MACHINE INTELLIGENCE

An Application of GA on Decision Trees for Optimal Results

Dr. Arti Arya

Department of Computer Science

artiarya@pes.edu

+080-66186629 Extn 6629

MACHINE INTELLIGENCE

An Application of GA on Decision Trees for Optimal Results

Dr. Arti Arya

Department of Computer Science and Engineering

Decision Trees Using Genetic Algorithm

Consider a problem statement as follows:

In order to serve mobile users efficiently, a mobile service provider wants to know the type of his customers in order to reach out to them better.

Can you help the mobile service provider??

Hint: Consider the title of the slide as a constraint!!!

Basically, as Computer Science Engineers, we could provide this Service provider with a system that classifies the Mobile User Dataset.

Decision Trees Using Genetic Algorithm

For Mobile Users classification, authors have found it different from General User Classification because:

1. Other than general attributes, context information providing attributes are considered.

2. Not all attributes contribute equally towards the classification.

Decision Trees Using Genetic Algorithm

So, a dataset is provided by the mobile service provider to you.

You will undergo the following steps:

1. Preprocessing
2. Rule generation from Decision Trees
3. Optimization of the Rules by Genetic Algorithm
4. Test the Optimized Rules

MACHINE INTELLIGENCE

Architecture of the DT Classifier

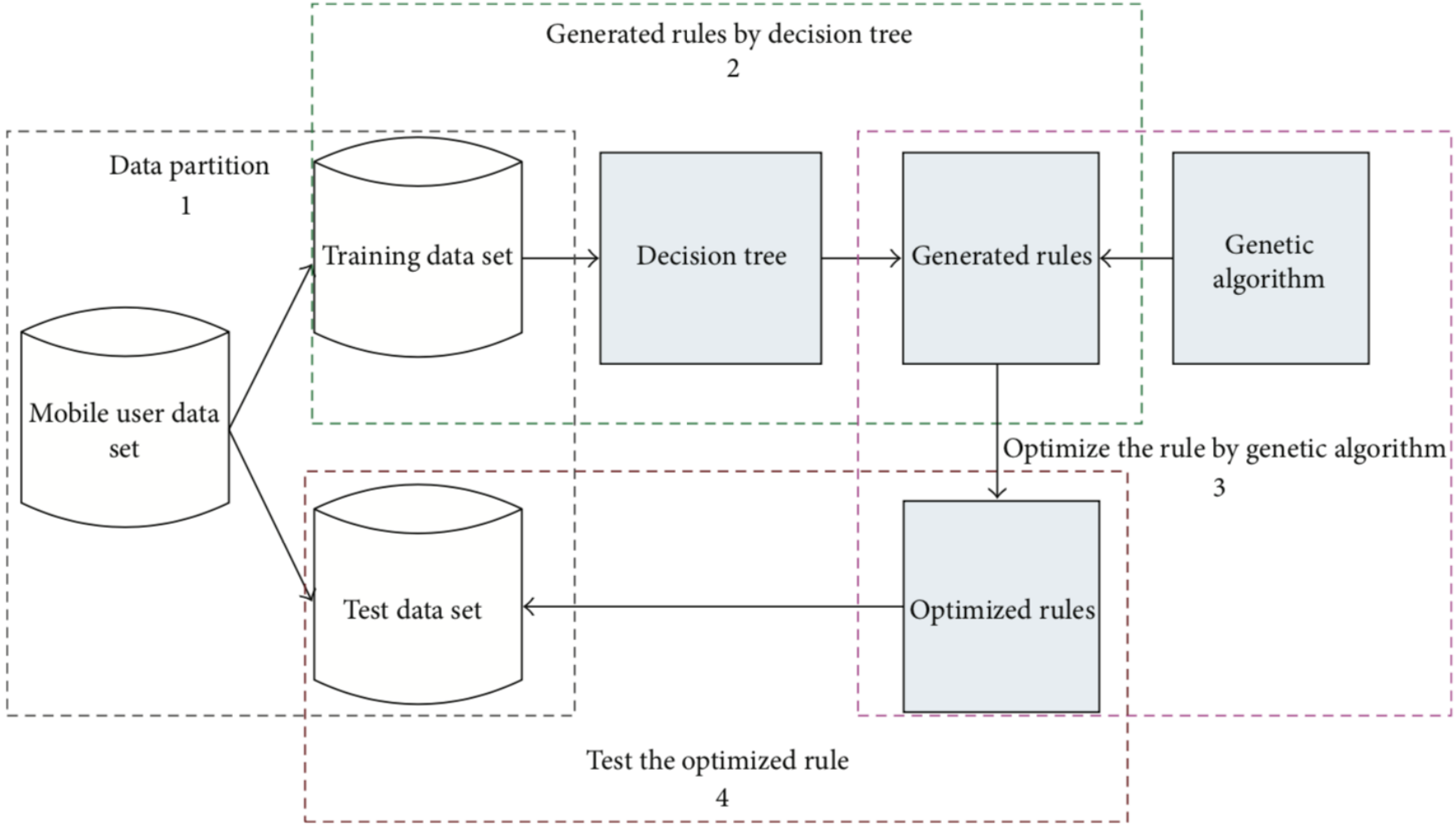

Fig Courtesy: [1] Liu DS, Fan SJ. A modified decision tree algorithm based on genetic algorithm for mobile user classification problem. The Scientific World Journal. 2014;2014

Decision Trees Using Genetic Algorithm

Entropy, Gain Ratio and Split Information is used to find the most appropriate tree.

$$I(s_1, s_2, \ldots, s_m) = -\sum_{i=1}^{m} p_i \log_2(p_i),$$

$$E(A) = \sum_{i=1}^{v} \frac{(s_{1j} + s_{2j} + \cdots + s_{mj})}{s} I(s_{1j}, s_{2j}, \ldots, s_{mj}).$$

$$I(s_{1j}, s_{2j}, \ldots, s_{mj}) = -\sum_{i=1}^{m} p_{ij} \log_2(p_{ij}),$$

$$\text{Split\_info}(A) = \sum_{j=1}^{v} \left|\frac{s_j}{s}\right| \log_2\left(\left|\frac{s_j}{s}\right|\right),$$

$$\text{gain\_ration}(A) = \frac{\text{Gain}(A)}{\text{Split\_info}(A)}.$$

MACHINE INTELLIGENCE

Coding for the Rule

Brush up off Binary Encoding

MACHINE INTELLIGENCE

Coding for the Rule

- Each chromosome represents a classification rule.
- Some chromosomes will become the solution of problem.
- The final rule set will be sorted by the quality of the rule.
- When the rule set recognizes a new sample, the best rule will be considered firstly; if the best rule cannot recognize the sample, then we can choose the next rule.
  (Rule based ordering…….Remember!!!!)
- If the rule in the rule set cannot recognize the sample either, the sample will be classified as default class.

References

[1] Liu DS, Fan SJ. A modified decision tree algorithm based on genetic algorithm for mobile user classification problem. The Scientific World Journal. 2014;2014

[2] Khatwani S, Arya A. A novel framework for envisaging a learner's performance using decision trees and genetic algorithm. In2013 International Conference on Computer Communication and Informatics 2013 Jan 4 (pp. 1-8). IEEE.

PES UNIVERSITY ONLINE

THANK YOU

Dr. Arti Arya
artiarya@pes.edu
+91 66186629 Extn 6629