


Apprentissage d'ensembles de règles



Références

Ces diapositives se basent en partie sur les sources suivantes :

- <http://www.cs.cmu.edu/~tom/mlbook-chapter-slides.html>
- Pour AQ :
https://www.lri.fr/~aze/Presentations/expose_aq.ppt □

Problème d'apprentissage d'un ensemble de règles

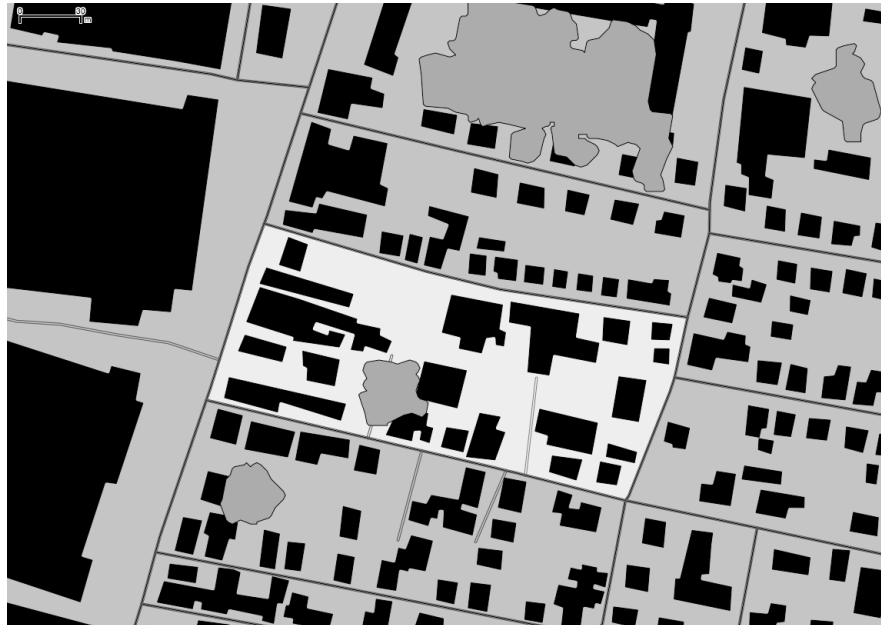
Etant donné

- une fonction cible
- un ensemble E d'exemples étiquetés pour cette fonction
- des connaissances du domaine (optionnel)
- un langage des hypothèses H

Trouver

un ensemble de règles de H qui discrimine les exemples de E (les classe correctement en fonction des étiquettes)

Exemple de données



Exemple de problème d'apprentissage

Etant donné

- une fonction (inconnue) qui classe des îlots urbains (habitat individuel/mixte/collectif, emprise spécialisée, ...)
- un ensemble E d'îlots étiquetés (i_1 : habitat individuel, i_2 : emprise spécialisée, ...). Ces îlots ont des caractéristiques (surface, élongation, densité, ...)
- des connaissances sur les bâtiments appartenant à ces îlots (surface, convexité, ... d'un bâtiment)
- un langage des hypothèses H

Trouver un ensemble de règles de H qui classe correctement les exemples de E

Règles

- Formules logiques
- **Si conditions alors conclusion**
(ex. *Si densité* < 0,45
alors classe = *emprise spécialisée*)
- Lisibilité
- Différents niveaux d'expressivité
(représentation attribut-valeur, restriction de la logique du premier ordre, ...)

Apprentissage d'ensembles disjonctifs de règles

- Exemple :

*Si densité $< 0,45$
alors classe = emprise spécialisée*

*Si élongation $> 0,9$
alors classe = emprise spécialisée*

Apprentissage d'ensembles disjonctifs de règles

- Intérêts
 - expressif
 - lisible
- Plusieurs façons de procéder
 - transformer un arbre de décision
 - utiliser des algorithmes génétiques
 - directement

Points de comparaison des systèmes d'apprentissage d'ensembles de règles

- Langage des hypothèses
- Stratégie d'apprentissage d'un ensemble de règles
 - Simultané / séquentiel
 - Une / plusieurs classe(s) à la fois
- Apprentissage d'une règle :
 - Stratégie de parcours de l'espace des hypothèses
 - Stratégie d'évaluation des hypothèses (de choix de la meilleure hypothèse)

Langage des hypothèses



Langage des hypothèses

- Formalisme logique pour représenter les règles
- Définit l'espace des hypothèses -> espace de recherche (espace à explorer)
Hypothèse = solution potentielle
- Choix dépend
 - du problème : ce qu'on a besoin de représenter
 - d'un compromis expressivité/complexité de la recherche dans l'espace des hypothèses

Structurer l'espace de recherche

- Définir un ordre entre les hypothèses :
opérateur de subsomption
- Subsomption = relation de généralité
- Pourquoi ? ...

Ensemble
d'apprentissage

Jour	Ciel	Température	Humidité	Vent	Jouer
J1	Soleil	Chaud	Elevée	Faible	Non
J2	Soleil	Chaud	Elevée	Fort	Non
J3	Couvert	Chaud	Elevée	Faible	Oui
J4	Pluie	Doux	Elevée	Faible	Oui
J5	Pluie	Froid	Normale	Faible	Oui
J6	Pluie	Froid	Normale	Fort	Non
J7	Couvert	Froid	Normale	Fort	Oui
J8	Soleil	Doux	Elevée	Faible	Non
J9	Soleil	Froid	Normale	Faible	Oui
J10	Pluie	Doux	Normale	Faible	Oui
J11	Soleil	Doux	Normale	Fort	Oui
J12	Couvert	Doux	Elevée	Fort	Oui
J13	Couvert	Chaud	Normale	Faible	Oui
J14	Pluie	Doux	Elevée	Fort	Non

Langage des hypothèses : le même que celui des exemples

Structurer l'espace de recherche

- Subsumption = relation de généralité
- Question : que veut dire qu'une hypothèse est plus générale qu'une autre ?
- Deux hypothèses :
 - H1 : Ciel = Soleil
 - H2 : Ciel = Soleil et Vent = Fort

Quelle est la plus générale ?

Ensemble
d'apprentissage

Jour	Ciel	Température	Humidité	Vent	Jouer
J1	Soleil	Chaud	Elevée	Faible	Non
J2	Soleil	Chaud	Elevée	Fort	Non
J3	Couvert	Chaud	Elevée	Faible	Oui
J4	Pluie	Doux	Elevée	Faible	Oui
J5	Pluie	Froid	Normale	Faible	Oui
J6	Pluie	Froid	Normale	Fort	Non
J7	Couvert	Froid	Normale	Fort	Oui
J8	Soleil	Doux	Elevée	Faible	Non
J9	Soleil	Froid	Normale	Faible	Oui
J10	Pluie	Doux	Normale	Faible	Oui
J11	Soleil	Doux	Normale	Fort	Oui
J12	Couvert	Doux	Elevée	Fort	Oui
J13	Couvert	Chaud	Normale	Faible	Oui
J14	Pluie	Doux	Elevée	Fort	Non

H1 : Ciel = Soleil

H2 : Ciel = Soleil et Vent = Fort

Structurer l'espace de recherche

- Un exemple est **couvert** par une hypothèse s'il vérifie les conditions de l'hypothèse
 - H1 : Ciel = Soleil couvre {J1,J2,J8,J9,J11}
 - H2 : Ciel = Soleil et Vent = Fort
couvre {J2,J11}

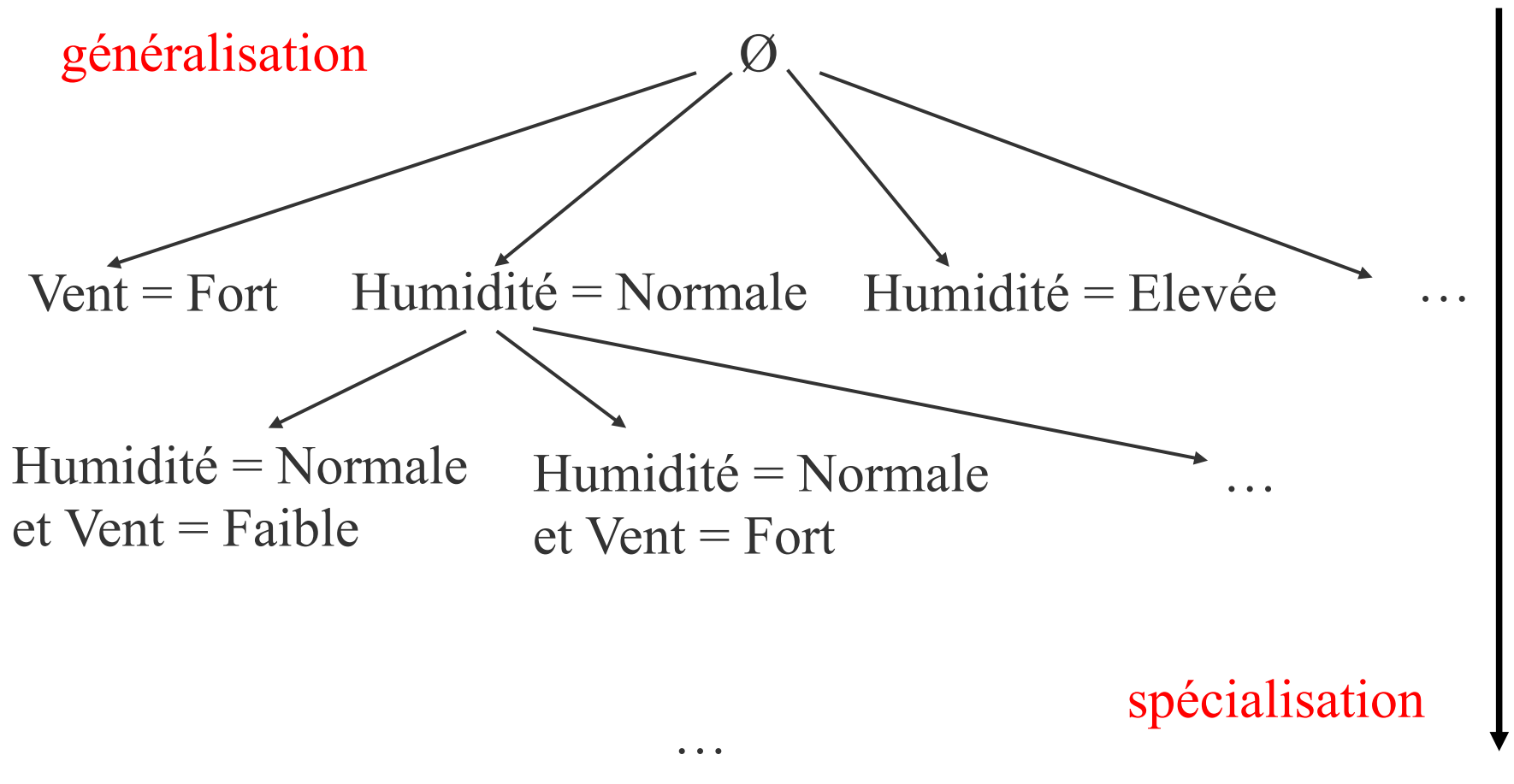
Structurer l'espace de recherche

- Un exemple est **couvert** par une hypothèse s'il vérifie les conditions de l'hypothèse
 - H1 : Ciel = Soleil
couvre {J1,J2,J8,J9,J11}
 - H2 : Ciel = Soleil et Vent = Fort
couvre {J2,J11}
- H1 est plus générale (ou égale) que H2
(noté $H1 \preceq H2$)
car $\text{couverture}(H2) \subseteq \text{couverture}(H1)$

Structurer l'espace de recherche

- Définir un ordre entre les hypothèses :
opérateur de subsomption
- Subsomption = relation de généralité
- On utilise cet ordre pour structurer l'espace

Généralisation/spécialisation



Structurer l'espace de recherche

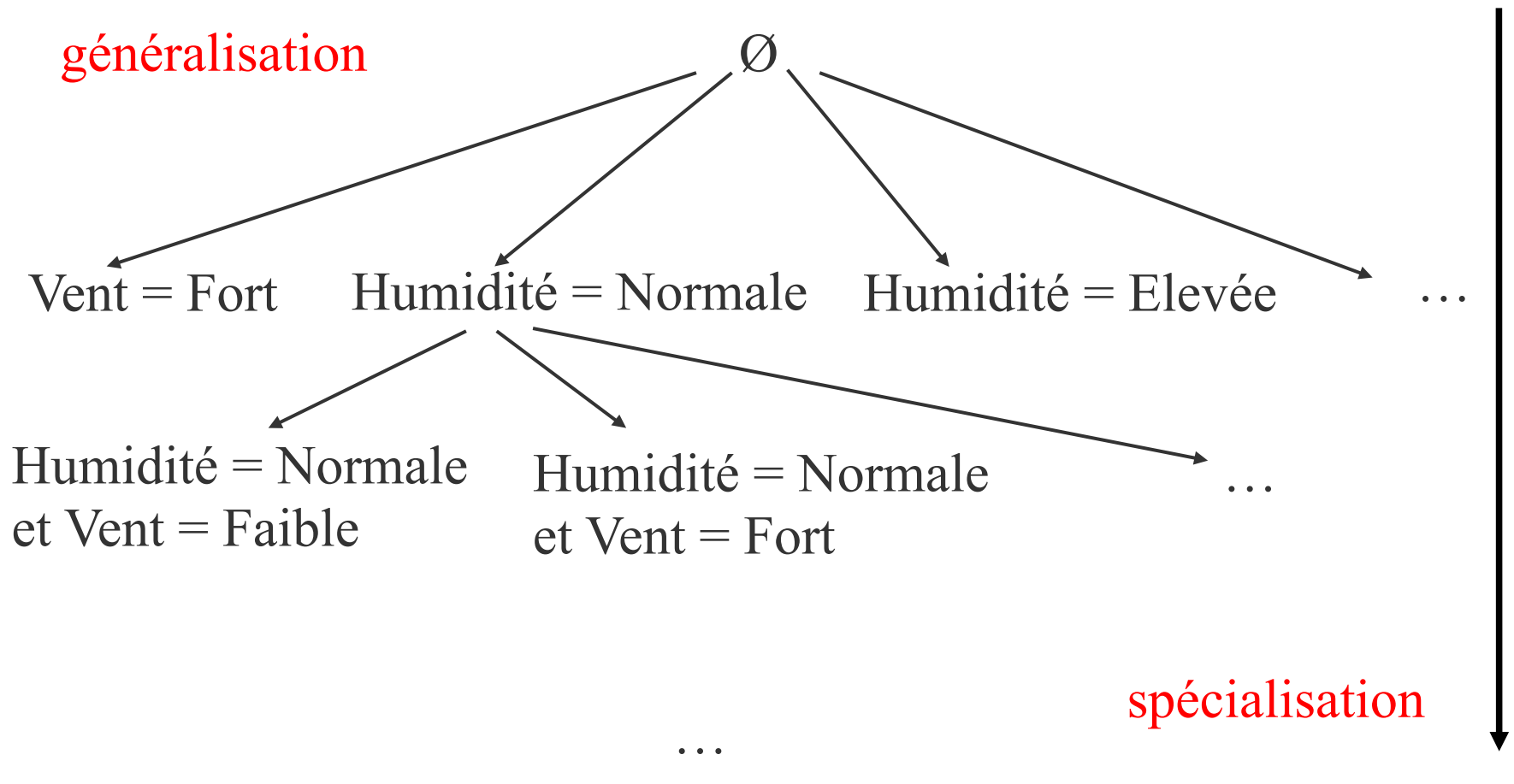
- Définir un ordre entre les hypothèses :
opérateur de subsomption
- On utilise cet ordre pour structurer l'espace

Question : En quoi cette structuration est-elle utile pour résoudre le problème ?

“Trouver

un ensemble de règles de H qui discrimine les exemples de E (les classe correctement en fonction des étiquettes)”

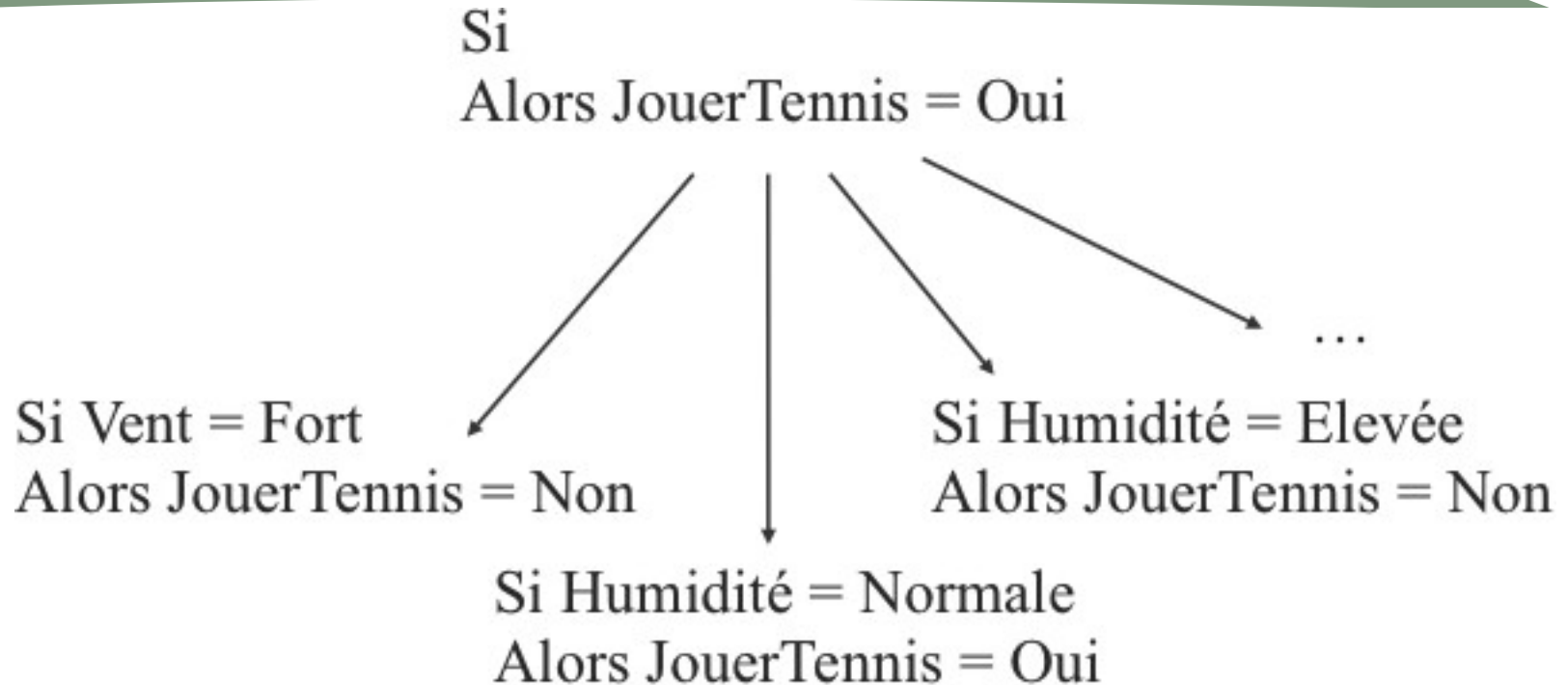
Généralisation/spécialisation



Structurer l'espace de recherche

- Définir un ordre entre les hypothèses :
opérateur de subsomption
- Parcours de l'espace :
 - Spécialiser une règle trop générale
 - Généraliser une règle trop spécifique

Espace des hypothèses



Expressivité du langage

- **Représentation attribut-valeur / logique du premier ordre**
- **Exemple 1 (attribut-valeur) :**
SI $(\text{pere}_1 = \text{Bob}) \wedge (\text{name}_2 = \text{Bob}) \wedge (\text{femme}_1 = \text{vrai})$
ALORS $\text{fille}_{12} = \text{vrai}$
- **Exemple 2 (logique du premier ordre) :**
SI $\text{pere}(Y, X) \wedge \text{femme}(Y)$
ALORS $\text{fille}(X, Y)$

Représentation attribut-valeur / logique du premier ordre

- Représentation attribut-valeur : impossibilité de représenter des relations entre des objets, des attributs

On ne manipule que

- ➔ un unique objet décrit par ses caractéristiques (ex. un champignon)
- ➔ ou un nombre fixe d'objets toujours soumis aux mêmes relations (ex. une molécule composée de trois atomes avec une liaison entre le 1er et le 2e et une liaison entre le 3e et le 4e)

Points de comparaison des systèmes d'apprentissage d'ensembles de règles

- Langage des hypothèses
- Stratégie d'apprentissage d'un ensemble de règles
 - Simultané / séquentiel
 - Une / plusieurs classe(s) à la fois
- Apprentissage d'une règle :
 - Stratégie de parcours de l'espace des hypothèses
 - Stratégie d'évaluation des hypothèses (de choix de la meilleure hypothèse)

Apprentissage d'un ensemble de règles



Nécessité d'apprendre un ensemble de règles

- Pour prédire plusieurs classes
- Car la représentation ne permet pas de résoudre le problème d'apprentissage avec une seule règle

Différentes stratégies

- Apprentissage simultané (arbres de décision, certains systèmes utilisant des algorithmes génétiques)
- Apprentissage séquentiel : algorithme de recouvrement (couverture) séquentiel

Problème d'apprentissage d'un concept

(cas où la fonction cible a 2 valeurs)

Etant donné

- un concept cible (une valeur donnée de la fonction cible)
- un ensemble d'exemples positifs P et négatifs N
- des connaissances du domaine (optionnel)
- un langage des hypothèses H

Trouver un ensemble de règles de H qui discrimine les exemples de P et N

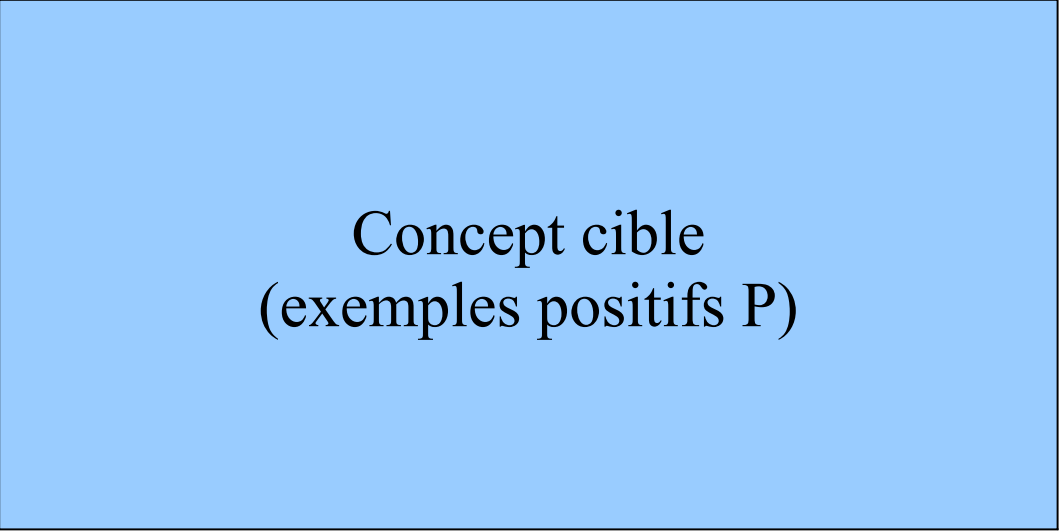
Algorithme de recouvrement séquentiel

- Apprendre une règle de haute précision mais de couverture quelconque
- Oter les exemples positifs couverts par cette règle
- Recommencer

Recherche impatiente : pas nécessairement le plus petit ni le meilleur ensemble de règles

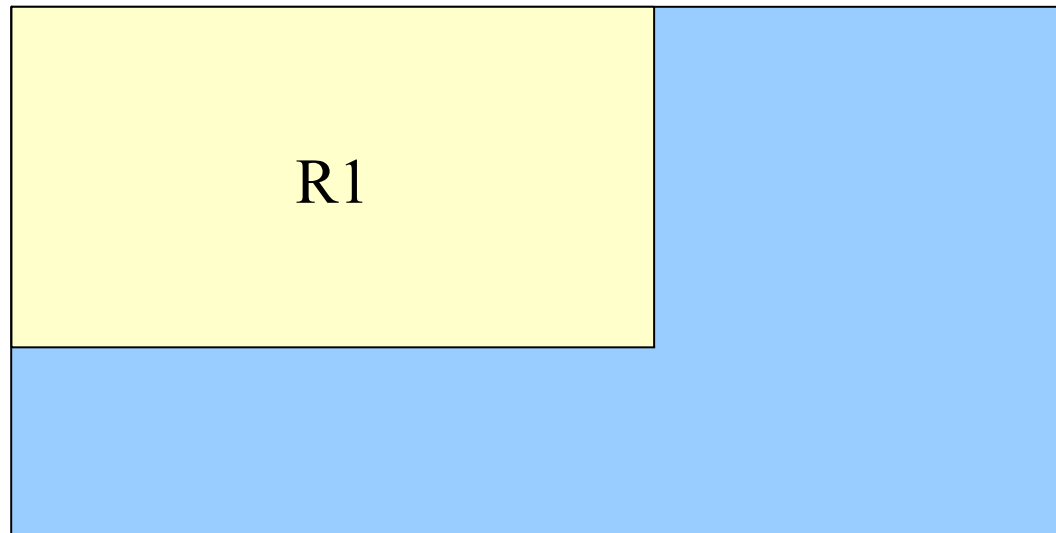
Espace des exemples

Exemples négatifs E

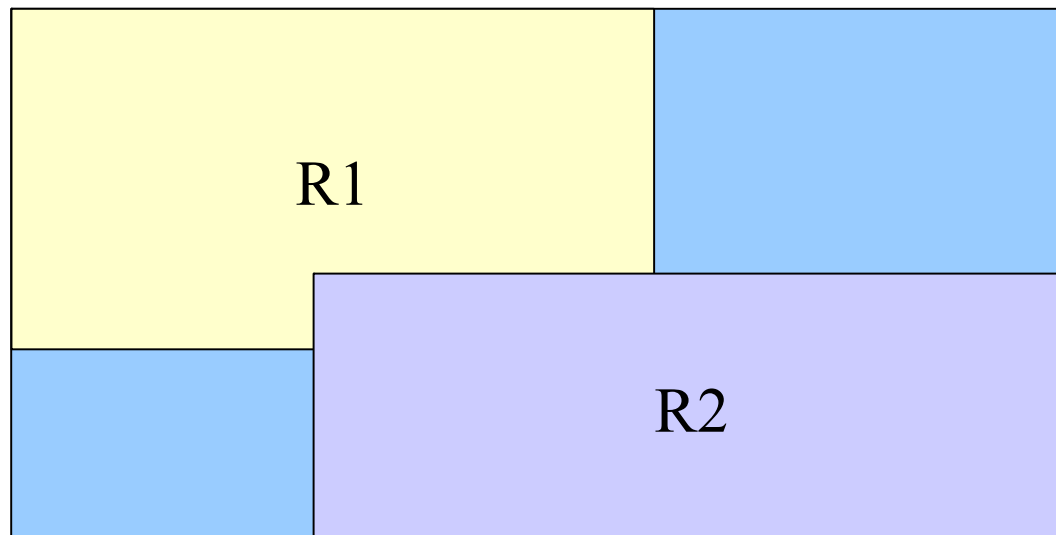


Concept cible
(exemples positifs P)

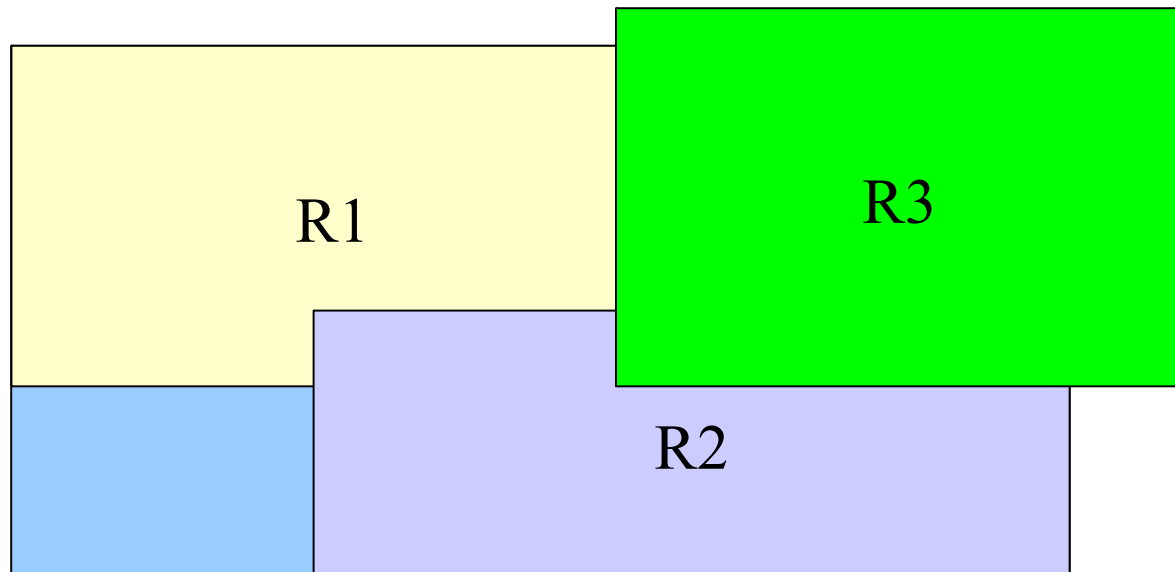
Espace des exemples



Espace des exemples



Espace des exemples



Apprentissage d'une règle



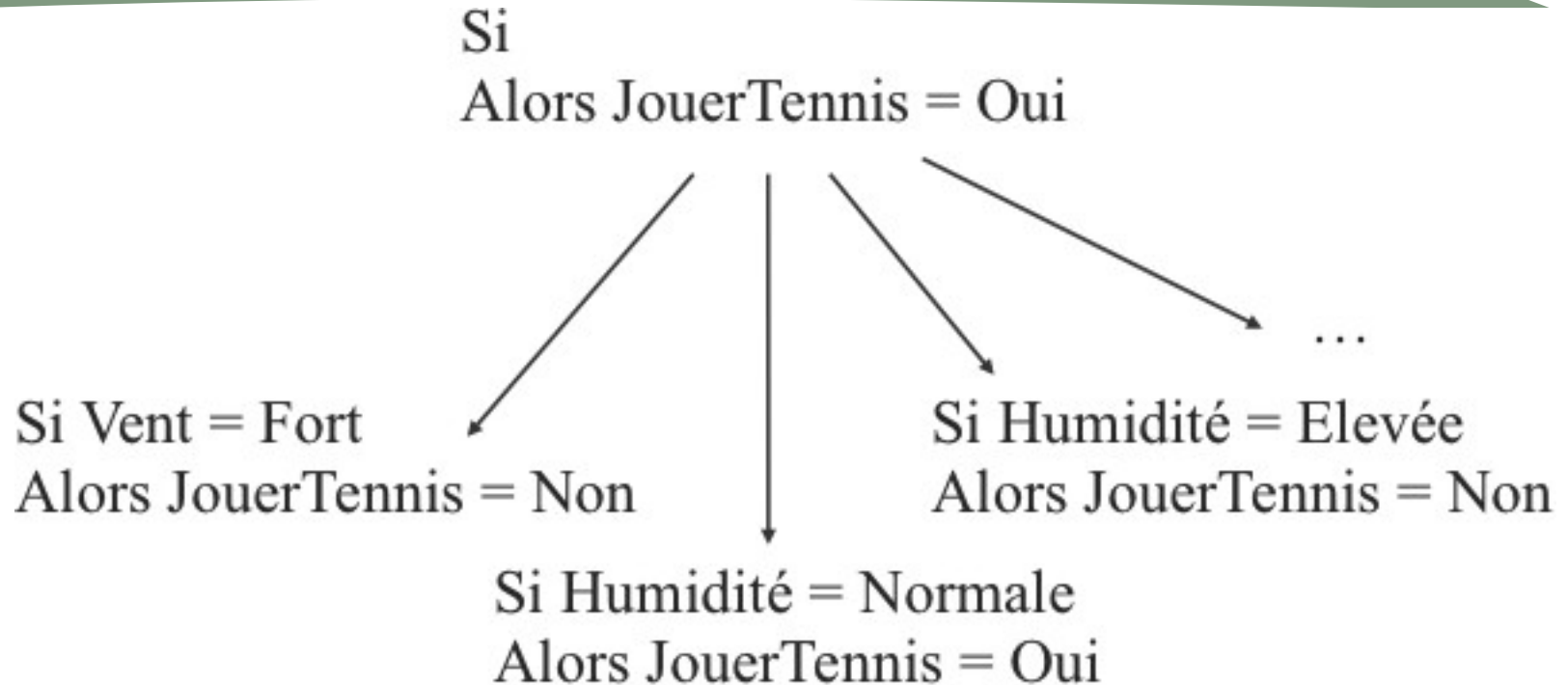
Méthode de recherche

- Hill-climbing : escalade de colline
- Beam search : recherche en faisceau
- Best-first search
- Recherche stochastique

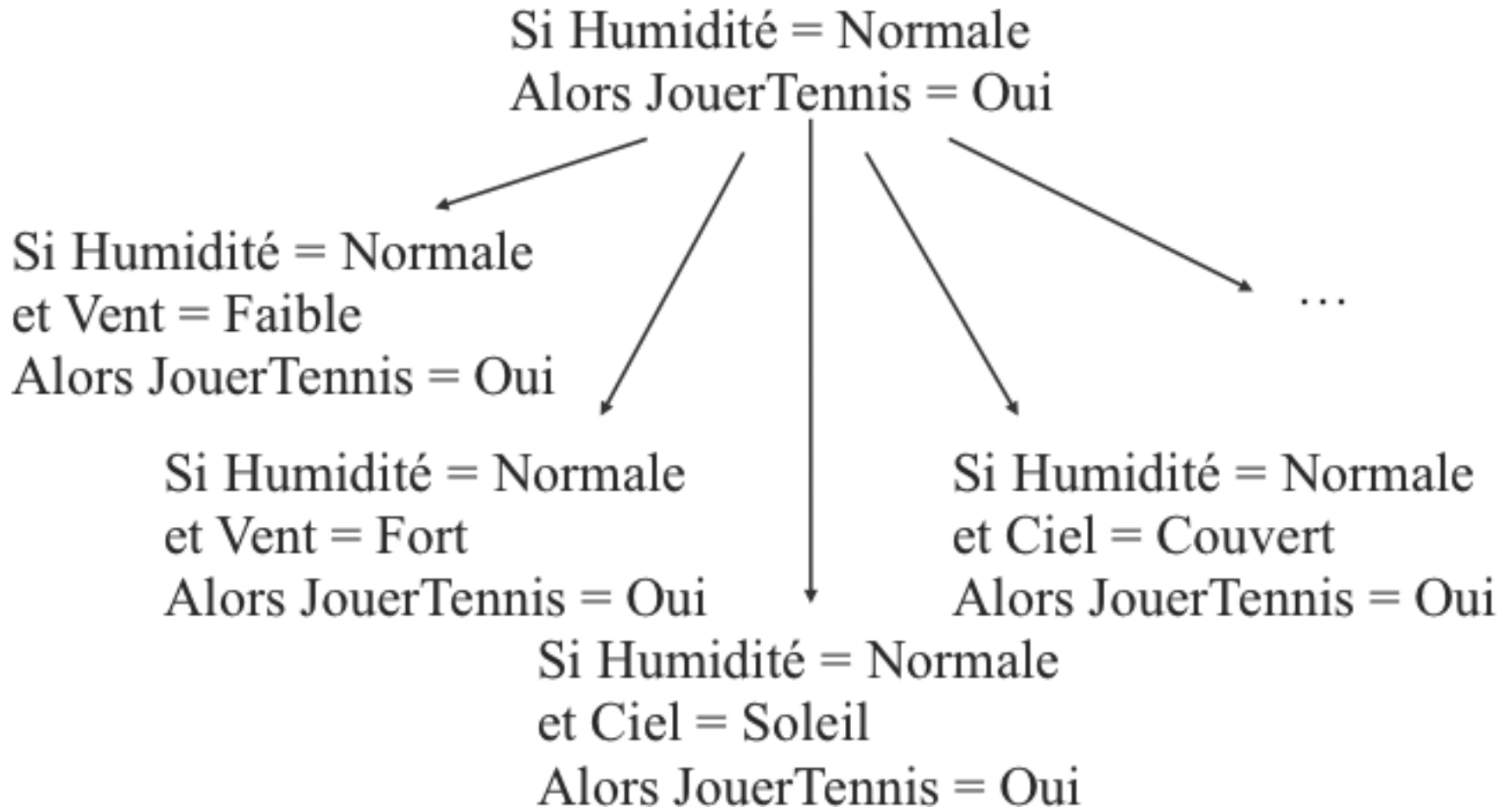
Direction de la recherche

- Descendant (du général au spécifique, top-down)
- Ascendant (du spécifique au général, bottom-up)
- Bidirectionnel

Recherche descendante



Recherche descendante



AQ

- “Algorithm quasi-optimal” : pour une solution quasi-optimale du problème de couverture = génère un nombre minimal ou presque de règles pour distinguer entre des exemples positifs et négatifs
- R. Michalski (1969)
- On the quasi-minimal solution of the general covering problem. In Proc. First International Symposium on Information Processing, pp. 125-128.
<http://www.mli.gmu.edu/papers/69-78/69-2.pdf>

AQ

- Représentation attribut-valeur
- Recherche ascendante
- Sensibilité au bruit

Principe d'apprentissage :

choisir un noyau

s'il couvre les exemples **FIN**

sinon, le faire croître jusqu'à ce qu'il les couvre

AQ

$G(p/N)$: règle la plus générale qui couvre p et rejette N

$i = 0$

Tant que P non vide faire

$i \leftarrow i+1$

Choisir un noyau e_i dans P

$R \leftarrow R \vee G(e_i / N)$

$P \leftarrow P - G(e_i / N)$

Fin

Sortie: R

Exemple AQ



Exemples	Att1	Att2	Att3	Classe
P1	Y	N	R	+
P2	X	M	R	+
P3	Y	N	S	+
P4	X	N	T	+
N1	X	M	S	-
N2	Y	M	T	-
N3	Y	N	T	-
N4	Z	N	T	-
N5	Z	N	R	-
N6	X	N	S	-

Exemple AQ : calcul de l'étoile

Etoile P1 = $G(P1/N) = \bigwedge_i G(P1/N_i)$

- $G(P1/N1) = (Att1 \neq X) \vee (Att2 \neq M) \vee (Att3 \neq S)$
- $G(P1/N2) = (Att2 \neq M) \vee (Att3 \neq T)$
- $G(P1/N3) = (Att3 \neq T)$
- $G(P1/N4) = (Att1 \neq Z) \vee (Att3 \neq T)$
- $G(P1/N5) = (Att1 \neq Z)$
- $G(P1/N6) = (Att1 \neq X) \vee (Att3 \neq S)$

Simplification : garder les règles les plus spécifiques :

$$G(P1/N) = (Att3 \neq T) \wedge (Att1 \neq Z) \wedge ((Att1 \neq X) \vee (Att3 \neq S))$$