Faire coopérer des agents hétérogènes par apprentissage de médiation

R. Charton, A. Boyer et F. Charpillet

MAIA - LORIA

Modèles Formels de l'Interaction (MFI'03) – Lille, France – 20-22 mai 2003

Contexte des travaux

Collaboration industrielle pour concevoir des services adaptatifs (multimédia, interactifs, grand public)

⇒ **Assistance à la recherche d'information**

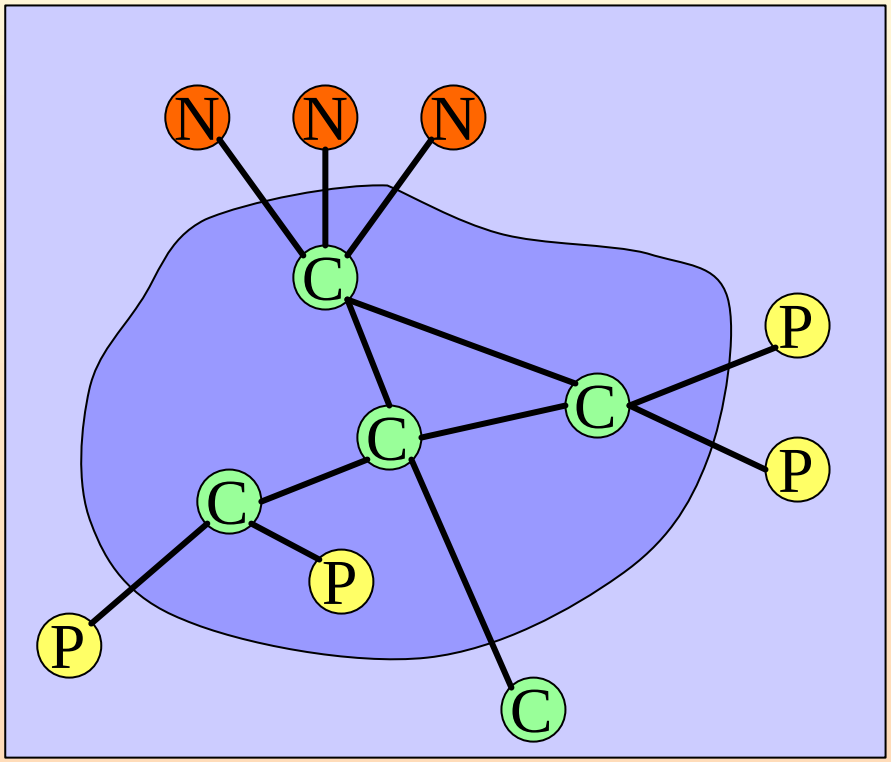
Contraintes :

- Utilisateur : occasionnel, novice
- Sources d'informations : propriété, coût

Objectif : Améliorer la qualité du service fourni

Coopération dans les systèmes multi-agents hétérogènes

Agents de natures différentes : humains, logiciels, robots, etc.



Agents Contrôlés

Agents Partiellement Contrôlés

Agents Non-Contrôlés

Environnement Virtuel

Environnement Physique

Liens d'Interaction

Comment faire coopérer ces agents?

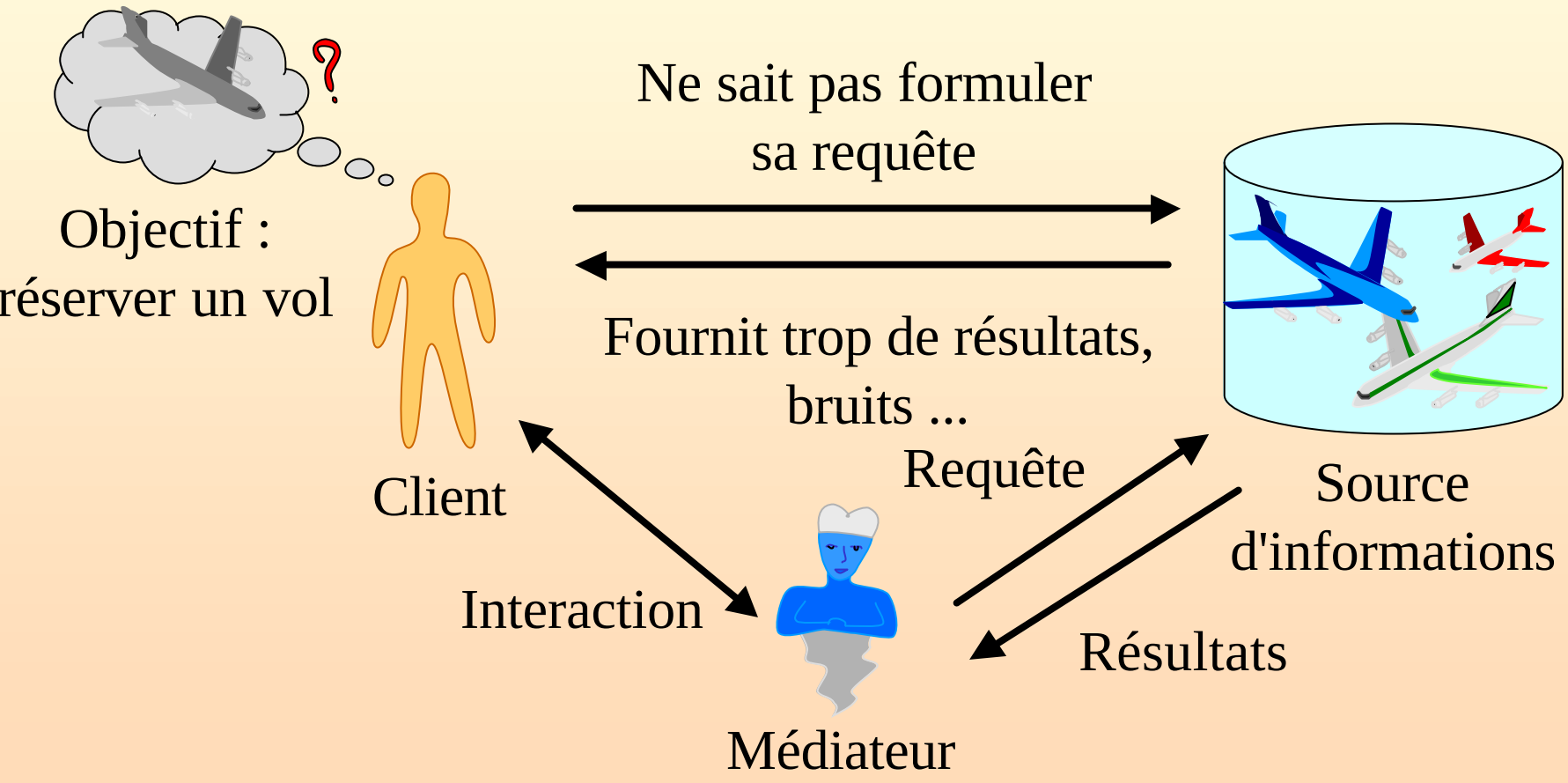
Réaliser ensemble des buts applicatifs qui satisfont un sous-ensemble des agents.

Plan de l'exposé

Faire coopérer des agents hétérogènes par apprentissage de médiation

- Exemple classique d'interaction
- Médiation à base de MDP
- Implantation du médiateur
- Expérimentation et résultat

Exemple de problème : le choix d'un vol



Notre objectif : **construire un médiateur**

Comment

- **Produire son comportement** ?
- **optimiser la qualité de service** ?

Sur l'exemple de choix de vol :

- Aider les utilisateurs à formuler des requêtes précises
- Trouver l'information pertinente

⇒ **Apprendre une stratégie de médiation**

(Trouver la meilleure séquence d'actions : questions à poser, requêtes, ...)

Caractéristiques du médiateur

Gérer :

- l'incertitude sur les connaissances imparfaites (agents, environnement)
- les ressources (RAP/SAP, BDD, e-Mail...)

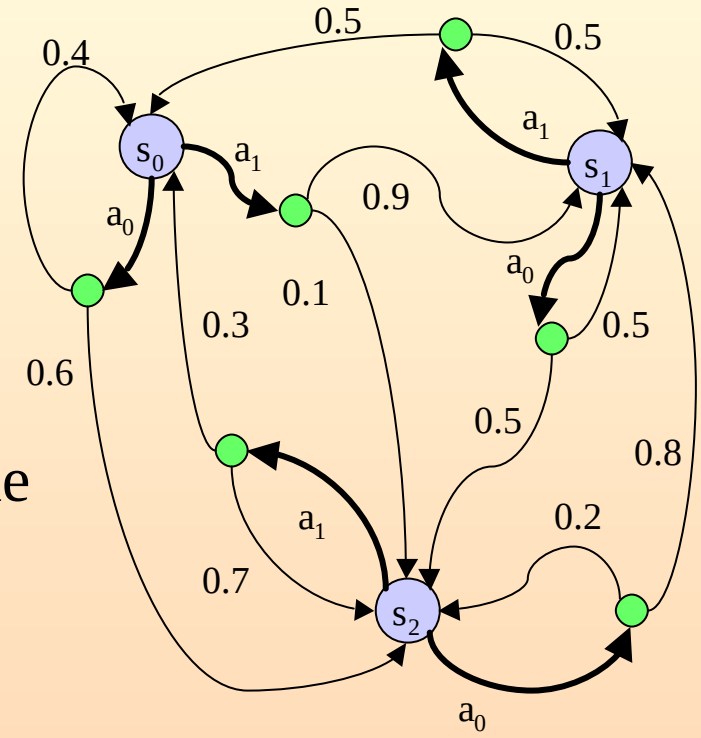
Tout ceci souligne le besoin d'un comportement adaptatif

$\Rightarrow$ Apprentissage Stochastique à base de MDP

\+ Modélisation probabiliste des agents et des ressources

Processus de Décision Markovien (MDP)

- Modèle Stochastique <S,A,T,R>
  - États $S=\{s_0,s_1,s_2\}$
  - Actions $A=\{a_0,a_1\}$
  - Transition $T : S \times A \times S \rightarrow [0;1]$ avec $T(s,a,s') = P(s'|s,a)$
  - Récompense $R : S \times A \times S \rightarrow IR$
- Prendre des décisions selon une politique $\pi : S \times A \rightarrow [0;1]$
- Optimiser la récompense espérée



Apprendre une stratégie de médiation

revient à **Apprendre une politique stochastique**

Modélisation sur l'exemple de choix de vol

Définir

- S : L'espace d'états
- A : Les actions du médiateur
- T : Les transitions
- R : Les récompenses

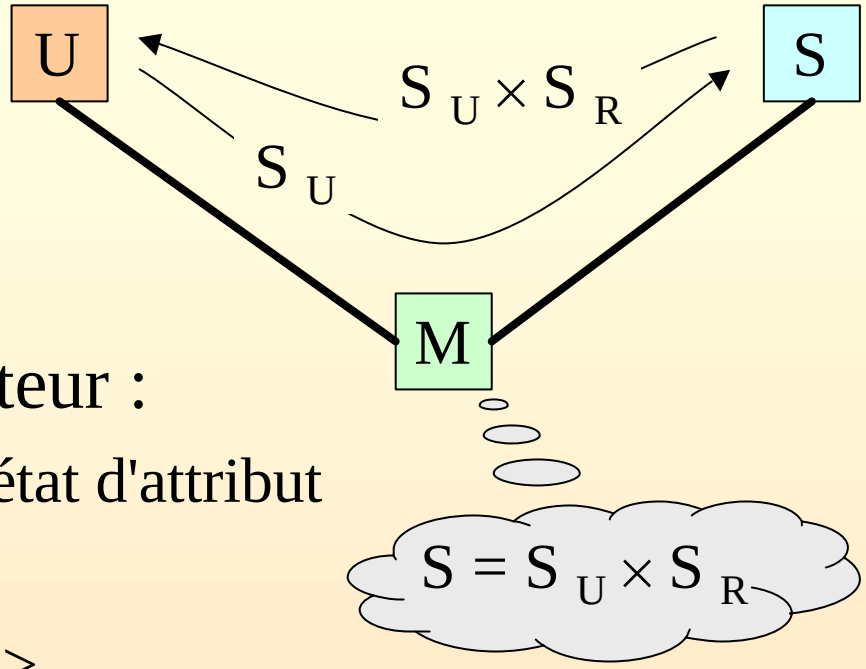
Etats : Décrire la tâche par des attributs

Les requêtes et les objets de la source et sont décrits par un ensemble d'attributs

Exemple de référentiel :

- Départ : { Londres, Genève, Paris, Berlin, … }
- Arrivée: { Pékin, Moscou, New-York, … }
- Classe : { Affaire, Normal, Économique, ... }

Espace d'états



Requête partielle de l'utilisateur :

$s_U = < ea_0, \ldots , ea_m >$ vecteur d'état d'attribut

Un état d'attribut $ea_i = < st_i , val_i >$

- Ouvert st = '?' val est libre
- Affecté st = 'A' val est affectée
- Fermée st = 'F' val est libre

Réponse de la source (pour la requête) :

$s_R = \{ vol_0; \ldots ; vol_r \}$

Abstraction de l'état

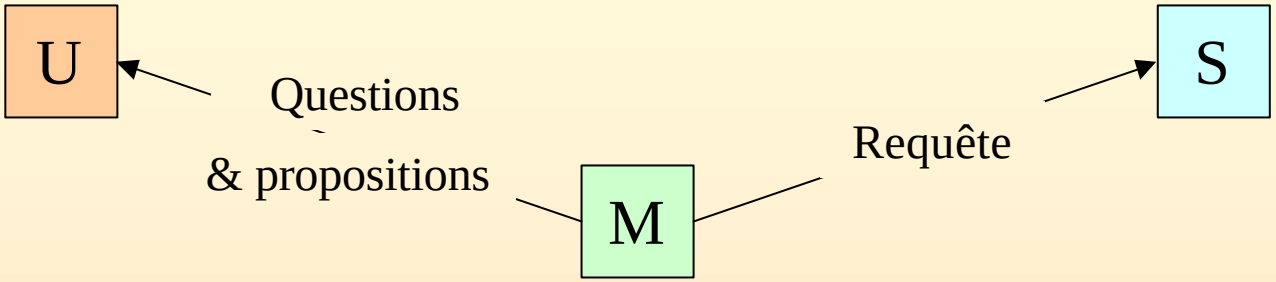
La taille de l'espace d'état est $2^n (2+i)^m$ **où**

- *n* : est le nombre total d'objets de la source d'information
- *m* : est le nombre d'attributs
- *i* : est le nombre moyen de valeurs par attribut

⇒ **Abstraction de l'état**

⇒ **Taille de l'espace d'états abstrait de : 4×3^m**

Actions du médiateur



Poser une question sur un attribut à l'utilisateur

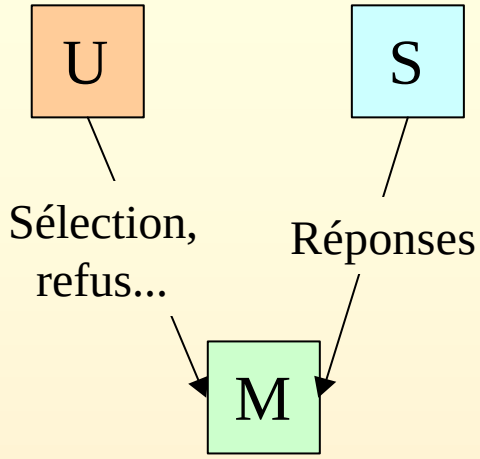
Exemple pour la classe de voyage :

- *“En quelle classe souhaitez-vous voyager ?”*
- *“Voulez-vous voyager en classe affaire ?”*
- *“Êtes-vous sûr de vouloir voyager en classe économique ?”*

+ interroger la source d'informations

+ proposer à l'utilisateur de sélectionner une réponse

Récompenses



Les récompenses sont obtenues…

- par l'interaction avec l'utilisateur
  - \+ $R_{selection}$ l'utilisateur sélectionne une proposition
  - \- $R_{noselect}$ l'utilisateur refuse toutes les propositions
  - \- $R_{timeout}$ l'interaction est trop longue
- par l'interaction avec la source d'informations
  - \+ R_{noresp} pas de réponses pour une requête pour une requête totalement spécifiée
  - \- $R_{overnum}$ trop de réponses

Calculer une stratégie

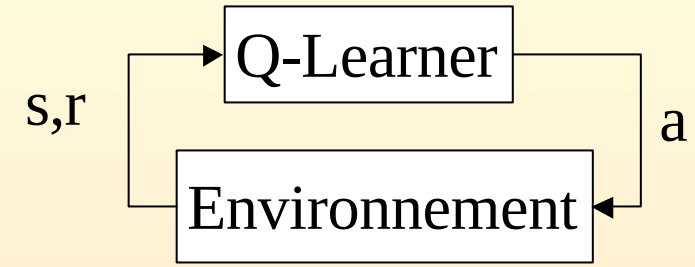
Le problème : Modèle inconnu !

- T = f (utilisateur, source d'informations)
- R = f (utilisateur, source d'informations)

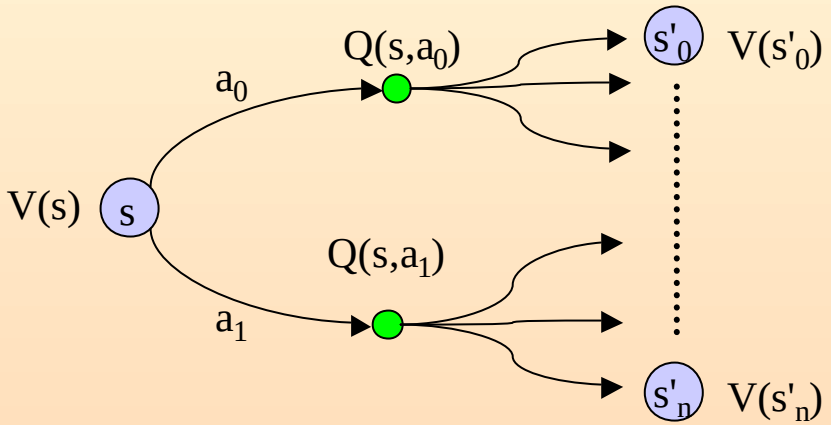
$\Rightarrow$ Apprentissage de la stratégie

Q-Learning *(Watkins 89)*

- Méthode d'Apprentissage par Renforcement
- Peut être utilisé "en ligne"



Q-Valeurs $Q : S \times A \rightarrow IR$

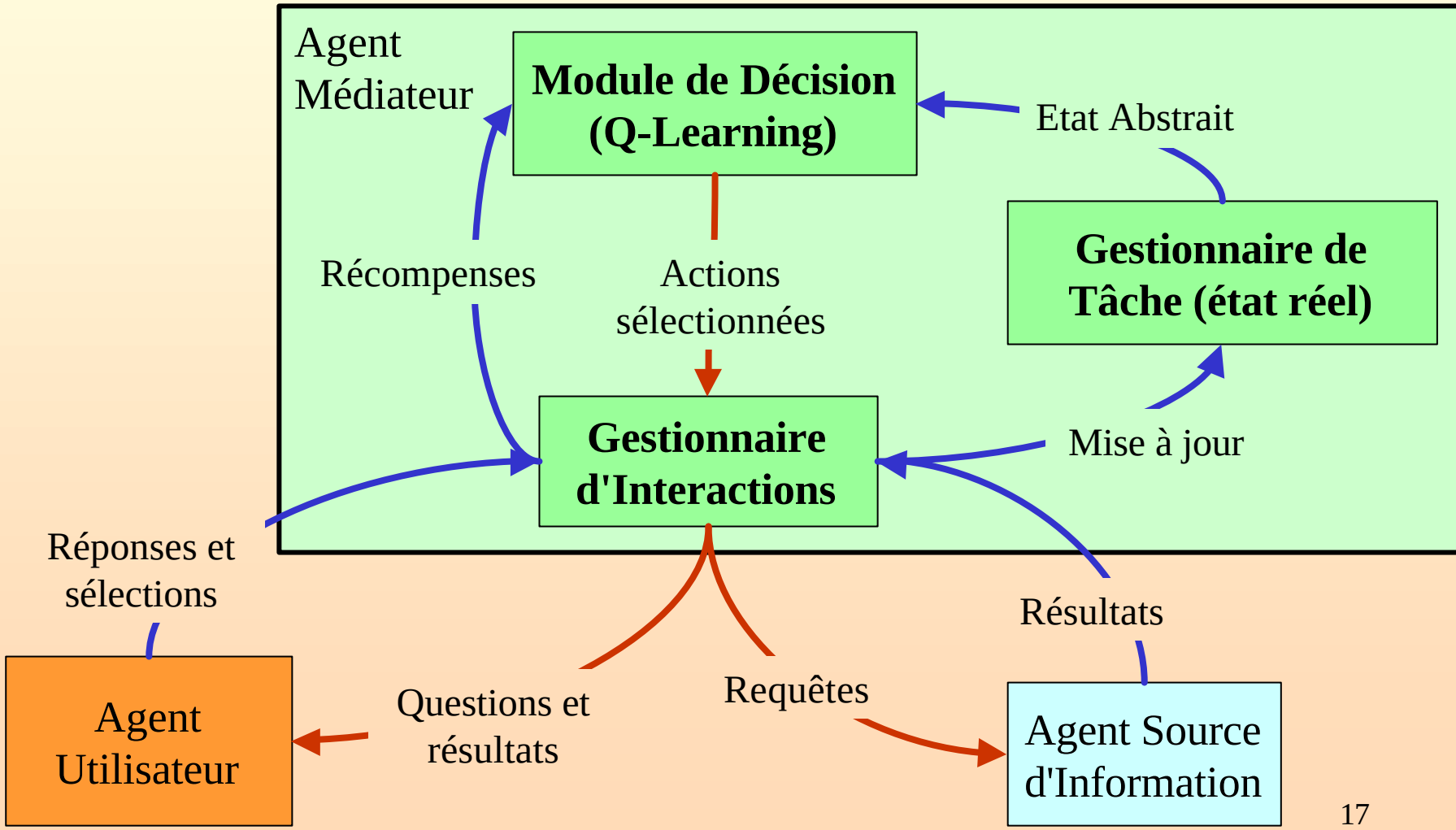


Mise à jour *(Bellman 57)*

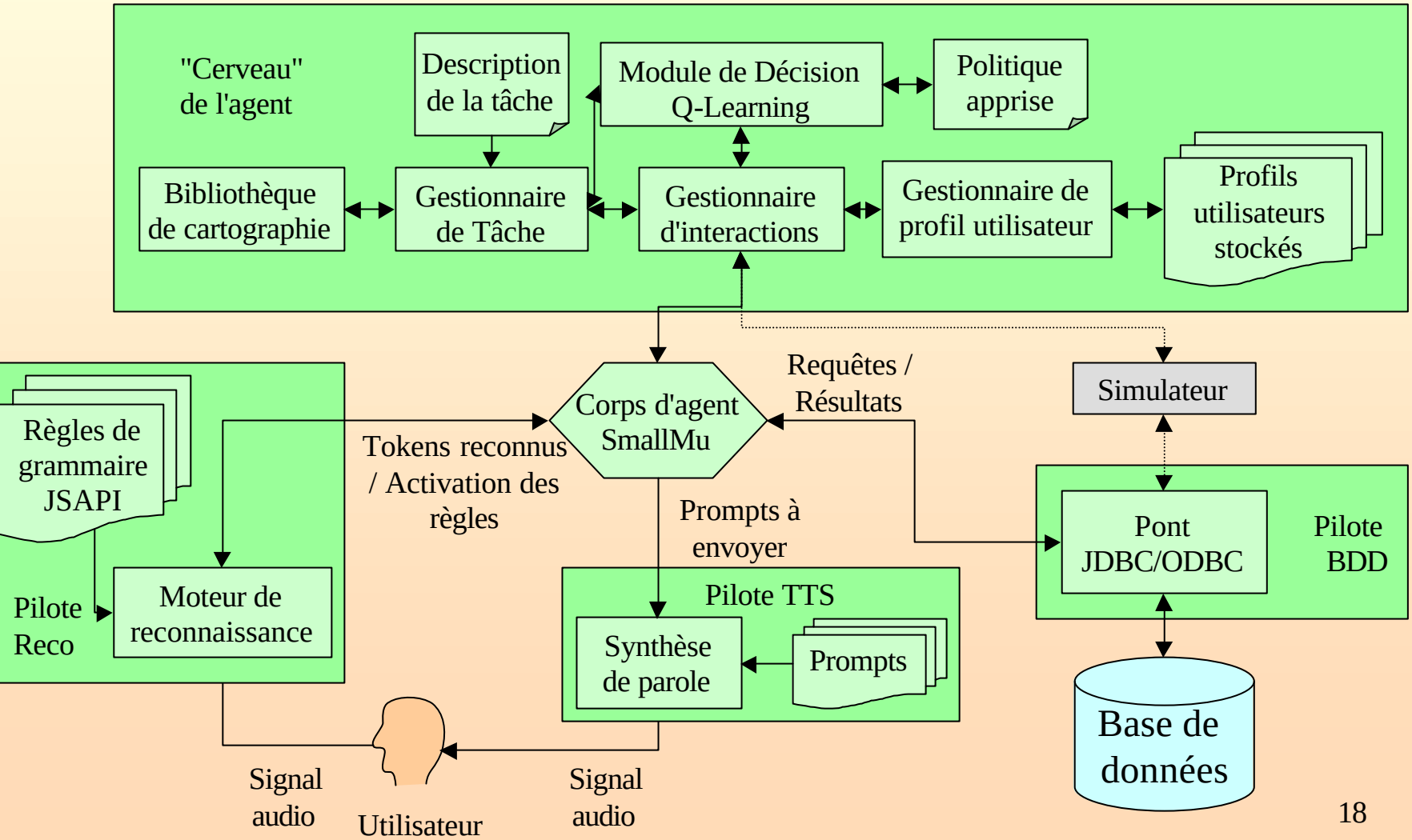
$$Q(s,a)_{t+1} = (1-\alpha)\ Q(s,a)_t + \alpha\left(R(s,a)_t + \gamma Max_{a' \in A} Q(s',a')_t\right)$$

α : Taux d'apprentissage

Architecture du médiateur

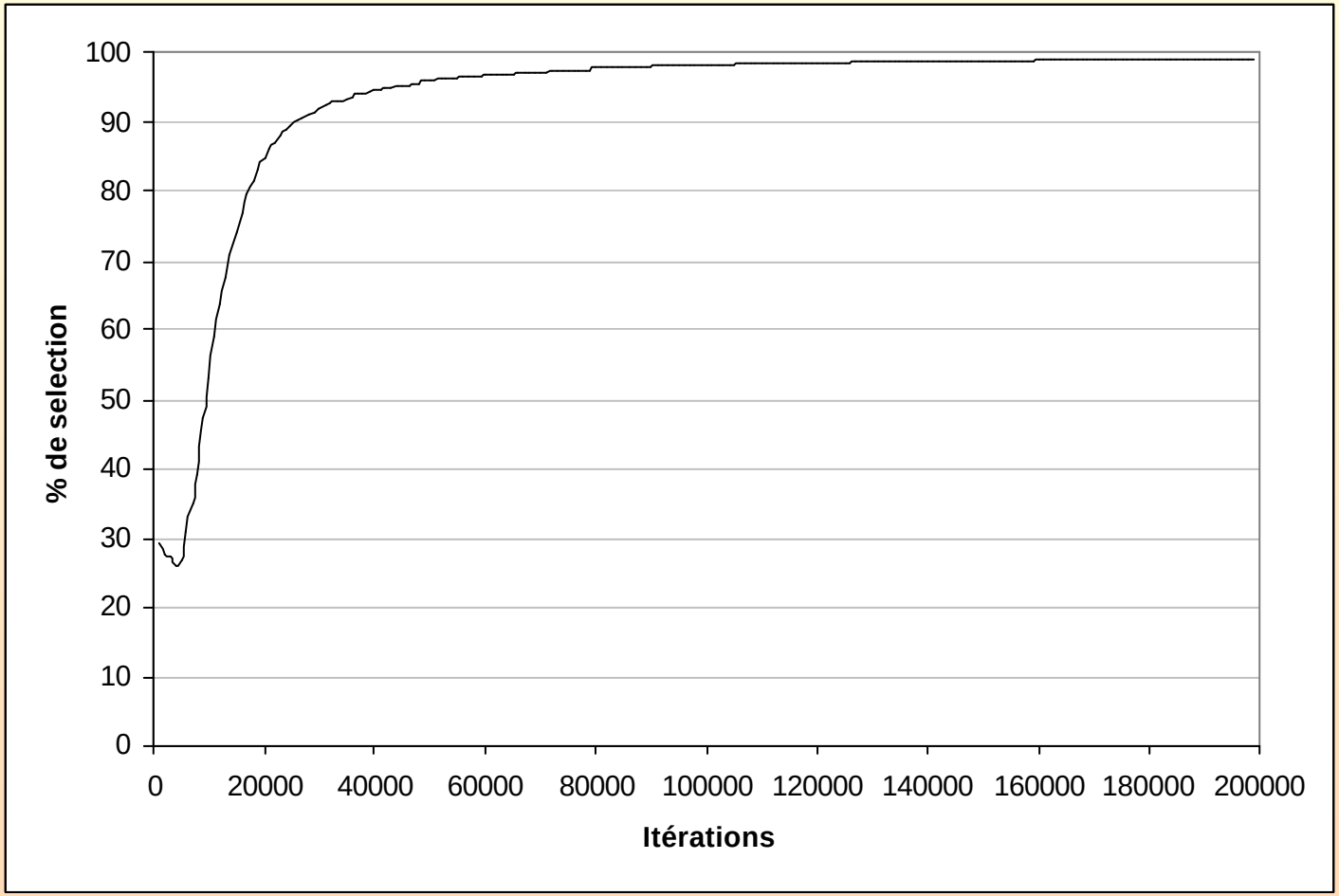


Implantation du médiateur



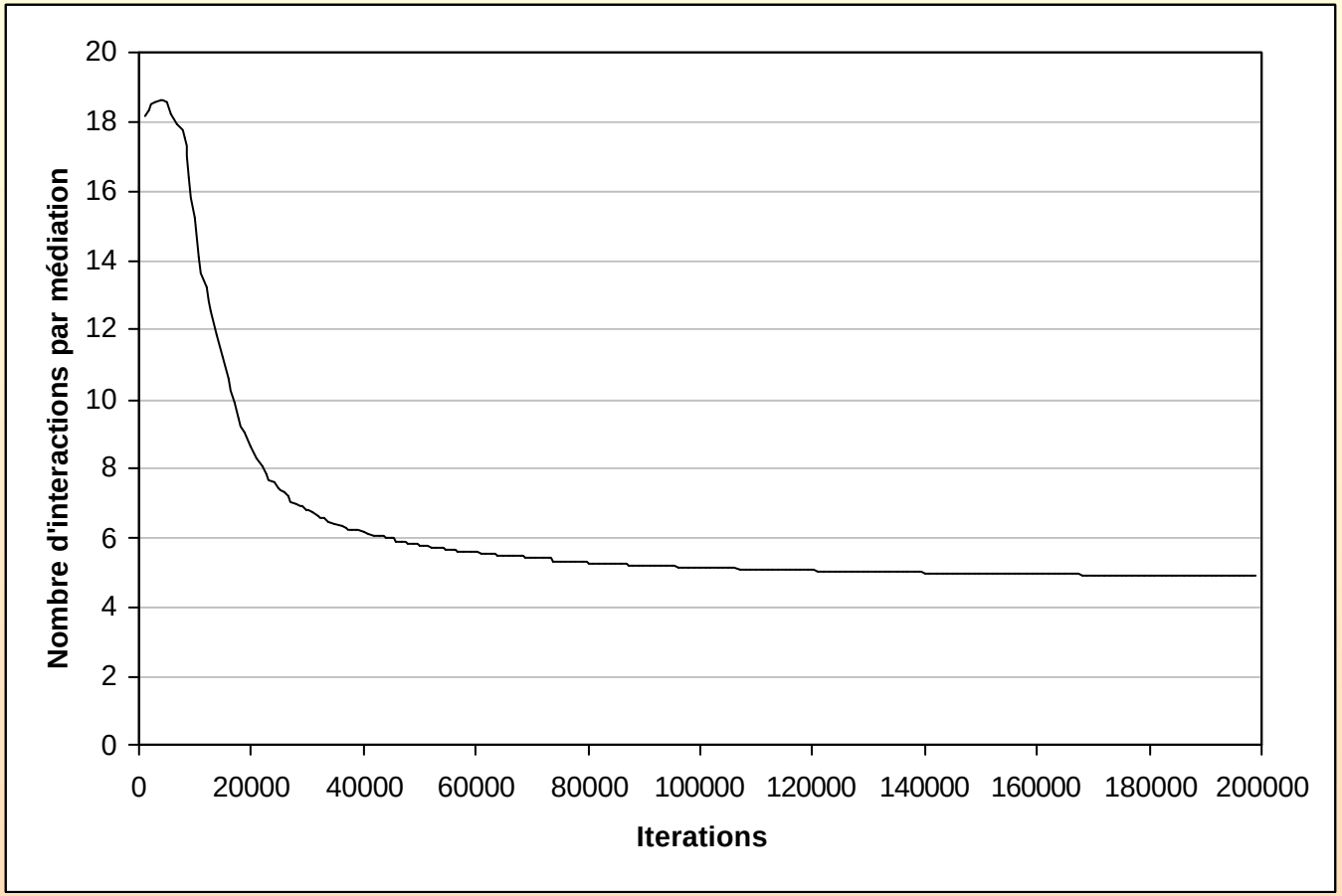
Résultats d'apprentissage pour 3 attributs

Application de choix de vol avec 3 attributs (départ, arrivée, classe)



Proche de l'optimal avec 99 % de sélection

Résultats d'apprentissage pour 3 attributs



5 actions pour atteindre le but

Pour 5 attributs : moins de sélections (80%) et des interactions plus longues

Conclusion

Apports

- MDP+AR permet d'apprendre des stratégies de médiation
- Répondre aux besoins du plus grand nombre (profils)
- Orienté concepteur $\Rightarrow$ orienté utilisateur
- Approche incrémentale
- Solution implantée

Limites

- L'utilisateur est partiellement observable, surtout si on utilise des capteurs comme la reconnaissance de parole
- Baisse des performances pour des tâches plus complexes

Travaux futurs

- Utiliser d'autres modèles/méthodes probabilistes :
  - Apprendre à partir de politique pré-établies
  - Apprendre le modèle (Sutton DynaQ, Classifieurs)
  - Approche POMDP (Q-learning modifié, Gradient de Baxter)
- Pour des tâches plus complexes
  - Décomposition Hiérarchique (H-MPD et H-POMDP) en gérant les dépendances entre les attributs

    (ex : Ville → Compagnies possibles → Options spécifiques)
  - Passage à l'échelle :

    Reformuler l'espace d'états abstrait pour mieux guider le processus dans l'espace réel

Rôles et Classes de Service

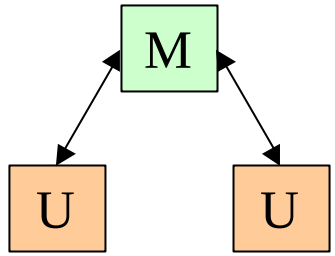
U Utilisateur

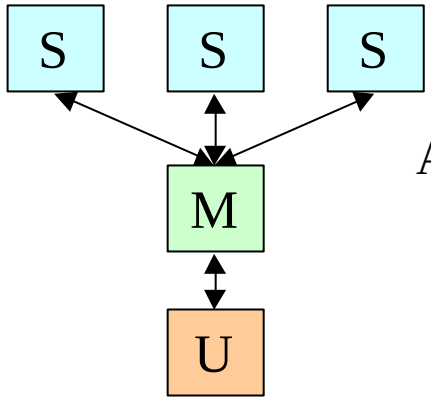
M Médiateur

S Source

T Traitement

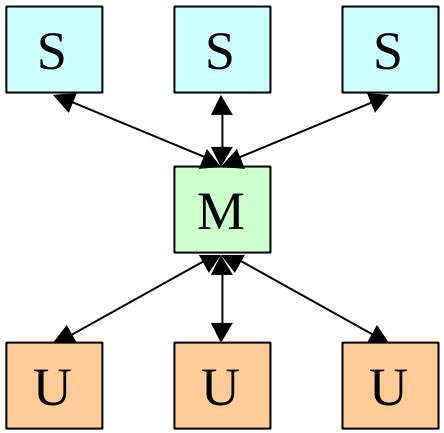
Médiation Simple



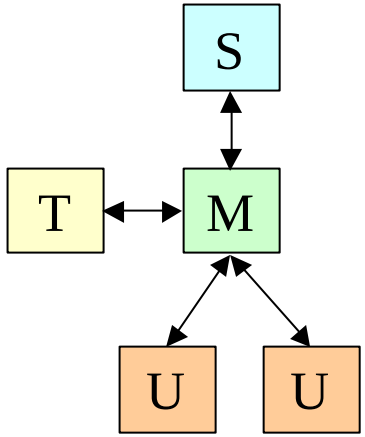


Assistance à la navigation

Télé-Réunion



Traitement intelligent de l'information



Diagnostic, filtrage ...

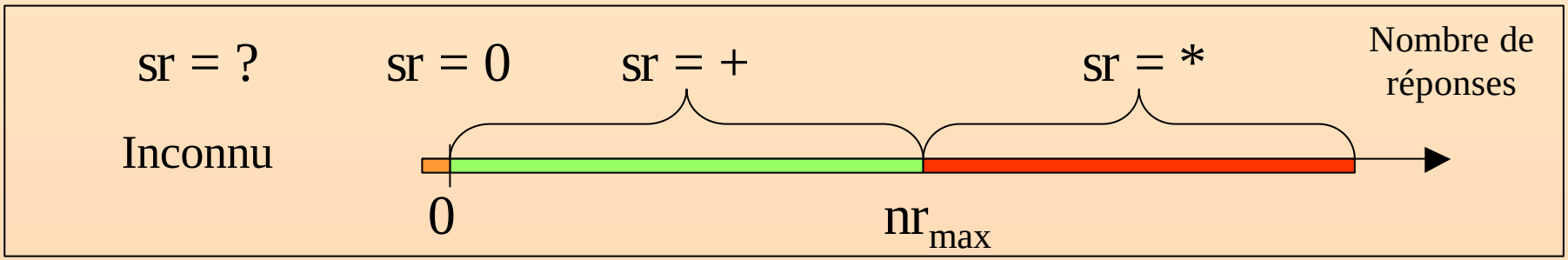
Abstraction de l'état

La taille de l'espace d'état est $2^{n} (2+i)^{m}$ où

- *n* : est le nombre total d'objets de la source d'information
- *m* : est le nombre d'attributs
- *i* : est le nombre moyen de valeurs par attribut

⇒ Abstraction de l'état

- Conserver seulement *st* ∈ {?, A, F} dans l'état d'attribut
- Qualifier la réponse de la source qr = {?, 0, +, *} selon le nombre de résultats :



⇒ Taille de l'espace d'états abstrait de : 4×3^{m}

Q-Learning - Algorithme

- Algorithme succinct

Utilisateurs occasionnels

Satisfaire le plus grand nombre

Sinon + fidèles alors gestion de profils