Tecniche di Apprendimento Mimetico per il Controllo di Sistemi di Refrigerazione

Stefano Zorzi

Dipartimento di Ingegneria

Università degli studi di Padova

Relatore:
Prof. Alessandro Beghi

Correlatore:
Ing. Mirco Rampazzo

Corso di Laurea Magistrale in Ingegneria dell'Automazione

Anno accademico 2014-2015

20 Aprile 2015

Nella vita non contano i passi che fai, né le scarpe che usi,
ma le impronte che lasci.

Sommario

I settori del condizionamento dell'aria e della refrigerazione (HVAC& R, Heating, Ventilating Air Conditioning and Refrigeration), sono tra i più energivori a livello globale, e hanno subito, nel corso degli ultimi anni, un vero e proprio cambio di paradigma a livello tecnologico e una evoluzione costante tesa al risparmio energetico e alla salvaguardia dell'ambiente. La maggior parte delle tecnologie che utilizzano refrigeranti naturali e funzionano con CO_2, idrocarburi, ammoniaca, acqua ed aria etc., possono ormai coprire tutti i tipi di applicazioni, dalla refrigerazione e condizionamento domestico, fino alle applicazioni per il settore commerciale, e i processi industriali. Un ruolo cruciale in questo tipo di applicazioni è svolto dai sistemi di controllo che devono assicurare che, ad esempio, grandezze come temperatura, pressione, umidità, etc. soddisfino assegnate specifiche limitando al tempo stesso i consumi energetici. Il progetto di un sistema di controllo può essere eseguito sfruttando metodi differenti, ad esempio: approcci model-based, che si giovano di un modello dinamico del sistema in esame, oppure modelli data-driven (model-free), che sfruttano l'informazione di misure presenti nell'impianto per descriverne il comportamento. In questa tesi si affronta il progetto di un sistema di controllo gerarchico basato su una architettura a due livelli (livello di supervisione e livello relativo ai controllori locali) per un sistema di refrigerazione commerciale (supermercato) che impiega l'anidride carbonica come refrigerante naturale. Il supervisore ha il compito di impostare in modo adeguato i valori di set-point per le variabili di interesse che poi devono essere regolati dai controllori locali. In particolare, si utilizza un approccio con apprendimento mimetico (Reinforcement Learining), di tipo data-driven, per la sintesi del controllore di alto livello (supervisore), mentre i controllori di basso livello sono regolatori standard. L'obiettivo del sistema di controllo è duplice: da un lato assicurare che le derrate siano conservate correttamente all'interno dei frigoriferi, dall'altro che il consumo di energia sia adeguato, al fine di assicurare prestazioni elevate del sistema. Per il progetto prestazioni del sistema di controllo e valutarne le prestazioni si fa riferimento ad un ambiente di simulazione (sviluppato utilizzando il software di calcolo scientifico Matlab/Simulink) relativo ad un sistema di refrigerazione a CO_2 per un supermercato costituito da un impianto di produzione del freddo a ciclo di compressione di vapore e da celle frigorifere di media

e bassa temperatura. Il lavoro di tesi è organizzato nel seguente modo. Nel primo capitolo è presentata una breve introduzione sui sistemi HVAC&R e al problema del loro controllo, nonché le metodologie comunemente utilizzate per il progetto di controllori. Nel secondo capitolo si illustra il modello di simulazione in ambiente Matlab/Simulink utilizzato per la progettazione del controllore e la verifica delle prestazioni. Il terzo capitolo è dedicato all'apprendimento mimetico e agli algoritmi di controllo ad esso ispirati. Nel quarto capitolo sono illustrati il progetto del sistema di controllo e le prestazioni dello stesso.

Ringraziamenti

Vorrei iniziare ringraziando il Professore Alessandro Beghi, relatore di questa tesi, per l'aiuto fornito e per la sempre grande disponibilità e cortesia dimostratami. Un ringraziamento speciale a Mirco Rampazzo. In questi lunghi mesi sei stata una guida esperta e illuminante, ti ringrazio veramente tanto per la tua grande pazienza, per i tuoi consigli mirati e per aver reso il lavoro di tesi un periodo piacevole e divertente. Non sarei riuscito a concludere il mio percorso di studi senza l'aiuto di molte persone. Vorrei cominciare ringraziando Andrea e Matteo che mi sono stati vicini nei primi tre anni, insieme abbiamo vissuto una triennale divertente e spensierata. Vi ringrazio per il vostro sostegno e aiuto. Un ringraziamento particolare a Giuliano, Precious e Marco. Vi ringrazio per tutte le ore passate insieme a studiare, per la pazienza avuta, per la vostra disponibilità, per la vostra gentilezza e per tutti i momenti divertenti vissuti. In voi ho trovato dei veri amici. Vorrei ringraziare la mia famiglia che mi ha accompagnato in questo cammino, i miei genitori Renato e Nilla per il sostegno morale. Un grazie anche a mia sorella Elisabetta, mio cognato Alberto e i miei tre splendidi nipoti Lorenzo, Ada Angela e Camilla che sono stati fonte di distrazione e rilassamento nei momenti difficili. Infine vorrei ringraziare la persona che più di tutti mi è stata vicina in questi lunghi anni scolastici. Grazie per la tua infinita pazienza, per i saggi consigli e per la tua presenza costante ed equilibratrice. Per questo ti ringrazio Tamara. Vorrei anche ringraziare tutti i compagni di studi e le persone che mi hanno aiutato in questi anni.

Indice

Elenco delle figure

Elenco delle tabelle

Capitolo 1

Introduzione

Il controllo di un sistema HVAC&R di grandi dimensioni è una sfida stimolante poichè vengono coinvolte dinamiche non lineari soggette a vincoli ed è necessario trovare un metodo per soddisfare due aspetti fondamentali:

- Consumo di energia del sistema
- Perdità di qualità degli alimentari conservati

Le tecniche di controllo possono essere di tipo model-based basate su un modello oppure model-free dove non è richiesto un modello accurato del sistema ma esclusivamente i dati ingresso-uscita. I supermercati sono un esempio di edifici commerciali che consumano un quantitativo di energia molto elevato. Il consumo di energia all'interno di un supermercato di grandi dimensioni ha una fabbisogno di energia per il riscaldamento, il condizionamento, l'illuminazione e la conservazione degli alimenti. Tale richiesta energetica dipende anche dalle tecniche di costruzione degli edifici e delle caratteristiche climatiche. Vale la pena sottolineare che gli utilizzi diversi necessitano di una qualità dell'energia strettamente connessa alle tecnologie utilizzate. Per esempio i gruppi frigoriferi classici a compressione di vapore, richiedono lavoro, cioè energia della qualità più elevata, sia sotto forma di lavoro meccanico sia di elettricità. L'energia consumata dipende dal numero e dalle tipologie delle componenti interne al supermercato (e.g celle frigorifere, compressori, condensatori, etc); in questo contesto i sistemi di refrigerazione (controllo temperatura banchi frigo e celle frigorifere, etc.) consumano più del 50% dell'energia totale, come si può vedere in figura 1.1. Per garantire che i diversi tipi di derrate siano conservate in maniera adeguata, le temperature all'interno dei frigoriferi devono essere mantenute all'interno di valori di riferimento, per esempio:

- Alimenti surgelati, temperatura massima a -18°C

- Uova, temperatura massima a 12°C
- Carne o Carne macinata, temperatura massima a 5°C
- Pesce fresco, temperatura massima a 2°C
- Latte, temperatura massima a 5°C

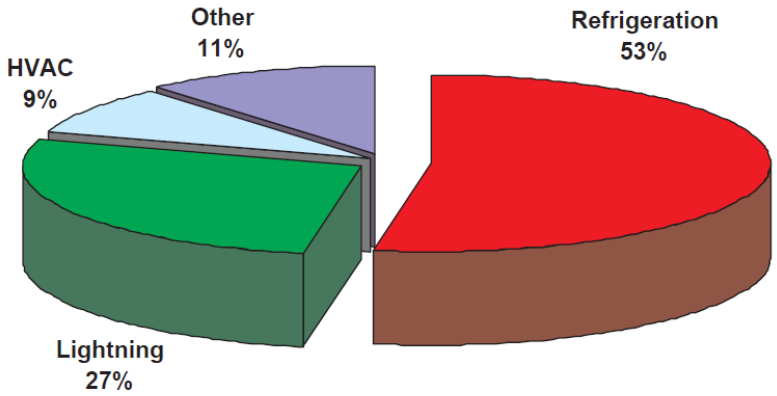

Figura 1.1 Ripartizione consumi energia elettrica tipici di un edificio commerciale

Un buon controllo dell'impianto aiuta a mantenere il cibo alle temperature corrette e a minimizzare i costi energetici relativi alla refrigerazione. Il sistema non solo deve essere in grado di funzionare bene in condizioni normali, ma deve anche essere in grado di mantenere a un livello minimo il rischio di deterioramento degli alimenti (e.g scongelamento, variazioni improvvise di temperatura etc.). Negli ultimi anni, soprattutto in ambito industriale, la continua richiesta di energia elettrica e l'introduzione di risorse rinnovabili hanno complicato tale problema poichè non è più sufficiente ottenere un ingresso di controllo che soddisfa determinate caratteristiche ma è necessario dare importanza anche ai consumi energetici. Il concetto di demand-response management è un aspetto molto importante per la gestione del consumo di energia, sia in campo industriale che privato. Un tipico andamento della domanda di energia elettrica per un edificio commerciale è rappresentata in figura 1.2 .
Il profilo riportato in figura 1.2 mostra l'energia consumata dagli impianti di illuminazione, raffreddamento (refrigerazione), ventilazione, e altri carichi in una giornata all'interno di un edificio commerciale come un supermercato. Il consumo energetico relativo alla refrigerazione in questa tipologia di edificio aumenta considerevolmente durante le ore

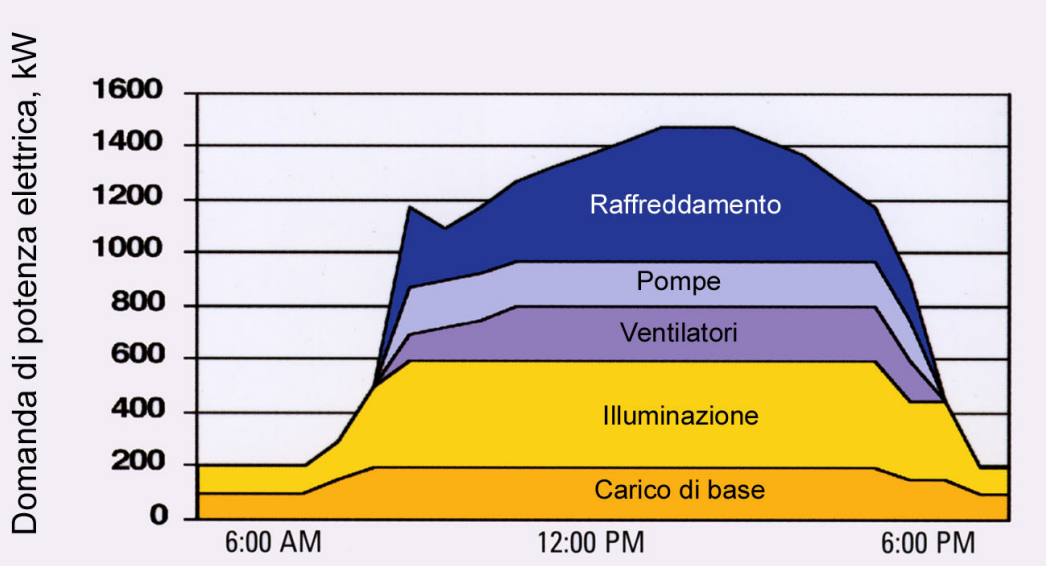

Figura 1.2 Andamento domanda elettrica giornaliero in un tipico edificio commerciale

centrali della giornata quando l'edificio è aperto al pubblico. La maggiore richiesta di energia dipende dall'aumento della temperatura all'interno del supermercato e all'interno dei banchi frigo. L'aumento di temperatura è dovuta alla presenza di clienti nell'edificio e alla continua apertura e chiusura delle celle frigo. Diminuire il consumo energetico di questi edifici commerciali rappresenta un mezzo importante per porre un limite al consumo energetico primario e ridurre l'inquinamento dovuto alle emissioni di gas nell'atmosfera. Uno degli obiettivi principale del futuro è lo sfruttamento più efficiente dell'energia, con l'introduzione di strategie (insieme a tecnologie) orientate al risparmio energetico in ogni fase, dalla produzione, alla trasmissione e alla distribuzione sia in ambito industriale che commerciale. L'energia elettrica è la "materia prima" disponibile più degradabile: infatti deve essere consumata subito dopo essere stata prodotta, ma deve anche essere prodotta nel momento in cui è richiesta. Solitamente la domanda sfugge a ogni controllo: per esempio, negli anni passati, i sistemi di refrigerazione commerciale hanno aumentato la produzione all'aumentare della domanda (ad esempio in riferimento all'andamento in figura 1.2 nelle ore centrali della giornata) e hanno ridotto i ritmi produttivi quando la domanda cala (nelle ore notturne). Con la tecnica del demand-response la domanda viene gestita in modo più attivo, permettendo un bilanciamento nella richiesta di energia. Per fare questo è possibile controllare i sistemi di refrigerazione, sulla base delle condizioni e dei prezzi, indirizzando i consumi non essenziali su periodi dove la richiesta di energia è minore oppure nei periodi in cui il costo dell'energia è più basso. Per esempio sempre in riferimento ai consumi di un sistema di refrigerazione, è possibile far fronte alla domanda di energia frigorifera necessaria

con un accumulo frigorifero realizzato nelle ore notturne così da diminuire la domanda durante il giorno come si può vedere in figura 1.3. Il "freddo" accumulato durante le ore notturne può essere distribuito nell'arco della giornata a seconda delle necessità in aggiunta alla capacità offerta dal sistema di refrigerazione.

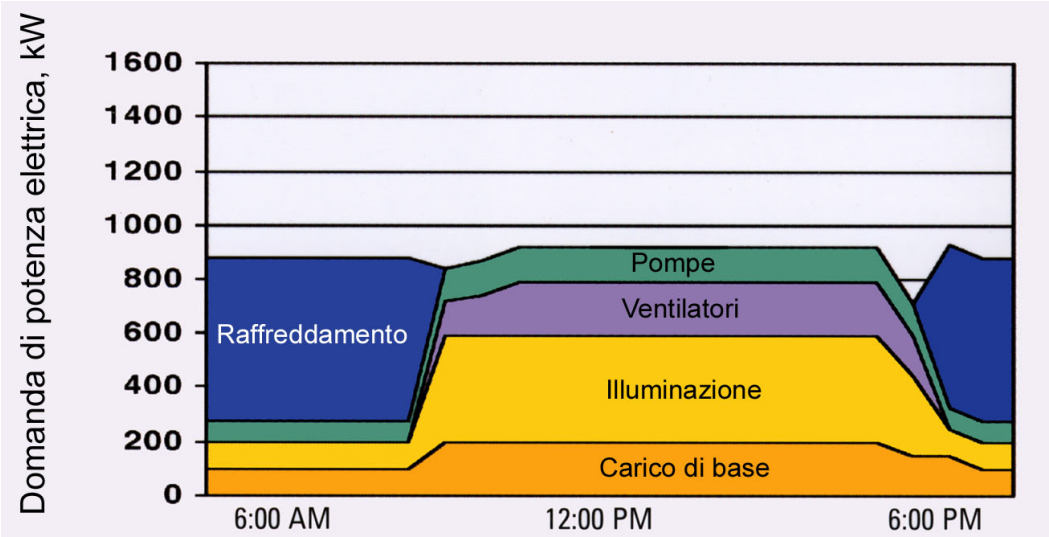

Figura 1.3 Andamento domanda elettrica giornaliero in un tipico edificio commerciale

Impostare i valori dei set-point di temperatura e pressione delle celle frigorifere nel sistema di refrigerazione è un aspetto cruciale non solo per la conservazione degli alimenti ma anche per il risparmio energetico. Le componenti principali all'interno del sistema di refrigerazione sono le celle frigorifere, i compressori, i condensatori e le valvole di espansione il cui funzionamento e la dinamica verranno presentati nel capitolo 2. In un sistema di refrigerazione solitamente le variabili controllate sono le temperature all'interno delle celle frigorifere, le pressioni dei compressori e dei condensatori presenti. La pressione del condensatore è controllata azionando a diverse velocità le ventole presenti all'interno del condensatore (nel nostro caso si ipotizza che si muovono sempre alla velocità massima), le temperature all'interno delle celle sono regolate da valvole termostatiche (e.g ad isteresi ON/OFF) mentre la pressione di aspirazione delle celle è regolata attraverso la frequenza dei compressori. Nella gestione di un sistema di refrigerazione possono sorgere problemi relativi alla perdita di refrigerante oppure può formarsi del ghiaccio sulla serpentina dell'evaporatore (situato all'interno della cella frigorifera) che possono causare:

- un malfunzionamento del compressore
- una breve conservazione degli alimenti

- un impianto di refrigerazione poco efficiente
- consumi energetici maggiori rispetto al normale funzionamento
- set-point di pressione e temperature delle celle non ottimali

Per risolvere il problema della formazione del ghiaccio sulla serpentina dell'evaporatore è necessaria una manutenzione accurata dell'impianto. In sistemi di refrigerazione avanzati esiste il controllo automatico delle operazioni di sbrinamento degli evaporatori. Il relativo sistema di controllo è costituito da due componenti, la prima segnala il livello di congelamento dell'evaporatore e inizia il processo di sbrinamento mentre la seconda componente ha il compito di eseguir le operazioni del processo. Per evitare la formazione di ghiaccio nell'evaporatore si possono applicare diverse strategie:

- Sbrinamento elettrico: la fusione del ghiaccio nell'evaporatore è causata dalla presenza di resistenze elettriche opportunamente posizionate.
- Sbrinamento a gas caldo: il gas caldo (l'apporto di calore è fornito dalla condensazione del refrigerante scaricato dal compressore) surriscaldato viene iniettato all'interno dell'evaporatore subito a valle della valvola termostatica

L'evoluzione tecnologica ha portato all'introduzione di nuove tecniche sempre più complicate a seconda del sistema considerato. Anche l'architettura di controllo negli ultimi anni, con l'obiettivo di gestire al meglio i consumi di potenza, ha subito una cambiamento passando da un'architettura centralizzata ad una struttura distribuita. Il sistema di controllo è centralizzato se esiste un unico centro di controllo cui fanno capo tutte le decisioni e di conseguenza tutte le informazioni necessarie ad assumere tali decisioni. Allo scopo di superare i limiti dell'architettura centralizzata, in molti sistemi si è implementata una forma gerarchica di controllo basata sull'idea di ripartire l'insieme delle responsabilità decisionali in livelli di importanza e di associare ciascun livello ad un determinato piano nella gerarchia del sistema. Localizzati sulle singole macchine, semplici controllori si limitano a eseguire i comandi ricevuti dal regolatore centrale. Tipicamente i sistemi di refrigerazione dei supermercati presentano una struttura di controllo di tipo gerarchico dove sono presenti:

- un controllo di alto livello (supervisore) che ha il compito di decidere i valori dei set-point di pressione e di temperatura delle singole componenti (celle frigorifere) per consumare un determinato valore di potenza.
- controllori locali con il compito di mantenere invariati i valori dei set-point forniti dal controllo di alto livello.

La struttura del controllo per un sistema di refrigerazione è riassumibile nel seguente grafico:

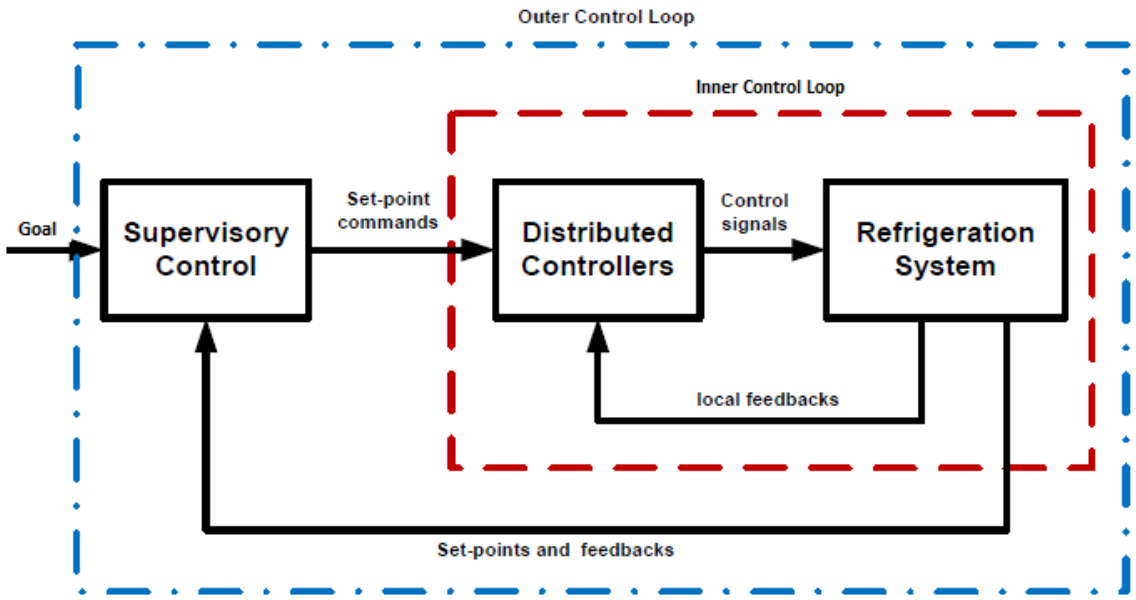

Figura 1.4 Struttura di controllo di un sistema di refrigerazione

Per la progettazione del controllo di alto e basso livello all'interno di un sistema di refrigerazione possono essere utilizzate diverse tipologie di controllori.

1.1 Controllori standard

I controllori standard sono PID insieme ai controllori ON/OFF sono molto diffusi. I controllori ON/OFF utilizzano una soglia inferiore ed una superiore per regolare il processo e l'ingresso di controllo può assumere un range di valori limitato. I controllori PID utilizzano l'errore dinamico e regolano una o più variabili per ottenere un accurato controllo del processo. I controllori standard sono sfruttati in numerosi campi grazie alla loro semplicità, al basso costo e alla facilità di implementazione. Per esempio trovano utilizzo tra le applicazioni relative ai sistemi HVAC&R, per il controllo di unità in un sistema di raffreddamento, per regolare la temperatura in una stanza, regolare la pressione all'intero di una cella frigorifera e altre applicazioni relative alla meccanica come il controllo di posizione. In [1] gli autori realizzano un controllo decentralizzato per un sistema di refrigerazione commerciale dove i controllori locali sono dei controllori ON/OFF che regolano il livello di apertura delle valvole di evaporazione mentre il controllo supervisore, che determina i set-point dei controllori locali, è udi tipo PI. Uno dei problemi principali nell'utilizzo dei PID sta nel fatto che è

richiesta una fase di taratura dei parametri del controllore che non sempre è semplice ed immediata. Può succedere che le performance del regolatore degradino con il passare del tempo poichè si possono verificare differenti condizioni al contorno rispetto alle condizioni rilevate in fase di taratura. E' necessario intervenire effettuando una nuova taratura dei parametri ma non sempre questo è possibile e dipende fortemente dal processo preso in considerazione. I controllori ON/OFF sono sicuramente più semplici ma il controllo del processo risulta essere meno preciso rispetto ad altri controllori a causa dei limitati valori (solitamente due) che può assumere la variabile di controllo.

1.2 Controllori non standard

Oltre ai controllori standard esistono anche altre tipologie di tecniche di controllo più complicate basate su:

- Controllo non lineare
- Controllo robusto
- Controllo ottimo
- Model Predictive Control (MPC)
- Controllori PID a parametri variabili

Per il controllo non lineare la regola di controllo viene derivata dalla teoria della stabilità di Lyapunov e dalla feedback linearization. La legge ricavata viene utilizzata per portare il sistema non lineare da uno stato arbitrario iniziale verso uno stato stazionario. Le tecniche non lineari sono efficienti ma necessitano di una identificazione degli stati stazionari, di una conoscenza generale del processo e di una complessa analisi matematica per la progettazione del controllore. In [2] gli autori presentano un metodo basato sulla feedback linearization per il controllo del clima all'interno di una serra per le operazioni di ventilazione, raffreddamento e idratazione. Il controllo robusto permette di ottenere prestazioni efficienti anche in presenza di disturbi tempo varianti e variazioni dei parametri. Il controllo ottimo risolve un problema di ottimizzazione minimizzando una funzione di costo. La determinazione della funzione di costo è molto importante e dipende dal sistema e da che aspetto si vuole ottimizzare, i consumi di potenza, le prestazioni, etc. Per esempio, in un sistema HVAC&R l'obiettivo dell'ottimizzazione può essere quello di minimizzare il consumo di energia del sistema rispettando i vincoli imposti. In [3] l'autore utilizza delle tecniche di controllo ottimo per regolare la temperatura all'interno di una stanza dimostrando come le prestazioni ottenute

con il controllo ottimo siano migliori, da un punto di vista del risparmio energetico, rispetto a quelle di un semplice controllore PI. Nel controllore a parametri variabili un sistema non lineare è suddiviso in due regioni lineari e per ogni regione vengono progettati due controllori PI o PID differenti ognuno con i propri parametri. Per questa tecnica di controllo è necessario identificare le regioni lineari e progettare uno switch logico per passare da una regione all'altra. Le tecniche del controllo robusto e del controllo non lineare sono affidabili poichè sono in grado di resistere ai disturbi e alle variazioni dei parametri garantendo prestazioni efficienti indipendentemente dai disturbi e dal deterioramento delle componenti del sistema. Il Model Predictive Control (MPC) è una tra le tecniche di progettazione maggiormente utilizzate. MPC utilizza un modello del sistema per determinare lo stato futuro e generare un vettore di controllo che minimizza una funzione di costo su un orizzonte finito in presenza di disturbi e vincoli. Solo il primo elemento del vettore viene utilizzato come ingresso di controllo mentre i restanti vengono scartati e il procedimento viene ripetuto ad ogni istante temporale. In [4] MPC viene applicata ad un sistema di refrigerazione commerciale per minimizzare il consumo di energia rispettando i vincoli imposti dalla temperatura per conservare gli alimenti. Con l'avvento di controllori digitali si sono sviluppate nuovi controllori come quelli basati sulla logica fuzzy (FL) e le reti artificiali neurali (ANN). Le reti neurali artificiali costruiscono un modello matematico non lineare basandosi esclusivamente sui dati di input/output del sistema ed effettuando una modellizzazione black box poiché nella determinazione del modello non è richiesta la conoscenza della fisica del processo. Le reti neurali artificiali possono essere utilizzate per il controllo feed-forward al posto di un controllore standard. La logica fuzzy è concettualmente semplice e permette di progettare un controllore in grado di manipolare dati affetti da disturbi. Per la progettazione del controllore FL è necessaria la conoscenza della dinamica del sistema che non sempre è semplice da ottenere mentre per le reti neurali artificiali sono indispensabili dati numerosi e significativi.

1.3 Altre tipologie di controllori

Tra le altre tipologie di controllo le più importanti sono la DFL (Direct feedback linearization), PMAC (Pulse Modulation Adaptive Control), PRAC (Pattern Recognition Adaptive Controller) e il Reinforcement learning (RL). Lo scopo del controllo DFL è di ottenere un disaccoppiamento tra diversi cicli di controllo e ottenere una stabilità globale del sistema. L'obiettivo del controllo DFL è di ottenere delle leggi di controllo disaccoppiate per diverse componenti del sistema al fine di raggiungere una stabilità globale. Le equazioni accoppiate vengono disaccoppiate attraverso la linearizzazione ingresso-uscita e controllate con tecniche di controllo lineare. Il controllo DFL può essere utilizzato per regolare la temperatura all'in-

terno di una stanza, come si può vedere da [5] i risultati ottenuti sono migliori rispetto al PID sia da un punto di vista del risparmio energetico che da un punto di vista della reiezione ai disturbi.

Il controllo PMAC viene utilizzata in sistemi discreti che accettano in ingresso solo due valori: tipicamente ON/OFF. Con questi sistemi la regolazione dell'uscita può avvenire solo attraverso lo switch degli ingressi ad una determinata frequenza. Queste leggi di controllo devono raggiungere un buon trade-off tra la frequenza di switching e la regolazione dell'output poichè aumentando la frequenza miglioro il controllo ma sovraccarico maggiormente il sistema. PMAC consente di progettare una legge di controllo digitale partendo da un ingresso continuo, per esempio in [6] l'uscita di un PID e viene convertita in un ingresso digitale per il sistema discreto.

Il controllo PRAC regola automaticamente i parametri di un controllore PI basandosi su uno schema che caratterizza la risposta del sistema. L'uscita del sistema viene misurata e inviata al controllore PRAC che stima il disturbo e modifica i parametri del PI in modo da regolare il segnale in uscita. Per esempio in [8] tale tecnica di controllo viene applicata ad un sistema HVAC&R ottenendo delle soluzioni quasi ottime con una buona resistenza ai disturbi.

L'apprendimento mimetico (Reinforcement Learning) è una tecnica che ricava l'ingesso di controllo attraverso la continua interazione tra il sistema e il controllore. Le prestazioni ottenute con l'apprendimento mimetico sono comparabili ai risultati ottenuti con i regolatori standard ma presenta il vantaggio di essere una tecnica di controllo model-free e non necessità di una conoscenza dettagliata del sistema.

Riassumendo le caratteristiche dei controllori utilizzati per i sistemi HVAC&R si può osservare:

- I controllori PID hanno bisogno di taratura e non sono sempre robusti ai disturbi tempo varianti mentre i regolatori ON/OFF hanno un controllo approssimativo

- I controllori non standard hanno bisogno di una analisi matematica accurata, sono complicati e necessitano di un identificazione degli stati stazionari

- I controllori basati sull'utilizzo delle reti neurali necessitano di un grande quantitativo di dati non sempre disponibili

- Controllori basati sull'apprendimento mimetico (Reinforcement Learning) hanno il vantaggio che non serve una conoscenza dettagliata del modello del sistema (model-free) ma il quantitativo di informazioni e di tempo necessarie per le tecniche di apprendimento (learning) possono essere significativi.

- Si nota una ulteriore suddivisione tra controllori che necessitano di una conoscenza del modello del processo da controllare (PID,ON/OFF e controllori non lineari) e controllori che hanno bisogno di dati ingresso-uscita per apprendere dall'esperienza (RL e reti neurali artificiali).

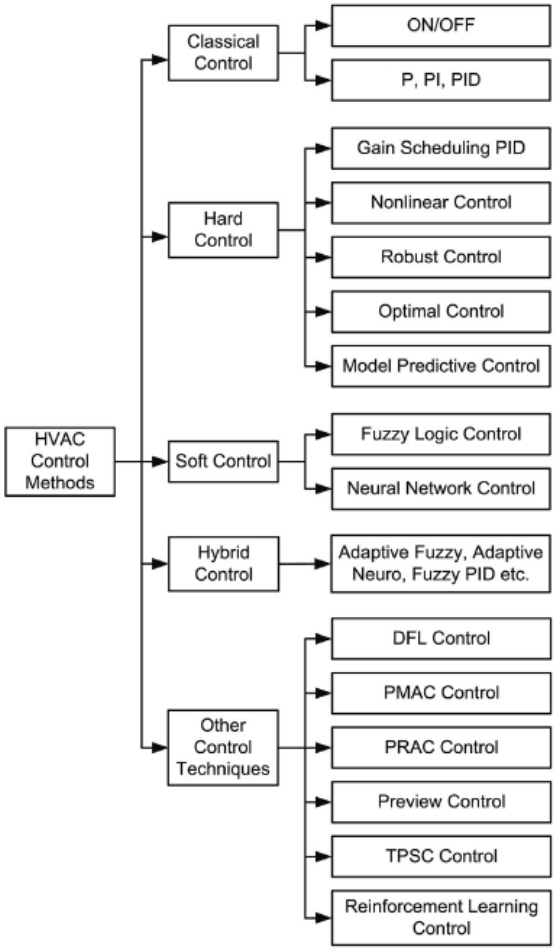

Figura 1.5 Tipologie controllori

Per esempio nei grandi sistemi di refrigerazione industriale l'implementazione di un controllore model-based, come MPC, richiede lo sviluppo di un modello accurato attraverso una procedura non semplice e molto dispendiosa da un punto di vista temporale. L'impiego di un controllore basato sui dati (model-free) come il Reinforcement Learning (RL) può essere più semplice e "imparando dall'esperienza" è in grado di auto-regolarsi in caso di variazioni delle condizioni iniziali e disturbi tempo varianti. In [10] gli autori confrontano i risultati ottenuti con un controllore basato su tecniche di apprendimento mimetico e con il Model Predictive Control su un sistema al fine di stabilizzare la potenza elettrica. I risultati ottenuti

con le due tecniche di controllo sono simili ma MPC ottiene prestazioni meno robuste con il vantaggio di avere una precisione maggiore rispetto all'apprendimento mimetico. Non esiste un controllore ottimale da utilizzare in qualsiasi applicazione ma la scelta del controllore da progettare dipende dal processo preso in considerazione. In sistemi dove non è richiesta una precisione elevatissima e i tempi di azione del processo sono piuttosto lenti è possibile progettare un controllore basato sui dati. Al contrario in sistemi dove la precisione è molto elevata e i tempi di reazione del sistema devono essere piuttosto rapidi allora è preferibile impiegare del tempo inizialmente per realizzare un modello del processo e solo successivamente progettare il controllore. Un sistema di refrigerazione commerciale come un supermercato è costituito da diverse componenti (celle frigorifere, compressori, valvole di espansione,etc.) e le singole componenti sono variabili sia per numero che per caratteristiche. Proprio a causa della diversità e della complessità delle singole componenti è molto difficile costruire un modello valido per ogni sistema di refrigerazione. La progettazione di un controllore model-based può risultare complicata in quanto il tempo speso inizialmente per la costruzione di un modello abbastanza accurato può essere molto elevato. L'approccio model-free mi permette di evitare la costruzione di un modello e di utilizzare come informazioni per determinare l'ingresso di controllo solo i dati ingresso-uscita del sistema. L'obiettivo del lavoro di tesi è quello di progettare un controllore di alto livello per un sistema di refrigerazione commerciale al fine di mantenere un livello di potenza elettrica assorbita. Oltre a mantenere la potenza, il controllo dovrà essere in grado di rispettare i vincoli di temperatura all'interno dei singoli scomparti. Il controllo di alto livello verrà prima implementato attraverso tecniche model-based impiegando i regolatori PID e successivamente attraverso tecniche model-free come l'apprendimento mimetico. Il controllo di alto livello (supervisore) ha il compito di assegnare i set-point di pressione e di temperatura delle rispettive celle frigo per regolare il consumo di potenza. Il sistema di refrigerazione, oltre al controllo supervisore, presenta dei controllori locali che mantengono con una certa precisione i set-point forniti dal controllo di alto livello. Nel capitolo successivo si introduce un modello di simulazione dinamico (sviluppato in Matlab/Simulink) relativo al sistema di refrigerazione di un supermercato. Questo modello sarà usato per progettare e vedere le prestazioni dei diversi tipi di architetture di controllo presentate

Capitolo 2

Sistema di refrigerazione commerciale

2.1 Descrizione del sistema

Negli ultimi anni i refrigeranti naturali sono diventati sempre più importanti dato che i refrigeranti sintetici, come CFC, HCFC e HFC sono tra i principali responsabili del surriscaldamento globale. L'anidride carbonica (CO_2), tra tutti i refrigeranti naturali, è un fluido che può operare in un ciclo di compressione del vapore attorno allo 0 °C e presenta i seguenti vantaggi rispetto ai refrigeranti sintetici:

- l'anidride carbonica è facilmente reperibile ad un costo piuttosto basso da scarti industriali.
- i danni ambientali provocati dalla CO_2 sono di gran lunga inferiori rispetto a quelli degli altri fluidi frigorigeni.
- É una sostanza non tossica, non infiammabile e inodore.
- Per quanto riguarda le proprietà termodinamiche presenta un'elevata conduttività termica, una densità maggiore rispetto agli altri fluidi frigorigeni, un elevato calore specifico[1] ed un elevato calore latente[2] volumetrico rispetto agli altri refrigeranti sintetici.

L'utilizzo di questo nuovo refrigerante naturale ha preso piede soprattutto nel settore della refrigerazione commerciale.
Una tipica configurazione di un sistema di refrigerazione per un supermercato di tipo CO_2

[1] Il calore specifico di una sostanza è la quantità di calore necessaria per aumentare di 1 °C la temperaura di 1 Kg della medesima sostanza

[2] Quantità di energia (quantitativo di calore) necessaria al passaggio di stato

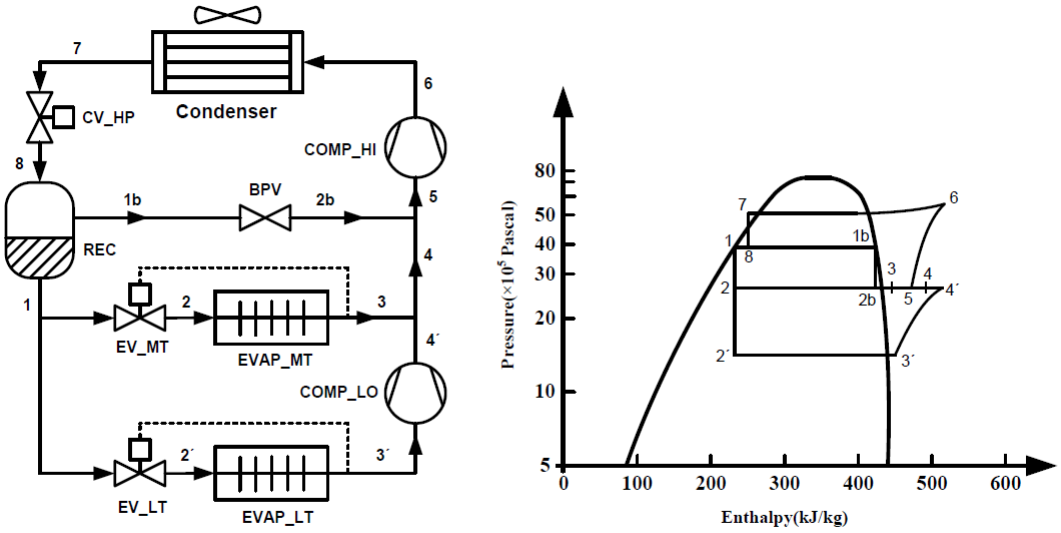

Figura 2.1 Sistema di refrigerazione a CO_2

Booster è rappresentata in figura 2.1(a). In figura 2.1(b) è rappresentato il diagramma pressione-entalpia (P-H) che descrive il ciclo termodinamico del sistema. Nella scala orizzontale è presente l'entalpia del fluido mentre nella scala verticale sono presenti i valori di pressione assoluta in pascal. L'entalpia costiuisce la quntità di calore contenuta in 1 Kg di una determinata sostanza ad una certa temperatura. L'entalpia aumenta all'aumentare della temperatura e diminuisce con il decrescere di questa. Per esempo per il gas refrigerante R22 l'entalpia a 0 °C è di 100 kcal/kg, mentre quella del vapor a 40 °C è di 152 kcal/kg. Le condizioni di entalpia costante sono rappresentate da curve verticali mentre quelle a pressione costante da rette orizzontali. Analizzando la figura 2.1(b) si possono identificare tre zone. La campana nel diagramma è l'insieme dei punti dove il liquido refrigerante è in saturazione a seconda dell'entalpia e della pressione a cui è sottoposto. A sua volta la campana può essere suddivisa in 2 parti:

- Curva di liquido saturo, dove il fluido ha il massimo valore di entalpia per ogni determinata pressione
- Curva di vapore saturo, dove il vapore ha il massimo valore di entalpia per ogni pressione

Le due curve si incontrano nel punto critico (vertice superiore della campana), dove il liquido refrigerante si trova indifferentemente sia nello stato liquido che gassoso. La zona racchiusa dentro la campana è la regione di miscuglio liquido-vapore, dove il refrigerante si trova sia

allo stato liquido che allo stato gassoso. All'esterno della campana sulla destra si trova la regione del vapore surriscaldato dove, a parità di pressione, la temperatura del refrigerante è superiore a quella di saturazione. Il termine di vapore surriscaldato rappresenta un gas la cui temperatura è superiore al punto di ebollizione. Sempre all'esterno della campana sulla sinistra si trova la zona di liquido sottoraffreddato, cioè con temperatura inferiore a quella di saturazione.
Il modello di figura 2.1(a) è costituito dalle seguenti componenti:

- ricevitore
- una valvola di espansione per ogni cella
- celle frigo MT (medium temperature)
- celle frigo LT (low temperature)
- compressore MT
- compressore LT
- condensatore
- valvola di By-Pass
- valvola per il controllo della pressione in ingresso al ricevitore

2.1.1 Descrizione Componenti Principali

- Il ricevitore di liquido non è altro che un recipiente contenente il liquido refrigerante allo stato liquido e gassoso inserito nel circuito frigorifero a valle del condensatore. La valvola di bypass garantisce il continuo funzionamento del sistema di refrigerazione infatti consente di mantenere il flusso di refrigerante necessario a garantire il raffreddamento del compressore ermetico o semiermetico.
- Le valvole di espansione termostatica regolano l'iniezione liquida di refrigerante nell'evaporatore. L'iniezione è regolata dal surriscaldamento del liquido frigogeno. La valvola termostatica può essere utilizzata in circuiti di raffreddamento o di riscaldamento. Nel caso di utilizzo in circuiti di raffreddamento, la valvola avrà un grado di apertura minore in corrispondenza di valori di temperatura più bassi e maggiore in corrispondenza a temperature più elevate. Sotto una temperatura minima (soglia

Figura 2.2 Esempio di un ricevitore (sx) e di una valvola di ByPass (dx)

d'intervento) sarà completamente chiusa mentre sarà completamente aperta per temperature ritenute ottimali per il sistema considerato. Nel caso invece di circuiti di riscaldamento il suo funzionamento sarà inverso, per cui l'apertura della valvola sarà maggiore in corrispondenza di valori di temperatura più bassi e minore in corrispondenza di valori di temperatura più alti.
Le valvole di espansione possono essere di due tipi:

- Valvole termostatiche meccaniche
- Valvole termostatiche elettroniche

Figura 2.3 Valvola termostatica per il controllo del liquido refrigerante

- Il compressore è una delle componenti principali del sistema e ha il compito di aumentare la pressione del liquido refrigerante riducendone il volume. Il compressore è caratterizzato dal rapporto di compressione che è il rapporto tra la pressione in uscita del compressore e la pressione assoluta di ingresso. I compressori si suddividono in due importanti categorie:

  1. Compressori volumetrici

2. Compressori dinamici

Nei compressori volumetrici, detti anche compressori a camere chiuse, la compressione avviene chiudendo, dopo l'aspirazione, il gas in un ambiente isolato, una camera, il cui volume viene progressivamente ridotto; il processo di compressione non è continuo ma ciclico.

Figura 2.4 Esempio di compressore volumetrico

A loro volta i compressori volumetrici si suddividono in:

1. Compressori alternativi
2. Compressori rotativi

I compressori alternativi sono costituiti da un cilindro nel quale scorre a tenuta uno stantuffo. Il moto alternativo di quest'ultimo è ottenuto mediante un meccanismo di biella e manovella. I compressori rotativi sono costituiti da capsulismi di vari tipi. Molto impiegati sono i compressori a palette, con i quali si raggiungono rapporti di compressione di 4-6. I compressori a ingranaggi sono formati da due rotori controrotanti costituiti da profili coniugati fra loro e con la superficie interna dello statore; il volume della camera chiusa è costante. Al contrario i compressori dinamici effettuano la compressione sfruttando l'energia cinetica impressa al gas da opportuni meccanismi. I compressori dinamici, detti anche turbocompressori o compressori a canali aperti, sono costituiti da un rotore palettato, in cui il gas viene accelerato, con susseguente incremento dell'energia cinetica e della pressione, e da uno statore anch'esso palettato, in cui il gas è decelerato con un ulteriore aumento della sua pressione. Il processo di compressione, che avviene nei canali delimitati dalle palettature, è continuo. A loro volta i compressori dinamici si suddividono in:

1. Compressori centrifughi
2. Compressori assiali

Nei compressori centrifughi l'ingresso del fluido è assiale e le palette gli imprimono un'accelerazione centrifuga. Per elevati rapporti di compressione si ricorre a compressori a più stadi, ottenendo rapporti di compressione dell'ordine delle decine. I compressori centrifughi vengono utilizzati nel campo dell'industria chimica e nel trasporto dei gas. Il compressore assiale è una turbomacchina operatrice a flusso assiale, dove il fluido refrigerante si comprime mentre scorre parallelamente all'asse di rotazione. Rispetto al compressore centrifugo gestisce maggiori portate a parità di superficie frontale, ma con un minore rapporto di compressione per singolo stadio, il che implica la necessità di disporre di più stadi in serie, potendo raggiungere, in questo modo, rapporti di compressione di alcune decine

- Il condensatore ha il compito di smaltire il calore immagazzinato dal liquido refrigerante durante il ciclo frigorifero. I fluidi ai quali il calore sottratto durante il ciclo può essere ceduto sono l'aria oppure l'acqua. Nel caso venga impiegata acqua questa può riscaldarsi e/o evaporare. Il condensatore ad acqua più conosciuto è lo scambiatore a fascio tubiero, dove l'acqua circola all'interno di tubi e il fluido frigorifero da condensare dal lato mantello. Il condensatore si dice ad acqua fluente se l'acqua viene eliminata dopo essere stata riscaldata dal condensatore. Il condensatore ad aria è costituito da una batteria di tubi alettati, entro i quali circola il fluido da condensare, esposta direttamente all'aria, in circolazione naturale o in circolazione forzata.

Figura 2.5 Esempio di condensatori ad aria

- Le celle frigorifere e i banchi frigo permettono la perfetta conservazione delle derrate alimentari mantenendo la loro temperatura costante. Questi due componenti rappresentano due elementi indispensabile all'interno di supermercati e magazzini. Ne esistono di tanti modelli, che si differenziano tra di loro per la capienza, per la disposizione, per la struttura. I banchi frigo possono essere orizzontali o verticali. Negli ultimi anni, sia

per banchi frigo orizzontali che verticali, sono stati introdotti nel mercato banchi frigo dotati di pannelli di chiusura per ridurre i consumi e proteggere l'ambiente.

Figura 2.6 Esempi di banchi frigo verticali e orizzontali

2.1.2 Funzionamento ciclo frigorifero

Il refrigerante che circola nel sistema, al punto 8, è costituito sia da liquido che da vapore e il ricevitore divide il refrigerante in liquido saturo (1) e gas saturo (1b). Il gas saturo passa attraverso la valvola di bypass (BPV) mentre il liquido saturo uscente dal ricevitore entra nelle valvole di espansione. Le valvole di espansione EV_{MT} e EV_{LT} sono regolate da 2 controllori locali ON/OFF ad isteresi e hanno il compito di mantenere costante la temperatura all'interno delle celle MT e LT rispettivamente. Il fluido frigogeno in ingresso alle celle MT (2) ha pressione superiore rispetto a quello in ingresso alle celle LT (2'). Passando attraverso gli evaporatori ($EVAP_{MT}$ e $EVAP_{LT}$), il refrigerante assorbe il calore dalle celle MT e LT. Successivamente sia l'entalpia che la pressione del refrigerante uscente dalla cella LT vengono incrementati dal compressore $COMP_{LO}$ da 3' a 4. L'intero flusso di refrigerante uscente dal compressore $COMP_{LO}$, dall'evaporatore $EVAP_{MT}$ e dalla valvola di Bypass viene raccolto nel suction manifold nel punto 5 dove la pressione e l'entalpia sono ulteriormente aumentate dal compressore $COMP_{HI}$ (punto 6). Infine il gas refrigerante entra nel condensatore per cedere il calore assorbito lungo il ciclo all'ambiente esterno e la sua entalpia decresce da 6 a 7 e la pressione diminuisce leggermente. All'uscita del condensatore è presente una valvola (CV_{HP}) che regola la pressione ad un valore più basso cosi da ottenere un flusso di refrigerante in ingresso al ricevitore costituito sia da liquido che da gas.

2.2 Modellizzazione

Come si può vedere in figura 2.1(a) il modello può essere suddiviso in tre sottosistemi:

- celle frigorifere
- suction manifold (collettore di aspirazione) e compressori
- condensatore

Il ricevitore e la valvola di controllo della pressione CV_{HP} non vengono modellati poichè la pressione intermedia all'uscita del ricevitore è costante.

2.2.1 Celle Frigorifere

All'interno delle celle il calore delle derrate $\dot{Q}_{foods/dc}$ viene trasferito inizialmente dagli alimenti all'evaporatore, e successivamente dall'evaporatore al fluido refrigerante. Il calore che viene trasferito dall'evaporatore al fluido refrigerante viene detto cooling capacity e si indica con $\dot{Q}_e$. Nella progettazione del modello bisogna anche prendere in considerazione il calore dell'ambiente circostante, $\dot{Q}_{load}$, che viene trattato come un disturbo. Le equazioni della dinamica del sistema sono le seguenti:

$$MCp_{foods}\frac{dT_{foods}}{dt} = -\dot{Q}_{foods/dc} \tag{2.1}$$

$$MCp_{dc}\frac{dT_{dc}}{dt} = \dot{Q}_{load} + \dot{Q}_{foods/dc} - \dot{Q}_e \tag{2.2}$$

dove MCp è il prodotto della massa per il calore specifico, T_{foods} è la temperatura degli alimenti all'interno della cella e T_{dc} è la temperatura all'uscita dell'evaporatore. I flussi di energia sono:

$$\dot{Q}_{foods/dc} = UA_{foods/dc}(T_{foods} - T_{dc}) \tag{2.3}$$

$$\dot{Q}_{load} = UA_{load}(T_{indoor} - T_{dc}) \tag{2.4}$$

$$\dot{Q}_e = UA_e(T_{dc} - T_e) \tag{2.5}$$

dove UA è la conducibilità termica mentre T_{indoor} è la temperatura all'interno del supermercato che è considerata come un disturbo. Il coefficiente UA può essere descritto come una

funzione lineare della massa di refrigerante liquido nell'evaporatore:

$$UA_e = k_m \cdot M_r \tag{2.6}$$

dove k_m è un valore costante. La massa del refrigerante M_r non può superare il limite $M_{r,max}$ ed è soggetta alla seguente dinamica:

$$\frac{dM_r}{dt} = \dot{m}_{r,in} - \dot{m}_{r,out} \tag{2.7}$$

dove $\dot{m}_{r,in}$ e $\dot{m}_{r,out}$ sono rispettivamente la variazione di massa del refrigerante in ingresso e in uscita all'evaporatore. Il flusso del refrigerante in ingresso all'evaporatore è regolato dalla valvola di espansione e dipende dal grado di apertura:

$$\dot{m}_{r,in} = OD \cdot KvA\sqrt{2\rho_{suc}(P_{rec} - P_{Suc})} \tag{2.8}$$

dove OD è il livello di apertura della valvola che nel nostro caso può assumere solo 2 valori, 0 se la valvola è completamente chiusa e 1 se è aperta. P_{rec} è la pressione del ricevitore e P_{suc} è la pressione del suction manifold. La densità del refrigerante che circola nel circuito è ρ_{suc} e KvA è un parametro costante dipendente dalle caratteristiche della valvola di espansione. La massa di refrigerante che lascia l'evaporatore è:

$$\dot{m}_{r,out} = \frac{\dot{Q}_e}{\Delta h_{lg}} \tag{2.9}$$

dove Δh_{lg} è il calore specifico latente del refrigerante nell'evaporatore ed è una funzione non lineare della P_{suc}

2.2.2 Suction manifold

Per la modellizzazione del suction manifold si utilizza come variabile di stato la pressione P_{suc}:

$$\frac{dP_{suc}}{dt} = \frac{\dot{m}_{dc} + \dot{m}_{dist} - \dot{V}_{comp}\rho_{suc}}{V_{suc}d\rho_{suc}/dP_{suc}} \tag{2.10}$$

La serie dei compressori è trattata come un unico compressore virtuale, $\dot{m}_{dc}$ è il flusso di massa totale delle celle, $\dot{m}_{dist}$ è il flusso di massa di disturbo che contiene al suo interno il fusso di massa delle celle a bassa temperatura (LT) e della valvola di bypass, mentre V_{suc} è il volume del suction manifold.

$\dot{V}_{comp}$ è il volume del flusso in uscita dal suction manifold:

$$\dot{V}_{comp} = f_{comp}\eta_{vol}V_d \tag{2.11}$$

dove f_{comp} è la frequenza del compressore in percentuale, V_d rappresenta la variazione del volume e η_{vol} è l'efficienza approssimata dello spazio volumetrico da:

$$\eta_{vol} = 1 - c((\frac{P_c}{P_{suc}})^{1/\gamma} - 1) \tag{2.12}$$

dove c è una costante volumetrica, γ un'esponente costante adiabatico e P_c è la pressione in uscita al compressore. Per quanto riguarda il consumo di potenza elettrica; essa è calcolata come:

$$\dot{W}_{comp} = \frac{1}{\eta_{me}}\dot{m}_{ref}(h_{o,comp} - h_{i,comp}) \tag{2.13}$$

dove $\dot{m}_{ref}$ è il flusso di massa totale nel suction manifold e $h_{o,comp}$ e $h_{i,comp}$ sono rispettivamente i valori delle entalpie all'uscita e all'ingresso del compressore. Il termine η_{me} rappresenta l'efficienza meccanica totale che tiene in considerazione anche l'attrito. L'entalpia del refrigerante all'ingresso del suction manifold è maggiore rispetto a quella in uscita dall'evaporatore poichè nel suction manifold bisogna tenere in considerazione anche il flusso di refrigerante proveniente dalla valvola di ByPass e dalle celle a bassa temperatura LT. L'entalpia in uscita è esprimibile come:

$$h_{o,comp} = h_{i,comp} + \frac{1}{\eta_{is}}(h_{o,is} - h_{i,comp}) \tag{2.14}$$

dove $h_{o,is}$ è l'entalpia in uscita quando il processo di compressione è isoentropico[3] e η_{is} è la relativa efficienza isoentropica[4]

$$\eta_{is} = c_0 + c_1(f_{comp}/100) + c_2(P_c/P_{suc}) \tag{2.15}$$

con c_i valori costanti.

2.2.3 Condensatore

La modellizzazione del condensatore è piuttosto delicata poichè è fortemente influenzata dai dettagli fisici e dalle dimensioni dei componenti. Analizzando la figura 2.1(b) si nota come il

[3]Un processo si definisce isoentropico se avviene ad entropia costante.

[4]Il rendimento isoentropico di un compressore è definito come il rapporto tra il lavoro necessario per compiere la compressione in maniera isoentropica e il lavoro reale necessario

condensatore operi in 3 regioni distinte:

- zona di vapore surriscaldato, situata all'esterno della campana, sulla destra. In questa zona, a parità di pressione, il vapore presente una temperatura superiore a quella di saturazione.
- zona di miscuglio liquido-vapore, situata all'interno della campana.
- zona di liquido sotto raffreddato, situata all'esterno della campana, sulla sinistra. Il liquido si trova ad una temperatura inferiore rispetto a quella di saturazione pur essendo alla stessa pressione.

Nel passaggio attraverso la prima regione (vapore surriscaldato) si verifica un leggero abbassamento di pressione:

$$\Delta P_c = P_c - P_{cnd} = (\frac{\dot{m}_{ref}}{A_c})^2(\frac{1}{\rho_{cnd}} - \frac{1}{\rho_c}) + \Delta P_f \tag{2.16}$$

dove A_c è l'area del condensatore mentre P_{cnd} e ρ_{cnd} sono rispettivamente la pressione e la densità all'uscita della regione di surriscaldamento. Il primo termine a destra dell'equazione (2.16) indica l'accelerazione della variazione di pressione mentre il secondo termine è costante. Il calore ceduto all'ambiente esterno per quanto riguarda la zona di surriscaldamento e di sotto raffreddamento è descritto dalla seguente equazione:

$$\dot{Q}_{c,k} = UA_{c,k}\frac{T_{i,k} - T_{o,k}}{ln[\frac{T_{i,k} - T_{outdoor}}{T_{o,k} - T_{outdoor}}]} \qquad k = \{1,3\} \tag{2.17}$$

Per quanto concerne invece la zona mista (liquido refrigerante sia allo stato liquido che gassoso) l'equazione relativa al calore ceduto all'ambiente esterno (2.17) diventa:

$$\dot{Q}_{c,2} = UA_{c,2}(T_{i,2} - T_{outdoor}) \tag{2.18}$$

dove UA_c è la conducibilità termica totale delle corrispondenti zone del condensatore, T_i e T_o sono le temperature in ingresso e in uscita di ogni regione e $T_{outdoor}$ è la temperatura esterna. Si può osservare di come la temperature d'ingresso e di uscita della zona mista è la stessa quando la pressione non varia. Il calore trasferito dal refrigerante attraverso la k-esima zona è:

$$\dot{Q}_c = \sum_{k=1}^{3} \dot{Q}_{c,k} \tag{2.19}$$

I parametri costanti utilizzati nelle formule precedenti sono espressi nelle seguenti tabelle:

A_c	ΔP_f	$UA_{c,1}$	$UA_{c,2}$	$UA_{c,3}$
0.0073	0.52	332	3185	148

Tabella 2.1 Parametri presenti nel modello

Cella	UA_{load}	$UA_{foods/dc}$	$MCp_{dc} * 10^5$	$MCp_{foods} * 10^5$	k_m	$KvA * 10^{-6}$
1	41.9	72.9	1.9	4.6	141.7	2.33
2	56.3	82.6	4.8	88.6	250.9	2.27
3	57.5	118.5	2.7	2.8	175.3	2.71
4	32.2	230.5	4.1	2.7	182.9	0.86
5	36.1	158.9	6.3	1.5	810.6	2.38
6	58.2	75.3	1.7	6.5	196.8	2.33
7	24.1	150.2	4.6	60	800	0.70

Tabella 2.2 Parametri per le celle MT

Cella	UA_{load}	$UA_{foods/fr}$	$MCp_{fr} * 10^5$	$MCp_{foods} * 10^6$	k_m	$KvA * 10^{-6}$
1	23	185	1.7	84.5	184.4	2.97
2	28	250	1.9	38	218.9	1.28
3	4	83	0.7	52	395.9	0.54
4	17	250	7.7	20.8	237.2	2.47

Tabella 2.3 Parametri per le celle LT

2.2.4 Condizioni nominali

Il sistema di refrigerazione è costituito da 7 celle a media temperatura MT con temperatura pari a 2.5°C, 4 celle a bassa temperatura LT con temperatura pari a -21°C. Le temperature devono mantenersi all'interno in un determinato intervallo al fine di permettere la conservazione degli alimenti.

La variazione di temperatura dal valore nominale non deve essere superiore ad 1°C:

- $1.5°C \leq T_{i,MT} \leq 3.5°C \qquad i = 1,...,7$

- $-22°C \leq T_{i,LT} \leq -20°C \qquad i = 1,...,4$

La temperatura interna del supermercato (T_{indoor}) viene assunta costante e pari a 26 °C e anche la temperatura esterna ($T_{outdoor}$) è costante pari a 12 °C. Il quantitativo di derrate è diverso per ogni cella, come si può vedere nelle tabelle precedenti (2.2, 2.3) dal parametro MCp_{foods}, e si assume costante.
L'architettura di controllo per il sistema di refrigerazione è di tipo gerarchico ed è suddiviso su due livelli:

- Controllori di basso livello

    1. Temperatura celle MT
        - variabile manipolata: Opening Degree (OD) valvola di espansione
    2. Temperatura celle LT
        - variabile manipolata: Opening Degree (OD) valvola di espansione
    3. Pressione di aspirazione MT
        - variabile manipolata: f_{comp} del compressore MT
    4. Pressione di aspirazione LT
        - variabile manipolata: f_{comp} del compressore LT

- Controllo di alto livello

    1. Set-point temperatura celle MT
    2. Set-point temperatura celle LT
    3. Set-point pressione di aspirazione MT
    4. Set-point pressione di aspirazione LT

I controllori locali hanno il compito di mantenere i set-point forniti dal controllo di alto livello che a sua volta decide i set-point al fine di soddisfare un obiettivo (e.g assicurare un assegnato riferimento di potenza elettrica consumata).

Per la progettazione del controllo supervisore viene inizialmente analizzato i comportamento del sistema con i set-point costanti:

1. Set-point temperatura celle MT = 2.5 °C

2. Set-point temperatura celle LT = -21 °C

3. Set-point pressione di aspirazione MT = 26×10^5 pascal

4. Set-point pressione di aspirazione LT = 12×10^5 pascal

Per le simulazioni del sistema è stato realizzato un modello in ambiente Matlab/Simulink (2.7).

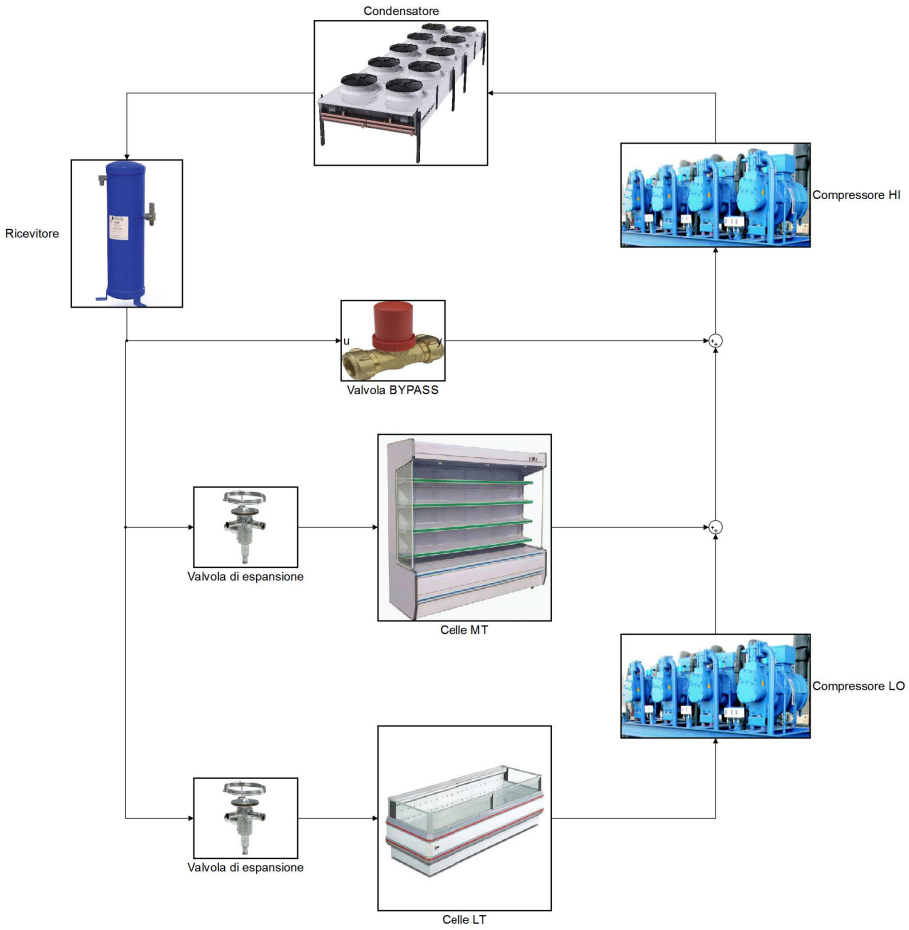

Figura 2.7 Sistema di refrigerazione di un supermercato: modello Simulink

In figura 2.8 è rappresentato l'andamento della potenza assorbita istantanea e la potenza assorbita media su 60 minuti dalla serie di compressori. Vale la pena notare che l'andamento nervoso della potenza istantanea assorbita dai compressori è dovuta alla tecnologia delle valvole di espansione impiegate che sono di tipo ON/OFF. L'utilizzo di valvole PWM oppure di valvole elettroniche assicurerebbero delle prestazioni migliori. La potenza media in 60 minuti, con i valore di set-point fissati ai valori indicati precedentemente, si aggira attorno ai 5 KW. Al variare dei set-point di pressione e temperatura anche la potenza assorbita cambia:

- aumentando i valori di set-point di pressione delle celle frigorifere e mantenendo costanti quelli delle temperature la potenza assorbita dai compressori diminuisce;
- aumentando i valori delle temperature delle celle frigorifere e mantenendo costanti quelli di pressione è necessario un consumo minore di potenza da parte dei compressori

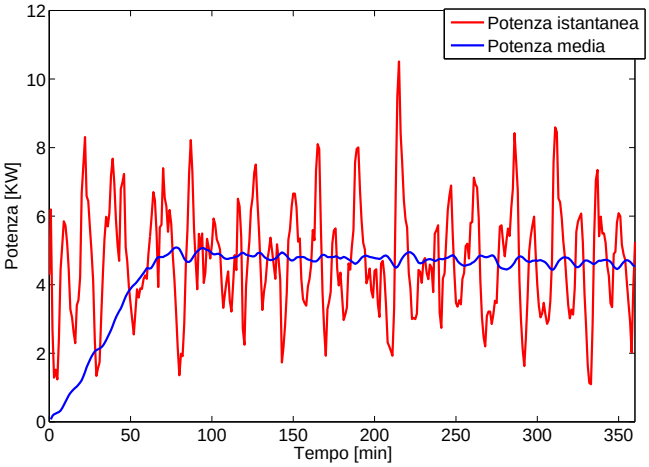

Figura 2.8 Potenza istantanea e potenza media assorbita dai compressori

Le temperature delle celle frigorifere di media e bassa temperatura, grazie all'azione dei controllori locali, soddisfano i vincoli imposti dal problema come si può vedere in figura 2.9. In figura 2.10 si può notare come le valvole di espansione delle celle frigo assumono solo due valori: un grado di apertura (OD) pari a 1 nel caso in cui siano completamente aperte, oppures pari a 0 se sono chiuse. Nelle figure 2.11 e 2.12 si può vedere l'andamento delle temperature e l'apertura delle valvole di espansione per le celle a media e bassa temperatura con i valori di set-point fissati precedentemente.

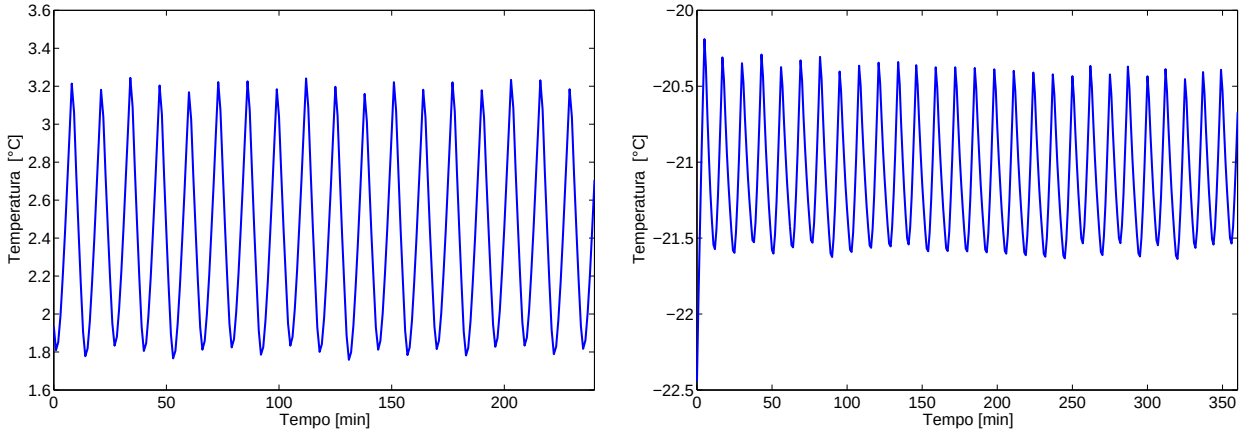

Figura 2.9 Temperature celle celle MT (sx) e LT (dx)

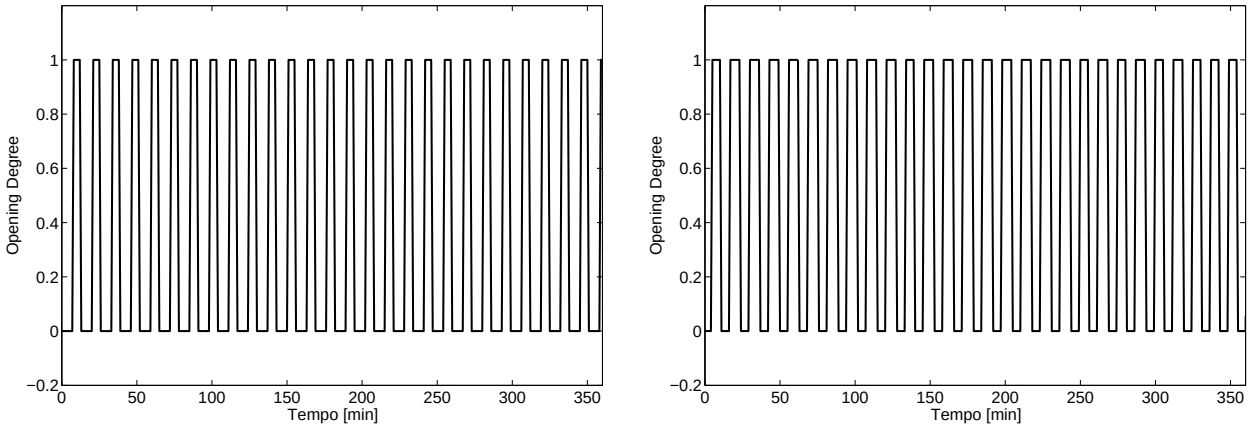

Figura 2.10 Opening degree delle celle MT (sx) e LT (dx)

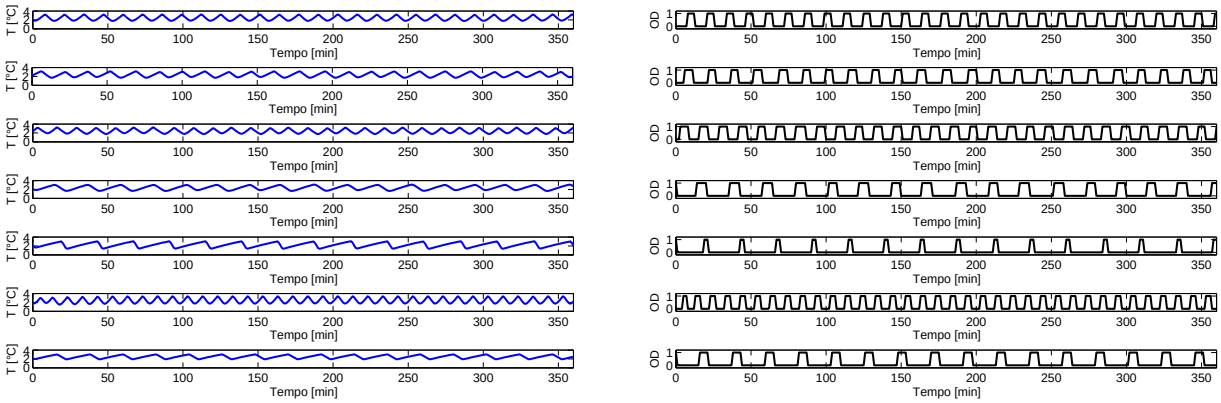

Figura 2.11 Temperature (sx) e Opening degree (dx) delle 7 celle MT

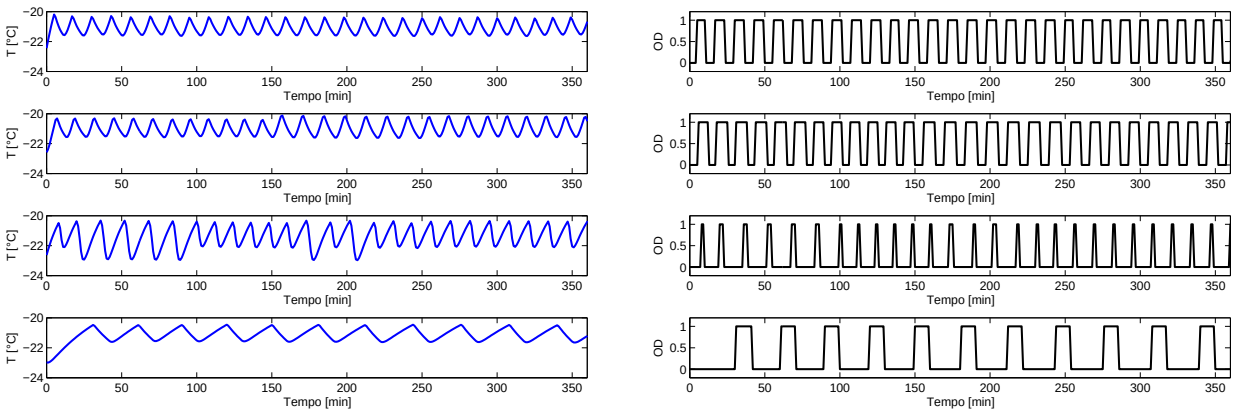

Figura 2.12 Temperature (sx) e Opening degree (dx) delle 4 celle LT

Capitolo 3

Apprendimento Mimetico

3.1 Introduzione

Le tecniche di apprendimento automatico si possono suddividere in tre grandi famiglie: apprendimento supervisionato, apprendimento non supervisionato e apprendimento mimetico. L'apprendimento supervisionato mira insegnare ad un sistema per permettergli di eseguire dei compiti in maniera autonoma sulla base di una serie di esempi. L'apprendimento non supervisionato è una tecnica di apprendimento automatico che sfrutta l'esperienza acquisita dall'interazione con l'ambiente per classificare ed organizzare sulla base di caratteristiche comuni e per cercare di effettuare ragionamenti e previsioni sui possibili risultati raggiungibili. Al contrario dell'apprendimento supervisionato, durante la fase di apprendimento non supervisionato non vengono forniti all'agente esempi ideali. L'apprendimento mimetico o Reinforcement Learning studia come diversi sistemi naturali o artificiali possono imparare ad ottimizzare il loro comportamento (massimizzando un premio o ricompensa) in un determinato ambiente attraverso azioni che li conducono in stati differenti. Tecniche di apprendimento mimetico trovano applicazioni in numerosi campi come l'economia, la psicologia e la teoria del controllo. L'apprendimento mimetico negli ultimi anni è diventato sempre più importante nel campo della teoria del controllo. L'apprendimento mimetico permette al sistema di controllo di comportarsi in maniere efficiente anche nel caso in cui ci siano delle variazioni delle condizioni al contorno dell'ambiente con il quale il controllore sta interagendo. Un aspetto fondamentale del controllore basato sull'apprendimento mimetico è la sua capacità di apprendere dalle esperienze passate aumentando l'autonomia del controllore.
Gli ambienti sono caratterizzati da alcune componenti fondamentali:

- uno spazio di stato
- un insieme di azioni

- un risultato/obiettivo finale da raggiungere

Le azioni causano uno spostamento da uno stato all'altro (ad esempio inducono una transizione di stato) e possono portare a determinati risultati. Tipicamente colui che prende le decisioni non possiede alcuna informazione a priori sull'ambiente con cui deve interagire e impara a conoscerlo in base all'esperienza acquisita dall'interazione con esso. I metodi basati sull'apprendimento mimetico si basano sull'interazione tra agente (controllore) e ambiente possono essere suddivisi in due grandi classi: i metodi model-based e i metodi model-free che portano ad ottenere risultati simili in maniera differente.

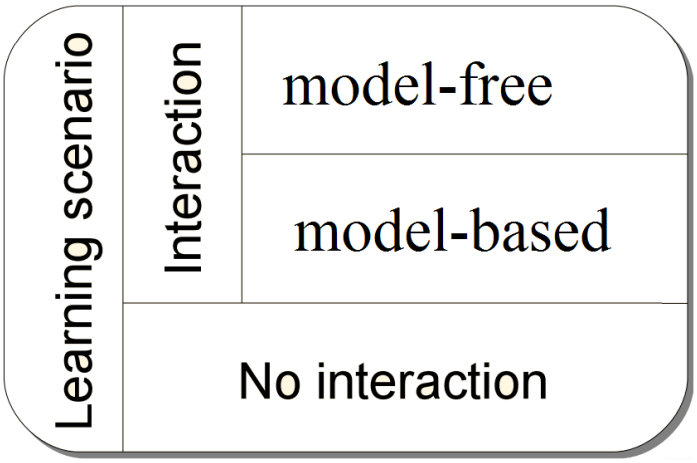

Figura 3.1 Classificazione tecniche di apprendimento mimetico

I metodi model-based utilizzano l'esperienza per costruire un modello delle transizioni tra gli stati e dei risultati ottenuti eseguendo determinate azioni. Le azioni che massimizzano la ricompensa vengono scelte attraverso lo sfruttamento del modello costruito. Questo è un modo statisticamente efficiente per usare l'esperienza accumulata in quanto ogni informazione trasmessa dall'ambiente può essere memorizzata nel modello. A patto che sia possibile una costante modifica e aggiornamento del modello questo permette di scegliere l'azione da compiere in maniera efficiente. I metodi model-free, al contrario, utilizzano l'esperienza per imparare direttamente la sequenza di ingressi (policy) che mi porta ad ottenere la soluzione ottimale senza modelli o stime. Data una policy, uno stato ha un valore, definito come la ricompensa attesa partendo dallo stato considerato. I metodi model-free sono più semplici da un punto di vista dell'implementazione rispetto a quelli model-based ma leggermente meno efficienti poichè l'informazione proveniente dall'interazione con l'ambiente può essere influenzata da stime precedenti errate o imprecise. Per comprendere meglio la differenza tra metodi model-based e model-free si consideri il seguente esempio: si vuole decidere la

strada da prendere per tornare a casa dal lavoro il venerdì sera rappresentato in figura 3.2

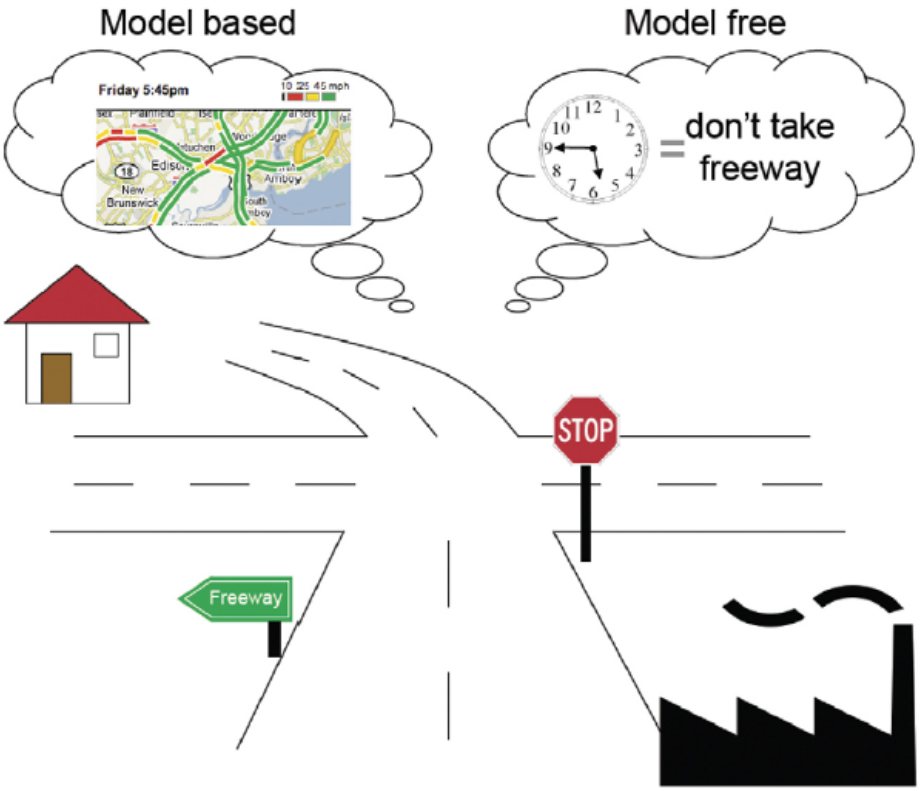

Figura 3.2 Problema della scelta della strada per tornare a casa

Basandosi sui concetti espressi precedentemente si può definire un insieme dei possibili stati (in questo caso luoghi,incroci o edifici particolari), azioni (e.g andare dritto oppure girare a sinistra o a destra ad un incrocio), probabilità di passare da uno stato all'altro quando viene eseguita una determinata azione e premi o ricompense (positive o negative) a seconda di ciò che succede lungo il percorso (per esempio semafori rossi, ingorghi, etc.). L'approccio model-based, come si può vedere in figura 3.2, è simile alla ricerca in una mappa mentale (modello costruito precedentemente) che è stato appreso dall'esperienza precedente. La selezione dell'azione migliore con l'approccio model-based avviene cercando nella mappa mentale (modello) l'azione che mi massimizza la ricompensa totale nel lungo periodo a partire dallo stato in cui mi trovo. Al contrario la selezione dell'azione che mi massimizza la ricompensa totale nel lungo periodo nell'approccio model-free avviene senza l'utilizzo di alcun modello. Per esempio, nel caso considerato, l'esperienza ha insegnato al guidatore che il venerdì sera la scelta migliore è quella di andare dritto all'incrocio e non prendere l'autostrada a causa del traffico. I metodi model-free sono notevolmente più semplici da un punto di vista di selezione dell'azione ottima ma, in alcuni casi, richiedono una fase di

apprendimento iniziale per ottenere una buona stima delle conseguenze che si ottengono eseguendo determinate azioni. Cambiando l'obiettivo finale (e.g l'indirizzo di casa) si può notare un ulteriore differenza tra i due approcci: mentre i metodi model-based si adattano al cambiamento in maniera piuttosto rapida, i metodi model-free richiedono una fase di adattamento leggermente maggiore in quanto non dispongono di alcun modello.

3.2 Concetti fondamentali

L'apprendimento mimetico (RL) è una tecnica di apprendimento automatico che permette ai sistemi di imparare ed adattarsi all'ambiente in cui si trovano al fine di raggiungere un determinato obiettivo. Colui che impara e prende le decisioni è detto agente mentre tutto ciò che è esterno all'agente e interagisce con esso viene interpretato come ambiente. Le due componenti interagiscono continuamente: l'agente seleziona l'azione da compiere e l'ambiente risponde all'azione presentando una nuova situazione all'agente. L'ambiente restituisce delle ricompense (reward), cioè dei valori numerici che l'agente deve massimizzare in un determinato intervallo di tempo. Ad ogni istante temporale la ricompensa è un semplice numero r_t e l'obiettivo dell'agente è quello di massimizzare le ricompense totali ricevute dall'ambiente. Questo significa che non bisogna massimizzare la ricompensa immediata ma la somma dei premi in un intervallo di tempo fissato.
Il principale aspetto dell'apprendimento mimetico è l'utilizzo di ricompense r_t per raggiungere l'obiettivo. Formalmente se la sequenza delle ricompense dopo l'istante t è:

$$r_{t+1} + r_{t+2} + r_{t+3} + r_{t+4} \ldots$$

allora si cerca di massimizzare la ricompensa totale attesa (Return), chiamata R_t, definita come:

$$R_t = r_{t+1} + r_{t+2} + r_{t+3} + r_{t+4} \ldots + r_N$$

dove N è l'istante finale. La notazione appena descritta viene usata nel caso di problemi ad orizzonte finito dove è presente un istante finale N. Ogni episodio termina in uno stato speciale detto stato terminale, ed è necessario distinguere l'insieme degli stati non terminali (S) dagli stati terminali (S^+). Nel caso di problemi ad orizzonte infinito viene introdotto il concetto di sconto e, scelta una azione da parte dell'agente, si vuole massimizzare la ricompensa totale:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} \ldots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

dove γ è un parametro $0 \leq \gamma \leq 1$ chiamato discount rate. Per $\gamma = 0$, l'agente cerca di massimizzare solo le ricompense immediate. Se γ è circa 1, i reward futuri vengono pesati maggiormente, e l'agente diventa più lungimirante. Non interessa sapere esattamente com'è fatto l'ambiente, interessa piuttosto fare delle ipotesi generali sulle proprietà che caratterizzano l'ambiente. Nell'apprendimento mimetico solitamente si assume che l'ambiente possa essere descritto da un Processo di Decisione Markoviano (Markov Decision Process o MDP). Un MDP è formalmente definito da:

- un insieme finito di stati S;
- un insieme finito di azioni A;
- una funzione di transizione T che assegna ad ogni coppia stato-azione una distribuzione di probabilità su S;
- una funzione di rinforzo (o reward) R che assegna un valore numerico ad ogni possibile transizione.

Considerando l'interazione agente ambiente abbiamo la seguente rappresentazione:

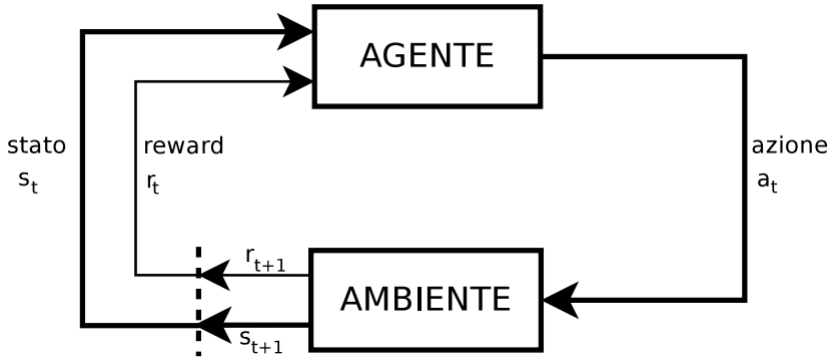

Figura 3.3 Interazione agente-ambiente

All'istante t l'agente percepisce l'ambiente come uno stato $s_t \in S$, e sulla base di s_t decide di agire con l'azione $a_t \in A$. L'ambiente, a seconda dello stato in cui si trova e dell'azione intrapresa, risponde inviando una ricompensa immediata $r_{t+1} = r(s_t, a_t)$ e producendo uno stato successivo s_{t+1}. In ogni istante l'agente implementa una mappa che va dagli stati ad ogni possibile azione, questa mappa viene chiamata policy (π_t), dove $\pi_t(s,a)$ è la probabilità che $a_t = a$ dato che $s_t = s$. I metodi basati sul reinforcement learning specificano come l'agente modifica la sua policy in base all'esperienza al fine di raggiungere una policy ottima (π^*) che massimizza la ricompensa totale attesa (R_t). Un problema soddisfa l'ipotesi Markoviana se

quello che succede all'istante $t+1$ dipende solo da ciò che è successo all'istante precedente e non dà tutto ciò che è successo precedentemente. Formalmente possiamo esprimere l'ipotesi Markoviana nel seguente modo:

$$P[s_{t+1} = s', r_{t+1} = r | s_t, a_t, s_{t-1}, a_{t-1}, ..., s_0, a_0] = P[s_{t+1} = s', r_{t+1} = r | s_t, a_t] \tag{3.1}$$

Se un problema soddisfa questa proprietà allora ha una dinamica che può essere espressa passo passo. Quasi tutti gli algoritmi di RL sono basati sulla stima delle value functions: funzioni di stati (o di coppie stato-azione) che stimano quanto utile è per l'agente essere in un determinato stato (o quanto utile è attuare una data azione in un dato stato). La nozione di "quanto utile" è qui definita in termini di ricompense future attese, più precisamente in termini di return attesi. Le value function sono definite rispetto ad una policy e dato uno stato s sotto la policy π, la value function si definisce come la ricompensa attesa (R_t) partendo da s e seguendo la policy π. Formalmente possiamo definire la value function $V^\pi(s)$:

$$V^\pi(s) = E_\pi\{R_t | s_t = s\} = E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\} \tag{3.2}$$

dove E_π è il valore atteso ottenuto nel caso in cui l'agente segua la policy π. Questa funzione prende il nome di state-value function per la policy π. Allo stesso modo si può definire la funzione stato-azione $Q(s,a)$ (action-state function) per una policy π:

$$Q^\pi(s,a) = E_\pi\{R_t | s_t = s, a_t = a\} = E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a\} \tag{3.3}$$

Le funzioni V^π e Q^π sono funzioni che vengono stimate in base all'esperienza. Una proprietà fondamentale delle value function utilizzata dai metodi del reinforcement learning è la ricorsività di V^π:

$$\begin{aligned}
V^\pi(s) &= E_\pi\{R_t | s_t = s\} && (3.4)\\
&= E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\} && (3.5)\\
&= \sum \pi(s,a) \sum_{s'} P^a_{ss'} \left[R^a_{ss'} + \gamma E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+2} | s_{t+1} = s'\} \right] && (3.6)\\
&= \sum \pi(s,a) \sum_{s'} P^a_{ss'} \left[R^a_{ss'} + \gamma V^\pi(s') \right] && (3.7)
\end{aligned}$$

Dove s è lo stato attuale, s' lo stato successivo, a l'azione scelta dall'agente, $P^a_{ss'}$ è la probabilità di transizione tra s e s' compiendo l'azione a e $R^a_{ss'}$ è la ricompensa attesa dalla transizione. L'ultima espressione trovata (3.7) è l'equazione di Bellman per V^π ed esprime la relazione tra la value function di uno stato e del suo successore. Risolvere un problema di RL significa trovare una policy che permetta di accumulare la maggior ricompensa possibile. Le value function definiscono un ordinamento parziale sulle policiy. Una policy π viene considerata migliore o uguale a una policy π' se il suo return atteso è maggiore o uguale a quello di π' per tutti gli stati:

$$\pi \geq \pi' \textit{ se e solo se } V^\pi(s) \geq V^{\pi'}(s) \tag{3.8}$$

La policy che risulta essere migliore o uguale a tutte le altre viene definita policy ottima π^*. Tutte le policies ottime condividono la stessa value function detta optimal state-value function:

$$V^*(s) = \max_\pi V^\pi(s) \quad \forall s \in S \tag{3.9}$$

Le policy ottime condividono anche la stessa funzione stato-azione ottima chiamata $Q^*(s,a)$:

$$Q^*(s,a) = \max_\pi Q^\pi(s,a) \quad \forall s \in S \tag{3.10}$$

$Q(s,a)$ restituisce il return atteso per l'azione a scelta nello stato s seguendo una politica ottima. Pertanto si può scrivere Q^* in termini di V^* nel seguente modo:

$$Q^*(s,a) = E\left\{r_{t+1} + \gamma V^*(s_{t+1}|s_t = s, a_t = a)\right\} \tag{3.11}$$

L'equazione di Bellman scritta per V^* (3.7) diventa la Bellman optimality equation (B.o.e). Intuitivamente essa esprime il fatto che il valore di uno stato sotto una policy ottima deve eguagliare la ricompensa totale attesa per la migliore azione in quello stato:

$$V^*(s) = \max_{a \in A(s)} Q^{\pi^*}(s,a) \tag{3.12}$$

Questo si può dimostrare attraverso i seguenti passaggi:

$$V^*(s) = \max_{a \in A(s)} Q^{\pi^*}(s,a) \tag{3.13}$$

$$= \max_a \; E_{\pi^*}\{R_t | s_t = s, a_t = a\} \tag{3.14}$$

$$= \max_a \; E_{\pi^*}\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a\} \tag{3.15}$$

$$= \max_a \; E_{\pi^*}\{r_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} | s_t = s, a_t = a\} \tag{3.16}$$

$$= \max_a \; E\{r_{t+1} + \gamma V^*(s_{t+1}) | s_t = s, a_t = a\} \tag{3.17}$$

$$= \max_a \; \sum_{s'} P^a_{ss'} \left[R^a_{ss'} + \gamma V^{\pi}(s')\right] \tag{3.18}$$

Allo stesso modo la Bellman optimality equation per Q^* è la seguente

$$Q^*(s,a) = E\left\{r_{t+1} + \gamma \max_{a'} Q^*(s_{t+1}, a') | s_t = s, a_t = a)\right\} \tag{3.19}$$

$$= \sum_{s'} P^a_{ss'} \left[R^a_{ss'} + \gamma \max_{a'} Q^*(s', a')\right] \tag{3.20}$$

Per MDP finiti, l'equazione di ottimalità di Bellman ha un'unica soluzione indipendente dalla politica. Infatti la B.o.e. conduce ad un sistema di equazioni, una per ciascun stato, così che se ci sono k stati, vi sono k equazioni in k incognite. Se le dinamiche del sistema sono note, in linea di massima si può risolvere questo sistema di equazioni per $V^*(s)$ usando un qualsiasi metodo per la risoluzione di sistemi di equazioni non lineari. Analoghe considerazioni possono essere fatte per $Q^*(s)$. Una volta che si dispone di $V^*(s)$, è facile determinare la politica ottima: per ogni stato s, esiste una azione a^* che fornisce il massimo nell'equazione di ottimalità di Bellmann. Si tiene a^* e si pongono a zero la probabilità associata alle altre azioni. Una policy che assegna probabilità non nulla solo a quelle azioni è una policy ottima. Una policy è greedy se sceglie le azioni che massimizzano la ricompensa totale sfruttando la conoscenza accumulata. Un altro modo per esprimere lo stesso concetto è che "ogni policy che è *greedy* rispetto alla funzione valore ottima V^* è una policy ottima". Vale la pena osservare che in 3.18 si esegue una ricerca locale ma allo stesso tempo si tiene conto dei risultati a lungo termine. Nota Q^* la ricerca dell'azione ottima è immediata: per ogni stato s si è già memorizzata l'azione che massimizza $Q^*(s,a)$.
L'utilizzo diretto delle equazioni di Bellman è poco pratico poichè richiede:

- una conoscenza accurata della dinamica dell'ambiente

- una buona risorsa di calcolo

- il rispetto della proprietà di Markov

In [11] vengono presentate tre classi di tecniche risolutive per i problemi di Reinforcement Learning:

- Programmazione dinamica (DP)
  La programmazione dinamica si basa sull'equazione di Bellmann e divide il problema principale in piccoli sottoproblemi. Suddividendo i problemi in sottoproblemi di complessità minore è necessaria una conoscenza accurata del modello dell'ambiente e i tempi di esecuzione sono molto elevati; per questi motivi non vengono utilizzati per risolvere problemi di RL. L'idea principale della programmazione dinamica, e del reinforcement learning in generale, è di usare la value function per trovare iterativamente la policy ottima π^*. Gli algoritmi di programmazione dinamica sono ottenuti modificando le equazioni di Bellman in regole di aggiornamento con lo scopo di migliorare l'approssimazione delle value function desiderate.

- Metodi Montecarlo
  I metodi Montecarlo, a differenza della programmazione dinamica, non hanno bisogno di un modello dell'ambiente e utilizzano le informazioni trasmesse dall'ambiente per ottenere una stima delle ricompense future. La value function viene aggiornata esclusivamente quando viene raggiunto lo stato finale quindi al termine dell'episodio:

$$V(s_t) \leftarrow V(s_t) + \alpha[R_t - V(s_t)] \tag{3.21}$$

- Temporal Difference Learning (TD-Learning)
  Temporal-Difference learning è una delle idee principali dell'apprendimento mimetico ed è una combinazione tra i metodi Montecarlo e la programmazione dinamica. Dai metodi Montecarlo mantiene la capacità di imparare direttamente dall'esperienza in un sistema dove la dinamica non è completamente nota. Inoltre il TD-Learning eredita dalla programmazione dinamica eredita l'abilità di aggiornare le stime delle funzioni utili alla risoluzione del problema senza attendere il raggiungimento di un stato finale.

$$V(s_t) \leftarrow V(s_t) + \alpha[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)] \tag{3.22}$$

3.2.1 Algoritmi Apprendimento Mimetico

Nel lavoro svolto la progettazione del supervisore avviene attraverso apprendimento mimetico. Si è deciso di utilizzare algoritmi di risoluzione appartenenti alla classe TD-Lerning poichè:

- non necessitano di un modello dettagliato dell'ambiente o sistema
- apprendono direttamente dall'esperienza
- l'aggiornamento della $Q(s,a)$ avviene anche se non si raggiunge uno stato finale

Gli algoritmi impiegati si suddividono in due classi distinte:

1. metodi on-line (e.g. SARSA)
2. metodi off-line (e.g. Q-Learning)

Gli algoritmi che appartengono a queste classi, come il SARSA e il Q-Learning, sono caratterizzati dallo stesso procedimento risolutivo anche se si differiscono nell'aggiornamento della funzione $Q(s,a)$.

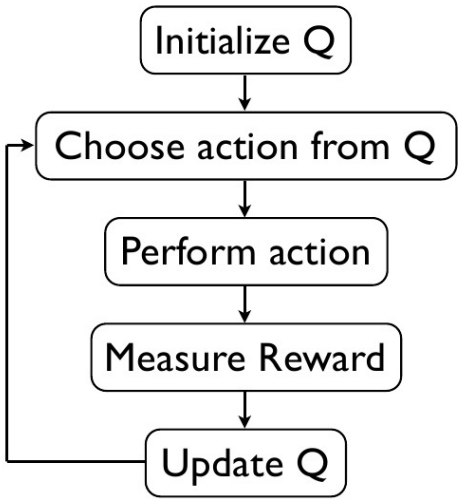

Figura 3.4 Procedimento algoritmi TD-Learning

Per i metodi on-line policy la funzione $Q(s,a)$ è aggiornata sulla base dei risultati di azioni determinate dalla policy selezionata. Per i metodi off-line policy $Q(s,a)$ è aggiornata sulla basa di azioni ipotetiche, non effettivamente intraprese.

Algoritmo SARSA

Una caratteristica fondamentale dell'algoritmo è l'utilizzo della funzione stato-azione $Q(s,a)$ al posto della value function $V(s)$. In particolare SARSA deve stimare $Q^{\pi}(s,a)$ per una determinata policy π e per ogni stato s e azione a . Questo avviene utilizzando la formula descritta in 3.22 solo che al posto di $V(s)$ si utilizza $Q(s,a)$:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)] \tag{3.23}$$

Nell'algoritmo vengono considerate transizioni da una coppia stato-azione ad un altra coppia stato-azione e l'aggiornamento avviene dopo ogni transizione indipendentemente se lo stato s_t è terminale o no. Il nome SARSA deriva dal fatto che in realtà gli aggiornamenti si effettuano utilizzando una quintupla $Q(s,a,r,s',a')$. Dove s e a rappresentano lo stato attuale e l'azione scelta nel passo corrente, invece s' e a' sono la nuova coppia stato-azione. Il primo passo consiste nell'apprendere il valore di Q^{π} per una determinata policy e successivamente, interagendo con l'ambiente, si può modificare e migliorare la policy. L'algoritmo viene riassunto nel seguente pseudocodice:

Initialize $Q(s,a)$ *arbitrarily*
Repeat (for each episode)
Initialize s
Choose a from s using policy derived from Q (e.g ε *-greedy)*
Repeat (for each step of the episode)
Take action a' from s' *using policy derived from Q (e.g* ε *-greedy)*
$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)]$
$s \leftarrow s'$
$a \leftarrow a'$
unitil s is terminal

I parametri γ e α rappresentano rispettivamente il discount factor e il learning rate. L'algoritmo SARSA converge con probabilità 1 ad una policy ottima e quindi ad un ottima funzione stato-azione $Q(s,a)$ se le coppie stato-azione vengono visitate un numero infinito di volte.

Algoritmo Q-learning

Uno degli algoritmi più importanti relativi alla risoluzione di problemi di apprendimento mimetico è il Q-Learning e si basa sull'utilizzo della funzione $Q(s,a)$. Apprendere la

funzione Q equivale ad apprendere la politica ottima, infatti la politica ottima è definita come:

$$\pi^* = \underset{a \in A}{\operatorname{argmax}} Q(s,a)$$

Per apprendere Q, occorre un modo per stimare i valori di addestramento per Q a partire solo dalla sequenza di ricompense immediate r in un lasso di tempo. La funzione Q viene aggiornata utilizzando la seguente formula:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)] \tag{3.24}$$

Lo pseudocodice relativo al Q-Learning è il seguente:

Initialize $Q(s,a)$ *arbitrarily*
Repeat (for each episode)
Initialize s
Repeat (for each step of the episode)
Choose a from s using policy derived from Q (e.g ε *-greedy)*
Take action a' and observe r e s'
$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)]$
$s \leftarrow s'$
unitil s is terminal

Si osservi che la regola di apprendimento 3.24 usa il valore corrente di Q dell'agente per il nuovo stato s_{t+1} per raffinare la sua stima di $Q(s,a)$ per il precedente stato s. L'agente deve solo eseguire l'azione nel suo ambiente e osservare il risultante nuovo stato s_{t+1} e la ricompensa r_{t+1}. Per quanto riguarda la convergenza, come riportato in [11], se le azioni sono scelte in modo che ogni coppia stato-azione sia visitata infinite volte allora è garantita la convergenza alla action-state function ottima Q^*.

3.2.2 Osservazioni aggiuntive

- L'implementazione degli algoritmi avviene attraverso l'utilizzo di una matrice avente un numero di righe pari al numero degli stati e un numero di colonne pari al numero delle possibili azioni. In ogni cella viene memorizzata una stima della funzione stato-azione $Q(s,a)$. Dato uno stato s per scegliere l'azione che massimizza la ricompensa totale deve scorrere la riga relativa allo stato s e scegliere l'azione avente $Q(s,a)$ maggiore.

- Durante la risoluzione di un problema di Reinforcement Learning bisogna essere in grado di bilanciare la fase di esplorazione e di sfruttamento. L'esplorazione (exploration) dello spazio delle azioni permette di scoprire azioni che portano a ricompense migliori. Un agente che esplora solamente difficilmente riuscirà a convergere ad una policy ottima. Le azioni migliori vengono scelte ripetutamente (sfruttamento) perché garantiscono una ricompensa massima (reward). Il bilanciamento può essere ottenuto attraverso la tecnica ε-greedy, come già riportato nei pseudocodici precedenti nella sezione relativa all'algoritmo SARSA (sezione 3.2.1) e Q-Learning (sezione 3.2.1). Questo metodo permette di scegliere l'opzione *greedy* per la maggior parte delle volte al fine di sfruttare l'informazione immagazzinata fino a quel momento e massimizzare la ricompensa totale e con probabilità ε di scegliere una azione in maniera casuale così da poter esplorare eventuali policy maggiormente produttive. Valori di ε troppo grandi portano a tempi di convergenza elevati mentre valori troppo piccoli non permettono di trovare la policy ottima.

- Il parametro $\alpha \in [0,1]$ influenza notevolmente le prestazioni degli algoritmi infatti per valori di α molto piccoli l'apprendimento è rallentato, mentre per valori di alpha troppo elevati l'algoritmo rischia di non convergere. Nell'implementazione degli algoritmi è consigliabile iniziare con un valore molto grande per poi farlo decrescere all'aumentare delle esplorazioni.

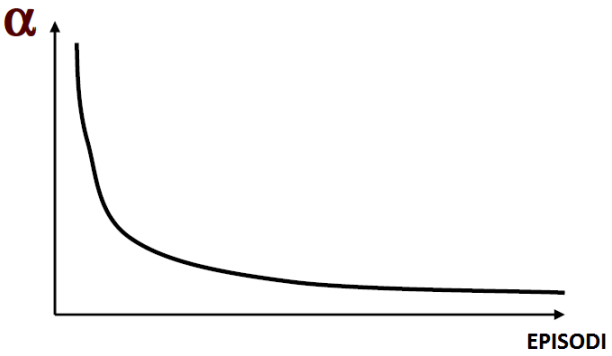

Figura 3.5 Andamento decrescente di α

- Le prestazioni degli algoritmi presentati sono influenzati anche dalla inizializzazione della matrice $Q(s,a)$ e dalla scelta delle ricompense nel caso di raggiungimento o non raggiungimento dell'obiettivo. Si è dimostrato come inizializzare una $Q(s,a)$ con valori tutti diversi da zero porti ad avere una maggiore velocità di convergenza della soluzione rispetto a $Q(s,a)$ con tutti valori nulli([12]).

- Gli algoritmi presentati sono robusti perché sopportano eventuali disturbi tempo varianti all'interno dell'episodio. Il tempo di convergenza per trovare la policy ottima in alcuni casi è abbastanza elevato. Per risolvere tale problema si può aggiungere un informazione a priori al controllore così da migliorare il tempo che l'agente (i.e. il supervisore nel caso in esame) impiega a trovare la policy ottima.

Capitolo 4

Progettazione del Supervisore

La gestione di sistemi di grandi dimensioni, costituiti da numerose componenti distinte, è spesso affidata ad un sistema di supervisione centralizzato al quale è richiesto un comportamento efficiente e resistente ai disturbi. L'architettura di controllo negli anni, con l'obiettivo di gestire al meglio i consumi di potenza (demand-side management), ha subito una cambiamento passando da un'architettura centralizzata ad una struttura gerarchica (supervisore e controllori locali). La formalizzazione di un supervisore che gestisca un sistema di refrigerazione è la difficile sfida intrapresa in questa parte del lavoro. In questo capitolo si affronta la progettazione di un controllore di alto livello utilizzando inizialmente un approccio model-based basato su un controllore PI (Proporzionale Integrale) e successivamente una tecnica model-free basata sull'apprendimento mimetico. Il controllo di alto livello ha il compito di impostare i valori dei set-point di pressione e di temperatura delle singole celle del sistema di refrigerazione al fine di assicurare un certo obiettivo. Nel caso specifico l'obiettivo è quello di garantire il rispetto di un assegnato riferimento di potenza media assorbita (in un dato intervallo di tempo) da parte dei compressori. Il controllo di basso livello mantiene i valori dei set-point inviati dal supervisore. L'architettura di controllo gerarchico può essere visualizzata in figura 4.1.
Nell'esempio affrontato si assume che il riferimento di potenza elettrica assorbita sia assegnato e pari a 6 KW e l'intervallo di tempo è pari a 4 ore. I set-point decisi dal controllo di alto livello sono:

- Temperatura delle celle MT
- Temperatura delle celle LT
- Pressione di aspirazione delle celle MT
- Pressione di aspirazione delle celle LT

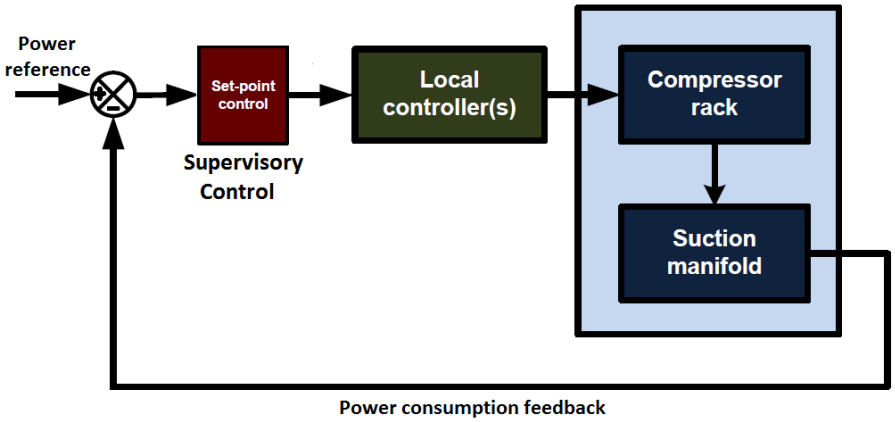

Figura 4.1 Architettura di controllo

Oltre a consumare un quantitativo di potenza fissata il controllo deve rispettare i vincoli di temperatura per ogni cella i-esima:

- $T_{i,MT} - 1°C \leq T_{i,MT} \leq T_{i,MT} + 1°C \qquad i = 1,...,7$
- $T_{i,LT} - 1°C \leq T_{i,LT} \leq T_{i,LT} + 1°C \qquad i = 1,...,4$
- $T_{food,MT} - 1°C \leq T_{food,MT} \leq T_{food,MT} + 1°C$
- $T_{food,LT} - 1°C \leq T_{food,LT} \leq T_{food,LT} + 1°C$

I primi due vincoli relativi alla temperatura interna alle celle frigorifere ($T_{i,MT}$ e $T_{i,LT}$) sono rispettati grazie alla presenza dei controllori locali mentre i veri vincoli del problemi di controllo sono quelli relativi alla temperatura delle derrate alimentari (T_{food}). La temperatura degli alimenti è fissata a $2°C$ per le celle MT ($T_{food,MT}$) e $-20°C$ per le celle LT ($T_{food,LT}$). Il controllo supervisore effettua la sua azione regolatrice con passo costante (tempo di supervisione) e può decidere se cambiare tutti i valori dei set-point oppure mantenerne alcuni costanti. Per la progettazione del controllore di alto livello si è deciso di considerare due diversi scenari:

- Determinazione dei set-point di pressione impostando ad un valore costante i set-point di temperatura. Le temperature delle celle MT e LT vengono fissati rispettivamente a 2.5 °C e -21 °C. Per quanto riguarda i valori dei set-point di pressione si assume costante la pressione di aspirazione delle celle LT (12×10^5 pascal) mentre la pressione di aspirazione delle celle MT viene modificata ogni 15 minuti.

- Determinazione dei set-point di temperatura impostando ad un valore costante i set-point di pressione. Le pressioni di aspirazione vengono fissate a 26×10^5 pascal per le celle MT e 12×10^5 pascal per le celle LT mentre i set-point delle temperature vengono regolati ogni 15 minuti.

Set-point	Valore
Temperatura MT	2.5 °C
Temperatura LT	-21 °C
Pressione MT	variabile decisionale
Pressione LT	12×10^5 pascal

Tabella 4.1 Scenario 1: Set-point controllo supervisore

Set-point	Valore
Temperatura MT	variabile decisionale
Temperatura LT	variabile decisionale
Pressione MT	26×10^5 pascal
Pressione LT	12×10^5 pascal

Tabella 4.2 Scenario 2: Set-point controllo supervisore

4.1 Progettazione controllo PI: Set-point pressione

Il controllo supervisore contiene un regolatore PI per decidere il valore del set-point della pressione di aspirazione delle celle MT al fine di avere un consumo di potenza media in un ora che segue il riferimento di 6 KW rispettando i vincoli imposti dal problema. Il controllore PI genera un ingresso $u(t)$ che dipende dall'errore dinamico del sistema e i gradi di libertà sono rappresentati dai guadagni del regolatore (K_P e K_I).
Il PI regola l'uscita in base a:

- l'errore dinamico all'istante t ($e(t)$) con l'azione proporzionale caratterizzata dal guadagno K_P
- l'errore dinamico negli istanti precedenti a t con l'azione integrale caratterizzata dal guadagno K_I

La componente proporzionale ha l'effetto di diminuire il tempo di salita[1] (parametro che caratterizza la prontezza del sistema), incrementare le sovraelongazioni e ridurre, ma non eliminare, l'errore a regime permanente. La componente integrale permette di eliminare l'errore a regime permanente ma causa una diminuzione della stabilità del sistema. L'uscita del controllore è il valore del set-point della pressione di aspirazione delle celle MT. L'ingresso di controllo viene cambiato ogni 15 minuti (tempo di supervisione) in modo tale da permettere anche alle grandezze più lente, come le temperature, di raggiungere un valore stazionario (o quasi-stazionario). Se il tempo di supervisione è pari a 15 minuti e l'intervallo temporale è di 4 ore allora il controllo supervisore assegna il valore del set-point della pressione per 17 volte. Aumentando il tempo di supervisione, per esempio a 20 minuti, le prestazioni peggiorano perchè a parità di intervallo temporale (240 minuti) il set-point verrà cambiato un numero inferiore di volte (13). Si noti che un tempo di supervisione troppo piccolo non permette alle pressioni e alle temperature di raggiungere un valore stazionario. Per determinare i guadagni del controllore PI è necessaria una fase di taratura che può essere onerosa in funzione delle prestazioni desiderate. Nel caso in esame la fase di tuning dei parametri del PI è stata effettuata giovandosi delle informazioni fornite da un modello approssimato del processo (Capitolo 2) (risposte indiciali, costanti di tempo dominanti, tempi di salita, tempi di assestamento,guadagni statici, etc.). Sulla base delle indicazioni fornite dal modello e giovandosi delle tecniche standard di taratura [17] i guadagni del controllore vengono impostati a:

$$K_P = 0.4 \qquad K_I = 0.2$$

[1] Si intende il tempo necessario al sistema per variare dal 10% al 90% del valore di regime dello stesso

Le simulazioni vengono effettuate utilizzando il modello Matlab/Simulink (2.7) per un tempo totale di 4 ore con un tempo di supervisione pari a 15 minuti.
I risultati ottenuti con il controllo PI sono i seguenti:

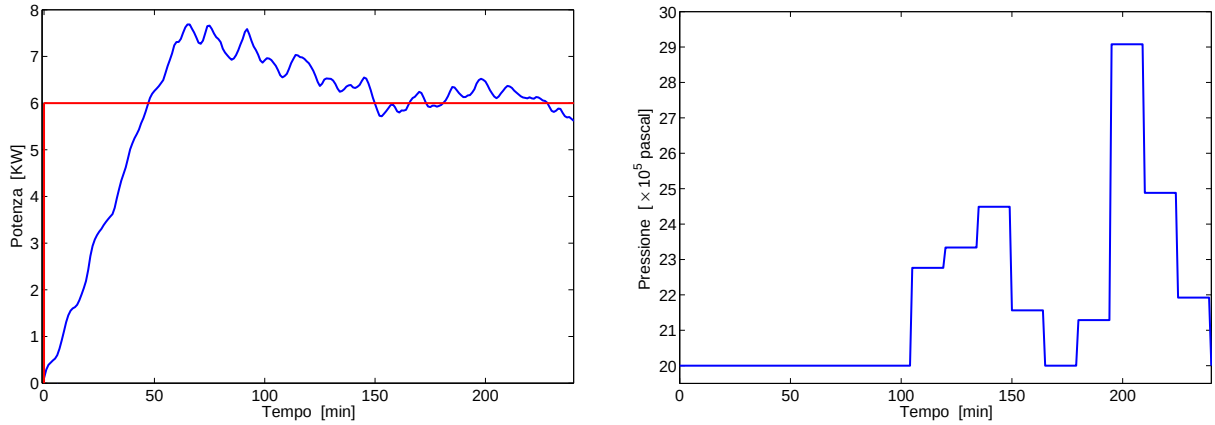

Figura 4.2 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e ingresso di controllo (dx) per il controllore PI

In figura 4.2 si nota che l'andamento della potenza media su 60 minuti è abbastanza regolare e riesce ad inseguire il riferimento nonostante i vincoli tecnologici imposti dalle valvole di espansione (di tipo ON/OFF). Come si può vedere in figura 4.3 e figura 4.4 la temperatura all'interno delle celle frigo e la temperatura del carico interno restano dentro limiti fissati inizialmente. Il vantaggio nell'utilizzo del controllo PI è la semplicità d'implementazione ma non è sempre robusto (ad esempio in caso di disturbi tempo varianti). Un'improvvisa variazione della temperatura interna al supermercato o di un'altra grandezza la prestazioni del controllore PI causa un peggioramento delle prestazioni del sistema ed è necessaria una nuova taratura dei parametri K_P e K_I.

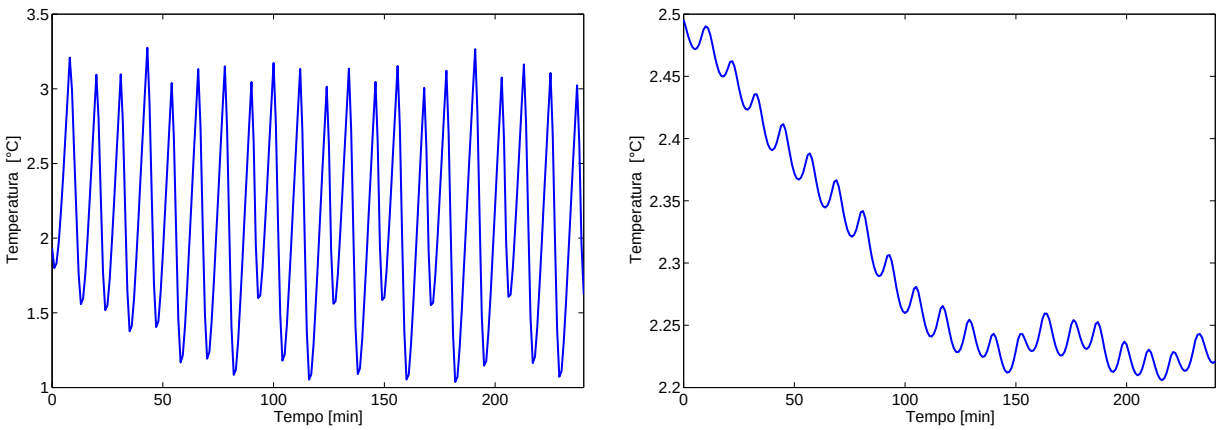

Figura 4.3 Temperatura interna (sx) e del carico (dx) in una delle celle MT con controllo PI

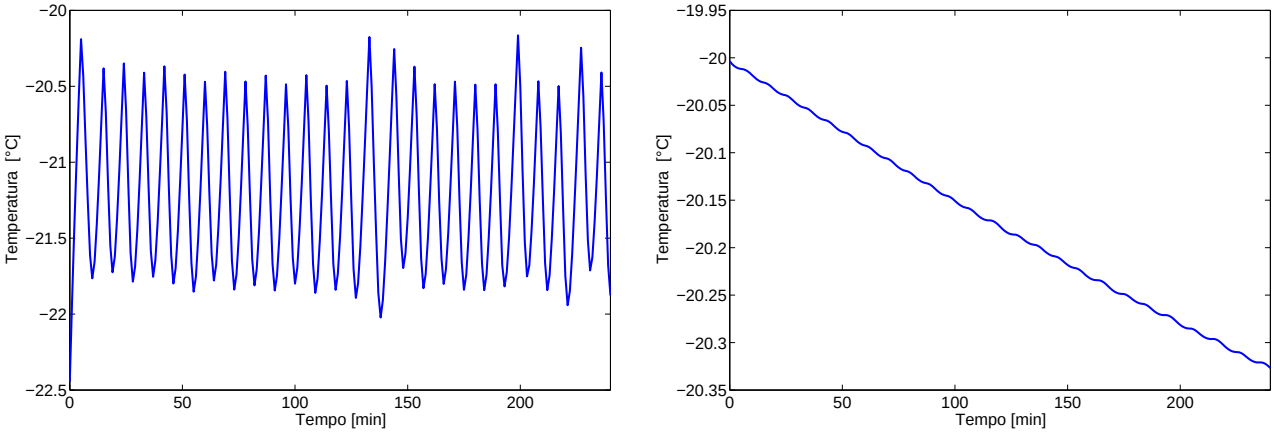

Figura 4.4 Temperatura interna (sx) e del carico (dx) in una delle celle LT con controllo PI

4.2 Progettazione controllo RL: Set-point pressione

L'obiettivo del controllo di supervisione resta quello di fornire i valori dei set-point delle pressioni e delle temperature delle celle LT e MT al fine di mantenere un consumo di potenza media in un ora pari a 6 KW. Per la progettazione del controllore con apprendimento mimetico è necessario considerare il sistema come un processo di decisione Markoviano.
Si definiscono:

- L'insieme delle possibili azioni, cioè l'insieme dei possibili valori di pressione delle celle MT, che il controllore può compiere (A_t):

$$A_t = \{20,\ 20.25,\ 20.5,\ 20.75,\ ....\ 33.25,\ 33.5,\ 33.75,\ 34\} \quad [\times 10^5 pascal]$$

- L'insieme degli stati del sistema (S_t):

$$S_t = \{Potenza\ assorbita\}$$

- Una funzione ricompensa che restituisce per un determinato stato la ricompensa r_t:

$$e_t = errore\ a\ regime$$

$$r_t = \begin{cases} 100 - 2 \cdot |e_t|, & \text{se } |e_t| \leq 0.5 \ [KW] \\ -1000 - 10 \cdot |e_t|, & \text{se } |e_t| > 0.5 \ [KW] \end{cases}$$

La scelta degli insiemi A_t e S_t è determinante per il buon funzionamento del controllore. Il range dei possibili valori di pressione è determinato dal modello utilizzato mentre il passo di discretizzazione è posto pari a 0.25×10^5 pascal. Un passo di discretizzazione troppo elevato determina soluzioni molto irregolari che non permettono alla potenza assorbita di restare all'interno dell'intervallo fissato. Al contrario un passo di discretizzazione troppo piccolo aumenta notevolmente il tempo di convergenza a π^*. Il passo di discretizzazione scelto per A_t è un buon compromesso tra precisione della soluzione e velocità di convergenza alla policy ottima. Anche all'aumentare della cardinalità di S_t aumenta il tempo di convergenza mentre al diminuire della cardinalità peggiora la precisione della soluzione trovata. La funzione ricompensa è una funzione che ad uno stato associa un valore numerico. Se la potenza è all'interno dell'intervallo assegnato (± 0.5 [KW]) allora il sistema premia il controllore con una ricompensa positiva mentre se è al di fuori dell'intervallo il sistema invia un premio negativo.
Il premio è proporzionale all'errore e_t:

- nel caso in cui $|e_t|$ è inferiore a 0.5 [KW] allora la ricompensa è tanto maggiore quanto più mi avvicino al riferimento di 6 KW
- nel caso in cui $|e_t|$ è superiore a 0.5 [KW] la ricompensa è tanto minore quanto più mi allontano dal riferimento

Definendo la funzione ricompensa in tale maniera il controllore cercherà di mantenere la potenza media su 60 minuti all'interno dell'intervallo fissato massimizzando la ricompensa totale. Fissata una finestra temporale di 4 ore (un episodio) e si fissano i parametri necessari per applicare gli algoritmi SARSA e Q-Learning (sezione 3.2.1). Il controllore agisce sul sistema ogni 15 minuti (tempo di supervisione). Partendo da uno stato s_t il controllo di alto livello sceglie un azione a_t nell'insieme A_t. L'azione viene passata al sistema che valuta lo stato s_{t+1} a cui si porta e infine invia al controllore la ricompensa r_{t+1}. Successivamente, in base alle informazioni ricevute, il controllore aggiorna la funzione stato-azione $Q(s,a)$ a seconda dell'algoritmo utilizzato.

4.2.1 Risultati con gli algoritmi SARSA e Q-Learning

Per eseguire le operazioni di controllo attraverso gli algoritmi di apprendimento mimetico è necessario impostare i parametri in maniera adeguata. L'inizializzazione di questi parametri è fondamentale per ottenere dei buoni tempi di convergenza per la policy ottima π^*. Il tasso di apprendimento (α) è inizialmente impostato ad un valore piuttosto elevato (0.7) poichè il controllore nella fase iniziale deve apprendere più informazioni possibili dal sistema e decresce ad ogni episodio fino a raggiungere un minimo di 0.125. Il discount factor è impostato ad un valore costante pari a 1 così da rendere il controllore il più lungimirante possibile. Anche la probabilità di esplorazione (ε) viene inizialmente impostata a 0.2 e decresce proporzionalmente al numero degli episodi. La probabilità di esplorazione dipende dal numero di azioni di controllo eseguite dal regolatore. Valori di ε troppo elevati causano un significativo rallentamento degli algoritmi ma permettono un'esplorazione più accurata e una probabilità maggiore di ottenere sicuramente la policy ottima. Probabilità di esplorazione troppo basse portano a policy sub-ottime. La funzione stato-azione viene inizializzata a valori nulli e nel codice Matlab è implementata come una matrice con m righe e n colonne dove m è la cardinalità dell'insieme S_t mentre n è la cardinalità di A_t. Ad ogni iterazione in posizione (i, j) viene memorizzato il valore della funzione stato-azione relativa alla coppia (s_i, a_j). I risultati ottenuti applicando l'algoritmo SARSA sono riportati nelle figure successive.

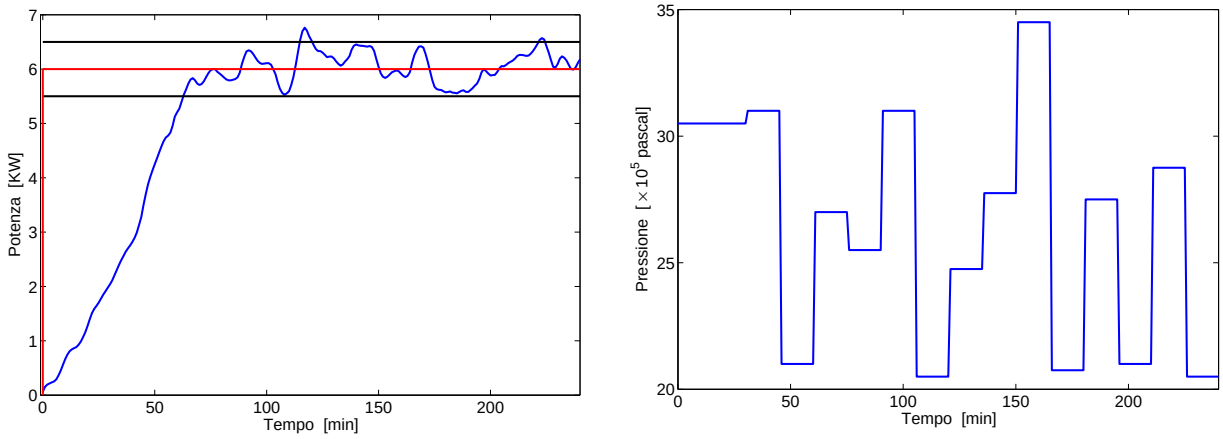

Figura 4.5 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e Set-point della pressione MT (dx)(SARSA)

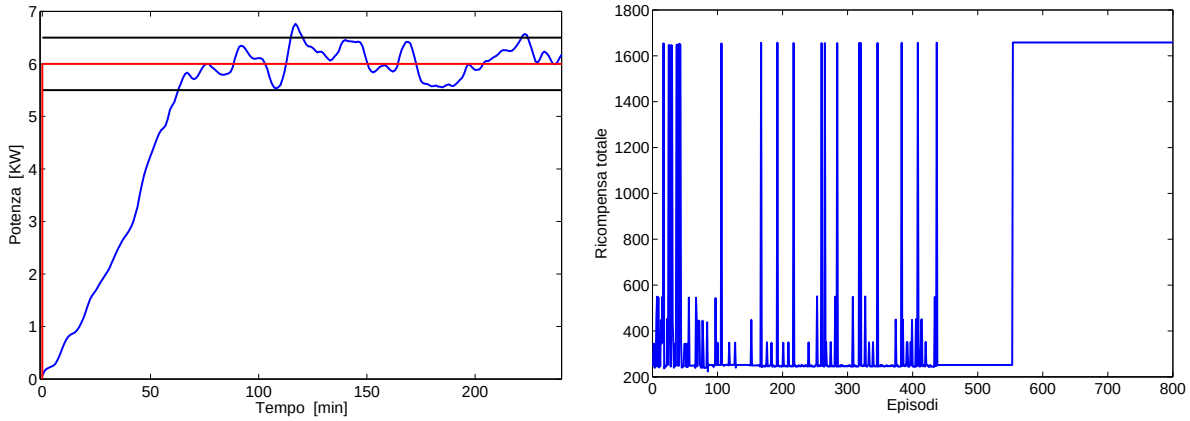

Figura 4.6 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e Ricompensa totale (dx)(SARSA)

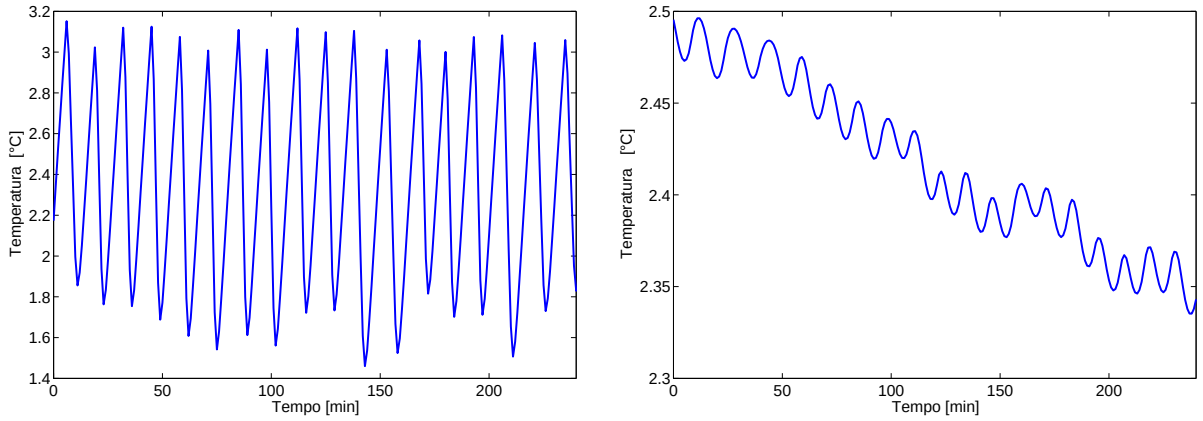

Figura 4.7 Temperatura interna (sx) e del carico (dx) in una delle celle MT con controllore RL (SARSA)

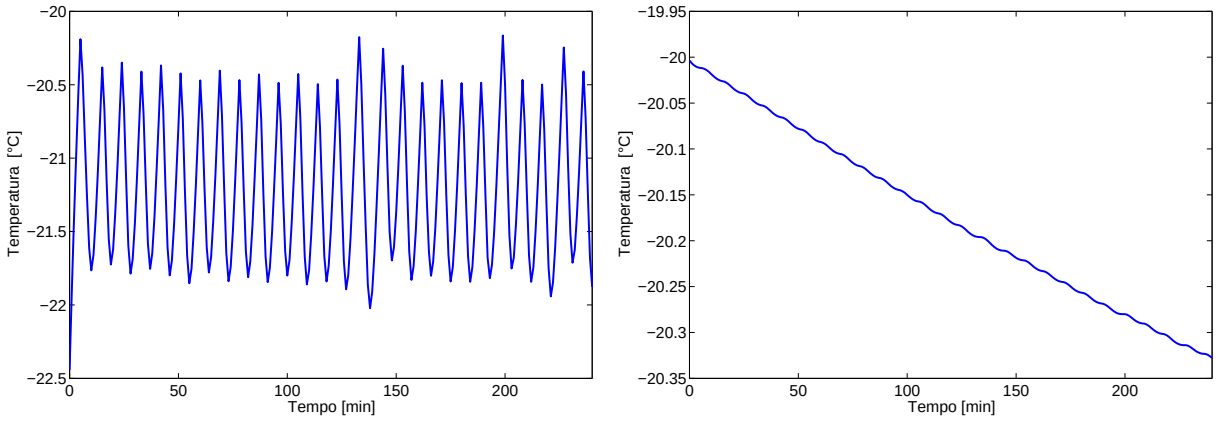

Figura 4.8 Temperatura interna (sx) e del carico (dx) in una delle celle LT con controllore RL (SARSA)

Vengono riportati ora i risultati ottenuti utilizzando l'algoritmo Q-Learning che a differenza del SARSA utilizza la ricompensa massima per lo stato successivo nell'aggiornamento della funzione stato-azione $Q(s,a)$.

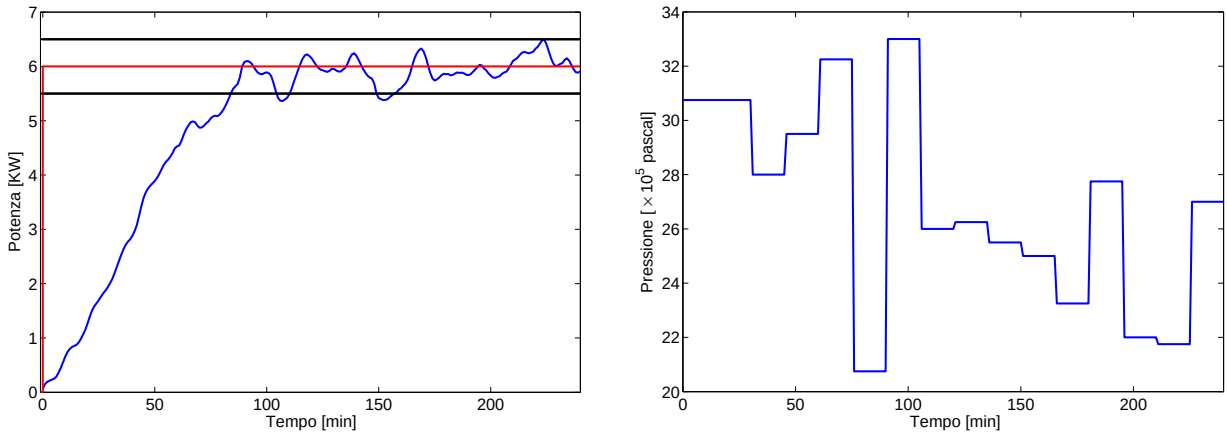

Figura 4.9 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e Set-point della pressione MT (dx)(Q-Learning)

Il set-point viene cambiato ogni 15 minuti come si può vedere dai grafici relativi all'ingresso di controllo in figura 4.5 e 4.9. In figure 4.6 e 4.10 si nota come inizialmente il controllore non riesce a massimizzare la ricompensa totale ma con il passare degli episodi, attraverso l'apprendimento e il continuo aggiornamento di $Q(s,a)$, riesce a raggiungere la policy ottima. Lo scostamento della ricompensa totale dal valore massimo è dovuto alla fase di esplorazione, infatti con il passare degli episodi la probabilità di esplorazione decresce e la ricompensa si stabilizza sul valore massimo. Si può osservare come, a parità di tasso di apprendimento e probabilità di esplorazione (ε), l'algoritmo Q-Learning presenta un apprendimento più rapido e una più netta convergenza alla policy ottimale rispetto all'algoritmo SARSA.

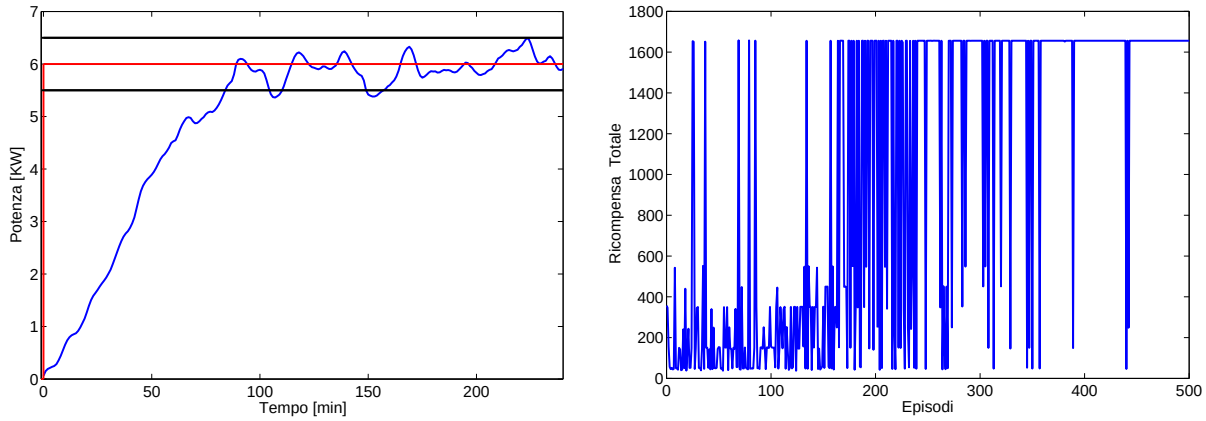

Figura 4.10 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e Ricompensa totale (dx) (Q-Learning)

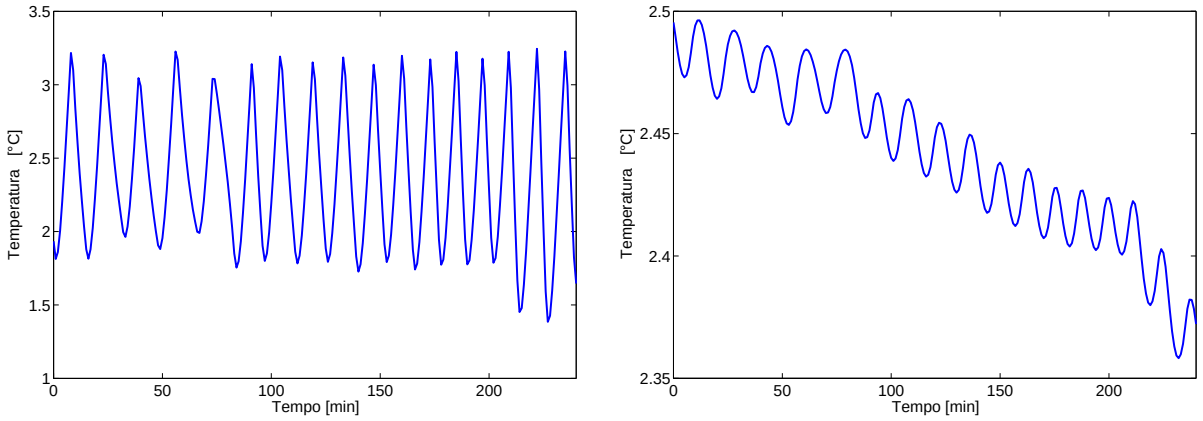

Figura 4.11 Temperatura interna (sx) e del carico (dx) in una delle celle MT con controllore RL (Q-Learning)

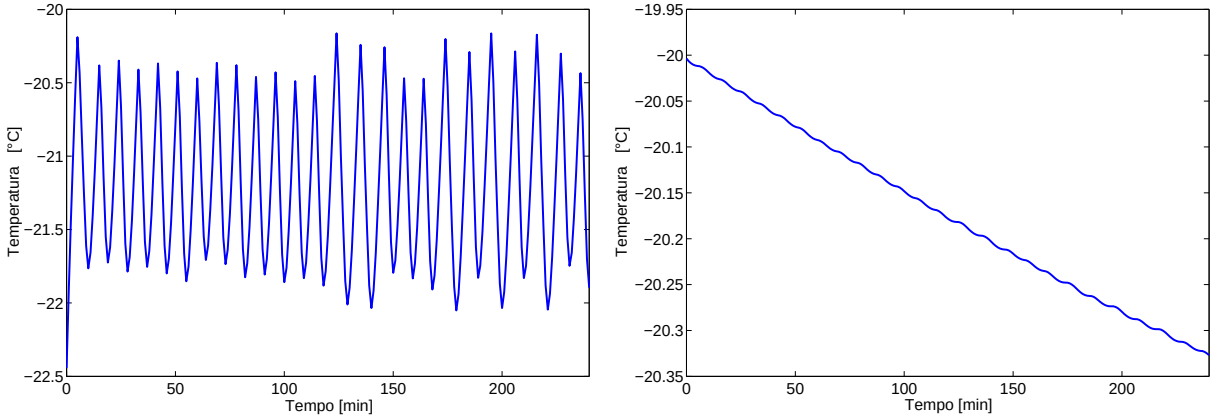

Figura 4.12 Temperatura interna (sx) e del carico (dx) in una delle celle LT con controllore RL (Q-Learning)

É, inoltre, possibile osservare dagli andamenti delle ricompense totali in figura 4.6 e 4.10 che l'algoritmo Q-Learning apprende più rapidamente, presentando una percentuale di successi più elevata rispetto all'algoritmo SARSA a parità di numero di episodi. Confrontando i risultati ottenuti con i due algoritmi:

- Le policy ottime ottenute con i due algoritmi sono diverse e la ricompensa ottenuta con il Q-learning (1655.8) è leggermente inferiore rispetto a quella ottenuta con il SARSA (1657.9).
- Il tempo di convergenza per il raggiungimento e la stabilizzazione della policy ottima è diversa a seconda dell'algoritmo utilizzato. In questo caso l'algoritmo Q-Learning in circa 200 episodi raggiunge la policy ottima mentre l'algoritmo SARSA impiega circa 500 episodi.
- Come si può vedere nelle figure 4.7, 4.8, 4.11 e 4.12, i vincoli di temperatura relativi alla temperatura dell'aria interna alla cella e del carico sono rispettato in entrambi i casi sia per le celle a media temperatura che per quelle a bassa temperatura.
- Le prestazioni degli algoritmi sono fortemente influenzate dagli insiemi delle possibili azioni che il controllore può compiere (A_t), dall'insieme degli stati (S_t) e dal tempo di supervisione. Aumentando il tempo di supervisione le prestazioni peggiorano e c'e il rischio che l'algoritmo non riesca a raggiungere la policy ottima mentre se si diminuisce eccessivamente il tempo di supervisione, il tempo di convergenza diventa troppo elevato.
- Diminuendo l'intervallo scelto per la funzione ricompensa (0.5 [KW]) si ottengono soluzioni più precise ma i tempi necessari per il raggiungimento della policy ottima saranno maggiori.

4.3 Progettazione controllo PI: Set-point Temperatura

Si passa ora al secondo scenario presentato nella tabella 4.2. Il controllo di alto livello contiene un PI per decidere i valori dei set-point delle temperature LT e MT al fine di assicurare un consumo di potenza medio su 60 minuti che segue il riferimento di 6 KW rispettando i vincoli del problema. Si ipotizza in questo caso che le pressioni di aspirazione delle celle MT e LT siano costanti e pari rispettivamente a 26×10^5 pascal e 12×10^5 pascal. Il controllore di alto livello modifica i valori dei set-point delle temperature delle celle LT e MT fornisce alla cella i-esima la variazione ΔT_i. L'aggiunta al riferimento del termine ΔT_i permette non solo di conservare in maniera adeguate il cibo (carico) all'interno delle celle ma anche di seguire il riferimento di potenza assorbita assegnata. Il controllo somma al set-point fissato $T_{i,0}$ il valore ΔT_i, come si può vedere in figura 4.13, determinando il nuovo set-point di temperatura per la cella i-esima:

$$T_i = T_{i,0} + \Delta T_i$$

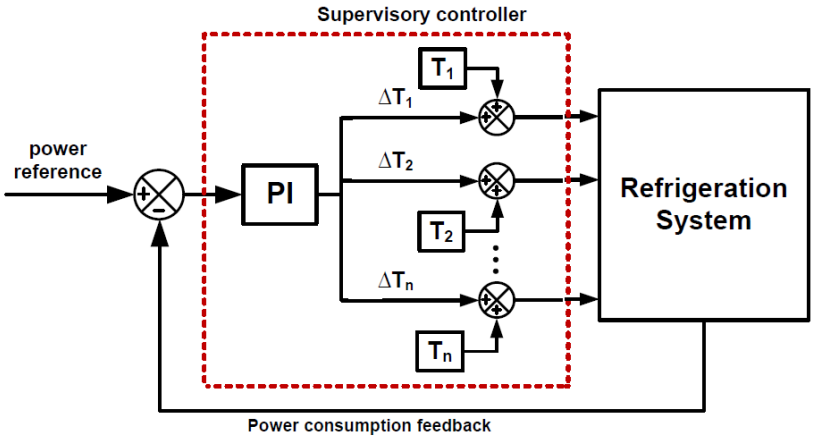

Figura 4.13 Architettura controllo temperatura

Per ottenere le prestazioni volute è necessaria una fase di taratura dei parametri del PI. I guadagni sono impostati a:

$$K_P = 0.02 \qquad K_I = 0.01$$

Il tempo di supervisione è costante e pari a 15 minuti e $T_{i,0}$ assume un valore costante pari a 2.5° C per le celle frigorifere a media temperatura e -21° C per le celle a bassa temperatura. Anche in questo caso non si considera la presenza di disturbi tempo-varianti quindi la temperatura interna del supermercato e la temperatura esterna sono fissate (26 °C e 12 °C).

I risultati ottenuti con il controllore PI sono i seguenti:

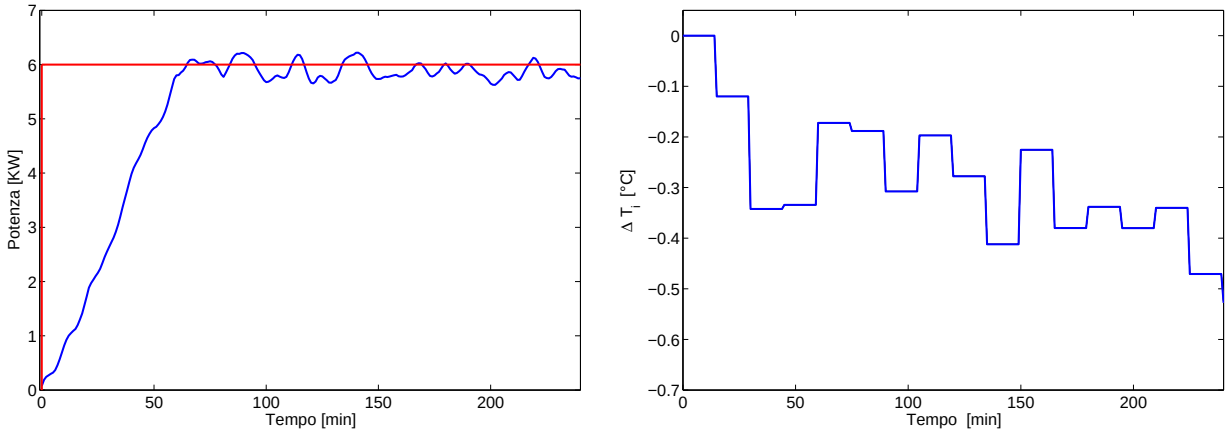

Figura 4.14 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e ΔT_i delle celle frigorifere per il controllore PI (dx)

In figura 4.14 si può notare come il ΔT_i assuma valori negativi quindi le temperature delle celle si abbassano e questo è comprensibile poichè diminuendo la temperatura si ottiene un consumo di potenza maggiore da parte dei compressori (a parità di carico e condizioni al contorno). Diversamente da ciò che succedeva nel primo scenario dove i valori dei set-point di temperatura e, conseguentemente anche i limiti del controllore locale sono fissi, in questo caso variando i valori dei set-point vengono modificati anche i limiti della temperatura dell'aria all'interno della cella come si può vedere in figura 4.15 e figura 4.16. Nonostante ciò la temperatura delle derrate soddisfa i vincoli del problema. Confrontando i risultati ottenuti con i PI nei due differenti scenari si può vedere come la potenza controllata agendo sui riferimenti di temperatura delle celle ha un andamento molto più regolare a differenza del consumo di potenza controllato agendo sul riferimento della pressione di aspirazione MT.

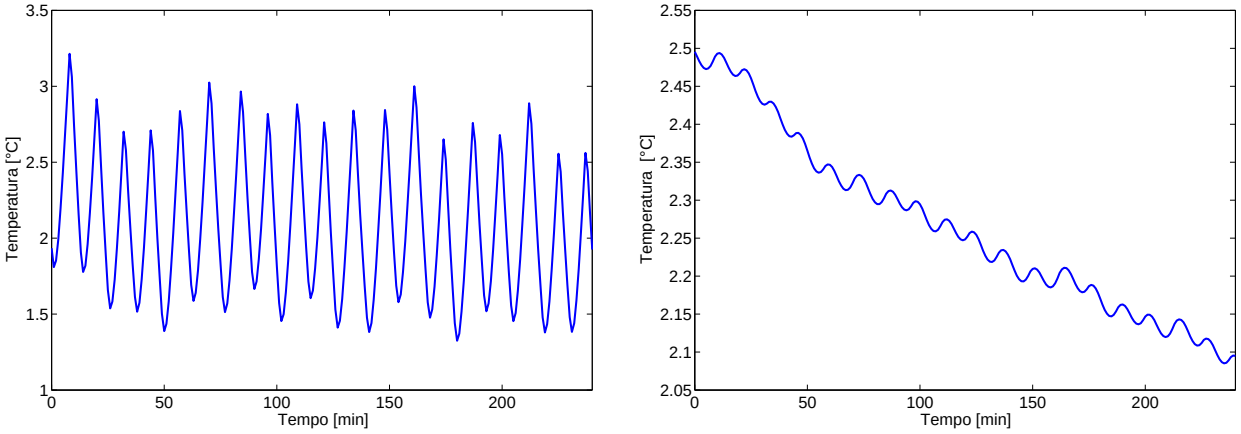

Figura 4.15 Temperatura interna (sx) e del carico (dx) in una delle celle MT con controllore PI

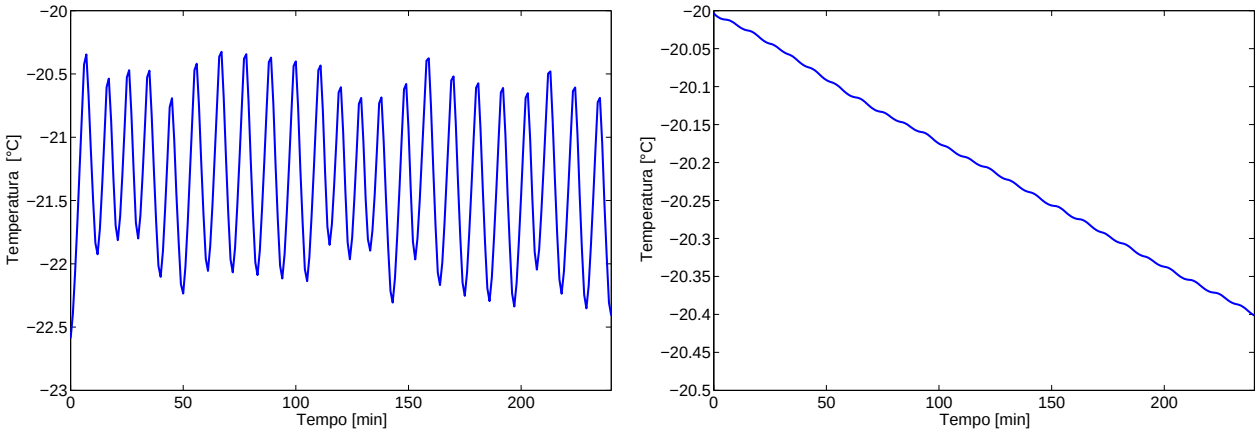

Figura 4.16 Temperatura interna (sx) e del carico (dx) in una delle celle LT con controllore PI

4.4 Progettazione controllo RL: Set-point Temperatura

Anche in questo caso, per la progettazione del controllore attraverso l'utilizzo di algoritmi di apprendimento mimetico, è necessario definire il sistema come un processo di decisione Markoviano. L'obiettivo del controllo di alto livello è di determinare un ΔT_i da sommare al valore del set-point fissato $T_{i,0}$ per mantenere il consumo di potenza.
Si definiscono:

- L'insieme delle possibili azioni, cioè l'insieme dei possibili valori di ΔT_i, tra cui il controllore può scegliere (A_t):

$$A_t = \{-1,\ -0.9,\ -0.8,\ \ldots\ 0.8,\ 0.9,\ 1\} \quad [^\circ C]$$

- L'insieme degli stati del sistema (S_t) resta invariato:

$$S_t = \{Potenza\ assorbita\ all'istante\ t\}$$

- Anche la funzione ricompensa resta invariata:

$$e_t = errore\ a\ regime$$

$$r_t = \begin{cases} 100 - 2 \cdot |e_t|, & \text{se } |e_t| \leq 0.5\ [KW] \\ -1000 - 10 \cdot |e_t|, & \text{se } |e_t| > 0.5\ [KW] \end{cases}$$

Si è deciso di applicare solo l'algoritmo Q-Learning visto la maggiore velocità di convergenza alla policy ottima esibite per questo tipo di problema. Per applicare il Q-Learning è necessario impostare il tasso di apprendimento (α), il discount factor (γ), la probabilità di esplorare nuove policy (ε) e l'inizializzazione della funzione stato-azione ($Q(s,a)$). Il tasso di apprendimento inizialmente è impostato a 0.7 e decresce ad ogni episodio fino a raggiungere un minimo di 0.125. La probabilità di esplorazione resta fissata a 0.2 e decresce col passare degli episodi. La funzione stato-azione viene inizializzata, come per la pressione, a valori nulli. I risultati ottenuti utilizzando l'algoritmo Q-Learning sono rappresentati nelle figure 4.17 e 4.18.

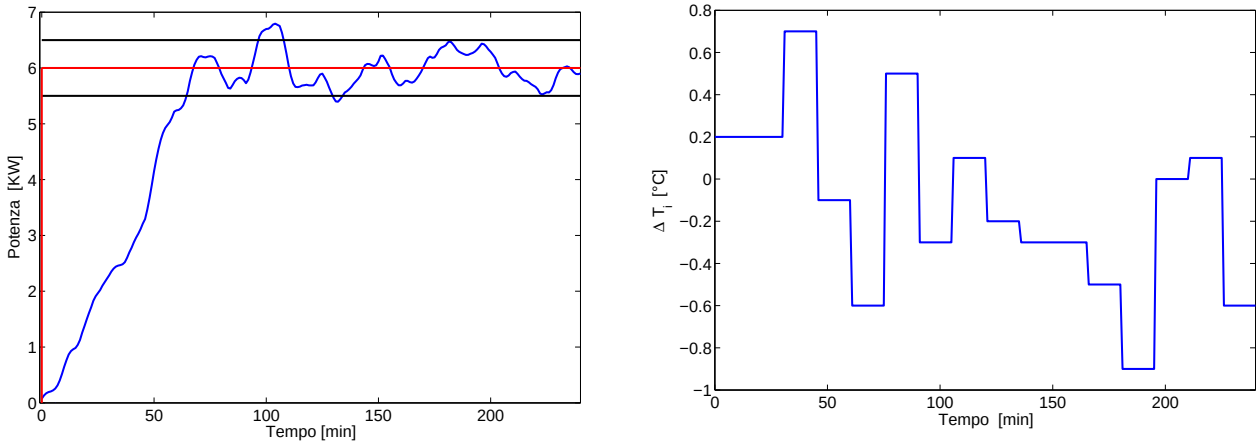

Figura 4.17 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e ΔT_i (dx) per il controllore RL (Q-Learning)

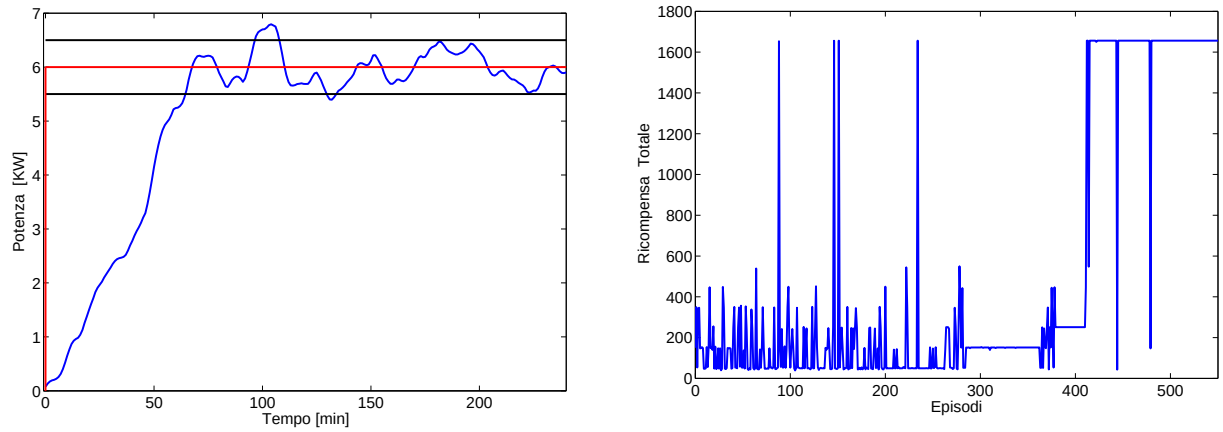

Figura 4.18 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e Ricompensa totale (Q-Learning) (dx)

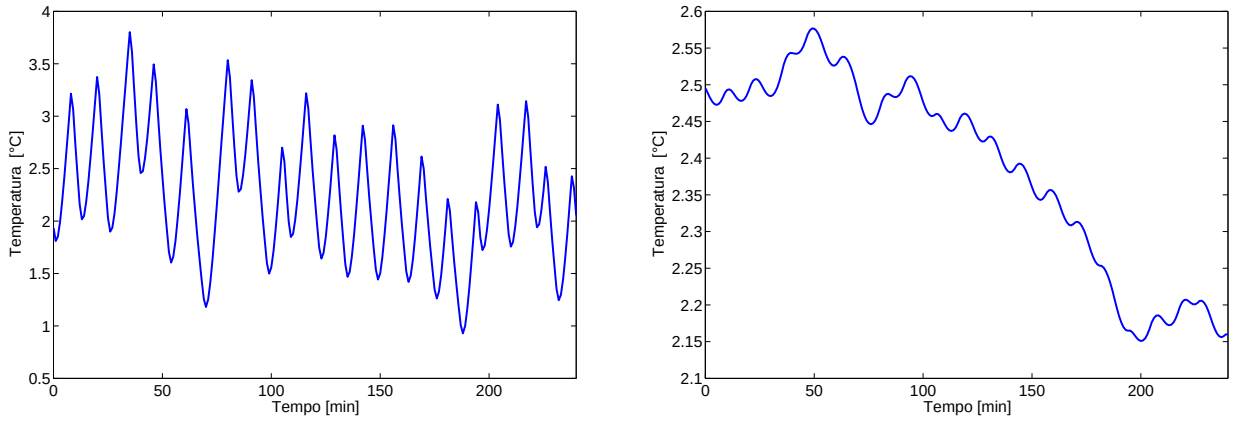

Figura 4.19 Temperatura interna (sx) e del carico (dx) in una delle celle MT con controllore RL (Q-Learning)

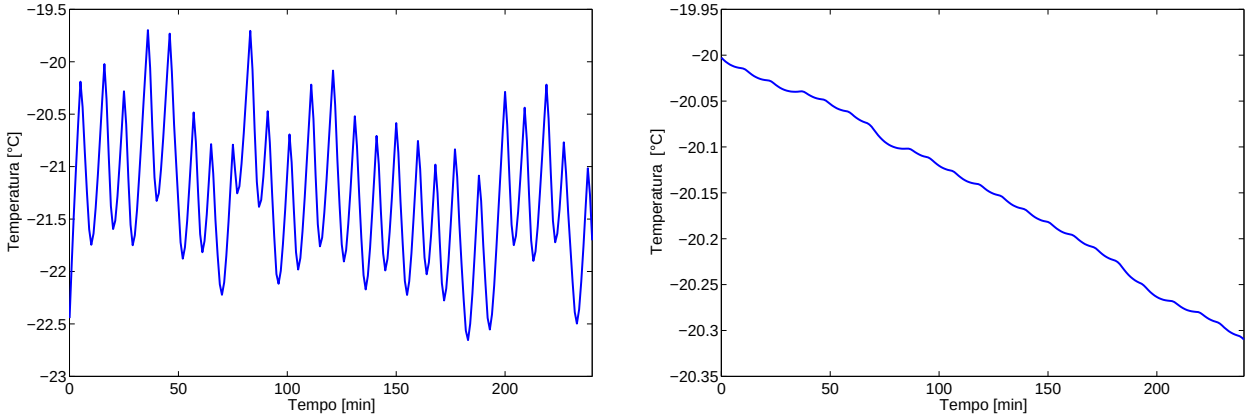

Figura 4.20 Temperatura interna (sx) e del carico (dx) in una delle celle LT con controllore RL (Q-Learning)

Si può osservare che:

- In figura 4.18 la ricompensa totale assume un valore massimo pari a 1656.15, inferiore alla ricompensa ottenuta controllando la pressione di aspirazione MT con l'algoritmo SARSA, ma superiore rispetto all'algoritmo Q-Learning (sezione 4.2.1)
- Il tempo di convergenza alla policy ottima controllando le temperature delle celle frigorifere è superiore (450 episodi) rispetto al tempo di convergenza ottenuto regolando la pressione di aspirazione delle celle MT (200 episodi)
- Come si può vedere in figura 4.19 l'andamento della temperatura interna della cella è irregolare ma i controllori locali mantengono la temperatura entro i vincoli imposti (vincoli dipendenti dal set-point) mentre la temperatura delle derrate alimentari in figura 4.20 è regolare e rispetta i vincoli dal problema.

4.5 Progettazione controllo RL con informazione a priori

I risultati ottenuti con un controllo di alto livello basato su tecniche di apprendimento mimetico portano a prestazioni comparabili con quelli dei controllori standard PI. Spesso si preferisce utilizzare i controllori standard perchè i controllori basati sull'apprendimento mimetico hanno inizialmente bisogno di apprendere dal sistema le informazioni necessarie per eseguire il controllo. L'interazione continua tra il controllore e il sistema attraverso lo scambio di ricompense porta il controllo a raggiungere la policy ottima dopo un certo periodo di tempo (che può essere considerevole). Infatti se all'istante iniziale il controllore non possiede alcuna informazione relativa al sistema da controllare allora la fase di apprendimento può essere piuttosto impegnativa e richiede un certo numero di episodi. É possibile velocizzare il tempo di convergenza a π^* introducendo un informazione a priori all'interno del controllore. L'aggiunta dell'informazione a priori può avvenire in diversi modi, per esempio in [19] gli autori suggeriscono delle modifiche all'algoritmo Q-Learning al fine di velocizzare l'apprendimento mentre in [20] gli autori introducono due metodi per l'aggiunta di informazione a priori basati sulla scomposizione del problema in sottoproblemi più semplici.
Consideriamo il caso della determinazione del valore del set-point di pressione di aspirazione delle celle MT

Set-point	Valore
Temperatura MT	2.5 °C
Temperatura LT	-21 °C
Pressione MT	variabile decisionale
Pressione LT	12×10^5 pascal

Tabella 4.3 Set-point controllo supervisore

Nel caso preso in considerazione l'informazione a priori da fornire a controllore è rappresentata dall'uscita del PI progettato nella sezione 4.1. Utilizzando l'ingresso di controllo del PI si può ridurre il range delle possibili azioni dell'insieme A_t aumentando il passo di discretizzazione. Applicando tale strategia le cardinalità di A_t aumenta leggermente ma diminuisce notevolmente la cardinalità dell'insieme degli stati raggiungibili S_t. É necessario ridefinire il problema:

- L'insieme delle possibili azioni, cioè l'insieme dei possibili valori di pressione delle celle MT, che il controllore può compiere (A_t):

$$A_t = \{20,\ 20.1,\ 20.2,\ 20.3,\ \ldots.\ 25.8,\ 25.9,\ 26\}$$

- L'insieme degli stati del sistema (S_t):

$$S_t = \{Potenza\ assorbita\ all'istante\ t\}$$

- una funzione ricompensa che restituisce per un determinato stato la ricompensa r_t:

$$e_t = errore\ a\ regime$$

$$r_t = \begin{cases} 100 - 2 \cdot |e_t|, & \text{se } |e_t| \leq 0.5\ [KW] \\ -1000 - 10 \cdot |e_t|, & \text{se } |e_t| > 0.5\ [KW] \end{cases}$$

I parametri vengono inizializzati esattamente con gli stessi valori delle sezioni precedenti, il tempo di supervisione è fisso a 15 minuti e l'algoritmo utilizzato è sempre il Q-Learning. I risultati ottenuti con l'aggiunta dell'informazione a priori sono notevolmente migliori rispetto a quelli senza informazione a priori:

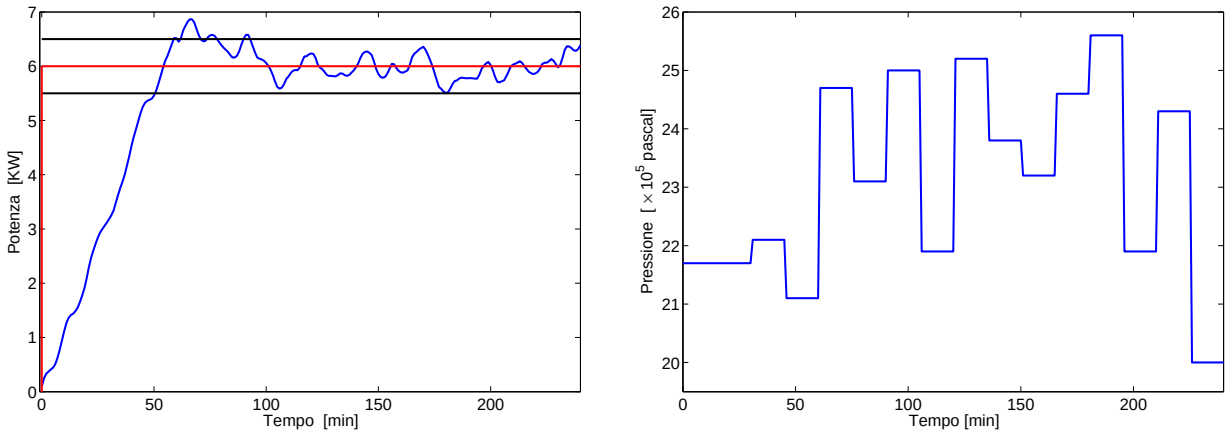

Figura 4.21 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e Set-point (dx) delle celle MT per il controllore RL con info a priori

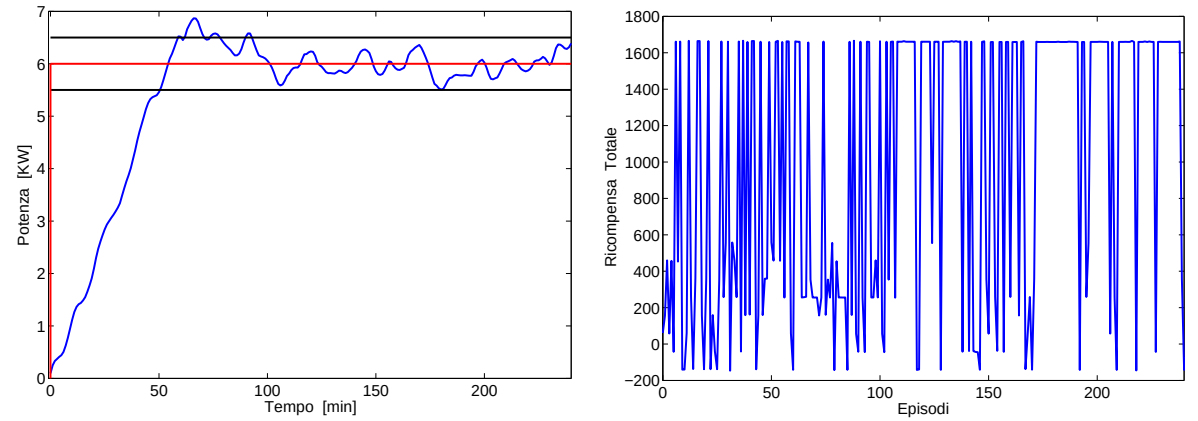

Figura 4.22 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e Ricompensa totale (dx) per il controllore RL con info a priori

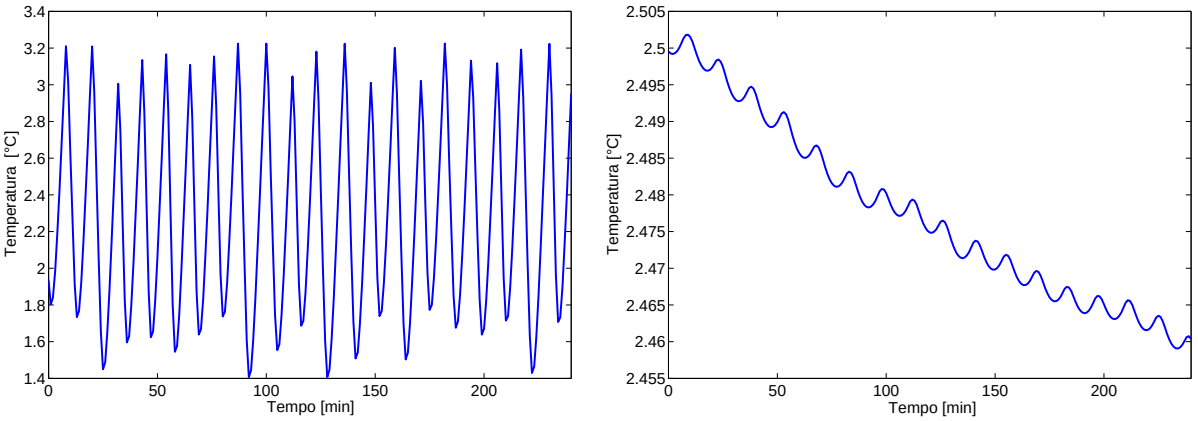

Figura 4.23 Temperatura interna (sx) e del carico (dx) in una delle celle MT per il controllore RL con info a priori

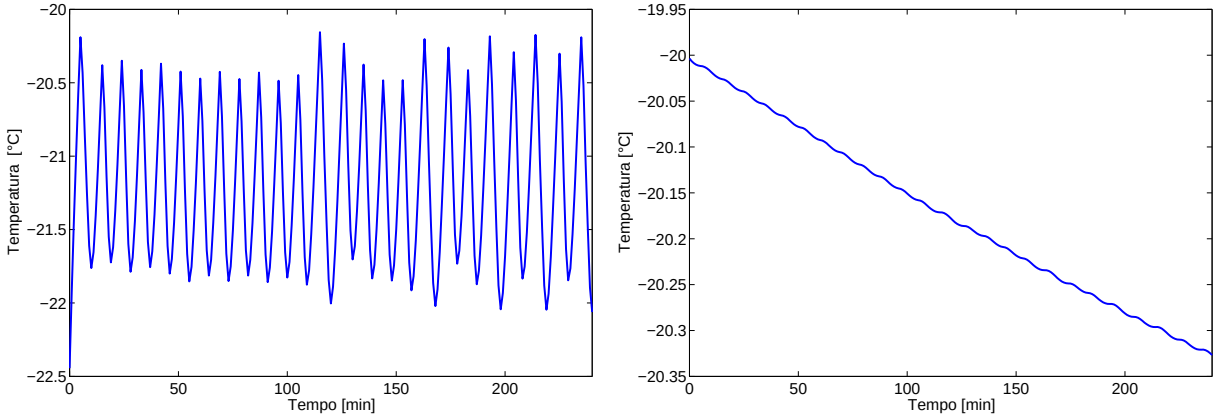

Figura 4.24 Temperatura interna (sx) e del carico (dx) in una delle celle LT per il controllore RL con info a priori

Si può osservare che:

- L'informazione a priori utilizzata modifica l'insieme A_t. Per scegliere il range delle possibili azioni si è preso il valore medio dell'ingresso di controllo ottenuto con il PI per il set-point di pressione. Una volta determinato il valore medio (circa 22.8×10^5 pascal) si è deciso di muoversi in un range del 10 % (circa) dal valore medio.

- Come si può vedere dall'andamento della ricompensa totale in figura 4.22 il numero di episodi necessario a raggiungere la policy ottima si riduce notevolmente rispetto al controllore senza informazione a priori infatti già dopo pochi episodi il controllo ha raggiunto π^*.

- A causa della fase di esplorazione la policy ottima non viene mantenuta costantemente ma il controllore cerca policy che portano a ricompense maggiori ma non trovandole non aggiorna π^*. Modificando la probabilità di esplorazione (ε) è possibile diminuire l'esplorazione e rendere più stabile il grafico relativo alla ricompensa totale in figura 4.22 con il rischio di non raggiungere la policy ottima.

- A parità di episodi, per esempio i primi 200 episodi, la percentuale di successi del controllore con informazione a priori è molto maggiore rispetto al supervisore senza alcuni informazione a priori riportato in figura 4.10.

- Aggiungendo informazione a priori si può pensare di diminuire il passo di discretizzazione del supervisore così da ottenere un ingresso di controllo che assicura prestazioni migliori. Infatti la ricompensa totale assume un valore maggiore (1660.84) rispetto ai risultati senza informazione a priori. La ricompensa è maggiore tanto più l'errore e_t è minore (per definizione della funzione ricompensa).

- I vincoli di temperatura dell'aria interna e delle derrate per le celle MT e LT sono rispettati anche in questo caso come è possibile vedere in figure 4.23 e 4.24.

Per confrontare le tecniche utilizzate per la progettazione del controllo di alto livello per il sistema di refrigerazione si introduce come parametro di valutazione l'errore medio pesato dal tempo nell'intervallo temporale di 4 ore. L'ITSE discreto (integral time square error) è definito come:

$$ITSE = \sum_{k=0}^{N} t_k \cdot e_k^2 \tag{4.1}$$

dove e_k è l'errore a regime al k-esimo intervallo di supervisione. Dalla tabella è chiaro come il controllo ottenuto attraverso tecniche di apprendimento mimetico ha ITSE (e il corrispondente errore medio) maggiore rispetto a quello ottenuto con il PI. Il vantaggio sta

nel fatto che il controllore, una volta inizializzati i parametri, ha autonomamente appreso dall'esperienza senza bisogno di un modello e di una fase di taratura dei guadagni. Purtroppo il tempo di convergenza alla policy ottima è significativo e per questo si è deciso di introdurre un informazione a priori nel controllore che ha portato non solo ad un miglioramento da un punto di vista temporale ma anche una diminuzione dell'errore medio.

Controllore	ITSE ($\times 10^4$)	Errore medio [KW]
PI	1.6239	0.7195
RL SARSA	2.0422	0.8403
RL Q-Learning	2.2328	0.8787
RL con info a priori	1.0790	0.6108

Tabella 4.4 Confronto controllori di alto livello

4.6 Effetti di un disturbo tempo variante

Uno dei principali punti di forza del controllore PI è la sua semplicità di implementazione e di taratura dei guadagni per ottenere le prestazioni desiderate. Con il passare del tempo spesso il deterioramento delle componenti del sistema o una variazione dalle condizioni al contorno può portare ad un peggioramento delle prestazioni del controllore. Un regolatore basato su tecniche di apprendimento mimetico, al contrario, grazie alla continua interazione con l'ambiente riesce ad adattarsi alle nuove condizioni modificando la policy ottima. In un sistema di refrigerazione commerciale difficilmente certi disturbi si mantengono costanti (e.g la temperatura esterna/interna può variare da un giorno all'altro o addirittura da un'ora all'altra). Considerando il modello presentato nel capitolo 2, la temperatura interna al supermercato è fissata pari a 26 °C. Nella realtà tale temperatura tende a non rimanere costante ma varia con il passare delle ore (anche in presenza di sistemi di condizionamento della temperatura ambiente). Nonostante le variazioni della temperatura dell'ambiente il controllo di alto livello dovrà essere in grado di fornire i valori dei set-point di pressione e temperatura alle celle per mantenere un determinato consumo di potenza media rispettando i vincoli del sistema.

Si consideri un intervallo temporale di 4 ore e di supponga che la temperatura interna dopo 3 ore sia soggetta ad una brusca variazione passando da 26 °C a 31°C. Come si può vedere nella figura 4.25 il controllore PI al variare della temperatura interna dopo 180 minuti ottiene prestazioni peggiori rispetto al controllore basato su tecniche di apprendimento mimetico. Il controllore di alto livello basato sull'apprendimento mimetico utilizza l'algoritmo

Q-Learning per la fase di apprendimento e non viene utilizzata nessuna informazione a priori. Il tempo di supervisione per entrambi i controllori è fissato a 15 minuti. La potenza assorbita con il controllo con apprendimento mimetico riesce a mantenere il consumo entro un intervallo stabilito di 0.5 KW mentre i controllo PI non riesce ad ottenere le stesse prestazioni. Prendendo come parametro di valutazione dei controllori l'errore medio pesato (il quale attribuisce una maggiore rilevanza all'errore a regime) si osserva che l'errore medio per il controllo PI è di 0.845 KW mentre per il controllo con apprendimento mimetico è di 0.79 KW. Riassumendo:

Controllore	Errore medio [KW]
PI	0.7195
PI con disturbo	0.84567
RL Q-Learning	0.8787
RL Q-Learning con disturbo	0.79

Tabella 4.5 Confronto controllori di alto livello soggetti a disturbo

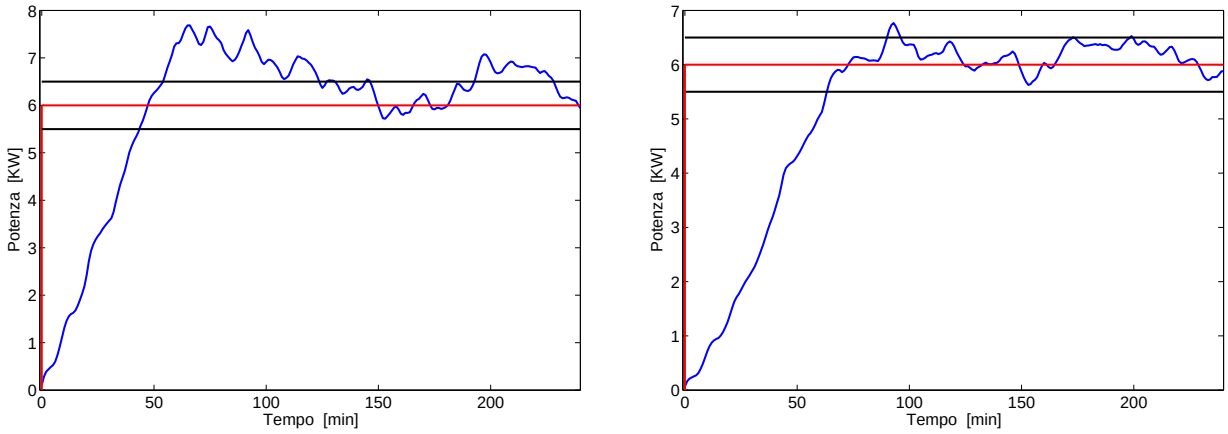

Figura 4.25 Confronto potenza assorbita con disturbo

Dai risultati ottenuti il controllore PI peggiora notevolmente le sue prestazioni a parità di guadagni mentre il controllore basato sul Q-Learning ottiene un errore medio minore perchè è riuscito a modificare opportunamente gli ingressi di controllo compensando l'effetto introdotto dal disturbo (i.e la temperatura interna del supermercato che è variata).

4.7 Progettazione controllo RL multivariabile

I controllori di alto livello progettati fino a questo punto determinano i set-point delle pressioni e delle temperature nelle singole celle. Nella sezione 4.2 e 4.4 per semplificare il problema di controllo alcuni set-point vengono fissati a valori costanti e la regolazione avviene attraverso una singola variabile di controllo;in questo modo un sistema MISO (multiple input single output) è trattato più semplicemente come un sistema SISO (single input single output).
Si procede ora alla progettazione di un controllo di alto livello per il sistema multivariabile (MISO). Il controllo supervisore deve fornire i set-point di temperatura delle singole celle. nonchè i valori di pressione delle celle MT:

Set-point	Valore
Temperatura MT	variabile decisionale
Temperatura LT	variabile decisionale
Pressione MT	variabile decisionale
Pressione LT	12×10^5 pascal

Tabella 4.6 Scenario 3: Set-point controllo supervisore

Infatti le variabili controllate sono sia le temperature delle celle a media e bassa temperatura che la pressione di aspirazione delle celle MT mentre il valore del set-point della pressione delle celle LT viene mantenuta costante pari a 12×10^5 pascal. Anche in questo caso si può decidere se utilizzare un approccio model-based oppure un approccio model-free come l'apprendimento mimetico. Il set-point delle temperature viene modificato applicando alla cella i-esima una variazione ΔT_i. Il controllo di alto livello somma al set-point fissato $T_{i,0}$ il valore ΔT_i per regolare il consumo di potenza, come succedeva nel caso SISO (sezione 4.3). $T_{i,0}$ è un valore costante pari a 2.5°C per le celle MT e -21°C per le celle frigorifere a bassa temperatura.

4.7.1 Controllore PI

Il controllo supervisore contiene un PI che effettua l'azione di controllo con un tempo di supervisione costante e pari a 15 minuti. L'uscita del PI viene moltiplicata per due guadagni, uno determinerà il valore di ΔT_i mentre l'altro la pressione di aspirazione della cella MT. L'utilizzo del PI, nel caso multivariabile, non solo presenta la difficoltà di tarare i guadagni interni al PI (K_P e K_I) ma è necessario determinare i guadagni ausiliari al fine di avere delle azioni di controllo (set-point) adeguate. I guadagni dei PI sono stati tarati a:

$$K_P = 1 \qquad\qquad K_I = 0.5$$

mentre i guadagni esterni al PI sono stati impostati a:

$$G_{Press} = 0.55 \qquad G_{Temp} = 0.02$$

Il guadagno relativo alla pressione è decisamente maggiore rispetto a quello di temperatura perchè a parità di errore di potenza è necessaria di una variazione maggiore di pressione rispetto alla temperatura delle celle per compensarlo.

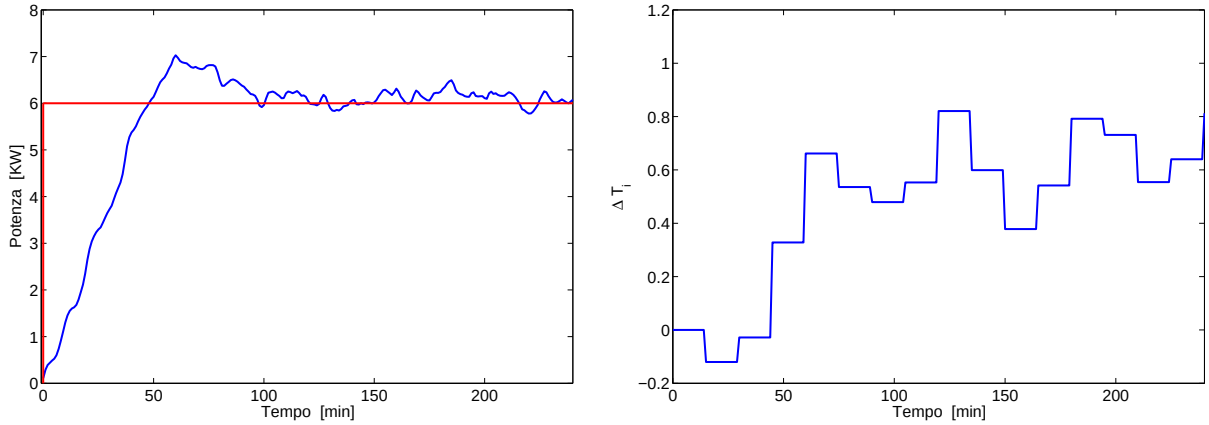

Figura 4.26 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e Set-point (dx) delle celle MT per il controllore PI

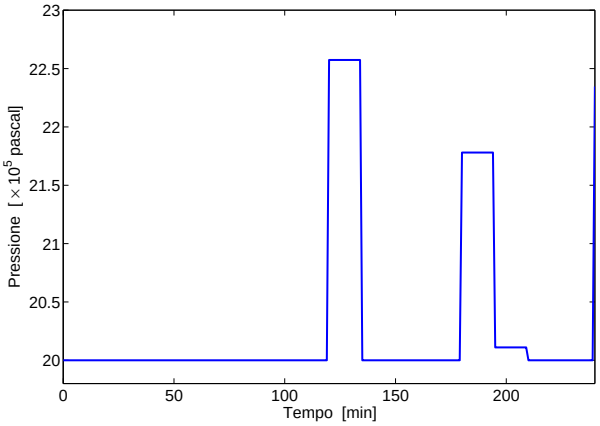

Figura 4.27 Set-point della pressione per le celle MT per il controllore PI

I vincoli di temperatura come si può vedere nelle figure 4.28 e 4.29 sono rispettati e le osservazioni sul comportamento del PI multivariabile sono, sostanzialmente, le stesse riportate per il PI con un unica variabile di controllo.

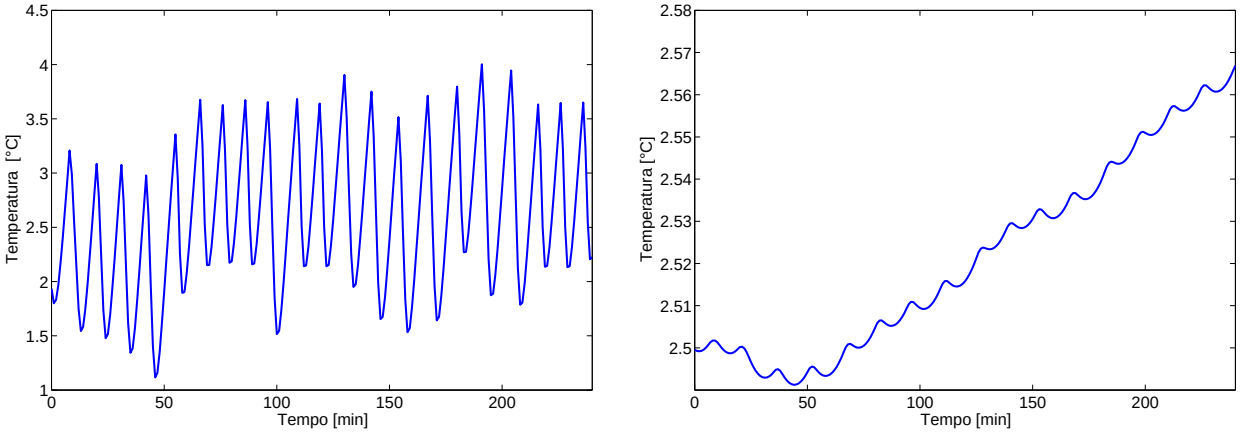

Figura 4.28 Temperatura interna (sx) e del carico (dx) in una delle celle MT con controllo PI

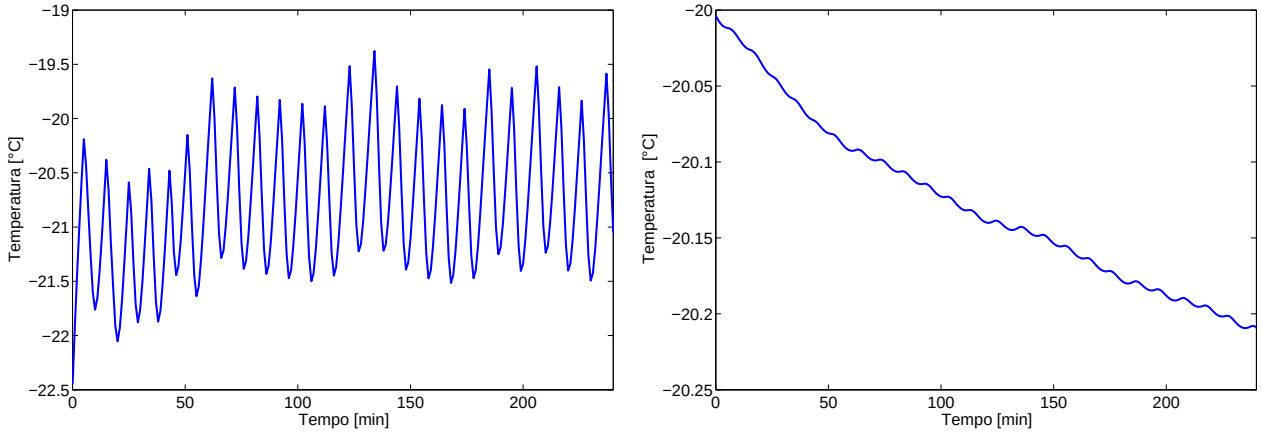

Figura 4.29 Temperatura interna (sx) e del carico (dx) in una delle celle LT con controllo PI

4.7.2 Apprendimento mimetico

Si procede ora alla progettazione di un controllo supervisore basato sull'apprendimento mimetico multivariabile. In un sistema SISO il controllore basato sull'apprendimento mimetico sceglieva un'azione da compiere in un insieme A_t costituito da singoli valori. In un sistema MISO il controllore dovrà inizialmente scegliere una coppia di valori, uno per la pressione di aspirazione MT e uno per le variazioni di temperatura ΔT_i, valutare lo stato a cui si porta il sistema s_{t+1} e restituire la ricompensa r_{t+1} al controllore.
L'insieme delle azioni A_t è definito come:

$$A_t = (p_i,\ \Delta T_j) \quad \forall \quad p_i \in P_t$$
$$\forall \quad \Delta T_j \in \Gamma_t$$

dove P_t è l'insieme di tutti i possibili valori di pressione pari a :

$$P_t = \{20,\ 20.1,\ 20.2,\ \ldots\ldots\ 22.8,\ 22.9,\ 23\} \quad [\times 10^5 pascal]$$

e Γ_t è l'insieme di tutti i possibili valori di ΔT:

$$\Gamma_t = \{-1,\ -0.9,\ \ldots..\ 0.8,\ 0.9,\ 1\} \quad [^\circ C]$$

L'insieme degli stati del sistema S_t è definito come nei casi precedenti:

$$S_t = \{Potenza\ assorbita\}$$

e anche la funzione ricompensa:

$$e_t = errore\ a\ regime$$

$$r_t = \begin{cases} 100 - 2 \cdot |e_t|, & \text{se } |e_t| \leq 0.5\ [KW] \\ -1000 - 10 \cdot |e_t|, & \text{se } |e_t| > 0.5\ [KW] \end{cases}$$

La policy ottima, in questo caso, è una sequenza di coppie che massimizzano la ricompensa totale in un intervallo di tempo fissato. Nel controllo multivariabile la cardinalità di A_t è molto maggiore rispetto al caso con un unica variabile di controllo, nonostante il range dei possibili valori di pressione di aspirazione sia limitatoo. L'aumento della cardinalità di A_t genera un aumento significativo del tempo di convergenza alla policy ottima π^*. La funziona stato-azione $Q(s,a)$ nel sistema SISO era inizializzata a valori nulli. Nel controllore multivariabile, inizializzando la $Q(s,a)$ a valori nulli, si ottiene un tempo di convergenza alla policy ottima troppo elevato e per questo si è deciso di inizializzare la funzione $Q(s,a)$ con valori casuali in modo tale da velocizzare la fase di apprendimento. La durata della fase di apprendimento, a causa dell'inizializzazione casuale, è variabile e influenza notevolmente il comportamento del sistema. Il tasso di apprendimento (α) all'inizio ha valore pari a 0.7 e decresce fino ad un minimo di 0.125. Il discout factor è unitario e la probabilità di effettuare esplorazione (ε) è inizializzata a 0.2 e decresce proporzionalmente a numero degli episodi. Il tempo di supervisione è sempre 15 minuti e l'algoritmo utilizzato è l'algoritmo Q-Learning visto che i risultati ottenuti con il SARSA presentano un tempo per il raggiungimento di π^* inaccettabile nel caso in esame.

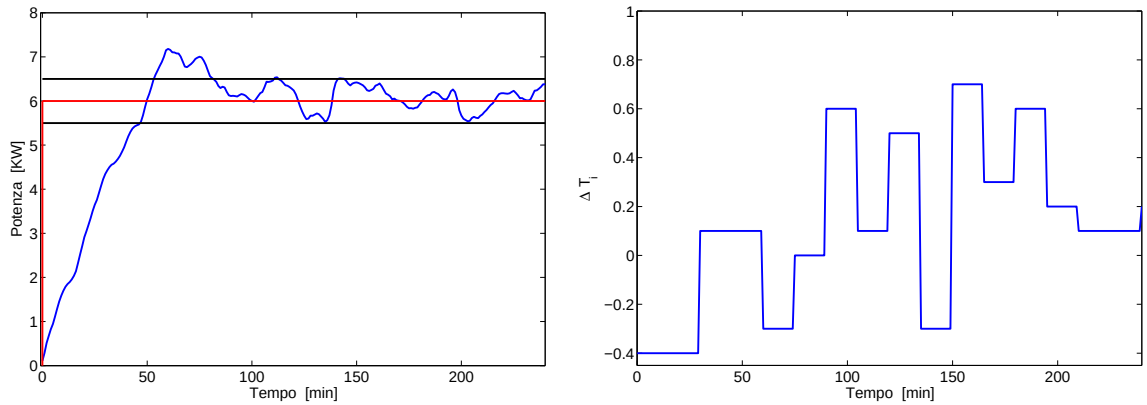

Figura 4.30 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e ΔT_i (dx) per il controllore RL multivariabile

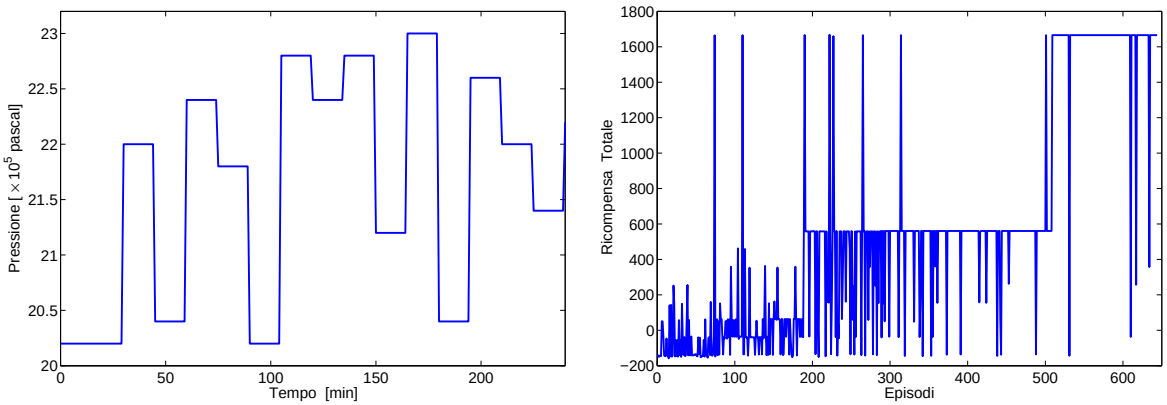

Figura 4.31 Risposta ad un riferimento di potenza media assorbita pari a 6 [KW] (sx) e Ricompensa totale (dx) con il controllore RL multivariabile

Si può osservare che:

- La policy ottima ottenuta con il Q-learning ha una ricompensa massima pari a 1665.86 superiore a qualsiasi ricompensa ottenuta precedentemente.

- Aumentando il numero delle variabili da controllare aumento la precisione ma aumenta significativamente anche il tempo di raggiungimento della policy ottima passando da 250 episodi a 500 episodi (4.31).

- I vincoli di temperatura relativi alla temperatura dell'aria interna alla cella e del carico sono rispettato in entrambi i casi sia per le celle a media temperatura che per quelle a bassa temperatura come si può vedere nelle figure 4.32 e 4.33 .

- Aumentando il tempo di supervisione o aumentando il range dei possibili valori della pressione di aspirazione le prestazioni peggiorano e c'e il rischio di una mancata convergenza a π^*.
- Anche in questo caso diminuendo l'intervallo scelto per la funzione ricompensa (0.5) si ottengono soluzioni più precise, a discapito però del tempo di convergenza.

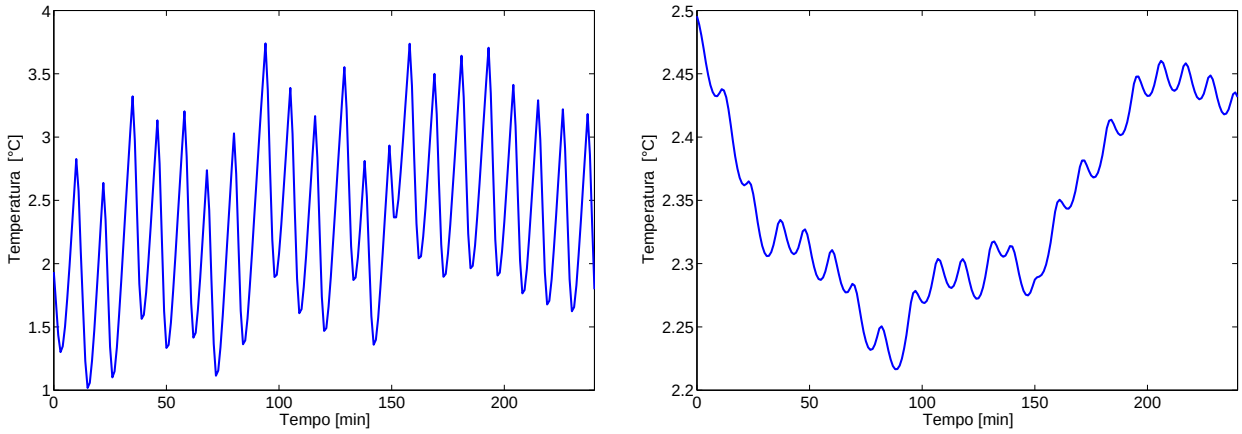

Figura 4.32 Temperatura interna (sx) e del carico (dx) in una delle celle MT con controllo RL multivariabile

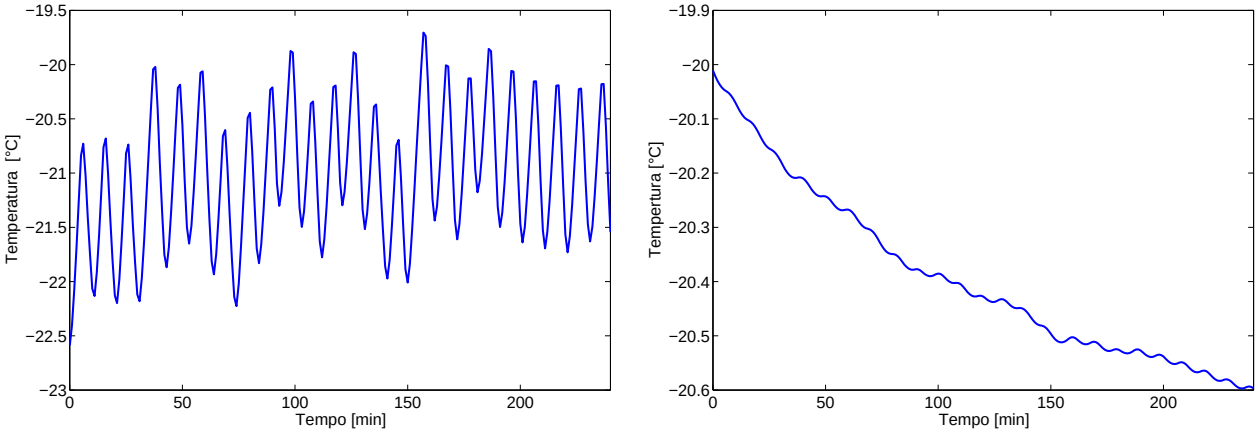

Figura 4.33 Temperatura interna (sx) e del carico (dx) in una delle celle LT con controllo RL multivariabile

Capitolo 5

Conclusioni

In questa tesi si è affrontato il progetto di un sistema di controllo di alto livello per un sistema di refrigerazione commerciale a CO_2. Inizialmente il controllore è stato realizzato utilizzando regolatori standard (PID) che sono stati tarati giovandosi di un modello relativo all'impianto di refrigerazione in esame; successivamente si è progettato un diverso algoritmo di controllo basato su un approccio di apprendimento mimetico che sfrutta tecniche di tipo model-free (Q-learning e SARSA). Si è verificato come il controllore PID fornisca dei buoni risultati con il vantaggio di essere semplice da implementare; per contro, la taratura dei relativi parametri non è stata agevole e ha sfruttato il modello dell'impianto di refrigerazione. Diversamente, l'apprendimento mimetico, che si basa sulla continua interazione tra controllore e sistema, imparando la sequenza di azioni da compiere per raggiungere un certo obiettivo, non necessita di un modello dell'impianto di refrigerazione e fornisce comunque delle buone prestazioni (paragonabili a quelle del regolatore PID). Il controllo che impiega il Reinforcement Learning può tuttavia richiedere un tempo di convergenza elevato. Per ovviare a questo problema si è utilizzata della informazione a priori per aiutare il controllo mimetico a determinare la sequenza di azioni di controllo ottimo in un tempo ragionevole. Le simulazioni condotte in ambiente Matlab/Simulink hanno utilizzato un modello di un impianto di refrigerazione commerciale di un tipico supermercato con celle di media e bassa temperatura. Le prestazione degli algoritmi di controllo sono state valutate assegnando un riferimento di potenza assorbita dal sistema di refrigerazione (seguendo una impostazione tipica dell'approccio demand-response) e impostando i limiti per le temperature all'interno delle celle frigorifere per garantire la corretta conservazione delle derrate. I risultati delle simulazioni indicano come il controllo che utilizza l'approccio mimetico consenta di ottenere buone prestazioni, in termini di errore sul riferimento e di reiezione ai disturbi rispetto al controllo di tipo PID. Il vantaggio principale dell'approccio con reinforcement learning è quello di non richiedere la costruzione di un modello del plant, che nel caso dell'applicazione in esame ha più ingressi e più uscite

ed esibisce tipicamente un comportamento non lineare, può essere molto oneroso.

5.1 Sviluppi futuri

Un possibili sviluppi del lavoro intrapreso possono essere:

- Miglioramento delle prestazioni del controllo multivariabile attraverso l'aggiunta di informazione a priori in modo tale da ottenere prestazioni accettabili anche con l'algoritmo SARSA.
- Confronto tra controllori basati su tecniche di apprendimento mimetico e controllori MPC.

Bibliografia

[1] Shafiei, Seyed Ehsan; Izadi-Zamanabadi, Roozbeh; Rasmussen, Henrik; Stoustrup, Jakob *A decentralized control method for direct smart grid control of refrigeration systems*,Decision and Control (CDC) 2013

[2] Pasgianos GD, Arvanitis KG, Polycarpou P, Sigrimis N. *A nonlinear feedback technique for greenhouse environmental control.* Comput Electron Agric 2003

[3] Dong B. *Non-linear optimal controller design for building HVAC systems.* In: Int Conf Control Appl (CCA). Yokohama, Japan: IEEE; 2010

[4] Tobias Gybel Hovgard, Lars F. S. Larsen, John Bagterp Jørgensen, Stephen Boyd *Fast Nonconvex Model Predictive Control for Commercial Refrigeration* 2012

[5] Li X, Shi Z, Hu S.*A novel control method of a variable volume air conditioning system for indoor thermal environment.* In: Int Conf Comput Eng Technol (ICCET). Chengdu, China: IEEE; 2010

[6] Salsbury TI. *A new pulse modulation adaptive controller (PMAC) applied to HVAC systems.* Control Eng Pract 2002

[7] Junping Cai, *Control of Refrigeration Systems for Trade-off between Energy Consumption and Food Quality Loss* Automation and Control Department of Electronic Systems Aalborg University Fredrik Bajers Vej 7C, 9220 Aalborg East, Denmark

[8] Seem JE. *A new pattern recognition adaptive controller with application to HVAC systems.* Autom 1998.

[9] Pal AK, Mudi RK. *Self-tuning fuzzy PI controller and its applications to HVAC systems*. Int J of Comput Cogn 2008;6:25e30.

[10] Ernst, D. ; Belgian Nat. Fund for Sci. Res., Brussels ; Glavic, M. ; Capitanescu, F. ; Wehenkel, L. *Reinforcement Learning Versus Model Predictive Control: A Comparison on a Power System Problem.* In Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on (Volume:39 , Issue: 2),2009

[11] R. Sutton and A. Barto, *Reinforcement Learning, An Introduction.* Cambridge, MA: MIT Press, 1998.

[12] Laetitia Matignon, Guillaume Laurent, Nadine Le Fort - Piat. *Reward function and initial values : Better choices for accelerated Goal-directed reinforcement learning.* Lecture notes in computer science, springer, 2006, 1 (4131), pp.840-849. <hal-00331752v2>

[13] S. E. Shafiei, H. Rasmussen, and J. Stoustrup, *Modeling supermarket refrigeration systems for demand-side management*, Energies, vol. 6, no. 2, pp. 900–920, 2013.

[14] Shi, Runfu; Fu, Degang; Feng, Yinshan; Fan, Junqiang; Mijanovic, Stevo; and Radcliff, Thomas, *Dynamic Modeling of CO2 Supermarket Refrigeration System* (2010). International Refrigeration and Air Conditioning Conference. Paper 1127. http://docs.lib.purdue.edu/iracc/1127

[15] Peter Dayana and Yael Nivb, *Reinforcement learning: The Good, The Bad and The Ugly* Current Opinion in Neurobiology 2008, 18:185–196

[16] *SRSim: A simulation benchmark for supermarket refrigeration systems using matlab.* http://www.es.aau.dk/projects/ refrigeration/simulation-tools/, Feb. 2013.

[17] K. J. Astrom and T. Hagglund, *PID Controllers: Theory, Design, and Tuning* Instrument Society of America, Research Triangle Park, NC, USA, 2nd edition, 1995.

[18] S. E. Shafiei, J. Stoustrup, and H. Rasmussen, *A supervisory control approach in economic mpc design for refrigeration systems*, in European Control Conference (Accepted paper), (Zurich, Switzerland), July 2013.

[19] Carlos H. C. Ribeiro *Embedding a Priori Knowledge in Reinforcement Learning* in Journal of Intelligent and Robotic Systems, 1998

[20] Kevin R. Dixon ; Richard J. Malak ; Pradeep K. Khosla *Incorporating Prior Knowledge and Previously Learned Information into Reinforcement Learning Agents* in Institute for Complex Engineered Systems Technical Report Series, 2000 31 January, 2000

[21] Simeng Liu and Gregor P. Henze, *Investigation of reinforcement learning for building thermal mass control* University of Nebraska – Lincoln, Architectural Engineering 1110 South 67th Street, Peter Kiewit Institute, Omaha, Nebraska 68182-0681 U.S.A

[22] M. Komareji, J. Stoustrup, H. Rasmussen, N. Bidstrup, P. Svendsen, and F. Nielsen *Optimal Set-point Synthesis in HVAC Systems* Proceedings of the 2007 American Control Conference Marriott Marquis Hotel at Times Square New York City, USA, July 11-13, 2007

[23] Daniel Nikovski and Alan Esenther *Construction of Embedded Markov Decision Processes for Optimal Control of Non-Linear Systems with Continuous State Spaces* 2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC) Orlando, FL, USA, December 12-15, 2011

[24] Frank L. Lewis and Draguna Vrabie *Reinforcement Learning and Adaptive Dynamic Programming for Feedback Control* IEEE CIRCUITS AND SYSTEMS MAGAZINE, 2009

[25] Damien Ernst, Member, IEEE, Mevludin Glavic, and Louis Wehenkel, Member, IEEE *Power Systems Stability Control: Reinforcement Learning Framework* IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 19, NO. 1, FEBRUARY 2004

[26] Gregor P. Henze and Jobst Schoenmann (2003) *Evaluation of Reinforcement Learning Control for Thermal Energy Storage Systems*, HVAC&R Research, 9:3, 259-275

[27] Lars F. S. Larsen ,Roozbeh Izadi-Zamanabadi, Rafael Wisniewski, Christian Sonntag *Supermarket refrigeration systems-A benchmark for the opyimal control of hybrid systems*

[28] M. Salazar, F. Méndez *PID control for a single-stage transcritical CO_2 refrigeration cycle* Facultad de Ingeniería, UNAM, México, D.F. 04510, Mexico 2013

[29] Marc Peter Deisenroth and Carl Edward Rasmussen *Efficient Reinforcement Learning for Motor Control* Department of Engineering, University of Cambridge Trumpington Street, Cambridge CB2 1PZ, UK