

Processus itératif d'extraction de classes en non supervisée

Alexandre Blansch  et Lydia Boudjeloud-Assala

Universit  de Lorraine,
Laboratoire d'Informatique Th orique et Appliqu e, LITA-EA 3097,
Metz, F-57045, France
alexandre.blanschellydia.boudjeloud-assala@univ-Lorraine.fr

R sum . Nous proposons dans cet article une nouvelle approche de classification non supervis e o  les classes sont obtenues les unes apr s les autres suivant un processus it ratif. L'approche utilise une m thode d'extraction de classes bas e sur la d tection de limite de classe, chaque classe  tant d finie par son centre. Nous avons  galement d fini des crit res d' valuation adapt s   la m thode propos e. Plusieurs exp rimentations ont montr  l'int r t de l'approche dans divers probl mes.

1 Introduction

La classification non supervis e est un probl me  tudi  depuis plusieurs d cennies, et r cemment de nouvelles approches ont  t  d velopp es pour s'adapter au challenge induit par les nouvelles m thodes d'acquisition automatique des donn es et le nombre croissant d'application produisant des donn es massives. Ces donn es doivent  tre  tudi es par des algorithmes suffisamment efficaces afin de pouvoir exploiter les connaissances qu'elles contiennent. En proc dant   une classification non supervis e, on cherche   construire des ensembles homog nes d'individus, c'est- -dire partageant un certain nombre de caract ristiques identiques. Classiquement, les m thodes fonctionnent de la fa on suivante : l'utilisateur fixe le nombre de classes, un partitionnement est ensuite g n r  puis  valu  par l'utilisateur m me ou par des crit res d'homog n it , le partitionnement obtenu pouvant  tre remis en question selon son  valuation. Nous proposons une approche diff rente qui consiste   pr senter et  valuer une par une les classes sans en fixer pr alablement le nombre. L'approche g n rique est bas e sur un processus it ratif qui va extraire les classes les unes apr s les autres permettant ainsi l'exploration pas   pas des donn es. L'approche propose   l'utilisateur en priorit  les classes les plus pertinentes (selon un crit re donn ) et lui laisse le soin de d cider quand arr ter le processus. L'approche peut  tre int gr e dans un syst me interactif qui lui permettra d' tudier les classes individuellement ou les unes par rapport aux autres. Cet article est organis  de la fa on suivante. Nous allons d'abord pr senter notre approche it rative ainsi que les crit res utilis s pour l'extraction de classes puis pr senter certains de nos r sultats exp rimentaux avant de conclure.