

RAISONNEMENT ANALOGIQUE ET SIMILARITE

BEN HASSENA Anouar
Anouar.benhassana@insa-lyon.fr

MEFTEH Wafa
Wafa.mefteh1@insa-lyon.fr

14 November 2006

INTRODUCTION :

La capacité de percevoir des similitudes et des analogies est l'un des aspects les plus fondamentaux d'une connaissance humaine. Elle joue un rôle important dans la découverte et la créativité scientifique. Ces dernières années, la similitude et l'analogie ont suscité l'attention croissante des scientifiques cognitifs.

Selon S. Vosniadou, A. Ortony [1], cette croissance d'intérêt de l'analogie est liée au fait que le raisonnement humain ne fonctionne pas toujours sur une base de règles d'inférence générales. mais, plutôt, souvent attachée aux objets particuliers de la connaissance. L'interprétation humaine est considérablement influencée par le contexte dans lequel elle se situe.

Ce rapport a pour objectif de présenter et de mettre en relation les notions de similarité et de raisonnement par analogie dans les différents domaines évoqués. Dans cette optique, on a divisé le travail en deux parties : une présentant la similarité, son utilité et ses différents types selon plusieurs auteurs. La deuxième s'intéresse à étudier le mécanisme du raisonnement analogique et son rapport avec la similitude, enfin on présente quelques exemples d'applications basés sur ces deux notions.

I. Similarité :

I.1 Qu'est ce que la similarité :

La similarité est définie par le degré de ressemblance entre deux objets. En effet, tout système ayant pour but d'analyser ou d'organiser automatiquement un ensemble de données ou de connaissances doit utiliser, sous une forme ou une autre, un opérateur de similarité dont le but est d'établir les ressemblances ou les relations qui existent entre les informations manipulées.

I.2 Utilisation de la similarité :

La notion de similarité a fait l'objet d'importantes recherches dans des domaines extrêmement divers tels que l'Analyse des Données AD, la Reconnaissance des Formes RF, les Sciences Cognitives SC. Nous allons décrire le cadre dans lequel les fonctions de similarité sont couramment utilisées.

I.2.1 Classification :

Le processus de classification vise à structurer les données contenues dans un univers U , en fonction de leurs ressemblances, sous la forme d'un ensemble de classes à la fois homogènes et contrastées. Par ailleurs, le lien entre classification et similarité est étudié en SC (on parle alors de catégorisation) puisqu'il s'agit d'un processus central dans la cognition humaine. Dans tous les cas, le critère de formation des classes consiste, de manière plus ou moins directe, à maximiser la mesure de similarité intra-classes et à minimiser la mesure de similarité inter-classes.

I.2.2 Identification :

Le processus d'identification a pour but de connaître la classe à laquelle un objet inconnu est susceptible d'appartenir ou encore, de trouver à quel(s) objet(s) de U il est le plus ressemblant. Dans ce dernier cas, la méthode dite des plus proches voisins, initialement introduite en AD, utilise directement la notion de similarité : en effet, l'objet de U qui est sélectionné est celui qui maximise la mesure de similarité avec l'instance inconnue. Cette approche est également utilisée en Apprentissage Symbolique AS, sous des formes plus ou moins complexes, dans le cadre de l'apprentissage à partir d'instances (IBL : Instance Based Learning, Par ailleurs, le processus d'identification est analysé en SC au travers de l'étude générale de la notion de similarité et de celle d'analogie [1].

I.2.3 caractérisation :

Le processus de caractérisation permet de construire une représentation explicite des informations qui sont communes à un ensemble de données. Ce processus correspond à l'opération de généralisation qui a fait l'objet de nombreuses recherches en AS et notamment dans le cadre de la programmation logique inductive.

I.3 Types de similarités

On peut définir des critères permettant de classer la similarité en plusieurs familles :

I.3.1 Similarités numérique et symbolique

Les mesures de similarités numériques s'avèrent être d'un emploi extrêmement souple. Elles sont capables de travailler sur un large spectre de types de données et il est assez facile d'introduire, lors du calcul, des approximations statistiques si les informations à traiter sont complexes. En outre, la quantification des ressemblances par une valeur continue implique qu'il est toujours possible et facile de comparer des couples d'objets entre eux. Par contre pour la similarité symbolique le traitement des valeurs numériques est alors souvent fait de manière peu satisfaisante en redécrivant ces valeurs sous une forme symbolique.

I.3.2 Similarités propositionnelle et relationnelle

Selon les données initiales sont exprimé soit sous une forme de conjonction attribut-valeur dans ce cas il suffit en première approximation de comparer deux à deux la valeur des attributs qui les composent c'est la similarité propositionnelle, soit les données sont exprimés sous la forme de conjonction de prédicats-instancié et dans ce cas il faut décider comment les objets vont être comparés entre eux.

I.3.3 Similarités informée et non-informée

A l'instar de Marcotorchino [4], on peut établir une distinction entre les similarités non-informées et les similarités informées. Dans le premier cas, le calcul se fait sur une base purement locale en ne prenant en compte que les informations qui sont explicitement présentes dans les exemples alors que dans le second cas on utilise en outre des informations d'ordre statistique ou encore symbolique portant sur l'ensemble des exemples de l'univers dans lequel on travaille. Ainsi, la similarité entre A et B est de 2/3 car ils ont deux modalités communes parmi les trois attributs qui les caractérisent.

	Forme	Percé	Couleur
A : 	carré	non	noir
B : 	carré	non	gris
C : 	rond	non	gris
D : 	rond	oui	gris

table 1 Similarités informées et non-informées

La ressemblance entre deux modalités du même attribut va être, comme précédemment, égale à 0 si elles sont différentes et va être égale à l'inverse du nombre total d'apparitions de la modalité sur l'ensemble des exemples, si elles sont identiques. Par exemple, la similarité entre A et B est égale à 5/6 car d'une part, les deux objets sont de même forme et la modalité carré est présente chez deux individus dans l'ensemble des exemples (1/2), d'autre part, ils sont tous deux non percés et cette modalité est présente chez trois individus (1/3).

Selon Gilles Bisson [3], On pourrait certes imaginer de calculer la ressemblance entre tous les couples d'objets et d'apparier les objets qui ont les N meilleures valeurs de similarité. Cette façon de faire n'est pas du tout satisfaisante car l'appariement que l'on obtiendrait ne serait basé que sur les attributs et pas du tout sur les relations entre les objets.

Sur cette base, Gentner et repris par S. Vosniadou et A. Ortony [1], propose une classification des méthodes d'appariement (tab. 2) en fonction du type d'information mise en œuvre.

	Attributs	Relations	Exemple
Sim. apparentes	Beaucoup	Peu	Une luciole ressemble à une lampe
Sim. littérales	Beaucoup	Beaucoup	Le lait ressemble à de l'eau
Analogie	Peu	Beaucoup	L'atome et le système solaire
Abstraction	Très peu	Beaucoup	La chaleur s'écoule par diffusion

table2 : Les différents types d'appariements

II. Analogie :

II.1 Analogie et appariements

Une similarité est, en fait, la comparaison d'une chose avec une autre. « *L'analogie opérationnalise cette similarité en mettant l'accent sur les relations entre items* » (Bunge, 1973).

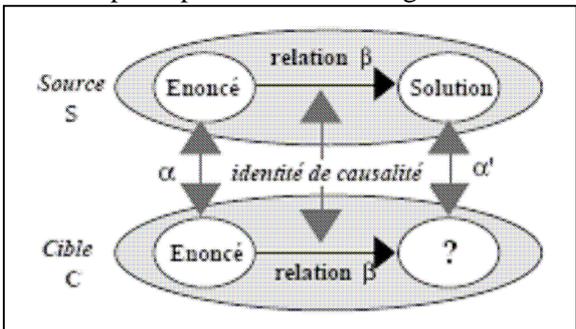
« *Un tel type de raisonnement fait habituellement appel à un appariement des éléments d'un domaine source avec ceux d'un domaine cible, domaines a priori différents.* » (Nicolas STROPPIA 2005).

“*Analogy will be described as a mapping between elements of a source domain and a target domain*». Hall (1989, p. 40)

En général, les connaissances disponibles au niveau du domaine source sont plus nombreuses que celles relatives au domaine cible. Un des objectifs du raisonnement par analogie est alors d'exploiter un appariement adéquat effectué entre les éléments des domaines source et cible, de manière à transférer un certain nombre de connaissances du premier vers le deuxième.

II.2 Principe de base : carré d'analogie

Les modèles de raisonnement par analogie partagent une vision de celui-ci impliquant tout ou partie des étapes représentées sur la figure 1



1. construction d'une représentation ;
2. recherche (remémoration, accès) d'une situation source analogue ;
3. construction d'un appariement entre les éléments des situations source et cible ;
4. transfert des éléments inappariés de la source vers la cible (adaptation, apprentissage, inférence).

Fig 1 Carré d'analogie

Par exemple, Falkenhainer et al. (1989) examine le cas de l'analogie entre deux configurations, l'une étant le siège d'un flux de liquide (le domaine source) et l'autre d'un flux thermique (le domaine cible).

1. Construction d'une représentation : Cette étape consiste à se munir d'un formalisme capable d'exprimer les connaissances disponibles relatives aux deux domaines. Dans notre exemple, pour le premier domaine :

- grand(vase), petit(fioule) ;
 - relié(vase, fioule) ;
 - supérieur(pression(vase), pression(fioule)) ;
 - cause(supérieur(pression(vase), pression(fioule)), flux),
- et pour le second :
- relié(café, glaçon) ;
 - supérieur(température(café), température(glaçon)).

2. Recherche d'analogues : La seconde étape consiste à rechercher dans une mémoire à long terme des situations potentiellement appariables avec la situation à analyser. Cette recherche peut être exhaustive, mais est souvent guidée par des connaissances relatives au contexte. Une position extrême est de ne considérer qu'une seule situation source : c'est ce qui est fait dans l'exemple. Dans un cas plus général, il s'agit d'effectuer cette recherche dans un ensemble potentiellement grand de situations sources. Cela est fréquemment le cas lorsque l'on cherche à résoudre un nouveau problème à l'aide de problèmes déjà résolus. On parle dans ce contexte de raisonnement à partir de cas (Kolodner, 1993 ; Aamodt & Plaza, 1994 ; Leake, 1996).

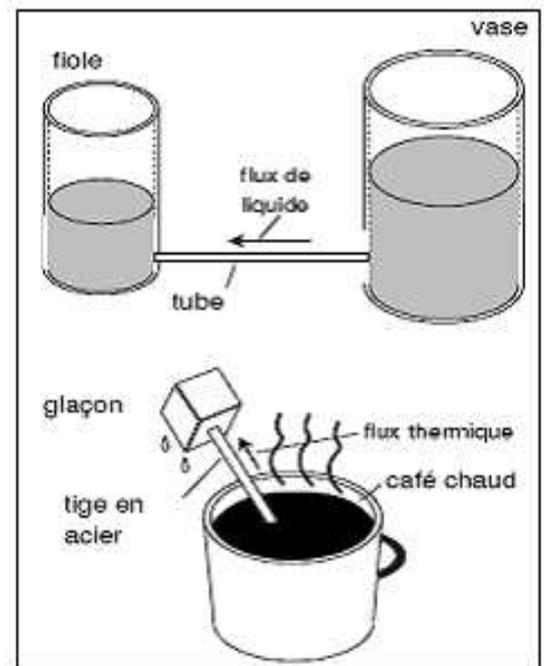


Fig 2 Analogie entre un flux de liquide et un flux thermique

3 Construction d'un appariement : La construction d'un appariement a pour objectif de mettre en correspondance les éléments des deux domaines, permettant de dire à quel point et en quoi les domaines sont analogues.

Dans l'exemple de l'analogie entre flux, les termes café et vase sont appariés, de même que les relations supérieures (pression(vase), pression(fiole)) et supérieur (température(café), température(glaçon)), donnant lieu au schéma représenté sur la partie haute de la figure 3.

4. Transfert : L'étape de transfert exploite quant à elle l'appariement effectué de manière à enrichir les connaissances disponibles sur le domaine cible. Elle peut être vue, d'une certaine façon, comme le rétablissement d'un équilibre, consistant à appairer les éléments encore inappariés avec des éléments nouveaux dont l'existence est supposée. Ce processus est illustré sur la partie basse de la figure 3.

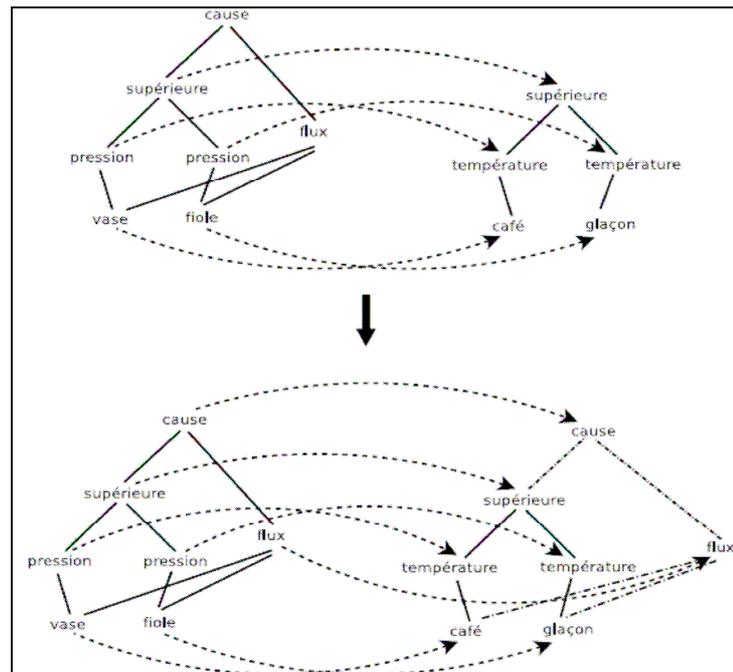


fig 3 Appariement et transfert

III. Exemples d'applications et d'outils :

III.1 ANALOGY

Le système ANALOGY, proposé par Evans (1968), est la première implantation d'un modèle de résolution d'équation analogique. L'objectif de ce programme est de pouvoir résoudre des équations analogiques géométriques à choix multiples, présentées sous la forme d'une équation analogique « *A est à B ce que C est à ?* », où les trois termes A, B et C sont des représentations de figures géométriques.

La figure 4 donne un exemple de tel problème : à partir des trois figures A, B et C, quelle est la figure D choisie parmi les cinq figures du dessous satisfaisant au mieux la relation « A est à B ce que C est à D » ? Ces équations proviennent en particulier de tests de QI et d'examens tels que les « Scholastic Aptitude Tests » utilisés fréquemment à l'entrée des universités américaines. Une particularité du système ANALOGY réside dans le fait que les termes de l'équation sont présentés sous forme d'une description bas-niveau (points, courbes, etc ...)

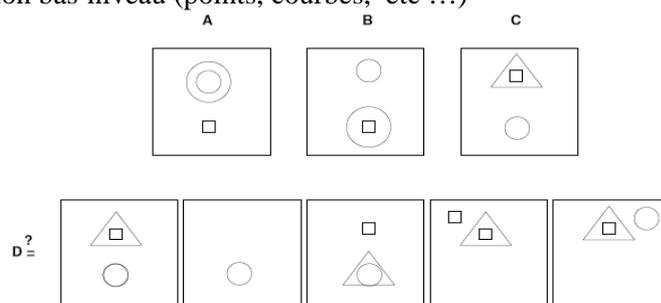


fig 4 Exemple d'équation analogique traitée par ANALOGY

Le procédé du système ANALOGY se décompose en quatre étapes consistant à :

- générer l'ensemble des règles permettant de transformer A en B ;
- générer l'ensemble des règles permettant de transformer C en chacune des solutions proposées ;
- comparer le premier et le deuxième ensembles de règles. Pour chaque paire de règles, il est possible de proposer une généralisation des deux règles, de manière à réconcilier des règles partiellement différentes : plus les règles sont différentes, plus la généralisation à effectuer est importante ;
- choisir la règle transformant C en un choix possible qui généralise le moins la règle transformant A en B, i.e. qui modifie le moins possible la règle originelle permettant de passer de A à B.

Ce modèle « simple » possède une caractéristique méritant d'être soulignée dès à présent : la représentation des règles de transformation sous la forme d'ensembles d'opérations consistant à éliminer, ajouter, substituer des éléments offre des similitudes remarquables avec les opérations sous-jacentes aux calculs de distances d'édition entre chaînes de symboles (Wagner & Fischer, 1974).

III.2 SME :

SME est un système basé sur *la théorie de l'appariement structurel* qui expose un certain nombre de critères : (i) les appariements entre relations doivent être préférés aux appariements entre attributs ; (ii) seules des relations identiques peuvent être appariées ; (iii) les appariements impliquant des relations d'ordre supérieur sont prioritaires. Ce dernier critère correspond au principe de systémativité : on recherche la mise en correspondance de systèmes de relations et non de relations isolées.

En effet, Le travail effectué par le système SME est alors, à partir d'une description des différentes relations et attributs de l'origine et du but de l'analogie, de calculer les liens possibles, d'en construire les *correspondances*.

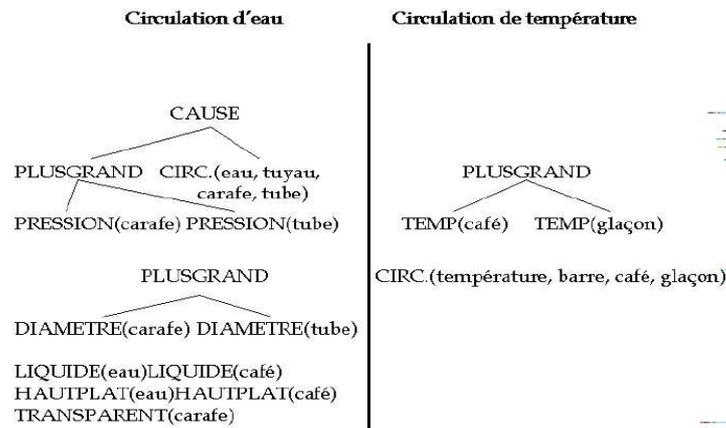


Figure 5 Représentation de l'eau et de la température donnée à SME

Pour reprendre l'exemple décrit plus haut, l'ensemble de relations et d'attributs présentés dans la figure 5 va être fourni à SME. Des relations hors du contexte de l'analogie, par exemple l'attribut HAUTPLAT ou la relation PLUSGRAND[DIAMETRE(carafe), DIAMETRE(tube)], ont intentionnellement été fournies à SME dans le but de symboliser l'approche approximative que peuvent avoir des personnes dans l'interprétation des analogies.

A partir de ces relations et attributs, SME effectue trois passes :

- 1) déterminer des correspondances locales entre attributs ou relations (par comparaison du nom par exemple),
- 2) la construction d'une correspondance globale, alors synonyme de corrélation structurelle entre les deux domaines,
- 3) une évaluation visant à discriminer les différents candidats possibles selon des critères comme le nombre de relations ou d'attributs partagés.

La conclusion alors fournie par SME est :

CAUSE{ PLUSGRAND[TEMP(café), TEMP(glaçon)],
CIRC.(température, barre, café, glaçon)}.

Conclusion :

La similarité et l'analogie ont fait l'objet d'importantes recherches dans des domaines extrêmement divers tels que l'Analyse des Données, la Reconnaissance des Formes, les Sciences Cognitives.

Ce travail vise d'une part à montrer aux lecteurs toute la richesse de la notion de similarité et analogie au travers d'une étude bibliographique.

D'autre part, il s'agit aussi de montrer l'importance de similarité dans la phase de recherche du raisonnement analogique.

A la lumière de cette étude, il apparaît clairement que l'usage du raisonnement analogique est du fondamentale car il résout plusieurs problèmes complexes à partir du souvenir des problèmes déjà résolus.

Bibliographie:

[1] GENTNER D. 1989. The mechanisms of analogical learning. *Similarity and Analogical Reasoning*. S. Vosniadou, A. Ortony (eds).

[2] STROPPIA N. 2005. Définitions et caractérisations de modèles à base d'analogies pour l'apprentissage automatique des langues naturelles

[3] Bisson G. LA SIMILARITE : UNE NOTION SYMBOLIQUE/NUMERIQUE

[4] MARCOTORCHINO F. 1991. La classification automatique aujourd'hui : bref aperçu historique applicatif et calculatoire. Publications Scientifiques et Techniques d'IBM. Numéro 2. 35-94.

