POLITECNICO DI TORINO

Corso di Laurea
in Matematica per l'Ingegneria

Tesi di Laurea

Reinforcement Learning per apprendere strategie di recupero crediti su NPL (Non-Performing Loans) e UTP (Unlikely To Pay)

Relatori
prof. Paolo Brandimarte

Candidato
Michele Vioglio

Anno Accademico 2022-2023

Sommario

Il recupero crediti su posizioni bancarie deteriorate è un tema economico-sociale diventato di grande importanza negli ultimi anni, soprattutto per la grande mole di rapporti con problemi di insolvenza che le banche si sono trovate nelle proprie passività. In questa tesi si cerca di trovare un metodo per suggerire alle società che gestiscono questi crediti, ricevuti dalle banche stesse, quali siano le migliori azioni da intraprendere su una pratica. Per fare ciò si è utilizzato un algoritmo di reinforcement learning (RL), caratterizzando le pratiche in alcune variabili discrete. I dati per supportare lo studio sono stati estrapolati da un database fornito da AMCO (Asset Management COmpany), una compagnia di gestione di asset finanziari. In particolare, lo stato di ogni pratica nasce dalla combinazione di dati provenienti da tabelle diverse, spesso comportando discretizzazione di valori originariamente nel continuo. Le azioni del RL rappresentano le strategie di recupero crediti più comuni o anche altre azioni salienti degli agenti. Si è fatto uso di un modello off-policy tabulare di RL, noto come *Q-Learning*, su una matrice di stati e azioni molto ampia. A tal proposito, una variante detta *smooth Q-Learning* viene utilizzata al fine di popolare maggiormente la suddetta matrice sfruttando la similarità tra stati coinvolti. È stato svolto un lavoro di tuning per gli iper-parametri più importanti al fine di garantire la massima performance all'apprendimento. Si è concluso che una policy viene appresa per le tipologie più frequenti di pratica, sfruttando la storia di pratiche già concluse o in corso di evoluzione. Per la visualizzazione dei risultati e il suggerimento di azioni su uno stato della pratica scelto dall'utente, é stato sviluppato un prototipo con la libreria Streamlit di Python.

Abstract

Credit recovery on banking exposures is a socio-economic theme that has grown in magnitude in the last years. That's because of the wide number of contracts suffering from insolvency that italian banks have in their liabilities. In this thesis, I try to figure out a way to suggest which are the best actions to be chosen on a file by a credit recovery agent. Reinforcement Learning (RL) is implemented for this purpose, characterizing files into discrete variables. Data are drawn from AMCO's database, where AMCO stays for Asset Management COmpany. Specifically, the state of each file comes from the combination of data from different tables, often binning originally continous variables. Strategies are composition of actions that can be taken to recover a credit. The model is an off-policy tabular Q-Learning. The Q-matrix is a wide one. A variant called smooth Q-Learning is applied to fill heavily the matrix exlpoiting similarity between states. Hyper-parameters tuning drove the model to more efficiency in learning. In conclusion, a policy is learned for most frequent states, exploiting the experience coming from on going files and archived ones. A Streamlit prototype had been developed to show results and suggest best actions to agents.

Indice

Di questi tempi, un conto è avere un credito, un altro è farselo pagare.

[Bud Spencer, I quattro dell'Ave Maria]

Capitolo 1

Introduzione generale

L'ambientazione di questo lavoro di tesi è economico-sociale. Negli ultimi anni, le banche hanno iniziato ad accumulare sempre più crediti deteriorati tra le loro passività, portando le stesse ad adottare decisioni più conservative in fase di bilancio, dal momento che non possono permettersi di avere troppi crediti inesigibili pendenti, ma che devono cercare in ogni modo di trarre la massima liquidità.

Da essi nascono quindi le società di asset management, le quali acquisiscono dalle banche portafogli di pratiche ad un prezzo che é spesso molto inferiore rispetto a quanto la banca avrebbe il diritto di recuperare da quei clienti, ma che è comunque a lei favorevole ai fini di togliere dalle passività i propri crediti inesigibili, o comunque generare liquidità immediata dalla vendita di questi asset finanziari.

Una delle più importanti società in Italia in questo campo è AMCO (Asset Management COmpany). Con l'ausilio di RAD Informatica, AMCO ha fornito un database di alimentazione a ORS Group (Operational Research System), per il lavoro di tesi in questione. Società come AMCO ricevono una mole considerevole di posizioni bancarie deteriorate, i cui intestatari risultano incapaci di portare a termine autonomamente la restituzione di denaro alla banca. Tali clienti rendono necessario da parte del creditore muoversi, in svariati modi, per tentare di recuperare almeno parte del credito del quale detiene il diritto.

Con l'aumento del numero di pratiche, e quindi di dati, a cui possono attingere le società di gestione dei crediti, cresce anche la possibilità di inserire questa branca dell'economia nel campo dei Big Data e di tutte le applicazioni ad essi riferite. In relazione al database di cui si dispone è possibile infatti avere centinaia di migliaia di pratiche da cui estrapolare conoscenza, applicando i classici algoritmi di analisi dei dati, ma anche guardare a metodi più innovativi.

In particolare, si è deciso oculatamente di utilizzare una tecnica di apprendimento nota come reinforcement learning (RL): è una tecnica basata sulla sequenzialità di azioni applicate ad un certo oggetto, in questo caso una pratica di recupero crediti, con lo scopo di raggiungere il maggior guadagno nel minor tempo possibile. Aspetti finanziari come i tassi di interesse per scontare i flussi di cassa entrano in gioco per dare al problema una motivazione logica. Il fine ultimo è quello di garantire alla società di recupero un aumento degli incassi dai propri clienti. Nel corpo di questo elaborato, si vedrà come tutti

Figura 1.1: Logo AMCO

gli elementi peculiari del RL vengano analizzati e inglobati nello studio più strettamente economico delle metriche. Esse caratterizzano nel miglior modo possibile le pratiche e ne danno un punto di partenza su cui costruire un modello decisionale, basato su azioni significative che possono essere prese sul campo da agenti incaricati da AMCO o da altre società preposte al recupero del credito.

Capitolo 2

Reinforcement learning

2.1 Elementi fondamentali

Il reinforcement learning (RL) è una branca del machine learning che si basa su alcune entità indispensabili per una corretta definizione degli algoritmi da utilizzare in questo ambito.

1. **Agente**: è il soggetto del RL che ha a disposizione un set di azioni che può intraprendere sull'ambiente. L'agente apprende dalle sue stesse azioni e dalla risposta dell'ambiente come comportarsi nel prendere le sue prossime decisioni, con lo scopo di massimizzare il rendimento a lungo termine, o di raggiungere un obiettivo nel più breve tempo possibile. Questo metodo di agire è noto come *trial and error*.

2. **Ambiente**: é il mondo in cui il processo di apprendimento avviene. Gli esempi più classici di ambiente sono quelli delle griglie bidimensionali in cui si può muovere un certo oggetto, con l'obiettivo di raggiungere un certo punto target. L'ambiente può essere anche qualcosa di più complicato, con molte dimensioni o di cui non si conosce la struttura, come ad esempio un organismo vivente, senza una forma regolare. L'ambiente riceve le azioni dell'agente e fornisce un reward in risposta, insieme ad una nuova istanza dello stato del sistema o oggetto che si sta considerando. Tuttavia, esistono problemi in cui non è possibile osservare l'intero ambiente, ma solo una parte di esso: si parla in questo frangente di *partially observable environment*(Loh and Raginsky [2022]). Un classico esempio è quello dei giochi di carte, in cui un giocatore vede le proprie carte, ma non può conoscere il resto dell'ambiente, cioè le carte degli avversari.

3. **Stato del sistema**: rappresenta la conoscenza dell'agente riguardo alla situazione attuale. Può essere la posizione di una pallina su una griglia, o dell'agente stesso in un labirinto, o il livello di emoglobina in un paziente. Non esiste un limite dimensionale o di forma; le variabili possono essere discrete, quindi anche categoriche, oppure continue, o ancora una commistione di questi due tipi. Da questo punto di vista, il RL lascia molta libertà a chi deve modellare il problema. Per ogni scelta di tipo

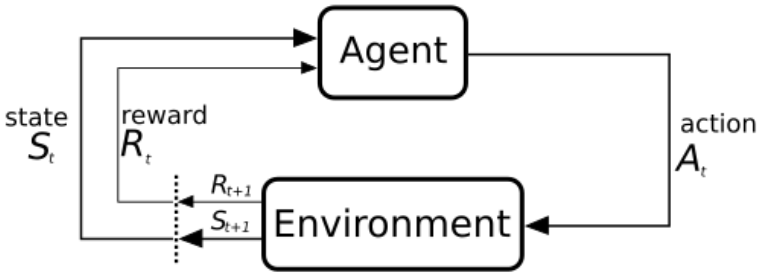


Figura 2.1: Diagramma di Reinforcement learning di un *Markov decision process* Sutton and Barto [2018]

delle variabili esistono soluzioni algoritmiche ad hoc. Lo stato subisce cambiamenti grazie all'interazione agente-ambiente.

4. **Reward**: si tratta della risposta da parte dell'ambiente all'azione eseguita dall'agente. Per definizione di RL, il reward può dipendere solamente dallo stato attuale del sistema e dall'azione. Consiste in una delle proprietà ereditata dai processi di Markov, su cui si basa tutta la teoria del RL classico. Violare questa regola, guardando a stati del sistema precedenti a quello attuale, significherebbe passare a modelli SMDP, cioè *semi-Markov decision process*(Fatemi et al. [2022]), in cui gli stati presi in considerazione sono molteplici, oppure dove l'azione non ha una durata deterministica sull'ambiente, ma stocastica. L'assunzione di Markovianità garantisce che algoritmi più semplici possano essere applicati; non è così stringente poichè i problemi più comuni non hanno bisogno di una storicità su cui basarsi, ma fanno affidamento solo sullo stato corrente, in questo caso dell'agente. Nella formulazione di problemi di RL questa assunzione può essere anche aggirata, inserendo nello stato metriche che registrino un certo valore, come può essere il tempo trascorso dall'inizio di una partita di scacchi, che porta il giocatore a muovere in modo più o meno frettoloso. Nell'applicazione oggetto di questa tesi, si vedrà come i comportamenti passati di un debitore possono essere riassunti nello stato del problema ed entrare negli aspetti tenuti in considerazione dall'algoritmo di RL.

5. **Azione**: è intrapresa dell'agente, in base ad una sua politica decisionale (*policy*). Anche l'azione, come le variabili di stato, può assumere forma discreta o continua a seconda delle "regole" imposte dall'ambiente o dal modellista. I più classici esempi sono quelli della rotazione di una manopola per determinare l'angolo di sparo di una certa arma, nel caso continuo, e del movimento su-giù-destra-sinistra di un agente in un labirinto a griglie, nel caso discreto.

Ognuna di queste entità può essere di vario genere e possedere determinate caratteristiche che guidano chi scrive il modello nella scelta dell'appropriato algoritmo di RL. Infatti, non tutti gli algoritmi di RL sono adatti ad ogni combinazione di caratteristiche degli elementi fondamentali.

2.2 Value function e rendimento

La *value function* dà un valore di bontà allo stato in cui l'agente si trova. Essa dipende sempre dallo stato del sistema, ma può dipendere anche dall'azione che si intraprende in un dato istante. Può quindi avere forma matriciale oppure di funzione continua sul dominio delle variabili di stato. La seguente rappresenta la più comune funzione valore nel RL:

$$v_\pi(s) = \mathbb{E}_\pi[G_t \,|\, S_t = s] \tag{2.1}$$

Condizionatamente allo stato s in cui si trova l'agente al tempo t, il valore atteso del rendimento G, seguendo la policy π è proprio il valore dello stato s. Questa è meglio definita come *state-value function* perché associa un valore allo stato. Come già anticipato, esiste anche una *state-action value function* che invece associa un valore all'accoppiata stato-azione, ottenuto condizionando oltre che sullo stato s, anche sull'azione a che l'agente prende in quello stato. Essa viene tipicamente in letteratura come *Q function*.

$$q_\pi(s, a) = \mathbb{E}_\pi[G_t \,|\, S_t = s,\ A_t = a] \tag{2.2}$$

Inoltre, é stato introdotto un altro importante elemento: il rendimento G_t. Questo si può ricavare guardando solamente al reward che intercorre tra lo stato attuale ed il successivo, oppure guardando ai prossimi n passi, o ancora al reward fino al termine dell'episodio. Un esempio per quest'ultima tipologia è il seguente: se ci si trova in un labirinto in cui ogni passo porta un penalizzazione di -1, e l'uscita dal labirinto porta un reward di +10, allora uscire dal labirinto in 5 passi comporta un rendimento pari a 5. Per altri modelli, come quelli studiati per la psicologia di certi animali, l'effetto di un'azione si associa ai reward che si verificano nei secondi immediatamente successivi. Passato un lasso di tempo definito, l'animale non associa più il reward ottenuto all'azione compiuta, quindi non apprende nulla.

Nella pratica esistono casi in cui l'agente continua potenzialmente all'infinito ad intraprendere azioni sull'ambiente: l'unica mossa possibile è quella di aggiustare il tiro in modo da ottenere il maggior reward possibile. È questo il caso di problema continuativo, in cui è necessario scontare le azioni future proprio per evitare il problema di un rendimento totale infinito.

Nel caso di problema episodico, in cui dopo un tempo finito l'agente interrompe le sue azioni sull'ambiente per ripartire eventualmente con un altro run di *trial and error*, si parla di rendimento cumulativo, il quale si presenta in forme come la seguente:

$$G_t = R_t + R_{t+1} + R_{t+2} + ... + R_{t+n} \tag{2.3}$$

Nel caso *continuing*, invece, si parla di rendimento scontato, potenzialmente su tutti i reward futuri, ottenuti sempre seguendo la policy di riferimento π, in cui entra in gioco un fattore di sconto γ, importante per garantire convergenza ai modelli e non lasciar esplodere il valore degli stati:

$$G_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + ... + \gamma^n R_{t+n} \tag{2.4}$$

2.3 Policy ed esplorazione delle azioni

Il RL è basato sull'apprendimento di azioni da intraprendere su stati di un certo tipo, con lo scopo di definire una *policy*, cioé una funzione che mappa ciascuno stato in un'azione, la migliore che possa essere presa. Per fare ciò, come metrica principale si utilizza sicuramente il valore dello stato appena introdotto, ma non solo. Infatti, le *policy* che utilizzano soltanto la *value function* per decidere quale azione intraprendere sono dette *greedy*, cioé avare. Esse puntano ad avere il maggior rendimento possibile con l'azione che si intraprende, senza un'ampia visione verso il futuro, in cui magari il passaggio ad uno stato più favorevole possa portare ad un vantaggio più grande, ma non nell'immediato. Per questa ragione è importante introdurre il concetto di esplorazione delle azioni, per consentire all'agente di scegliere strade che non portino necessariamente al miglior reward immediato, ma che magari possono portare vantaggi in futuro. La tecnica più famosa e semplice in tal senso è la cosiddetta ϵ-*greedy policy*, secondo la quale si massimizza il reward immediato con probabilità 1-ϵ, mentre si sceglie un'azione non "avara" con probabilità ϵ, puntando ad esplorare meglio lo spazio delle azioni. Questa è la tecnica utilizzata in alcuni algoritmi di RL per evitare di ricadere in ottimi (minimi o massimi) locali a causa di una scarsa esplorazione dell'ambiente.

Un esempio è quello di una strada lunga 10 passi per uscire da un labirinto: senza esplorazione sarebbe ripercorsa all'infinito dall'agente, dopo averla provata la prima volta, non scoprendo mai che in realtà esiste una strada più breve. Quest'ultima porta quindi ad un reward finale maggiore, a costo di passare qualche ciclo dell'apprendimento in stati con valori più bassi perché ancora poco esplorati. Questo tipo di esempio si può visualizzare in figura 2.2 in cui la strada con maggiore rendimento è quella passante per il centro della griglia, ma esistono altre strade possibili, corrispondenti a massimi locali di rendimento, da cui l'agente rischia di non spostarsi mai, nel caso di mancata esplorazione delle azioni.

Un'alternativa per esplorare le azioni è quella dell' *Upper-Confidence-Bound* (UCB), utilizzato ad esempio nello studio di problemi come il *multi-armed bandit* in Sutton and Barto [2018]. L'oggetto del problema è la scelta della migliore leva da attivare in una slot machine, osservando i reward restituiti dopo ogni scelta, cioè la leva che ha il maggior ritorno atteso. Qui l'azione scelta non è sempre quella con il *q-value* massimo, parlando di *state-action value function*, così come per la ϵ-*greedy policy*. Per scegliere in che direzione esplorare, tale tecnica tiene conto sia del valore della funzione Q che di un altro fattore, il cui valore aumenta logaritmicamente con il tempo e diminuisce linearmente con il numero di volte in cui l'azione in oggetto è già stata intrapresa($N_t(a)$):

$$A_t = \underset{a}{\operatorname{argmin}} \left[Q_t(a) + c\sqrt{\frac{\ln t}{N_t(a)}} \right] \tag{2.5}$$

In questo modo, le azioni che sono già state visitate più volte a partire da un dato stato, vengono penalizzate rispetto a quelle meno esplorate, mantenendo un'esplorazione uniforme, a lungo termine, su tutte le possibili azioni, e non solamente legata al caso come avviene nell'altra tecnica di esplorazione vista in precedenza.

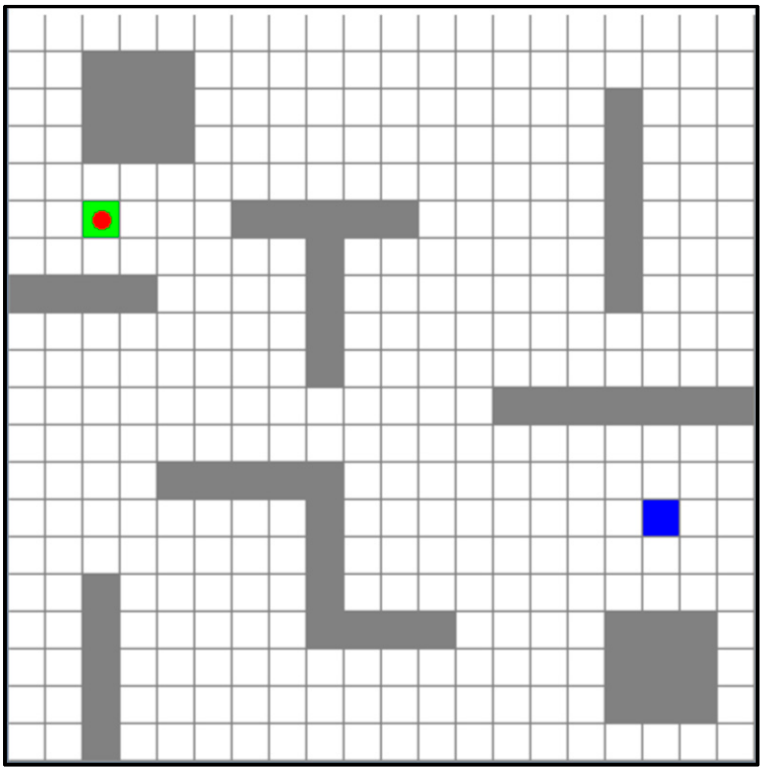


Figura 2.2: Problema di uscita da un labirinto da parte di un agente ("A Self-Adaptive Reinforcement-Exploration Q-Learning Algorithm", Lieping Zhang et al., 2021)

2.4 Metodi di soluzione tabulare

Il RL è una branca del machine learning piuttosto recente, ma già molto esplorata specialmente per i primi metodi utilizzati, cioè quelli tabulari. Quest'ultimi sono non-parametrici, nel senso che la policy non si aggiorna facendo variare uno o più parametri riferiti alla policy stessa, ma i modelli si basano soltanto sul valore degli stati possibili nell'ambiente in cui si svolge il problema e non da pesi associati ad ogni variabile, ed allenati nel RL stesso. Questo li rende maggiormente comprensibili e interpretabili, lasciando spazio uguale a tutte le variabili di stato, senza aggregarle in un qualunque tipo di approssimazione lineare della *value-function*

2.5 Q-Learning

Uno dei più utilizzati metodi tabulari è il *Q-Learning*. Si tratta di un algoritmo*off-policy* che consente di scegliere ad ogni iterazione l'azione più promettente in termini di rendimento atteso, senza dover seguire alcuna policy di riferimento, come avverrebbe, nel caso *on-policy*. Un esempio per quest'ultimo tipo di apprendimento è SARSA: esso

utilizza come rendimento G_t il reward immediato, come il *Q-learning*, ma anche uno o più ritorni ottenuti scegliendo come azioni negli stati successivi quelle presenti nella *policy* π ottenuta nel precedente ciclo di *policy improvement*.

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha[R_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \tag{2.6}$$

s_t = stato al tempo t
a_t = azione al tempo t
R_t = reward in seguito all'azione a_t s_{t+1} = stato al tempo $t+1$
a_{t+1} = azione al tempo $t+1$, scelta in base alla policy π α = tasso di apprendimento
γ = tasso di sconto

L'apprendimento off-policy comporta un'elevata varianza per lo stimatore dei ritorni G_t, il cui valore atteso condizionato compone la *state-value function* dell'equazione 2.2. Tuttavia sono garantite una maggiore velocità di convergenza e computazionale. Un altro esempio simile per quanto riguarda il RL off-policy è rappresentato dal metodo **Monte-Carlo** in cui l'assegnazione del valore di uno stato avviene guardando alla media dei reward avuti a partire da quello stato; procedura che comporta una varianza elevata nello stimatore G_t in quanto ogni traiettoria delle azioni può produrre ritorni profondamente diversi, ma il *bias* dalla politica ottimale è minimo in quanto non è presente una policy che influenza la direzione prevalente delle traiettorie stesse.

Per salvare i valori di ogni stato, come anticipato, si utilizza la *state-action value function* in quanto sia gli stati che le azioni sono discreti e quindi costituiscono una matrice con gli stati sulle righe e le azioni sulle colonne. Quando la matrice risulta molto ampia, è utile realizzare degli aggiornamenti asincroni della matrice, favorendo gli stati più visitati nella simulazione, tralasciando in parte l'esplorazione di stati più remoti rispetto alle traiettorie più comuni.

Le entrate della matrice vengono inizializzate con valori arbitrari, scelti da chi scrive il modello. La formula di riferimento per l'aggiornamento della matrice Q è la seguente:

$$Q(s, a) = Q(s, a) + \alpha[R + \gamma \max_a Q(s', a) - Q(s, a)] \tag{2.7}$$

s = stato
a = azione
s' = stato successivo nell'episodio
α = tasso di apprendimento
γ = tasso di sconto

Questo tipo di aggiornamento prosegue per ogni episodio simulato, o basato su un'esperienza reale, restituendo già durante questo processo una policy attuabile, ovvero quella che prende per ogni stato s l'argomento con il maggior valore in quella riga della matrice Q. Non è necessario perciò aspettare il termine dell'apprendimento per avere una politica sub-ottimale da seguire. Ciò che succede invece in metodi on-policy è l'attesa per la fine della fase di *policy evaluation* in cui si itera per migliorare la funzione valore di ogni stato, seguendo la policy di riferimento. Nella fase di *policy improvement* la politica da seguire viene dunque aggiornata e si ripete la fase di evaluation, proseguendo nel ciclare queste due fasi fino al raggiungimento di una convergenza della stessa.

Nel *Q-learning*, l'informazione di una specifica entrata $Q(s', a)$ si propaga quindi in altri punti della matrice quando questa rappresenta il valore massimo in quella riga (s'). Viene utilizzata, in aggiunta al reward, per dare un termine di paragone con lo stato attuale dell'entrata considerata. Se questa quantità è maggiore rispetto al vecchio valore della matrice, allora $Q(s, a)$ subisce un incremento positivo soppesato dal fattore di apprendimento α, altrimenti diminuisce in proporzione.

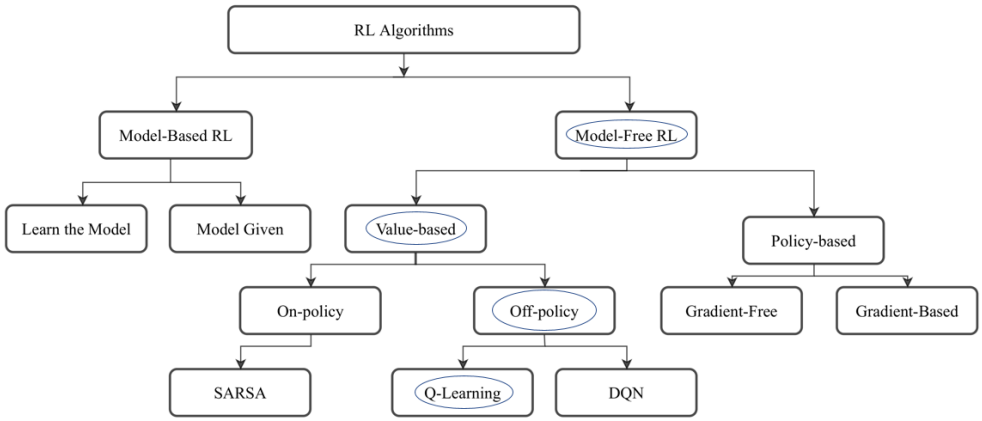


Figura 2.3: Albero di scelta degli algoritmi applicabili nel RL. Le scelte cerchiate portano alla foglia del *Q-learning*, algoritmo *model-free*, basato sull'aggiornamento della *value function*, *off-policy*

2.6 Smooth Q-Learning

Esistono molte varianti per ogni tipo di algoritmo in questa branca. Per il *Q-Learning* in ambienti in cui lo stato del sistema è molto grande, è importante trovare come popolare nel modo più efficiente possibile la matrice Q per non ricadere in modelli poco efficaci. A differenza di altri modelli di reinforcement learning in cui è obbligatorio visitare ogni stato della matrice più volte per arrivare convergenza, in questo caso si può pensare di aggiornare la matrice in modo asincrono: gli stati più frequenti vengono valutati molte volte prima che altri siano visitati per la prima volta. È quello che viene definito *trajectory sampling*, cioé un modo di procedere da stati di partenza ritenuti più importanti, visitando solo stati raggiungibili. Per quegli stati che difficilmente sono visitabili seguendo le azioni campionate o derivate da una qualche esperienza reale, viene assegnata una *policy* da chi scrive il modello oppure non viene definita in quanto non sarebbe frutto dell'apprendimento.

Tuttavia, può capitare che non sia sufficiente valorizzare gli stati lungo la traiettoria perché i campioni sono troppo pochi o lo spazio degli stati è molto grande. Si può scegliere allora di adottare una variante chiamata *Smooth Q-Learning*(Liao et al.). Questo algoritmo consente di aggiornare più stati contemporaneamente, a patto che siano simili in qualche senso. È allora possibile valorizzare lo stato adiacente a quello che si sta trattando nel caso in cui, secondo una certa metrica di riferimento, si trovi all'interno del

vicinato di quello stato. Oppure ancora, nel caso in cui le variabili di stato fossero difficilmente aggregabili in un'unica metrica significativa, si può pensare di dare dei termini di somiglianza ad hoc per singole parti dello stato o per sottogruppi di esse.

Ad esempio, nel caso di un ambiente bidimensionale quadrato detto *gridworld*, l'agente che si trova nel centro della griglia e che vuole intraprendere un'azione vedrà sia il cambiamento di valore dello stato in cui si trova, sia quello di tutti gli stati adiacenti a lui. Resta solo da definire il concetto di vicinanza che può rappresentare i quattro spazi in direzione dei punti cardinali relativamente a quello stato, o anche quelli in direzione diagonale. É una scelta puramente modellistica che ha maggiore utilità ovviamente in casi in cui la griglia è piuttosto estesa ed è difficile poter visitare tutti gli stati per poter dare loro una valore. L'aggiornamento della matrice Q risulta così aumentato e la conoscenza per uno stato viene condivisa con stati vicini, anche se in modo più diluito. Verrà infatti attribuito un coefficiente di smorzamento alle informazioni provenienti da stati vicini, in quanto la similarità non può implicare uguaglianza di trattamento tra due stati. Il coefficiente in questione può assumere valori tra 0 e 1, tenendo presente che per valori vicino allo zero l'influenza degli stati vicini sarà minima, mentre con valori prossimi all'unità la condivisione di informazioni sarà molto elevata.

La formula di questo tipo di algoritmo viene applicata solo nei casi in cui uno stato sia compreso nell'intorno dello stato appartenente alla traiettoria corrente, secondo la metrica di riferimento. Per tutti gli stati lontani, in un qualche senso, dallo stato considerato non c'è alcun contributo.

$$Q(s_n, a) = (1 - \beta)\, Q(s_n, a) + \beta\, \alpha[R + \gamma \max_a Q(s', a) - Q(s, a)] \tag{2.8}$$

s_n = stato nell'intorno di s
s = stato nella traiettoria corrente
a = azione
s' = stato successivo nell'episodio
α = tasso di apprendimento
γ = tasso di sconto
β = tasso di smorzamento

Capitolo 3

Definizioni e meccanismi del recupero crediti

3.1 Inquadramento economico-finanziario e ambientazione

La natura di questo problema è economico-finanziaria. Oltre a trattare strettamente di ricavi, guadagni e perdite, quindi l'aspetto economico, si parla anche di asset finanziari e di portafogli di pratiche, che vengono considerati come fonte di investimento finanziario. I creditori, che la banca ritiene di dover inserire in portafogli da cedere, vengono associati a delle pratiche in cui vengono inseriti tutti i rapporti bancari riferiti alla persona, fisica o giuridica che sia. I rapporti possono essere conti correnti, mutui, conti di anticipo, o forme tecniche simili. In questo settore si parla molto spesso di cartolarizzazione, in quanto le banche possono fornire tranches di crediti inesigibili a società come AMCO, prezzate anche in base al rating che viene dato a questi portafogli di pratiche. D'altra parte, anche AMCO fornisce pacchetti di investimento fondati di fatto sugli incassi derivanti dai flussi di cassa dei debitori, le cui pratiche vengono cartolarizzate sotto forma di ABS, *asset-backed securities*, cioé titoli garantiti da collaterali.

La totalità dei creditori che una banca vuole cedere viene dunque suddivisa in portafogli di qualità variabile che vengono acquistati da società come AMCO. Questa pratica é diventata sempre più frequente nelle banche e sono nate leggi per regolamentarla, al fine di evitare bolle speculative come quella di Lehmans Brothers del 2008. La più nota é la "legge di conversione del 21 giugno 2017 n. 96 del D.L. n. 50/2017" che ha semplificato il meccanismo di cessione dei crediti. Ogni banca deve infatti poter garantire la nota *senior* BBB su tali portafogli.

La legge non impone limiti sulla condizione in cui si possa trovare il debitore; si può parlare anche di posizioni *in bonis*, cioé persone che pagano regolarmente le proprie rate del mutuo o che non hanno sconfini sul conto corrente. Esistono infatti portafogli in cui vengono inseriti clienti che non sono cattivi debitori, affinché la banca ottenga una liquidità immediata e faccia proseguire la gestione del credito ad un ente esterno.

3.2 Crediti deteriorati

A parte rari casi, i tipi più frequenti di crediti ceduti riguardano quelli deteriorati, quindi debitori che hanno sconfini in conti correnti, o che hanno avuto dei ritardi nel saldo di una rata del proprio mutuo.

I crediti deteriorati si dividono in due categorie (cre [2017], Iannilli):

1. **NPE**(Non Performing Exposures): esposizioni debitorie deteriorate, senza concessioni particolari che ne possano giustificare il cattivo andamento.

2. **NPE con misure di tolleranza** (*forborne*): sono esposizioni non performanti oggetto di concessioni come moratorie covid, ottenute da molti debitori tra il 2020 e il 2022, o per spese mediche, o ancora per cambi recenti di mansione lavorativa di una persona fisica. Questo tipo di pratiche gode di tassi di interesse agevolati, o dilazione delle rate nei casi più frequenti. Per ragioni legislative, vanno trattate in modo differente dalle altre, portando spesso a lunghi tempi di attesa da parte degli agenti prima di poter agire in caso di insolvenza comprovata.

Esistono poi delle sotto-categorie che specificano la posizione deteriorata dell'uno o dell'altro tipo:

1. **Esposizioni scadute** e/o **sconfinanti** (*past due*): ne sono esempi una rata non pagata o lo sconfino di un fido, da più di 90 giorni. Nonostante i ritardi, la banca non ritiene ancora necessaria la classificazione a sofferenza del cliente, avendo fiducia che questo possa ritornare a breve *in bonis*. È un tipo di classificazione oggettiva, ma che non porta di per sé una pratica a essere segnalata al CRIF, la Centrale Rischi di Intermediazione Finanziaria.

2. **Inadempienze probabili** (UTP, *unlikely to pay*): posizioni per cui la banca giudica, soggettivamente, che il rientro completo dell'inadempienza sia impossibile senza un'escussione delle garanzie o di una riformulazione del contratto in seno al rapporto bancario coinvolto. Anche in questo caso la pratica non viene ancora dichiarata insolvente. A differenza del *past due*, non esiste un modo di classificare in modo oggettivo le pratiche che rientrano in questa categoria. Piuttosto è il responsabile della banca a dare questa indicazione e ad iniziare il processo di cessione del credito, se necessario.

3. **Sofferenze** (NPL, *Non Performing Loans*): esposizioni scadute da 90 giorni che la banca ritiene ormai (quasi) impossibili da risanare integralmente e che segnala quindi al CRIF. Da questo punto in poi, il cliente subisce una morsa più stringente sul credito e non può più accedere a nuove fonti di credito. Può accadere che anche questo tipo di pratiche possa essere ceduto dall'istituto di credito verso un SPV, società veicolo.

Figura 3.1: Tipi di posizioni debitorie in possesso dalle società di asset management pun

3.3 Strategie di recupero crediti

Sono molti gli aspetti interessanti trattabili in un campo come questo, a livello di analisi dei dati, decision-making, pricing di portafogli e altri. L'analisi delle pratiche può partire ex-ante, quando esse appartengono ancora alla banca, per effettuare una stima dei prezzi applicabili ad ogni pratica sulla base di crediti analizzati in precedenza. Si parla perciò di distribuzioni di probabilità per quanto riguarda il recupero atteso. Le pratiche possono poi essere divise in un numero di cluster arbitrario ed assegnate in modo proporzionale alla difficoltà ai vari asset manager. Inoltre si possono compiere analisi su entità quali avvocati, tribunali e regioni d'Italia per valutare il rischio di procedimenti civili che non portano ad un buon risultato. E possono esistere molti altri modi di sfruttare la mole di dati presenti.

In questo progetto l'attenzione si focalizza su AMCO, società che acquista i portafogli di crediti deteriorati e che deve gestirli in modo da ottimizzare il rendimento sul lungo periodo su ogni singola pratica. La granularità del problema quindi passa da un aggregato di pratiche acquisite in blocco, alla pratica riguardo il singolo cliente.

L'obiettivo primario che si persegue è quello di suggerire un'azione da parte di un agente, o di un team di agenti di recupero crediti, e impostare una possibile sequenza di azioni, identificabili in una strategia.

Un primo set di azioni è a carico prevalentemente degli asset manager di AMCO: questi possono scegliere se **gestire internamente** la pratica (azione frequente quando il debito è dell'ordine delle centinaia di migliaia di euro), **affidare** la pratica ad uno special **servicer** (una società esterna piuttosto che un'altra in base alla specializzazione, via legale o rinegoziazione dei debiti, o in base all'area geografica di riferimento), o non gestire la pratica e dichiarare un **recupero nullo**, perché la pratica presenta aspetti particolarmente critici che comporterebbero un'inutile spreco di risorse (ad esempio, pratiche che hanno un valore recuperabile massimo molto basso, inteso come poche migliaia di euro).

Un secondo insieme di azioni sono gestite dagli asset manager di AMCO o da società esterne: un agente si trova a dover prendere una decisione riguardo il tipo di recupero che si vuole tentare. Le due strade percorribili sono del tipo *gone* oppure *going*; il primo consiste nel procedere per vie legali e tentare di escutere le garanzie associate ad una pratica, mentre con il secondo si prova prima una mediazione con il cliente, e si concorda,

tipicamente, un piano di ristrutturazione del debito.

Tra le strategie di tipo *going* rientrano:

1. **Saldo e stralcio**: si concorda con il debitore un valore, inferiore al debito totale, che dovrà pagare in una o più soluzioni.

2. **Ristrutturazione del debito**: ridefinizione delle rate di un mutuo, ad esempio, con rate più agevolate e/o più spalmate sull'asse temporale.

3. **Classificazione a sofferenza**: azione specifica per crediti *in bonis* o UTP, che passando a sofferenza, subiranno una morsa più stringente da tutti gli istituti di credito, i quali non concederanno più ulteriori prestiti.

Tra le strategie *gone* figurano invece:

1. **Esecuzione forzata**: espropriazione di un bene in seguito ad un iter giudiziario che parte dal decreto ingiuntivo, attraversa le varie udienze e culmina con il pignoramento del bene mobile o immobile, se il tribunale accoglie la richiesta da parte dell'SPV.

2. **Procedura concorsuale**: azione specifica sulle ditte o persone giuridiche che porta a eventi come il fallimento e/o la liquidazione di un'azienda; tramite le insinuazioni al passivo fallimentare, le società di recupero crediti sperano di trarre profitto da questi eventi. Anche in questo caso si tratta di un iter giudiziario lungo e complesso che può impiegare molto tempo a volgere ad un conclusione.

3. **Escussione di garanzia consortile**: si cerca di escutere una garanzia presso un consorzio (Confidi tipicamente) che aveva garantito per l'azienda richiedente di un certo prestito. L'iter è simile a quello dell'esecuzione forzata con la differenza che non viene pignorato un bene ma si escutono le garanzie monetarie garantite dal consorzio stesso.

Capitolo 4

Analisi del database

4.1 Tabelle principali

RAD Informatica ha fornito per questo progetto un database di alimentazione del suo software (EPC, Ex Parte Creditoris) con più di 100 tabelle ed in continua evoluzione per via di ammodernamenti del software stesso. Molte tabelle fungono solo da supporto tecnico al funzionamento del programma, mentre da alcune si possono estrapolare informazioni importanti riguardo a tutte le pratiche di recupero crediti che sono gestite su EPC da AMCO o da altre SPV(Società veicolo), anche note come *special servicer*. Il numero di pratiche è dell'ordine delle centinaia di migliaia. Ogni pratica è caratterizzata in modo approfondito, se l'inserimento dei dati è stato effettuato correttamente dal responsabile che segue la pratica.

La prima tabella analizzata è "PRATICHE": questa contiene un'istantanea di tutte le pratiche alla loro data di caricamento nel database e dà quindi indicazioni riguardo la SPV che ha attualmente in carico la pratica e specifica l'asset manager responsabile interno ad AMCO. Fornisce inoltre una prima descrizione sommaria dello stato in cui si trova la pratica stessa: se ha garanzie reali sottostanti (immobili in primis), da quale banca proviene il credito deteriorato ed eventualmente se la pratica è già stata archiviata. Le metriche più importanti che si possono dedurre da questa tabella riguardano alcuni aspetti chiave delle pratiche:

1. Stato della pratica: i 3/4 delle pratiche sono ancora in corso, mentre 1/4 di queste sono già state archiviate. La componente di storicità è quindi abbastanza presente.
2. Presenza di garanzie reali: un quinto delle pratiche presenta garanzie reali, quindi ipoteche su immobili, mentre la restante parte presenta garanzie personali o non ha alcuna garanzia.
3. Stato del debitore: più della metà dei debitori sono in uno stato conclamato di sofferenza; una piccola quota, circa il 3% sono *in Bonis*, e la restante parte é composta da inadempienze probabili (UTP).

La seconda tabella analizzata riguarda la "DEBITORIA", cioé il collegamento tra le pratiche e i relativi rapporti. Ogni pratica deve possedere almeno un rapporto: questo

può essere un conto corrente, un mutuo, oppure appartenere a forme più particolari quali il conto anticipo fatture o spese legali associate alla stessa pratica. Ad ogni rapporto è associato anche il proprio valore lordo (GBV), gli interessi maturati finora sulla pratica e la previsione di perdita fatta da un agente di AMCO in modo soggettivo, basandosi sulla propria esperienza. La previsione di perdita è la richiesta da parte di un asset manager di passare a perdita una percentuale del GBV totale della pratica in oggetto perché ritenuta non più recuperabile in alcun modo.

La terza tabella riguarda le "GARANZIE". Qui sono salvate quasi 100 tipologie di garanzie, dalle ipoteche alle fidejussioni, passando per pegni e avalli (che sono cambiali in cui il debitore è garantito da terzi riguardo i pagamenti dovuti). Sono presenti più di un milione di ipoteche riferite a svariate migliaia di immobili su cui gravano. Questo fornisce un'accurata descrizione riguardo il massimo valore recuperabile da una certa pratica, campo che è utilizzato anche in una variabile di stato del RL (*max_recovery_value*). Il tipo di garanzia sottostante un prestito è importante per capire quale sia l'iter da seguire: per pratiche cosiddette immobiliari, l'agente è più incentivato ad avviare azioni legali per pignorare l'immobile in ipoteche e procedere alla rivendita, mentre per gli altri tipi di garanzia è più usuale tentare una rinegoziazione del debito. In base al tipo di garanzia è anche possibile stimare il valore del recupero: per gli immobili ci si può basare sia sulle perizie che sui valori di immobili simili all'asta; per gli altri tipi garanzia, la varianza riguardo il ricavo previsto è molto più elevata in quanto non sempre chi firma una fidejussione è poi in grado di adempiere ai suoi doveri.

La quarta tabella riguarda gli "IMMOBILI" e contiene il numero delle aste andate deserte per quell'immobile ed informazioni catastali quali regione, via, CAP e categoria catastale. Di questa tabella si sfrutta in particolare la provincia di locazione. Data la vastità del database, tutte le province sono presenti, con picchi di più di 20,000 occorrenze per le province di Perugia, Macerata e Treviso. Questa tabella è fondamentale per imputare la macro-regione di riferimento della pratica, come si vedrà più nello specifico in seguito.

La quinta tabella è "STATI CONTROPARTE" e descrive l'andamento dello stato del debitore al passare del tempo. Si tratta di un'informazione imprescindibile per associare al giusto tipo di cliente le azioni intraprese nel RL. Senza l'uso di questa tabella, l'informazione della variabile di stato *npe_type* (NPL/UTP/*in Bonis*) non potrebbe che essere un'assegnazione statica. In tal caso, non varierebbe nel tempo, causando il verificarsi di azioni su stati che in realtà non potrebbero subire tale azione da parte dell'agente.

La sesta tabella utilizzata è quella relativa alle "PERIZIE". Raccoglie una serie di valutazioni di periti, sia di parte che d'ufficio, su molti immobili riconducibili a pratiche già in EPC. Gli importi di queste perizie sono utili per dare un valore indicativo alle proprietà di tipo reale che fanno parte delle garanzie e quindi di stimare un valore di recupero nel caso in cui venissero assegnate all'asta, sperabilmente per un valore prossimo a quello stimato all'interno dei documenti peritali. Data l'importanza di una simile voce, in tutte quelle pratiche con garanzie reali sottostanti, è valorizzata all'interno dello stato del RL.

Esistono poi tabelle in cui vengono storicizzati tutti i passaggi automatici e manuali che si susseguono in una pratica. Per ogni debitore, si possono decidere alcune strategie di recupero, come la ristrutturazione del debito o l'esecuzione forzata. Ogni strategia si

compone di una o più procedure: le *going concern* sono spesso più semplici, mentre le *gone concern* sono scomponibili in più procedure come il decreto ingiuntivo o altri passi fondamentali del processo civile telematico (PCT). La granularità più bassa è rappresentata dalle fasi, contenute nell'omonima tabella, e indicanti ad esempio l'inizio di una procedura, una delibera o una trattativa con un cliente. Da qui si ricava un'importante data che rappresenta l'inizio ufficiale di una lavorazione in EPC, idealmente quella dell'on-boarding documentale riguardo ad una pratica e che serve nel RL come termine di paragone per la costruzione della metrica temporale presente nella variabile di stato *time_from_start_bins*.

Fin qui si è parlato prevalentemente di tabelle da cui derivare variabili di stato, ma grazie alla profondità del database si può avere anche una descrizione approfondita delle azioni intraprese da agenti di AMCO o di altri SPV. Le tabelle di riferimento sono "ESITI AFFIDAMENTI" e "PROPOSTE DI DELIBERA". Nella prima si trovano informazioni riguardo ad agenti sul campo che compiono azioni come rintracci o chiamate a clienti, ma anche azioni da parte di AMCO di revoca dell'affidamento ad un particolare SPV e assegnazione ad un altro. La seconda tabella rappresenta più nel concreto le decisioni da parte degli agenti riguardo quali strategie di recupero intraprendere, che vanno approvate da profili superiori a quello del proponente. Qui trovano spazio tutte le strategie più comuni, analizzate in precedenza e vanno a formare il set di azioni del RL.

Le ultime due tabelle sfruttate riguardano i flussi di cassa che si verificano nel processo di recupero del credito. Per i flussi in entrata (e raramente in uscita) si fa riferimento alla tabella "WATSON": qui si trovano importi e causali, prevalentemente derivanti da pagamenti dai clienti ad AMCO. Per i flussi in uscita si attinge alla tabella "FATTURAZIONE ELETTRONICA", in cui sono presenti sia fatture *corporate*, difficilmente allocabili alle singole pratiche, che altre associate esplicitamente alla singola pratica tramite l'identificativo univoco.

Capitolo 5

Reinforcement Learning in ambito recupero crediti

In questo capitolo si analizzano tutti gli elementi fondamentali del RL in tema recupero crediti. Quindi si vedrà come le varie componenti possono essere tradotte in uno specifico ambito, finanziario in questo caso, e trovare la loro modellizzazione in un algoritmo che prende ispirazione dal classico Q-Learning.

Si discute anche la motivazione per cui si è scelto di utilizzare il Reinforcement Learning in questo problema e non un classificatore più standard, sfruttando la storicizzazione delle azioni prese in passato e i loro rendimenti scontati ottenuti.

5.1 Motivazione per l'utilizzo del Reinforcement Learning

La motivazione che sta dietro questa specifica scelta, concordata a monte del progetto, risiede principalmente nella sequenzialità delle azioni. Il *modus operandi* degli addetti ai lavori può provocare un cambio di stato della pratica comportando una nuova valutazione delle possibili alternative procedurali. Una pratica può passare da uno stato in cui è necessaria un'azione prudente, quale la dilazione dei pagamenti, ad uno in cui la scelta è più mirata ad escutere le garanzie. Questo cambio continuo di stato rende ogni volta disponibile un nuovo modo di agire sulla pratica. Si tratta di un tipo di ambiente di lavoro in cui azioni successive possono portare ad un valore aggiunto se eseguite con certe tempistiche e questo tipo di dinamicità viene difficilmente catturata dalle classiche regressioni lineari o da clustering effettuati sulle pratiche. Se il problema fosse più iscrivibile in un perimetro di classificazione delle pratiche, si potrebbe pensare a metodi come *random forests* o SVM, come in Bellotti and Brigo [2019], ma non è questo il caso, vista la difficoltà ad assegnare etichette di buono o cattivo recupero riguardo una pratica.

Un'altra criticità trattata dal RL è l'individuazione del risultato di un'azione: si parla spesso di reward ritardati all'interno di una serie di azioni su una pratica. Grazie al RL siamo in grado di imparare, dalla ripetizione di quelle azioni in sequenze diverse su pratiche

id_file	deliberation_type	deliberation_date	servicer_name
1625636	lawsuit_procedure	2020-10-14	other_servicer
1625636	lawsuit_procedure	2021-01-11	other_servicer
1625636	credits_classification	2021-06-15	other_servicer
1625636	bank_credit_withdrawal	2021-06-15	other_servicer

Figura 5.1: Azioni compiute in successione su una stessa pratica, di tipo diverso, e che rendono idea della dinamicità delle azioni nell'ambito del recupero crediti

simili, se il risultato finale in termini di rendimento scontato sia frutto di un'azione presa all'inizio o alla fine dello storico di una pratica.

Per fornire un esempio, il RL è in grado di discernere l'alto valore di un rendimento atteso per la revoca di un fido prima dell'escussione di garanzie, per evitare che il debitore amplifichi il suo debito con AMCO, contro il basso valore che questa azione avrebbe nell'ambito di un piano di rientro in cui la limitazione di liquidità rende ancora più difficile da parte del cliente portare a termine il piano stipulato.

Infine, si può precisare che l'utilizzo di un classificatore in questo ambito sarebbe sconsigliato perché, come già trattato in precedenza, non è possibile definire a priori se l'esito di una pratica sia positivo o negativo, il che farebbe pensare ad una tecnica di *supervised learning*. Spesso non è neppure possibile identificare quale sia stata la strategia prevalente su una pratica perchè le azioni di tipo diverso si intervallano tra di loro, rendendo ambigua un'identificazione univoca. L'unico tipo di classificatore che si potrebbe pensare è di tipo *unsupervised*, quindi un clustering sulle pratiche in modo da suddividere il totale di esse in sottoinsiemi con caratteristiche comuni, ma a quel punto si vedrebbe come molte di queste subiscano tipi diversi di azioni (esempio in figura 6.1) Questo renderebbe di nuovo complicato l'assegnazione di una stessa strategia a tutto un cluster.

5.2 Variabili di Stato

Descrivere lo stato di una pratica non è semplice in quanto molti attori figurano all'interno del processo di recupero crediti: asset manager, special servicer, garanzie, garanti, debitori, tribunali, avvocati, periti e molte altre entità. Purtroppo non è possibile farli partecipare tutti all'interno dello stato che identifica in maniera sintetica ed esaustiva una pratica. I dati numerici più importanti per fotografare una pratica in qualsiasi momento sono sicuramente il **debito totale**, il **recupero parziale** effettuato e il tempo già passato a lavorare su di una pratica.

Esistono poi caratteristiche come il **servicer** che sta lavorando alla pratica, che in molti casi fa la differenza. Infatti ogni SPV compie molto spesso azioni caratterizzanti il suo stile nel recupero dei crediti. Alcuni di essi prediligono la negoziazione del debito, mentre altri sono più inclini alla discussione dei loro diritti in tribunale con l'obiettivo di

escutere garanzie e rivenderle. In un'altra applicazione, lo stile di recupero potrebbe essere utilizzato per assegnare pratiche a servicer che si pensa possano essere più predisposti ad operare in modo proattivo su quel tipo di pratiche.

Anche l'**area geografica** di appartenenza fa da discriminante, specialmente in combinazione con il tempo trascorso dall'inizio della lavorazione. Infine la **forma tecnica** prevalente (conti correnti o mutui) è una caratterizzazione molto importante, specie nell'ambito UTP (inadempienze probabili), visto che le procedure di recupero dipendono molto dal tipo di rapporto in essere. Le variabili di stato sono le seguenti:

1. **GBV_bins**: GBV sta per Gross Book Value, ossia il valore lordo totale dell'ammontare dovuto dal debitore nei confronti del titolare del credito della pratica. Questo valore si ottiene tramite *binning* del GBV totale della pratica, ricavato come somma dei singoli GBV associati ai rapporti (o contratti) di una pratica. È un valore che non cambia durante lo studio della pratica, ma il cui recupero viene aggiornato in un'altra variabile di stato (*revenues_over_gbv_bins*).

2. **NPE_type**: NPE sta per *Non Performing Exposure* e identifica in maniera generale una pratica con rapporti deteriorati. La suddivisione è pressoché binaria tra crediti NPL (sofferenze) e UTP (inadempienze probabili), ma in questo lavoro si è deciso di unire una terza voce che è quella dei crediti *in Bonis*: questi entrano sempre più spesso a far parte dei crediti detenuti dalle società di recupero crediti, anche se non presentano situazioni problematiche.

3. **technical_form**: rappresenta la forma tecnica prevalente dei contratti. Per ogni pratica viene considerato il rapporto con GBV più alto. I tipi di forme tecniche più frequenti e significativi, come individuato anche con esperti di settore, sono: conti correnti, mutui ipotecari, mutui chirografari (garantiti da fidejussioni) e mutui fondiari (mutui che riguardano la costruzione o ristrutturazione della prima casa). A causa di una carenza di descrizione nel database, esistono poi altre due categorie: *other_form*, per forme tecniche non riconducibili a quelle di cui sopra, e *minor_form*, per tutte quelle che forme tecniche a bassa occorrenza nel database il cui GBV ammonta allo 0,3% del totale di tutti i crediti.

4. **max_recovery_value_bins**: rappresenta il massimo valore di recupero di una pratica e lo si ottiene sommando l'importo di tutte le garanzie personali (fidejussioni) di una pratica e del valore delle ipoteche. Così come la variabile GBV, anche questa ha dovuto subire un *binning* per rientrare nel dominio delle variabili discrete. La distinzione tra tipi di garanzie non viene catturata da questa variabile, ma piuttosto rientra nel tipo di forma tecnica (*technical_form*) che differenzia pratiche con garanzie reali da garanzie personali o da conti correnti privi di garanzie.

5. **macroregions**: indica la macro-regione geografica italiana di riferimento per quella pratica. Questo valore è stato ricavato dalla tabella "Immobili" guardando quale provincia, e di conseguenza macro-regione, avesse il maggior numero di occorrenze di immobili associati ad una specifica pratica. Se una pratica non ha alcun immobile associato, allora si dà il valore *no_macro_region*.

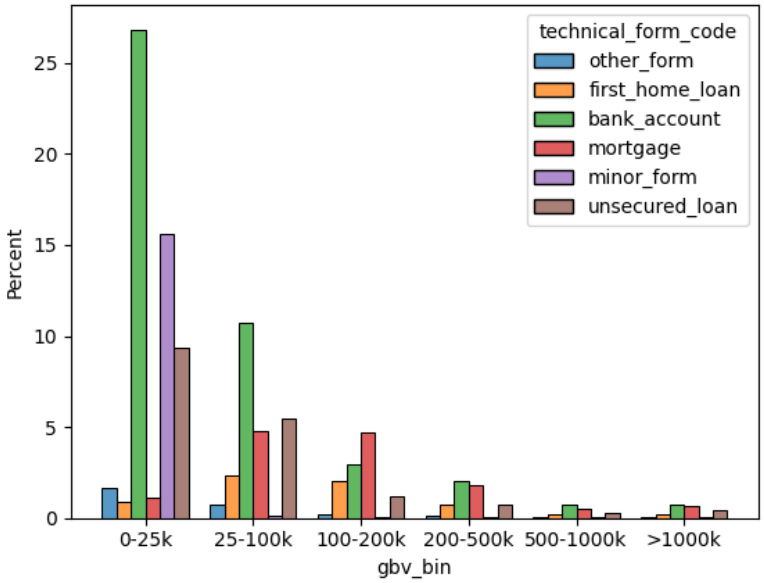


Figura 5.2: Grafico a barre combinato per valore lordo della pratica e forma tecnica dei rapporti. Percentuali espresse sul totale delle pratiche.

6. **time_from_start_bins**: è una metrica temporale che indica quanto tempo sia passato dal momento in cui la pratica è stata presa in carico per la prima volta da un agente di AMCO. È di nuovo una variabile che ha subìto una discretizzazione in intervalli semestrali o annuali. L'indicazione temporale è quella riguardante l'azione intrapresa dall'agente, ma il termine di paragone resta sempre quello di cui sopra.

7. **revenues_over_gbv_bins**: è la metrica di riferimento per il valore recuperato allo stato attuale e come dice il nome è di nuovo una discretizzazione. Si ricava dividendo il reward totale ottenuto fino al momento presente sulla pratica per il suo GBV totale. Gli intervalli sono "schiacciati" verso lo 0% perché è molto improbabile recuperare più del 50% del GBV di una pratica, ma soprattutto perché i cambi di strategia avvengono molto piú spesso nella fase in cui il recupero sulla pratica è ancora molto basso.

8. **servicers**: è una variabile categorica che rappresenta gli SPV più presenti nel database, valorizzata in ogni azione intrapresa da un agente. Per i servicer considerati di minor importanza si utilizza la voce *other_servicer*.

5.2.1 Vicinati delle variabili di stato

Un aspetto peculiare dello *smooth Q-learning* è lo studio dei vicinati. Le variabili di stato appena elencate sono di tipo discreto e tutte molto diverse tra loro. Esse non sono assimilabili in una metrica comune, ma vanno trattate in modo separato. Serve sfruttare la conoscenza di dominio per definire quando due stati siano simili.

Due stati sono considerati simili e quindi subiscono l'aggiornamento da parte dello *smooth Q-learning* se:

1. i valori di *GBV_bins* sono adiacenti. Ad esempio, una pratica con *GBV_bins* = "0-25k€" è nel vicinato di una pratica con *GBV_bins* = "25-100k€".

2. i valori di *max_recovery_value_bins* sono adiacenti. Ad esempio, una pratica con *max_recovery_value_bins* = "25-100k€" è nel vicinato di una pratica con *max_recovery_value_bins* = "0-25k€" o con *max_recovery_value_bins* = "100-250k€" .

3. la forma tecnica prevalente espressa in *technical_form* riguarda mutui con garanzie reali. Un mutuo fondiario (*first home loan*) è considerato simile ad un mutuo ipotecario (*mortgage*). L'unica differenza reale tra i due tipi di contratto è la qualifica dell'immobile in garanzia, che passa dall'essere obbligatoriamente una prima casa nel primo tipo, ad un qualunque immobile nel secondo.

4. appartengono a due regioni geografiche confinanti. Una pratica con maggioranza di immobili a garanzia al nord è considerata simile ad una pratica con maggioranza di immobili a garanzia al centro. Per mutui diversi da quelli con garanzie reali, no si applica questa regola di vicinato.

5. sono ad una distanza temporale simile dall'*on-boarding*. Una pratica con uno storico di 12-24 mesi è considerata simile ad una con più di 2 anni di lavorazione.

5.3 Azioni

Il discorso fatto per le variabili di stato può essere ripetuto in questo frangente. Sono molte le azioni che possono essere prese su una pratica da parte di un agente, ma per una buona riuscita degli esperimenti ci si dovrà accontentare di una ventina di opzioni. Questo perchè altrimenti la grandezza della matrice diventa insostenibile a livello computazionale, ma anche per una questione di significato: molte azioni sono di scarsa importanza e spesso ripetitive, come i contatti telefonici ai clienti, e vengono quindi inglobate nella categoria dei solleciti di pagamento per dare più spazio di incidere ad azioni che indirizzano la pratica pesantemente in un verso piuttosto che in un altro.

Come già analizzato in precedenza, le strategie di recupero del credito si possono classificare grossolanamente in due gruppi: la rinegoziazione del debito con il cliente oppure l'avvio di procedure legali atte all'escussione delle garanzie sottostanti i rapporti coinvolti. Gran parte delle azioni riguardano sfaccettature di questi due tipi di approccio, anche detti rispettivamente *going concern* e *gone concern*. Le restanti azioni individuate sono punti critici del processo come la classificazione di un debito a sofferenza oppure un *warning* da parte di un agente.

Segue una breve descrizione di tutte le azioni possibili:

1. **agent_warning**: sollecito da parte di un agente su un cliente che non sta rispettando il piano di rientro pattuito.

2. **outsourcing_withdrawal**: revoca da parte di AMCO di una pratica ad un servicer che non è riuscito a completare il recupero.
3. **credits_classification**: classificazione di un credito, per motivi fondati, che passa da inadempienza probabile a sofferenza.
4. **re_entry_plan**: l'agente propone un piano rateale di rientro al cliente.
5. **credit_renunciation**: rinuncia di effettuare altre azioni per il recupero del credito.
6. **full_and_final_settlement**: proposta di saldo e stralcio con pagamento in un'unica soluzione.
7. **write_off_with_deferred_payments**: saldo e stralcio con dilazione dei pagamenti.
8. **new_recovery_strategy**: proposta generica di cambio della strategia finora adottata.
9. **forbearance_or_covid**: passaggio di una pratica in stato di *forbearance*, molto spesso associato ad una moratoria covid. Questa particolare condizione è trasversale alle tipologie di debito esistenti (NPL, UTP, *in Bonis*) e consiste in una speciale concessione da parte del servicer di condizionei più favorevoli nel piano di rientro, spesso l'abbassamento degli interessi o la dilazione dei pagamenti.
10. **bank_credit_withdrawal**: revoca di un fido associato ad una pratica come misura preliminare di un procedimento giudiziale.
11. **credit_collection**: richiesta di escussione pegni o garanzie.
12. **credit_cession**: proposta di cessione di un credito ad altro servicer.
13. **info_request**: richiesta di informazioni alla banca originatrice della pratica, magari riguardo possibile documentazione mancante.
14. **begin_lawsuit**: inizio di una causa legale, solitamente tramite un decreto ingiuntivo.
15. **lawsuit_procedure**: attuazione di una procedura legale spesso riconducibile ad esecuzione immobiliare, o comunque una delle ultime fasi dell'iter giudiziale.
16. **confidi_liquidation**: escussione di crediti consortili, con garante di riferimento che è spesso il consorzio Confidi. Alternativamente si tratta di garanzie erogate da consorzi regionali, soprattutto riferite a partite IVA.
17. **extrajudicial_appraiser**: richiesta di una perizia su uno o più immobili associati alla pratica.
18. **claim**: la traduzione letterale è "indennizzo" e può presentarsi in casi in cui si trovino delle non conformità all'interno della documentazione conferita dalla banca, o in generale su vizi di forma relativi alla pratica in oggetto.

5.4 Reward

Il reward di un'azione in questo problema è quanto guadagna la società da quel tipo di azione, in termini monetari, senza tener conto di altre metriche o valori. Se i dati dei pagamenti sono stati inseriti correttamente e in modo completo dagli utenti, si può ricavare in modo immediato dal database la cifra che ogni cliente versa ad AMCO. Anche l'indicazione sul giorno in cui viene effettuato il pagamento risulta importante per scontare in modo corretto, con un tasso giornaliero, tutti i flussi di cassa. Bisogna specificare che i reward possono essere anche negativi, nel caso in cui su una pratica non si siano visti ancora risvolti positivi, e quindi versamenti da parte del cliente o vendite all'asta degli immobili pignorati. Inoltre esistono casi in cui il reward è nullo, a causa di uno svolgimento ancora in atto della pratica che non ha portato ancora né spese né benefici in seguito all'azione in oggetto.

È stato necessario scegliere anche quali reward associare ad una particolare azione. Purtroppo i dati disponibili sui flussi di cassa, molto descrittivi e difficilmente analizzabili, hanno reso più prudente considerare tutti i reward successivi all'azione in gioco, nel tempo che intercorre tra questa e l'azione successiva. Esistono casi in cui il reward associabile ad una certa azione non avviene prima che si compia un'altra azione. Ad esempio, se si stipula un piano di rientro e in seguito si fa un'azione di sollecito dei pagamenti, gran parte del merito è della prima azione, ma viene associato in parte maggiore alla seconda. Tuttavia, grazie alla dinamicità che riesce a cogliere il RL, questo reward ritardato va ad impreziosire lo stato successivo alla stipula del piano di rientro, portando questa azione ad essere consigliata all'agente non per il reward immediato, ma per quello a lungo termine. Si parla in questo frangente di reward ritardati rispetto all'azione, il che complica notevolmente il lavoro del RL nel riconoscere da dove provenga la buona riuscita della pratica analizzata.

Infine, occorre evidenziare la differenza tra problemi con reward ritardati come questo, e problemi in cui i reward sono direttamente associabili ad azioni appena prese. Giochi di carte come scopa ne sono un chiaro esempio: giocare una carta uguale ad una già presente sul tavolo comporta un reward immediato e ben identificabile (2 carte raccolte, che magari concorrono a bonus finali più complessi). Tuttavia, anche in questi esempi possono esistere giocate il cui reward immediato è nullo, ma fatte in vista di un rendimento finale o di medio termine più promettente. In ogni caso, spesso si è consapevoli che una certa giocata sia stata complice di un rendimento finale elevato, mentre nell'ambito del problema trattato in questa tesi è difficile, anche al termine di un episodio, capire quali azioni abbiano contribuito e in che quantità alla buona riuscita di un recupero crediti.

Capitolo 6

Implementazione

L'implementazione si compone di più linguaggi e tool di programmazione. I primi step di estrazione dei dati dal database sono query SQL su tutte le tabelle del database precedentemente citate. L'algoritmo è sviluppato in Python, scelto per la semplicità e la presenza di librerie note e utili per maneggiare dati e vettori, anche di grandi dimensioni. L'ultimo tool utilizzato nel progetto è Streamlit, noto per ideare prototipi di applicazioni. Si vedrà come questi strumenti abbiano concorso per trasformare il dato grezzo ottenuto dal database in un'applicazione utile per l'agente di recupero crediti, al fine di selezionare le migliori strategie di lavorazione delle pratiche.

6.1 Librerie Python in uso

Le librerie Python utilizzate sono le più comuni e non sono molte in quanto contengono già al loro interno tutte le funzioni utili per trasformare il dato e renderlo fruibile al RL vero e proprio. Le principali librerie sono:

1. **numpy**: è la libreria più utilizzata nel progetto. È una tra le più conosciute e affidabili per la manipolazione di **array** di dati omogenei, cioé con elementi dello stesso tipo. L'incredibile **velocità** di scorrimento e **calcolo** dei vettori di questa libreria la rende indispensabile, se l'obiettivo è quello di lavorare con un database molto esteso come quello di EPC. Inftti, il pattern ricorrente nel codice è quello di trasformare ogni matrice di informazioni riguardo una pratica in un array multidimensionale in Numpy per avere prestazioni anche 100 volte migliori rispetto ad altre librerie.

2. **pandas**: è utilizzata per estrarre i dati dai file .csv, grazie all'ampia scelta di opzioni riguardo il trattamento, ad esempio, dei nomi delle colonne o dei separatori utilizzati nei file in lettura. Tale libreria ha la comodità di indicizzare righe e colonne con nomi significativi, a differenza di Numpy, ma per come è progettata, non consente di manipolare grosse moli di dati ed effettuare calcoli estesi su di essi. Nonostante ciò, é molto utile per avere un'organizzazione iniziale ordinata delle tabelle in **dataframe**.

3. **datetime**: è la libreria di riferimento per il *parsing* delle **date**: permette di uniformare formati di date diversi, trovati all'interno del database, in un unico formato

scelto per permettere facili e veloci operazioni tra date. Principalmente verranno utilizzati il calcolo del tempo tra due date e il confronto tra di esse.

4. **seaborn**: è una libreria per disegnare tipi di **grafico** diversi in modo intuitivo.

6.2 Query da SQL server

Le *query* SQL sono utilizzate per estrarre i dati più importanti e che saranno utili all'algoritmo di RL. Ciò che viene selezionato tramite le query viene salvato in file .csv perché si è riscontrata una maggiore velocità di esecuzione nel leggere dati direttamente da .csv piuttosto che utilizzare librerie come *Pyodbc*, che rendono molto macchinosi e lenti i processi.

La fase di filtraggio dei dati inizia già in questa fase, infatti non tutte le pratiche vengono utilizzate: esistono delle pratiche con numero identificativo basso e con titolare del credito diverso da AMCO che presentano pattern diversi rispetto a tutte le altre. Si tratta di circa 10000 pratiche che non presentano in alcun caso un servicer a cui è affidata la pratica e per cui la qualità dei dati è più scadente. Tali pratiche non sono state considerate per mantenere un pattern del database il più omogeneo possibile. Vengono eliminate anche tutte quelle pratiche, poche centinaia, in cui per errore il GBV ha valore negativo, sintomo di un errore di importazione del dato da parte del responsabile.

Un altro problema nell' utilizzare tabelle così variegate, risiede nei dati mancanti: è difficile compiere un'imputazione ad hoc per ogni colonna, quindi si è scelto di associare in modo univoco una stringa '00' a tutti i dati mancanti al fine di lavorare quest'ultimi in modo puntuale nel momento in cui siano richiesti.

Infine, non tutti i dati di una tabella sono importanti per un RL, ad esempio i dati più descrittivi e che non portano ad un'informazione utilizzabile, perciò i campi selezionati riguardano principalmente date, importi e codici identificativi.

6.3 Preprocessing dei dati

Vista la disomogeneità del database, è necessario effettuare un *preprocessing* dei dati in input per renderli fruibili all'algoritmo in modo univoco e consentire ad esso di lavorare su dati omogenei, in modo che non impieghi più tempo del dovuto a manipolare dati all'interno dei cicli di apprendimento o di test.

Il principale trattamento dei dati è stato effettuato sulle date, presenti in tabelle e strutture diverse, ma rese uniformi attraverso il comando *strptime* della libreria *datetime*. Questo lavoro di *parsing* ha dunque permesso di rendere confrontabili date che provenissero indifferentemente da incassi, spese o azioni degli agenti. In questo frangente è stato trattato uno dei primi *missing values*: non tutte le pratiche compaiono nella tabella "Fasi", in cui è salvata ogni piccola azione effettuata su una pratica, e in tal caso presentano un dato mancante a riguardo. Tuttavia, vista l'importanza di associare una data di partenza alla pratica per poter scontare i flussi di cassa, si è deciso di affidare una data di partenza alternativa, ricavata dalla tabella "Pratiche", che non rappresenta la data di inizio dei lavori da prte di AMCO, ma da parte del servicer a cui è affidata la pratica stessa.

id_file	date_first_phase	gbv	technical_form_code	gbv_bin
1589026	2018-12-19	1992675.94	other_form	>1000k
1732342	2017-08-26	21245.47	first_home_loan	0-25k
1636574	2020-07-28	54553.99	bank_account	25-100k
1715736	2017-09-21	2055507.44	first_home_loan	>1000k
1747601	2006-01-31	33096.66	bank_account	25-100k

Figura 6.1: Piccolo esempio di pratiche presenti nella tabella *df_state_static* con proprietà principali visualizzate per ogni pratica. Le voci presenti provengono dalle tabelle "Pratiche" e "Debitoria", o da dati processati su di esse, ad esempio tramite *binning*.

6.3.1 Estrapolazione di azioni per il RL

Una parte di analisi del database in questa fase è stata incentrata sull'individuazione delle azioni, a partire da dati molto descrittivi e classificati soggettivamente dagli utenti stessi di EPC. L'iter decisionale in AMCO e in tutte le SPV che utilizzano EPC ha un funzionamento non lineare né omogeneo, a seconda dei casi d'uso. L'agente che desidera applicare un'azione di un certo peso sul cliente deve prima farne richiesta scritta tramite una delibera al proprio superiore e solo dopo la sua approvazione può procedere all'azione. Molto spesso le delibere sono lunghi testi in cui viene ripercorsa la storia della pratica in oggetto, dai quali sarebbe difficile ricavare quale sia il fulcro dell'azione. Fortunatamente in EPC l'agente seleziona prima la tipologia di delibera che intende proporre e grazie a quel dato ci si riconduce all'azione principale richiesta dal proponente.

Noto il dato da cui si vuole estrarre l'azione, cioé la fattispecie di delibera legata ad un'azione, il *preprocessing* si occupa di selezionare i primi 5 caratteri e codificare, per mezzo di quest'ultimi, tutte le azioni possibili. È stata classificata, in maniera puntuale per ogni tipologia di delibera, una descrizione funzionale allo scopo di questo progetto. Ad esempio, per la stringa "02.Piano di rientro", si mantiene solo "02.Pia" e lo si decodifica nell'azione *re_entry_plan* che è una delle colonne della matrice Q, cioé un'azione del RL. Questo processo di estrazione delle stringhe avviene non solo per le delibere di cui sopra, ma anche per altre azioni presenti nella tabella "Esiti Affidamenti" che riguardano la revoca dell'affidamento di una pratica o il sollecito di pagamento da parte di un agente, azioni queste che non necessitano di una delibera scritta.

6.4 Validazione del modello: tuning degli iper-parametri

Il punto di partenza del modello vero e proprio di RL si apre con la scelta degli iper-parametri ottimali. A tale scopo, si attua una fase ispirata all'*Imitation Learning*, metodo di apprendimento per casi in cui un agente esperto intraprende azioni, anche sub-ottimali, ma comunque prese come punto di riferimento per l'algoritmo(Stanford). Per un numero limitato di pratiche si valuta quindi quali parametri portano l'algoritmo a consigliare

azioni più simili possibile a quelle scelte dall'agente nel mondo reale. In questo modo, si inizializza il RL ad uno standard realistico, nell'ambito della valorizzazione delle coppie stato-azione. Ciò significa che il Q-Learning impara ad agire in modo simile agli agenti, diventando il più possibile un emulatore di quest'ultimi. Tuttavia, già in questa fase si nota come molte decisioni prese dall'algoritmo si discostino da quelle più comuni intraprese dagli agenti reali nelle medesime condizioni.

La motivazione di questa fase risiede nel fatto che già di per sé il RL *off-policy* presenta un elevata varianza negli stimatori dei ritorni e quindi nei valori della *Q-function*: con questo approccio si cerca di indirizzare l'apprendimento in un verso più realistico, per poi andare a distinguere azioni più esotiche, rispetto alla norma, nelle fasi più avanzate dell'apprendimento, dopo che il RL ha capito intrinsecamente le caratteristiche dell'ambiente in cui deve muoversi.

Gli iper-parametri presenti nel modello vanno quindi selezionati in modo da portare il RL a suggerire strategie che l'agente esperto sarebbe molto propenso a scegliere. I 3 iper-parametri sono:

1. **alpha**: il fattore di apprendimento
2. **gamma**: il fattore di sconto
3. **beta**: il fattore di smorzamento utilizzato nello *smooth Q-Learning*. Esso rappresenta quanto l'informazione calcolata nel passo di *Q-learning* classico influenzi uno stato simile. Più è alto il valore di beta, più un'entrata della matrice Q è influenzata dagli aggiornamenti di stati vicini.

Il collo di bottiglia in questa fase è sicuramente il costo computazionale: è molto dispendioso, a livello di tempo, effettuare tutte le combinazioni possibili di parametri. Il training set deve essere sufficientemente ampio per coprire almeno gli stati più frequenti. La stima per allenare una matrice di un RL di questo tipo su 10000 pratiche in fase di training e lo stesso numero in fase di validazione é di 15 minuti su un computer di media potenza (4GB di RAM, processore i5) per ogni combinazione dei tre iper-parametri.

Prima di partire con la validazione del modello, si precisa che non tutte le pratiche sono adatte al RL, infatti sono state selezionate solamente quelle con almeno un'azione compiuta da un agente. Il totale è di quasi 120000 pratiche. Il tipo di validazione che si è scelto è una *k-fold cross validation* con $k = 6$, in modo da ottenere, per ogni split di validazione, poco meno di 20000 pratiche e circa 100000 per ogni training set. Purtroppo operare un training su così tante pratiche renderebbe lungo il tempo di svolgimento di questa fase e inoltre non è questo lo scopo dell'*imitation learning*, dato che deve solamente indirizzare il RL basandosi su un set limitato di pratiche. Si è scelto, quindi, di campionare sia dal training che dal validation set 10000 pratiche per ciascun insieme.

Un altro problema da affrontare quando si parla di RL off-policy su ambienti non stazionari riguarda la mancanza di un target da parte della policy stessa. Questo porta ad una difficile comprensione della buona riuscita dell'apprendimento, e in particolare della scelta degli iper-parametri. Risulta difficile attuare una serie di simulazioni per valutare il reward medio ottenuto con tutte le combinazioni degli iper-parametri, ma ci si deve affidare a metriche più *custom* e basate su obiettivi non ben definiti.

6.4.1 Metriche di riferimento

Il problema affrontato in questo studio è difficilmente iscrivibile in una categoria di problemi standard. Non è presente un target ideale da raggiungere poiché mentre una pratica può essere considerata di successo se raggiunge il 20% del GBV recuperato, altre pratiche sono considerate dei fallimenti se vengono recuperate in percentuale anche maggiore.

In virtù di questo fatto, si è pensato ad un paio di metriche che potessero essere significative e facilmente ricavabili. Per comprendere le metriche, conviene ricordare che su ogni pratica possono essere intraprese più azioni in sequenza e che lo stato della pratica può cambiare in seguito all'esecuzione di un'azione (ad esempio un recupero che porta la pratica ad un *revenue_over_gbv* più grande).

Su ogni pratica sono registrate una serie di azioni {a,b,c,..} con le loro tempistiche, attuate dagli agenti e il RL le ripercorre una alla volta. Questo processo viene chiamato *trajectory sampling* da esperienze reali, cioé la successione di stati e azioni e l'osservazione dei relativi reward. Grazie a questi elementi tipici del RL, si ha quanto necessario per aggiornare i valori della matrice Q, ma una volta popolata in modo anche sommario quest'ultima serve stabilire quando l'apprendimento porta maggiori frutti in termini di realismo delle azioni. Per fare ciò, si possono applicare due metriche simili tra loro:

1. **single_action_suggested**: èuna metrica che considera la sequenzialità delle azioni di una pratica. Se all'istante t è stata intrapresa l'azione *re entry plan* (proposta di piano di rientro), allora si valuta se questa sia una delle 3 azioni con *q-value* più elevato riferite a quello stato. Se lo è, la metrica aumenta il proprio valore di 1, altrimenti si aggiorna solamente il counter delle azioni totali sulla pratica. Al termine dell'analisi di quella pratica, si divide il valore della metrica per il counter di azioni totali e si ottiene uno score compreso tra 0 e 1. Lo score dà un'indicazione puntuale, stato per stato, di come il RL riesca ad emulare le azioni di un agente esperto nella realtà.

2. **batch_action_suggested**: simile alla metrica precedente, ma considera la pratica nell'insieme e valuta quante delle 3 azioni considerate migliori, in termini di *q-value*, siano state applicate alla pratica, assegnando uno score da 0 (nessuna azione suggerita) a 3(tutte le migliori azioni sono state attuate dall'agente reale). Lo stato della pratica considerato è quello iniziale.

6.4.2 Risultati del tuning

I valori dei 3 parametri da valutare (α, β, γ) hanno dominio $[0, 1]$, ma si può restringere notevolmente il campo grazie ad alcune indicazioni ottenute dalla letteratura e dai primi test effettuati:

1. **alpha**: a seguito di test preliminari, il valore del tasso di apprendimento va mantenuto piuttosto alto. Si è scelto, perciò, di effettuare molti esperimenti su quelli che sono i valori più comuni, cioé $\alpha \in [0.9, 0.95]$. In particolare, si nota come già per piccoli set di pratiche in training (1000 pratiche) i valori delle metriche risultino indirizzati verso valori di alpha più grandi. All'aumentare di α, aumenta il valore di

	date	action	servicer
0	2020-02-05	lawsuit_procedure	other_servicer
1	2020-03-24	extrajudicial_appraiser	other_servicer
2	2021-01-13	lawsuit_procedure	other_servicer
3	2021-04-30	full_and_final_settlement	other_servicer
4	2021-06-14	write_off_with_deferred_payments	other_servicer
5	2022-08-25	full_and_final_settlement	other_servicer

	action
0	credit_renunciation
1	lawsuit_procedure

	action
0	begin_lawsuit
1	agent_warning
2	credit_cession

	action
0	credit_renunciation
1	lawsuit_procedure

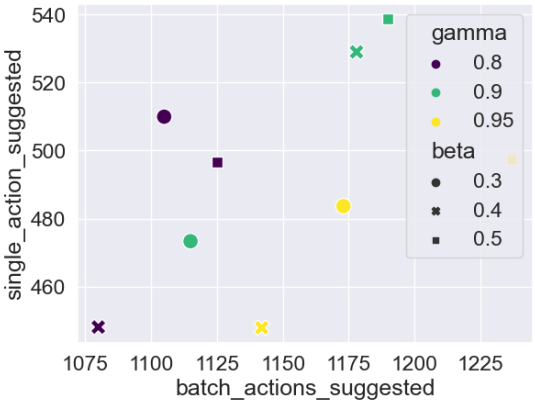
	action
0	begin_lawsuit
1	agent_warning

	action
0	begin_lawsuit
1	agent_warning
2	credit_cession

Figura 6.2: In alto, tutte le azioni prese su una pratica con associate date e servicer per ognuna. Sotto, le 5 azioni suggerite dal *reinforcement learning* nel momento in cui l'agente sceglie come comportarsi nella realtà. Tutti gli stati erano già stati valorizzati, tranne quello corrispondente all'azione *extrajudicial_appraiser*. Si nota come la prima metrica, in nero, componga uno score di 2 azioni suggerite su 5 totali: $single\_action\_suggested = 2/5$. La seconda metrica, applicata solo al primo stato della pratica, ha $batch\_action\_suggested = 1$, dal momento che solo una delle azioni con maggior *q-value* viene scelta dall'agente reale all'interno della pratica, *lawsuit_procedure*

scala di entrambe le metriche in figura 6.3 e ciò significa che l'apprendimento è più performante per valori di α attorno allo 0,9.

2. **gamma**: in letteratura è consigliato associare un fattore di sconto prossimo all'1 (Fernandez and Caarls [2018]). Vengono scelti tre valori su cui effettuare un'analisi: $\gamma \in [0.85, 0.9, 0.95]$. L'indicazione del paper di riferimento sopra citato è confermata anche dalla figura 6.3 in cui i due valori che hanno performance migliori per tutti e tre i valori di α sono $\gamma = 0.9$ e $\gamma = 0.95$, nonostante l'apprendimento avvenga su sole 1000 pratiche. È così giustificata la scelta di utilizzare un fattore di sconto che



(a) $\alpha = 0.7$

(b) $\alpha = 0.8$

(c) $\alpha = 0.9$

Figura 6.3: Varie combinazioni di γ e β per tre valori fissati di α. Le pratiche considerate nei training set sono 1000, così come nei set di validazione

si avvicina all'unità. Un'altra considerazione necessaria riguarda il business a cui è applicato questo algoritmo: essendo in ambito finanziario, uno sconto, detto *daily rate*, viene applicato ad ogni reward ed è dovuto alla distanza temporale del reward in oggetto dal momento di inizio dei lavori sulla pratica, ossia dal *date_first_phase*. Questo tasso potrebbe essere soggetto a fluttuazioni dovute al mercato, portando quindi anche a influenzare il valore degli stati.

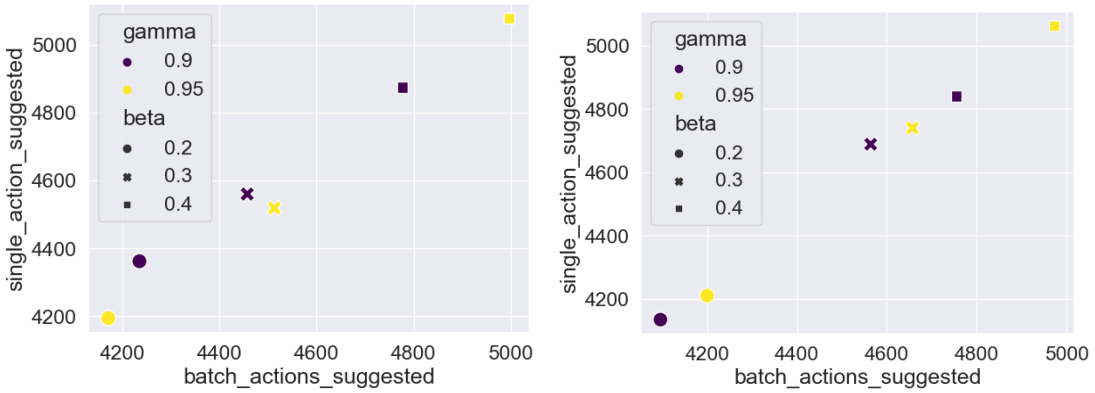
Il valore di gamma sconta ulteriormente i flussi di cassa in modo che l'associazione di un incasso dopo 10 azioni dia poca importanza alla prima azione compiuta, che è stata verosimilmente poco influente nello sviluppo del recupero crediti.

3. **beta**: Conviene escludere dallo studio $\beta > 0.5$, come si può anche logicamente dedurre, perchè i *q-value* provenienti da stati vicini diventano più pesanti del valore attuale dello stato. Si decide perciò di valutare $\beta \in [0.2, 0.3, 0.4]$. Per quanto riguarda questo iper-parametro, non si evince un'indicazione di massima dalle analisi

preliminari in 6.3 e lo si studierà in maniera più approfondita all'aumentare del numero di pratiche nel training set.

Un importante aspetto da considerare è che la varianza nei *q-value* portata dal metodo *off-policy* applicato portano i risultati sui vari set di validazione a dipendere in gran parte dalla composizione dell'insieme stesso. Lo si nota quando, al variare di un singolo parametro, cambiano solo leggermente i valori delle due metriche e molto spesso non cambia la gerarchia di preferenza, come si può evincere dalle figure 6.4 e 7.1, comparando i grafici con *seed* 35. È necessario perciò far variare il seme delle estrazioni random, che vanno a formare i vari *fold*, in modo da valutare in più casi possibili quali parametri portano a risultati mediamente migliori rispetto agli altri.

Aumentando il numero delle pratiche analizzate in fase di training, e limitando il numero di split per la *k-fold cross validation* a 6, contro i 9 utilizzati in fase preliminare (figura 6.3) si trovano i valori delle metriche come definiti nelle figure 6.4 e 7.1.

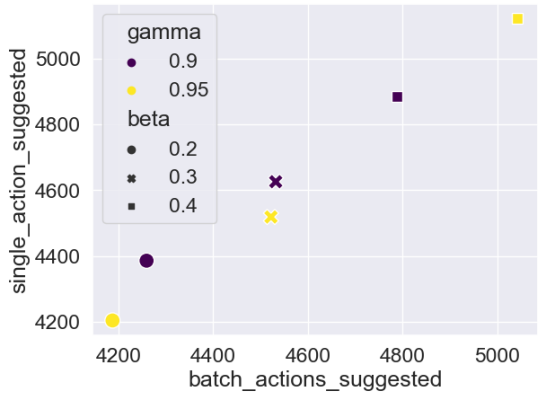


(a) *seed* = 35 (b) *seed* = 46

Figura 6.4: Varie combinazioni di γ e β per $\alpha = 0.9$, al variare del seme del generatore di numeri casuali utilizzato per campionare gli elementi dei *validation set*

In figura 6.4 si osserva come, al variare del seme, cambia abbastanza il range delle metriche, ma le gerarchie rimangono pressoché invariate. Visivamente, c'è chiarezza riguardo i due parametri da scegliere, fissato α. Si può notare infatti che $\gamma = 0.95$ è il valore da scegliere che si guardi sia al valore più alto in assoluto sia alla dominanza quasi assoluta del colore giallo rispetto al viola, separatamente per ogni coppia di parametri. Per quanto riguarda il fattore β di smorzamento, relativo allo *smooth* Q-Learning, l'impressione è che abbia prestazioni peggiori quando $\beta = 0.2$.

Dalla figura 7.1 si mantiene l'impressione che i valori $\beta = 0.2$ e $\gamma = 0.9$ siano meno performanti. Tuttavia, questo è in parte dovuto all'impiego di 2 semi già utilizzati in precedenza, il che non permette alle metriche di discostarsi di molto. D'altra parte, utilizzando lo stesso seme perciò gli stessi insiemi di training e test set, si nota come la variazione del parametro porti a cambiamenti minimi a livello di metriche.



(a) $seed = 35$

(b) $seed = 21$

(c) $seed = 46$

Figura 6.5: Varie combinazioni di γ e β per $\alpha = 0.95$, al variare del seme del generatore di numeri casuali utilizzato per separare i vari set di validazione

Fatte queste considerazioni, secondo cui si può intuire quali siano i parametri migliori senza avere però una significatività chiara a livello statistico, si procede al training sull'intero set di pratiche con i seguenti valori degli iper-parametri:

$$\alpha = 0.95 \quad \beta = 0.4 \quad \gamma = 0.95 \tag{6.1}$$

Una considerazione può essere fatta anche riguardo alle performance generali nei confronti delle metriche per questa tripla di valori nel caso della figura 7.1b. Il totale delle pratiche analizzate in fase di test su ognuno dei 6 *fold* è di circa 10000, su un totale di circa 20000. Considerando a posteriori i valori degli iper-parametri scelti e disegnando un grafico a barre, si ha già un'idea di quanto il RL abbia appreso bene le singole azioni da intraprendere su una pratica, a seconda dello stato. Il grafico 6.6 evidenzia come più di 4000 delle 10000 pratiche del test set vengano emulate in modo preciso dal RL. Bisogna ricordare che, affinché la prima metrica aumenti, serve che l'azione ricavata dall'esperienza sia una delle 3 migliori azioni in termini di *q-value*, riguardo a quello specifico stato della

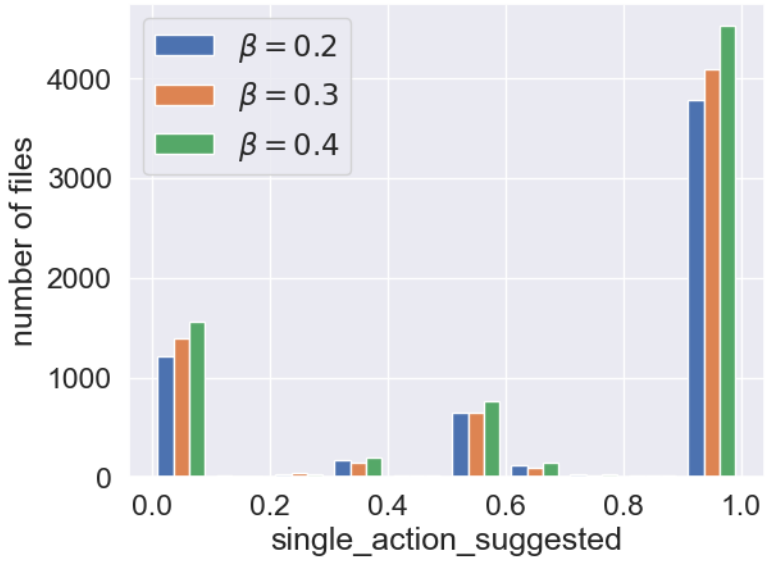


Figura 6.6: $\gamma = 0.95$ $\alpha = 0.95$, mentre β è lasciato libero di variare nei valori indicati nella legenda. Le pratiche su cui si effettua il test sono 10 000

pratica. Risulta infine evidente da questo grafico come (6.6) il valore di $\beta = 0.4$ sia quello con più alta predisposizione ad emulare i comportamenti degli agenti esperti, in quanto domina gli altri nella colonna all'estrema destra riferita a quelle pratiche in cui le azioni sulle pratiche in oggetto rientrano sempre tra le 3 con maggior *q-value*. D'altro canto, si può notare come la dominanza di $\beta = 0.4$ ci sia anche per le pratiche in cui nessuna azione viene emulata correttamente. Questi migliori risultati sono sintomo di un set di validazione più ricco, con meno stati poco visitati e quindi non ancora valorizzati dal RL.

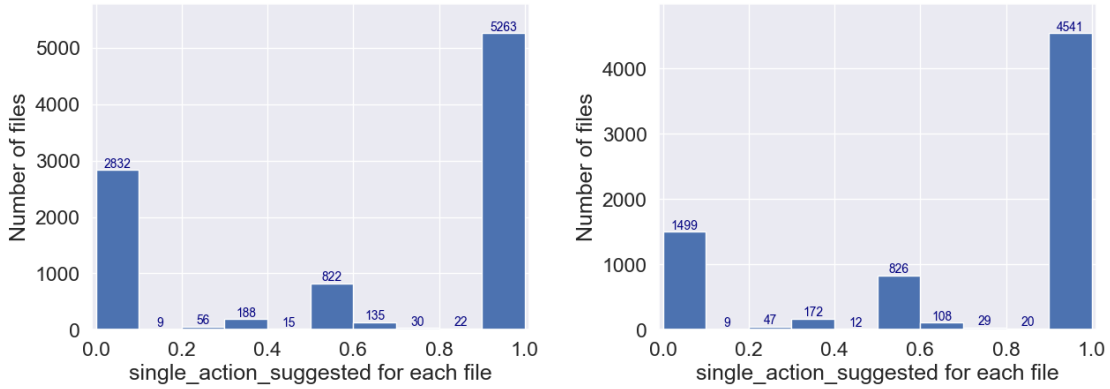
Capitolo 7

Risultati

Dopo aver definito i valori degli iper-parametri per l'algoritmo, è necessario applicarli a tutte le pratiche disponibili su cui sia stata eseguita almeno un'azione ritenuta significativa, nel corso della propria storia. Lo scopo è quello di imparare dalla storia delle pratiche disponibili e di valutare i risultati in termini di interpretabilità e tendenza a superare in termini di performance gli agenti esperti sulle pratiche in cui lo spettro di azioni valorizzate è più ampio. Si ricorda che nella fase di tuning venivano considerate solo 10000 pratiche alla volta, con l'obiettivo di arrivare ad avere i primi valori della matrice Q in modo che fossero più ispirati alle azioni, magari sub-ottimali, degli agenti reali e dei servicer. Un passaggio questo che segue la via tracciata su alcuni paper (Seita [2019]): implementare in prima battuta un *imitation learning* sulle pratiche analizzate per cercare di dare un riferimento iniziale alla matrice Q per quanto riguarda i valori sulle azioni più comuni, ripercorse più volte già in fase di tuning.

Al termine di questa implementazione, per un ulteriore paragone, si è scelto di mantenere lo stesso set di test di 10000 pratiche utilizzato in precedenza, con un seme ben definito (il numero 46), per valutare come cambia il comportamento del RL quando il training set aumenta considerevolmente di dimensione. Partendo proprio da ciò, si osserva come il cambiamento di performance sia sorprendente, da un certo punto di vista. Il RL non ha imparato ad emulare in modo ancora più preciso il comportamento degli agenti, nonostante la direzione imposta nella scelta degli iper-parametri fosse quella. Piuttosto la matrice Q finale suggerisce azioni differenti per migliorare le prestazioni degli agenti stessi. In figura 7.1a si nota infatti come il numero di pratiche per cui il RL imita l'asset manager esperto *in toto* sia maggiore in minima parte rispetto alla fase di tuning: si passa da 4541 a 5263 pratiche emulate in ogni azione rispetto all'esperienza reale. Il numero di pratiche in cui il RL si differenzia completamente dall'agente esperto è invece quasi raddoppiato, passando da 1499 a 2832.

Dal grafico 7.1a si osserva inoltre che il totale delle pratiche su cui il RL riesce ad agire questa volta è di gran lunga superiore, attestandosi a quasi il 99% delle pratiche, lasciandone solamente 167 senza un'indicazione riguardo la prossima migliore azione. Nei casi in cui non si riesce a suggerire nemmeno un'azione, significa che la matrice Q non è valorizzata per nessuna azione in corrispondenza di quello stato. Questo risultato non è sorprendente, dal momento che nonostante le pratiche su cui si sta allenando il motore di



(a) training set sulla quasi totalità delle pratiche

(b) training set su sole 10000 pratiche

Figura 7.1: i valori $\gamma = 0.95$, $\alpha = 0.95, \beta = 0.4$ sono utilizzati per allenare il RL in entrambi i casi, valutando in base alla metrica *single_action_suggested* quante delle 10000 pratiche con *seed* 46 vengono emulate (1.0) oppure in cui il RL sceglie altre azioni(0.0).

RL siano circa 110000, la matrice Q rappresenta ben 311040 stati differenti. Probabilmente continueranno ad esistere particolari stati mai visitati da alcun agente, né aggiornati grazie allo *smooth Q-learning*, in cui non sarà presente un'indicazione riguardo alla decisione da prendere. Ci si aspetta comunque che, aumentando il numero di pratiche su cui si possa allenare l'algoritmo, la matrice venga popolata in modo sempre più capillare e che la previsione di recupero atteso a partire da uno stato sia sempre più precisa e verosimile. Un altro dato osservabile è proprio la copertura della matrice Q, che risulta valorizzata per meno di un campo su 100, ma nonostante ciò riesce a dare indicazioni su una percentuale così alta di stati possibili di una pratica. Infatti gran parte delle pratiche si troverà spesso negli stati più comuni e già percorsi più volte da altri asset manager.

Una delle caratteristiche che spesso viene a mancare nel RL è **l'interpretabilità** dei risultati. Come visto in un paper riguardante decisioni da prendere su debiti privi di garanzia (Mark et al. [2022]), è importante che la decisione che prende un dato agente sia interpretabile *ex-post* da chi subisce l'azione. Ad esempio, se un agente prende la decisione di escutere le garanzie di un piccolo debitore con una situazione di inadempienza probabile, cioé una scadenza non grave dei termini di pagamento, e non lo fa riguardo un grosso debitore in una situazione pesante di sofferenza, il metodo risulta non comprensibile né da un esterno né tanto meno dall'agente. Quest'ultimo molto probabilmente non terrà nemmeno in considerazione l'azione di non muoversi in modo deciso sul secondo debitore appena descritto. All'atto pratico, si vuole analizzare su pratiche con uno stato simile quale tipo di azioni presentino una stima del ritorno finale atteso a partire da quello stato, guardando al valore della matrice Q. Per fare ciò si utilizza un prototipo Streamlit, libreria di Python per applicativi prototipali, simulando la richiesta di azione da suggerire all'utente di recupero crediti.

La struttura del prototipo si compone di una parte dedicata agli input da parte degli agenti, rappresentata dall'immagine 7.2. L'input soddisfa a pieno la richiesta di qualsiasi

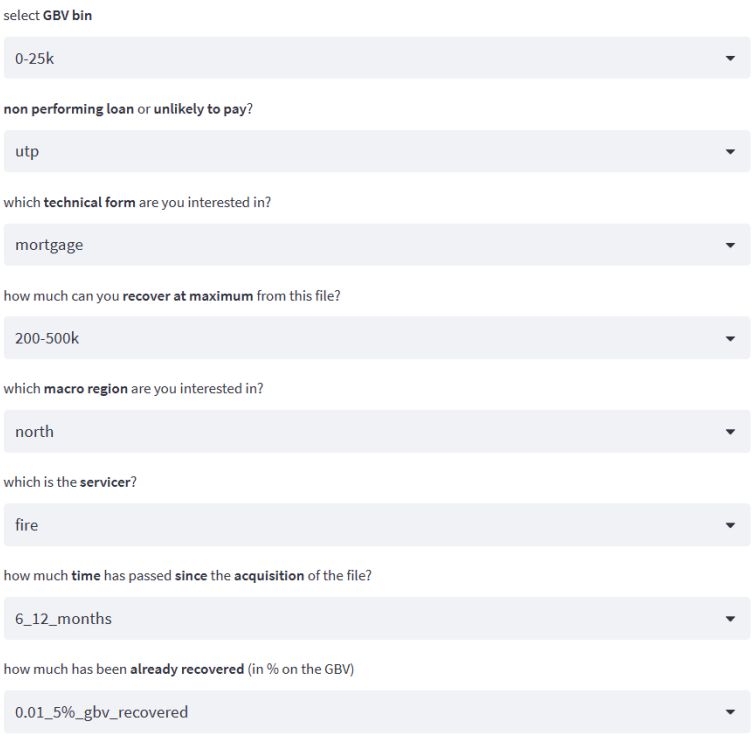


Figura 7.2: Schermata di scelta dei parametri di input da parte dell'agente. Ad ogni cambiamento di uno di questi parametri, il prototipo ricalcola i valori di output in tempo reale

tipo di stato di una pratica in quanto composto di 8 variabili con varie opzioni che servono a comporre uno dei più di 300000 stati. Per comprendere l'interpretabilità di questo algoritmo e valutare quanto e come si discosta dalle tipiche azioni intraprese dagli agenti esperti, ci si serve proprio dell'output di questa schermata.

7.0.1 Interpretabilità del Reinforcement Learning

L'interpretabilità si può evincere dai risultati degli output del prototipo. Si visualizzano nella figura 7.3 i valori di recupero medio atteso su pratiche con variabili di stato impostate in figura 7.2.

Occorre precisare che la scelta degli input è stata effettuata guardando a quegli stati con almeno 8 azioni valorizzate nella matrice Q, portando quindi a stati più visitati e significativi. In questo caso la pratica è un UTP, cioé non ancora una sofferenza. Questo di solito porta gli agenti ad avere una più ampia gamma di azioni a disposizione, essendo percorribile una trattativa extragiudiziale più ampia, basata proprio sul desiderio del debitore di non passare ad uno stato di sofferenza conclamata, con tutte ciò che ne deriva: segnalazione alla centrale rischi e blocco di ogni nuovo rapporto con le banche, come ad esempio la richiesta di nuovi finanziamenti. La figura 7.3 mostra 10 azioni valorizzate

	q-values	description
lawsuit_procedure	87098€	take a key legal procedure (usually real estate execution)
new_recovery_strategy	83286€	propose a generic new strategy of recovery
agent_warning	55109€	send a letter to invite the customer to pay, if he/she is late in doing so
credit_collection	39120€	enforce pledges or warranties
begin_lawsuit	38864€	begin a legal procedure
credits_classification	31097€	pass the loan to non-performing from another state
bank_credit_withdrawal	20716€	revoke the bank credit to the customer
write_off_with_deferred_payments	12979€	propose a final convenient solution to the customer, with deferred payments
confidi_liquidation	9102€	enforce warranties from a consortium (usually Confidi)
full_and_final_settlement	4516€	propose a final convenient solution to the customer in order to recover at least a little part of the debt

Figura 7.3: Output del prototipo fissate le condizioni della figura 7.2. Le azioni non presenti in questa tabella non sono state mai visitate da agenti reali e quindi hanno *q-value* pari all'iniziaizzazione, cioé "-1".

	q-values	description
lawsuit_procedure	35077€	take a key legal procedure (usually real estate execution)
begin_lawsuit	32287€	begin a legal procedure

Figura 7.4: Output del prototipo fissate le condizioni della figura 7.2, tranne la variabile di stato *npe_type* che passa da UTP (inadempienza probabile) a NPL (sofferenza). Le azioni non presenti in questa tabella non sono state mai visitate da agenti reali e quindi hanno *q-value* pari all'inizializzazione, cioé "-1".

di cui la più consigliata, con un recupero atteso di 87000€, è la *lawsuit_procedure*, che comunemente si traduce in una procedura esecutiva immobiliare. Tuttavia, è presente un folto range di azioni che non coinvolgono l'azione legale da parte del servicer in questione, come le due opzioni di saldo e stralcio o il sollecito di pagamento.

Si passa allora alla figura 7.4, che rappresenta l'output delle stesse variabili di input utilizzate in 7.2 ad eccezione del tipo di esposizione, che passa da UTP a NPL, quindi sofferenza. Qui si nota subito come le azioni fornite dal RL si riducono a due: *lawsuit_procedure* e *begin_lawsuit*. Questo significa che l'agente, a quasi parità di condizioni di partenza, cambiando solamente una variabile di stato, riesce a capire come le azioni consigliate diventino diverse e drasticamente indirizzate verso l'escussione delle garanzie presenti nella debitoria del cliente.

7.0.2 Dall'emulazione delle azioni reali, al suggerimento di azioni mai intraprese

Resta da discutere se esistano casi in cui il RL suggerisce azioni molto diverse da quelle intraprese storicamente dagli agenti. Per fare ciò, è stato innanzitutto salvato tutto lo storico delle azioni degli agenti: dato uno stato di partenza, la matrice delle occorrenze conta il numero di volte in cui un'azione è stata intrapresa in quello stato. In questo modo, è possibile risalire a quegli stati visitati più volte.

Come primo esperimento, si è scelto di valutare gli stati con almeno 600 visite. Fissato questo vincolo, 6 possibili stati di una pratica sono risultati conformi. In figura 7.5 sono evidenziate le azioni più scelte dagli agenti in tali stati. A parte alcune visite su azioni minori, in questo tipo di stati gli agenti hanno molto chiaro l'iter da seguire. In 4 casi, concedono quasi sempre condizioni di *forbearance* al debitore: ad esempio speciali agevolazioni, come tassi d'interesse favorevoli o dilazioni dei pagamenti. Negli altri due casi, l'azione scelta principalmente è la revoca della pratica al servicer a cui era stata affidata. Il confronto con le azioni suggerite dal RL si ha guardando alla figura 7.6, analizzando uno stato alla volta:

1. Nel primo stato, il RL consiglia all'agente misure meno leggere della *forberance_or_covid*, azione con più occorrenze nella vita reale. La direzione è quella di attuare una procedura esecutiva sulle garanzie o interrompere il fido del debitore, con ritorno atteso molto più alto rispetto all'azione più popolare tra gli agenti. Questo tipo di approccio è spiegabile perchè il debito ha già alle spalle una storia di più di 2 anni e metodi di trattativa stragiudiziale saranno già stati probabilmente tentati. La terza azione più consigliata è però quella che ha più senso rispetto alle altre ovvero il "cambio di servicer" che in questo caso passerebbe da AMCO ad uno special servicer. Si tratta quasi di routine per la grande società che raramente tiene in lavorazione interna pratiche con GBV sotto i 25000 €.

2. Il secondo stato cambia rispetto al primo solo nella forma tecnica. Passa da un conto corrente ad un credito senza garanzie reali. Il comportamento suggerito rimane lo stesso, con procedure legali e affidamento ad altro servicer tra le azioni consigliate.

3. Il terzo stato è un UTP affidato ad un servicer minore. La decisione nella vita reale è quella di revocare la pratica, visto che il tempo di lavorazione supera già l'anno, senza alcun risultato tangibile. Tuttavia, il RL consiglia di non revocare, bensì passare il credito a sofferenza oppure di revocare il fido dal conto corrente.

4. Nel quarto stato, RL e agenti scelgono a maggioranza la concessione di *forbearance* al debitore. Questo succede, a differenza degli stati precedenti, perchè il debitore ha garanzie piuttosto importanti alle spalle nonostante il piccolo debito. La speranza che possa tornare a pagare con regolarità è alta, anche perché non è in una condizione di sofferenza. D'altra parte, il RL consiglia l'azione di rinuncia ai crediti per evitare di perdere altre risorse su una pratica con uno storico di più di due anni.

5. Il quinto stato viene solitamente trattato con una revoca dell'affidamento, mentre il RL non suggerisce strade con elevati ritorni attesi. A giudicare dagli '0' presenti

nella matrice Q, significa che lo stato viene sì visitato, ma non porta in media ad alcun guadagno.

6. Il sesto stato è quello con valori più alti di debito, oltre i 25000€. Si tratta di pratiche in cui l'agente reale prende molto spesso la decisione di concedere *forbearance*, sbagliando però secondo il RL. L'algoritmo assegna infatti un valore molto negativo a quel tipo di azione, sintomo che le spese in quelle circostanze superano di gran lunga i ricavi. Le azioni consigliate sono invece quelle di procedura legale e affidamento ad altro servicer. Quest'ultima azione si spiega grazie al basso importo del debito.

	0-25k	0-25k	0-25k	0-25k	0-25k	25-100k
	npl	npl	utp	utp	utp	utp
	bank_account	unsecured_loan	bank_account	bank_account	minor_form	bank_account
	0-25k	0-25k	0-25k	25-100k	0-25k	0-25k
	no_macro_region	no_macro_region	no_macro_region	no_macro_region	no_macro_region	no_macro_region
	24+_months	24+_months	12_24_months	24+_months	12_24_months	24+_months
	no_revenue_over_gbv	no_revenue_over_gbv	no_revenue_over_gbv	no_revenue_over_gbv	no_revenue_over_gbv	no_revenue_over_gbv
	amco	amco	other_servicer	amco	other_servicer	amco
begin_lawsuit	0	0	0	0	0	0
claim	0	0	0	0	0	0
re_entry_plan	1	0	1	0	0	5
forbearance_or_covid	3747	1799	0	719	0	1226
outsourcing_withdrawal	0	0	3309	0	1418	0
confidi_liquidation	0	0	0	0	0	0
extrajudicial_appraiser	0	0	1	0	0	2
write_off_with_deferred_payments	0	0	0	0	0	0
bank_credit_withdrawal	0	0	0	4	0	18
credit_cession	0	0	0	0	0	1
credit_renunciation	3	0	0	0	0	0
full_and_final_settlement	0	0	1	0	2	0
credit_collection	0	0	1	0	0	0
lawsuit_procedure	5	2	8	0	0	11
agent_warning	0	0	0	0	1	0
info_request	0	0	0	0	0	0
credits_classification	0	0	7	3	2	25
new_recovery_strategy	0	0	0	0	0	0

Figura 7.5: Sulle colonne, gli stati su cui sono state intraprese almeno 600 azioni da agenti esperti. Sulle righe, i tipi di azioni. Le entrate della matrice rappresentano le occorrenze per ogni coppia (stato, azione)

Come secondo esperimento, si considerano invece quegli stati con più di 9 azioni valorizzate nella matrice Q. Tra questi si selezionano ancora quelli per cui l'azione con più occorrenze nella realtà non è tra le tre con il *q-value* più alto. In tutti questi casi, le occorrenze sono molto basse. La matrice Q è stata quindi pesantemente influenzata da aggiornamenti di stati simili, per i casi in questione:

1. Il primo stato non presenta occorrenze. Su garanzie piuttosto alte (200-500 mila euro) il RL consiglia una procedura esecutiva immobiliare, trattandosi di *mortgage*, cioé mutui immobiliari. Anche il sollecito di pagamento è tra le misure più indicate, poichè il debitore ha già restituito parte del debito, è un UTP e non è passato nemmeno un anno dall'on-boarding.

	0-25k npl bank_account 0-25k no_macro_region 24+_months no_revenue_over_gbv amco	0-25k npl unsecured_loan 0-25k no_macro_region 24+_months no_revenue_over_gbv amco	0-25k utp bank_account 0-25k no_macro_region 12_24_months no_revenue_over_gbv other_servicer	0-25k utp bank_account 25-100k no_macro_region 24+_months no_revenue_over_gbv amco	0-25k utp minor_form 0-25k no_macro_region 12_24_months no_revenue_over_gbv other_servicer	25-100k utp bank_account 0-25k no_macro_region 24+_months no_revenue_over_gbv amco
begin_lawsuit	-1	-1	-1	-1	-1	-1
claim	-1	-1	-1	-1	-1	-1
re_entry_plan	1400	-1	-1	-1	-1	2864
forbearance_or_covid	426	-1	696	2506	-1	-2727720
outsourcing_withdrawal	1793	250	-1	-237	-1	1231
confidi_liquidation	-1	-1	425	-1	0	-1
extrajudicial_appraiser	-1	-153	579	-1	-1	4147
write_off_with_deferred_payments	-1	-1	65	-1	-1	-1
bank_credit_withdrawal	1916	-1	714	1466	0	1366
credit_cession	100	-1	-1	-1	-1	7957
credit_renunciation	1745	-1	-1	2664	-1	2584
full_and_final_settlement	-1	-1	45	-1	0	-1
credit_collection	-1	-1	480	-1	-1	-1
lawsuit_procedure	1807	264	351	353	0	12784
agent_warning	-1	-1	-1	-1	-1	-1
info_request	-1	-1	-1	-1	-1	-1
credits_classification	1381	-1	1390	1033	0	1658
new_recovery_strategy	-1	-1	611	-1	0	-1

Figura 7.6: Sulle colonne, gli stati su cui sono state intraprese almeno 600 azioni da agenti esperti. Sulle righe, i tipi di azioni. Le entrate della matrice rappresentano i valori della matrice Q, allenata sulla quasi totalità delle pratiche

2. Il secondo stato rappresenta pratiche pluriennali, con inadempienze probabili. Storicamente, la scelta più utilizzata è quella della revoca del fido, insieme al passaggio a sofferenza. Quest'ultima è consigliata anche dal RL e spiegabile dal molto tempo passato senza ricevere pagamenti. Tuttavia, la scelta che sembra più indicata è quella del "cambio di servicer", in quanto potrebbe non aver svolto un buon lavoro, in oltre due anni, su pratiche di questo tipo.

3. Il terzo stato presenta 6 occorrenze sulla procedura giudiziale. I valori della matrice Q suggeriscono invece azioni non ben inquadrabili in una strategia precisa. Il ritorno atteso è molto basso, ma la più quotata è la stima di un perito per valutare se esistono garanzie mal valutate o non segnalate sulla pratica. La presenza di sole 6 occorrenze reali mostra come questo caso sia stato pesantemente influenzato dallo *smooth Q-learning*.

4. Il quarto stato ha visto azioni contrastanti da parte degli asset manager: concessioni come moratorie covid o procedure giudiziali. Anche in questo caso, i valori della matrice Q sono molto bassi, infatti si nota la somiglianza con il terzo stato in figura. Gli unici tentativi consigliabili potrebbero essere richiedere un indennizzo alla banca (*claim*) per un vizio di forma, oppure avviare azioni legali, nonostante le prospettive non sembrino essere le migliori.

5. Il quinto stato rappresenta pratiche più recenti rispetto al quarto, e con inadempienze probabili. Il RL consiglia quindi la classificazione del credito a sofferenza come premessa di avvio di atti giudiziali. Anche qui il recupero previsto si attesta attorno ai 20000€, nonostante un debito oltre i 200000€, ma purtroppo la forma tecnica del conto corrente molto spesso non presenta grosse garanzie a copertura. Trattandosi appunto di un conto corrente, un'altra opzione è quella della revoca del fido. Trova spazio anche la rinuncia del credito poiché non è così scontato riuscire a recuperare molto con poche garanzie.

	0-25k	**25-100k**	**100-200k**	**200-500k**	**200-500k**
	utp	**utp**	**npl**	**npl**	**utp**
	mortgage	**bank_account**	**bank_account**	**bank_account**	**bank_account**
	200-500k	**25-100k**	**100-200k**	**200-500k**	**25-100k**
	north	**no_macro_region**	**no_macro_region**	**no_macro_region**	**no_macro_region**
	6_12_months	**24+_months**	**12_24_months**	**12_24_months**	**6_12_months**
	0.01_5%_gbv_recovered	**no_revenue_over_gbv**	**no_revenue_over_gbv**	**no_revenue_over_gbv**	**no_revenue_over_gbv**
	fire	**fire**	**other_servicer**	**amco**	**amco**
begin_lawsuit	0	0	0	0	0
claim	0	0	0	0	0
re_entry_plan	0	0	0	0	0
forbearance_or_covid	0	1	0	12	0
outsourcing_withdrawal	0	0	0	0	0
confidi_liquidation	0	1	0	0	0
extrajudicial_appraiser	0	2	0	0	0
write_off_with_deferred_payments	0	0	0	0	0
bank_credit_withdrawal	0	12	0	0	0
credit_cession	0	0	0	0	0
credit_renunciation	0	0	0	1	0
full_and_final_settlement	0	1	0	0	0
credit_collection	0	0	0	0	0
lawsuit_procedure	0	3	6	6	0
agent_warning	0	0	0	0	0
info_request	0	0	0	0	0
credits_classification	0	9	0	0	0
new_recovery_strategy	0	0	0	0	0

Figura 7.7: Sulle colonne, gli stati su cui sono valorizzate almeno 9 entrate della matrice Q ed in cui l'azione con più occorrenze non si trova tra i 3 *q-value* maggiori. Sulle righe, i tipi di azioni. Le entrate della matrice rappresentano le occorrenze per ogni coppia (stato, azione)

	0-25k	25-100k	100-200k	200-500k	200-500k
	utp	utp	npl	npl	utp
	mortgage	bank_account	bank_account	bank_account	bank_account
	200-500k	25-100k	100-200k	200-500k	25-100k
	north	no_macro_region	no_macro_region	no_macro_region	no_macro_region
	6_12_months	24+_months	12_24_months	12_24_months	6_12_months
	0.01_5%_gbv_recovered	no_revenue_over_gbv	no_revenue_over_gbv	no_revenue_over_gbv	no_revenue_over_gbv
	fire	fire	other_servicer	amco	amco
begin_lawsuit	38864	-1	189	247	1149
claim	-1	-1	-1	205	1350
re_entry_plan	-1	3499	24	91	569
forbearance_or_covid	-1	-1	168	185	410
outsourcing_withdrawal	-1	4912	-1	226	375
confidi_liquidation	9101	586	-1	-1	-1
extrajudicial_appraiser	-1	758	846	189	354
write_off_with_deferred_payments	12978	3269	477	-1	-1
bank_credit_withdrawal	20716	1073	246	92	1505
credit_cession	-1	-1	-1	-1	-1
credit_renunciation	-1	-1	-1	121	7411
full_and_final_settlement	4515	3288	0	-1	-1
credit_collection	39119	1098	-1	-1	-1
lawsuit_procedure	87097	2597	239	114	795
agent_warning	55108	-1	-1	-1	1261
info_request	-1	-1	-1	-1	-1
credits_classification	31097	3413	13	101	20143
new_recovery_strategy	83285	4605	582	-1	-1

Figura 7.8: Sulle colonne, gli stati su cui sono valorizzate almeno 9 entrate della matrice Q ed in cui l'azione con più occorrenze non si trova tra i 3 *q-value* maggiori. Sulle righe, i tipi di azioni. Le entrate della matrice rappresentano i valori della matrice Q, allenata sulla quasi totalità delle pratiche

Capitolo 8

Conclusione e sviluppi futuri

Dagli esperimenti compiuti, si è potuto notare come il RL abbia dato risultati soddisfacenti in quanto a interpretabilità delle azioni consigliate. Si è osservato che, per stati in cui era già presente una forte indicazione da parte degli agenti reali, l'algoritmo ha interpretato i reward ottenuti e ha saputo distinguere quando era giusto seguire l'esempio storicamente contenuto nel database, rispetto a quando era il caso di prendere decisioni diverse.

I risultati sono buoni anche dal punto di vista della copertura della matrice. Basti pensare al fatto che il 99% delle pratiche presenti nel test set presentano valori non nulli di *q-value* per almeno un'azione, dando quindi, almeno in parte, un'indicazione riguardo la strategia da adottare.

Gli sviluppi possibili per un progetto di questo tipo possono viaggiare in direzioni perpendicolari tra loro, in quanto si potrebbe ampliare molto. Sono state fatte alcune importanti assunzioni basate più sul ragionamento che sui dati, come la scelta degli iperparametri.

Uno sviluppo importante si avrebbe sicuramente estendendo le variabili di stato al fine di garantire ancora più diversità tra le pratiche e migliorare quelle già esistenti. Questo comporterebbe un aumento ancor più massiccio della matrice Q, richiedendo una rivisitazione dell'algoritmo. Un importante miglioramento si potrebbe avere inserendo una rete neurale che traduca le variabili discrete dello stato in layer intermedi e infine in azioni discrete da intraprendere. Credo che questo gioverebbe in prestazioni, anche se il numero di pratiche non dovesse aumentare; tuttavia, è possibile che l'interpretabilità dei risultati diventi più debole, difetto proprio delle reti neurali in tutte le applicazioni.

In una direzione leggermente diversa, si potrebbe pensare di aumentare le prestazioni passando da un modello tabulare a uno parametrico, basato cioè su approssimazioni lineari, quali polinomi, o attraverso aggregazione di stati simili, il cosiddetto *tile coding*.

Infine, durante l'analisi del database, si nota fin da subito che tutto il sistema dei cash-flow non è funzionale allo svolgimento di un algoritmo di machine learning, in quanto spesso i dati sono mancanti o associati in modo complicato alle pratiche, rendendo difficile la valutazione delle azioni su alcune di esse. Capita, ad esempio, che alcuni pagamenti siano registrati in modo errato, o su tabelle diverse da quelle considerate, ma difficilmente identificabili perchè magari ricavabili soltanto da una descrizione testuale di un qualche attributo della pratica.

Bibliografia

URL www.ghostcfo.it/formazione/articoli/crediti-deteriorati.

I crediti deteriorati (non-performing loans - npls) del sistema bancario italiano, giugno 2017. URL www.bancaditalia.it/media/views/2017/npl/.

Anthony Bellotti and Damiano Brigo. Forecasting recovery rates on non-performing loans with machine learning. agosto 2019.

Mehdi Fatemi, Mary Wu, Jeremy Petch, and Walter Nelson. Semi-markov onine reinforcement learning for healthcare, 2022.

Franklin Cardeñoso Fernandez and Wouter Caarls. Parameters tuning and optimization for reinforcement learning algorithms using evolutionary computing, 2018.

Dott. Luigi Iannilli. Non-performing exposures: definizioni e regolamentazione.

Wei Liao, Xiaohui Wei, and Jizhou Lai. Smooth q-learning: accelerate convergence of q-learning using similarity.

Po-Ling Loh and Maxim Raginsky, editors. *When Is Partially Observable Reinforcement Learning Not Scary?*, 2022.

Michael Mark, Naveed Chehrazi, Huanxi Liu, and Thomas A. Weber. Optimal recovery of unsecured debt via interpretable reinforcement learning. *Elsevier*, 2022.

Daniel Seita. Combining imitation learning and reinforcement learning using dqfd, aprile 2019. URL danieltakeshi.github.io/2019/04/30/il-and-rl/.

Stanford. Principles of robot autonomy. URL web.stanford.edu/class/cs237b/pdfs/lecture/lecture_10111213.pdf. School of Engineering, Lecture 10-Imitation Learning.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, Cambridge, Massachussets, 2018.