

Analyse de la Vraisemblance des Liens Relationnels : Une méthodologie d'analyse classificatoire des données

Israël-César Lerman

Université de Rennes 1-IRISA Projet *Symbiose*,
Campus Universitaire de Beaulieu, 35042 Rennes Cedex
lerman@irisa.fr
<http://www.irisa.fr/symbiose/>

Résumé. La méthodologie de classification des données par l'Analyse de la Vraisemblance des Liens Relationnels a pris naissance vers la fin des années soixante. Elle s'est très largement développée. De nombreux chercheurs et praticiens ont pris part à son développement. De nombreuses applications d'envergure provenant des domaines les plus divers (Bioinformatique, Informatique, Sciences sociales, Traitement d'images, Traitement de langues naturelles, ...) ont validé cette méthodologie. Elle s'adresse à n'importe quel type mathématicologique du tableau de description des données. L'objet de notre article est de présenter de façon illustrée les principes fondamentaux de cette approche. Ces principes se situent d'une part au niveau de la représentation de la description et d'autre part, au niveau de l'évaluation quantifiée des ressemblances entre les structures mathématiques à comparer. Dans notre cas la représentation de la description sera ensembliste et relationnelle et la ressemblance sera évaluée au moyen d'une similarité qui se réfère à une échelle de probabilité établie par rapport à une hypothèse statistique d'absence de liaison. Le texte ci-dessous se veut le reflet de ma présentation orale aux "3-èmes Journées Thématiques Apprentissage Artificiel et Fouille des Données", 8-9 avril 2008.

1 Préambule

Mon texte s'apparente à une introduction. Il se veut le reflet exact de ma présentation orale aux "3-èmes Journées Thématiques Apprentissage Artificiel et Fouille des Données", 8-9 avril 2008. Ce texte ne correspond pas à la forme classique d'un article.

Il s'agit pour nous de présenter une méthodologie d'analyse classificatoire des données sur laquelle nous avons travaillé de très longues années. De nombreuses collaborations ont contribué de façon importante à sa mise au point. Cette méthode n'est pas vraiment connue dans ses différentes facettes ; alors qu'elle repose sur deux principes fondamentaux. Le premier consiste en la représentation ensembliste et relationnelle des variables (attributs) de description de l'ensemble des objets ou individus et le second, en l'évaluation probabiliste de la similarité entre les structures mathématiques à comparer (Indices de la *vraisemblance des liens*). Ces principes sont concrétisés quelle que soit la complexité de la description des données. Il peut s'agir