

ATELIER 2 / WORKSHOP 2

" SIMILARITÉ ET RESSEMBLANCE "

" SIMILARITY AND RESEMBLANCE "

Bernadette BOUCHON-MEUNIER¹ & Jean-François OMHOVER

Aussi répandu que soit l'usage des mesures de similarité, le sens et la définition apposés sur ces mesures diffèrent selon les communautés. Les présentations faites lors de l'atelier " similarité et ressemblance " témoignaient de ces différences : aussi bien au plan théorique, concernant les axiomes qui servent de cadre aux mesures de similarité, que dans la façon dont ces mesures sont utilisées en pratique.

Successivement, nous avons abordé :

- l'utilisation des mesures de similarité de Tversky pour **la comparaison d'images**, et ainsi permettre **la recherche automatique d'images** par leur contenu.

- les propriétés nécessaires des mesures de similarité et de distance pour **l'extraction de prototypes** (représentants d'un ensemble d'éléments dans une population) à partir de données numériques.

- **la comparaison entre documents** au moyen de leur ressemblance sémantique (mots-clés) et structurelle (agencement des sections, paragraphes, phrases)

- la mise en évidence des propriétés des indices de ressemblance permettant, dans le cadre de l'analyse relationnelle, **le regroupement des éléments d'un ensemble**

- le mécanisme cognitif qui sous-tend **la catégorisation** en fonction du contexte qui l'entourne et des descriptions d'objets.

Lors de la discussion qui a suivi l'exposé des résultats scientifiques rassemblés dans cet atelier, nous avons abouti aux conclusions qui suivent.

Bien que des divergences aient été observées dans les approches et les définitions exposées lors de l'atelier, de nombreux parallèles ont pu être dressés relatifs non seulement à l'utilisation des mesures de similarité, mais aussi à leurs propriétés théoriques.

C'est dans la classification, ou le regroupement d'éléments que les mesures de similarité, ou de ressemblance, trouvent un terrain d'application commun. Que ce soit des mesures de distance, des fonctions de noyaux, des mesures de similarité entre ensembles flous ou des indices de ressemblance, il s'agit d'être capables de rassembler des éléments selon leur ressemblance commune et leurs caractéristiques en tant que groupe.

Au travers des différents travaux, nous avons pu observer que la définition de cette ressemblance **dépend toujours d'un but, ou d'un contexte donnés**.

¹ Bernadette Bouchon-Meunier, LIP6 (Laboratoire d'Informatique de l'Université de Paris 6), anime les séminaires "Apprentissage" et "Intelligence artificielle et processus de décision", du groupe de travail Intelligence Artificielle et Processus de Décision, de l'AFSCET
<http://www.afscet.asso.fr/grpeBB.html>

Ces méta-connaissances indiquent, selon les cas : la façon d'équilibrer les mesures de **similarité interne** et de **dissimilarité externe** en fonction de l'éparpillement des données, la pondération de la similarité textuelle par rapport à la similarité structurelle d'un document web, ou le contexte indiqué par la question qui sous-tend la recherche d'un élément similaire dans un ensemble d'éléments catégorisés.

Nous avons aussi pu valider les propriétés des mesures de similarité dans différents contextes. Elles dépendent, dans la plupart des cadres exposés, d'une évaluation quantitative des caractéristiques communes et distinctives des deux éléments comparés.