



Università
Ca'Foscari
Venezia

Corso di Laurea
magistrale
in Economia e Finanza

Tesi di Laurea

**Particle Swarm
Optimization e
Fireworks Algorithm
per l'ottimizzazione di
un Trading System**

Relatore

Prof. Claudio Pizzi

Correlatore

Prof. Marco Corazza

Laureanda

Bajame Petritaj
Matricola 861062

Anno Accademico

2019 / 2020

Sommario

1	Introduzione	1
2	Analisi Tecnica	3
2.1	I mercati finanziari moderni.....	3
2.2	Presupposti dell'analisi tecnica.....	3
2.3	I Trend	4
2.4	Le Trendline	5
2.4.1	<i>Uptrend line</i>	5
2.4.2	<i>Downtrend line</i>	6
2.5	I Canali.....	7
2.6	I livelli di Supporto e Resistenza	8
2.7	Il Volume	10
2.8	Fattori che devono essere considerati nella lettura di un grafico.....	11
2.9	Tipologia di grafici	12
2.10	Chart Pattern.....	13
2.11	Analisi Tecnica e analisi fondamentale.....	14
2.11.1	<i>Le differenze tra analisi tecnica e analisi fondamentale</i>	15
2.12	Indicatori tecnici per l'analisi tecnica	16
2.12.1	<i>La media mobile</i>	17
2.12.2	<i>Sentiment Indicators</i>	20
2.12.3	<i>Accumulation/Distribution Line</i>	21
2.12.4	<i>Average Directional Index ADX (Indice direzionale medio)</i>	21
2.12.5	<i>Moving Average Convergence/Divergence</i>	24
2.12.6	<i>Relative Strength Index (RSI)</i>	28
2.12.7	<i>Bollinger Bands</i>	29
2.12.8	<i>Average true range (ATR)</i>	30
2.12.9	<i>Money Flow Index (MFI)</i>	32
2.12.10	<i>Commodity Channel Index (CCI)</i>	33
2.12.11	<i>Stochastic oscillator(KDJ)</i>	36
2.12.12	<i>Know Sure Thing (KST)</i>	38
2.12.13	<i>Williams Percent Range (WilliamR)</i>	40
2.12.14	<i>The Ultimate Oscillator indicator (UO)</i>	41
2.13	Critiche dell'analisi tecnica	43

3	Particle Swarm optimization e Fireworks algorithms	45
3.1	Optimization Era	45
3.1.1	<i>Teoria dell'algoritmo</i>	45
3.2	L'algoritmo meta-euristico	48
3.3	Particle Swarm Optimization development.....	48
3.3.1	<i>Swarm Explosion e Velocity Clamping</i>	58
3.3.2	<i>Il concetto di peso inerziale</i>	60
3.3.3	<i>Il concetto del neighborhood</i>	62
3.4	PSO Standard contemporanea	66
3.4.1	<i>Lavori correlati alla Particle Swarm Optimization</i>	67
3.4.2	<i>Due macro-categorie per la performance del PSO</i>	68
3.4.3	<i>Limiti della PSO</i>	70
3.4.4	<i>Modifiche del PSO</i>	71
3.4.5	<i>Potenziali ricerche da poter sviluppare</i>	72
3.5	Fireworks Algorithm.....	73
3.5.1	<i>Caratteristiche e vantaggi di FWA</i>	75
3.5.2	<i>Principi e implementazione di FWA</i>	77
3.6	Enhanced Fireworks Algorithm.....	84
3.6.1	<i>Un nuovo controllo dell'ampiezza minima di esplosione (MEAC)</i>	85
3.6.2	<i>Un nuovo operatore per la generazione delle scintille esplosive</i>	87
3.6.3	<i>Un nuovo operatore di mappatura</i>	87
3.6.4	<i>Un nuovo operatore per la generazione di scintille gaussiane</i>	88
3.6.5	<i>Un nuovo operatore di selezione</i>	92
4	Case Study	95
4.1	Fireworks Algorithms and PSO Case Study	95
4.1.1	<i>Applicazione</i>	97
4.1.2	<i>Analisi risultati</i>	102
5	Conclusioni	105
6	Bibliografia.....	110
7	Sitografia.....	113

1 INTRODUZIONE

L'obiettivo di questo lavoro è di esporre in modo sintetico i principali concetti e strumenti legati all'approccio evolutivo applicati all'analisi di sistemi di trading.

In particolare, nel primo capitolo sarà dedicato all'analisi tecnica e in particolare si presenteranno i principali indicatori con l'intento di poter fornire al lettore una referenza basica per supportare lo studio dell'analisi tecnica. L'analisi tecnica è costruita nelle seguenti assunzioni base, le quali si estendono nei prossimi capitoli:

- 1) l'interazione tra offerta e domanda determina il prezzo;
- 2) l'offerta e la domanda sono influenzate dalle emozioni degli investitori in particolare le aspettative dei venditori e acquirenti che determinano la domanda e l'offerta, infatti, le aspettative sono causa delle decisioni umane, e le decisioni sono basate a) sulle informazioni b) sulle emozioni come speranza, paura, avidità c) e limitazioni cognitive come le deviazioni comportamentali, emozioni, e sentimenti che provengono dalla chimica e dalle connessioni elettriche all'interno del nostro cervello;
- 3) i prezzi scontano tutto e riflettono tutte le informazioni disponibili – questa assunzione è coerente con la teoria dei mercati efficienti - assunzione questa articolata da Charles Dow e rinforzata da William Peter Hamilton: *“The averages discount everything: The fluctuations of the daily closing prices of the DowJones rail and industrial averages afford a composition index of all the hopes, disappointments, and knowledge of everyone who knows anything of financial matters, and for that reason the effects of coming events (excluding acts of God) are always properly anticipated in their movement.”*;
- 4) i prezzi non sono random perché i mercati hanno una tendenza;
- 5) schemi riconoscibili all'interno dei trend – l'assunzione è che la storia tenderà a ripetersi - come Mark Twain disse *“la storia non si ripete, ma fa rima”* – e tutti gli esseri umani tenderanno ad avere un comportamento simile al modo in cui lo hanno avuto nel passato in circostanze simili ed è proprio questo comportamento simile che tende a formarsi in schemi che hanno risultati prevedibili.
- 6) Le emozioni sono inoltre influenzate da precedenti emozioni attraverso il feedback emotivo – quindi il comportamento nei mercati non è indipendente ma piuttosto è correlato a come si sta comportando il mercato stesso.

In questo capitolo si discute non solo dei principi base dell'analisi tecnica ma anche delle controversie e dibattiti tra accademici sul merito dell'analisi tecnica e sull'efficienza che questo approccio ha nel mercato finanziario.

Considerando l'evoluzione intervenuta nella diffusione degli algoritmi metaeuristici per l'analisi dei vari fenomeni, questo argomento trova una trattazione più ampia. Infatti, nel terzo capitolo si introducono 2 metaeuristiche: la Particle Swarm Optimization e il Fireworks Algorithm. Nella prima parte si introdurranno i concetti base dell'ottimizzazione, dell'evoluzione computazionale, e della swarm intelligence. Dopodiché si introduce la nuova categoria di algoritmi proposta negli anni 90, dove, i nuovi metodi hanno modellato la popolazione ad un livello macroscopico, cioè in termini di strutture sociali e comportamenti aggreganti – anche in questo caso la natura ha offerto ispirazione agli scienziati, e le società gerarchicamente organizzate da organismi semplici, con una gamma limitata di risposte individuali, mostrano comportamenti affascinanti con tratti identificabili di intelligenza sociale nel suo complesso.

La seconda parte invece è devota alla presentazione dell'algoritmo particle swarm optimization (PSO), dove inizialmente si descrive la storia, lo sviluppo e di seguito vengono anche evidenziate carenze delle prime varianti, e, le loro soluzioni sono riportate in un ordine storico, portando il lettore agli sviluppi contemporanei considerati oggi varianti PSO "all'avanguardia".

Nell'ultima parte del terzo capitolo viene introdotto il firework algorithm (FWA) – il quale è sulla simulazione del processo di esplosione dei fuochi d'artificio. In analogia con i veri fireworks che esplodono e illuminano il cielo notturno, i fireworks in FWA vengono "esplosi" nello spazio parametrico di ricerca, e per ogni firework viene avviato un processo di esplosione e poi una "pioggia" di scintille che riempiono lo spazio locale circostante. Inizialmente si introducono i principi base, l'implementazione, la modellizzazione, e l'analisi teorica di FWA e di seguito viene presentato l'algoritmo Enhanced fireworks algorithm (EFWA) il che supera alcune limitazioni di FWA ottenendo miglioramenti significativi.

Nel capitolo 4 si presenta un case study dove si considerano le 2 metaeuristiche di ottimizzazione: Particle Swarm Optimization e Enhanced Fireworks Algorithm, con l'obiettivo di confrontare la loro capacità di esplorazione nello spazio delle soluzioni, e una volta individuata una soluzione promettente, le loro abilità di "perlustrare a fondo"

quella determinata area identificata. Queste metaeuristiche sono state applicate con lo scopo di ottimizzare i parametri di un Trading System basato su un unico indicatore di analisi tecnica: le bande di Bollinger.

2 ANALISI TECNICA

L'analisi tecnica è definita da Pring (1991) come: *“A reflection of the idea that prices move in trends which are determined by the changing attitudes of investors towards a variety of economy, monetary, political and psychological forces.”*

2.1 I MERCATI FINANZIARI MODERNI

I mercati finanziari moderni sono complessi e allo stesso tempo affascinanti. Una caratteristica importante del sistema finanziario moderno è la presenza di un mercato organizzato per gli acquisti e le vendite delle attività finanziarie; per esempio nel mercato azionario i prezzi delle azioni e delle commodities sono registrate giornalmente. Molto prima dell'avvento dell'ipotesi dell'efficienza dei mercati, i professionisti del mercato hanno iniziato a creare metodi statistici semplici per analizzare il mercato azionario.

Nel 1884 Charles Dow sviluppò la sua teoria e propose un indice, il Dow Industrial Index, per tracciare i movimenti del mercato azionario statunitense, 29 anni prima che Louis Bachelier (1900) utilizzasse i processi Random Walk per descrivere i movimenti dei prezzi azionari. Basato sul lavoro di Dow e altri pionieri, un nuovo campo di studio è cresciuto rapidamente, che utilizzando i dati di prezzo e volume, cerca di prevedere l'andamento futuro dei prezzi azionari: ci si riferisce all' "Analisi Tecnica".

In finanza, l'analisi tecnica è una metodologia di analisi per prevedere la direzione dei prezzi attraverso lo studio dei dati passati e presenti del mercato; dati come prezzo e volume. Questo studio è un metodo di valutazione dei titoli che implica un'analisi statistica dei dati. Esistono diverse forme di analisi tecnica: alcune si basano sull'analisi dei patterns dei prezzi, altre utilizzano indicatori tecnici e oscillatori e la maggior parte utilizza una combinazione di queste tecniche.

2.2 PRESUPPOSTI DELL'ANALISI TECNICA

1) Il mercato sconta tutto

Molti esperti criticano l'analisi tecnica perché considera solo i movimenti dei prezzi e dei volumi mentre ignora i fattori fondamentali. La contro argomentazione si basa sull'ipotesi del mercato efficiente, in cui si afferma che il prezzo di un titolo riflette

già tutto ciò che ha o potrebbe influenzare un'azienda, inclusi i fattori fondamentali. L'unica cosa che rimane è l'analisi dei movimenti dei prezzi, che gli analisti tecnici la considerano come il prodotto della domanda e dell'offerta per un particolare stock nel mercato.

2) Il prezzo si muove seguendo un determinato trend

Gli analisti tecnici ritengono che i prezzi si muovano in un trend di breve, medio e lungo termine. Secondo gli analisti è più probabile che un prezzo azionario continui a seguire una tendenza passata piuttosto che muoversi in modo irregolare. La maggior parte delle strategie di trading si basano su questa assunzione.

3) La storia tende a ripetersi

Gli analisti tecnici ritengono che la storia tenda a ripetersi. La natura ripetitiva dei movimenti dei prezzi è spesso attribuita alla psicologia del mercato, che tende ad essere molto prevedibile sulla base di emozioni come per esempio: la paura o l'eccitazione. L'analisi tecnica utilizza chart patterns per analizzare queste emozioni e successivi movimenti del mercato in modo da comprendere le tendenze. Molte forme di analisi tecnica sono state utilizzate per più di 100 anni, e si ritiene ancora che siano rilevanti, perché illustrano modelli di movimenti dei prezzi che spesso si ripetono.

2.3 I TREND

Il trend è un concetto molto importante, è il “cuore” dell'analisi tecnica; è una direzione generale nella quale un mercato è diretto. I trend non sono sempre facili da individuare perché i prezzi non si muovono quasi mai in linea retta. Piuttosto, i prezzi tendono a muoversi in una serie di alti e bassi nel tempo. In analisi tecnica, è la direzione di questi alti e bassi che costituiscono una tendenza. Un trend al rialzo è classificato come una serie di higher highs e higher lows, mentre una tendenza al ribasso è costituita da lower lows e lower highs.

Esistono tre tipi di trend:

- 1) Trend al rialzo (uptrend);
- 2) Trend al ribasso (downtrend);

3) Tendenze orizzontali (sideways): si verificano quando nei peaks ci sono spostamenti piccoli verso l'alto o verso il basso. Si può dire che un trend orizzontale è in realtà l'assenza di un trend ben definito in entrambe le direzioni.

I trend possono essere classificati anche in base alla loro lunghezza. I trend di lungo periodo si verificano in un arco di tempo superiore a un anno; i trend a medio termine si verificano da uno a tre mesi; e le tendenze a breve termine si verificano in meno di un mese. I trend sono integrati l'uno con l'altro. In alte parole, i trend a lungo termine consistono in una serie di trend a medio termine che consistono in una serie di tendenze a breve termine. I trend al rialzo a lungo termine possono avere diversi trend discendenti a breve e medio termine. Quando si analizza un trend, è importante che il grafico sia costruito per riflettere al meglio il tipo di trend analizzato. In generale si suppone che i grafici giornalieri o settimanali siano i migliori per identificare i trend a lungo termine, mentre i grafici a minuti o ore sono i migliori per le tendenze a breve termine.

2.4 LE TRENDLINE

L'analisi tecnica viene costruita nell'assunzione che i prezzi hanno un trend, e l'utilizzo delle trendlines è importante per l'identificazione e conferma di questi trend. Un trendline è una linea retta che connette due o più punti di prezzi che si estendono nel futuro come una linea di supporto o resistenza. È una tecnica di creazione di grafici semplici in cui delle linee vengono aggiunte a un grafico per rappresentare la tendenza presente in un mercato. Queste linee sono utilizzate per mostrare dove è diretto il prezzo.

2.4.1 Uptrend line

Un uptrend line ha una pendenza positiva, formata collegando due o più punti bassi. Che la linea abbia una tendenza positiva il secondo punto basso deve essere più alto del primo punto. Si nota che ci deve essere una connessione di almeno 3 punti prima che la linea si consideri una valida trendline. L'uptrend agisce come supporto e indica che la domanda netta (domanda meno offerta) è crescente anche se il prezzo aumenta. Un prezzo in aumento combinato con una domanda crescente è molto bullish, e mostra una determinazione forte da parte dei buyers. Fino a quando il prezzo resta sopra il trendline, l'uptrend viene considerato solido e intatto. Una interruzione sotto l'uptrend

indica che la domanda netta si è indebolita e che potrebbe essere imminente un cambiamento di trend.



Figura 2.1: Up Trendline (Fonte: StockCharts.com)

2.4.2 Downtrend line

Una linea downtrend ha una pendenza negativa ed è formata collegando due o più punti alti. In modo che la linea venga definita con pendenza negativa il secondo punto alto deve essere minore del primo punto. Almeno 3 punti devono essere connessi prima che la linea venga considerata una trendline valida. Le linee downtrend agiscono come resistenza, e indicano che l'offerta netta (offerta meno domanda) cresce anche se i prezzi diminuiscono. Un prezzo decrescente combinato con l'aumento dell'offerta è molto bearish, e mostra la forte determinazione degli venditori. Fintanto che i prezzi rimangono sotto la linea downtrend, il downtrend rimane solido e intatto.

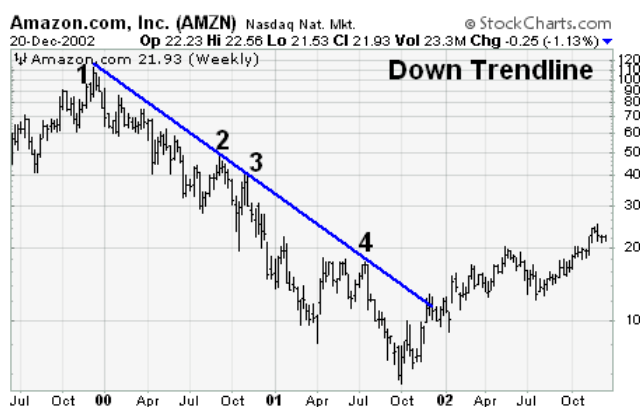


Figura 2.2: Down Trendline (Fonte: StockCharts.com)

Per una spiegazione dettagliata riguardo ai cambiamenti dei trend, si consiglia The Dow Theory. Molti dei principi applicabili ai livelli di supporto e resistenza possono essere applicati alle trendline, in modo da validare le interruzioni dei trendline.

2.5 I CANALI

Un canale è costituito da due trendlines che si comportano da aree di supporto e resistenza con il prezzo che rimbalza tra loro. Il trendline superiore consiste in una serie di massimi, mentre il trendline inferiore consiste in una serie di minimi. Un canale può presentare una inclinazione verso l'alto, verso il basso o lateralmente, ma indipendentemente dalla direzione, l'interpretazione è sempre la stessa. I traders si aspettano che la tendenza di un prezzo si inverta quando il prezzo si sposta e si muove al di fuori dei uno dei 2 livelli (di resistenza e supporto), in quel caso gli operatori possono aspettarsi una forte mossa nella direzione del breakout (il breakout fa riferimento al prezzo azionario che si muove sopra l'area di resistenza, oppure sotto l'area di supporto).



Figura 2.3: Price Channel (Fonte: StockCharts.com)

La figura 2.3 illustra un side channel dove il trendline superiore collega una serie di massimi e il trendline inferiore collega una serie di minimi. Quando il prezzo 'rompe' la trendline superiore (resistenza), il trendline superiore diventa un nuovo livello di supporto.

2.6 I LIVELLI DI SUPPORTO E RESISTENZA

Supporto e resistenza rappresentano giunture chiave laddove le forze della domanda e dell'offerta si incontrano. I livelli di supporto sono dei prezzi che rispetto ai quali la domanda è percepita come sufficientemente forte da impedire che il prezzo scenda ulteriormente, mentre i livelli di resistenza sono i prezzi in cui si ritiene che le vendite siano abbastanza forti da impedire che i prezzi aumentino.

Le frecce vicino alle trendline in alto e in basso mostrano i livelli in cui il prezzo raramente ha superato fino a quando non è andato in breakout più in alto. Dopo il breakout, il trendline superiore è passato da un livello di resistenza a un livello di supporto per il nuovo trend. Perché succede ciò? Perché i livelli di supporto e resistenza sono livelli psicologicamente importanti in cui molti acquirenti e/o venditori sono disposti a negoziare il titolo. Quando i trendline si rompono, "la psicologia" del mercato si sposta e si stabiliscono nuovi livelli di supporto e resistenza. Una volta che il prezzo inizia a scendere in maniera significativa gli acquirenti inizieranno ad acquistare grandi quantità di stock; il che rende più difficile per le azioni di scendere al di sotto di tale livello. D'altra parte, i venditori iniziano a vendere un titolo mentre si sposta verso un picco significativo, rendendo difficile superare il livello superiore. Questa grande pressione di acquisto-vendita rende importanti i punti di supporto e resistenza, punti psicologicamente importanti per la behavioral finance. Un trendline non cessa di essere un'area importante di supporto e resistenza quando è rotto; piuttosto, il suo ruolo è semplicemente invertito. Se un prezzo breaksout da un trendline, il trendline diventa un livello di supporto. È da sottolineare che l'inversione deve essere una vera inversione piuttosto che un falso breakout o breakdown, In altri termini il picco dei prezzi è accompagnato da un volume significativo.



Figura 2.4: Support and Resistance (Fonte: StockCharts.com)

Ad esempio, la figura 2.4 mostra una situazione in cui un livello di resistenza diventa un livello di supporto, prima di tornare a un livello di resistenza dopo una serie di breakout. La forza del trendline diventa sempre più grande nel tempo, dato il numero di 'tocchi' e rimbalzi. Purtroppo, molti trader sottovalutano l'importanza di queste inversioni e non riescono a rendersi conto di quanto spesso si verificano - anche in titoli popolari. In quasi tutti i casi, un titolo avrà sia un livello di resistenza e di supporto che si scambiano tra di loro quando vi è una rottura dall'alto o dal basso. Il breakout trasformerà il trendline che è stato spezzato nel ruolo opposto e verrà stabilito un nuovo price channel. I livelli di supporto e resistenza sono una parte fondamentale dell'analisi dei trend poiché possono essere utilizzati per prendere decisioni di trading e aiutano nell'identificare quando un trend si sta per invertire. Ad esempio, un trader potrebbe identificare un prezzo vicino al livello di supporto e decidere di iniziare a comprare l'azione sapendo che probabilmente il prezzo rimbalzerà più in alto. Tuttavia, una rottura al di là del supporto o della resistenza non sempre indica un'inversione. Ad esempio, un breakout più alto potrebbe essere l'inizio di un trend rialzista più veloce e viceversa per una rottura al di sotto del trendline support. I trader dovrebbero essere consapevoli dei livelli di supporto e resistenza e evitare di piazzare ordini in questi punti principali poiché sono solitamente caratterizzati da molta volatilità. Se ci si sente sicuri di fare uno scambio vicino a questi livelli, è importante evitare di piazzare gli ordini direttamente al livello, poiché sono raramente raggiunti.

2.7 IL VOLUME

Il volume è il numero di azioni che vengono scambiate in un dato periodo, solitamente un giorno. I grafici del volume possono anche essere analizzati per mostrare i trend di aumento o diminuzione del volume nel tempo.



Figura 2.5: Volume on Chart (Fonte: StockCharts.com)

Il volume viene utilizzato dagli analisti tecnici per confermare il trend e i *chart patterns*. La forza di ogni movimento di prezzo è misurata principalmente dal volume. In effetti, un aumento del 50% in un prezzo azionario potrebbe non essere tutto pertinente se si verifica su un volume molto piccolo. Se il volume è inferiore alla media, è probabile che la mosse continuo. Se il volume era significativamente superiore alla media, potrebbe essere l'inizio di un'inversione. Oltre ai movimenti di un giorno, l'andamento del volume nel tempo può essere correlato all'andamento dei prezzi per determinare se uno stock sta guadagnando o perdendo 'momentum'. Il volume è importante per confermare i chart patterns, come 'head and shoulders', triangoli, bandiere e altri patterns. I pattern grafici cercano di prevedere momenti cruciali, come le inversioni. Se il volume non è presente accanto a questi schemi grafici, il segnale di trading risultante non è così affidabile. Gli analisti tecnici guardano attentamente il volume per vedere quando è probabile che si verifichino inversioni, il che significa che le variazioni di volume possono essere un precursore delle variazioni di prezzo. Se il volume è decrescente in un uptrend, potrebbe segnalare che l'uptrend si sta avvicinando alla fine cioè è probabile una inversione.

2.8 FATTORI CHE DEVONO ESSERE CONSIDERATI NELLA LETTURA DI UN GRAFICO

Ci sono diversi fattori di cui gli investitori dovrebbero essere a conoscenza quando si guarda un grafico: la scala temporale, la scala del prezzo e le proprietà dei punti di prezzo utilizzate nella creazione del grafico.

Le scale temporali più frequentemente utilizzate sono intraday, giornalieri, settimanali, mensili, tri mensili e annuali. I Tempi più brevi tendano ad avere più "rumore" ciò può rendere i trend più difficili da individuare. I grafici intraday tracciano i movimenti dei prezzi nel corso di un singolo giorno e vengono spesso utilizzati esclusivamente dai trader giornalieri.

Ogni punto in questi grafici è una versione condensata di ciò che è accaduto in un determinato periodo. Ad esempio, il punto di un grafico settimanale rappresenta il movimento del prezzo nel corso di un'intera settimana o il prezzo di chiusura dell'ultimo giorno di negoziazione.

La scala dei prezzi appare generalmente sul lato destro del grafico e mostra gli intervalli di prezzo dello stock. I prezzi possono essere di natura lineare (aritmetica) o logaritmica.

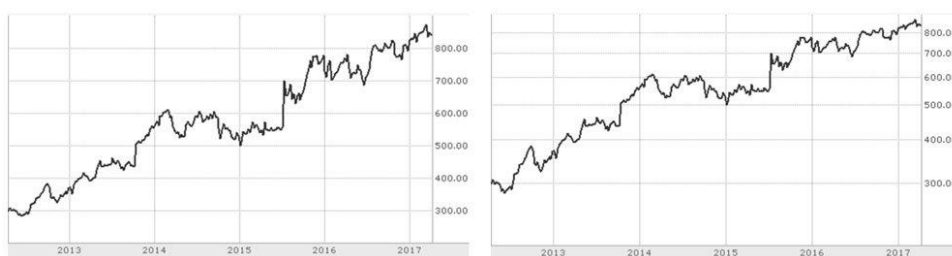


Figura 2.6: Linear v. Logarithmic (Fonte: TD Ameritrade Charts)

Le scale di prezzo lineari (figura 2.6) hanno una spaziatura uniforme tra ciascun punto di prezzo, il che significa che un prezzo che passa da \$10 a \$20 ha la stessa distanza di un prezzo che passa da \$40 a \$50. La scala dei prezzi misura le variazioni assolute e non mostra gli effetti delle variazioni percentuali del prezzo nel tempo. Le scale logaritmiche (figura 2.6- lato destro) esaminano i movimenti dei prezzi in termini percentuali, il che significa che lo spazio tra ogni punto è uguale alla variazione percentuale. Molti professionisti usano i grafici logaritmici perché è più facile individuare il movimento dei prezzi su base percentuale, mentre i grafici lineari possono essere leggermente distorti nei mercati di rapido movimento.

2.9 TIPOLOGIA DI GRAFICI

Nei grafici a linee, ciascun punto rappresenta il prezzo di chiusura del giorno. Nei grafici a candele (candlestick), ogni punto rappresenta l'apertura, il massimo, il minimo e la chiusura del giorno¹. Il grafico a barre è costituito da una serie di linee verticali i cui estremi sono il prezzo minimo e massimo per un dato periodo con un trattino orizzontale su ciascun lato che rappresenta i prezzi di apertura e di chiusura. I prezzi di apertura è il trattino orizzontale sul lato sinistro della linea orizzontale e il prezzo di chiusura si trova sul lato destro della linea. Come in un grafico a barre, i grafici a candele hanno una sottile linea verticale che mostra l'intervallo di prezzo per un dato periodo che è ombreggiato con colori diversi in base al fatto che il titolo si è concluso più in alto o più in basso. La differenza è: una barra o un rettangolo più ampio che rappresenta la differenza tra i prezzi di chiusura e di apertura. I *falling periods* in genere hanno un candlestick body rosso o nero, mentre i *rising periods* avranno un candlestick body bianco o verde. I giorni in cui i prezzi di apertura e chiusura sono gli stessi non avranno affatto un corpo o un rettangolo ampio.

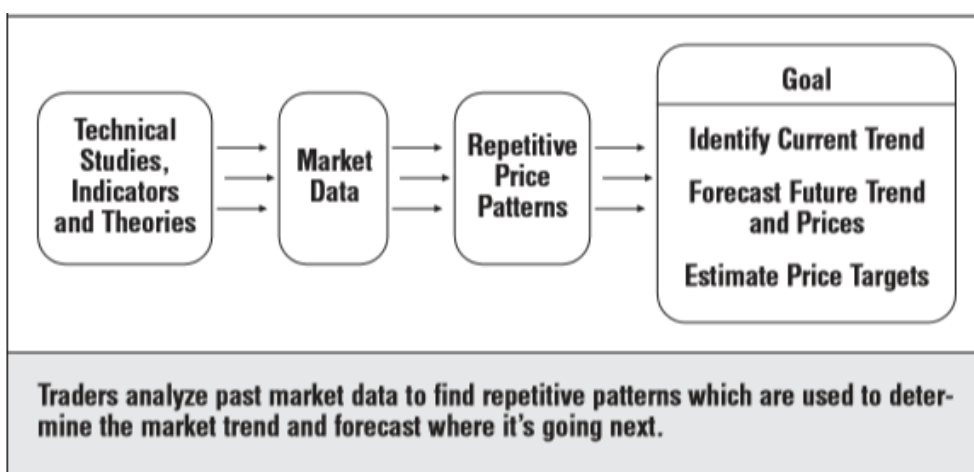


Figura 2.7: Candlestick Chart Example (Fonte: StockCharts.com)

¹ Si fa riferimento a dati giornalieri ma quanto descritto può essere facilmente riferito a periodi diversi (settimane o ore).

2.10 CHART PATTERN

I partecipanti al mercato acquistano e vendono titoli per una vasta gamma di motivi: speranza di guadagno, paura della perdita, conseguenze fiscali, copertura breve, hedging, innescare target di prezzo, analisi fondamentale, analisi tecnica, consigli dei broker e qualche altra dozzina di motivi. I modelli grafici mettono in prospettiva tutti gli acquisti e le vendite consolidando le forze di domanda e dell'offerta in un quadro conciso. I modelli di grafico forniscono un framework per analizzare "la battaglia" tra bulls e bears. Di preciso, le chart patterns e l'analisi tecnica aiutano a determinare chi sta vincendo "la battaglia", permettendo ai trader e agli investitori di posizionarsi adeguatamente. I dati possono essere giornalieri, settimanali, o mensili e i patterns possono essere sia brevi di un giorno che lunghi di alcuni anni per esempio. Le analisi dei Chart Patterns possono essere utilizzate per fare short-term o long-term forecasts.



Source: Market Technologies Corporation

The *head and shoulders* (testa – spalle) è uno schema grafico di inversione del trend. Un "head and shoulder" è caratterizzata da tre picchi con il picco medio che è il picco più alto (head) e gli altri due inferiori e approssimativamente uguali (shoulders). I minimi tra questi picchi sono collegati tramite una linea (neckline) che rappresenta il livello di supporto chiave per controllare un breakdown e un'inversione di trend.

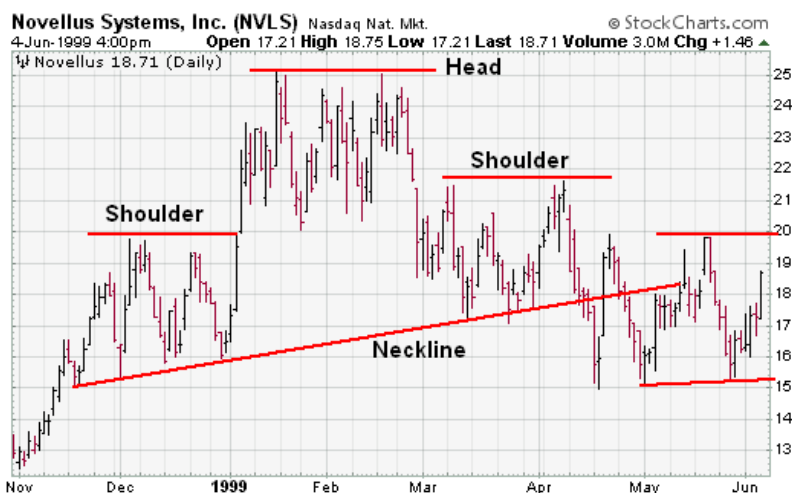


Figura 2.8: Head and Shoulders Chart Pattern (Fonte: StockCharts.com)

Il Cup and handle è uno schema bullish - continuativa rialzista, in cui un trend al rialzo è stato sospeso, ma continuerà quando il pattern sarà confermato. La parta “cup” del pattern dovrebbe essere una forma a “u” con uguali altezze su entrambi i lati della tazza. Le forme “handle” si collocano sul lato destro della “cup”. Una volta che l’handle è completo, il titolo potrebbe superare nuovi massimi e riprendere il trend più in alto.



Figura 2.9: Cup and handle Chart Pattern (Fonte: StockCharts.com)

2.11 ANALISI TECNICA E ANALISI FONDAMENTALE

Quando si tratta di analizzare i mercati finanziari ci sono 2 tipologie di analisi che principalmente si tengono in considerazione: l’analisi tecnica e l’analisi fondamentale. L’analisi tecnica è lo studio e l’analisi dei movimenti dei prezzi nel mercato azionario con lo scopo di cercare di indovinare quali saranno i movimenti futuri del mercato. È una

disciplina di trading utilizzata per valutare gli investimenti e identificare le opportunità di trading analizzando le tendenze statistiche raccolte dall'attività di trading, come la variazione dei prezzi e il volume. L'analisi tecnica analizza prezzo, volume, psicologia degli investitori – un'area conosciuta come finanza comportamentale, flusso di denaro, e altre informazioni di mercato, mentre l'analisi fondamentale esamina i fatti dell'azienda, del mercato, della valuta o delle commodity, considera i fattori economici e finanziari che influenzano un'azienda. Quest'ultima prevede l'uso di dati economici (ad es. produzione, consumo, reddito disponibile per prevedere i prezzi), mentre l'analisi tecnica si basa principalmente sullo studio dei modelli dei dati sui prezzi stessi.

Ci si chiede quale metodo sia migliore? Questa domanda è oggetto di un grande dibattito ma, non c'è un lato "giusto" al dibattito del analisi fondamentale versus l'analisi tecnica; il metodo più appropriato dipende dal individuo perché: alcuni trader sono confortevoli solo quando i segnali vengono generati da un sistema automatico, altri invece trovano questo metodo scomodo, altri invece hanno successo solo se le loro decisioni vengono prese in un ufficio silenzioso, altri trader invece trovano l'analisi fondamentale oppure l'analisi tecnica un approccio naturale, mentre altri utilizzano sia l'analisi tecnica che l'analisi fondamentale.

2.11.1 Le differenze tra analisi tecnica e analisi fondamentale

1) Strumenti

Gli analisti tecnici tipicamente iniziano la loro analisi con i grafici; mentre gli analisti fondamentali iniziano con le dichiarazioni finanziarie di un'azienda. Gli analisti fondamentali cercano di determinare il valore di un'azienda esaminando il conto finanziario, il bilancio e il rendiconto economico. In termini finanziari, l'analista cerca di misurare il valore intrinseco di una società scontando il valore dei flussi di cassa futuri a un valore attuale netto. Un prezzo azionario che opera al di sotto del valore intrinseco di un'azienda è considerato una buona opportunità di investimento e viceversa.

Gli analisti tecnici ritengono che non vi sia alcun motivo per analizzare i rendiconti finanziari di una società poiché il prezzo delle azioni include già tutte le informazioni pertinenti. L'analista si concentra sull'analisi del grafico azionario stesso per suggerimenti su dove può essere diretto il prezzo.

2) Tempo di riferimento

L'analisi fondamentale adotta un approccio di investimento a lungo termine rispetto all'approccio a breve termine adottato dall'analisi tecnica. Mentre i stock charts possono essere limitati in settimane, giorni, o anche minuti, l'analisi fondamentale spesso considera orizzonti temporali più lunghi: trimestri o anni.

Gli investitori si affidano anche a rendiconti finanziari trimestrali, nonché a variazioni degli utili per azione che non emergono quotidianamente come le informazioni sui prezzi e volume. Dopo tutto, una società non può attuare cambiamenti radicali da un giorno all'altro e ci vuole tempo per creare nuovi prodotti, campagne di marketing e altre strategie per cambiare o migliorare un'azienda. Parte del motivo cui gli analisti fondamentali utilizzano un arco temporale a lungo termine, quindi, è perché i dati che usano per analizzare uno stock sono generati molto più lentamente dei dati di prezzo e di volume utilizzati dagli analisti tecnici.

3) Obiettivi

L'analisi tecnica e l'analisi fondamentale hanno diversi obiettivi. Gli analisti tecnici cercano di identificare molte operazioni a breve - medio termine, mentre gli analisti fondamentali cercano di effettuare investimenti a lungo termine nel business sottostante di un titolo.

2.12 INDICATORI TECNICI PER L'ANALISI TECNICA

Gli indicatori rappresentano un approccio quantitativo all'analisi tecnica rispetto a un approccio soggettivo. Guardando il flusso di denaro, le tendenze, la volatilità e il momentum, forniscono una misura secondaria ai movimenti effettivi dei prezzi e aiutano i trader a confermare la qualità dei pattern grafici o a formare i propri segnali di acquisto o di vendita.

Tipologia di indicatori tecnici

Indicatori principali (Leading Indicators). I leading indicators precedono i movimenti dei prezzi e cercano di fornire indicazioni sul futuro andamento dei prezzi. Questi indicatori sono più utili durante i periodi di movimenti laterali (assenza di trend) poiché possono aiutare a identificare breakouts oppure breakdowns.

Lagging Indicators (Indicatori di ritardo). Gli indicatori di ritardo seguono i movimenti dei prezzi e fungono da strumento di conferma. Questi indicatori sono più utili durante

i periodi di trend in cui possono essere utilizzati per confermare che un trend è ancora in fase di collocamento o se si sta indebolendo.

Categorie di indicatori tecnici

Gli *oscillatori*: sono il tipo più comune di indicatore tecnico e sono generalmente delimitati (bound) all'interno di un intervallo. Ad esempio, un oscillatore può avere un minimo di 0 e un massimo di 100 dove zero rappresenta le condizioni di ipervenduto e 100 rappresenta condizioni di ipercomprato.

Non-bounded indicators (indicatori non delimitati: sono meno comuni, ma aiutano a formare segnali di acquisto e vendita, nonché a mostrare forza o debolezza nelle tendenze. Tuttavia, lo fanno in molti modi senza l'uso di un intervallo impostato.

Gli indicatori generano segnali di acquisto e vendita attraverso *crossover* o *divergenze*.

I *crossover* sono la tecnica più popolare per cui il prezzo incrocia una media mobile o quando due medie mobili si incrociano.

La *divergenza* si verifica quando la direzione dell'andamento dei prezzi e la direzione di un indicatore si muovono in direzioni opposte, il che tende a suggerire che la direzione dell'andamento dei prezzi si sta indebolendo.

Gli indicatori possono essere estremamente utili per identificare il momentum, i trend, la volatilità e altri aspetti di un asset. Tuttavia, per massimizzare le probabilità di successo, è importante notare che gli indicatori funzionano meglio se combinati con altre forme di analisi tecnica.

2.12.1 La media mobile

È un indicatore che mostra il valore medio del prezzo in un determinato periodo. È un "lagging" indicator (indicatore di ritardo) che rende più visibile la tendenza attuale. I pattern grafici possono essere difficili da leggere, data la volatilità dei movimenti dei prezzi. Le medie mobili possono aiutare a compensare questi movimenti irregolari rimuovendo le fluttuazioni quotidiane e rendendo più facile individuare i trend.

I tre tipi più comuni di medie mobili sono le medie mobili semplici (SMA), le medie mobili ponderate lineari e le medie mobili esponenziali (EMA). Sebbene il calcolo di queste medie mobili sia diverso, esse vengono utilizzate allo stesso modo per aiutare gli operatori a identificare gli andamenti dei prezzi a breve, medio e lungo termine.

Media mobile semplice (SMA)

Il tipo più comune di media mobile è la media mobile semplice, che prende semplicemente la somma di tutti i prezzi di chiusura riferiti ad un periodo di tempo e divide il risultato per il numero totale di prezzi utilizzati nel calcolo.



Figura 2.10: Simple Moving Averages (Fonte: StockCharts.com)

La Figura 2.10 mostra un grafico azionario con una media mobile di 50 giorni e 200 giorni. La media mobile a 50 giorni è più reattiva alle variazioni di prezzo rispetto alla media a 200 giorni. In generale, i traders possono aumentare la reattività di una media mobile diminuendo il periodo viceversa si attenuano i movimenti aumentando il periodo. Ad esempio, un salto di prezzo di 199 giorni fa ha un impatto su una media mobile di 200 giorni come quello di un giorno fa. Queste critiche hanno spinto i traders a identificare altri tipi di medie mobili progettate per risolvere questi problemi e creare una misura più accurata.

Media ponderata lineare

La media ponderata (WMA) lineare è una media mobile che assegna ad ogni prezzo di chiusura un peso. La somma dei prezzi moltiplicati per i pesi è divisa per la somma dei pesi. Per esempio, una media ponderata lineare di cinque giorni prenderà il prezzo di chiusura corrente e lo moltiplicherà per cinque, il prezzo di chiusura ieri e lo moltiplicherà per quattro, e così via, e quindi dividerà il totale per la somma dei pesi (5+4+...).

Media mobile esponenziale

La media mobile esponenziale (EMA) sfrutta un calcolo più complesso per sintetizzare i dati assegnando un peso più elevato ai dati più recenti. EMA è più reattivo alle nuove informazioni relative alla media mobile semplice. Ecco perché molti trader tecnici scelgono questa media mobile.



Figura 2.11: EMA v. SMA Moving Average (Fonte: StockCharts.com)

La figura 2.11 mostra come l'EMA (linea rossa) reagisce più rapidamente della SMA (linea blu) quando si verificano movimenti improvvisi dei prezzi. Ad esempio, il breakout a fine novembre ha causato un aumento più rapido dell'EMA rispetto a SMA, anche se entrambi misurano lo stesso periodo di 50 giorni. La differenza può sembrare lieve, ma può influire notevolmente sui rendimenti. Quando una media mobile a breve termine è superiore alla media mobile a lungo termine, significa che il trend è più alto o rialzista e viceversa per le medie mobili a breve termine al di sotto delle medie mobili a lungo termine.

Le medie mobili possono essere utilizzate per verificare le inversioni dei trend in vari casi

Crossover del prezzo. Il superamento da parte dei prezzi della media mobile può essere un segnale forte di un'inversione di trend, indica un breakout rialzista. Spesso i trader utilizzano una media mobile a lungo termine per misurare questi crossover poiché il prezzo interagisce spesso con le medie mobili a breve termine, il che crea troppi possibili falsi segnali per l'uso pratico.

Crossover tra media mobili. La media mobile a breve termine che attraversa dall'alto (basso) verso il basso (alto) la media mobile a lungo termine è spesso un segnale di inversione del mercato al ribasso (rialzo), mentre un crossover della media mobile a breve termine sopra una media mobile a lungo termine potrebbe precedere un breakout più alto. Ad esempio, un crossover medio mobile di 50 giorni sopra una media mobile di 200 giorni è un segnale più forte di un crossover media mobile di 10 giorni sopra una media mobile di 20 giorni.



Figura 2.12: Crossover and Support Illustrations (Fonte: StockCharts.com)

E infine, le medie mobili possono essere utilizzate per identificare le aree di supporto e resistenza. Le medie mobili a lungo termine, come la media mobile a 200 giorni, sono aree di supporto e resistenza strettamente osservate. I timeframes più comuni utilizzati durante la creazione di medie mobili sono a 200, 100, 50, 20, e 10 giorni. La media mobile a 200 giorni è una buona misura per un intervallo temporale di un anno, mentre le medie mobili più brevi vengono utilizzate per periodi di tempo più brevi. Queste medie mobili aiutano i trader a ridurre il rumore riscontrato nei movimenti dei prezzi giornalieri e a dare loro un'immagine più chiara del trend.

2.12.2 Sentiment Indicators

Questi indicatori tentano di misurare la psicologia del mercato. Gli indicatori di sentimento comprendono il rapporto di breve periodo, il raggruppamento di servizi di consulenza, mutual funds cash/asset ratio, l'analisi del margin debt, il rapporto put / call,

le indagini dei manager degli investimenti, investment newsletter sentiment, lo short interest, Barron's confident index and CBOE volatility index (indicatore di paura).

2.12.3 Accumulation/Distribution Line

La linea di accumulo/distribuzione è uno degli indicatori di volume più popolari che misura il flusso di denaro in una security. L'indicatore tenta di misurare il rapporto tra acquisto e vendita confrontando il movimento del prezzo di un periodo con il volume per quel periodo. La formula è relativa al così detto Money flow volume (MFV). La ADLine è la somma cumulata degli MFV

$$Acc/Dist = ((Close - Low) - (High - Close)) / (High - Low) * Period's Volume$$

I traders utilizzano l'indicatore per ottenere informazioni sull'ammontare degli acquisti rispetto alla vendita in un dato titolo. Se la linea di accumulo/distribuzione tende verso l'alto, è segno che ci sono più acquisti che vendite e viceversa.

2.12.4 Average Directional Index ADX (Indice direzionale medio)

L'indice direzionale medio (ADX) è un indicatore di trend utilizzato per identificare e misurare la direzione di un trend. L'indice direzionale medio (ADX), l'indicatore direzionale negativo (-DI) e l'indicatore direzionale positivo (+DI) rappresentano un gruppo di indicatori di movimento direzionale che formano un sistema di trading - sviluppato da Welles Wilder. Quest'ultimo determino i movimenti direzionali comparando la differenza tra due consecutivi lows e la differenza tra i loro rispettivi highs.

L'ADX comprende l'indicatore direzionale positivo (+DI) e l'indicatore direzionale negativo (-DI). Il +DI misura la forza del trend rialzista mentre il -DI misura la forza del trend al ribasso. Questi due indicatori sono spesso indicati come Directional Movement Indicator (DMI). ADX è a sua volta derivato dalle smoothed averages della differenza tra +DI e -DI; misura la forza del trend (indipendentemente dalla direzione) nel tempo. Utilizzando questi tre indicatori insieme, i chartisti possono determinare sia la direzione che la forza del trend.

Il movimento direzionale è positivo quando il massimo corrente meno il massimo precedente è maggiore del minore precedente meno il minimo corrente. Questo

cosiddetto movimento direzionale positivo (+DM) equivale quindi al massimo corrente meno il massimo precedente, a condizione che sia positivo. Un valore negativo verrebbe semplicemente inserito come zero.

Il movimento direzionale è negativo quando il minimo precedente meno il minimo corrente è maggiore del massimo corrente meno il massimo precedente. Questo cosiddetto movimento direzionale negativo (-DM) è uguale al minimo precedente meno il minimo corrente, a condizione che sia positivo. Un valore negativo verrebbe inserito come zero.



Figura 2.13 Calculation examples for directional movement (Fonte: StockCharts.com)

La figura 2.13 mostra quattro esempi di calcolo di movimento direzionale. Il primo abbinamento mostra una grande differenza positiva tra gli “high” per un forte movimento direzionale (+DM). Il secondo abbinamento mostra una giornata “outside” per causa del movimento direzionale negativo (-DM) che ottiene “vantaggio”. Il terzo abbinamento mostra una grande differenza tra i minimi per un forte movimento direzionale negativo (-DM). L’accoppiamento finale mostra un giorno “inside”, che equivale a nessun movimento direzionale (zero). Sia il movimento direzionale positivo (+ DM) che il movimento direzionale negativo (-DM) sono negativi e ritornano a zero, quindi si annullano a vicenda. Tutti i giorni “inside” avranno zero movimenti direzionali.

DX viene utilizzato per misurare la forza o la debolezza di un trend, non la direzione attuale. I movimenti direzionali vengono definite da +DI e -DI. Questi indicatori direzionali possono essere combinati con ADX per avere un sistema di trading completo. L'indice direzionale medio misura la forza del trend (indipendentemente dalla direzione) nel tempo. Utilizzando insieme questi 3 indicatori, i grafici possono determinare sia la direzione che la forza del trend. L'indice medio direzionale (ADX) viene utilizzato per misurare la forza o la debolezza di un trend, e non per la direzione effettiva. Il movimento direzionale è definito da +DI e -DI. In generale, i "tori" (i bulls) prevarranno quando +DI è maggiore di -DI, mentre gli orsi (i bears) prevarranno quando -DI è maggiore. Gli incroci di questi indicatori due direzionali possono essere combinati con ADX per avere la visione di un sistema di trading completo. Average Directional Index (ADX) può essere utilizzato per determinare se un titolo ha una tendenza oppure no. Wilder suggerisce che un trend forte è presente quando ADX è maggiore di 25 e che non c'è un trend presente quando ADX è minore di 20. Sembra che ci sia una zona grigia tra 20 e 25 e per questo si può avere bisogno di aggiustare le impostazioni per aumentare la sensibilità dei segnali. ADX ha anche un discreto ritardo a causa di tutte le tecniche di smoothing. Molti analisti tecnici usano 20 come livello chiave per ADX.

2.12.4.1 Direzione del trend e i Corssovers

Wilder ha messo a punto un semplice sistema di negoziazione con questi indicatori di movimento direzionale. Il primo requisito è che ADX negozi al di sopra di 25. Ciò garantisce che i prezzi siano in trend. Molti trader, tuttavia, usano 20 come livello chiave. Un segnale di acquisto si verifica quando +DI attraversa sopra -DI. Wilder ha basato l'arresto iniziale sul minimo del segnale giornaliero. Il segnale rimane in vigore finché rimane basso, anche se +DI torna indietro sotto -DI. Questo segnale rialzista è rafforzato quando l'ADX si alza e il trend si rafforza. Una volta che il trend si sviluppa e diventa profittevole, i traders devono incorporare uno stop-loss e un trailing stop se la tendenza continua. Si innesca un segnale di vendita invece quando -DI passa sopra +DI. Bisogna tenere presente che Wilder era un trader di materie prime (commodity trader) e valute (currency trader). Gli esempi nei suoi libri sono basati su questi sottostanti, non sulle quotazioni delle azioni. I titoli con bassa volatilità potrebbero non generare segnali basati sugli indicatori di Wilder.

2.12.5 Moving Average Convergence/Divergence

Ci sono 3 componenti usati nel calcolo del MACD; le media mobili “lunga” e “corta” in uno specifico periodo e il signal line. La differenza delle due medie mobili è conosciuto come MACD line (o differential line). Il signal line è la media mobile del MACD line.

La *moving average convergence-divergence* (MACD) è uno degli indicatori più potenti e noti nell'analisi tecnica. L'indicatore è composto da due medie mobili esponenziali che aiutano a misurare il momentum in una security. Il MACD è semplicemente la differenza tra queste due medie mobili tracciate contro una linea centrale, dove la linea centrale è il punto in cui le due medie mobili sono uguali. L'indice confronta il momentum a breve termine e il momentum a lungo termine per segnalare l'attuale direzione del momento piuttosto che la direzione del prezzo. Quando il MACD è positivo, segnala che la media mobile a breve termine è superiore alla media mobile a lungo termine e il momentum di una security è al rialzo. Viceversa, quando il MACD è negativo indica che la media mobile a breve termine è inferiore alla media a lungo termine e suggerisce un momentum verso il basso.



Figura 2.14: some of the major MACD crossovers that suggest a change in price momentum over time (Fonte: StockCharts.com)

Ci sono 3 componenti usati per il calcolo della MACD: il numero di termini per la media mobile corta, il numero di termini per la media mobile lunga, e il numero di termini per il calcolo della signal line. L'istogramma MACD il quale è stato sviluppato da Thomas

Asprey, viene utilizzato come uno strumento visuale; è la differenza della linea MACD e del signal line.



Figura 2.15: MACD, signal line and MACD histogram (Fonte: StockCharts.com)

Le medie mobili corta e lunga evidenziano 2 diversi aspetti del prezzo: La media mobile corta rifletterà il cambiamento dei prezzi più rapidamente invece la media mobile lunga farà sì che la situazione attuale del trend sia più visibile. In questo contesto, l'indicatore MACD mostra la forza e riflette i cambiamenti nella direzione del trend attuale. La linea MACD è una Exponential Moving Average (EMA) di 12 giorni meno una EMA di 26 giorni. I prezzi di chiusura vengono utilizzati per queste medie mobili. L'istogramma MACD rappresenta la differenza tra MACD e EMA di 9-giorni, la signal line. L'istogramma è positivo quando la linea MACD è sopra il signal line e negativo quando la linea MACD è sotto il signal line. I valori di 12,26 e 9 sono le impostazioni tipiche utilizzate con MACD, anche se possono essere sostituiti da altri valori, a seconda dello stile di trading e degli obiettivi. MACD si basa sulle convergenze e sulle divergenze di due medie mobili. Le convergenze succedono quando le medie mobili si muovono uno verso l'altro. La divergenza succede quando le medie mobili si allontanano gli uni dagli altri. La media mobile più corta (12-giorni) è più veloce e responsabile per la maggior parte dei movimenti di MACD. La media mobile più lunga (26-giorni) è più lenta e meno reattiva alle variazioni di prezzo del titolo sottostante. La linea MACD oscilla sopra e sotto la linea zero, che è anche conosciuta come la linea centrale. La direzione dipende dalla direzione dell'incrocio delle medie mobili: MACD positiva indica che le EMA di 12 giorni è sopra

EMA di 26 giorni. Valori positivi aumentano la EMA più breve si discosta ulteriormente dall'EMA più lungo. Questo significa che il momentum al rialzo è in aumento. Valori negativi di MACD indica che EMA di 12 giorni è sotto l'EMA di 26 giorni. Valori negativi aumentano quando l'EMA più breve diverge ulteriormente al di sotto dell'EMA più lungo. Questo significa che il momentum downside sta crescendo. Nella figura seguente l'area gialla mostra la linea MACD in un territorio negativo mentre l'EMA di 12 giorni diverge ulteriormente dal EMA di 26-giorni. Il cross iniziale succede a fine settembre (freccia nera) e la MACD si muove ulteriormente in un territorio negativo fino a quando l'EMA di 12-giorni diverge ulteriormente da EMA di 26-giorni. L'area arancione sottolinea un periodo di valori positivi di MACD, cioè quando l'EMA di 12 giorni fu sopra l'EMA di 26-giorni. Da notare che la linea MACD rimane sotto un determinato periodo (linea rossa). Questo significa che la distanza tra un EMA di 12 giorni e un EMA di 26-giorni fu minore di 1 punto, il quale non è una grande differenza.



Figura 2.16: Example MACD (Fonte: StockChart.com)

2.12.5.1 Signal line crossovers

I signal line crossovers sono i segnali più comuni dei segnali MACD. Il signal line è un'EMA di 9-giorni della linea MACD. Agendo come una media mobile dell'indicatore, guida MACD e semplifica l'individuazione delle svolte del MACD. Un crossover bullish succede quando MACD si rialza e performa sopra il signal line. Un crossover bearish succede quando MACD si abbassa e performa sotto il signal line. I crossover possono durare per

giorni, o per settimane; dipendendo dalla forza del movimento. Da sottolineare però che i crossover delle linee di segnale agli estremi positivi o negativi devono essere visti con cautela; ci vuole una forte mossa del titolo per spingere lo “slancio” all’estremo. Anche se la mossa può continuare, è probabile che lo slancio rallenti e questo di solito produrrà un signal line crossover alle estremità. La volatilità del titolo può incrementare il numero dei crossover.

2.12.5.2 Centerline Crossovers

Una linea centrale bullish succede quando la linea MACD si muove sopra la linea zero e torna ad avere un valore positivo. Questo succede quando l’EMA di 12-giorni si muove sopra EMA di 26 giorni. Una linea centrale bearish crossover succede quando la MACD si muove sotto la linea zero per tornare ad essere negativa. Questo succede quando EMA di 12 giorni si muove sotto EMA di 26 giorni. I crossover della linea centrale possono durare dei giorni e dei mesi, dipendendo dalla forza del trend. MACD rimane positivo fino a quando c’è una tendenza al ribasso sostenuta.

2.12.5.3 Le divergenze

Una divergenza bullish si forma quando un titolo registra un minimo minore e il MACD un minimo maggiore. Il minimo minore afferma l’attuale downtrend, però il minimo maggiore in MACD mostra meno downside momentum. Fintanto che il MACD rimane in territorio negativo, nonostante il calo, lo slancio al ribasso supera ancora lo slancio al rialzo. Il rallentamento del momento negativo a volte può prefigurare un’inversione di trend o una considerevole rally.

2.12.5.4 Conclusioni

L’indicatore MACD è speciale perché mette assieme momentum e trend in un unico indicatore. Questa miscela tra trend e momentum può essere applicata ai grafici giornalieri, settimanali o mensili. L’impostazione standard per MACD è la differenza tra gli EMA a 12 e 26 periodi. Se si cerca una maggiore sensibilità dell’indice si può provare una media mobile a breve termine e una media mobile a lungo termine. MACD (5,35,5) è più sensibile del MACD (12,26,9) e può essere più adatto per i grafici settimanali. Se si cerca meno sensibilità si può prendere in considerazione l’allungamento delle medie mobili. Un MACD meno sensibile continuerà a oscillare

sopra/sotto lo zero, ma i crossover della linea centrale e i crossover della linea del segnale saranno meno frequenti. MACD non è particolarmente buono per identificare livelli di ipercomprato e ipervenduto. Anche se è possibile identificare livelli ipercomprati o ipervenduti storici, il MACD non ha limiti superiori o inferiori per vincolarne il movimento. Durante mosse brusche, il MACD può continuare a estendersi oltre i suoi estremi storici. Riassumendo, la linea MACD viene calcolata utilizzando la differenza effettiva tra due medie mobili. Ciò significa che i valori MACD dipendono dal prezzo del titolo sottostante. Non è possibile confrontare i valori MACD per un gruppo di titoli con prezzi variabili. Se si desidera confrontare le letture del momento, è necessario utilizzare l'oscillatore di prezzo percentuale (PPO)², anziché MACD.

2.12.6 Relative Strength Index (RSI)

Il Relative Strength Index (RSI) è un indicatore sviluppato da Wilder. L'RSI è definito come il rapporto tra guadagno medio e la perdita media in un determinato periodo. Esso è un indicatore di momentum che riflette la velocità e le variazioni di prezzo. Viene anche utilizzato per identificare i livelli di prezzo in ipercomprato /ipervenduto. I valori dell'Indice oscillano in un intervallo costante tra 0 e 100; 0 indica che nel periodo di riferimento si sono registrate solo variazioni negative il mercato è in una fase di ipervenduto mentre 100 indica che si sono registrate solo variazioni positive e il mercato è in fase di ipercomprato. Un RSI sopra 70 suggerisce che una security è in fase di ipercomprato, mentre un valore dell'indice inferiore a 30 suggerisce che la security è in fase di ipervenduto. Spesso, l'indicatore viene utilizzato dai trader per determinare se il prezzo è stato spinto a livelli irragionevolmente più alti o bassi dopo una reazione improvvisa alle notizie. Il calcolo RSI standard utilizza come base 14 giorni di trading, sebbene tale valore possa essere adattata in base alle esigenze individuali di un trader.

² L'oscillatore di prezzo percentuale (PPO) è un oscillatore di momentum che misura la differenza tra due medie mobili come percentuale della media mobile più grande.

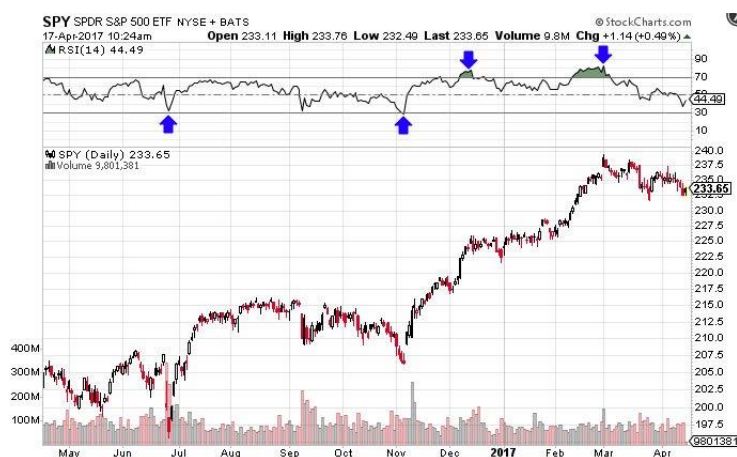


Figura 2.17: RSI Overbought and Oversold Points (Fonte: StockCharts.com)

$$\begin{cases} +1 \text{ (Buy signal) if } (RSI(t-1) < +30) \wedge \text{Current } (RSI(t) > +30) \\ -1 \text{ (Sell signal) if } (RSI(t-1) < +70) \wedge \text{Current } (RSI(t) > +70) \\ 0 \text{ (Stay out from the market) - signal } (RSI(t-1)) \text{ otherwise} \end{cases}$$

2.12.7 Bollinger Bands

Le Bande di Bollinger sono costituite da 3 linee: la linea centrale ottenuta come media mobile del prezzo in un determinato periodo e 2 bande posizionate sopra e sotto la linea centrale (Bande di Bollinger). Il calcolo delle bande di Bollinger superiore e inferiore è ottenuto aggiungendo (e sottraendo) m volte la deviazione standard al prezzo nel periodo specificato della media mobile. Le bande superiore e inferiore si allargano o si restringono a seconda della volatilità del prezzo. Il %b deriva dall'indicatore Bollinger Bands e mira a indirizzare la posizione relativa del prezzo rispetto alle bande superiore e inferiore dell'indicatore Bollinger Bands; l'indicatore %B mostra la relazione tra prezzo e deviazione standard delle bande bollinger.

$$\text{Middle Band} = 20\text{-day simple moving average (SMA)}$$

$$\text{Upper Band} = 20\text{-day SMA} + (20\text{-day standard deviation of price} \times 2)$$

$$\text{Lower Band} = 20\text{-day SMA} - (20\text{-day standard deviation of price} \times 2)$$

$$\text{Bandwidth}^3 = ((\text{Upper Band} - \text{Lower Band}) / \text{Middle Band}) \times 100$$

³ Bandwidth misura la differenza percentuale tra la banda superiore e banda inferiore. La larghezza della banda diminuisce quando le bande di Bollinger si restringono e aumenta all'aumentare delle bande di Bollinger. Poiché le bande di Bollinger si basano sulla deviazione standard, la riduzione della larghezza di

2.12.8 Average true range (ATR)

True Range è un indicatore utilizzato per misurare la volatilità del prezzo; la volatilità cresce/decrece direttamente proporzionalmente all'attività di mercato, ATR invece è la media mobile livellata dei valori TR. Il primo step per calcolare ATR è di trovare una serie di valori true range per un titolo. Il range del prezzo per un titolo in un dato giorno è semplicemente il suo massimo meno il suo minimo. TR viene definito come il valore maggiore dei seguenti metodi:

Method 1: Current High less the current Low

Method 2: Current High less the previous Close (absolute value)

Method 3: Current Low less the previous Close (absolute value)

$$TR = \text{Max}[(H-L), \text{Abs}(H-Cp), \text{Abs}(L-Cp)]$$

$$ATR = \frac{1}{n} \sum_{i=1}^n TR_i$$

Dove:

$TR_i =$ A particular true range

$n =$ The time period employed

Tipicamente, la ATR è basata in 14 periodi e può essere calcolata quotidianamente, settimanalmente, mensilmente. Il primo valore True Range è semplicemente il massimo corrente meno il minimo corrente e il primo ATR è una media dei primi 14 valori True Range. La vera formula ATR non si avvia fino al giorno 15. Anche così, i resti di questi due calcoli "indugiano" per influenzare leggermente i successivi valori ATR. I titoli con prezzi alti generalmente hanno valori ATR maggiori; invece i titoli con prezzi bassi generalmente hanno valori ATR minori. ATR non è un indicatore direzionale come MACD o RSI, ma piuttosto un indicatore di volatilità che riflette il livello di interesse o disinteresse in un movimento. Movimenti notevoli, in ogni direzione, sono spesso accompagnati da True Range Ampii; questo è specialmente vero a inizio di un

una banda riflette una volatilità decrescente e l'aumento della larghezza della banda riflette una volatilità crescente.

movimento. Un'inversione rialzista con un aumento dell'ATR mostrerebbe una forte pressione d'acquisto e rafforzerebbe l'inversione. Un supporto bearish con un aumento di ATR mostrerebbe una forte pressione di vendita. Una rottura al ribasso del supporto (bearish support break) con un aumento di ATR mostrerebbe una forte pressione di vendita e confermerebbe l'interruzione del supporto. Si sottolinea che come per la maggior parte dei suoi indicatori, Wilder ha progettato l'ATR tenendo conto delle commodities e dei prezzi giornalieri. Le commodities sono spesso più volatili delle azioni. Una formula di volatilità basata solo sull'intervallo alto-basso non riuscirebbe a catturare la volatilità dal gap o da mosse limitate; infatti Wilder ha creato Average True Range per catturare questa volatilità "mancante". ATR non fornisce indicazioni sull'orientamento dei prezzi, ma solo sulla volatilità.



Figura 2.18: Average True range (Fonte: StockCharts.com)

I titoli con prezzi alti generalmente hanno valori ATR maggiori; invece i titoli con prezzi bassi generalmente hanno valori ATR minori. I valori assoluti sono usati per assicurare numeri positivi. L'utilizzo di ATR è comunemente utilizzato come "l'exit method" che può essere applicato indipendentemente dalla decisione di entrata. Una tecnica popolare è conosciuta come "chandelier exit", ed è stata sviluppata da Chuck LeBeau. Il

“Chandelier exit” pone un trailing stop⁴ sotto il massimo più alto del titolo raggiunto da quando si è entrato nel trade. ATR può anche dare a un trader un’indicazione di quale misura di trade applicare ai mercati dei derivati. E ‘possibile utilizzare l’approccio ATR per posizionare il dimensionamento che tiene conto della volontà di un singolo operatore di accettare il rischio come anche la volatilità del mercato sottostante. Ci sono due limitazioni nell’utilizzare l’indicatore ATR. La prima è che l’ATR è una misura soggettiva – il che significa che è aperta a delle interpretazioni. Non c’è un singolo valore ATR che indica con certezza il movimento di un trend. Secondo, ATR misura solo la volatilità e non la direzione del prezzo di un titolo.

2.12.9 Money Flow Index (MFI)

Il Money Flow Index (MFI) è un oscillatore che utilizza sia il prezzo che il volume per misurare la pressione di vendita e acquisto. Creata da Gene Quong e Avrum Soudack, MFI è anche nota come RSI ponderata per il volume. Il calcolo dell’MFI inizia con il calcolo del *Typical Price*⁵ per ciascun periodo. Il flusso di denaro è positivo quando il *typical price* aumenta (pressione di acquisto) e negativo quando il prezzo tipico diminuisce (pressione di vendita). Un rapporto tra flusso negativo e positivo viene inserita in una formula RSI per creare un oscillatore che si muove tra zero e cento. Essendo un oscillatore di momentum legato al volume, il MFI è il più adatto per identificare inversioni e prezzi estremi con una varietà di segnali. I valori del MFI possono variare da 0 a 100. Gli autori hanno definito MFI come un RSI volume weighted quando l'indicatore è stato pubblicato per la prima volta; quindi ha delle somiglianze con RSI. RSI si basa su RS che è il rapporto tra guadagno medio diviso per perdita media. Allo stesso modo la MFI utilizza il rapporto di flusso monetario, che è il rapporto tra il flusso di denaro positivo diviso per il flusso di denaro negativo. L’MFI può essere utilizzato per rilevare i livelli di ipercomprato/ipervenduto; i valori vicino a 100 mostrano un livello di ipercomprato e

⁴ Un *trailing stop* è uno *stop order* che può essere setato ad una determinata percentuale o ammontare di soldi – lontano dal prezzo del mercato attuale del titolo. Per una posizione lunga, il *trailing stop* loss va collocato sotto il prezzo del mercato attuale. Per una posizione corta, il *trailing stop* va collocato sopra il prezzo del mercato corrente. Un *trailing stop* è stato disegnando per proteggere i guadagni permettendo un trade di rimanere aperti e continuare ad avere guadagni fino a quando il prezzo si muove a favore dell’investitore.

⁵ *Typical Price* = $\frac{High + Low + Close}{3}$

segnalano opportunità di vendita mentre i valori vicini a 0 segnalano un livello di ipervenduto ed è un segnale di acquisto. Il volume offre una possibilità precoce nel rilevare i livelli di ipercomprato/ipervenduto poiché il volume “conduce” i prezzi.

$$\text{Money Flow Index} = 100 - \frac{100}{1 + \text{Money Flow Ratio}}$$

dove:

$$\text{Money Flow Ratio} = \frac{14 \text{ Period Positive Money Flow}}{14 \text{ Period Negative Money Flow}}$$

$$\text{Raw Money Flow} = \text{Typical Price} * \text{Volume}$$

$$\text{Typical Price} = \frac{\text{High} + \text{Low} + \text{Close}}{3}$$

Quong e Soudack hanno identificato dei segnali di base usando l'MFI. In primo luogo, i chartists possono cercare livelli di ipercomprato o ipervenduto per mettere in guardia da prezzi estremi insostenibili. In secondo luogo, le divergenze rialziste e ribassiste possono essere utilizzate per anticipare le inversioni di tendenza.

2.12.10 Commodity Channel Index (CCI)

L'indicatore Commodity Channel Index, sviluppato da Lambert, è uno strumento per identificare andamenti ciclici nei mercati delle materie prime e nei mercati delle azioni e delle valute. Può essere utilizzato per identificare un nuovo trend o per avvertire delle condizioni estreme. In generale CCI misura il livello di prezzo attuale relativamente ad un livello di prezzo medio in un determinato periodo di tempo. CCI è relativamente alto quando i prezzi sono molto lontano dalla loro media. In questa maniera, CCI può essere usato per identificare livelli di ipercomprato e ipervenduto. Valori molto positivi indicano che i prezzi sono ben al di sopra della loro media, il che è una dimostrazione della forza del trend; valori bassi e negativi indicano che i prezzi sono nettamente inferiori alla loro media, il che è una dimostrazione della debolezza del trend. Il Commodity Channel Index può essere utilizzato come indicatore principale (leading indicator) o come conferma dei segnali provenienti da altri indicatori. In quest'ultimo caso valori superiori a +100 riflettono un movimento forte sui prezzi che può indicare

l'inizio di un trend rialzista. valori inferiori a -100 riflettono un movimento debole sui prezzi che può segnalare l'inizio di un trend al ribasso.

$$CCI = \frac{\text{Typical Price} - MA}{0.015 \times \text{Mean Deviation}}$$

Dove:

$$\text{Typical Price} = \sum_{i=1}^p ((\text{High} + \text{Low} + \text{Close}) \div 3)$$

P = Number of Periods

MA = Moving Average

$$\text{Moving Average} = (\sum_{i=1}^p \text{Typical Price}) \div P$$

$$\text{Mean Deviation} = (\sum_{i=1}^p |\text{Typical Price} - MA|) \div P$$

L'indice CCI misura la differenza tra il prezzo corrente e il prezzo medio storico. Quando CCI è superiore a zero indica che il prezzo è superiore alla media storica; quando invece CCI è inferiore a zero, il prezzo è inferiore alla media storica. Letture elevate superiori di 100 ad esempio, indicano che il prezzo è al di sopra della media storica e che c'è stato un trend forte al rialzo; invece letture basse al di sotto di -100 ad esempio, indicano che il prezzo è ben al di sotto della media storica e che il trend è stato fortemente negativo. Il passaggio da letture negative o quasi zero a +100 può essere utilizzato come un segnale per tenere conto di un possibile trend rialzista. Passare da letture positive o quasi zero a -100 può indicare un possibile emergente trend al ribasso. Il CCI viene utilizzato principalmente per trovare nuovi trend, cercando livelli di ipercomprato e ipervenduto, individuando debolezze nei trend quando l'indicatore diverge con il prezzo. Non ci sono dei livelli di ipercomprato e ipervenduto prefissati, e quindi i traders guardano letture passate dell'indicatore in modo di avere un'idea di dove il prezzo si è invertito.

Ci sono altrettanto le divergenze. Questo è quando il prezzo si muove in un modo e l'indicatore in un altro. Se per esempio il prezzo cresce e l'indicatore decresce, questo può indicare una debolezza nel trend. La divergenza può essere un buon warning per gli

traders - indicandoli che ci può essere una possibilità di una inversione. CCI può essere utilizzato sia come "coincidenza" che come indicatore principale. Come indicatore di coincidenza, se il trend si rialza sopra +100 riflette una forte azione di prezzo che può segnalare l'inizio di un uptrend. Un ribasso sotto -100 riflette un'azione debole di prezzo che può segnalare l'inizio di un downtrend. Come un indicatore principale, i traders possono cercare condizioni di ipervenduto o ipercomprato, che possono prefigurare una inversione media. Similmente, le divergenze bullish and bearish possono essere utilizzate per rilevare i primi cambiamenti di slancio del trend e altrettanto prevedere inversioni di trend. La maggior parte delle volte i movimenti dell'indicatore CCI occorrono tra -100 e +100. Un movimento che eccede questo range mostra una forza insolita o una debolezza che può prefigurare una mossa significativa del trend. Questi livelli vanno visti come "filtri" bearish o bullish. Tecnicamente CCI favorisce i "bulls" quando positivo e i "bears" quando negativo. Identificare livelli di ipercomprato e ipervenduto con CCI può risultare complicato. In primo luogo, CCI è un indicatore senza limiti, teoricamente non ci sono limiti al rialzo o al ribasso. Cioè, rende soggettiva una valutazione ipercomprata o ipervenduta. In secondo luogo, i titoli possono continuare a salire dopo che un indicatore diventa ipercomprato. Allo stesso modo, i titoli possono continuare a scendere dopo che un indicatore diventa ipervenduto. La definizione di ipercomprato o ipervenduto varia per CCI. + o - 100 possono funzionare in un intervallo di trading, ma sono necessari livelli più estremi per altre situazioni. + o - 200 è un livello molto più difficile da raggiungere e più rappresentativo di un vero estremo. La selezione dei livelli di ipercomprato/ipervenduto dipende anche dalla volatilità del titolo sottostante. Le divergenze segnalano un potenziale punto di inversione perché il momento direzionale non conferma il prezzo. Si verifica una divergenza rialzista quando il titolo sottostante fa un minimo più basso e CCI indica un minimo più alto. Una divergenza ribassista si forma quando il titolo registra un massimo più alto e il CCI indica un massimo più basso, il che mostra uno slancio inferiore. Si nota che le divergenze possono essere fuorvianti in un trend forte. Una divergenza ribassista può essere confermata con una rottura sotto lo zero in CCI o una rottura di resistenza sul grafico dei prezzi. Un forte trend rialzista può mostrare numerose divergenze ribassiste prima che un punto estremo si realizzi effettivamente. Una divergenza rialzista invece può essere

confermata con una rottura sopra lo zero in CCI o una rottura di resistenza sul grafico dei prezzi.

Quindi riassumendo CCI è un oscillatore di momentum versatile che può essere utilizzato per identificare livelli di ipercomprato/ipervenduto o inversioni di trend. L'indicatore diventa ipercomprato o ipervenduto quando raggiunge un estremo relativo. Tale estremo dipende dalle caratteristiche del titolo e dall'intervallo storico di CCI. E' probabile che i titoli volatili tendino a richiedere estremi maggiori rispetto ai titoli docili. I cambiamenti del trend possono essere identificati quando CCI attraversa una soglia specifica tra zero e 100. Indipendentemente dal modo in cui viene utilizzato CCI, i grafici dovrebbero usare CCI in combinazione con altri indicatori o analisi di prezzi. Un altro oscillatore del momento sarebbe ridondante, ma On Balance Volume (OBV) o Accumulation Distribution Line possono aggiungere valore ai segnali CCI.

2.12.11 Stochastic oscillator (KDJ)

L'oscillatore stocastico KDJ fu proposto da George Lane negli anni 50. È un indicatore di momentum che utilizza livelli di supporto e resistenza. Il valore dell'indicatore è determinato dalla posizione del prezzo corrente in relazione alla sua fascia di prezzo per un periodo di tempo. L'oscillatore stocastico viene visualizzato con due linee. La linea principale si chiama %K. La seconda riga, chiamata %D, è una EMA iterativa di %K. A volte i trader considerano anche un'altra linea chiamata %J, che è una combinazione lineare di %K e %D. Da nome delle tre linee, l'indicatore dell'oscillatore stocastico è anche chiamato KDJ.

Un oscillatore stocastico è un indicatore di momentum che compare un determinato prezzo di chiusura di un titolo con un determinato range di prezzi in un determinato periodo di tempo. La sensibilità dell'oscillatore versus i movimenti del mercato viene utilizzato per generare segnali di ipervenduto/ipercomprato di trading, utilizzando un intervallo di valori limitato 0-100.

La formula dell'oscillatore stocastico è:

$$\%K = \frac{(\text{Current Close} - \text{Lowest Low})}{(\text{Highest High} - \text{Lowest Low})} * 100$$

$$\%D = 3\text{-day SMA of \%K}$$

Dove:

%K= Il valore attuale dell'indicatore stocastico.

Lowest Low= è il minimo più basso per il period look-back

Highest High= è il massimo più alto per il period look back.

%K viene moltiplicato per 100 per spostare il punto decimale di due posizioni.

%K viene talvolta indicato come indicatore stocastico lento: L'indicatore stocastico "veloce" è preso come %D= media mobile a 3 periodi di %K. La teoria generale che funge da base per questo indicatore è che in un mercato che tende al rialzo, i prezzi si chiuderanno vicino al massimo, e in un mercato che tende al ribasso, i prezzi si avvicinano al minimo. I segnali di transazione vengono creati quando %K attraversa una media mobile di tre periodi, chiamata %D. L'oscillatore stocastico ha un range limitato, il che significa che è sempre compreso tra 0 e 100. Questo lo rende un utile indicatore delle condizioni di ipercomprato e ipervenduto. Tradizionalmente, le letture oltre gli 80 sono considerate nel range ipercomprato e le letture sotto i 20 sono considerate ipervendute. Tuttavia, questi non sono sempre indicativi di un'inversione imminente; trend molto forti possono mantenere condizioni di ipercomprato o ipervenduto per un periodo prolungato. Invece, i trader, dovrebbero cercare cambiamenti nell'oscillatore stocastico per indizi sui futuri cambiamenti di trend.

Il grafico dell'oscillatore stocastico è generalmente costituito da due linee: una che riflette il valore effettivo dell'oscillatore per ciascuna sessione e una che riflette la sua media mobile semplice di tre giorni. Poiché si ritiene che il prezzo segua il momentum, l'intersezione di queste due linee è considerata un segnale che indica che potrebbe essere in atto un'inversione, poiché indica un grande spostamento di momentum di giorno in giorno. La divergenza tra l'oscillatore stocastico e l'azione del trend dei prezzi viene altrettanto considerata come un importante segnale di inversione. Ad esempio, quando un trend ribassista raggiunge un nuovo minimo più basso, ma l'oscillatore segnala un minimo più alto, potrebbe essere un indicatore del fatto che si sta verificando un'inversione rialzista. L'oscillatore stocastico presenta la posizione del prezzo di chiusura di un titolo in relazione alla fascia alta bassa del prezzo di un titolo per un periodo di tempo, in genere un periodo di 14 giorni. L'oscillatore segue la velocità o il momentum del prezzo. Lane rivela che, di norma, il momentum o la velocità del prezzo

di un titolo cambia prima che il prezzo stesso cambi. In questo modo, l'oscillatore stocastico può essere utilizzato per rivelare le inversioni quando l'indicatore rivela divergenze rialziste o ribassiste. In generale si acquista quando l'oscillatore (o% K o% D) scende al di sotto un livello specifico (ad es. 20) e poi sale sopra quel livello. Si vende quando l'oscillatore sale sopra un livello specifico (ad es. 80) e poi scende al di sotto di quel livello. Invece si acquista quando la linea %K sale sopra la linea %D e vende quando la linea% K scende sotto la linea% D.

2.12.12 Know Sure Thing (KST)

Il Know Sure Thing (KST) è stato sviluppato da Martin Pring. È un oscillatore basato sul momento, basato sulla velocità di variazione per quattro diversi intervalli di tempo.

KST misura la dinamica dei prezzi per quattro diversi cicli di prezzi, combinandoli in un unico oscillatore di momento. KST può essere utilizzato per osservare/cercare divergenze, linee di segnale crossover. Pring applica frequentemente linee di trend a KST. Sebbene i segnali di trendline non si verificano spesso, Pring osserva che tali interruzioni rafforzano le linee di segnale crossover. KST viene calcolato con la seguente equazione:

$$KST = (RCMA \#1 \times 1) + (RCMA \#2 \times 2) + (RCMA \#3 \times 3) + (RCMA \#4 \times 4)$$

Dove:

RCMA #1=10-period SMA of 10-period ROC

RCMA #2=10-period SMA of 15-period ROC

RCMA #3=10-period SMA of 20-period ROC

RCMA #4=15-period SMA of 30-period ROC

Il signal line viene calcolato prendendo il 9-period SMA del valore KST.

Se il KST supera la linea dello zero e fa un crossover sopra la sua linea del segnale viene attivato un segnale di acquisto. Se il KST rimane sopra la linea zero per un lungo periodo, viene confermato un trend rialzista. Se invece il KST attraversa sotto la linea zero, viene attivato un segnale di vendita. Inoltre, se il KST rimane al di sotto della linea zero, viene confermata la tendenza al ribasso. L'impostazione predefinita dell'indicatore KST è KST (10, 15, 20, 30, 10, 10, 10, 15, 9). L'indicatore KST fornisce segnali anche in presenza di

divergenze: se i prezzi diminuiscono e l'indicatore KST aumenta, vuol dire che c'è una divergenza rialzista. Se il prezzo delle azioni aumenta mentre l'indicatore KST diminuisce, allora si tratta di una divergenza ribassista. L'indicatore KST può essere uno strumento efficace per i trader intraday anche se a volte l'indicatore tende a dare falsi segnali come qualsiasi altro strumento di trading. Il KST insieme all'indicatore del volume è uno dei modi efficaci per conoscere e convalidare i trend. Utilizzando i tre indicatori ROC, Pring seleziona le negoziazioni utilizzando le seguenti linee guida:

- 1) Una trendline basata su circa la metà del periodo di ROC più lungo determina la direzione del trade.
- 2) Il ROC più lungo di 24 periodi identifica il movimento principale.
- 3) I movimenti di prezzo più forti si verificano quando tutti e tre i valori ROC si muovono nella stessa direzione.
- 4) Se il ROC a 24 periodi raggiunge il picco mentre gli altri due indicatori sono in aumento, il sell-off è minore. Allo stesso modo, quando gli indicatori a 12 e 6 periodi raggiungono il picco ma il periodo a 24 aumenta, anche il sell-off è lieve.
- 5) Tutti e tre gli indicatori devono raggiungere il picco contemporaneamente, seguiti da tutti e tre i valori in calo, per vedere un significativo sell-off.



Figura 2.19: KST (Fonte: StockCharts.com)

2.12.13 Williams Percent Range (WilliamR)

Williams %R (Williams Percent Range), è un tipo di indicatore di momentum che si sposta tra 0 e -100 e misura i livelli di ipercomprato e ipervenduto. William %R può essere utilizzato per trovare punti di entrata e di uscita sul mercato. L'indicatore è stato sviluppato da Larry Williams e confronta il prezzo di chiusura di un titolo con la fascia alta-bassa in un periodo specifico, in genere 14 giorni o periodi.

La formula di Williams %R è:

$$Williams\%R = \frac{(Highest\ High - Close)}{(Highest\ High - Lowest\ Low)} * 100$$

Dove:

Highest High = Prezzo più alto nei "N" periodi di ricerca "lookback", in genere 14 giorni.

Close = Prezzo di chiusura più recente.

Lowest Low = Prezzo più basso nei "N" periodi di ricerca "lookback", in genere 14 giorni.

Williams%R registra il massimo e il minimo per ciascun periodo in 14 periodi. Nel 14° periodo, nota il prezzo corrente, il prezzo più alto e il prezzo più basso. Nel 15° periodo, annota il prezzo corrente, il prezzo più alto e il prezzo più basso, ma solo per gli ultimi 14 periodi (non gli ultimi 15). Alla fine di ciascun periodo, si calcola la nuova Williams%R, utilizzando solo gli ultimi 14 periodi di dati. Quando l'indicatore è compreso tra -20 e zero, il prezzo è ipercomprato o vicino al massimo del suo price range recente. Quando l'indicatore è compreso tra -80 e -100 il prezzo è ipervenduto, o lontano dal massimo dal suo price range recente. Durante un uptrend, i trader possono stare all'allerta che l'indicatore si sposti sotto -80. Quando il prezzo inizia a salire e l'indicatore ritorna al di sopra di -80, potrebbe segnalare che l'uptrend del prezzo sta ricominciando. Una indicazione di fase di ipercomprato o ipervenduto non significa che il prezzo si inverta. L'ipercomprato significa semplicemente che il prezzo è vicino ai massimi del suo range recente, e l'ipervenduto significa che il prezzo è nella parte bassa del suo range recente. Può essere utilizzato per generare segnali di trend quando il prezzo e l'indicatore si spostano dal territorio ipercomprato o ipervenduto. Lo stesso concetto potrebbe essere utilizzato per trovare posizioni corte in un downtrend. Quando l'indicatore è al di sopra di -20, si osserva che il prezzo inizierà a scendere insieme a Williams %R che scende al di sotto di -20 per segnalare una potenziale continuazione del downtrend. Durante un

forte uptrend, il prezzo raggiungerà spesso -20 o superiore. Se l'indicatore scende e quindi non può tornare al di sopra di -20 prima di ricadere, questo indica che lo slancio dei prezzi al rialzo è in difficoltà e potrebbe seguire un calo dei prezzi maggiore. Lo stesso concetto si applica a un downtrend. Quando l'indicatore non riesce più a raggiungere quei livelli bassi prima di salire, potrebbe indicare che il prezzo salirà. Se si visualizza una lettura zero sull'indicatore significa che durante il periodo selezionato il mercato sta facendo trading vicino o al di sopra del massimo più alto. Al contrario, se si visualizza una lettura di -100, significa che durante il periodo selezionato il mercato sta facendo trading vicino o al di sotto del minimo più basso.

Con il passare dei periodi, il prezzo corrente relativo agli alti e ai bassi nel periodo lookback cambia, anche se il prezzo non si è realmente mosso. L'indicatore ci fornisce anche informazioni sulla forza o debolezza dell'andamento di un qualsiasi strumento finanziario che ha un prezzo associato.

2.12.14 The Ultimate Oscillator indicator (UO)

Ultimate Oscillator è un indicatore tecnico sviluppato da Larry Williams nel 1976 per misurare la dinamica dei prezzi di un asset in più periodi. Utilizzando la media ponderata di tre diversi intervalli di tempo, l'indicatore ha meno volatilità e meno segnali di trend rispetto ad altri oscillatori che si basano su un singolo intervallo di tempo. I segnali di acquisto e vendita sono generati seguendo le divergenze. L'UO genera meno segnali di divergenza rispetto ad altri oscillatori a causa della sua costruzione multi-timeframe. L'indicatore utilizza tre intervalli di tempo per il calcolo: 7, 14 e 28 periodi. Il periodo più breve ha il maggior peso nel calcolo, mentre il periodo più lungo ha il minor peso. I segnali di acquisto si verificano quando vi è una divergenza rialzista (divergenza bullish), il livello di divergenza bassa è inferiore a 30 sull'indicatore e quindi l'oscillatore sale al di sopra del livello di divergenza alta. Un segnale di vendita si verifica quando vi è una divergenza ribassista (bearish), il livello di divergenza alta è superiore a 70 e quindi l'oscillatore scende al di sotto del livello di divergenza bassa. Questo indicatore è un metodo per misurare la pressione di acquisto. L'UO aumenta quando la pressione di acquisto è forte mentre quando la pressione di acquisto è debole, l'UO diminuisce. La caratteristica più importante dell'UO è che dà il massimo peso al più breve lasso di tempo, eppure tiene anche conto dell'asso temporale più lungo. L'obiettivo di ciò è di

evitare false divergenze. Quasi tutti gli oscillatori del momento generano segnali basati su divergenze rialziste e ribassiste. Pertanto, se si presentassero frequenti falsi segnali basati su queste divergenze sarebbero ovviamente un problema. Includendo più intervalli di tempo e ponderandoli di conseguenza, l'UO mira a ridurre questi falsi segnali. L'ultimate oscillator è un indicatore di intervallo con un valore che oscilla tra 0 e 100. Simile all'indice RSI, livelli inferiori a 30 sono considerati ipervenduti e quelli superiori a 70 sono considerati ipercomprati. I segnali di trading vengono generati quando il prezzo si muove nella direzione opposta rispetto all'indicatore.

Bearish UO Divergence, tale configurazione si osserva quando:

- 1) La divergenza ribassista: il prezzo forma un massimo più alto mentre l'UO fa un massimo più basso.
- 2) In secondo luogo, il primo massimo nella divergenza deve essere superiore a 70. Ciò significa che la divergenza è iniziata dal territorio ipercomprato ed è più probabile che si traduca in un'inversione del prezzo al ribasso.
- 3) UO scende al di sotto del minimo della divergenza. La divergenza bassa è il punto più basso tra i due massimi della divergenza.
- 4) In questo caso l'indicatore genera segnali di vendita.



Figure 2.20: Example of Bearish UO Divergence (Fonte: TradingView.com)

2.13 CRITICHE DELL'ANALISI TECNICA

Ci sono alcune critiche pragmatiche all'analisi tecnica; alcuni investitori credono erroneamente che l'analisi tecnica sia solo per il trading a breve termine e non sia utile per l'investitore a lungo termine. Poiché le informazioni fondamentali e pertinenti su un titolo non cambiano minuto per minuto o giorno per giorno, un trader di breve termine deve fare affidamento maggiormente sull'analisi tecnica. Il trader a breve termine deve fare affidamento ad una interpretazione del comportamento dei prezzi di mercato piuttosto che a notizie e annunci aziendali. In questi casi, l'analisi tecnica offre al professionista un vantaggio maggiore rispetto all'analisi fondamentale. Ma l'analisi tecnica è lo studio del prezzo che le persone determinano. Le persone influenzano i prezzi a lungo termine tanto quanto influenzano i prezzi a breve termine. L'analisi del comportamento dei prezzi a lungo termine è altrettanto preziosa per l'investitore quanto l'analisi a breve termine è per il trader. Per quanto riguarda le regole tecniche che funzioneranno in futuro sono di difficile previsione. Tutte le regole sono soggette a modifiche. Almeno l'analisi tecnica funziona con dati affidabili; la maggior parte delle regole sono state testate, e i livelli di rischio possono essere stabiliti per limitare la perdita di capitale. Un'ulteriore critica all'analisi tecnica è che una parte delle regole tecniche richiede un giudizio soggettivo. D'altra parte, quale forma di analisi degli investimenti non richiede un giudizio soggettivo? Ma una caratteristica fondamentale da sottolineare è che l'analisi tecnica fa riferimento anche ad una informazione utilissima che è l'equity – informazione questa utilissima perché è strutturata, oggettiva, certificata e per la teoria dei mercati efficienti il prezzo rappresenta tutte le informazioni. È vero che l'analisi tecnica dei grafici è soggettiva e alcuni la chiamano "arte" o "abilità", ma la rappresentazione dei dati su un grafico è solo un altro mezzo di analisi delle serie storiche. Molti teorici usano anche grafici per chiarire le loro ipotesi. Un aspetto interessante dell'analisi tecnica, a differenza di altri, è che i dati utilizzati sono tempestivi e precisi. Però, guardando solo i numeri, non si ha sotto controllo il problema, perché ci danno sì informazioni ma ci dicono abbastanza poco. In tutte le analisi formali vanno riportati solo i numeri ma per non avere una percezione limitata bisogna andare a vedere anche i grafici. Limitandosi a guardare solo i numeri gli operatori di mercato prendono tanti granchi - non si riesce a sintetizzare. Per esempio:

la conoscenza dello shape di una distribuzione a parità di numero mi può portare a conclusioni diametralmente opposte.

3 PARTICLE SWARM OPTIMIZATION E FIREWORKS ALGORITHMS

3.1 OPTIMIZATION ERA

L'era dell'ottimizzazione iniziò con i primi tempi di Newton, Lagrange e Cauchy. In particolare, lo sviluppo delle basi matematiche - come i metodi di calcolo differenziale che sono in grado di muoversi verso l'ottimo di una funzione - è stato possibile grazie ai contributi di Newton, Gauss, e Leibnitz.

È stato Cauchy che ha proposto il primo *metodo di discesa più ripida* (detto anche metodo di discesa del gradiente – tecnica per determinare massimo o minimo di una funzione con n variabili) per risolvere problemi di ottimizzazione senza vincoli. Inoltre, Bernoulli, Eulero, Lagrange, Fermat e Weistrass hanno sviluppato le basi della minimizzazione delle funzioni di calcolo mentre Lagrange ha inventato il metodo di ottimizzazione per problemi vincolati usando i moltiplicatori chiamati oggi "Moltiplicatori di Lagrange". Dopo la seconda metà del ventesimo secolo, con l'invenzione dei computer digitali, sono state sviluppate nuove tecniche e algoritmi per risolvere problemi complessi di ottimizzazione e tali sforzi hanno stimolato ulteriori ricerche su aree diverse e completamente nuove.

Nel 1957 Bellman ha introdotto il principio di ottimalità per la programmazione dinamica con una strategia di ottimizzazione basata sulla suddivisione del problema in sotto-problemi minori. L'equazione data dal suo nome descrive la relazione tra questi sotto-problemi.

L'ottimizzazione combinatoria (combinatorial optimization) è un termine generico per un insieme di metodi di ottimizzazione che includono la ricerca operativa, la *teoria dell'algoritmo* e la teoria della complessità computazionale. Questi ultimi metodi cercano possibili soluzioni fattibili in spazi discreti con l'obiettivo di trovare la soluzione ottimale quando la ricerca esaustiva (o sequenziale) non è fattibile.

3.1.1 Teoria dell'algoritmo

3.1.1.1 Storia e origine di un algoritmo

La parola "algoritmo" deriva dal nome di un grande matematico persiano Abd Allāh Muhammad ibn Mūsā al-Khwārizmī, che visse nella prima metà del IX secolo. Questo matematico ha praticamente creato il sistema di numeri arabi e l'algebra di base (in

particolare metodi per risolvere equazioni lineari e quadratiche). Il suo nome è stato trasformato in latino come "algorismus", nel tempo come "algoritmo", ciò che originariamente significava "l'implementazione dell'aritmetica utilizzando numeri arabi" (*Wikipedia.com*).

3.1.1.2 Definizione di un algoritmo

Un algoritmo viene spesso definito come istruzioni o passaggi precisi che vengono utilizzati per risolvere un determinato tipo di attività. Questa definizione sembra essere vera, comunque, da ricerche ulteriori si potrebbe scoprire che non è molto preciso. La definizione accurata di un algoritmo: un algoritmo è una procedura che può essere eseguita dalla macchina di Turing. La macchina di Turing è un modello teorico di computer descritto dal matematico Alan Turing. Quest'ultimo è definito il padre dell'intelligenza artificiale, in base soprattutto alla teorizzazione dell'omonima macchina. Che cos'è la macchina di Turing e perché è così importante?

È una macchina di calcolo logico, consistente in: a) un nastro infinito in entrambe le direzioni, diviso in caselle ciascuna delle quali può contenere il simbolo 0 oppure il simbolo 1. Il nastro rappresenta la memoria della macchina; b) una testina che può leggere il simbolo, 0 oppure 1, contenuto in una casella e scrivere un simbolo in una casella, e può muoversi lungo il nastro, una casella per volta. Si tratta soltanto di un modello teorico, poiché prevede un tempo e uno spazio (cioè il nastro) infinito. Un algoritmo è una procedura schematica per risolvere un certo tipo di problema, che viene implementato usando un numero finito di passaggi ben definiti. In pratica, un algoritmo è intelligente, semplifica il lavoro e si applica nelle attività ripetitive.

Creare un algoritmo è certamente un lavoro mentale molto duro. Non ha senso sprecarlo in cose uniche per una persona, che non si ripeterà, e che nessuno ha bisogno. Un algoritmo di solito funziona con alcuni input, variabili, e attività. Gli input hanno un definito set di valori che possono essere acquisiti. Un algoritmo ha almeno un output, necessario in relazione a input specifici e costituisce quindi la risposta a un problema che l'algoritmo dovrebbe risolvere. Gli algoritmi devono essere limitati, definiti, efficaci e generali. Ogni algoritmo deve terminare dopo un numero finito di passaggi. Questo numero può essere arbitrariamente grande (a seconda dell'entità e dei valori dei dati di input) ma per ogni input deve essere finito. Ogni step di un algoritmo deve essere

definito in modo chiaro e preciso. Poiché in genere un linguaggio semplice non fornisce assoluta precisione e chiarezza di espressione; il linguaggio di programmazione fu disegnato e in questo linguaggio, ogni comando ha un significato chiaramente definito. L'espressione di metodi computazionali in linguaggio di programmazione viene chiamato "programma". Riassumendo, un algoritmo non risolve un problema specifico (ad esempio come calcolare $8+9$), ma la classe generale di problemi simili (ad esempio come calcolare la somma di due numeri interi).

L'ottimizzazione combinatoria ha importanti applicazioni in diversi campi, tra cui intelligenza artificiale, machine learning, matematica e ingegneria del software. Viene inoltre applicato a determinati problemi di ottimizzazione che comportano incertezza. Dantzig, Charnes e Cooper hanno sviluppato la programmazione stocastica, che studia il caso in cui alcuni dei vincoli o parametri dipendono da variabili casuali.

La programmazione stocastica è un approccio per modellare i problemi di ottimizzazione che contengono incertezza. Mentre i problemi di ottimizzazione deterministica sono formulati con parametri noti, i problemi del mondo reale includono quasi sempre parametri che sono sconosciuti al momento in cui una decisione deve essere presa. Quando i parametri sono incerti, ma si presume che si trovino in un determinato insieme di possibili valori, si potrebbe cercare una soluzione che sia fattibile per tutte le possibili scelte di parametri e ottimizzare una determinata funzione oggettiva.

La famiglia degli algoritmi evolutivi (EA) sono talvolta chiamati meta-euristici⁶, perché fanno poche o nessuna ipotesi sul problema da ottimizzare e possono quindi cercare l'ottimale globale su un ampio set di soluzioni. Tuttavia, non vi è alcuna garanzia che la soluzione ottimale sarà mai trovata.

Tutti siamo ottimizzatori poiché prendiamo decisioni con lo scopo di massimizzare la qualità della nostra vita, la produttività nel tempo, e il nostro benessere in un modo o nel altro. L'ottimizzazione è stata, è, e sarà sempre il requisito fondamentale della vita umana, e questo fatto produce lo sviluppo di un numero enorme di tecniche in quest'area, a partire dai primi di anni di civiltà fino ad ora.

⁶ Una metaeuristica è un algoritmo che, dato uno specifico problema di ottimizzazione, "approfondisce" la ricerca della soluzione ottima nelle zone più "promettenti" nello spazio delle soluzioni. (M. Corazza (2017), Department of Economics, Ca' Foscari University of Venice)

La maturità dei mercati finanziari sta dando forma a un sistema più complesso, non lineare, rumoroso, dinamico e caotico. Pertanto, il problema della previsione dell'andamento dei mercati finanziari è stato un lavoro difficile e richiede nuovi modelli in grado di catturare il comportamento delle tendenze del settore delle securities e di contribuire a prendere decisioni di trading migliori.

3.2 L'ALGORITMO META-EURISTICO

Gli algoritmi meta-euristico sono uno strumento importante per i problemi di ottimizzazione globale. Si possono distinguere ottimi locali dagli ottimi globali. Gli ottimi locali sono la soluzione migliore trovata in un sottoinsieme di soluzioni, ma non necessariamente la migliore per l'intero spazio del problema. Gli ottimi globali invece sono la soluzione migliore nell'intero spazio del problema. Trovare gli ottimi globali nella maggior parte dei problemi complessi è estremamente difficile e quindi soluzioni abbastanza soddisfacenti sono spesso accettate. Gli algoritmi meta-euristici sono implementati per raggiungere questi obiettivi.

Ci sono tre ragioni per utilizzare gli algoritmi meta-euristici:

- 1) *Semplicità*: La maggior parte degli algoritmi sono semplici e facili da implementare. La parte iniziale/primaria dell'algoritmo può essere scritto in 100 linee in un linguaggio di programmazione.
- 2) *Flessibilità*: Questi algoritmi sono flessibili in quanto sono in grado di risolvere diversi problemi di ottimizzazione che non possono essere affrontati dagli algoritmi classici.
- 3) *Ergodicità*: Gli algoritmi meta-euristici hanno un alto grado di ergodicità il che significa che possono ricercare spazi multimodali con una varietà sufficiente ed allo stesso tempo evitare ottimi locali. L'ergodicità è spesso il risultato di una tecnica di casualizzazione che deriva da sistemi naturali come il crossover e mutation (negli algoritmi genetici) oppure da modelli statistici come il Random Walk e Lévy flights.

3.3 PARTICLE SWARM OPTIMIZATION DEVELOPMENT

La Particle Swarm Optimization è stata introdotta da Kennedy ed Eberhart nel 1995 come un processo stocastico di ricerca e ottimizzazione basato sulla popolazione. È

originato dalle simulazioni al computer del comportamento degli individui (particelle o organismi viventi) di uno stormo di uccelli che mostrano un comportamento particolare quando cercano del cibo (obiettivo).

Come strumento di ottimizzazione il PSO ha visto un successo nel campo dell'ingegneria civile dove è stato utilizzato per risolvere problemi di ottimizzazione di forma e dimensione nella progettazione strutturale, valutazione delle condizioni strutturali e monitoraggio della salute, caratterizzazione e modellizzazione del materiale strutturale, progettazione della rete di trasporto, previsione flusso traffico, controllo del traffico, previsione degli incidenti stradali, previsione della fase fluviale, ottimizzazione della progettazione delle reti di distribuzione acqua/acque reflue, risoluzione dei problemi di funzionamento del giacimento, stima/calibrazione dei parametri dei modelli idrologici, pianificazione della costruzione, pianificazione e gestione, contenzioso edilizio, stima dei costi di costruzione e previsione, identificazione inversa dei parametri e calibrazione del modello geotecnico, analisi della stabilità del pendio, ingegneria della pavimentazione e così via.

Il PSO in termini elementari appartiene al paradigma dello swarm intelligence che studia il comportamento collettivo e le caratteristiche sociali di sistemi organizzati, decentralizzati e complessi, noti come "sciame". Uno sciame (swarm) è una collezione (popolazione) apparentemente disorganizzata di individui (particelle) in movimento che tendono a raggrupparsi insieme mentre ogni individuo sembra muoversi in una direzione casuale. Ogni individuo nello sciame ha la capacità di interagire con gli altri individui sebbene le capacità di ciascuna particella siano piuttosto limitate da un certo insieme di regole. Pertanto, il comportamento di una particella in uno sciame è spesso insignificante, però il loro comportamento collettivo e sociale è di fondamentale importanza, in quanto l'intelligenza dello sciame deriva sia dall'adattamento collettivo che dalla natura stocastica dello sciame. La motivazione principale deriva direttamente dagli sciame biologici (organic swarms) presenti in natura come stormi di uccelli, banchi di pesci, colonie di formiche, che esibiscono una straordinaria auto-organizzazione e capacità di adattamento collettivo/sociale. Ciò non può essere spiegato semplicemente dal comportamento aggregato di ogni singolo membro nello sciame ma dal loro adattamento collettivo all'ambiente, che a sua volta rende possibile la sopravvivenza in

natura. Uno swarm può quindi essere definito come una popolazione di individui interagenti che raggiungono un obiettivo adattandosi collettivamente all'ambiente locale e/o globale.

Ogni individuo, indicato come una particella, segue due comportamenti molto semplici:

- 1) segue l'individuo con le migliori prestazioni;
- 2) si muove verso le migliori condizioni trovate dall'individuo stesso.

L'ottimizzazione è una procedura per rilevare attributi, configurazioni, o parametri di un sistema, per produrre risposte desiderabili. La modellazione offre una traslazione del problema originale in una struttura matematica che può essere gestita attraverso procedure di ottimizzazione algoritmica. I modelli prodotti sono solitamente formulati come funzioni, chiamate "funzioni obiettivo", in una o più variabili che corrispondono a parametri adattabili del sistema.

Il modello è costruito in modo tale che, sulla base di criteri casuali di ottimalità, le configurazioni più desiderabili di sistema corrispondano ai valori estremi della funzione obiettivo. Pertanto, il problema di ottimizzazione del sistema originale viene trasformato in un problema di minimizzazione o massimizzazione della funzione equivalente. La difficoltà nel risolvere questo problema è dipendente fortemente dalla forma e dalle proprietà matematiche della funzione obiettivo.

Molti degli algoritmi proposti per la soluzione di problemi di ottimizzazione richiedono l'assunzione di alcune proprietà matematiche della funzione obiettivo, come la differenziabilità e la continuità di Lipschitz.

Questi approcci raggiungono elevati tassi di convergenza. Tuttavia, le ipotesi necessarie per la loro applicazione non sono generalmente soddisfatte nella pratica. In realtà si presentano problemi come l'esistenza di discontinuità, la mancanza di rappresentazione analitica della funzione obiettivo e la diffusione del rumore.

In queste circostanze, l'applicabilità e l'efficienza degli algoritmi di ottimizzazione classici sono discutibili, dando origine alla necessità di sviluppare diversi metodi di ottimizzazione; per esempio la combinazione di algoritmi classici e casuali hanno offerto risultati buoni, tuttavia, sono necessarie delle ipotesi matematiche stringenti sulla funzione obiettivo. I primi approcci, come la "simulated annealing", offrivano la possibilità di risolvere problemi che per altri algoritmi erano laboriosi. Tuttavia, per un

certo numero di anni, sono rimasti al margine della letteratura, a causa dello sviluppo teorico limitato a quel tempo. Allo stesso tempo, un nuovo tipo di algoritmo stava lentamente emergendo. L'ispirazione derivava dallo studio dei meccanismi di adattamento nei sistemi naturali, come il DNA. Le operazioni sottostanti che supportano i meccanismi evolutivi secondo la "Darwinian biological theory" sono stati modellati e utilizzati per sviluppare soluzioni basate su criteri di ottimalità definiti dall'utente.

Gruppi di ricerca negli Stati Uniti e in Europa sono riusciti a perfezionare le prime varianti di questi algoritmi, introducendo una serie di approcci efficienti sotto il nome generico di "algoritmi evolutivi".

Gli studi teorici condotti hanno reso questi algoritmi promettenti alternative nei casi in cui gli approcci classici non fossero applicabili.

Negli anni 90 una nuova categoria di algoritmi è stata proposta. Piuttosto di modellare procedure evolutive a livello **microscopico** (DNA), i nuovi metodi hanno modellato la popolazione ad un livello **macroscopico**, cioè in termini di strutture sociali e comportamenti aggreganti. Anche in questo caso la natura ha offerto ispirazione agli scienziati; le società gerarchicamente organizzate di organismi semplici, come formiche, api e pesci, con una gamma limitata di risposte individuali, mostrano comportamenti affascinanti con tratti identificabili di intelligenza sociale nel suo complesso. La mancanza di un meccanismo di sintonizzazione e controllo centrale in tali sistemi ha scatenato la curiosità degli studiosi. Modelli semplificati di questi sistemi sono stati sviluppati e studiati attraverso simulazioni. Le loro dinamiche erano approssimate da metodi matematici simili a quelli usati nella fisica delle particelle, mentre la teoria della probabilità e i processi stocastici offrivano un solido background teorico per lo sviluppo di una nuova categoria di algoritmi classificati con il nome di "swarm intelligence" (intelligenza dello sciame). Particle swarm optimization appartiene a questa categoria. I suoi primi precursori erano simulatori di comportamenti sociali che implementavano regole come: la corrispondenza della velocità del nearest-neighbor, l'accelerazione per distanza, al fine di riprodurre comportamenti "swarming" in gruppi di agenti semplici. La prima versione di PSO fu pubblicata nel 1995 (*Eberhart & Kennedy, 1995*) e presentato come un algoritmo di ottimizzazione stocastica basato su modelli di simulazione sociale.

Fin dal suo sviluppo, PSO ha guadagnato un ampio riconoscimento grazie alle sue abilità nel fornire soluzioni efficienti, richiedendo solo uno sforzo minimo di implementazione. L'algoritmo utilizza una popolazione di individui che si muovono stocasticamente in uno spazio di ricerca. Allo stesso tempo, la posizione raggiunta da ciascun individuo tiene in considerazione della sua esperienza, conservata nella memoria. Questa esperienza viene quindi comunicata a una parte o a tutta la popolazione spingendo il suo movimento verso le regioni rilevanti e più promettenti. Lo schema di comunicazione è determinato da una rete sociale fissa o adattiva, la quale svolge un ruolo cruciale nella proprietà di convergenza dell'algoritmo.

Lo sviluppo dello PSO fu basato in concetti e regole che governano popolazioni socialmente organizzate in natura, popolazioni gerarchicamente organizzate di organismi semplici, come formiche, api e pesci, con una gamma molto limitata di risposte individuali; dove comportamenti aggregati si incontrano, producendo movimenti impressionanti, sincronizzati, senza collisioni. In tali sistemi/popolazioni, il comportamento di ciascun membro del gruppo si basa su risposte intrinseche semplici, sebbene il loro esito sia piuttosto complesso da un punto di vista macroscopico.

Ad esempio, il volo di uno stormo di uccelli può essere simulato con relativa precisione mantenendo semplicemente una distanza target tra ciascun uccello e i suoi vicini immediati. Questa distanza può dipendere dalle sue dimensioni e dal suo comportamento desiderabile.

Altro esempio è quello del pesce che mantiene una certa distanza dagli altri pesci quando nuota spensierato, mentre si concentra in gruppi molto densi in presenza di predatori. I gruppi possono anche reagire alle minacce esterne cambiando rapidamente la loro forma, rompendo le parti più piccole e riorganizzandosi, dimostrando una notevole capacità di rispondere collettivamente a stimoli esterni al fine di preservare l'integrità personale.

Fenomeni simili sono osservati nei sistemi fisici. Un tipico esempio è l'aggregazione delle particelle causata dall'attrazione diretta tra le particelle a causa del moto browniano o del taglio del fluido.

Tuttavia, queste interazioni possono diventare molto complesse, specialmente nello spazio di credenza, dove, in contrasto con lo spazio fisico, lo stesso punto (una credenza

o un'idea) può essere occupato contemporaneamente da ampi gruppi di persone senza collisioni. I suddetti comportamenti aggreganti, caratterizzati dalla semplicità dei sistemi animali e fisici o dall'astrattezza del comportamento sociale umano, hanno affascinato i ricercatori e hanno motivato la loro ulteriore indagine attraverso estese sperimentazioni e simulazioni (Heppner & Grenander, 1990; Reynolds, 1987; Wilson, 1975).

Nonostante le loro differenze fisiche o strutturali, tali sistemi condividono proprietà comuni, riconosciute come i cinque principi base dell'intelligenza dello sciame (Millonas, 1994):

- 1) Prossimità: capacità di eseguire calcoli dello spazio e del tempo.
- 2) Qualità: capacità di rispondere ai fattori di qualità ambientale.
- 3) Diverse risposte: capacità di produrre una pluralità di risposte diverse.
- 4) Stabilità: capacità di mantenere comportamenti robusti in presenza di lievi cambiamenti ambientali.
- 5) Adattabilità: capacità di cambiare comportamento quando è dettata da fattori esterni.

Inoltre, la condivisione sociale delle informazioni tra gli individui in una popolazione può fornire un vantaggio evolutivo.

Mettendolo in un quadro matematico, supponiamo che $A \subset R^n$ sia lo *spazio di ricerca*, e, $f: A \rightarrow Y \subseteq R$, sia la *funzione obiettivo*. Per mantenere le descrizioni il più semplici possibile, assumiamo che A sia anche lo *spazio ammissibile* del problema in questione, cioè non ci sono ulteriori vincoli espliciti posti sulle soluzioni candidate. Inoltre, non sono necessarie ulteriori ipotesi sulla forma della funzione obiettivo e dello spazio di ricerca perché PSO è un algoritmo basato sulla popolazione, cioè sfrutta una popolazione di potenziali soluzioni per indagare sulle probabilità di successo di un'azione. La popolazione è chiamata "sciame" (swarm) ed i suoi individui sono chiamati "particelle". Lo sciame è definito come un insieme $S = \{x_1, x_2, \dots, x_N\}$ di N particelle, definite come:

$$x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T \in A, i = 1, 2, \dots, N.$$

Gli indici sono assegnati arbitrariamente alle particelle, mentre N è il numero di parametri dell'algoritmo scelto dal ricercatore. Ogni particella ha un unico valore di

funzione, $f_i = f(x_i) \in Y$. Si assume che le particelle possano muoversi all'interno dello spazio di ricerca, A , in maniera iterativa.

Ciò è possibile modificando la loro posizione, controllando direzione e velocità dello spostamento.

La velocità verrà indicata con

$$v_i = (v_{i1}, v_{i2}, \dots, v_{in})^T, i = 1, 2, \dots, N.$$

La velocità è altrettanto adattata in modo iterativo per rendere le particelle capaci di visitare qualsiasi regione di A (potenzialmente). Se t indica il ciclo di iterazione, allora la posizione corrente della particella i -esima e la sua velocità saranno indicati con $x_i(t)$ e $v_i(t)$, rispettivamente.

La velocità viene aggiornata in base alle informazioni ottenute nei passaggi precedenti dell'algoritmo; questo è implementato in termini di memoria individuale (migliore posizione ottenuta dalla particella) e memoria sociale (migliore posizione globale trovata dallo sciame).

A questo scopo, oltre allo sciame S , che contiene le posizioni correnti delle particelle, l'algoritmo PSO mantiene anche un set di memoria $p = \{p_1, p_2, \dots, p_N\}$, che contiene le posizioni migliori mai visitate da ogni particella $i = (p_{i1}, p_{i2}, \dots, p_{in})^T \in A, i = 1, 2, \dots, N$.

Poiché l'algoritmo PSO si basa su modelli di simulazione del comportamento sociale, esso considera anche un meccanismo di scambio di informazioni per consentire alle particelle di comunicare reciprocamente la propria esperienza. Ciò si traduce memorizzando la migliore posizione mai visitata da tutte le particelle.

Sia g l'indice della posizione migliore con il valore più basso della funzione in P ad una determinata iterazione t , cioè:

$$p_g(t) = \arg \min_t f(p_i(t))$$

Quindi, la versione iniziale di PSO è definita dalle seguenti equazioni (Eberhart & Kennedy, 1995; Eberhart et al., 1996; Kennedy & Eberhart, 1995):

$$v_{ij}(t+1) = v_{ij}(t) + c_1 R_1 (p_{ij}(t) - x_{ij}(t)) + c_2 R_2 (p_{gj}(t) - x_{ij}(t)) \quad (1)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (2)$$

dove $i=1, 2, \dots, N$ e $j=1, 2, \dots, N$ e t denota l'iterazione; R_1 e R_2 sono delle realizzazioni di variabili casuali distribuite uniformemente tra $[0,1]$; e c_1, c_2 sono fattori di ponderazione (weighting factors), chiamati rispettivamente parametro cognitivo e sociale. Nella prima versione dell'algoritmo, veniva usato un singolo peso, $c = c_1 = c_2$, chiamato costante di accelerazione ma la possibilità di distinguere il parametro cognitivo da quello sociale ha offerto un controllo migliore dell'algoritmo.

Ad ogni iterazione, dopo l'aggiornamento della posizione della particella e la valutazione dei valori della funzione di fitness, sono aggiornate anche le migliori posizioni (memoria). Pertanto, la nuova migliore posizione (in caso di minimizzazione) di x_i all'iterazione $t+1$ è definita come segue:

$$p_i(t+1) = \begin{cases} x_i(t+1) & \text{if } f(x_i(t+1)) \leq f(p_i(t)), \\ p_i(t), & \text{otherwise.} \end{cases}$$

La determinazione della migliore posizione globale g completa un'iterazione di PSO.

Le particelle sono inizializzate casualmente, seguendo una distribuzione uniforme sullo spazio di ricerca, A . Questa scelta tratta ogni regione di A in modo equivalente; quindi è preferibile nei casi in cui non ci siano informazioni sulla forma della funzione obiettivo, che richiede un diverso schema di inizializzazione. Tale strategia è facilitata dal fatto che tutti i moderni sistemi informatici possono essere dotati di un generatore di numeri casuali uniforme.

Il precedente termine di velocità, $v_{ij}(t)$, nella parte destra dell'equazione (1), offre un mezzo di movimento inerziale alla particella prendendo in considerazione il suo spostamento dalla posizione precedente.

Il termine di velocità serve anche a "perturbare" la posizione della particella migliore globale. Infatti, se la particella, x_i , scopre una nuova posizione con un valore della funzione di fitness migliore, allora diventa la migliore posizione globale e la sua migliore posizione, p_i , coinciderà con p_g . Pertanto, i due termini stocastici nell'equazione (1) svaniranno (se $c_1 = c_2$).

Se nell'equazione (1) non esistesse un termine di velocità, allora la particella sopra menzionata rimarrebbe nella stessa posizione per diverse iterazioni, finché una nuova posizione migliore non venisse rilevata da un'altra particella. Contrariamente a ciò, il termine velocità consente a questa particella di continuare la ricerca, seguendo il suo

precedente spostamento di posizione. I valori di c_1 e c_2 possono influenzare la capacità di ricerca di PSO polarizzando le nuove posizioni campionate di una particella, x_i , verso le posizioni migliori, p_i e p_g , rispettivamente, nonché modificando la grandezza/magnitudine della ricerca.

Ad esempio, consideriamo i due casi illustrati in figura 3.1. Sia $x_i = (0,0)^T$, indicato con il simbolo della croce, la posizione corrente di una particella. Inoltre, sia, $p_i = (2,1)^T$ e $p_g = (1,3)^T$, la sua migliore posizione e la migliore posizione complessiva rispettivamente (indicati con una stella e un simbolo quadrato). Inoltre, per semplicità, si supponga che la sua velocità attuale, v_i , sia uguale a zero. La figura 3.1 rappresenta 1000 possibili nuove posizioni di x_i per $c_1 = c_2 = 1.0$ (parte sinistra) e $c_1 = c_2 = 2.0$ (parte destra).

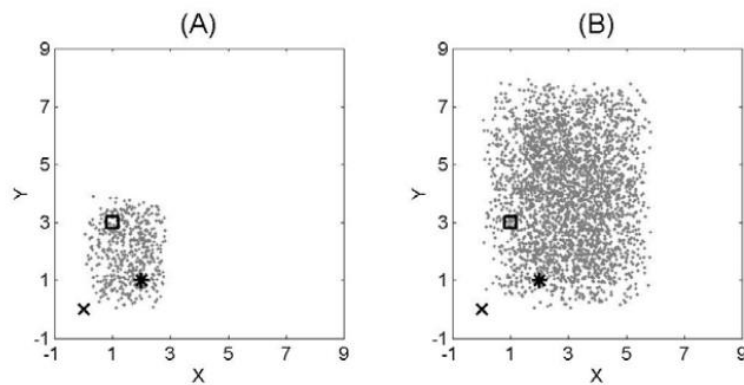


Figura 3.1: Candidate new positions of the particle $x_i = (0, 0)^T$ (cross) with $p_i = (2, 1)^T$ (star) and $p_g = (1, 3)^T$ (square), for the cases (A) $c_1 = c_2 = 1.0$, and (B) $c_1 = c_2 = 2.0$ (Fonte: PSO and Intelligence - Konstantinos E. Parsopoulos, Michael N. Vrahatis)

Apparentemente, la magnitudine della ricerca differisce significativamente nei due casi. Se è necessaria una migliore esplorazione globale, valori alti di c_1 e c_2 possono fornire nuovi punti in regioni relativamente distanti dello spazio di ricerca. D'altra parte, una ricerca locale più raffinata attorno alle migliori posizioni raggiunte finora richiederebbe la selezione di valori più piccoli per i due parametri. Inoltre, se $c_1 > c_2$, la ricerca va verso la direzione di p_i , mentre nel caso opposto, $c_1 < c_2$, sarebbe favorito il campionamento verso la direzione di p_g .

Questo effetto può essere utile nei casi in cui esistono informazioni riguardo alla forma della funzione obiettivo. Per esempio, nelle funzioni unimodali convesse, una scelta che promuove il campionamento più vicino a p_g dovrebbe essere più efficiente se combinata con un'adeguata magnitudine di ricerca.

Si sottolinea che il PSO opera indipendentemente su ciascuna direzione delle coordinate-un tipico errore commesso da diversi ricercatori, specialmente quando le equazioni di PSO sono state considerate nella loro forma vettoriale:

$$v_i(t+1) = v_i(t) + c_1 R_1 (p_i(t) - x_i(t)) + c_2 R_2 (p_g(t) - x_i(t)), \quad (3)$$

$$x_i(t+1) = x_i(t) + v_i(t+1), \quad (4)$$

$$i = 1, 2, \dots, N.$$

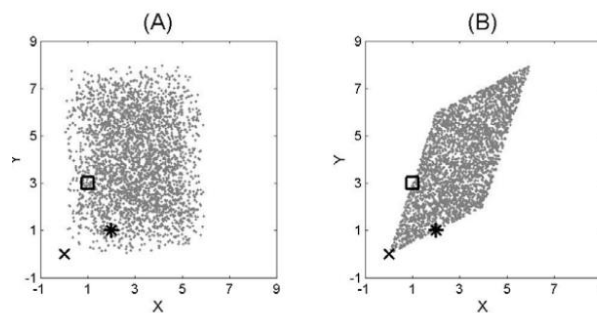


Figura 3.2: Example - Candidate new positions generated by equation (3) per la particella $x_i = (0,0)^T$ (croce), with $p_i = (2,1)^T$ (star), $p_g = (1,3)^T$ (square), e $c_1 = c_2 = 2.0$, quando (A) R_1 e R_2 sono vettori n -dimensionali casuali e (B) R_1 e R_2 sono valori monodimensionali casuali. (Fonte: PSO and Intelligence - Konstantinos E. Parsopoulos, Michael N. Vrahatis)

In questo caso, R_1 e R_2 devono essere considerati come vettori randomizzati n -dimensionali con i loro componenti distribuiti uniformemente all'interno di $[0,1]$. Invece, R_1 e R_2 sono spesso considerati come valori unidimensionali casuali, in modo simile all'equazione (1), risultando in uno schema che utilizza lo stesso numero casuale per tutte le componenti di direzione del vettore.

L'effetto di questa scelta è illustrato in figura 3.2, dove 1000 possibili nuove posizioni della stessa particella, x_i , come in figura 3.1, sono generate con la (corretta) configurazione dell'equazione (1) (parte sinistra) e la (errata) configurazione dell'equazione (3) con valori casuali usati al posto di vettori casuali (parte destra). Ovviamente, quest'ultimo caso limita il campionamento all'interno di una regione a parallelogramma tra le due migliori posizioni, p_i e p_g . Nella maggior parte delle applicazioni di ottimizzazione, è auspicabile o inevitabile considerare solo le particelle che si trovano nello spazio di ricerca. A questo scopo, i limiti sono imposti sulla posizione di ogni particella, x_i , per restringere lo spazio di ricerca, A. Se una particella assume un valore indesiderato, cioè esterno allo spazio di ricerca, dopo l'applicazione

dell'equazione (2), il suo valore viene immediatamente riportato al suo confine. Nel caso più semplice, lo spazio di ricerca può essere definito come la seguente “box”:

$$A = [a_1, b_1] * [a_2, b_2] * \dots * [a_n, b_n],$$

$$a_i, b_i \in R, i=1, 2, \dots, n, \text{ e le particelle come di seguito:}$$

$$x_{ij}(t+1) = \begin{cases} a_j, & \text{se } x_{ij}(t+1) < a_j, \\ b_j, & \text{se } x_{ij}(t+1) > b_j, \end{cases} \quad (5)$$

$$i=1, 2, \dots, N; j=1, 2, \dots, n.$$

In alternativa, è stato considerato un movimento di rimbalzo dal confine nello spazio di ricerca. La popolarità di questo approccio è limitata, poiché richiede la modellazione del movimento delle particelle con equazioni fisiche complesse. Nei casi in cui A non può essere definito come una scatola (box), per limitare le particelle è necessario porre delle condizioni speciali dipendenti dal problema.

Le prime varianti di PSO sono state eseguite in modo soddisfacente per problemi di ottimizzazione semplici. Tuttavia, le loro carenze cruciali sono state rivelate non appena sono state applicate a problemi più complessi con ampi spazi di ricerca e una moltitudine di minimi locali. Nei paragrafi seguenti vengono riportati e discussi i perfezionamenti sviluppati per risolvere le carenze del modello PSO originale.

3.3.1 Swarm Explosion e Velocity Clamping

Il primo problema significativo, verificato da diversi ricercatori, è stato l'effetto dell'esplosione dello sciame. Ci si riferisce all'incremento incontrollato della magnitudine delle velocità, con conseguente divergenza di sciame. Questo problema è dovuto alla mancanza di un meccanismo per restringere le velocità nelle prime versioni di PSO, ed è stata affrontata direttamente usando limiti rigorosi per limitare la velocità a livelli desiderabili, impedendo alle particelle di compiere spostamenti estremamente grandi dalla loro posizione corrente. Più specificamente, viene considerata una soglia di velocità massima definita dall'utente, $v_{max} > 0$. Dopo aver determinato la nuova velocità di ogni particella con l'equazione (1), prima dell'aggiornamento della posizione con l'equazione (2) si applicano le seguenti restrizioni:

$$|v_{ij}(t+1)| \leq v_{max} \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, n. \quad (6)$$

In caso di violazione, la velocità corrispondente del componente è posta direttamente nel limite di velocità più vicina, cioè:

$$v_{ij}(t+1) = \begin{cases} v_{max}, & \text{se } v_{ij}(t+1) > v_{max} \\ -v_{max}, & \text{se } v_{ij}(t+1) < -v_{max} \end{cases}$$

Se necessario, è possibile utilizzare diversi limiti di velocità per componente di direzione. Il valore di v_{max} viene generalmente considerato come una frazione della dimensione dello spazio di ricerca per direzione. Pertanto, se lo spazio di ricerca è definito come nell'equazione (5), una velocità massima comune per tutti i componenti di direzione può essere definita come segue:

$$v_{max} = \frac{\min_i \{b_i - a_i\}}{k}$$

In alternativa, le soglie di velocità massima separate per componente possono essere definite come:

$$v_{max,i} = \frac{b_i - a_i}{k}, \text{ dove } i = 1, 2, \dots, n,$$

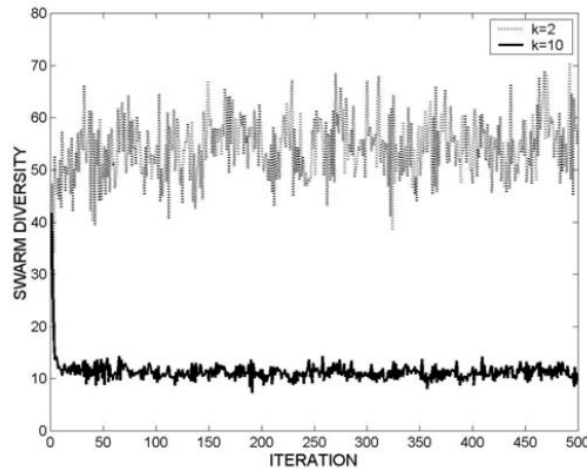


Figura 3.3: Example - Swarm diversity during search for $k=2$ (dotted line) and $k=10$ (solid line). The plots pertain to a swarm of 20 particles with $c_1=c_2=2$, minimizing the 2-dimensional instance of a test problem, range $[-100, 100]^2$ for 500 iterations. (Fonte: PSO and Intelligence - Konstantinos E. Parsopoulos, Michael N. Vrahatis)

Supponendo con $k=2$. Se il problema richiede degli step più piccoli della particella, allora devono essere utilizzati k con valori maggiori. Per esempio, se lo spazio della ricerca ha una moltitudine di minimi con regioni ristrette di attrazioni vicina l'una dall'altra, allora

la k deve assumere dei valori sufficientemente grandi per prevenire alle particelle di sorvolare su di loro.

D'altra parte, k non deve assumere valori molto piccoli o molto grandi che possono impedire di avere un soddisfacente avanzamento della ricerca.

Il limite della velocità offre una soluzione semplice ma efficace al problema dell'esplosione dello sciame. Tuttavia, non risolve il problema della convergenza. Le particelle sono in grado di fluttuare attorno alle loro posizioni migliori, ma non sono in grado né di raggiungere la convergenza su una posizione promettente né di eseguire una raffinata ricerca attorno ad essa. Questo problema è stato risolto introducendo un nuovo parametro nel modello PSO originale, come descritto di seguito.

3.3.2 Il concetto di peso inerziale

Sebbene la soglia di velocità massima abbia migliorato le prestazioni delle prime versioni dell'algorithm PSO, questa non era adeguata a rendere l'algorithm efficiente in problemi di ottimizzazione complessi. L'utilizzo della soglia, infatti, evita l'esplosione di sciame, ma non è in grado di concentrare le particelle attorno alle soluzioni più promettenti. In altre parole, ha evidenziato una incapacità nel controllare le velocità. La ricerca nelle regioni promettenti, cioè intorno alle posizioni migliori, richiede una forte attrazione delle particelle verso di esse, e ciò è possibile riducendo le perturbazioni che allontanano le particelle stesse dalle posizioni migliori. Pertanto, l'effetto della velocità al tempo t deve svanire con il passare del tempo. A tale scopo, un nuovo parametro, w , chiamato *peso inerziale* è stato introdotto nell'equazione (1), risultando in una nuova variante PSO (Eberhart & Shi, 1998):

$$v_{ij}(t+1) = wv_{ij}(t) + c_1R_1(p_{ij}(t)-x_{ij}(t)) + c_2R_2(p_{gj}(t)-x_{ij}(t)), \quad (7)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1), \quad (8)$$

$$i = 1, 2, \dots, N, \quad j = 1, 2, \dots, n.$$

Il resto dei parametri rimane lo stesso della prima variante delle equazioni (1) e (2). Il peso inerziale deve essere selezionato in modo tale che l'effetto di $v_{ij}(t)$ svanisca durante l'esecuzione dell'algorithm. Pertanto, è preferibile un valore di w decrescente nel tempo. Una scelta molto comune è l'inizializzazione di w su un valore leggermente superiore a 1 (ad esempio 1,2) per promuovere l'esplorazione nelle prime fasi di

ottimizzazione e una diminuzione lineare verso zero per eliminare comportamenti oscillatori nelle fasi successive. Di solito viene fissato anche un valore minimo strettamente positivo di w (ad esempio 0.1) per evitare che il precedente termine di velocità svanisca. In generale, uno schema lineare decrescente per w può essere matematicamente descritto come di seguito:

$$w(t) = w_{up} - (w_{up} - w_{low}) \frac{t}{T_{max}} \quad (9)$$

dove t indica il contatore di iterazioni; w_{low} e w_{up} sono i limiti inferiori e superiori desiderabili di w ; and T_{max} è il numero totale di iterazioni consentito. L'equazione (9) produce un peso inerziale dipendente dal tempo linearmente decrescente con valore iniziale, w_{up} , all'iterazione, $t=0$ e valore finale, w_{low} , all'ultima iterazione, $t=T_{max}$. La figura 3.4 illustra diversità per uno sciame di 20 particelle aggiornato con le equazioni (7) e (8), con $c1 = c2 = 2$, riducendo al minimo l'istanza bidimensionale di un problema testato (esempio nel grafico sotto), nell'intervallo $[-100,100]$, con e senza peso inerziale decrescente.

In figura 3.4, la diversità è definita come il valore medio delle deviazioni standard delle particelle per direzione delle coordinate. L'uso del peso inerziale ha un enorme effetto sulla diversità degli sciami, che svanisce dopo quasi 300 iterazioni, in contrasto con il caso del semplice "bloccaggio" della velocità (velocity clamping), che mantiene quasi gli stessi livelli di diversità durante la ricerca. Inoltre, osserviamo che, nel caso di utilizzo del parametro inerziale, si verifica un aumento della diversità degli sciami per le prime quasi 100 iterazioni. Questo effetto può essere attribuito al valore iniziale, $w_{up}=1.2$, del peso inerziale. Poiché questo valore è maggiore di 1.0 il termine di velocità precedente ha un impatto maggiore nell'equazione (7) rispetto all'equazione (1). Ciò si verifica in un'esplosione temporanea di sciami che migliora la capacità di esplorazione anche di uno sciame poco inizializzato. Dopo quasi 90 iterazioni, il peso inerziale assume valori inferiori a 1.0 e la diversità inizia a diminuire verso lo zero, promovendo così lo sfruttamento (exploitation).

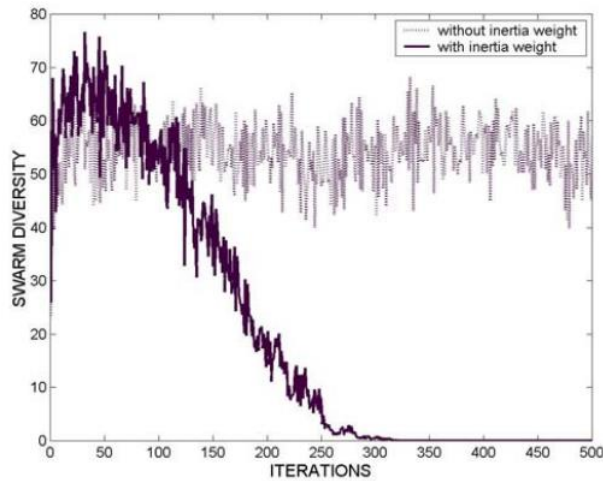


Figura 3.4: Example – Swarm diversity during search, with (solid line) and without (dotted line) inertia weight, The plots pertain to a swarm of 20 particles with $c_1=c_2=2$, minimizing the 2-dimensional instance of a test problem, in the range $[-100,100]^2$ for 500 iterations and $v_{max}=50$. The inertia weight decreases linearly from 1.2 to 0.1 (Fonte: PSO and Intelligence - Konstantinos E. Parsopoulos, Michael N. Vrahatis)

Per produrre comportamenti diversi dell'algoritmo è possibile utilizzare diversi schemi per l'adattamento del peso inerziale. Il miglioramento delle prestazioni ottenute utilizzando la variante del PSO appena descritta, ha reso l'approccio PSO popolare per alcuni anni. Tuttavia, rimane non risolto il problema del rischio di convergenza in minimi locali, specialmente in problemi complessi. Questa carenza è stata risolta introducendo uno schema più sofisticato di condivisione delle informazioni tra le particelle, come descritto nella sezione seguente.

3.3.3 Il concetto del neighborhood

Come evidenziato alla fine del precedente paragrafo, in ambiti complessi e multimodali l'introduzione del peso inerziale non aumenta l'efficienza a livelli soddisfacenti. Come illustrato nella figura 3.4, dopo una serie di iterazioni, lo sciame collassa a causa della completa perdita di diversità. Questo implica che non è possibile un'ulteriore esplorazione e che le particelle possono eseguire solo ricerche locali intorno al loro punto di convergenza, che molto probabilmente si trova nelle vicinanze della migliore posizione generale. Sebbene l'effetto della convergenza rapida possa essere lieve in problemi semplici di ottimizzazione specialmente nelle funzioni unimodali e convesse, diventa dannoso in condizioni complesse. Questa carenza può essere attribuita allo schema globale di scambio di informazioni che consente a ciascuna particella di conoscere istantaneamente la migliore posizione generale ad ogni iterazione.

Utilizzando questo schema, tutte le particelle assumono nuove posizioni in regioni legate alla stessa migliore posizione generale, riducendo la capacità di esplorazione dello sciame. Il problema è stato affrontato introducendo il concetto del "*neighborhood*". L'idea principale era la riduzione dello schema globale di scambio di informazioni a uno locale, in cui le informazioni sono diffuse solo in piccole parti dello sciame ad ogni iterazione. Più precisamente, ogni particella assume che un altro insieme di particelle siano i suoi vicini e, ad ogni iterazione, comunica la sua posizione migliore solo a queste particelle, anziché all'intero sciame. Pertanto, le informazioni relative alla migliore posizione vengono inizialmente comunicate solo al "vicinato" della migliore particella e successivamente al resto attraverso i loro vicini.

Sia x_i la i -esima particella di uno sciame, $S = \{x_1, x_2, \dots, x_N\}$. Quindi un vicinato di x_i è definito come un insieme: $NB_i = \{x_{n1}, x_{n2}, \dots, x_{ns}\}$. Dove, $\{n1, n2, \dots, ns\} \subseteq \{1, 2, \dots, N\}$ è l'insieme di indici dei suoi vicini. La cardinalità, $|NB_i|$, di questo insieme è chiamato "*neighborhood size*". Sia g_i l'indice della particella migliore in NB_i , allora possiamo scrivere:

$$p_{g_i} = \underset{j \text{ such that } x_j \in NB_i}{\operatorname{argmin}} f(p_j)$$

le equazioni sono modificate nel seguente modo:

$$v_{ij}(t+1) = wv_{ij}(t) + c_1R_1(p_{ij}(t) - x_{ij}(t)) + c_2R_2(p_{g_{ij}}(t) - x_{ij}(t)) \quad (10)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1), \quad (11)$$

$$i = 1, 2, \dots, N, \quad j = 1, 2, \dots, n.$$

Secondo equazione (10), la particella si sposterà verso la sua posizione migliore, così come verso la migliore posizione del suo vicinato, piuttosto che nella posizione migliore generale come avveniva nell'equazione (7).

Lo schema per determinare i "vicini" di ciascuna particella è chiamato "*neighborhood topology*". Uno schema comune che viene in mente è la formazione dei neighborhoods basato sulle distanze attuali delle particelle nello spazio di ricerca. In base a ciò, a ciascuna particella sarebbe assegnato un "vicinato" costituito da un numero, s , delle particelle che si trovano più vicino alla sua posizione attuale. Sebbene semplice, questo schema richiede il calcolo di $N(N + 1)/2$ distanze tra le particelle ad ogni iterazione, dato che lo spazio di ricerca è dotato di una metrica corretta. Si mostra una tendenza generale

a formare "particle clusters" che possono essere facilmente intrappolati nei minimi locali. Per queste ragioni, la "neighborhood topology" non è stata stabilita come la soluzione più promettente.

Per alleviare gli effetti di raggruppamento delle particelle prodotte dalle "neighborhood topology" basate sulla distanza è stata promossa l'idea di formare vicinati basati su criteri arbitrari. L'alternativa più semplice e direttamente applicabile era la formazione di quartieri basati su indici delle particelle. In base a ciò, la i -esima particella assume i vicini con indici vicini.

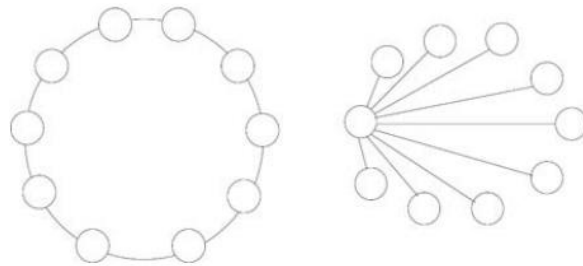


Figura 3.5: Common neighborhood topologies of PSO - ring (left) and star(right) (Fonte: PSO and Intelligence - Konstantinos E. Parsopoulos, Michael N. Vrahatis)

Pertanto, il vicinato di x_i può essere definito come: $NB_i = \{x_{i-r}, x_{i-r+1}, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{i+r-1}, x_{i+r}\}$ come se le particelle fossero adagiate su un anello e ognuna fosse collegata solo ai suoi vicini immediati. Questo schema è illustrato nella figura 3.5 (a sinistra) e si chiama topologia ad anello, mentre il parametro r che determina la dimensione del vicinato viene chiamato "raggio del vicinato". Ovviamente, gli indici sono considerati riciclabili in questa topologia, i.e., l'indice $i=1$ segue immediatamente dopo l'indice $i=N$ sull'anello. La variante PSO che utilizza la posizione migliore complessiva dello sciame può essere considerata un caso speciale del suddetto schema ad anello, in cui ogni vicinato è l'intero sciame, vale a dire $NB_i \equiv S$, per tutti gli $i=1, 2, \dots, N$. Per distinguere tra i due approcci, la variante che utilizza la migliore posizione complessiva è chiamata variante PSO globale (spesso indicata come $gbest$), mentre quella con vicinati strettamente più piccolo è chiamato variante PSO (indicata, rispettivamente, come migliore). Lo schema $gbest$ è anche chiamato topologia a stella (star topology), ed è rappresentato graficamente nella figura 3.5 (a destra) in cui tutte le particelle comunicano con la migliore. L'effetto dell'utilizzo di $lbest$ invece di $gbest$ sulla diversità degli sciame è illustrato nella figura 3.6 per uno sciame di 20 particelle con $c_1=c_2=2$ e peso inerziale decrescente, con $T_{max} = 500$,

$w_{up} = 1.2$, w_{low} e $v_{max} = 5.12$, minimizzando la funzione tridimensionale di Rastrigin⁷. La linea continua rappresenta la diversità di sciami (come definito prima) per lbest con topologia anello e raggio $r=1$; la linea tratteggiata rappresenta il caso gbest corrispondente. Esiste un'apparente differenza nelle diversità degli sciami tra gbest e lbest, che diventa più intensa all'aumentare della dimensione del problema. Sebbene la topologia anello sia semplice ed efficiente, in letteratura sono state proposte diverse topologie (Kennedy,1999; Mendes et al., 2003). Inoltre, la topologia può cambiare con il tempo anziché rimanere fissa durante un processo di ottimizzazione. Ogni particella può avere una propria topologia individuale (fissa o dinamica), offrendo una elevata flessibilità all'utente, con capacità di adattarsi a qualsiasi requisito speciale del problema. Tuttavia, la maggioranza dei migliori modelli in letteratura si basa sulla topologia anello; quindi, può essere considerata una scelta standard per le varianti di PSO locali.

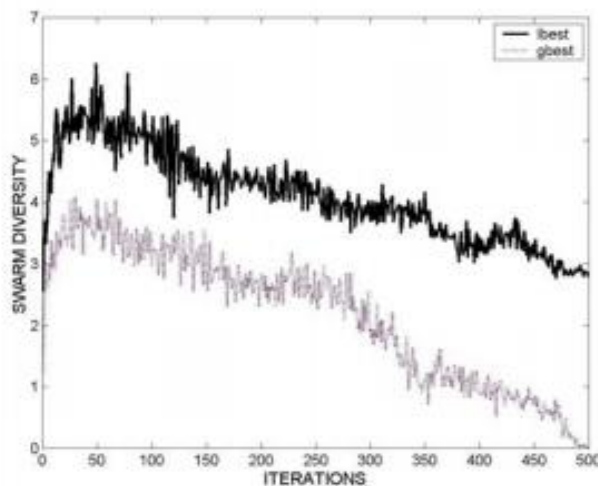


Figura 3.6: Example – Swarm diversity during search using lbest (solid line) and the gbest (dotted line) PSO variant. The plots pertain to a swarm of 20 particles with $c_1 = c_2 = 2$, minimizing the 10-dimensional Rastrigin function in the range $[-5.12, 5.12]$ ¹⁰ for 500 iterations and $v_{max} = 5.12$. The inertia weight decreases linearly from 1.2 to 0.1, while ring topology with radius $r = 1$ is used for the lbest case (Fonte: PSO and Intelligence - Konstantinos E. Parsopoulos, Michael N. Vrahatis)

⁷ La funzione di Rastrigin è una funzione non convessa utilizzata come una performance test problem per algoritmi di ottimizzazione. È un tipico esempio di funzione multimodale non lineare. Su un dominio n-dimensionale è definito da: $f(x) = An + \sum_{i=1}^n [x_i^2 - A \cos(2\pi x_i)]$ dove $A=10$ e $x_i \in [-5.12, 5.12]$. Ha un massimo globale in $x=0$ dove $f(x)=0$.

3.4 PSO STANDARD CONTEMPORANEA

L'efficienza e la semplicità dell'algoritmo PSO hanno consentito agli scienziati di varie discipline, con un background limitato in informatica e programmazione, di utilizzare PSO come uno strumento di ottimizzazione efficiente in applicazioni in cui i metodi di ottimizzazione classici erano inefficienti.

Fare un'analisi tecnica sarebbe un compito difficile: il PSO incorpora elementi stocastici ma la ricerca non si basa su distribuzioni di probabilità. Pertanto, un'analisi probabilistica diretta basata su distribuzioni adattive non è possibile. A tal fine, sono state inizialmente studiate approssimazioni deterministiche del modello originale di PSO, mentre la componente stocastica è stata introdotta nei modelli studiati come fattore di perturbazione dei sistemi deterministici considerati.

Ozcan and Mohan (1999) hanno pubblicato la prima indagine teorica in spazi multidimensionali, fornendo equazioni a forma chiusa per traiettorie di particelle. Il loro studio si è concentrato sul primo modello PSO di equazioni (1) e (2), e hanno mostrato che le particelle si muovono su onde sinusoidali per ogni coordinata dello spazio di ricerca, mentre la componente stocastica offre un mezzo per manipolarne la frequenza e l'ampiezza. Successivamente Clerc e Kennedy (2002), hanno preso in considerazione diversi modelli generalizzati di PSO e hanno condotto un'analisi dinamica della loro convergenza.

L'analisi di Clerc e Kennedy ha offerto un background teorico solido all'algoritmo e ha stabilito uno dei modelli investigati come variante PSO contemporanea di default.

Questo modello è definito dalle seguenti equazioni:

$$v_{ij}(t+1) = \chi [v_{ij}(t) + c_1 R_1 (p_{ij}(t) - x_{ij}(t)) + c_2 R_2 (p_{gj}(t) - x_{ij}(t))], \quad (12)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1), \quad (13)$$

$$i = 1, 2, \dots, N, \quad j = 1, 2, \dots, n.$$

dove χ è un parametro chiamato "coefficiente di costrizione". Questa variante di PSO è algebricamente equivalente alla variante con peso inerziale definita dalle equazioni (7) e (8).

Tuttavia, si distingue in letteratura per le sue proprietà teoriche che implicano la seguente selezione esplicita dei suoi parametri (Clerc & Kennedy, 2002):

$$X = \frac{2}{|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|}$$

dove $\varphi = c_1 + c_2$, e $\varphi > 4$.

Il meccanismo di apprendimento della PSO tradizionale è semplice e facile da implementare, ma soffre di alcuni problemi, come i fenomeni di "oscillazione" e "due passi in avanti, un passo indietro" (*Oscillatory Particle Swarm Optimizer - Haiyan Shi, Shilong Liu, Hongkun Wu, Ruowei Li, Sanchi Liu, Ngaiming Kwok, Yeping Peng, 2018*).

3.4.1 Lavori correlati alla Particle Swarm Optimization

Dalla sua prima introduzione, sono stati risolti molti problemi di ottimizzazione ingegneristica utilizzando l'algoritmo PSO. Ad esempio, il problema del "Profilo aerodinamico" è stato completato con successo utilizzando un PSO accoppiato con una mutazione ispirata alla natura in modo adattivo. Inoltre, è stato utilizzato un PSO binario per massimizzare la potenza in uscita da un parco eolico, collocando le turbine eoliche nelle migliori posizioni. Il PSO è stato utilizzato anche per la correzione del colore ottimale nelle immagini digitali e il miglioramento del contrasto.

In molte varianti del PSO, lo scambio di conoscenze è stato migliorato incorporando una maggiore complessità, ma questa strategia non è sempre auspicabile. Una modifica semplice ed efficace è quella di regolare il peso inerziale in base ad alcune funzioni predeterminate. Tuttavia, i cambiamenti nel peso inerziale potrebbero influenzare in modo critico le traiettorie che le particelle seguono. Diverse scelte potrebbero far seguire alla particella una traiettoria esponenziale di decadimento o oscillazione verso la soluzione ottimale.

Un *Oscillatory Particle Swarm Optimizer* è stato proposto per migliorare ulteriormente le prestazioni dell'algoritmo PSO. A differenza degli approcci precedentemente illustrati, si prevede che per trovare buone soluzioni, le particelle debbano coprire il più possibile lo spazio di ricerca invece di convergere verso la soluzione sub-ottimale istantanea. Questo comportamento è ottenuto mediante una corretta impostazione dei parametri PSO. Le scelte dei fattori cognitivi e di apprendimento sociale sono fatte assicurando che la ricerca non diverga dalla migliore soluzione. Inoltre, un percorso oscillatorio è garantito da un'adeguata impostazione del peso inerziale.

Il PSO è stato applicato con successo anche a problemi di ottimizzazione meccanica e strutturale e di ottimizzazione multi-obiettivo, l'addestramento artificiale della rete neurale e sistemi di controllo fuzzy. C'è stata una crescita esponenziale del numero di pubblicazioni che riportano le applicazioni della PSO in diversi campi scientifici tra cui elaborazione dei segnali, grafica, robotica e così via.

Il PSO è stato utilizzato anche per la soluzione di problemi di selezione di portafoglio che sono problemi di ottimizzazione vincolata. Il problema di ottimizzazione vincolata può essere riformulato in termini di problema di ottimizzazione non vincolata mediante il ricorso ad una tecnica detta di penalizzazione. Ciò richiede di riformulare opportunamente i vincoli (*Marco Corazza, Giacomo di Tollo, Giovanni Fasano, Raffaele Pesenti A PSO-based framework for nonsmooth portfolio selection problems, Neural Advances in Processing Nonlinear Dynamic Signals, Cham, Springer Science and Business Media Deutschland GmbH, vol. 102, pp. 265-275*).

3.4.2 Due macro-categorie per la performance del PSO

Al fine di migliorare la performance del PSO, sono state proposte numerose varianti dell'algoritmo.

Questi varianti possono essere classificati in 2 categorie:

1) self - change strategy (strategie di auto-cambiamento). Per esempio, *Shi e Eberhart* nel 1998 hanno proposto il peso inerziale con linearità decrescente. Il peso inerziale con controllo fuzzy è stato proposto nel 2001 da *Shi e Eberhart*. Strategie di controllo per i parametri incorporati nell'equazione del movimento delle particelle sono state proposte da *Shi, Y., Eberhart, R.C. (Fuzzy adaptive particle swarm optimization Proceedings of the 2001 Congress on Evolutionary Computation, 2001. Vol. 1. IEEE, pp. 101–106)*.

Un altro esempio è il comprehensive learning PSO (CLPSO) il quale prevede che la regola di aggiornamento della velocità di ogni particella sia basata su diverse posizioni storiche personali migliori (*Liang, J.J. Qin, A.K., Suganthan, P.N., Baskar, S., 2006. Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. IEEE Trans. Evol. Comput. 10 (3), 281–295*);

2) combinazione PSO e altre strategie. Per esempio, la strategia ortogonale viene introdotta in una PSO per scoprire informazioni più utili; più utili perché il disegno

sperimentale ortogonale (OED) offre una capacità di scoprire i migliori livelli di combinazione per diversi fattori, con un numero ragionevolmente piccolo di campioni sperimentali. Gli esperimenti hanno mostrato che la strategia OL (orthogonal learning) può dare alla PSO più robustezza e affidabilità nei problemi di ottimizzazione numerica (Zhan, Z.-H., Zhang, J., Li, Y., Shi, Y.-H., 2011. *Orthogonal learning particle swarm optimization. IEEE Trans. Evol. Comput.* 15 (6), 832–847).

Un altro esempio è l'aging mechanism; in natura, quasi tutti gli organismi invecchiano e hanno una durata limitata. L'invecchiamento è stato esplorato dai biologi come un meccanismo importante per mantenere la diversità. In una colonia animale sociale, l'invecchiamento rende debole il vecchio leader della colonia, offrendo agli altri individui l'opportunità di sfidare la posizione di comando. Ispirato da questo fenomeno naturale i ricercatori inseriscono il meccanismo dell'invecchiamento nel PSO e propongono una PSO con un aging leader e sfidanti. L'aging mechanism (il meccanismo dell'invecchiamento) introdotto nella PSO è stato utilizzato per cambiare la migliore posizione storica nell'intero sciame ed è stato mostrato che questo metodo migliora la performance di ottimizzazione del PSO (Chen, W.-N., Zhang, J., Lin, Y., Chen, N., Zhan, Z.-H., Chung, H.S.-H., Li, Y., Shi, Y.-H., 2013. *Particle swarm optimization with an aging leader and challengers. IEEE Trans. Evol. Comput.* 17 (2), 241–258).

I meccanismi di insegnamento sociale sono stati introdotti nel 2015. In questa versione, altre particelle migliori basate nella current swarm sono state utilizzate per guidare ogni particella a muoversi. È stato mostrato che questo approccio era migliore dello PSO standard. Le tecniche multi-swarm sono altrettanto un buon metodo. Grazie alle informazioni scambiate tra i diversi gruppi, le tecnologie multi-swarm possono equilibrare l'abilità di exploiting e di esplorazione dell'algoritmo (Cheng, R., Jin, Y., 2015. *A social learning particle swarm optimization algorithm for scalable optimization. Inform. Sci.* 291, 43–60).

Nel 2005 Liang e Suganthan hanno proposto una multiswarm PSO dinamica (DMS-PSO). In DMS-PSO, gli swarms sono stati raggruppati continuamente, l'intera swarm ha mantenuto una diversità migliore (Liang, J., Suganthan, P.N., 2005. *Dynamic multi-swarm particle swarm optimizer. In: Swarm Intelligence Symposium, 2005. SIS 2005. Proceedings 2005 IEEE. IEEE, pp. 124–129*).

Una PSO multistrato (MLPSO) è stato proposto nel 2014. La MLPSO può evitare la caduta nella local optimum crescendo il numero di strati di sciame (swarm layers). La strategia multistrato può essere considerata una tecnica a sciame multiplo (Wang, L., Yang, B., Chen, Y., 2014. *Improving particle swarm optimization using multilayer searching strategy. Inform. Sci.* 274, 70–94).

L'operatore crossover si considera come uno dei principali operatori di ricerca nell'algoritmo genetico, perché sfrutta le informazioni disponibili in precedenti campioni per influenzare le ricerche future. L'operatore di crossover può combinare parti di buone soluzioni per formare nuove soluzioni potenziali. L'operatore di crossover può intensificare la condivisione di informazione tra le particelle e prevenire la convergenza precoce della swarm (Singh, D., Singh, V., Ansari, U., 2011. *Binary particle swarm optimization with crossover operation for discrete optimization. Int. J. Comput. Appl.* 28 (11), 15–20).

3.4.3 Limiti della PSO

L'ultima revisione dettagliata e comprensiva del PSO è stata pubblicata nel 2017 da Zbigniew Michalewicz e da Mohammad Reza Bonyadi ("*Particle Swarm Optimization for single objective continuous space problems*"). Per stimare la crescita del numero di pubblicazioni in questo campo, gli autori sopra citati hanno condotto una ricerca per l'esatta corrispondenza di "particle swarm optimization" nel titolo dei documenti in cinque database scientifici riconosciuti (Scopus, Google Scholar, Springer, Web of Science, and IEEE Xplore). Il rapporto tra il numero di articoli pubblicati dal 2008 e quelli pubblicati prima del 2008 in Scopus, Google scholar, Springer, Web of science e IEEE Xplore è 3.07, 2.39, 3.09, 2.34 e 2.23, rispettivamente.

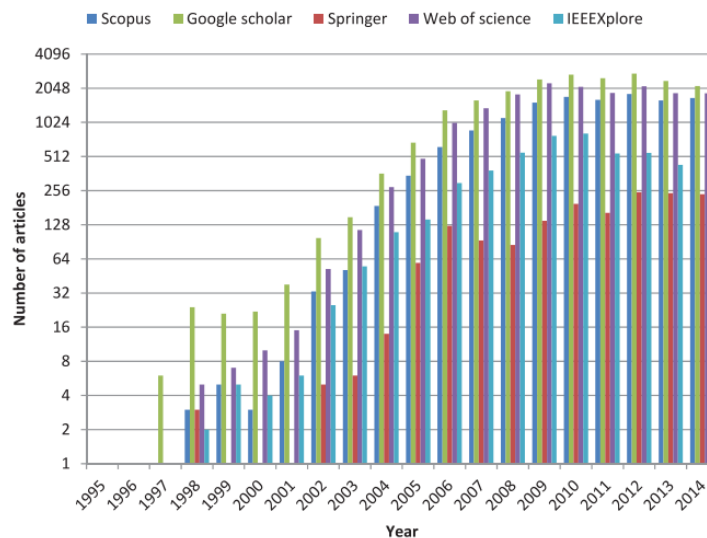


Figura 3.7: The number of published articles with “particle swarm optimization” in their title in Scopus, Google Scholar, Springer, Web of Science, and IEEE Xplore databases (Fonte: Zbigniew Michalewicz, Mohammad Reza Bonyadi “Particle Swarm Optimization – a review”)

A causa della sua facilità di implementazione e di rapida convergenza verso soluzioni accettabili, la PSO ha ricevuto ampia attenzione negli ultimi anni.

Però diverse limitazioni sono state identificate negli standard PSO. Queste limitazioni sono legate alla sensibilità della ricerca alle trasformazioni (rotazione, traslazione e scala), convergenza locale, stabilità, first hitting time, e bias. Pertanto, vi è la necessità di rivedere i progressi che affrontano questi limiti in dettaglio. Il termine "limitazione" si riferisce a un problema che è stato dimostrato che impedisce l'algoritmo di funzionare bene in diversi aspetti dell'operazione, come trovare soluzioni di alta qualità o ad essere stabile.

Ci sono due aree principali correlate ai limiti del PSO: convergenza e invarianza di trasformazione (per dettagli di queste limitazioni vedere Wilke, D. (2005). *Analysis of the particle swarm optimization algorithm. Master's thesis, Department of Mechanical and Aeronautical Engineering, University of Pretoria, South Africa & Bonyadi, M. R., and Michalewicz, Z. 2014. A locally convergent rotationally invariant particle swarm optimization algorithm. Swarm Intelligence, 8(3):159–198).*

3.4.4 Modifiche del PSO

I vari studi che sono stati svolti per quanto riguarda la modifica dello PSO si possono dividere in 3 diverse categorie: 1) impostazione dei parametri - si riferisce

all'impostazione di topologia, coefficienti (coefficienti di accelerazione o del peso inerziale), e dimensioni della popolazione; 2) modifica dei componenti dell'algoritmo - si riferisce alle variazioni della regola di aggiornamento della velocità o della posizione (inclusa l'aggiunta di nuovi componenti; la modifica del modo in cui vengono calcolati); 3) e combinazione dell'algoritmo con altri algoritmi – si riferisce all'ibridazione del PSO con altri metodi.

3.4.5 Potenziali ricerche da poter sviluppare

Per quanto riguarda le proprietà di convergenza di PSO: analizzare diversi comportamenti di particella prima della convergenza e dell'analisi della convergenza di altre varianti di PSO costituisce una potenziale ricerca da sviluppare. Inoltre, non ci sono molti studi sull'analisi "first hitting time" delle varianti del PSO, un'altra potenziale ricerca. Un'altra importante area di studio in PSO è la proprietà di invarianza della trasformazione dell'algoritmo. In effetti, non ci sono molte varianti PSO che sono invarianti di trasformazione, il che rende questo argomento uno spazio aperto per nuove idee e analisi.

3.5 FIREWORKS ALGORITHM

In questo paragrafo si introduce un nuovo swarm intelligence algorithm chiamato Fireworks Algorithm (FWA). Come il PSO, la maggior parte dei swarm intelligence algorithms sono ispirati da alcuni comportamenti intelligenti delle colonie di animali o insetti in natura.

L'algorithmo dei fireworks, abbreviato in FWA, ispirato dall'esplosione dei fireworks nel cielo notturno, è stato proposto da Ying Tan nel 2010. FWA è un algorithmo di ottimizzazione dell'intelligenza dello sciame, che sembra efficace nel trovare una soluzione abbastanza buona dell'ottimo globale, di un problema complesso di ottimizzazione. Nell'algorithmo FWA, quando esplose un firework, nell'area adiacente viene mostrata una " pioggia " di scintille. Queste scintille esplodono di nuovo e generano altre piogge di scintille in un'area più piccola. Gradualmente, le scintille cercano sull'intero spazio della soluzione e si concentrano su una piccola regione per trovare una soluzione ragionevolmente buona. FWA è in grado di soddisfare tre requisiti forniti da Storn e Price nel 1997.

Innanzitutto, FWA può elaborare funzioni di test lineari, non lineari e multi-modello.

In secondo luogo, FWA può essere parallelizzato per affrontare complicati problemi del mondo reale.

In terzo luogo, FWA ha buone proprietà di convergenza e può sempre trovare soluzioni sufficienti per un problema di minimizzazione globale.

FWA è un algorithmo iterativo composto da quattro componenti, ovvero:

- 1) operatore di esplosione
- 2) operatore di mutazione
- 3) regola di mappatura
- 4) strategia di selezione.

L'operatore di esplosione può essere suddiviso ulteriormente in: intensità/forza di esplosione, ampiezza di esplosione, operazione di spostamento. La mutazione gaussiana è l'operatore di mutazione più utilizzata. In particolare, il flusso di lavoro di FWA può essere indicato come segue.

Innanzitutto, N fireworks sono generati casualmente come sciame iniziale. Quindi, ogni firework conduce operazioni di esplosione e operazioni di mutazione e, se necessario,

viene attivata la regola di mappatura. Infine, gli N fireworks vengono selezionati da tutti gli fireworks e generano scintille in base alla strategia di selezione. L'iterazione continua fino a quando non viene soddisfatto un determinato criterio di terminazione. Man mano che l'iterazione procede, può essere possibile ottenere risultati di ottimizzazione migliori.

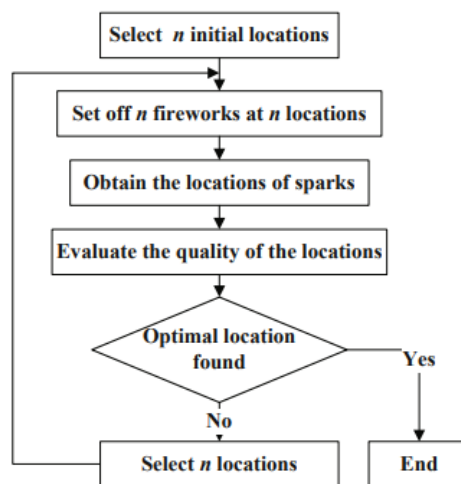


Figura 3.8: Framework of fireworks algorithm. (Fonte: Fireworks algorithm for optimization – Ying Tan and Yuanchun Zhu, 2010)

Dalla figura si può capire che il successo di FWA risiede in una buona progettazione del processo di esplosione e in un metodo adeguato nel selezionare le posizioni.

Lo studio di FWA si concentra principalmente su quattro aspetti: analisi teorica, studio dell'algoritmo, risoluzione dei problemi e applicazioni.

1) Analisi teorica che coinvolge il meccanismo, la qualità della convergenza, la traiettoria e i parametri di FWA. Aiuta a progettare nuovi algoritmi e fornisce informazioni utili per migliorare algoritmi già stabiliti.

2) Studio dell'algoritmo che analizzando e regolando i componenti di FWA. Le persone cercano di migliorare le prestazioni di FWA (convergenza, qualità della soluzione ed efficienza nel tempo) e escogitano varianti FWA migliorate. Nel frattempo, sarà utile integrare FWA con altri metodi, finendo con algoritmi ibridi efficienti.

3.5.1 Caratteristiche e vantaggi di FWA

FWA contiene le seguenti caratteristiche: esplosione, istantaneità, semplicità, località, proprietà emergenti, parallelismo distribuito, diversità ed estensibilità. I dettagli sono riportati di seguito.

1) Esplosività: dopo la prima iterazione di FWA, i fireworks esplodono all'interno dell'ampiezza e producono una pioggia di scintille. Alla fine dell'iterazione, l'algoritmo seleziona N scintille per la generazione successiva. Le N scintille selezionate vengono trattate come nuovi fireworks, preparandosi all'esplosione nella prossima iterazione. In ogni iterazione esploderanno dei fireworks, indicando la caratteristica esplosiva di FWA.

2) Istantaneità: le scintille sono istantanee, il che significa che le scintille non selezionate svaniranno. Ad ogni iterazione, FWA calcola il numero di scintille e l'ampiezza dell'esplosione, a seconda del valore fitness dei fireworks. Quindi, le scintille sono prodotte dagli operatori di esplosione e mutazione. Infine, la scintilla migliore viene inizialmente preservata e poi altre scintille ($N-1$) vengono selezionate in base a una strategia di selezione. Le N scintille selezionate vengono trattate come dei fireworks per la generazione successiva, mentre il resto delle scintille non vengono più considerate.

3) Località: l'ampiezza di esplosione di un firework è inferiore al limite consentito, quindi può esplorare solo localmente nello spazio di ricerca. In FWA, le scintille sono confinate entro un certo intervallo. Le funzionalità localizzate di FWA riflettono potenti capacità di ricerca locale, poiché l'algoritmo può essere utilizzato per la ricerca locale nell'ultima fase del processo di ricerca. Quindi, FWA contiene località.

4) Parallelismo: non esiste alcun meccanismo di controllo centrale tra i fireworks, quindi i fireworks sono indipendenti, molto adatti alla parallelizzazione. In ogni iterazione di FWA, ogni firework esplosione e ricerca all'interno di spazi diversi, ovvero ogni firework conduce una ricerca in dimensioni diverse. Infine, le scintille e i fireworks vengono combinati insieme per scegliere gli N fireworks per la generazione successiva. Ad ogni iterazione, FWA cerca lo spazio parallelo, mostrando le caratteristiche del parallelismo di distribuzione.

5) Diversità: Le diversità della popolazione sono vitali per le prestazioni di qualsiasi swarm intelligence algorithm. Mantenendo la diversità della popolazione, l'algoritmo può saltare fuori dall'ottimo locale, il che fa convergere l'algoritmo nel punto di ottimo

globale, che le tecniche tradizionali difficilmente riescono a raggiungere. Migliore è la diversità della popolazione, più ampia è la distribuzione degli individui. Il valore ottimale potrebbe essere più facile da trovare se una popolazione è fortemente diversificata, poiché la convergenza dell'algoritmo non sarà influenzata in modo significativo. Pertanto, la diversità della popolazione è una parte importante dell'FWA. Le diversità di FWA possono essere le seguenti:

a) diversità del numero di scintille e ampiezza di esplosione. Secondo l'operatore dell'esplosione e i valori di fitness, ogni firework genera un numero diverso di scintille, con una diversa intensità. I fireworks con valori di fitness più alti producono più scintille all'interno di intervalli più piccoli, mentre i fireworks con valori di fitness più bassi producono meno scintille all'interno di intervalli più grandi. In tal modo, la diversità della popolazione sarebbe garantita;

b) diversità dello spostamento e della mutazione gaussiana. Il FWA ha due operatori, operatore di esplosione e operatore di mutazione. Nell'operatore dell'esplosione, lo spostamento viene calcolato secondo l'ampiezza. Nella mutazione gaussiana, i fireworks selezionati devono moltiplicare un numero casuale gaussiano nella posizione di una dimensione. L'operatore dell'esplosione è relativo ai valori fitness dei fireworks mentre l'operatore della mutazione è relativo alla posizione dei fireworks. I due operatori sono diversi l'uno dall'altro, ma entrambi garantiscono che l'esplosione sia diversa;

c) la diversità dei fireworks. Attraverso un certo meccanismo di selezione, le coordinate conservate dei fireworks sono diverse. Di conseguenza, questi fenomeni assicurano la diversità della popolazione. Inoltre, nella strategia di selezione, è più probabile che siano selezionate scintille con una distanza maggiore dalle altre, il che a loro volta garantiscono la diversità della popolazione.

6) Estendibilità: In FWA, il numero di scintille è incerto e può essere determinato in base alla complessità del problema in questione. Il numero dei fireworks e delle scintille può essere di più o di meno, poiché sia l'aumento che la diminuzione degli individui possono risolvere efficacemente il problema. Pertanto, FWA ha estensibilità.

7) Adattabilità. Quando si risolvono i problemi utilizzando FWA, non è necessario che il problema abbia un'espressione esplicita. Il problema può essere risolto calcolando solo i valori fitness. Il FWA può anche risolvere problemi con espressioni esplicite, indicandone le capacità. Pertanto, FWA è adattabile e può essere considerato un algoritmo adattivo.

3.5.2 Principi e implementazione di FWA

FWA inizia a funzionare in modo iterativo fino a quando non vengono soddisfatte una delle seguenti condizioni di arresto: si soddisfano i requisiti di precisione ovvero si raggiunge il numero massimo di valutazioni delle funzioni (iterazioni).

FWA prevede le seguenti fasi:

- 1) Generazione casuale dei fireworks nello spazio possibile.
- 2) Calcolo del valore della funzione di fitness per ogni firework. Il numero di scintille viene calcolato in base alla teoria della concentrazione immunitaria in immunologia e i fireworks con valori fitness migliore producono più scintille.
- 3) Generazione delle scintille entro una certa ampiezza. L'ampiezza dell'esplosione è determinata dal valore fitness di quel firework. L'ampiezza dell'esplosione per il firework con un valore fitness migliore è minore (e viceversa). Ogni scintilla rappresenta una possibile soluzione. Per mantenere la diversità nella popolazione, sono necessarie operazioni di mutazione gaussiana che consiste nel calcolare il miglior valore fitness. Se la condizione di arresto viene soddisfatta, l'algoritmo si interrompe, altrimenti continua il processo di iterazione. La migliore scintilla e le scintille selezionate formano una nuova popolazione.

3.5.2.1 L'operatore di esplosione

Come detto precedentemente nell'inizializzazione, il FWA genera N fireworks in modo casuale. Ogni fireworks genera scintille mediante l'operatore esplosione che è caratterizzato dalla forza dell'esplosione, dall'ampiezza dell'esplosione e dalle operazioni di spostamento. Per ogni fireworks viene calcolata il valore della funzione di fitness. Quanto più il valore della fitness è alto (nel caso di problemi di minimizzazione), tanto maggiore è l'ampiezza dell'esplosione scintilla e il numero di scintille è inferiore. In questo caso, le scintille con valori di fitness migliori cercheranno più attentamente nelle aree più piccole, mentre le scintille con valori di fitness peggiori cercheranno in

aree più ampie; pertanto, FWA ha maggiori possibilità di trovare l'ottimo globale. L'operatore dell'esplosione ha due parametri: il primo parametro m viene utilizzato per limitare il numero totale di scintille, e il secondo parametro N è il numero di fuochi d'artificio.

3.5.2.1.1 La forza di esplosione

La forza dell'esplosione è un'operazione centrale dell'algoritmo in quanto simula il modo di esplodere dei fireworks nella vita reale. Quando un firework esplode esso svanisce in un brevissimo tempo e poi molte piccole esplosioni appaiono attorno ad esso. L'algoritmo di Fireworks determina innanzitutto il numero di scintille, quindi calcola l'ampiezza di ogni esplosione. I fireworks con valori di fitness migliori producono più scintille; viceversa per i fireworks con valori di fitness peggiori, le scintille generate sono meno numerose e sparse nella regione d'indagine. I fireworks con valori di fitness peggiori vengono utilizzati per impedire che l'algoritmo converga prematuramente. Si può vedere dalla figura che i fireworks con valori di fitness migliori producono più scintille all'interno di un'ampiezza minore (buona esplosione) rispetto a quelli con valori di fitness peggiori (esplosione non buona). Dopo aver determinato il numero delle scintille, è necessario calcolare l'ampiezza delle scintille nell'esplosione di un fireworks.

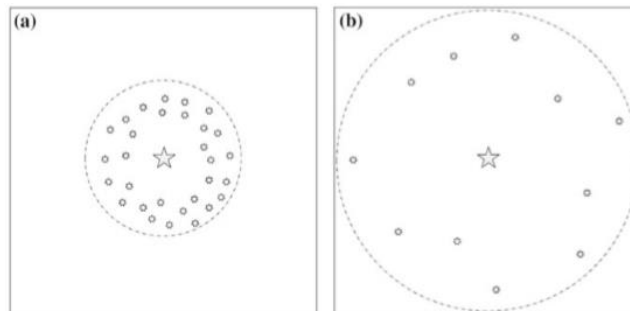


Figura 3.9: The good and bad explosions of a firework. a) Good explosion. b) Bad explosion (Fonte: Fireworks Algorithm-Ying Tan-2015)

Il numero di scintille è determinato come segue:

$$S_i = m * \frac{Y_{max} - f(x_i) + \varepsilon}{\sum_{i=1}^N (Y_{max} - f(x_i)) + \varepsilon}, \quad (1)$$

Dove S_i è il numero di scintille per ciascun individuo o firework, m è una posizione costante per il numero totale di scintille e Y_{max} indica il valore fitness dell'individuo peggiore tra gli N individui della popolazione. La funzione $f(x_i)$ rappresenta la

precisione di un singolo x_i , mentre l'ultimo parametro ε viene utilizzato per evitare che il denominatore diventi zero.

Di seguito la limitazione del numero di scintille:

$$\hat{s}_i = \begin{cases} \text{round}(a \cdot m), & \text{if } s_i < am \\ \text{round}(b \cdot m), & \text{if } s_i > bm, a < b < 1, \\ \text{round}(a \cdot m), & \text{otherwise} \end{cases} \quad (2)$$

dove a e b sono costanti, \hat{s}_i invece è la limitazione del numero di scintille, e $\text{round}()$ è la funzione di arrotondamento.

3.5.2.1.2 L'ampiezza di esplosione

Osservando le curve di alcune tipiche funzioni da ottimizzare, i punti attorno all'ottimo locale e all'ottimo globale hanno sempre valori di fitness migliori. Pertanto, controllando l'ampiezza dell'esplosione, l'ampiezza dei fireworks con valori di fitness migliori si riduce gradualmente, portando l'algoritmo fireworks a trovare l'ottimo locale e globale. Al contrario, i fireworks con valori di fitness peggiori esplorano l'ottimo in una grande ampiezza.

Di seguito l'ampiezza di esplosione:

$$A_i = \hat{A} * \frac{f(x_i) - Y_{min} + \varepsilon}{\sum_{i=1}^N (f(x_i) - Y_{min}) + \varepsilon}, \quad (3)$$

dove A_i denota l'ampiezza di ogni individuo, \hat{A} è una costante come somma di tutte le ampiezze, mentre Y_{min} indica il valore fitness migliore tra tutti gli N individui. Il significato della funzione di $f(x_i)$ e del parametro ε è come menzionato nel eq. (1).

3.5.2.1.3 Operazione di spostamento

Dopo il calcolo dell'ampiezza di esplosione, è necessario determinare lo spostamento all'interno dell'ampiezza dell'esplosione. Il FWA utilizza lo spostamento casuale. In questo modo, ogni firework ha il suo numero specifico di esplosioni e ampiezza delle scintille. Per garantire la diversità della popolazione, il FWA genera diversi spostamenti casuali all'interno di ciascuna ampiezza. Attraverso l'operatore dell'esplosione, ogni firework genera una pioggia di scintille, contribuendo a trovare l'ottimo globale di una data funzione. L'operazione di spostamento consiste nello spostamento su ogni dimensione di un firework e può essere definita come:

$$x_i^k = x_i^k + U(-A_i, A_i), \quad (4)$$

Dove $U(-A_i, A_i)$ indica il numero casuale uniforme all'interno degli intervalli dell'ampiezza A_i . Il seguente algoritmo è lo pseudocodice dell'operatore di esplosione descritto nelle equazioni (1) - (4).

Algorithm 3.1 Generate Sparks (Fireworks Algorithm-Ying Tan-2015)

1: Initialization, calculate the fitness value $f(x_i)$ for each firework.

2: Calculate the number of sparks S_i .

3: Calculate the amplitude of sparks A_i .

4: $z = \text{rand}(1, \text{dimension})$ //randomly choose z dimensions

5: for $k=1 \rightarrow \text{dimension}$ do

6: if $k \in z$ then

7: $x_i^k = x_i^k + U(-A_i, A_i)$

8: end if

9: end for

3.5.2.2 Operatore di mutazione gaussiana

Per migliorare ulteriormente la diversità di una popolazione, in FWA viene introdotta la mutazione gaussiana. Il modo per produrre scintille dalla mutazione gaussiana è il seguente: si sceglie un firework dalla popolazione attuale, e quindi poi applicare la mutazione gaussiana al firework in dimensioni selezionate casualmente. Per la mutazione gaussiana, le nuove scintille vengono generate tra il firework migliore e il firework selezionato (figura seguente). Tuttavia, la mutazione gaussiana può produrre scintille che superano lo spazio consentito. Quando una scintilla si trova oltre il limite superiore o inferiore, la regola di mappatura verrà eseguita per mappare la scintilla in una nuova posizione all'interno dello spazio consentito.

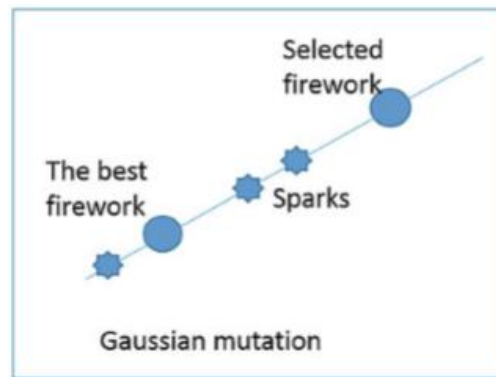


Figura 3.10: Gaussian mutation. (Fonte: Fireworks Algorithm-Ying Tan-2015)

Supponiamo che la posizione individuale corrente sia indicata come x_i^k , dove i varia da 1 a N e k indica la dimensione corrente. Le scintille dell'esplosione gaussiana sono calcolate da:

$$x_i^k = x_i^k * g, \quad (5)$$

Dove g è un numero random con media 1 e varianza 1:

$$g = N(1,1) \quad (6)$$

Algorithm 3.2 Gaussian Mutation (Fireworks Algorithm-Ying Tan-2015)

- 1: Calculate the fitness value $f(x_i)$ for each firework.
- 2: Calculate the coefficient $g = N(1,1)$.
- 3: $z = \text{rand}(1, \text{dimension})$ // randomly selected z dimensions
- 4: for $K=1 \rightarrow \text{dimension}$ do
- 5: if $k \in z$ then
- 6: $x_i^k = x_i^k * g$
- 7: end if

L'operatore di mutazione gaussiana può aumentare la diversità dell'algoritmo in quanto le scintille generate dalla mutazione gaussiana non si limitano a quelle attorno ai fireworks. Inoltre, poiché le scintille sono generate tra la posizione corrente e l'origine

dalla mutazione gaussiana ($x_i^k \leq x_i^k * g$), per le funzioni che hanno l'ottimale all'origine, le prestazioni sono piuttosto buone.

3.5.2.3 Regole di mappatura

Se un firework si trova vicino al limite dello spazio consentito, mentre la sua ampiezza di esplosione copre sia lo spazio consentito sia lo spazio non consentito, allora le scintille generate possono trovarsi fuori dallo spazio consentito. Come tale, la scintilla oltre lo spazio consentito è inutile. Pertanto, deve tornare nello spazio consentito. La regola di mappatura viene utilizzata per gestire questa situazione. La regola di mappatura garantisce che tutte le scintille si trovino nello spazio consentito. Se c'è una scintilla che viene generata da un firework oltre lo spazio consentito, verrà mappata nello spazio consentito. La regola di mappatura garantisce che tutti gli individui rimangano nello spazio consentito. Se ci sono alcune scintille fuori dal confine, verranno ricondotte nel loro ambito consentito. La regola di mappatura utilizza un'operazione modulare, come indicato di seguito:

$$x_i^k = X_{LB,k} + x_i^k \% (X_{UB,k} - X_{LB,k}), \quad (7)$$

Dove x_i^k presenta la posizione di ogni scintilla che è collocata fuori dai confini, mentre $X_{UB,k}$ e $X_{LB,k}$ indicano i limiti massimi e minimi della posizione di una scintilla. Il segno % rappresenta l'aritmetica modulare.

La regola di mappatura viene utilizzata per garantire che tutte le scintille vengano generate nell'ambito dello spazio consentito. Quando un firework si trova vicino al confine e l'ampiezza dell'esplosione è grande, le scintille generate potrebbero essere fuori dai confini dello spazio consentito. Pertanto, le scintille fuori dai confini verranno mappate nello spazio consentito per evitare calcoli non necessari. Tuttavia, la regola di mappatura ha i suoi svantaggi. Ad esempio, può facilmente portare una scintilla nelle posizioni vicino all'origine. La regola di mappatura adotta un'aritmetica modulare per garantire che le scintille fuori dai confini vengano tirate indietro nello spazio consentito.

3.5.2.4 La strategia di selezione

Dopo aver applicato l'operatore di esplosione, l'operatore di mutazione e la regola di mappatura, alcune delle scintille generate devono essere selezionate e trasmesse alla

generazione successiva. Nell'algoritmo FWA viene utilizzata la strategia basata sulla distanza. Innanzitutto, per selezionare le scintille per la prossima generazione, per prima, la scintilla migliore viene sempre mantenuta per la prossima generazione. Quindi, gli altri (N-1) vengono selezionati in base alla distanza mantenendo la diversità della popolazione.

Nella strategia di selezione, viene utilizzata la misurazione della distanza euclidea, dove $d(x_i, x_j)$ indica la distanza euclidea tra due individui x_i e x_j .

$$R(x_i) = \sum_{j=1}^K d(x_i, x_j) = \sum_{j=1}^K \|x_i - x_j\|, \quad (8)$$

Dove $R(x_i)$ rappresenta la somma delle distanze tra gli individui x_i e tutti gli altri individui. $j \in K$ significa che la posizione j appartiene all'insieme K , dove K è l'insieme di combinazione sia delle scintille generate dall'operatore di esplosione che dell'operatore di mutazione.

Il modo "roulette" (vedi fig. 3.8) viene utilizzato per scegliere gli individui per la prossima generazione, poiché la possibilità di scegliere il singolo x_i dovrebbe essere $p(x_i)$, che è data da:

$$p(x_i) = \frac{R(x_i)}{\sum_{j \in K} R(x_j)}, \quad (9)$$

Dalla equazione 9, si può vedere che gli individui con distanze maggiori avranno più possibilità di essere selezionati per la generazione successiva.

Algorithm 3.3 Pseudo code of FWA (Fireworks Algorithm-Ying Tan-2015)

```

1: Randomly select  $N$  locations for fireworks
2: while terminal condition is not met do
3: Set off  $N$  fireworks, respectively, at the  $N$  locations:
4: for all fireworks  $x_i$  do
5: Calculate the number of sparks as  $S_i$ 
6: Calculate the amplitude of sparks as  $A_i$ 
7: end for //  $\hat{m}$  is the number of sparks generated by Gaussian mutation
8: for  $k=1 \rightarrow \hat{m}$  do
9: Randomly select a firework  $x_i$  and generate a spark
10: end for
11: select the best spark and the other sparks according to selection strategy
12: end while

```

Quindi riassumendo: la strategia di selezione è quella di scegliere gli individui per la prossima generazione. L'individuo migliore viene sempre mantenuto per la generazione successiva, mentre gli altri ($N - 1$) vengono selezionati in base alla distanza euclidea e le scintille più lontane da altre scintille possono essere selezionate con maggiori possibilità. Pertanto, la diversità di FWA è garantita in questo modo.

3.6 ENHANCED FIREWORKS ALGORITHM

FWA funziona molto bene su funzioni che hanno il loro ottimo all'origine; quando invece viene applicato su funzioni spostate, la qualità dei risultati di FWA convenzionale peggiora gravemente e peggiora con l'aumento dei valori di spostamento, ovvero con l'aumentare della distanza tra la funzione ottimale e l'origine dello spazio di ricerca. Inoltre, rispetto ad altri algoritmi di ottimizzazione metaeuristica, FWA ha un elevato costo computazionale per iterazione. Per queste limitazioni di seguito si introducono cinque importanti miglioramenti su FWA:

- 1) un nuovo controllo dell'ampiezza minima dell'esplosione
- 2) un nuovo operatore per la generazione di scintille di esplosione
- 3) una nuova strategia di mappatura per scintille che non rientrano nello spazio di ricerca

- 4) un nuovo operatore per la generazione di scintille gaussiane e
 - 5) un nuovo operatore per la selezione della popolazione per la successiva iterazione.
- L'algoritmo risultante si chiama Enhanced Fireworks Algorithm (EFWA).

3.6.1 Un nuovo controllo dell'ampiezza minima di esplosione (MEAC)

L'equazione (3) del paragrafo 3.5.2 mostra come viene calcolata l'ampiezza dell'esplosione per ogni firework nel FWA. Un firework con miglior fitness avrà un'ampiezza di esplosione maggiore. Sebbene questa idea sembri ragionevole, l'ampiezza dell'esplosione dei fireworks con fitness migliore (o molto buona) generalmente sarà molto piccola cioè vicino a zero - secondo l'equazione (3) del paragrafo 3.5.2. Se l'ampiezza dell'esplosione è (vicino a) zero, le scintille dell'esplosione si troveranno nella (quasi) stessa posizione dei fireworks stessi. Di conseguenza, può succedere che la posizione del miglior firework non può essere migliorata fino a quando un altro firework non abbia trovato un posto migliore. In modo da evitare questo problema, di seguito si introduce un limite inferiore A_{min} dell'ampiezza dell'esplosione, che si basa sullo stato di avanzamento dell'algoritmo. Durante la fase iniziale della ricerca, A_{min} è impostato su un valore più alto per facilitare l'esplorazione, con l'aumentare del numero di valutazioni, A_{min} viene ridotto al fine di consentire migliori capacità di esplorazione in buone posizioni. Per ogni dimensione k , l'ampiezza dell'esplosione A_i^k è vincolata come di seguito:

$$A_{ik} = \begin{cases} A_{min,k} & \text{if } A_{i,k} < A_{min,k} \\ A_{i,k} & \text{otherwise,} \end{cases} \quad (1)$$

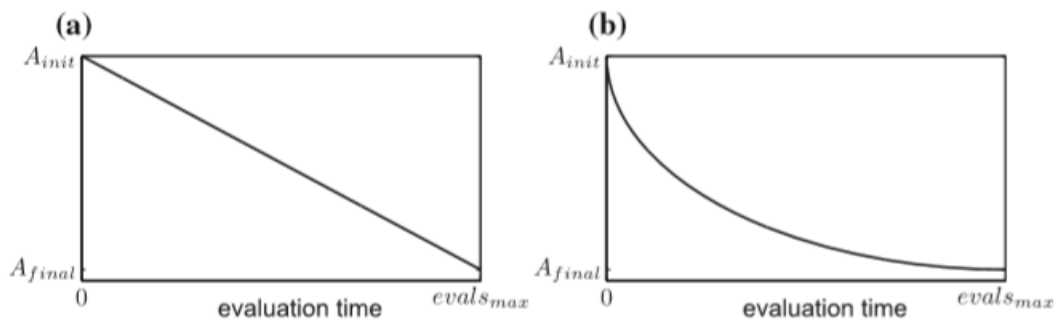


Figura 3.11: Linearly and nonlinearly decreasing minimal explosion amplitude. a) Linear decrease; b) Nonlinear decrease (Fonte: Fireworks Algorithm-Ying Tan-2015)

Per ogni iterazione viene calcolato un nuovo valore A_{min} : Ying Tan nel suo studio “Fireworks Algorithm - A Novel Swarm Intelligence Optimization Method” propone di utilizzare due diversi modi per calcolare A_{min} . Il primo approccio si basa in una funzione lineare decrescente (eq.2 seguente), e l’altro approccio è basato in una funzione non-lineare decrescente (eq.3 seguente):

$$A_{min,k}(t) = A_{init} - \frac{A_{init} - A_{final}}{evals_{max}} * t, \quad (2)$$

$$A_{min,k}(t) = A_{init} - \frac{A_{init} - A_{final}}{evals_{max}} * \sqrt{(2 * evals_{max} - t)t}. \quad (3)$$

In entrambe le equazioni, t si riferisce al numero di valutazione della funzione all’inizio dell’iterazione corrente e $evals_{max}$ è il numero massimo di valutazioni. A_{init} e A_{final} sono rispettivamente l’ampiezza minima iniziale e finale dell’esplosione. Rispetto alla diminuzione lineare di A_{min} , la diminuzione non-lineare migliora l’efficienza il prima possibile, ovvero dopo poche iterazioni. La figura 3.12 mostra una rappresentazione grafica delle equazioni (3) e (4).

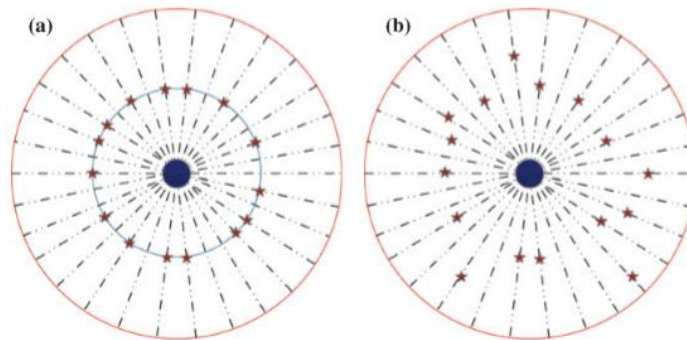


Figura 3.12: Generation of explosion sparks in FWA and EFWA. a) FWA b) EFWA (Fonte: Fireworks Algorithm-Ying Tan-2015)

Algorithm 3.4 Generating “explosion sparks” in EFWA.

- 1: Initialize the location of the “explosion sparks”: $x_i = X_i$,
- 2: Set $z^k = \text{round}(U(0,1))$, $k=1,2,\dots, D$.
- 3: **for** each dimension of x_{ik} , where $z^k = 1$ **do**
- 4: Calculate offset displacement: $\Delta X_{ik} = A_i \times U(-1,1)$,
- 5: $x_{ik} = x_{ik} + \Delta X_{ik}$,
- 6: **if** x_{ik} out of bounds **then**

7: map x_{ik} to the potential space (guarda la sezione seguente)

8: **end if**

9: **end for**

3.6.2 Un nuovo operatore per la generazione delle scintille esplosive

In FWA, lo spostamento per le scintille di esplosione di ciascun firework viene calcolato una sola volta e lo stesso valore viene aggiunto alla posizione delle dimensioni selezionate. Ovviamente, l'aggiunta dello stesso valore in ogni dimensione comporta una cattiva capacità di ricerca locale. Per evitare questo problema, si suggerisce di calcolare uno spostamento offset diverso per le dimensioni selezionate (quelle dimensioni, dove z^k è uguale a 1). L'algoritmo 3.4 mostra il processo proposto per generare scintille di esplosione nell'EFWA. La figura 3.12 mostra la differenza tra la generazione di scintille di esplosione in FWA e EFWA. Quindi, lo spostamento offset in FWA è simile per tutte le dimensioni selezionate, invece in EFWA viene calcolato uno spostamento offset diverso in ogni dimensione.

3.6.3 Un nuovo operatore di mappatura

Nel FWA standard, quando la posizione di una nuova scintilla supera l'intervallo della dimensione k , la nuova scintilla verrà mappata in un'altra posizione secondo la seguente equazione:

$$x_{ik} = X_{LB,k} + |x_{ik}| \% (X_{UB,k} - X_{LB,k}), \quad (4)$$

Dove $X_{UB,k}$ $X_{LB,k}$ indicano il limite superiore e il limite inferiore di ricerca nell'intervallo consentito. In molti casi, una scintilla supera lo spazio di ricerca consentito solo di un valore piuttosto piccolo. Inoltre, siccome lo spazio di ricerca è spesso equamente distribuito ($X_{LB,k} = -X_{UB,k}$), la posizione regolata x_{ik} sarà molto vicina all'origine in molti casi. Esempio: consideriamo un problema di ottimizzazione nello spazio di ricerca $[-20,20]$; se, nella dimensione k , viene creata una nuova scintilla nel punto $x_{ik}=21$, verrà mappata nella posizione $[x_{ik}=-20 + |21| \% (40)]$. Siccome il risultato dell'operazione modulo $21 \% (40) = 21$; x_{ik} sarà mappato nella posizione $x_{ik}=1$, che è già molto vicino all'origine. Nei casi in cui $X_{LB,k} = -X_{UB,k}$, questo operatore di mappatura è in parte responsabile degli

svantaggi menzionati nei paragrafi precedenti. Al fine di evitare i problemi causati dall'operatore di mappatura convenzionale, nel EFWA questo metodo si sostituisce con un operatore di mappatura casuale uniforme che mappa le scintille in qualsiasi posizione nello spazio di ricerca con distribuzione uniforme.

3.6.4 Un nuovo operatore per la generazione di scintille gaussiane

Insieme all'operatore di mappatura, l'operatore di mutazione gaussiana è la ragione principale per cui la FWA convenzionale funziona in modo significativamente migliore rispetto ad altri algoritmi di ottimizzazione per funzioni che hanno il loro ottimale all'origine (risultati dello studio Y. Tan, Y. Zhu, Fireworks algorithm for optimization, in *Advances in Swarm Intelligence* - Springer, Berlin, 2010 - pp. 355–364). La figura 3.13 (a), (b) mostra la posizione delle scintille gaussiane per una funzione bidimensionale di Ackley con l'optimum a $[0,0]$ e $[-70, -55]$, rispettivamente. In ogni iterazione, la posizione delle (solo) scintille gaussiane viene tracciata e non eliminata. La posizione delle scintille gaussiane in Fig 3.13 (a) indica che la maggior parte delle scintille si trova all'origine, cioè vicino all'ottimale. Inoltre, si può vedere che le aree vicine agli assi delle coordinate sono anche più affollate rispetto ad altre parti dello spazio di ricerca.

La figura 3.13 (b) rileva un fatto interessante sulla posizione delle scintille gaussiane per la funzione di Ackley; anche se l'ottimale è ormai lontano dall'origine, molte scintille si trovano vicino al centro. Ovviamente, molte scintille nella figura 3.13 (a) non è che non si trovano vicino al centro a causa dell'intelligenza dell'algoritmo, ma piuttosto perché molte scintille gaussiane vengono create vicino all'origine dello spazio di ricerca, indipendentemente dalla posizione della funzione ottimale. Si sottolinea che la figura 3.13 (a), (b) è stata creata usando l'operatore di mappatura di FWA convenzionale.

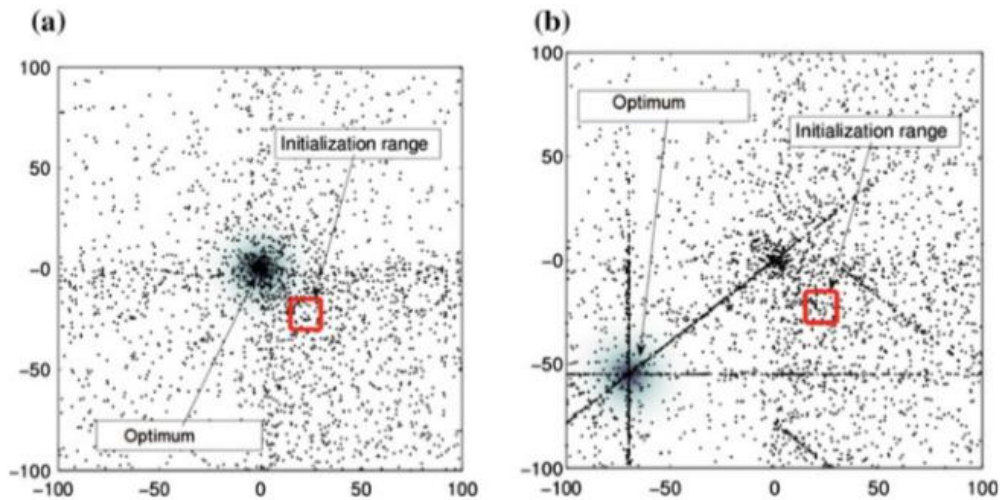


Figura 3.13: The locations of the Gaussian sparks using the conventional FWA (Ackely function using 100000 function evaluations). (a) No shift – optimum at origin. (b) Shift – optimum at (-70,-55). (Fonte: Fireworks Algorithm-Ying Tan-2015).

La ragione di questo comportamento viene calcolato dalle scintille gaussiane, come mostrato nell'equazione seguente:

$$x_{ik} = X_{ik} * g \quad (5)$$

Qui, g è impostato come un valore casuale con distribuzione normale con valore atteso e varianza pari a uno. Nei casi in cui g è vicino a 0, anche x_{ik} sarà vicino a zero. Di conseguenza, molte scintille gaussiane saranno posizionate vicino all'origine dello spazio di ricerca di dimensione k . Inoltre, per ampie g , molte scintille gaussiane vengono create in punti che sono fuori limite, fuori dalle bande. In questo caso, l'operatore di mappatura FWA convenzionale mapperà la scintilla appena creata in una posizione che in molti casi è vicino all'origine.

Un altro problema dell'operatore convenzionale delle scintille gaussiane è il fatto che i fireworks che si trovano già vicino all'origine dello spazio di ricerca non possono sfuggire a questa posizione; se per esempio il firework X_{ik} è vicino allo zero, anche la posizione della scintilla x_{ik} deve essere vicina allo zero, poichè $x_{ik} = X_{ik} * g$.

Inizializzazione

Poiché la FWA convenzionale è in grado di convergere verso l'ottimale già dopo pochissime valutazioni delle funzioni (facendo riferimento ai risultati dello studio Y. Tan, Y. Zhu, *Fireworks_algorithm for optimization, in Advances in Swarm Intelligence* -

Springer, Berlin, 2010 - pp. 355–364), è stato anche analizzato il comportamento di FWA durante la prima iterazione. La Figura 3.14 (a)-(d) mostra la distribuzione dell'esplosione e delle scintille gaussiane, rispettivamente, direttamente dopo l'inizializzazione con diversi intervalli di inizializzazione. I fireworks sono stati inizializzati ripetutamente, sono stati creati due tipi di scintille, e tracciate le loro posizioni fino a quando non sono stati raggiunti 5000 valutazioni delle funzioni (circa 100 ripetizioni). La distribuzione delle scintille è (prevedibilmente) indipendente dalla funzione ottimale e mostra un comportamento simile per diversi intervalli di inizializzazione che sono stati impostati su dim 1: [15,30]; dim 2: [15,30] per la figura 3.14 (a), (b) e dim (1): [60,75]; dim 2: [30,45] per la figura 3.14 (c), (d). Ovviamente, alcune scintille gaussiane si trovano molto vicino all'origine della funzione, indipendentemente dall'intervallo di inizializzazione. Questa è un'altra indicazione del motivo per cui il FWA convenzionale è in grado di trovare l'ottimale di funzioni centrate in poche iterazioni.

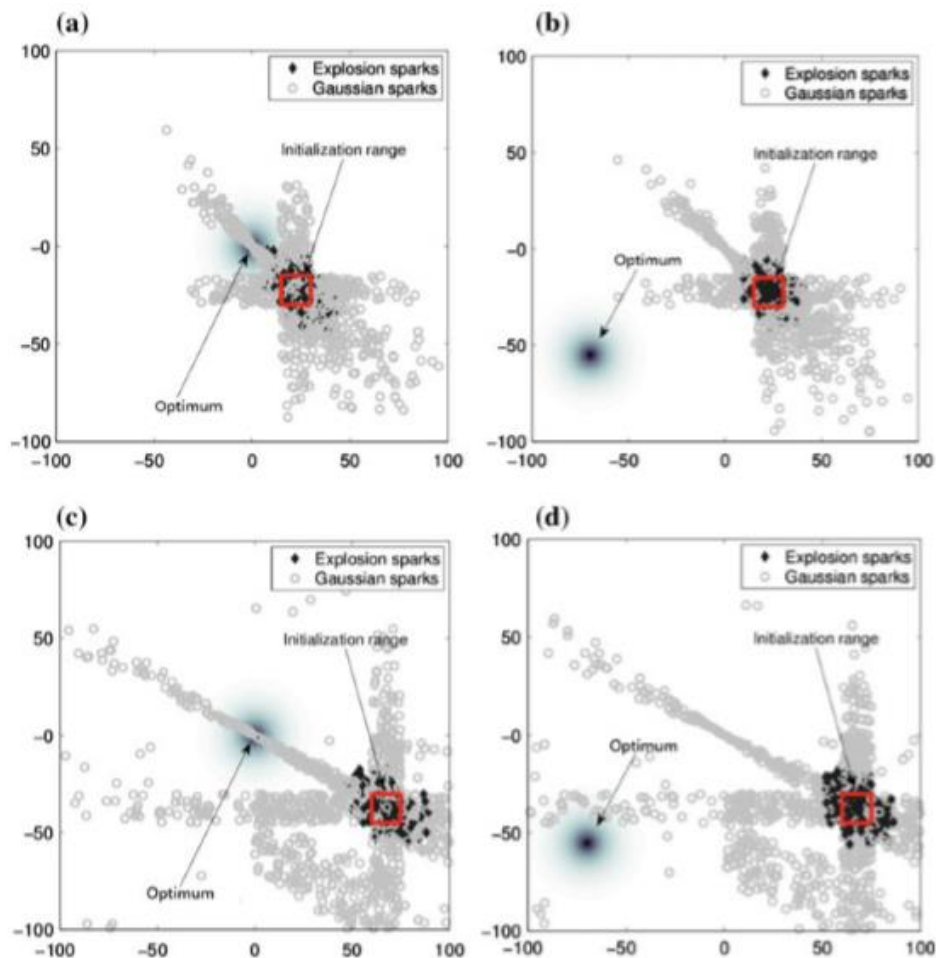


Figura 3.14: Sparks after initialization in conventional FWA. (a) No shift—optimum at origin. (b) Shift— optimum at $(-70,-55)$. (c) No shift—optimum at origin. (d) Shift— optimum at $(-70,-55)$. (Fonte: Fireworks Algorithm-Ying Tan-2015).

Il nuovo operatore di scintilla gaussiana

Al fine di evitare i problemi dell'operatore convenzionale di mutazione gaussiana, si propone un nuovo operatore di mutazione gaussiana che viene calcolato da:

$$x_{ik} = X_{ik} + (X_{Bk} - X_{ik}) * e, \quad (6)$$

dove, X_B è la posizione attuale della migliore scintilla del fireworks di esplosione trovata finora, e $e = N(0,1)$: I dettagli sono riportati nell'Algoritmo 3.5. Come mostrato nella figura 3.15, il nuovo operatore di mutazione di estenderà lungo la direzione tra la direzione corrente del firework e la posizione del firework migliore. Ciò garantisce la diversità della ricerca ma comporta anche alcuni movimenti globali verso la migliore posizione trovata finora. Questo nuovo operatore comporta un movimento verso l'origine dello spazio di ricerca se il firework attuale migliore si trova all'origine.

Algorithm 3.5 Generating "Gaussian sparks" in EFWA

- 1: Initialize the location of the "Gaussian sparks": $x_i = X_i$
- 2: Set $z^k = \text{round}(U(0,1))$, $k = 1, 2, \dots, D$
- 3: Calculate offset displacement: $e = N(0,1)$
- 4: for each dimension x_{ik} , where $z^k = 1$ do
- 5: $x_{ik} = X_{ik} + (X_{Bk} - X_{ik}) * e$, where X_B is the position of the best firework found so far.
- 6: if x_{ik} out of bounds then
- 7: mapping the position according to Eq. (4)
- 8: end if
- 9: end for

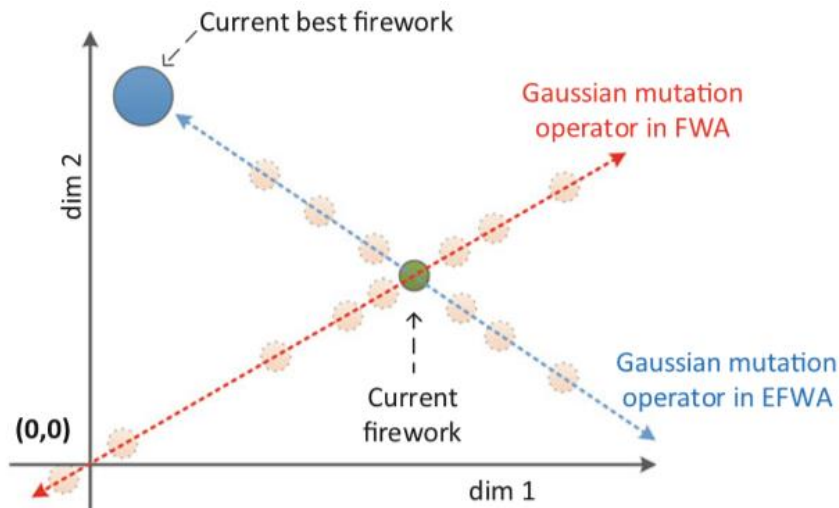


Figura 3.15: Difference between the Gaussian sparks operator in FWA and EFWA. (Fonte: *Fireworks Algorithm-Ying Tan-2015*)

3.6.5 Un nuovo operatore di selezione

FWA prevede una strategia di selezione basata sulla distanza che favorisce la selezione di fireworks/scintille in regioni meno affollate dello spazio di ricerca. Sebbene la selezione di posizioni in regioni poco affollate con una maggiore probabilità aumenti la diversità, questo operatore di selezione ha lo svantaggio di essere computazionalmente molto costoso. Una profilazione di runtime del codice FWA originale ha rilevato che l'operatore di selezione di FWA convenzionale è responsabile della maggior parte del runtime. Al fine di accelerare il processo di selezione della popolazione per la generazione successiva, si applica un altro metodo di selezione, che è indicato come metodo di selezione casuale ellittica (ERP) (S. Zheng, A. Janecek, Y. Tan, *Enhanced fireworks algorithm, in 2013 IEEE Congress on Evolutionary Computation (CEC) IEEE, 2013, pp. 2069–2077*). In questo processo di selezione, l'ottima dell'insieme verrà selezionata per prima. Quindi poi, gli altri individui vengono selezionati casualmente. Ovviamente, la complessità computazionale di ERP è solo lineare rispetto al numero dei fireworks e quindi, riduce la durata dell'EFWA significativamente.

“La valutazione sperimentale su 12 funzioni benchmark con valori diversi di spostamento mostra che l'EFWA supera le FWA convenzionali in termini di capacità di convergenza,

riducendo significativamente il tempo di esecuzione” (Fireworks Algorithm-Ying Tan-2015).

In questo capitolo è stata presentata una versione migliorata di FWA. Sulla base dell'analisi dettagliata degli operatori di FWA convenzionali, sono stati presentati cinque nuovi operatori per superare i limiti di FWA, tra cui la strategia di controllo dell'ampiezza minima dell'esplosione, il nuovo operatore di mappatura, il nuovo operatore per la generazione di scintille di esplosione, il nuovo operatore di mutazione gaussiana e l'operatore di selezione. Risultati sperimentali (in Fireworks Algorithm-Ying Tan-2015) hanno dimostrato che l'EFWA è una versione di miglioramento significativo di FWA. Nel paragrafo seguente si riporta un esperimento e delle osservazioni riguardo a questo argomento.

4 CASE STUDY

4.1 FIREWORKS ALGORITHMS AND PSO CASE STUDY

In uno studio (Claudio Pizzi, Irene Bitto, Marco Corazza (forthcoming) *Exploration and Exploitation in Optimizing a Basic Financial Trading System: A Comparison Between FA and PSO Algorithms*, *Progresses in Artificial Intelligence and Neural Systems*, Singapore, Springer Nature, vol. 184, pp. 293-303), sono state confrontate le due metaeuristiche di ottimizzazione presentate nel precedente capitolo: Fireworks Algorithm (FA), e la Particle Swarm Optimization (PSO). Entrambe le metaeuristiche sono basate sulla "Swarm Intelligence" e il lavoro succitato ha come obiettivo il confronto della loro capacità di esplorazione dello spazio delle soluzioni, e poi, quando viene identificata una soluzione promettente, le loro abilità di "perlustrare a fondo" quella determinata area identificata. FA e PSO sono caratterizzate da una popolazione di operatori e ogni operatore costituisce una possibile soluzione del problema di ottimizzazione globale per il quale si effettua l'indagine. Un operatore rappresenta un punto dello spazio di soluzioni. Questa popolazione evolve nel tempo secondo le regole specifiche della metaeuristica utilizzata e l'evoluzione finisce quando un criterio di arresto viene soddisfatto. La qualità di ogni operatore/soluzione è valutata dalla funzione fitness ovvero del valore assunto dalla funzione da ottimizzare. Più precisamente minore (maggiore) è il punteggio fitness in caso di minimizzazione (massimizzazione), migliore è la soluzione dell'operatore.

Una differenza importante tra FA e PSO consiste nel numero di punti nel quale si valuta (ad ogni iterazione) la funzione da ottimizzare: tale numerosità varia per FA mentre è costante per PSO. Queste metaeuristiche sono state applicate con lo scopo di ottimizzare i parametri di un semplice Trading System (TS) basato su un unico indicatore di analisi tecnica: le bande di Bollinger (BBs). Questo indicatore tiene in considerazione la media mobile dei prezzi di un asset finanziario, la cosiddetta linea centrale. L'ampiezza delle bande fornisce una misura della volatilità dei prezzi di un asset. Nella versione standard, le bande di Bollinger dipendono da due parametri: n , il numero di termini da utilizzare per calcolare la media, e m , un numero reale che indica l'ampiezza di ogni banda in termini di moving standard deviation.

$$\text{Formalmente: } \begin{cases} CL(t) = MA(t, n) \\ BBup(t) = CL(t) + m * MSD(t, n), \\ BBdn(t) = CL(t) - m * MSD(t, n). \end{cases}$$

Dove $CL(t)$ indica la linea centrale al tempo t , $MA(t, n)$ è la media mobile calcolata al tempo t utilizzando le ultime n osservazioni, $BBup(t)$ e $BBdn(t)$ indicano rispettivamente la banda superiore e inferiore al tempo t , $MSD(t, n)$ è la moving standard deviation calcolata al tempo t utilizzando le ultime n osservazioni. Come è evidente, questo indicatore implicitamente assume che variazioni dei prezzi di un asset finanziario si distribuiscano simmetricamente. Lo studio delle serie storiche dei prezzi molto spesso mostra una autocorrelazione negativa nei rendimenti finanziari. Cosicché, è generalmente accettato che la distribuzione delle variazioni dei prezzi di un asset finanziario sia asimmetrica. Per queste ragioni, in questo studio le Bande di Bollinger sono state generalizzate permettendo che la distanza delle $BBup$ e $BBdn$ dalla linea centrale possa essere diversa.

In particolare, questa generalizzazione è implementata modificando le formule per il calcolo delle BBs nel seguente modo:

$$BBup(t) = CL(t) + m.up * MSD(t, n)$$

$$BBdn(t) = CL(t) - m.dn * MSD(t, n)$$

Dove $m.up$ e $m.dn$ indicano l'ampiezza rispettivamente della banda superiore e di quella inferiore.

La stima corretta di n , $m.up$ a $m.dn$ è cruciale per ogni TS affinché si generino segnali di acquisto e vendita convenienti. In particolare, i segnali indicati con -1 in caso di vendita, $+1$ in caso di acquisto, sono generati nel seguente modo:

$$\begin{cases} -1 & \text{if } P(t) > BBup(t) \wedge P(t-1) < BBup(t-1) \\ +1 & \text{if } P(t) < BBdn(t) \wedge P(t-1) > BBdn(t-1), \\ & signal(t-1) \text{ otherwise} \end{cases} \quad (1)$$

L'obiettivo di un TS consiste nel massimizzare il guadagno cumulativo alla fine del periodo di trading $G(T)$, dove $G(T) = G(t-1)[1 + signal(t-1)*r(t)]$, con $t=0, \dots, T$ e $G(0)=G$, dove $r(t)$ è il rendimento del asset finanziario al tempo t , e G denota il capitale investito inizialmente. Quindi, il problema dell'ottimizzazione globale al quale si fa fronte è:

$$\begin{aligned} & \max_{n,m.up,m.dm} G(T) \\ & \text{s.t.} \begin{cases} n \in \{1,2,\dots,99,100\} \\ m.up \in [1,10] \\ m.dm \in [1,10] \end{cases} \end{aligned} \quad (2)$$

Si sottolinea che il problema di ottimizzazione globale è fortemente non lineare a causa della funzione obiettivo ed è NP-Hard a causa del vincolo sul parametro n . Quindi si tratta in un problema di ottimizzazione complessa da risolvere.

4.1.1 Applicazione

FA e PSO sono stati originariamente disegnati per risolvere problemi di ottimizzazione globale senza vincoli. Per risolvere un problema di ottimizzazione vincolato come quello presentato nel paragrafo precedente è possibile utilizzare lo schema di penalità proposto da “Corazza, Fasano e Gusso (Corazza et al. 2013)”.

Per quanto riguarda la serie storica finanziaria dei prezzi di chiusura, si sono stati considerati 5 importanti titoli inseriti nel calcolo dello S&P 500 index. Essi coprono diversi settori economici e sono: Bank of America Corporation (BAC), Boeing Company (BC), Booking Holdings Inc. (BKNG), Microsoft Corporation (MC), and NIKE Inc. (NKE). L’analisi è stata condotta su tre differenti periodi: il primo dal 2 gennaio 2002 fino al 29 dicembre 2006 (1259 osservazioni); il secondo dal 3 gennaio 2006 fino a 30 dicembre 2011 (1511 osservazioni); il terzo dal 3 gennaio 2011 fino il 30 dicembre 2015 (1257 osservazioni). La prima meta di ogni periodo viene utilizzato per ottenere le stime dei parametri del TS, mentre la seconda meta viene utilizzata per valutare le performance out-of-sample del TS.

L’utilizzo delle serie storiche dei prezzi (di chiusura) di 5 diversi titoli, e tre periodi differenti permette di considerare un ampio range di situazioni. Infatti, i tre periodi mostrano diversi comportamenti del mercato finanziario: il primo periodo è caratterizzato da una fase iniziale di ribasso seguito da una ripresa e da una successiva fase laterale (2002-2006); il secondo periodo caratterizzato da una forte crisi finanziaria preceduta e seguita da fasi rialziste (2006-2011); infine, il terzo periodo è caratterizzato da una forte fase rialzista (2011-2015). Le figure da 4.1 a 4.3 mostrano il comportamento dei prezzi dei titoli considerato nel periodo di trading (linea nera) e nel periodo di

validazione (linea rossa). Come si può notare l'evoluzione dei prezzi delle azioni non necessariamente seguono quella del mercato finanziario nel suo insieme. Da sottolineare che: lo short-selling non viene praticato, non si considerano costi di transazione e neanche altri aspetti frizionali.

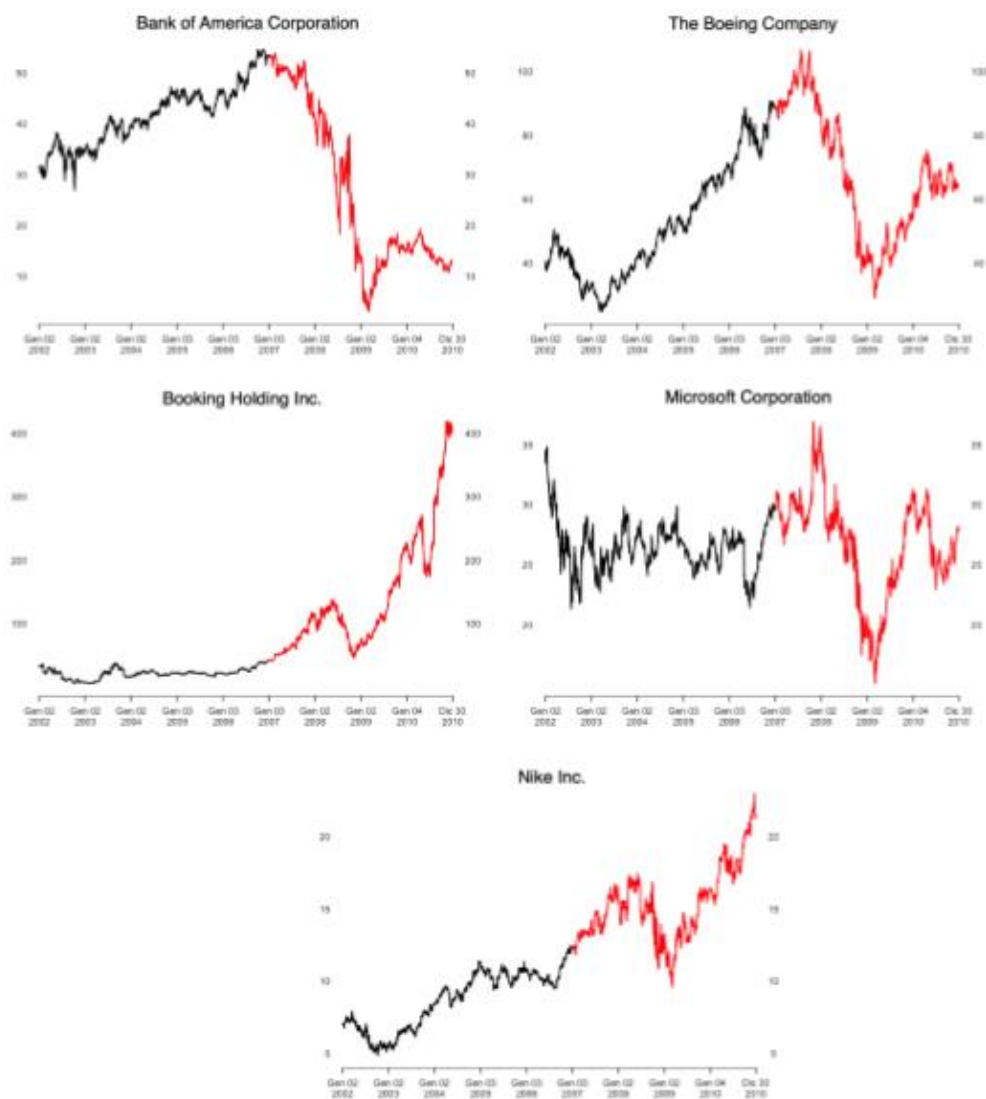


Figura 4.1: Closing prices time series of the five stocks from **02-01-2002** to **30-12-2010**: black line in the training period and red line in the validation period. (Fonte: Pizzi, Bitto e Corazza (2019))

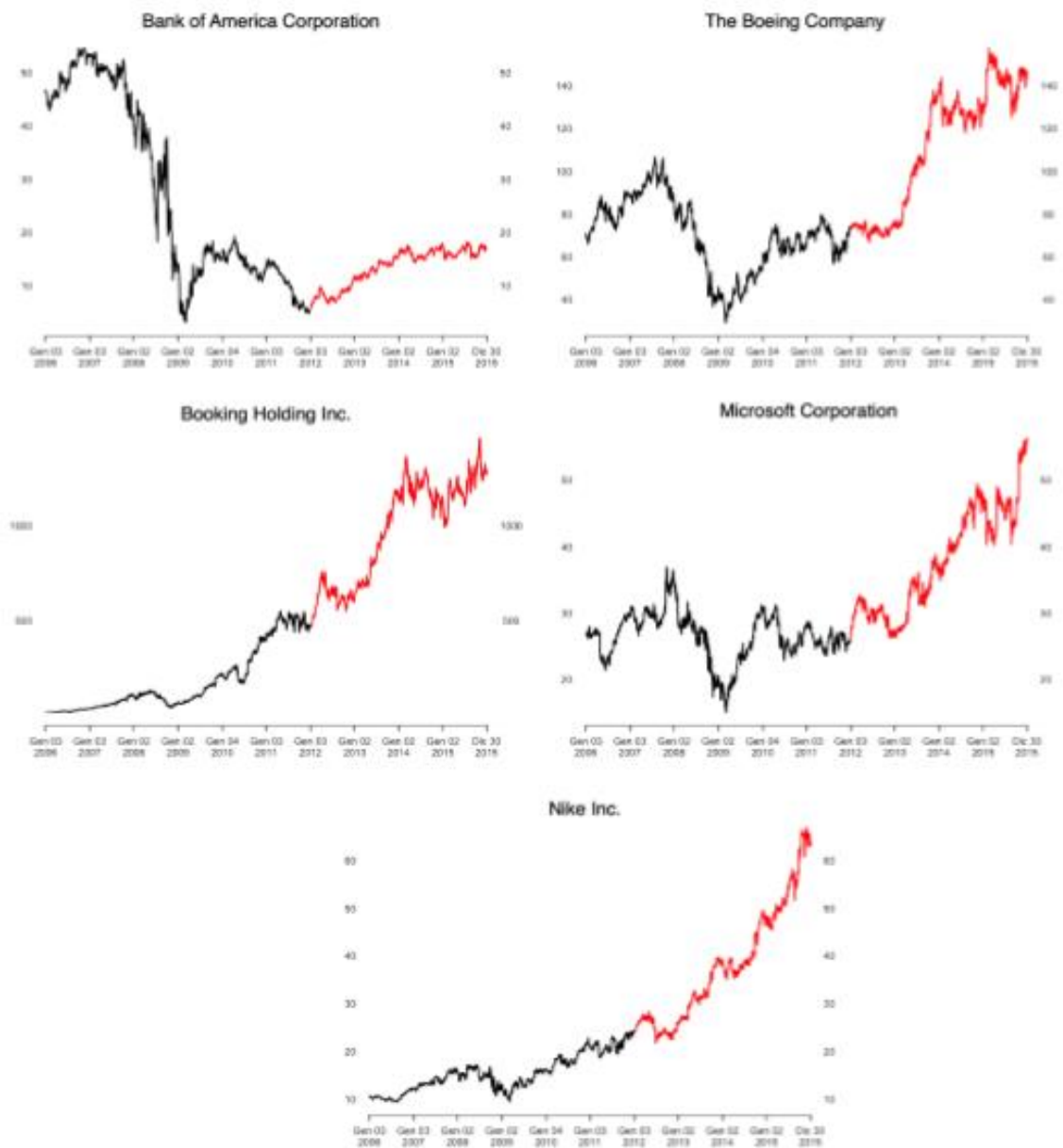
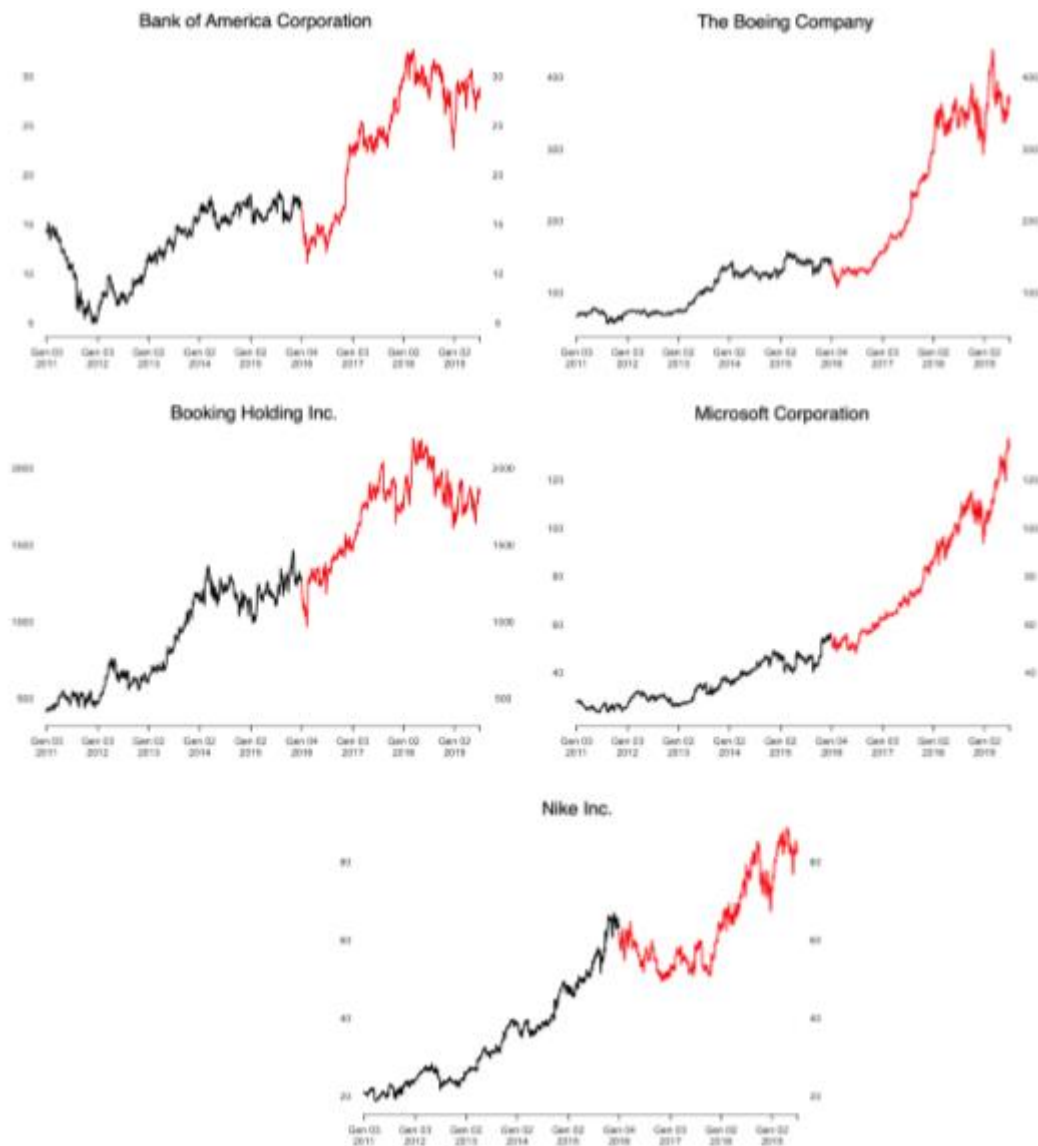


Figura 4.2: Closing prices time series **02-01-2006** to **30-12-2015**: black line in the training period and red line in the validation period. (Fonte: Pizzi, Bitto e Corazza (2019))



*Figura 4.3: Closing prices time series of the five stocks from **02-01-2011** to **30-06-2019**: black line in the training period and red line in the validation period. (Fonte: Pizzi, Bitto e Corazza (2019))*

Per quanto riguarda le impostazioni delle due metaeuristiche, per FA si considerano 5 fireworks, 50 sparks per firework, mentre per l'algoritmo PSO si considerano 10 particelle. Per entrambi gli algoritmi si fissa il numero massimo di iterazioni a 200. Gli altri parametri necessari per le due metaeuristiche sono stati settati seguendo l'indicazione della letteratura. L'applicazione di ciascuna metaeuristica (PSO e FA) è stata replicata 100 volte.

Stock	Approach	N	Fitness score	Parameters			Out-of-sample rate of return	
				n	m.up	m.dn	One year	Four years
BAC	FA	1532	1.761	5	2.068	2.073	-0.054 (14)	0.382 (43)
	PSO	910	1.816	5	2.073	2.070	-0.054 (14)	0.382 (43)
	Stan.		0.509	20	2.000	2.000	-0.120 (2)	-0.690 (10)
BC	FA	1052	2.532	96	3.791	2.830	0.000 (0)	0.008 (1)
	PSO	600	2.532	96	5.000	2.824	0.000 (0)	0.008 (1)
	Sta		0.352	20	2.000	2.000	-0.025 (3)	-0.117 (12)
BKNG	FA	715	7.356	5	2.155	1.566	2.108 (11)	5.958 (40)
	PSO	990	8.955	5	2.152	1.305	2.223 (12)	7.496 (45)
	Sta		0.082	20	2.000	2.000	0.325 (3)	0.207 (10)
MC	FA	1169	1.242	83	2.546	2.288	0.246 (1)	0.279 (3)
	PSO	380	1.242	84	2.546	2.267	0.246 (1)	0.279 (3)
	Sta		0.133	20	2.000	2.000	0.010 (2)	-0.402 (9)
NKE	FA	1210	2.669	32	3.874	2.979	0.000 (0)	1.031 (3)
	PSO	720	2.669	32	3.885	2.944	0.000 (0)	1.031 (3)
	Sta		0.314	20	2.000	2.000	0.064 (3)	0.951 (15)
BAC	FA	1605	5.485	6	2.405	2.730	0.044 (3)	0.709 (14)
	PSO	440	4.627	6	1.953	2.592	0.013 (3)	0.248 (19)
	Sta		-0.830	20	2.000	2.000	0.383 (5)	1.291 (17)
BC	FA	805	1.498	98	4.740	3.037	0.107 (1)	1.169 (1)
	PSO	800	1.941	5	1.041	1.432	0.151 (18)	0.507 (73)
	Sta		0.182	20	2.000	2.000	0.109 (3)	0.564 (22)
BKNG	FA	1060	48.479	6	3.170	2.516	-0.118 (2)	0.870 (4)
	PSO	450	37.889	6	3.158	2.258	-0.091 (2)	1.196 (5)
	Sta		0.579	20	2.000	2.000	0.129 (4)	0.664 (12)
MC	FA	738	1.382	28	3.018	3.401	0.000 (0)	0.446 (4)
	PSO	350	1.382	28	3.035	3.460	0.000 (0)	0.446 (4)
	Sta		-0.415	20	2.000	2.000	-0.122 (3)	0.075 (11)
NKE	FA	740	3.972	27	2.551	1.669	0.089 (3)	0.854 (11)
	PSO	360	3.972	27	2.559	1.676	0.089 (3)	0.854 (11)
	Sta		1.705	20	2.000	2.000	0.022 (5)	0.369 (11)
BAC	FA	659	1.880	9	3.508	3.179	0.963 (2)	1.335 (3)
	PSO	190	1.814	10	4.743	3.225	0.999 (6)	1.396 (18)
	Sta		0.358	20	2.000	2.000	0.424 (3)	0.851 (10)
BC	FA	764	1.563	54	3.802	3.068	0.000 (0)	0.000 (0)
	PSO	250	2.830	25	3.464	3.207	0.000 (0)	0.000 (0)
	Sta		0.807	20	2.000	2.000	0.227 (2)	0.799 (10)
BKNG	FA	941	3.541	36	1.247	1.182	0.596 (12)	1.541 (37)
	PSO	360	4.343	5	2.960	2.542	0.644 (16)	1.370 (42)
	Sta		0.148	20	2.000	2.000	0.700 (4)	1.469 (12)
MC	FA	568	2.485	6	1.043	1.038	0.435 (4)	2.491 (20)
	PSO	280	2.638	5	1.911	1.000	0.492 (4)	2.036 (17)
	Sta		0.148	20	2.000	2.000	0.146 (4)	0.687 (11)
NKE	FA	991	2.647	31	3.301	1.961	0.088 (2)	0.589 (6)
	PSO	660	2.587	28	3.258	1.913	0.056 (2)	0.490 (6)
	Sta		0.556	20	2.000	2.000	0.042 (4)	0.516 (11)

Table 4.1: Results obtained by the two FTSS optimized by FA and PSO, respectively, and by the FTSS set following the standard indications from the professional practice. Trading periods: 2002-2006 (table above); 2006-2011 (table in the middle); 2011-2015 (table below). (Fonte: Pizzi, Bitto e Corazza (2019))

4.1.2 Analisi risultati

I risultati ottenuti dai due TS ottimizzati utilizzando le metaeuristiche FA e PSO sono comparate con il risultato ottenuto da una TS configurato utilizzando i parametri comunemente usati dalla pratica professionale, cioè $n=20$, $m.up=2$ e $m.dn=2$. Considerando il valore della funzione di fitness, i TS ottimizzati ottengono sempre risultati migliori rispetto al TS settato seguendo la parametrizzazione standard (colonna 4 della tabella 1).

Da sottolineare che, in alcuni casi FA e PSO sebbene raggiungano punteggio fitness simile, i parametri ottimizzati sono diversi, specialmente nell'ultimo periodo di trading (colonna 4, 5, 7 della tabella 1). Questa può essere per causa dell'appiattimento della funzione obiettivo vicina al punto massimo vincolato.

Da notare che, in ogni TS ottimizzato, $m.up$ e $m.dn$ sono generalmente diversi. Questo tende a confermare l'ipotesi fatta inizialmente sull'asimmetria delle variazioni di prezzo che richiede parametri per il calcolo delle bande di Bollinger diversi. In particolare, nel 80% dei casi considerati, si ha $m.up > m.dn$, che indica una prevalenza di una asimmetria giustificata nelle variazioni dei prezzi dei titoli. Infine, una differenza operativa significativa tra FA e PSO consiste nel numero di punti dello spazio delle soluzioni valutati tramite la funzione fitness. Infatti, FA ottimizza la funzione fitness valutando, in media, più del doppio dei punti dello spazio di soluzioni valutato da PSO (colonna 3, tabella 1). Questa evidenza è graficamente esemplificata nella figura 4.5 (titolo Microsoft) durante il primo periodo di training. In questa figura, si riporta la stima della densità del numero dei punti dello spazio di soluzione: a colore più "caldo" corrisponde un maggiore numero di valutazione.

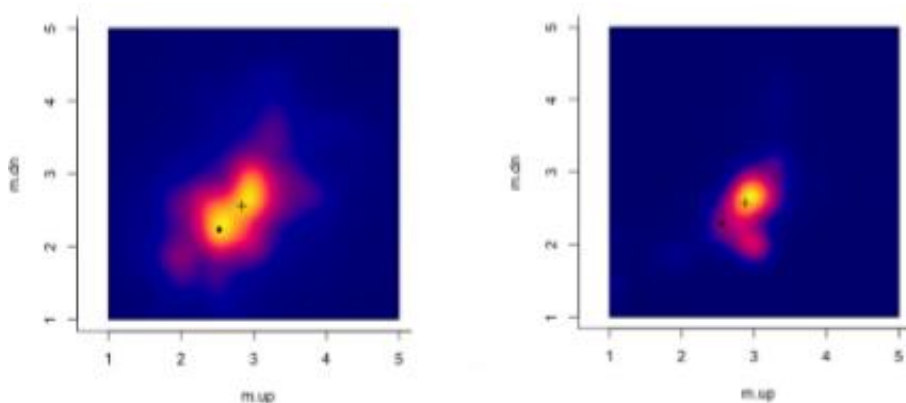


Figura 4.5: Coverage of the solution space restricted to m.up and m.dn by FA (graph on the left) and by PSO (graph on the right) for the Microsoft Corporation stock. Training period: 2002-2006. The symbol “+” indicates the mean best global optimum, the symbol “●” denotes the best global optimum. (Fonte: Pizzi, Bitto e Corazza (2019))

Questa peculiarità permette all’algoritmo Fireworks di performare meglio. ciononostante, non sembra che ciò abbia particolari impatti positivi nella performance del TS. Per quanto riguarda le performance degli algoritmi PSO e FA nell’analisi out-of-sample, i risultati sono presentati in termini di tasso di rendimento medio annualizzato. In particolare, ogni TS è stato valutato utilizzando un anno di operatività out-of-sample (Colonna 8, tabella 1) e 4 anni di operatività out-of-sample (colonna 10, tabella 1). nelle colonne 9, e 11, della stessa tabella, sono riportate anche il numero delle operazioni chiuse (acquisto e successiva vendita).

Generalmente il TS con parametri determinati utilizzando le metaeuristiche performano meglio rispetto a quelli settati con parametri standard. In particolare: nel periodo di validazione di un anno i TSs basati su parametri ottenuti con le metaeuristiche FA e PSO performano meglio rispetto al settaggio standard rispettivamente nel 60% e nel 66.7% dei casi; invece nel periodo di validazione di 4 anni, i TSs basati su FA e PSO ottengono risultati migliori rispetto al TS con settaggio standard nel 86.67% e in 73.33% dei casi, rispettivamente.

Infine, si sottolinea che in 5 casi su 30 le configurazioni ottenute con le metaeuristiche non generano segnali di acquisto e vendita nei periodi out-of-samples, mentre il TS usato come benchmark ottiene rendimenti negativi o vicini allo zero.

Riassumendo i TS i cui parametri sono determinati ottimizzando la funzione di fitness mediante FA o PSO performano meglio che quelli standard. Questi risultati sottolineano le buone capacità di “exploration” ed “exploitation” degli algoritmi FA e PSO. In particolare, viene sottolineata un’attitudine migliore di exploitation di FA rispetto a quella di PSO.

5 CONCLUSIONI

I mercati finanziari, per loro stessa natura, attirano un gran numero di individui di alto livello che sono genuinamente entusiasti dalla professione scelta e il loro entusiasmo e disponibilità a condividere le conoscenze, è, un grande privilegio per tutti i trader, analisti, investitori, per tutti noi. *“There are not many people around who can both walk their talk and talk their walk”* (Tony Plummer).

Il metodo base con cui gli analisti tecnici fanno profitto è quello dell'identificare un trend nei prezzi e “cavalcare” quel trend. Per poter “limare” le fluttuazioni giornaliere dei prezzi, e per rendere più facile la distinzione e l'identificazione del trend genuino si utilizzando gli strumenti dell'analisi tecnica. Ci sono due pilastri importanti dell'analisi tecnica: 1) i patterns, che comprendono anche trendlines e channels, 2) indicatori. In particolare, molti esperti criticano l'analisi tecnica perché considera solo i movimenti dei prezzi e volumi e ignora i fattori fondamentali – la contro argomentazione si basa sull'ipotesi del mercato efficiente, in cui si afferma che il prezzo di un titolo riflette già tutto ciò che ha o potrebbe influenzare un'azienda, inclusi i fattori fondamentali. Gli analisti tecnici ritengono che i prezzi si muovano in un trend di breve, medio e lungo termine, e, secondo gli analisti è più probabile che un prezzo azionario continui a seguire una tendenza passata piuttosto che muoversi in modo irregolare e la maggior parte delle strategie di trading si basano su questa assunzione. I movimenti dei prezzi delle attività non sono solo fluttuazioni casuali guidate dal comportamento razionale di trader indipendenti. Gli esseri umani sono per natura socievoli e comunicativi, e, hanno una spinta interiore ad appartenere ai gruppi e la psicologia dei gruppi fornisce un'influenza di controllo sull'attività individuale, questo però trasforma una grande quantità di decisioni apparentemente non correlate in un risultato più certo. Importante sottolineare che questo risultato si rivela sotto forma di movimenti ritmici e strutturati dei prezzi che non solo hanno una relazione naturale tra loro ma sono anche sostanzialmente prevedibili, una volta compresi. Gli analisti tecnici ritengono che la storia tende a ripetersi; la natura ripetitiva dei movimenti dei prezzi è spesso attribuita alla psicologia del mercato che tende ad essere molto prevedibile. Il riconoscimento dei patterns è l'aspetto più visivo dell'analisi tecnica che, molto probabilmente, trae beneficio dalla ripetuta visione dei grafici passati e presenti e dal follow-up.

Inoltre, poiché i mercati tendono essenzialmente ad anticipare i movimenti dei fenomeni economici e sociali, l'uso accurato dell'analisi tecnica implica un'abilità di prevedere questi fenomeni. In particolare, un nuovo campo di studio chiamato *neurofinanza* - uno studio interdisciplinare sull'applicazione della neuroscienza alle attività di investimento - sta trovando notevoli connessioni tra il funzionamento del nostro cervello, il modo in cui prendiamo le decisioni e il modo in cui investiamo. Secondo la neurofinanza il processo decisionale degli investitori comprende sia un aspetto quantitativo (oggettivo) che qualitativo (soggettivo) che si basa sulle caratteristiche specifiche del prodotto di investimento. Il comportamento degli investitori esamina i fattori cognitivi (processi mentali) e le questioni affettive (emotive) che individui, esperti finanziari e operatori rivelano durante il processo di pianificazione finanziaria e di gestione degli investimenti. Questo nuovo campo è un interessante sfumatura del mondo finanziario da poter approfondire.

Nel terzo capitolo si presentano 2 algoritmi metaeuristici: la Particle Swarm Optimization (PSO) e Firework Algorithm (FWA). Risulta che gli algoritmi metaeuristici siano facili da implementare, flessibili - in quanto sono in grado di risolvere diversi problemi di ottimizzazione che non possono essere affrontati dagli algoritmi classici, e, che abbiano un alto grado di ergodicità, il che significa che possono ricercare spazi multimodali e allo stesso tempo evitare gli ottimi locali. Per quanto riguarda il contributo del PSO, ha visto un notevole successo in vari campi, non solo in quello finanziario, specialmente nel campo dell'ingegneria civile, dove è stato utilizzato per risolvere problemi di ottimizzazione di forma e dimensione nella progettazione strutturale; per pianificazione della costruzione; per valutazioni delle condizioni strutturali; per progettazione delle reti di distribuzione acqua; per stima, previsione dei costi di costruzione e così via.

A causa della sua facilità di implementazione e di rapida convergenza verso soluzioni accettabili, la PSO ha ricevuto ampia attenzione negli ultimi anni. Però sono state identificate delle limitazioni dello PSO standard; il termine "limitazione" si riferisce a un problema che è stato dimostrato che impedisce all'algoritmo di funzionare bene in diversi aspetti dell'operazione, come per esempio nel trovare soluzioni di alta qualità o di essere stabile. Ci sono due principali aree correlate ai limiti del PSO: convergenza e

invarianza di trasformazione. I vari studi svolti a migliorare il PSO standard possono essere distinti in 3 diverse categorie: 1) modifica delle impostazioni dei parametri – si riferisce all'impostazione di topologia, coefficienti (coefficienti di accelerazione o del peso inerziale), e dimensioni della popolazione; 2) modifica dei componenti dell'algoritmo – si riferisce alle variazioni della regola di aggiornamento della velocità o della posizione (la modifica del modo in cui vengono calcolati); 3) combinazione del algoritmo con altri algoritmi – si riferisce all'ibridazione del PSO con altri metodi. Non ci sono tanti studi per le seguenti tematiche: analisi di comportamento di una particella prima della convergenza; proprietà di invarianza della trasformazione dell'algoritmo – il che rende questi argomenti uno potenziale spazio aperto per nuove idee e analisi.

Il Firework Algorithm (FWA) offre un nuovo modo per cercare l'ottimo globale nei problemi di ottimizzazione complessi, novità questa che arricchisce lo studio dell'intelligenza dello sciame. L'attuale FWA e le sue applicazioni dimostrano che potrebbe anche essere utilizzato per risolvere efficacemente molti problemi di ottimizzazione complessi. Inoltre, FWA è abbastanza adatto per le funzioni di ottimizzazione multimodale, il che significa che troveranno una vasta gamma di applicazioni nel mondo reale. Secondo l'idea del FWA, un buon firework indica il miglior fitness, il che significa che il firework potrebbe essere vicino alla posizione ottimale. Pertanto, questo firework dovrebbe generare più scintille nella più piccola ampiezza dell'esplosione. Inoltre, FWA può anche essere parallelizzato ed è quindi adatto per affrontare i problemi dei big data; il termine "parallelismo" indica che non esiste alcun meccanismo di controllo centrale tra i fireworks, quindi i fireworks sono indipendenti, quindi molto adatti alla parallelizzazione - in ogni iterazione di FWA, ogni firework esplose e ricerca all'interno di spazi diversi, ovvero ogni firework conduce una ricerca in dimensioni diverse. FWA trova applicazione e sviluppi nei: single objective, multi-objective, implementazioni parallele, hybrid FWA più altri algoritmi. I ricercatori hanno utilizzato FWA insieme a particle swarm optimization, genetic algorithms, differential evolution e i loro risultati indicano che FWA non può performare meglio degli altri algoritmi di ottimizzazione quando il numero di dimensioni è piccolo, ma ottiene buoni risultati quando il numero di dimensioni viene aumentata. L'algoritmo di Fireworks non è ancora perfetto: in primo luogo, ha bisogno di basi matematiche e analisi teoriche; in

secondo luogo, le prospettive delle applicazioni dell' algoritmo sono ancora agli inizi e richiedono ulteriori esplorazioni; però è adatto per la combinazione con altri algoritmi per produrre un nuovo algoritmo ibrido, quindi può “imparare” dagli altri algoritmi.

FWA funziona molto bene su funzioni che hanno l'ottimo all'origine; quando invece viene applicato su funzioni spostate, la qualità dei risultati di FWA convenzionale peggiora gravemente e peggiora con l'aumentare della distanza tra la funzione ottimale e l'origine nello spazio di ricerca. E quindi per migliorare la performance della FWA (convergenza, qualità della soluzione, costo computazione) è stato introdotto un controllo dell'ampiezza minima dell'esplosione, un operatore per la generazione di scintille di esplosione, una strategia di mappatura per scintille che non rientrano nello spazio di ricerca e un nuovo operatore per la selezione della popolazione per la successiva iterazione. Esperimenti svolti da ricercatori mostrano che l'EFWA supera le FWA convenzionali in termini di capacità di convergenza, e tempo di esecuzione.

Nel case study riportato nel capitolo 4, si sottolinea il valore aggiunto e le potenzialità che hanno le metaeuristiche, se applicate ad un trading system i cui parametri sono determinati ottimizzando la funzione di fitness mediante FWA o PSO. Le performance migliorano rispetto ai Trading system standard. Questi risultati evidenziano le buone capacità di “exploration” ed “exploitation” degli algoritmi FWA e PSO. In particolare, viene sottolineata un'attitudine migliore di exploitation di FWA rispetto a quella di PSO.

6 BIBLIOGRAFIA

Charles D. Kirkpatrick, Julie R. Dahlquist (2015). *Technical Analysis, The Complete Resource for Financial Market Technicians*, FT Press, pp. 25-483.

Welles Wilder Jr. (1978). *New Concepts in Technical Trading Systems*, *Commodities Magazine*, *Now Futures Magazine*.

H. Kent Baker, Victor Ricciardi (2014). *Investor Behavior, The Psychology of Financial Planning and Investing*, Wiley, pp. 3-58, pp. 327-452.

Thomas N. Bulkowski (2014). *Encyclopedia of Chart Patterns*, Wiley.

John Bollinger (1992). *Using Bollinger Bands*, *Technical Analysis of Stock & Commodities*.

Louis B. Mendelsohn (2000). *Trend Forecasting with Technical Analysis*, *Market Technologies Corporation*, pp. 45-95.

Konstantinos E. Parsopoulos, Michael N. Vrahatis (2010). *Particle Swarm Optimization and Intelligence Advances and Applications*, *Information Science Publishing, Premier Reference Source*, pp. 1-164.

Mohammad Reza Bonyadi, Zbigniew Michalewicz (2017). *Particle Swarm Optimization for Single Objective Continuous Space Problems, a Review*, *MIT Press Journals*.

Serkan Kiranyaz, Turker Ince, Moncef Gabbouj (2012). *Multidimensional Particle Swarm Optimization for Machine Learning and Pattern Recognition, Adaptation, Learning, and Optimization*, vol. 15, pp. 1-85.

Corazza Marco, Parpinel Francesca, Pizzi Claudio (2018). *Can PSO improve TA-based trading systems?* *Smart Innovation, Systems and Technologies, Cham, Springer International Publishing*, vol. 102, pp. 277-288.

Amaresh Saha, Sushanta Kumar Panigrahi, Sabyasachi Pattnaik (2011). *Fast Convergence Particle Swarm Optimization for Functions Optimization*, *Elsevier*.

Wei-Neng Chen, Jun Zhang, Ying Lin, Ni Chen, Zhi-Hui Zhan, Henry Shu-Hung Chung, Yun Li, Yu-Hui Shi (2013). Particle swarm optimization with an aging leader and challengers, *IEEE Transactions on Evolutionary Computation*, vol. 17, pp. 241–258.

Zhi-Hui Zhan, Jun Zhang, Yun Li, Yu-Hui Shi (2011). Orthogonal learning particle swarm optimization, *IEEE Transactions on Evolutionary Computation*, vol.15, pp. 832–847.

Liang, J.J. Qin, A.K., Suganthan, P.N., Baskar, S. (2006). Comprehensive learning particle swarm optimizer for global optimization of multimodal functions, *IEEE Transactions on Evolutionary Computation*, vol.10, pp. 281–295.

Cheng, R., Jin, Y., (2015). A social learning particle swarm optimization algorithm for scalable optimization, *Information Sciences, Elsevier*, pp. 43–60.

Haiyan Shi, Shilong Liu, Hongkun Wu, Ruowei Li, Sanchi Liu, Ngaiming Kwok, Yeping Peng (2018). Oscillatory Particle Swarm Optimizer, *Applied soft computing, Elsevier*, vol.73, pp. 316-327.

Yuhui Shi, R.C. Eberhart (2001). Fuzzy adaptive particle swarm optimization, *Proceedings of the 2001 Congress on Evolutionary Computation*, vol.1, pp. 101–106.

Liang, J., Suganthan, P.N., (2005). Dynamic multi-swarm particle swarm optimizer, *Proceedings 2005 IEEE Swarm Intelligence Symposium*, pp. 124–129.

Mohammadreza Bonyadi (2014). Particle Swarm Optimization: Theoretical analysis, modifications, and applications to constrained optimization problems, A thesis for the degree of doctor in philosophy, *School of computer Science, The University of Adelaide*.

Wang, L., Yang, B., Chen, Y., (2014). Improving particle swarm optimization using multilayer searching strategy, *Information Sciences 274*, pp. 70–94.

Marco Corazza, Giacomo di Tollo, Giovanni Fasano, Raffaele Pesenti (2019). A PSO-based framework for non-smooth portfolio selection problems, *Neural Advances in Processing Nonlinear Dynamic Signals, Cham, Springer Science and Business Media Deutschland GmbH*, vol. 102, pp. 265-275.

Ying Tan (2015). Fireworks Algorithm, A Novel Swarm Intelligence Optimization Method, *Springer-Verlag Berlin Heidelberg*.

Claudio Pizzi, Irene Bitto, Marco Corazza (2019). Exploration and Exploitation in Optimizing a Basic Financial Trading System: A Comparison Between FA and PSO Algorithms, *Progresses in Artificial Intelligence and Neural Systems*, Singapore, *Springer Nature*, vol. 184, pp. 293-303.

Ying Tan, Yuanchun Zhu (2010). Fireworks Algorithm for Optimization, Peking University Department of Machine Intelligence, *School of Electronics Engineering and Computer Science*.

Shaoqiu Zheng, Andreas Janecek and Ying Tan (2013). Enhanced Fireworks Algorithm, *Department of Machine Intelligence, School of Electronics Engineering and Computer Science, Peking University*.

Ying Tan, Yuhui Shi, Hongwei Mo (2013). Advances in Swarm Intelligence, *4th International Conference ICSI*.

7 SITOGRAFIA

- 1) www.sciencedirect.com
- 2) www.investopedia.com
- 3) www.stockcharts.com
- 4) www.edx.org
- 5) www.tradingview.com
- 6) www.investing.com
- 7) www.wikipedia.com
- 8) <https://www.cil.pku.edu.cn/fwa/developments/index.htm>
- 9) <https://www.exfinsis.com/>

Ringraziamenti

A Te.

Alla mia Famiglia che mi è sempre stata vicina nonostante la lontananza, soprattutto a mia Mamma.

A Professor Pizzi per la collaborazione, pazienza, per la fiducia e l'entusiasmo che mi ha trasmesso, e infine per aver stimolato il mio interesse per l'argomento qui discusso.

A Professor Vinelli, Maria Barbara, Alice, Gianluca, Giorgio per avermi "vista" e creduto nelle mie capacità.

A C. Zarabara, Flavio per il senso dell'umorismo che mi salvano le giornate.

Ad Alessandro.

A Pellegrino che con la sua precisione e creatività, mi ha insegnato tanto.

A Vilma, Valentina, Davide, Simone, Abo, Maestro, Enrico, Moretto, Pana, Vjola per esserci stati, professionalmente e umanamente.

A Gio per essere com'è, e per lo spirito collaborativo.

A Sarish, Klajdi, Irma, Cinzia, Nini, Koli, Andrea, con cui si sta bene assieme senza dover spiegare niente, a voi che siete "casa".

A tutte le persone che conosco ma non ho elencato.

E...a me stessa :) .

