TMCOSS: Thresholded Multi-Criteria Online Subset Selection for Data-Efficient Autonomous Driving - Supplementary Material

Soumi Das
IIT Kharagpur

Harikrishna Patibandla
IIT Kharagpur

Suparna Bhattacharya
Hewlett Packard Labs, HPE Enterprise

Kshounis Bera
HPE Enterprise

Niloy Ganguly
IIT Kharagpur

Sourangshu Bhattacharya
IIT Kharagpur

1. Proof of submodularity

For every set S, $f(S)$ can be defined as:

$$f(S) = \sum_{i \in X} min\{min_{j \in R} Q_{ij}, min_{j \in S} Q_{ij}\} \tag{1}$$

where X is the incoming set, R is the existing set and S is the representative set, $Q_{ij} = \rho d_{ij} - (1-\rho)L_j$, d_{ij} is the pairwise metric and L_j is the pointwise metric.

-f(S) is submodular.

Proof: According to definition of submodularity, for $-f(S)$ to be submodular,

$$-f(S \cup \{x\}) - (-f(S)) \geq -f(T \cup \{x\}) - (-f(T)) \tag{2}$$

where $S \subseteq T$, $x \notin S$, $x \notin T$. Let us define the notations:

1. $D_i(S)$: Function value for finding a representative for an element i $\in$ X with past representative set as S. Formally, $D_i(S) = min\{min_{j \in R} Q_{ij}, min_{j \in S} Q_{ij}\}$
2. $D_i(T)$: Function value for finding a representative for an element i $\in$ X with past representative set as T. Formally, $D_i(T) = min\{min_{j \in R} Q_{ij}, min_{j \in T} Q_{ij}\}$
3. D_{ix} : Function value after x has been found as a representative of $i \in X$.

Note that there can be 3 possibilities after addition of x.

1. x does not become a representative of $i \in X$.
2. x becomes a representative of $i \in X$ with past representative set as only S and not as T. Since $S \subseteq T$, if x is a representative with past representative set only as T, by set property, x also becomes a representative with S as the past representative set. So we exclude the possibility of having representative set as only T.
3. x becomes a representative of $i \in X$ with both S and T as past representative sets.

To prove:

$$f(S) - f(S \cup \{x\}) \geq f(T) - f(T \cup \{x\}) \tag{3}$$

or

$$D_i(S) - D_{ix} \geq D_i(T) - D_{ix} \tag{4}$$

Under the first possibility, D_{ix} retains the value before addition of x ; $D_i(S), D_i(T)$ for sets S and T. So, both sides of inequality lead to 0.

Since $S \subseteq T$, by the form of $f(S)$, $D_i(T) \leq D_i(S)$. Addition of representatives can either decrease the function value or keep it the same.

$$\therefore D_i(T) - D_{ix} \geq 0, D_i(S) - D_{ix} \geq 0 \tag{5}$$

Hence, under the second possibility, $D_i(S) - D_{ix} \geq D_i(T) - D_{ix}$ or $D_i(S) - D_{ix} \geq 0$.

Now, $\because D_i(T) \leq D_i(S), \therefore D_i(T) - D_{ix} \leq D_i(S) - D_{ix}$ (under the third possibility) or $D_i(S) - D_{ix} \geq D_i(T) - D_{ix}$.

Therefore, $-f(S)$ is submodular.

2. Conditions for MCOSS

The optimisation problem for MCOSS is:

$$\min_{z_{ij}^o, z_{ij}^n} \sum_{i=1}^{m} \sum_{j=1}^{|R_t|} z_{ij}^o Q_{ij}^o + \sum_{i,j=1}^{m} z_{ij}^n Q_{ij}^n + \lambda \sum_{j=1}^{m} \|[z_{1,j}^n \dots z_{m,j}^n]\|_p$$
$$s.t. \sum_{j=1}^{|R_t|} z_{i,j}^o + \sum_{j=1}^{m} z_{i,j}^n = 1, \; \forall i \in X_{t+1}$$
$$z_{i,j}^n, z_{i,j}^o \in [0,1], \; \forall i, j \tag{6}$$

where $Q_{ij} = \rho d_{ij} - (1-\rho)L_j$, d_{ij} is the pairwise metric and L_j is the pointwise metric.

Theorem 1 *Let z_{ij}^o and z_{ij}^n be the optimal solution for formulation 6. A new frame $j \in X_{t+1}$ is selected as a representative frame for at least one incoming frame $i \in X_{t+1}$, i.e. $z_{ij}^n = 1$, only if the following conditions hold:*

- *For some incoming frame $i \in X_{t+1}$, $Q_{ij}^n < Q_{ij'}^n$, for all $j' \in X_{t+1}$ and $j' \neq j$*
- *For some incoming frame $i \in X_{t+1}$, $Q_{ij}^n < \frac{\sum_{i'=1}^{m} z_{i',k}^o Q_{i'k}^o + \lambda \|[z_{1,j}^n \dots z_{m,j}^n]\|_p}{\|_j^n\|_1}$*

where $k = argmin_j \sum_{i'=1}^{m} z^o_{i',j} Q^o_{i',j}$, *and* $\|^n_j\|_1 = \sum_{i'=1}^{m} z^n_{i'j}$

Proof: In order for an element to become a representative of atleast one incoming frame $i \in X_{t+1}$, it should have the minimum function value. The first condition states that a new frame $j \in X_{t+1}$ will be selected as a representative in place of $j' \in X_{t+1}$ if it holds a lower function value ($Q^n_{ij} < Q^n_{ij'}$). If j' would have been a representative, $z^n_{ij'} = 1$ would have to be true. This would not be possible due to a higher function value of Q^n_{ij}. By contradiction, *Condition 1* holds.

A point $j \in X_{t+1}$ will get selected if $\sum_{i'=1}^{m} z^n_{i'j} Q^n_{i'j} < \sum_{i'=1}^{m} z^o_{i'k} Q^o_{i'j} + \lambda \|[z^n_{1,j} \dots z^n_{m,j}]\|_p$. The second condition states that an element can be a representative from X_{t+1} if its cost is atmost λ times higher than the best representative from the existing set R_t (k = $argmin_j \sum_{i'=1}^{m} z^o_{i',j} Q^o_{i',j}$). From the above condition, we can see that $\sum_{i'=1}^{m} Q^n_{i'j} < \frac{\sum_{i'=1}^{m} z^o_{i',k} Q^o_{i'k} + \lambda \|[z^n_{1,j} \dots z^n_{m,j}]\|_p}{\|^n_j\|_1}$. This is true for any $i \in X_{t+1}$ which has j as its representative. Hence, we can rewrite the condition as $Q^n_{ij} < \frac{\sum_{i'=1}^{m} z^o_{i',k} Q^o_{i'k} + \lambda \|[z^n_{1,j} \dots z^n_{m,j}]\|_p}{\|^n_j\|_1}$ where $i \in X_{t+1}$ has $j \in X_{t+1}$ as a representative. Thus, *Condition 2* holds true.

3. Comparison of methods: Relative Angle affordance for 100:7 compression ratio

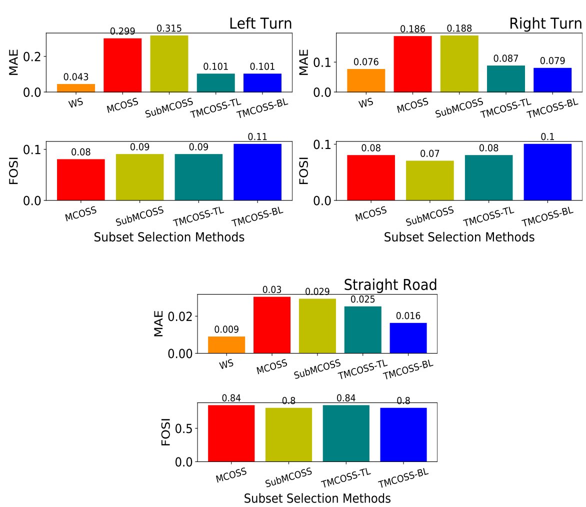

Figure 1: **Fraction of selected instances (FOSI) and Mean Absolute Error (MAE) for Relative Angle affordance.**

We draw a comparison among the subset selection methods on the basis of prediction of the affordance 'Relative Angle' for the compression ratio 100:7. We show here the results of the following selection methods : MCOSS, SubMCOSS, TMCOSS-TL and TMCOSS-BL.

We can clearly observe that the proposed methods TMCOSS-TL and TMCOSS-BL have a lower Mean Absolute Error (MAE) compared to MCOSS and SubMCOSS. This is essentially due to the effective selection of significant instances during turns which help in better prediction of relative angle values.

Table 1: **Time complexity of subset selection methods**

Compression Ratio	WS		100:20	100:7
Training Time (hours)	42		11	7
Subset Selection (hours)	0	MCOSS	2	1
		SubMCOSS	5	2.5
		TMCOSS	1.5	0.75

4. Training time for selection methods

Table 1 shows the amount of total time taken for training and subset selection for each compression ratio. Compression ratio of 100:20 takes about 1/3rd the total time taken to train the whole set (WS). This follows for the other compression ratio too and hence saves both time and space, essential in the current IoT setting with massively huge data.

5. Episode completion in diverse conditions

In this experiment, we report the number of successfully completed episodes (out of 10) by CAL [1] driving model in new weather and new town scenario. We use the selected subsets obtained from each method for training the driving model, and then use the model on the simulator to record the episode numbers in diverse conditions. We observe that the proposed method TMCOSS-TL and TMCOSS-BL perform better than any other baselines, especially for the more difficult turn scenario.

Table 2: **Episode completion in diverse conditions for 100:20 compression ratio**

Methods	**New Weather**		**New Town**	
	Straight	**One Turn**	**Straight**	**One Turn**
WS	10	7	10	6
US	9	0	10	2
OL	9	1	10	2
OSS	10	2	9	2
MCOSS	7	2	8	3
SubMCOSS	10	2	8	2
TMCOSS-TL	10	7	10	5
TMCOSS-BL	10	7	10	6

6. Analysis of semantic segmentation

Here, we extend the results shown in the main paper for semantic segmentation on Cityscapes. Table 3 lists the Mean IoU for all the classes (*class numbers in brackets*) in the Cityscapes dataset. We can observe that the proposed method differs by quite a narrow margin from the performance of WS on almost all the classes, while the baselines are quite a few margins apart. We also show the intuitive analogy in Figure 2 where we can observe that TMCOSS has a higher fraction of instances and pixels than one of the baselines MCOSS in the more important classes viz. *Wall* (Class 3), *Person* (Class 11) and few more, thus justifying its higher IoU values.

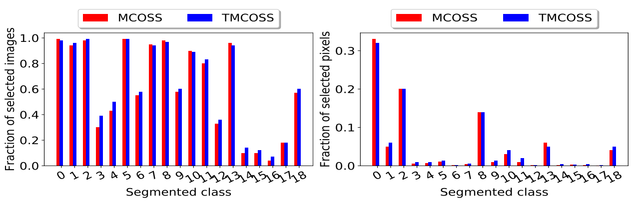

Figure 2: **Fraction of selected instances and pixels for the task of semantic segmentation on Cityscapes**

6.1. Effects on performance by varying ρ

We took different values of ρ = 0.2, 0.5, 0.7 and performed subset selection using the proposed method TMCOSS for the task of semantic segmentation on Cityscapes dataset. Table 4 shows that $\rho = 0.5$ is a fairly appropriate value to be considered for the mentioned tasks.

References

[1] Axel Sauer, Nikolay Savinov, and Andreas Geiger. Conditional affordance learning for driving in urban environments. *arXiv preprint arXiv:1806.06498*, 2018. 2

Table 3: **Semantic segmentation results for all the classes on Cityscapes**

Method	Road (0)	Sidewalk (1)	Building (2)	Wall (3)	Fence (4)	Pole (5)	Traffic Light (6)	Traffic Sign (7)	Vegetation (8)	Terrain (9)	Sky (10)	Person (11)	Rider (12)	Car (13)	Truck (14)	Bus (15)	Train (16)	Motor Cycle (17)	Bicycle (18)	Mean IoU
WS	98.0	83.0	92.0	50.0	59.0	62.0	66.0	76.0	92.0	63.0	94.0	81.0	61.0	94.0	69.0	84.0	65.0	62.0	76.0	75.0
OL	96.0	75.0	88.0	35.0	43.0	51.0	51.0	64.0	90.0	55.0	90.0	73.0	42.0	90.0	52.0	57.0	53.0	45.0	68.0	65.0
OSS	96.0	75.0	88.0	31.0	41.0	52.0	53.0	64.0	89.0	50.0	91.0	74.0	47.0	91.0	25.0	60.0	28.0	50.0	71.0	62.0
MCOSS	96.0	76.0	88.0	29.0	38.0	52.0	53.0	64.0	89.0	52.0	91.0	74.0	47.0	90.0	29.0	64.0	26.0	41.0	70.0	61.5
TMCOSS	**98.0**	**82.0**	**91.0**	**50.0**	**55.0**	**60.0**	**65.0**	**75.0**	**92.0**	**63.0**	**93.0**	**79.0**	**58.0**	**93.0**	**62.0**	**75.0**	**55.0**	**61.0**	**74.0**	**73.0**

Table 4: **Semantic segmentation results for all the classes on Cityscapes on varying ρ**

Method	Road (0)	Sidewalk (1)	Building (2)	Wall (3)	Fence (4)	Pole (5)	Traffic Light (6)	Traffic Sign (7)	Vegetation (8)	Terrain (9)	Sky (10)	Person (11)	Rider (12)	Car (13)	Truck (14)	Bus (15)	Train (16)	Motor Cycle (17)	Bicycle (18)	Mean IoU
WS	98.0	83.0	92.0	50.0	59.0	62.0	66.0	76.0	92.0	63.0	94.0	81.0	61.0	94.0	69.0	84.0	65.0	62.0	76.0	75.0
TMCOSS-0.2	97.0	80.0	90.0	44.0	53.0	57.0	61.0	71.0	91.0	58.0	93.0	77.0	54.0	92.0	57.0	69.0	58.0	56.0	73.0	70.0
TMCOSS-0.5	**98.0**	**82.0**	**91.0**	**50.0**	**55.0**	**60.0**	**65.0**	**75.0**	**92.0**	**63.0**	**93.0**	**79.0**	**58.0**	**93.0**	**62.0**	**75.0**	55.0	**61.0**	**74.0**	**73.0**
TMCOSS-0.7	97.0	80.0	90.0	44.0	52.0	56.0	59.0	69.0	90.0	58.0	93.0	77.0	54.0	92.0	55.0	61.0	57.0	50.0	72.0	68.8