Shapley Deep Learning: A Consensus for General-Purpose Vision Systems [Supplementary Materials]

Youcef Djenouri
University of South-Eastern Norway, Norway
NORCE Norwegian Research Centre, Norway
youcef.djenouri@usn.no, yodj@norceresearch.no

Ahmed Nabil Belbachir
NORCE Norwegian Research Centre
Norway
nabe@norceresearch.no

Tomasz Michalak
IDEAS NCBR, Warsaw, Poland
tomasz.michalak@ideas-ncbr.pl

Anis Yazidi
OsloMet University, Oslo, Norway
anisy@oslomet.no

Optimized Shapley Learning

In this section we sketch the changes that need to be made the Shapely algorithm more efficient for computing the contribution of each model in $\mathcal{M}$. We use intermediate values $\phi(\cdot,\cdot,\cdot)$ analogous to the values $\beta(\cdot,\cdot)$ that constituted the base of the Banzhaf algorithm. For any vertex $v \in T$, set $G \subseteq U$ and integer $k = 0, \ldots, |G|$, let

$$\phi(v, G, k) := \frac{1}{|G|+1} \sum_{\substack{S \subseteq G \\ |S|=k}} \binom{|G|}{k}^{-1} \times P[v, S]. \quad (1)$$

Let us also consider $\phi(v, G)$ as a vector containing all of the values $\phi(v, G, \cdot)$:

$$\phi(v, G) = (\phi(v, G, k))_{k=0}^{|G|}. \quad (2)$$

For the sake of clarity, let us interpret $\phi(v, G, k) = 0$ for $k < 0$ or $k > |G|$. Let $v \in T$, $G \subseteq U$ and $k \in \{0, \ldots, |G|\}$. Let $y \in U \setminus G$. we get:

$\phi(v, G \cup \{y\}, k) = \frac{|G|+1-k}{|G|+2} \times \phi(v, G, k) + \frac{k}{|G|+2} \times \Delta_{v,y} \times \phi(v, G, k-1)$.

proof Let $m = |G| + 1$, we have:

$\phi(v, G \cup \{y\}, k) = \sum_{\substack{S \subseteq G \cup \{y\} \\ |S|=k}} \frac{1}{m+1} \binom{m}{k}^{-1} P[v, S]$

$= \left(\sum_{\substack{S \subseteq G \\ |S|=k}} \frac{1}{m+1} \binom{m}{k}^{-1} P[v, S] \right) + \left(\sum_{\substack{y \in S \subseteq G \cup \{y\} \\ |S|=k}} \frac{1}{m+1} \binom{m}{k}^{-1} P[v, S] \right)$

$= \left(\sum_{\substack{S \subseteq G \\ |S|=k}} \frac{m-k}{m+1} \cdot \frac{1}{m} \binom{m-1}{k}^{-1} P[v, S] \right) + \left(\sum_{\substack{S \subseteq G \\ |S|=k-1}} \frac{k}{m+1} \cdot \Delta_{v,y} \cdot \frac{1}{m} \binom{m-1}{k-1}^{-1} P[v, S] \right)$

$= \frac{m-k}{m+1} \cdot \phi(v, G, k) + \frac{k}{m+1} \cdot \Delta_{v,y} \cdot \phi(v, G, k-1)$.

Shapely value recursive equation designed to demonstrate that the complexity of computing each $\phi(v, F_v)$ is in $O(|F_v|) = O(D)$ time. This overhead is a result of the fact that the vectors $\phi(\cdot,\cdot)$ used have coordinates up to D. Keep in mind that similar values had only one coordinate when computing Banzhaf value explanations, so a similar transition could be conducted in constant time. As a result, the basic algorithm of Lunderg revised the calculation of the vectors $\phi(v, F_v)$where it tends to take $O(LD^2)$ time. We need to create a Shapley formula to get an asymptotically accelerated $O(LD)$ time algorithm for calculating Shapley interpretations using our approach. Consider the following funding to that end:

$$\Psi(v, k) = \sum_{l \in L_v} f(l) \cdot \phi(l, F_l, k), \quad (3)$$

A bottom-up computation may be employed to calculate all the values $\Psi(v, k)$ for $v \in T$ in $O(LD)$ time by proceeding player similarly.

We also have:

$$\gamma(v, k) = \sum_{l \in L_v} f(l) \cdot \phi(l, F_l \setminus \{d_{p_v}\}, k) \quad (4)$$

and

$$\Gamma(v) := \sum_{l \in L_v} f(l) \cdot \Phi(l, F_l \setminus \{d_{p_v}\}). \quad (5)$$

As a result, we can reformulate Shapely as follows:

$$
\begin{aligned}
\phi_i &= \sum_{\substack{l\in L(T)\\ i\in F_l}} f(l)\cdot(\Delta_{l,i}-1)\cdot\Phi(l,F_l\setminus\{i\})\\
&= \sum_{\substack{v\in T\\ d_{p_v}=i}}\sum_{l\in L_v} f(l)\cdot(\Delta_{l,i}-1)\cdot\Phi(l,F_l\setminus\{i\})\\
&= \sum_{\substack{v\in T\\ d_{p_v}=i}}(\Delta_{v,i}-1)\sum_{l\in L_v} f(l)\cdot\Phi(l,F_l\setminus\{i\})\\
&= \sum_{\substack{v\in T\\ d_{p_v}=i}}(\Delta_{v,i}-1)\cdot\Gamma(v).
\end{aligned}
$$

It is worth noting that the preceding derivation provides a $O(L)$ time reduction from computing all ϕ_i to determining all $\Gamma(v)$ values. Once we possess the entire values $\gamma(v,k)$, we can clearly obtain them by addition operation performed in $O(LD)$ time. The following lemma establishes a connection between the values $\Psi(\cdot,\cdot)$ and $\gamma(\cdot,\cdot)$.

lemma. *Let $v\in T$, $v\neq\rho$. Suppose the sets F_l have equal sizes s for all $l\in L_v$. Then, for any $k=0,\ldots,s$, we have:*

$$\Psi(v,k)=\frac{s-k}{s+1}\cdot\gamma(v,k)+\frac{k}{s+1}\cdot\Delta_{v,d_{p_v}}\cdot\gamma(v,k-1).$$

proof. For any $l\in L_v$ we have:

$$
\begin{aligned}
\phi(l,F_l,k)&=\frac{|F_l|-k}{|F_l|+1}\phi(l,F_l\setminus\{d_{p_v}\},k)-\frac{k}{|F_l|+1}\cdot\Delta_{l,d_{p_v}}\cdot\phi(l,F_l\setminus\{d_{p_v}\},k-1)\\
\phi(l,F_l,k)&=\frac{s-k}{s+1}\phi(l,F_l\setminus\{d_{p_v}\},k)-\frac{k}{s+1}\cdot\Delta_{l,d_{p_v}}\cdot\phi(l,F_l\setminus\{d_{p_v}\},k-1).
\end{aligned}
$$

We get the preferred equality by adding the above to all $l\in L_v$ and then using the followin equality:

$$\Delta_{l,d_{p_v}}=\Delta_{v,d_{p_v}}$$