NEURIPS 2026 / RESEARCH PRESENTATION

A Cross-Interaction Neural Architecture for Submodular Functions

Soutrik Sarangi · Aditya Singh · Vansh Maheshwari · Abir De

Indian Institute of Technology Bombay

PAIRWISE INTERACTIONS / CONTROLLED CONCAVITY / SUBMODULARITY

Motivation: Learning Submodular Set Functions

A structural prior for subset selection, representation, and diversity.

THE DEFINING INEQUALITY

$$F(e \mid S) \geq F(e \mid T)$$

$S \subseteq T, \quad e \notin T$

The marginal gain of an element decreases as the selected set grows.

WHERE IT MATTERS

01 Data selection

02 Product recommendation

03 Image / text summarization

The challenge **Learn element interactions while preserving monotonicity and diminishing returns.**

Related Work: The Missing Interactions

Three architectural families address complementary parts of the problem.

FAMILY	WHAT IT ENABLES	RELEVANT LIMITATION
Deep Submodular Functions	Concave-over-modular hierarchy; submodular by design	No explicit pairwise interaction block
FlexSubNet	Trainable concave transformations; submodular by design	No explicit pairwise interaction block
Deep Sets / Set Transformer	Permutation-invariant set learning; attention models interactions	No guaranteed monotone submodularity by design

Missing combination: explicit pairwise interactions + submodularity by construction

Pairwise Interactions and the Supermodularity Challenge

A natural choice for interactions creates a conflict between monotonicity and diminishing returns.

01 / NORMALIZED ATTENTION

Set-dependent normalization

$$w_{s,e}(S) = \exp(q_s^\top k_e) / Z_s(S)$$

Adding an element changes the normalization of existing scores.

Monotonicity is not guaranteed.

02 / UNNORMALIZED PAIRS

Nonnegative interaction scores

$$I(S) = \Sigma\, a(z_s, z_e)\,, \quad (s,e) \in S \times S$$

New elements contribute additional nonnegative interaction terms.

Monotone, but supermodular.

Key question: can a concave transformation reverse increasing marginal gains?

Theoretical Foundations: Limits of Concave Transformations

Within a bounded range, Theorem 3 identifies a necessary condition for transformation.

THEOREM 3 / BOUNDED-RANGE IMPOSSIBILITY

$$\rho(f) = \inf \; f(e \mid S) \,/\, f(e \mid T)$$

over $S \subset T$ and $e \notin T$

$$\rho(f) = 0 \quad \Rightarrow \quad \rho(\varphi \circ f) = 0$$

The transformed function is not submodular.

$f : 2^V \to [0,1]$ normalized, monotone; $\varphi \in C^1$, $\varphi'(t) > 0$.

INTERPRETATION

Vanishing early marginals can coexist with positive later gains.

A smooth strictly increasing outer map cannot remove this obstruction.

We enforce a positive marginal floor.

Architectural implication / Add a modular term $m(z) \geq \delta > 0$ before concave transformation.

From Supermodularity to Submodularity

Derivative-rate $D\varphi(t) = \varphi''(t) / \varphi'(t)$ measures relative slope decay.

THEOREM 4 / FINITE GROUND SET

$$f : 2^V \to [0, 1]$$

$$t\, D\varphi(t) \le -1 / \delta_f$$

$t \in [\delta_f, 1]$

All marginal gains $\ge \delta_f > 0$.
Then $\varphi \circ f$ is monotone submodular.

THEOREM 6 / INFINITE V, $|S| \le k$

$$U = k\, \Delta_f / \beta$$

$$t\, D\varphi(t) \le -U / \delta_f$$

$t \in [\delta_f, U]$

Supermodular f; $0 < \delta_f \le f(\{e\}) \le \Delta_f$
and $\rho(f) \ge \beta > 0$.

Common conditions: f normalized, monotone; $\varphi \in C^2$, concave, $\varphi(0)=0$ and $\varphi'>0$.

Generalizing to Infinite Sets and α-Submodularity

Theorem 8 relaxes the target to α-submodularity; Theorem 9 removes the set-size limit.

THEOREM 9 / ALL FINITE SUBSETS

$$F(e \mid S) \geq \alpha\, F(e \mid T)$$

defines α-submodularity, $\alpha \in (0, 1]$.

$$D\varphi(t) \leq -\alpha / \delta_f$$

for every $t \geq \delta_f$

⇒ $F = \varphi \circ f$ is normalized, monotone and α-submodular.

KEY ASSUMPTIONS

f normalized and monotone; all marginals ≥ $\delta_f > 0$.

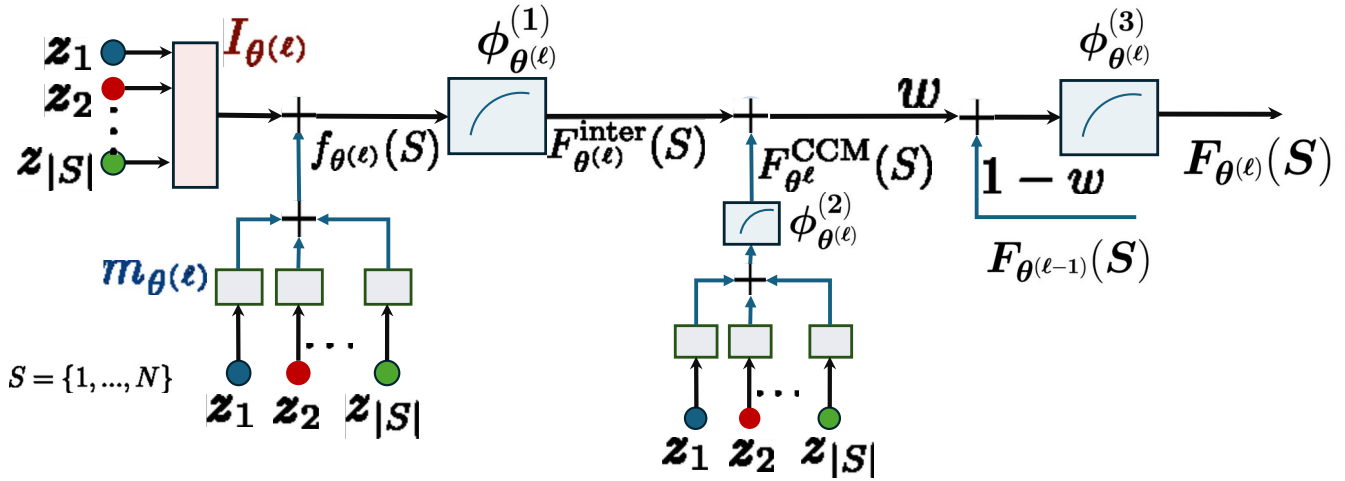
$\varphi \in C^2([0,\infty))$, concave, $\varphi(0)=0$, and $\varphi'(t)>0$.

If f is supermodular, it suffices to lower-bound singleton scores.

Set α = 1 for exact submodularity. No cardinality or upper-range bound is needed.

SubmodInt: The Proposed Architecture

At each layer: explicit interactions + concave-over-modular branch + hierarchical composition.



01 PAIRWISE + MODULAR

$f_l(S) = I_l(S) + M_l(S)$

Positive modular scores guarantee the marginal lower bound.

02 CONTROLLED CONCAVITY

$F_l^{inter}(S) = \varphi_l^{1}(f_l(S))$

Derivative-rate control gives a submodular interaction branch.

03 LAYERWISE COMPOSITION

$CCM + interactions + F_{l-1}$

Positive combination, followed by increasing concavity.

Neural Parameterization of Concave Transformations

Parameterizing φ through its positive derivative makes the theoretical constraint trainable.

TRAINABLE INCREASING CONCAVE MAP

$$\varphi(x) = \int_0^x \varphi'(v)\, dv$$

$$\varphi'(x) = \eta\, e^{-Bx}\, \mathrm{softplus}(-\Lambda\theta(x))$$

$$\Lambda\theta(x) = \int_0^x \mathrm{NN}\theta(v)\, dv$$

NNθ is positive; η > 0 and B ≥ 0.

WHY IT IS CERTIFIED

$$\varphi'(x) > 0$$

Increasing transformation

$$D\varphi(x) \leq -B$$

Choose $B = 1/\delta_l$ for the exact interaction branch (or α/δ_l).

Other concave maps: B = 0.

The modular lower bound and this concave parameterization jointly implement Theorem 9.

Expressivity: Interactions and Depth

Two strict inclusions establish a representational advantage within the compared recursive classes.

THEOREM 10 / INTERACTIONS

$$\mathbf{DSF}_\ell \subsetneq \mathbf{SubmodInt}_\ell$$

At the same depth, pairwise interactions allow additional submodular set functions.

THEOREM 11 / DEPTH

$$\mathbf{SubmodInt}_\ell \subsetneq \mathbf{SubmodInt}_{\ell+1}$$

Adding a layer strictly increases the representable family of set functions.

Scope: fixed-width recursive function classes with admissible concave maps; not arbitrary-width universality.

Experiment I: Submodular Function Regression

INPUT

2,000 vectors in R^3; 20,000 sampled sets ($|S| \leq 200$).

PREDICTION

Learn set value F(S) from labeled pairs (S, F(S)).

EVALUATION

70/15/15 split; test RMSE ↓ across six targets.

TARGET FUNCTION	BEST OTHER MODEL	RMSE ↓	SUBMODINT ↓
Facility Location	Set Transformer	0.164	**0.127**
Graph Cut	Set Transformer	0.047	**0.012**
Log	FlexSubNet	0.006	0.044
LogDet	DSF	0.147	**0.020**
Sqrt	FlexSubNet	0.008	**0.008**
Interaction (INT)	Deep Sets	2.094	**0.137**

ACROSS SIX TARGETS

4 / 6

lowest test RMSE

1 tie · 1 loss

Largest gains: GCut, LogDet and INT, where target functions involve interactions.

Ablation Studies: Parameterization and Depth

Same four synthetic targets (FL, GCut, LogDet, INT); held-out set-value RMSE ↓

A / CONCAVE TRANSFORMS

Replace all trainable φ maps with fixed, admissible functions.

TARGET	FIXED	LEARNED
FL	0.136	**0.127**
GCut	0.348	**0.012**
LogDet	0.211	**0.020**
INT	65.852	**0.137**

B / ARCHITECTURE DEPTH

Compare one-layer and two-layer SubmodInt models.

TARGET	1 LAYER	2 LAYERS
FL	0.136	**0.127**
GCut	0.020	**0.012**
LogDet	0.028	**0.020**
INT	0.294	**0.137**

Learning the concave maps and increasing depth each improves RMSE on all four targets.

Experiment II: Product Recommendation

DATA

8 Amazon Baby Registry categories (>50 items); BERT text embeddings.

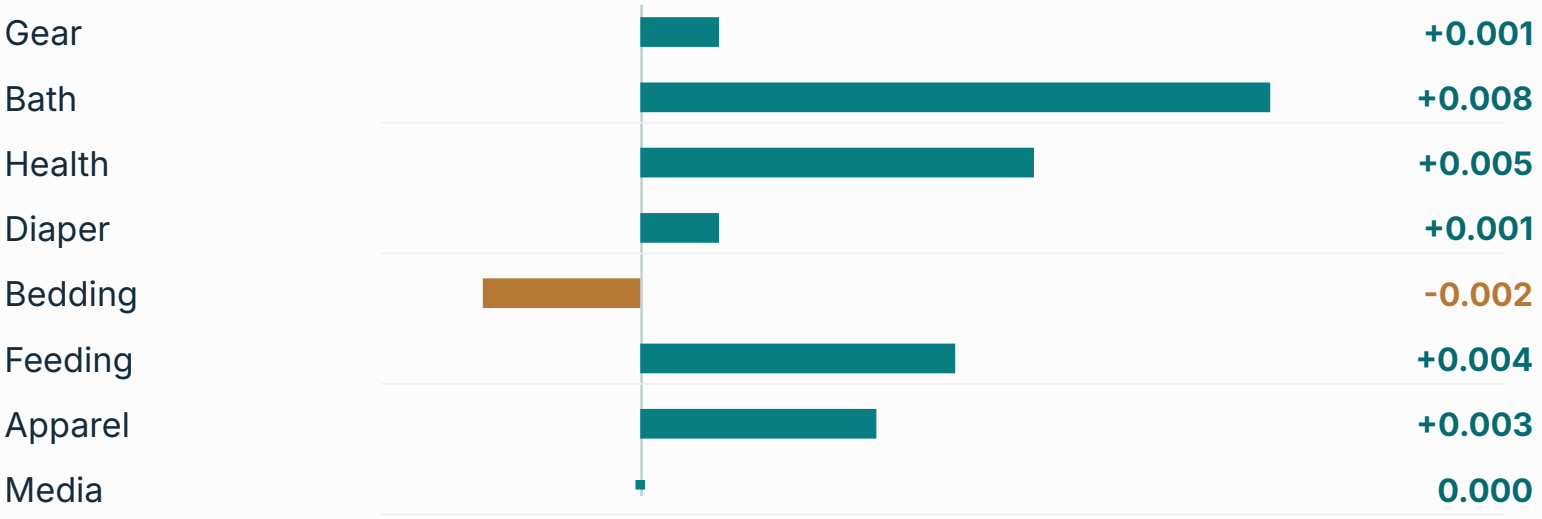
SELECTION

Learn observed sets; greedily choose next product by marginal gain.

TEST METRICS

Jaccard: set overlap; NDCG@10: ranking relevance.

Δ NDCG@10 / SUBMODINT − BEST OTHER MODEL



Positive = SubmodInt outperforms the best alternative

MEAN JACCARD

5 / 8

outright wins · 2 ties

NDCG@10

6 / 8

outright wins · 1 tie

Bedding: the only category where SubmodInt trails the best baseline on both metrics.

Experiment III: Submodular Distillation

DATA

ImageNet100; disjoint held-out pool; choose 100 images.

TRAINING SIGNAL

Oracle gives signed Facility Location score differences.

EVALUATION

FL Eval: ratio to direct FL greedy; offline/online: classifier accuracy.

ROOF MODEL	FL EVAL ↑	OFFLINE ACC ↑	ONLINE ACC ↑
SubmodInt	**0.918 ± 0.024**	**0.68 ± 0.02**	**0.67 ± 0.02**
DSF	0.908 ± 0.016	0.67 ± 0.01	0.66 ± 0.02
SubmodInt (fixed)	0.869 ± 0.003	0.65 ± 0.02	0.63 ± 0.02
Deep Sets	0.581 ± 0.006	0.09 ± 0.01	0.03 ± 0.01
Set Transformer	0.335 ± 0.010	0.13 ± 0.01	0.59 ± 0.01

Online = non-IID, class-incremental stream. Best reported means; DSF gaps are small vs. uncertainties.

Limitations and Future Directions

AREA	CURRENT LIMITATION	NEXT STEP
Interaction cost	Explicit pairwise terms scale as $O(\|S\|^2)$	Explore sparse pair scoring; investigate r-ary interactions
Theory	Strict positive marginal lower bounds and controlled derivative decay	Find less conservative sufficient conditions
Expressivity	Strict comparisons are for fixed-width recursive classes	Study broader architectural families and widths
Evaluation	Limited task families; efficiency not benchmarked here	Broaden benchmarks and report runtime / significance

Controlled concavity enables explicit cross-interactions without sacrificing submodularity.