Supplementary Material for CVPR 2023 paper #11206

Anonymous CVPR submission

Paper ID 11206

1. Summary

This material presents the supplementary paper from the main paper due to the space limitation. In Section 2, the detailed calculation process of gradient for coefficients are presented. Additional details for ASAP architecture that were omitted from the main manuscript are presented in Section 3. In Section 4, hyperparameters of ASAP and baselines for both node classification and graph classification and more results on brain network classification on ADNI data are shown.

2. Gradients of Polynomial Coefficients with Scale

In this section, we are going to show the detailed derivation of gradient $\frac{\partial c_{s,n}}{\partial s}$ for the three polynomials, i.e., Chebyshev, Hermite and Laguerre.

Chebyshev Polynomial. The expansion coefficient for Chebyshev polynomial is given as

$$c^T_{s,n} = (2-\delta_{n0})(-1)^n e^{-\frac{sb}{2}} I_n\left(\frac{sb}{2}\right), \tag{1}$$

and the derivative of $c^T_{s,n}$ with respect to s is obtained as

$$\begin{aligned}\frac{\partial c^T_{s,n}}{\partial s} &= 2(-1)^n e^{\frac{-sb}{2}}\left(\frac{-b}{2}\right) I_n\left(\frac{sb}{2}\right) + 2(-1)^n e^{\frac{-sb}{2}} I_n\left(\frac{sb}{2}\right)\frac{b}{2} \\ &= 2(-1)^n e^{\frac{-sb}{2}}\left(-\frac{b}{2} I_n\left(\frac{sb}{2}\right) + \frac{b}{2} I'_n\left(\frac{sb}{2}\right)\right) \\ &= (-1)^n b e^{\frac{-sb}{2}}\left(I'_n\left(\frac{sb}{2}\right) - I_n\left(\frac{sb}{2}\right)\right) \\ &= (-1)^n b e^{\frac{-sb}{2}}\left(I_{n-1}\left(\frac{sb}{2}\right) - \frac{n}{\frac{sb}{2}} I_n\left(\frac{sb}{2}\right) - I_n\left(\frac{sb}{2}\right)\right) \\ &= (-1)^n b e^{\frac{-sb}{2}}\left(I_{n-1}\left(\frac{sb}{2}\right) - \left(\frac{2n}{sb}+1\right) I_n\left(\frac{sb}{2}\right)\right) \\ &\left(\because I'_n(x) = I_{n-1}(x) - \frac{n}{x} I_n(x)\right).\end{aligned} \tag{2}$$

Hermite Polynomial. The expansion coefficient for Hermite polynomial is written as

$$c^H_{s,n} = \frac{1}{n!}\left(\frac{-s}{2}\right)^n e^{\frac{s^2}{4}}, \tag{3}$$

and the derivative of $c^H_{s,n}$ with respect to s is computed as

$$\begin{aligned}
\frac{\partial c^H_{s,n}}{\partial s} &= \frac{n}{n!}\left(\frac{-s}{2}\right)^{n-1}\left(\frac{-1}{2}\right)e^{\frac{s^2}{4}} + \frac{1}{n!}\left(\frac{-s}{2}\right)^{n} e^{\frac{s^2}{4}}\frac{s}{2} \\
&= \frac{1}{n!}e^{\frac{s^2}{4}}\left(\frac{-s}{2}\right)^{n}\left(\frac{-n}{2}\left(\frac{-s}{2}\right)^{-1} + \frac{s}{2}\right) \\
&= \frac{1}{n!}e^{\frac{s^2}{4}}\left(\frac{-s}{2}\right)^{n}\left(\frac{n}{s} + \frac{s}{2}\right) \\
&= \frac{1}{n!}e^{\frac{s^2}{4}}\left(\frac{-s}{2}\right)^{n}\frac{n}{s}\frac{s}{2}\frac{2}{s} + \frac{1}{n!}e^{\frac{s^2}{4}}\left(\frac{-s}{2}\right)^{n}\frac{s}{2} \\
&= \frac{se^{\frac{s^2}{4}}}{2n!}\left(\frac{-s}{2}\right)^{n}\left(\frac{2n}{s^2}+1\right).
\end{aligned} \tag{4}$$

Laguerre Polynomial. The expansion coefficient for Laguerre polynomial is written as

$$c^L_{s,n} = \frac{s^n}{(s+1)^{n+1}}, \tag{5}$$

and the derivative of $c^L_{s,n}$ with respect to s is calculated as

$$\begin{aligned}
\frac{\partial c^L_{s,n}}{\partial s} &= \frac{ns^{n-1}(s+1)^{n+1} - s^n(n+1)(s+1)^n}{\left\{(s+1)^{n+1}\right\}^2} \\
&= \frac{ns^{n-1}(s+1)^{n+1} - s^n(n+1)(s+1)^n}{(s+1)^{2n}(s+1)^2} \\
&= \frac{ns^{n-1}(s+1) - s^n(n+1)}{(s+1)^{n}(s+1)^2} \\
&= \frac{ns^n + ns^{n-1} - ns^n - s^n}{(s+1)^{n+2}} \\
&= \frac{ns^{n-1} - s^n}{(s+1)^{n+2}} \\
&= \frac{\left(\frac{n}{s}-1\right)s^n}{(s+1)^{n+2}} \\
&= \frac{s^{n-1}(n-s)}{(s+1)^{n+2}}.
\end{aligned} \tag{6}$$

3. ASAP Architecture

Readout for Graph Classification. The $f_R(\cdot)$ with weights W_R takes H_K from the convolution layers as an input and returns h_G. In our model architecture, h_G is chosen as 2-layer Multi-layer perceptron (MLP) as

$$h_G = \sigma_{R_2}(\sigma_{R_1}(H_K W_{R_1})W_{R_2}) \tag{7}$$

where R_1 and R_2 correspond to first and second layer for MLP structure, respectively. W_{R_1} and W_{R_2} denote weights and Rectified Linear Unit (ReLU) was used for the non-linear activation functions σ_{R_1} and σ_{R_2} for each layer. To make our model end-to-end trainable, the derivative of h_G with respect to H_K is computed as

$$h_G^{'} = \sigma_{R_2}^{'}(\sigma_{R_1}(H_K W_{R_1})W_{R_2})W_{R_2}\sigma_{R_1}^{'}(H_K W_{R_1})W_{R_1}. \tag{8}$$

Table 1. Hyperparameters for classification tasks.

Task	Dataset	Model	Hidden Units	Learning Rate	Dropout Rate	Regularization (α)	Scale's Learning Rate (β_s)	ASAP-C (b)
Node Classification	Cora	**ASAP** & Exact	64	0.01	0.5	0.1	1	1.48
	Citeseer	**ASAP** & Exact	32	0.01	0.5	1	10	1.50
	Pubmed	**ASAP** & Exact	64	0.1	0.5	1	10	1.65
	Amazon Computer	GAT	32	0.01	0.5	-	-	-
		GDC	32	0.0001	0.5	-	-	-
		GraphHeat	32	0.01	0.5	-	-	-
		ASAP & Exact	32	0.001	0.5	1	1	1.60
	Amazon photo	GAT	32	0.01	0.5	-	-	-
		GDC	16	0.0001	0.5	-	-	-
		GraphHeat	32	0.01	0.5	-	-	-
		ASAP & Exact	32	0.01	0.5	1	10	1.59
	Coauthor CS	GAT	32	0.01	0.5	-	-	-
		GDC	16	0.0001	0.5	-	-	-
		GraphHeat	32	0.01	0.5	-	-	-
		ASAP & Exact	32	0.01	0.5	1	1	1.41
Graph Classification	Cortical Thickness	SVM (Linear)	-	-	-	-	-	-
		MLP (2-layers)	16	0.01	0.5	-	-	-
		GCN	16	0.01	0.5	-	-	-
		GAT	16	0.01	0.1	-	-	-
		GDC	16	0.01	0.5	-	-	-
		GraphHeat	32	0.01	0.5	-	-	-
		ASAP & Exact	16	0.01	0.5	1	1	1.20
	FDG	SVM (Linear)	-	-	-	-	-	-
		MLP (2-layers)	16	0.01	0.5	-	-	-
		GCN	16	0.01	0.5	-	-	-
		GAT	16	0.01	0.1	-	-	-
		GDC	16	0.01	0.5	-	-	-
		GraphHeat	16	0.01	0.5	-	-	-
		ASAP & Exact	16	0.01	0.5	1	1	1.20

4. Experiments

In the main experiments, we compared the performances between ASAP and baselines on node classification and graph classification. For node classification, standard benchmarks (Cora, Citeseer and Pubmed) and additional datasets (Amazon Computers, Amazon Photo, and Coauthor CS) were used to validate the performance of ASAP. For graph classification, the ADNI dataset containing a number of subjects with brain network was used to classify AD-specific labels. Hyperparameters should be set so that the model can properly learn data in each experiment. Table 1 shows the detailed parameters of ASAP and baselines for the main experiments.

The trained scales on the brain network classification with Exact and ASAP are visualized in Fig. 1 and Fig. 2 containing additional interpretable results visualized from various views. In Table 2, based on Exact, smallest scales that appear in common across Exact and ASAPs are listed. 7 ROIs were detected in Exact and ASAPs, 6 ROIs were detected in 3 models,

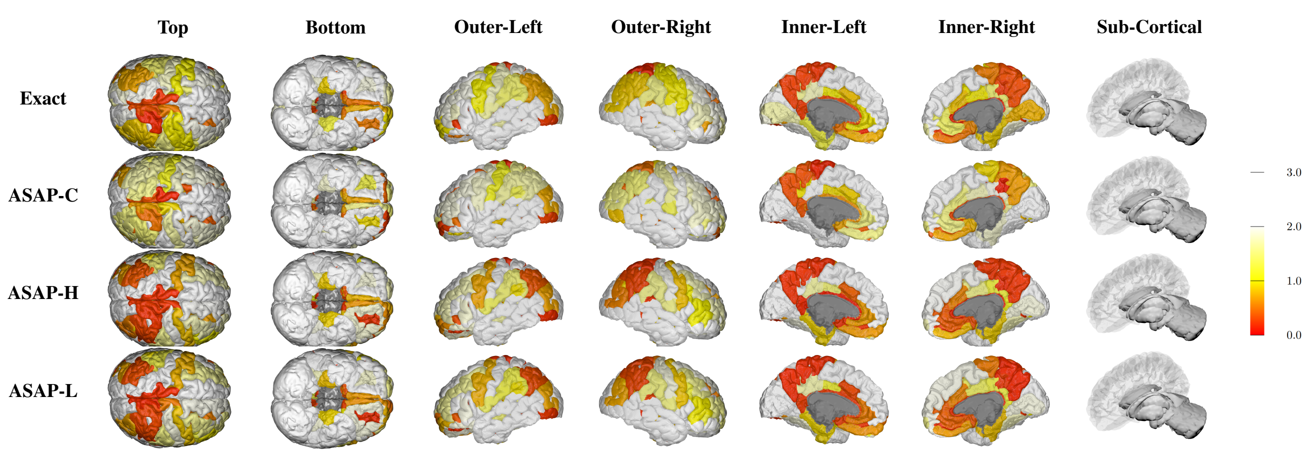

Figure 1. Visualization of learned scales on the cortical and sub-cortical regions of a brain. This visualization shows the scale of each RoI through the classification result using Cortical Thickness feature.

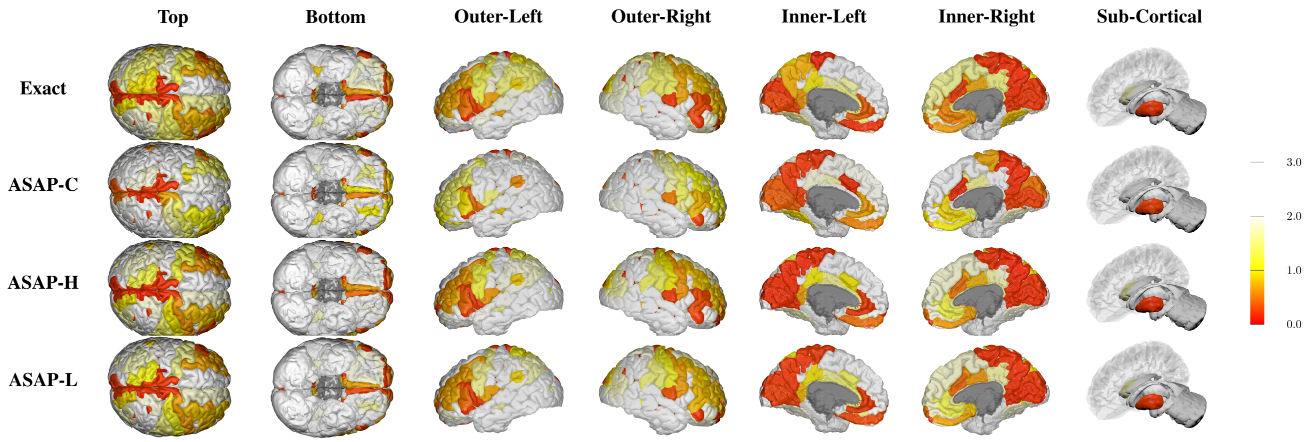

Figure 2. Visualization of learned scales on the cortical and sub-cortical regions of a brain. This visualization shows the scale of each RoI through the classification result using FDG feature.

Table 2. ROIs with the smallest trained scales for AD classification. (L) and (R) denote left and right hemisphere.

Cortical Thickness						FDG					
# of common models	**ROI**	**Exact**	**ASAP-C**	**ASAP-H**	**ASPA-L**	**# of common models**	**ROI**	**Exact**	**ASAP-C**	**ASAP-H**	**ASPA-L**
4	(R) S.pericallosal	0.035	0.097	0.033	0.050	**4**	(L) G&S.paracentral	0.034	0.052	0.049	0.066
	(R) Lat.Fis.ant.Horizont	0.048	0.074	0.039	0.037		(L) G.front.inf.Orbital	0.036	0.071	0.060	0.043
	(L) Lat.Fis.ant.Horizont	0.049	0.083	0.038	0.051		(R) G.precuneus	0.041	0.044	0.034	0.051
	(L) G&S.paracentral	0.059	0.034	0.046	0.044		(R) S.ortibal.med.olfact	0.047	0.078	0.054	0.059
	(L) G&S.occipital.inf	0.062	0.071	0.043	0.055		(R) G.cingul.Post.ventral	0.055	0.056	0.055	0.051
	(R) Lat.Fis.ant.Vertical	0.064	0.108	0.046	0.041		(R) S.oc.temp.lat	0.055	0.065	0.045	0.063
	(L) G.cingul.Post.ventral	0.068	0.127	0.035	0.066		(R) G.oc.temp.med.Lingual	0.055	0.076	0.043	0.040
3	(L) S.suborbital	0.036	-	0.054	0.060		(L) Sub.put	0.058	0.077	0.047	0.060
	(L) S.temporal.inf	0.045	0.129	0.039	0.057		(L) S.postcentral	0.061	0.069	0.060	0.018
	(L) S.occipital.ant	0.050	0.119	-	0.064		(R) G.front.inf.Orbital	0.063	0.093	0.069	0.050
	(R) S.precentral.inf.part	0.054	0.052	-	0.057	**3**	(R) S.intraparietl&P.trans	0.046	0.095	-	0.054
	(R) S.oc.temp.med&Lingual	0.055	-	0.052	0.062		(L) G.cuneus	0.049	-	0.044	0.062
	(L) G.temp.sup.G.T.transv	0.068	0.104	0.049	-		(R) S.temporal.sup	0.049	-	0.028	0.056
2	(R) G.parietal.sup	0.035	-	0.056	-		(R) S.calcarine	0.053	0.059	-	0.047
	(R) Lat.Fis.post	0.049	0.067	-	-		(R) G&S.paracentral	0.055	-	0.051	0.048
	(L) Lat.Fis.ant.Vertical	0.059	-	-	0.040		(R) G.cuneus	0.057	-	0.057	0.066
	(R) S.suborbital	0.064	-	-	0.048	**2**	(R) G&S.cingul.Mid.Ant	0.062	0.041	-	-
							(R) Sub.put	0.065	0.037	-	-
	RMSE for all ROIs	-	**0.5358**	**0.4193**	**0.4072**		**RMSE** for all ROIs	-	**0.5827**	**0.2904**	**0.2899**

and only 2 ROIs were in 2 models for cortical thickness feature. Using FDG, 10 ROIs were detected in every model, 6 ROIS were detected in 3 models, and 2 ROIs were detected in 2 models. In addition, similar ROIs overall showed similar values of scale in all models.