TD(0), TD(n), Monte Carlo et SARSA dans la Vraie Vie : Exemples Concrets (Fraude, Cyber, Décision Humaine)

1 Introduction

Les méthodes TD(0), TD(n), Monte Carlo, SARSA et Q-Learning sont souvent décrites en termes mathématiques abstraits. Ici, nous montrons comment ces modèles correspondent à des situations réelles : fraude bancaire, cybersécurité, conduite, prévision météo, relations humaines et prise de décision générale.

L'idée centrale : selon la situation, nous devons prendre une décision **immédiatement**, **après quelques étapes** ou **à la fin d'une séquence**. C'est exactement la différence entre TD(0), TD(1–3), TD(n) et Monte Carlo.

2 1. Détection de fraude (banque, carte de crédit)

2.1 TD(0) : décision immédiate

Objectif : bloquer le danger tout de suite, même avec information partielle.

Exemples :

— Transaction de 3000$ à Tokyo alors que l'utilisateur est à Montréal : le système gèle la carte immédiatement.

— Tentative de paiement via une IP jamais vue ou un pays très risqué.

Le système met à jour son estimation du risque après **1 seul événement**. Il agit immédiatement.

Analogie RL : TD(0) = décision en fonction du prochain état uniquement.

2.2 TD(1–3) : attendre 2–3 événements

Exemple concret :

— 1re transaction suspecte : drapeau mais pas de blocage.

— 2e transaction similaire : score de risque augmente.

— 3e transaction : blocage automatique.

On attend quelques étapes avant de conclure à la fraude. C'est un compromis entre réactivité et réduction des faux positifs.

Analogie RL : TD(1), TD(2), TD(3).

2.3 Monte Carlo : analyse complète

Dans les équipes de conformité/audit :

— on analyse 6 mois d'historique,

— on reconstruit les patterns complets,

— on conclut "fraude avérée" après toute la séquence.

Monte Carlo = décision uniquement après la fin de la séquence.

3 Attaques sur un serveur (cyber)

3.1 TD(0) : IDS/IPS en temps réel

Exemples :
— Snort/Suricata/WAF bloque une seule requête malveillante.
— 100 tentatives de login en 5 secondes ⇒ bannissement immédiat.

Un exploit peut compromettre un serveur en **une seule requête**. Réaction instantanée obligatoire.

C'est du TD(0) pur : décision sur l'instant.

3.2 TD(1–3) : corrélation légère d'événements

Exemple :

1. Scan de ports,
2. Brute force SSH,
3. Requête anormale sur /wp-login.php.

Actions possibles :
— log au premier,
— alerte au deuxième,
— blocage au troisième.

TD(2)–TD(3) = décision après 2–3 étapes observées.

3.3 TD(n) : kill-chain longue

Un kill-chain MITRE ATT&CK typique :

1. Reconnaissance,
2. Phishing,
3. Accès VPN volé,
4. Mouvement latéral,
5. Exfiltration des données.

On choisit d'intervenir après n étapes clefs (souvent 3 ou 4). C'est une logique de TD(n).

3.4 Monte Carlo : forensic / post-mortem

Après l'incident :
— analyse complète,
— reconstitution de timeline,
— rapport final (assurance, direction, judiciaire).

Monte Carlo = décision post-incident complète.

4 SARSA dans les environnements de sécurité

SARSA est on-policy : il apprend la politique réellement suivie.
Utile pour :
— systèmes d'auto-blocage qui doivent rester prudents,

— environnements où un excès d'agressivité cause des faux positifs,
— filtres dynamiques sensible à la stabilité du service.

SARSA = comportement prudent. Q-Learning = comportement agressif optimal.

5 Exemples humains (vie quotidienne)

5.1 TD(0) : décision immédiate

— Conduire : un piéton apparaît ⇒ freinage immédiat.
— Météo instantanée : une goutte ⇒ parapluie.
— Urgence médicale : personne inconsciente ⇒ action immédiate.

5.2 TD(1–3) : attendre 2–3 étapes

— Choisir un film : bande-annonce + premières minutes.
— Juger le comportement d'une personne : 2–3 interactions.
— Jauger un restaurant : accueil + odeur + carte.

5.3 Monte Carlo : attendre la fin

— Évaluer un étudiant : verdict en fin de session.
— Juger une relation : après plusieurs années.
— Bilan financier : à la fin de l'année.

6 Tableau récapitulatif

Méthode	**Quand l'utiliser ?**
TD(0)	Décision immédiate : fraude instantanée, IDS/IPS, conduite, urgence.
TD(1–3)	Décision après quelques signaux : corrélation légère, jugement humain court terme.
TD(n)	Scénarios multi-étapes : kill-chain cyber, prévision météo 3–5 jours.
Monte Carlo	Décision à la fin : forensic, audit, évaluation complète.
SARSA	Stratégie prudente : sécurité sensible, éviter sur-blocage.
Q-Learning	Stratégie optimale agressive : trading, jeux compétitifs.

7 Conclusion

Les méthodes TD(0), TD(n), Monte Carlo et SARSA ne sont pas abstraites : elles correspondent directement à des façons humaines ou techniques de prendre des décisions. La clé est de savoir si la situation exige une décision immédiate, une décision après quelques étapes, ou une décision basée sur l'ensemble complet des événements.