WiCV 2016: Women in Computer Vision

Ambiguity Helps: Classification with Disagreements in Crowdsourced Annotations

Viktoriia Sharmanska

University of Sussex

Sussex Machine Learning for Computational Linguistics, Network analysis, and Computer vision

Joint work with Daniel Hernández-Lobato, José Miguel Hernández-Lobato, and Novi Quadrianto

Ambiguity

Examples of ambiguous tasks: deciding whether a place is "fun" or "not fun" from an image.

*Is this a **fun** place to be?*

©by Lisa, Milhouse, and Homer

Collecting attribute annotations using Amazon Mechanical Turk

What Do We Propose?

- To re-think the common practice in crowdsourcing (take the majority vote among trusted annotators and disregard disagreements).
- **Technical contribution**:
  A framework to incorporate annotation disagreements into the learning process of a classifier.
- **Setup**:
  We are given data instances $\mathbf{x}_n$, their associated labels y_n, and label confidence $\mathbf{x}_n^{\text{conf}}$, for example, agreement among annotators (in the *cartoon* example, **it is $^2/_3$ for CVPR as a fun place to be**).

Ambiguity Model GPC$^{\text{conf}}$

Gaussian process classification (GPC) Under this model $p(y_n|\mathbf{x}_n, f) = \Theta(y_n f(\mathbf{x}_n))$ for class label $y_n \in \{-1, 1\}$, where $\Theta(\cdot)$ denotes Heaviside step function and f is assumed to be generated by a Gaussian process, *i.e.*, $f(\mathbf{x}_n) \sim \mathcal{GP}(0, k(\mathbf{x}_n, \cdot))$, for some covariance function $k(\mathbf{x}_n, \cdot)$.

GPC with annotation disagreements (GPC$^{\text{conf}}$)
We introduce another latent function g that takes into account the confidence in label annotations $\mathbf{x}_n^{\text{conf}}$, $g(\mathbf{x}_n^{\text{conf}}) \sim \mathcal{GP}(0, k(\mathbf{x}_n^{\text{conf}}, \cdot))$.

Ambiguity Model GPC$^{\text{conf}}$

The GPC$^{\text{conf}}$ model is:

$$p(y_n|\mathbf{x}_n, \mathbf{x}_n^{\text{conf}}, f, g) = \Theta\Big(y_n f(\mathbf{x}_n)\Big)^{1-\Theta\left(g(\mathbf{x}_n^{\text{conf}})\right)} \left({}^1\!/_2\right)^{\Theta\left(g(\mathbf{x}_n^{\text{conf}})\right)} .$$

- For un-ambiguous data points, the standard likelihood is used ($g(\mathbf{x}_n^{\text{conf}})$ is *negative*);
- For ambiguous data points **CVPR is a fun place to be**, the influence is reconsidered when learning the concept **fun** ($g(\mathbf{x}_n^{\text{conf}})$ is *positive*).

Inference: Confidence in Annotations

For a particular instance $\mathbf{x}_n$, $\mathbf{x}_n^{\text{conf}}$, y_n, by marginalizing g, the associated term in the likelihood function of f is:

$$p(g(\mathbf{x}_n^{\text{conf}}) > 0)\,\frac{1}{2} \;+\; (1 - p(g(\mathbf{x}_n^{\text{conf}}) > 0))\,\Theta(y_n f(\mathbf{x}_n)).$$

During inference, an instance with less **confidence** will have its likelihood being **ignored** ($1/2$), having reduced influence (a mixture of $1/2$ and step likelihood), or being as informative as confident instances (a step likelihood).

All you need in this life is **ignorance** *and* **confidence**, *and then success is sure.*

Mark Twain

Posterior Inference: Expectation Propagation for GPC$^{\text{conf}}$

The posterior is approximated by the product of two Gaussians:

$$\underbrace{\frac{\prod_{n=1}^{N} p(y_n|\mathbf{f}, \mathbf{g}, \mathbf{x}_n, \mathbf{x}_n^{\text{conf}}) p(\mathbf{f}) p(\mathbf{g})}{p(\mathbf{y}|\mathbf{X}, \mathbf{X}^{\text{conf}})}}_{\text{posterior}} \approx \mathcal{N}(\mathbf{f}|\mathbf{m}_f, \mathbf{\Sigma}_f)\mathcal{N}(\mathbf{g}|\mathbf{m}_g, \mathbf{\Sigma}_g)\,.$$

Each factor $p(y_n|\mathbf{x}_n, \mathbf{x}_n^{\text{conf}}, f, g)$ is approximated as:

$$\overline{z}_n \mathcal{N}(f(\mathbf{x}_n)|\overline{m}_f, \overline{v}_f)\mathcal{N}(g(\mathbf{x}_n^{\text{conf}})|\overline{m}_g, \overline{v}_g)\,.$$

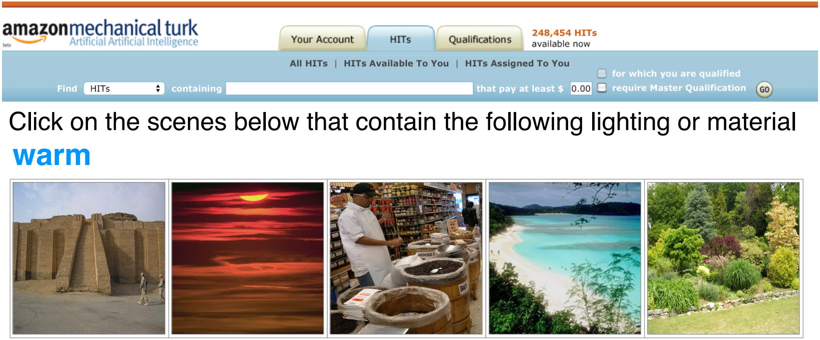
The parameters $\overline{z}_n$, $\overline{m}_f$, $\overline{m}_g$, $\overline{v}_f$ and $\overline{v}_g$ can be obtained from the log of:

$$Z_n = \underbrace{\Phi\left({}^{m^{-n}}\!/\!\sqrt{v^{-n}}\right)\Phi\left({}^{-\mu^{-n}}\!/\!\sqrt{\nu^{-n}}\right) + \Phi\left({}^{\mu^{-n}}\!/\!\sqrt{\nu^{-n}}\right)/2}_{\text{novelty: prior work GPC+ requires a quadrature approach}},$$

where $m^{-n}, v^{-n}, \mu^{-n}, \nu^{-n}$ are parameters of a (cavity) distribution, a posterior minus the approximate factor.

Code is available at author's homepage.

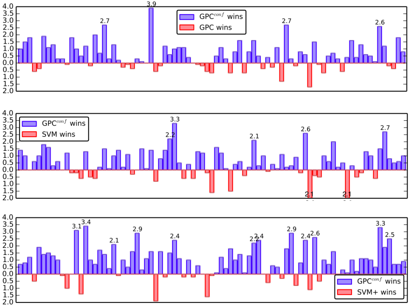
Results: Ambiguity in Recognizing Semantic Attributes

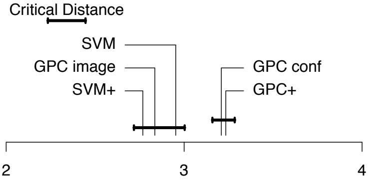


- SUN Attribute dataset: 83 attributes, as confidence we use MTurk annotations of attributes being present in the images.

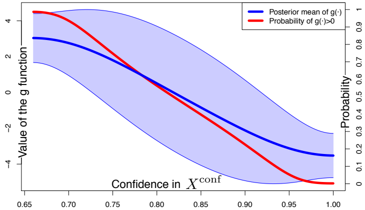
Results: Ambiguity in Recognizing Semantic Attributes

- Pairwise comparison in terms of difference in accuracies and statistical comparison of all methods using Demšar:

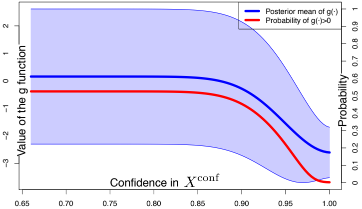




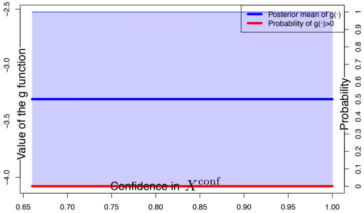
Analysis of the confidence in annotations



Ignored



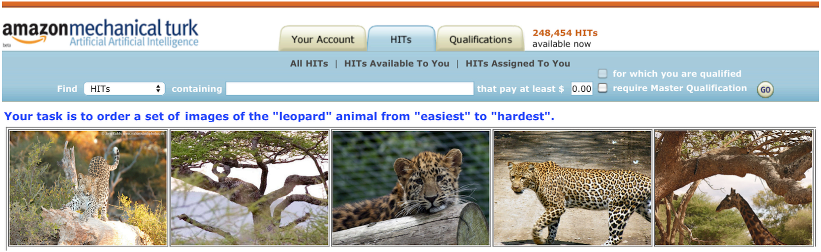
Reduced influence



Informative

Representative posterior mean of the g function and 1-std confidence interval (solid blue curve) and the probability of $g > 0$ (solid red curve) for three different cases.

Results: Ambiguity to Distinguish Easy from Hard Images



- AwA dataset: 8 animal classes; easy-hard score annotation is available per image that shows how easy/hard it is to spot the animal based on MTurk user study

Results: Ambiguity to Distinguish Easy from Hard Images

- The binary task is to distinguish easy from hard images of the class, where label confidence reflects the easy-hard score:

	GPC image	GPC^{conf} (ours) image+conf	SVM+ image+conf	SVM image
Chimp.	74.86 ± 0.8	74.93 ± 0.7	$\mathbf{75.07 \pm 0.7}$	73.71 ± 0.9
G.panda	80.64 ± 0.5	81.17 ± 0.6	$\mathbf{81.33 \pm 0.5}$	80.53 ± 0.6
Leo	81.67 ± 0.7	$\mathbf{82.00 \pm 0.7}$	80.58 ± 0.6	80.42 ± 0.8
Pers.cat	79.72 ± 0.4	$\mathbf{80.14 \pm 0.4}$	79.15 ± 0.7	78.17 ± 1.0
Hippo	72.85 ± 1.0	72.78 ± 1.1	$\mathbf{73.33 \pm 1.4}$	73.06 ± 1.1
Raccoon	78.57 ± 1.0	$\mathbf{78.81 \pm 0.8}$	76.98 ± 0.8	76.51 ± 0.6
Rat	$\mathbf{84.33 \pm 1.5}$	84.00 ± 1.5	83.50 ± 1.8	81.50 ± 1.8
Seal	48.00 ± 1.4	48.10 ± 1.2	48.50 ± 0.8	49.20 ± 0.8

- Running time

	GPC	GPC^{conf}	GPC+	SVM	SVM+
SUNAttribute	27m.	32m.	51m.	6m.	106m.
AwA	32m.	42m.	73m.	10m.	252m.

Summary

- We propose to incorporate annotation disagreements when learning a classifier for inherently ambiguous tasks.
- We do not remove ambiguous instances, and we do not redefine data collection process
- Future direction: deep disagreement, or how to incorporate ambiguos labels into deep neural networks.

Thank You!