Regionalized optimisation and message passing algorithms

Gregoire Sergeant-Perthuis

Laboratoire de Mathématiques de Lens
Université d'Artois

Séminaire du CRIL, March 31 2022

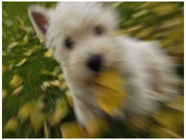
Reference

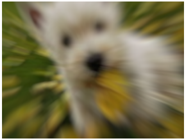
Presentation based on work in:

Regionalized optimization, arXiv:2201.11876 [SP22]

Motivation (1)

Data with multiple point of view on it: for example images of Dog with two **types of blurs at different intensity of blurring**





Motivation (2)

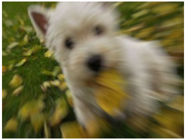
Cat with two **types of blur at different intensity of blurring**:

How to classify dogs and cats taking into account the extra data given by the different point of views?

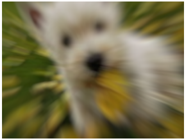
Framework: Structured Data (1)

Data: collection of images $(u_{i,j}, i \in \{0,1,2\}, j \in \{0,1\})$

$u_{1,0}$



$u_{1,1}$



$u_{0,0}$

$u_{2,0}$

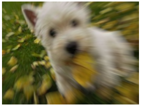
$u_{2,1}$

Framework: Structured Data (2)

Denote first kind of blurring as $B_1(\lambda)$ and second kind as $B_2(\lambda)$,

$u_{0,0}$

$u_{1,0}$



To go from $u_{0,0}$ to $u_{1,0}$,

$$u_{1,0} = B_1(0.3)[u_{0,0}]$$

Compatibility relations:

$$u_{0,0} = id[u_{0,0}]$$

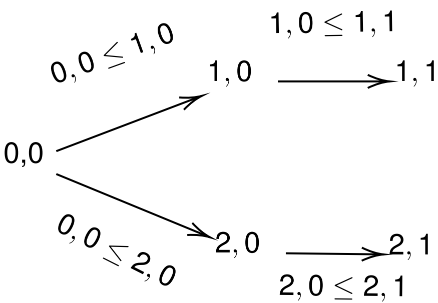
$$u_{1,0} = B_1(0.3)[u_{0,0}] \qquad u_{1,1} = B_1(0.5)[u_{1,0}]$$

$$u_{2,0} = B_2(0.1)[u_{0,0}] \qquad u_{2,1} = B_2(0.4)[u_{2,0}]$$

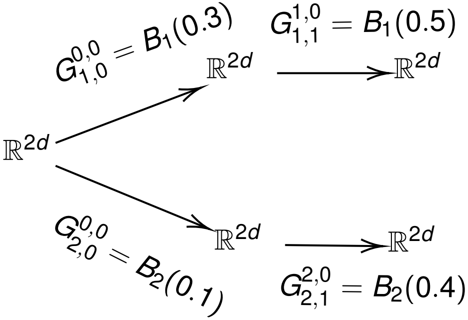
Framework: Structured Data (3)

Formally, compatibility relations are equivalent to saying that: $(u_{i,j}, i \in \{0,1,2\}, j \in \{0,1\})$ is a *section* of a *functor G* over a *partially ordered set* (poset) $\mathscr{A}$.

Poset $\mathscr{A}$:



Functor G:



where $\mathbb{R}^{2d}$ is the space in which the images live.

Framework: Structured Data (4)

Partially ordered set $\mathscr{A}$: a relation $\leq$ ($\subseteq \mathscr{A} \times \mathscr{A}$) such that,

1. $a \leq a$
2. (Transitivity) $b \leq a$ and $c \leq b$ then $c \leq a$
3. $b \leq a$ and $a \leq b$ then $a = b$

Functor G over a poset:

1. sends elements $a \in \mathscr{A}$ to a (vector) space $G(a)$
2. relations $b \leq a$ to (linear) morphisms between spaces

$$G_a^b : G(b) \to G(a)$$

3. Respects Transitivity:

$$G_a^b G_b^c = G_a^c$$

Framework: Structured Data (5)

Limit of a functor: ($\lim G$) set of collections ($u_a \in G(a), a \in \mathscr{A}$) that are compatible under the functor :

$$\forall b \leq a, \quad G_a^b(u_b) = u_a$$

Now: Data is the limit of a functor over a poset.

Framework: Loss (1)

To classify cats or dogs over a dataset $D = [(x^i, y^i), i = 1..N]$ of size N:
Cross entropy

$$l(\theta) = \frac{1}{N} \sum_{i=1..N} \ln p_\theta(y^i | x^i)$$

where $y = 0$ for a cat and $y = 1$ for a dog.

Framework: Loss (2)

In our case there are multiple points of views on the images: the dataset is a collection of samples $[(x^i_{a(i)}, y^i), i = 1..N]$ over **different view points** $a \in \mathscr{A}$ where $a(i)$ is the view point on the image, recall that possible values are:

$$(0,0),(1,0),(1,1),(2,0),(2,1)$$

For example for the following sample

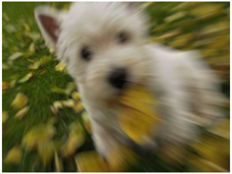
$$a(i) = (1,0)$$

Framework: Loss (3)

Dataset can be reorganized as collection of datasets $[(x_a^i, y^i), i = 1..N_a]$ for $a \in \mathscr{A}$.

The expression of the loss does not change with the point of view on the data,

$$l_{0,0} = \frac{1}{N_{0,0}} \sum_{i=1..N_{0,0}} \ln p_{\theta_{0,0}}(y^i | x_{0,0}^i)$$



$$l_{1,0} = \frac{1}{N_{1,0}} \sum_{i=1..N_{1,0}} \ln p_{\theta_{1,0}}(y^i | x_{1,0}^i)$$

Framework: Loss (4)

For a given point of view a, the previous loss is simply the cross entropy for the dataset restricted to this point of view:

$$l_a(p_{\theta_a}) = \frac{1}{N_a} \sum_{i=1..N_a} \ln p_{\theta_a}(y^i | x_a^i)$$

Formally, for each element of the poset $a \in \mathscr{A}$, we consider a collection of losses (functions) $l_a : G(a) \to \mathbb{R}$. We now call the points of view 'local' ($\to$ inspired by topology).

Framework: Loss (4)

Problem: How to optimize l_a for all points of view at the same time?
Answer?: Total loss is the sum of the losses? $l = \sum_{a \in \mathscr{A}} l_a$.

- Very redundant!
- $u \in \lim G$ is a 'global' reconstruction of 'local' points of view $u_a, a \in \mathscr{A}$ we want the loss to 'behave' the same way
- In our example, the non blurred image $u_{0,0}$ is enough to index the sections of G:
$$G \cong \mathbb{R}^{2d}$$
- However $l \neq l_{0,0}$. This loss **does not** behave well under 'global reconstruction'
- **NOT an answer**

Framework: Loss (5)

We follow the construction of Yedidia,Freeman, Weiss in the celebrated article *Constructing free-energy approximations and generalized belief propagation algorithms*[YFW05]. They use inclusion–exclusion principle to build an entropy on probability distribution compatible by marginalization.

- Good properties under 'global reconstruction' Proposition 2.2 [SP22]

Framework: Loss (6)

Inclusion–exclusion principle: simplest version for two set A, B then,

$$|A \cup B| = |A| + |B| - |A \cap B| \tag{0.1}$$

Rota in his celebrated article *On the foundations of combinatorial theory I. Theory of Möbius functions* [Rot64], extended inclusion–exclusion to any poset by introducing Möbius inversion.

Framework: Loss (7)

Two important functions for poset $\mathscr{A}$:

- ζ funtion of the poset, for any $f \in \bigoplus_{a \in \mathscr{A}} \mathbb{Z}$,

$$\forall a \in \mathscr{A} \quad \zeta(f)(a) = \sum_{b \leq a} f(b)$$

- Its inverse (Proposition 2 [Rot64]), Möbius inversion μ,

$$\forall a \in \mathscr{A} \quad \mu(f)(a) := \sum_{b \leq a} \mu(a, b) f(b)$$

Framework: Loss (7)

Proposed global loss (called 'Regionalized loss' in reference to Yedidia,Freeman, Weiss construction):

for a functor F from $\mathscr{A}^{op}$ (the poset with inverse relation) to vector spaces, and $u = (u_a \in F(a), a \in \mathscr{A})$:

$$l(u) = \sum_{a\in\mathscr{A}} \sum_{b\leq a} \mu(a,b) l_b(u_b) \qquad \text{(RLoss)}$$

Solve

$$\min_{u\in\lim F} l(u)$$

Framework: Loss (8)

The Regionalized loss can be rewritten as,

$$l(u) = \sum_{a \in \mathscr{A}} c(a) l_a(u_a) \tag{0.2}$$

where $c(a) = \sum_{b \geq a} \mu(b, a)$.

In the inclusion-exclusion principle for two sets A, B, $c(A) = 1$, $c(B) = 1$, $c(A \cap B) = -1$.

$$|A \cup B| = |A| + |B| - |A \cap B| \tag{0.3}$$

Critical points of Regionalized loss

When G is a functor from $\mathscr{A}$ to vector spaces, the collection of dual maps

$$G_a^{b^*} : G(a)^* \to G(b)^*$$

defines a functor from $\mathscr{A}^{op}$ to vector spaces denoted as G^*

Theorem (GSP)

F a functor from $\mathscr{A}^{op}$ to vector spaces. An element $u \in \lim F$ is a critical point of the 'global' loss l if and only if there is $(m_{a\to b} \in \bigoplus_{\substack{a,b:\\ b\leq a}} F(b)^*)$ such that for any $a \in \mathscr{A}$,

$$d_u l_a = \sum_{b\leq a} F_b^{a^*} \left(\sum_{c\leq b} F_c^{b^*} m_{b\to c} - \sum_{c\geq b} m_{c\to b} \right) \tag{CP}$$

Message passing algorithms (1)

Assume that the local losses $l_a, a \in \mathscr{A}$ are such that there is a collection of functions $g_a, a \in \mathscr{A}$ that inverses the relation induced by differentiating the local losses, i.e.

$$d_{u_a} l_a = y \iff x = g_a(y_a)$$

Messages:

$$m(t) \in \bigoplus_{\substack{a,b: \\ b \leq a}} F(b)^*: m_{a \to b} \text{ for } b \leq a$$

Auxiliary variables,

$$A(t) \in \bigoplus_{a \in \mathscr{A}} F(a)^*$$

Message passing algorithms (2)

For any $a, b \in \mathscr{A}$ such that $b \leq a$, the update rule is given by,

$$A_a(t) = \sum_{b:b\leq a} \sum_{c:b\geq c} F_c^{a*} m_{b\to c}(t) - \sum_{b:b\leq a} \sum_{c:c\geq b} F_b^{a*} m_{c\to b}(t)$$

$$m_{a\to b}(t+1) = m_{a\to b}(t) + F_b^a g_a(A_a(t)) - g_b(A_b(t)) \qquad \text{(MSP)}$$

Fix points of MSP $\leftrightarrow$ Critical points CP

Theorem (GSP)

Fix points of message passing algorithm (MSP) are critical points of 'global' Regionalized loss (Rloss): if $MSP(m^*) = m^*$, *then let* $\forall a \in \mathscr{A}$,

$$u_a^* = g_a \left[\sum_{b \leq a} F_b^{a*} \left(\sum_{c \leq b} F_c^{b*} m_{b \to c}^* - \sum_{c \geq b} m_{c \to b}^* \right) \right]$$

Then u^* *satisfies (CP).*

Extends previous result of Yedidia, Freeman, Weiss, Peltre (Theorem 5 [YFW05], Theorem 5.15 [Pel20]) stating that:

Fix points of General Belief Propagation $\leftrightarrow$ critical points of Region based approximation of free energy.

Understanding expression of critical points:

Zeta function ζ and Möbius functions μ for functors:

- for $u \in \bigoplus_{a\in\mathscr{A}} G(a)$, and $a \in \mathscr{A}$,

$$\zeta_G(u)(a) = \sum_{b\leq a} G_a^b(u_b)$$

- $$\mu_G(u)(a) = \sum_{b\leq a} \mu(a,b) G_a^b(v_b)$$

μ_G is the inverse of ζ_G

To go further (2)

Understanding expression of critical points:

For F a functor from $\mathscr{A}^{op}$ to vector spaces, critical points u of 'global' regionalized loss are such that:

$$\mu_{F*} d_u l|_{\lim F} = 0$$

To go further (3)

Understanding expression of critical points:

$$0 \to \lim F \to \bigoplus_{a \in \mathscr{A}} F(a) \xrightarrow{\delta_F} \bigoplus_{\substack{a,b \in \mathscr{A} \\ a \geq b}} F(b)$$

where for any $v \in \bigoplus_{\substack{a,b \in \mathscr{A} \\ a \geq b}} F(b)$ and $a, b \in \mathscr{A}$ such that $b \leq a$,

$$\delta_F(v)(a,b) = F^a_b(v_a) - v_b$$

This is simply stating that $\ker \delta = \lim F$.

To go further (4)

Understanding expression of critical points:

$$0 \leftarrow (\lim F)^* \leftarrow \bigoplus_{a \in \mathscr{A}} F(a)^* \overset{d_F}{\leftarrow} \bigoplus_{\substack{a,b \in \mathscr{A} \\ a \geq b}} F(b)^*$$

Pose $d = \delta^*$. For any $l_{a \to b} \in \bigoplus_{\substack{a,b \in \mathscr{A} \\ a \geq b}} F(b)^*$ and $a \in \mathscr{A}$,

$$dm(a) = \sum_{a \geq b} F_b^{a*}(m_{a \to b}) - \sum_{b \geq a} m_{b \to a}$$

To go further (5)

Rewriting condition on fix points:

$$\mu_F^* d_u f \in \operatorname{im} \mathrm{d}$$

is the same as the fact that there is $(m_{a\to b} \in F(b)^* | a, b \in \mathscr{A}, b \leq a)$ such that,

$$d_x f = \zeta_{F^*} \mathrm{d} m$$

To go further (6)

Understanding this choice of message passing algorithm:

g Lagrange multipliers m to $u \in \bigoplus_{a \in \mathscr{A}} F(a)$. $\delta_F(u) = 0$ defines the constraints on u.
$\delta_F g \zeta_{F^*} \mathrm{d}_F$ sends a Lagrange multiplier $m \in \bigoplus_{\substack{a,b \in \mathscr{A} \\ a \geq b}} F(b)^*$ to a constraint $c \in \bigoplus_{\substack{a,b \in \mathscr{A} \\ a \geq b}} F(b)$ defined as, for $a, b \in \mathscr{A}$ such that $b \leq a$,

$$c(a,b) = \delta_F g \zeta_{F^*} \mathrm{d}_F m(a,b) = F_b^a g_a(\zeta_{F^*} \mathrm{d}_F m(a)) - g_b(\zeta_{F^*} \mathrm{d}_F l = m(b))) \tag{0.4}$$

We are interested in $c = 0$, i.e.

$$\delta_F g \zeta_{F^*} \mathrm{d}_F m = 0$$

To go further (7)

Understanding this choice of message passing algorithm:

Choice of algorithm on the Lagrange multipliers so that $\delta_F g \zeta_{F*} \mathrm{d}_F m = 0$,

$$m(t+1) - m(t) = \delta_F g \zeta_{F*} \mathrm{d}_F m(t)$$

Any other choice would also be a good candidate!

Example of applications

- Extension of General Belief Propagation to noisy channel networks
- PCA for filtered data like time series
- Inference (learning) with multimodal integration, inference on scenes with multiple views.

Thank you very much for your attention

Thank you very much for your attention!

References I

- Olivier Peltre, *Homology of Message-Passing Algorithms*, `http://opeltre.github.io`, 2020, Ph.D. thesis (preprint).
- Gian-Carlo Rota, *On the foundations of combinatorial theory I. Theory of Möbius functions*, Probability theory and related fields **2** (1964), no. 4, 340–368.
- Grégoire Sergeant-Perthuis, *Regionalized optimization*, arXiv:2201.11876, 2022.
- Jonathan S Yedidia, William T Freeman, and Yair Weiss, *Constructing free-energy approximations and generalized belief propagation algorithms*, IEEE Transactions on information theory **51** (2005), no. 7, 2282–2312.