

Visualisation de l'intra et inter structure des groupes en classification non supervisée

Guénaël Cabanes*, Younès Bennani*

*LIPN-CNRS, UMR 7030
99 Avenue J-B. Clément, 93430 Villetaneuse, France
cabanes@lipn.univ-paris13.fr

Résumé. La croissance exponentielle des données engendre des volumétries de bases de données très importantes. Une solution couramment envisagée est l'utilisation d'une description condensée des propriétés et de la structure des données. De ce fait, il devient crucial de disposer d'outils de visualisation capables de représenter la structure des données, non pas à partir des données elles mêmes, mais à partir de ces descriptions condensées. Nous proposons une méthode de description des données à partir de prototypes enrichis puis segmentés à l'aide d'un algorithme adapté de classification non supervisée. Nous introduisons ensuite un procédé de visualisation capable de mettre en valeur la structure intra et inter-groupes des données.

1 Introduction

La croissance exponentielle des données engendre des volumétries de bases de données très importantes. Des études montrent que la quantité des données engendrées double chaque année. Parfois, l'évolution et la masse des données sont tellement importantes qu'il est impossible de les stocker dans une base et que seule une analyse "à la volée" est possible. Une solution couramment envisagée est l'utilisation d'une description condensée des propriétés et de la structure des données (Gehrke et al., 2001; Manku et Motwani, 2002; Aggarwal et al., 2003). De ce fait, il devient crucial de disposer d'outils de visualisation capables de représenter la structure des données, non pas à partir des données elles mêmes, mais à partir de ces descriptions condensées.

Dans cet article nous proposons une méthode de description des données à partir de prototypes enrichis puis segmentés à l'aide d'un algorithme adapté de classification non supervisée, puis nous introduisons des procédés de visualisation capables de mettre en valeur la structure intra et inter groupes des données.

Le reste de cet article est organisé comme suit. La section 2 présente l'apprentissage de la structure des données sous la forme d'une description condensée. La section 3 décrit l'outil de visualisation et montre quelques exemples. Une conclusion et des perspectives sont données dans la section 4.