

Nouvelle mesure de degré de relation sémantique pour une meilleure modularisation d'ontologies

Nourhène ALAYA et Sadok BEN YAHIA

Université de Tunis El Manar, Faculté des Sciences de Tunis, LR11ES14 Laboratoire d'Informatique programmation algorithmique et heuristique, 2060, Tunis, Tunisie, alaya.nourhene@etudiant.tun.tn, sadok.benyahia@fst.rnu.tn

Résumé : Dans ce poster, nous allons introduire une nouvelle mesure de degré de relation sémantique. Elle ne mesure plus la ressemblance entre deux concepts, mais plutôt la force de leur connectivité du point de vue sémantique. La nouvelle mesure est intégrée dans un algorithme de classification hiérarchique faisant partie de notre approche de modularisation.

Mots-clés : Ontologie, modularisation, classification hiérarchique, degré de relation sémantique.

1 Introduction

L'objectif initial du présent travail est d'introduire une nouvelle mesure sémantique, qui améliore les résultats de notre approche de modularisation d'ontologies volumineuses (Alaya *et al.*, 2012), qui se base sur un algorithme de classification hiérarchique. La première version de notre travail a utilisé la mesure de similarité sémantique décrite dans (Wu & Palmer, 1994). Cependant, les résultats obtenus ont montré des faibles scores en terme d'indicateur de "*modularité*" (Newman, 2006). En effet, les mesures classiques de similarité sémantique se limitent à la détermination des ressemblances en considérant la quantité d'attributs partagés par les concepts. Cependant, deux concepts peuvent être dissimilaires mais reliés sémantiquement par l'intermédiaire de relations fonctionnelles comme "Étudiant" et "Cours". Alors, nous nous sommes orientés vers l'étude des mesures de degré de relation sémantique "*semantic relatedness*", qui cherchent à identifier la similarité relationnelle entre concepts plutôt que la similarité "attributionnelle". Cependant, la majorité des mesures existantes,

e.g., (Sussna, 1993) et (Hirst & St-Onge, 1998) ne sont définies que pour WORDNET. Dans ce poster, nous proposons d'introduire une nouvelle mesure de degré de relation, qui essaye de quantifier le degré de connectivité entre deux concepts d'une ontologie de domaine. La mesure présentée ici prend en compte toute sorte de relations entre deux concepts sans faire appel à des sources de connaissances externes.

2 Nouvelle mesure de calcul du degré de relation sémantique

Nous avons choisi de modéliser l'ontologie sous forme d'un graphe orienté libellé. Les sommets de ce graphe correspondent aux concepts de l'ontologie alors que les arêtes présentent les relations sémantiques. Nous avons aussi distingué entre les liens du langage, dont les libellés sont exprimées par les constructeurs du langage de l'ontologie, et les liens de domaine dont les libellés sont attribuées par le constructeur de cette ontologie. L'objectif de notre mesure est d'estimer à quel point deux concepts de l'ontologie sont dépendants ou similaires. Pour satisfaire cette finalité, nous avons commencé par définir un ordre total sur l'ensemble des relations de l'ontologie en considérant le type de chaque relation. Une fonction de pondération *PRS* a été proposée afin d'attribuer un poids numérique à chaque type de lien. Cette fonction attribue des poids plus importants aux liens hiérarchiques de langage, que à ceux des liens non-hiérarchiques de domaine. Toutefois, les liens de domaine varient d'une ontologie à une autre, d'où la difficulté de les pondérer. Ainsi, nous avons choisi de calculer la probabilité d'apparition d'un lien sémantique de domaine dans une ontologie et la considérer comme le poids de ce type de lien.

Ensuite après avoir calculé les différentes pondérations, notre algorithme de calcul de degré de relation commence à traiter les paires de concepts adjacents. Afin de quantifier la force du lien, la mesure proposée considère à la fois le type de ce lien et la densité locale des sommets qui présentent les extrémités du lien. La densité locale d'un nœud du graphe \mathcal{G} est égale au degré de ce nœud, autrement dit le nombre de ses liens entrants et sortants. Par exemple, le degré de relation sémantique entre un concept fils et son père diminue au fur et à mesure que le nombre de concepts fils augmente. Soit c_i un nœud (concept) appartenant à notre graphe \mathcal{G} , $\text{DEGRES}(c_i)$ désigne le degré sortant de c_i ; respectivement $\text{DEGREE}(c_i)$ son degré entrant : $\text{DEGRES}(c_i) = \sum_k \text{PRS}(c_i \xrightarrow{r_{ik}} d_k)$. Nous proposons la fonction $DREL_{LS}(c_s \xrightarrow{r_{sd}} c_d)$ de calcul du degré de relation sémantique entre deux concepts adjacents c_s et c_d . La fonction $DREL_{LS}$ décrite par l'équation 1 considère le degré sortant du concept source c_s , le degré entrant du concept cible c_d , ainsi que le poids de la relation r_{sd} , calculée

par la fonction $PRS(r_{sd})$. Sa formule est décrite par l'équation 1.

$$DREL_{LS}(c_s, c_d) = \frac{1}{2} \times \left(\frac{PRS(r_{sd})}{DEGRES(c_s)} + \frac{PRS(r_{sd})}{DEGREE(c_d)} \right) \quad (1)$$

Toutefois, notre objectif n'est pas de comparer un couple particulier de concepts, mais plutôt de parcourir le graphe d'ontologie en respectant l'orientation des liens et en déterminant le degré de relation entre les différents concepts adjacents ou éloignés. Un parcours particulier est considéré comme un chemin sémantique CS . Le chemin débute par un concept source choisi aléatoirement. Les liens sémantiques de langage ou de domaine sortant de C_s seront explorés afin de déterminer les concepts cibles. Ces derniers seront à leur tour, considérés comme des points de départ pour explorer des nouveaux liens. Cette procédure se poursuit de manière itérative jusqu'à atteindre une distance prédéfinie RAY . La distance indique la longueur du chemin sémantique CS ($LONGUEUR(CS)$) en terme du nombre de liens, qui le construisent. Pendant la construction des chemins sémantiques, les motifs définis par HIRST et ST ONGE sont appliqués afin de ne retenir que les chemins sémantiquement corrects. Dans la suite, nous proposons la fonction $DREL_{CS} : \mathcal{C} \times \mathcal{C} \rightarrow [0, 1]$, qui calcule le degré de relation d'un chemin sémantique entre deux concepts éloignés c_s et c_d :

$$DREL_{CS}(c_s \xrightarrow{CS} c_d) = \prod_{h=s}^{h=d} DREL_{LS}(c_h \xrightarrow{r_k} c_{h+1}) \quad (2)$$

Toutefois, il pourrait exister des chemins multiples entre les deux nœuds éloignés c_s et c_d . Dans ce cas, nous gardons la valeur, qui maximise le degré de relation entre ces deux nœuds.

Références

- ALAYA N., YAHIA S. B. & LAMOLLE M. (2012). MODLSON : une nouvelle approche de modularisation d'ontologies à grande échelle. In *Actes d'EGC'12, Bordeaux, France*, p. 279–284.
- HIRST G. & ST-ONGE D. (1998). Lexical chains as representations of context for detection and correction of malapropisms. In *WordNet : An Electronic Lexical Database*, p. 305–332. Cambridge, Massachusetts.
- NEWMAN M. E. J. (2006). Modularity and community structure in networks. *Comput. Linguist.*, **32**(3), 379–416.
- SUSSNA M. (1993). Word sense disambiguation for free-text indexing using a massive semantic network. In *Actes de CIKM'93*, p. 67–74, USA.
- WU Z. & PALMER M. (1994). Verb semantics and lexical selection. In *Actes de the Annual Meeting of the Association for Computational Linguistics, Mexique*, p. 133–138.