Learning Attributes from the Crowdsourced Relative Labels (Supplementary Materials)

Tian Tian, Ning Chen, Jun Zhu
Tsinghua University, Beijing, China
{tiant16@mails, ningchen@, dcszj@}tsinghua.edu.cn

A Derivation Details of the Bayesian Aggregating Model

In this section, we first present the derivation details about the evidence lower bound of the Bayesian aggregating model, then we show the gradients for updating the variational distribution and the model parameters.

A.1 Evidence Lower Bound

According to our Bayesian aggregating model, the expansion of the evidence lower bound (ELBO) is:

$$\begin{aligned}\mathcal{L}(\gamma,\phi,\psi,\tau) &= \mathbb{E}_q[\log p(\boldsymbol{c},\boldsymbol{z},\boldsymbol{y},\boldsymbol{b}|\boldsymbol{\theta})] - \mathbb{E}_q[\log q(\boldsymbol{c},\boldsymbol{z},\boldsymbol{y})] \\ &= \mathbb{E}_q[\log p(\boldsymbol{b}|\boldsymbol{y},\boldsymbol{c})] + \mathbb{E}_q[\log p(\boldsymbol{y}|\boldsymbol{z},\sigma)] + \mathbb{E}_q[\log p(\boldsymbol{z}|\boldsymbol{\alpha})] + \mathbb{E}_q[\log p(\boldsymbol{c}|\boldsymbol{\beta})] - \\ &\quad \mathbb{E}_q[\log q(\boldsymbol{c})] - \mathbb{E}_q[\log q(\boldsymbol{z})] - \mathbb{E}_q[\log q(\boldsymbol{y})].\end{aligned} \tag{1}$$

In the formula above, we have seven terms to specify. We will present the detailed derivations for each of them.

The first term is the expected log-likelihood. Since our likelihood is defined on the distribution of $\boldsymbol{y}$, this expectation can be calculated by plug the variational distribution $q(\boldsymbol{y})$ into the likelihood as

$$\begin{aligned}\mathbb{E}_q[\log p(\boldsymbol{b}|\boldsymbol{y},\boldsymbol{c})] &= \sum_t \mathbb{E}_{q(c_t)} \sum_{(i,j,k)} \mathbb{E}_{q(\boldsymbol{y})}\left[\log p(b^t_{i,j,s}|\boldsymbol{y}_i,\boldsymbol{y}_j,\boldsymbol{y}_s,c_t)\right] \\ &= \sum_t \mathbb{E}_{q(c_t)} \sum_{(i,j,s)} \left[\log \int p(b^t_{i,j,s}|\boldsymbol{y}_i,\boldsymbol{y}_j,\boldsymbol{y}_s,c_t)q(\boldsymbol{y})d\boldsymbol{y}\right] \\ &= \sum_t \mathbb{E}_{q(c_t)} \sum_{(i,j,s)} \log p\left(b^t_{i,j,s}|q(\boldsymbol{y}_i,\boldsymbol{y}_j,\boldsymbol{y}_s),c_t\right) \\ &= -\sum_t\sum_k \gamma_{t,k} \sum_{(i,j,s)} \Big[b^t_{i,j,s}D^t_{i,j} + (1-b^t_{i,j,s})D^t_{i,0} + \\ &\quad \log\left(e^{-D^t_{i,j}} + e^{-D^t_{i,s}} + e^{-D^t_{i,0}}\right)\Big],\end{aligned} \tag{2}$$

where $D^t_{i,j}$ is defined as the Jensen-Shannon divergence between two distributions $P = q(y_{i,k})$ and $Q = q(y_{j,k})$. It has a form of $D^t_{i,j} = \mathrm{JS}(P\|Q) = [\mathrm{KL}(P\|H) + \mathrm{KL}(Q\|H)]/2$, where $H = (P+Q)/2$. $D^t_{i,0} = \mathrm{JS}(P\|U)$, where $U = \mathcal{N}(0,\sigma^2)$.

For the second term, we have

$$\begin{aligned}
&\mathbb{E}_q[\log p(\boldsymbol{y}|\boldsymbol{z},\sigma)] \qquad (3)\\
= &-\frac{1}{2\sigma^2}\sum_n \mathbb{E}_q\big[(\boldsymbol{y}_n-\boldsymbol{\mu}_n)^\top(\boldsymbol{y}_n-\boldsymbol{\mu}_n)\big] - NK\log\sigma - \frac{NK}{2}\log(2\pi)\\
= &-\frac{1}{2\sigma^2}\left(\sum_n \mathbb{E}_{q(\boldsymbol{y}_n)}[\boldsymbol{y}_n^\top\boldsymbol{y}_n] - 2\sum_n \mathbb{E}_{q(\boldsymbol{\mu}_n)}[\boldsymbol{\mu}_n]^\top\mathbb{E}_{q(\boldsymbol{y}_n)}[\boldsymbol{y}_n] + \sum_n \mathbb{E}_{q(\boldsymbol{\mu}_n)}[\boldsymbol{\mu}_n^\top\boldsymbol{\mu}_n]\right) -\\
&NK\log\sigma - \frac{NK}{2}\log(2\pi)\\
= &-\frac{1}{2\sigma^2}\Big(\sum_n [K\tau_n^2 + \boldsymbol{\psi}_n^\top\boldsymbol{\psi}_n] - 2\sum_n\sum_k [(\phi_{l_n,k,1}\cdot 1 + \phi_{l_n,k,-1}\cdot -1)\psi_{n,k}] +\\
&\sum_n\sum_k(\phi_{l_n,k,-1}+\phi_{l_n,k,1})\Big) - NK\log\sigma - \frac{NK}{2}\log(2\pi).
\end{aligned}$$

For the third term, we have:

$$\begin{aligned}
\mathbb{E}_q[\log p(\boldsymbol{z}|\boldsymbol{\alpha})] &= \sum_{m,k}\mathbb{E}_{q(z_{m,k})}[\log p(z_{m,k}|\boldsymbol{\alpha})] \qquad (4)\\
&= \sum_{m,k}[\phi_{m,k,-1}\log\alpha_{-1} + \phi_{m,k,0}\log\alpha_0 + \phi_{m,k,1}\log\alpha_1].
\end{aligned}$$

For the fourth term, we have:

$$\mathbb{E}_q[\log p(\boldsymbol{c}|\boldsymbol{\beta})] = \sum_t \mathbb{E}_{q(c_t)}[\log p(c_t|\boldsymbol{\beta})] = \sum_t\sum_k[\gamma_{t,k}\log\beta_k]. \qquad (5)$$

For the fifth term, we have:

$$\mathbb{E}_q[\log q(\boldsymbol{c})] = \sum_t\sum_k \gamma_{t,k}\log\gamma_{t,k}. \qquad (6)$$

For the sixth term, we have:

$$\mathbb{E}_q[\log q(\boldsymbol{z})] = \sum_{m,k}[\phi_{m,k,-1}\log\phi_{m,k,-1} + \phi_{m,k,0}\log\phi_{m,k,0} + \phi_{m,k,1}\log\phi_{m,k,1}]. \qquad (7)$$

For the seventh term, we have:

$$\begin{aligned}
\mathbb{E}_q[\log q(\boldsymbol{y})] &= \sum_n\sum_k \mathbb{E}_{q(y_{n,k})}[\log q(y_{n,k})] \qquad (8)\\
&= -\sum_n K\log\tau_n - \frac{NK}{2}\log(2\pi) - \frac{NK}{2}.
\end{aligned}$$

With above derivations for these seven terms, we can have an expanded formula for the ELBO.

We have

$$
\begin{aligned}
\mathcal{L}(\gamma,\phi,\psi,\tau) = & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (9)\\
-\sum_t \sum_k \gamma_{t,k} \sum_{(i,j,s)} & \Big[b^t_{i,j,s} D^t_{i,j} + (1-b^t_{i,j,s}) D^t_{i,0} + \log\big(e^{-D^t_{i,j}} + e^{-D^t_{i,s}} + e^{-D^t_{i,0}}\big)\Big] - \\
& \frac{1}{2\sigma^2}\Big(\sum_n [K\tau_n^2 + \boldsymbol{\psi}_n^\top \boldsymbol{\psi}_n] - 2\sum_n\sum_k [(\phi_{l_n,k,1}\cdot 1 + \phi_{l_n,k,-1}\cdot -1)\psi_{n,k}] + \\
& \sum_n\sum_k (\phi_{l_n,k,-1} + \phi_{l_n,k,1})\Big) - NK\log\sigma - \frac{NK}{2}\log(2\pi) + \\
& \sum_{m,k} [\phi_{m,k,-1}\log\alpha_{-1} + \phi_{m,k,0}\log\alpha_0 + \phi_{m,k,1}\log\alpha_1] + \\
& \sum_t\sum_k \gamma_{t,k}\log\beta_k - \sum_t\sum_k \gamma_{t,k}\log\gamma_{t,k} - \\
& \sum_{m,k} [\phi_{m,k,-1}\log\phi_{m,k,-1} + \phi_{m,k,0}\log\phi_{m,k,0} + \phi_{m,k,1}\log\phi_{m,k,1}] + \\
& \sum_n K\log\tau_n + \frac{NK}{2}\log(2\pi) + \frac{NK}{2}.
\end{aligned}
$$

For computational simplicity, we fix $\tau_n = \sigma_y$, then we have

$$
\begin{aligned}
\mathcal{L}(\gamma,\phi,\psi) = & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (10)\\
-\sum_t \sum_k \gamma_{t,k} \sum_{(i,j,s)} & \Big[b^t_{i,j,s} D^t_{i,j} + (1-b^t_{i,j,s}) D^t_{i,0} + \log\big(e^{-D^t_{i,j}} + e^{-D^t_{i,s}} + e^{-D^t_{i,0}}\big)\Big] - \\
& \frac{1}{2\sigma^2}\sum_n\sum_k \big[\psi_{n,k}^2 - 2(\phi_{l_n,k,1} - \phi_{l_n,k,-1})\psi_{n,k} + (\phi_{l_n,k,1} + \phi_{l_n,k,-1})\big] - \\
& \sum_t\sum_k \gamma_{t,k}\log\gamma_{t,k} + \sum_t\sum_k \gamma_{t,k}\log\beta_k + \\
& \sum_{m,k} [\phi_{m,k,-1}\log\alpha_{-1} + \phi_{m,k,0}\log\alpha_0 + \phi_{m,k,1}\log\alpha_1] - \\
& \sum_{m,k} [\phi_{m,k,-1}\log\phi_{m,k,-1} + \phi_{m,k,0}\log\phi_{m,k,0} + \phi_{m,k,1}\log\phi_{m,k,1}].
\end{aligned}
$$

where $D^t_{i,j} = \frac{1}{2\sigma^2}(\psi_{i,k} - \psi_{j,k})^2$ and $D^t_{i,0} = \frac{1}{2\sigma^2}\psi^2_{i,k}$.

A.2 Learning the Variational Parameters

In this part, we take the derivative of the ELBO with respect to the variational parameters. Then we can learn the variational distribution.

For updating $\boldsymbol{\phi}$, we have

$$
\begin{aligned}
\frac{\partial}{\partial\phi_{m,k,1}}\mathcal{L} &= \frac{1}{\sigma^2}\sum_n \mathbb{I}(l_n = m)(\psi_{n,k} - \frac{1}{2}) + \log\alpha_{z,1} - \log\phi_{m,k,1} - 1, \qquad (11)\\
\frac{\partial}{\partial\phi_{m,k,-1}}\mathcal{L} &= \frac{1}{\sigma^2}\sum_n \mathbb{I}(l_n = m)(-\psi_{n,k} - \frac{1}{2}) + \log\alpha_{z,-1} - \log\phi_{m,k,-1} - 1,\\
\frac{\partial}{\partial\phi_{m,k,0}}\mathcal{L} &= \log\alpha_{z,0} - \log\phi_{m,k,0} - 1.
\end{aligned}
$$

By setting $\partial\mathcal{L}/\partial\boldsymbol{\phi}$ to zero, we can get

$$\begin{aligned}
\phi_{m,k,1} &= \alpha_1 \exp(\frac{1}{\sigma^2}\sum_n \mathbb{I}(l_n = m)(\psi_{n,k} - \frac{1}{2}) - 1), \\
\phi_{m,k,-1} &= \alpha_{-1} \exp(\frac{1}{\sigma^2}\sum_n \mathbb{I}(l_n = m)(-\psi_{n,k} - \frac{1}{2}) - 1), \\
\phi_{m,k,0} &= \alpha_0 \exp(-1).
\end{aligned} \tag{12}$$

For updating $\boldsymbol{\gamma}$, we have

$$\begin{aligned}
\frac{\partial}{\partial\gamma_{t,k}}\mathcal{L} &= -\sum_{(i,j,s)\in\mathcal{T}_t}\left[b^t_{i,j,s}D^t_{i,j} + (1-b^t_{i,j,s})D^t_{i,0} + \log\left(e^{-D^t_{i,j}} + e^{-D^t_{i,s}} + e^{-D^t_{i,0}}\right)\right] + \\
& \quad \log\beta_k - \log\gamma_{t,k} - 1.
\end{aligned} \tag{13}$$

By setting $\partial\mathcal{L}/\partial\boldsymbol{\gamma}$ to zero, we can get

$$\gamma_{t,k} \propto \beta_k \exp\left(-\sum_{(i,j,s)\in\mathcal{T}_t}\left[b^t_{i,j,s}D^t_{i,j} + (1-b^t_{i,j,s})D^t_{i,0} + \log\left(e^{-D^t_{i,j}} + e^{-D^t_{i,s}} + e^{-D^t_{i,0}}\right)\right] - 1\right). \tag{14}$$

So these two kinds of parameters have close-form updating expressions.

For updating $\boldsymbol{\psi}$, we have

$$\begin{aligned}
\frac{\partial}{\partial\psi_{n,k}}\mathcal{L} = -\frac{1}{\sigma^2}\Bigg\{ \sum_t \gamma_{t,k} \Bigg(\\
\sum_{(n,j,s)\in\mathcal{T}_t}\left[\psi_{n,k} - b^t_{n,j,s}\psi_{j,k} - \frac{e^{-D^t_{n,j}}(\psi_{n,k}-\psi_{j,k}) + e^{-D^t_{n,s}}(\psi_{n,k}-\psi_{s,k}) + e^{-D^t_{n,0}}\psi_{n,k}}{e^{-D^t_{i,j}} + e^{-D^t_{i,s}} + e^{-D^t_{i,0}}}\right] + \\
\sum_{(i,n,s)\in\mathcal{T}_t}\left[b^t_{i,n,s}(\psi_{n,k}-\psi_{i,k}) - \frac{e^{-D^t_{i,n}}(\psi_{n,k}-\psi_{i,k})}{e^{-D^t_{i,j}} + e^{-D^t_{i,s}} + e^{-D^t_{i,0}}}\right] + \\
\sum_{(i,j,n)\in\mathcal{T}_t}\left[-\frac{e^{-D^t_{i,n}}(\psi_{n,k}-\psi_{i,k})}{e^{-D^t_{i,j}} + e^{-D^t_{i,s}} + e^{-D^t_{i,0}}}\right]\Bigg) + \psi_{n,k} - (\phi_{l_n,k,1} - \phi_{l_n,k,-1})\Bigg\}.
\end{aligned} \tag{15}$$

Since it is hard to find the close-form solution for $\partial\mathcal{L}/\partial\boldsymbol{\psi} = \mathbf{0}$, we use gradient based optimization methods to update these parameters.

A.3 Estimating the Model Parameters

Similar to the variational inference, we take the derivative of the ELBO with respect to the model parameters, and then update these values.

When updating $\boldsymbol{\alpha}$, since $\alpha_{-1} + \alpha_0 + \alpha_1 = 1$, we import the Lagrange multipliers to the lower bound, and then we have

$$\begin{aligned}
\frac{\partial}{\partial\alpha_1}\mathcal{L}_{[\boldsymbol{\alpha}]} &= \frac{\sum_{m,k}\phi_{m,k,1}}{\alpha_1} - \lambda, \\
\frac{\partial}{\partial\alpha_{-1}}\mathcal{L}_{[\boldsymbol{\alpha}]} &= \frac{\sum_{m,k}\phi_{m,k,-1}}{\alpha_{-1}} - \lambda, \\
\frac{\partial}{\partial\alpha_0}\mathcal{L}_{[\boldsymbol{\alpha}]} &= \frac{\sum_{m,k}\phi_{m,k,0}}{\alpha_0} - \lambda.
\end{aligned} \tag{16}$$

By setting $\partial \mathcal{L}_{[\boldsymbol{\alpha}]} / \partial \boldsymbol{\alpha}$ to zero, we can get

$$\begin{aligned} \alpha_1 &= \frac{\sum_{m,k} \phi_{m,k,1}}{MK}, \\ \alpha_{-1} &= \frac{\sum_{m,k} \phi_{m,k,-1}}{MK}, \\ \alpha_0 &= \frac{\sum_{m,k} \phi_{m,k,0}}{MK}. \end{aligned} \tag{17}$$

When updating $\boldsymbol{\beta}$, since $\sum_k \beta_k = 1$, similarly we import the Lagrange multipliers to the lower bound, and then we have

$$\frac{\partial}{\partial \beta_k} \mathcal{L}_{[\boldsymbol{\beta}]} = \frac{\sum_{t,k} \gamma_{t,k}}{\beta_k} - \lambda. \tag{18}$$

By setting $\partial \mathcal{L}_{[\boldsymbol{\beta}]} / \partial \boldsymbol{\beta}$ to zero, we can get

$$\beta_k = \frac{\sum_{t,k} \gamma_{t,k}}{TK}. \tag{19}$$

When updating σ, since it is hard to find a close-form expression for updating, we use grid search to find good values.

B More Details about the Nonparametric Initialization

The initialization strategy described in our paper involves the following nonparametric optimization problem:

$$\max_{\mathcal{R}, \{l_k\}} \sum_{k=1}^{K} \sum_{t \in l_k} f(\boldsymbol{h}_t, \boldsymbol{r}_k) - \lambda K, \tag{20}$$

where $\lambda = (1 - \rho)M$ is a similarity threshold, and $\rho \in [0, 1]$ is a relaxation factor. This problem admits two properties: (1) The final amount of the crowdsourced attributes K is potentially unbounded and automatically determined by the data, and (2) the similarity between each signature to its most similar crowdsourced attribute must be larger than ρM.

To see the first property, since the amount of the crowdsourced attributes K is also a variable in this objective, we need to learn this value in real time. So K will increases when data grows, which is potentially unbounded. To see the second property, once the similarity between a signature $\boldsymbol{h}_t$ to its most similar crowdsourced attribute $\boldsymbol{r}_k$ is smaller than ρM, we can create a new crowdsourced attribute whose signature is $\boldsymbol{h}_t$, then the similarity between $\boldsymbol{h}_t$ and this new crowdsourced attribute will be M, and the total change on the objective is $M - f(\boldsymbol{h}_t, \boldsymbol{r}_k) - \lambda = \rho M - f(\boldsymbol{h}_t, \boldsymbol{r}_k) > 0$.

These two properties provide us an iterative way to find a local optimum of the problem. The final algorithm is similar as K-means' two-step update. For each iteration, we first assign each signature into its nearest crowdsourced attribute, then update the centers based on the assignments. The main difference between our algorithm and K-means is the assignment step. When dealing with a new signature $\boldsymbol{h}_t$, we find its most similar crowdsourced attribute $\boldsymbol{r}_k$. If the similarity between $\boldsymbol{h}_t$ and $\boldsymbol{r}_k$ is smaller than ρM, we generate a new crowdsourced attribute whose signature $\boldsymbol{r}_{K+1}$ is $\boldsymbol{h}_t$. Then we assign the signature $\boldsymbol{h}_t$ to this new crowdsourced attribute $\boldsymbol{r}_{K+1}$.

C Attribute Analysis for the Animals Dataset

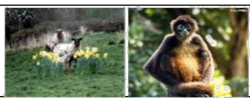
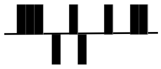
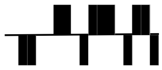
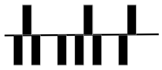
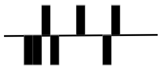
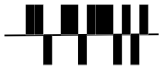
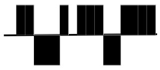
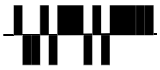
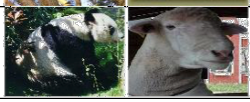
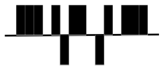
No.	Pos Examples	Neg Examples	Signatures	Worker Descriptions
1				Live on land or in water
2				Not hooved or hooved
3				Common or rare
4				Can climb trees or not
5				Stand out colors or blend in colors
6				Ground or tree dweller
7				Solid or patterned fur
8				Tropical or cool climate dweller
9				Herbivore or omnivorous

Figure 1: Crowdsourced attributes learned from the Animals dataset (Best viewed in color).

We visualize the crowdsourced attributes we learned from the Animals dataset in Fig. 1. The meanings of the columns are the same to them in the yellow flowers dataset experiment. Signature legend is also shown in Appendix C. When $\rho = 0.55$ we have 10 attributes in the results. Since 1 attribute is compatible with less than 4 categories, it is removed and then we show 9 crowdsourced attributes here. The results show that the 9 discovered attributes cover a wide range of animal properties. Some attributes are widely recognized, such as the living habitat, hooved or not and ubiquitous or rare. Other attributes, such as the ability to climb trees, herbivore versus omnivorous, etc, are more subtle. An interesting observation is that most of these attributes have meanings beyond the visual information expressed by the pictures. They may come from the diverse knowledge backgrounds of different annotators.

D Signature Legends for the Two Nature Scene Datasets

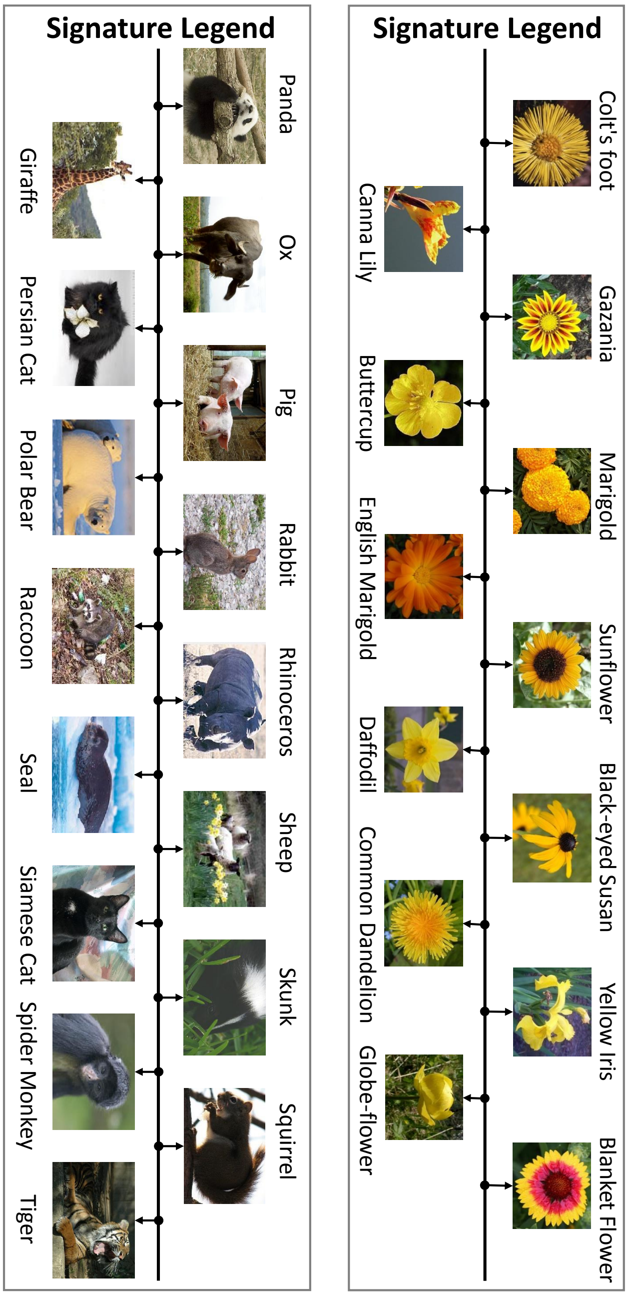
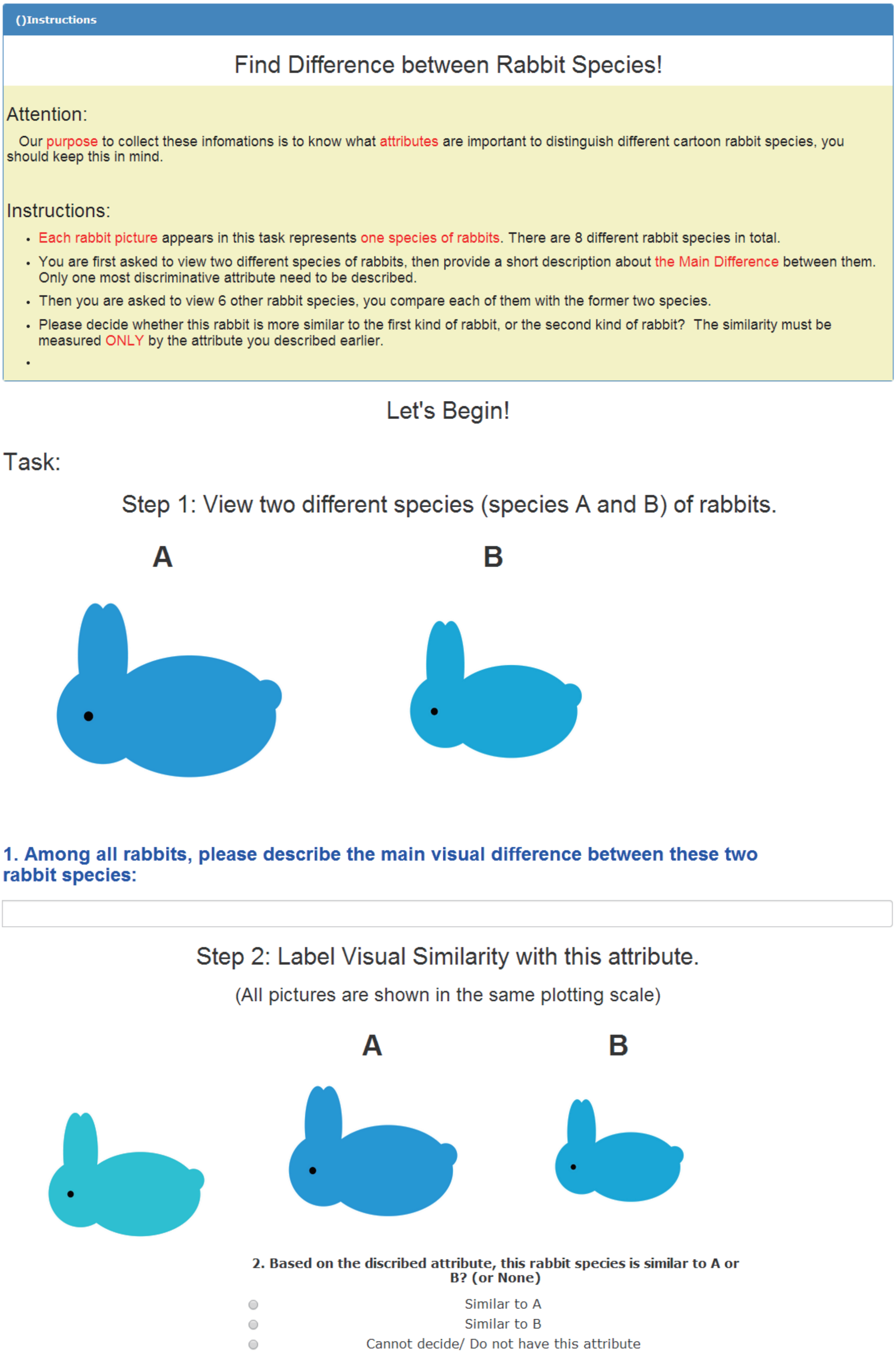
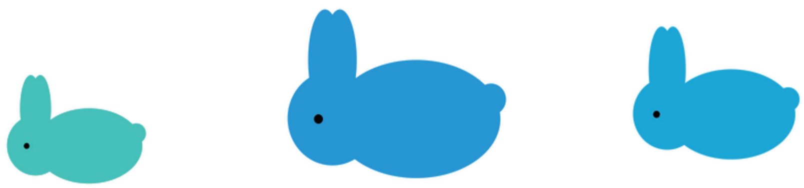


Figure 2: Signature legends for the yellow flowers dataset and the animals dataset (Best viewed in color).

E Analogical Interface for the Rabbits dataset



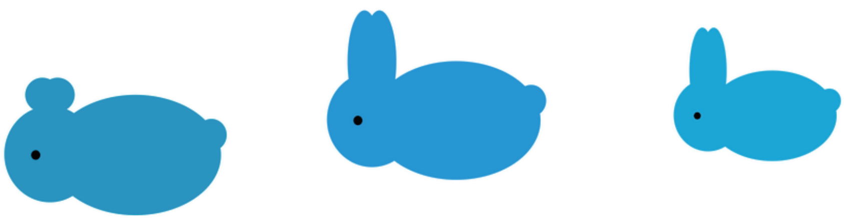
A B



3. Based on the discribed attribute, this rabbit species is similar to A or B? (or None)

- Similar to A
- Similar to B
- Cannot decide/ Do not have this attribute

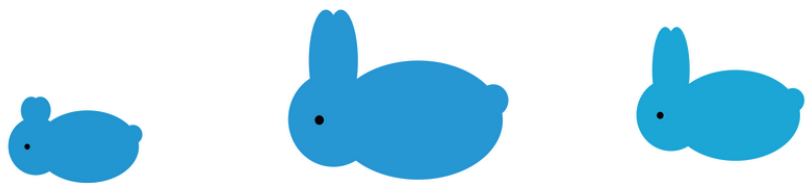
A B



4. Based on the discribed attribute, this rabbit species is similar to A or B? (or None)

- Similar to A
- Similar to B
- Cannot decide/ Do not have this attribute

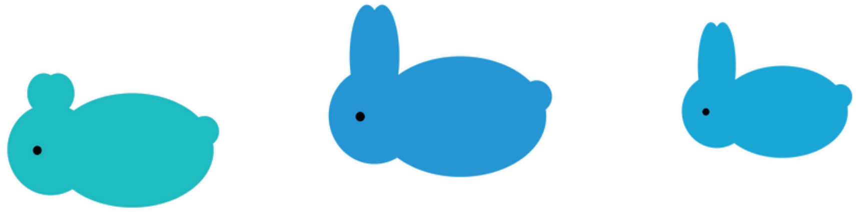
A B



5. Based on the discribed attribute, this rabbit species is similar to A or B? (or None)

- Similar to A
- Similar to B
- Cannot decide/ Do not have this attribute

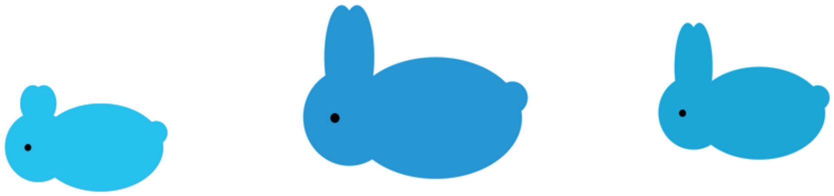
A B



6. Based on the discribed attribute, this rabbit species is similar to A or B? (or None)

- Similar to A
- Similar to B
- Cannot decide/ Do not have this attribute

A **B**



7. Based on the discribed attribute, this rabbit species is similar to A or B? (or None)

- Similar to A
- Similar to B
- Cannot decide/ Do not have this attribute

Submit