

UNIVERSITÀ DEGLI STUDI DI PADOVA



Dipartimento di Ingegneria dell'Informazione  
Corso di Laurea Magistrale in Ingegneria dell'Automazione

Tesi di laurea

**TECNICHE DI APPRENDIMENTO AUTOMATICO PER LA  
RILEVAZIONE DI MALFUNZIONAMENTI IN SISTEMI HVAC:  
IL CASO DI UN REFRIGERATORE DI LIQUIDO CENTRIFUGO**

*Relatore:* **Prof. Alessandro Beghi**

*Correlatore:* **Ing. Mirco Rampazzo**

*Laureando:* **Chiara Corazzol**

A.A. 2013/2014



Il futuro appartiene a coloro che credono nella bellezza dei propri sogni.

*Eleanor Roosevelt*

## Ringraziamenti

Ringrazio anzitutto il professor Alessandro Beghi, Relatore e l'ingegner Mirco Rampazzo, Co-relatore: sempre disponibili a dirimere i miei dubbi durante la stesura di questo lavoro, senza il loro supporto e la loro guida sapiente questa tesi non esisterebbe. Vorrei ringraziare anche il dottor Gian Antonio Susto per il grandissimo sostegno ricevuto durante le mie ricerche. Ringrazio inoltre l'Università di Padova per il supporto tecnico concessomi e la disponibilità delle infrastrutture.

Un ringraziamento particolare ai miei genitori, Enrica e Alberto: per non avermi mai fatto mancare nulla e per il loro incondizionato supporto durante tutto il mio percorso. Un ringraziamento più ampio va a tutta la mia famiglia per avermi sempre sostenuto e motivato. Vorrei ringraziare soprattutto mia sorella Martina per aver condiviso con me tantissimi momenti importanti e avermi sempre ispirato con il suo comportamento e la sua dedizione allo studio.

Un grazie a tutti quelli che hanno sempre creduto in me; quelli che ci sono sempre stati e mi hanno reso una persona migliore: gli amici. Ringrazio Adriano e Andrea per aver condiviso questo percorso con me e per avermi compreso e aiutato nei momenti di difficoltà; Ilaria Ilari Lara e Caterina per avermi concesso qualche distrazione e aver reso meno impegnativa la strada. Inoltre un pensiero per Nicholas e Federico per avermi tollerato negli ultimi mesi della mia esperienza accademica. Infine un grazie di cuore a mia cugina Carlotta per aver condiviso con me questi mesi di grande impegno e avermi concesso il suo pc senza cui non avrei potuto ultimare il mio lavoro.

A tutti loro va la mia gratitudine, anche se a me spetta la responsabilità per ogni errore contenuto in questa tesi.

## Abstract

Guasti negli impianti di riscaldamento, ventilazione e condizionamento (HVAC) possono determinare una perdita di prestazioni con conseguente aumento del consumo di energia, ridurre il tempo di vita degli elementi che compongono l'impianto e creare disagio agli occupanti (discomfort). Questi malfunzionamenti devono essere tempestivamente individuati per evitare un aggravarsi delle condizioni di guasto, perché potenzialmente pericolose, ed evitare inutili sprechi di energia. Vale la pena osservare che alcuni fenomeni si presentano molto raramente oppure non possono essere conosciuti al momento dell'installazione del sistema, non è facile di conseguenza classificare tutte le condizioni di cattivo funzionamento. Per questi motivi è utilizzato un metodo di classificazione semi-supervisionato, One-Class SVM come un sistema di novelty detection per identificare stati di funzionamento sconosciuti (per lo più con accezione negativa) e possibili guasti. L'approccio proposto, che sfrutta l'analisi delle componenti principali (PCA) per evidenziare anomalie rispetto al comportamento normale, è stato testato su un dataset relativo ad un chiller centrifugo.



# Indice

<b>Indice</b>	<b>v</b>
<b>Elenco delle figure</b>	<b>vii</b>
<b>1 Introduzione</b>	<b>1</b>
<b>2 Dati Sperimentali</b>	<b>3</b>
2.1 Chiller Centrifugo . . . . .	3
2.1.1 Carico Simulato . . . . .	4
2.2 Dati utilizzati per l'analisi . . . . .	6
2.2.1 Dati provenienti dal progetto di ricerca ASHRAE RP-1043 . . . . .	7
2.2.2 Acquisizione dati e controllo . . . . .	7
2.2.3 Variabili misurate e calcolate . . . . .	10
2.2.4 La sequenza dei punti operativi considerata . . . . .	10
2.3 Tipi di guasti considerati . . . . .	12
2.4 Dati ottenuti dai test . . . . .	13
2.4.1 Dati di riferimento per il comportamento normale . . . . .	13
2.4.2 Riduzione della portata dell'acqua al condensatore . . . . .	13
2.4.3 Riduzione della portata dell'acqua all'evaporatore . . . . .	14
2.4.4 Perdita di refrigerante . . . . .	14
2.4.5 Sovraccarica di refrigerante . . . . .	15
2.4.6 Eccesso di olio . . . . .	15
2.4.7 Sporramento del condensatore . . . . .	15
2.4.8 Presenza materia non condensabile nel refrigerante . . . . .	15
2.4.9 Defective pilot-valve . . . . .	16
2.5 Stati stazionari e filtraggio . . . . .	16
2.5.1 Individuazione degli stati stazionari . . . . .	17
2.5.2 Presentazione dei risultati nel caso di funzionamento normale . . . . .	19
2.5.3 Filtraggio dei dati . . . . .	20
2.6 Selezione delle variabili . . . . .	23

---

<b>3</b>	<b>Novelty Detection</b>	<b>25</b>
3.1	Tipi di Apprendimento . . . . .	27
3.2	Tipi di anomalie . . . . .	30
3.2.1	Anomalie Puntuali . . . . .	30
3.2.2	Anomalie Contestuali . . . . .	30
3.2.3	Anomalie Collettive . . . . .	31
3.3	Anomaly Detection . . . . .	32
3.3.1	Basate sulla classificazione . . . . .	32
3.3.2	Basate sulla vicinanza . . . . .	33
3.3.3	Clustering . . . . .	33
3.3.4	Approcci statistici . . . . .	34
<b>4</b>	<b>One-Class SVM</b>	<b>37</b>
4.1	Formulazione . . . . .	38
4.2	Tightness Detection . . . . .	40
4.3	Analisi delle prestazioni . . . . .	43
4.4	Esempio . . . . .	44
4.5	Analisi delle componenti principali . . . . .	49
4.6	Andamenti dell'errore . . . . .	50
<b>5</b>	<b>Simulazioni e Risultati</b>	<b>53</b>
5.1	Dati Normali . . . . .	53
5.2	Dati relativi ad Anomalie . . . . .	58
5.2.1	Riduzione della portata di acqua all'evaporatore . . . . .	58
5.2.2	Riduzione della portata di acqua al condensatore . . . . .	65
5.2.3	Perdita di refrigerante e sovraccarica di refrigerante . . . . .	67
5.2.4	Eccesso di olio . . . . .	68
5.2.5	Sporcamento del condensatore e materia non condensabile nel refrigerante . . . . .	73
5.2.6	Confronto dei risultati ottenuti . . . . .	74
<b>6</b>	<b>Conclusioni e Sviluppi Futuri</b>	<b>79</b>
<b>A</b>	<b>Dati</b>	<b>81</b>
A.1	Dati Sperimentali dal progetto di ricerca 1043-RP . . . . .	81
A.2	Conversione delle misure . . . . .	81
<b>B</b>	<b>Classificatori</b>	<b>87</b>
B.1	Support Vector Machines . . . . .	87
<b>C</b>	<b>LIBSVM</b>	<b>91</b>
C.1	Utilizzo del pacchetto LIBSVM . . . . .	91
	<b>Bibliografia</b>	<b>93</b>

# Elenco delle figure

2.1	Schema di principio di un chiller a compressione di vapore: evaporatore, compressore, condensatore, valvola di espansione. . . . .	4
2.2	Schema del sistema di raffreddamento. . . . .	5
2.3	Schema che mostra il chiller di test e l'interfaccia di controllo. . . . .	8
2.4	Sensori situati nel circuito dell'acqua dell'evaporatore e fornitura del vapore. . . . .	9
2.5	Sensori situati nel circuito dell'acqua del condensatore e fornitura dell'acqua proveniente dall'acquedotto. . . . .	9
2.6	Regressione lineare su un intervallo di TEO. A sinistra è riportato un particolare della figura a destra. . . . .	20
2.7	TEO attraversa i primi 9 setpoint, sono riportati anche gli intervalli considerati steady state. . . . .	21
2.8	PRC attraversa i primi 9 setpoint, sono riportati anche gli intervalli considerati steady state. . . . .	21
2.9	TEO, TCI, PRE, PRC in steady state. . . . .	22
2.10	Valori di TEO e PRC trovati tramite l'algoritmo (puntini blu) e quelli forniti dal report (crocette rosse). . . . .	22
2.11	Verifica del comportamento di steady state su otto variabili significative. . . . .	23
3.1	Esempio di anomalie in un data set a due dimensioni ([Chandola et al., 2009]).	27
3.2	Componenti chiave associati alla tecnica di individuazione delle anomalie ([Chandola et al., 2009]). . . . .	29
3.3	Anomalia contestuale $t_2$ in una serie di dati temporali. Si noti che la temperatura $t_1$ è uguale a $t_2$ ma non è considerata un'anomalia in quanto si presenta in un diverso contesto ([Chandola et al., 2009]). . . . .	31
4.1	Campioni estratti da una distribuzione di Rosenbrock (in blu) e da una distribuzione normale (in rosso). . . . .	45
4.2	Valori dell'accuratezza ottenuta sui dati di training in funzione dei parametri del classificatore $\gamma$ e $\nu$ . . . . .	46
4.3	Rappresentazione della regione normale e della distanza dal bordo del classificatore, per i valori del parametro $\gamma$ trovati dall'algoritmo durante le varie iterazioni . . . . .	47

4.4	Convergenza dell'algoritmo di tightness detection. . . . .	48
4.5	(a) Dati di training e regione normale individuata. (b) Risultato della classificazione. . . . .	48
4.6	(a) Curva ROC. (b) Matrice di confusione. . . . .	49
4.7	Andamento dell'errore in funzione delle PC scartate partendo da quelle associate ad autovalori maggiori. . . . .	51
4.8	Andamento dell'errore in funzione delle PC scartate partendo da quelle associate ad autovalori minori. . . . .	52
5.1	(a) Convergenza dell'algoritmo di tightness detection. (b) Dati proiettati in $\mathbf{p}_1$ e $\mathbf{p}_2$ e proiezione delle feature originali. . . . .	55
5.2	Convergenza dell'algoritmo OCSVM (dati in 8 dimensioni). . . . .	56
5.3	(a) Regioni individuate per tre diversi valori di $\gamma$ . (b) Valori del parametro $\gamma$ in funzione del numero di iterazioni dell'algoritmo di tightness detection. . . . .	57
5.4	Le 15 feature originali plottate in funzione del carico, fwe (in rosso) e normal (in blu). . . . .	59
5.5	Le 15 feature ottenute dalla trasformazione PCA, fwe (in rosso) e normal (in blu). . . . .	60
5.6	Curve ROC relative alla riduzione della portata di acqua all'evaporatore a differenti livelli di severità. . . . .	62
5.7	(a) Media ogni 100 campioni della distanza dei punti dalla regione. (b) Proiezione in due dimensioni dei dati normal e fwe colorati secondo la distanza dalla regione a 8 dimensioni. . . . .	63
5.8	Boxplot associati alla distanza con segno corrispondenti al guasto fwe per diversi livelli di severità. . . . .	63
5.9	(a) Regione normale. (b) Risultato della classificazione. . . . .	64
5.10	Curve ROC relative alla riduzione della portata di acqua al condensatore a differenti livelli di severità. . . . .	66
5.11	(a) Proiezione dati normal, fwe e livello di severità del guasto. (b) Distanza con segno relativa ai punti corrispondenti al guasto fwc per differenti livelli di severità . . . . .	66
5.12	Boxplot associati alla distanza con segno corrispondenti al guasto fwc per diversi livelli di severità. . . . .	67
5.13	(a) Curve ROC relative alla perdita del refrigerante a differenti livelli di severità. (b) Curve ROC relative all'eccesso del refrigerante a differenti livelli di severità. . . . .	69
5.14	(a) Distanza con segno relativa ai punti corrispondenti al guasto rl per differenti livelli di severità (proiettati lungo $\mathbf{p}_{12}$ e $\mathbf{p}_{13}$ ). (b) Distanza con segno relativa ai punti corrispondenti al guasto ro per differenti livelli di severità (proiettati lungo $\mathbf{p}_{14}$ e $\mathbf{p}_{15}$ ). . . . .	70
5.15	Boxplot associati alla distanza con segno corrispondenti ai guasti rl e ro per diversi livelli di severità. . . . .	70

*ELENCO DELLE FIGURE*

---

5.16	Curve ROC relative all'eccesso del olio a differenti livelli di severità. . . . .	71
5.17	Distanza con segno relativa ai punti corrispondenti al guasto eo per differenti livelli di severità (proiettati lungo $p_{12}$ e $p_{13}$ ). . . . .	72
5.18	Boxplot associati alla distanza con segno corrispondenti al guasto eo per diversi livelli di severità. . . . .	72
5.19	Curve ROC relative al guasto cf. . . . .	74
5.20	Curve ROC relative al guasto nc. . . . .	75
5.21	Distanza con segno relativa ai punti corrispondenti ai guasti cf e nc per differenti livelli di severità. . . . .	76
5.22	Boxplot associati alla distanza con segno corrispondenti ai guasti cf e nc per differenti livelli di severità. . . . .	76
5.23	Distanza con segno relativa ai punti corrispondenti ai guasti fwe, fwc e rl, ro per differenti livelli di severità. . . . .	77



# Capitolo 1

## Introduzione

Il tema dell'efficienza e della sostenibilità ambientale applicata agli edifici, rappresenta uno degli aspetti prioritari delle attuali politiche energetiche. Basti ricordare il piano 20-20-20 della UE che prevede di ridurre le emissioni di gas serra del 20%, alzare al 20% le quote di energia prodotta da fonti rinnovabili e portare al 20% il risparmio energetico entro il 2020. Problemi associati al riscaldamento, ventilazione e raffreddamento (HVAC) riguardano molti edifici del settore terziario quali ospedali, scuole, alberghi, uffici e spesso sono il risultato del deterioramento degli elementi, sensori difettosi, installazioni sbagliate, scarsa manutenzione e controlli inadeguati. Il malfunzionamento di questi impianti può portare a situazioni disagiati per gli occupanti, consumo di energia, scarsa efficienza o breve durata dei componenti del sistema.

Quando si verificano malfunzionamenti nei sistemi HVAC, le conseguenze possono essere molto dannose e costose da essere trattate.

Il sistema-edificio contiene tecnologie sempre più sofisticate (*BMS, Building management systems*) per la gestione intelligente dell'edificio. In questo contesto i sistemi di *Fault Detection* (FD) ricoprono un ruolo fondamentale nel monitoraggio di impianti HVAC complessi e nella rilevazione di comportamenti anomali.

Nonostante l'utilizzo in altri campi (i.e. automotive, aerospaziale, manifatturiero ecc), lo sviluppo di sistemi FD nel mondo HVAC rimane tuttavia una sfida a causa delle difficoltà nel reperire dati corrispondenti a malfunzionamenti/anomalie (*labelled data*). I dati etichettati contengono informazioni qualitative relative alle condizioni di funzionamento del sistema (normali o anomale per esempio) e queste sono spesso:

- costose perché l'etichettatura è fatta manualmente da un operatore esperto;
- difficili da ottenere in quanto i dati anomali non sono disponibili preventivamente nella maggior parte degli impianti HVAC moderni, dove nessuno, o più in generale, poche delle potenziali anomalie si sono verificate in passato.

Per queste ragioni in questo campo è utile affrontare problemi di *novelty detection*, che mirano a monitorare il comportamento del sistema e a identificare la presenza di deviazioni dal comportamento nominale [Chandola et al., 2009].

In questa tesi è proposto un metodo per l'individuazione di anomalie al fine di rilevare in anticipo potenziali situazioni di malfunzionamento. In particolare si considera un refrigeratore di liquido (chiller) centrifugo (il cui nome deriva dal tipo di compressore utilizzato), comune in strutture con medi e grandi carichi [Stanford, 2012]: questo tipo di macchine refrigeranti sono unità di volume variabile che utilizzano giranti per comprimere il vapore del refrigerante e la capacità di raffreddare è regolata attraverso l'utilizzo di un vano di ingresso che riduce l'afflusso di refrigerante alla girante. Vale la pena sottolineare che i guasti ai componenti di un chiller incidono per il 42% sulle richieste di servizio e per il 26% dei costi di riparazione [Mathew C. Comstock, 1999]: per questi motivi il mantenimento dei sistemi HVAC in condizioni ottimali e l'individuazione tempestiva di anomalie sono vantaggiosi sia per il risparmio energetico che per assicurare il comfort agli occupanti. In questo contesto si considerano le anomalie come guasti, sconosciuti a priori; i tipi di guasti che possono insorgere sono ad esempio la riduzione della portata di acqua al condensatore e all'evaporatore, perdita o sovraccarica di refrigerante, perdita di olio, sporcamento del condensatore, ecc.

Il problema principale che si riscontra nel procedimento di rilevazione delle anomalie in sistemi HVAC è che la maggior parte della variabilità dei dati è dovuta al funzionamento normale del sistema; questo sfortunatamente maschera il comportamento del sistema e rende complicata l'individuazione di situazioni anomale (i.e. la variabilità è dominata dal cambiamento tra i diversi punti operativi). Per affrontare questo problema il sistema di rilevamento delle anomalie proposto si basa su un classificatore One-Class Support Vector Machine, e si avvale della tecnica Principal Component Analysis (PCA) per aiutare a mostrare la separazione tra il comportamento normale e la variabilità associata alle anomalie.

Per testare e convalidare la tecnica di *anomaly detection* proposta, sono stati utilizzati dataset provenienti dal progetto ASHRAE 1043-RP [Mathew C. Comstock, 1999]; il progetto svolge uno studio sperimentale al fine di generare dati relativi a guasti e funzionamento normale per un chiller centrifugo al fine di sviluppare e validare metodi di *fault detection and diagnosis*. Come risultato finale degli esperimenti si forniscono informazioni riguardanti la severità delle anomalie riscontrate tramite l'utilizzo della distanza dei nuovi dati dal limite di decisione fornito dal classificatore OCSVM.

## Capitolo 2

# Dati Sperimentali

In questa tesi, per svolgere le analisi dei metodi proposti nell'ambito della novelty detection si utilizzano dei dati sperimentali ottenuti dal progetto ASHRAE 1043-RP, che descrivono il comportamento di un chiller centrifugo; questi dati sono stati raccolti al fine di studiare il comportamento di un chiller centrifugo e l'effetto dell'introduzione di guasti sul sistema con lo scopo di sviluppare e analizzare tecniche di rilevazione e diagnosi di guasti.

In questo capitolo si descrive il sistema utilizzato per ottenere i dati sperimentali, le componenti del chiller centrifugo e l'impianto utilizzato per simulare il carico. Inoltre è fornita la descrizione dei dataset presenti nel report 1043-RP al quale si fa riferimento e sono descritte le variabili e i dati provenienti dal sistema. I dati utilizzati si riferiscono al comportamento del sistema attraverso vari punti operativi.

Si presenta un algoritmo che individua gli stati stazionari del sistema utilizzati per le analisi. I dati ottenuti sono in seguito filtrati prima di essere utilizzati per le simulazioni.

### 2.1 Chiller Centrifugo

Un chiller (o refrigeratore) è una macchina che rimuove calore da un liquido attraverso un ciclo di compressione del vapore o un ciclo frigorifero ad assorbimento. Il liquido raffreddato passa attraverso uno scambiatore di calore e raffredda un liquido secondario. La refrigerazione crea un eccesso di calore all'interno del circuito primario che viene smaltito verso l'ambiente esterno. Aspetti da considerare durante la progettazione o selezione di un refrigeratore riguardano: il dimensionamento, le prestazioni, l'efficienza, la manutenzione, la vita del prodotto e l'impatto ambientale de esso procurato. Il sistema considerato (Figura 2.1) si compone in quattro elementi principali, un evaporatore a fascio tubiero e mantello, un condensatore dello stesso tipo una valvola di espansione controllata da una valvola pilota e un compressore centrifugo:

- scambiatori di calore (heat exchanger): sia l'evaporatore che il condensatore sono scambiatori di calore a fascio tubiero e mantello, l'acqua scorre nei tubi e il refrigerante scorre all'esterno,

- compressore (compressor): si utilizza un compressore centrifugo azionato da un motore elettrico,
- valvola di espansione (expansion valve): valvola principale nella linea del liquido e una pilot valve parallela a questa.

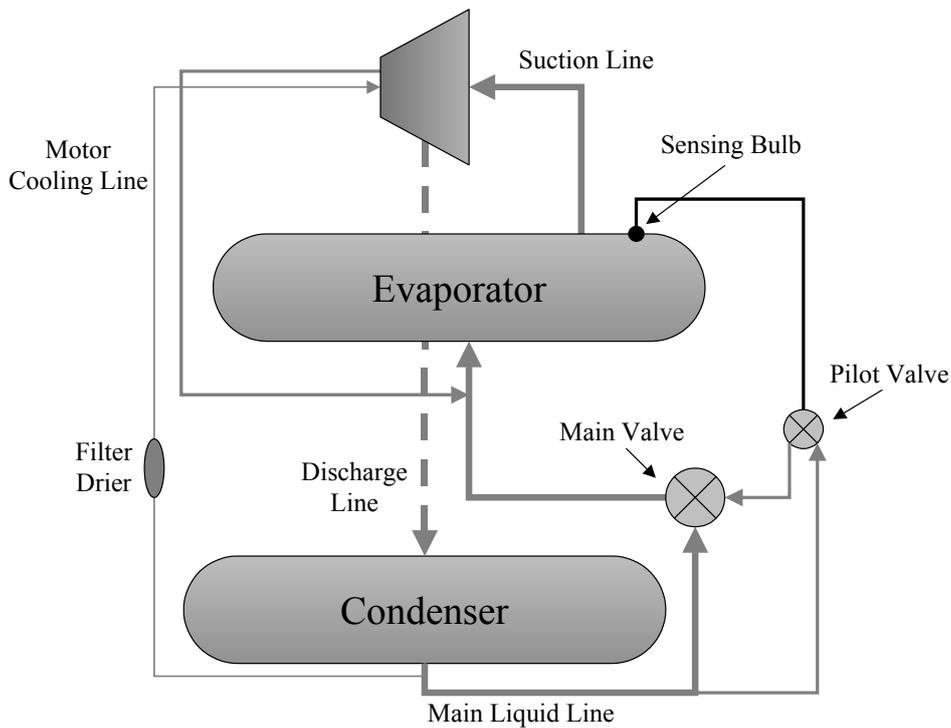


Figura 2.1: Schema di principio di un chiller a compressione di vapore: evaporatore, compressore, condensatore, valvola di espansione.

Il refrigerante utilizzato è l'R134a e il fluido secondario è l'acqua. In Figura 2.1 è mostrato lo schema del sistema e il percorso del refrigerante.

All'uscita dell'evaporatore il fluido frigorifero si trova allo stato gassoso, il compressore tramite la girante comprime il gas e determina un aumento della pressione del vapore e di conseguenza un aumento della sua temperatura. Il gas in uscita dal compressore passa attraverso il condensatore dove cede calore e ritorna allo stato liquido [Stanford, 2012].

### 2.1.1 Carico Simulato

In questa sezione si descrive il sistema di test impiegato come carico per il chiller il cui compito è quello di simulare il comportamento che avrebbe tipicamente un edificio. Il sistema di test è caratterizzato dai seguenti fattori [Mathew C. Comstock, 1999]:

- il circuito dell'acqua del condensatore e dell'evaporatore devono essere chiusi,

## 2.1. CHILLER CENTRIFUGO

- il sistema di riscaldamento a vapore utilizzato come carico per il chiller non deve essere posto a rischio diretto di contaminazione.

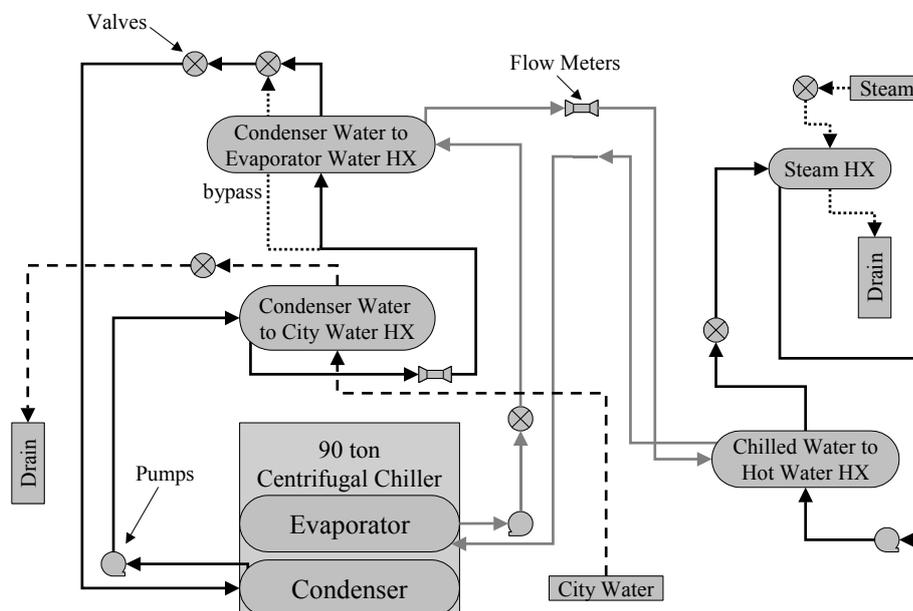


Figura 2.2: Schema del sistema di raffreddamento.

La Figura 2.2 illustra gli elementi più importanti che compongono la struttura dell'intero impianto di prova. Il chiller di prova utilizza i seguenti elementi:

- tre scambiatori di calore acqua-acqua,
- uno scambiatore di calore acqua-vapore,
- una valvola a tre vie ad attuazione elettrica,
- sei valvole a due vie ad attuazione elettrica,
- tre pompe per l'acqua,
- due misuratori di portata vortex.

In Figura 2.2 sono riportati i versi di entrata e uscita per ogni scambiatore di calore e ogni valvola. Non è comunque necessario conoscere l'esatta descrizione fisica dell'impianto per avere un'idea della filosofia del design che lo caratterizza.

Il sistema è caratterizzato da quattro percorsi di flusso:

- circuito dell'acqua dell'evaporatore,
- circuito dell'acqua del condensatore,

- circuito dell'acqua calda,
- apporto dell'acquedotto,
- apporto di vapore.

I tre circuiti interni sono isolati e scambiano solamente energia termica tramite i rispettivi scambiatori senza che avvengano mescolamenti tra i liquidi. I due circuiti dell'evaporatore e del condensatore scambiano energia tra loro. Questa interazione tra i due circuiti rappresenta il 70-80 % del carico visto dal chiller. Il circuito del condensatore cede calore all'acquedotto (city water) raffreddandosi. Il circuito dell'evaporatore invece scambia calore con il circuito dell'acqua calda che si trova sempre ad una temperatura maggiore in quanto mantenuta alta dallo scambio termico con il vapore. Il ciclo dell'acqua calda è utilizzato solo per evitare un contatto diretto tra il vapore e il circuito dell'evaporatore.

L'ordine con cui avvengono gli scambi di calore tra i circuiti è scelto per assicurare la massima differenza di temperatura ottenibile. Per quanto concerne il ciclo dell'evaporatore il primo scambio è quello effettuato con il condensatore in modo che la bassa temperatura dell'acqua in uscita dal condensatore favorisca un maggior trasferimento di calore, inoltre il secondo scambio che avviene lungo questo circuito garantisce sempre una differenza di temperatura tale da assicurare un trasferimento di calore dal circuito dell'acqua calda a quello dell'evaporatore.

Nel caso del circuito del condensatore è stato necessario posizionare lo scambiatore con l'acquedotto prima di quello con l'evaporatore in quanto la temperatura dell'acqua proveniente dall'evaporatore è sempre inferiore a quella dell'acquedotto. A contatto con l'acqua dell'acquedotto il liquido nel circuito del condensatore perde qualche grado ma rimane sempre ad una temperatura superiore all'acqua in uscita dall'evaporatore.

Una valvola a tre mandate posta lungo il circuito del condensatore permette di oltrepassare lo scambiatore comune tra i due circuiti principali (condensatore-evaporatore) in modo da poter ridurre il carico sul chiller all'occorrenza. I cambiamenti che riguardano questo elemento influiscono sia sulla temperatura dell'acqua nel circuito del condensatore sia quella dell'acqua nel circuito dell'evaporatore.

Lo scopo dell'acquedotto è duplice. Da una parte serve per raffreddare il motore del compressore e delle pompe dell'acqua, dall'altra per controllare la temperatura dell'acqua del condensatore senza però intaccare la stabilità della temperatura dell'acqua all'evaporatore. Similmente, l'impiego del vapore permette di variare il carico del chiller senza indurre cambiamenti nella temperatura dell'acqua nel circuito del condensatore [Mathew C. Comstock, 1999].

## 2.2 Dati utilizzati per l'analisi

In letteratura sono illustrate numerose metodologie per la rilevazione e la diagnosi (FDD methods) di guasti con riferimento a diverse applicazioni (automotive, aerospaziali ecc.), tuttavia, pochi studi affrontano applicazioni FDD in campo HVAC.

Oggi giorno la maggior parte dei produttori di chiller installano una varietà di sensori sulla macchina per attingere informazioni. Le informazioni vengono poi passate ad un microprocessore che effettua il controllo e l'individuazione di semplici irregolarità come la presenza di valori di pressioni e temperature anomale. Con il progressivo aumento della strumentazione nei chiller, l'interesse per i metodi di individuazione dei guasti è in rapido aumento.

Inoltre molti chiller dispongono di modem o schede di rete che consentono il collegamento ad internet rendendo le misure disponibili ovunque nel mondo. La prospettiva futura è quella di riuscire in pochi anni a pianificare la manutenzione di questi impianti basandosi solamente sui dati disponibili online. La manutenzione preventiva ricopre un aspetto fondamentale che giustifica l'impiego di metodi FDD nei chiller [Mathew C. Comstock, 1999].

### 2.2.1 Dati provenienti dal progetto di ricerca ASHRAE RP-1043

I dati utilizzati in questa tesi provengono da un progetto sponsorizzato da ASHRAE per lo studio dei guasti nei chiller. Lo studio sperimentale è stato condotto su un chiller centrifugo di 90 ton<sup>1</sup>. Lo scopo del report è quello di fornire un database con il quale testare l'accuratezza e la sensibilità dei metodi di rilevazione e diagnosi dei guasti. I vari test effettuati hanno prodotto un ricco database sul quale si può basare lo sviluppo e l'analisi di tecniche per l'individuazione di guasti.

Il chiller è stato testato su 27 differenti punti operativi (setpoint) sia in condizioni normali, cioè quando il sistema funziona correttamente secondo gli standard attesi, che di guasto, cioè in situazioni in cui qualche componente del sistema compromette il corretto funzionamento della macchina.

Si considerano solamente chiller al di sopra delle 80 ton e non include sistemi come le unità soprattetto. I sistemi di larghe e medie dimensioni sono muniti di elaborati meccanismi di controllo che garantiscono una adeguata sicurezza nei confronti di eventi critici o catastrofici. In questo studio non si considerano guasti di grossa entità (come ad esempio la rottura della cinghia della ventola o il motore bruciato) ma ci si focalizza su guasti incipienti che gradualmente portano al consumo maggiore di energia e al deterioramento dei pezzi.

I dati sperimentali sono utilizzati per la derivazione di modelli per la simulazione che vengono impiegati per l'individuazione dei guasti. Il lavoro sperimentale è caratterizzato da:

- considerazione di diverse tipologie di guasto,
- considerazione di diversi livelli di gravità del guasto,
- considerazione di guasti per diversi punti operativi,
- disponibilità dei dati provenienti dal sistema grazie ai sensori utilizzati.

### 2.2.2 Acquisizione dati e controllo

Nel pacchetto di installazione del chiller è compreso un controllore MicroTech che utilizza una serie di dati provenienti dai sensori. I dati sono passati al PC dal controllore MicroTech

---

<sup>1</sup>la conversione al sistema internazionale SI è riportata in Appendice A.2

attraverso una connessione RS-232. Il banco di prova è controllato da un gruppo di tre controllori Johnson Controle Inc. Air Handling Unit (JCI AHU) con un N2 bus (RS-485 network) a sua volta connesso al PC tramite un convertitore RS-485 a RS-232, come illustrato in Figura 2.3.

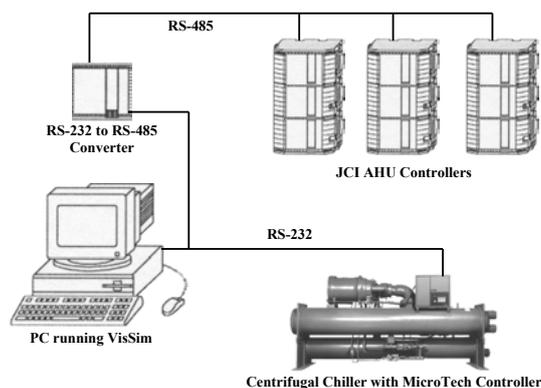


Figura 2.3: Schema che mostra il chiller di test e l'interfaccia di controllo.

Entrambi i controllori sono interfacciati con il PC tramite VisSim<sup>1</sup>, un software di simulazione con caratteristiche personalizzabili che permette di utilizzare vari protocolli di comunicazione abilitati simultaneamente. Il programma sviluppato in ambiente VisSim colleziona tutti i dati provenienti dai controllori e li esporta in un file di testo. Inoltre il programma importa i dati necessari per il controllo automatico della sequenza di test e dei punti di lavoro. VisSim campiona i dati e aziona i controlli ad intervalli di 10 secondi.

Il controllore MicroTech è in grado di controllare la temperatura dell'acqua in uscita dall'evaporatore e altre variabili interne del chiller. I controllori JCI AHU sono utilizzati per controllare le pompe dell'acqua e le posizioni delle valvole. Le posizioni influenzano direttamente la temperatura dell'acqua all'interno del sistema compresa la temperatura dell'acqua in uscita dall'evaporatore.

Il controllore MicroTech montato sul chiller è in grado di fornire tutte le misure necessarie eccetto la potenza istantanea in ingresso al compressore e le portate dell'acqua. L'impianto di test è fornito di sensori addizionali che aiutano nella misurazione della intensità dei guasti prodotti e aiutano la risoluzione dei problemi. Questi sensori sono interfacciati con i controllori JCI AHU. Alcuni sensori supplementari misurano la temperatura dell'acqua all'entrata e all'uscita del condensatore e dell'evaporatore. Tutte le informazioni dei sensori sono compilate in VisSim utilizzando protocolli di comunicazione speciali per comunicare attraverso la rete composta dai due tipi di controllori.

<sup>1</sup><http://www.vissim.com>

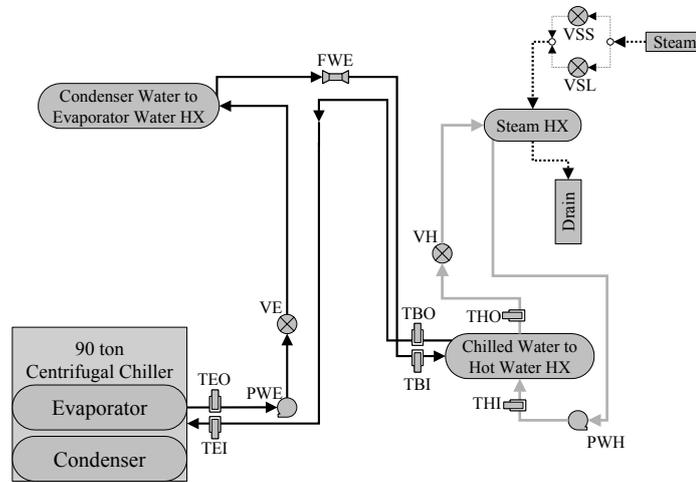


Figura 2.4: Sensori situati nel circuito dell'acqua dell'evaporatore e fornitura del vapore.

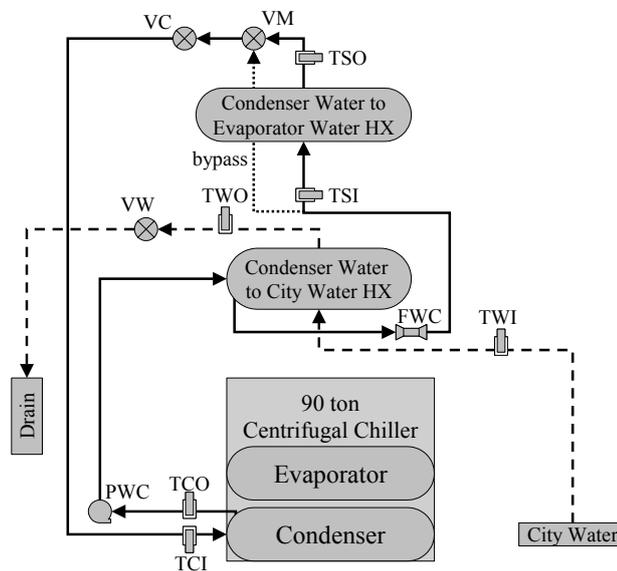


Figura 2.5: Sensori situati nel circuito dell'acqua del condensatore e fornitura dell'acqua proveniente dall'acquedotto.

### 2.2.3 Variabili misurate e calcolate

Per azionare il banco di prova e raccogliere i dati è utilizzato il programma VisSim che permette anche di eseguire calcoli in tempo reale, i quali sono esportati insieme ai dati di misura provenienti dai sensori. Tutte le informazioni ottenute dal sistema sono elencate in Tabella A.3. Anche i controllori MicroTech internamente eseguono dei calcoli utilizzando le misure a disposizione (i.e. temperatura all'evaporatore e al condensatore, superheat, subcooling, approach temperature ecc.) Tra i calcoli svolti nell'ambiente VisSim ci sono anche:

$$\text{EvapTons} = \frac{FWE \times (TEI - TEO)}{24}, \quad (2.1)$$

$$\text{CondTons} = \frac{FWC \times (TCO - TCI)}{24}, \quad (2.2)$$

$$\text{COP} = \frac{\text{EvapTons} \times 12000}{\text{kW} \times 3413}, \quad (2.3)$$

$$\text{kW/Ton} = \frac{\text{kW}}{\text{EvapTons}}. \quad (2.4)$$

### 2.2.4 La sequenza dei punti operativi considerata

Per determinare il campo operativo del chiller varie prove sono state svolte portando la macchina a temperature e carichi estremali. Le prove sono state eseguite seguendo le linee guida indicate dallo standard ARI 550<sup>1</sup>. Per testare il chiller in condizioni standard nominali si porta la temperatura dell'acqua del chiller (TEO) a 44°F mentre la temperatura dell'acqua in entrata al condensatore (TCI) a 85°F. Per testare il chiller in condizioni operative TEO deve variare nel range di valori tra 40°F e 48°F e TCI tra 60°F e 105°F.

A causa dei limiti di raffreddamento dell'acqua in entrata al condensatore TCI non può mai raggiungere una temperatura di 60°F. Inoltre il chiller non può ricevere in ingresso al condensatore liquido con temperatura superiore ai 90°F. Se infatti la temperatura dell'acqua in ingresso al condensatore è troppo elevata c'è il rischio che la differenza di pressione tra evaporatore e condensatore superi i limiti critici. Se il compressore non è in grado di opporsi alla resistenza del sistema in questi casi si rischia di incorrere in reversioni di flusso o altri eventi catastrofici per l'impianto. La temperatura limite considerata viene quindi abbassata rispetto allo standard e portata a 85°F in modo da essere robusta anche rispetto agli imprevedibili scostamenti provocati durante le simulazioni dei guasti.

Per eseguire i test viene utilizzata sempre la stessa sequenza di punti operativi. In questa maniera risulta più facile il confronto tra i diversi casi considerati. Le tre variabili di controllo considerate sono:

- la temperatura dell'acqua in uscita dall'evaporatore (TEO),
- la temperatura dell'acqua in ingresso al condensatore (TCI),

<sup>1</sup><http://www.ari.org>

Tabella 2.1: Successione dei 27 punti operativi.

TEO	TCI	Capacità %
50	85	90-100
50	85	50-60
50	85	25-40
50	75	90-100
50	75	50-60
50	70	70-80
50	65	45-50
50	62	25-35
45	85	90-100
45	85	50-60
45	85	25-40
45	75	90-100
45	75	50-60
45	75	25-40
45	70	70-90
45	62	50-60
45	62	25-40
40	80	90-100
40	80	50-60
40	80	25-40
40	70	90-100
40	70	50-60
40	70	25-40
40	65	70-90
40	62	50-60
40	62	25-40

- il carico di raffreddamento del chiller (EvapTons).

Per ogni variabile vengono considerati tre possibili valori ottenendo  $3 \times 3 \times 3$  possibili combinazioni. Tutte le 27 combinazioni sono riportate in Tabella 2.10. Nell'ultima colonna sono utilizzati degli intervalli in quanto il carico da raffreddare non può essere direttamente controllato e il valore effettivo può variare negli intervalli riportati.

Ad ogni test eseguito il sistema deve passare attraverso i punti operativi (setpoint) elencati in Tabella 2.1, ogni punto operativo è impostato a distanza di 30 minuti per permettere al sistema di raggiungere lo steady-state; ogni test ha una durata di circa 15 ore. Ogni volta che viene cambiato il valore di una variabile il sistema impiega dai 5 ai 10 minuti per tornare in una condizione di stabilità. Si considera che il transitorio si esaurisca in 5-15 minuti e il sistema raggiunga i punti operativi e rimanga stabile dai 15 ai 25 minuti.

La successione dei punti operativi elencati in 2.1 è scelta in modo da far variare meno frequentemente quelle variabili che causano una risposta più lenta sul sistema e quindi inducono un tempo di assestamento maggiore. In questo caso la risposta più lenta è dovuta al cambiamento di temperatura all'uscita dall'evaporatore, per questo si impone che la sua variazione abbia frequenza minore rispetto alle altre.

Durante ogni test le portate dell'acqua sono state impostate secondo lo standard ARI; la

temperatura dell'acqua dall'evaporatore (TEO) inizialmente è posta a 50°F, in seguito ridotta a 45°F e 40°F. Per ogni temperatura di uscita dall'evaporatore la temperatura di ingresso al condensatore viene variata dal maggiore valore ottenibile a quello minore. La variabile che cambia con frequenza maggiore è il carico all'evaporatore che va dal più alto al più basso ottenibili.

### 2.3 Tipi di guasti considerati

Per individuare i guasti più frequenti e onerosi in termini economici è stato fatto un sondaggio tra le società americane produttrici di chiller più importanti [Mathew C. Comstock, 1999]. Durante questo studio sono stati individuati un grande numero di guasti differenti. Non tutti i guasti però si prestano all'analisi tramite metodi di analisi dei guasti (FDD). Per esempio la maggior parte dei guasti elettrici o riguardanti il motore (come il motore bruciato) non richiedono metodi sofisticati di individuazione [Mathew C. Comstock, 1999].

Molti guasti possono essere facilmente individuati con tecniche semplici, anche se a volte costose. Tuttavia i guasti collegati alla degradazione dell'equipaggiamento, che portano a perdite di prestazioni, non sono facili da individuare in quanto il chiller mantiene il proprio funzionamento. Per individuare questi guasti normalmente non basta un singolo sensore e quindi è richiesta un'analisi più complessa. Conseguentemente, i guasti scelti per l'analisi sono:

- Riduzione della portata dell'acqua al condensatore
- Riduzione della portata dell'acqua all'evaporatore
- Perdita del liquido refrigerante
- Sovraccarica di liquido refrigerante
- Eccesso di olio
- Sporciamento del condensatore
- Presenza di materia non condensabile nel refrigerante
- Rottura della valvola pilota
- Guasti multipli

Alcuni dei guasti considerati non sono molto frequenti ma l'insieme di tutti i guasti costituisce il 42% delle richieste di assistenza e il 26% dei costi di riparazione [Mathew C. Comstock, 1999]. Ai guasti ottenuti dal sondaggio ne è stato aggiunto uno, il guasto di sovraccarica del refrigerante in quanto è teoricamente rilevabile a seguito di un'operazione impropria durante la manutenzione.

Un guasto non considerato invece è un bloccaggio nella linea del liquido, normalmente associato al filtro di disidratazione in quanto non situato nella linea principale del chiller

Tabella 2.2: Benchmark e guasti.

Guasto considerato	Test	Riferimento
Diminuzione della portata dell'acqua al condensatore	fwe	normal
Diminuzione della portata dell'acqua all'evaporatore	fwc	normal
Perdita di refrigerante	rl	normal r
Sovraccarica di refrigerante	ro	normal r
Eccesso di olio	eo	normal eo
Sporcamento del condensatore	cf	normal cf
Materia non condensabile	nc	normal nc
Valvola difettosa	dpv	normal dpv

di test. Questo guasto è stato dunque trascurato contando che ha una scarsa frequenza e incidenza economica.

Ci si aspetta che questi guasti vengano individuati tramite il monitoraggio degli stati termodinamici. Inoltre è possibile che molti dei guasti rilevabili termodinamicamente possano servire da spie per potenziali danneggiamenti che possono essere la conseguenza di guasti non opportunamente trattati [Mathew C. Comstock, 1999].

## 2.4 Dati ottenuti dai test

### 2.4.1 Dati di riferimento per il comportamento normale

Basandosi sulle informazioni presenti in Tabella A.3 si può decidere quale test sia il migliore per il confronto con gli esperimenti di guasto (benchmark). Se avviene un cambiamento non invertibile sul sistema allora i dati raccolti dopo quel cambiamento non possono essere confrontati con quelli raccolti prima. Cambiamenti introdotti dalla posizione di una valvola o dalla rimozione di tubi (ad esempio dal condensatore) non influenzano il sistema, quindi i test fatti prima e dopo possono essere confrontati. Invece cambiamenti riguardanti la quantità di olio, refrigerante, la valvola di espansione e simili, non permettono il ritorno alle condizioni originali del sistema dopo la rimozione del guasto. Si riporta una tabella con l'indicazione dei test utilizzati come riferimento in relazione al guasto considerato (Tabella 2.2).

### 2.4.2 Riduzione della portata dell'acqua al condensatore

Il flusso dell'acqua è regolato da una valvola elettronica. La portata nominale è di 216 gpm e per ogni livello di guasto viene ridotto del 10%. L'accuratezza con cui è rilevato il livello dipende dall'accuratezza della misurazione.

Le riduzioni imposte alla portata non sono state raggiunte in pieno. I livelli effettivi di riduzione ottenuti sono 19% per il secondo livello 28% per il terzo e 36% per il quarto. I test considerati per questo guasto sono riportati in Tabella 2.3.

Tabella 2.3: Livelli di guasto per la riduzione della portata dell'acqua al condensatore.

Caso	Condizione desiderata	Intervallo portata	Test
Condizioni normali	270 gpm	264-270 gpm	normal
Livello guasto 1	Riduzione 10% (243 gpm)	234-250 gpm	fwc10
Livello guasto 2	Riduzione 20% (216 gpm)	209-219 gpm	fwc20
Livello guasto 3	Riduzione 30% (189 gpm)	187-190 gpm	fwc30
Livello guasto 4	Riduzione 40% (162 gpm)	159-166 gpm	fwc40

### 2.4.3 Riduzione della portata dell'acqua all'evaporatore

Il flusso dell'acqua è regolato da una valvola elettronica. La portata nominale è di 270 gpm e per ogni livello di guasto viene ridotto del 10%. L'accuratezza con cui è rilevato il livello dipende dall'accuratezza della misurazione, che ha un'incertezza di  $\pm 2.8$  gpm. I livelli di guasto con i relativi test effettuati sono riportati in Tabella 2.4.

Tabella 2.4: Livelli di guasto per la riduzione della portata dell'acqua all'evaporatore.

Caso	Condizione desiderata	Intervallo portata	Test
Condizioni normali	216 gpm	214-216 gpm	normal
Livello di guasto 1	Riduzione del 10% (194gpm)	194-196 gpm	fwe10
Livello di guasto 2	Riduzione del 20% (173 gpm)	155-156 gpm	fwe20
Livello di guasto 3	Riduzione del 30% (151 gpm)	175-177 gpm	fwe30
Livello di guasto 4	Riduzione del 40% (130 gpm)	137-141 gpm	fwe40

### 2.4.4 Perdita di refrigerante

La rimozione di una parte del liquido refrigerante simula una perdita nell'impianto. La quantità nominale di refrigerante presente nel sistema è di 300 pound. Ogni livello di guasto riduce il carico di refrigerante del 10%. Inizialmente il sistema viene posto sottovuoto e poi riempito con 180 pound (Livello 4). I test svolti in presenza di guasto sono stati eseguiti in ordine inverso aumentando ogni volta il carico con 30 pound di refrigerante. Il quantitativo aggiunto è stato misurato con una bilancia industriale di incertezza  $\pm 0.5$  pound.

Tabella 2.5: Livelli di guasto per la perdita di refrigerante.

Caso	Condizione desiderata	Carica	Test
Condizioni normali	300 lbs	300 lbs	normal
Livello di guasto 1	Riduzione della carica del 10%	270 lbs	rl10
Livello di guasto 2	Riduzione della carica del 20%	240 lbs	rl20
Livello di guasto 3	Riduzione della carica del 30%	210 lbs	rl30
Livello di guasto 4	Riduzione della carica del 40%	180 lbs	rl40

L'errore per ogni livello è meno dell' 1%. Il totale di refrigerante aggiunto è calcolato rispetto ad un singolo livello e non rispetto alla carica precedente. I test considerati per questo guasto sono riportati in Tabella 2.5.

### 2.4.5 Sovraccarica di refrigerante

L'aggiunta di una parte del liquido refrigerante simula una sovraccarica nell'impianto. La quantità nominale di refrigerante presente nel sistema è di 300 pound. Ogni livello di guasto aumenta il carico di refrigerante del 10%. Questo guasto è un'estensione di quello precedente, vengono aggiunte 30 pound ad ogni livello. Il quantitativo aggiunto è stato misurato con una bilancia industriale di incertezza  $\pm 1$  pound. L'errore per ogni livello è meno del 1%. Il totale

Tabella 2.6: Livelli di guasto per il sovraccarica di refrigerante.

Caso	Condizione desiderata	Carica reale	Test
Condizioni normali	300 lbs	300 lbs	normal r
Livello di guasto 1	Riduzione della carica del 10%	330 lbs	ro10
Livello di guasto 2	Riduzione della carica del 20%	360 lbs	ro20
Livello di guasto 3	Riduzione della carica del 30%	390 lbs	ro30
Livello di guasto 4	Riduzione della carica del 40%	420 lbs	ro40

di refrigerante aggiunto è calcolato rispetto ad un singolo livello e non rispetto alla carica precedente.

### 2.4.6 Eccesso di olio

L'aggiunta di olio simula una sovraccarica di questo nell'impianto. La quantità nominale di olio presente nel sistema è di 22 pound come riportato dalle specifiche del produttore. Ogni livello di guasto aumenta l'olio nel sistema di varie quantità come mostrato in Tabella 2.7. Il quantitativo aggiunto è stato misurato con una bilancia di precisione con di incertezza  $\pm 0.1$  pound. La quantità di olio presente nel sistema è fornita con un'accuratezza di meno del 2%.

Tabella 2.7: Livelli di guasto per l'eccesso di olio.

Caso	Condizione desiderata	Carica reale	Test
Condizioni normali	22 lbs	22 lbs	normal eo
Livello di guasto 1	Riduzione della carica del 10%	25 lbs	eo14
Livello di guasto 2	Riduzione della carica del 20%	29 lbs	eo32
Livello di guasto 3	Riduzione della carica del 30%	33 lbs	eo50
Livello di guasto 4	Riduzione della carica del 40%	37 lbs	eo68

### 2.4.7 Sporciamento del condensatore

Vengono rimossi dei tubi per simulare lo sporciamento del condensatore. Il condensatore ha 164 tubi. Ogni livello di guasto riduce il numero di tubi come mostrato in 2.7. Il quantitativo aggiunto è stato misurato con una bilancia di precisione con incertezza  $\pm 0.1$  pound.

### 2.4.8 Presenza materia non condensabile nel refrigerante

Al sistema è stato aggiunto del nitrogeno per simulare la presenza di materia non condensabile nel refrigerante. La quantità di nitrogeno presente per ogni livello di guasto è mostrata in

Tabella 2.8: Livelli di guasto per lo sporcamento del condensatore.

Caso	Condizione desiderata	Condizione reale	Test
Condizioni normali	164 tubi	tubi non bloccati	normal cf
Livello di guasto 1	12% dei tubi bloccati	20 tubi bloccati	cf6
Livello di guasto 2	20% dei tubi bloccati	33 tubi bloccati	cf20
Livello di guasto 3	30% dei tubi bloccati	49 tubi bloccati	cf30
Livello di guasto 4	45% dei tubi bloccati	74 tubi bloccati	cf45

Tabella 2.9: Livelli di guasto per la presenza di materia non condensabile nel refrigerante.

Caso	Condizione desiderata	Condizione reale	Test
Condizioni normali	No nitrogeno	No nitrogeno	normal nc
Livello di guasto 1	1% in volume di nitrogeno	0.10 pound (1.0%)	nc1
Livello di guasto 2	2% in volume di nitrogeno	0.16 pound (1.7%)	nc2
Livello di guasto 3	3% in volume di nitrogeno	0.22 pound (2.4%)	nc3
Livello di guasto 4	5% in volume di nitrogeno	0.54 pound (5.7%)	nc4

Tabella 2.9. Anche la più piccola presenza di nitrogeno è facilmente riscontrabile. L'incertezza è di  $\pm 0.05$  pound.

La parte non condensabile è in genere aria che entra nel circuito del liquido refrigerante durante le manutenzioni.

#### 2.4.9 Defective pilot-valve

Questo guasto era già esistente nel sistema ed è stato risolto sostituendo la valvola. La valvola pilota è una valvola di laminazione (thermal expansion valve) con un sensore posto nell'evaporatore. Questa valvola regola il flusso del refrigerante nella conduttura minore che scorre parallelamente a quella principale. Il refrigerante dirottato torna poi nella conduttura principale attraverso la valvola principale. La posizione della valvola principale è controllata dalla quantità di liquido che proviene dalla condotta secondaria controllata dalla valvola pilota.

### 2.5 Individuazione degli stati stazionari e filtraggio dei dati

Il rilevamento degli stati stazionari (steady state detection) è un passaggio fondamentale per l'ottimizzazione, il controllo e la valutazione delle prestazioni dei processi. Molte tecniche di analisi ed elaborazione dei dati come l'ottimizzazione dei processi, il rilevamento guasti e l'identificazione di modelli (ad esempio per applicazioni in impianti chimici e industriali) richiedono che il sistema sia in stato stazionario (o a regime) per ottenere prestazioni ottimali.

Il sistema è considerato in stato stazionario quando tutte le sue variabili si sono stabilizzate attorno ad un valore operativo e in queste non avvengono cambiamenti significativi a meno dell'applicazione di un ingresso, come la variazione dei punti di lavoro (setpoint) o l'azione di un disturbo.

### 2.5.1 Individuazione degli stati stazionari

I dati del report 1043 forniscono una descrizione del sistema sia in regime stazionario che transitorio. Per l'analisi dei guasti si è scelto di riferirsi solamente agli istanti temporali in cui il sistema si trova in steady state. La scelta di non considerare i transitori è dovuta alla complessità dell'analisi del comportamento dinamico. In questo caso infatti le relazioni tra le variabili diventano non lineari e si deve tener conto dei ritardi introdotti dal sistema.

Il report 1043 fornisce per ogni test eseguito sull'impianto la serie dei 27 steady state raggiunti. Si è scelto però di estrarre i dati del sistema in regime stazionario direttamente dal dataset completo di ogni esecuzione (con campionamento ogni 10 secondi). Questa scelta è stata fatta in base a due considerazioni. La prima considerazione riguarda il fatto che non si conoscono i parametri di scelta che hanno determinato l'individuazione dei 27 stati stazionari proposti nel report. La seconda è legata all'utilizzo che si vuole fare dei dati, infatti per ottenere un buon modello è preferibile avere più informazioni possibili sul funzionamento del sistema.

Il chiller viene testato nell'arco di circa 15 ore (5000 campioni per ogni test eseguito). Se si assume che mediamente il sistema si trova in steady state per un arco di 20 minuti ogni volta che raggiunge un setpoint, si ottengono circa 3240 campioni, invece che 27 corrispondenti a steady state. Quindi si rende necessario l'utilizzo di un algoritmo che riconosca il comportamento statico del sistema. Una volta determinati questi campioni si effettua un'ulteriore filtraggio dei dati per eliminare valori non consistenti con il comportamento generale del sistema.

### Utilizzo delle variabili TEO, TCI, PRE e PRC

Poiché il condizionamento su tutte le variabili porterebbe all'ottenimento di pochi punti in steady state, si è scelto di considerare quattro variabili significative per la descrizione del sistema. Le quattro variabili scelte sono:

- TEO (la temperatura dell'acqua all'uscita dall'evaporatore),
- TCI (la temperatura dell'acqua in ingresso al condensatore),
- PRC (la pressione del refrigerante nel condensatore),
- PRE (la pressione del refrigerante nell'evaporatore).

La scelta di queste quattro variabili è dovuta al fatto che si vuole basare l'analisi del sistema su delle quantità che ne descrivano l'andamento complessivo. Le due temperature descrivono il comportamento dei cicli dell'acqua. Le due pressioni sono indicatrici dello stato del refrigerante nel circuito principale.

Una volta ottenuti gli istanti di campionamento in steady state per le quattro variabili selezionate si estraggono solo quei campioni per i quali tutte le variabili risultano essere in steady state.

Un'altra variabile controllata nel sistema, oltre a TEO è la temperatura di surriscaldamento del refrigerante all'uscita dell'evaporatore  $T_{sh\_suc}$ . Il sistema difficilmente riesce a

stabilizzarsi prima che questa variabile raggiunga lo steady state in quanto se il sistema non ha raggiunto il setpoint questa sarà sottoposta ad un'azione di controllo. Si è visto però che considerare questa variabile non aggiunge molta informazione rispetto a quella ottenuta.

### Procedimento utilizzato per individuare gli stati stazionari

Prima di procedere con la descrizione dell'algoritmo utilizzato si vogliono descrivere brevemente alcuni approcci che possono essere considerati per affrontare questo problema.

In alcune strategie si calcola la media di una finestra di dati e la si confronta con la successiva utilizzando una soglia di  $\pm 3$  volte la deviazione standard. Altre strategie calcolano la differenza tra la misura corrente e la media delle misure precedenti, successivamente la differenza è comparata con la deviazione standard. Un altro metodo considera la differenza tra due campioni e confronta questa con le differenze successive.

Il metodo che si è scelto di seguire è basato sulla regressione lineare ai minimi quadrati su una finestra di dati mobile. L'analisi viene svolta per una sola variabile alla volta. Per ogni finestra di dati l'algoritmo esegue una regressione lineare ai minimi quadrati per trovare la retta che meglio li descrive. La retta è indice del cambiamento dei dati all'interno dell'intervallo. L'equazione della retta ottenuta è:

$$y = mx + b, \quad (2.5)$$

dove  $b$  indica l'intercetta e  $m$  il coefficiente angolare. Il coefficiente angolare viene confrontato in seguito con una soglia. Se l'inclinazione è minore della soglia allora significa che per quel intervallo di tempo la variabile ha mantenuto un andamento stazionario. La lunghezza dell'intervallo da considerare e l'inclinazione della retta interpolatrice sono due parametri che dipendono dal tipo di variabile, in particolare dalla sua costante di tempo e dall'incertezza di misura ad essa associata.

### La scelta della soglia

L'individuazione della soglia ( $T$ ) utilizzata dall'algoritmo può avvenire attraverso diverse modalità. La soglia è considerata una variabile dipendente dal setpoint (SP) e dalla deviazione standard del rumore ( $\sigma$ ). L'ottenimento del parametro  $T$  avviene in seguito a vari esperimenti durante i quali viene eseguito l'algoritmo per differenti combinazioni di setpoint e  $\sigma$ . In seguito si esegue una regressione multipla per descrivere i differenti comportamenti ottenuti. L'equazione per la soglia è:

$$T = a_0 + a_1\sigma + a_2SP \quad (2.6)$$

Il metodo adottato in questa tesi è quello della scelta manuale dei parametri. Prima ci si basa su considerazioni generali riguardanti i setpoint e la deviazione standard del rumore, poi eseguendo delle prove ripetute si individuano i valori più adatti. Nel secondo caso si è scelto di mantenere costante il valore di  $T$  in quanto si ritiene che in questo caso il valore

dei setpoint e la varianza del rumore incidano in maniera minima. Nella prossima sezione si illustrano i motivi che hanno portato alla scelta dei parametri utilizzati.

### Individuazione dei parametri

Le soglie sono state calcolate tenendo conto che il sistema quando si trova in stato stazionario non deve subire variazioni nelle sue variabili di stato. Questo in situazioni reali ovviamente non è verosimile in quanto bisogna considerare del rumore sovrapposto. Supponendo che il rumore delle variabili derivi principalmente dall'incertezza di misura a loro associata, allora si può individuare un intervallo di tolleranza all'interno del quale si considera che la variabile sia ancora in stato stazionario.

Durante i test tra un setpoint e il successivo intercorrono indicativamente 30 minuti. In questo arco di tempo 5-15 minuti corrispondono al transitorio e 15-25 al funzionamento a regime. In media quindi si hanno 20 minuti di funzionamento a regime. Si sceglie di utilizzare finestre di 12 campioni (2 minuti) per eseguire le interpolazioni.

Per avere una stima del caso peggiore di  $m$  si divide l'intervallo di tolleranza per il tempo di durata della finestra e si considera il valore assoluto del risultato ottenuto. Effettuando delle prove sui dati a disposizione si è trovato che l'algoritmo fornisce risultati accettabili utilizzando le soglie  $m_1 = \pm 0.003$ ,  $m_2 = \pm 0.006$ ,  $m_3 = \pm 0.01$ ,  $m_4 = \pm 0.005$ . Risulta che i valori  $m_1$ ,  $m_2$ ,  $m_3$  e  $m_4$  permettono di individuare tutti gli steady state presenti nei test. Per questo motivo si è scelto di mantenere questi valori per svolgere l'intera analisi.

#### 2.5.2 Presentazione dei risultati nel caso di funzionamento normale

In questa sezione si presentano i risultati ottenuti applicando l'algoritmo ai dati riferiti al test svolto in condizioni normali. Si considerino i dati del test normal: in Figura 2.6 è riportato un intervallo del test corrispondente al passaggio tra il terzo e il quarto setpoint; per ogni gruppo di 12 campioni si individua una retta di regressione. Confrontando la pendenza con la soglia si può stabilire se i punti corrispondono a steady state o a un comportamento transitorio.

In Figura 2.7 è riportato l'andamento della variabile TEO attraverso i primi 9 setpoint; le linee rosse indicano gli intervalli considerati come steady state dall'algoritmo. In questo caso l'algoritmo individua abbastanza bene gli steady state del sistema: non riesce propriamente a distinguere il passaggio dal primo al secondo perché le condizioni sono molto simili e il transitorio è troppo breve. Questo non è un problema in quanto si vedrà che nel caso della pressione il passaggio è ben distinto (Figura 2.8) e contribuisce alla corretta individuazione.

In Figura 2.8 è riportato l'andamento della variabile PRC attraverso i primi 9 setpoint. Le linee rosse indicano gli intervalli considerati come steady state dall'algoritmo, tralasciando il transitorio iniziale si vede che l'algoritmo individua tutti e 9 gli steady state.

In Figura 2.9 sono mostrati gli steady state individuati dall'algoritmo per le variabili TEO, PRC, TCI e PRE.

Come controprova si plottano i 27 steady state (in rosso) individuati nel report (Figura 2.10) insieme agli steady state trovati (in blu). Si nota che i campioni selezionati si distribuiscono attorno agli steady state originali.

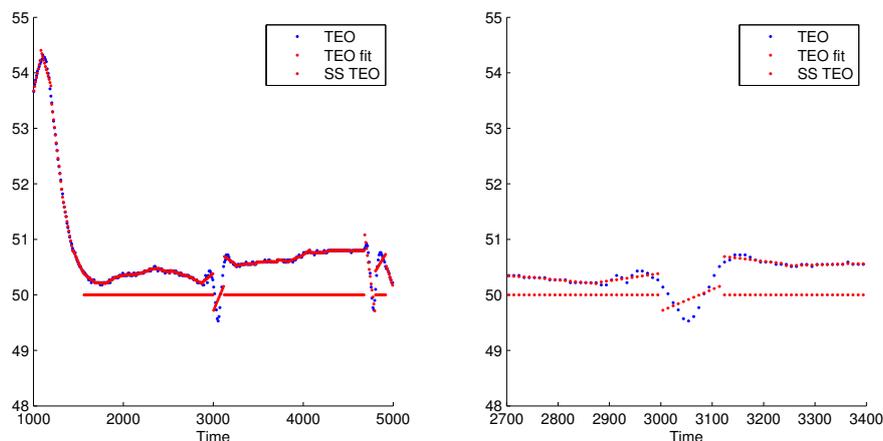


Figura 2.6: Regressione lineare su un intervallo di TEO. A sinistra è riportato un particolare della figura a destra.

Per provare che il sistema si trova in uno stato stazionario tutte le variabili considerate devono essere in stato stazionario. Si seleziona un insieme di otto variabili significative per il sistema e si fa un controllo su queste. Quattro sono riferite al ciclo secondario: il flusso dell'acqua all'evaporatore (FWE) e al condensatore (FWC) e le differenze di temperatura tra ingresso e uscita dell'acqua al condensatore e all'evaporatore (TWED e TWCD). Le altre quattro sono riferite al ciclo frigorifero: le pressioni del refrigerante al condensatore e all'evaporatore (PRC e PRE) e le temperature del refrigerante al condensatore e all'evaporatore (TRC e TRE).

In Figura 2.11 sono riportati gli andamenti delle variabili che vengono considerati steady state dall'algorithm. Si nota che in tutti i casi i punti individuati formano delle nuvole concentrate attorno ad un valore nominale.

### 2.5.3 Filtraggio dei dati

Una volta ottenuti i nuovi dataset con solo i dati in regime stazionario è necessario un filtraggio per eliminare eventuali valori non coerenti con il resto dei dati che compaiono nella variabili non sottoposte al trattamento precedente. Questi valori di solito si scostano abbondantemente dai valori operativi delle variabili quindi è sufficiente applicare delle soglie corrispondenti alle soglie medie di variazione della variabile.

Questa operazione viene svolta sia sui dati riguardanti il funzionamento normale sia sui dati di fault. Per il modo in cui è eseguita non assicura però l'assenza totale di dati non coerenti con l'andamento complessivo del sistema. La presenza di valori non coerenti, errori nelle misure e disturbi di vario genere presenti sul sistema contribuiscono a creare del rumore sovrapposto ai valori delle variabili considerate.

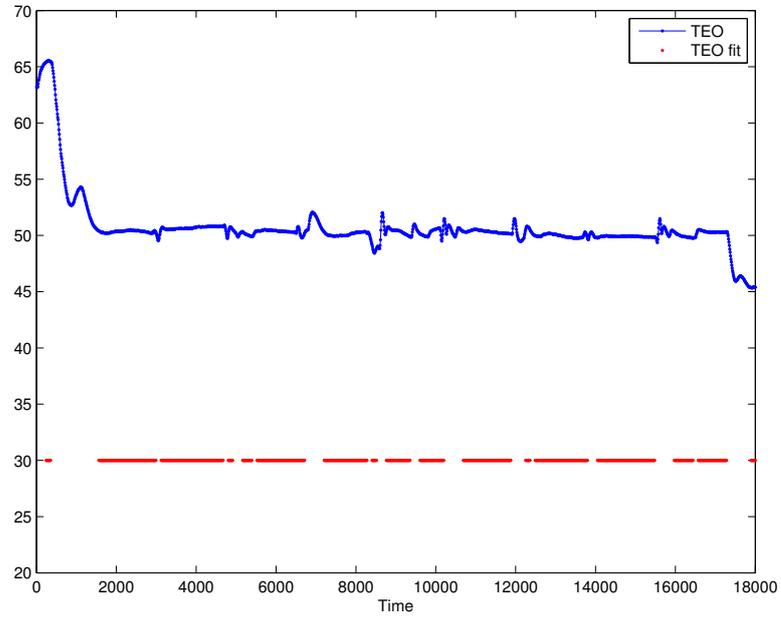


Figura 2.7: TEO attraversa i primi 9 setpoint, sono riportati anche gli intervalli considerati steady state.

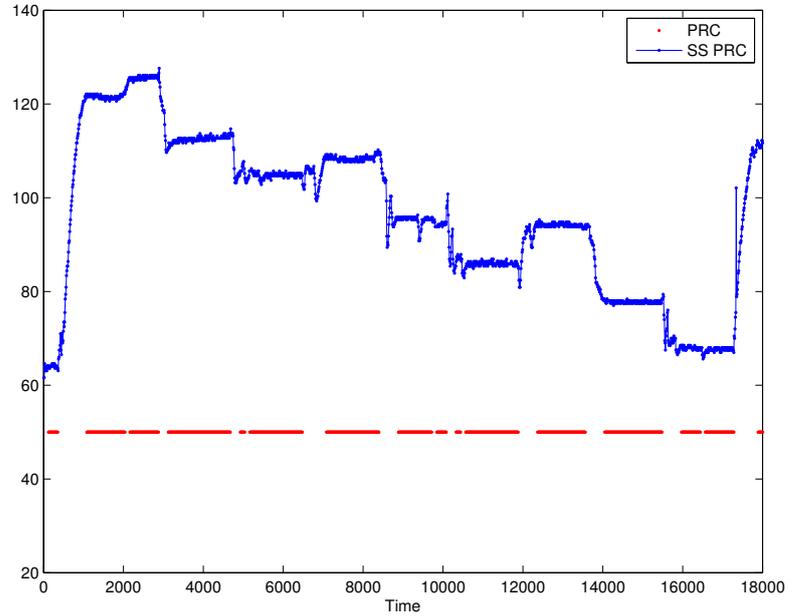


Figura 2.8: PRC attraversa i primi 9 setpoint, sono riportati anche gli intervalli considerati steady state.

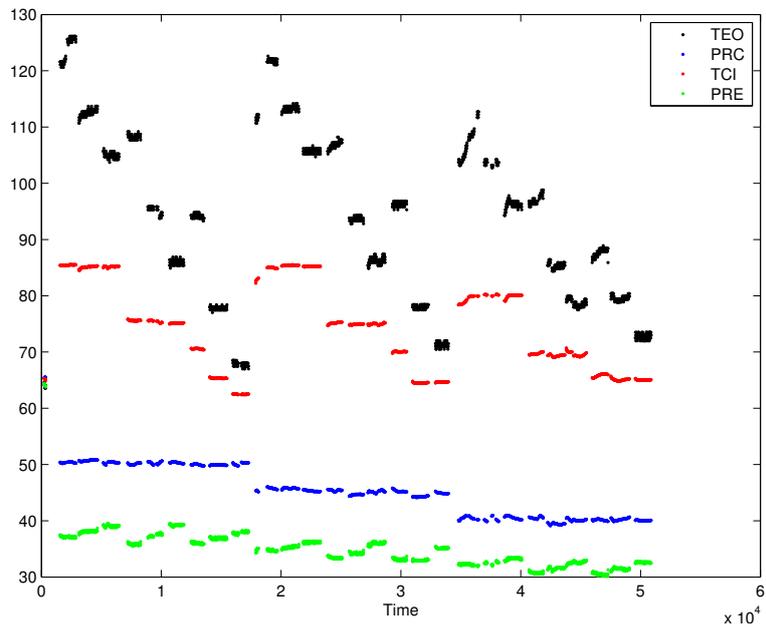


Figura 2.9: TEO, TCI, PRE, PRC in steady state.

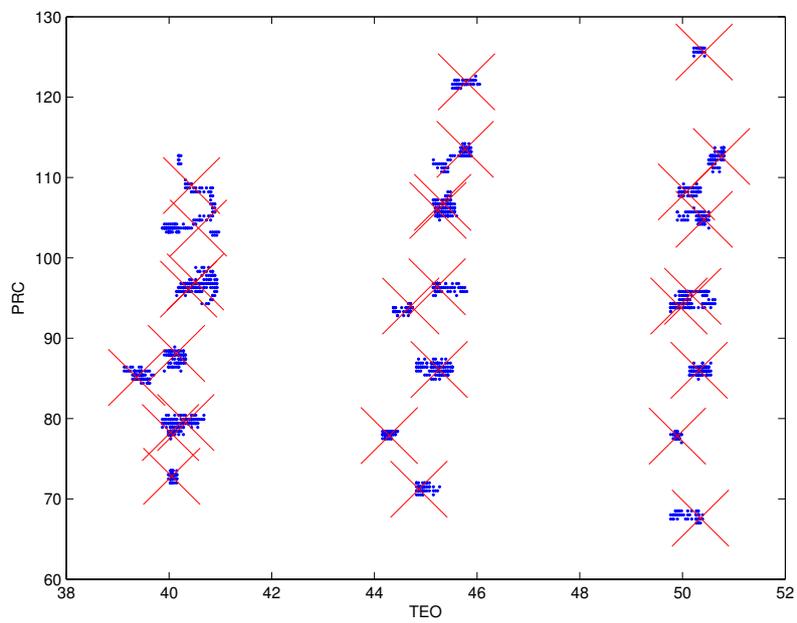


Figura 2.10: Valori di TEO e PRC trovati tramite l'algorithm (puntini blu) e quelli forniti dal report (crocette rosse).

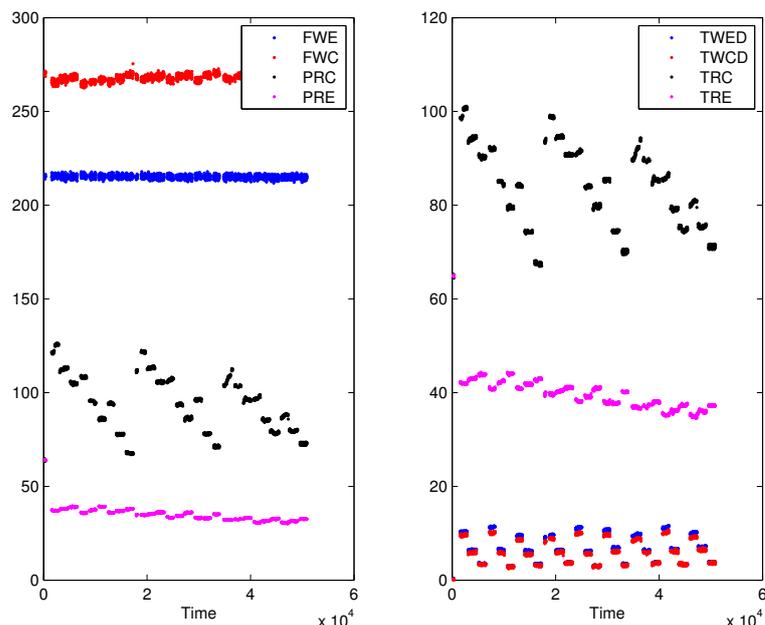


Figura 2.11: Verifica del comportamento di steady state su otto variabili significative.

## 2.6 Selezione delle variabili

La selezione delle variabili (feature selection) è stata utilizzata in molte applicazioni della computer science compresa la visione computazionale, pattern recognition e machine learning. Questa tecnica è molto importante per la riduzione di dimensionalità di campioni con alte dimensioni. La selezione delle variabili ha due vantaggi principali: rende il metodo applicato computazionalmente più efficiente e permette un incremento dell'accuratezza.

Le variabili del sistema fornite dal report RP-1043 ASHRAE sono considerate caratteristiche (o feature) che descrivono il comportamento del chiller ed è su queste che si basa l'analisi per l'individuazione dei guasti. In totale sono 66 come riportato in Tabella A.3. Di queste 66 feature disponibili si fa una prima selezione per individuare quelle più importanti al fine dell'analisi in modo da ridurne il numero.

In questo caso si sono individuate le variabili che indicano meglio lo stato del sistema, tra queste ci sono sia variabili primarie, direttamente derivate da sensori (e.g. TEA, TCA), variabili derivate da manipolazioni (e.g. EvapTons, TWED), variabili di controllo (e.g. Tsh\_suc) variabili legate alle performance (e.g. kW/ton). Tra tutte se ne sono individuate 15, elencate in Tabella 2.10.

Le caratteristiche individuate descrivono l'andamento generale del sistema.

Tabella 2.10: Le 15 caratteristiche considerate per l'analisi.

Caratteristica	Descrizione	Unità di misura
CondTons	Calculated heat rejection rate	Tons
EvapTons	Calculated evaporator cooling rate	Tons
kW	Instantaneous compressor power	kW
kW/Ton	Calculated Compressor Efficiency	-
TEA	Evaporator Approach Temperature	F
TCA	Calculated Heat Rejection Rate	F
PRE	Calculated Evaporator Cooling Rate	PSIG
PRC	Instantaneous Compressor Power	PSIG
TRC sub	Calculated Compressor Efficiency	F
Tsh suc	Refrigerant Suction Temperature	F
Tsh dis	Refrigerant Discharge Temperature	F
TO net	Oil in Sump minus Oil Feed Temperature	F
PO net	Oli Feed minus Oil Vent Pressure	PSIG
TWCD	Condenser Water Temperature Difference	F
TWED	Evaporator Water Temperature Difference	F

## Capitolo 3

# Novelty Detection

Con il termine *novelty detection* si intende la procedura di individuazione di dati nuovi o sconosciuti attraverso l'ausilio di algoritmi di apprendimento automatico (*machine learning*). Nel caso di sistemi HVAC, le tecniche di *novelty detection* possono essere utilizzate per la rilevazione di guasti (i.e comportamenti del sistema non conformi a quello nominale) soprattutto nel caso in cui questi siano rari e/o non siano disponibili dati caratterizzanti i malfunzionamenti.

In questo capitolo si descrivono sinteticamente alcune tecniche di *novelty detection* presenti in letteratura che possono essere impiegate per applicazioni di *machine learning*. Quando si parla di *novelty detection* applicata alla diagnosi di guasti in impianti HVAC si fa riferimento a elementi di novità nei dati (i.e anomalie), che rappresentano tipicamente situazioni potenzialmente dannose per il sistema.

In questa sede i termini *novelty* e *anomaly* sono considerati equivalenti e utilizzati come sinonimi, tuttavia va precisato che se in altre applicazioni possono essere utilizzati con accezioni diverse; taluni infatti distinguono il procedimento di *novelty detection* dall'*anomaly detection* in quanto nel primo caso, oltre all'individuazione dell'elemento di novità il sistema, incorpora l'informazione e la utilizza nelle analisi successive. In generale le soluzioni proposte per il problema della *novelty detection* sono utilizzate anche in quello dell'*anomaly detection*.

Un buon sistema di *novelty detection* deve poter gestire informazioni riguardanti i dati di test sconosciuti oppure non presenti/disponibili al momento del training del modello. L'abilità di un sistema nell'individuare eventi ignoti rispetto a tutti quelli noti in precedenza è una caratteristica dei sistemi di *machine learning* cosiddetti non supervisionati. Di fatto non è possibile addestrare un sistema di apprendimento automatico rispetto a tutte le possibili situazioni con cui il sistema potenzialmente si può trovare, quindi è importante che durante la fase di test il sistema sia in grado di discernere situazioni già note da novità.

La letteratura in questo campo è molto vasta: in questo capitolo si fornisce una breve descrizione delle tecniche utilizzate più frequentemente. Questi metodi si prestano a innumerevoli applicazioni nei più svariati campi tra i quali *signal processing*, *computer vision*, *pattern recognition*, *data mining* e *robotics*. Nel seguito sono elencati gli aspetti più significativi quando si considera un metodo di *novelty detection*:

- *Robustezza e trade-off.* Un metodo di novelty detection deve essere robusto rispetto ai dati di test. Con questo si intende che deve massimizzare l'esclusione dei dati che portano novità e allo stesso tempo devono minimizzare l'esclusione dei campioni già noti. Questo trade-off deve essere predicibile e sotto controllo sperimentale.
- *Scalatura uniforme dei dati.* Per migliorare le prestazioni a volte è preferibile normalizzare i dati.
- *Minimizzazione dei parametri.* Un metodo di novelty detection deve minimizzare il numero di parametri che vengono impostati dall'utente.
- *Generalizzazione.* Il sistema deve essere in grado di generalizzare senza confondere gli elementi di novità.
- *Indipendenza.* Il metodo di novelty detection deve essere indipendente dal numero di feature (i.e le variabili considerate) o classi, inoltre non deve mantenere buone prestazioni anche nel caso si utilizzino dataset sbilanciati, vi siano pochi campioni o i dati siano affetti da rumore.
- *Adattabilità.* Un sistema che riconosca dei dati come novelty deve essere in grado di aggiornarsi.
- *Complessità computazionale.* Un gran numero di queste applicazioni sono applicazioni online per questo motivo la loro complessità computazionale deve essere la più bassa possibile.

Con il termine anomaly detection ci si riferisce al problema di trovare comportamenti nei dati non conformi a quanto ci si aspetta. Questi comportamenti non conformi sono spesso indicati anche come outlier, osservazioni discordanti, aberrazioni, novità in differenti applicazioni. I termini anomalie, novelty e outlier sono utilizzati per indicare queste aree di ricerca. L'importanza dell'anomaly detection sta nel fatto che a livello dei dati si traducono in informazioni significative e a volte critiche per il sistema in osservazione.

Alcuni esempi di applicazioni sono:

- *Monitoraggio web.* Un traffico anomalo in una rete di computer può nascondere un attacco da parte di un hacker che sta sottraendo dati sensibili per spedirli verso una postazione non autorizzata [Kumar, 2005].
- *Transizioni con carte di credito.* Anomalie nei trasferimenti di denaro tramite carte di credito possono indicare furti di identità o delle carte stesse [Aleskerov et al., 1997].

La Figura 3.1 illustra un insieme di dati in due dimensioni. Ci sono due regioni di dati normali  $N_1$  e  $N_2$  dal momento che la maggior parte dei dati è situata in questa regione. I punti che non si trovano in queste regioni sono considerati anomalie, come ad esempio  $o_1$  e  $o_2$  e i punti nella regione  $O_3$ .

L'attività di ricerca di anomalie è legata ma distinta dal *noise removal* [Teng et al., 1990] e dal *noise accommodation* [Piepel, 1989] che riguardano la gestione del rumore nei dati. Il

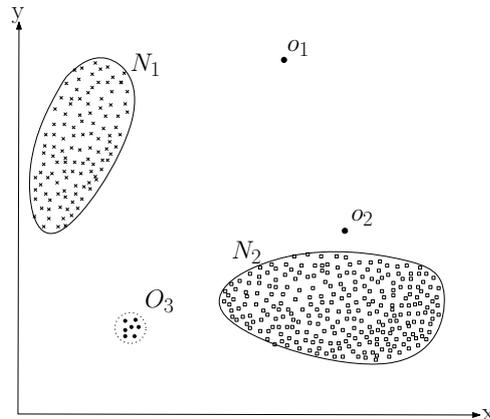


Figura 3.1: Esempio di anomalie in un data set a due dimensioni ([Chandola et al., 2009]).

rumore può essere considerato un fenomeno dei dati che non è di interesse per gli analisti ma costituisce un ostacolo per l'analisi dei dati. Durante un'operazione di *noise removal* l'intento è quello di rimuovere dati indesiderati prima di compiere qualsiasi attività di analisi dei dati. Il termine *noise accommodation* si riferisce invece ad immunizzare il modello statistico impiegato per l'identificazione di anomalie.

### 3.1 Apprendimento supervisionato e non supervisionato

Quando si forniscono i dati da utilizzare per una classificazione si fa la distinzione tra dati ritenuti *normali* e dati ritenuti *anomalie*. Non sempre però risulta facile ottenere dati etichettati, in quanto di solito l'etichettatura viene fatta manualmente da un operatore esperto e quindi risulta notevole lo sforzo per ottenere un dataset completo. Difficilmente si riesce a ottenere una panoramica completa di tutti i comportamenti anomali nei quali può incorrere il sistema. Inoltre il comportamento delle anomalie è per sua natura dinamico e nuovi tipi di anomalie possono sempre insorgere. Le tecniche di rilevazione delle anomalie si distinguono in tre differenti casi: tecniche supervisionate, semi-supervisionate e non supervisionate.

- *Supervised Anomaly detection.* Le tecniche supervisionate (supervised) si basano sul presupposto che i dati siano etichettati, cioè che ogni anomalia sia classificata con un apposita etichetta. Questa modalità ha il vantaggio di poter sfruttare l'informazione a priori sulle anomalie durante la fase di costruzione del classificatore. Quando si presenta un nuovo dato, questo viene confrontato attraverso tecniche statistiche con le due classi (dati normali e anomalie) per stabilire a quale appartenga. I problemi che sorgono con l'utilizzo di questo approccio sono di due tipi. Il primo è legato alla differenza del numero di campioni disponibili per le due classi, tipicamente si hanno più dati normali che anomali disponibili. Il secondo riguarda la difficoltà di ottenere le etichette per le anomalie.

- *Semi-supervised anomaly detection.* Nel caso semi-supervisionato si fa l'assunzione che siano disponibili solo le etichette dei dati normali e che invece non si abbiano a disposizione informazioni sul tipo di anomalie che possono insorgere. Il fatto di non richiedere una conoscenza delle anomalie che possono presentarsi rende questa tecnica più facilmente utilizzabile nelle applicazioni pratiche rispetto alle tecniche supervisionate. L'approccio tipico utilizzato per questo genere di tecniche è quello di costruire un modello per il comportamento normale e utilizzare l'informazione ricavata dai dati normali per riconoscere le anomalie.
- *Unsupervised anomaly detection.* L'approccio non supervisionato non richiede nessuna conoscenza o classificazione del comportamento normale né delle anomalie che possono insorgere. Le tecniche non supervisionate non richiedono dati di training e quindi sono le più utilizzabili nelle applicazioni pratiche. In questa modalità si assume che i dati normali siano molto più frequenti delle anomalie, se questo non è vero si otterrà un tasso alto di falsi allarmi. Alcune tecniche semi-supervisionate possono essere riadattate in scenari non supervisionati utilizzando come dati di training i campioni non etichettati assumendo che i dati di test contengano poche anomalie rispetto ai dati normali e che il modello ottenuto sia robusto rispetto a quelle poche anomalie considerate.

Un semplice approccio per l'individuazione delle anomalie è quello di definire una regione che rappresenti il comportamento normale al di fuori della quale ogni dato osservato viene considerato anomalo. Molti fattori rendono questo approccio controverso:

- Il modello non deve solo descrivere bene i dati utilizzati per il training, cioè fornire soluzioni 'overfitted' ma deve descrivere il comportamento generale.
- In molti casi il comportamento normale tende a modificarsi nel tempo e la nozione iniziale di normalità può perdere significatività nel futuro.
- L'esatta definizione di anomalia cambia da contesto a contesto. Per esempio in applicazioni mediche una piccola variazione rispetto al comportamento normale (e.g. fluttuazioni della temperatura corporea) possono essere considerate anomalie, mentre variazioni simili in altre situazioni (e.g. nel commercio una variazione delle quantità) possono essere ritenute normali.

A causa di queste problematiche il problema dell'individuazione delle anomalie, nelle forme più generali non è semplice da risolvere. Molte delle tecniche esistenti per questi problemi sono infatti formulazioni specifiche per i dati contesti. La formulazione del problema inoltre è dovuta a diversi fattori come la natura dei dati, la disponibilità di dati già classificati, tipi di anomalie da trovare ecc. In Figura 3.2 si riportano le componenti chiave associate alle tecniche di anomaly detection.

Un altro aspetto da considerare sono i risultati ottenuti dopo l'applicazione del metodo di rilevazione delle anomalie. Le risposte ottenute sono di due tipi:

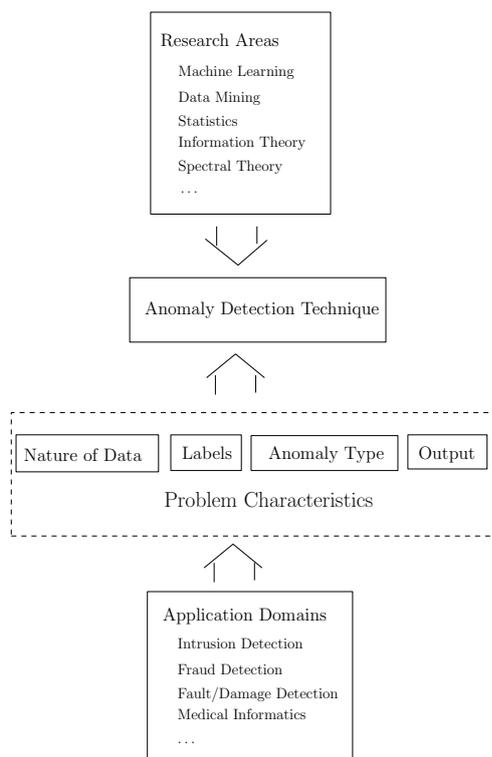


Figura 3.2: Componenti chiave associati alla tecnica di individuazione delle anomalie ([Chandola et al., 2009]).

- *Punteggi.* Ad ogni dato viene affibbiata un'etichetta con la quale si descrive il grado di certezza che il dato sia veramente un'anomalia. I dati classificati quindi possono essere ordinati secondo una scala che può essere poi utilizzata per l'analisi, ad esempio prendendo solo le anomalie con grado di probabilità maggiore o inserendo una soglia di tolleranza.
- *Label.* Ad ogni dato si mette un'etichetta (label) con due sole distinzioni, normale e anomalia. Questo approccio non permette un controllo diretto come nel caso precedente ma in modo indiretto si possono modificare i parametri del metodo di classificazione per ottenere i risultati voluti.

## 3.2 Tipi di anomalie

Un aspetto importante che riguarda le tecniche di individuazione delle anomalie è la natura di queste. Le anomalie possono essere classificate nelle tre seguenti categorie: *anomalie puntuali*, *anomalie contestuali* e *anomalie collettive*. Nelle prossime sezioni se ne da una breve descrizione.

### 3.2.1 Anomalie Puntuali

Se un singolo campione o dato può essere considerato anomalo rispetto al resto dei campioni allora viene detta anomalia puntuale. Questa è la tipologia più semplice di anomalie ed è il tipo di l'anomalia più considerata negli studi in questo campo. Per esempio, in Figura 3.1, i punti  $o_1$  e  $o_2$  come anche i punti nella regione  $O_3$  si trovano all'esterno dei limiti delle regioni normali, quindi sono ritenuti anomalie puntuali in quanto differiscono dai dati normali.

Come esempio pratico si pensi alla contraffazione delle carte di credito. Si supponga che il data set corrisponda alle transazioni di un unico individuo e si consideri il credito speso come unica feature. Se la somma spesa risulta essere maggiore rispetto a quello che di solito la persona spende potrebbe trattarsi di un'anomalia puntuale.

### 3.2.2 Anomalie Contestuali

Se un dato campione può essere ritenuto un'anomalia solo in certi contesti (e non altrimenti) allora si parla di anomalie contestuali. La nozione di contesto deve far parte della formulazione del problema. Ogni campione è descritto da due tipi di attributi:

- *Contestuali.* Essi determinano il contesto in cui il dato si manifesta. Ad esempio in una serie temporale di dati il tempo è un attributo contestuale in quanto determina la sua posizione all'interno della serie.
- *Di comportamento.* Questi attributi non riguardano il contesto in cui appaiono. Ad esempio in una serie temporale di dati che descrivono le precipitazioni piovose in tutto il mondo, il livello raggiunto in ogni località considerata è un attributo legato al comportamento.

I comportamenti anomali vengono individuati considerando gli attributi di comportamento all'interno di un dato contesto. Un dato campione può essere un'anomalia se considerato all'interno di un certo contesto mentre lo stesso dato può essere ritenuto normale all'interno di un altro contesto. Questa è una proprietà molto importante nell'individuazione di anomalie contestuali. La Figura 3.3 mostra un esempio di serie temporale di temperature mensili di

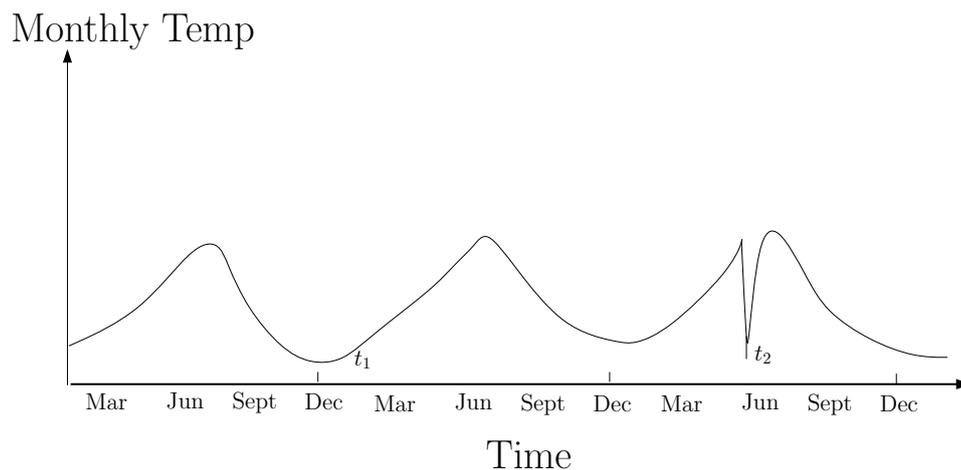


Figura 3.3: Anomalia contestuale  $t_2$  in una serie di dati temporali. Si noti che la temperatura  $t_1$  è uguale a  $t_2$  ma non è considerata un'anomalia in quanto si presenta in un diverso contesto ([Chandola et al., 2009]).

un luogo lungo l'arco di qualche anno. Una temperatura di  $2^\circ\text{C}$  può essere normale durante il periodo invernale (al tempo  $t_1$ ) in quel luogo mentre la stessa temperatura in estate (al tempo  $t_2$ ) sarà un'anomalia.

La scelta di usare una tecnica basata sulle anomalie contestuali è determinata dalla significatività di considerare un determinato contesto nell'applicazione in esame. Un fattore chiave da considerare è la disponibilità di attributi di contesto, in molti casi infatti definire dei contesti è naturale e quindi utilizzare anomalie contestuali diviene semplice. In altri casi diviene difficile definire dei contesti significativi e l'analisi perde di efficacia.

#### 3.2.3 Anomalie Collettive

Se un insieme di dati è anomalo rispetto al dataset viene detta anomalia collettiva. I singoli dati di un'anomalia complessiva possono non essere delle anomalie ma la loro comparsa collettiva è anomala. Si noti che mentre le anomalie puntuali compaiono in qualsiasi dataset, le anomalie collettive possono trovarsi solo in dataset in cui tra i campioni corrono delle relazioni. Invece le anomalie contestuali possono verificarsi solo se si dispone di attributi di contesto per i dati. Un'anomalia puntuale o un'anomalia collettiva possono essere anomalie contestuali se analizzate rispetto ad un dato contesto. Quindi un problema di individuazione

di anomalie puntuali o collettive può essere trasformato in un problema riguardante anomalie contestuali se si considerano informazioni di contesto.

### 3.3 Tecniche di rilevazione delle anomalie

Si riporta in seguito un breve resoconto delle tecniche che possono essere utilizzate quando si affronta il problema dell'anomaly detection. Per maggiori dettagli si consulti Chandola et al. [2009]. Le tecniche si dividono in quattro tipologie di tecniche principali:

- Tecniche di rilevazione delle anomalie basate sulla classificazione.
- Tecniche di rilevazione delle anomalie basate sulla vicinanza.
- Tecniche di rilevazione delle anomalie basate sul clustering.
- Tecniche statistiche.

#### 3.3.1 Basate sulla classificazione

Le tecniche basate sulla classificazione operano in due fasi: costruiscono un modello a partire dai dati di training opportunamente etichettati, confrontano i dati di test con il modello che li classifica come normali o anomali. Di seguito sono elencate alcune tecniche tra le più importanti che utilizzano il metodo della classificazione:

- *Reti neurali (Neural Networks Based)*. Queste tecniche sono utilizzate sia per problemi supervisionati che non supervisionati. Una tecnica semplice (di tipo supervisionato) che utilizza le reti neurali allo scopo di individuare un'anomalia opera in due passi. Prima la rete impara a inferire le relazioni che legano i dati a partire da un insieme di dati di addestramento (*training set*). Poi ogni dato utilizzato per la validazione (*validation set*) viene processato dalla rete. Se la rete rifiuta il dato in ingresso è considerato un'anomalia, De Stefano et al. [2000].
- *Bayesian Networks Based*. La rilevazione di anomalie tramite questa tecnica si applica principalmente al caso di gestione di classi multiple. Nel caso univariato con variabili categoriche si può applicare il classificatore di Bayes che stima la probabilità a posteriori di osservare una certa classe di dati data una certa osservazione. La classe con probabilità a posteriori più alta viene scelta. La probabilità di osservare un certo campione data la classe e le probabilità a priori delle classi sono stimate dai dati di addestramento.
- *Macchine a vettori di Supporto (Support Vector Machines)*. Le macchine a vettori di supporto (SVMs) Vapnik [2000] sono state ampiamente sfruttate per scopi di classificazione delle anomalie sia nel caso supervisionato che in quello non supervisionato. In particolare nel caso semi-supervisionato si considera una sola classe e si cerca di individuare una regione che contenga i dati di training. Per descrivere queste regioni

si possono impiegare kernel di tipo *radial basis function* (RBF). Per ogni campione utilizzato per il test, il classificatore determina se appartiene o no alla regione. Se il campione ricade al di fuori di questa è considerato un'anomalia.

- *Rule Based.* Le tecniche di individuazione delle anomalie basate sull'utilizzo di regole individuano delle regole che descrivono il comportamento dei dati normali; i dati che non rispettano queste regole sono considerati come anomalie. Questo approccio è utilizzato sia in campo supervisionato che non supervisionato. Il metodo per individuare anomalie utilizzando multiclassi consiste in due passaggi. Costruisce le regole utilizzando i dati di training e un algoritmo di apprendimento come gli alberi di decisione. Ad ogni regola è associato un valore di confidenza basato sul numero di campioni di training che riesce a classificare correttamente. Per ogni dato di test, si individua la regola che ha il più alto valore di confidenza, il suo inverso costituisce un indice per il punteggio dell'anomalia.

I vantaggi principali nell'usare le tecniche di classificazione sono la possibilità di sfruttare algoritmi efficaci e la velocità nella fase di test una volta ottenuto il classificatore. Tra gli svantaggi si citano la necessità di disporre di dati etichettati e la mancanza di una misura diretta che dia l'idea della bontà della classificazione.

#### 3.3.2 Basate sulla vicinanza

Questo tipo di tecniche si basano sull'assunzione che i dati normali si distribuiscono in una regione densa mentre le anomalie assumono una configurazione più sparsa. In molti casi è richiesta la definizione di una distanza (o una similarità) tra i dati compatibile con il tipo di attributi considerato. Le tecniche che utilizzano un approccio nearest neighbor si possono distinguere in due categorie che utilizzano parametri diversi per ottenere un punteggio da associare ad un'anomalia:

- *Distanza dai k-esimi vicini.* Questa tecnica utilizza la distanza di un dato dai suoi vicini per ottenere un indice di anomalia.
- *Densità relativa.* Si utilizza la densità relativa per ottenere un indice analogo (ad esempio il Local Outlier Factor LOF Breunig et al. [2000]).

Il vantaggio di questa tecnica è che non ha bisogno di nessuna conoscenza a priori sui dati. Nei casi semi-supervisionati si ottengono prestazioni migliori rispetto ai casi non supervisionati in quanto difficilmente nei dati di training le anomalie hanno grande densità. Nei casi non supervisionati possono insorgere problemi se i dati normali non sono abbastanza vicini o se le anomalie hanno molti dati distribuiti attorno. Inoltre altri svantaggi sono che aumenta la complessità computazionale e le prestazioni in genere dipendono dalla definizione della distanza.

#### 3.3.3 Clustering

Il clustering è una tecnica principalmente non supervisionata. Consiste nel raggruppare dati *simili*. Le tecniche che utilizzano cluster si distinguono sulla base di tre assunzioni principali:

1. *I dati normali appartengono a dei cluster, mentre le anomalie non appartengono a nessun cluster.*
2. *I dati normali si dispongono attorno al centroide del cluster più vicino, mentre le anomalie sono distanti rispetto al centroide del cluster più vicino.*
3. *I dati normali appartengono a cluster grandi e densi, mentre le anomalie appartengono a cluster piccoli e sparsi.*

Le tecniche di clustering possono operare in modo non supervisionato, e sono facilmente riadattabili ad altri tipi di dati. La fase di test è veloce in quanto un dato deve venir confrontato con un numero di cluster basso. D'altra parte le prestazioni dipendono dal tipo di algoritmo di clustering utilizzato. Molte tecniche funzionano solo se le anomalie non formano grandi cluster tra loro. La complessità computazionale spesso è un collo di bottiglia.

### 3.3.4 Approcci statistici

Le tecniche di anomaly detection di tipo statistico si basano sull'assunzione che le anomalie siano osservazioni di scarsa rilevanza in quanto non generate dal modello stocastico preso in considerazione.

In particolare le anomalie hanno una probabilità bassa di presentarsi nella regione del modello stocastico, mentre i dati nominali (o normali) hanno un'alta probabilità di presentarsi nella regione del modello stocastico.

Le tecniche statistiche calcolano un modello (tipicamente del comportamento nominale) dei dati di training e in seguito eseguono dei test per determinare se l'osservazione appartiene a quel dato modello o no. Le osservazioni che hanno una scarsa probabilità di appartenere al modello, basandosi sui risultati ottenuti dal test considerato, sono ritenute anomalie.

Le tecniche utilizzabili si dividono in due categorie tecniche parametriche e tecniche non parametriche.

#### Tecniche parametriche

In questo caso si assume che i dati normali siano generati da una distribuzione nota, e si debbano stimare i parametri a partire dai dati di training. Oppure attraverso un test di ipotesi si assume che l'ipotesi nulla ( $H_0$ ) sia che i dati sono stati generati da una distribuzione con certi parametri. Si può fare un'ulteriore distinzione tra le tecniche parametriche:

- *Modelli gaussiani.* Questo approccio assume che i dati sino generati da una distribuzione gaussiana. I parametri del modello gaussiano sono stimati utilizzando il metodo della massima verosomiglianza. La distanza dei dati dalla media stimata determina il punteggio associato all'anomalia, quindi tramite la definizione di una soglia si può utilizzare il punteggio per definire l'anomalia. Ad esempio nel caso gaussiano si prende  $3\sigma$  come distanza di riferimento dalla media oltre la quale un dato è considerato anomalia in quanto si trova oltre la regione che contiene il 99,7% dei dati.

- *Modelli di regressione.* I modelli di regressione sono ottenuti a partire dai dati. Il punteggio associato alle anomalie viene calcolato tramite i residui che costituiscono la parte dei campioni che non è spiegata dal modello. Attraverso dei test statistici si può ottenere un livello di confidenza per le anomalie [Anscombe, 1960].
- *Mixture of parametric distribution.* Queste tecniche fanno uso di più modelli parametrici per descrivere i dati. Il modello può essere utilizzato sia per modellare solamente i dati normali sia entrambi i dati e le anomalie. Un esempio di questo metodo sono le Gaussian Mixture Modelling (GMM) nelle quali si considerano densità di probabilità gaussiane. Per sviluppare il modello si impiegano algoritmi di stima come EM.

#### Tecniche non parametriche

Le tecniche non parametriche sono quelle tecniche in cui la densità di probabilità si deduce a partire dai dati senza fare l'ipotesi di conoscere il suo modello parametrico.

- *Istogrammi.* La tecnica parametrica più semplice che si può considerare fa uso dell'istogramma come stima della frequenza dei valori delle variabili di training. Una volta costruito l'istogramma se il valore da analizzare ricade in uno dei livelli è ritenuto un dato normale. Una misura del punteggio dell'anomalia è data dalla frequenza (altezza) dell'istogramma.
- *Metodi che utilizzano funzioni kernel.* Un metodo non parametrico per la stima della densità di probabilità sono le finestre di Parzen [Parzen, 1962]. Queste tecniche sono simili a quelle descritte per i metodi parametrici, la differenza sta nella tecniche di stima della densità di probabilità utilizzata.

Se le assunzioni riguardanti le distribuzioni che generano i dati sono valide, allora le tecniche statistiche forniscono una soluzione giustificata dell'individuazione delle anomalie. Il punteggio associato all'anomalia è accompagnato da un intervallo di confidenza che può essere utilizzato come informazione addizionale per la decisione finale. Inoltre se la stima della distribuzione è robusta rispetto alle anomalie presenti nei dati i metodi statistici possono operare in modo non supervisionato.

Lo svantaggio di questi metodi è che difficilmente è individuabile una distribuzione che genera i dati, soprattutto in casi in cui si considerano dimensioni elevate. Inoltre anche se le assunzioni fatte sono giustificabili non è facile scegliere il test statistico migliore.

## Gestione delle anomalie contestuali

Per individuare un'anomalia contestuale i dati devono avere attributi che ne definiscono il contesto e attributi che rendono possibile l'individuazione delle anomalie nel dato contesto. I contesti possono essere divisi in quattro categorie:

- *Spaziale.* Il contesto è definito da attributi spaziali.

- *Basato su grafo*. Ricorre alle relazioni specificate dalla struttura di un grafo
- *Sequenziale*. Si riferisce a sequenze di dati come serie temporali o serie di campioni.
- *Basato su dei profili*. Quando non esiste un contesto specifico spaziale o temporale ma il contesto è dato dal raggruppamento di più attributi.

I possibili approcci per rilevare un'anomalia contestuale sono due. Il primo riduce il problema contestuale in un problema puntuale, il secondo modella la struttura dei dati e utilizza il modello per trovare le anomalie.

## Gestione delle anomalie collettive

Trovare questo tipo di anomalie è più complesso rispetto al caso di anomalie puntuali o contestuali perché bisogna considerare la struttura dei dati e le relazioni che li legano. Ci sono tre tipi di relazioni che sono più frequenti:

- *Sequenziali*. In questo caso si considerano le anomalie come sotto-sequenze di una sequenza principale.
- *Spaziali*. Le anomalie spaziali collettive sono sotto-regioni dei dati che hanno un comportamento anomalo.
- *Basate su grafo*. Le anomalie individuate sono sotto-grafi connessi che hanno un comportamento anomalo.

La presenza di dati sequenziali caratterizza molte applicazioni. Anche in questo caso si può cercare di riportare il problema al caso di un'anomalia puntuale utilizzando le sequenze come feature in uno spazio finito. Altre tecniche costruiscono un modello della sequenza con il quale creano poi delle regole generali di associazione che utilizzano per riconoscere le anomalie. Un altro metodo si basa sul riconoscimento di sotto sequenze anomale all'interno di una sequenza più lunga supponendo che la sequenza segua un pattern predefinito, dove una sequenza anomala è una sequenza che non si conforma con questo pattern.

## Capitolo 4

# Il Classificatore One-Class SVM

Il problema dell'anomaly detection, per un'applicazione peculiare in ambito HVAC, è affrontato giovandosi di tecniche di apprendimento automatico. Specificatamente si è optato per classificatori a massimo margine (SVM). Il principale vantaggio di questo approccio risiede nel fatto che non richiede la conoscenza di un modello del sistema da supervisionare. Nel caso in esame si sfruttano algoritmi di apprendimento semi-supervisionato che agiscono in due fasi: la prima fase è dedicata alla raccolta dei dati di training riferiti al funzionamento normale, la seconda fase utilizza i dati di training raccolti per costruire un modello in grado di definire il supporto dei dati riferiti al comportamento normale (detti normali).

L'algoritmo di classificazione ottenuto può lavorare in linea con i dati e restituire il risultato della classificazione in corrispondenza di ogni nuovo campione. Il classificatore è basato sul modello a partire dai dati di training e il risultato della classificazione è fornito sotto forma di etichetta binaria: i possibili valori in uscita dal classificatore sono -1 e 1, dove con -1 si indicano le anomalie mentre con 1 si indica il comportamento normale. Il segno delle etichette è determinato valutando il segno di una funzione di decisione ( $f(x)$ ). In molti casi si è invece interessati più ad un livello di confidenza che alla semplice classificazione; a quest'ultima quindi può essere utile associare un valore sulla bontà di decisione (e.g. la distanza che ogni dato osservato ha dal bordo della funzione di decisione).

In questo capitolo si descriverà in dettaglio una tecnica di classificazione tramite Support Vector Machines [Cortes and Vapnik, 1995]. In particolare, si presenta un caso semi-supervisionato che fa uso di un classificatore di tipo One-Class Support Vector Machines (OCSVM) per risolvere il problema della Novelty Detection. L'utilizzo di questa tecnica fornisce alcuni vantaggi come ad esempio la possibilità di lavorare con dati in elevate dimensioni e la possibilità di utilizzare funzioni kernel [Schölkopf et al., 2001] che permettono di ottenere:

- regioni di decisione non lineari,
- problemi di ottimizzazione convessi,
- flessibilità nelle soluzioni.

Nella trattazione successiva si intenderà implicitamente l'utilizzo di kernel quando si applica OCSVM.

Una volta scelto il kernel da utilizzare per il classificatore OCSVM si pone il problema della scelta dei parametri che definiscono il kernel e di quelli che definiscono le prestazioni del classificatore. Questi parametri possono essere determinati in maniera automatica tramite l'impiego di un algoritmo che utilizza un approccio euristico per valutare la bontà del classificatore.

## 4.1 Formulazione del Problema One-Class SVM

Il classificatore a minimo margine One-Class SVM, costituisce una particolarizzazione del più generale problema delle Support Vector Machines nel caso in cui i dati disponibili per la creazione del modello appartengano ad un'unica classe.

Il classificatore ad una classe ha lo scopo di identificare oggetti provenienti da una classe specifica rispetto a tutti gli altri oggetti, e lo fa attraverso l'apprendimento su un set di oggetti di allenamento appartenenti ad un'unica classe. Questo classificatore può essere visto come un classificatore SVM standard a due classi, dove tutti gli oggetti utilizzati per l'allenamento appartengono ad una sola classe e solo l'origine rappresenta l'unico oggetto appartenente alla seconda classe. Questa tecnica utilizza i vettori di supporto (SV, [Hastie et al., 2001]). I SV sono le osservazioni identificate come strutturali necessarie alla costruzione del classificatore: lo stesso classificatore verrebbe costruito allo stesso modo se tutte le osservazioni a parte i support vector venissero cancellati dal dataset (Appendice B).

Data una serie di campioni  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ ,  $\mathbf{x}_i \in \mathbb{R}^p$  si vuole stimare un supporto  $S$  che contenga la maggior parte dei dati generati dalla distribuzione  $P$  e sia il più piccolo possibile. Il supporto individuato deve definire una regione piccola, con una forma semplice che contenga la maggior parte dei dati di training ma allo stesso tempo non dipenda troppo da questi, deve cioè generalizzare il problema. Inoltre l'intento, come già sottolineato, non è quello di stimare tutta la densità di probabilità ma solamente la regione in cui è disposta una frazione pre-specificata dei dati di training.

L'algoritmo OCSVM mappa i dati in uno spazio di dimensione maggiore:

$$\Phi : \mathbb{R}^p \rightarrow F \quad (4.1)$$

$$x \rightarrow \Phi(x) \quad (4.2)$$

Dove  $\Phi$  è una trasformazione non lineare dallo spazio dei dati  $\mathbb{R}^p$  allo spazio delle caratteristiche  $F$  (feature) che è determinato implicitamente (e usualmente non conosciuto) nei metodi che utilizzano kernel. L'algoritmo One-Class SVM calcola i support vector in  $T$  (dove con  $T$  si indicano i dati di training) attraverso un problema di ottimizzazione quadratico vincolato

#### 4.1. FORMULAZIONE

---

la cui forma primaria è:

$$\min_{w \in F, \xi_i \in \mathbb{R}^p, \rho \in \mathbb{R}} \frac{1}{2} \|w\|^2 + \frac{1}{n\nu} \sum_{i=1}^n \xi_i - \rho,$$

subject to :

$$w \times \Phi(\mathbf{x}_i) \geq \rho - \xi_i,$$

$$\xi_i \geq 0 \quad \forall i = 1, \dots, n. \quad (4.3)$$

$$(4.4)$$

Ad ogni dato è associata una variabile di slack  $\xi_i$ , che misura il grado di errore di classificazione del dato. Con  $\rho$  si indica la distanza dall'origine nello spazio delle caratteristiche,  $\omega$  è la parametrizzazione dell'iperpiano che separa l'origine dai dati in  $F$  e  $\nu \in (0, 1]$  rappresenta un limite superiore alla frazione dei dati che ricade all'esterno del supporto stimato e un limite superiore al numero di vettori di supporto. Generalmente il precedente problema di minimizzazione è risolto nella sua forma duale [Hastie et al., 2001] ricorrendo ai moltiplicatori di Lagrange, il problema duale è nella forma:

$$\min_{\alpha_i \in \mathbb{R}^n} \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j),$$

subject to :

$$0 \leq \alpha_i \leq \frac{1}{n\eta} \quad \forall i = 1, \dots, n,$$

$$\sum_{i=1}^n \alpha_i = 1, \quad (4.5)$$

dove la funzione  $k(x_i, x_j) = \langle \Phi(\mathbf{x}_i) \Phi(\mathbf{x}_j) \rangle$  è la funzione kernel. Risolvendo il problema duale per gli  $\alpha_i$  e ricostruendo  $\rho$  per il problema primario si ottiene la funzione di decisione:

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i k(\mathbf{x}_i, \mathbf{x}) - \rho, \quad (4.6)$$

un valore negativo di  $f(\mathbf{x})$  indica che  $\mathbf{x}$  è un outlier. I punti  $\mathbf{x}_i$  per i quali vale  $0 < \alpha_i < \frac{1}{n\nu}$  sono i support vector [Dreiseitl et al., 2010]. I kernel più utilizzati nell'ambito delle SVM sono:

- *d<sup>th</sup>-Degree polynomial*:  $k(\mathbf{x}, \mathbf{x}') = (1 - \langle \mathbf{x}, \mathbf{x}' \rangle)^d$ ,
- *Radial Basis*:  $k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \| \mathbf{x} - \mathbf{x}' \|^2)$ ,
- *Neural Network*:  $k(\mathbf{x}, \mathbf{x}') = \tanh(k_1 \langle \mathbf{x}, \mathbf{x}' \rangle + k_2)$ .

La scelta della funzione kernel influenza il supporto stimato. Il kernel adottato in questo caso è di tipo gaussiano; comparato con altre funzioni risulta essere più efficace nella maggior parte dei casi ed è più conveniente in quanto possiede il solo parametro  $\gamma$ .

I parametri associati all'algoritmo One-Class SVM sono dunque due:  $\nu$  e  $\gamma$  e da questi dipendono le prestazioni del classificatore; la scelta corretta dei due parametri ha una forte influenza sulla qualità del modello ottenuto. Il parametro  $\gamma$  del kernel Gaussiano ha un impatto significativo sull'accuratezza ottenuta e sulle prestazioni di generalizzazione. Diminuendo  $\gamma$  il numero di SV decresce e i contorni della funzione di decisione si rilassano e divengono larghi, allo stesso tempo diminuisce il numero di falsi positivi (dati normali classificati come anomalie) e aumenta quello di falsi negativi (anomalie classificate come dati normali). Anche il parametro  $\nu$  incide sulla forma del limite di decisione con l'aumento di  $\nu$  aumenta il numero di SV e il numero di errori di classificazione tra i dati di training. Siccome il parametro  $\nu$  è principalmente legato al numero di campioni esterni alla regione normale viene impostato ad un valore basso in modo da assicurare un valore di falsi positivi basso. Per queste ragioni scegliere un valore appropriato per il parametro  $\gamma$  rappresenta la sfida maggiore per costruire un buon classificatore OCSVM.

In problemi supervisionati, la taratura dei parametri può essere effettuata tramite tecniche di cross-validazione che, basandosi su criteri di minimizzazione dell'errore, selezionano i valori ottimi; nel contesto supervisionato tecniche di:

- grid search,
- curve caratteristiche,
- ottimizzazione (come Particle Swarm Optimisation PSO),

possono essere utilizzate.

Nel contesto semi-supervisionato considerato si utilizza l'approccio euristico proposto da Wang et al. [2012] per selezionare il parametro  $\gamma$  tramite un algoritmo che giudica il grado di descrizione della regione rispetto ai dati.

## 4.2 Algoritmo di tightness detection

Il parametro  $\gamma$  è scelto in modo da ottenere una regione né troppo larga, per mantenere la sensibilità rispetto alle anomalie, né troppo stretta attorno ai dati in modo da mantenere la generalizzazione del classificatore. L'algoritmo si basa su un approccio iterativo e la scelta di  $\gamma$  è effettuata attraverso un indice di aderenza o strettezza, *tightness*. La relazione tra la strettezza e il parametro  $\gamma$  è monotona, al crescere di gamma anche la regione diventa più stretta attorno ai dati; di seguito sono presentati i passi dell'algoritmo:

1. si determinano i valori superiore  $\gamma_u$  e inferiore  $\gamma_l$  iniziali,
2. considerato il valore intermedio  $\gamma = \frac{(\gamma_u + \gamma_l)}{2}$  per un prefissato  $\nu$ , si applica l'algoritmo OCSVM per ottenere il contorno della regione normale,
3. si valuta la strettezza del limite rispetto ai dati: se il limite risulta stretto si impone  $\gamma_u = \gamma$ , se non è né stretto né largo si interrompe l'iterazione,

#### 4.2. TIGHTNESS DETECTION

---

4. se  $\gamma_u$  e  $\gamma_l$  è inferiore ad una data soglia si interrompe l'iterazione, altrimenti torna si al passo 2.

I parametri iniziali sono scelti in modo da essere uno decisamente largo e l'altro decisamente stretto, il valore sub-ottimo si troverà nell'intervallo definito dai due valori limite.

Di seguito è riportato un metodo per la rilevazione del grado di avvolgimento della curva di decisione attorno ai dati ripreso dall'articolo di Wang et al. [2012]. Le definizioni di limite largo o stretto sono date come segue:

- una regione è definita da un limite largo se al suo interno si può disegnare una sfera sufficientemente grande nella quale non è presente alcun dato,
- al contrario è definita da un limite stretto quando il limite tra due vicini è concavo.

Partendo da questa idea l'algoritmo è implementato come segue. Data una soglia  $\theta_d$  e una funzione di decisione  $f(x)$  si definiscono:

$$\chi_1 = (\mathbf{x}_i, \mathbf{x}_j) \mid \|\mathbf{x}_i - \mathbf{x}_j\| > \theta_d, \mathbf{x}_i, \mathbf{x}_j \in T, i \neq j, \quad (4.7)$$

$$\chi_2 = (\mathbf{x}_i, \mathbf{x}_j) \mid \|\mathbf{x}_i - \mathbf{x}_j\| \leq \theta_d, \mathbf{x}_i, \mathbf{x}_j \in T, i \neq j, \quad (4.8)$$

dove  $T$  indica il training set.

1. *Controllo se i limiti sono larghi.* Si verifica per ogni coppia  $(\mathbf{x}_i, \mathbf{x}_j) \in \chi_1$  se esistono una o più coppie di punti il cui punto medio  $\bar{x}$  soddisfi:

$$f(\bar{x}) \geq 0 \quad (4.9)$$

e

$$T \cap x \mid \|x - \bar{x}\| \leq \frac{D}{2}, x \in \mathbb{R}^m = \emptyset, \quad (4.10)$$

il limite è considerato largo e quindi si va al passo successivo. Se dopo aver controllato tutte le coppie di valori il limite non viene considerato largo si va al passo successivo.

2. *Controllo se i limiti sono stretti.* Se esiste una coppia  $(x_i, x_j) \in \chi_2$  per la quale si ha:

$$f(\bar{x}) < 0, \quad (4.11)$$

il limite sarà giudicato stretto e quindi si passerà allo step 3. Se per nessuna delle coppie è soddisfatta la condizione, il limite non è giudicato stretto e si passa allo step 3.

3. *Giudizio di strettezza.* In base al giudizio raccolto nei due passi precedenti, possono verificarsi due situazioni critiche:

- il limite può essere giudicato sia largo che stretto allo stesso tempo. In questo caso si preferisce dare più peso al secondo giudizio e considerare il limite stretto.

- il limite non è ritenuto né largo né stretto. Si è ottenuto dunque un valore ritenuto ottimo.

Le condizioni (4.9), (4.11) significano rispettivamente che il punto è all'interno o all'esterno della regione. La condizione (4.10) significa che non ci sono campioni utilizzati per il training all'interno di una sfera di raggio  $D$ .

Il diametro  $D$  può essere calcolato a partire dai dati. Prima si calcola la distanza da tra ogni coppia di dato. Poi si prende la distanza minore associata ad ogni dato  $D_i^*$  in modo che:

$$D = \max D_i^*. \quad (4.12)$$

Inoltre  $D$  è utilizzato anche per valutare se i due campioni selezionati sono dei vicini, cioè come  $\theta_d$ . Allorché il limite sia largo, si scelgono due punti con distanza maggiore di  $D$ . Per verificare se il limite è stretto si scelgono due punti con distanza minore di  $D$ . L'algoritmo in pseudocodice è riportato in Algorithm 1.

```

Data:  $X = (x_1, \dots, x_n) \in \mathbb{R}^{n \times m}$ ,  $\gamma_l$  e  $\gamma_u$  iniziali, le soglie  $\delta_u$ ,  $\nu$ 
Result:  $\gamma$ 
Calcolare  $D$ ;
while  $|\gamma_u - \gamma_l| > \Delta\gamma$  do
   $\text{flag\_loose} = 0$ ,  $\text{flag\_tigth} = 0$ 
  sia  $\gamma = \frac{\gamma_u + \gamma_l}{2}$  e si costruisca il modello con OCSVM;
  for  $i = 1, \dots, n$  do
    for  $j = 1, \dots, n$  do
      if  $\text{flag\_loose} == 0$  then
         $\text{flag\_loose} = \text{Is\_tigth}(x_i, x_j, D)$ 
      end
      if  $\text{flag\_tigth} == 0$  then
         $\text{flag\_tigth} = \text{Is\_loose}(x_i, x_j, D)$ 
      end
      if  $\text{flag\_loose} == 0 \ \&\ \text{flag\_tigth} == 0$  then
         $\text{go to}$ 
      end
    end
  end
  if  $\text{flag\_tigth} == 1$  then
     $\gamma_u = \gamma$ 
  else if  $\text{flag\_loose} == 1$  then
     $\gamma_l = \gamma$ 
  else
     $\text{break}$ ;
  end
end

```

**Algorithm 1:** Algoritmo di tightness detection.

### 4.3 Analisi delle prestazioni

La capacità diagnostica di un test rappresenta la capacità di classificare come negativi i dati che sono generati veramente dalla distribuzione di riferimento (normale) e come positivi quelli invece che sono generati da altre distribuzioni e quindi devono essere considerati delle anomalie.

Per l'analisi delle prestazioni si può utilizzare una matrice di confusione, *confusion matrix*, che confronta l'output del test in esame con il risultato vero. L'etichetta vera assegnata ai dati non è in realtà conosciuta dal sistema quindi quest'analisi non può essere impiegata per migliorare le prestazioni del classificatore. In Tabella 4.1  $a$  rappresenta il numero di veri

Tabella 4.1: Matrice di Confusione

	Predizioni positive	Predizioni negative
Osservazioni positive	$a$ (VP)	$b$ (FN)
Osservazioni negative	$c$ (FP)	$d$ (VN)

positivi,  $b$  il numero di falsi negativi,  $c$  il numero di falsi positivi e  $d$  il numero di veri negativi. Il confronto fra i risultati ottenuti dal test e le etichette vere dei dati permette di definire dei parametri significativi:

- *la sensibilità*: è la probabilità che un dato anomalo risulti test positivo; l'errore del I tipo è dato da  $1 - S_e$ ,

$$S_e = \frac{a}{a + b} \quad (4.13)$$

- *la specificità*: è la probabilità che un dato normale risulti test negativo; l'errore del II tipo è dato da  $1 - S_p$ ,

$$S_p = \frac{d}{d + c} \quad (4.14)$$

- *il valore predittivo positivo*: indica la probabilità che un dato classificato come positivo sia positivo,

$$VP^+ = \frac{a}{a + c} \quad (4.15)$$

- *il valore predittivo negativo*: indica la probabilità che un dato classificato come negativo sia negativo.

$$VP^- = \frac{d}{b + d} \quad (4.16)$$

$S_e$  e  $S_p$  esprimono delle probabilità pre-test mentre  $VP^+$  e  $VP^-$  esprimono probabilità post test;  $S_e$  e  $S_p$  sono tra loro inversamente correlate.

Nei casi supervisionati, un metodo empirico utilizzabile per la scelta del classificatore è quello ad esempio di imporre un valore per  $S_p$  e calcolare  $S_e$ . Questo approccio presenta due inconvenienti: il primo è che non è detto che i valori così trovati siano ottimali per gli scopi contingenti, inoltre è difficile fare un confronto fra più test basandosi su un solo valore dei parametri. Altresì, i risultati dei valori predittivi dipendono dal rapporto tra dati positivi e negativi.

## Analisi del ROC e dell' AUC

L'analisi ROC (*Receiver Operating Characteristic*) è una metodologia applicata inizialmente per l'analisi delle immagini radar e lo studio del rapporto segnale/disturbo [E. Bottarelli, 2003]. In un test quantitativo la curva ROC rappresenta il legame tra la probabilità di ottenere un risultato vero positivo nella classe delle anomalie ( $S_e$ ) e la probabilità di ottenere un risultato falso positivo nella classe normale. Cioè vengono studiati i rapporti fra allarmi veri (hit rate) e falsi allarmi (false alarm).

La relazione tra i suddetti parametri è rappresentata sul piano cartesiano da una curva; in ordinata è riportato il tasso di veri positivi e in ascissa il tasso di falsi positivi. Se i valori sono continui si possono calcolare i valori di  $S_e$  e  $1 - S_p$  e disegnare una curva continua. Oppure si possono suddividere i valori in  $k$  intervalli ( $k$  dipendente dal numero di dati disponibili) l'unione dei punti genera una curva spezzata con andamento a scaletta. Per ottenere l'interpolazione è possibile eliminare la scalettatura e ottenere la curva basata sui parametri del data set sperimentale [E. Bottarelli, 2003].

La capacità discriminante, ossia l'attitudine a separare i dati negativi da quelli positivi è proporzionale all'estensione dell'area sottesa dalla curva ROC (AUC area under curve). L'AUC equivale alla probabilità che un'anomalia risulti positiva sia superiore rispetto a quella che risulti positivo un dato normale.

Nel caso di una classificazione perfetta (nessun falso positivo né falso negativo) la capacità discriminante è del 100%. In questo caso l'AUC passa per il valore (0,1) e l'area della curva è unitaria (100 % classificazione corretta). Come regola generale si può affermare che il punto sulla curva ROC più vicino all'angolo superiore sinistro rappresenta il miglior compromesso tra sensibilità e specificità.

Se invece la classificazione è priva di valore informativo la curva ROC sarà una diagonale e l'area  $AUC = 0.5$ . Ai fini della valutazione dell'attitudine discriminante ci sono due segmenti di scarsa importanza, le curve sovrapposte agli assi delle ascisse e delle ordinate. Infatti, esistono altri valori dei parametri che forniscono una migliore  $S_p$  senza perdita di  $S_e$  o, viceversa, una migliore  $S_e$  senza perdita di  $S_p$ . Infine è da ricordare che la valutazione di un test attraverso l'AUC viene compiuta attribuendo ugual importanza alla  $S_e$  e alla  $S_p$ , mentre in molti casi è necessario, nella pratica, dati i costi associati alle diverse tipologie di errore, differenziare il peso da assegnare ai suddetti parametri [E. Bottarelli, 2003].

## 4.4 Esempi di applicazione del metodo OCSVM

In questa sezione si presenta un esempio di classificazione tramite il metodo One-Class SVM. Per mostrare il funzionamento del metodo e fornire un'idea di come vengono scelti i parametri è presentato un esempio semplice in due dimensioni, in modo che i risultati possano essere facilmente visualizzati. Come distribuzione di riferimento per il caso normale si prende una distribuzione di Rosenbrock. La funzione di Rosenbrock è una funzione concava e viene impiegata per testare il comportamento dell' algoritmo di tightness detection nel selezionare regioni con forme allungate o non regolari. Per simulare il comportamento delle anomalie

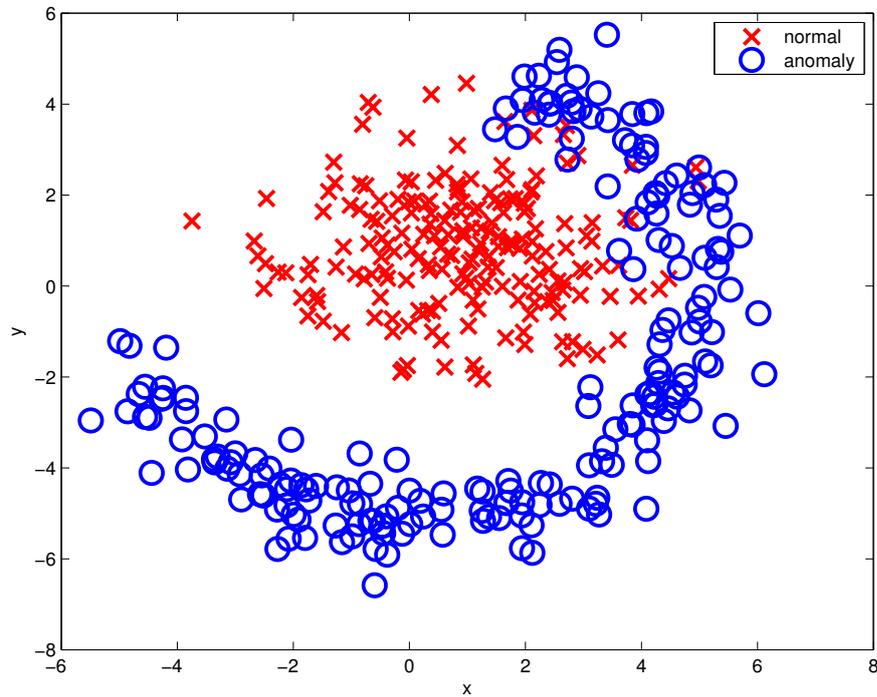


Figura 4.1: Campioni estratti da una distribuzione di Rosenbrock (in blu) e da una distribuzione normale (in rosso).

si sceglie invece una distribuzione gaussiana che si sovrappone in parte con la prima distribuzione. La scelta delle medie e delle varianze viene fatta in modo che le due distribuzioni generino punti sovrapposti parzialmente come mostrato in Figura 4.6.

### Ottimizzazione dei parametri

Si dividono i dati appartenenti alla distribuzione che descrive il comportamento normale in dati di training e dati di test. La divisione viene fatta utilizzando  $2/3$  dei dati per il training e  $1/3$  per i test. Tra i dati di training sono scelti casualmente 50 dati da utilizzare per il tuning dei parametri giovandosi dell'Algorithm 1. Le soglie superiore ed inferiore sono impostate  $\gamma_l = 0.01$  e  $\gamma_u = 5$ . Per  $\gamma = 0.01$  si considera la regione larga mentre per  $\gamma = 5$  la si considera stretta. Al primo passo l'algoritmo valuta il valore intermedio  $\gamma = 0.25$  e stabilisce se la regione è stretta oppure no. Se la regione è considerata stretta la soglia superiore viene reimpostata al valore corrente  $\gamma_l = \gamma$ , se invece è considerata larga il valore corrente viene assegnato alla soglia inferiore  $\gamma_l = \gamma$ . L'algoritmo procede fino a che non raggiunge un valore sub-ottimo oppure finché la differenza tra un valore e il successivo non è inferiore alla soglia pre-fissata.

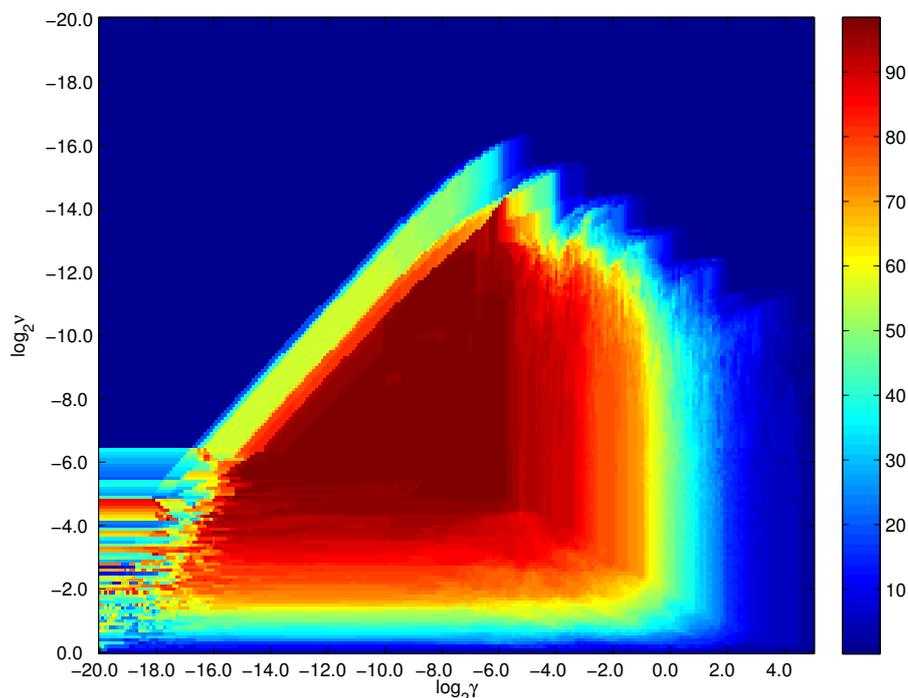


Figura 4.2: Valori dell'accuratezza ottenuta sui dati di training in funzione dei parametri del classificatore  $\gamma$  e  $\nu$ .

I due parametri dell'algoritmo da selezionare vengono impostati dopo alcune prove a  $trs = 0.05$  e  $D = 0.03$ . I risultati delle iterazioni sono riportati in Figura 4.3: nella prima figura è illustrata la regione normale individuata dal classificatore, mentre nella seconda la regione circostante ai dati è colorata in relazione alla distanza dei punti dal bordo della regione normale.

La scelta dei parametri dell'algoritmo One-Class SVM suggerita dal pacchetto LIBSVM si basa sull'analisi dell'accuratezza. In Figura 4.2 si mostra come cambia l'accuratezza del modello a seconda della scelta dei parametri.

A partire dai parametri individuati si costruisce il modello sui dati di training e si valutano le prestazioni con i dati di test. In Figura 4.5(a) sono rappresentati tramite dei cerchi i dati di training sui quali viene costruito il modello, i support vector sono colorati di nero. In Figura 4.6 sono rappresentati i dati di test con dei cerchi e le anomalie con delle crocette. I dati appartenenti al caso normale classificati correttamente dal modello sono rappresentati con cerchietti blu mentre quelli classificati erroneamente (i falsi positivi) sono rappresentati con cerchietti neri. Le anomalie classificate correttamente sono rappresentate con crocette rosse mentre gli errori (i falsi negativi) sono rappresentati con crocette verdi (e si trovano all'interno della regione normale).

In Figura 4.6(a) è riportato l'andamento del ROC per il classificatore ottenuto. Esso

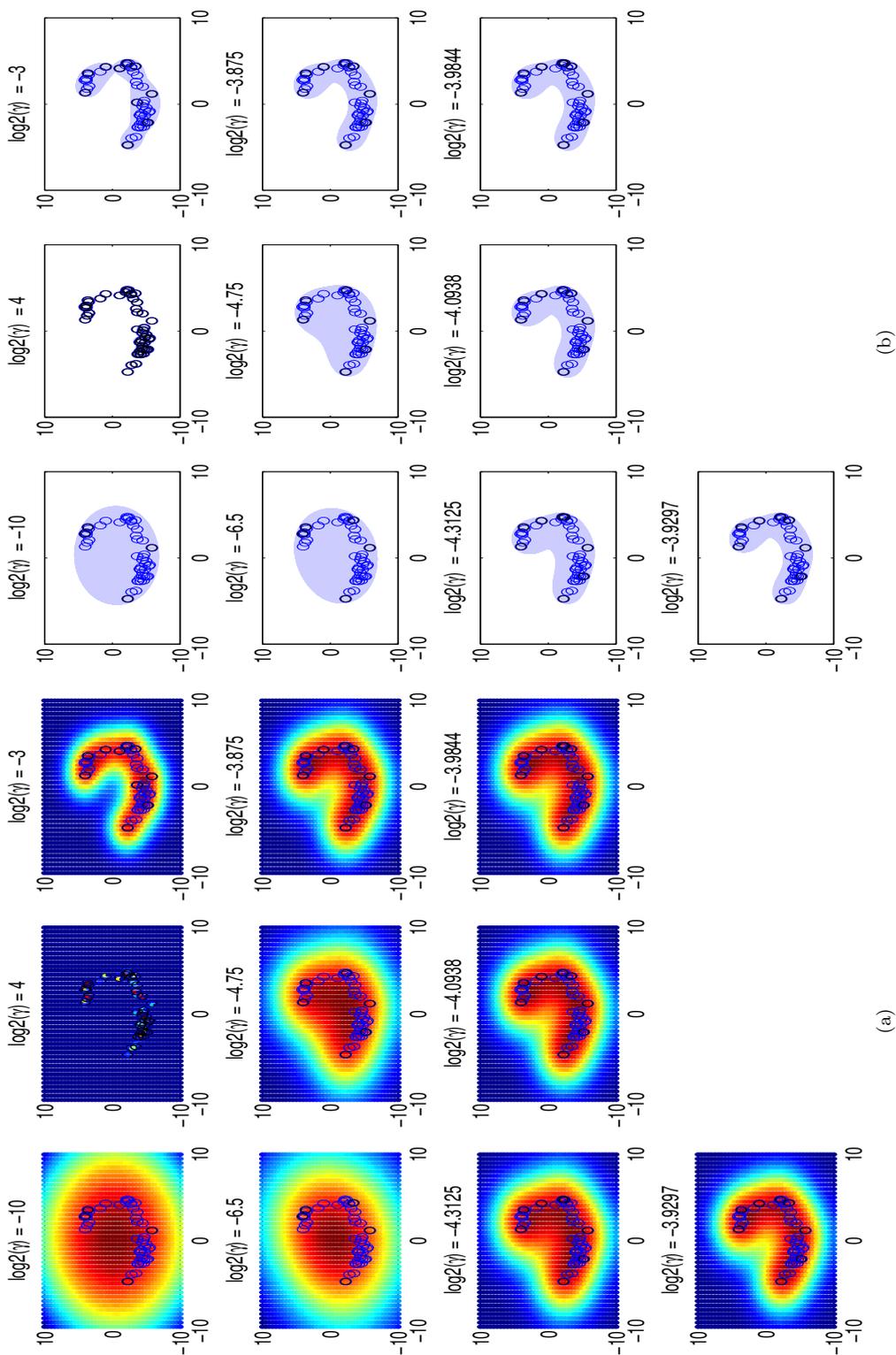


Figura 4-3: Rappresentazione della regione normale e della distanza dal bordo del classificatore, per i valori del parametro  $\gamma$  trovati dall'algoritmo durante le varie iterazioni

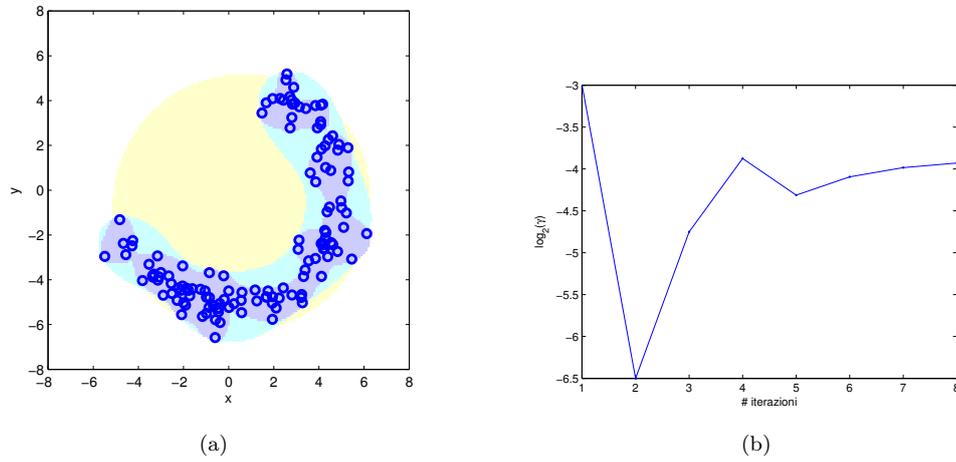


Figura 4.4: Convergenza dell'algorithmo di tightness detection.

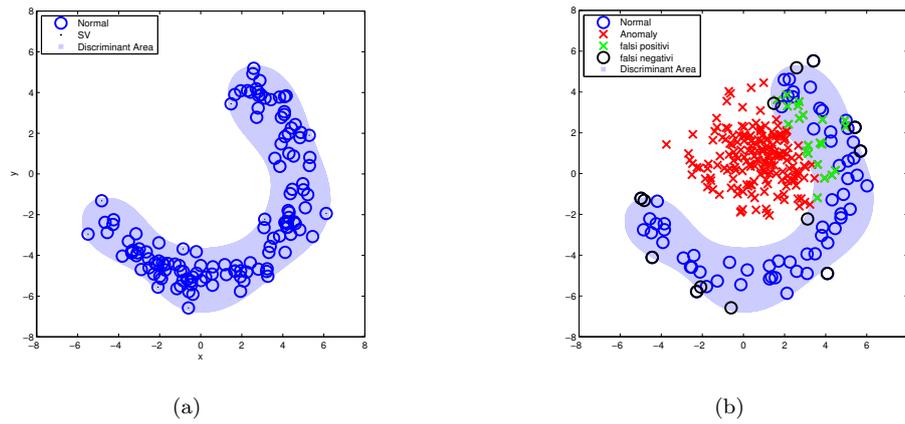


Figura 4.5: (a) Dati di training e regione normale individuata. (b) Risultato della classificazione.

differisce dall'andamento ottimale in quanto sono commessi alcuni errori di classificazione, il valore dell'AUC corrispondente è di 0.9 che sta ad indicare un test moderatamente accurato.

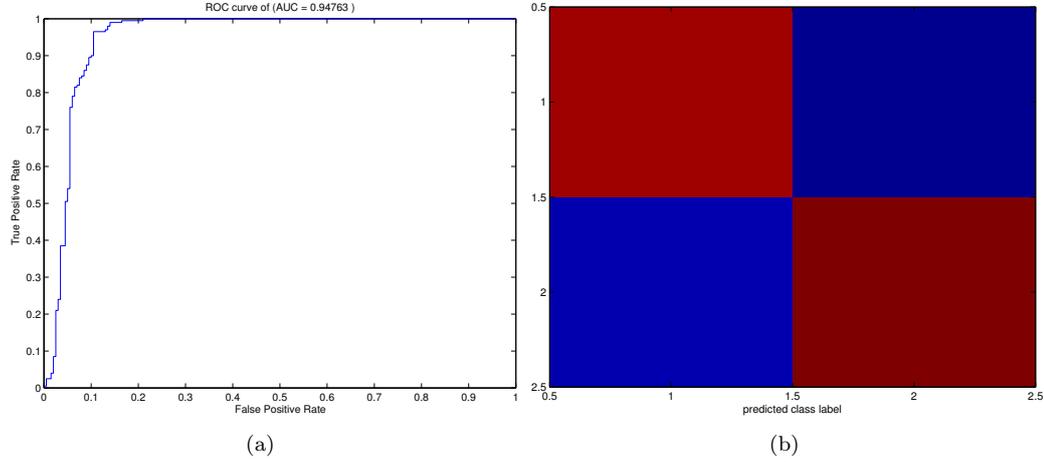


Figura 4.6: (a) Curva ROC. (b) Matrice di confusione.

## 4.5 Analisi delle componenti principali

I problemi di classificazione diventano più ardui con l'aumentare delle dimensioni dei dati. Inoltre l'elevata dimensionalità accentua il carattere di sparsità dei dati e questo determina delle difficoltà nel caso di apprendimento non supervisionato. In letteratura questo fenomeno è conosciuto come *curse of dimensionality* [Hastie et al., 2001]. La dimensionalità elevata dei dati è un problema per molti algoritmi di classificazione in quanto determina una maggiore complessità computazionale e necessita di un maggiore spazio in memoria. Inoltre elevate dimensionalità dei dati possono portare a una scarsa comprensione del modello descritto [Tan et al., 2006] [Janecek et al., 2008].

La *Principal Component Analysis* (PCA) può essere utilizzata come tecnica di riduzione delle dimensioni. La PCA è un metodo di proiezione lineare che trasforma un set di variabili in un nuovo set di variabili non correlate tra loro chiamate componenti principali (PC). La PCA si applica a dataset definiti tramite una matrice  $X$  di dimensioni  $n \times p$ , dove le  $p$  colonne sono le variabili mentre le  $n$  righe sono le osservazioni:

$$X = TP^T + E = \sum_{i=1}^l \mathbf{t}_i \mathbf{P}_i^T + E, \quad (4.17)$$

dove  $\mathbf{t}_i = X\mathbf{p}_i$ . I vettori  $\mathbf{p}_i$  sono le PC e se  $l = p$  allora  $E = 0$ . Le componenti principali sono ordinate in funzione della varianza spiegata di  $X$ : la prima componente  $\mathbf{p}_1$  può essere interpretata geometricamente come la direzione dove risiede la maggior variabilità e le altre sono direzioni ortogonali nelle alle quali è associata sempre meno varianza. La tecnica dell'analisi delle componenti principali è anche chiamata *Eigenvalue Decomposition* in quanto ogni

componente principale è associata ad un autovalore della matrice in esame, ordinati secondo la grandezza. La trasformazione indotta dalla PCA può dunque essere utilizzata per ridurre le dimensioni del problema, infatti si possono selezionare solo  $l < p$  variabili che esprimono un certo valore di variabilità del dataset di ingresso.

In questa tesi la PCA viene impiegata per raggruppare lungo le prime componenti la variabilità dominante del problema associata al comportamento del sistema attraverso i vari punti operativi che nasconde i cambiamenti dovuti alla presenza di anomalie.

Diversamente da tutti gli altri impieghi dell'analisi delle componenti principali nei quali si estraggono gli autovettori associati agli autovalori maggiori come direzioni di proiezione, si considerino gli autovettori associati agli autovalori più piccoli e in particolare lo spazio nullo degli autovalori, in quanto è lo spazio nullo che caratterizza le parti comuni dei dati [Lian, 2012].

## 4.6 Andamenti dell'errore con il variare delle componenti principali

Per analizzare le prestazioni sono stati costruiti differenti classificatori in funzione della scelta delle dimensioni dei dati in ingresso. Si sono costruiti dei classificatori  $\{f^{(i)}\}_{i=0}^{p-1}$  utilizzando come ingressi differenti spazi di dati:

$$f^{(i)}(U_i), \quad U_i = [\mathbf{p}_{i+1} \dots \mathbf{p}_p] \quad (4.18)$$

per  $i = 0, \dots, p - 1$  con  $p = l$ . Le prestazioni sono in Figura 4.7 in termini di errore di classificazione per tutti i tipi di guasto. L'errore di classificazione è definito come il rapporto tra i dati predetti erroneamente (falsi negativi e positivi) sul numero totale di dati. In ascissa è riportato il numero di PC scartate dallo spazio dei dati. Il punto più a sinistra corrisponde all'errore senza applicare la riduzione delle dimensioni. Il punto successivo corrisponde all'errore ottenuto togliendo allo spazio la componente associata all'autovalore più grande e così via. Si nota che l'andamento delle curve ha una forma a U nella maggior parte dei casi: è più difficile ottenere un errore basso quando si considerano le componenti associate a maggior varianza; si ottengono risultati migliori quando queste vengono scartate. Questo può essere considerato come un miglioramento delle prestazioni quando le componenti con maggior varianza sono scartate. La varianza delle componenti principali è dovuta principalmente alla variazione delle condizioni operative.

Questo risultato motiva l'utilizzo della PCA come strumento per ordinare la variabilità del sistema in esame e permettere di evidenziare la varianza nascosta relativa alle anomalie. Proiettando i dati lungo le direzioni associate ai più piccoli autovalori (e.g dall' 8<sup>a</sup> fino all'ultima) prima di applicare il classificatore, l'errore risulta minore rispetto a quello ottenuto senza riduzione delle dimensioni.

#### 4.6. ANDAMENTI DELL'ERRORE

La conferma che siano le prime componenti maggiormente implicate nella descrizione della variabilità del sistema dovuta al cambiamento dei punti operativi è supportata anche da un secondo tipo di esperimento. In questo esperimento lo spazio in ingresso al classificatore è modificato in maniera differente. Il nuovo classificatore  $\{g^{(i)}\}_{i=0}^{p-1}$  utilizza uno spazio di ingresso  $V_i$  differente dal precedente.

$$g^{(i)}(V_i), \quad V_i = [\mathbf{p}_1 \dots \mathbf{p}_{p-i}] \quad (4.19)$$

per  $i = 0, \dots, p-1$  con  $p = l$ . I risultati sono riportati in Figura 4.8: si nota come l'errore di classificazione cresca con l'eliminazione delle componenti principali più basse. Questo prova che le prime componenti non sono informative per il classificatore.

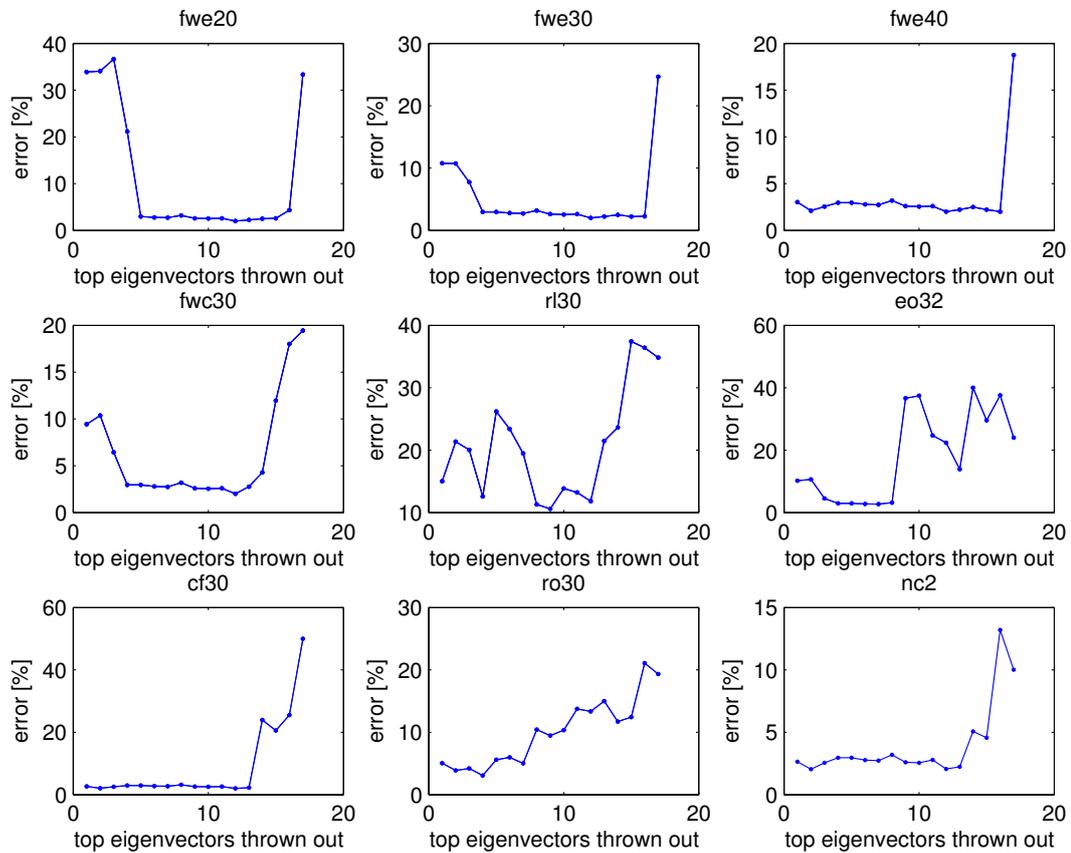


Figura 4.7: Andamento dell'errore in funzione delle PC scartate partendo da quelle associate ad autovalori maggiori.

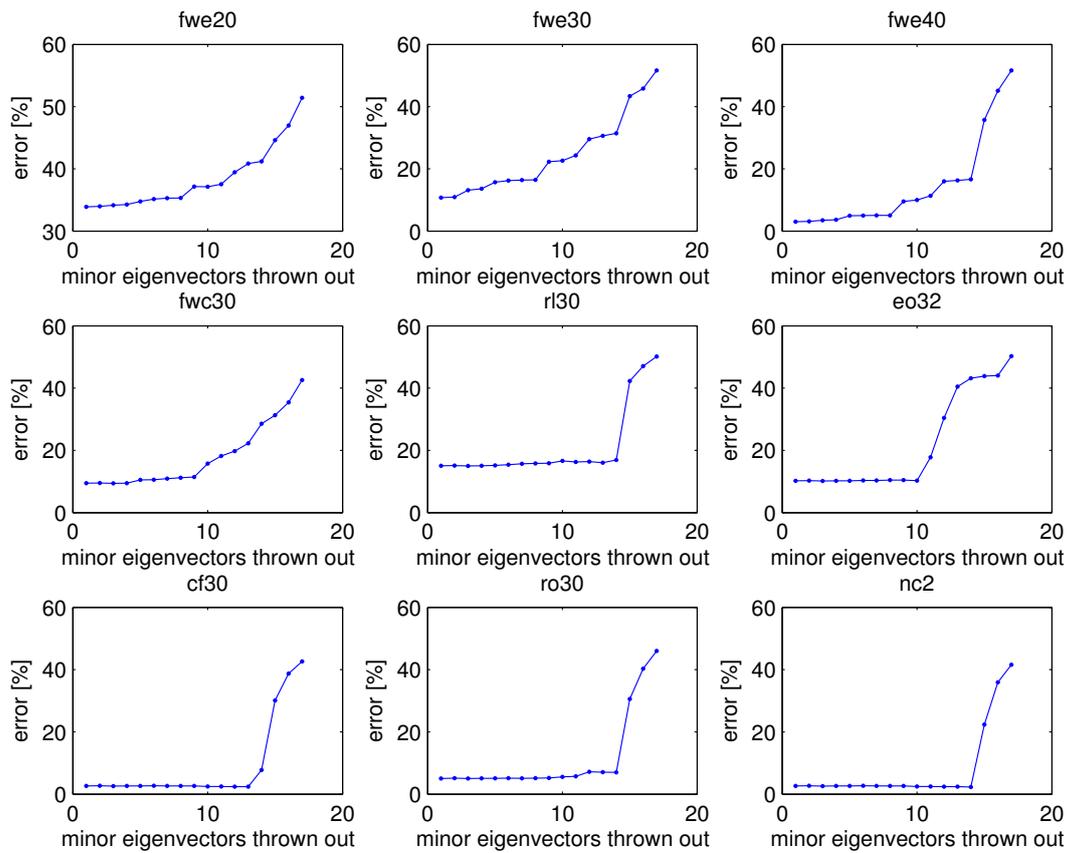


Figura 4.8: Andamento dell'errore in funzione delle PC scartate partendo da quelle associate ad autovalori minori.

## Capitolo 5

# Simulazioni e Risultati

I risultati ottenuti applicando le tecniche di classificazione ai dati del report 1043-RP allo scopo di individuare situazioni anomale sono presentati di seguito. Le simulazioni sono condotte tramite l'ausilio del software di calcolo scientifico MATLAB. I dati raccolti nel report descrivono il comportamento del sistema attraverso vari punti operativi. Quando il sistema lavora in condizioni normali il suo comportamento nominale è preso come riferimento attraverso il quale confrontare i successivi dati per individuare eventuali variazioni nel comportamento del sistema (e.g. in seguito all'insorgenza di un guasto). Avendo a disposizione le informazioni riguardanti i 27 differenti punti di lavoro raggiunti durante le simulazioni, si sfrutta l'informazione portata dalle 15 variabili selezionate per l'analisi in modo da descrivere e caratterizzare il supporto dei dati riferiti al comportamento normale.

A questo scopo si impiega il classificatore One-Class SVM descritto nel capitolo precedente, utilizzando i parametri ottenuti a seguito di una fase di tuning. Tramite questo modello è possibile valutare il comportamento che il chiller esibisce durante le prove effettuate. I guasti analizzati sono: la riduzione della portata di acqua all'evaporatore e al condensatore, la perdita e la sovraccarica di refrigerante, l'eccesso di olio, lo sporco del condensatore e la presenza di materia non condensabile nel refrigerante.

Ad ogni guasto è riservata una sezione dove sono presentati i risultati della classificazione, per i vari livelli di severità, attraverso alcuni indici di prestazione come la curva ROC, la matrice di confusione e la distanza dalla regione di classificazione. Come spiegato nei capitoli precedenti i dati originali presenti nei dataset vengono filtrati per renderli idonei all'analisi e per rimuovere eventuale rumore presente. In seguito si effettua una selezione delle variabili e una riduzione delle dimensioni attraverso la tecnica della PCA.

### 5.1 Dati che caratterizzano il comportamento normale del sistema

I test svolti in condizioni normali sono presi come riferimento (benchmark) per caratterizzare il comportamento del chiller durante il funzionamento nominale: per ogni guasto sono

indicati i test di riferimento con i quali è lecito eseguire il confronto (Tabella 2.2). Per descrivere il procedimento utilizzato per la costruzione del classificatore si usa il test normal, per gli altri benchmark valgono le stesse considerazioni fatte su questo dataset (il dataset normal è considerato un benchmark per la riduzione delle portate di acqua al condensatore e all'evaporatore).

Il dataset originale è costituito da 5191 campioni ottenuti ad intervalli di 10 secondi circa l'uno dall'altro. Come già visto nel Capitolo 2, i dati sono selezionati corrispondono ad istanti di tempo in cui il sistema ha un comportamento stazionario (si escludono i transitori).

I dataset utilizzati per le simulazioni sono costituiti solamente dalle feature selezionate per l'analisi (Tabella 2.10), scelte in base a dei criteri che garantiscono una descrizione adeguata del sistema. In Figura 5.4 sono riportate le 15 feature utilizzate in funzione del carico all'evaporatore (in blu).

Una volta selezionate le feature si rende necessaria un'ulteriore riduzione delle dimensioni. Questa riduzione è ottenuta con l'applicazione della PCA. Prima di procedere con la proiezione dei dati provenienti dal dataset normal nello spazio PCA è opportuno compiere una standardizzazione delle variabili. Questa operazione permette di riportare ogni variabile ad un valore comparabile con le altre indipendentemente dai valori assunti da ogni grandezza.

A partire dai dati normali si costruisce il classificatore OCSVM. Un buon classificatore deve risultare il più robusto possibile rispetto ai disturbi: il modello che descrive il comportamento normale del sistema è costruito in modo da non considerare una percentuale dei punti che, per varie ragioni, si discostano maggiormente dall'andamento nominale, i cosiddetti outlier.

## Lo spazio generato dalla trasformazione PCA

La trasformazione indotta dipende solamente dal dataset normale di riferimento, e non dai dati relativi al funzionamento con guasti né da altre situazioni normali. Per individuare lo spazio PCA si sono utilizzati tutti i campioni del dataset. La Figura 5.5 illustra i dati provenienti dal dataset normal proiettati lungo le componenti generate dalla PCA. Si nota come lungo le prime componenti i dati hanno grande varianza mentre nelle ultime componenti i dati hanno varianza più piccola e il comportamento del sistema si concentra attorno a un punto.

La Figura 5.1 mostra alcuni dati relativi alla trasformazione PCA. In Figura 5.1(a) si evidenzia come le prime componenti siano quelle che maggiormente spiegano la varianza del sistema, oltre il 90% della varianza complessiva dei dati è spiegata nelle prime tre componenti. La Figura 5.1(b) esibisce i dati proiettati lungo le prime due componenti insieme alle proiezioni delle feature che contribuiscono a queste.

In Tabella 5.1 sono riportati gli autovalori associati alle componenti principali individuate dalla PCA in ordine decrescente. Si nota che gli autovalori associati alle prime tre componenti sono di un ordine di grandezza o più superiori rispetto ai successivi. La scelta di considerare solo le componenti corrispondenti alla seconda metà degli autovalori più piccoli è giustificata dal fatto che è ragionevole ritenere che le dimensioni associate descrivano in modo

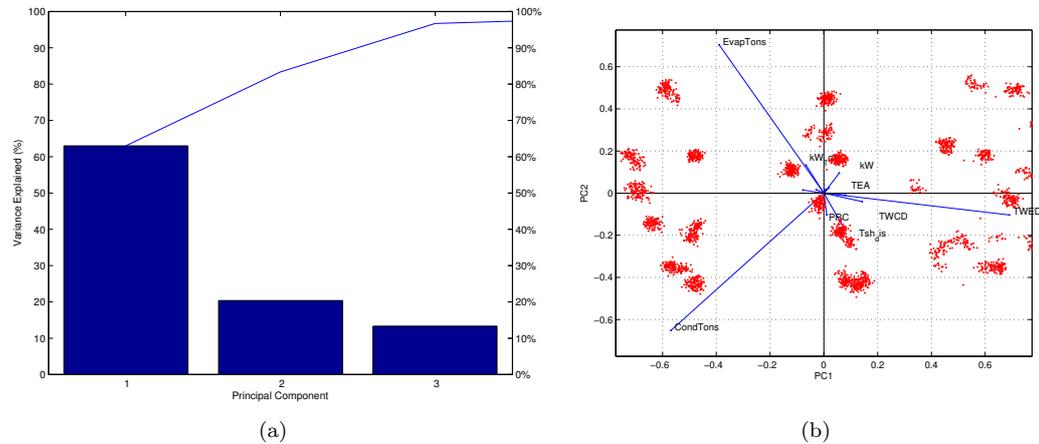


Figura 5.1: (a) Convergenza dell’algoritmo di tightness detection. (b) Dati proiettati in  $\mathbf{p}_1$  e  $\mathbf{p}_2$  e proiezione delle feature originali.

più significativo il comportamento del sistema in situazioni anomale. Durante il funzionamento normale i fattori che inducono più variazione tra le variabili del sistema sono il carico posto all’evaporatore e il cambiamento delle condizioni al contorno. Queste variazioni però sono proprie del comportamento del sistema e quindi non sono discriminanti in situazioni di guasto, con ciò si intende che queste variabili non caratterizzano il comportamento comune del sistema.

## Scelta dei parametri

I parametri del modello generato dal classificatore OCSVM sono, come già visto,  $\nu$  legato al numero di outlier e di support vector e  $\gamma$  il parametro del kernel. Il parametro  $\nu$  è scelto in modo che sia piccolo per garantire una bassa percentuale di outlier, si pone un valore  $\nu = 0.05$ , costante. Per individuare il parametro  $\gamma$  si applica l’algoritmo di tightness detection, descritto in Sezione 4.2, che opera su dati in 8 dimensioni. I valori estremi di  $\gamma$  che definiscono l’intervallo all’interno del quale è eseguita la ricerca sono  $\gamma_l = -5$  e  $\gamma_u = 5$ ; dove  $\gamma_l = -5$  è ritenuto un valore che determina una regione larga, mentre a  $\gamma_l = 5$  corrisponde una regione stretta. Le iterazioni dell’algoritmo e i valori del parametro  $\gamma$  sono riportati in Figura 5.2. In Tabella 5.2 è riportato il valore finale di  $\gamma$  e i relativi valori di accuratezza raggiunti dal modello nel caso si considerino i dati di training oppure i dati di test.

## Il classificatore in due dimensioni

Per mostrare il comportamento del classificatore, la regione di decisione e i dati classificati si presenta un classificatore bidimensionale ottenuto a partire dai dati del dataset normal. Nell’esempio si considera il guasto della portata all’evaporatore visualizzato lungo le ultime due componenti principali (quelle che hanno meno varianza).

Tabella 5.1: Autovalori associati alle componenti principali.

PC	autovalori
<b>p<sub>1</sub></b>	9.4484
<b>p<sub>2</sub></b>	3.0571
<b>p<sub>3</sub></b>	1.9979
<b>p<sub>4</sub></b>	0.1920
<b>p<sub>5</sub></b>	0.1345
<b>p<sub>6</sub></b>	0.0657
<b>p<sub>7</sub></b>	0.0384
<b>p<sub>8</sub></b>	0.0265
<b>p<sub>9</sub></b>	0.0172
<b>p<sub>10</sub></b>	0.0097
<b>p<sub>11</sub></b>	0.0066
<b>p<sub>12</sub></b>	0.0029
<b>p<sub>13</sub></b>	0.0017
<b>p<sub>14</sub></b>	0.0011
<b>p<sub>15</sub></b>	0.0004

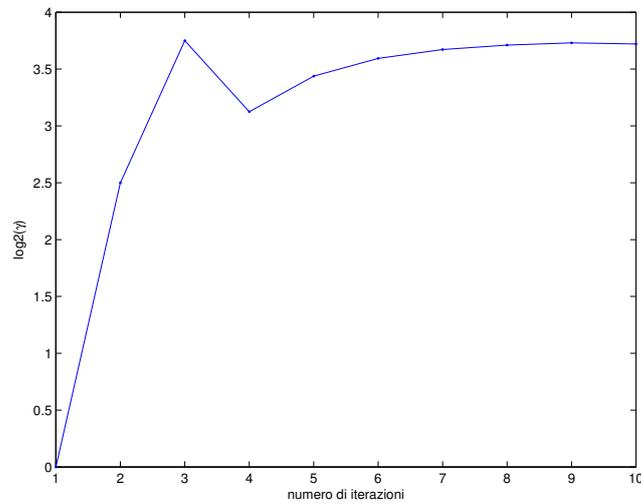


Figura 5.2: Convergenza dell'algoritmo OCSVM (dati in 8 dimensioni).

Tabella 5.2: Risultati del tuning di  $\gamma$ .

Parametro	Dati di training	Dati di test
$\log_2 \gamma = 3.6$	95%	94%

## 5.1. DATI NORMALI

I dati standardizzati provenienti dal dataset normal sono proiettati nello spazio generato dalla trasformazione indotta dalla PCA. Le proiezioni sulle ultime due componenti sono riportate in Figura 5.3(a).

Per ottenere il modello si deve prima effettuare un tuning dei parametri dell'algoritmo One-Class SVM implementato tramite il pacchetto LIBSVM. La forma della regione, il numero di SV e quello degli outlier individuati dal modello sono legati alle variazioni dei parametri dell'algoritmo  $\gamma$  e  $\nu$ . Il parametro  $\nu$ , responsabile del numero di outlier viene tenuto fisso al valore di 0.05 (fissando questo parametro si fissa un limite superiore alla percentuale degli outlier). Il parametro  $\gamma$  è invece variato, si può notare che al variare di  $\gamma$  varia la forma della

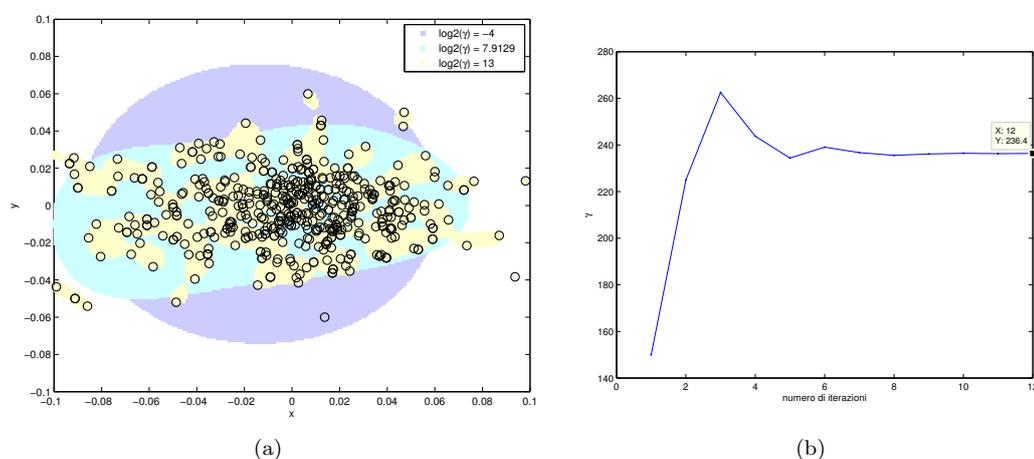


Figura 5.3: (a) Regioni individuate per tre diversi valori di  $\gamma$ . (b) Valori del parametro  $\gamma$  in funzione del numero di iterazioni dell'algoritmo di tightness detection.

regione di decisione, in particolare all'aumentare del parametro la regione si stringe attorno ai punti normali determinando un overfitting. Al contrario, diminuendo il valore di  $\gamma$  la regione si rilassa e perde la capacità di descrivere opportunamente la regione interessata dai dati normali. Tramite l'algoritmo di tightness detection si cerca di raggiungere un risultato sub-ottimo che determina una regione né troppo stretta né troppo larga. Il numero di dati utilizzato per le iterazioni è 400, scelti casualmente con probabilità uniforme, tra i dati di training.

I parametri dell'algoritmo sono scelti sulla base dell'ordine della distribuzione spaziale dei dati e della loro densità. L'algoritmo compie varie iterazioni prima di convergere al valore di  $\gamma = 238$  come mostrato in Figura 5.3(b).

In Figura 5.3(a) si illustrano le regioni ottenute; il primo parametro  $\log_2 \gamma = -3$  è stato scelto per dare un esempio di regione troppo larga (regione blu) mentre  $\log_2 \gamma = 13$  definisce una regione troppo stretta (regione gialla). Il parametro ottenuto dall'algoritmo è  $\log_2 \gamma = 7.9$  e costituisce una via di mezzo tra le due precedenti situazioni (regione azzurra).

## 5.2 Classificazione dei dati relativi alle anomalie

I dati riguardanti il funzionamento in caso di guasto sono raccolti nei relativi dataset. Prima di applicare il classificatore ottenuto utilizzando i dati di allenamento dei dataset riferiti al comportamento normale è necessario compiere alcune operazioni sui dati. Dal dataset originale, riferito ad un guasto con un adeguato livello di severità, si estraggono le 15 feature selezionate per l'analisi (Tabella 2.10); da ogni dataset vengono estratti i campioni corrispondenti ai 27 steady states raggiunti dal sistema tramite lo steady state detector descritto nella Sezione 2.5. Una volta ottenuti i dati in steady state vengono filtrati per rimuovere eventuali outlier.

I dati così ottenuti sono standardizzati rispetto al dataset con il quale devono essere confrontati: ai dati viene sottratta la media e vengono divisi per la varianza.

A questo punto i dati sono pronti per essere trasformati nello spazio della PCA determinato dal dataset di riferimento. I dati ottenuti sono in 15 dimensioni (come lo spazio di partenza) per i motivi spiegati in Sezione 4.6 si sceglie di utilizzare solamente la metà delle componenti ottenute, quelle che vanno dall'ottava all'ultima.

Questo procedimento è utilizzato per preparare i dati alla classificazione tramite l'impiego del metodo One-Class SVM; di seguito si presentano i risultati di classificazione ottenuti per ogni tipologia di guasto considerata.

### 5.2.1 Riduzione della portata di acqua all'evaporatore

In questa sezione si descrivono i risultati ottenuti utilizzando il classificatore OCSVM ottenuto come descritto nelle sezioni precedenti. La riduzione della portata di acqua all'evaporatore (Sezione 2.4.3) provoca un comportamento anomalo del sistema che influenza i valori assunti dalle feature durante i test di prova. Il guasto (fwe Tabella 2.2) è simulato sul sistema con quattro differenti livelli di gravità ad ognuno dei quali è associato un dataset di riferimento, come riportato in Tabella 2.4.

In Figura 5.4 sono mostrati gli andamenti delle 15 diverse feature in funzione del carico. Gli andamenti ottenuti dal sistema durante il test normal sono riportati in blu, mentre gli andamenti relativi al guasto sono riportati in rosso. Una volta ottenuti i dati standardizzati si applica la trasformazione che porta i punti dallo spazio delle feature originali allo spazio delle componenti principali.

Per dare un'idea dello spazio che si considera sono riportate in Figura 5.5 le proiezioni su due dimensioni. Non potendo riportare tutte le coppie di dimensioni si è scelto di visualizzare solo quelle consecutive (quindi  $\mathbf{p}_1$  e  $\mathbf{p}_2$ ,  $\mathbf{p}_2$  e  $\mathbf{p}_3$  e così via). Si nota che nelle prime componenti ( $\mathbf{p}_1$ ,  $\mathbf{p}_2$  e  $\mathbf{p}_3$ ) i campioni si sovrappongono considerevolmente, facendo intuire come la classificazione sarebbe ardua lungo queste dimensioni. Inoltre come già osservato il comportamento normale tende a in una regione limitata nello spazio generato dalle ultime componenti e le anomalie invece rimangono sparse e distribuite lungo alcune dimensioni.

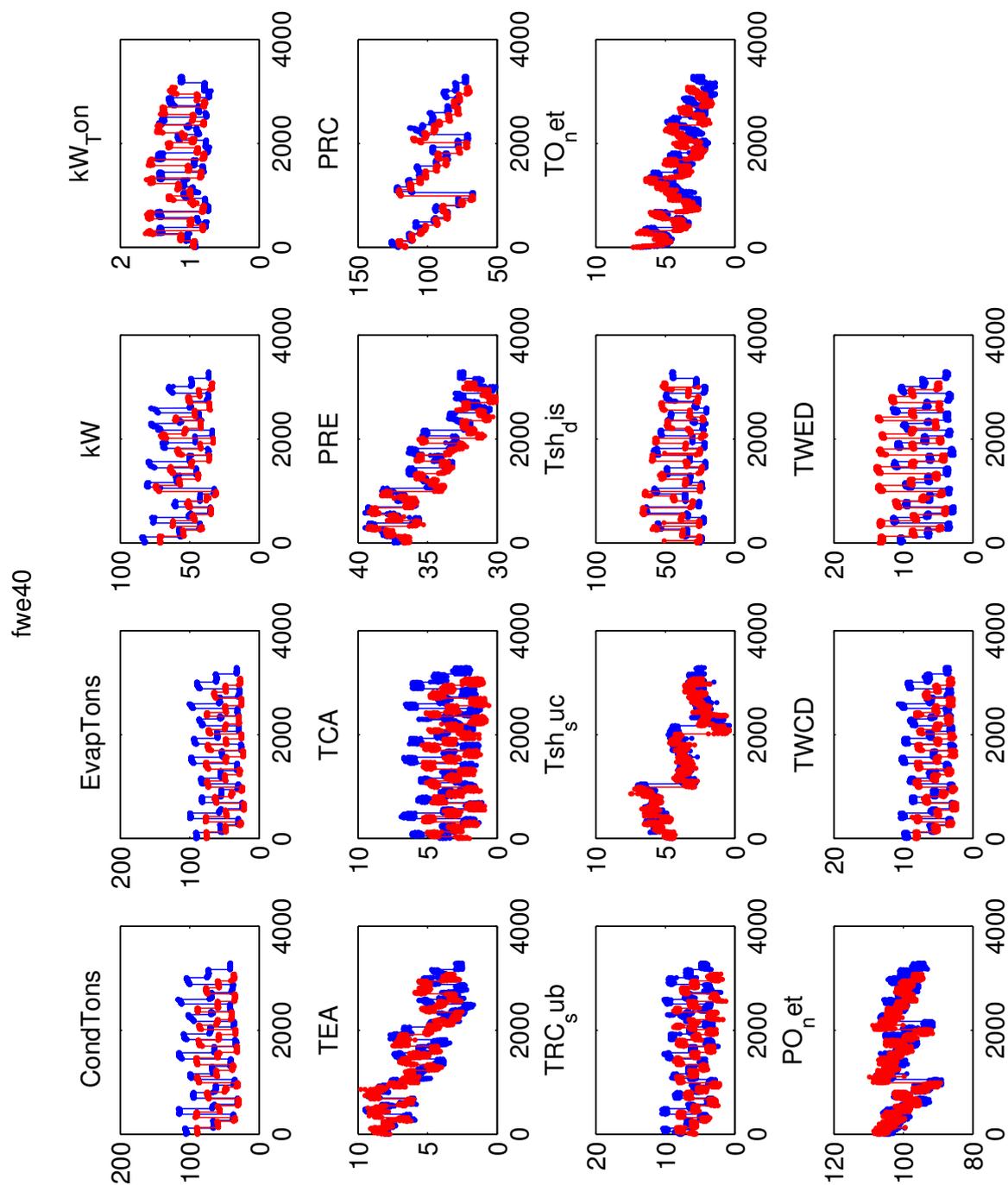


Figura 5.4: Le 15 feature originali plottate in funzione del carico, fwe (in rosso) e normal (in blu).

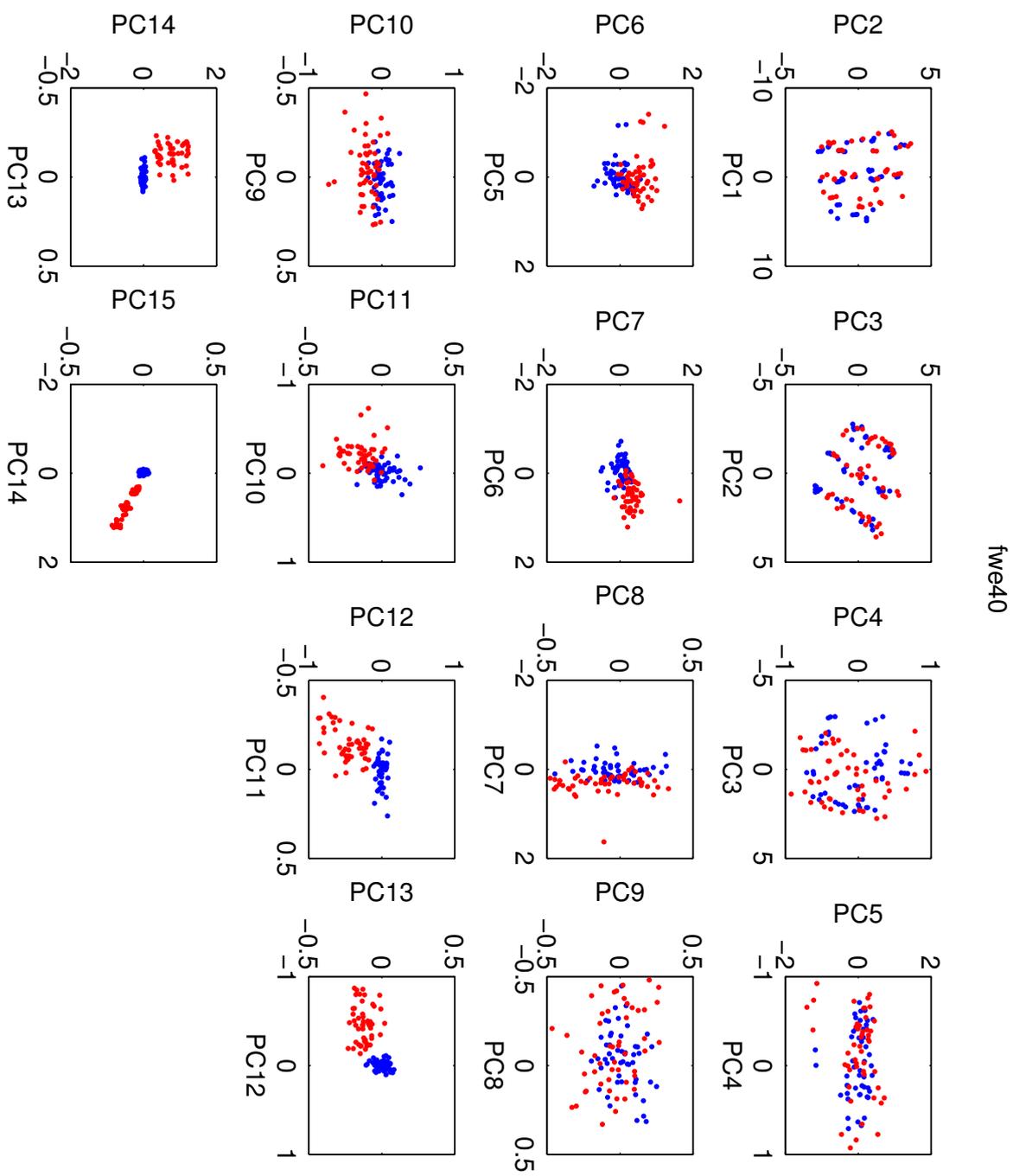


Figura 5.5: Le 15 feature ottenute dalla trasformazione PCA, fwe (in rosso) e normal (in blu).

### Le prestazioni

Il classificatore OCSVM opera in 8 dimensioni, ad ogni dato da processare fornisce un'etichetta a seconda che il punto appartenga o no alla regione normale da esso individuata. Se il punto appartiene alla regione viene classificato come normale e l'etichetta assegnata è 1, mentre se il punto ricade all'esterno della regione normale l'etichetta assegnata è  $-1$ . Di seguito sono riportati alcuni indici di prestazione utilizzati per l'analisi dei risultati:

- *Matrice di confusione.* Dato un numero di campioni appartenenti al dataset relativo al comportamento normale e lo stesso numero di campioni di un dataset relativo al guasto il risultato della classificazione può essere presentato tramite una matrice di confusione (Tabella 5.3).

Tabella 5.3: Risultati della classificazione per fwe.

Dataset	Predizioni		Accuratezza
	Positivo	Negativo	
normal	63	1035	94%
fwe10	781	317	71%
fwe20	1064	34	97%
fwe30	1098	0	100%
fwe40	1098	0	100%

- *ROC.* In Figura 5.6 si riportano gli andamenti di queste curve in riferimento ad ogni livello di guasto. Si nota che per il guasto meno severo, la prestazione è peggiore in quanto sono commessi più errori di classificazione mentre man mano che aumenta l'entità del guasto anche la curva ROC si appiattisce quindi migliorano i risultati della classificazione. Dal ROC si può ricavare il valore dell'AUC.
- *Distanza dalla regione normale.* La distanza rispetto alla regione normale viene fornita tramite un valore di decisione riferito alla classificazione. Il valore ottenuto non è da intendersi come valore effettivo della distanza ma come un modo per giudicare la bontà della classificazione. Tale valore risulta saturato dopo un certo limite di sicurezza per le anomalie. L'informazione utile per discriminare dati normali da anomalie è il segno. In Figura 5.7(b) è rappresentato il risultato della classificazione per i dati appartenenti ai dataset confrontati, normal e fwe, relativo a quattro diverse gravità. L'ascissa e l'ordinata sono le ultime due componenti principali e i punti rappresentati sono le proiezioni dei dati su queste. Il colore associato ad ogni punto indica la distanza che il punto (inteso come punto in 8 dimensioni) ha dalla regione determinata dalla regione normale individuata dal classificatore. Si vede che i dati posti vicino al punto  $(0, 0)$  hanno valore positivo e cioè appartengono alla regione di discriminazione mentre man mano che ci si allontana dal comportamento normale il valore associato ai punti è negativo e diminuisce. Tutti i punti con valore negativo sono classificati come anomalie.
- *Boxplot.* Un boxplot graficamente rappresenta i dati usando alcuni attributi di riassunto come il valore minimo ritenuto parte della distribuzione, il quartile inferiore, la mediana,

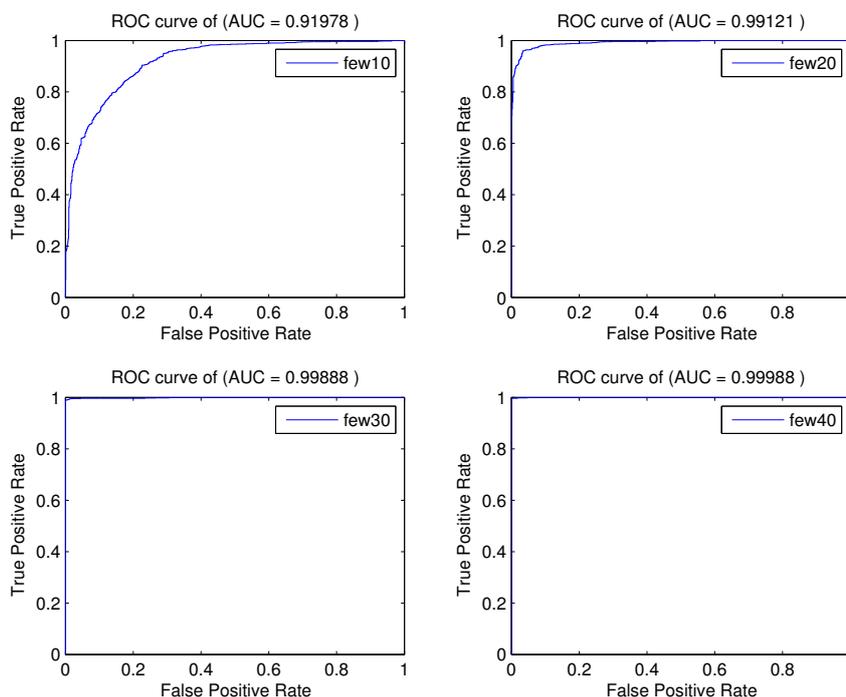


Figura 5.6: Curve ROC relative alla riduzione della portata di acqua all'evaporatore a differenti livelli di severità.

il quartile superiore e il valore massimo ritenuto parte della distribuzione; gli outlier sono riportati singolarmente. Un esempio di boxplot è riportato in Figura 5.8 dove si mostrano le proprietà statistiche dei dati relative alla riduzione della portata di acqua all'evaporatore, per differenti livelli di severità. Si nota che dopo una certa distanza dai dati viene assegnato un valore che tende a saturare e quindi i boxplot sono schiacciati sui valori negativi. Utilizzando semplicemente il valore associato alla classificazione si può comunque sfruttare l'informazione ricavata per assegnare al guasto un grado di severità e quindi sfruttare questo indice per scopi inerenti alla manutenzione predittiva.

### Classificatore in due dimensioni

Il classificatore in 8 dimensioni ha una buona capacità discriminante, ma come si vede dalla Figura 5.5, si potrebbero ottenere buone prestazioni anche considerando un classificatore solamente sulle ultime componenti principali. Vogliamo mostrare ora come risulterebbe un classificatore costruito solo per due dimensioni.

In Figura 5.9(a) sono rappresentate le proiezioni dei dati di training del dataset normal lungo le ultime componenti principali ( $\mathbf{p}_{14}$  e  $\mathbf{p}_{15}$ ). A partire da questi dati viene costruito il modello che determina la regione di decisione (regione blu). I punti che ricadono all'interno della regione sono classificati come normali mentre gli altri come anomalie. I punti in nero in figura rappresentano i support vector del sistema.

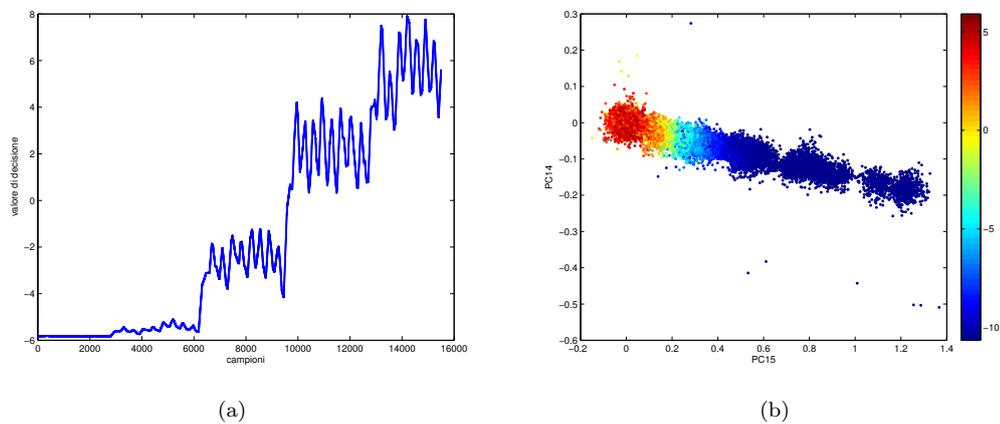


Figura 5.7: (a) Media ogni 100 campioni della distanza dei punti dalla regione. (b) Proiezione in due dimensioni dei dati normal e fwe colorati secondo la distanza dalla regione a 8 dimensioni.

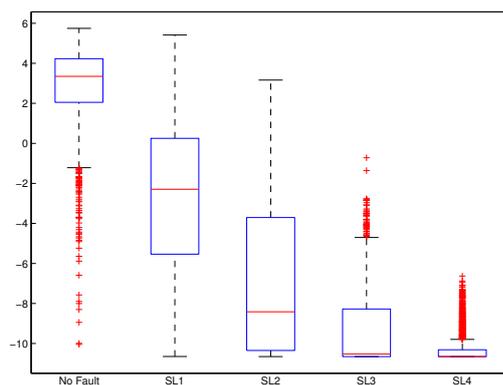


Figura 5.8: Boxplot associati alla distanza con segno corrispondenti al guasto fwe per diversi livelli di severità.

In Figura 5.9(b) sono rappresentati i dati normal (cerchietti blu) proiettati lungo le due ultime dimensioni e i dati relativi al guasto proiettati sulle stesse dimensioni (crocette rosse). La regione di decisione è colorata di blu chiaro, i punti appartenenti al dataset normal che ricadono in questa regione sono correttamente classificati mentre quelli che ricadono all'esterno sono i falsi positivi (cerchietti neri). I punti appartenenti al dataset fwe che si trovano all'esterno della regione sono classificati correttamente mentre quelli che ricadono all'interno sono falsi negativi.

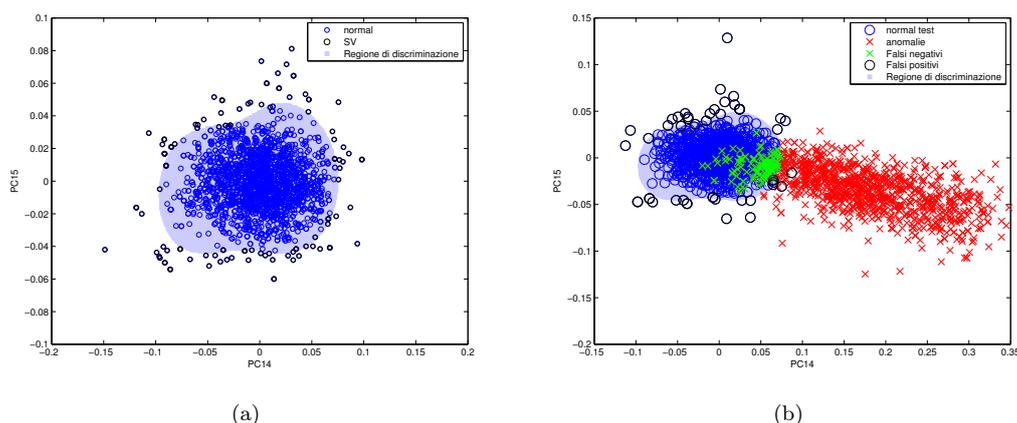


Figura 5.9: (a) Regione normale. (b) Risultato della classificazione.

Tabella 5.4: Risultati della classificazione per fwe in 2 dimensioni.

Dataset	Predizioni		Accuratezza
	Positivo	Negativo	
normal	55	1034	94%
fwe10	983	115	90%
fwe20	1098	0	100%
fwe30	1098	0	100 %
fwe40	1098	0	100%

In Figura 5.4 sono riportati i risultati della classificazione dei dati normal utilizzati per i test e dello stesso numero di dati provenienti da ognuno dei quattro dataset riferiti al guasto della portata di acqua all'evaporatore, uno per ogni livello di severità. Si nota che la classificazione è buona per tutti i livelli di intensità, l'accuratezza della classificazione si mantiene infatti sempre superiore al 90%; per i livelli di intensità più alti l'accuratezza della classificazione delle anomalie è del 100%. La classificazione comunque, non può però essere considerata perfetta in quanto l'errore di classificazione dei dati normali è del 6%.

Rispetto ai risultati ottenuti utilizzando il classificatore a 8 dimensioni, si nota che le prestazioni sono migliori in quanto, a parità di livello di severità del guasto, il classificatore in 2 dimensioni fornisce valori di accuratezza più alti. Inoltre l'impiego di un classificatore con dimensioni ridotte porta dei vantaggi a livello pratico, come la riduzione dell'onere computazionale. Da questo fatto non si può però concludere che sia preferibile l'impiego di un

Tabella 5.5: Risultato della classificazione per fwc.

Dataset	Predizioni		Accuratezza
	Positivo	Negativo	
normal	63	1035	94%
fwc10	986	112	90%
fwc20	1090	8	99 %
fwc30	1098	0	100%
fwc40	1098	0	100%

classificatore in due dimensioni in quanto non tutti i guasti potrebbero essere correttamente classificati utilizzando solamente le ultime due componenti (come confermato dalla Figura 4.7).

### 5.2.2 Riduzione della portata di acqua al condensatore

Si presentano i risultati ottenuti dalla classificazione dei dati relativi alla riduzione della portata di acqua al condensatore, fwc descritto in Sezione 2.4.2. Sono considerate quattro diverse intensità di guasto e i dataset a cui si fa riferimento sono riportati in Tabella 2.3.

Una volta che i dati sono stati standardizzati e riportati nello spazio della PCA, i campioni ottenuti sono classificati tramite il classificatore ottenuto dai campioni estratti dal dataset normal.

I risultati della classificazione sono illustrati in Tabella 5.5; si nota che la classificazione migliora con l'aumentare dell'intensità del guasto: il classificatore classifica con un'accuratezza maggiore del 90% in tutti i casi considerati, per le intensità maggiori le anomalie sono classificate con un'accuratezza del 100%.

Il ROC (Figura 5.10) riassume le prestazioni del classificatore, si mostra che con l'aumentare del livello di severità le curve sono sempre più schiacciate verso gli assi (lontane dalla diagonale passante per i punti (0,0) e (1,1)) e anche il valore dell'AUC conferma il buon risultato ottenuto per i quattro diversi livelli di severità, tutti i valori sono superiori a 0.97.

In Figura 5.11 sono riportate le proiezioni dei dati lungo le componenti  $\mathbf{p}_{11}$  e  $\mathbf{p}_{12}$ . Nella prima figura il colore rappresenta l'intensità del guasto (dal verde, la più bassa, fino al nero, la più alta); si vede che con l'aumentare dell'intensità l'anomalia si allontana dal comportamento normale e diventa più sparsa. Nella seconda figura il livello cromatico rappresenta la distanza del punto dal bordo del classificatore (ottenuta per la classificazione in 8 dimensioni). Si nota che in Figura 5.11 la regione normale (punti rossi e gialli) è larga attorno ai punti normali (punti blu in Figura 5.23(a)) questo perché si è scelto di visualizzare i dati lungo componenti con poca varianza, mentre il classificatore è costruito sui dati in 8 dimensioni tra le quali ci sono anche dimensioni con varianza maggiore; se il classificatore fosse costruito basandosi solo sulle componenti visualizzate la regione normale risulterebbe più stretta attorno ai dati normali.

I Boxplot in Figura 5.12 confermano che con l'aumentare della severità del guasto le distanze medie dei dati rispetto al bordo della regione normale aumentano, e quindi che i guasti diventano più facilmente separabili e la classificazione restituisce risultati migliori. Si

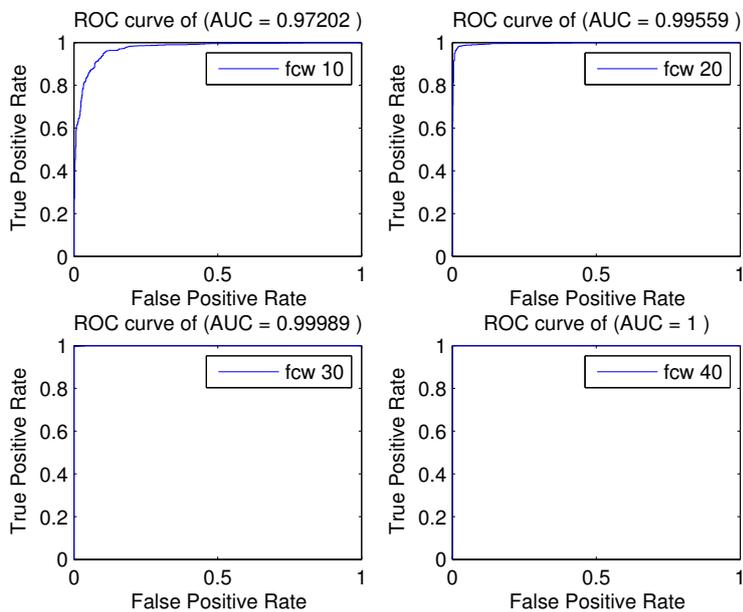


Figura 5.10: Curve ROC relative alla riduzione della portata di acqua al condensatore a differenti livelli di severità.

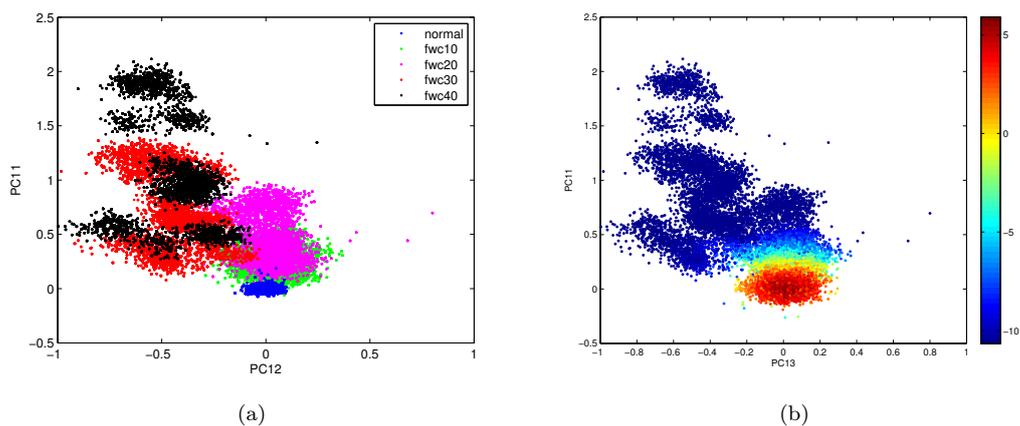


Figura 5.11: (a) Proiezione dati normal, fwc e livello di severità del guasto. (b) Distanza con segno relativa ai punti corrispondenti al guasto fwc per differenti livelli di severità

osserva che per le ultime due intensità il classificatore ritiene tutte le distanze associate ai punti negativi, ciò significa che tutti i punti sono classificati come anomalie.

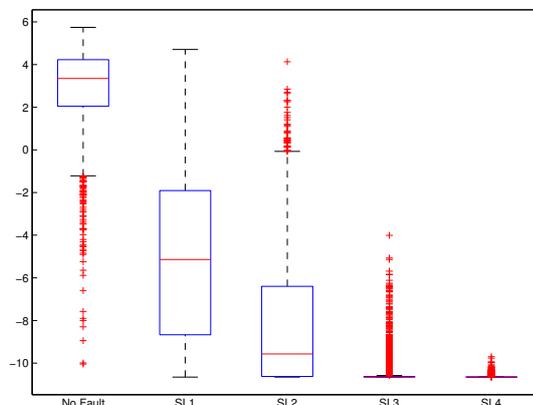


Figura 5.12: Boxplot associati alla distanza con segno corrispondenti al guasto fwc per diversi livelli di severità.

### 5.2.3 Perdita di refrigerante e sovraccarica di refrigerante

Questi due tipi di guasti sono indotti nel sistema agendo sulla stessa quantità (la carica di refrigerante), quindi sono trattati in un'unica sezione. I guasti sono descritti in Sezione 2.4.4 e in Sezione 2.4.5. Per ogni guasto si considerano quattro diverse intensità, i dataset a cui si fa riferimento sono riportati nelle Tabelle 2.5 e 2.6. Il dataset di riferimento utilizzato per la classificazione è normal r.

Per i guasti lievi si ha una bassa capacità di discriminazione (Tabella 5.6). Sia nel caso di perdita che di eccesso del refrigerante, per severità basse l'accuratezza è inferiore al 50%. In entrambi i casi, con l'aumento del livello di severità, la classificazione migliora fino ad arrivare ad oltre il 99% di classificazioni corrette.

Tabella 5.6: Risultati della classificazione per rl e ro.

data-set	Predizioni		accuratezza
	positivo	negativo	
normal r	68	918	93%
rl10	179	807	18%
rl20	418	568	42%
rl30	582	404	59%
rl40	907	79	92%
ro10	152	834	15%
ro20	311	675	32%
ro30	795	191	81%
ro40	962	24	98%

Gli andamenti delle curve ROC (Figura 5.13) confermano quanto appena osservato: per livelli di severità bassi le curve sono vicine alla diagonale, man mano che l'intensità aumenta

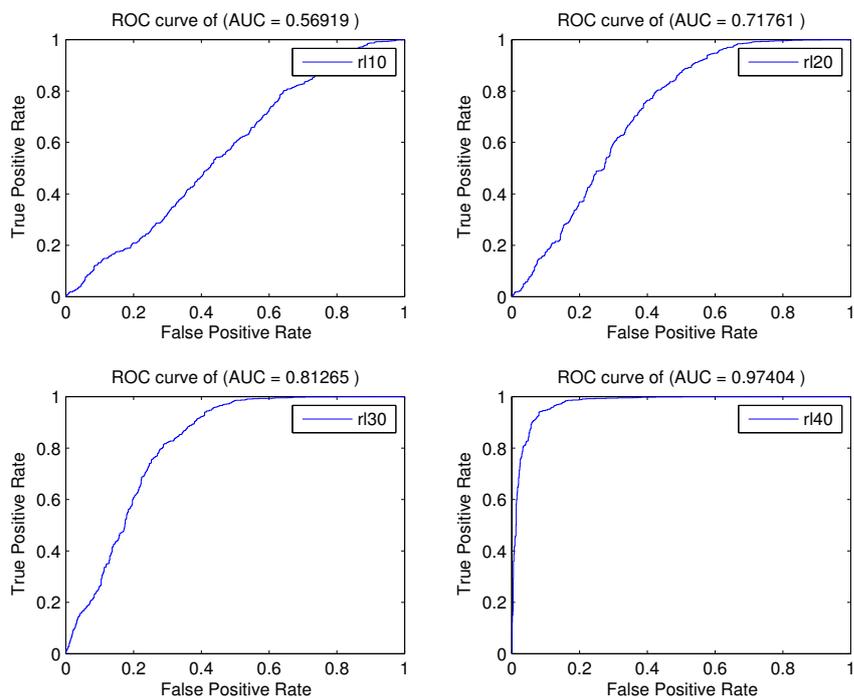
le prestazioni migliorano e le curve si alzano, i valori dell'AUC (riportati in figura) passano da 0.7 (valore che indica prestazione scarsa) a oltre 0.9 (valore che indica prestazioni buone). Si nota inoltre che le prestazioni raggiunte per il secondo guasto sono migliori rispetto a quelle raggiunte nel primo guasto.

Per visualizzare i risultati si mostrano le proiezioni dei dati nelle componenti  $\mathbf{p}_{12}$  e  $\mathbf{p}_{13}$  per il primo guasto, e nelle componenti  $\mathbf{p}_{14}$  e  $\mathbf{p}_{15}$  per il secondo guasto. Lungo queste componenti i guasti sembrano distinguersi bene dal comportamento normale. In Figura 5.14(a) si riportano le proiezioni dei dataset normal r, rl10, rl20, r30 e rl40 (Tabella 2.5) lungo le componenti  $\mathbf{p}_{12}$  e  $\mathbf{p}_{13}$ ; il colore associato ad ogni punto indica la distanza dal bordo del classificatore. Come si vede i punti centrali sono classificati come interni alla regione normale (distanza positiva) e hanno un colore rosso mentre man mano che ci si allontana dal bordo i punti vengono classificati come anomalie (distanza negativa) e il colore tende al blu. In Figura 5.14(b) sono riportate le proiezioni dei punti appartenenti ai dataset normal r, ro10, ro20, ro30, ro40 (Tabella 2.6); anche in questo caso si nota come all'aumentare dell'intensità del guasto il comportamento si discosti da quello normale e la distanza dal bordo diventi negativa (punti blu). Tramite i boxplot (Figura 5.15), si mostrano le distribuzioni delle distanze dal bordo per ogni dataset di guasto considerato. Si nota che il sistema segue un'evoluzione: parte dal livello di intensità più grave del primo guasto, passa per il comportamento normale e va al livello più grave del secondo guasto. Con l'aumentare dell'intensità la distanza media dei dati cresce (in modulo) e quindi i punti si allontanano dal bordo del classificatore. Si vede però, come già notato, che il comportamento non è perfettamente simmetrico, nel secondo caso, per le ultime due intensità si manifesta un distacco più netto dai dati normali, per questo motivo la classificazione risulta migliore.

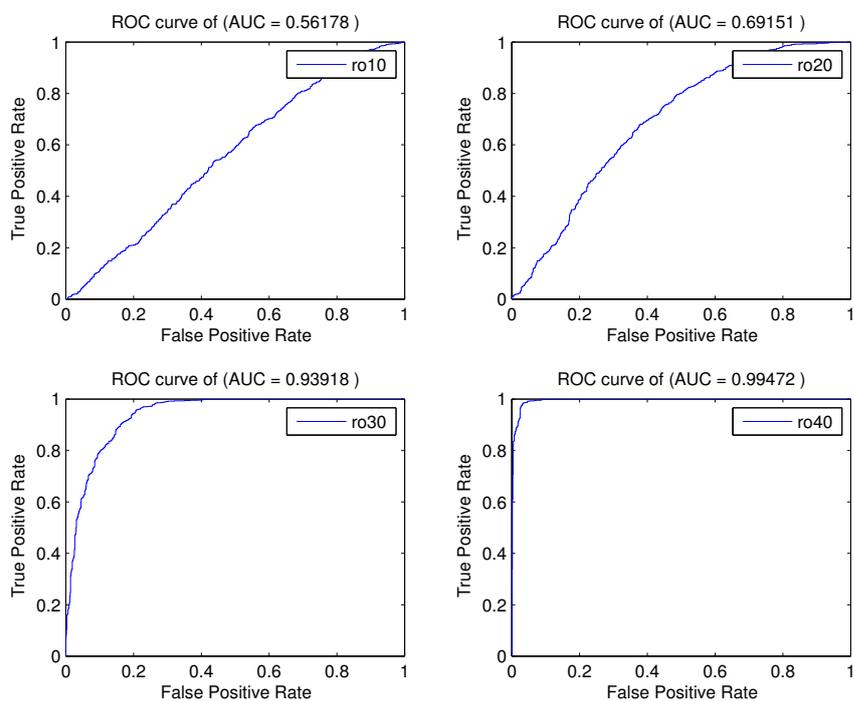
#### 5.2.4 Eccesso di olio

Di seguito si presentano i risultati ottenuti dalle simulazioni delle condizioni di guasto riferite all'eccesso di olio (descritto in Sezione 2.4.6). I test svolti sul sistema sono riferiti a quattro diversi livelli di intensità, in questa analisi però si considerano solamente i primi tre: infatti nel quarto test non si sono raggiunti tutti gli steady state e si è preferito non utilizzare i relativi dati. I data set di riferimento sono riportati in Tabella 2.7. Il test di riferimento per questo guasto è normal eo. A partire dai dati ottenuti dal data set normal eo si costruisce il classificatore OCSVM utilizzando il procedimento descritto precedentemente.

I risultati ottenuti dalla classificazione dei dati normali di test e dai dati relativi al funzionamento in caso di guasto sono riportati in Tabella 5.7. Il numero di dati considerati è pari al numero di dati di test del dataset normal eo, mentre per il calcolo dell'accuratezza si sono utilizzati lo stesso numero di dati utilizzati per il test normal eo provenienti dal dataset relativo al guasto. Il classificatore per il livello di severità basso non riesce a raggiungere un'accuratezza maggiore del 25%; mentre negli altri due casi classifica correttamente le anomalie. Questo significa che tra il primo livello e gli altri due c'è una separazione netta confermata anche dagli andamenti delle curve ROC e dai valori di AUC (Figura 5.16).



(a)



(b)

Figura 5.13: (a) Curve ROC relative alla perdita del refrigerante a differenti livelli di severità. (b) Curve ROC relative all'eccesso del refrigerante a differenti livelli di severità.

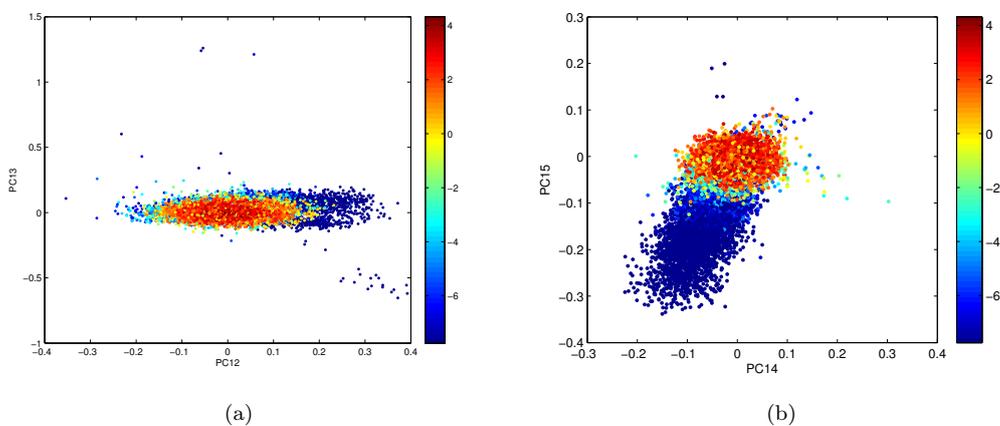


Figura 5.14: (a) Distanza con segno relativa ai punti corrispondenti al guasto rl per differenti livelli di severità (proiettati lungo  $\mathbf{p}_{12}$  e  $\mathbf{p}_{13}$ ). (b) Distanza con segno relativa ai punti corrispondenti al guasto ro per differenti livelli di severità (proiettati lungo  $\mathbf{p}_{14}$  e  $\mathbf{p}_{15}$ ).

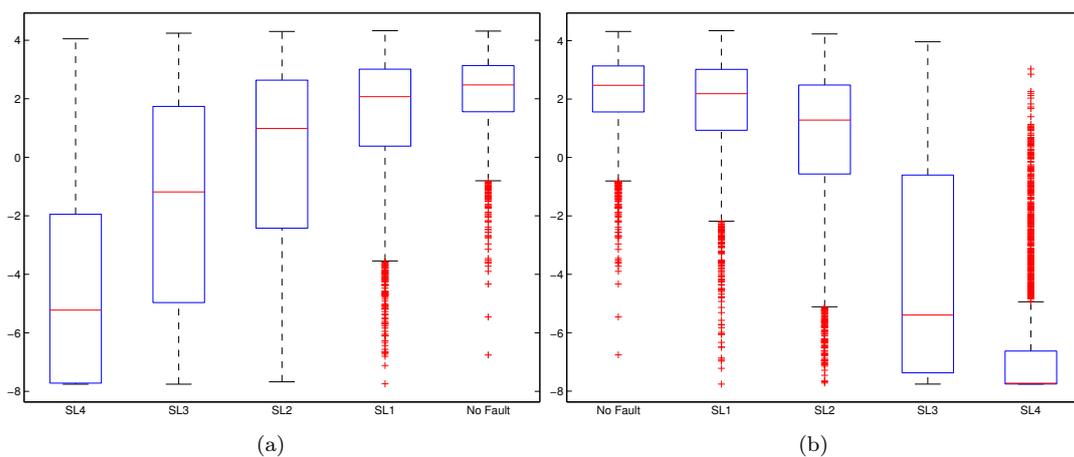


Figura 5.15: Boxplot associati alla distanza con segno corrispondenti ai guasti rl e ro per diversi livelli di severità.

Tabella 5.7: Risultati della classificazione per eo.

Dataset	Predizioni		Accuratezza
	Positivo	Negativo	
normal eo	52	972	95%
eo14	262	762	26%
eo32	1024	0	100%
eo50	1024	0	100%

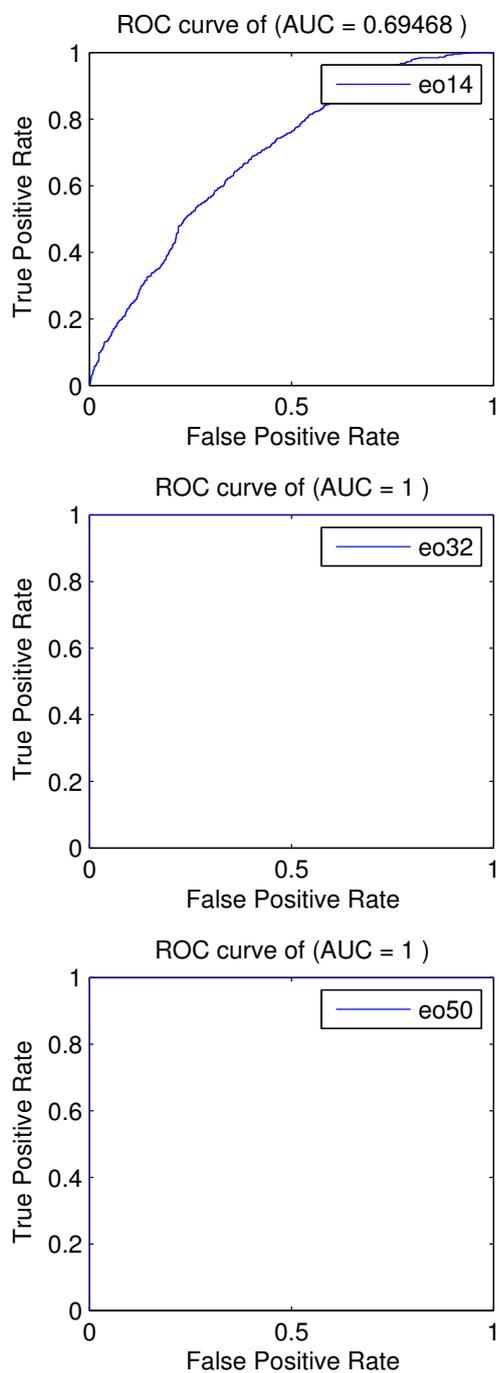


Figura 5.16: Curve ROC relative all'eccesso del olio a differenti livelli di severità.

In Figura 5.17 si mostra la distanza con segno (nello spazio 8-dimensionale) di tutti i punti di test appartenenti all'anomalia causata dall'eccesso di olio. I punti sono proiettati lungo le componenti  $\mathbf{p}_{12}$  e  $\mathbf{p}_{13}$  per poter visualizzare il risultato della classificazione. Una distanza positiva dal limite della regione di discriminazione indica che il punto in analisi appartiene alla regione normale, mentre un valore negativo classifica il dato come anomalo; con l'aumentare della distanza negativa dal bordo il livello di severità del guasto aumenta. Questo è confermato anche dalla rappresentazione tramite boxplot riportata in Figura 5.18. Si nota che con l'aumentare della severità del guasto le descrizioni statistiche tramite boxplot variano e in particolare si spostano verso valori negativi rendendo più marcata la differenza che il classificatore riscontra come distanza dal limite del classificatore. Si nota inoltre che gli ultimi due livelli sono schiacciati verso i valori negativi e questo conferma quanto notato prima.

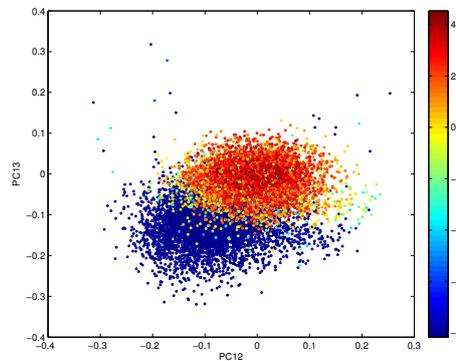


Figura 5.17: Distanza con segno relativa ai punti corrispondenti al guasto eo per differenti livelli di severità (proiettati lungo  $p_{12}$  e  $p_{13}$ ).

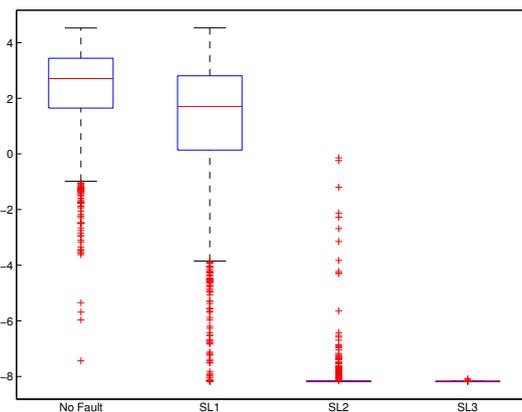


Figura 5.18: Boxplot associati alla distanza con segno corrispondenti al guasto eo per diversi livelli di severità.

### 5.2.5 Sporciamento del condensatore e materia non condensabile nel refrigerante

Si presentano i risultati ottenuti dalla classificazione per i guasti che riguardano lo sporciamento del condensatore, cf descritto in Sezione 2.4.7) e l'introduzione di materia non condensabile nel refrigerante, nc descritto in Sezione 2.4.8. Le tabelle di riferimento sono la Tabella 2.8 e la Tabella 2.9. Ogni data set viene standardizzato secondo il proprio dataset di riferimento, normal cf per il primo guasto e normal nc per il secondo. Il risultato della classificazione è riportato in Tabella 5.8. Il classificatore del guasto cf è basato sui dati normal cf mentre il classificatore del guasto nc è basato sul dataset normal nc. I campioni considerati per le classificazioni sono 983 nel primo caso e 1043 nel secondo. Per tutti i livelli di severità si raggiunge una classificazione con accuratezza superiore al 97%; anche gli andamenti ROC mostrano che in tutti i casi si ottiene una buona classificazione in quanto tutte le curve sono schiacciate verso l'alto (Figura 5.2.5). Anche guardando i Boxplot associati alla distanza (con segno) dalla regione normale si nota che sono molto separati dal dataset di riferimento, e soprattutto nel secondo caso i valori si distribuiscono tutti vicino ad un valore negativo molto distante da 0.

Tabella 5.8: Risultati classificazione per cf e nc.

Dataset	Predizioni		Accuratezza
	Positivo	Negativo	
normal cf	43	940	96%
cf6	906	77	99%
cf12	980	3	100%
cf20	955	28	97%
cf30	982	1	100%
normal nc	51	992	88%
nc1	1010	33	97%
nc2	1024	19	98%
nc3	1043	0	100%

Per visualizzare l'effetto della classificazione i punti sono proiettati lungo le componenti  $\mathbf{p}_{10}$  e  $\mathbf{p}_{11}$  per il guasto cf e lungo le componenti  $\mathbf{p}_8$  e  $\mathbf{p}_9$  per il guasto nc. La rappresentazione tramite boxplot riportata in Figura 5.22 evidenzia un distacco brusco tra il comportamento normale e il comportamento del sistema affetto dal guasto; questo è probabilmente dovuto al fatto che anche le variazioni applicate al sistema necessarie per simulare i guasto hanno un grande effetto a livello pratico sulle prestazioni dell'impianto e provocano grandi cambiamenti nelle variabili.

In Figura 5.21 sono mostrati i risultati della classificazione (in 8 dimensioni) lungo le componenti  $\mathbf{p}_{10}$  e  $\mathbf{p}_{11}$  per il guasto cf e  $\mathbf{p}_8$  e  $\mathbf{p}_9$  per il guasto nc. Come già osservato, la classificazione risulta netta (Figura 5.21): i dati normali sono principalmente rossi mentre le anomalie blu. Questo significa che in qualche componente i dati risultano marcatamente separati. Anche in Figura 5.21(b) la separazione è molto marcata ma il distacco è più graduale, questo può significare che non c'è nessuna componente che presenta una separazione

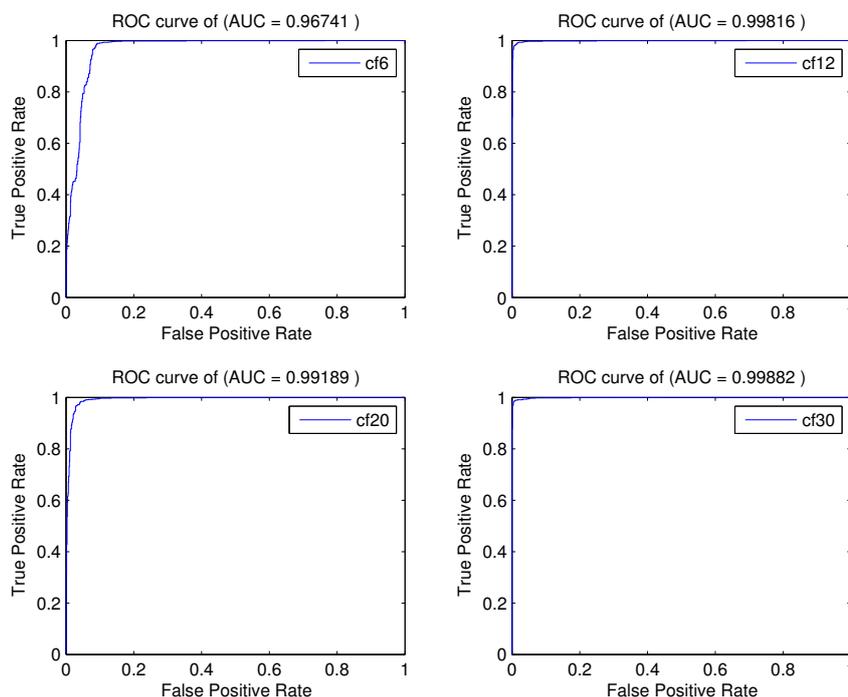


Figura 5.19: Curve ROC relative al guasto cf.

netta ma in molte componenti si manifesta un distacco graduale.

### 5.2.6 Confronto dei risultati ottenuti

La classificazione ha restituito buoni risultati per tutti i guasti considerati: il classificatore OCSVM, costruito con i dati di allenamento normali, è riuscito a classificare correttamente oltre il 90% delle anomalie riferite ai livelli di guasto più alti. Non è agevole fare un confronto tra i risultati ottenuti per i diversi tipi di guasto, infatti le cause che li generano sono differenti e i livelli di severità non hanno una comune scala di comparazione, inoltre gli effetti provocati sul sistema possono variare di intensità a seconda del guasto e non è stato sviluppato un indice comune che descrive l'effetto in termini di costi o consumo di energia per valutare l'importanza della rilevazione di ciascun tipo di anomalia. I risultati comunque confermano quanto ci si aspettava, e cioè che i guasti riguardanti le portate di acqua al condensatore e all'evaporatore sono facilmente individuabili [Comstock and Braun, 1999], mentre i guasti che interessano il circuito primario (perdita ed eccesso di liquido refrigerante) sono più nascosti dal comportamento normale. Anche nel caso in cui vi sia materia non condensabile all'interno del refrigerante il sistema reagisce prontamente ed è sensibile a piccole variazioni come risulta dalla classificazione.

Guardando i risultati ottenuti nello spazio della PCA si intuisce che lungo alcune dimensioni c'è più separazione rispetto ad altre tra guasti e comportamento normale: a seconda del guasto considerato la separazione appare più evidente lungo le componenti che più descrivono le variabili interessate dal guasto.

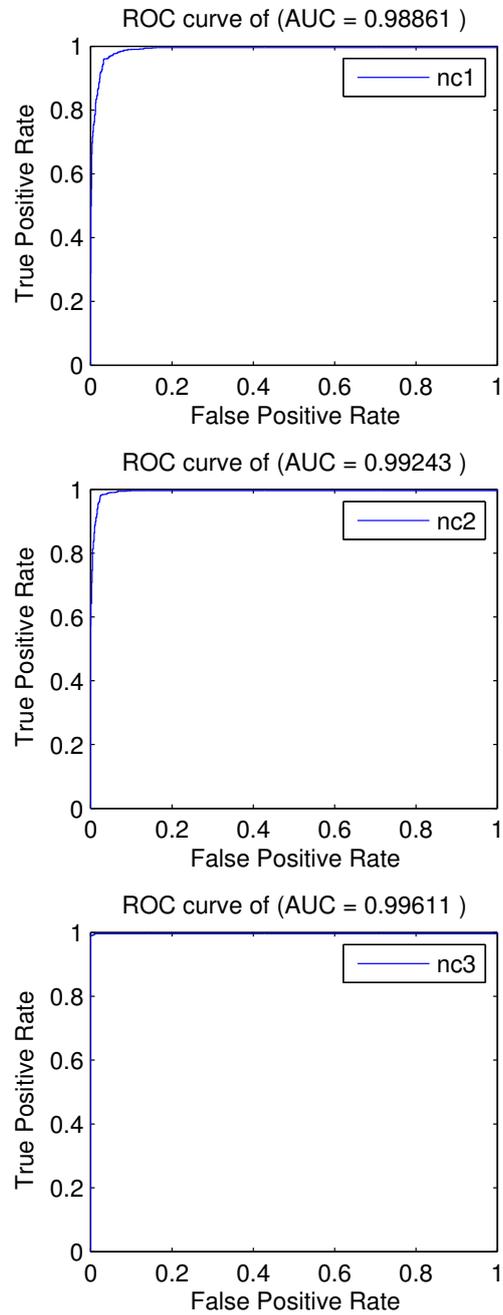


Figura 5.20: Curve ROC relative al guasto nc.

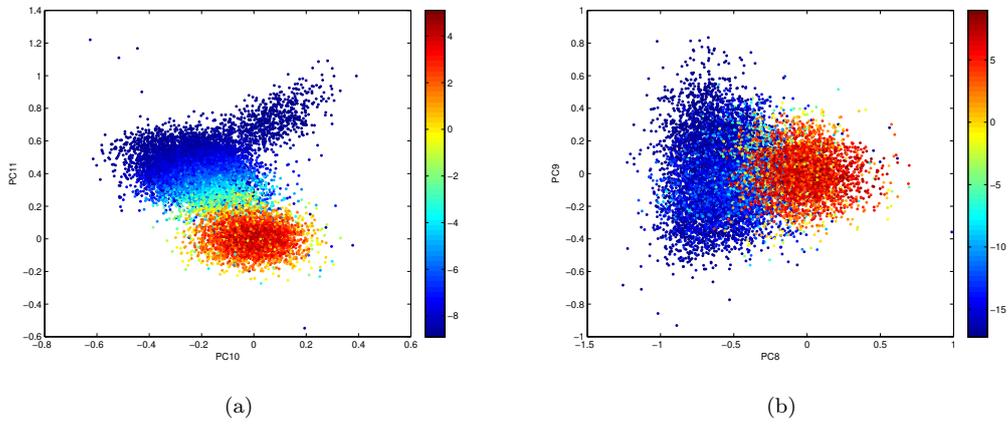


Figura 5.21: Distanza con segno relativa ai punti corrispondenti ai guasti cf e nc per differenti livelli di severità.

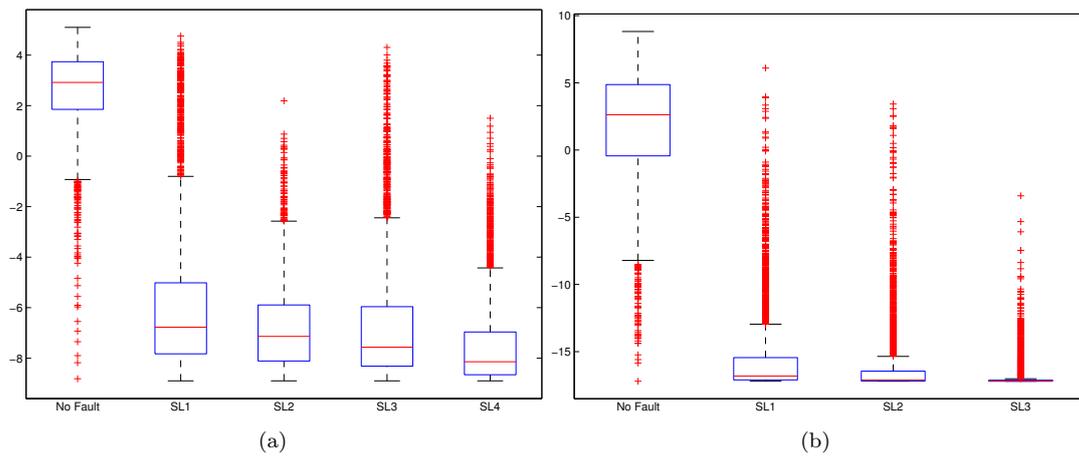


Figura 5.22: Boxplot associati alla distanza con segno corrispondenti ai guasti cf e nc per differenti livelli di severità.

Ogni guasto ha un comportamento diverso nello spazio delle componenti principali, l'evoluzione della severità crea direzioni di allontanamento dal comportamento normale la cui intensità è variabile a seconda del guasto e delle componenti. Questo fatto è evidente se si proiettano più guasti lungo coppie di componenti; ad esempio se si proiettano i dataset relativi ai guasti che modificano le portate di acqua all'evaporatore e al condensatore insieme al dataset di riferimento (normal) lungo le componenti  $\mathbf{p}_{12}$  e  $\mathbf{p}_{13}$  si ottengono due direzioni di evoluzione (Figura 5.23(a)); il guasto che si evolve lungo la parte positiva della componente  $\mathbf{p}_{13}$  è fwc mentre l'altra direzione è quella definita dal guasto fwe. Lo stesso può essere fatto per i guasti che riguardano la carica di refrigerante: in questo caso il guasto evolve lungo un'unica direzione; il guasto segue un'evoluzione, la carica del refrigerante passa da bassa (180 lbs, corrispondente a valori negativi delle componenti principali) a normale (300 lbs) ad alta (420 lbs, corrispondente a valori positivi delle componenti principali). Ogni direzione selezionata per costruire il classificatore porta dell'informazione utile per la discriminazione di qualche tipo di guasto, la capacità di una componente di discriminare il comportamento normale dal guasto dipende dalla capacità della componente di descrivere la variazione causata al sistema dal guasto in questione.

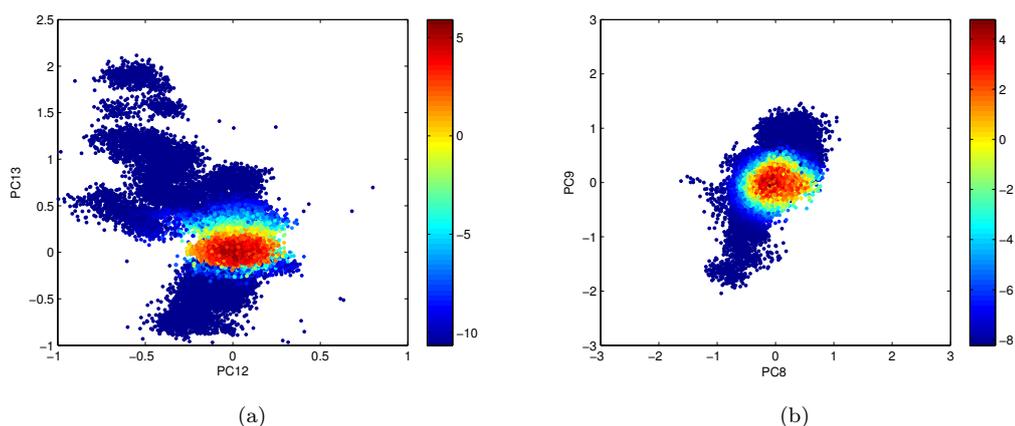


Figura 5.23: Distanza con segno relativa ai punti corrispondenti ai guasti fwe, fwc e rl, ro per differenti livelli di severità.



## Capitolo 6

# Conclusioni e Sviluppi Futuri

La manutenzione e l'efficienza energetica di impianti HVAC possono essere migliorate attraverso l'utilizzo di sistemi di fault detection. La problematica centrale nello sviluppare sistemi di fault detection efficienti sta nella difficoltà di reperire dati che diano informazione sullo stato del sistema.

In questo lavoro è stato presentato un sistema di novelty detection capace di identificare situazioni anomale senza fare uso di dati classificati a priori che portassero informazione sullo stato del sistema. Il lavoro ha mostrato che il problema dell'anomaly detection non è banale nel contesto di impianti HVAC in quanto il comportamento normale tende a nascondere i cambiamenti nei dati dovuti alle anomalie: la combinazione del metodo OCSVM per la classificazione e del metodo PCA per elidere la variabilità dovuta al funzionamento del sistema hanno dimostrato di essere validi per gli scopi di rilevazione.

La metodologia proposta ha un certo grado di generalità. Uno degli sviluppi futuri è quello di utilizzare dei dati ottenuti tramite simulazioni di un modello dinamico di un chiller centrifugo. Il comportamento del chiller potrebbe essere simulato per differenti tipologie di guasto. Dai dati ottenuti, in modo analogo a quanto fatto in questa tesi, si potrebbe sviluppare un modello per il comportamento normale e testare le prestazioni del classificatore OCSVM per un differente tipo di sistema. Per questo scopo si può impiegare il modello sviluppato in Passuello [2013]. I risultati ottenuti potrebbero essere confrontati con quelli riportati in questa tesi e il metodo potrebbe essere testato su altre tipologie di guasto per verificarne l'affidabilità.



# Appendice A

## Dati

### A.1 Dati Sperimentali dal progetto di ricerca 1043-RP

I dati utilizzati in questa tesi fanno parte di un progetto sponsorizzato da ASHRAE per lo studio di guasti nei chiller. I dati sperimentali sono necessari per costruire modelli da utilizzare per scopi di rilevamento e diagnosi di guasti nei chiller. Tutti i dati collezionati dal sistema sono elencati nelle tabelle A.3 e A.4. I valori ottenuti sono per ogni test effettuato sul sistema sono riportati in formato Excel. Le sequenze con cui sono stati eseguiti i test sono riportate in Tabella A.1 e in Tabella A.2. Molti test sono stati svolti nottetempo; la data indicata si riferisce al termine del test. Alcuni test sono stati eseguiti più di una volta a causa di problemi con la parte iniziale dei run. In molti casi i test individuali sono buoni, solamente in qualche paio di volte il sistema non ha raggiunto lo steady state atteso.

### A.2 Conversione delle misure

Le unità di misura delle grandezze utilizzate nel report 1043-RP sono espresse secondo lo standard del sistema imperiale britannico. L'analisi dei dati è stata svolta senza utilizzare conversioni. Di seguito sono riportate le formule per effettuare le conversioni al sistema internazionale di unità di misura (SI). La misura della temperatura è espressa in Fahrenheit:

$$C = \frac{F - 32}{1,8} \quad (\text{A.1})$$

La misura della potenza è espressa in Tons:

$$1W = 2,84 \times 10^{-4} \text{ Tons} \quad (\text{A.2})$$

File name	Date of Test	Comments
Defective Pilot Valve	8/19/99	The last test before the pilot valve was replaced; condenser water flow rate was not corrected
Normal DPV	8/25/99	The first test where the new pilot valve was properly adjusted; condenser water flow rate was not corrected
FWC30	9/2/99	Condenser water flow rate reduced by 30%; was able to perform test with old pump
FWC40	9/3/99	Condenser water flow rate reduced by 40%; was able to perform test with old pump
Near Normal3	9/9/99	The worst of the corrected condenser water flow rate tests; hence it was numbered #3 even though it was the first
Near Normal1	9/10/99	Test run under normal conditions; however, not all the desired operating conditions were met satisfactorily
Near Normal2	9/11/99	Test run under normal conditions; however, not all the desired operating conditions were met satisfactorily
Normal	9/12/99	The first test run where all operating conditions were met
FWC10	9/13/99	Condenser water flow rate reduced by 10%
FWC20 alt	9/14/99	Condenser water flow rate reduced by 20%; not all operating conditions met satisfactorily
FWC20	9/15/99	Condenser water flow rate reduced by 20%
Normal1	9/16/99	Another test under normal conditions; became the reference standard by meeting all the operating conditions exceptionally well
FWE10	9/17/99	Evaporator water flow rate reduced by 10%
FWE20 alt	9/18/99	Evaporator water flow rate reduced by 20%; test number 5 never reached steady state
FWE20	9/18/99	Evaporator water flow rate reduced by 20%; test performed all day long immediately after overnight test run FWE20 alt
FWE30	9/19/99	Evaporator water flow rate reduced by 30%
FWE40 alt	9/20/99	Evaporator water flow rate reduced by 40%; test number 8 never reached steady state
FWE40	9/21/99	Evaporator water flow rate reduced by 40%
Normal2	9/21/99	Another test under normal conditions created for comparison purposes
FWE20FWC20	9/22/99	Evaporator and Condenser water flow rates reduced by 20%; a compound fault
RL40 alt	9/24/99	Refrigerant charge 40% less than nominal; 5 of the 27 tests did not reach steady state
RL40	9/25/99	Refrigerant charge 40% less than nominal
RL30	9/26/99	Refrigerant charge 30% less than nominal
RL20	9/27/99	Refrigerant charge 20% less than nominal
RL10	9/28/99	Refrigerant charge 10% less than nominal
Normal R	9/29/99	Refrigerant charge at nominal (300 lbs)
RO10	9/30/99	Refrigerant charge 10% more than nominal
RO20	10/1/99	Refrigerant charge 20% more than nominal
RO30	10/2/99	Refrigerant charge 30% more than nominal
RO40	10/3/99	Refrigerant charge 40% more than nominal
Normal R1	10/5/99	Refrigerant charge at nominal (300 lbs)
Normal EO	10/6/99	Changed control program to improve ability to reach desired operating conditions

Tabella A.1: Sequenza dei test e dei nomi dei file.

A.2. CONVERSIONE DELLE MISURE

Normal B	10/7/99	Control program reversed so that tests were run in opposite sequence; data compares favorably to other normal test runs but is normally excluded from analysis because test sequence numbers do not match and may therefore cause confusion
EO14	10/8/99	Oil charge 14% more than nominal (added 3.25 lbs of oil, bringing total charge to 25 pounds)
EO32	10/9/99	Oil charge 32% more than nominal (added 4 lbs of oil, bringing total charge to 29 pounds)
EO50	10/10/99	Oil charge 50% more than nominal (added 4 lbs of oil, bringing total charge to 33 pounds)
Aborted EO86	10/11/99	Oil charge 86% more than nominal (added 8 lbs of oil, bringing total charge to 41 pounds); chiller stopped in middle of third test
EO73 alt	10/12/99	Oil charge 73% more than nominal (removed 3 lbs of oil, bringing total charge to 38 pounds); first test never reached steady state
EO68	10/13/99	Oil charge 68% more than nominal (removed 1 lb of oil, bringing total charge to 37 pounds); first test never reached steady state
Normal CF	10/14/99	Oil charge back to normal (21 pounds, 1 less than before); oil pressure regulator turned CW (1/8th turn) boosting oil pressures
CF6	10/15/99	Plugged 10 tubes in the condenser (out of 164)
CF12	10/16/99	Plugged 20 tubes in the condenser
CF20	10/19/99	Plugged 33 tubes in the condenser
CF30	10/20/99	Plugged 49 tubes in the condenser
Normal CF2	10/21/99	Unplugged all the tubes
Normal CF3	10/22/99	Another normal test run; 16th test had about a 5% reduction in condenser water flow
Normal CF4	10/23/99	Another normal test run
Normal CF5	10/24/99	Another normal test run; started soon after the previous one had finished (starting water temperatures were therefore cooler)
Normal CF6	10/25/99	Another normal test run; the fifth in a row
CF45	10/26/99	Plugged 74 tubes in the condenser
Normal NC	10/27/99	Unplugged all the tubes
Aborted NC	10/29/99	Too much Nitrogen (even after trying to purge some); stopped after 10th test due to surge
Modified NC	10/31/99	Modified low load tests to determine if complete test sequence could be completed by eliminating the surge-prone low load tests
NC5	11/3/99	Approximately 0.54 lbs Nitrogen (displacing about 5.6% of the volume at room temperature); could not reach all desired operating conditions
NC3	11/4/99	Approximately 0.22 lbs Nitrogen (displacing about 2.4% of the volume at room temperature)
NC2	11/6/99	Approximately 0.16 lbs Nitrogen (displacing about 1.8% of the volume at room temperature)
NC1	11/7/99	Approximately 0.10 lbs Nitrogen (displacing about 1.0% of the volume at room temperature)
NC Trace	11/8/99	Trace amount of Nitrogen present
NC Trace 2	11/9/99	Trace amount of Nitrogen present (some purged after previous test)

Tabella A.2: Sequenza dei test e dei nomi dei file.

<b>Designation</b>	<b>Source</b>	<b>Description</b>	<b>Units</b>
Time	VisSim	Real time counter	Seconds
TWE_set	MicroTech	Chilled water setpoint—control variable	F
TEI	JCI AHU (RTD)	Temperature of Evaporator Water In	F
TWEI	MicroTech (Thermistor)	Temperature of Evaporator Water In	F
TEO	JCI AHU (RTD)	Temperature of Evaporator Water Out	F
TWEO	MicroTech (Thermistor)	Temperature of Evaporator Water Out	F
TCI	JCI AHU (RTD)	Temperature of Condenser Water In	F
TWCI	MicroTech (Thermistor)	Temperature of Condenser Water In	F
TCO	JCI AHU (RTD)	Temperature of Condenser Water Out	F
TWCO	MicroTech (Thermistor)	Temperature of Condenser Water Out	F
TSI	JCI AHU (RTD)	Temperature of Shared HX Water In (in Condenser Water Loop)	F
TSO	JCI AHU (RTD)	Temperature of Shared HX Water Out (in Condenser Water Loop)	F
TBI	JCI AHU (RTD)	Temperature of Building Water In (in Evaporator Water Loop)	F
TBO	JCI AHU (RTD)	Temperature of Building Water Out (in Evaporator Water Loop)	F
Cond Tons	VisSim	Calculated Condenser Heat Rejection Rate	Tons
Cooling Tons	VisSim	Calculated City Water Cooling Rate	Tons
Shared Cond Tons	VisSim	Calculated Shared HX Heat Transfer (only valid with no water bypass)	Tons
Cond Energy Balance	VisSim	Calculated 1 <sup>st</sup> Law Energy Balance for Condenser Water Loop (only valid with no water bypass)	Tons
Evap Tons	VisSim	Calculated Evaporator Cooling Rate	Tons
Shared Evap Tons	VisSim	Calculated Shared HX Heat Transfer (should equal Shared Cond Tons with no water bypass)	Tons
Building Tons	VisSim	Calculated Steam Heating Load	Tons
Evap Energy Balance	VisSim	Calculated 1 <sup>st</sup> Law Energy Balance for Evaporator Water Loop	Tons
kW	JCI AHU	Watt Transducer Measuring Instantaneous Compressor Power	kW
COP	VisSim	Calculated Coefficient of Performance	--
kW/ton	VisSim	Calculated Compressor Efficiency	kW/ton
FWC	JCI AHU	Flow Rate of Condenser Water	GPM
FWE	JCI AHU	Flow Rate of Evaporator Water	GPM
TEA	MicroTech	Evaporator Approach Temperature	F
TCA	MicroTech	Condenser Approach Temperature	F
TRE	MicroTech	Saturated Refrigerant Temperature in Evaporator	F
PRE	MicroTech	Pressure of Refrigerant in Evaporator	PSIG
TRC	MicroTech	Saturated Refrigerant Temperature in Condenser	F
PRC	MicroTech	Pressure of Refrigerant in Condenser	PSIG
TRC_sub	MicroTech	Liquid-line Refrigerant Subcooling from Condenser	F

Tabella A.3: Variabili misurate e calcolate.

## A.2. CONVERSIONE DELLE MISURE

T_suc	MicroTech	Refrigerant Suction Temperature	F
Tsh_suc	MicroTech	Refrigerant Suction Superheat Temperature	F
TR_dis	MicroTech	Refrigerant Discharge Temperature	F
Tsh_dis	MicroTech	Refrigerant Discharge Superheat Temperature	F
P_lift	MicroTech	Pressure Lift Across Compressor	PSI
Amps	MicroTech	Current Draw Across One Leg of Motor Input	Amps
RLA%	MicroTech	Percent of Maximum Rated Load Amps	%
Heat Balance (kW)	VisSim	Calculated 1 <sup>st</sup> Law Energy Balance for Chiller	kW
Heat Balance%	VisSim	Calculated 1 <sup>st</sup> Law Energy Balance for Chiller	%
Tolerance%	VisSim	Calculated Heat Balance Tolerance According to ARI 550	%
Unit Status	MicroTech	Consult Table B.4 in Appendix	0 – 27
Active Fault	MicroTech	Consult Table B.3 in Appendix	0 – 44
TO_sump	MicroTech	Temperature of Oil in Sump	F
TO_feed	MicroTech	Temperature of Oil Feed	F
PO_feed	MicroTech	Pressure of Oil Feed	PSIG
PO_net	MicroTech	Oil Feed minus Oil Vent Pressure	PSI
TWCD	MicroTech	Condenser Water Temperature Delta	F
TWED	MicroTech	Evaporator Water Temperature Delta	F
VSS	JCI AHU	Small Steam Valve Position	% Open
VSL	JCI AHU	Large Steam Valve Position	% Open
VH	JCI AHU	Hot Water Valve Position	% Open
VM	JCI AHU	3-way Mixing Valve Position	% Mix
VC	JCI AHU	Condenser Valve Position	% Open
VE	JCI AHU	Evaporator Valve Position	% Open
VW	JCI AHU	City Water Valve Position	% Open
TWI	JCI AHU (RTD)	Temperature of City Water In	F
TWO	JCI AHU (RTD)	Temperature of City Water Out	F
THI	JCI AHU (RTD)	Temperature of Hot Water In	F
THO	JCI AHU (RTD)	Temperature of Hot Water Out	F
FWW	VisSim	Calculated City Water Flow Rate	GPM
FWH	VisSim	Calculated Hot Water Flow Rate	GPM
FWB	VisSim	Calculated Condenser Water Bypass Flow Rate	GPM

Tabella A.4: Variabili misurate e calcolate.

La tonnellata di refrigerazione (abbreviata anche come RT) è un'unità di misura della potenza utilizzata in alcuni paesi (specialmente in nord America) per esprimere la capacità di estrazione del calore di una macchina frigorifera. Questa unità di potenza equivale al calore di fusione assorbito per sciogliere 1 short ton (907.18474 kg) di ghiaccio a temperatura di 0°C in 24 ore <sup>1</sup>. 1Ton = 12 BTU/h = 3.517 W. La misura della portata è espressa in GPM (gallon per minute):

$$1GPM = 15853 \text{ m}^3/\text{s} \quad (\text{A.3})$$

La misura della pressione è espressa in PSIG (pound/square inch [gauge]):

$$1PSIG = 14,5 \text{ Pa} \quad (\text{A.4})$$

<sup>1</sup>[http://en.wikipedia.org/wiki/Ton\\_of\\_refrigeration](http://en.wikipedia.org/wiki/Ton_of_refrigeration)



# Appendice B

## Classificatori

### B.1 Support Vector Machines

Le macchine a vettori di supporto (SVM, dall'inglese Support Vector Machines), sono un insieme di metodi di apprendimento supervisionato per la regressione e la classificazione di pattern, sviluppati negli anni '90 da Vladimir Vapnik. Il metodo permette di separare due classi tramite un iperpiano di separazione ottimo, *optimal separating hyperplane* che divide le due classi massimizzando la distanza dei punti più vicini di ciascuna classe [Vapnik, 2000]. L'idea di base è quella di mappare i dati in ingresso in uno spazio caratteristico, *higher-dimensional feature space* nel quale sia facile individuare un iperpiano che divida i dati.

### Iperpiano di separazione ottimo

L'iperpiano di separazione ottimo divide due classi di dati e massimizza la distanza dal punto più vicino per ciascuna classe [Vapnik, 2000]. Non solo si ricava una soluzione unica al problema di separazione ma viene anche massimizzato il margine tra le due classi per i dati di allenamento. Si consideri un set di dati di allineamento composto da  $N$  coppie  $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ , dove  $x_i \in \mathbb{R}^l$  è il vettore di train e  $y_i \in \{-1, 1\}$  sono le etichette associate ai dati. Un iperpiano si definisce tramite:

$$x : f(x) = x^T \beta + \beta_0, \quad \|\beta\| = 1. \quad (\text{B.1})$$

La funzione  $f(x)$  rappresenta la distanza con segno tra il punto  $x$  e l'iperpiano, quindi per la classificazione dei punti si può utilizzare la funzione:

$$G(x) = \text{sign}[x^T \beta + \beta_0]. \quad (\text{B.2})$$

Nel caso i dati siano separabili è possibile individuare una funzione  $y_i f(x_i) > 0 \forall i$ . Si consideri il problema di ottimizzazione:

$$\max_{\beta, \beta_0, \|\beta\|=1} M \quad (\text{B.3})$$

$$\text{subject to: } y_i(x_i^T \beta + \beta_0) \geq M, \quad i = 1, \dots, N. \quad (\text{B.4})$$

Le condizioni assicurano che sono distanti almeno  $M$  dal limite di decisione definito da  $\beta$  e  $\beta_0$  e si vuole trovare  $M$  e i parametri ad esso associati. Si può rilassare l'ipotesi  $\|\beta\|=1$  rimpiazzando la condizione di vincolo con:

$$\frac{1}{\|\beta\|} y_i(x_i^T \beta + \beta_0) \geq M, \quad (\text{B.5})$$

o equivalentemente:

$$y_i(x_i^T \beta + \beta_0) \geq M \|\beta\|. \quad (\text{B.6})$$

Dato che per ogni  $\beta$  e  $\beta_0$  che soddisfano la (B.6) anche una loro riscalatura ad opera di moltiplicatori positivi rimane verificata è possibile imporre  $\|\beta\| = \frac{1}{M}$ . Quindi la (B.4) è equivalente a:

$$\max_{\beta, \beta_0} \frac{1}{2} \|\beta\|^2 \quad (\text{B.7})$$

$$\text{subject to: } y_i(x_i^T \beta + \beta_0) \geq 1, \quad i = 1, \dots, N. \quad (\text{B.8})$$

I limiti definiscono una regione vuota attorno al confine di decisione di spessore  $\frac{1}{\|\beta\|}$ , dunque si scelgono  $\beta$  e  $\beta_0$  in modo che sia massimizzato tale margine. Questo è un problema di ottimizzazione convessa. La funzione lagrangiana da minimizzare (rispetto a  $\beta$  e  $\beta_0$ ) è:

$$L_p = \frac{1}{2} \|\beta\|^2 - \sum_{i=1}^N \alpha_i [x_i^T \beta + \beta_0 - 1]. \quad (\text{B.9})$$

Uguagliando le derivate a 0 si ottiene:

$$\beta = \sum_{i=1}^N \alpha_i y_i x_i, \quad (\text{B.10})$$

$$0 = \sum_{i=1}^N \alpha_i y_i, \quad (\text{B.11})$$

che sostituite in (B.9) portano al cosiddetto duale di Wolfe:

$$L_d = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^N \alpha_i \alpha_k y_i y_k x_i^T x_k.$$

*subject to* :  $\alpha_i \geq 0.$  (B.12)

La soluzione si ottiene massimizzando  $L_D$  ovvero risolvendo un semplice problema di ottimizzazione convessa; soluzione ottenibile con un qualunque software di ottimizzazione. Il risultato deve soddisfare le condizioni di Karush-Kuhn-Tucker (KKT) che comprendono la (B.10), (B.11), (B.12 ) e:

$$\alpha y_i [x_i^T \beta + \beta_0 - 1] = 0 \quad \forall i. \quad (\text{B.13})$$

Da queste si può osservare che:

- se  $\alpha_i < 0$ , allora  $y_i(x_i^T \beta + \beta_0) = 1$  cioè  $x_i$  appartiene al confine della zona di separazione ed è denominato *support point*;
- $y_i(x_i^T \beta + \beta_0) > 1$ , allora  $x_i$  non giace sul confine e  $\alpha_i = 0$

Dalla (B.10) si vede che il vettore  $\beta$  si ottiene dalla combinazione lineare dei *support point*. Allo stesso modo  $\beta_0$  si ottiene risolvendo (B.13) per ognuno dei punti di supporto.



# Appendice C

## LIBSVM

### C.1 Utilizzo del pacchetto LIBSVM

Per l'implementazione del metodo One-Class SVM si è utilizzato il pacchetto software LIBSVM Chang and Lin [2011] per MATLAB. LIBSVM è un software per la classificazione tramite i vettori di supporto (C-SVC, nu-SVC), la regressione (epsilon-SVM e nu-SVR) e la stima di distribuzioni (one-class SVM). Le caratteristiche principali di questo pacchetto sono:

- differenti formulazioni delle SVM,
- cross-validazione per la scelta del modello,
- stima della probabilità,
- vari tipi di kernel selezionabili.

LIBSVM mette a disposizione due funzioni principali, una per costruire il modello, *svmtrain*, e l'altra per analizzare la prestazione del classificatore, *svmpredict*. I campioni del data set di riferimento vanno divisi in campioni di addestramento e campioni di test. Con i campioni di addestramento si costruisce il modello tramite il comando *svmtrain* passando come argomento oltre ai dati una stringa di parametri che definiscono il tipo di modello voluto (one-class SVM, nu-SVM) e altre opzioni riguardanti il tipo di kernel selezionato e i relativi parametri che si vogliono impostare.

Il comando *svmtrain* genera il modello e restituisce alcune informazioni come:

- *obj*: il valore ottimo della funzione obiettivo per il problema duale SVM,
- *nSV*: numero dei vettori di supporto.

Una volta costruito il modello la validazione è svolta sui dati di test tramite il comando *svmpredict* che come argomento riceve i dati con le rispettive label e il modello. Dopo aver risolto il problema di ottimizzazione l'utilizzatore può applicare la funzione di decisione per

predire le etichette dei dati di test. Siano  $\mathbf{x}_1, \dots, \mathbf{x}_n$  i dati di test  $f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)$  i valori di decisione predetti da LIBSVM. Se le etichette vere dei dati di test sono note e indicate come  $\mathbf{y}_1, \dots, \mathbf{y}_n$  possiamo valutare il risultato della predizione tramite la sua accuratezza:

$$\text{accuratezza} = \frac{\# \text{ etichette corrette}}{\# \text{ dati di test}} \times 100\% \quad (\text{C.1})$$

Oltre l'accuratezza il comando *svmpredict* restituisce le label predette dal modello, un punteggio associato a queste che indica la distanza dei dati dal limite della regione (nel caso one-class SVM) l'errore quadratico medio (RMSE) e il coefficiente di errore quadratico ( $r^2$ ) [Chang and Lin, 2011].

## Scelta dei parametri con LIBSVM

Per fare il training del classificatore l'utilizzatore deve scegliere alcuni parametri. LIBSVM fornisce un metodo semplice per provare una griglia di parametri. L'utilizzatore può fornire un intervallo di  $\nu$  (o  $\gamma$ ). Per ogni coppia di parametri appartenenti all'intervallo, LIBSVM restituisce l'accuratezza ottenuta tramite un cross-validazione e seleziona i parametri che hanno la maggiore accuratezza tramite la cross-validazione:

- separa i dati in k gruppi,
- usa k-1 gruppi per la validazione e 1 per i test,

una volta ottenuto il classificatore si usano i parametri ottenuti per costruire il modello utilizzando i dati di addestramento.

# Bibliografia

- E. Aleskerov, B. Freisleben, and B. Rao. Cardwatch: a neural network based database mining system for credit card fraud detection. In *Computational Intelligence for Financial Engineering (CIFER), 1997., Proceedings of the IEEE/IAFE 1997*, pages 220–226, 1997. doi: 10.1109/CIFER.1997.618940. 26
- F. J. Anscombe. Rejection of outliers. *Technometrics*, 2(2):123–146, 1960. doi: 10.1080/00401706.1960.10489888. URL <http://amstat.tandfonline.com/doi/abs/10.1080/00401706.1960.10489888>. 35
- Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: Identifying density-based local outliers. *SIGMOD Rec.*, 29(2):93–104, May 2000. ISSN 0163-5808. doi: 10.1145/335191.335388. URL <http://doi.acm.org/10.1145/335191.335388>. 33
- Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3):15:1–15:58, July 2009. ISSN 0360-0300. doi: 10.1145/1541880.1541882. URL <http://doi.acm.org/10.1145/1541880.1541882>. vii, 1, 27, 29, 31, 32
- Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. 91, 92
- M.C. Comstock and J.E. Braun. Development of analysis tools for the evaluation of fault detection and diagnostics in chillers. Technical report, ASHRAE Research Project RP-1043, Report HL 99-20 4036-3, 1999. 74
- C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. 37
- C. De Stefano, C. Sansone, and M. Vento. To reject or not to reject: that is the question-an answer in case of neural classifiers. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 30(1):84–94, 2000. ISSN 1094-6977. doi: 10.1109/5326.827457. 32
- Stephan Dreiseitl, Melanie Osl, Christian Scheibböck, and Michael Binder. Outlier detection with one-class svms: An application to melanoma prognosis. In *AMIA Annual Symposium Proceedings*, volume 2010, page 172. American Medical Informatics Association, 2010. 39

- S. Parodi E. Bottarelli. *Un approccio per la valutazione della validità dei test diagnostici. Le curve ROC (Receiver Operating Characteristic)*. Annali della Facoltà di Medicina Veterinaria - Università di Parma, 2003. 44
- T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer New York, 2001. 38, 39, 49
- A. Janecek, W. Gansterer, M. A. Demel, and G. F. Ecker. On the relationship between feature selection and classification accuracy. In *Journal of Machine Learning Research: Workshop and Conference Proceedings*, volume 4, pages 90–105, 2008. 49
- V. Kumar. Parallel and distributed computing for cybersecurity. *Distributed Systems Online, IEEE*, 6(10):–, 2005. ISSN 1541-4922. doi: 10.1109/MDSO.2005.53. 26
- Heng Lian. On feature selection with principal component analysis for one-class svm. *Pattern Recognition Letters*, 33(9):1027–1031, 2012. 50
- James E. Braun Mathew C. Comstock. Development of analysis tools for the evaluation of fault detection and diagnostics for chillers. Technical report, ASHRAE, 1999. 2, 4, 6, 7, 12, 13
- Emanuel Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):pp. 1065–1076, 1962. ISSN 00034851. URL <http://www.jstor.org/stable/2237880>. 35
- F. Passuello. Modellazione dinamica di un rfrigeratore di liquido per applicazioni di diagnostica guasti. Technical report, Università degli Studi di Padova, 2013. 79
- Gregory F. Piepel. Robust regression and outlier detection. *Technometrics*, 31(2):260–261, 1989. doi: 10.1080/00401706.1989.10488524. URL <http://amstat.tandfonline.com/doi/abs/10.1080/00401706.1989.10488524>. 26
- B. Schölkopf, J.C. Platt, J.C. Shawe-Taylor, A.J. Smola, and R.C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001. 37
- H.W. Stanford. *HVAC Water Chillers and Cooling Towers: Fundamentals, Application, and Operation*. CRC Press Taylor & Francis Gropu, 2012. 2, 4
- P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison-Wesley, 2006. 49
- H.S. Teng, K. Chen, and S.C. Lu. Adaptive real-time anomaly detection using inductively generated sequential patterns. In *Research in Security and Privacy, 1990. Proceedings., 1990 IEEE Computer Society Symposium on*, pages 278–284, 1990. doi: 10.1109/RISP.1990.63857. 26
- Vladimir Vapnik. *The nature of statistical learning theory*. springer, 2000. 32, 87

Huangang Wang, Lin Zhang, Yingchao Xiao, and Wenli Xu. An approach to choosing gaussian kernel parameter for one-class svms via tightness detecting. In *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on*, volume 2, pages 318–323, 2012. doi: 10.1109/IHMSC.2012.172. 40, 41