

Titre de l'article

S.G.CLARO

Date à mettre

1 Pertinence

No problem classic plan (Intro, Method, Result, Discu and result)

2 Qualité

2.1 Où ?

PLos One <https://journals.plos.org/plosone/>

2.2 Qui ?

Recent author, associate professor of computer Science, University of Maribor(second University of Slovenia, 1961). A lot of article written.

2.3 Quand ?

2012 : recent.

2.4 Avis

Cited : 53 times.

2.5 Pertinent ?

It's relevant !

3 Lecture article

3.1 Introduction

Gregor Stiglic (2012) :

"Decision trees are one of the most popular classification techniques in data mining. One of the main reasons for this is decision trees' ability to represent the results in a simple decision tree format which is easy to interpret for experts, as they can see the structure of decisions in the classifying process."

- Idea of decision tree format : leaves = labeled value of the class attribute; inner nodes = descriptive attributes.
- In an inner node, the children correspond to values of the associated descriptive attribute. (scheme needed, remember that to reach a leaf, sometimes need to pass through multiple descriptive attribute).
- Tree built -> determine the class value for a new instance (final leaf). Need to "follow the path from the root to a leaf according to the values of the descriptive attributes of the instance." (G. Stiglic, 2012).
- simple process -> can extract a classification rule that human can understand.
- Advantages : simple built, less time consuming classification process and decision tree rule directly used as statements in database access language (SQL).
- Built approaches : C4.5, CART, ID3(Quinlan). Field of application : bioinfo, medicine, image classification.
- Decision trees = "group of classifiers that perform classification by a sequence of simple, [...] tests whose semantics are intuitively clear [...]" (G. Stiglic, 2012).
- Decision tree classification model outperformed by SVM (Support Vector Machines) and ensemble classifier Random Forest or Rotation Forest in term of accuracy and performance metrics. But DT are the best to the process of discovering knowledge.
- Complexity of decision trees : measured as the number of nodes or the number of rules that can be extracted from a tree. Important factor that can influence the evaluation of the discovered knowledge.
- Studies in term of reducing complexity and improving the accuracy for decision trees.
- Oates and Jensen : tree size strongly depends on the training size set size.

Interesting site : <http://scikit-learn.org/stable/modules/tree.html>