

CSC480: Introduction to Data Mining

Fall 2018

Assignment 1

Decision Trees are naturally suited for discrete attributes while Multi-Layer Perceptrons (MLP) are appropriate for continuous attributes. Yet, continuous attributes can be discretized in a way that can be made appropriate for Decision Trees and discrete attributes can be transformed into continuous ones.

Question 1 (Written): What techniques are used in WEKA to deal with the continuous versus discrete attribute issue in the case of C4.5 (J48) and MLP?

Question 2 (Experimental/Written): This is a hypothesis testing question. The hypothesis is: “It is best to use classifiers well-suited to the natural knowledge structure of a domain than to convert the domain’s knowledge structure to a less natural one but one that is appropriate for the classifier in use” (In other words, one is better off using DTs on discrete domains and MLPs on continuous domains than to convert continuous domains to discrete ones for use with DTs and discrete ones to continuous ones for use with MLPs).

You are asked to test this hypothesis by selecting appropriate domains from the UCI Repository for Machine Learning (Google it!) and running J48 and MLPs on these domains to test the hypothesis. (Hint: Look at the Attributes Types and select domains from the categorical section on the one hand, and numerical section, on the other hand). You may also want to vary the kind of discretizers and converters to continuous domains that you use to ensure that the effect you are observing is linked to the type of conversion rather than to the particular converter you choose (or the default converter).

You need to think carefully about your experimental set up: what experiments will you run? What evaluation metrics, error estimation or re-sampling methods, and statistical tests will you use? And why?

Notes:

1. This is an open-ended question. Think of it as a starting point and make it your own by refining it in whichever way you feel is appropriate
2. I am not interested in seeing Weka outputs. Instead, present your results in tables or graphs as you see it done in the research papers you have been reading.

3. It is extremely important for you to construct a logical argument that explains to what extent you feel that your experiments support or disconfirm the hypothesis (or the sub-hypothesis you have formulated).
4. Your assignment should be presented somewhat like a research paper. It should have
 - a. an introduction in which the hypothesis and sub-hypothesis is/are described,
 - b. a section that describes the types of techniques that have been around to transform attributes from discrete to continuous and vice-versa and that explains which of these techniques are available in Weka [this is a sort of Literature review section but not quite.
 - c. a section that describes the experimental set-up and justifies why it is appropriate to test the hypothesis. (including a list of domains chosen, evaluation methodology, algorithms chosen and their parameter settings)
 - d. a section that present the results and discusses them. (Including their limitations: what d they not show)
 - e. a conclusion that discusses the major findings of your work and suggests avenues for future research.

Assignment Project Exam Help
I hope you enjoy it. I sure look forward to reading your papers! [you can talk to each other about the assignment, but don't all do the same thing (it sure would be boring to read! And you wouldn't learn as much if you didn't each think of how to go about answering the questions). i.e., use different domains, different sub-hypotheses, evaluation methods, etc. if relevant...] The assignment is not a team assignment. Instead, it should be done on an individual basis.

<https://powcoder.com>
Add WeChat powcoder