P-Law: Predicting Quantitative Scaling Law with Entropy Guidance in Large Recommendation Models

Tingjia Shen[1], Hao Wang[1(✉)], Chuhan Wu[2], Chin Jin Yao[2], Wei Guo[2], Yong Liu[1],Huifeng Guo[2], Defu Lian[1], Ruiming Tang[2], Enhong Chen[1]

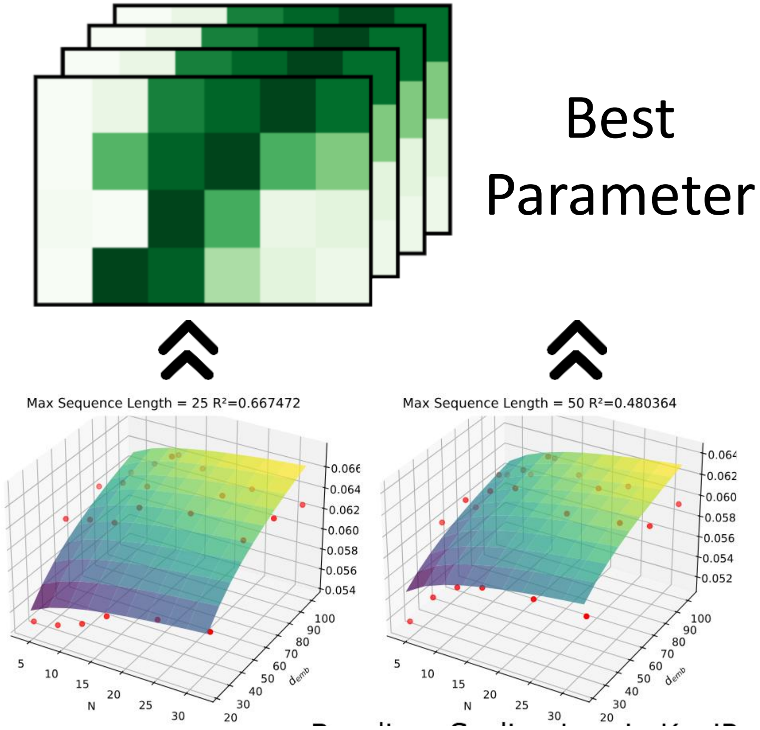
[1] State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China, Hefei, China

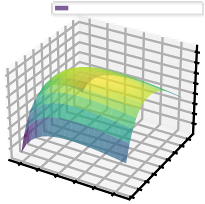
[2] Shenzhen Huawei Technologies Co.Ltd.

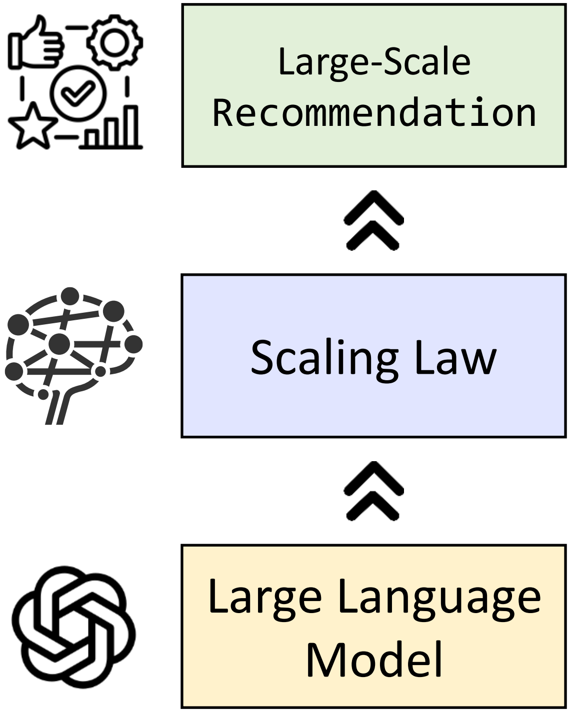
Code: https://github.com/USTC-StarTeam/DLF

Background

- The Scaling Law (SL) achieved significant success by predicting model loss when scaling model size.
- **Recent advances:** Directly utilizing Scaling Law to large-scale recommendation models, as they share similar Transformer architectures





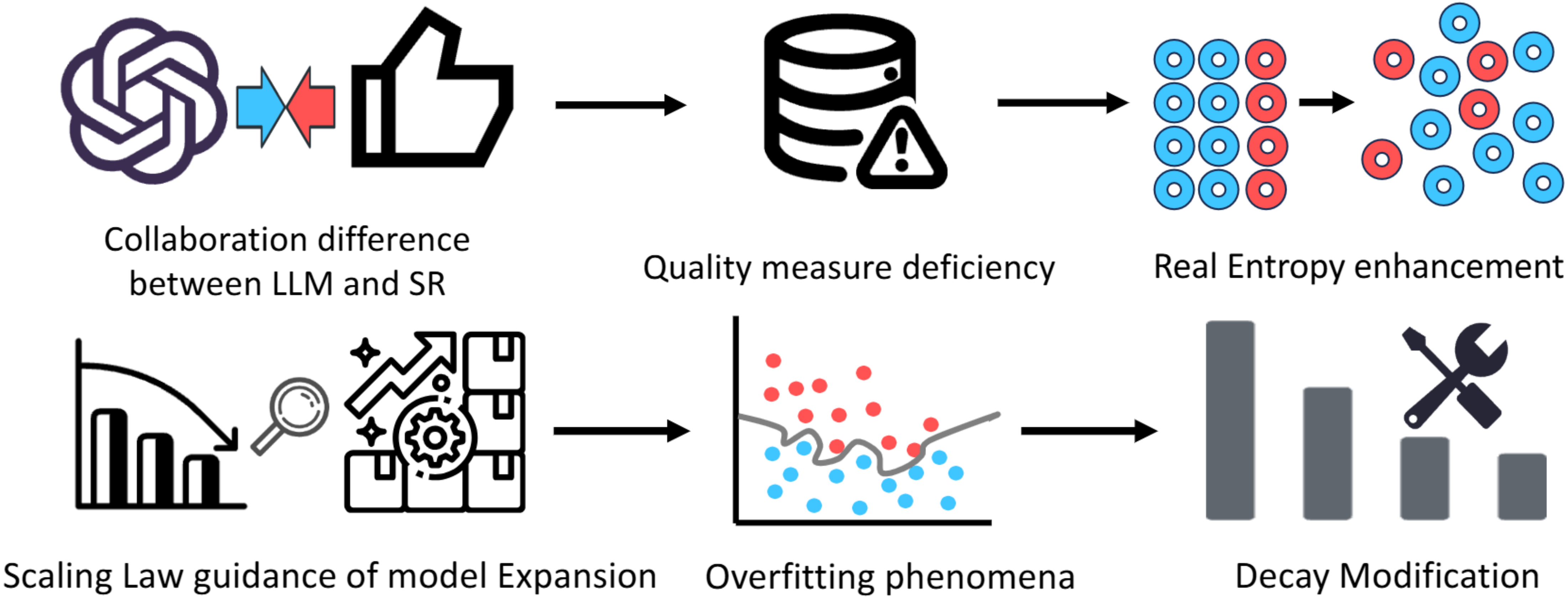


Common Usage of Scaling Law: Parameter Guidance

A Common Pipeline of SL in Large-Scale Recommendation

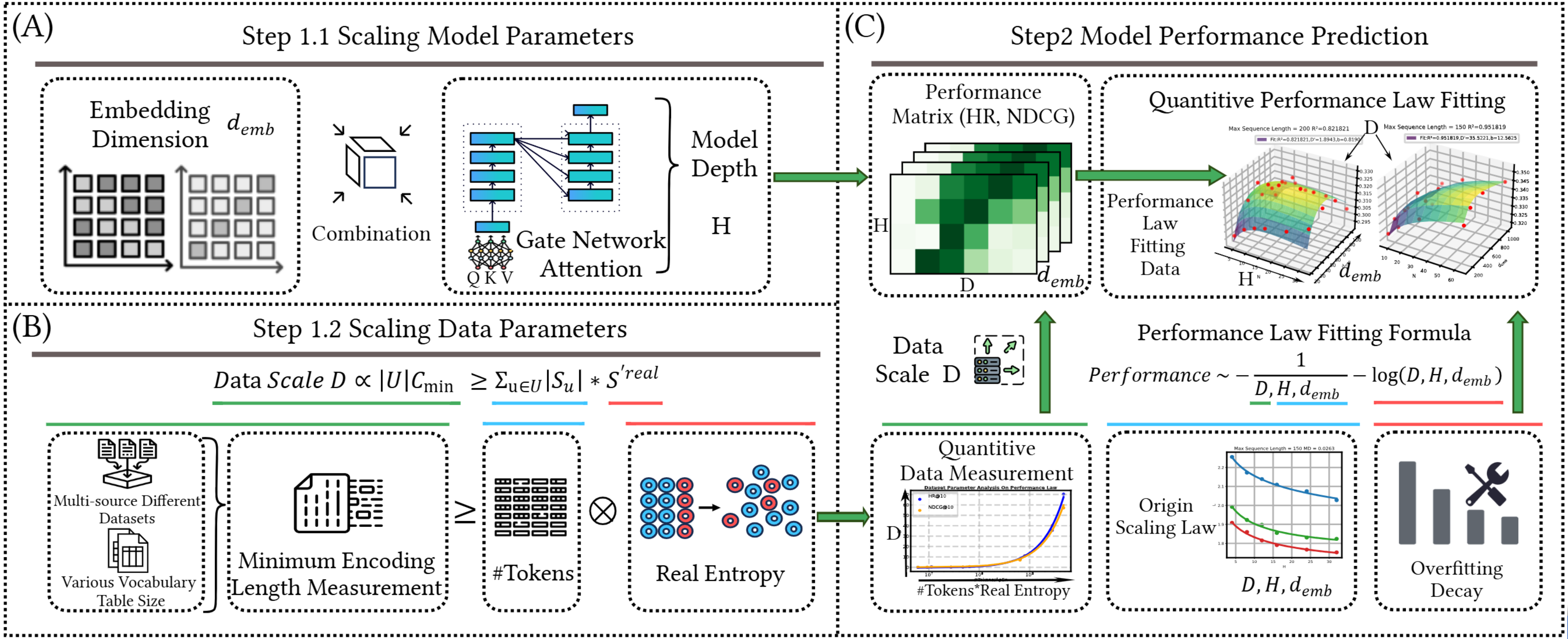
Challenges

- **Quality Measure Deficiency.**
- **Loss-Performance Discrepancy.**



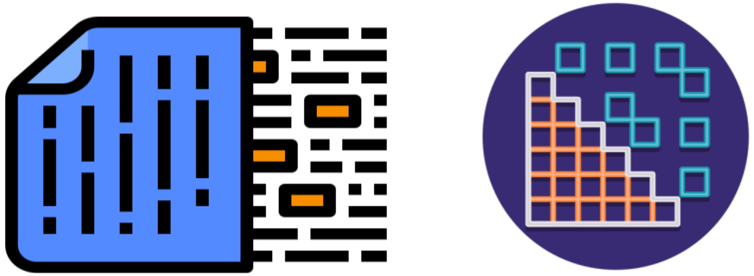
P-Law: Predicting Quantitative Scaling Law with Entropy Guidance in Large Recommendation Models

➢Preview of the overall methodology:

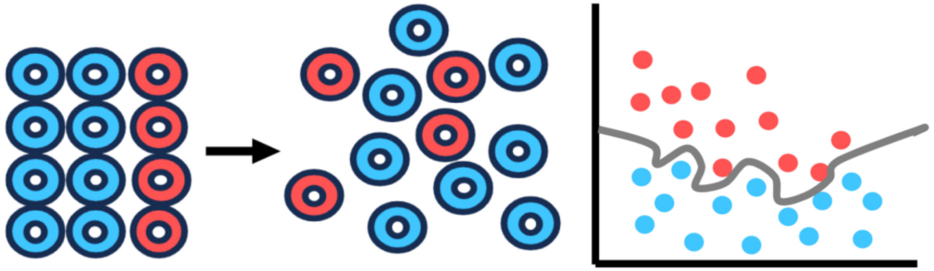


Establish the correlation between data quality S'^{real} and model loss L: $S'^{\mathrm{real}} \propto L$

① Longest Repeated Sequence Calculation



② Calculation of Real Entropy for Data Samples



$$S'^{real} = \left(\frac{1}{|S_u|}\sum_j \Lambda_j\right)^{-1} \ln|S_u| = -\sum_{T' \subset T} P(T')\log_2[P(T')]$$

③ Introduce Real Entropy S^'real to measure sample quality

The higher the Real Entropy S'^{real}, the lower the data repetition rate.

P-Law: Predicting Quantitative Scaling Law with Entropy Guidance in Large Recommendation Models （NeurIPS'2025）
https://openreview.net/pdf/3942d1cbe812cad056ae47f7160ec6bf70f63b4f.pdf

Establish the correlation between data quality S'^{real} and model loss L: $S'^{real} \propto$ L

$$S'^{real} = \left(\frac{1}{|S_u|}\sum_j \Lambda_j\right)^{-1} \ln|S_u| = -\sum_{T'\subset T} P(T')\log_2[P(T')]$$

③ Introduce Real Entropy S^′real to measure sample quality

Large Language Model Scaling Law

$$L(N,D) = \left[\left(\frac{N_c}{N}\right)^{\frac{\alpha_N}{\alpha_D}} + \frac{D_c}{D}\right]^{\alpha_D} \quad 1/D \propto 1/Tokens$$

Quantified Only Data Amount (Tokens) and Model Loss (L)

Formula Derivation and Extension

Large Recommendation Model Scaling Law

$$L(N,D) = \left[\left(\frac{N_c}{N}\right)^{\frac{\alpha_N}{\alpha_D}} + \frac{1}{D'}\right]^{\alpha_D} \quad 1/D' \propto \boxed{1/S'^{real}}\, Tokens$$

Quantifying quality data S^′real is inversely proportional to model loss L.

P-Law: Predicting Quantitative Scaling Law with Entropy Guidance in Large Recommendation Models (NeurIPS'2025)
https://openreview.net/pdf/3942d1cbe812cad056ae47f7160ec6bf70f63b4f.pdf

Establish the correlation between data quality S'^{real} and model loss L: $S'^{\text{real}} \propto$ L

Large Language Model Scaling Law

$$L(N,D) = \left[\left(\frac{N_c}{N}\right)^{\frac{\alpha_N}{\alpha_D}} + \frac{D_c}{D}\right]^{\alpha_D} \quad 1/D \propto 1/Tokens$$

Quantified Only Data Amount (Tokens) and Model Loss (L)

Formula Derivation and Extension

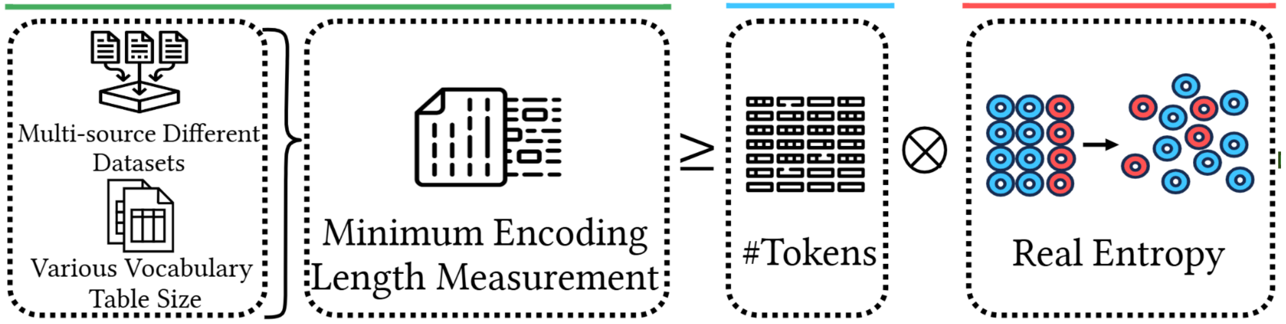
Large Recommendation Model Scaling Law

$$L(N,D) = \left[\left(\frac{N_c}{N}\right)^{\frac{\alpha_N}{\alpha_D}} + \frac{1}{D'}\right]^{\alpha_D} \quad 1/D' \propto \boxed{1/S'^{real}}\,Tokens$$

Quantifying quality data S^'real is inversely proportional to model loss L.

Theoem1: Scaling Data Parameters

$$Data\ Scale\ D \propto |U|C_{\min} \geq \Sigma_{u\in U}|S_u| * S'^{real}$$



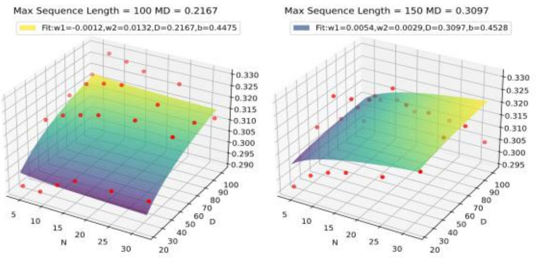
P-Law: Predicting Quantitative Scaling Law with Entropy Guidance in Large Recommendation Models (NeurIPS'2025)
https://openreview.net/pdf/3942d1cbe812cad056ae47f7160ec6bf70f63b4f.pdf

Establish the correlation between model loss L and domain performance P: L ∝ P

① Directly applying the Scaling Law to model loss L will lead to the unlimited increase of model parameters N, resulting in overfitting.

$$P(N,D) \propto -L(N,D) = -\left[\left(\frac{N_c}{N}\right)^{\frac{\alpha_N}{\alpha_D}} + \frac{D_C}{D}\right]^{\alpha_D}$$

Model performance P increases monotonically with model parameters N and data scale D



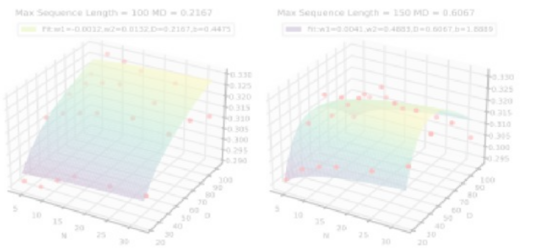
Scaling Law Fitting correlation coefficient R^2 =0.189586

Due to overfitting, it is difficult to associate domain performance P with model loss L

② Introduce a decay term to avoid overfitting and establish a Performance Law with domain performance P: S^′real ∝ L ∝ P

$$P(N,D) \propto -[L(N,D) + \underbrace{\alpha_N ln(N) + \alpha_D ln(D)}_{\text{Decay Term}}]$$

Model performance P first increases and then decreases with model parameter scale N and data scale D.



Performance Law Fitting correlation coefficient R^2 =0.189586

Introducing the decay term avoids unlimited model scaling and successfully fits the Performance Law curve.

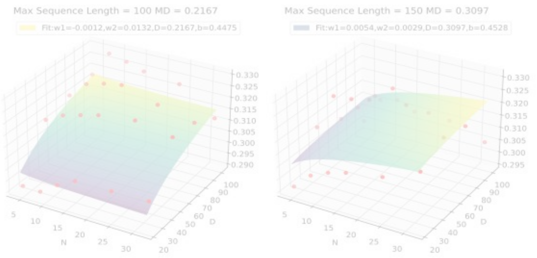
Predictive Models in Sequential Recommendations: Bridging Performance Laws with Data Quality Insights   Paper: https://arxiv.org/pdf/2412.00430

Establish the correlation between model loss L and domain performance P: L ∝ P

① Directly applying the Scaling Law to model loss L will lead to the unlimited increase of model parameters N, resulting in overfitting.

$$P(N,D) \propto -L(N,D)$$
$$= -\left[\frac{N_c}{N}^{\frac{\alpha_N}{\alpha_D}} + \frac{D_C}{D}\right]^{\alpha_D}$$

Model performance P increases monotonically with model parameters N and data scale D



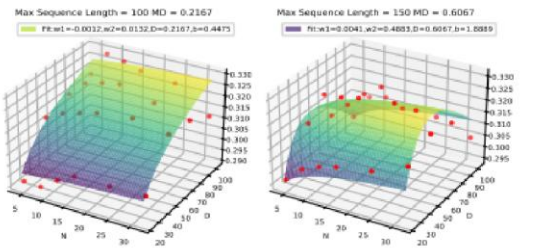
Scaling Law Fitting correlation coefficient R^2 =0.189586

Due to overfitting, it is difficult to associate domain performance P with model loss L

② Introduce a decay term to avoid overfitting and establish a Performance Law with domain performance P: S^′real ∝ L ∝ P

$$P(N,D) \propto -[L(N,D) + \underbrace{\alpha_N ln(N) + \alpha_D ln(D)]}_{\text{Decay Term}}$$

Model performance P first increases and then decreases with model parameter scale N and data scale D.

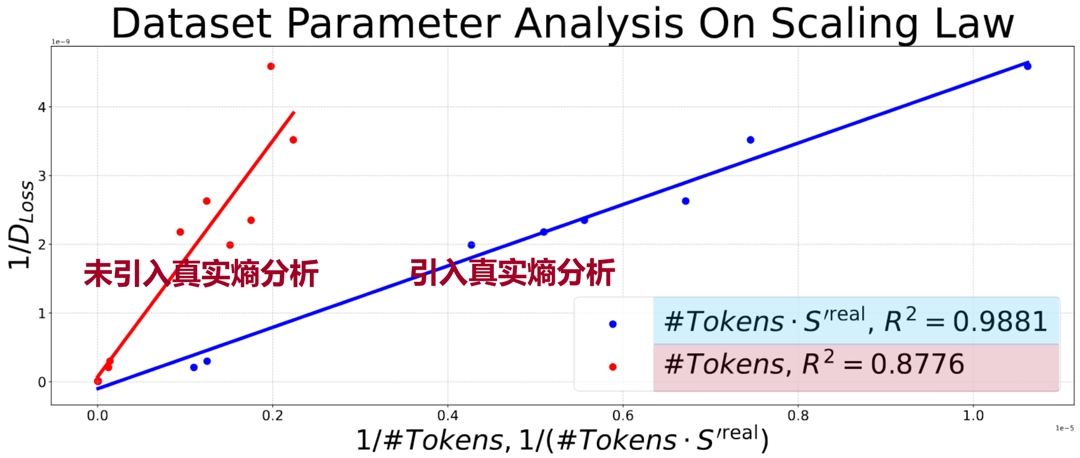


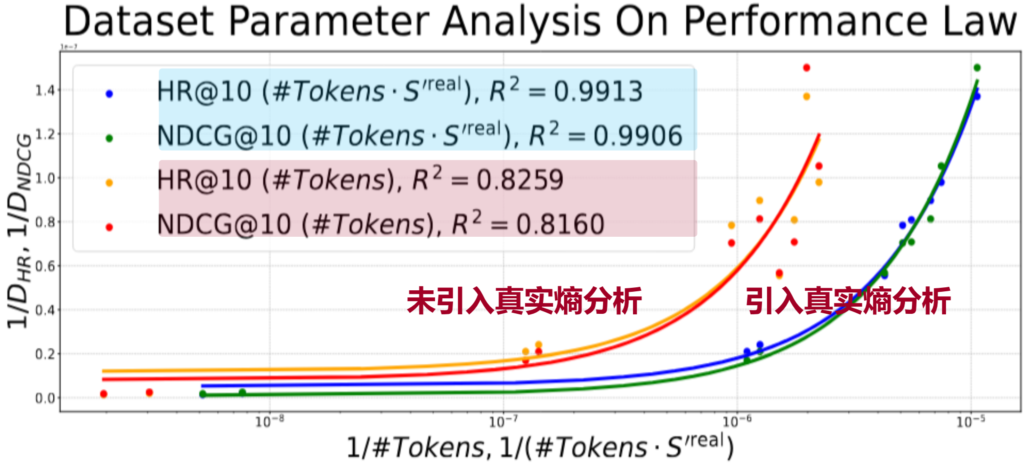
Performance Law Fitting correlation coefficient R^2 =0.189586

Introducing the decay term avoids unlimited model scaling and successfully fits the Performance Law curve.

Experiment Validation on Quality Measure Extension

- 1. Real Entropy delivers more accurate data quality assessment than token count alone, yielding better fits in both SL and Performance Law models.
- 2. Using #Tokens $\cdot S'^{real}$ achieves very high predictive power $R^2 > 0.99$ for HR and NDCG metrics, quantitatively supporting performance analysis.





未引入真实熵分析

引入真实熵分析

Application 1: Global and Local Optimal Parameter Search

- 1. The Performance Law accurately identifies globally optimal parameters, outperforming other configurations.
- 2. It remains robust in predicting locally optimal parameters, offering reliable guidance across various scenarios.

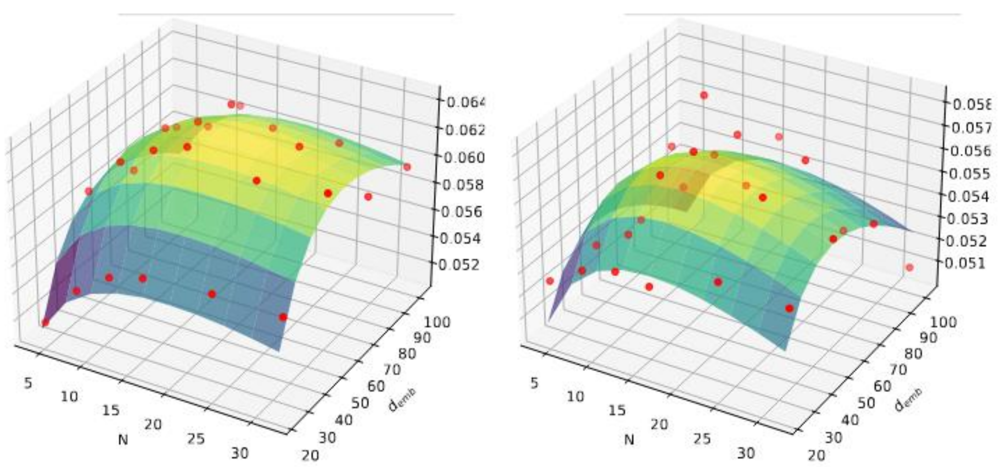
	Global optimal solution						Global optimal solution					
	H	d_{emb}	NDCG@10	NDCG@50	HR@10	HR@50	H	d_{emb}	NDCG@10	NDCG@50	HR@10	HR@50
Smallest Dataset ML-1M	8	54	0.1831	0.2418	0.3265	0.5916	28	25	0.1732	0.2326	0.3101	0.5791
	12	54	0.1824	0.2409	0.3271	0.5913	28	50	0.1866	0.2437	0.3311	0.5917
	16	54	0.1853	0.2434	0.3286	0.5903	28	75	0.1810	0.2408	0.3203	0.5882
	32	54	0.1810	0.2387	0.3216	0.5837	28	100	0.1726	0.2307	0.3102	0.5741
Prediction	28	54	**0.1878**	**0.2443**	**0.3322**	**0.5924**	28	54	**0.1878**	**0.2443**	**0.3322**	**0.5924**
	Optimal solution (with constraint H =64)						Lptimal solution (with constraint $H \cdot d_{emb} \simeq 512$)					
	H	d_{emb}	NDCG@10	NDCG@50	HR@10	HR@50	H	d_{emb}	NDCG@10	NDCG@50	HR@10	HR@50
Largest Dataset Industrial	64	256	0.2019	0.2623	0.3481	0.6205	4	128	0.1758	0.2371	0.3111	0.5854
	64	370	0.2035	0.2639	0.3504	0.6226	8	64	0.1773	0.2381	0.3118	0.5858
	64	512	0.2032	0.2636	0.3502	0.6226	10	51	0.1758	0.2365	0.3092	0.5840
	64	1024	0.1981	0.2590	0.3448	0.6195	16	32	0.1704	0.2305	0.3007	0.5732
Prediction	64	603	**0.2040**	**0.2644**	**0.3512**	**0.6235**	12	44	**0.1777**	**0.2383**	**0.3121**	**0.5863**

Application 2: Exploring Performance Law Potential Among Framework

- 1. The Performance Law's fitted parameters (w_1, w_2)accurately reflect a model's scaling-up potential, as confirmed across multiple frameworks.
- 2. This enables effective guidance for model structure configuration, improving efficiency in memory and time when adapting frameworks.

Table 4: Comparison of Model Parameters and Performance Across Different Precisions in Movielens-1M with NG denotes NDCG. All results are statistically significant with p<0.05.

Precision	Float32			Bfloat16		
Model	HSTU	LLaMA2	SASRec	HSTU	LLaMA2	SASRec
$w_1 \uparrow$	0.009	**0.036**	0.007	0.003	0.015	-0.014
$w_2 \uparrow$	0.083	**0.159**	0.001	0.034	0.086	0.008
HR@10↑	0.332	**0.346**	0.302	0.332	0.336	0.293
HR@50↑	0.585	**0.598**	0.573	0.594	0.598	0.561
NG@10↑	0.185	**0.194**	0.172	0.187	0.188	0.162
NG@50↑	0.242	**0.252**	0.231	0.247	0.249	0.221



Conclusion & Future Work

- Performance Law has been thoroughly explored in the SR domain, and, with appropriate metrics, our theoretical framework can also be applied to other recommendation tasks.

- We will extend Performance Law to larger datasets and a broader range of recommendation scenarios such as ranking and retrieval in future work.

- Performance Law quantitatively predicts SR model performance, surpasses traditional Scaling Laws, and its strong applicability is illustrated by experiments on model parameter and potential prediction.

THANK YOU