)

0.863 0.885 0.89 0.83 0.86