

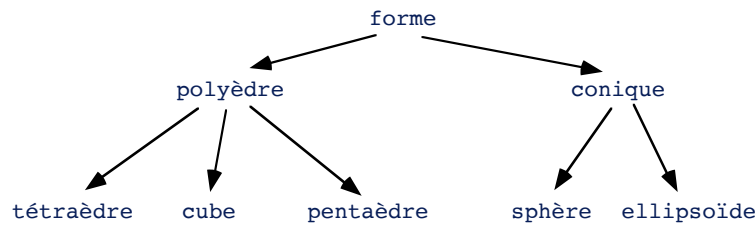
DA IODAA : Cours d'Apprentissage Artificiel

(À remettre le 25 septembre 2023 par email ou sur papier)

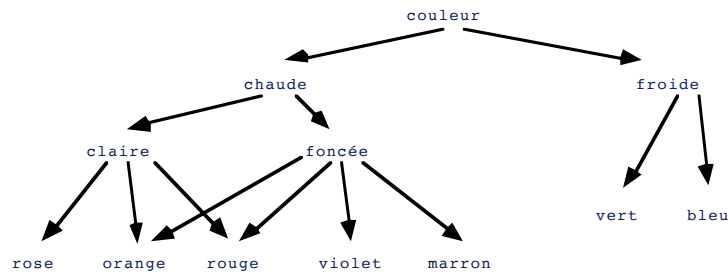
1- On suppose qu'une sonde est envoyée sur Mars, et qu'elle collecte des échantillons décrits à l'aide de trois descripteurs : leur forme (tétraèdre, cube, pentaèdre, sphère, ellipsoïde), leur couleur (vert, bleu, rose, orange, rouge, violet, marron) et leur composition chimique partielle (cétone, amine, organo-phosphoré, oxyde, sulfure).

On suppose de plus que les descripteurs sont organisés sous forme de hiérarchies dont le sommet correspond au plus grand niveau de généralité, tandis que les feuilles correspondent au niveau le moins général (ainsi, dans l'arbre des descripteurs des formes ci-dessous, *tétraèdre* est moins général que *polyèdre*, qui lui-même est moins général que *forme*, par contre, il n'y a pas de lien de généralité entre *tétraèdre* et *ellipsoïde* car ils appartiennent à des branches différentes de l'arbre. De même, *orange* est à la fois moins général que *claire* et que *foncée*).

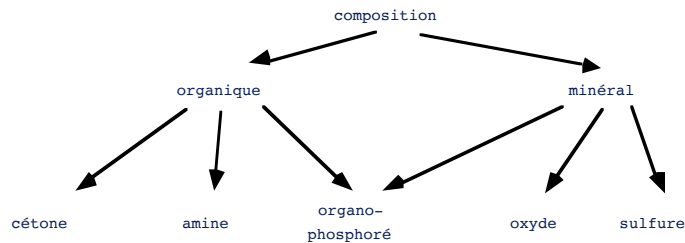
Taxonomie des formes



Taxonomie des couleurs



Taxonomie des compositions chimiques



Soit la séquence d'exemples suivante :

- | | | | | |
|----|--------------|----------|--------------------|--------------|
| E1 | : Ellipsoïde | & orange | & amine | --> classe + |
| E2 | : Sphère | & rouge | & organo-phosphoré | --> classe + |
| E3 | : Pentaèdre | & vert | & sulfure | --> classe - |
| E4 | : Sphère | & rose | & amine | --> classe - |

1) Spécifiez le S-set et le G-set après la prise en compte de chacun d'eux.

2) Comment seront alors classés les exemples suivants et pourquoi ?

Sphère & marron & cétone

Sphère & bleu & oxyde

Pentaèdre & violet & cétone

2- Réflexion sur la fragmentation du G-set.

1) Pourquoi le G-set peut facilement être de taille exponentielle en le nombre d'exemples d'apprentissage ?

Essayez d'illustrer la possibilité de ce phénomène sur un exemple.

2) Quel remède pourrait permettre de limiter de phénomène ? Réfléchissez en particulier à l'utilisation de near-miss (*contre-exemples critiques*). Quel effet ont-ils sur le G-set ?

On rappelle ci-dessous l'**algorithme d'élimination des candidats** :

Initialiser S et G par les ensembles des généralisations les plus spécifiques et les plus générales cohérentes avec le premier exemple positif

Pour chaque exemple suivant **i** :

Début

Si **i** est un exemple négatif **alors**

- ne retenir dans S que les généralisations ne couvrant pas **i**
- rendre plus spécifiques les généralisations de G couvrant **i** juste assez pour qu'elles ne couvrent plus **i**, et seulement de manière que chacune reste plus générale que certaines généralisations de S
- Retirer de G tout élément plus spécifique que d'autres éléments de G

fin si

sinon, si **i** est un exemple positif **alors**

- ne retenir dans G que les généralisations couvrant **i**
- généraliser les éléments de S ne couvrant pas **i** juste assez pour leur permettre de couvrir **i**, et seulement de telle manière que chacune reste plus spécifique que certaines généralisations de G
- retirer de S tout élément plus général que d'autres éléments de S

fin si

fin pour chaque