Comparaison et Évaluation de Mesures de Similarité entre Concepts d'un Treillis

Florent Domenach*, George Portides**

* Akita International University, Yuwa, Akita-city 010-1292, Japan
fdomenach@aiu.ac.jp,
** University of Nicosia, 46 Makedonitissas Ave., PO Box 24005, 1700 Nicosia, Cyprus

Résumé. Cet article se situe dans le cadre de l'analyse de concepts formels (ACF) qui fournit des classes (les extensions) d'objets partageant des caractères similaires (les intensions), une description par des attributs étant associée à chaque classe. Dans un article récent, une nouvelle mesure de similarité entre deux concepts dans un treillis de concepts a été introduite, permettant une normalisation par la taille du treillis. Dans cet article, nous comparons cette mesure de similarité avec des mesures existantes, soit basées sur la cardinalité des ensembles ou issues de la conception d'ontologies et basées sur la structure hiérarchique du treillis. Une comparaison statistique avec des méthodes existantes est effectuée et testée pour leur consistance.

1 Introduction

Cet article est un résumé de Domenach et Portides (2016). Les mesures de similarité on été largement utilisées, en particulier dans le domaine biomédical (Nguyen et Al-Mubaid, 2006) ou dans le web sémantique pour le traitement du langage naturel (Seco et al., 2004). Cependant, la plupart de ces applications reposent sur une structure ontologique arborescente pour quantifier le degré de similarité de deux concepts. Le but de cet article est d'étendre ces mesures de similarité au cadre plus général fourni par les treillis et l'analyse de concepts formels, et d'évaluer et de comparer la mesure introduite dans Domenach (2015).

2 Définitions

Analyse de Concepts Formels Nous rappelons ici les notations standards utilisées en analyse de concepts formels (ACF) et nous renvoyons le lecteur à Ganter et Wille (1999) pour plus de détails. Un *contexte formel* (O, A, I) est défini comme un ensemble O d'objets, un ensemble A d'attributs, et une relation binaire $I \subseteq O \times A$. $(o, a) \in I$ signifie que "l'objet o est lié à l'attribut a par la relation I". Deux opérateurs de dérivation peuvent être définis sur les ensembles d'objets et d'attributs comme suit, $\forall O_1 \subseteq O, A_1 \subseteq A : O_1' = \{a \in A : \forall o \in O_1, (o, a) \in I\}$, $A_1' = \{o \in O : \forall a \in A_1, (o, a) \in I\}$. Les deux opérateurs $(\cdot)'$ définissent une correspondance de Galois entre l'ensemble des parties des objets $\mathcal{P}(O)$ et l'ensemble des parties des attributs