EVOLUTIONARY COMPUTATION AND MULTIOBJECTIVE OPTIMIZATION

Gary G. Yen, *FIEEE*, *FIET, FIAPR*

gyen@okstate.edu

***Regents Professor*, Oklahoma State University**

四川大学高端外籍讲座教授

Summer Course at

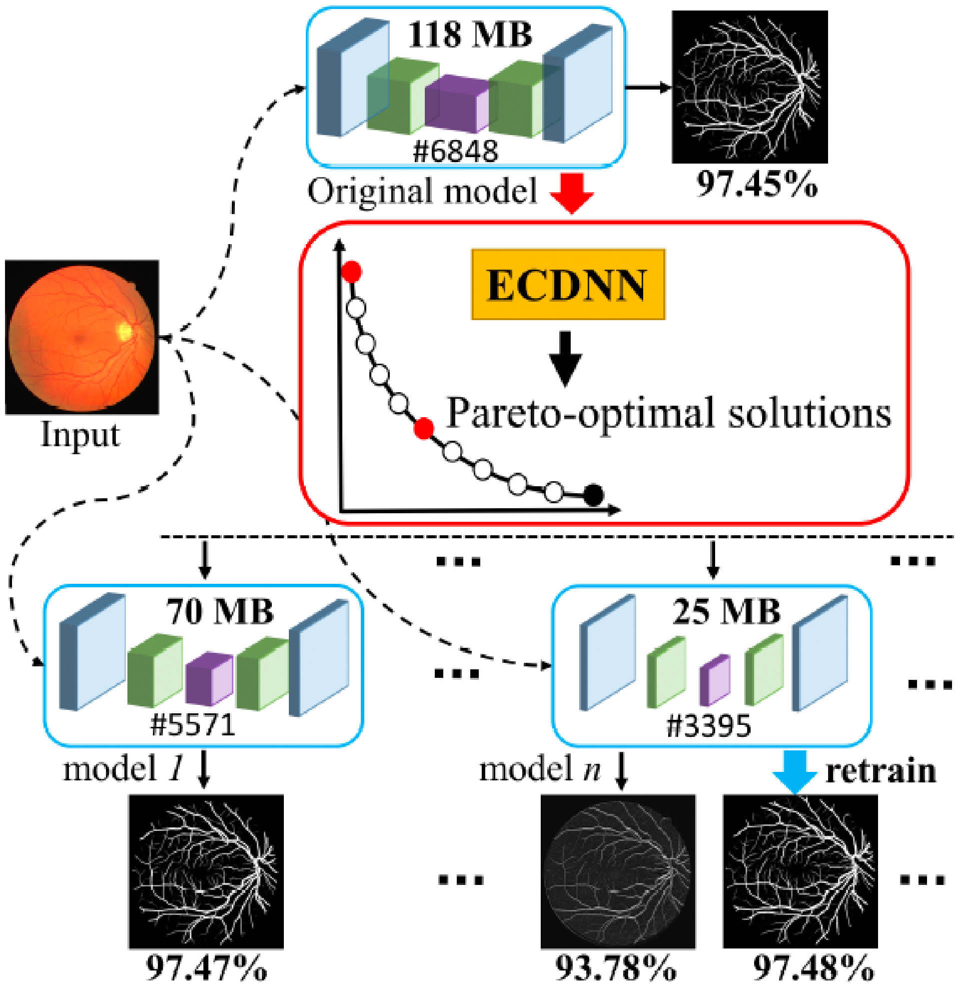
四川大学计算机学院

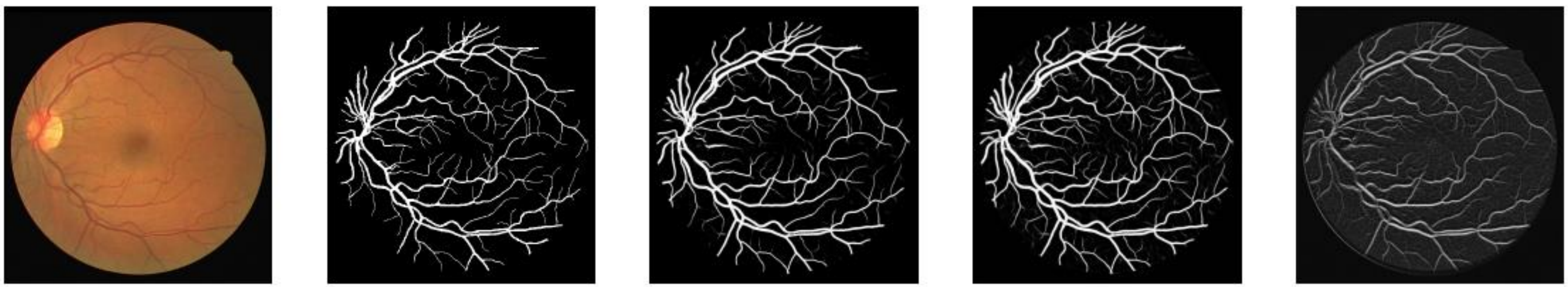
Day Two of EIGHT, June 28, 2022

Case Study 2: Evolutionary Compression of DNNs

- **Motivation:** Biomedical image segmentation is lately dominated by deep neural networks due to their surpassing expert-level performance. However, the existing DNN models for biomedical image segmentation are generally highly parameterized, which severely impede their deployment on real-time platforms and portable devices.
- **Approach:** We propose an evolutionary compression method to automatically discover efficient DNN architectures for biomedical image segmentation. Different from the existing studies, ECDNN can optimize network loss and number of parameters simultaneously during the evolution, and search for a set of Pareto-optimal solutions in a single run, which is useful for quantifying the tradeoff in satisfying different objectives, and flexible for compressing DNN when preference information is uncertain.

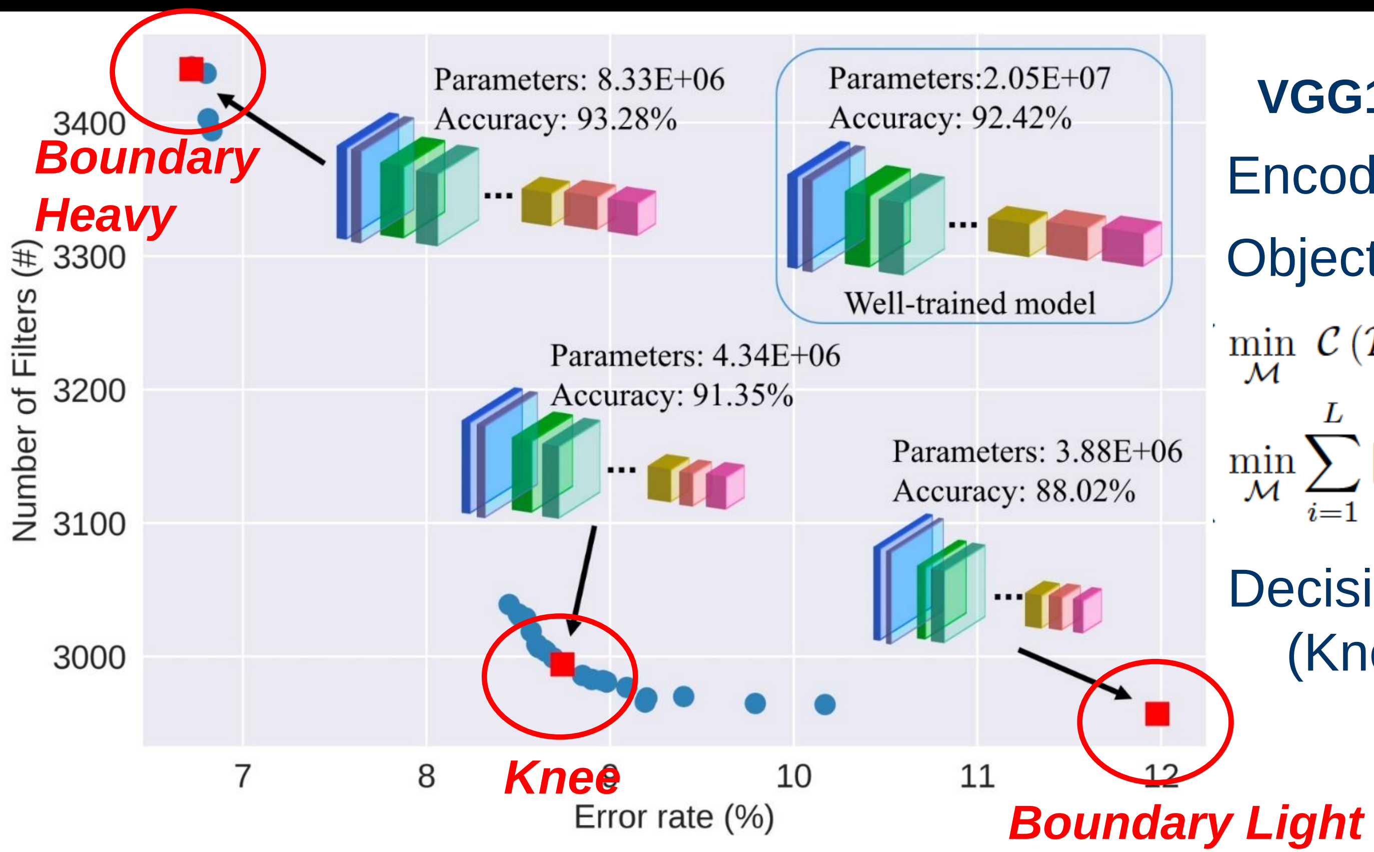
- Experiments carried out on compressing DNN for retinal vessel and neuronal membrane segmentation tasks show that ECDNN can not only improve the performance without any retraining but also discover efficient network architectures that well maintain the performance. The superiority of the proposed method is further validated by comparison with the state-of-the-art methods.





(a) (b) (c) (d) (e)

Fig. 9. Test retinal image, its manual annotation and predicted probability maps from uncompressed and compressed models. (a) is the input retinal image, and its manual annotation is shown in (b). The predicted probability map obtained from the well-trained model listed as *u-net-retinal* in Table I is shown in (c), and those obtained from boundary solution and knee solution are depicted in (d) and (e), respectively. (a) Input retinal image. (b) Manual annotation. (c) u-net-retinal. (d) Boundary solution. (e) Knee solution.



VGG19 on CIFAR10

Encoding: binary string

Objectives:

$$\min_{\mathcal{M}} \ \mathcal{C}\left(\mathcal{D}; \mathcal{M} \circ W\right)$$ Error rate

$$\min_{\mathcal{M}} \sum_{i=1}^{L} \|\mathcal{M}_i\|_1$$ Model size

Decision making (Knee selection)

Zhou, Yen&Zhang, IEEE CYB, *2021, 51(3), 1626-1638*

Zhou, Yen&Zhang, IEEE TNNLS, 2020, 31(8), 2916-2929

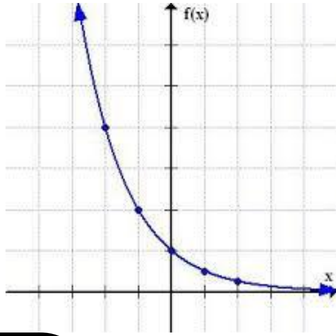
3. Simulated Annealing

模拟退火

Fundamental Concept

- Motivation by an analogy to the statistical mechanics of annealing in solids. => to coerce a solid (i.e., in a poor, unordered state) into a low energy thermodynamic equilibrium (i.e., a highly ordered defect-free state) such as a crystal lattice.
- The material is annealed by heating the material to a temperature that permits many atomic rearrangements, then cooled *slowly* until the material freezes into a good crystal- ***Metropolis procedure***
- Different from greedy algorithm, allows perturbation to move uphill in a controlled fashion to escape from local minima.

Design Metaphor



- Simulated annealing offers a physical analogy for the solution of optimization problems.
- Boltzmann distribution

$$\frac{N_{current\_state}}{N_{new\_state}} = \exp\left(\frac{E_{new\_state} - E_{current\_state}}{kT}\right)$$

N_i: number of atoms with energy

E_i, i = new_state or current_state

k is the Boltzmann constant

T is the absolute temperature

- In simulated annealing the left hand side is interpreted as a probability
- The probability of a uphill move of size ΔE at temperature T is $\Pr(accept) = e^{-\Delta E / kT}$
- If ΔE<0, the new configuration is accepted and

  if ΔE>0, probability of accepting the worse configuration is calculated based on Boltzmann distribution

- At higher temperatures, the probability of large uphill move in energy is large (permits an aggressive, essentially random search of the configuration space)
- At lower temperatures the probability is small (few uphill moves are allowed)
- By successfully lowering the temperature (cooling schedule) and running the algorithm, we can simulate the material coming into equilibrium at each newly reduced temperature.

Cooling Schedule

- A starting hot temperature and rule to determine when the temperature should be lowered and how much the temperature should be lowered and when annealing should be terminated.
- $T = \alpha T, \quad \alpha < 1$
- If α is very small, the temperature reduces very fast and there is high possibility of being trapped in a local minimum
- If α is large the energy decreases very slowly
- Many schemes to reduce temperature
  - Fast Cauchy: $T_k = T_0 \dfrac{1}{k}$
  - Geometric: $T_k = T_0 \alpha k$
  - Boltzmann: $T_k = T_0 \dfrac{1}{\ln(k)}$

Where T_0 is the initial temperature, and T_k is the temperature after the k'th temperature decrement

Pseudo Codes

- M = number of moves (perturbations) to attempt
- T = current temperature
- Generate initial move
- For m = 1 to M

  Generate NEXT move that is *related to the previous move*

  Evaluate the change in energy ΔE

  IF ($\Delta E < 0$)

    Accept this move and update configuration /* downhill move

  ELSE

    Accept with probability, $P = e^{-\Delta E/T}$

    Update configuration if accepted

  ENDIF

  Update temperature $T = \alpha T$
- ENDFOR

Key Design Components

- M = *number of moves (perturbations) to attempt* (or stopping criteria)
- T = *current temperature*
- *Generate initial solution*
- FOR m = 1 to M

    Generate next solution that is *related to the previous move*

    Evaluate the change in energy ΔE

    IF ($\Delta E < 0$)

        Accept this move and update configuration /* downhill move

    ELSE

        Accept with probability, $P = e^{-\Delta E/T}$

        Update configuration if accepted

    ENDIF

    Update temperature $T = \alpha T$

- ENDFOR

Simulated Annealing for TSP

- Initial configuration: permutation =>1, 2, 3, 4, …, N

  temperature $T = 2$

  cooling rate $\alpha = 0.99$

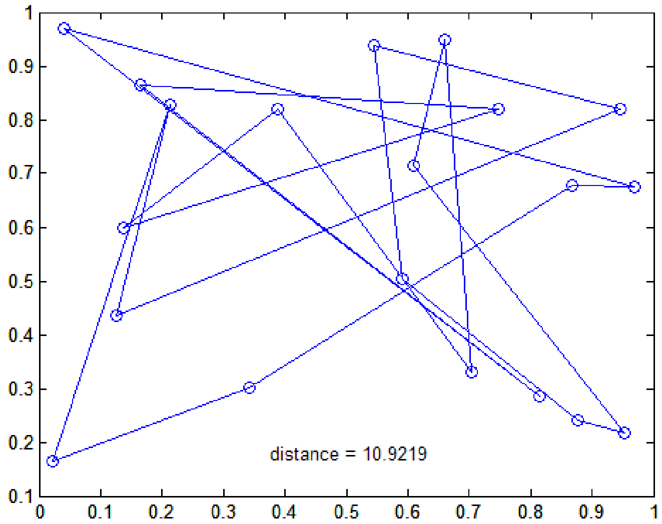
  energy = $d(1,2)+d(2,3)+\ldots+d(N,1)$

- Generate a new configuration from the current one at random
- Evaluate ΔE = current energy – previous energy
- If $\Delta E < 0$ accept the current configuration (downhill)

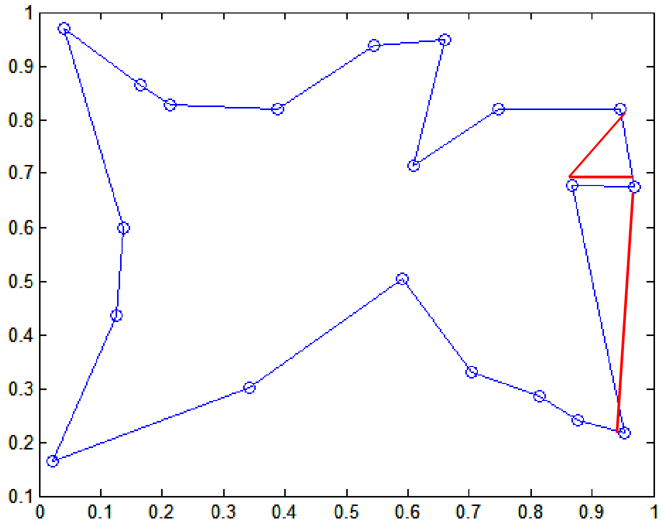
  Else accept configuration with probability $P = e^{-\Delta E / kT}$

- $T = \alpha T$
- Stopping criteria if energy < threshold or number of iterations is reached.

Case Study: 20-city problem

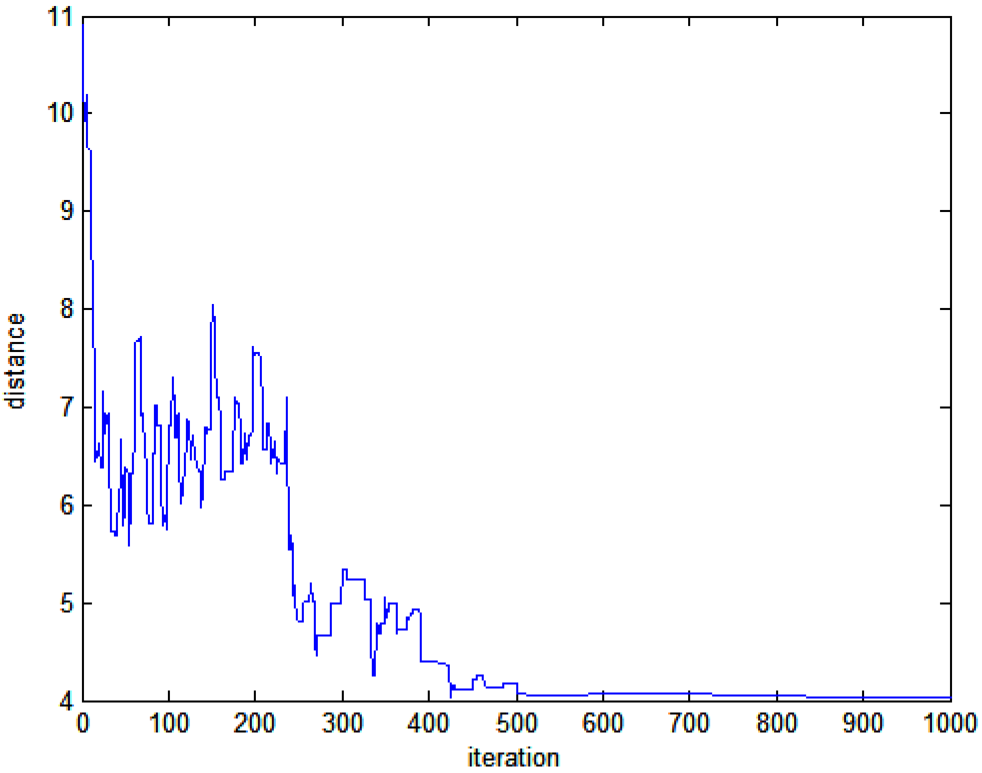


Initial random permutation



Best route found for 20-city problem
d=4.0185 (d=4.0192)

Initial temperature	2	2	2	2	2
Cooling rate	0.8	0.9	0.95	0.99	0.999
Converged iteration	87	155	298	851	5982
Minimal energy obtained	4.1056	4.0920	4.0427	4.0185	4.0185



T_0= 2
α=0.99
Iteration = 851
d = 4.0185

Case Study: 101-city problem (ali101)

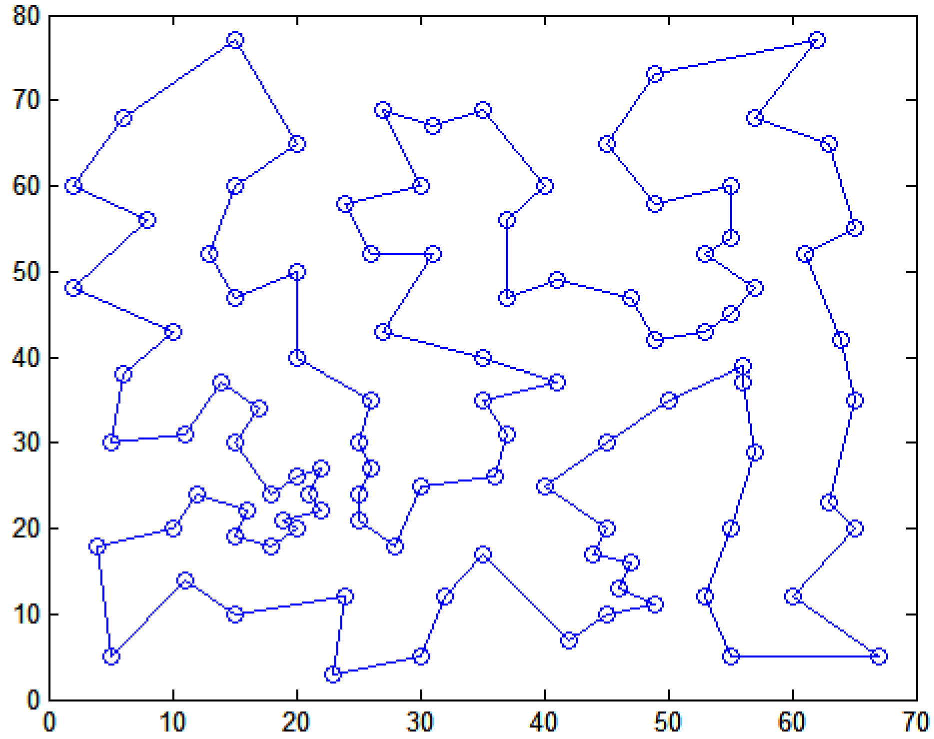
T_0=200

α= 0.999

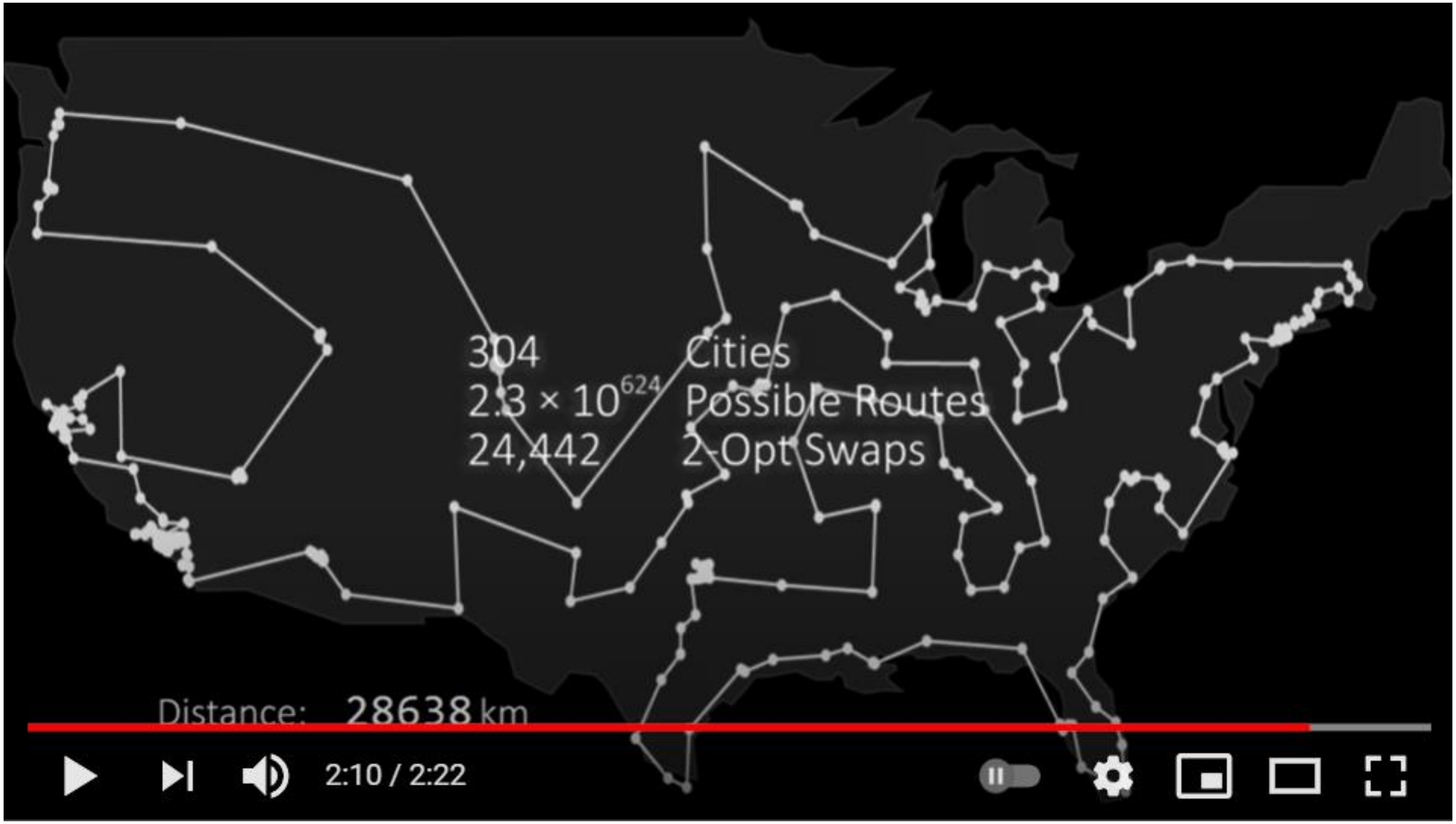
Iterations = 100,000

d = 661.25

Global minima = 629



TSP Demo

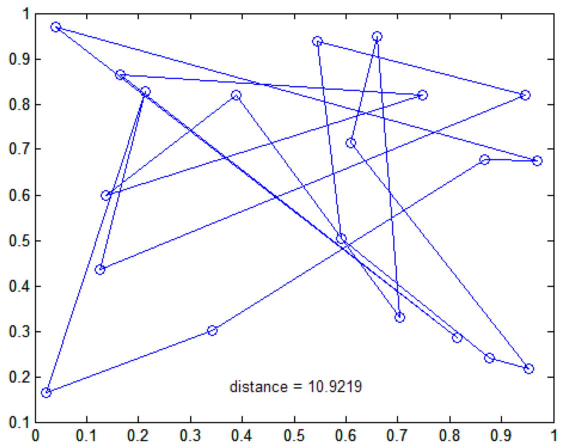

304 Cities
2.3 × 10^624 Possible Routes
24,442 2-Opt Swaps
Distance: 28638 km
2:10 / 2:22

Research Issues

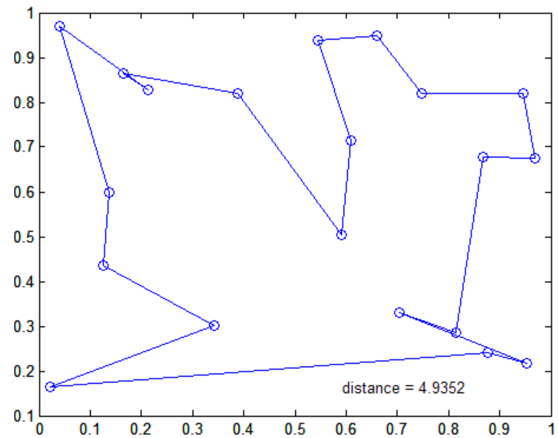
- Initial Temperature
- Initial Configuration
- Next Move (neighborhood size)
  - is it possible to design a truly local search for combinatorial optimization problem (as oppose to numerical optimization problem)?
- Cooling Schedule (cooling rate)
- Stopping Criteria
- Acceptance Probability

Effect on Initialization

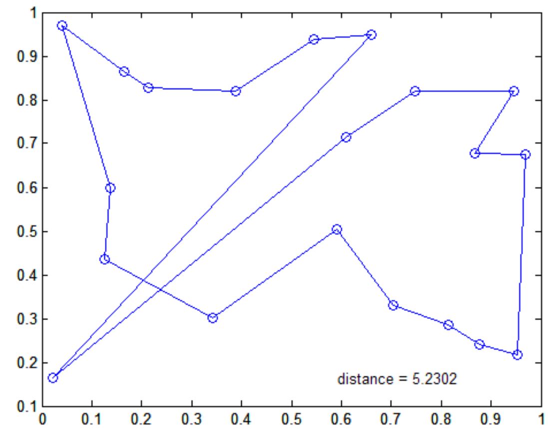
Random sequence



Hilbert space filling curve



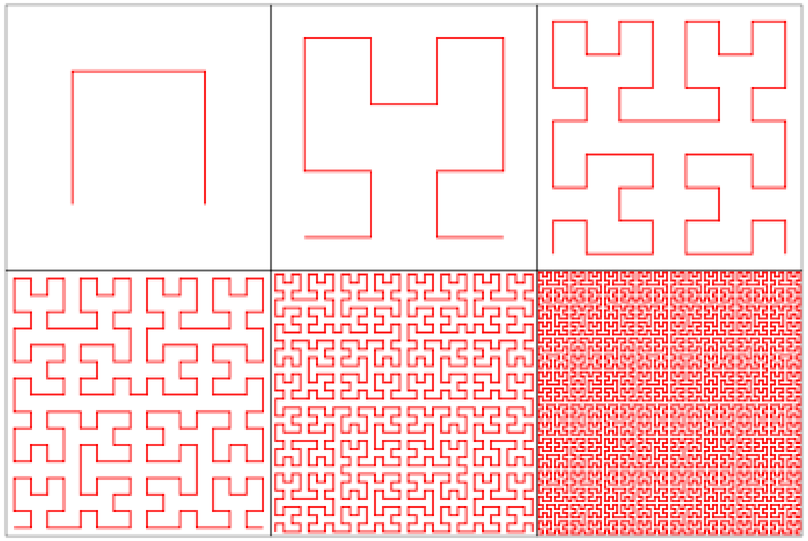
Fastest closest path

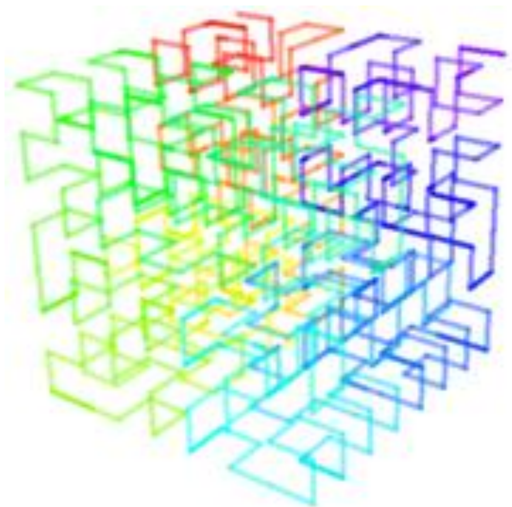


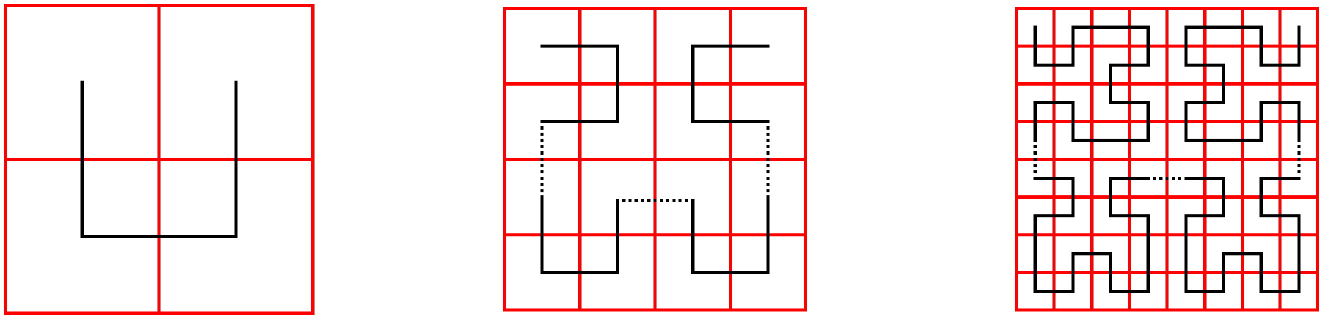
Experimental results indicate that the initialization method does not affect much the probability of the system to converge to the global optimal point.

Space Filling Curve

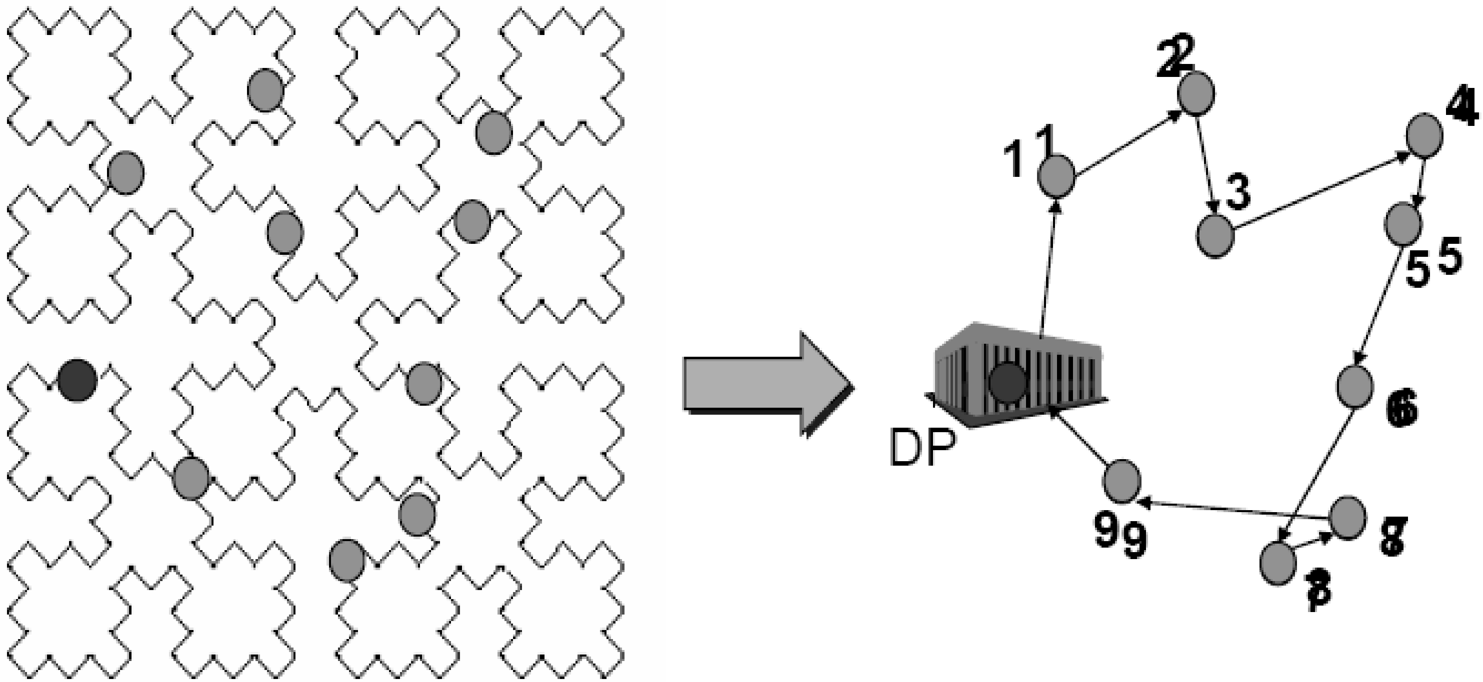
- A ***space filling curve*** is a continuous mapping from a lower dimensional space into a higher one. A spacefilling curve is formed by repeatedly copying and shrinking a simple pattern.
- A ***Hilbert curve*** (also known as a Hilbert space-filling curve) is a continuous fractal space-filling curve first described by the German mathematician David Hilbert in 1891.







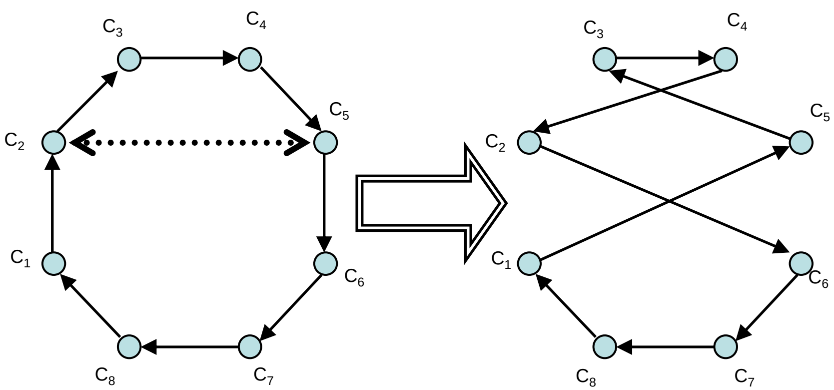
- Divide the square into four smaller squares $\mathbf{I}_{00}$, $\mathbf{I}_{01}$, $\mathbf{I}_{10}$, and $\mathbf{I}_{11}$
- Define $\mathbf{f}_0$ so that it maps [0,1/4] into $\mathbf{I}_{00}$, [1/4,1/2] into $\mathbf{I}_{01}$ and so on…



1
2
3
4
5
6
9
DP

Effect on Next Move Modification

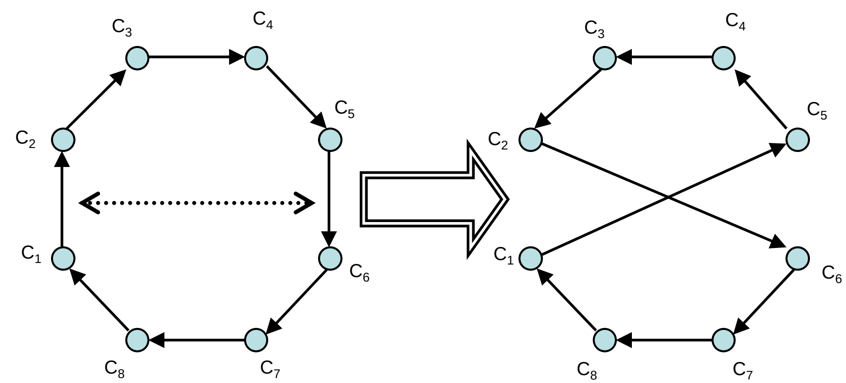
Random switch of two vertices



1 2 3 4 5 6 7 8

1 5 3 4 2 6 7 8

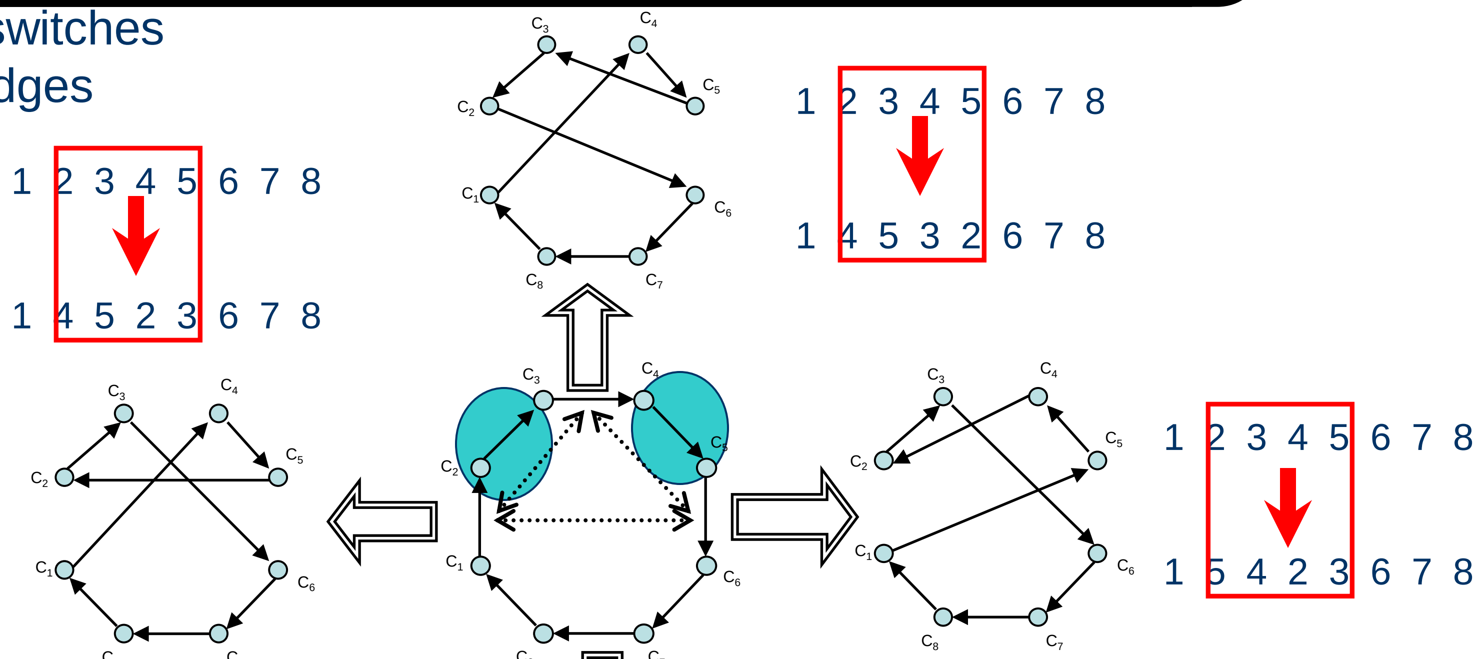
Random switch of two edges

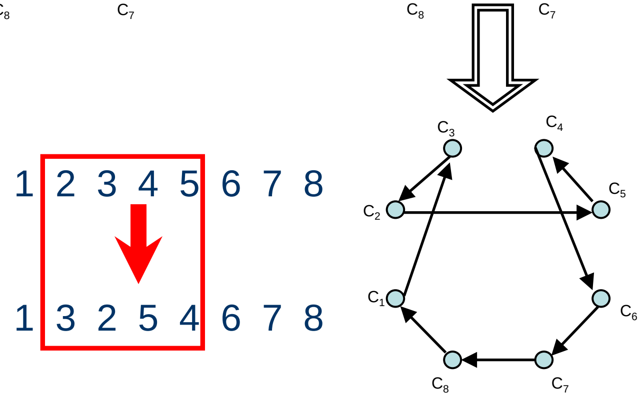


1 2 3 4 5 6 7 8

1 5 4 3 2 6 7 8

Random switches
of three edges

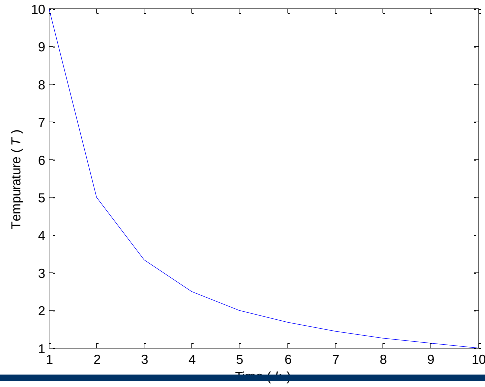




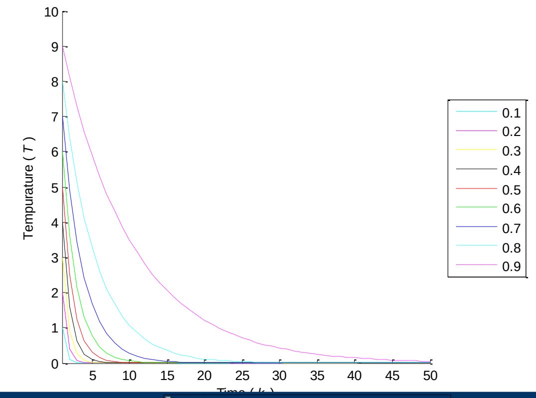
Experimental results show that the three edge change is much more powerful in making the method converge to the global minimum than the other methods. However, its processing cost is also higher, as one would expect.

Effect on Cooling Schedule

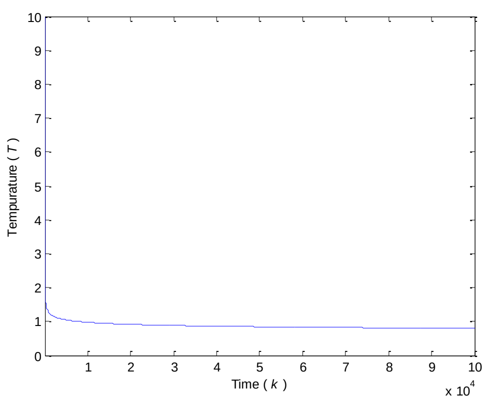
Fast Cauchy



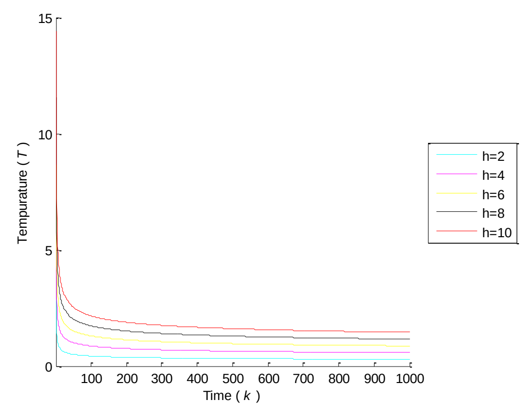
Geometric



Boltzmann



Logarithmic



Applications

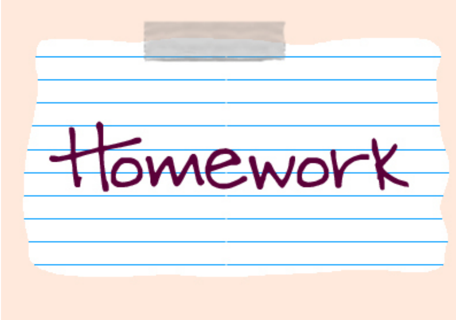
- Computer-aided VLSI design
  - Simulated annealing for VLSI design, Kluwer Academic, 1988
- Combinatorial optimization
  - Simulated annealing: theory and application, Reidel Publishing, 1987
- Neural network training
  - A learning algorithm for Botzman machine, Cognitive Science, 9: 147-169, 1985
- Image processing
  - Image processing by simulated annealing, IBM Journal of Research and Development, 29: 569-579, 1985
- Code design
  - Using simulated annealing to design good codes, IEEE Trans Information Theory, 33: 116-123, 1987
- Function optimization
  - An empirical study of bit vector function optimization, in Genetic Algorithm and Simulated Annealing, Chapter 13, 170-204, 1987
- and etc

Critical Message Conveyed

Simulated Annealing, a thermodynamic inspired optimization algorithm, encourages greedy search,

but allows controlled up-hill moves to jump out of local minima.

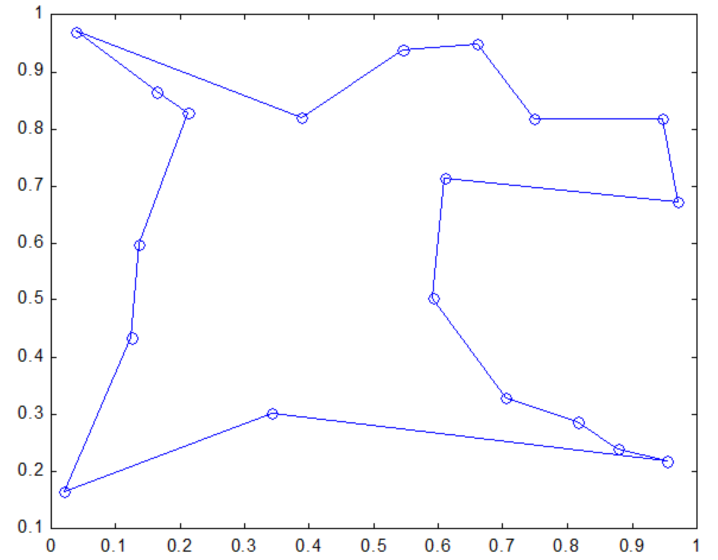
Homework #3



Problem 1: Develop a generic simulated annealing algorithm to solve the traveling salesman problem with 20 cities that are uniformly distributed within a unit square in a 2-dimensional plane. The coordinates of 20 cities are given below in a 2×20 matrix:

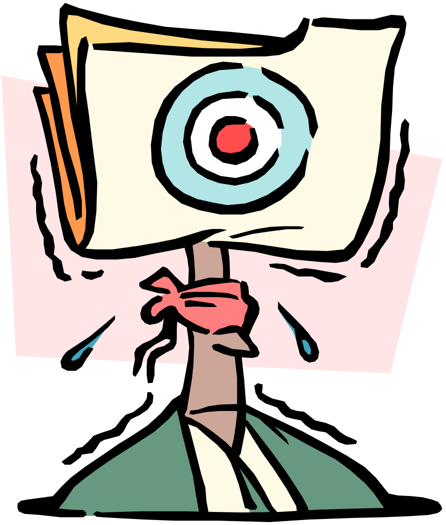
$$\text{cities} = \begin{bmatrix} 0.6606,0.9695,0.5906,0.2124,0.0398,0.1367,0.9536,0.6091,0.8767,0.8148 \\ 0.9500,0.6740,0.5029,0.8274,0.9697,0.5979,0.2184,0.7148,0.2395,0.2867 \\ 0.3876,0.7041,0.0213,0.3429,0.7471,0.5449,0.9464,0.1247,0.1636,0.8668 \\ 0.8200,0.3296,0.1649,0.3025,0.8192,0.9392,0.8191,0.4351,0.8646,0.6768 \end{bmatrix}.$$

Show the “best” route you find and the associated distance with attached computer coding (with *documentation*). An example is given below for reference.

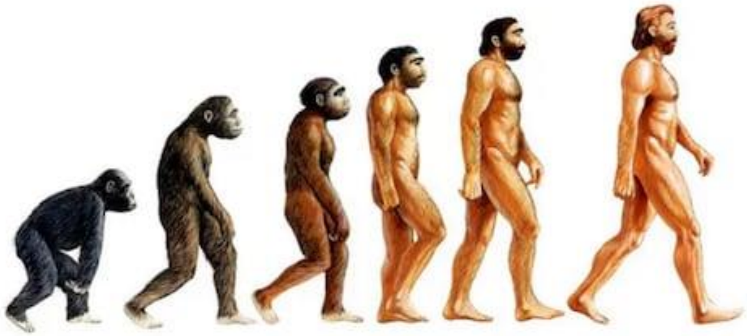


- Due: next Monday on 7/4/22

Q&A



4. EVOLUTION BIOLOGY
进化生物学



Biology

- *Biology* is a natural science concerned with the study of **life** and **living organisms**, including their structure, function, growth, origin, evolution, distribution, and taxonomy. Biology is a vast subject containing many subdivisions, topics, and disciplines. Among the most important topics are five unifying principles that can be said to be the fundamental axioms of modern biology:

  Cells are the basic unit of life

  New species and inherited traits are the product of *evolution*

  Genes are the basic unit of heredity

  An organism *regulates* its internal environment to maintain a stable and constant condition

  Living organisms consume and transform *energy*.

Evolution Biology

- Evolution is a cornerstone of modern biology. It unites all the fields of biology under one theoretical umbrella.
- ***Evolution*** is a change in the gene pool (i.e., the set of all genes) of a population over time.
- A gene is a hereditary unit that can be passed on unaltered for many generations. The *gene pool* is the set of all genes in a species or population.
- *ENGLISH MOTH* (2% in dark prior to 1848 grows to 95% in 1898)
- Populations evolve. In order to understand evolution, it is necessary to view populations as a collection of individuals, each harboring a different set of traits. A single organism is never typical of an entire population unless there is no variation within that population. Individual organisms do not evolve, they retain the same genes throughout their life. When a population is evolving, the ratio of different genetic types is changing, yet each individual organism within a population does not change.
- The process of evolution can be summarized in three sentences: *gene mutate*, *individual are selected*, and *population evolve*.

Common Misconceptions

- Microevolution vs. Macroevolution (different time scales)
- Phenotypic changes induced solely by changes in environment do not count as evolution because they are not inheritable. E.g., human are larger now than in past
  - *An organism's phenotype is determined by its genes and its environment.*
  - Most changes due to environment are fairly subtle.
  - Large scale phenotypic changes are obviously due to genetic changes, and therefore are evolution.
- Evolution is not progress. Populations simply adapt to their current surroundings. They do not necessarily become better in any absolute sense over time. A trait or strategy that is successful at one time may be unsuccessful at another.
- Species do not simply change to fit their environment; they modify their environment to suit them as well.

Genetic Variations

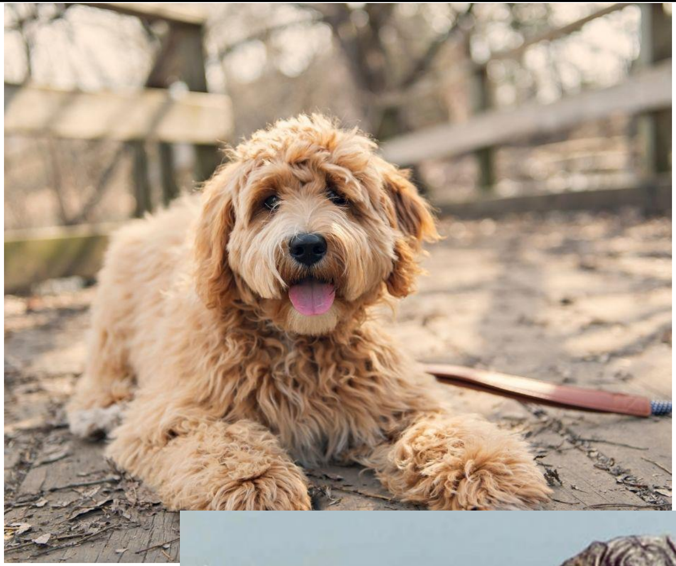
- Evolution requires genetic variation. English moth
- Genetic Variation
  - In order for continuing evolution there must be mechanisms to increase or create genetic variation (i.e., mutation, recombination and gene flow) and mechanism to decrease it (natural selection, sexual selection and genetic drift).
- Natural Selection
  - Some types of organisms within a population leave more offspring than others. Over time, the frequency of the more prolific type will increase. The difference in reproductive capability is called natural selection. It is defined as differential reproduction success of pre-existing classes of genetic variants in the gene pool.
  - The most common action of natural selection is to remove unfit variants as they arise via mutation. Natural selection usually prevents new alleles (alternate version of a gene) from increasing in frequency.
  - Natural selection can maintain or deplete genetic variation.
  - Organisms do not perform any behaviors that are for the good of their species. An individual organism competes primarily with others of its own species for its reproductive success

- Sexual Selection
  - Sexual selection is natural selection operating on factors that contribute to an organism’s mating success. Traits that are a liability to survival can evolve when the sexual attractiveness of a trait outweighs the liability incurred for survival.
  - A male who lives a short time but produces many offspring is much more successful than a long-lived one that produces few. The former’s genes will eventually dominate the gene pool of his species.
- Genetic Drift
  - Allele ***frequencies*** can change due to chance alone. Drift is a binomial sampling error of the gene pool.
  - The allele that form the next generation’s gene pool are a sample of the alleles from the current generation. When sampled from a population, the frequency of alleles differs slightly due to chance alone.

- Example of Genetic Drift

  - Of the two pink monkeys in the world – one male, one female – the female dies, ensuring that there will never be a pure-bred pink monkey again.
  - A random succession of births results in all other hair colors going extinct within a village full of redheaded people.
  - The last green-eyed person in a small town dies, leaving only brown-eyed and blue-eyed people.
  - A man steps on a group of beetles, randomly killing most of the green ones but leaving most of the brown ones alive, resulting in fewer green beetles being produced in the population.
  - A wildflower population consisting of blue, purple, and pink flowers is subjected to a mudslide that kills most of the blue ones. As time progresses, blue flowers eventually die out, leaving only purple and pink wildflowers.
  - Due to random successions of births, a town has an unusually high population of people with strawberry blonde hair, a trait that increases over time and leaves very few people with different hair colors.

- <u>Mutation</u>
  - The cellular machinery that copy DNA sometimes makes mistakes. These mistakes alter the sequence of a gene. A point mutation is a mutation in which one “letter” of the genetic code is changed to another. Lengths of DNA can also be deleted or inserted in a gene. Finally genes or parts of genes can become inverted or duplicated. Typical rates of mutation are between 10^{-10} and 10^{-12} mutations per base pair of DNA per generation.
  - Mutation creates new alleles. Each new allele enters the gene pool as a single copy amongst many. Most are lost from the gene pool, the organism carrying them fails to reproduce, or reproduces but does not pass on that particular allele.
- <u>Recombination</u>
  - Recombination can be thought of as gene shuffling. In most sexually reproducing organisms, there are two of each chromosome type in every cell (one from mother and another from father). When an organism produces gametes, the genes end up with only one of each chromosome per cell. Recombination is a mechanism of evolution because it adds new alleles and combinations of alleles to the gene pool.
- <u>Gene Flow</u>
  - New organisms may enter a population by migration from another population. If they mate within the population, they can bring new alleles to the local gene pool.
  - Gene flow between more distantly related species occurs infrequently.




MINSK-NOVOSTI

Darwinian’s Evolution

- 4 postulates
  - Individuals within species are variables
  - Some of the variations are passed onto offsprings
  - In every generation, more offspring are produced than can survive
  - The survival and reproduction of individuals are not random: the individuals who survive and go on to reproduce or who reproduce the most, are those with the most favorable variations. They are “naturally selected.”

    (new variations arise continually within populations. A small percentage of these variants cause their bearers to produce more offspring than others. These variants thrive and supplant their less productive competitors.)

Natural Selection

- Natural evolution acts…
  - On individuals, but the consequences occur in the population
  - On individuals, not groups
  - On phenotypes, but evolution consist of changes in the genotype.
  - On existing traits, but can produce new traits.
- Evolution…
  - Is backward looking
  - Is not perfect
  - Is nonrandom
  - Is not progressive

- Results of Evolution are…
  - Creative, surprising, unexpected
  - Highly adapted to environmental niches
  - Unsupervised (no conscious design, no knowledge involved)
- Natural evolution had an extremely long time (3.7B years)
- Natural evolution acts in parallel

Terms from Genetics

- DNA (DeoxyribonucleicAcid): very large linear self-replicating molecules found in all living cells, the physical carrier of genetic information
- Chromosome: a single, very long molecule of DNA
- Gene: the base unit of inheritance, a length of DNA which exerts its influence on an organisms form and function by encoding and directing the synthesis of a protein
- Allele: one of a number of alternative forms of a gene that can occupy a given genetic locus on a chromosome
- Locus: location of a gene on a chromosome

- Evolution computation…
  - Is based on biological metaphors
  - Has great practical potentials
  - Is getting popular in many fields
  - Yield powerful, diverse applications
  - Give high performance against low costs
  - And its fun.
- Evolution Algorithms are inspired by natural evolution with four key elements:
  - Group of individuals- population
  - Reproductive fitness- fitness
  - Survival of fittest- selection
  - Source of variations- genetic operators

EC Variants and Their Origins

- **<u>Genetic Algorithm (GA)</u>** by John Holland in 1962 for numerical optimization
  - Simulate Darwinian evolution
  - Essential recombination (crossover) operation
  - Mainly binary representation
  - Schemata theory
- **<u>Evolutionary Programming (EP)</u>** by Larry Fogel for simulated intelligence in 1962; finite state machine representation
  - Close to Lamarckian representation
  - No recombination
  - Adaptive mutation
  - Apply to phenotypes directly
- **<u>Evolutionary Strategy (ES)</u>** by I. Rochenberg and H.P. Schwefel for numerical optimization in 1965
  - Real-valued representation
  - Mutation based
  - Adaptive mutation

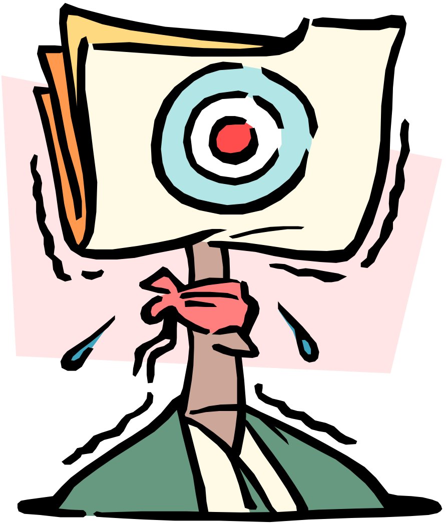
- **Genetic Programming (GP)** by John Koza in 1989
  - Evolve LISP programs
  - Tree representation
- **Particle Swarm Optimization (PSO)** by james Kennedy and Russell Eberhart in 1995
  - Based on a metaphor of social interactions, searches a space by adjusting the trajectories of individual vectors called particle as they are conceptualized as moving points in multidimensional space. The individual particles are drawn stochastically toward the positions of their own previous best performance and the best previous performance of their neighbors.
- **Ant Colony System (ACS)** by Marco Dorigo in 1996
  - Inspired by how real ants are capable of finding the shortest path from a food source to their nest without using visual cues by exploiting pheromone information

Critical Message Conveyed

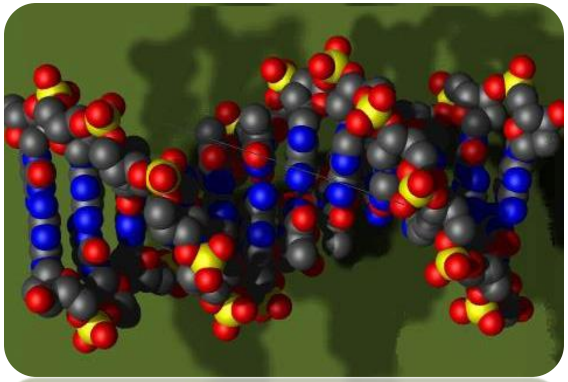
Evolution Biology

is the cornerstone of biology and provide an inspiration for modern day's Genetic Algorithms

Q&A



5. GENETIC ALGORITHMS- Part 1
遗传算法



Introduction

- Genetic algorithms are inspired by Darwin's theory of evolution (i.e., survival of the fittest);

"The new breeds of classes of living things come into existence through the process of <u>reproduction</u>, <u>crossover</u>, and <u>mutation</u> among existing organisms."

- The above concept in the theory of evolution has been translated into an algorithm to search for solutions to problems in a more 'natural' way.

History

- Genetic algorithms are originated from the studies of cellular automata, conducted by John Holland and his colleagues at the University of Michigan.
- Until the early 1980s, the research in genetic algorithms was mainly theoretical, with few real applications.
- From the early 1980s the community of genetic algorithms has experienced an abundance of applications which spread across a large range of disciplines. Each and every additional application gave a new perspective to the theory.

- In the process of improving performance as much as possible via tuning and specializing the genetic algorithm operators, new and important findings regarding the generality, robustness, and applicability of genetic algorithms became available.

- Following the last couple of years of furious development of genetic algorithms in the sciences, engineering and the business world, these algorithms in various guises have now been successfully applied to optimization problems, scheduling, data fitting and clustering, trend spotting and path finding.

Biological Background

- Chromosome: are strings of DNA (genes) and serve as a model for the whole organism
- Each gene encodes a particular protein. Basically, it can be said that each gene encodes a trait, for example color of eyes.
- Possible settings for a trait (e.g. blue, brown) are called alleles.
- Each gene has its own position in the chromosome. This position is called locus.

- Complete set of genetic material (all chromosomes) is called *genome*.
- Particular set of genes in genome is called *genotype*.
- The genotype is the after birth base for the organism's *phenotype*, its physical and mental characteristics, such as eye color, intelligence, etc.

- During reproduction, ***recombination*** (or ***crossover***) first occurs. Genes from parents combine to form a whole new chromosome.

- The newly created offspring can then be mutated. ***Mutation*** means that the elements of DNA are a bit changed. This changes are mainly caused by errors in copying genes from parents.

- The ***fitness*** of an organism is measured by success of the organism in its life (survival).

Biological Metaphor

- Algorithm begins with a set of solutions (represented by chromosomes) called population.
- Solutions from one population are evaluated for their goodness (fitness) and used to form a new population, (reproduction). This is motivated by a hope, that the new population will be better than the old one.
- Solutions which are then selected to form new solutions. (Offspring) are selected according to their fitness - the more suitable they are the more chances they have to reproduce.
- This is repeated until some condition (for example number of populations or improvement of the best solution) is satisfied.