

Induction Extensionnelle: définition et application à l'acquisition de concepts à partir de textes

Yves Kodratoff

CNRS, U. Paris-Sud, LRI, Bât. 490 91405 Orsay

yk@lri.fr

<http://www.lri.fr/~yk/>

Résumé. Lorsque des outils inductifs sont inclus dans un système d'acquisition des connaissances, on dit que l'on construit un *système apprenti*. C'est dans le but de soulager la charge de travail de l'expert du domaine que cette forme d'apprentissage comporte des outils inductifs. La difficulté tient en ce que l'énumération des connaissances expertes produit des données peu bruitées mais très incomplètes que les itérations successives d'induction vont compléter, toutefois en y ajoutant de grandes quantités de bruit. Il en résulte qu'on doit utiliser des procédures inductives spéciales, adaptées à l'apprentissage par croissance de noyaux de connaissance supervisée. En particulier, pour résoudre le problème difficile de la reconnaissance de concepts dans les textes, nous avons défini une forme d'apprentissage qui intègre *l'apprentissage à partir d'instances* et les systèmes apprentis, que nous nommons '**Induction Extensionnelle**', un oxymoron qui souligne que malgré l'absence de création d'un modèle explicite, une induction prend effectivement place.

1. Introduction

Parmi les nombreuses techniques utilisées pour faire réaliser un certain apprentissage par un ordinateur, deux occupent une place particulière, pour des raisons très différentes. Il s'agit d'une part des Systèmes Apprentis comme le système DISCIPLE de Tecuci [Kodratoff et Tecuci, 1987; Tecuci, 1998], particuliers à cause de leur complexité et de leur nombre réduit d'applications, et d'autre part de l'Apprentissage à partir d'Instances (souvent désigné par IBL: "Instance-Based Learning") [Kibler et Aha, 1988], très simple et très appliqué, mais particulier parce qu'il ne propose pas de généralisation des exemples.

Les Systèmes Apprentis se construisent en observant et analysant les étapes de résolution de problème introduites par les experts du domaine. Un autre exemple de tels systèmes, en plus de DISCIPLE déjà cité, est le premier de tels systèmes: LEAP [Mitchell et al., 1985].

L'Apprentissage à partir d'Instances a connu de très nombreuses applications, en particulier en linguistique. Par exemple, il a été utilisé pour l'analyse syntaxique [Cardie et Pierce, 1998], la catégorisation de textes [Riloff et Lehnert, 1994], l'étiquetage grammatical [Daelemans et al., 1996], pour la levée d'ambiguïté du sens des mots [Hoste et al., 2002]. L'apprentissage à partir d'instances repose sur une technique calculatoire appelée les *k* plus proches voisins. Cette technique est particulièrement bien adaptée aux techniques permettant le passage d'un apprentissage supervisé à un non supervisé: les quelques exemples classés de départ forment un noyau autour duquel on agglomère les exemples inconnus, augmentant