Electronic Transactions on Artificial Intelligence
Vol. 4 (2000), Section C: pp. *1-20*
http://www.ep.liu.se/ej/etai/2000/001/

Refereed Article

Guy Mineau, Akshay Bissoon, and Robert Godin:

Simple Pre- and Post-Pruning Techniques for Large Conceptual Clustering Structures

Authors' affiliations:
GM, AB: Université Laval, Quebec City, Québec, Canada
RG: Université du Québec à Montréal, Montréal, Québec, Canada.

Abstract: In (Godin et al, 1995a) we proposed an incremental conceptual clustering algorithm, derived from lattice theory (Godin et al., 1995b), which is fast to compute (Mineau & Godin, 1995). This algorithm is especially useful when dealing with large data or knowledge bases, making classification structures available to large size applications like those found in industrial settings.

However, in order to be applicable on large data sets, the analysis component of the algorithm had to be simplified: the thoroguh comparison of objects normally needed to fully justify the formation of classes had to be cut down. Of course, from less analysis results classes which carry les semantics, or which should not have been formed in the first place. Consequently, some classess are useless in terms of the information needs of the applications that will later on interact with the data. Pruning techniques are thus needed to eliminate these classes and simplify the classification structure.

However, since these classification structures are huge, the pruning techniques themselves must be simle so that they can be applied in reasonable time on large classification structures. This paper presents three such techniques: one is based on the definition of constraints over the generalization language, the other two are based on discrimination metrics applied on links between classes or on the classes themselves. Because the first technique is applied before the classification structure is built, it is called the *pre-pruning* technique, while the other two are called *post-pruning* techniques.

Publication and review history

1. *Received* by the research area "Concept Based Knowledge Representation" of the Electronic Transactions on Artificial Intelligence, posted on its web site, and advertised in its Newsletter on 10.3.2000.

2. Available for *discussion* in the ETAI area of Concept Based Knowledge Representation until June, 2000. The discussion protocol is available and will be retained in the Article Interaction Page at
   http://www.etaij.org/ra/cbkr/003/

3. First version conditionally *accepted* after due refereeing and according to scientific journal standards by the Electronic Transactions on Artificial Intelligence (ETAI) on 28.9.2000.

4. Revised version *accepted* on 24.10.2000.

5. Revised version appears in Electronic Transactions on Artificial Intellligence, Volume 4 (2000), pages 1–20. The section and the annual volume are made permanently available at
   http://www.ep.liu.se/ej/etai/2000/

6. Paper editions of the ETAI issue and the ETAI volume containing this artice *republished* by the Royal Swedish Academy of Sciences.

7. First and final versions are also *published* as individual articles by Linköping University Electronic Press in December 2000 and January 2001, and are permanently available at
   http://www.ep.liu.se/ea/cis/2000/002/

The review policy and the quality requirements for acceptance are documented at http://www.etaij.org/info/

Article maintenance

The Electronic Transactions on Artificial Intelligence maintains an Article Interaction Page for this article at
http://www.etaij.org/ra/cbkr/003/
Besides the publication and review history, it contains links for contacting the author(s), as well as to amendments and other related information for the article. It is intended to keep these links up-to-date in the future.