Evolutionary Algorithm Applications

Presented By Penghuan Ni

What is Data Mining?

- Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information.
- What tasks does it involve?
  - Feature Selection
  - Classification
  - Clustering
  - Some other tasks

Multi-Objective Optimization (MOO)

- In many real-word situations, there may be several objectives that must be optimized simultaneously in order to solve a certain problem.
- It is difficult to compare one solution with another one.
- In general, these problems admit multiple solutions, each of which is considered acceptable and equivalent when the relative importance of the objectives is unknown.
- The best solution is subjective and depends on the need of the designer or decision maker.

Feature Selection

- Feature Selection problem deal with selection of an optimum relevant set of features or attributes that are necessary for the recognition process (classification or clustering).
- It helps reduce the dimensionality of the measurement space.

Feature Selection

- Example: Automobile dataset

Automobile Data Set

Download: Data Folder, Data Set Description

Abstract: From 1985 Ward's Automotive Yearbook

Data Set Characteristics:	Multivariate	**Number of Instances:**	205	**Area:**	N/A
Attribute Characteristics:	Categorical, Integer, Real	**Number of Attributes:**	$\boxed{26}$	**Date Donated**	1987-05-19
Associated Tasks:	Regression	**Missing Values?**	Yes	**Number of Web Hits:**	204078

Feature Selection

Attribute Information:

Attribute: Attribute Range

1. symboling: -3, -2, -1, 0, 1, 2, 3.
2. normalized-losses: continuous from 65 to 256.
3. make:
alfa-romero, audi, bmw, chevrolet, dodge, honda,
isuzu, jaguar, mazda, mercedes-benz, mercury,
mitsubishi, nissan, peugot, plymouth, porsche,
renault, saab, subaru, toyota, volkswagen, volvo

4. fuel-type: diesel, gas.
5. aspiration: std, turbo.
6. num-of-doors: four, two.
7. body-style: hardtop, wagon, sedan, hatchback, convertible.
8. drive-wheels: 4wd, fwd, rwd.
9. engine-location: front, rear.
10. wheel-base: continuous from 86.6 120.9.
11. length: continuous from 141.1 to 208.1.
12. width: continuous from 60.3 to 72.3.
13. height: continuous from 47.8 to 59.8.
14. curb-weight: continuous from 1488 to 4066.
15. engine-type: dohc, dohcv, l, ohc, ohcf, ohcv, rotor.
16. num-of-cylinders: eight, five, four, six, three, twelve, two.
17. engine-size: continuous from 61 to 326.
18. fuel-system: 1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi.
19. bore: continuous from 2.54 to 3.94.
20. stroke: continuous from 2.07 to 4.17.
21. compression-ratio: continuous from 7 to 23.
22. horsepower: continuous from 48 to 288.
23. peak-rpm: continuous from 4150 to 6600.
24. city-mpg: continuous from 13 to 49.
25. highway-mpg: continuous from 16 to 54.
26. price: continuous from 5118 to 45400.

Feature Selection

- The goal of feature selection is mainly threefold.
    - It is practically and computationally difficult to work with all the features if the number of features is too large.
    - Many of the given features may be noisy, redundant, and irrelevant to the classification or clustering task at hand.
    - It is a problem when the number of features becomes much larger than the number of input data points.
- Curse of Dimensionality

Feature Selection (Chromosome Representation)

- Almost all the MOEA-based feature selection algorithms have used a binary chromosome to encode a feature subset.
- The length of each chromosome is taken as d, where d is the total number of features. Each position of the chromosome can take either a 1 or 0 value.
- If position i has value 1, then the feature i has a value of part of the selected feature subset. If position i has a value of 0, then the corresponding feature is ignored.

Feature Selection (Objective Function)

- Supervised Case
  - Usually some classification algorithms are used to measure the goodness of the selected feature subset based on how well they can classify the training examples using a certain classifier.
  - Two objective functions: reducing the misclassification rate as much as possible, with a minimum set of features.

Feature Selection (Objective Function)

- Unsupervised Case (no training set, like clustering)
  - Usually, a clustering algorithm is used to evaluate a feature subset on the basis of how well these features are able to identify the clustering structure of the dataset.
  - In this regard, some cluster validity index is used to evaluate the goodness of the clustering solution generated by the feature subset.

Feature Selection (Operators)

- Single-point and uniform crossover
- Bit-flip mutation
- In most of cases, we use standard crossover and mutation operators, without having to rely on more sophisticated operators.

Classification

- There are mainly three different approaches:
  - The most commonly studied approach is the use of MOEAs for evolving a good set of classification rules.
  - The second approach is to employ MOEAs to define the class boundaries (hyperplanes) in the training data.
  - The final approach is to use MOEAs for training and model the construction of well known classifiers such as neural networks and decision tree classifiers.

Evolving Classification Rules

- A classification rule can be represented as an if-then rule of the form *If <condition> and <condition> and … Then <class>*
- If some attribute consists of continuous values, it must be first discretized to make it categorical before using the attribute in a classification rule.
- The objective of any rule-based classification system is to identify a good set of classification rules that can properly represent the training dataset, i.e., that provides a good classification performance on the training data.

Evolving Classification Rules (Encoding)

- There are mainly two approaches to encode the classification rules in the chromosomes.
  - The first one is Pittsburgh approach, in which a set of rules is encoded in a single chromosome.
  - The second one is the Michigan approach, where each chromosome encodes one rule.
- From the classification point of view, the Pittsburgh approach is more useful since here each chromosome represents a complete classifier system.

Evolving Classification Rules (Objective Functions)

- In different works, different sets of objective functions have been considered to be optimized.
- The general notion is to achieve a tradeoff between accuracy and complexity of the candidate rule set.

Evolving Class Boundaries (Hyperplanes)

- Any nonlinear surface can be approximated by using a number of hyperplanes. Therefore, the classification problem can be viewed as that of searching for a number of linear surfaces that can appropriately model the class boundaries.
- Binary chromosomes of variable length are used to encode the parameters of a variable number of hyperplanes.
- Three objectives are: to minimize the number of misclassified patterns and the number of hyperplanes and to maximize the classification accuracy.

Model Building of Standard Classifiers

- Use MOEAs for model building or training of standard classifiers such as, ANN, SVM, and decision trees.
- Most of those model building of standard classifiers adopt binary encoding.
- Different objective functions have been used in different studies.
  - SVM: minimize the false positive rate, the false negative rate, and the number of support vectors.
  - RNN: minimize the output error while minimizing the number of hidden units and the number of connections in order to reduce the complexity.
  - Decision tree: maximize the classification accuracy and minimize the size of decision tree.

Clustering

- A straightforward way to pose clustering as an optimization problem is to optimize some cluster validity index that reflects the goodness of the clustering solutions.
- Cluster validity index includes: external index, internal index, relative index.
- Multiobjective clustering techniques optimize more than one cluster validity index simultaneously.

Clustering (Encoding)

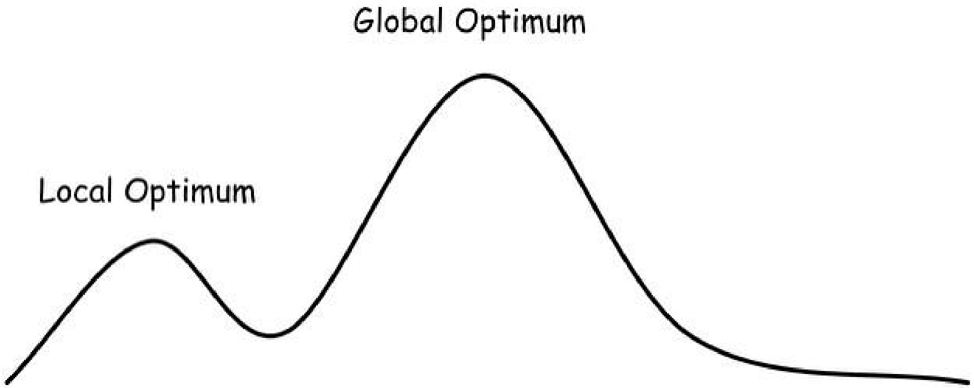
- Two major chromosome representation approaches: prototype-based approach and point-based approach.
- In the prototype-based approach, we use chromosome to represent cluster centroids, medoids and modes.
- In the point-based approach, a complete clustering solution is encoded in the chromosome.

Some other tasks

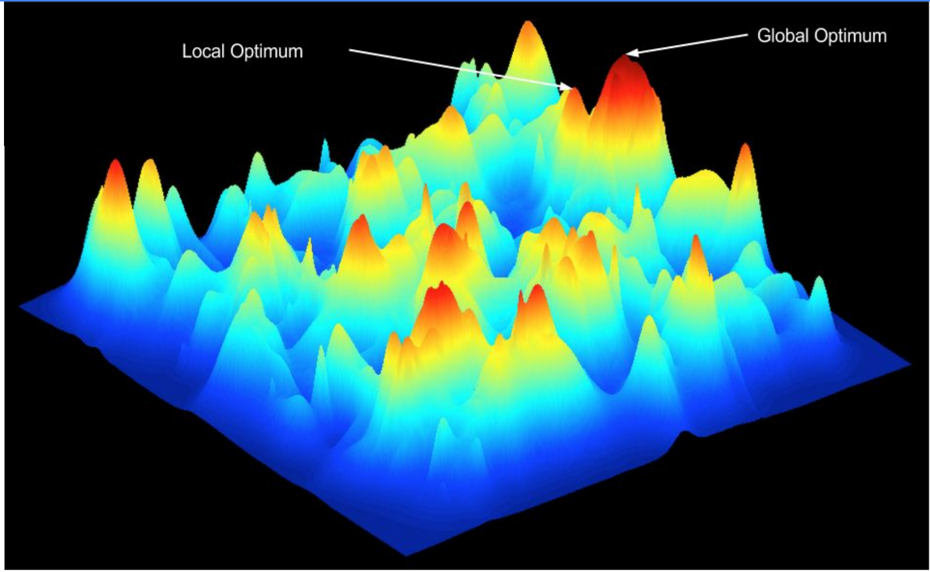
- Association Rule Mining
- Ensemble Learning
- Biclustering

Crowding

Crowding is a big problem for GA.



one variable: f(x)



two variables: f(x, y)

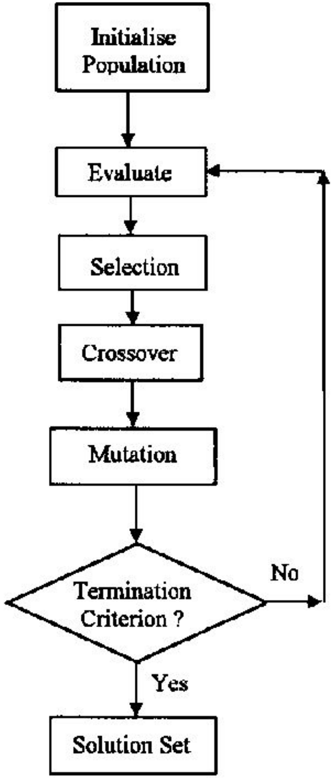
What if many variables?
f(x1, x2, x3, … , xn)

Summary

- Translate a ML problem into an Optimization or Search problem
- Use GA to solve it
- Find an encoding method
- Decide what your chromosome stand for
- Make fitness function (what is your objectives)
- Choose GA variables and operations
- Increase the diversity (deal with crowding problem)

An Experiment

- Start with random characters of string and try to approach target string. (eg. “hello, world!”)
- GA variables and operators
- Searching space: 128 * length



An Experiment

- Change string into image:
- Searching space
- Cheating on the mutation