A Memetic and Bayesian Optimization Perspective of ‘Artificial General Intelligence’

presented by

Yew-Soon Ong

Professor, School of Computer Science and Engineering
Nanyang Technological University, Singapore

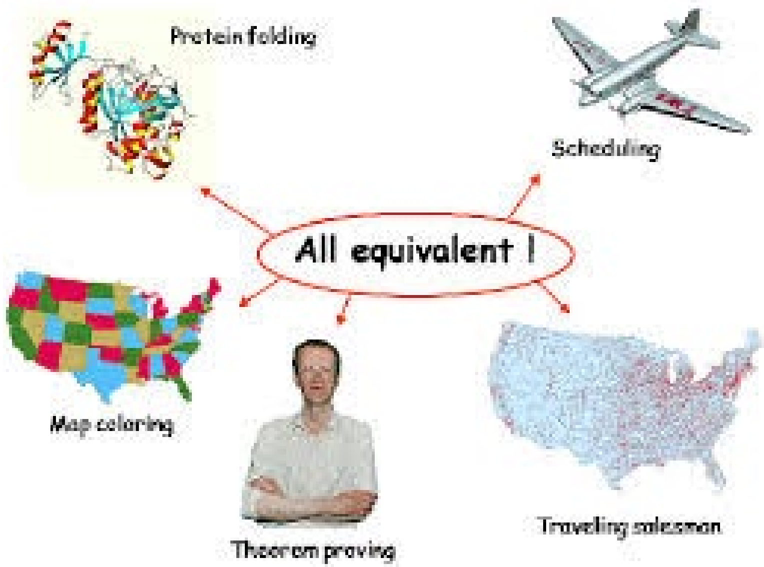
March, 2019

Present Day Optimization Solvers

zero knowledge

- **Observations**

  - **Often start a Search from scratch & at "Ground Zero" Knowledge State.**
  - **No learning. Capability does not grow or evolve along with problems solved or experiences**
  - **BUT Problems seldom exist in isolation & hence Humans do not search from Scratch.**



- **Useful information exist between tasks & problems…**
- **Learning** & properly harnessing the past knowledge are effective for future problem-solving.

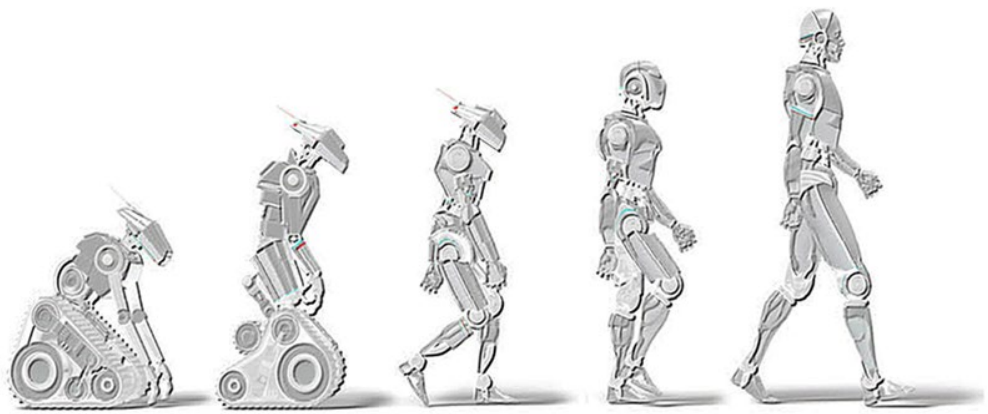
In Global Optimization…

- **Artificial General Intelligence: Ability to accomplish any cognitive task at least as well as humans** ✗

- **General Intelligence: Ability to accomplish virtually any goal, including learning** ✓

→ **General Optimization Intelligence**

What Constitutes 'General Optimization Intelligence' (GOI)?



- Universal machines that are capable of solving a variety of problems.
- Little/no problem-dependent tweaking or redesign required.
- Machines that get smarter with problem solved, just like the way humans do… Learning…

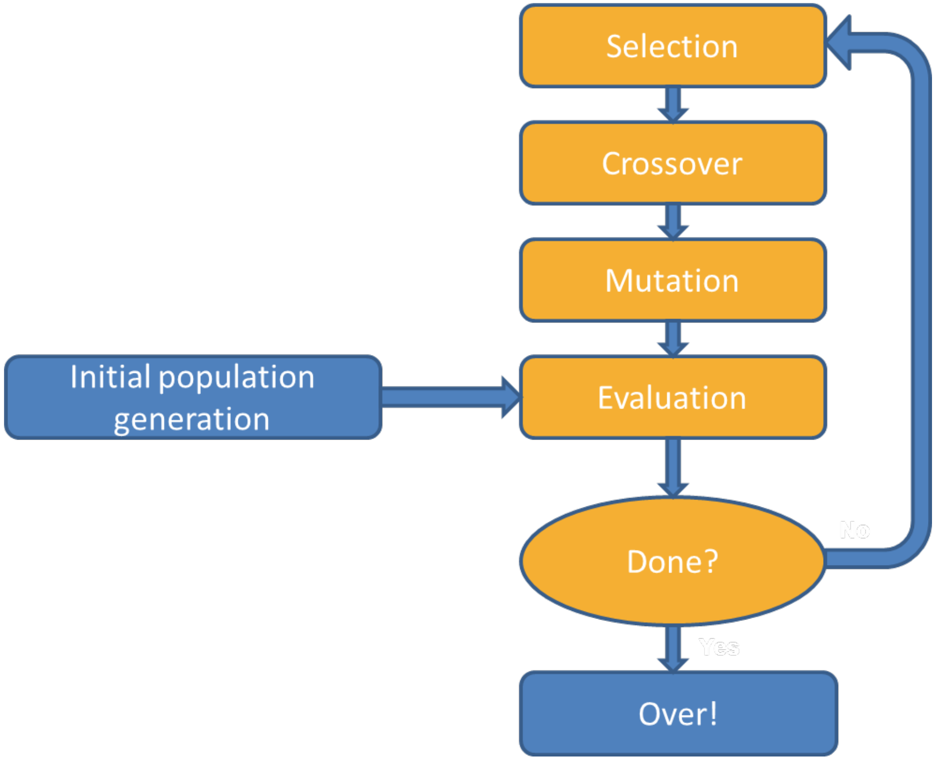
Shall we work towards a future in GOI harnessing MACHINE's ability to LEARN & THINK:

To automatically SELECT, ADAPT and INTEGRATE knowledge from past problems for efficient future problem-solving!

NANYANG TECHNOLOGICAL UNIVERSITY

How about Evolutionary Computation?

Viewing Conventional EC as a Stepping-Stone to 'GOI'



A largely **PROBLEM-INDEPENDENT** framework that can be applied across a variety of optimization tasks

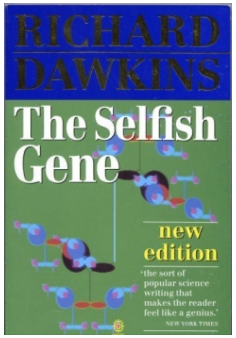
However, biological evolution & hence conventional EC is often deemed to be **TOO SLOW** in practice!

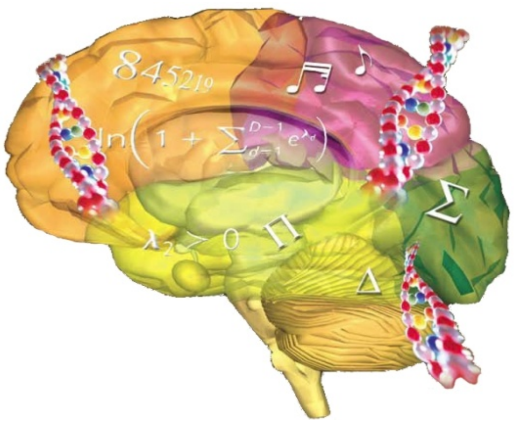
NANYANG TECHNOLOGICAL UNIVERSITY

Incorporating Knowledge in Search: Rise of "Memetics" in Computing

Dawkins in 1976, Book "**The Selfish Gene**", Chapter 11:

"Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so **memes propagate themselves in the meme pool by leaping from brain to brain** via a process which, in the broad sense, can be called **imitation**."



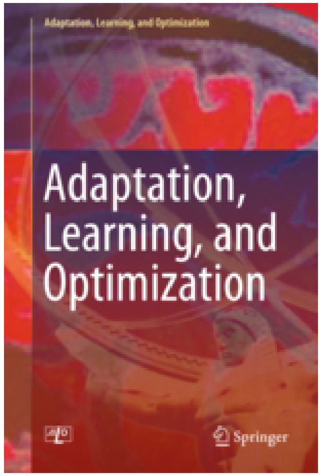


In Computational Intelligence, *Memes* are viewed as *computationally encoded* 'Knowledge Building Blocks' that appear in the form of *recurring information patterns (cultural evolution) for problem solving*.

NANYANG TECHNOLOGICAL UNIVERSITY

"**Research Frontier: Memetic Computation - Past, Present & Future**", **IEEE Computational Intelligence Magazine**, Vol. 5, No. 2, pp. 24 -36, 2010.

A new book in the Springer series **“Studies in Adaptation, Learning, and Optimization”**



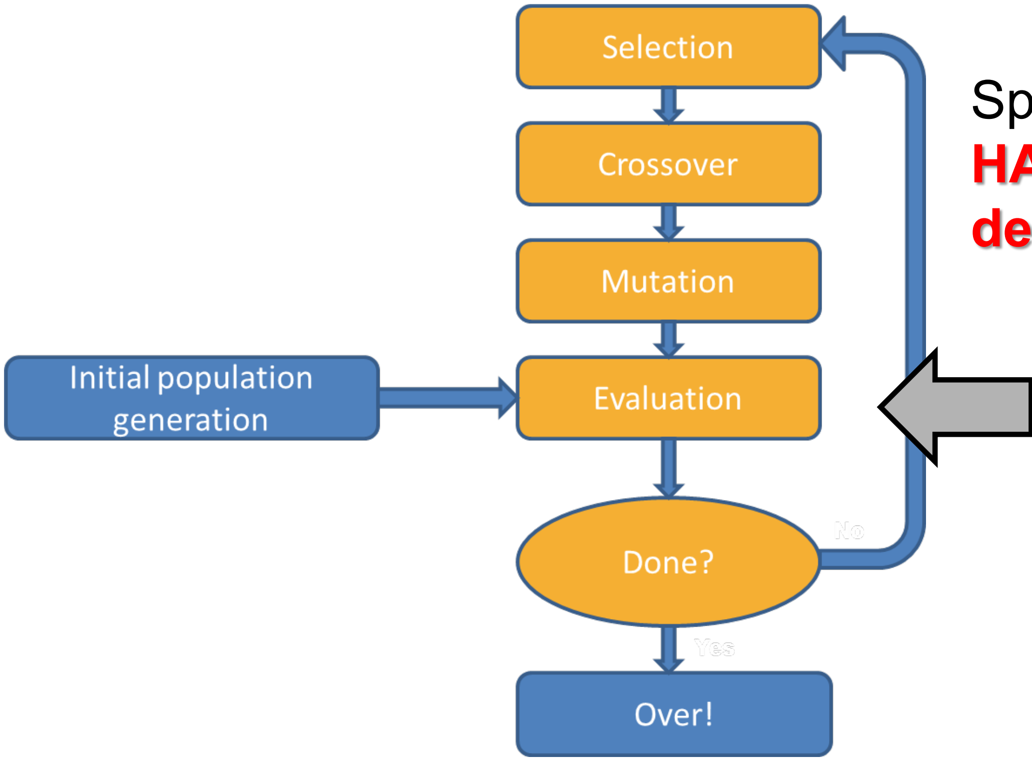
Book Title ➔
Memetic Computation:
The Mainspring of Knowledge Transfer in the Data-Driven Optimization Era (2019)

Offers a comprehensive overview of modern research activities in memetic computation, spanning:

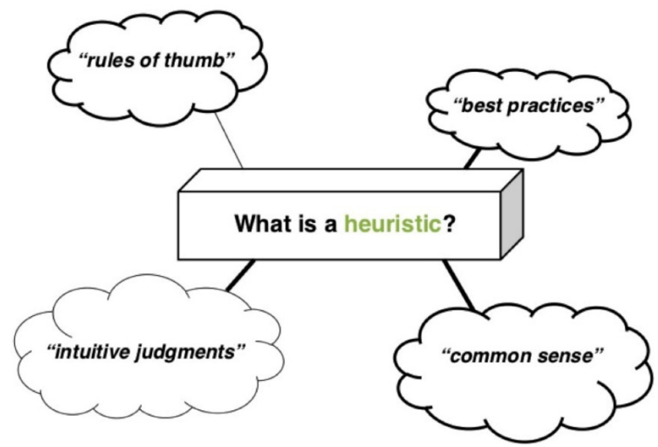
1) **First generation canonical memetic algorithms**
2) **Adaptive meme selection and integration**
3) **Formalization of memetic automatons**
4) **Sequential knowledge transfers across problems**
5) **Multitask knowledge transfers across problems**
6) **Meme-space evolutions for large-scale optimization**

NANYANG TECHNOLOGICAL UNIVERSITY

First Generation Memetic Algorithms:
Hybridizing EC with Local Search



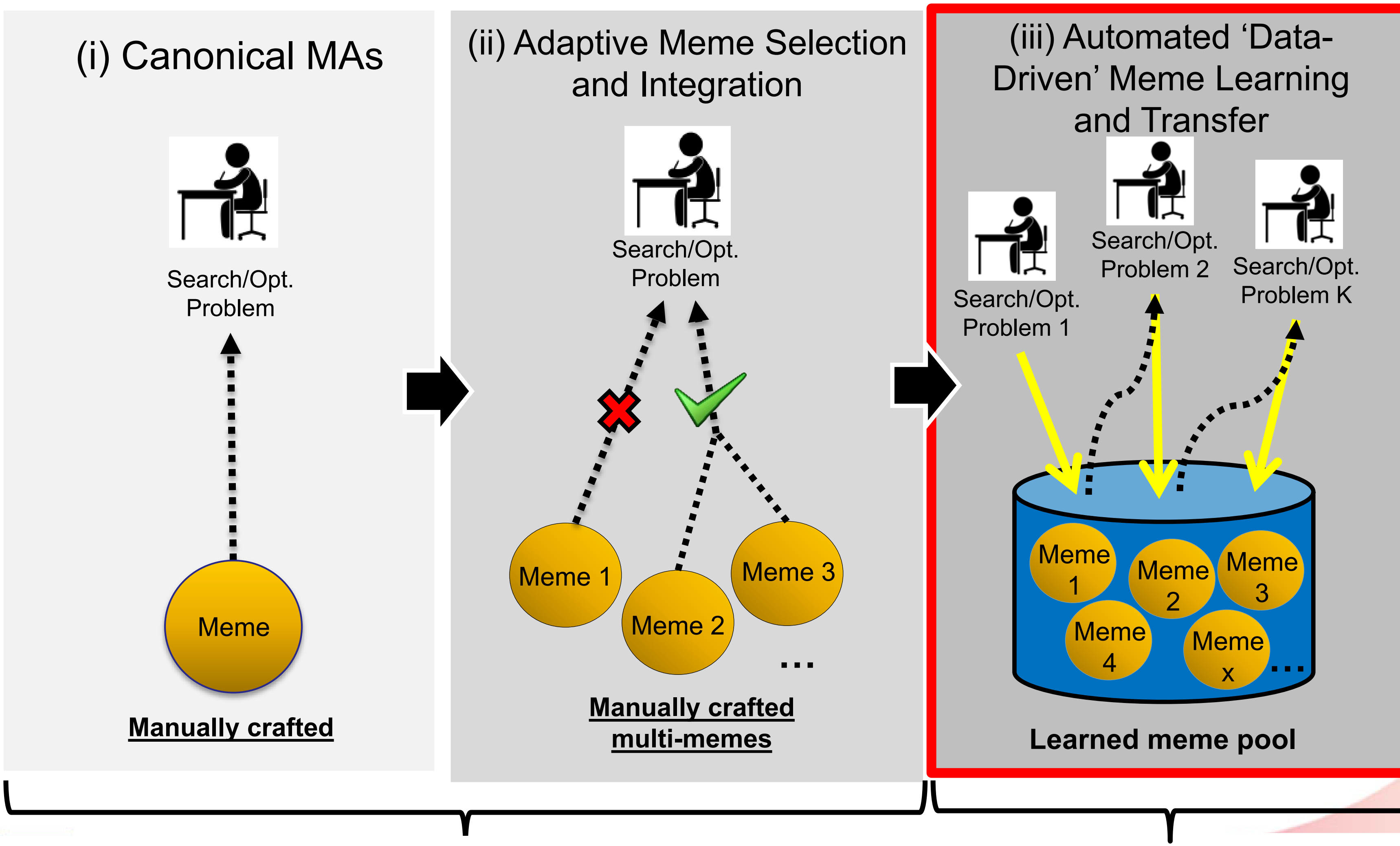
Speeding up search by incorporating **HANDE-CRAFTED “problem-dependent” local search heuristics**



Narrow Intelligence: ability to accomplish a narrow set of goals

NANYANG TECHNOLOGICAL UNIVERSITY

Modern Memetic Computation: Automated Learning & Knowledge Transfer Across Problems



Early Generations memetic algorithms (hybrid optimization algorithms)
Narrow Intelligence

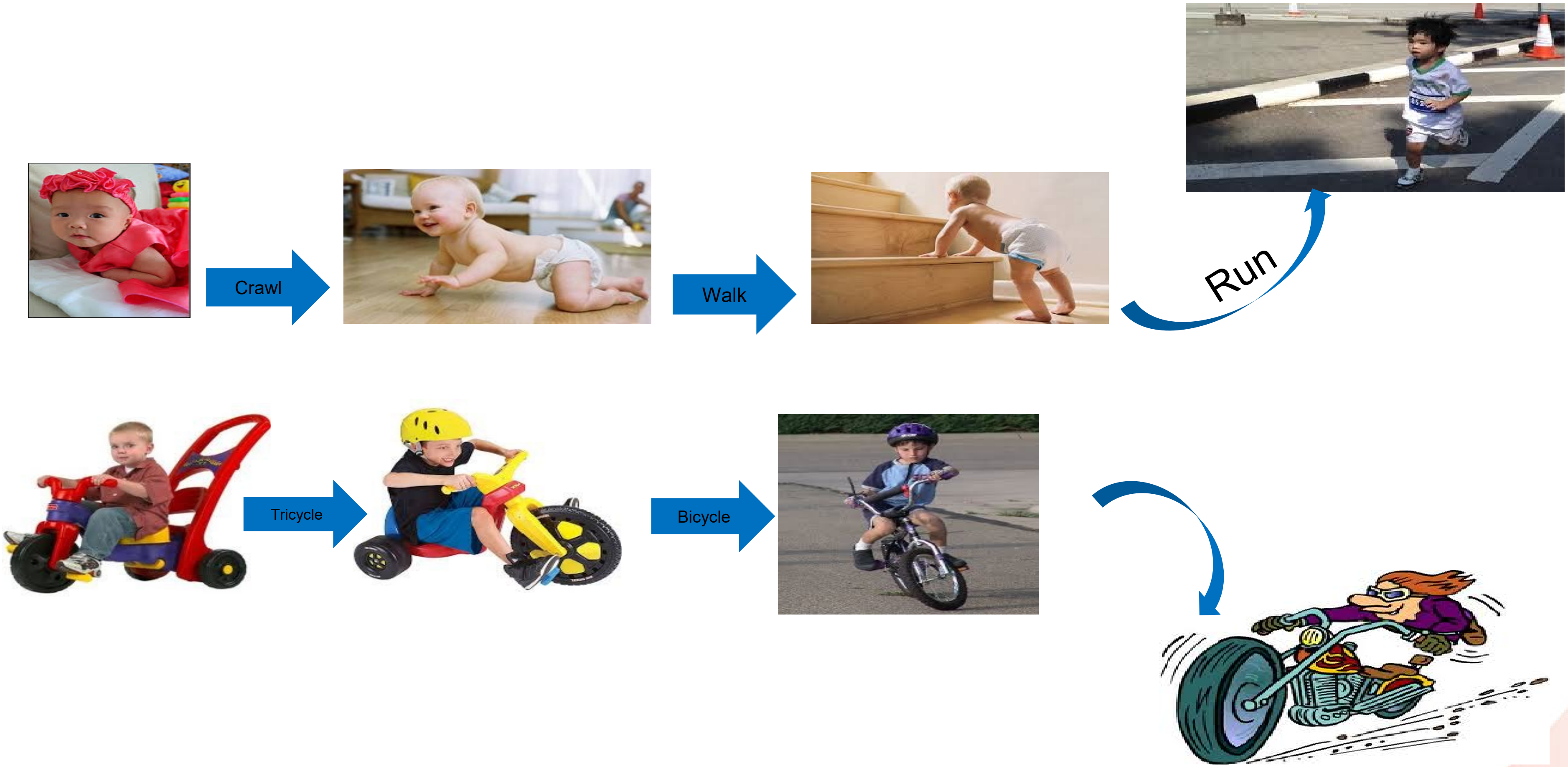
Towards 'General Optimization Intelligence'

TRANSFER OPTIMIZATION

General Optimization Intelligence

NANYANG TECHNOLOGICAL UNIVERSITY

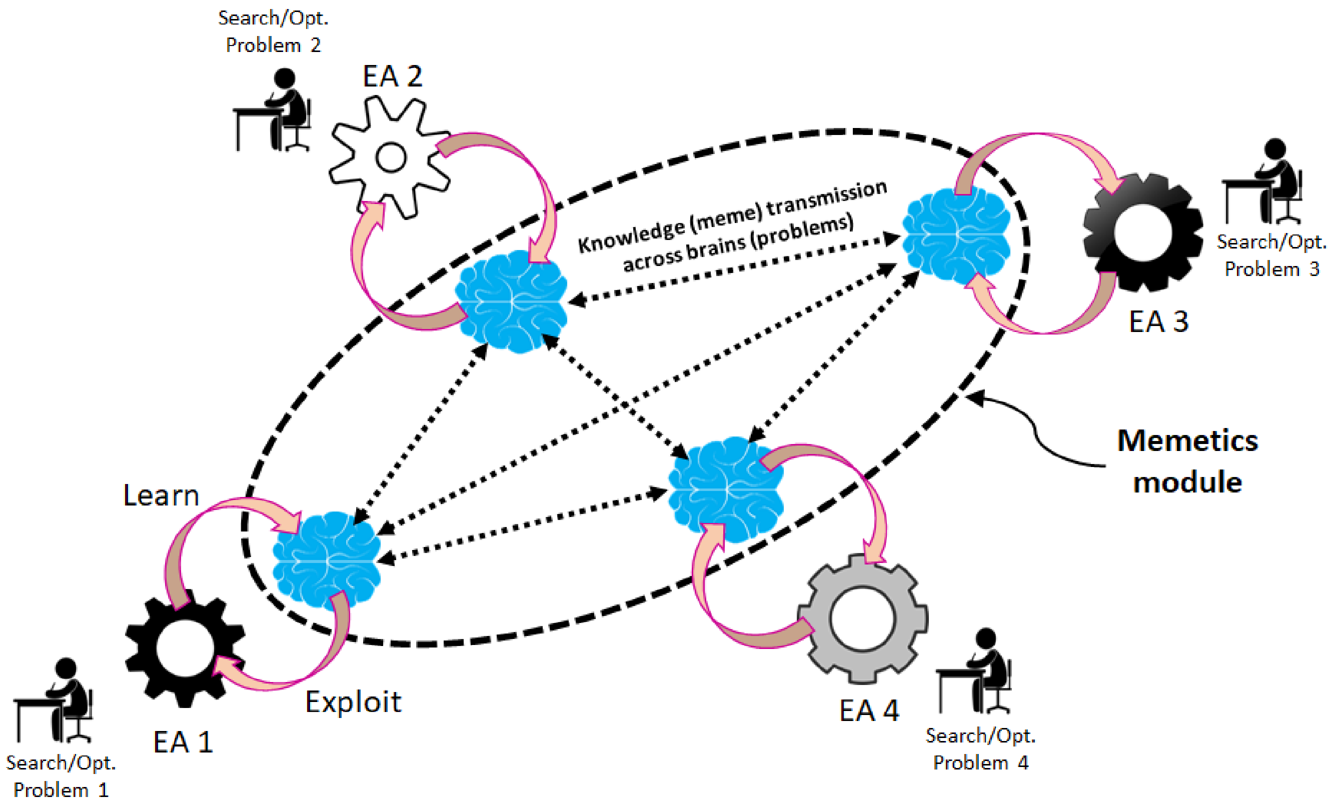
Key Inspiration of Knowledge Transfer via memetic computation



This is how humans operate to solve real-world problems more efficiently !!!

NANYANG TECHNOLOGICAL UNIVERSITY

A Modern Interpretation of "Memes"



In modern memetic computation, the notion of a meme is set free from the narrow scope of a local search scheme, and takes flight to embody potentially diverse forms of **problem-solving knowledge**.

Such memes are expressible in arbitrary computational representations that can be learned from previous source tasks and transferred to a related target task.

Formalizing the Existence of Multiple *Related* Problems (Multi-Problems)

Explicitly stating the *joint* existence of *K* optimization tasks:

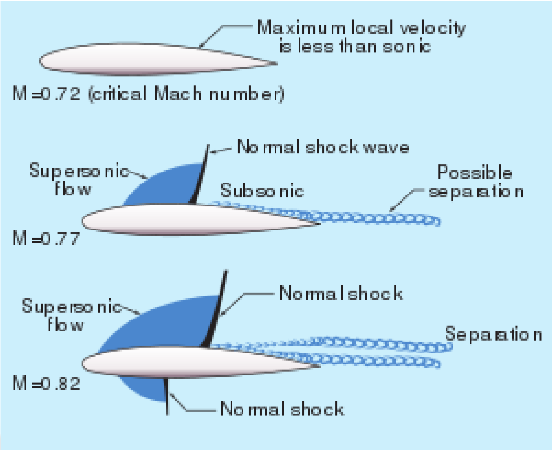
$$\mathcal{T}_k, \forall k \in \{1, 2, \ldots, K\}: \max_{\boldsymbol{x}} f_k(\boldsymbol{x}, \boldsymbol{y}_k),$$

$$\text{such that}, g_{ki}(\boldsymbol{x}, \boldsymbol{y}_k) \leq 0, \text{ for all } i$$

$$\text{and, } h_{kj}(\boldsymbol{x}, \boldsymbol{y}_k) = 0, \text{ for all } j$$

$\boldsymbol{y}_k$:= Environmental variables / operating conditions of the optimization task

Illustrative example: The flight speed (Mach number) marks the operating conditions in aircraft wing design

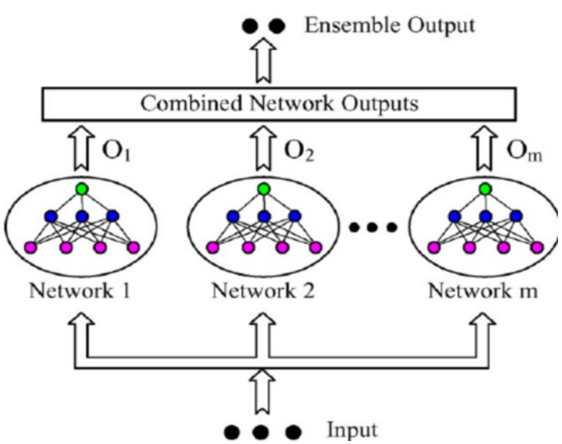


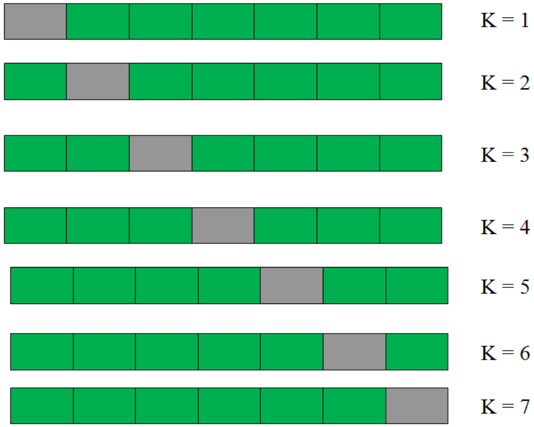
Different operating conditions naturally give rise to a plethora of distinct but related designs!

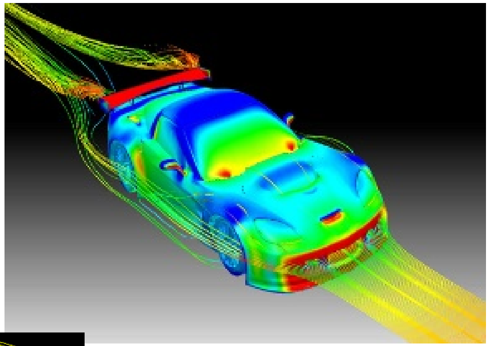
“GOI: main idea is to synergize knowledge of related domains for Multi-Problems Optimization”

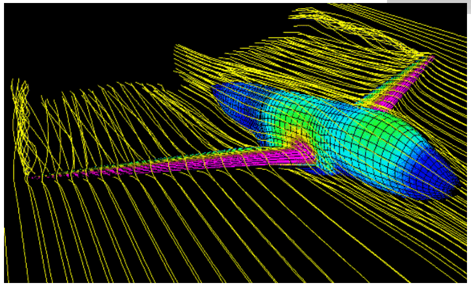
Swarm robotics navigations

Accelerating Model Selection with K-fold Cross-Validation









Sharing knowledge across domains in engineering design

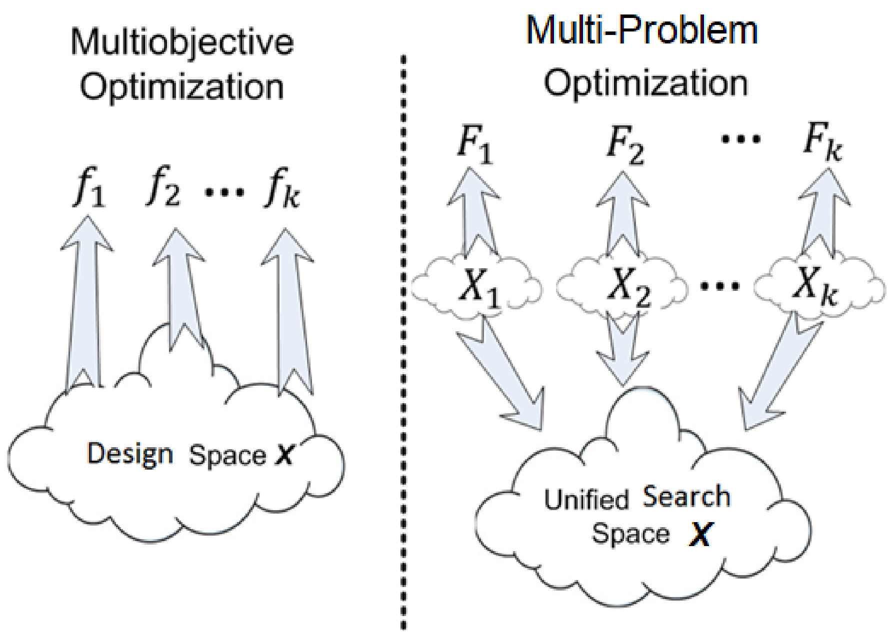
What is the Difference Between Multi-Problems & Multi-Objectives?

➔ In **MULTI-OBJECTIVE OPTIMIZATION**, all functions are inherently defined in a common input / design / search space.

➔ In **MULTI-PROBLEM SETTINGS**, distinct search spaces corresponding to different optimization problems exist. All these are subsequently **mapped into a unified space** in which knowledge (meme) transfer can occur.

➔ In **MULTI-PROBLEM SETTINGS**, we do not seek trade-offs between problems.

➔ We aim to globally optimize each problem more efficiently through the use of related problem-solving experiences.

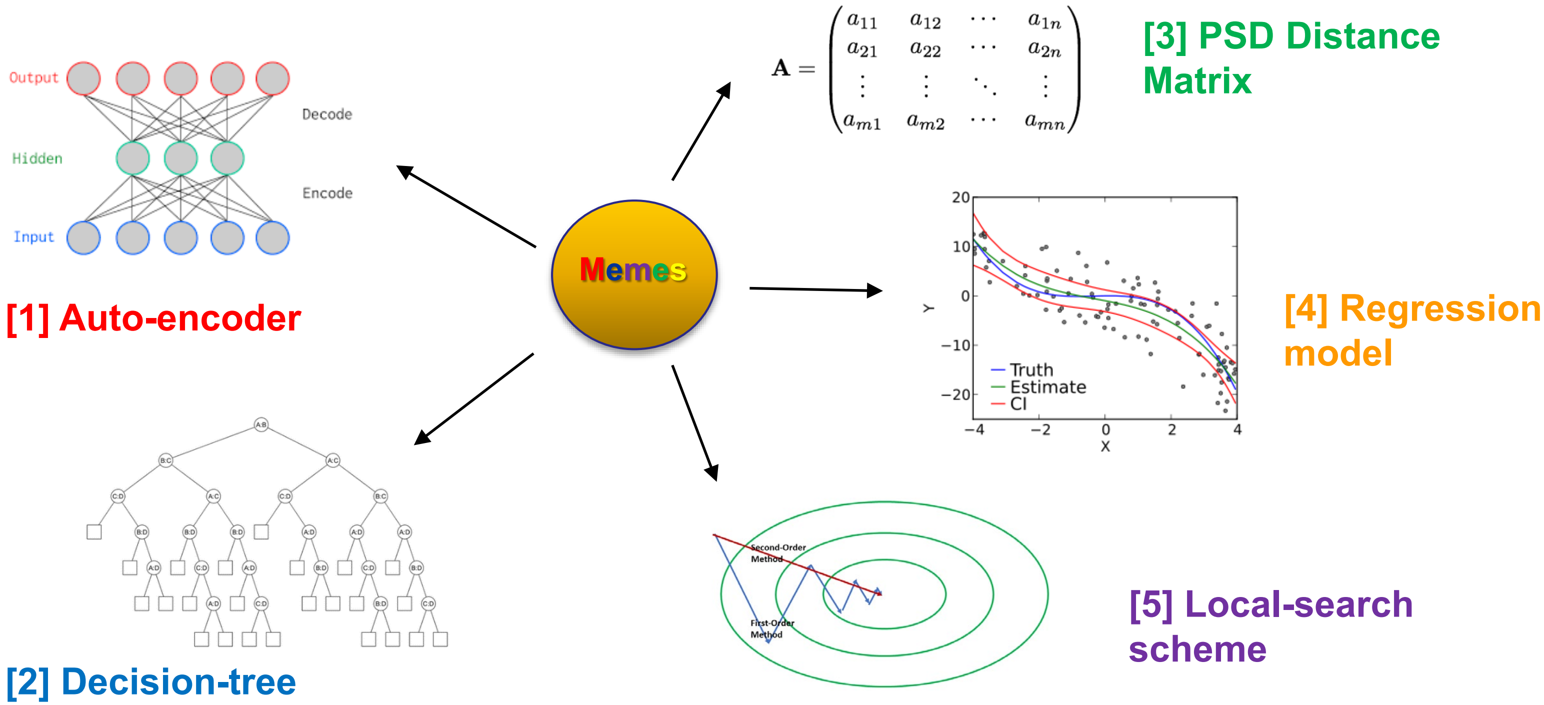


NANYANG TECHNOLOGICAL UNIVERSITY

Meme realizations for Multi-Problem Settings

Many alternate representations of Memes exist!

➔ Memes have been used to induce an informed search bias



[1] Autoencoding evolutionary search with learning across heterogeneous problems. **IEEE Transactions on Evolutionary Computation**

[2] Towards a new Praxis in optinformatics targeting knowledge re-use in evolutionary computation: simultaneous problem learning & optimization. **Evol. Intell.**

[3] Memes as building blocks: a case study on evolutionary optimization+ transfer learning for routing problems. **Memetic Computing**

[4] Multi-Problem Surrogates: Transfer Evolutionary Multiobjective Optimization of Computationally Expensive Problems. **IEEE Transactions on Evol. Comp.**

[5] A probabilistic memetic framework. **IEEE Transactions on Evolutionary Computation**

A Probabilistic Interpretation of memes

NANYANG TECHNOLOGICAL UNIVERSITY

A Unifying Probabilistic Interpretation of Memes

➔ Consider an abstract interpretation of memes as '**Probabilistic models'** that explicitly capture the search distribution bias

➔ **Specifically, given a series of optimization tasks $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_K\}$, we define a meme drawn from task $\mathcal{T}_k$ as $m_k \rightarrow p_k^{t_{budget}}(x)$, such that,**

$$\int f_k(\boldsymbol{x}, \boldsymbol{y}_k) \cdot p_k^{t_{budget}}(\boldsymbol{x}) \cdot \mathrm{d}\boldsymbol{x} \geq f_k^* - \varepsilon_k$$

Where,

(1) f_k^* is the global optimum of $\mathcal{T}_k$, and ε_k is the expectation gap

(2) t_{budget} is the temporal budget of the base optimizer, e.g. EA

NANYANG TECHNOLOGICAL UNIVERSITY

SEQUENTIAL TRANSFER OF MEMES ACROSS PROBLEMS

An evolutionary search paradigm that learns & gets smarter with problems solved…

NANYANG TECHNOLOGICAL UNIVERSITY

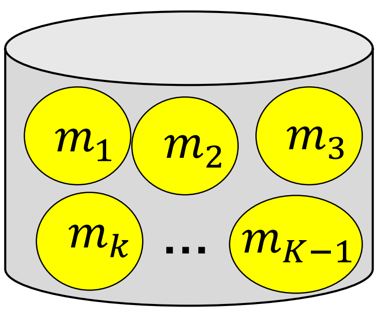
Meme-based Knowledge Transfer across Problems

Consider the following transformation of the Kth optimization task $\mathcal{T}_K$:

$$\underset{\boldsymbol{x}}{\text{maximize}}\; f_K(\boldsymbol{x}, \boldsymbol{y}_k) \quad \Rightarrow \quad \underset{p_K(\boldsymbol{x})}{\text{maximize}} \int f_K(\boldsymbol{x}, \boldsymbol{y}_k) \cdot \boxed{p_K(\boldsymbol{x})} \cdot \mathrm{d}\boldsymbol{x}$$

Population distribution model

Using available knowledge base, we can write $p_K(x)$ as a mixture model:



Knowledge Base of stored memes from past problems solved ➔ {$\mathcal{T}_1$, $\mathcal{T}_2$, …, $\mathcal{T}_{K-1}$ }

$$p_K(\boldsymbol{x}) \sim \sum_{k=1}^{K-1} \alpha_k \cdot p_k^{t_{budget}}(\boldsymbol{x}) + \alpha_K \cdot p'_K(\boldsymbol{x})$$

$$\text{such that, } \alpha_k \geq 0 \wedge \textstyle\sum_{\forall k} \alpha_k = 1$$

Source population distribution biases are explicitly used to direct the target opt. search!

NANYANG TECHNOLOGICAL UNIVERSITY

Insights on Transfer Optimization: Because Experience is the Best Teacher. ***IEEE Transactions on Emerging Topics in Computational Intelligence***, 2(1), 51-64. (2018).

Adapting the Transfer of Knowledge across Problems

➔ **We need to learn α_k values in:** $$p_K(\boldsymbol{x}) \sim \sum_{k=1}^{K-1} \alpha_k \cdot p_k^{t_{budget}}(\boldsymbol{x}) + \alpha_K \cdot p'_K(\boldsymbol{x})$$

High values of α_k automatically imply high inter-task relevance / transfer.

Low values of α_k automatically imply low inter-task relevance / transfer.

The learning problem (maximizing log-likelihood of population dataset) solved via Expectation-Maximization (EM) algorithm:

$$\boxed{\text{maximize}_{\forall \alpha_k} \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K-1} \alpha_k \cdot p_k^{t_{budget}}(\boldsymbol{x}_i) + \alpha_K \cdot p'_K(\boldsymbol{x}_i) \right)}$$

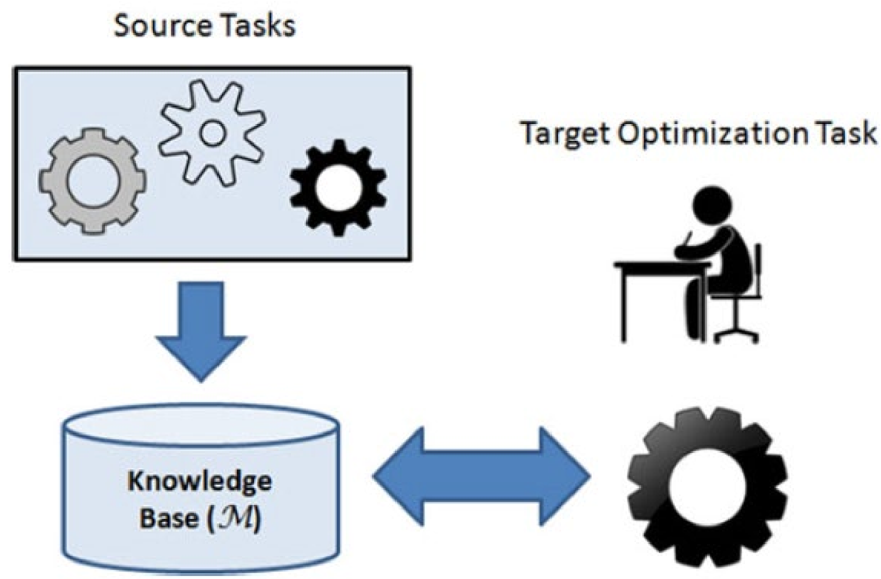
Where, $\boldsymbol{x}_i$ is the *i*th member of the population

NANYANG TECHNOLOGICAL UNIVERSITY

Curbing negative influences online for seamless transfer evolutionary optimization. **IEEE Transactions on Cybernetics.**

Summarizing the Knowledge Transfer Framework

➔ **We are interested in solving the target optimization task $\mathcal{T}_K$, aided by the knowledge acquired from source tasks $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_{K-1}\}$**



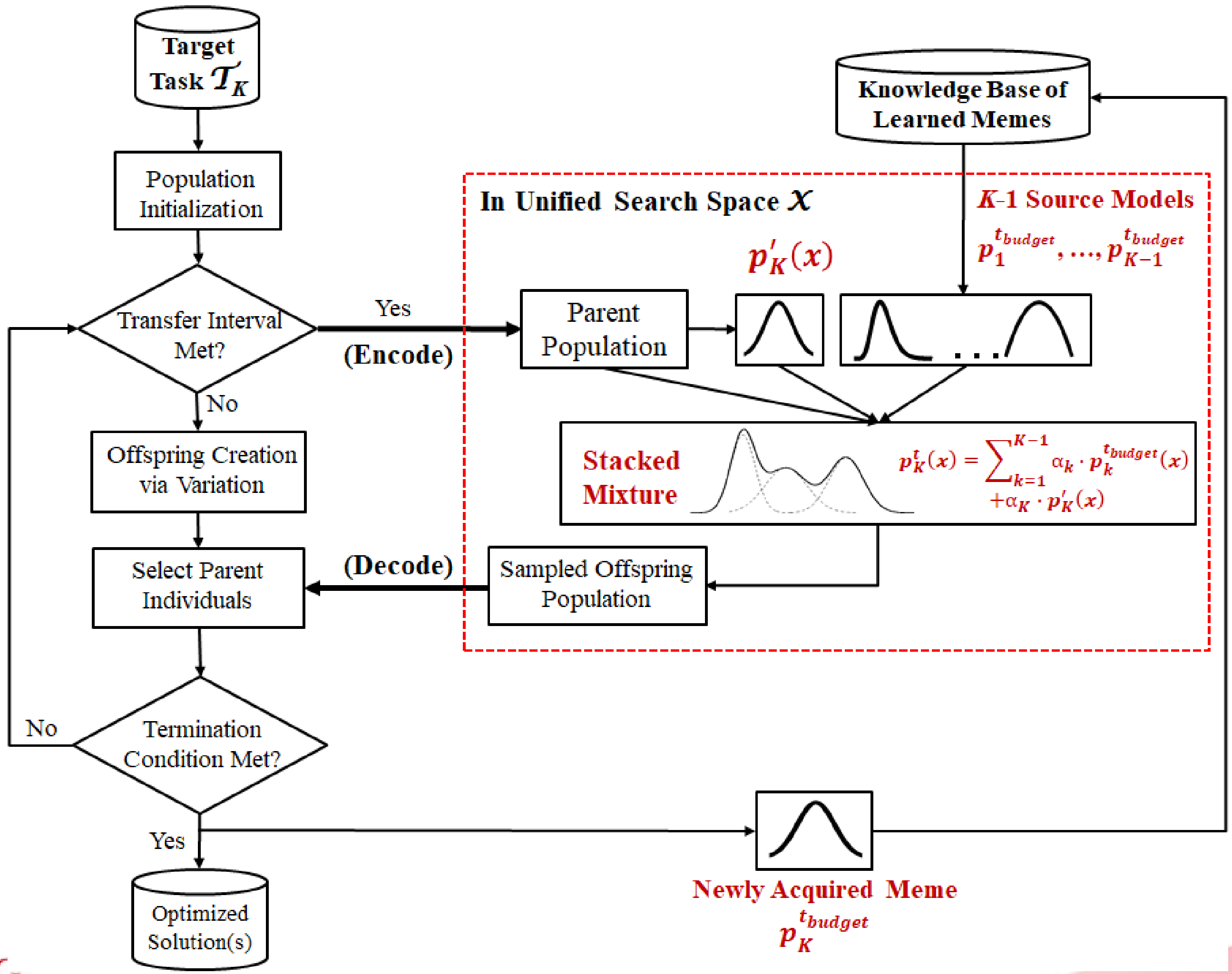
A novel reformulation of $\mathcal{T}_K$, given source models (memes):

$$\underset{\{\alpha_1,\ldots,\alpha_{K-1},\alpha_K,p'_K(\boldsymbol{x})\}}{\text{maximize}} \int f_K(\boldsymbol{x}, \boldsymbol{y}_K) \cdot \left[\boxed{\sum_{k=1}^{K-1} \alpha_k \cdot p_k^{t_{budget}}(\boldsymbol{x})} + \alpha_K \cdot p'_K(\boldsymbol{x})\right] \cdot \mathrm{d}\boldsymbol{x}$$

$$\text{s.t., } \textstyle\sum_{k=1}^{K} \alpha_k = 1, \text{ and } \alpha_k \geq 0, \forall k.$$

Source models induced search bias on $\mathcal{T}_K$

Algorithmic Realization of an Adaptive Memetic Transfer EA



Curbing negative influences online for seamless transfer evolutionary optimization. **IEEE Transactions on Cybernetics**.

Theoretical Rationale of the Proposed Framework

Theoretical Result 1. *In the limiting case of infinitely large population size, asymptotic global convergence of a probabilistic model-based EA is guaranteed if the learned model – from which offspring are sampled – precisely captures the true underlying distribution of the parent population.*

In theory, the performance of transfer optimization should be no worse than "conventional optimization without transfer" – as long as the computational cost associated with learning the mixture of models is affordable.

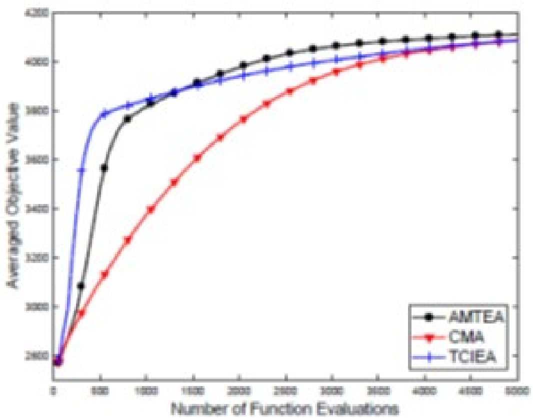
Theoretical Result 2. *The EM algorithm is guaranteed to find the globally minimum distribution gap (measured by the Kullback-Leibler divergence) between the stacked mixture model and the true underlying distribution of the parent population*

Proof. Curbing negative influences online for seamless transfer evolutionary optimization. **IEEE Transactions on Cybernetics**.

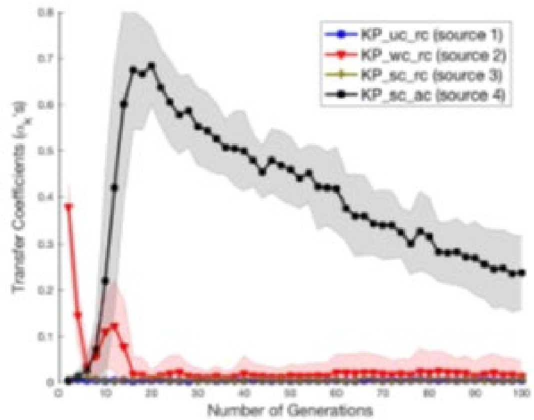
NANYANG TECHNOLOGICAL UNIVERSITY

Combinatorial Application: Knapsack Problem

*** The distribution model used in these examples is the *factored Bernoulli distribution***



(c) KP_uc_ac



(d) Transfer coefficients learned for KP_uc_ac

AMTEA: Adaptive memetic transfer EA

CMA: Canonical MA

TCIEA: Transfer case-injected EA

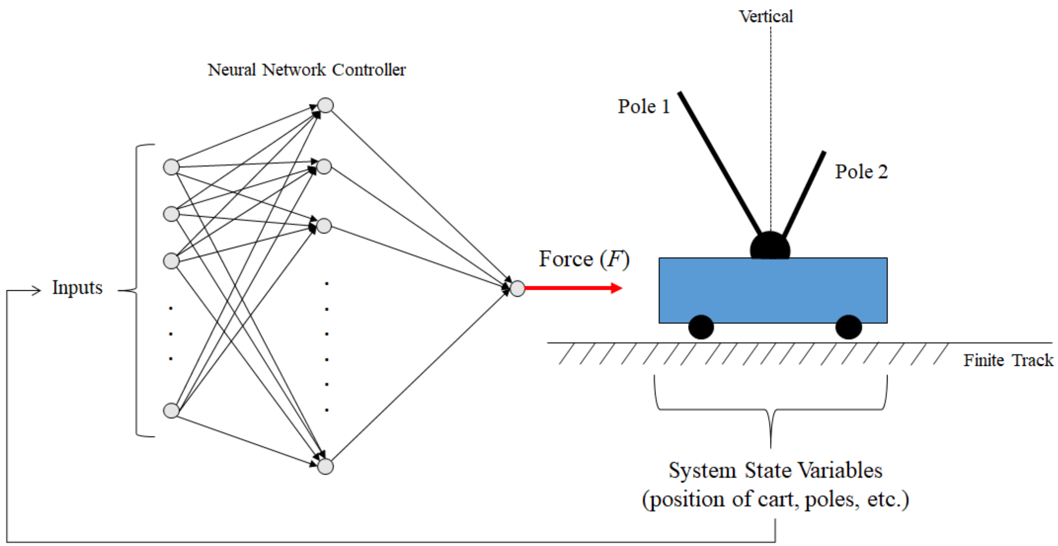
*** Key observation (right panel):**

Meaningful inter-task relevance coefficients (α_k) are found to be automatically learned – without any human intervention!

NANYANG TECHNOLOGICAL UNIVERSITY

Reinforcement Learning Example: Neuro-Evolutionary Pole-Balancing Controller

- ➔ Many distinct problems are formed by altering the lengths of the poles
- ➔ **Length of Pole 1 (l_1) is fixed as 1 meter. Length of Pole 2 (l_2) is altered to construct multiple related tasks**



Source Tasks: $l_2 = 0.6, 0.65, 0.7, 0.75, 0.775$ meters, respectively

Target Task: $l_2 = 0.8$ meters

Methods	Successes	Function Evaluations
CEA	0/50	NA
NES	1/50	8977
TCIEA	1/50	8900
AMTEA	22/50	7918 ± 1241

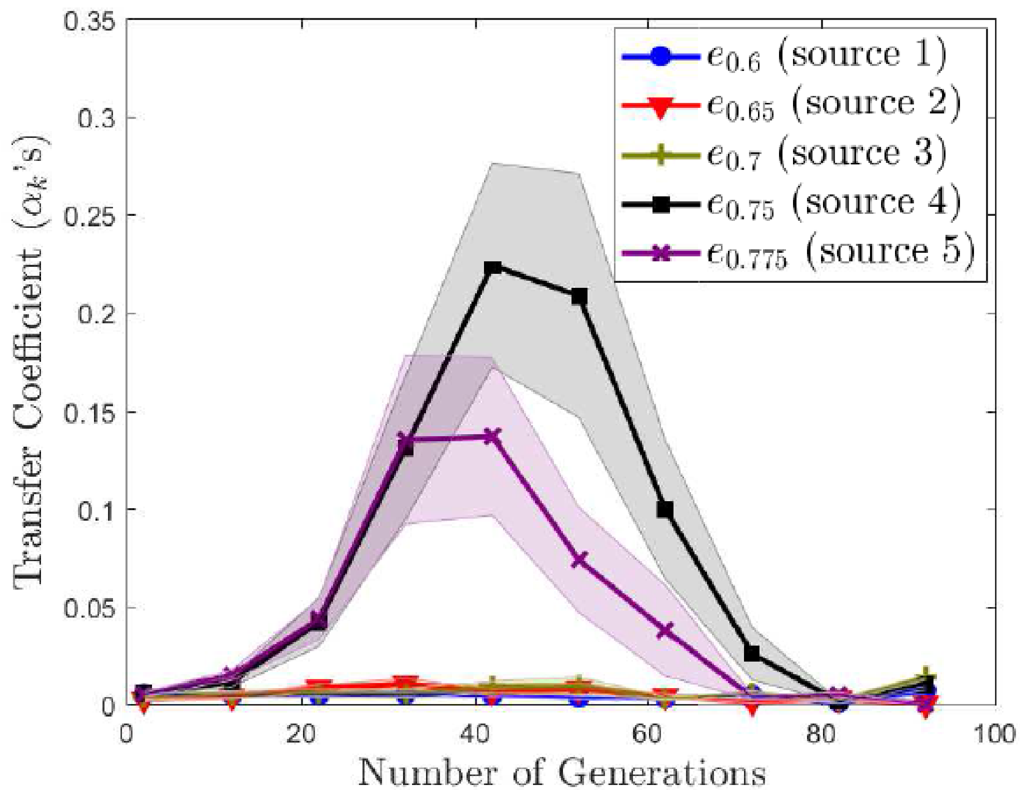
NES: Natural evolution strategies – popular for RL

TCIEA: Transfer case-injected EA

CEA: Canonical EA

NANYANG TECHNOLOGICAL UNIVERSITY

Reinforcement Learning Example: Neuro-Evolutionary Pole-Balancing Controller



*** Key observation:**

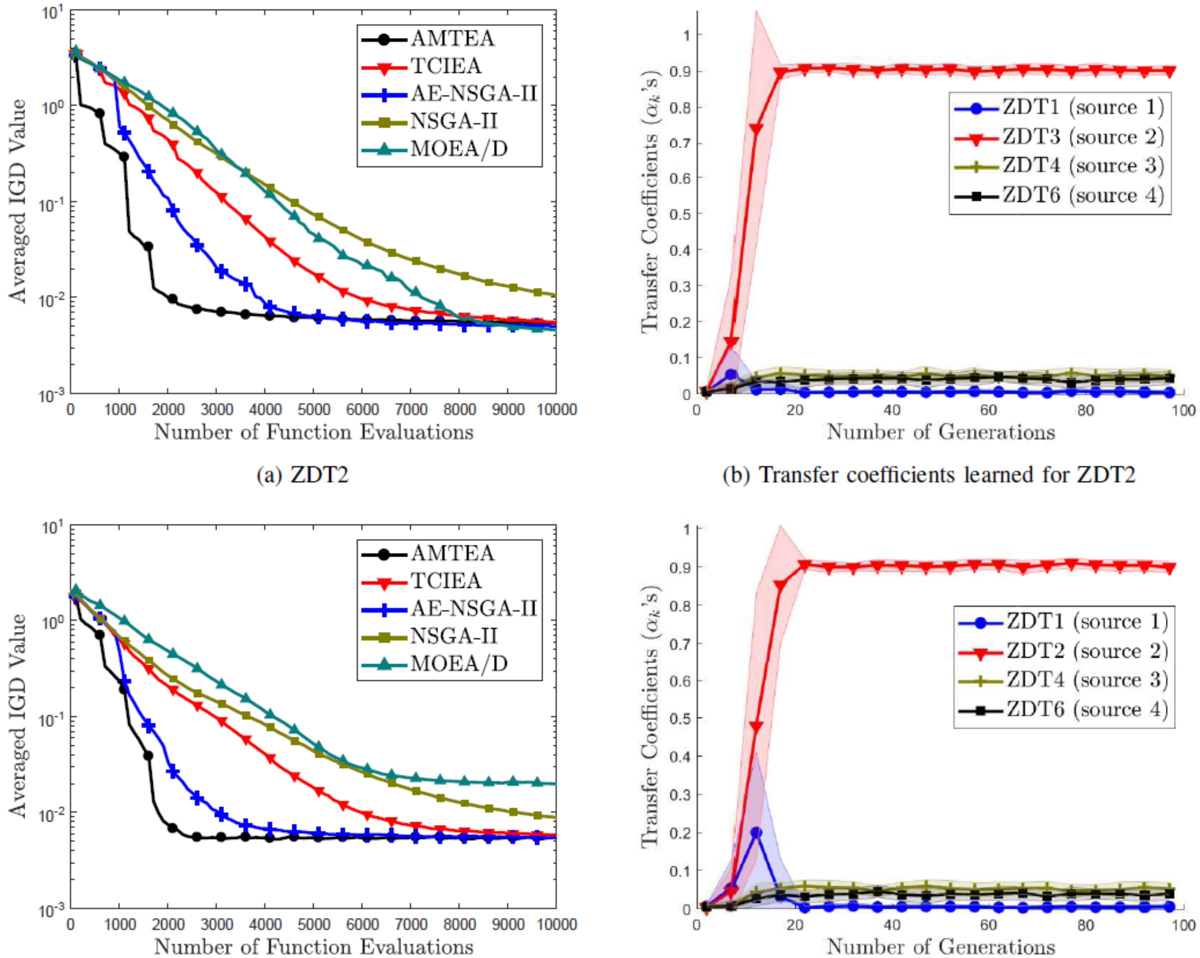
Meaningful inter-task relevance coefficients (α_k) are found to be automatically learned – without any human intervention!

NANYANG TECHNOLOGICAL UNIVERSITY

Curbing negative influences online for seamless transfer evolutionary optimization. **IEEE Transactions on Cybernetics**.

Example in Benchmark Multi-Objective Optimization

*** The distribution model used in these examples is the *multivariate normal distribution***



(a) ZDT2

(b) Transfer coefficients learned for ZDT2

(c) ZDT3

(d) Transfer coefficients learned for ZDT3

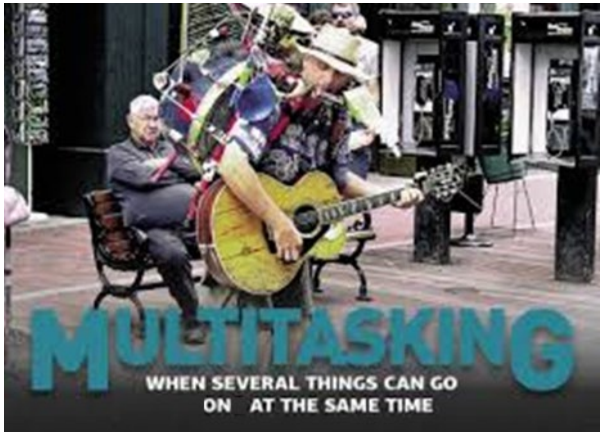
AMTEA: Adaptive memetic transfer EA

AE-NSGA-II: Autoencoder-based transfer + NSGA-2

TCIEA: Transfer case-injected EA

*** Key observation (right panels):**

Meaningful inter-task relevance coefficients (α_k) are found to be automatically learned – without any human intervention!



Multitask Meme transfer across Problems

A framework for knowledge exploitation across multiple optimization problems solved in tandem

NANYANG TECHNOLOGICAL UNIVERSITY

The ability of the human mind to manage and execute multiple tasks in what seems like apparent simultaneity is perhaps once of its most remarkable features…

Evolutionary Multitasking

Can I solve multiple Optimization problems at the same time

&

yet faster???

A. Gupta, Y. S. Ong, L. Feng, "**Multifactorial Evolution: Toward Evolutionary Multitasking**" **IEEE Transactions in Evolutionary Computation**, 2016.

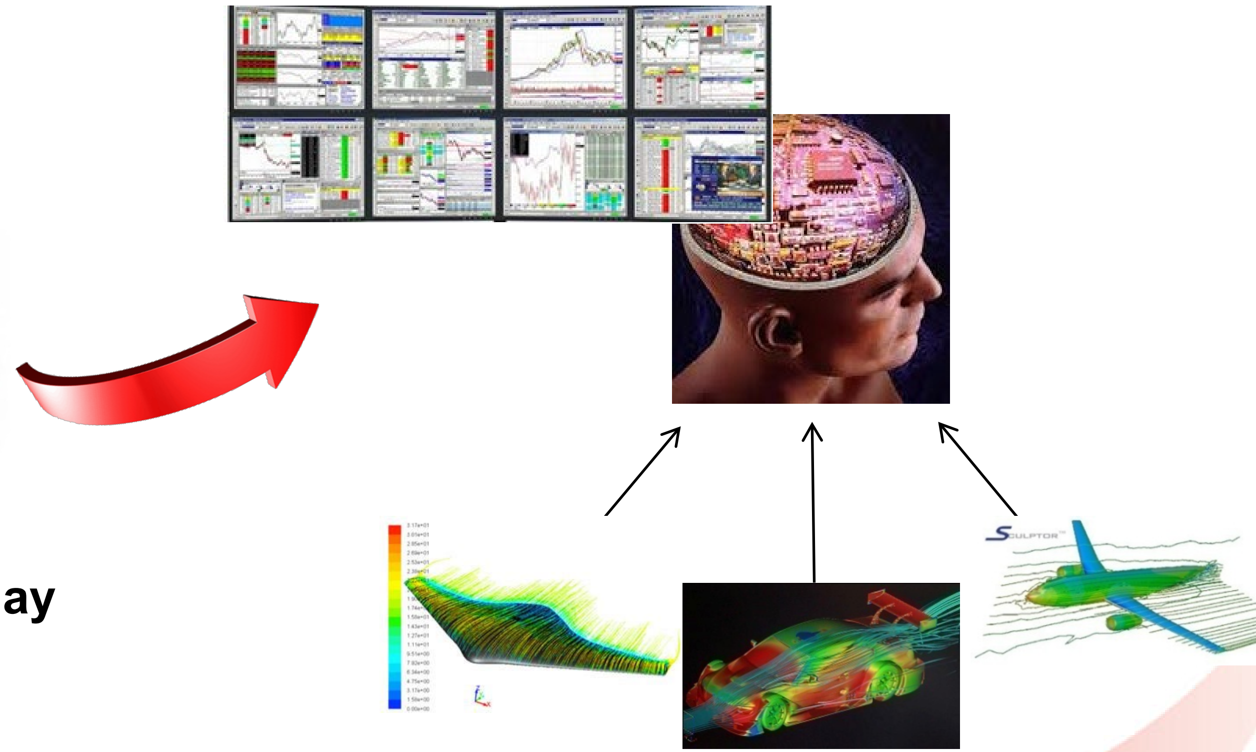
A. Gupta, Y. S. Ong, L. Feng and K. C. Tan, "**Multi-Objective Multifactorial Optimization in Evolutionary Multitasking**", **IEEE Transactions on Cybernetics**, 2017.

K. K. Bali, Y. S. Ong, A. Gupta and P. S. Tan, "**Multifactorial Evolutionary Algorithm with Online Transfer Parameter Estimation: MFEA-II**" **IEEE Transactions in Evolutionary Computation**, 2019.

NANYANG TECHNOLOGICAL UNIVERSITY

Evolutionary Multitasking

"in this fast-paced, technologically-driven world that we live in, **multitasking** is perhaps the best way to fit in all of our priorities…"



The ubiquity of present-day
Cognitive Multitasking

Y. S. Ong et al., **"Evolutionary Multitasking: A Computer Science View of Cognitive Multitasking"**, **Cognitive Computation**, Vol. 8, No. 2, pps. 125-142, 2016.

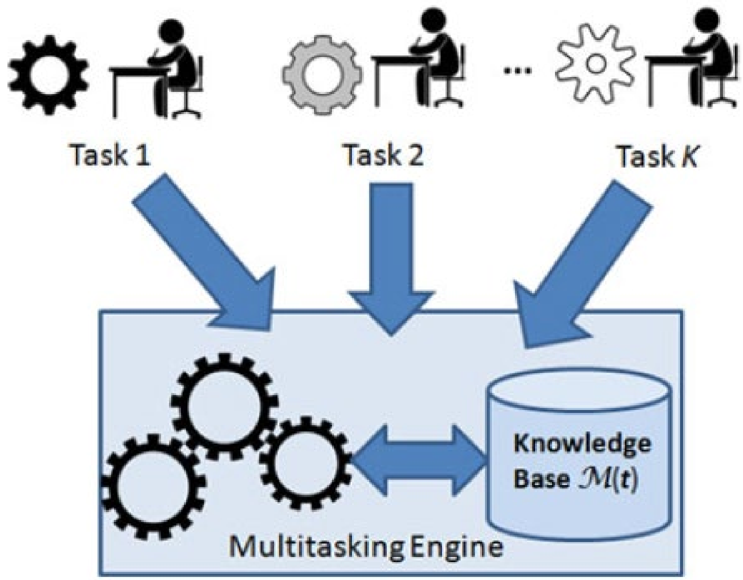
An Explicit Probabilistic Interpretation of memes for transfer

Evolutionary Multitasking

NANYANG TECHNOLOGICAL UNIVERSITY

The Multitask Knowledge Transfer Framework

- ➔ A set of K optimization tasks are assumed to **occur at the same time**
- ➔ We are interested in solving all $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_K\}$ tasks in tandem with the scope of "**adaptive**" **knowledge transfer**
- ➔ **We store and harness a dynamic knowledge base $\mathcal{M}(t)$ of partially evolved memes**



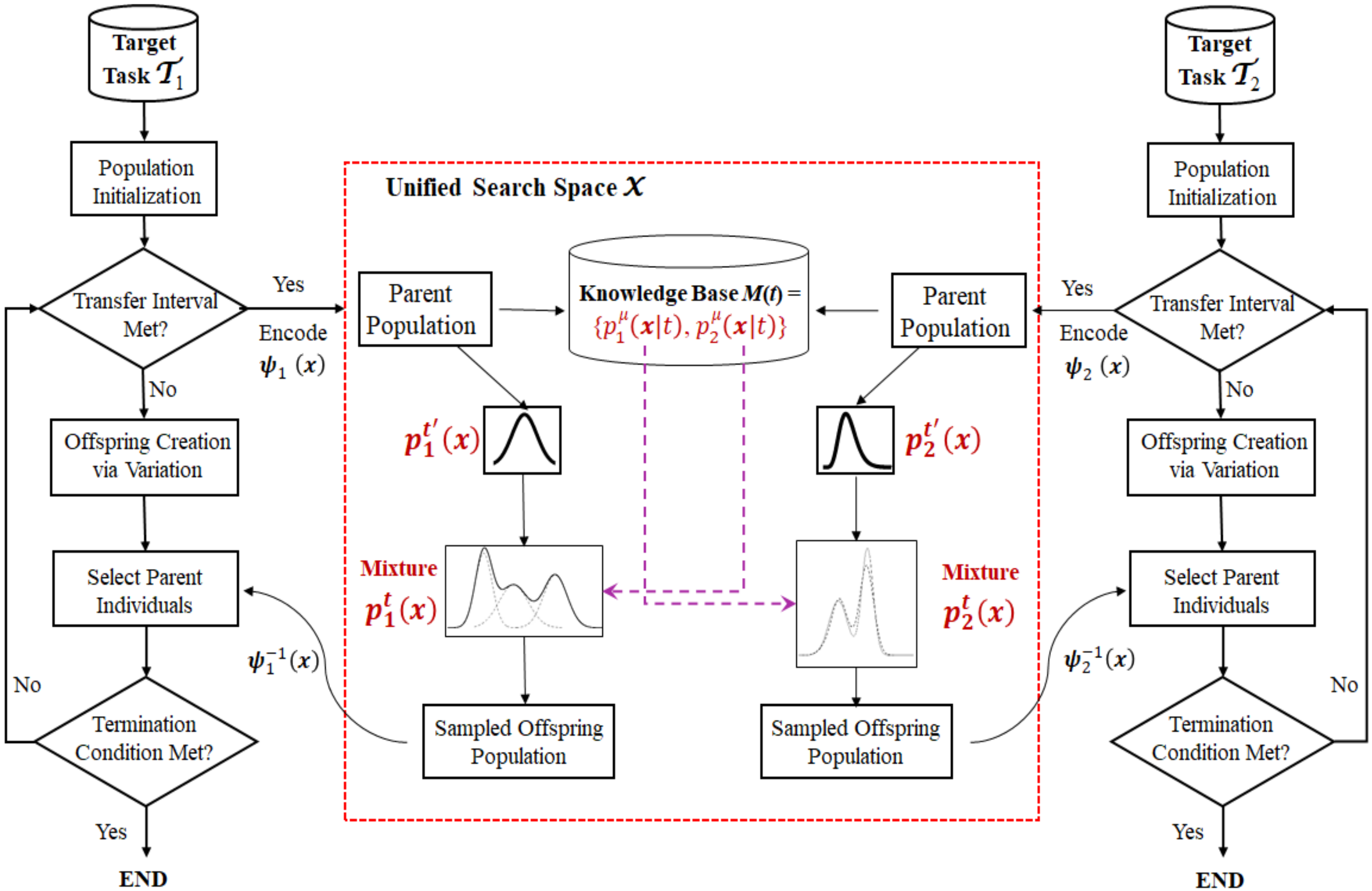
A novel mathematical formulation of *jointly* solving $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_K\}$:

$$\max_{\{w_{jk}, p_j(\boldsymbol{x})\ \forall j,k\}} \sum_{k=1}^{K} \int f_k(\boldsymbol{x}, \boldsymbol{y}_k) \cdot \left[\sum_{j=1}^{K} w_{jk} \cdot p_j(\boldsymbol{x})\right] \cdot \mathrm{d}\boldsymbol{x},$$

$$\text{such that, } \sum_{j=1}^{K} w_{jk} = 1, \forall k, \text{ and } w_{jk} \geq 0, \forall j, k$$

NANYANG TECHNOLOGICAL UNIVERSITY

Algorithmic Realization of an Adaptive Multitask EA (simple 2 task case)

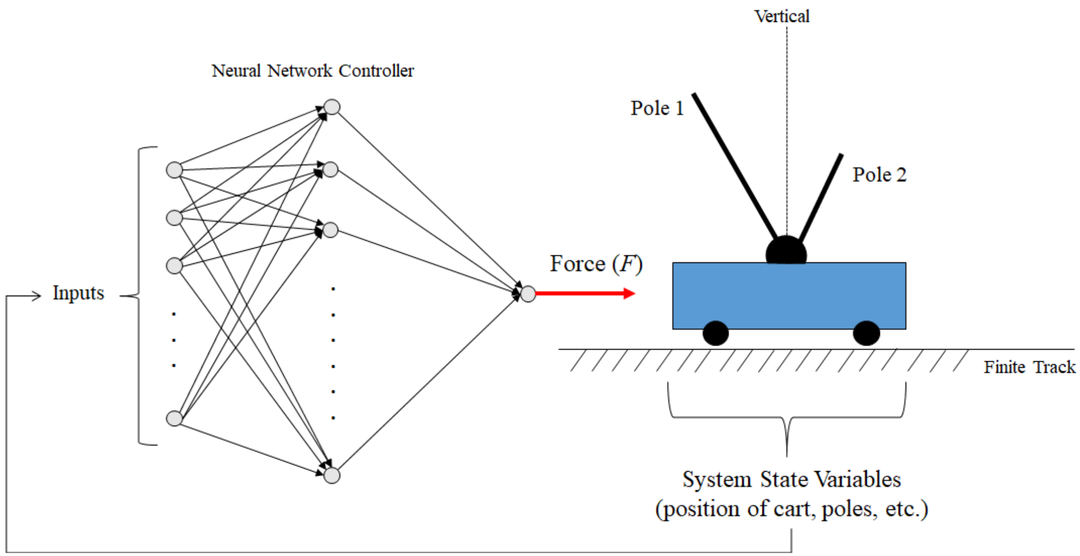


NANYANG TECHNOLOGICAL UNIVERSITY

Memetic Computation: The Mainspring of Knowledge Transfer in the Data-Driven Optimization Era. **in "Studies in Adaptation, Learning, and Optimization". Chapter 6.**

Reinforcement Learning Example: Neuro-Evolutionary Pole-Balancing Controllers

➔ Length of Pole 1 (l_1) is fixed as 1 meter. Length of Pole 2 (l_2) is altered to construct multiple related tasks



*** Results of simultaneously solving 3 distinct double-pole balancing tasks with multitask knowledge transfer**

Short pole length (l_2)	Success rate	
	Simple EA	Adaptive Memetic Multi-task EA
0.60 meters	85%	**90%**
0.65 meters	45%	**75%**
0.70 meters	10%	**35%**

NANYANG TECHNOLOGICAL UNIVERSITY

Evolutionary Multitasking

An Implicit transfer of knowledge via genetic mechanisms

A. Gupta, Y. S. Ong, L. Feng, "***Multifactorial Evolution: Toward Evolutionary Multitasking***," **IEEE Transactions in Evolutionary Computation**, Vol. 20, No. 3, pp. 343-357, **2016**.

K. K. Bali, Y. S. Ong, A. Gupta and P. S. Tan, "**Multifactorial Evolutionary Algorithm with *Online Transfer Parameter Estimation: MFEA-II***" **IEEE Transactions in Evolutionary Computation**, *In Press*, **2019**.

NANYANG TECHNOLOGICAL UNIVERSITY

Multitasking Problem Formulation

- Consider a situation where K optimization tasks are to be performed simultaneously.

- The i^{th} task, denoted T_i, has a scalar objective function $F_i : X_i \rightarrow \mathbb{R}$ to be minimized.

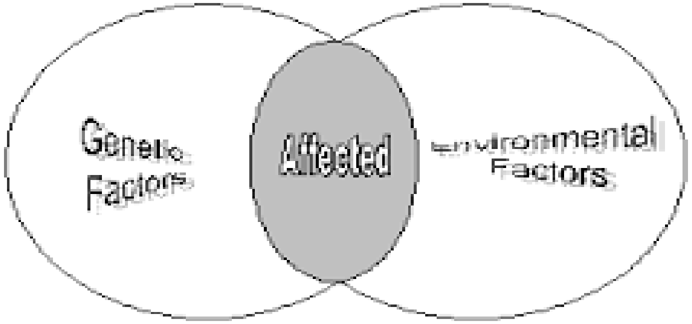
- Evolutionary Multitasking builds on the implicit parallelism of population-based search with the aim to simultaneously find

$$\{x_1, x_2, \ldots, x_{K-1}, x_K\} = argmin \{F_1(x), F_2(x), \ldots, F_{K-1}(x), F_K(x)\}.$$

NANYANG TECHNOLOGICAL UNIVERSITY

Multifactorial Optimization (MFO) for Multitasking

- $\{x_1, x_2, \ldots, x_{K-1}, x_K\} = argmin\ \{F_1(x), F_2(x), \ldots, F_{K-1}(x), F_K(x)\}$.
- Each F_i is treated as an additional *factor* influencing evolution.
- The problem is referred to as a ***K*-factorial problem** and the formulation is labelled as



Multifactorial Optimization (MFO)

NANYANG TECHNOLOGICAL UNIVERSITY

A. Gupta, Y. S. Ong, L. Feng, "***Multifactorial Evolution: Toward Evolutionary Multitasking***," **IEEE Transactions in Evolutionary Computation**, Vol. 20, No. 3, pp. 343-357, 2016.

Some definitions in *Multifactorial Evolution*

- For every individual p_i in a population *P we* define:

  - (*Factorial rank*): Factorial rank r_{ij} is the rank of p_i on task T_j, relative to all other individuals in P

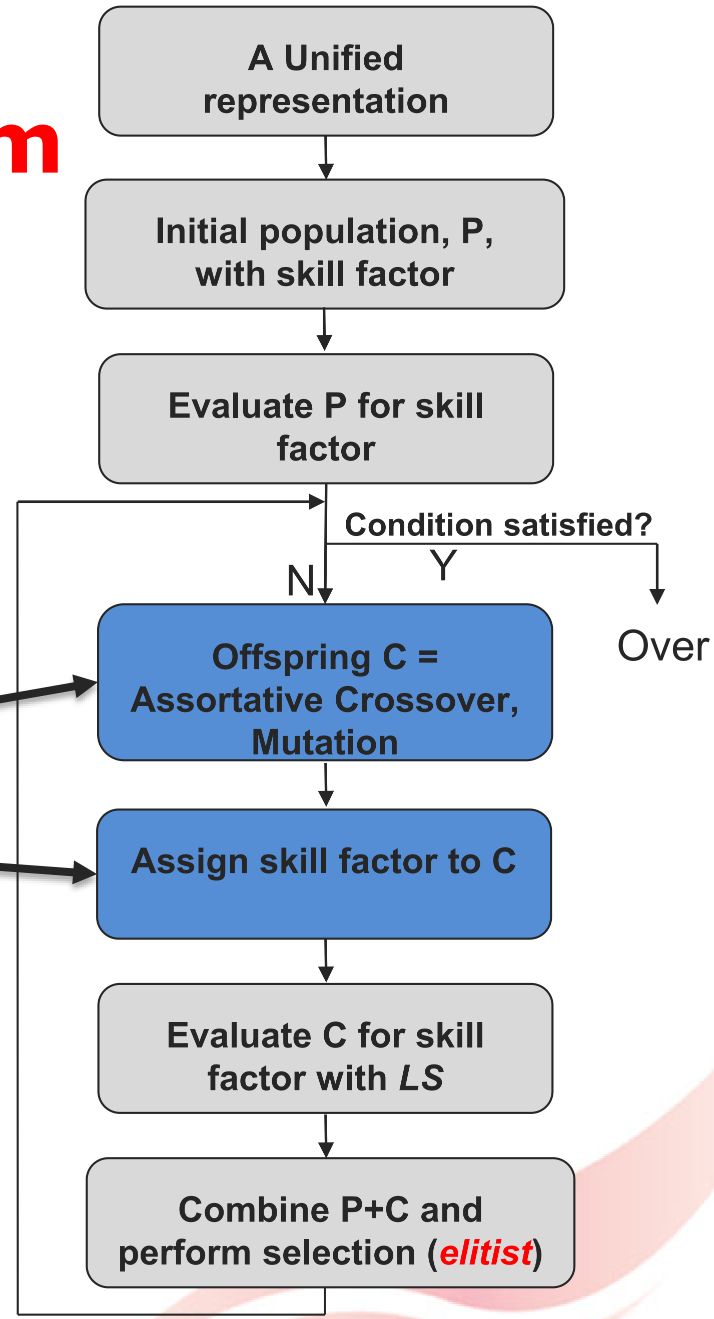
  - (*Scalar fitness*): Scalar fitness φ_i of p_i is based on its best rank over all tasks; i.e. $\varphi_i = 1/\min\{ r_{i1}, r_{i2}, \ldots, r_{iK}\}$.

  - (*Skill factor*): Skill factor τ_i of p_i is the one task, amongst all other tasks in MFO, with which the individual is associated. This may be defined as $\tau_i = argmin_j\{ r_{ij} \}$.

NANYANG TECHNOLOGICAL UNIVERSITY

The Multifactorial Evolutionary Algorithm (MFEA)

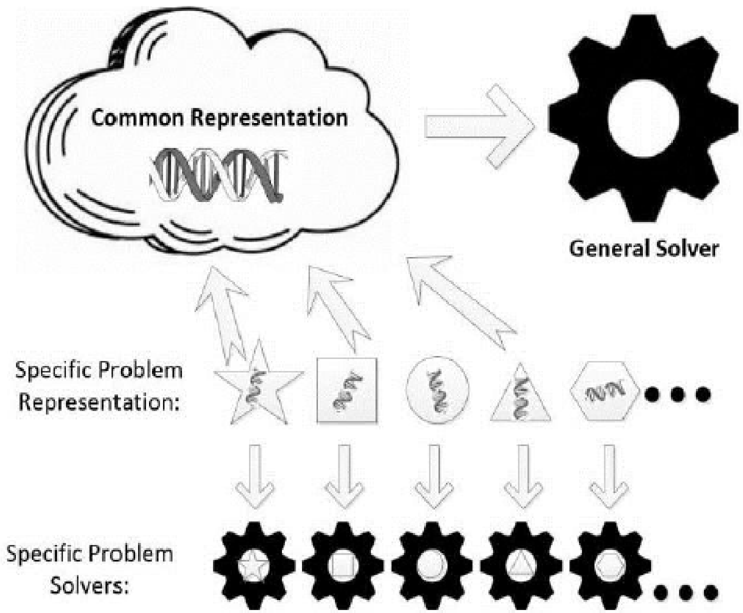
- **An algorithm inspired by the biological concept of Multifactorial Inheritance.**
- **Gene-cultural interaction forms the crux of evolutionary algorithms**
  - **Assortative mating**
  - **Vertical cultural transmission**



NANYANG TECHNOLOGICAL UNIVERSITY

(1) Important Ingredient: Population initialization with a 'Unified Chromosome Representation'

- With *K* optimization tasks to be performed simultaneously, the dimensionality of the i^{th} task is given by D_i.
- Accordingly, we define a **Unified search space** with **dimensionality ($D_{multitask}$) equal to $max_i\{D_i\}$**.
- During the population initialization step, every individual is assigned a vector of $D_{multitask}$ **random keys** that lie in the fixed range [0, 1].
- While addressing task T_i, we consider the 'Random-key chromosome representation', with D_i denoting the first of the random-keys of the chromosome.

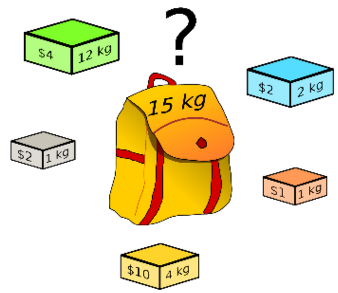


We do not append, we unify

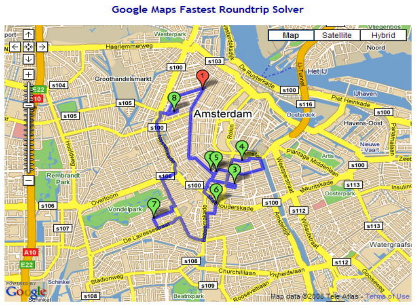
NANYANG TECHNOLOGICAL UNIVERSITY

(1.1) Random-Key Encoding + Decoding Exemplar

Multitasking Problem



&



Task 1: 12-D Knapsack

Task 2: 6-D TSP

Sample 12-D Chromosome

0.81	0.90	0.13	0.91	0.63	0.09	0.28	0.55	0.96	0.45	0.16	0.97

Binary Decoding

1	1	0	1	1	0	0	1	1	0	0	1

12-D Knapsack Solution

Rank	1	2 …				6
	6	3	5	1	2	4

6-D TSP Solution

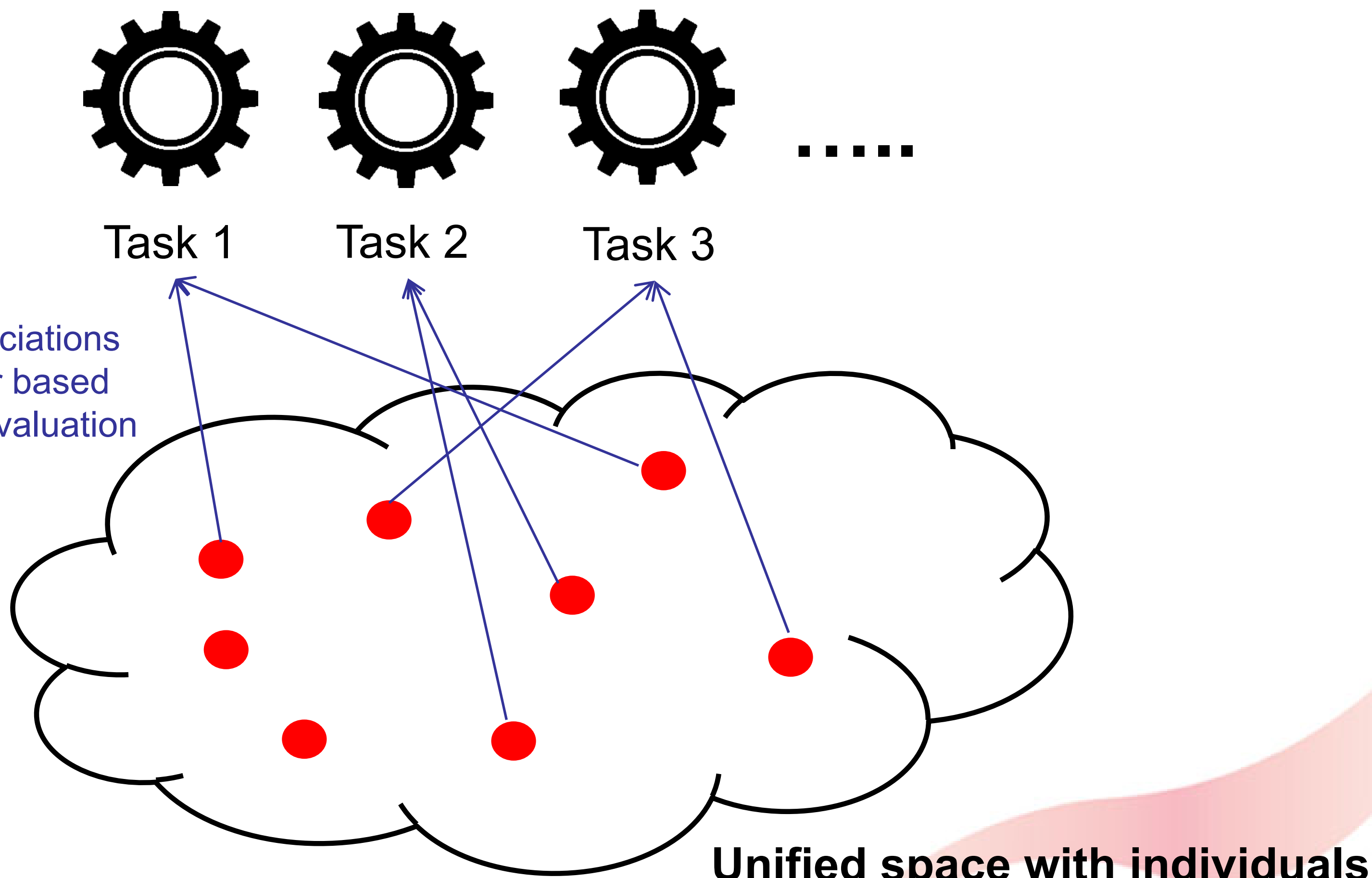
Sequence-Based Decoding

NANYANG TECHNOLOGICAL UNIVERSITY

(1.2) Assigning Skill Factors to Initial Population Members

- **Skill factor is the task with which an individual is associated.**

Create initial associations either randomly or based on performance evaluation



(2) Assortative Mating

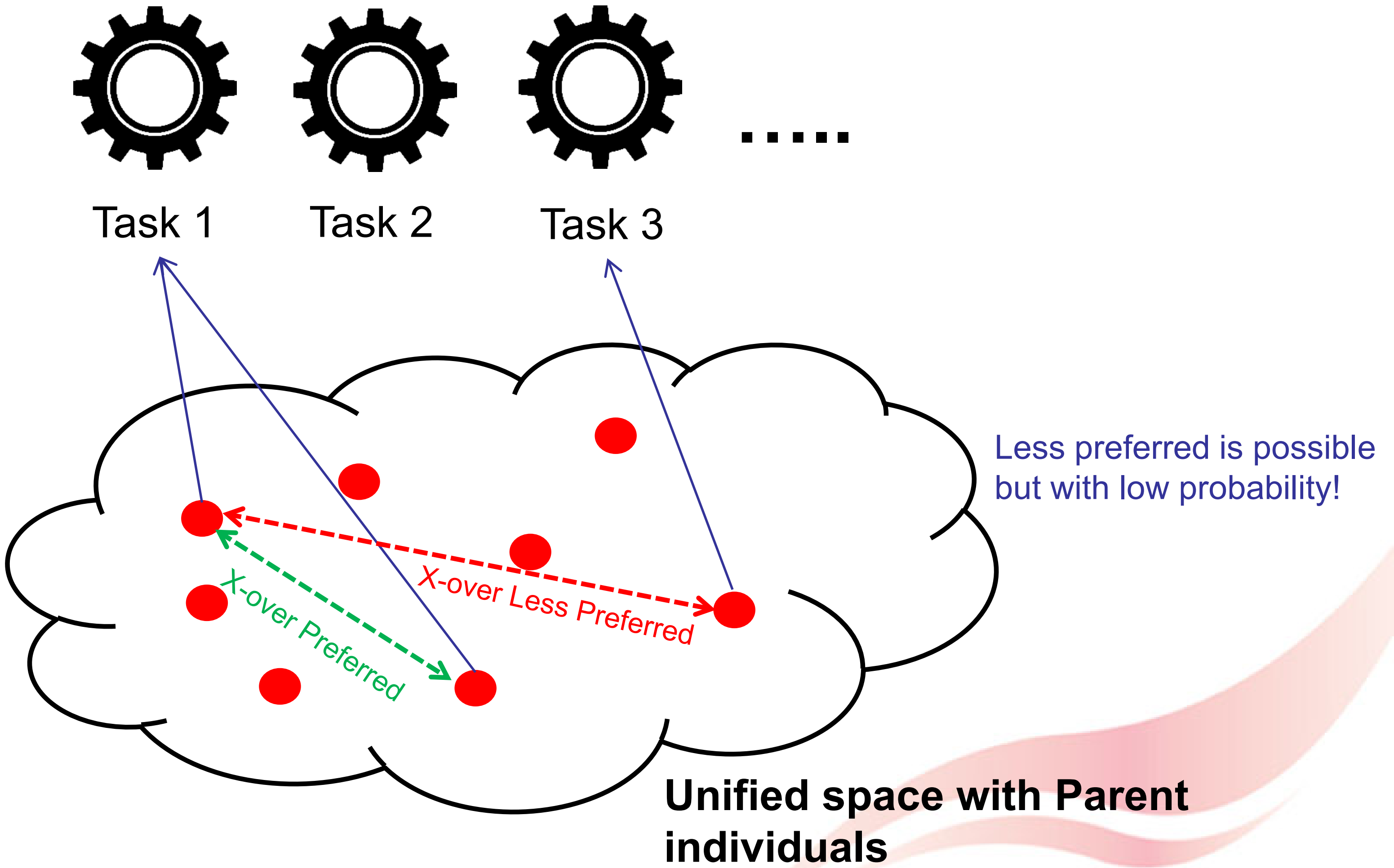
The principle of Assortative Mating suggests that biological entities prefer to *mate with those sharing similar characteristics or similar cultural backgrounds*.

In the MFEA, the above is realized with *individuals preferring to mate with those possessing the same skill factor*.

NANYANG TECHNOLOGICAL UNIVERSITY

(2.1) Assortative Crossover

- **Crossover Preferred with "similar" individuals (having same skill factor), as a means of preventing excessive gene mixing.**



(3) Vertical Cultural Transmission

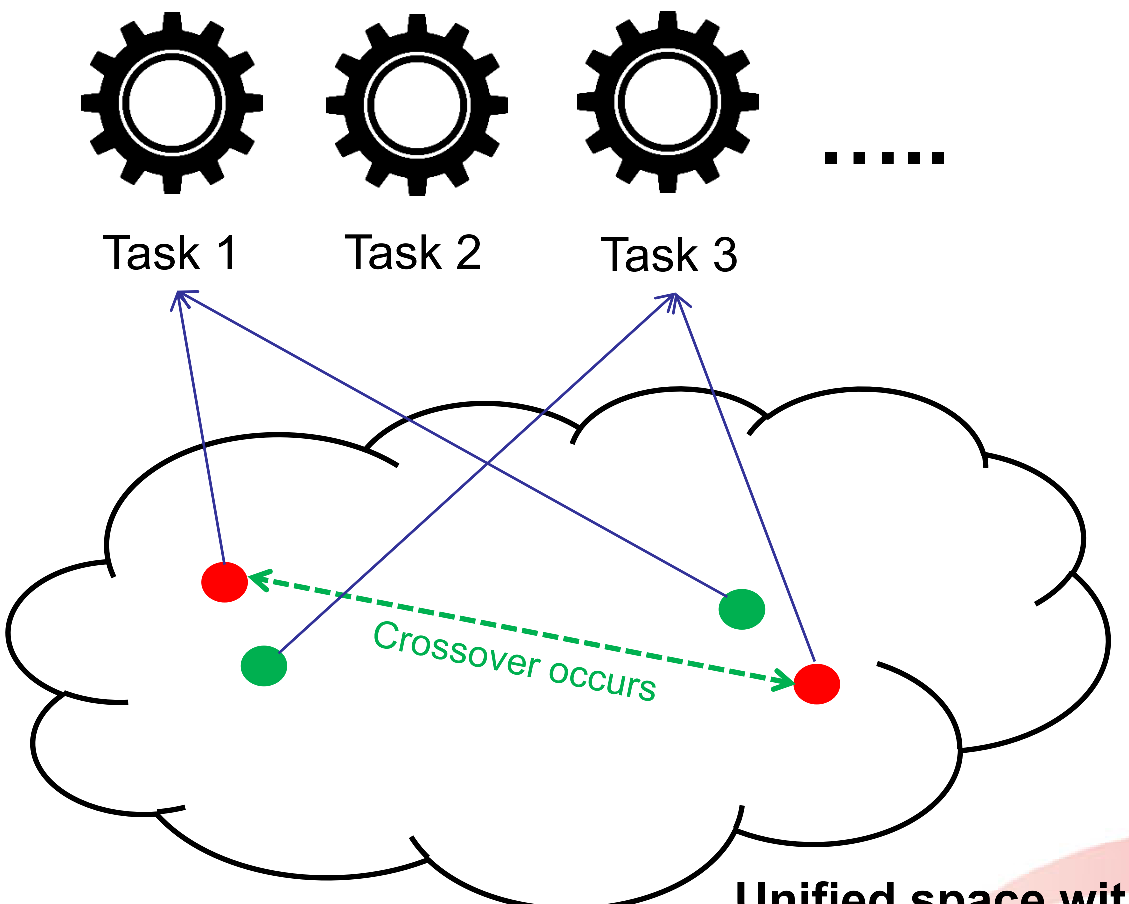
One of the most prevalent forms of Vertical Cultural Transmission is *Offspring Imitating Parents*

In the MFEA, the above is realized with *offspring imitating the skill factor of any one parent at random*.

NANYANG TECHNOLOGICAL UNIVERSITY

(3) Vertical Cultural Transmission

- Each offspring randomly *imitates* the skill factor of any one of its parents.



Unified space with Parent individuals

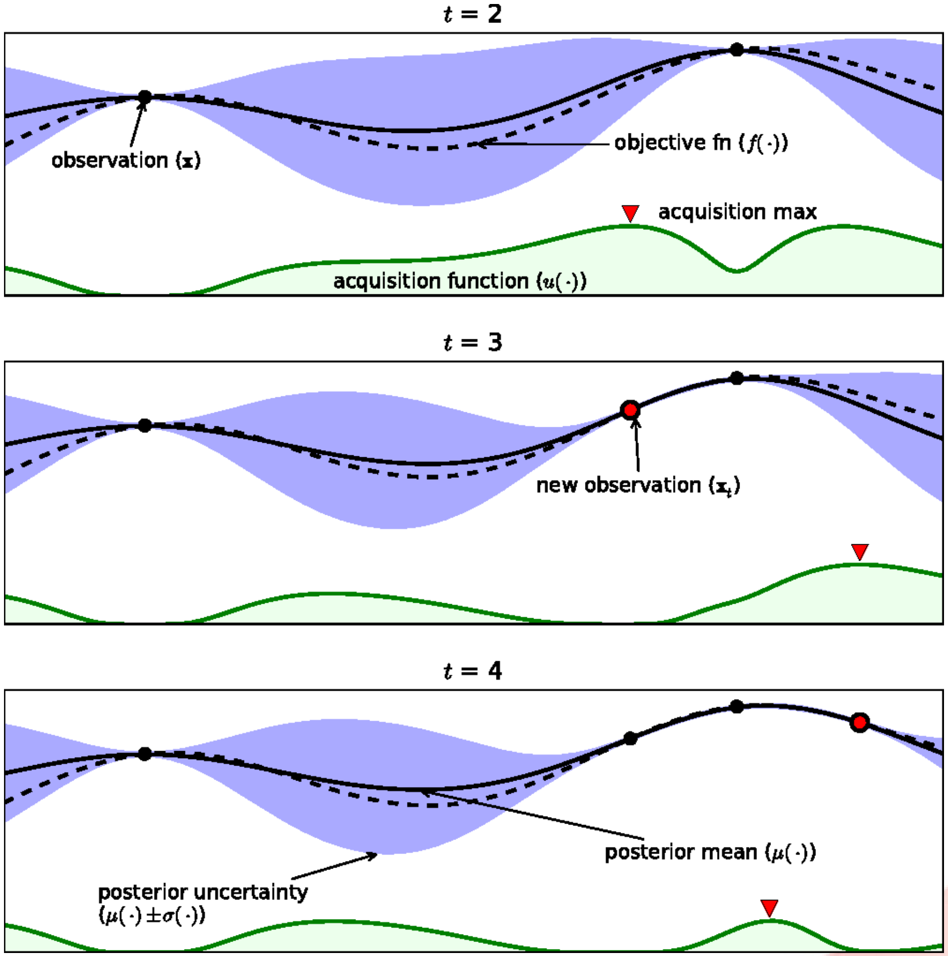
Knowledge Transfers in Computationally Expensive Domains

Transfer Bayesian optimization algorithms and applications

NANYANG TECHNOLOGICAL UNIVERSITY

Bayesian Optimization: A Basic Overview

- Surrogate-assisted optimization utilizing models trained on optimization data.
- Predictive distributions typically obtained from **Gaussian Process Surrogate Models**.
- The method naturally accommodates experiential knowledge incorporation via **Regression Transfer Learning → Transfer/Multi-task Gaussian Process**.



NANYANG TECHNOLOGICAL UNIVERSITY

Case (1)

Characteristic:

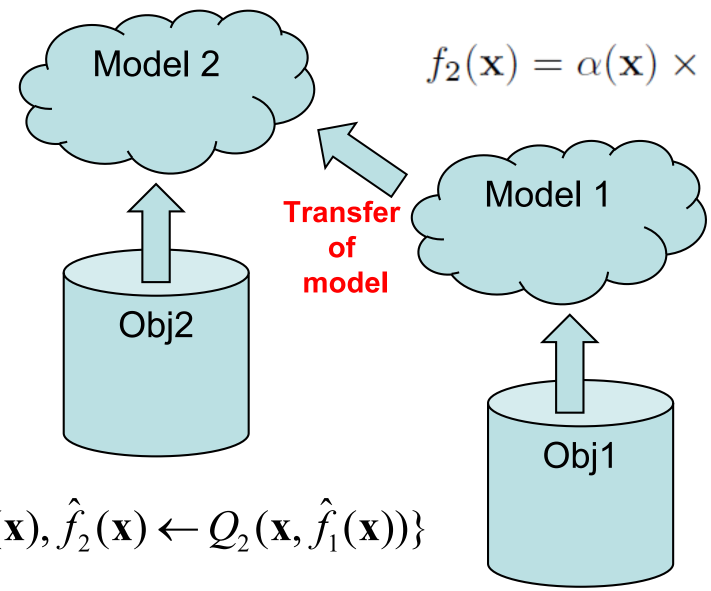
Knowledge transfer across <u>Distinct Objectives</u> of a single MOP

NANYANG TECHNOLOGICAL UNIVERSITY

Transferring Knowledge ‘Across Objectives’ in Multi-objective Optimization

- ➢ ***Exchange and Reuse of Data & Knowledge among Objectives, i.e., Transfer*** **between objectives.**

- ➢ **E.g. Using one a simpler (cheaper) Objective to model the more complex (expensive) one**



$$f_2(\mathbf{x}) = \alpha(\mathbf{x}) \times f_1(\mathbf{x}) + Q(\mathbf{x})$$

$$\{\hat{f}_1(\mathbf{x}) \leftarrow Q_1(\mathbf{x}), \hat{f}_2(\mathbf{x}) \leftarrow Q_2(\mathbf{x}, \hat{f}_1(\mathbf{x}))\}$$

NANYANG TECHNOLOGICAL UNIVERSITY

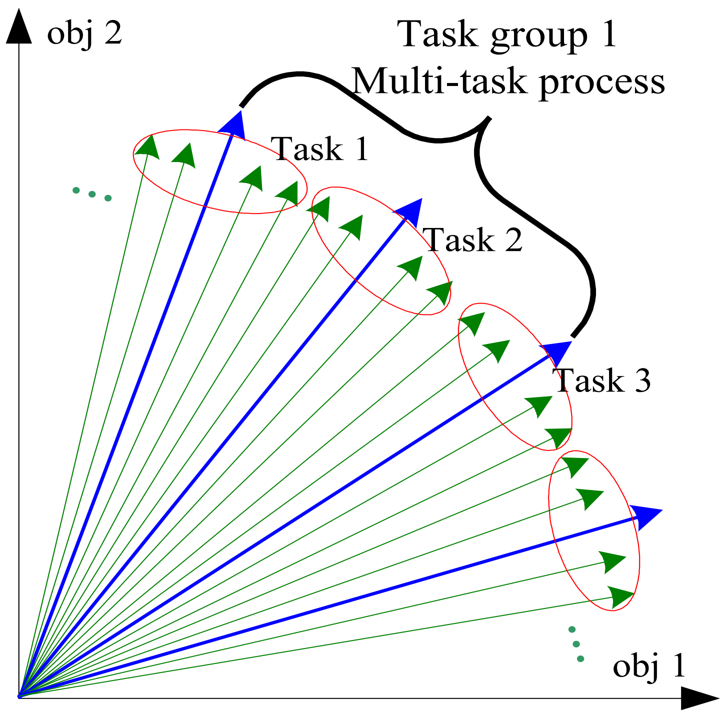
'Multi Co-objective Evolutionary Optimization: Cross Surrogate Augmentation for Computationally Expensive Problems' , IEEE Congress on Evolutionary Computation, **2012**.

Case (2)

Characteristic:

Knowledge transfer across <u>Distinct Sub-problems</u> (PF sectors) of a single MOP

Transferring Knowledge 'Across Subproblems' in Multi-objective Optimization



- □ MOP is decomposed into several subproblems (green arrows)
- □ Subproblems are divided into "*Tasks*" that represent sectors or portions of the *PF* ➔ a.k.a **subPFs**
- □ Pick a group of *adjacent* (neighboring) subPFs to form a Task Group
- □ Augment surrogate modelling via joint learning of adjacent subPFs ➔ *multi-task GP-based Co-SubPF (GCS) modelling*
- □ Optimize surrogate fitness landscapes of all adjacent subPFs to find most promising solutions to exactly evaluate
- □ Append dataset with evaluated solutions

$$\mathrm{cov}[f_l(\mathrm{x}), f_k(\mathrm{x'})] = K^f_{lk} k^x(\mathrm{x},\mathrm{x'})$$

Inter-task similarity capture in MTGP kernel

Standard GP kernel

"Evolutionary Optimization of Expensive Multi-objective Problems with Co-sub-Pareto Front Gaussian Process Surrogates", IEEE Transactions on Cybernetics, 2018.

NANYANG TECHNOLOGICAL UNIVERSITY

Distinguishing Feature of the Multi-Task Gaussian Process (MTGP) model

Cross-Task Covariance Function

Output of lth task   Output of kth task

$$\mathrm{cov}[f_l(\mathrm{x}), f_k(\mathrm{x}')] = K^f_{lk} k^x(\mathrm{x},\mathrm{x}')$$

← MTGP kernel

Inter-task similarity capture between *l*th and *k*th tasks

Conventional GP kernel

$$k^x(\mathrm{x},\mathrm{x}') = \exp(-\sum_{i=1}^{d} \theta_i \,|\, x_i - x_i' \,|^2)$$

Given a total of '*T*' tasks, and '*n*' data points, we construct the following matrices based on the following ➔

$$K^f = \begin{bmatrix} K^f_{11} & \cdots & K^f_{1T} \\ \vdots & \ddots & \vdots \\ K^f_{T1} & \cdots & K^f_{TT} \end{bmatrix}; \quad K^x = \begin{bmatrix} k^x(x_1, x_1) & \cdots & k^x(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k^x(x_n, x_1) & \cdots & k^x(x_n, x_n) \end{bmatrix}$$

Inter-task similarity matrix   Conventional covariance matrix

NANYANG TECHNOLOGICAL UNIVERSITY

Predictive Distribution of the MTGP model

➔ **Surrogate fitness landscape depends on predicted mean and predicted variance at unknown point x***

$$\overline{f_l}(\mathbf{x}^*) = (\mathbf{k}_l^f \otimes \mathbf{k}_*^x)^T C^{-1} \mathbf{y}$$ **Predicted mean for *l*th task**

$$\mathbf{V}(f_l(\mathbf{x}^*)) = K_{ll}^f k^x(\mathbf{x}^*, \mathbf{x}^*) - (\mathbf{k}_l^f \otimes \mathbf{k}_*^x)^T C^{-1} (\mathbf{k}_l^f \otimes \mathbf{k}_*^x)$$ **Predicted variance for *l*th task**

Where, MTGP covariance matrix ➔ $$C = K^f \otimes K^x + \boldsymbol{\sigma} \otimes I$$

Noise term

Kronecker product

Conventional covariance matrix

Inter-task similarity matrix

Transferring Knowledge 'Across Subproblems' in Multi-objective Optimization

Performance of the proposed Gaussian process-based co-sub-PF (GCS) MOE algorithm

Ins.	GCS-MOE	ParEGO		MOEA/D-EGO		NSGA-II		SAMO		K-RVEA	
ZDT1	**0.01325(0.00293)**	0.11230(0.06765)	↑	0.01503(0.00068)	~	0.85724(0.29670)	↑	0.07582(0.02408)	↑	0.02877(0.00920)	↑
ZDT2	**0.00792(0.00063)**	0.10932(0.06396)	↑	0.01525(0.00272)	~	1.49606(0.58739)	↑	0.08898(0.03343)	↑	0.04648(0.02779)	↑
ZDT3	**0.01206(0.00126)**	0.35032(0.09552)	↑	0.06541(0.00986)	↑	1.09520(0.25585)	↑	0.21676(0.09550)	↑	0.03696(0.01186)	↑
ZDT4	1233.95204(388.51472)	39.14445(7.30210)	↓	53.82244(15.77437)	↓	800.89210(417.48926)	~	42.60012(12.36217)	↓	**25.01193(7.76258)**	↓
ZDT6	**0.02365(0.01358)**	0.79778(0.10725)	↑	0.05799(0.01109)	↑	6.13114(0.58605)	↑	1.24964(0.13651)	↑	2.12807(0.46917)	↑
UF1	**0.02625(0.00755)**	0.05947(0.02181)	↑	0.06249(0.03698)	↑	0.07282(0.01130)	↑	0.13539(0.06825)	↑	0.03950(0.01337)	↑
UF2	**0.02538(0.00519)**	0.03694(0.00794)	↑	0.02932(0.00127)	~	0.04481(0.00545)	↑	0.07039(0.02478)	↑	0.03402(0.01107)	~
UF3	**0.30028(0.15409)**	0.40941(0.18354)	~	0.37680(0.00341)	~	0.70364(0.33857)	↑	0.55279(0.16370)	~	0.56284(0.35552)	↑
UF4	**0.04238(0.00155)**	0.06083(0.00600)	↑	0.05037(0.00273)	~	0.07804(0.00888)	↑	0.09624(0.00731)	↑	0.06949(0.02046)	~
UF5	**0.24214(0.07070)**	0.33065(0.10330)	↑	0.82013(0.41080)	↑	0.62764(0.29619)	↑	1.27160(0.41311)	↑	0.75542(0.34913)	↑
UF6	**0.01859(0.01058)**	0.30368(0.06210)	↑	0.56596(0.16924)	↑	0.12482(0.04979)	↑	1.72130(0.60145)	↑	1.28608(0.66696)	↑
UF7	**0.04526(0.00623)**	0.07534(0.05045)	↑	0.07999(0.01933)	↑	0.14890(0.07119)	↑	0.23967(0.06393)	↑	0.11683(0.08174)	↑
UF8	**0.11258(0.04505)**	0.22472(0.04757)	↑	0.12508(0.00528)	~	0.90236(0.77881)	↑	0.40166(0.10030)	↑	0.12571(0.01177)	~
UF9	0.15283(0.02385)	0.20260(0.03131)	↑	0.16880(0.09537)	~	0.74322(0.50683)	↑	0.42848(0.20722)	↑	**0.12248(0.01886)**	↓
UF10	**1.36080(0.50432)**	1.69478(0.60456)	↑	1.64695(0.96512)	↑	2.85992(0.85683)	↑	3.05480(1.13977)	↑	1.60132(0.52306)	↑

NANYANG TECHNOLOGICAL UNIVERSITY

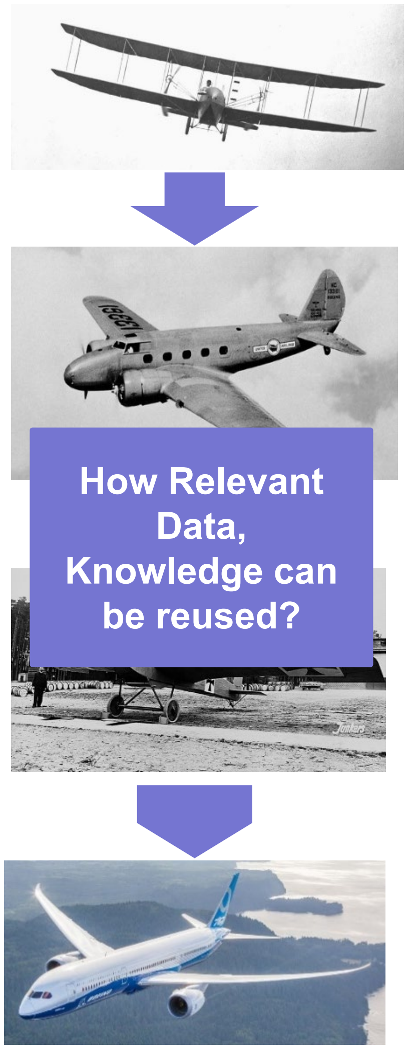
"Evolutionary Optimization of Expensive Multi-objective Problems with Co-sub-Pareto Front Gaussian Process Surrogates", IEEE Transactions on Cybernetics, 2018.

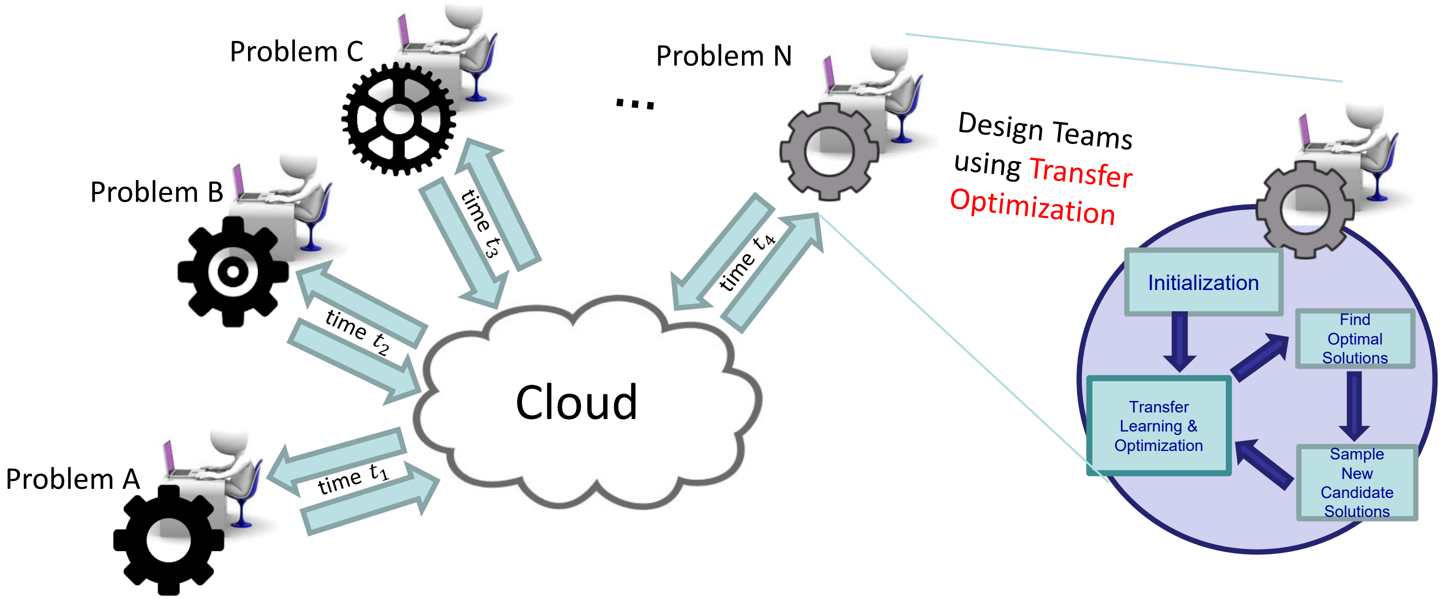
Case (3)

Characteristic:

Knowledge transfer across Distinct MOPs

Knowledge Transfer Across Distinct Expensive Problems



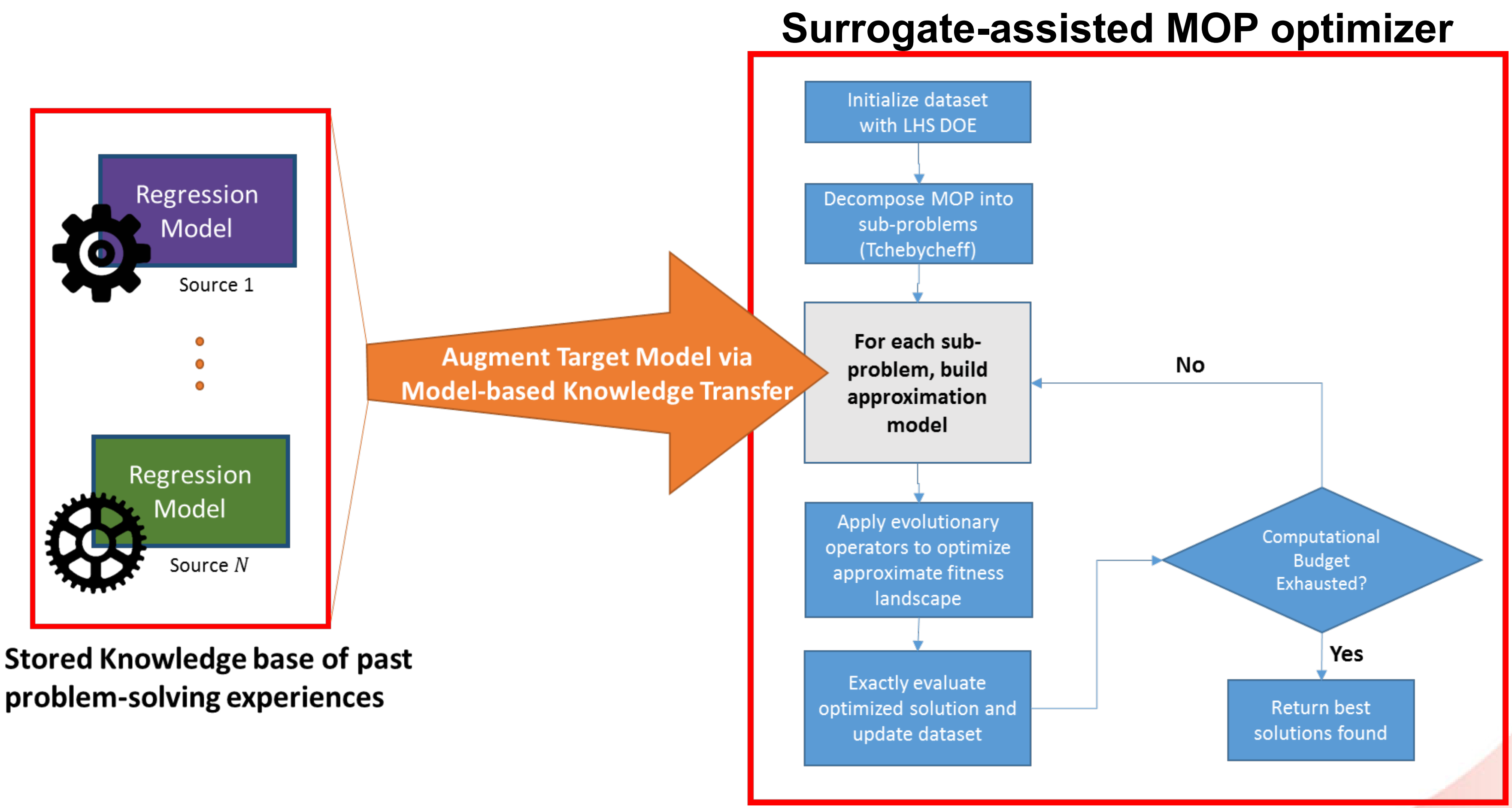


Data, Knowledge exchanges between teams

Improved Designs/Solutions

NANYANG TECHNOLOGICAL UNIVERSITY

Transferring Knowledge ‘Across MOPs’, Multi-Problem Surrogates (TEMO-MPS)



NANYANG TECHNOLOGICAL UNIVERSITY

W. M. Tan, Y. S. Ong, A. Gupta and C. K. Goh, "Multi-Problem Surrogates: Transfer Evolutionary Multiobjective Optimization of Computationally Expensive Problems", IEEE Transactions on Evolutionary Computation, 2018.

Learning the Transfer Stacking 'Multi-Problem Surrogates'

Step 1: Learning the Transfer Stacking coefficients a

Prediction of Source Model j at $x^{(i)}$

Ground Truth

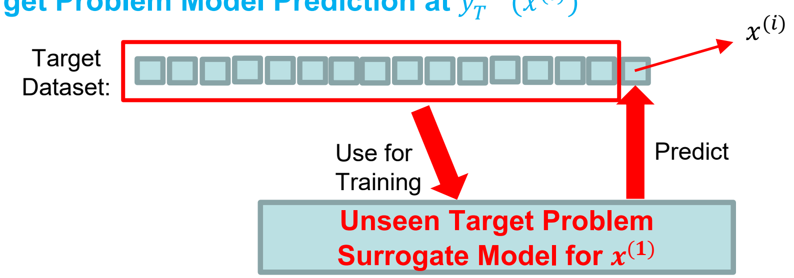
$$\text{Minimize:} \quad SE(\boldsymbol{a}) = \sum_{i=1}^{n_T} \Big(\sum_{j=1}^{B} a_{S,j} \hat{y}_{S,j}^{(i)} + a_T \hat{y}_T^{(i)} - y^{(i)} \Big)^2$$

$$\text{Subject to:} \quad \sum_{j=1}^{B} a_{S,j} + a_T = 1$$

$$a_{S,j} \geq 0, \text{for } j = 1, \ldots, B$$

$$a_T \geq 0$$

Target Problem Model Prediction at $\hat{y}_T^{(i)}\ (x^{(i)})$

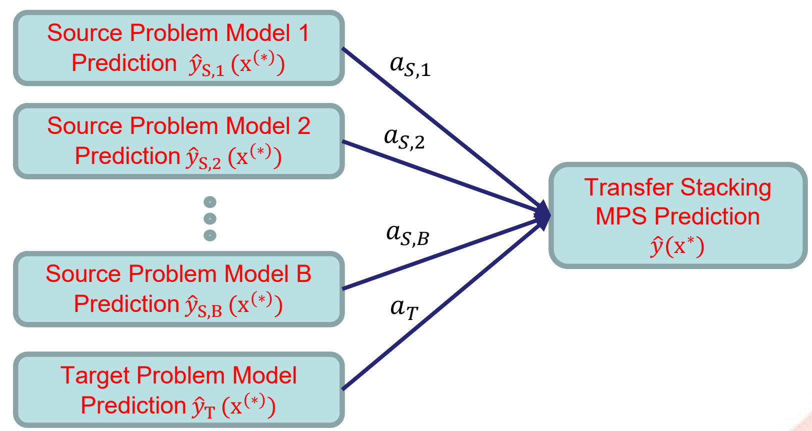


Step 2: Making Predictions at unknown point $x^{(*)}$

$$\hat{y}(\boldsymbol{x}^{(*)}) = \sum_{j=1}^{B} a_{S,j} \hat{y}_{S,j}(\boldsymbol{x}^{(*)}) + a_T \hat{y}_T(\boldsymbol{x}^{(*)})$$

$$\hat{\sigma}^2(\boldsymbol{x}^{(*)}) = \sum_{j=1}^{B} a_{S,j}^2 \hat{\sigma}_{S,j}^2(\boldsymbol{x}^{(*)}) + a_T^2 \hat{\sigma}_T^2(\boldsymbol{x}^{(*)})$$

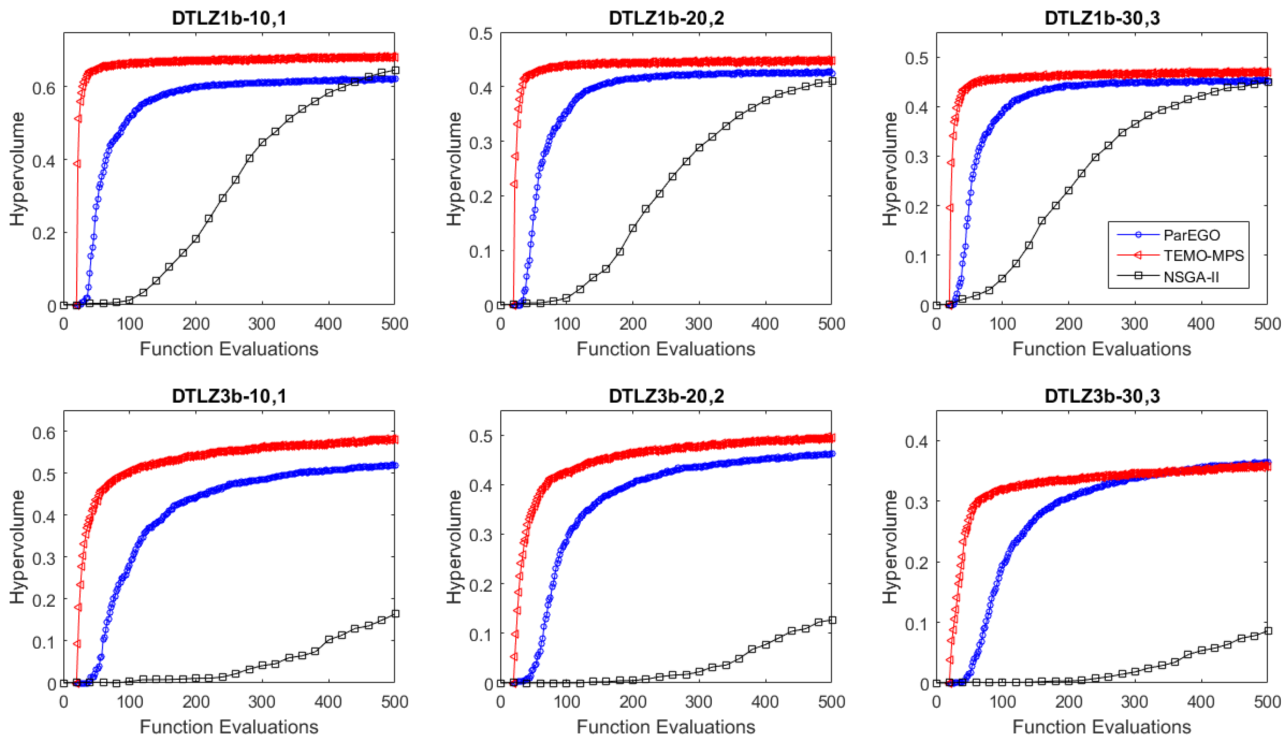
Illustration



NANYANG TECHNOLOGICAL UNIVERSITY

Multi-Problem Surrogates: Knowledge Transfer Across Expensive MOPs

Performance comparison of the proposed Transfer Evolutionary Multi-objective Optimizer with Multi-Problem Surrogates (TEMO-MPS)

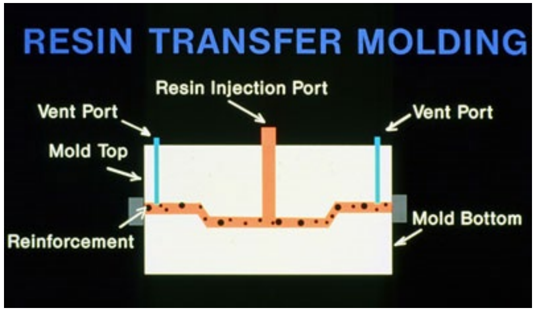


NANYANG TECHNOLOGICAL UNIVERSITY

W. M. Tan, Y. S. Ong, A. Gupta and C. K. Goh, "Multi-Problem Surrogates: Transfer Evolutionary Multiobjective Optimization of Computationally Expensive Problems", IEEE Transactions on Evolutionary Computation, 2018.

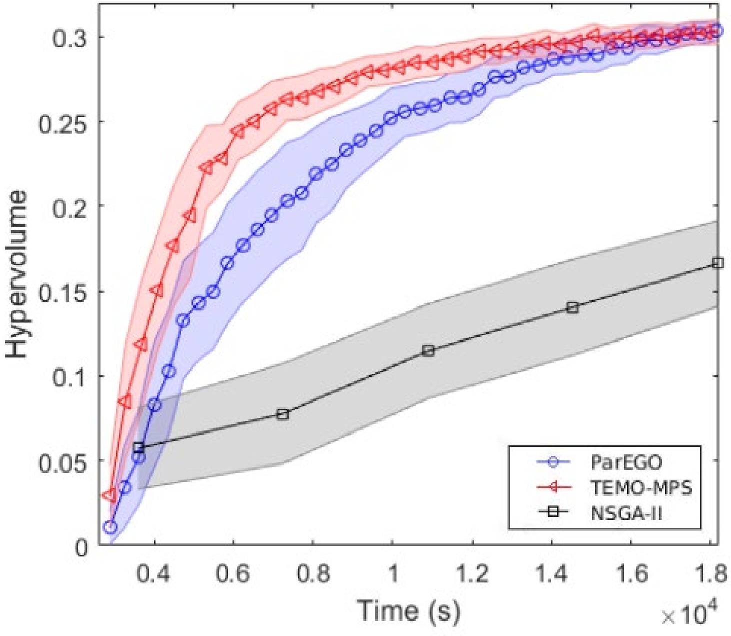
Engineering Design Application: Composite Materials Manufacturing

Case study: Simulation-based manufacturing process optimization of glass-fibre + epoxy composite parts of similar shape but different size and material configuration.



➔ **Part 1 (Source task)**: disc of 0.8 m dia. with 50% fibre volume fraction

➔ **Part 2 (Target task)**: disc of 1 m dia. with 35% fibre volume fraction



- In order to achieve a target HV measure of 0.25 TEMO-MPS takes 6700 seconds in comparison to 9950 seconds for ParEGO.

Conclusions

We described approaches towards smarter search, enabling 'General Optimization Intelligence' by learning from and exploiting related problem-solving experiences.

Data-driven Optimization:

- Modern Memetic Computation: Knowledge incorporation in search without human intervention!
- Memetics in Expensive Domains: Transfer Bayesian Optimization

In summary, it is proposed that memes (occurring as search distribution models / regression models / in any other computational representation) be perceived as entities capturing some form of problem-solving knowledge that can be directly learned from data and transmitted across problems. In turn, it becomes possible for future optimization exercises to harness the acquired memes to tailor custom search behaviours on the fly!

IEEE TRANSACTIONS ON

EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE

A PUBLICATION OF THE IEEE COMPUTATIONAL INTELLIGENCE SOCIETY

IEEE Computational Intelligence Society

Yew-Soon Ong
Founding Editor-in-Chief
School of Computer Science and Engineering
Nanyang Technological University, Singapore

The IEEE Transactions on Emerging Topics in Computational Intelligence (TETCI) publishes original articles on emerging aspects of computational intelligence, including theory, applications, and surveys.

- **TETCI is an electronics only publication. Publishes 6 issues per year**
- **Submissions have increased from 278 in 2017 to 327 in 2018.**
- **2018 submissions showcase diverse authorship from 38 different countries: with highest publications from China, US, and UK.**
- **Published papers in new topics spanning CI for e-Governance, privacy in fog computing, medical computing, neuro-chips for CI, etc.**

Four Types of Contributions - IEEE two-column style

- **Survey papers** (max 15 pages)*
- **Full papers** (max 10 pages)*
- **Short papers** (max 6 pages)*
- **Letters** (Comments on Published Papers, max 3 pages)*

***charges applies for additional pages.**

NANYANG TECHNOLOGICAL UNIVERSITY

http://cis.ieee.org/ieee-transactions-on-emerging-topics-in-computational-intelligence.html frontiers in Neuroscience IF (3.877) frontiers in Neurorobotics IF (2.606) frontiers in Computational Neuroscience IF (2.703)

Call for Paper

Special Issue on Cognitive Multitasking– Towards Augmented Intelligence

Topics of primary interest are centered on cognitive multitasking, including but not being limited to:

- Cognitive abilities in multitask learning
- Theoretical studies that enhance our understandings on the behaviors of multitasking
- Individual learning and social learning inspired memetic computation
- Memetic automaton, cognitive and brain inspired agent based multitasking algorithms
- Multitask Optimization for expensive and complex real-world problems
- Evolutionary multitasking algorithm design
- Multitask learning in image classification, natural language processing, speech recognition, etc.
- Theoretical study of task similarity towards enhanced multitasking performance
- Deep learning for multitasking
- Transfer learning in multitasking.
- Collaborative robotic systems, autonomous unmanned systems
- Brain-inspired mechanism for multitask learning

Submission link:
https://www.frontiersin.org/research-topics/9389/cognitive-multitasking---towards-augmented-intelligence#research-topic-articles

Important Dates:

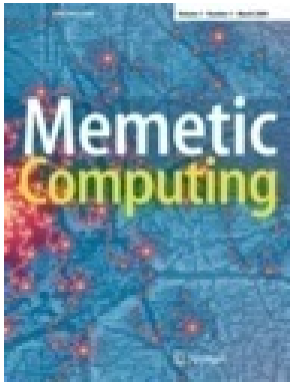
30 April 2019 Abstract
31 Auguest 2019 Manuscript
1 Dec 2019 Decision to authors

Note: The authors can choose only 1 journal when they submit their manuscripts. That is, Frontiers in Neuroscience, Frontiers in Neurorobotics **OR** Frontiers in Computational Neuroscience. This choice **cannot** be changed upon submission.

Memetic Computing Journal, Springer

SCI-Indexed

Springer
the language of science



Managing Editor: Meng-Hiot Lim
Technical Editors-in-Chief
Yew Soon Ong, Nanyang Technological University, Singapore

Memetic Computing is an avenue for the latest results in natural computation, artificial intelligence, machine learning, operational research and natural sciences, which are combined in novel ways so as to transcend the intrinsic limitations of a single discipline.

- Outlet for high quality research in hybrid metaheuristics for optimization, control and design in continuous and discrete optimization domains. We seek to dissolve the barriers separating metaheuristics, exact and approximation algorithms research and to bring forth a renewed impetus towards the investigation and understanding of promising new hybrid algorithmic technologies.

- Ultimately, Memetic Computing aspires to serve as a focal publication where the latest results in Natural Computation, Artificial Intelligence, Machine Learning, Operational Research and Natural Sciences (e.g. cognitive, animal and insect’s behavior, etc.) are fuzzed together in novel ways.

Reviews and short research communications are also welcomed.

Thank you！

…. Questions ?