



Recommendation System



AutoML 연구실
학부연구생 조성현
21.01.04

Contents

A many-object feature selection for multi-label classification

1. Matters to be considered in Multi-label versus Single-label
2. Advantages of Wrapper method compared to Filter method in Feature Selection
3. Genetic Algorithm(GA) component description
4. Meaning of Diversity & Convergence in GA
5. Key points and limitations of NSGA III
6. Matters and methods to be improved in this paper (in connection with 3/4/5)
7. Evaluation Measure description (Multi-label Accuracy, Subset Accuracy, Hamming Loss, One-error)

Matters to be considered in Multi-label

- Two ways to solve the multi-label problem

	Problem Transformation Method (Data is transformed to fit the algorithm)	Algorithm Adaption Method (Algorithm is transformed to fit data)
Main Idea	Solving multi-label data with single-label problem	Solving by replacing the existing single-label classification algorithm
Classifier	Any single-label classifier can be used	Benefit from specific classifier advantages
Disadvantages	Can hinder exploring label dependence The amount of computation required by the model can be large	There are not many machine learning models that can be modified to enable multiple classifications
Types	Binary Relevance(one vs. all), Label Powerset, Pairwise, Copy-Weight	K-nearest neighbors, Decision Trees, Neural Networks, Support Vector Machines

Advantages of Wrapper Method

- Filter Method vs. Wrapper Method

	Filter Method	Wrapper Method
Main Idea	Although it may not be the Best Feature Subset in the model, it can be seen that it plays a role in helping users	It is a way to find the Best Feature Subset while performing Machine Learning
Algorithm	Generic set of methods which does not incorporate a specific machine learning algorithm	Evaluates on a specific machine learning algorithm to find optimal features
Time Complexity	Much faster	High computation time for a dataset with many features
Over-fitting	Less prone to over-fitting	High chances of over-fitting
Examples	Correlation, Chi-Square test, ANOVA, Information gain, etc.	Forward Selection, Backward elimination, Stepwise selection, etc.

Genetic Algorithm components description

- Components

Components	Description
Chromosomes	Made up of genes to model them. Expressed as N data consisting of binary (0 or 1), where a gene becomes one bit. (e.g. [A B C])
Gene	A component of a chromosome, representing a single genetic information. (e.g. A, B, C)
Population	Refers to a set consisting of a plurality of chromosomes. In genetic algorithms, the number remains constant.
Fitness Function	Determines whether chromosomal information is the optimum solution and determines the direction of evolution according to how the user sets the fitness function.
Generation	A process in which individuals give birth to children and change into the next generation group. Determine whether the convergence speed is fast or slow depending on how many algorithmic operations we will achieve the desired evolutionary goal.
Offspring	Chromosomes created from chromosomes that existed at a specific time are called descendants of chromosomes that existed at the time. The offspring have similar genetic information to the previous generation.

Genetic Algorithm components description

- Calculation

Components	Description
Selection	The most essential operation and is an operation process to preferentially select genetically good individuals from the solution group . (e.g. quality proportional roulette wheel selection, tournament selection, and ranking-based selection)
Crossover	An operation is used to pass genes from parent generation to child generation . Since the characteristics or properties of an individual are inherited as parents, it can be considered that it is modeling what it inherits from the parents. (e.g. one-dimensional cross operation, multi-dimensional cross operation, and normalization)
Mutation	Difficult to explain in parents because genetic modification occurs naturally. The probability is very low , and although it can move away from the desired solution, it also acts as a dynamic factor for convergence to the desired solution. It is important to determine the appropriate probability of mutation according to the problem. In addition, it is robust against the local minima problem known as a disadvantage of optimization algorithms such as the gradient descent rule. (e.g. typical mutation, non-uniform mutation, and repair)
Replacement	Refers to replacing the population to move to the next generation . You can simply change all the elements of the population, or you can replace only those that are better than before. The imputation operation must be determined in conjunction with Crossover and Mutation operations. If the object is deformed a lot due to crossover/mutation, a highly convergent imputation operation may be used, but if the transformation occurs less, the imputation operation should be selected considering the diversity of populations.

Meaning of Diversity & Convergence in GA

- **Convergence**

- The model is said to converge when **further training with current data does not improve the model.**
- Full convergence might be seen in genetic algorithms the using only crossover. Premature convergence is when a population has converged to a single solution, but that solution is not as high of quality as expected.
- The loss value remains almost constant, and it may be mistaken that the model is temporarily converging due to numerous iterations, but it may eventually descend.
- It **can be avoided with a variety of diversity-generating techniques.**

- **Diversity**

- **The number of different behaviors, the number of different structures, the edit distance between structures in the population, and other complex or composite measures.**
- There are methods that attempt to control or promote diversity during evolution. Since the best yet found is not necessarily the global optimum, **if one exists, some diversity is usually provided by mutation to allow the population to explore the fitness landscape further.** Getting the **balance right between exploring and exploiting** is thus a matter of setting the correct level of mutation within the population.
- c.f.)The balance between convergence and diversity in the objective space requires higher exploration and development ability of the decision space. Exploration is to access new areas in the search space. Exploitation is to access the neighborhood space of previously accessed points.

Key points and limitations of NSGA III

- **NSGA(Non-dominated Sorting Genetic Algorithm)**

- It is a genetic algorithm that is widely used in the field of multipurpose optimization and is widely used in fields that require optimization. Evolutionary computation has shown great performance in solving many multi-objective optimization problems; non-dominated sorting **plays an important role in determining the relative quality of solutions in a population.**
- 1. However, **problems such as the complexity of the operation, insufficient selection conditions and difficulty in determining the value of shared variables** due to excellent prior.
- 2. The implementation of non-dominated sorting can be **computationally expensive**, especially for populations with a large number of solutions and many objectives.
 - ① The main reason is that most existing non-dominated sorting algorithms need to compare one solution with almost all others to determine its front, and many of these comparisons are redundant or unnecessary.
 - ② Another reason is that as the number of objectives increases, more and more candidate solutions become non-dominated solutions, and most existing time-saving approaches cannot work effectively.

- **NSGA II**

- Deb et al. (2002) have developed the NSGA-II that is the second version of the NSGA. An NSGA-II (Non-dominated Sorting Algorithm-II) algorithm was proposed to compensate for these shortcomings.
- 1. However, the conventional NSGA-II algorithm is optimized in the same way as the multi-purpose optimization algorithm in the direction of maximizing or minimizing the target value, making it **difficult to perform selective optimization.**

Key points and limitations of NSGA III

- **NSGA III**

- K. Deb et al. (2004) proposed NSGA III (the third version of the non-dominated sorting genetic algorithm) which improves NSGAI to make it suitable for Multi-objective Optimization Problems(MOPs) and Many-objective Optimization Problems(MaOPs). NSGA III **selects elite solutions based on a reference point method** instead of crowding distance in NSGAI. NSGA III is one of the most famous algorithms that mainly focuses on solving continuous many-objective problems.
 1. However, first of all, **feature selection is a combinational optimization problem in discrete space**. Therefore, the coding strategy, crossover and mutation operator for feature selection require a special design.
 2. Secondly, although NSGA III improves the maintaining diversity strategy of NSGAI, **how to effectively combine diversity and convergence is worthy** of further study.
 3. Thirdly, the traditional NSGA III has a **small selection pressure** on the non-dominated solution and **cannot distinguish the quality of the non-dominated solution**. So, it is necessary to increase the pressure of population selection, because the algorithm based on the Pareto dominance relationship produces many non-dominated solutions on many-objective problems.

Matters and methods to be improved

- **Matters and methods to be improved in this paper**
 - In this study, researchers propose **a many-objective optimization based multi-label feature selection method**. The optimization objectives of the algorithm include six multi-label evaluation criteria and the feature subset size. They **adopt the real number coding method** and **design new crossover operator and mutation operator** based on the label frequency difference.
 - Meanwhile, they adaptively **determine the selection threshold in CA** to explore the feature space as much as possible. And they **increase the selection pressure** on the non-dominated solutions in terms of the convergence information in CA. The positions of excellent individuals can be exchanged through the crossover operation.
 - However, the calculation cost of the wrapper feature selection algorithms is relatively large. In the future, they will investigate how **to reduce the computational complexity** of wrapper feature selection algorithms.
- **CA** : mainly focuses on local search to achieve rapid convergence
- **DA** : mainly maintains global search to improve the diversity of the population

Evaluation Measure

• Multi-label Accuracy

- Ratio of predicted label that is correct among all predicted and actual labels.

$$Acc = \frac{1}{N} \sum_{i=1}^N \frac{\hat{y}_i \cap y_i}{\hat{y}_i \cup y_i} = \frac{TP + TN}{TP + FP + FN + TN}$$

	Actually Correct	Actually Incorrect
Predicted Correct	True Positive (TP)	False Positive (FP)
Predicted Incorrect	False Negative (FN)	True Negative (TN)

• Subset Accuracy

- Indicates **the percentage of samples that have all their labels classified correctly.**
- Computes the fraction of the prediction region being identical to the true label set.
- Disadvantage of ignoring partially hit credit because all of them have to be hit. This is why dependence on labels is considered.

$$SubAcc = \frac{1}{N} \sum_{i=1}^N I(\hat{y}_i \neq y_i)$$

Evaluaiion Measure

- **Hamming Loss**

- **Evaluates the fraction of misclassified instance-label pairs among the total labels**, i.e. a relevant label is missed or an irrelevant is predicted. Because each is calculated for a label, dependence on the label may not be considered when minimized.
- Hamming loss assesses performance on a partial basis, while Exact match calculates as an instance. Therefore, it is not possible to improve both performance metrics simultaneously.

$$Hloss = \frac{1}{NL} \sum_{i=1}^N \sum_{j=1}^L I(\hat{y}_j^i \neq y_j^i) = 1 - Acc$$

- **One-error**

- **Evaluates the fraction of examples whose top-ranked label is not in the true label set.**

$$one-error = \frac{1}{N} \sum_{i=1}^N (argMax F(x_i) \notin Y_i^+)$$