Permutation-Invariant Consensus over Crowdsourced Labels

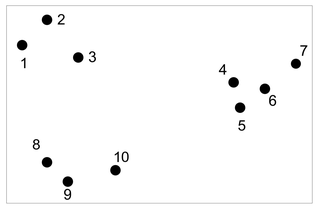
Michael Giancola (CS/MA)
Professor Jacob Whitehill (CS)
Professor Randy Paffenroth (MA)

Abstract

This Major Qualifying Project introduces a novel crowdsourcing consensus model and inference algorithm – which we call PICA (Permutation-Invariant Crowdsourcing Aggregation) – that is designed to recover the ground-truth labels of a dataset while being *invariant* to the class permutations enacted by the different annotators. This is particularly useful for settings in which annotators may have systematic confusions about the meanings of different classes, as well as clustering problems (e.g., dense pixel-wise image segmentation) in which the names/numbers assigned to each cluster have no inherent meaning. The PICA model is constructed by endowing each annotator with a doubly-stochastic matrix (DSM), which models the probabilities that an annotator will perceive one class and transcribe it into another. We conduct simulations and experiments to show the advantage of PICA compared to Majority Vote for three different clustering/labeling tasks. We also explore the conditions under which PICA provides better inference accuracy compared to a simpler but related model based on right-stochastic matrices by [4]. Finally, we show that PICA can be used to crowdsource responses for dense image segmentation tasks, and provide a proof-of-concept that aggregating responses in this way could improve the accuracy of this labor-intensive task.

Motivation

Objective: Determine a clustering of the points by aggregating the annotators responses.



	Labels									
	1	2	3	4	5	6	7	8	9	10
Annotator 1	1	1	1	2	2	2	2	3	3	3
Annotator 2	1	1	1	2	2	2	1	3	3	2
Annotator 3	3	3	2	1	1	1	1	2	2	2
Majority Vote	1	1	1	2	2	2	**1**	3	3	**2**
PICA	1	1	1	2	2	2	2	3	3	3
Ground-truth	1	1	1	2	2	2	2	3	3	3

Figure 1: Entries give the cluster which each annotator assigns to the given point.

Model Description

We define the probability that annotator i assigns label $l \in \Omega$ to example j, given the ground-truth label z, style matrix S, and accuracy a as:

$$p(\mathsf{L}_j^{(i)} = l \mid \mathsf{z}_j = z, \mathsf{S}^{(i)} = S, \mathsf{a}^{(i)} = a)$$
$$\doteq \quad a \times S_{zl} + (1-a) \times \frac{\sum_{z' \neq z} S_{z'l}}{n-1}$$
$$= \quad \left(S \begin{bmatrix} a & \frac{1-a}{n-1} & \cdots & \\ \frac{1-a}{n-1} & a & \frac{1-a}{n-1} & \cdots \\ \vdots & & \ddots & \\ \frac{1-a}{n-1} & & & a \end{bmatrix} \right)_{zl}$$

Proof: Product of Doubly-Stochastic Matrices

If the dot product of two DSMs is a DSM, we can "fold" the accuracy matrix into the style matrix (see Model Description) and thus reduce our model to a single DSM.

Let A and B be two arbitrary DSMs.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \vdots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ & & \vdots & \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix}$$

Column j of AB has the form:

$$AB_{*j} = \begin{bmatrix} a_{11}b_{1j} + a_{12}b_{2j} + \cdots + a_{1n}b_{nj} \\ a_{21}b_{1j} + a_{22}b_{2j} + \cdots + a_{2n}b_{nj} \\ \vdots \\ a_{n1}b_{1j} + a_{n2}b_{2j} + \cdots + a_{nn}b_{nj} \end{bmatrix}$$

Taking the sum of the elements in column j, we can reorder the terms and find:

$$\begin{aligned} & b_{1j}(a_{11} + \cdots + a_{n1}) \\ & + b_{2j}(a_{12} + \cdots + a_{n2}) \\ & + \dots \\ & + b_{nj}(a_{1n} + \cdots + a_{nn}) \\ & = b_{1j} + b_{2j} + \cdots + b_{nj} \\ & = 1 \end{aligned}$$

A similar argument shows that each row of AB sums to 1.

Expectation-Maximization

Objective: Simultaneously infer the ground-truth labels (for each example) and style matrices (for each annotator).

E-Step:

$$\begin{aligned} & p(\mathsf{z}_j = z_j \mid \mathsf{L}^{(1)} = L^{(1)}, \dots, \mathsf{L}^{(m)} = L^{(m)}, \\ & \qquad \mathsf{S}^{(1)} = S^{(1)}, \dots, \mathsf{S}^{(m)} = S^{(m)}) \\ \propto \quad & p(z_j) \prod_{i : L_j^{(i)} \neq \epsilon} S^{(i)}_{z, L_j^{(i)}} \end{aligned}$$

where $S^{(i)}_{z,L_j^{(i)}}$ is the zth row and $L_j^{(i)}$th column of $S^{(i)}$.

M-Step:

$$\begin{aligned} & Q(S^{(1)}, \dots, S^{(m)}) \\ = \quad & E[\log p(L^{(1)}, \dots, L^{(m)}, z_1, \dots, z_d | S^{(1)}, \dots, S^{(m)})] \\ = \quad & E\left[\log \prod_j \left(p(z_j) \prod_{i : L_j^{(i)} \neq \epsilon} p(L_j^{(i)} | z_j, S^{(i)}) \right) \right] \\ = \quad & \sum_{ij} \sum_{z_j} \log \left(S^{(i)}_{z, L_j^{(i)}} \right) \widetilde{p}(z_j) + \text{const.} \end{aligned}$$

Optimization over Doubly-Stochastic Matrices

Goal: Optimize a function of a doubly-stochastic matrix (DSM).

Problem: Standard gradient ascent procedures don't ensure that the gradient update is a DSM.

Sinkhorn Normalization [3]

- $A \in \mathbb{R}^{n \times n} \Rightarrow DSM$
- Iterative row/column normalizations

$$R_{i,j}(M) = \frac{M_{i,j}}{\sum_{k=1}^n M_{i,k}} \qquad C_{i,j}(M) = \frac{M_{i,j}}{\sum_{k=1}^n M_{k,j}}$$

$$\mathrm{SP}^s(A) = \begin{cases} A & \text{if } s = 0 \\ C(R(\mathrm{SP}^{s-1}(A))) & \text{otherwise} \end{cases}$$

Example:

$$A = \begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix}$$
$$R(A) = \begin{bmatrix} \frac{2}{7} & \frac{5}{7} \\ \frac{3}{7} & \frac{4}{7} \end{bmatrix}$$
$$C(R(A)) = \mathrm{SP}^1(A) = \begin{bmatrix} \frac{2}{5} & \frac{5}{9} \\ \frac{3}{5} & \frac{4}{9} \end{bmatrix}$$
$$\mathrm{SP}^2(A) \approx \begin{bmatrix} 0.421 & 0.577 \\ 0.578 & 0.422 \end{bmatrix}$$
$$\mathrm{SP}^3(A) \approx \begin{bmatrix} 0.422 & 0.577 \\ 0.577 & 0.422 \end{bmatrix}$$

Sinkhorn Propagation [1]

- Gradient of a function f of a DSM w.r.t. A
- Propagate derivative through normalizations

$$\frac{\partial f}{\partial A} = \frac{\partial f}{\partial \mathrm{vec}[C]} \frac{\partial \mathrm{vec}[C]}{\partial \mathrm{vec}[R]} \frac{\partial \mathrm{vec}[R]}{\partial \mathrm{vec}[C]} \cdots \frac{\partial \mathrm{vec}[R]}{\partial \mathrm{vec}[A]}$$

$$\frac{\partial f}{\partial \mathrm{vec}[C]} = \begin{bmatrix} \frac{\partial f}{\partial C_{11}} & \cdots & \frac{\partial f}{\partial C_{1n}} & \frac{\partial f}{\partial C_{21}} & \cdots & \frac{\partial f}{\partial C_{nn}} \end{bmatrix}$$

$$\frac{\partial \mathrm{vec}[C]}{\partial \mathrm{vec}[R]} = \begin{bmatrix} \frac{\partial C_{11}}{\partial R_{11}} & \cdots & \frac{\partial C_{1n}}{\partial R_{11}} & \frac{\partial C_{21}}{\partial R_{11}} & \cdots & \frac{\partial C_{nn}}{\partial R_{11}} \\ \vdots & & & & & \\ \frac{\partial C_{11}}{\partial R_{1n}} & & & & & \\ \frac{\partial C_{11}}{\partial R_{21}} & & & \ddots & & \vdots \\ \vdots & & & & & \\ \frac{\partial C_{11}}{\partial R_{2n}} & & & & & \\ \vdots & & & & & \\ \frac{\partial C_{11}}{\partial R_{nn}} & & & \cdots & & \frac{\partial C_{nn}}{\partial R_{nn}} \end{bmatrix}$$

$$\frac{\partial C_{ij}}{\partial R_{xy}} = \delta(i,x) \left(\frac{\delta(j,y)}{\sum_{k=1}^n R_{ky}} - \frac{R_{xy}}{(\sum_{k=1}^n R_{ky})^2} \right)$$

where $\delta(\cdot)$ is the Kronecker delta function.

Comparison to Naïve Algorithm

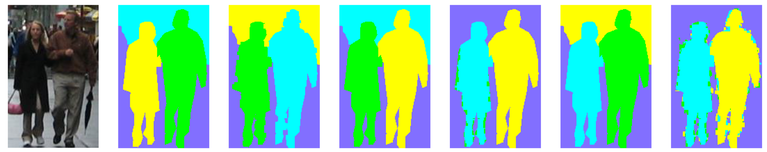
Goal: Describe a naive approach which could perform as well as our model in some situations.

Algorithm:

- Randomly pick an annotator i as the "leader".
- Using i's labels,
- For each annotator $j \neq i$,
  - For each class $k = 1, ..., n$
    - Find the symbol used most frequently by j to express class k according to i's labels
    - Un-permute j's labels based on the inferred permutation
- Conduct majority vote over the unpermuted labels of all annotators

Image Segmentation

Goal: Evaluate our model's ability to decrease the cost of generating image segmentations.
Problem: Reconstruct a (pixel-wise) image segmentation from multiple noisy segmentations.
Note: In this experiment, each *pixel* is a label.

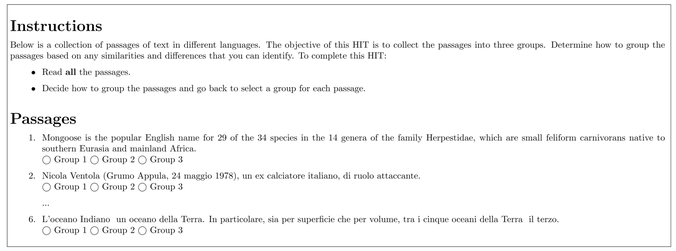


(a) Original Image [2,5] (b) Ground-truth (c) Noisy labeler (d) PICA (DSM) (e) RSM (f) Baseline (g) MV

Model	Image							
	"couple"		"flag"		"light"		"people"	
	%	C.E.	%	C.E.	%	C.E.	%	C.E.
PICA	100	65.5	**99.4**	**.077**	99.8	**.020**	99.0	48.2
RSM	80.4	**13.9**	98.7	.139	99.1	.040	99.0	**39.6**
BASE	100	–	99.2	–	99.8	–	98.9	–
MV	77.5	–	81.5	–	88.9	–	86.0	–

Text Passage Clustering

Problem: Infer an aggregate clustering from those of several annotators.



Instructions

Below is a collection of passages of text in different languages. The objective of this HIT is to collect the passages into three groups. Determine how to group the passages based on any similarities and differences that you can identify. To complete this HIT:

- Read **all** the passages.
- Decide how to group the passages and go back to select a group for each passage.

Passages

1. Mongoose is the popular English name for 29 of the 34 species in the 14 genera of the family Herpestidae, which are small feliform carnivorans native to southern Eurasia and mainland Africa.
   ○ Group 1 ○ Group 2 ○ Group 3
2. Nicola Ventola (Grumo Appula, 24 maggio 1978), un ex calciatore italiano, di ruolo attaccante.
   ○ Group 1 ○ Group 2 ○ Group 3

...

6. L'oceano Indiano un oceano della Terra. In particolare, sia per superficie che per volume, tra i cinque oceani della Terra il terzo.
   ○ Group 1 ○ Group 2 ○ Group 3

Figure 1: The text passage clustering task we posted on Amazon Mechanical Turk.

Model	%	C.E.
PICA	93	2.54
RSM	**96**	**1.65**
BASE	90.5	–
MV	89	–

Rare Class Simulation

Goal: Illustrate the benefit of PICA over the similar RSM-based model [4].
Problem: Cluster arbitrary data where one class is highly unrepresented.

- Generated 100 examples and assigned ground-truth labels
- $z_j \in \Omega = \{$ 'a', 'b', 'c' $\}$, with $p(\text{'a'}) = 0.5$, $p(\text{'b'}) = 0.45$, $p(\text{'c'}) = 0.05$

Model	%	C.E.
PICA	**88**	**4.01**
RSM	79	6.32
BASE	62	–
MV	47	–

References

[1] Ryan Prescott Adams and Richard S. Zemel. Ranking via Sinkhorn Propagation. *arXiv preprint arXiv:1106.1925*, 2011.

[2] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, and William T. Freeman. LabelMe: a database and web-based tool for image annotation. *International journal of computer vision*, 77(1-3):157–173, 2008.

[3] Richard Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2):876–879, 1964.

[4] Padhraic Smyth, Usama M. Fayyad, Michael C. Burl, Pietro Perona, and Pierre Baldi. Inferring ground truth from subjective labelling of venus images. pages 1085–1092, 1995.

[5] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. 2017.