Agent Skill Acquisition for Large Language Models via CycleQD

So Kuroki, Taishi Nakamura, Takuya Akiba, Yujin Tang

NEURAL INFORMATION PROCESSING SYSTEMS

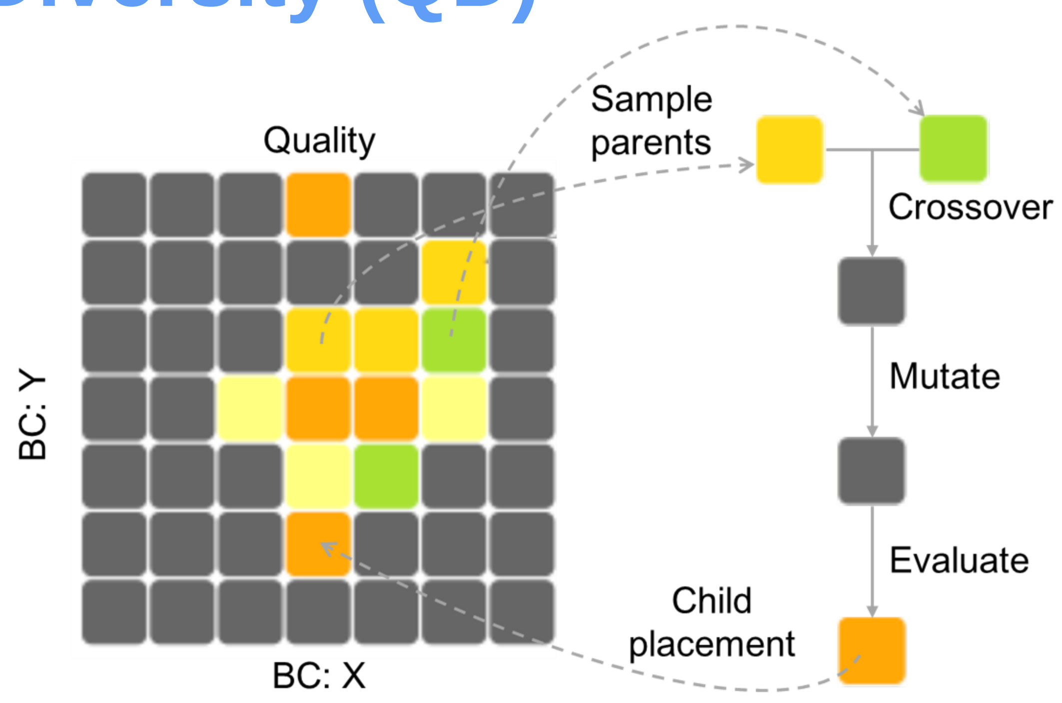
sakana.ai

Problem of Agentic Skill Acquisition

While **fine-tuning** remains the approach, acquiring diverse agentic skills poses significant challenges:

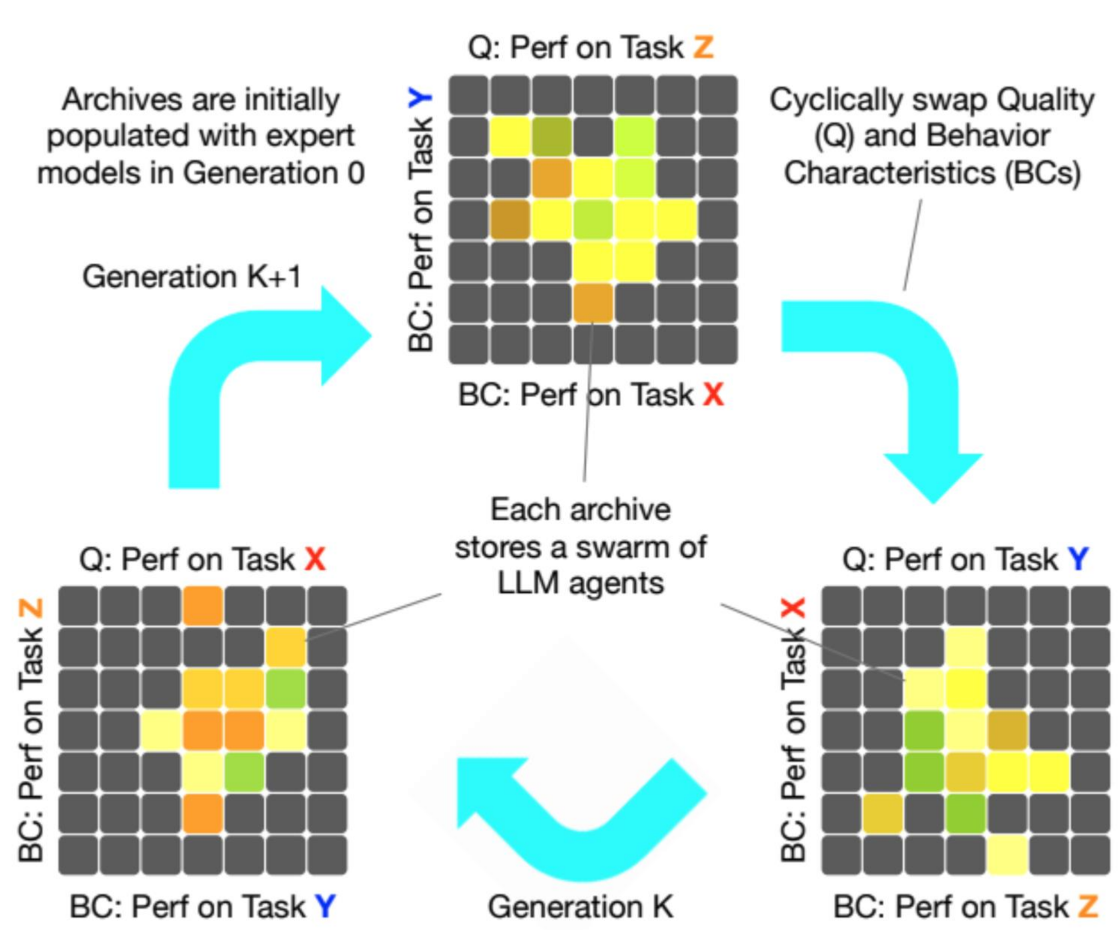
- **Data Ratio Imbalance**: Learning across skill datasets, leading to overfitting or neglect of tasks.
- **Ineffective Objective Functions**: Loss functions, such as next token prediction, fail to align with performance, hindering the acquisition of skills.

Quality Diversity (QD)



QD optimizes **performance** and **diversity** to find solutions. It stores solutions archive, where each cell represents **Behavioral Characteristics (BCs)** and contains the highest **Quality** solution.

Cyclically Alternating BCs and Quality



In **CycleQD**, with **BCs** and **Quality** cyclically swapping, two challenges were solved:

- Task-specific periodic optimization **eliminated the need** for manual **data ratio adjustments**
- **Task-specific** optimization enabled learning tailored to each agent's skills

Evolutionary Process in CycleQD

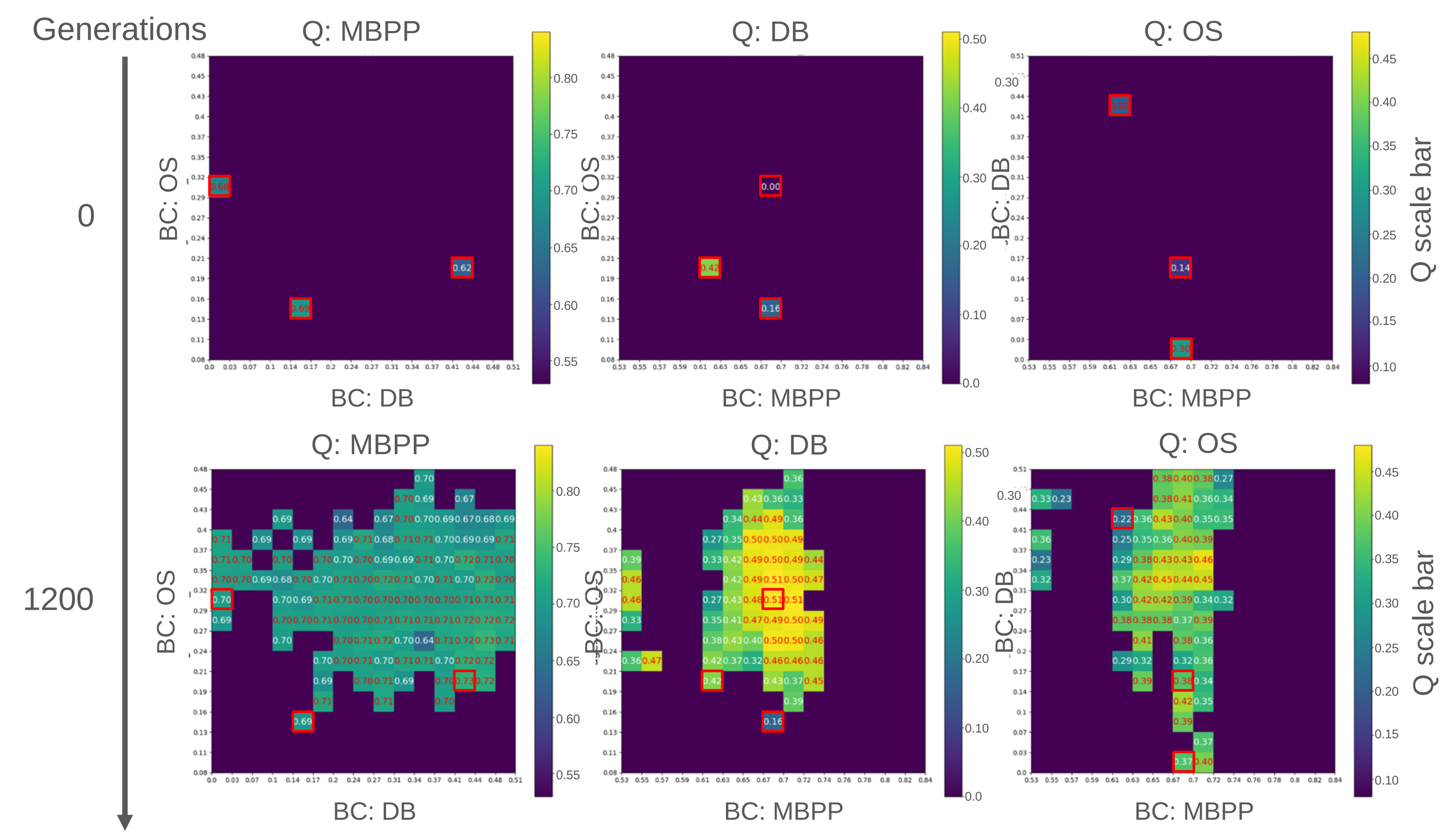
Model Merge as crossover

$$\theta_{\text{child}} = \theta_{\text{base}} + (\omega_1/(\omega_1+\omega_2))\tau_{p_1} + (\omega_2/(\omega_1+\omega_2))\tau_{p_2}$$

SVD-based mutation

$$h(\theta_{\text{child}}) = \theta_{\text{base}} + \text{concat}([U_l(\Sigma_l w)V_l^{\mathsf{T}}]_{l=1}^{L})$$

Archives Development



Evaluation on computer science tasks

#	Methods	MBPP	DB	OS	Avg
0	GPT-4	83.6	36.5	63.7	61.3
1	GPT-3.5-TURBO	82.0	41.6	38.5	53.7
2	Llama3-8B-Instruct (base model)	67.3	5.3	25.2	32.6
	Fine-tuning Based Methods				
3	Fine-tuning (Coding expert)	70.4	21.2	20.7	37.4
4	Fine-tuning (DB expert)	65.8	**42.4**	28.5	45.6
5	Fine-tuning (OS expert)	66.3	0.0	30.4	32.2
6	Fine-tuning (All)	67.3	37.1	36.7	47.0
	Merging Based Methods				
7	Merging (w/o learning)	72.9	24.7	**42.6**	46.7
8	Merging (learning w/ GD)	69.3	41.2	29.6	46.7
9	Merging (learning w/ CMA-ES)	69.3	41.2	30.2	46.9
10	Merging (learning w/ NSGA-II)	75.9	**42.4**	36.4	51.6
11	CycleQD (Ours)	**76.4**	38.2	**42.6**	**52.4**

Ablation Studies

#	Trials	MBPP	DB	OS	Avg
0	QD + No mutation + Random sampling	70.4	28.8	**43.7**	47.6
1	CycleQD + No mutation + Random sampling	72.9	33.5	41.9	49.4
2	CycleQD + Gaussian mutation + Random sampling	73.4	30.0	42.2	48.5
3	CycleQD + SVD mutation + Random sampling	75.9	**38.2**	41.1	51.7
4	CycleQD + SVD mutation + Elite sampling	**76.4**	**38.2**	42.6	**52.4**

Generalization performance

Model	Coding Tasks		Language Tasks				Avg
	HUMANEVAL+	BigCodeBench	Reasoning	GSM8K	RC	CommonSense	
MBPP expert	1.18	0.97	0.57	0.82	0.94	1.03	0.92
DB expert	0.80	0.84	0.84	0.87	0.98	0.98	0.89
OS expert	0.94	0.90	0.98	0.93	0.99	0.99	0.95
CycleQD	1.10	1.03	0.95	0.88	0.98	1.02	0.99

CycleQD to Sagment Anything Model

#	Expert A	Expert B	Score A	Score B	Avg Score	Model Similarity
0	CAM	POL	0.95	0.99	0.97	0.98
1	CAM	SKL	0.85	0.99	0.92	0.97
2	CAM	LEA	0.51	0.89	0.70	0.88
3	POL	SKL	0.98	0.95	0.96	0.99
4	POL	LEA	0.40	0.84	0.62	0.93
5	SKL	LEA	0.83	0.84	0.83	0.95

Future works

Life-long Learning: Systems continually grow and adapt, with CycleQD enabling diverse foundation.

Swarm of Agents: Diverse agents collaborate, expanding AI capabilities for real-world problems.