Hybrid quantum-classical metaheuristics for automated machine learning applications

Von Dollen, D.J.

Curriculum Vitae

David Von Dollen, born in 1978 in Berkeley, California, USA, received his Bachelor of Arts in Music from William Paterson University in 2007. He worked for four years as a mathematics teacher, completing his certificate in education from the state of New Jersey, USA, before transitioning into a career in data science, working in the finance and insurance industries. In 2014, he completed his professional certification in data science from the University of Washington. In 2018, he earned his Master of Science in Computer Science from The Georgia Institute of Technology while working as a data scientist for Audi of America and the Volkswagen Group. While at the Volkswagen Group, David co-authored three patents and seven publications and worked on projects related to artificial intelligence, quantum computing, intelligent transportation, mobility, and information technology. He started his Ph.D. at Leiden University in 2019. During his Ph.D. studies, he supervised two Master's students and coordinated the Quantum Computing for Industrial Applications Group. In 2023, he traveled to Leiden, where he gave a guest lecture on quantum machine learning, and attended classes on practicing ethical science. Since 2022, David has held technology leadership and advisory positions at several startups operating in the intelligent transportation, IoT, computer vision, and mobile application spaces. He currently resides in the Portland, Oregon, USA, metro area, where he enjoys spending time with his family and friends outside of his scientific and professional work.

Acronyms

AGI

Artificial General Intelligence

AI

Artificial Intelligence

AutoML

Automated Machine Learning

BPP

Bounded Error Probabilistic Polynomial Time

BQP

Bounded Error Quantum Polynomial Time

EDQA

Evolutionary Designed Quantum Algorithms

ELBO

Evidence Lower Bound

ES

Evolution Strategies

FS

Feature Selection

GBR

Gradent Boosted Regression tree

GMIC

Generalized Mean Information Coefficient

GPR

Gaussian Process Regression

HHL

Harrow, Hassidim, Lloyd algorithm

HPO

Hyper-Parameter Optimization

LHS

Leap Hybrid Sampler

LR

Linear Regression

MDP

Markov Decision Process

MI

Mutual Information

MIC

Maximal Information Coefficient

ML

Machine Learning

MRMR

Maximum Relevancy Minimum Redundancy

NAS

Neural Architecture Search

NFL

No Free Lunch

NISQ

Noisy Intermediate Scale Quantum

NP

Nondeterministic Polynomial Time

P

Polynomial Time

PCC
: Pearson Correlation Coefficient

PQK
: Projected Quantum Kernel

PSD
: Positive Semi-Definite

QA
: Quantum Assisted

QE
: Quantum Enhanced

QEA
: Quantum Evolutionary Algorithms

QIEA
: Quantum-Inspired Evolutionary Algorithms

QMA
: Quantum Merlin-Arthur

QML
: Quantum Machine Learning

QPU
: Quantum Processing Unit

QUBO
: Quadratic Unconstrained Binary Optimization

RFE
: Recursive Feature Elimination

RFF
: Random Fourier Features

RL
: Reinforcement Learning

SA
: Simulated Annealing

SD

Steepest Descent

TT

Tensor train

VQC

Variational Quantum Circuit

Symbols

$\mathcal{A}$ learning algorithm (AutoML process in preliminaries chapter)

$\mathbf{a}$ chromosome/individual

a activation function

$\hat{\alpha}$ acquisition function parameter

α fitness weight in QUBO

$\boldsymbol{\alpha}$ GP weight vector

$\mathbf{a}^+$ current best point

argmax argmax operator

argmin argmin operator

$\mathbf{a}^*$ best solution found

A_t time coefficient for driver Hamiltonian

b_d phase shift (indexed)

β diversity weight in QUBO

$\mathcal{O}$ big-O notation

$\{0, 1\}$ binary set

$B(n)$ maximum grid size parameter

b phase shift

$\langle\cdot|$ bra

$\langle\cdot|\cdot\rangle$ quantum bra-ket notation

B_t time coefficient for target Hamiltonian

$\mathbf{b}$ bias vector

CE cross-entropy loss

Cholesky Cholesky decomposition

C characteristic matrix

cov covariance function

C^* maximal characteristic matrix

d dimension of feature vectors

$\mathcal{D}_c$ classical dataset

$\textbf{distance}(\cdot,\cdot)$ distance metric

$\mathcal{D}'$ processed dataset

$\mathcal{D}_q$ quantum dataset

D number of random Fourier features

$\mathcal{D}$ dataset

E energy function

EI expected improvement

ELBO ELBO function

ℓ feature index (second)

E_1 first excited state energy

$\boldsymbol{\epsilon}$ error term

$\mathbb{E}$ expectation

E_0 ground state energy

$f(\cdot)$ function

f_Q QUBO objective function

$\mathbf{f}_*$ function at test points

$f(\mathbf{x}^+)$ current best observation

γ RBF kernel length scale parameter

$\gamma_{\mathbf{RL}}$ RL discount factor

G generations to convergence

GD genotype diversity

g generation index

GMIC generalized mean information coefficient

$g_{\mathbf{min}}$ minimum energy gap

$\mathcal{H}$ Hilbert space

Hamming$(\cdot, \cdot)$ Hamming distance

H_f target/final Hamiltonian

h learning hypothesis

H_I driver/initial Hamiltonian

hit_score hit score

$\mathbf{h}$ neural network layer

h^* optimal hypothesis

H_t time-dependent Hamiltonian

i iteration index

$\mathbf{I}$ identity matrix

$\mathbb{I}$ indicator function

$\langle \cdot | \cdot \rangle$ inner product

I^* maximum mutual information

j feature index

κ UCB exploration parameter

κ_{nn} number of nearest neighbors (MI estimation)

kernel$(\cdot, \cdot)$ kernel function

$|\cdot\rangle$ ket

k number of features in reduced space

KL KL divergence

$\mathbf{K}_m$ reduced kernel matrix

$\mathbf{K}$ kernel/Gram matrix

$\tilde{\mathbf{K}}$ reduced quantum kernel matrix

$\mathbf{K}_{X^*X}$ kernel matrix (test-train)

$\mathbf{K}_{X^*X^*}$ kernel matrix (test-test)

$\mathbf{K}_{XX}$ kernel matrix (train-train)

$\mathbf{K}_{XX}^{-1}$ kernel matrix inverse

λ number of offspring

λ_s penalty weight for QUBO size constraint

$\mathbf{L}$ Cholesky lower triangular matrix

length_score length score

m number of inducing points

MAE mean absolute error

MIC maximum information coefficient

MI mutual information

MSE mean squared error

μ number of selected parents

$\hat{\mu}$ predicted mean

$\tilde{\mu}$ mean function

n number of samples/population size

$\mathcal{N}$ Gaussian/normal distribution

$\|\cdot\|$ Euclidean norm

n_q number of qubits

N number of samples for MSE

$\hat{\mathbf{o}}$ optimal binary solution

o_i binary decision variable

ω_d frequency vector (indexed)

$\boldsymbol{\omega}$ frequency vector

$\tilde{\mathbf{o}}$ binary solution for inducing points

$\mathbf{o}$ binary decision variable vector

PCC Pearson correlation coefficient

$P_{\mathcal{D}}$ preprocessing function

Φ standard normal CDF

$\phi(\cdot)$ quantum feature map

ϕ standard normal PDF

P_{λ} offspring population

P_{μ} parent population

p_m mutation probability

$P(\boldsymbol{\omega})$ spectral density

P population

p prior distribution

PQK projected quantum kernel

ψ digamma function

$\hat{\mathbf{Q}}$ QUBO matrix for features

q qubit index

$\mathbf{Q}$ QUBO matrix

$\tilde{\mathbf{Q}}$ QUBO matrix for inducing points

ρ_q quantum state for qubit q

ρ density operator

$\mathbb{R}$ real numbers

SA subset accuracy

σ mutation strength

$\hat{\sigma}$ predicted standard deviation

σ_n^2 noise variance

s performance score

$\otimes$ tensor product

T evolution timescale

θ hyperparameters (rotation parameters when applied to quantum circuits)

θ^* optimal hyperparameters

$\theta_{\mathbf{var}}$ variational parameters (VGPR)

t maximum generations

UCB upper confidence bound

$\mathbf{u}$ inducing point values

U unitary operator

U_{Z_j} unitary rotation for qubit j

$U_{Z_j Z_\ell}$ controlled-Z unitary

val_acc validation accuracy

$\mathbf{V}$ intermediate matrix for covariance

$\mathbf{W}$ weight matrix

$\mathbf{w}$ weights for linear regression

x_j individual feature component

$\mathbf{X}_m$ inducing point subset

$\mathbf{X}$ set of input data

$\mathbf{x}^+$ current best point

$\mathbf{X}^*$ test data matrix

$\mathbf{x}^*$ test feature vector

$\tilde{\mathbf{X}}$ reduced training data

$\mathbf{x}$ feature vector

$\hat{\mathbf{y}}$ predicted observation vector

y label/observation

y^+ current observation

$\mathbf{y}$ observation vector

z_d RFF basis function

$\mathbf{z}$ inducing points

Z_j Pauli-Z operator for qubit j

$Z_{\mathbf{norm}}$ normalization constant (GMIC)

Z Pauli-Z operator