



Ricerca di Sistema elettrico

Analisi multiscala e multiactors decision
making per il trattamento di grandi moli di
dati in contesti smart city

Gerardo Iovane



UNIVERSITÀ DEGLI STUDI DI SALERNO



UNIVERSITÀ DEGLI STUDI DI SALERNO
DIPARTIMENTO DI INFORMATICA
DIPARTIMENTO DI ECCELLENZA

ANALISI MULTISCALE E MULTIACTORS DECISION MAKING PER IL TRATTAMENTO DI GRANDI MOLI DI DATI IN CONTESTI SMART CITY

Gerardo Iovane (Università di Salerno, Dipartimento di Informatica)

Settembre 2018

Report Ricerca di Sistema Elettrico

Accordo di Programma Ministero dello Sviluppo Economico - ENEA

Piano Annuale di Realizzazione 2017

Area: Efficienza energetica e risparmio di energia negli usi finali elettrici e interazione con altri vettori energetici

Progetto: D.6 Sviluppo di un modello integrato di smart district urbano

Obiettivo: D.6a Piattaforma ICT per la gestione di smart district

Responsabile del Progetto: Claudia Meloni, ENEA

Il presente documento descrive le attività di ricerca svolte all'interno dell'Accordo di collaborazione "Analisi multiscale e multiactors decision making per il trattamento di grandi moli di dati in contesti smart city"

Responsabile scientifico ENEA: Marta Chinnici

Responsabile scientifico UNISA: Gerardo Iovane

Indice

INDICE FIGURE	4
SOMMARIO	5
INTRODUZIONE	6
1 ANALISI DI TECNICHE NON LINEARI PER IL SIGNAL PROCESSING	7
1.1 METODOLOGIE MULTISCALE IN SPAZI TRASFORMATI	7
1.1.1 <i>Fast Fourier Transform</i>	7
1.1.2 <i>La Trasformata di Walsh-Hadamard</i>	8
1.1.3 <i>La Trasformata di Hartley</i>	8
1.1.4 <i>La Trasformata di Laplace</i>	9
1.1.5 <i>La Trasformata di Hilbert</i>	10
1.2 WAVELETS	12
1.1.6 <i>La Trasformata Wavelet continua</i>	12
1.1.7 <i>La Trasformata Wavelet discreta: Frames</i>	15
1.1.8 <i>Wavelets ortogonali e l'analisi multirisoluzione</i>	18
1.1.9 <i>MultiWavelerts</i>	22
1.1.10 <i>Dizionari Tempo-Frequenza e Wavelet Packets</i>	25
2 LE PRINCIPALI APPLICAZIONI PRESENTI IN LETTERATURA RELATIVE ALLE METODOLOGIE ANALIZZATE IN DATA FUSION IN CONTESTI URBANI	30
2.1 RICHIAMI AL PAR 2016	30
2.2 INTRODUZIONE ALLA DATA FUSION	30
2.3 OUTLINE	32
3 IL TRATTAMENTO DELLE INFORMAZIONI INCERTE O INCOMPLETE IN PROSPETTIVA PROBABILISTICA CON APPLICAZIONE IN SMART CITY	34
3.1 LA PLAUSIBILITÀ	34
3.1.1 <i>Caratteristiche</i>	34
3.1.2 <i>Riferimenti storici: la tradizione greca</i>	35
3.1.3 <i>Incerteza dei metodi</i>	35
3.1.4 <i>Incerteza e allocazione della credenza globale</i>	35
3.2 LA DEFINIZIONE PROBABILISTICA	36
3.2.1 <i>La nascita della concezione probabilistica</i>	36
3.2.2 <i>Brevi cenni sull'accezione di probabile</i>	36
3.2.3 <i>Il teorema di Cox</i>	37
3.2.4 <i>Il teorema di Bayes</i>	38
3.2.5 <i>Barriere e limiti</i>	38
3.3 LA LOGICA DEL PLAUSIBILE DI POLYA	39
3.3.1 <i>L'inferenza plausibile</i>	39
3.3.2 <i>Interpretazione probabilistica dei modelli d'inferenza plausibile</i>	42
3.3.3 <i>Analogia e scoperta</i>	43
3.3.4 <i>Analogia e Probabilità</i>	43
3.3.5 <i>Il credibile</i>	43
3.3.6 <i>Barriere e limiti</i>	44
3.4 LA CONCEZIONE DI DEMPSTER-SHAFER	45
3.4.1 <i>La teoria dell'evidenza</i>	45
3.4.2 <i>Plausibilità e credenza</i>	45
3.4.3 <i>Barriere e Limiti</i>	48
3.5 LA CONCEZIONE DELLA TEORIA DEZERT SMARANDACHE	49

3.5.1	<i>La teoria del ragionamento plausibile</i>	49
3.5.2	<i>Plausibilità e credenza</i>	49
3.5.3	<i>Barriere e limiti</i>	49
3.6	LA CONCEZIONE NON PROBABILISTICA	50
3.6.1	<i>Motivazioni e obiettivi</i>	50
3.6.2	<i>Le deduzioni di Rescher</i>	50
3.6.2.1	<i>Le inferenze plausibili</i>	50
3.6.2.2	<i>Plausibilità e probabilità a confronto</i>	51
3.6.3	<i>L'Entimema</i>	51
3.6.3	<i>Barriere e limiti</i>	51
3.7	PLAUSIBILITÀ E NON MONOTONIA	52
3.7.1	<i>Logiche non monotone</i>	52
3.7.2	<i>Barriere e limiti</i>	52
3.8	COGNITIVISMO E PLAUSIBILITÀ	54
3.8.1	<i>L'approccio cognitivista</i>	54
3.8.2	<i>Concetti e basi</i>	54
3.8.3	<i>La rappresentazione della conoscenza</i>	54
3.8.4	<i>I modelli d'inferenza plausibile</i>	54
3.8.5	<i>Barriere e limiti</i>	56
3.9	VERSO UNA TEORIA DEL RAGIONAMENTO PLAUSIBILE	57
3.9.1	<i>Sistemi aperti e reti</i>	57
3.9.2	<i>Reti di inferenza plausibile</i>	59
3.10	DECISIONE IN DOMINI DI INFO-INCERTEZZA E/O DI INFO-INCOMPLETEZZA	60
3.10.1	<i>Lo scenario</i>	60
3.10.2	<i>Il continuo verificarsi dell'improbabile possibile</i>	63
3.10.3	<i>Necessità di metodi d'inferenza per strumenti finanziari dominati dall'incertezza: interazione operatore finanziario- mercato</i>	66
3.10.4	<i>Trattamento dell'incertezza via fusione delle informazioni</i>	67
3.10.5	<i>Information Fusion e modellazione della funzione di aspettazione</i>	68
3.10.6	<i>Modello della media</i>	68
3.10.7	<i>Modello del prodotto</i>	68
3.10.8	<i>Modello della media pesata</i>	68
3.10.9	<i>Modello del prodotto pesato</i>	69
3.10.10	<i>Modello basato sulle regole di composizione di Dempster</i>	69
3.10.11	<i>Modello di sovrapposizione con shift basato sulla probabilità</i>	69
3.10.12	<i>Modello di sovrapposizione con shift basato sulle Pi considerate in modo gerarchico</i>	70
3.10.13	<i>L'affidabilità</i>	70
3.10.14	<i>Attendibilità</i>	71
4	CONCLUSIONI & PROSPETTIVE	73
5	RIFERIMENTI BIBLIOGRAFICI	75

Indice Figure

Figura 1: Decomposizioni Wavelet Packet	29
Figura 2: Grafico di esempio per dati incompleti	60

Sommario

Il presente studio trae origine dall'Accordo di Programma tra il Ministero dello Sviluppo Economico (MiSE) e l'Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile (ENEA) sulla Ricerca di Sistema Elettrico relativamente al Piano Annuale di Realizzazione 2017. Il presente documento rientra nell'ambito dell'obiettivo D.6a Piattaforma ICT per la gestione di smart districts che prevede la realizzazione di una piattaforma software in grado di raccogliere grandi moli di dati provenienti da distretti urbani. La gestione e l'estrazione delle informazioni dai dati possono aiutare nella creazione di servizi capaci di migliorare la qualità delle città e indirizzarle alla smartizzazione. Con particolare riferimento alle informazioni derivanti dalle solution verticali della smart city platform, lo studio inerente al tema in esame è costituito da un rapporto tecnico che si articola in tre sezioni principali:

- Metodologie multiscala in spazi trasformati in relazione;
- Principali applicazioni presenti in letteratura relative alle metodologie analizzate in data fusion riguardanti i contesti urbani;
- Trattamento delle informazioni incerte o incomplete in prospettiva probabilistica.

L'obiettivo della prima sezione è di offrire una panoramica prima generale e poi contestualizzata dell'analisi multiscala e degli spazi trasformati per il trattamento di Big Data in contesti complessi come le smart cities dove l'informazione più trovare un trattamento più efficace qualora la si relazioni alla scala.

La seconda sezione concerne l'information fusion ed i sistemi di supporto alle decisioni. I risultati indicati in essa sono i richiami a quanto presentato nel PAR 2016 di cui questo studio è la naturale continuazione. Infatti, lo studio del PAR2016 approfondiva due temi fondamentali, ovvero: a) la Definizione di metodologie, metodi e modelli della complessità per il decision making per la smart city, b) gli Esempi di procedure ed algoritmi per il decision making in smart cities context. Grazie all'analisi multiscala offerta nella prima sezione si rende quindi capace una soluzione tecnologica di estendere i risultati ottenuti con una analisi efficace che espone diverse viste dell'informazione con una granularità diversa a seconda della scala a cui si è interessati ovvero dell'Ente/attore considerato.

La terza sezione completa il presente studio affrontando un ulteriore tema di interesse, ovvero la possibilità di gestire informazioni incerte o incomplete che sono presenti all'interno del contesto smart city e che caratterizzano di fatto le grandi moli di dati da esso provenienti. Per raggiungere un tale obiettivo si è reso necessario operare in un contesto di probabilità estesa. Infatti, è consueto che in ambiti propri della complessità il concetto classico di probabilità non sia più sufficiente. Spesso si è avviato alla costruzione di nuovi modelli, considerando il fenomeno in studio descritto attraverso una distribuzione gaussiana e modellando le discrepanze tra il modello ed i dati reali, aggiungendo delle opportune code pesanti alla distribuzione normale. Data l'importanza del lavoro e la possibilità di impiego in molti altri contesti, in questo studio è stata costruita una nuova metodologia della attendibilità/aspettazione che coniuga in modo sinergico il concetto di probabilità, plausibilità, credibilità e possibilità all'interno di uno scenario che tratti informazione incerta ovvero incompleta.

Le conclusioni riassumono i risultati conseguiti, ovvero la capacità di costruire un sistema di supporto alle decisioni che si basi sull'information fusion per il trattamento di big data in contesti smart cities, prevedendo la possibilità di analisi multiscala/multiente ed il trattamento di informazione incerta e/o incompleta.

Introduzione

Nella maggior parte dei casi, in cui ci si appropria a risolvere problemi, ci si trova a creare conoscenza. I metodi per analizzare i grandi moli di dati, ogni talvolta si sono posti in prospettiva di creare inferenza e costruire previsioni che hanno fatto riferimento alla probabilità. Tuttavia però, spesso le informazioni dipendono dalla scala di osservazione e non seguono regole probabilistiche. Varie volte troviamo intervalli di tempo in cui l'informazione appare causale, ovvero che nelle distribuzioni ci sono cosiddette "code pesanti", cioè non rappresentabili con la distribuzione normale. In questo studio costruiremo una logica dell'incertezza e dell'info-incompletezza considerando due diverse metodologie che possono essere usate singolarmente o fuse insieme:

1. Da un lato l'analisi multiscala in spazi trasformati;
2. Dall'altro affiancando al concetto di probabilità, quelli di plausibilità, credibilità, possibilità, affidabilità, attendibilità.

Ci sarà una distinzione tra i seguenti concetti: incertezza dei dati iniziali, incertezza degli obiettivi e infine dei metodi.

Relativamente al punto 1 lo scopo di questo studio riguarda l'analisi e l'applicazione di metodologie multiscala non lineari in spazi trasformati in contesti dove è necessario fondere e sintetizzare i dati. In particolare, verranno analizzate varie tecniche in cui grazie ad una fusione dei dati adattiva che analizzi i dettagli o le approssimazioni dei segnali provenienti dai sensori distribuiti in campo si potrà avere una rappresentazione più efficace dei dati.

Lo studio riguarderà una esaustiva introduzione agli strumenti matematici utili in contesti di signal processing in spazi trasformati; in dettaglio, ci si focalizzerà sullo studio delle basi wavelets, multiwavelets, wavelets vettoriali, dizionari tempo-frequenza e pacchetti wavelets, dapprima analizzandone, da un punto di vista puramente analitico, le proprietà matematiche, e successivamente cercando di calare tali studi in contesti di interesse quali appunto la fusione dei dati raccolti da sensori in contesti smart cities.

A tal riguardo nel primo capitolo, daremo una analisi estemporanea inerente le principali trasformate comunemente utilizzate in letteratura in contesti di signal processing. In particolare, verranno analizzate la trasformata di Fourier veloce (Fast Fourier Transform - FFT), la trasformata di Walsh-Hadamard, di Hartley, di Laplace, e di Hilbert, ed infine la trasformata Wavelet. Sulla cui trattazione ci soffermeremo in modo particolare, analizzando non solo la classica teoria delle basi wavelets, ma anche le naturali generalizzazioni che condurranno all'analisi delle Multivavelets, dei pacchetti wavelets e dei generici dizionari tempo-frequenza.

Nel secondo capitolo, verranno introdotte le tecniche e le metodologie già utilizzate in letteratura e relative alla fusione dei dati in spazi trasformati con tecnologie wavelet-based.

Nel terzo ed ultimo capitolo le metodologie multiscala per la fusione dei dati verranno contestualizzate nel settore delle smart cities in condizioni di info-incertezza e/o info-incompletezza.

Relativamente al punto 2 partendo dalla considerazione che una pura analisi multiscala, seppure spesso utile per guardare alle informazioni da una prospettiva diversa (si pensi all'analisi frequenziale), non sempre permette di avere informazioni complementari circa la probabilità che è associato al dato ed atteso che spesso i dati con cui si ha a che fare non seguono una distribuzione di probabilità nota, ovvero mostrano cosiddette code pesanti, che implicano che eventi rari si verificano più frequentemente di quanto atteso, considereremo versione estese del concetto di accadibilità o aspettazione in cui alla probabilità classica di un evento si accompagna la sua plausibilità, la credibilità e la possibilità di accadimento.

Il risultato finale sarà un information fusion ed una sintesi efficace delle informazioni in contesti di informazione incompleta e/o incerta.

1 Analisi di Tecniche non Lineari per il Signal Processing

Come premesso, l'obiettivo di questo capitolo consiste nel fornire una sintetica, ma allo stesso tempo esaustiva review inerente alle principali trasformate comunemente utilizzate in letteratura in contesti di analisi di dati bidimensionali come immagini ma anche come dati di misure scalari accompagnati da un secondo dato scalare legato alla sua accadibilità, ovvero alla precisione di misura, ecc. In particolare, verranno analizzate la trasformata di Fourier veloce (Fast Fourier Transform - FFT), la trasformata di Walsh-Hadamard, di Hartley, di Laplace, e di Hilbert, ed infine la trasformata Wavelet. Sulla cui trattazione ci soffermeremo in modo particolare, analizzando non solo la classica teoria delle basi wavelets, ma anche le naturali generalizzazioni che condurranno all'analisi delle Multivavelets, dei pacchetti wavelets e dei generici dizionari tempo-frequenza. In particolare, dopo una breve panoramica riguardante la FFT, la trasformata di Walsh-Hadamard, di Hartley, di Laplace, e di Hilbert, nel secondo paragrafo, a partire dalla definizione della trasformata wavelet continua, definiremo la trasformata wavelet ortogonale attraverso lo studio dell'analisi multirisoluzione. Successivamente generalizzeremo tali studi al caso vettoriale analizzando la teoria delle Multiwavelet, delle Wavelet Packet ed infine dei dizionari tempo frequenza.

1.1 Metodologie multiscala in spazi trasformati

Come premesso, in questo paragrafo, forniremo una breve panoramica relativa alle principali trasformate, comunemente utilizzate in contesti di analisi dei segnali e delle immagini. In particolare, analizzeremo la trasformata di Fourier veloce (Fast Fourier Transform - FFT), la trasformata di Walsh-Hadamard, di Hartley, di Laplace, ed infine la trasformata di Hilbert.

1.1.1 Fast Fourier Transform

Assegnato un segnale discreto $X = \{x_j\}_{j=0, \dots, N-1}$, la trasformata di Fourier discreta (DFT) e la sua inversa (DFT^{-1}) entrambe applicate al segnale X , sono definite attraverso le seguenti identità:

$$\hat{x}_k = \frac{1}{N} \sum_{j=0}^{N-1} x_j e^{-\frac{2\pi i k j}{N}},$$

$$x_j = \sum_{k=0}^{N-1} \hat{x}_k e^{\frac{2\pi i k j}{N}} \quad (1.1)$$

L'implementazione della DFT per un segnale di dimensione N richiede un numero di operazioni pari a $O(N)$, ma nel caso in cui il segnale da trasformare abbia una dimensione pari ad una potenza di 2, il numero delle operazioni si riduce a $O(N \log_2 N)$.

In realtà sebbene, la Fast Fourier Transform (FFT), sia un algoritmo veloce per il calcolo della DFT di un segnale avente un numero di campioni pari ad una potenza di 2, lo si può applicare anche a segnali discreti che non soddisfano tale condizione; basta, infatti, in base al cosiddetto metodo dello *zero padding*, aumentare i campioni del segnale in esame, aggiungendo il minimo numero di zeri sufficiente ad ottenere un segnale avente una dimensione pari ad una potenza di 2.

1.1.2 La Trasformata di Walsh-Hadamard

La trasformata di Walsh-Hadamard ha caratteristiche simili alla trasformata di Fourier, ma ammette un costo computazionale inferiore. Essa è generalmente usata per l'elaborazione dei segnali e delle immagini, per il riconoscimento di pattern e filtraggio di sequenze.

Definizione 1.1 Sia $X = [x_1, \dots, x_N]^T$ un segnale campionato a dimensione $N = 2^n$, dove n rappresenta un numero intero, ebbene la trasformata di Walsh-Hadamard del segnale è data dal seguente prodotto:

$$WH\{X\} = H_N \cdot X, \quad (1.2)$$

dove H_N rappresenta la matrice di Hadamard di dimensione $N \times N$, soddisfacente la seguente formula ricorsiva:

$$H_N = H_{2^n} = \begin{bmatrix} H_{N/2} & H_{N/2} \\ H_{N/2} & -H_{N/2} \end{bmatrix}. \quad (1.3)$$

Di conseguenza, applicando in modo ricorsivo la (1.3) alla (1.2) otteniamo

$$WH\{X\} = \left(\bigotimes_{i=1}^n H_2 \right) \cdot X, \quad (1.4)$$

dove

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad (1.5)$$

rappresenta la 2×2 matrice di Hadamard e \otimes indica il prodotto tensoriale o di Kronecker.

1.1.3 La Trasformata di Hartley

La trasformata di Hartley rappresenta una valida alternativa non solo alla trasformata di Fourier discreta, ma anche alla trasformata Coseno, i vantaggi che spingono ad utilizzare una trasformata piuttosto che l'altra vanno analizzati non solo in termini di costo computazionale, ma soprattutto in vista del fatto che la trasformata di Hartley è a nucleo reale.

Come vedremo a breve, infatti, a differenza della trasformata di Fourier, il cui nucleo risulta pari a $\cos(\cdot) - i \sin(\cdot)$, la trasformata di Hartley ammette *Kernel* reale, dato da $\cos(\cdot) + \sin(\cdot)$. In particolare vale la seguente definizione:

Definizione 1.2 Assegnata una funzione a valori reali $f(\mathbf{x})$, $\mathbf{x} \in \mathfrak{R}^n$, allora la trasformata continua multidimensionale di Hartley è definita come segue:

$$H(\omega) = \int_{\mathfrak{R}^n} f(\mathbf{x}) \text{cas}(2\pi \mathbf{x} \cdot \omega) dx_1 \dots dx_n, \quad (1.6)$$

dove $\text{cas}(\mathcal{G}) = \cos(\mathcal{G}) + \sin(\mathcal{G})$.

In realtà, nella maggior parte delle applicazioni, all'uso della trasformata di Hartley continua, si predilige la variante discreta. Assegnata una sequenza di n valori reali, $X(t), t \in [0, n-1]$ la *trasformata di Hartley discreta* è ancora un vettore di n componenti reali dato da:

$$H(s) = \sum_{t=0}^{n-1} X(t) \text{cas}\left(\frac{2\pi st}{n}\right). \quad (1.7)$$

Infine, l'inversa della trasformata discreta di Hartley risulta simile alla $H(s)$ a meno di un fattore di normalizzazione $\frac{1}{n}$:

$$X(t) = \frac{1}{n} \sum_{s=0}^{n-1} H(s) \text{cas}\left(\frac{2\pi st}{n}\right). \quad (1.8)$$

1.1.4 La Trasformata di Laplace

La trasformata di Laplace reale e continua di una funzione $f(x)$ è definita come segue:

$$L(t) = \int_0^{\infty} f(x) e^{-tx} dx, \quad (1.9)$$

dove t , è un numero reale ed si assume che la funzione $f(x)$ soddisfi le ipotesi del teorema di esistenza.

La trasformata di Laplace discreta, ottenuta a partire dalla relativa versione continua, visto l'alto costo computazionale, non trova particolari applicazioni in contesti di elaborazione di segnali e di immagini. In realtà è possibile utilizzare un metodo che consenta di calcolare l'approssimazione numerica della trasformata di Laplace, utilizzando la ben nota trasformata di Fourier frazionata (Fractional Fourier transform - FrFT). In letteratura diversi autori giungono all'approssimazione numerica della trasformata di Fourier continua a partire dalla FrFT, lasciando al lettore la possibilità di utilizzare la medesima metodologia per ottenere una versione numerica della trasformata di Laplace continua. A tal scopo, a partire dall'analisi del suddetto metodo sviluppato da Bailey e Swartztrauber, cercheremo di definire una versione discreta della trasformata di Laplace. In particolare, sia $f(x)$ una funzione puntualmente continua, nulla fuori dall'intervallo $\left[\frac{1}{3}a, \frac{3}{2}a\right]$, in tal modo è possibile definire la trasformata di Laplace per la funzione

assegnata come in (1.9). Sia, inoltre, $\beta = \frac{a}{m}$, per $m \in \mathbb{N}_d$, l'intervallo in x degli m input di ascissa

$x_j = \left(j + \frac{m}{2}\right)\beta$ tutti centrati in a ed a cui applicare la funzione assegnata ottenendo gli m output $f(x_j)$

per $j = 0, \dots, m-1$. In tal modo, indicheremo con γ l'intervallo degli m output, le cui ascisse risulteranno

$t_k = \left(k + \frac{m}{2}\right)\gamma$, per $k = 0, \dots, m-1$.

Di conseguenza per la sequenza $f[x] = \{f(x_j)\}_{j=0, \dots, m-1}$ è possibile definire la trasformata di Fourier frazionaria, come segue:

$$G_k(f[x], \alpha) = \sum_{j=0}^{m-1} f(x_j) e^{-2\pi i j k \alpha}, \quad (1.10)$$

dove α può essere sia reale che complesso.

A questo punto, detto $\delta = \frac{\beta\gamma}{2\pi}$, e scelto $\alpha = -i\delta$, ne risulta l'approssimazione numerica della trasformata di Laplace continua:

$$\begin{aligned} L(t_k) &\approx \sum_{j=0}^{m-1} f(x_j) e^{-t_k x_j} \beta \\ &= \beta e^{-\pi \left(k + \frac{m}{2}\right) m \delta} \sum_{j=0}^{m-1} \{f(x_j) e^{-\pi \delta j m}\} e^{-2\pi i j k \alpha} \\ &= \beta e^{-\pi \left(k + \frac{m}{2}\right) m \delta} G_k(\{f(x_j) e^{-\pi \delta j m}\}, \alpha) \end{aligned} \quad (1.11)$$

dove $k = 0, \dots, m-1$.

1.1.5 La Trasformata di Hilbert

La trasformata di Hilbert è usata in vari settori dell'elaborazione dell'immagine, quali ad esempio la teoria della modulazione dei segnali. La trasformata di Hilbert, $Hf(x)$ di una funzione è il valore principale di Cauchy definito come segue:

$$\begin{aligned} Hf(x) &= \frac{1}{\pi} PV \left(\int_{-\infty}^{+\infty} \frac{f(y)}{x-y} dy \right) \\ &= \lim_{\substack{R \rightarrow \infty \\ \varepsilon \rightarrow 0}} \frac{1}{\pi} \left(\int_{-R}^{\tau-\varepsilon} + \int_{\tau+\varepsilon}^R \right) \frac{f(y)}{x-y} dy \end{aligned} \quad (1.12)$$

dove il PV è utilizzato per garantire la convergenza dell'integrale.

Assumendo che $f \in L^p(R)$, $1 < p < \infty$ la trasformata di Hilbert risulta definita puntualmente dalla (1.12) ed è un operatore limitato sullo spazio delle funzioni p -sommabili. In particolare per $f \in L^1(R)$, la trasformata di Hilbert rimane comunque definita dalla (1.12) quasi ovunque su R , ma in tal caso essa non rappresenta una mappa di $L^1(R)$ in se stesso.

La trasformata di Hilbert, può essere anche definita come la convoluzione della funzione assegnata $f(x)$ e la funzione kernel $l(x) = \frac{1}{\pi x}$, ottenendo quanto segue:

$$Hf(x) = (l * f)(x);$$

in tal modo, in base al teorema di convoluzione, ed applicando l'operatore lineare di Fourier definito in $L^1(\mathbb{R})$ risulta:

$$\widehat{Hf}(\omega) = \hat{l}(\omega) \cdot \hat{f}(\omega) = -i \operatorname{sign}(\omega) \hat{f}(\omega). \quad (1.13)$$

Allo scopo di definire la trasformata di Hilbert discreta per un generico segnale digitale $X = \{x_j\}_{j=0, \dots, N-1}$, verrà applicata alla (1.13) la trasformata di Fourier discreta ottenendo:

$$\hat{H}_k = \hat{x}_k \cdot \hat{l}_k \quad (1.14)$$

dove \hat{x}_k è la trasformata di Fourier discreta del segnale X , \hat{H}_k rappresenta la FFT della versione discreta della trasformata di Hilbert ed infine \hat{l}_k è la discretizzazione della $\hat{l}(\omega) = -i \operatorname{sign}(\omega)$. In particolare \hat{l}_k per un segnale avente un numero di campioni pari o dispari ammette rispettivamente le due seguenti definizioni:

$$\hat{l}_k = \begin{cases} -i & k = 1, 2, \dots, \left(\frac{N}{2}\right) - 1 \\ 0 & k = 0, \left(\frac{N}{2}\right) \\ i & k = \left(\frac{N}{2}\right) + 1, \left(\frac{N}{2}\right) + 2, \dots, N - 1 \end{cases} \quad (1.15)$$

$$\hat{l}_k = \begin{cases} -i & k = 1, 2, \dots, \left(\frac{N-1}{2}\right) \\ 0 & k = 0 \\ i & k = \left(\frac{N+1}{2}\right), \left(\frac{N+2}{2}\right), \dots, N - 1 \end{cases}$$

Di conseguenza, la trasformata di Hilbert discreta del segnale $X = \{x_j\}_{j=0,\dots,N-1}$ può essere definita come l'inversa della trasformata di Fourier discreta applicata alla (1.14), ottenendo:

$$H_j = \frac{1}{N} \sum_{k=0}^{N-1} \hat{I}_k \sum_{m=0}^{N-1} x_m e^{\frac{2\pi i k}{N}(j-m)}. \quad (1.16)$$

1.2 Wavelets

In letteratura è ormai comune l'analogia tra le wavelets ed i frattali, in realtà tale associazione non è per niente fuori luogo basti, infatti, osservare che i frattali sono delle figure geometriche la cui caratteristica consiste nel poterle vedere come delle piccole copie di se stesse. In particolare, un frattale non cambia aspetto se viene visto con una lente di ingrandimento (si pensi al frattale di Von Koch o all'insieme di Julia) ed allo stesso modo, immaginando di poter effettuare un zoom su una funzione wavelet, si ritroverà l'andamento grafico di una wavelet "principale". Ebbene, rimanendo in contesti di frattalità o self-similarità possiamo senz'altro affermare che la struttura delle wavelet è self-similare, esse infatti possono essere viste come una versione scalata di un'unica funzione che in letteratura viene (a volte) indicata con il nome di *wavelet madre*. Nei paragrafi a seguire cercheremo di formalizzare quanto detto introducendo la trasformata wavelet continua vista come preludio per lo studio delle basi wavelets ortogonali, biortogonali ed ortonormali.

1.1.6 La Trasformata Wavelet continua

Supponiamo di considerare lo spazio di Hilbert delle funzioni a quadrato sommabile, $L^2(\mathfrak{R}) = \{f : \|f\|_2 < \infty\}$, dove la norma $\|f\|_2 = \langle f, f \rangle^{1/2}$ è definita tramite l'usuale prodotto scalare in $L^2(\mathfrak{R})$, vale a dire $\langle f, g \rangle = \int_{\mathfrak{R}} f(x)g^*(x)dx$, con $g^*(x)$ il complesso coniugato nel caso di funzioni a valori complessi.

Il passo principale per definire la trasformata wavelet è la scelta di una funzione $\psi(x)$ (a valori reali o complessi) detta wavelet madre, la quale soddisfa la seguente "admissibility condition":

$$C_\psi = \int_0^\infty |\hat{\psi}(k)|^2 \frac{dk}{k} < \infty, \quad (1.17)$$

dove $\hat{\psi}(k) = \int_{-\infty}^{+\infty} \psi(x)e^{-i2\pi kx} dx$ denota la trasformata di Fourier della $\psi(x)$ e k rappresenta il numero d'onda.

E' possibile provare che se la wavelet madre è integrabile essa soddisfa la precedente condizione di ammissibilità, ed in particolare vale il seguente teorema.

Teorema 1.1 La funzione $\psi \in L^1(\mathfrak{R})$ soddisfa la condizione di ammissibilità se e solo se $\int \psi(x) dx = 0$.

Dimostrazione. Sia $\psi \in L^1(\mathfrak{R})$, allora $\hat{\psi}(k)$ è continua ed, inoltre, C_ψ è finito solo se $\hat{\psi}(0) = 0$, ossia se $\int \psi(x) dx = 0$.

Viceversa, se $\int \psi(x) dx = 0$, applicando una condizione di integrabilità forte a $\psi(x)$, segue che $\int (1+|x|)^\alpha |\psi(x)| dx < \infty$, per qualche $\alpha > 0$, allora $|\hat{\psi}(k)| \leq C|k|^\beta$, con $\beta = \min\{\alpha, 1\}$.

Dal teorema precedente discendono delle proprietà della wavelet madre, in particolare:

1. $\psi(x)$ è *ben localizzata*, nel senso che essa ammette un decadimento veloce per $|x|$ tendente ad infinito.
2. $\psi(x)$ è *smooth*, nel senso che la sua trasformata di Fourier ammette un decadimento "smoothly" nello spazio delle frequenze in $k = 0$.
3. $\psi(x)$ è *oscillante*, nel senso che i primi m , momenti sono nulli.

Assumiamo ad esempio, che la wavelet madre $\psi(x)$ ammetta M momenti evanescenti, vale a dire

$$\int_0^\infty x^m \psi(x) dx = 0, \quad \text{per } m=1, \dots, M-1. \quad (1.18)$$

Da cui passando allo spazio delle frequenze, ed osservando che vale la seguente eguaglianza:

$$\int_0^\infty x^m \psi(x) dx = \left. \frac{d^m \hat{\psi}(k)}{dk^m} \right|_{k=0}, \quad \text{per } m=1, \dots, M-1, \quad (1.19)$$

a questo punto mettendo insieme la (1.18) con la (1.19) segue che

$$\left. \frac{d^m \hat{\psi}(k)}{dk^m} \right|_{k=0} = 0, \quad \text{per } m=1, \dots, M-1. \quad (1.20)$$

Da cui effettivamente se la wavelet madre risulta oscillante, la relativa trasformata di Fourier ammette un decadimento smooth per $k = 0$.

A partire da $\psi(x)$ si genera una famiglia di wavelet continue dilatate e traslate rispetto alla wavelet madre. In particolare, dalla dilatazione (parametro a legato all'ampiezza del supporto della particolare wavelet) e dalla traslazione (parametro b corrispondente alla posizione della particolare wavelet) della wavelet madre ne risulterà una famiglia di wavelets dipendente da una coppia di indici così come segue:

$$\psi_{a,b}(x) = |a|^{-1/2} \psi\left(\frac{x-b}{a}\right), \quad (1.21)$$

dove $a > 0$ e $b \in \mathfrak{R}$. In genere si suppone che le wavelets siano normalizzate in L^2 , vale a dire:

$$\|\psi_{a,b}(x)\|_2 = \left(\int_{\mathfrak{R}} |\psi(x)|^2 dx \right)^{1/2} = 1.$$

L'equazione (1.21) nello spazio di Fourier diviene:

$$\hat{\psi}_{a,b}(k) = \int_{\mathfrak{R}} |a|^{-1/2} \psi\left(\frac{x-b}{a}\right) e^{-2\pi i k x} dx$$

che per le proprietà di scala (relativa alla contrazione di $1/a$ nello spazio ordinario) e traslazione (relativa alla traslazione nello spazio ordinario del termine b) della trasformata di Fourier risulta

$$\hat{\psi}_{a,b}(k) = \sqrt{a} \hat{\psi}(ak) e^{-2\pi i b k} \tag{1.22}$$

In conclusione la trasformata wavelet continua di una funzione $f \in L^2(\mathfrak{R})$ può essere definita come la convoluzione della funzione stessa con la famiglia di wavelet $\psi_{a,b}(x)$ così come segue:

$$\tilde{f}(a,b) = \int_{-\infty}^{+\infty} f(x) \psi_{a,b}^*(x) dx, \tag{1.23}$$

dove $\psi_{a,b}^*$ denota il complesso coniugato, nel caso in cui si considerino wavelets a valori complessi.

A questo punto usando l'identità di Parseval¹, otteniamo che

$$\tilde{f}(a,b) = \int_{-\infty}^{+\infty} \hat{f}(k) \hat{\psi}_{a,b}^*(k) dk. \tag{1.24}$$

Di conseguenza, la trasformata wavelet può essere interpretata come una decomposizione in frequenza della funzione, usando un filtro passabanda $\hat{\psi}_{a,b}^*$ centrato alle frequenze $k = k_{\psi}/a$; il numero d'onda k_{ψ} denota il baricentro del supporto della wavelet nello spazio di Fourier, così come segue:

¹ Ricordiamo che l'identità di Parseval ci garantisce che in condizioni di sufficiente regolarità per la funzione e per le sue derivate, l'area sottesa dal modulo al quadrato della funzione è uguale a quella sottesa dal modulo al quadrato della sua trasformata di Fourier.

$$k_{\psi} = \frac{\int_0^{+\infty} k |\hat{\psi}(k)| dk}{\int_0^{+\infty} |\hat{\psi}(k)| dk} . \quad (1.25)$$

Si noti che tali filtri hanno ampiezza variabile $\Delta k / k$, per cui aumentando il numero d'onda la banda diventa sempre più ampia e viceversa diminuendo il numero d'onda la banda si riduce.

Convien, infine, soffermarsi sulla fase di sintesi o di ricostruzione della funzione a partire dai corrispondenti coefficienti wavelets. La condizione di ammissibilità (1.17), rappresenta una condizione necessaria per poter ricostruire la funzione f a partire dai relativi coefficienti wavelets \tilde{f} ; per cui, in tali condizioni segue che

$$f(x) = \frac{1}{C_{\psi}} \int_0^{+\infty} \int_{-\infty}^{+\infty} \tilde{f}(a,b) \psi_{a,b}(x) \frac{da db}{a^2} . \quad (1.26)$$

la quale rappresenta *la trasformata wavelet inversa*. Di conseguenza l'energia totale di un segnale può essere calcolata sia nello spazio fisico che nello spazio wavelet così come segue:

$$E = \int_{-\infty}^{+\infty} |f(x)|^2 dx = \frac{1}{C_{\psi}} \int_0^{+\infty} \int_{-\infty}^{+\infty} |\tilde{f}(a,b)|^2 \frac{da db}{a^2} . \quad (1.27)$$

La (1.27) consente altresì di definire lo spettro wavelet di un generico segnale.

1.1.7 La Trasformata Wavelet discreta: Frames

Nel paragrafo precedente abbiamo definito la trasformata wavelet continua come la convoluzione di funzione a quadrato sommabile con la famiglia $\psi_{a,b}(x)$; si osservi che il termine "continuo" dipende dalla variabilità nel continuo dei valori assunti dai parametri di scala e di traslazione.

In genere, i segnali per i quali è interessante effettuare un'analisi wavelet based ammettono una larghezza della banda di frequenza finita ed una durata temporale limitata, il che consente una "riduzione" degli intervalli di variabilità dei parametri di scala e di traslazione. In particolare è possibile ottenere una famiglia di wavelets discrete campionando l'intervallo di appartenenza dei parametri a e b secondo delle opportune leggi di campionamento. Ad esempio, per il parametro di scala, si è soliti utilizzare una discretizzazione logaritmica, vale a dire ad a si sostituisce il parametro discreto $a_j = a_0^{-j}$, dove

$a_0 = \Delta(\log a)$ rappresenta la frequenza di campionamento² dell'asse $\log a$, mentre j rappresenta l'indice di scala. Per quanto riguarda l'indice di traslazione b ad esso si applicherà una discretizzazione lineare, secondo cui a b si sostituisce il parametro discreto $x_{jk} = kb_0a_0^{-j}$, dove b_0 è la frequenza di campionamento dell'asse di traslazione ed k rappresenta l'indice di traslazione. Si osservi che la frequenza di campionamento del parametro di traslazione dipende dall'indice di scala, in particolare per scale fini, ossia al diminuire di j e dunque all'aumentare di a_j , essa cresce; mentre per larghe scale, vale all'aumentare di j e dunque al diminuire di a_j , la frequenza di campionamento decresce. Di conseguenza, a partire da tali assunzioni è possibile definire le wavelets discrete $\psi_{jk}(x)$ come segue:

$$\psi_{jk}(x) = a_j^{-1/2} \psi\left(\frac{x - x_{jk}}{a_j}\right). \quad (1.28)$$

E' dunque chiaro che in tali condizioni la trasformata wavelet discreta non è altro che una versione campionata della trasformata wavelet continua infatti, dalla (1.28) discende direttamente la seguente *formula di decomposizione discreta*:

$$\tilde{f}_{jk} = \langle \psi_{jk}, f \rangle = a_j^{-1/2} \int_{-\infty}^{+\infty} f(x) \psi^*\left(\frac{x - x_{jk}}{a_j}\right) dx, \quad (1.29)$$

dove \tilde{f}_{jk} , al variare degli indici j e k in opportuni insiemi, rappresentano i coefficienti wavelets della funzione $f(x)$ rispetto alla famiglia di wavelets discrete $\psi_{jk}(x)$.

Si può semplicemente verificare che per funzioni a quadrato sommabile, scelta un'opportuna wavelet (ossia una wavelet che ammetta il medesimo decadimento sia nel tempo che in frequenza ed inoltre che soddisfi

l'ipotesi del Teorema 1.1, vale a dire $\int_{-\infty}^{+\infty} \psi(x) dx = 0$) allora esisteranno delle costanti A e B , tali che $A > 0$

e $B < \infty$ per cui i relativi coefficienti wavelets soddisfino la seguente stima:

$$A\|f\|^2 \leq \sum_{i,j} |\langle f, \psi_{jk} \rangle|^2 \leq B\|f\|^2, \quad (1.30)$$

Per ogni $f \in L^2(\mathfrak{R})$. In altre parole (cfr. Definizione 1.3) la famiglia delle wavelet discrete $\{\psi_{jk} : j, k \in \mathbb{Z}\}$ costituisce un *frame*.

² La frequenza di campionamento è la misura espressa in Hertz del numero di volte al secondo in cui un segnale *analogico* viene misurato e memorizzato in forma *digitale*. Ogni singola misurazione del segnale corrisponde ad un numero memorizzato, detto appunto *campione*.

Definizione 1.3 Una famiglia di funzioni $\{\varphi_m\}_{m \in M}$ definite in uno spazio di Hilbert H è detta *frame* se esiste $A > 0$ e $B < \infty$ tale che per ogni f in H risulti:

$$A\|f\|^2 \leq \sum_m |\langle f, \varphi_m \rangle|^2 \leq B\|f\|^2; \quad (1.31)$$

In tali condizioni le costanti A e B sono dette *frame bounds*; se i frame bounds risultano uguali tra loro, ossia se $A = B$, si parlerà di *tight frame*. In quest'ultimo caso, per ogni funzione $f \in H$ la (1.31) si riscrive come segue

$$\sum_m |\langle f, \varphi_m \rangle|^2 = A\|f\|^2;$$

che in base all'identità di polarizzazione³ diviene:

$$A \langle f, g \rangle = \sum_m \langle f, \varphi_m \rangle \langle \varphi_m, g \rangle;$$

da cui

$$f = A^{-1} \sum_m \langle f, \varphi_m \rangle \varphi_m. \quad (1.32)$$

Come vedremo a breve la (1.32) ricorda la formula di ricostruzione discreta di una funzione rispetto ad una base ortonormale, ma è importante osservare che in generale i frames (siano essi bounds oppure tight) non costituiscono una base ortonormale per lo spazio di Hilbert a cui appartengono. In realtà, è possibile provare che in determinate condizioni (cfr.

Proposizione 1.1) dei particolari tight frames rappresentano una base ortonormale.

Proposizione 1.1 Sia $\{\varphi_m\}_{m \in M}$ un tight frame con frame bound e norma unitaria, vale a dire $A = 1$ e $\|\varphi_m\| = 1$ per ogni m , allora $\{\varphi_m\}_{m \in M}$ costituisce una base ortonormale.

In generale, la *formula di ricostruzione discreta* per una funzione a quadrato sommabile rispetto ad una famiglia di wavelet discrete $\{\psi_{jk} : j, k \in \mathbb{Z}\}$ risulta essere la seguente:

$$f(x) = C \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \tilde{f}_{jk} \psi_{jk}(x) + R(x), \quad (1.33)$$

dove C è una costante ed $R(x)$ è un "residuo" della ricostruzione della funzione attraverso la famiglia di wavelet stessa. Di conseguenza sia la costante C sia la funzione $R(x)$ dipenderanno dalla scelta della

³ Identità di polarizzazione: $\langle f, g \rangle = \frac{1}{4} [\|f + g\|^2 - \|f - g\|^2 + i\|f + ig\|^2 - i\|f - ig\|^2]$.

particolare wavelet e dalla frequenza di campionamento dell'asse di scala e di traslazione. In particolare, supponendo di scegliere $a_0 = 2$ e $b_0 = 1$, otterremo un campionamento diadico; ebbene in tal caso ne risulteranno delle particolari wavelets $\psi_{jk} = 2^{j/2} \psi(2^j x - k)$ che costituiranno al variare degli indici i e j una base ortonormale per lo spazio $L^2(\mathfrak{R})$, vale a dire:

$$\langle \psi_{jk}, \psi_{mn} \rangle = \delta_{jm} \delta_{kn},$$

dove δ denota il simbolo di Kronecker.

Questo equivale a dire che assegnato un campionamento diadico le wavelets risultano ortogonali alle relative versioni traslate di $2^{-j}k$ e dilatate di 2^{-j} ; inoltre, in tal caso la formula di ricostruzione **(1.33)** è esatta, nel senso che la costante C è unitaria ed il residuo $R(x)$ è pari a zero.

1.1.8 Wavelets ortogonali e l'analisi multirisoluzione

Fondamentale per la costruzione di basi wavelet ortogonali è l'introduzione dell'analisi multirisoluzione (MRA), la quale fu introdotta da S.Mallat nel 1988. Come chiariremo meglio a breve, la formulazione della MRA rivoluzionò il settore dell'immagine analysis poiché essa permetteva lo studio simultaneo di un'immagine (vale a dire di un segnale bidimensionale) a differenti scale, ossia a differenti frequenze; inoltre la conseguente comparazione dei risultati ottenuti dagli studi alle diverse scale consentiva una analisi più o meno dettagliata della tramatura dell'immagine stessa evidenziandone o trascurandone dei particolari.

Generalizzando al caso di una funzione a quadrato sommabile $f \in L^2(\mathfrak{R})$ l'idea, alla base della MRA, consiste nel considerare delle approssimazioni f_j della f a differenti scale j ; e così l'analisi wavelet ortogonale può essere interpretata come la decomposizione della funzione in corrispondenti approssimazioni ottenute proiettando la funzione stessa su particolari spazi wavelet via via sempre più "fini" al diminuire del parametro di scala j .

L'analisi multirisoluzione ortogonale e monodimensionale di $L^2(\mathfrak{R})$ è definita come una sequenza di spazi di approssimazione V_j dipendenti dal parametro di scala $j \in \mathbb{Z}$. Più precisamente essi rappresentano dei sottospazi chiusi di $L^2(\mathfrak{R})$ verificanti le seguenti proprietà:

$$1. \quad V_j \subset V_{j+1}, \quad j \in \mathbb{Z}; \tag{1.34}$$

$$2. \quad \bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathfrak{R}); \tag{1.35}$$

$$3. \quad \bigcap_{j \in \mathbb{Z}} V_j = \{0\}; \tag{1.36}$$

$$4. \quad f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}. \tag{1.37}$$

Inoltre, si è soliti assumere l'esistenza di una funzione $\phi(x)$, detta *funzione di scala*, le cui traslate rispetto all'indice k definiscono una base ortonormale per ogni spazio V_j :

$$V_j = \overline{\text{span}} \{ \phi_{jk}, k \in \mathbb{Z} \}, \quad (1.38)$$

dove

$$\phi_{jk} = 2^{j/2} \phi(2^j x - k), \quad j, k \in \mathbb{Z}. \quad (1.39)$$

In particolare, ad ogni valore dell'indice di scala j , la base definita in (1.39) risulta ortonormale rispetto alle corrispondenti traslate di uno step pari a $k/2^j$:

$$\langle \phi_{jk}, \phi_{lm} \rangle = \delta_{km}. \quad (1.40)$$

A questo punto, a partire dall'analisi delle proprietà caratteristiche della MRA di $L^2(\mathfrak{R})$ (cfr. (1.34)-(1.37)) cercheremo di formalizzare quanto descritto in precedenza in relazione alla possibilità di applicare ad una funzione assegnata una decomposizione wavelet ortogonale, in modo tale che le informazioni necessarie per la ricostruzione della funzione stessa siano distribuite in vari sottospazi di $L^2(\mathfrak{R})$ ottenuti al variare del parametro di scala j .

In particolare, a partire dalla equazione (1.37) segue che la funzione di scala $\phi(x)$ soddisfa la ben nota *equazione di raffinamento*:

$$\phi_{j-1,k}(x) = \sum_{n=-\infty}^{+\infty} h_{n-2k} \phi_{j,n}(x) \quad (1.41)$$

dove $h_n = \langle \phi_{j,n}(x), \phi_{j-1,0}(x) \rangle$ rappresentano i coefficienti di scala.

Di conseguenza, analizzando l'equazione di raffinamento e generalizzandone l'applicazione ad una qualunque funzione di $L^2(\mathfrak{R})$ si conclude che l'approssimazione della medesima funzione alle scale meno fini può essere espressa come combinazione lineare delle traslate di un'ulteriore sua approssimazione ottenuta per scale più fini.

Come premesso, le approssimazioni delle funzioni a differenti scale, rappresentano le proiezioni della funzione stessa sui corrispondenti sottospazi chiusi di $L^2(\mathfrak{R})$; si definisce, infatti, la seguente proiezione ortogonale di f su V_j :

$$P_j : f \in L^2(\mathfrak{R}) \rightarrow f_j \in V_j \quad (1.42)$$

da cui

$$f_j(x) = \sum_{k \in \mathbb{Z}} \langle f, \phi_{j,k} \rangle \phi_{j,k}(x) \quad (1.43)$$

Di conseguenza, la (1.43) consente di esprimere l'approssimazione della funzione al J -simo livello attraverso un filtraggio tra la funzione stessa e quella di scala $\phi(x)$; la quale, alla stregua di un comune filtro,

$$\int_{-\infty}^{+\infty} \phi(x) dx = 1$$

ammette norma unitaria, vale a dire

A questo punto sembra naturale aspettarsi, sempre nel contesto dell'analisi multirisoluzione di $L^2(\mathfrak{R})$, l'esistenza di un'ulteriore sequenza di sottospazi chiusi di $L^2(\mathfrak{R})$ stavolta aventi come basi ortonormali le traslate di una particolare wavelet, detta wavelet *madre*.

A tal proposito, infatti, a partire dalla considerazione che lo spazio V_{J-1} è incluso in V_J è possibile definire il complemento ortogonale di V_{J-1} in V_J , come segue:

$$V_J = V_{J-1} \oplus W_{J-1}; \tag{1.44}$$

di conseguenza l'approssimazione della funzione f alla scala 2^{-J} definita in V_J , può essere decomposta nella somma delle corrispondenti proiezioni ottenute rispettivamente sugli spazi V_{J-1} e W_{J-1} :

$$P_J(f) = P_{J-1}(f) + Q_{J-1}(f), \tag{1.45}$$

dove $Q_{J-1}(f)$, detto dettaglio della funzione alla scala $2^{-(J+1)}$, rappresenta la proiezione ortogonale di f su W_{J-1} ; $Q_{J-1}: f \in L^2(\mathfrak{R}) \rightarrow Q_{J-1}(f) \in W_{J-1}$.

In tal modo l'approssimazione della funzione alla scala 2^{-J} sarà descritta dall'approssimazione della medesima funzione ad una scala più fine aggiungendovi delle informazioni di *dettaglio* contenuti nella corrispondente proiezione della funzione stessa sullo spazio W_{J-1} .

In tali condizioni si assume l'esistenza di una funzione detta wavelet madre definita a partire dalla funzione di scala così come indicato di seguito:

$$\psi_{jk}(x) = \sum_{n \in \mathbb{Z}} g_{n-2k} \phi_{jn}(x), \tag{1.46}$$

dove $g_n = \langle \phi_{jn}, \psi_{j-1,0} \rangle$ rappresentano i coefficienti wavelet, mentre $\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$ con $j, k \in \mathbb{Z}$ rappresentano le traslate e dilatate della funzione wavelet madre ottenute rispettivamente al variare degli indici di traslazione e di scala in \mathbb{Z} .

I coefficienti wavelet possono essere ottenuti a partire da quelli di scala tramite la seguente eguaglianza:

$$g_n = (-1)^{1-n} h_{1-n}. \quad (1.47)$$

Infine, come premesso, le wavelets $\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$ definiscono una base ortonormale per lo spazio W_j

$$W_j = \overline{\text{span}}\{\psi_{jk}\}_{k \in \mathbb{Z}}. \quad (1.48)$$

A questo punto, allo scopo di comprendere l'importanza delle informazioni codificate attraverso le wavelet ortogonali utili ai fini della ricostruzione di una funzione $f \in L^2(\mathbb{R})$, consideriamone delle approssimazioni a due differenti scale:

$$\begin{aligned} f_j(x) &= \sum_{k=-\infty}^{+\infty} \bar{f}_{jk} \phi_{jk}(x), \\ f_{j-1}(x) &= \sum_{k=-\infty}^{+\infty} \bar{f}_{j-1k} \phi_{j-1k}(x), \end{aligned} \quad (1.49)$$

dove i coefficienti di scala $\bar{f}_{jk} = \langle f, \phi_{jk} \rangle$ rappresentano i valori medi locali della funzione f alla posizione $k2^{-j}$ ed alla scala 2^{-j} .

Ebbene la differenza delle approssimazioni ai due livelli di scala è definita tramite wavelet ortogonali come segue:

$$f_j(x) - f_{j-1}(x) = \sum_{k=-\infty}^{+\infty} \tilde{f}_{j-1,k} \psi_{j-1,k}(x), \quad (1.50)$$

dove i coefficienti wavelet $\tilde{f}_{jk} = \langle f, \psi_{jk} \rangle$ corrispondono alle differenze locali della funzione alla posizione $(2k+1)2^{-(j+1)}$ tra le approssimazioni alle scale 2^{-j} e $2^{-(j+1)}$.

Iterando questo ragionamento, dalla (1.50) segue che ogni funzione $f \in L^2(\mathbb{R})$ può essere espressa come la somma della approssimazione della funzione stessa ottenuta rispetto una scala di riferimento, fissata in j_0 e delle successive differenze o dettagli calcolati in relazione ai precedenti livelli di scala, così come descritto dalla seguente relazione:

$$f(x) = \sum_{k=-\infty}^{+\infty} \bar{f}_{0,k} \phi_{0,k}(x) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{+\infty} \tilde{f}_{j,k} \psi_{j,k}(x), \quad (1.51)$$

la quale rappresenta la formula di decomposizione ortogonale della funzione assegnata rispetto alla base wavelet $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$.

1.1.9 MultiWavelets

La teoria delle Multiwavelets rappresenta una naturale generalizzazione, al caso vettoriale, di quella relativa alle classiche funzioni wavelet. Sviluppata negli ultimi anni, essa è di straordinario interesse applicativo in contesti in cui sia fondamentale l'utilizzo di dati derivanti da metodi di pre e post-filtraggio di segnali bidimensionali; quali, ad esempio, la codifica, la riduzione del rumore (tecniche di denoising) e la compressione delle immagini.

La nascita di tale teoria è attribuibile a Geronimo, Hardin, e Massopust, i quali costruirono le funzioni di multiscala e le corrispondenti multiwavelet a partire da funzioni di interpolazione frattali. In particolare, in letteratura gli autori mostrano come, a partire dalle funzioni di multiscala, si pervenga alla formalizzazione di un'opportuna analisi multirisoluzione per segnali ad energia finita (o in generale per funzioni reali a quadrato sommabile); attraverso tale analisi verranno successivamente identificate le multiwavelets, le quali dovranno soddisfare alcune considerevoli proprietà, quali l'ortogonalità, la simmetria, la compattezza del minimo supporto e l'alto numero di vanishing moments, caratteristiche non comuni nel caso scalare.

Analizziamo nei dettagli l'analisi multirisoluzione di molteplicità r e le funzioni multiscala, cercando, di volta in volta, di mettere in luce le differenze sostanziali e le analogie esistenti tra le multiwavelets e le wavelet.

La teoria delle multiwavelets reali è connessa con la teoria delle funzioni reali a valori vettoriali, vale a dire delle mappe da \mathfrak{R} in \mathfrak{R}^r , di conseguenza $\forall x \in \mathfrak{R}$, $f(x)$ è un vettore colonna di r componenti reali $[f_1(x), f_2(x), \dots, f_r(x)]^T$, dove f_1, f_2, \dots, f_r sono funzioni reali a variabile reale.

A partire da tali assunzioni, indicheremo con $L^2(\mathfrak{R})^r$, $l^2(\mathbb{Z})^r$, $l^2(\mathbb{Z})^{r \times r}$ i seguenti spazi funzionali:

$$L^2(\mathfrak{R})^r = \{f = [f_1, \dots, f_r]^T : f_i \in L^2(\mathfrak{R}), i = 1, \dots, r\} \quad (1.52)$$

$$l^2(\mathbb{Z})^r = \{c = \{c_k\}_{k \in \mathbb{Z}} = [c_1, \dots, c_r]^T : c_i = \{c_i^k\}_{k \in \mathbb{Z}} \in l^2(\mathbb{Z}), i = 1, \dots, r\} \quad (1.53)$$

$$l^2(\mathbb{Z})^{r \times r} = \{C = \{C_k\}_{k \in \mathbb{Z}} = [C_{i,j}]_{i,j \in \{1, \dots, r\}} : C_{i,j} = \{C_{i,j}^k\}_{k \in \mathbb{Z}} \in l^2(\mathbb{Z}), i, j = 1, \dots, r\} \quad (1.54)$$

Ognuno di essi sarà uno spazio di Banach con le rispettive norme:

$$\|f\|_{L^2(\mathfrak{R})^r}^2 = \sum_{i=1}^r \|f_i\|_{L^2(\mathfrak{R})}^2 \quad (1.55)$$

$$\|c\|_{l^2(\mathbb{Z})^r}^2 = \sum_{i=1}^r \|c_i\|_{l^2(\mathbb{Z})}^2 = \sum_{k \in \mathbb{Z}} \|c_k\|_r^2, \quad (1.56)$$

$$\|C\|_{l^2(\mathbb{Z})^{r \times r}}^2 = \sum_{i,j=1}^r \|C_{i,j}\|_{l^2(\mathbb{Z})}^2 = \sum_{i,j=1}^r \|C_{i,j}\|_{r \times r}^2, \quad (1.57)$$

dove $\|\cdot\|_r$ e $\|\cdot\|_{r \times r}$ rappresentano le norme su \mathfrak{R}^r e sullo spazio matriciale $M^{r \times r}$, definite a partire da una naturale estensione del prodotto scalare in $L^2(\mathfrak{R})^r$, vale a dire

$$\begin{aligned} \langle f, g \rangle &= \int_{\mathfrak{R}} f(x) g^T(x) dx \\ &= [\langle f_i, g_j \rangle]_{i,j \in \{1, \dots, r\}}. \end{aligned} \quad (1.58)$$

Ebbene, come nel caso scalare le multiwavelets sono costruite a partire da funzioni di scala, che per analogia al caso in questione chiameremo funzioni di multiscala. Tali funzioni si definiscono, così come avviene nel caso scalare, a partire dalla corrispondente analisi multirisoluzione, ottenuta generalizzando al caso vettoriale quella analizzata nel paragrafo 1.2.3 e relativa al caso scalare. In particolare indicheremo con

$$V_j = \overline{\text{span}} \{ 2^{j/2} \phi_i(2^{j/2} x - k) : 1 \leq i \leq r, k \in \mathbb{Z} \}, \quad \forall j \in \mathbb{Z}, \quad (1.59)$$

gli analoghi degli spazi definiti in (1.38), e su cui costruiremo la MRA vettoriale. Si osservi che in tal caso i sottospazi sono definiti non solo al variare del parametro di traslazione, ma anche al variare dell'indice i nell'insieme $\{1, \dots, r\}$, da cui seguirà la scelta della i -sima funzione di scala.

Di conseguenza segue la seguente definizione:

Definizione 1.4 : Sia $\Phi = [\phi_1, \dots, \phi_r]^T$, una funzione a valori vettoriale definita in $L^2(\mathfrak{R})^r$, dove per ogni i le $\{\phi_i\}_{i=1, \dots, r}$ rappresentano rispettivamente delle funzioni di scala per $L^2(\mathfrak{R})$, allora la Φ è detta funzione di multiscala se gli spazi definiti dalla (1.59) per ogni j appartenente a \mathbb{Z} , soddisfano le analoghe delle (1.34)-(1.37).

In tal modo, la funzione di multiscala Φ genera un'analisi multirisoluzione di molteplicità r per lo spazio $L^2(\mathfrak{R})$, ed in particolare, poiché $\phi_1(\cdot), \dots, \phi_r(\cdot) \in V_0 \subset V_1$ esisterà una sequenza di matrici $\{\mathbf{P}_n\}_{n \in \mathbb{Z}} \in l^2(\mathbb{Z})^{r \times r}$ tale che

$$\Phi(x) = \sum_{n \in \mathbb{Z}} \mathbf{P}_n \Phi(2x - n), \quad (1.60)$$

la quale esplicitata diviene

$$\begin{bmatrix} \phi_1(x) \\ \phi_2(x) \\ \vdots \\ \phi_r(x) \end{bmatrix} = \sum_{n \in \mathbb{Z}} \begin{bmatrix} P_{11}^n & P_{12}^n & \dots & P_{1r}^n \\ P_{21}^n & \dots & \dots & P_{2r}^n \\ \vdots & & & \vdots \\ P_{r1}^n & \dots & \dots & P_{rr}^n \end{bmatrix} \begin{bmatrix} \phi_1(2x - n) \\ \phi_2(2x - n) \\ \vdots \\ \phi_r(2x - n) \end{bmatrix} \quad (1.61)$$

La (1.60) è detta relazione matriciale bi-scala, o anche *equazione di raffinamento matriciale* della funzione di multiscala Φ . In realtà, introducendo le matrici a blocchi è possibile dare una forma più compatta della (1.60).

La (1.60) è dunque l'estensione al caso vettoriale della (1.41) per $j = 1$ e $k = 0$; dunque per ogni n la matrice \mathbf{P}_n sarà in qualche modo comparabile ai coefficienti di scala $h_n = \langle \phi_{j,n}(x), \phi_{j-1,0}(x) \rangle$ definiti nel caso scalare. Fondamentale è osservare che trattandosi in tal caso non più di coefficienti, ma di matrici di scala il numero dei gradi di libertà è maggiore e di conseguenza è possibile imporre alle funzioni di multiscala diverse condizioni o richiedere che esse soddisfino contemporaneamente differenti proprietà, quali ad esempio la simmetria, l'ortogonalità, la compattezza del supporto o infine l'elevato numero di vanishing moments. E' dunque questa coesistenza di differenti aspetti, fondamentali ai fini applicativi, il vantaggio di utilizzare le multiwavelets piuttosto che le wavelets scalari.

A questo punto, affrontiamo il tema della introduzione delle multiwavelets a partire dalle funzioni di multiscala. Assegnata una MRA di molteplicità r tramite gli spazi $\{V_j\}_{j \in \mathbb{Z}}$, è possibile definire il complemento ortogonale W_j di V_j rispetto a V_{j+1} , in modo tale che $V_{j+1} = V_j \oplus W_j, \forall j \in \mathbb{Z}$. Da cui segue direttamente la seguente definizione.

Definizione 1.4: Sia $\Psi = [\psi_1, \dots, \psi_r]^T \in L^2(\mathbb{R})^r$ una funzione a valori vettoriali⁴ tale che

$$W_j = \overline{\text{span}} \{2^{j/2} \psi_i(2^{j/2}x - k) : 1 \leq i \leq r, k \in \mathbb{Z}\}, \quad \forall j \in \mathbb{Z}, \quad (1.62)$$

allora essa è detta una multiwavelets semi-ortogonale se, per ogni j appartenente a \mathbb{Z} , sono soddisfatte le seguenti condizioni:

- i. $V_j \perp W_j$;
- ii. la famiglia $\{2^{j/2} \psi_i(2^{j/2}x - k) : 1 \leq i \leq r, k \in \mathbb{Z}\}$ è una base 2-stabile per W_j , o equivalentemente che W_j è l'insieme di tutte le combinazioni lineari di tali funzioni con coefficienti in $l^2(\mathbb{Z})$.

Ebbene dalla condizione i. segue che $W_j \perp W_k, j \neq k$ da cui le multiwavelets soddisfano la seguente condizione di ortogonalità:

$$\langle \Psi(2^j \cdot -k), \Psi(2^l \cdot -n) \rangle = 0, \quad j \neq l, j, k, l, n \in \mathbb{Z}, \quad (1.63)$$

Inoltre, si osservi che anche le multiwavelets soddisfano l'equazione di raffinamento matriciale o relazione bi-scala (two-scale relation):

$$\Psi(x) = \sum_{k \in \mathbb{Z}} \mathbf{Q}_k \Phi(2x - k), \quad (1.64)$$

dove la matrice \mathbf{Q}_k è una matrice $r \times r$ equiparabile ai coefficienti wavelets del caso scalare.

Si osservi, infine, che, come nel caso scalare, è possibile, a partire dalle proprietà della MRA di molteplicità r , definire un algoritmo, computazionalmente efficiente, che consenta il calcolo dei coefficienti di multiscala, vale a dire, $\mathbf{c}_k^{-j} = [c_1^{k,-j}, c_2^{k,-j}, \dots, c_r^{k,-j}]^T$ e $\mathbf{d}_k^{-j} = [d_1^{k,-j}, d_2^{k,-j}, \dots, d_r^{k,-j}]^T$,

⁴ Anche stavolta le $\{\psi_i\}_{i=1, \dots, r}$ sono delle wavelets scalari.

corrispondenti alle proiezioni della generica funzione $f \in L^2(\mathfrak{R})$ sugli spazi V_{-j} e W_{-j} rispettivamente. Ebbene tali proiezioni, $f_{-j} \in V_{-j}$ e $g_{-j} \in W_{-j}$, saranno esprimibili come combinazioni lineari delle funzioni di multiscala le prime e delle mutiwavelets le seconde, così come mostrato:

$$\begin{aligned} f_{-j}(x) &= \sum_{k \in \mathbb{Z}} (\mathbf{c}_k^{-j})^T \Phi(2^{-j}x - k); \\ g_{-j}(x) &= \sum_{k \in \mathbb{Z}} (\mathbf{d}_k^{-j})^T \Psi(2^{-j}x - k). \end{aligned} \quad (1.65)$$

Pertanto, dalla sequenza dei coefficienti di multiscala $\{\mathbf{c}_k^0\}$ della funzione $f = f_0 \in V_0$, è possibile calcolare ricorsivamente i coefficienti multiwavelets utilizzando le seguenti formule di decomposizione:

$$\begin{aligned} \mathbf{c}_k^{-j} &= \sum_{m \in \mathbb{Z}} \mathbf{P}_{m-2k} \mathbf{c}_m^{-j+1}; \\ \mathbf{d}_k^{-j} &= \sum_{m \in \mathbb{Z}} \mathbf{Q}_{m-2k} \mathbf{d}_m^{-j+1}, \end{aligned} \quad j = 1, \dots, L; \quad k \in \mathbb{Z}. \quad (1.66)$$

1.1.10 Dizionari Tempo-Frequenza e Wavelet Packets

Come più volte detto nei paragrafi precedenti, la decomposizione di un segnale attraverso una famiglia di funzioni ben localizzate nel tempo ed in frequenza, trova diverse applicazioni nel campo dell'elaborazione dei segnali ed in generale in analisi armonica.

Dalla scelta delle funzioni (note in letteratura come *atomi tempo-frequenza*), attraverso le quali decomporre i segnali, discende la possibilità di ricavare determinate proprietà intrinseche al segnale stesso ed utili ai fini della elaborazione. La trasformata di Fourier a finestra (Window Fourier Transforms - WFT) e la trasformata wavelet, entrambe analizzate nei paragrafi precedenti, rappresentano degli esempi fondamentali di decomposizione dei segnali in tempo-frequenza. Ambedue, infatti, sono atomi tempo-frequenza che descrivono basi o frames per lo spazio $L^2(\mathfrak{R})$.

Una generica famiglia di atomi tempo-frequenza può essere generata per effetto di scala, di traslazione e di modulazione di una singola funzione finestra $g(t) \in L^2(\mathfrak{R})$, che per semplicità supporremo reale, differenziabile ed $O(1/(t^2 + 1))^5$; di conseguenza, imporremo che $\|g(t)\| = 1$, e che essa sia non nulla in zero, ossia $g(0) \neq 0$.

Definiamo, dunque, per ogni valore del parametro di scala s , di modulazione in frequenza ξ ed infine di traslazione u la seguente funzione finestra:

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) e^{i\xi t}, \quad (1.67)$$

⁵ Ci si riferisce al simbolo di Landau secondo il quale la $g(t)$ è un infinitesimo di ordine superiore rispetto a $1/(t^2 + 1)$.

dove $\gamma = (s, u, \xi) \in \Gamma = \mathfrak{R}^+ \times \mathfrak{R}^2$. A partire dalla (1.67) è possibile definire la trasformata di Fourier di $g(t)$ come segue:

$$\hat{g}_\gamma(\omega) = \sqrt{s} \hat{g}(\omega - \xi) e^{-i(\omega - \xi)u}, \quad (1.68)$$

ed essendo $|\hat{g}(\omega)|$ una funzione pari, la $|\hat{g}_\gamma(\omega)|$ sarà centrata alla frequenza $\omega = \xi$.

Dunque, essendo la famiglia $D = (g_\gamma(t))_{\gamma \in \Gamma}$ estremamente ridondante, allo scopo di descrivere efficacemente una funzione $f(t)$ è possibile selezionare un appropriato sottoinsieme numerabile di atomi $D_n = (g_{\gamma_n}(t))_{\gamma_n \in N}$ con $\gamma_n = (s_n, u_n, \xi_n)$, per il quale valga la seguente decomposizione della funzione assegnata:

$$f(t) = \sum_{n=-\infty}^{+\infty} a_n g_{\gamma_n}(t) \quad (1.69)$$

I coefficienti di espansione a_n , dipendenti dalla particolare scelta degli atomi tempo-frequenza $g_{\gamma_n}(t)$, forniscono particolari informazioni relative al contenuto frequenziale della funzione in esame. Ad esempio, nel caso della trasformata di Fourier a finestra, tutti gli atomi $g_{\gamma_n}(t)$ sono caratterizzati da un parametro di scala costante $s_n = s_0$ (da cui dipende l'ampiezza della finestra); di conseguenza, essi saranno particolarmente utili ad analizzare segnali che ammettano un andamento, ben localizzato nello spazio tempo-frequenza, dell'ordine di $s_n = s_0$. In tal caso, infatti, gli atomi a scala fissa riusciranno ad analizzare tutte le componenti del segnale nello spazio tempo-frequenza ed i coefficienti di espansione a_n forniranno importanti informazioni relative al contenuto frequenziale. Viceversa, per segnali che non rispettino tali requisiti, ossia che abbiano componenti a dimensione frequenziale variabile, l'applicazione degli atomi a scala fissa comporterebbe un'estrema perdita di informazioni poiché verrebbero essenzialmente analizzate le componenti frequenziali ben localizzate, rispetto al parametro di scala fissato, tralasciando tutte le altre. In tal caso risulteranno particolarmente adatti gli atomi tempo-frequenza a scala variabile, vale a dire le wavelets. Esse infatti rappresentano una famiglia di atomi $(g_{\gamma_n}(t))_{n \in N}$ per cui il parametro di scala è correlato alla frequenza tramite la seguente eguaglianza:

$$\xi_n = \frac{\xi_0}{s_n},$$

dove ξ_0 rappresenta una costante. I coefficienti di espansione a_n , che in tal caso corrispondono ai coefficienti wavelets, caratterizzano il comportamento di scala del segnale analizzato, fornendo utili informazioni relative al più complesso contenuto frequenziale. In particolare a valori piccoli del parametro di scala corrisponderanno atomi atti ad analizzare le componenti del segnale alle alte frequenze e viceversa per valori maggiori di s_n quelle alle basse frequenze.

In generale per segnali $f(t)$ a struttura di scala estremamente oscillante è possibile, utilizzando particolari algoritmi di selezione (quali ad esempio l'algoritmo di Matching Pursuit), individuare a priori da un set ridondante di atomi appartenenti ad un dizionario D quelli che meglio si adattano ad analizzare il segnale in questione.

Un ulteriore esempio di atomi tempo-frequenza è rappresentato dalle Wavelet Packet. I dizionari Wavelet Packet sono un insieme ridondante di funzioni con buone proprietà di ortogonalità, a partire dal quale è possibile ottenere un sottoinsieme di basi ortonormali, composti da vettori ben localizzati nello spazio tempo-frequenza.

Come di consueto, la definizione rigorosa delle basi Wavelet Packet monodimensionali, si ottiene a partire dall'analisi multirisoluzione per $L^2(\mathfrak{R})$.

Definizione 1.5 Sia $\Psi_0(x) = \phi(x)$ e $\Psi_1(x) = \psi(x)$, è possibile definire il pacchetto wavelet, associato alla funzione base $\Psi_0(x)$ come segue:

$$\begin{aligned}\Psi_{2n}(x) &= \sqrt{2} \sum_k l_k \Psi_n(2x - k), \\ \Psi_{2n+1}(x) &= \sqrt{2} \sum_k h_k \Psi_n(2x - k).\end{aligned}\tag{1.70}$$

Di conseguenza i dizionari Wavelet Packet saranno ottenuti a partire dalle $\Psi_n(x)$, come una collezione di funzioni definite tramite la seguente uguaglianza:

$$\Psi_{jnk}(x) = 2^{j/2} \Psi_n(2^j x - k),\tag{1.71}$$

dove $j, k, n \in \mathbb{Z}$.

In particolare, come descritto dettagliatamente in [20] è possibile definire tra tutte le funzioni del tipo $\{\Psi_{jnk}(x)\}_{j,n,k \in \mathbb{Z}}$ una base ortonormale $\{\Psi_{jnk} : (j, n, k) \in I\}$ ottenuta al variare della terna di indici (j, n, k) in un opportuno sottoinsieme $I \subset \mathbb{Z}^3$, per cui gli intervalli diadici $\{[2^{-j}n, 2^{-j}(n+1)]: (j, n, k) \in I\}$, formino un ricoprimento disgiunto per l'intervallo $[0,1]$.

A differenza delle classiche basi wavelets, una funzione appartenente ad un generico dizionario wavelets packet è caratterizzata da tre indici: l'indice di scala j , l'indice di traslazione o localizzazione k , ed infine l'indice di oscillazione o numero d'onda n . Di conseguenza la funzione $\Psi_{jnk}(x) = 2^{j/2} \Psi_n(2^j x - k)$, sarà centrata in $2^{-j}k$, avrà un supporto di dimensione 2^{-j} ed infine oscillerà n volte nell'unità di tempo. Si osservi, inoltre, che in base al principio di indeterminazione di Heisenberg, non è possibile localizzare con una buona precisione una funzione sia rispetto al parametro di localizzazione k sia rispetto a quello di oscillazione n , infatti da tale principio discende che

$$\Delta k \Delta n \geq 1,$$

dove Δk e Δn rappresentano rispettivamente gli errori in termini di posizione e numero d'onda. Ebbene nel caso dei dizionari wavelet packet, risulta che, in una medesima normalizzazione, le incertezze in posizione e numero d'onda dipendono dal fattore di scala, secondo le seguenti relazioni:

$$\Delta k \approx 2^{-j}, \quad \Delta n \approx 2^j,$$

di conseguenza essendo il prodotto delle indeterminazioni molto piccolo sarà possibile, con buona approssimazione, localizzare una funzione di un generico dizionario wavelet packet sia in posizione che in numero d'onda.

Supposta la famiglia $\{\Psi_{jnk} : (j, n, k) \in I\}$ una base ortonormale per un generico spazio funzionale, ed assegnata una funzione f appartenente a tale spazio, sarà possibile esprimere tale funzione attraverso la base Wavelet Packet come segue:

$$f = \sum_{(j,n,k) \in I} Wp_{jnk} \Psi_{jnk}, \tag{1.72}$$

dove $Wp_{jnk} = \langle \Psi_{jnk}, f \rangle$ rappresentano i coefficienti Wavelet Packet.

Di conseguenza, ogni funzione f appartenente allo spazio funzionale, per il quale le Wavelet Packet siano una base ortonormale, potrà essere espressa come una sovrapposizione di "blocchi elementari" o in termini analitici una combinazione lineare di funzioni base Ψ_{jnk} rispetto alle coordinate k ed n .

In particolare, attraverso una base Wavelet Packet è possibile decomporre la funzione o il segnale di partenza in sottosegnali; tale scomposizione può essere schematizzata attraverso un grafo ad albero in cui i livelli di decomposizione corrispondono ai differenti canali frequenziali rispetto ai quali si differenzia il segnale. L'algoritmo di decomposizione Wavelet Packet consente la scelta del canale frequenziale o del ramo dell'albero, lungo cui decomporre il segnale definendo un opportuno sottoalbero del grafo originario. In particolare, è noto che nel caso delle wavelet la decomposizione avviene sempre lungo il ramo delle basse frequenze, poiché si presuppone che il maggiore contenuto energetico del segnale sia in esso contenuto; nel caso, invece, delle Wavelet Packet la decomposizione avviene in modo adattivo decomponendo il segnale lì dove sono effettivamente localizzate le maggiori informazioni, o in generale, nella regione in cui un particolare funzionale, associato al segnale, risulti minimo o massimo. In particolare, in Figura 1 sono illustrati due differenti alberi di decomposizione Wavelet Packet: quello a sinistra indica una classica decomposizione binaria del segnale, in cui esso sarà suddiviso indistintamente in entrambi i canali frequenziali; nella parte destra dell'immagine invece sono rappresentati due differenti esempi di grafo ad albero Wavelet Packet adattivi in cui la decomposizione è arbitraria e dipendente dal particolare segnale a cui gli atomi Wavelet Packet sono applicati.

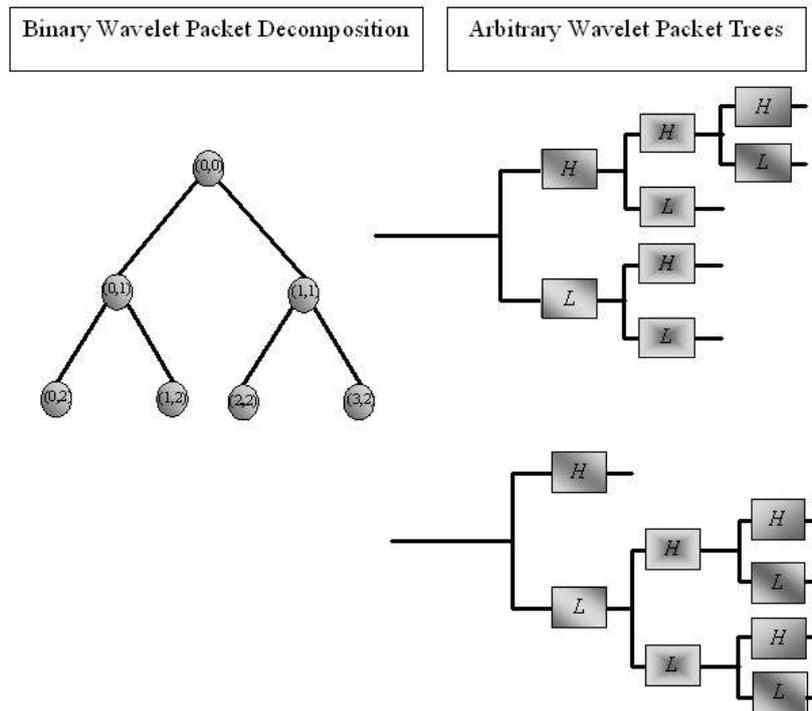


Figura 1: Decomposizioni Wavelet Packet.

In ognuno di queste strutture ad albero, ogni nodo definisce il livello di decomposizione indicizzato tramite il fattore di scala j ed a cui è possibile associare i sottospazi V_j o W_j rispetto ai quali proiettare il segnale a seconda che si decomponga secondo il canale ad alte (come nel grafo in alto a destra) o basse frequenza (grafo in basso a destra).

Tutte le considerazioni fatte per i dizionari Wavelet Packet nel caso monodimensionale trovano una naturale generalizzazione al caso bidimensionale, in cui le basi Wavelet Packet sono definite tramite il prodotto tensoriale tra le corrispondenti wavelet packet monodimensionali. Di conseguenza, assegnato uno spazio delle immagini, quale ad esempio $L^2(\mathbb{R}^2)$, esso può essere decomposto in un insieme \mathbf{W} di sottospazi $\{V_i\}_{i \in I}$, che corrisponderanno alle sottoimmagini a differenti frequenze. In tal modo da un albero \mathbf{W} sarà costituito da un insieme ridondante di sottospazi, o meglio da una famiglia di sottospazi dei quali soltanto alcuni conterranno una maggiore quantità di informazioni utili ai fini della ricostruzione dell'immagine stessa. Appare quindi chiaro, attraverso una corretta selezione dei sottospazi di \mathbf{W} si potrebbe pervenire ad una buona descrizione del segnale bidimensionale riducendo il costo di calcolo. E' questo infatti l'obiettivo dell'algoritmo *best Wavelet Packet*, il quale consente di individuare la migliore base wavelet packet, rendendo minimo un funzionale (quale ad esempio l'entropia) scelto ad hoc per l'immagine selezionata e per lo scopo richiesto della particolare applicazione.

2 Le principali applicazioni presenti in letteratura relative alle metodologie analizzate in data fusion in contesti urbani

2.1 Richiami al PAR 2016

In questo capitolo verranno introdotte e richiamate le tecniche e le metodologie già utilizzate nel PAR2016; inizialmente tali metodologie non riguarderanno necessariamente approcci in spazi trasformati; in particolare l'aspetto di interesse che emergerà riguarderà essenzialmente i risultati conseguiti attraverso differenziate metodologie di data fusion, in modo tale che rimandando al PAR2016 (Soluzione Tecnologica di supporto alle decisioni ed Information Fusion per grandi moli di dati in contesto smart cities: Parte I – Definizione di metodologie, metodi e modelli della complessità per il decision making per la smart city) si potrà approfondire il tema in un'ottica di continuità senza correre il rischio di essere qui inutilmente prolissi.

Effettuare la data fusion basata su analisi multiscala non costituisce alcuna difficoltà, poiché ad esempio la wavelet analisi non necessita di specificità dei dati, ma piuttosto permette di operare in uno spazio di frequenza indipendentemente dal tipo di dato che riceve in input. Ciò vuol dire che a seconda del tipo di analisi che si intende effettuare sui dati raccolti in campo in un contesto complesso quale quello dei big data e delle smart cities, si potrà effettuare un'analisi multiscala o su dati nativi o su dati già fusi.

2.2 Introduzione alla Data Fusion

I sistemi di fusione delle informazioni hanno suscitato notevole interesse nel tempo, offrendo sempre nuovi stimoli, vista la loro applicabilità nei domini applicativi più disparati.

Dal momento che spesso le terminologie adottate derivano dal campo di applicazione, in letteratura sono presenti diverse definizioni di Information Fusion.

L'Information Fusion è un insieme di tecniche e metodi che consentono di combinare informazioni provenienti da sorgenti diverse al fine di determinare un quadro dello scenario osservato più preciso e fornire strumenti di supporto alle decisioni ancora più affidabili. In definitiva, i sistemi di fusione delle informazioni mirano ad emulare il comportamento del cervello umano, che è in grado di fondere esperienze passate e dati raccolti dai diversi sensi al fine di acquisire conoscenza e consapevolezza dell'ambiente ed intraprendere azioni con la capacità di prevederne gli effetti. Nell'ambito di una sfida così complessa ma, al tempo stesso, affascinante, diversi passi in avanti sono stati compiuti dalla comunità scientifica, specialmente negli ultimi anni.

Le sorgenti possono differire sotto diversi punti di vista; ad esempio possono essere misure effettuate in punti diversi dello spazio (o anche in momenti diversi del tempo) su una stessa entità o su entità di differente natura. La tipologia di osservazioni che si possono effettuare su tali entità può spaziare, ad esempio, dalla misura effettuata da un comune sensore di campo all'analisi comportamentale effettuato per mezzo di tecniche automatiche di video analisi nell'ambito di un sistema di videomonitoring.

I sistemi di Information Fusion trovano applicazione nei contesti applicativi più disparati, dalle reti di sensori o la robotica fino all'elaborazione dei segnali. Ultimamente di specifico interesse applicativo è diventato l'ambito dei Big Data e dei contesti smart come le smart cities a cui siamo specificamente interessati. Molte applicazioni, specie quelle risalenti alle prime fasi del processo di Information Fusion, sono riconducibili all'area militare. Basti pensare ad esempio a sistemi di tracking per seguire le traiettorie di aerei, dove si sfrutta uno degli aspetti più interessanti legati all'Information Fusion: la capacità di acquisire rapidamente un quadro della situazione tale da consentire la predizione delle evoluzioni future del sistema.

Uno dei campi di applicazione, che ha maggiormente stimolato la ricerca negli ultimi anni, è la Homeland Security, con particolare riferimento alla lotta al crimine legata alla crescita di internet ed alla diffusione dei social network: in questi casi, infatti, gli investigatori hanno a disposizione una quantità di informazioni da elaborare elevatissima. Tale mole di dati è ingestibile se non si dispone di strumenti automatici in grado di filtrare e correlare le informazioni, cosa invece resa possibile da sistemi di fusione delle informazioni. L'obiettivo più stimolante ed al tempo stesso complesso per l'Information Fusion è quindi la possibilità di trasformare grandi quantità di informazioni in conoscenza. E' evidente che quanto detto diventa ancora più significativo in cui si desidera avere un controllo intelligente di aree come è il caso delle smart city.

La progettazione di un sistema di fusione delle informazioni può ispirarsi ai diversi modelli presenti in letteratura. Tali modelli sono classificati sulla base di cosa viene considerato come elemento base del processo di fusione. L'elemento base del processo di fusione può essere il livello di astrazione delle informazioni, come nel JDL model, oppure le attività, come nell'OODA model. Si possono effettuare altre classificazioni, individuando così la categoria dei modelli basati sulle regole e la famiglia dei modelli che pongono particolare attenzione sulla situation awareness. La scelta del modello (o dei modelli) di riferimento viene in genere dettata da considerazioni su quale elemento (astrazione delle informazioni, attività, regole o consapevolezza della situazione) ricopra un ruolo fondamentale all'interno del dominio applicativo considerato.

Inoltre, la topologia dell'architettura con la quale si implementa il sistema di fusione delle informazioni gioca un ruolo importante sulle performance complessive. La ricerca negli ultimi anni si è orientata soprattutto rispetto a scenari basati su rete (net-centric environment). In tali scenari è stato dimostrato come si ottengano prestazioni migliori mediante l'utilizzo di reti oloniche e market-based. L'utilizzo di architetture oloniche porta ad una maggiore robustezza, scalabilità ed efficienza del sistema di Information Fusion. Le architetture market-based, invece, hanno come principale vantaggio la capacità di orientare in maniera adattativa le sorgenti, al fine di ridurre le moli di dati in gioco quando non necessario.

Infine, il processo di fusione è fortemente influenzato dai metodi adoperati per la fusione dei dati in input. Tali prestazioni sono generalmente legate alla capacità dei metodi impiegati per trattare le problematiche legate alla natura dei dati. In letteratura sono presenti diversi metodi per la fusione delle informazioni, anche se generalmente non esistono algoritmi in grado di rispondere al meglio a tutte le problematiche relative ai dati. Ovviamente, a seconda dell'applicazione o al campo di impiego, possono essere combinati più metodi al fine di migliorare le prestazioni globali del sistema di Information Fusion.

Con tale premessa nel PAR2016 è stato avviato uno studio sui più utili paradigmi di *Information Fusion al contesto della smart city*. In particolare, sono stati presentati i principali modelli e tecniche per la realizzazione di sistemi di fusione delle informazioni, fornendo altresì gli utili elementi innovativi per le diverse scelte tecnologiche da adottare all'interno del progetto.

Grazie al capitolo precedente ed avendo la conoscenza del PAR2016 adesso si è in grado di trattare e fondere i dati non solo relativi ad una data scala ma anche a scale di misura ed osservazione diversa. ciò come di consueto, per chi è abituato a lavorare in spazi di frequenza o spazi trasformati, avviene portando il dato dallo spazio di acquisizione o misura ad uno spazio di frequenza, fondendo nello spazio trasformato, ed anti trasformando per tornare allo spazio ordinario in cui è avvenuta l'acquisizione o la misura. Il risultato è una conoscenza o un'informazione più efficace, sintetica e multiscala proprio perché la fusione delle informazioni è avvenuta in uno spazio di frequenza trasformato via FFT o wavelet ad esempio.

Nonostante un tale risultato sia interessante ed originale, esso può essere reso ancora più ampio in termini di portata cognitiva ed in termini di generalizzazione d'impiego all'orquando si abbia a che fare con misure in campo o acquisizioni reali in scenari complessi come le smart cities in cui spesso ci si interseca con il trattamento di big data. In questo scenario è comune operare in condizioni di info-incertezza o info-incompletezza. Questo tema verrà trattato nel capitolo successivo, mentre nel paragrafo successivo offriremo un'outline sull'Information Fusion, rimandando al PAR2016, ogni ulteriore approfondimento.

2.3 Outline

I sistemi di fusione delle informazioni hanno suscitato notevole interesse nel tempo, offrendo sempre nuovi stimoli, vista la loro applicabilità nei domini applicativi più disparati.

Dal momento che spesso le terminologie adottate derivano dal campo di applicazione, in letteratura sono presenti diverse definizioni di *Information Fusion*. L'*Information Fusion* è un insieme di tecniche e metodi che consentono di combinare informazioni provenienti da sorgenti diverse al fine di determinare un quadro dello scenario osservato più preciso e fornire strumenti di supporto alle decisioni ancora più affidabili. In definitiva, i sistemi di fusione delle informazioni mirano ad emulare il comportamento del cervello umano, che è in grado di fondere esperienze passate e dati raccolti dai diversi sensi al fine di acquisire conoscenza e consapevolezza dell'ambiente ed intraprendere azioni con la capacità di prevederne gli effetti. Nell'ambito di una sfida così complessa ma, al tempo stesso, affascinante, diversi passi in avanti sono stati compiuti dalla comunità scientifica, specialmente negli ultimi anni.

Le sorgenti informative possono differire sotto diversi punti di vista; ad esempio possono essere misure effettuate in punti diversi dello spazio (o anche in momenti diversi del tempo) su una stessa entità o su entità di differente natura. La tipologia di osservazioni che si possono effettuare su tali entità può spaziare, ad esempio, dalla misura effettuata da un comune sensore di carica elettrica passante o temperatura all'analisi comportamentale effettuato per mezzo di tecniche automatiche di video analisi nell'ambito di un sistema di videomonitoring.

I sistemi di *Information Fusion* trovano applicazione nei contesti applicativi più disparati, dalle reti di sensori o la robotica fino all'elaborazione dei segnali. Molte applicazioni, specie quelle risalenti alle prime fasi del processo di *Information Fusion*, sono riconducibili all'area militare. Basti pensare ad esempio a sistemi di *tracking* per seguire le traiettorie di aerei, dove si sfrutta uno degli aspetti più interessanti legati all'*Information Fusion*: la capacità di acquisire rapidamente un quadro della situazione tale da consentire la predizione delle evoluzioni future del sistema.

Uno dei campi di applicazione, che ha maggiormente stimolato la ricerca negli ultimi anni, è la *Homeland Security*, con particolare riferimento alla lotta al crimine legata alla crescita di internet ed alla diffusione dei social network: in questi casi, infatti, gli investigatori hanno a disposizione una quantità di informazioni da elaborare elevatissima. Tale mole di dati è ingestibile se non si dispone di strumenti automatici in grado di filtrare e correlare le informazioni, cosa invece resa possibile da sistemi di fusione delle informazioni. L'obiettivo più stimolante ed al tempo stesso complesso per l'*Information Fusion* è quindi la possibilità di trasformare grandi quantità di informazioni in conoscenza. Da ciò si comprende quanto tali tecniche possano essere utili ed efficaci nel contesto dei big data legati alla Smart City.

La progettazione di un sistema di fusione delle informazioni può ispirarsi ai diversi modelli presenti in letteratura. Tali modelli sono classificati sulla base di cosa viene considerato come elemento base del processo di fusione. L'elemento base del processo di fusione può essere il *livello di astrazione delle informazioni*, come nel *JDL model*, oppure le *attività*, come nell'*OODA model*. Si possono effettuare altre classificazioni, individuando così la categoria dei modelli basati sulle *regole* e la famiglia dei modelli che pongono particolare attenzione sulla *situation awareness*. La scelta del modello (o dei modelli) di riferimento viene in genere dettata da considerazioni su quale elemento (astrazione delle informazioni, attività, regole o consapevolezza della situazione) ricopra un ruolo fondamentale all'interno del dominio applicativo considerato.

Inoltre, la topologia dell'architettura con la quale si implementa il sistema di fusione delle informazioni gioca un ruolo importante sulle *performance* complessive. La ricerca negli ultimi anni si è orientata soprattutto rispetto a scenari basati su rete (*net-centric environment*). In tali scenari è stato dimostrato come si ottengano prestazioni migliori mediante l'utilizzo di reti oloniche e *market-based*. L'utilizzo di architetture oloniche porta ad una maggiore robustezza, scalabilità ed efficienza del sistema di *Information Fusion*. Le architetture *market-based*, invece, hanno come principale vantaggio la capacità di orientare in maniera adattativa le sorgenti, al fine di ridurre le moli di dati in gioco quando non necessario.

Infine, il processo di fusione è fortemente influenzato dai metodi adoperati per la fusione dei dati in input. Tali prestazioni sono generalmente legate alla capacità dei metodi impiegati di trattare le problematiche legate alla natura dei dati. In letteratura sono presenti diversi metodi per la fusione delle informazioni, anche se generalmente non esistono algoritmi in grado di rispondere al meglio a tutte le problematiche relative ai dati. Ovviamente, a seconda dell'applicazione, possono essere combinati più metodi al fine di migliorare le prestazioni globali del sistema di *Information Fusion*.

3 Il trattamento delle informazioni incerte o incomplete in prospettiva probabilistica con applicazione in smart city

In molti contesti come le smart cities si sperimentano situazioni in cui i dati acquisiti non rispondono ad una distribuzione di probabilità attesa o modellata. In altri casi addirittura eventi poco o scarsamente probabili accadono con una frequenza inusuale. Per risolvere tali problemi spesso ci si accontenta di modellare i fenomeni in modo gaussiano e descrivere le discrepanze tra il misurato e l'atteso grazie all'impiego di code pesanti, ovvero code più massicce di quanto previsto dalla distribuzione gaussiana. Seppure ciò possa rappresentare una patch ai singoli problemi di fatto manca una descrizione ed un modello in grado di trattare eventi definiti rari ma che di fatto accadono più spesso di quanto si immagini nella realtà.

Questo capitolo intende offrire una metodologia che sicuramente è impiegabile nel contesto smart cities ed in particolare nel contesto SCP-ENEA soprattutto per la gestione e la valorizzazione dei dati provenienti dalle solutions verticali. Infatti, sebbene sia una metodologia applicabile ai Big Data, di fatto mostra una generalità ed un'applicabilità molto più ampia, in grado di affrontare con una metodologia robusta i temi della info-incertezza ed info-incompleta in ogni ambito in cui tali tematiche si presentano.

Oltre al concetto di probabilità, come vedremo più avanti, fin dall'antichità sono emersi altri concetti per descrivere l'incerto, quali la plausibilità, la credibilità e la possibilità.

Per plausibilità intendiamo una condizione che viene verificata in un gran numero di processi inferenziali: naturalmente quindi esistono diversi modi di intendere il ragionamento plausibile, di definirne la forma, le funzioni e le principali forme inferenziali. Nella formulazione di una teoria dell'inferenza plausibile in primo luogo esiste una definizione preliminare di cosa sia 'incerto' e di come esso intervenga nel corso del processo inferenziale. Inoltre è da non sottovalutare che un'inferenza plausibile può essere invalidata, corretta o anche solo aggiornata in corrispondenza dell'ingresso di nuovi dati o informazioni.

George Polya nel 1954 in una delle sue opere più influenti, definisce il ragionamento plausibile in aperta contrapposizione con quello strettamente dimostrativo: l'inferenza plausibile serve ad acquisire nuova conoscenza, essa ha un carattere fluido, provvisorio, controverso e azzardato, e le sue forme tipiche sono le induzioni e l'analogia che per loro natura non sono certe e non trasmettono la verità.

Il filosofo Resche nel 1976 definisce il ragionamento plausibile come una deduzione a partire da varie premesse incerte, finalizzato al trattamento delle "dissonanze cognitive": sono dunque le premesse e le conclusioni a essere provvisorie, controverse, fluide.

Dezert invece nel 2002 ipotizza il ragionamento plausibile come un'estensione della teoria dell'evidenza di Dempster-Shafer che permette di trattare informazioni incerte, ma anche paradossali, grazie al cosiddetto insieme "iper-potenza".

3.1 La plausibilità

3.1.1 Caratteristiche

La plausibilità è un concetto che troviamo in svariati contesti: dalle previsioni sull'andamento della Borsa, alla generazione di ipotesi scientifiche, fino alla valutazione di una giustificazione data da un bambino in seguito ad una malefatta, come dicevano Connell e Keane nel 2003: "il ricorso a questa nozione sembra infatti davvero un fenomeno ineluttabile".

Tutte le conoscenze sono condizionate dalla natura non certa della plausibilità, soprattutto le conoscenze inerenti alla matematica. Possiamo dire che nessuna delle nostre conoscenze matematiche è assolutamente certa, è soltanto plausibile, cioè compatibile con la conoscenza esistente, e tale compatibilità non ne assicura la certezza, poiché la conoscenza esistente non è assolutamente certa, è soltanto plausibile.

La plausibilità può essere molto generalmente, concepita come una nozione che tratta le forme di ragionamento che candidano conclusioni e valutano ipotesi incerte.

3.1.2 Riferimenti storici: la tradizione greca

La plausibilità possiamo trovarla sin dall'antica Grecia, dove viene sviluppata nella retorica, a partire dal cosiddetto argomento *eikotico*, attribuito da Aristotele e successivamente Platone alla scuola sofista. Codesto concetto, così come si configura nel mondo greco antico, è dunque collegato all'apparenza, alla soggettività e dà luogo ad inferenze condotte su assunzioni implicite.

Nel caso della plausibilità, il giudizio viene formato attraverso il riferimento agli ascoltatori e alla loro assunzione sulle cose, piuttosto che in riferimento alle cose in sé.

Secondo Aristotele, nell'esempio della verità, il giudizio viene formato attraverso il riferimento alla cosa in sé con cui l'opinione ha a che fare.

Ma la formulazione più importante della plausibilità più nota nell'antichità probabilmente è quella di Sesto Empirico, nella quale egli attribuisce a Carneade di Cirene e che descrive nel suo saggio "Contro i logici". In questo caso 'Plausibile' vuol significare 'persuadere'. I criteri sono i seguenti:

- criterio 1)** *qualcosa è plausibile se sembra essere vera, oppure*
- criterio 2)** *è ancora più plausibile se sembra essere vera ed è compatibile con altre cose che sembrano essere vere, oppure*
- criterio 3)** *è stabile ed è testata*

Questa teoria si sviluppa con l'intento ben preciso di mostrare che sebbene la realtà sia inconoscibile, possono esistere dei gradi di conoscibilità, per cui esistono cose più vere o meno vere.

3.1.3 Incertezza dei metodi

L'incertezza può essere di due tipi:

- 1) Stocastica, ovvero quando riguarda l'imprevedibilità intrinseca del sistema oggetto di conoscenza;
- 2) *Soggettiva*, ovvero quando si riferisce allo stato di conoscenza del soggetto rispetto al sistema.

Questi due tipi danno luogo ad una delle principali distinzioni, quella che è meglio identificata come la distinzione tra il "defeasible reasoning" e il "plausible reasoning". La prima esprime un approccio secondo il quale "un argomento è considerato come un frammento fallibile di ragionamento fondato su premesse ferme" (Vreewsjk 1992,1), dunque su metodi incerti ma premesse certe. Con l'espressione metodi incerti ci si riferisce alle forme di inferenza fallibili in generale - come l'analogia, l'induzione, l'abduzione, etc. - che per loro natura non sono in grado di trasmettere la verità dalle premesse alle conclusioni, e che possono dunque solo conferire un certo grado di "probabilità", o meglio di attendibilità, alle conclusioni candidate.

3.1.4 Incertezza e allocazione della credenza globale

L'incertezza epistemica è espressione di una relazione soggettiva, e quindi una teoria della plausibilità è concezione della soggettività. Tale concezione è caratterizzata da un insieme di ipotesi e assunzioni sul comportamento e le caratteristiche proprie del soggetto che compie valutazioni e previsioni sotto condizioni di incertezza.

Una trattazione della nozione di plausibilità e lo sviluppo di una teoria sull'inferenza plausibile passa attraverso l'indagine del controverso rapporto che esiste tra le nozioni di plausibilità, credenza, probabilità, e delle loro differenze concettuali e funzionali. La possibilità di distinguere chiaramente ciò che è plausibile da ciò che è probabile o credibile è infatti uno dei presupposti per uno sviluppo di una chiara teoria sul ragionamento plausibile e non a caso è un tema che ricorre spesso nelle trattazioni dell'inferenza plausibile.

3.2 La definizione probabilistica

3.2.1 La nascita della concezione probabilistica

Per potere formalizzare e trattare la plausibilità vi è l'importante nozione della probabilità. Alla base della concezione della plausibilità c'è perciò l'idea che la plausibilità, intesa quale ragionamento sotto condizioni d'incertezza, esprima e tratti la credenza e i suoi gradi, e che la probabilità sia lo strumento adeguato a fornire, in tal senso, una trattazione esauriente.

3.2.2 Brevi cenni sull'accezione di probabile

La teoria della probabilità può essere studiata tramite due principali interpretazioni: quella *ontologica* e quella *epistemologica*.

L'interpretazione ontologica è caratterizzata dalla descrizione della causalità ad essa connessa. Essa si divide in tre principali interpretazioni: la *classica*, la *frequentista relativa* e infine la *propensitiva*.

Nell'interpretazione classica, la nozione di probabilità è sviluppata a partire dal "principio di ragion insufficiente" e dal "principio di ragione cogente": il primo asserisce che in caso di impossibilità di determinare il risultato più probabile, bisogna porre tutti i risultati come egualmente probabili; il secondo sostiene che le simmetrie fisiche implicano uguali probabilità. Ciò conduce alla tradizionale definizione della probabilità quale "rapporto del numero dei casi favorevoli su quello di tutti i casi possibili" (Laplace 1812, xi).

Tuttavia, questa concezione, e i principi su cui risiede, va incontro a molti problemi e limitazioni, che sono legati principalmente al significato attribuito alla nozione di simmetria e alle possibili conseguenze non additive e contro-intuitive del principio di ragion insufficiente.

Per ovviare a queste lacune, l'interpretazione classica è stata sottoposta a numerose riformulazioni e varianti, di cui la più celebre è quella costruita da Von Mises (1928), nota con il nome di concezione *frequentista relativa*. Essa si basa sulla nozione di "*Kollektiv*"; un'astrazione matematica, che esprime semplicemente una serie infinita di prove indipendenti. La nozione di probabilità viene così definita a partire da quella di *Kollektiv*: la probabilità di un dato evento, nel corso di una particolare prova, è la frequenza relativa all'occorrenza di quel evento in un *Kollektiv*. In un certo senso, essa è collegata alla famosa legge dei grandi numeri di Bernoulli. Questa, infatti, afferma che se un evento si ripete per un certo numero di volte, k , in n prove indipendenti e identiche, allora, se il numero delle prove è sufficientemente ampio, il rapporto k/n tende ad avvicinarsi sufficientemente alla probabilità oggettiva di quel evento. Tuttavia l'idea di *Kollektiv* fa riferimento alla possibilità di ripetizione infinita, che come tale non solo è una pura idealizzazione, ma crea problemi quando ci si trova a dover rendere conto di eventi intrinsecamente unici (ad esempio le elezioni presidenziali in Italia in un dato anno).

Il tentativo di superare questo limite della visione frequentista costituisce il punto di partenza della terza concezione di natura ontologica della probabilità: quella nota come propensitiva. Essa venne inizialmente proposta da Peirce (1910) e quindi diffusa da Popper (1959) e sostiene che la probabilità rappresenta la disposizione, o la "propensione", della Natura a produrre un particolare evento in una singola prova, senza che esso sia associato necessariamente ad una frequenza a lungo corso.

La concezione epistemologica, che affonda le proprie radici nel lavoro di Bayes (1763), concepisce la nozione di probabilità partendo da un'ipotesi del tutto diversa da quella dell'approccio ontologico: infatti essa nega che la causalità possa essere un fenomeno misurabile in modo oggettivo e la concepisce come un fenomeno meramente conoscitivo: Secondo tale approccio, infatti, anche il semplice lancio di un dado, ad esempio, non è necessariamente, di per sé, causale. Se fossimo in grado di conoscere tutti i valori delle variabili che concorrono alla determinazione del risultato di un lancio (come la forma esatta ed il peso del dado, le condizioni atmosferiche dell'ambiente in cui il lancio viene eseguito, la distanza tra la mano del lanciatore e il terreno, etc.) saremmo in grado di predire con certezza, per mezzo dell'applicazione delle sole leggi della fisica, il risultato del lancio: "se il reale ragionamento probabilistico è condotto fino alle sue estreme conclusioni, sembra che non ci sia " realmente" alcuna probabilità, ma solo certezza, se la conoscenza è completa" (Knight 1921, 219). Poiché questa mole d'informazione non è comunemente accessibile, si preferisce assumere l'evento come casuale e assegnare valori probabilistici ad ognuna delle

sei facce del dado. Alla luce della concezione epistemologica, o soggettiva, la probabilità non attiene dunque alla sfera dell'essere (*ontos*) ma a quella conoscitiva (*episteme*): essa esprime una *manca di conoscenza* rispetto alle condizioni che influenzano l'evento ed esprime le nostre personali, soggettive credenze e opinioni circa l'esito della prova. All'interno della concezione epistemologica è possibile individuare almeno tre grandi versanti teorici: il *relazionismo logico*, il *soggettivismo* e l'*intuizionismo*.

Il relazionismo logico, posizione che ha una sua rappresentazione compiuta nei lavori di Keynes (1921) e Carnap (1950), sostiene che esiste una "oggettiva" *relazione*, anche se non sempre misurabile, tra la conoscenza e la probabilità che sono derivate da essa.

Il rifiuto di questa asserzione rappresenta proprio il punto da cui muove la visione soggettiva della probabilità, elaborata in modo indipendente da Ramsey (1926) e de Finetti (1931 e 1937). Secondo questa posizione, dunque, la probabilità non si lega alla conoscenza in sé e per sé, ma quella posseduta da un particolare individuo: sono le conoscenze personali che determinano le probabilità e quindi la probabilità è di natura strettamente soggettiva.

Tuttavia anche questa asserzione non solo si rivela controversa, ma la sua messa in questione costituisce il punto di partenza di un altro approccio di natura epistemologica alla probabilità; l'*intuizionismo*. Secondo questa posizione, che ha le sue radici nei lavori di Koopman (1940 a e 1940b) e Good (1950, 1962), l'approccio di de Finetti e Ramsey è troppo dogmatico. Esso infatti implica, come abbiamo sottolineato, che una credenza non è tale se non viene espressa (rivelata) da una scelta comportamentale. L'aspetto da sottolineare rimane che la concezione probabilistica della plausibilità muove da una visione epistemologica della teoria della probabilità per costruire una teoria dell'inferenza plausibile.

3.2.3 Il teorema di Cox

Il teorema di Cox (dal nome del fisico e statistico Richard Threlkeld Cox) è una derivazione delle leggi della probabilità a partire da un certo insieme di postulati.

Offre una motivazione formale e precisa, sul legame che può essere istituito tra la nozione e il calcolo della probabilità, e la plausibilità. Esso stabilisce infatti, sotto certe condizioni, l'esistenza di una precisa relazione tra la nozione di plausibilità, *pl*, intesa come ragionamento su gradi di opinione, credenza, confidenza, credibilità, e la teoria del calcolo della probabilità, *pr*. In particolare, dimostra che qualsiasi sistema di inferenze plausibili, che non sia isomorfo alla teoria della probabilità, debba violare almeno una delle seguenti "ragionevoli" e "intuitive" condizioni:

$$pl(A|X) = 0 \leftrightarrow A \text{ è falsa in } X$$

$$pl(A|X) = 1 \leftrightarrow A \text{ è vera in } X$$

$$0 \leq pl(A|X) \leq 1$$

$$pl(A \wedge B|X) = pl(A|X) \cdot pl(B|X)$$

$$pl(\neg A|X) = 1 - pl(A|X)$$

dove, *pl*(A|X) indica la plausibilità di A rispetto a X.

Nel caso in cui il sistema in questione non violi alcuna delle condizioni i) – v), si può dunque sostituire a *pl*, senza alcuna perdita di informazione, il concetto *pr* e fare ricorso al calcolo delle probabilità per modellare la nozione di plausibilità. Il teorema è dunque uno degli argomenti più forti cui si ricorre per giustificare l'uso di una distribuzione probabilistica per trattare sistemi di inferenza su gradi di credenze: in base il teorema, ogni misura di credenza è isomorfa a misure probabilistiche. Se quindi si intende la plausibilità come espressione di una credenza e il ragionamento plausibilistico come calcolo di gradi di credenza, il teorema di Cox giustifica l'uso della teoria della probabilità come mezzo per modellare l'incertezza epistemica.

Per gradi di credenza (plausibilità) si intende che si ha che fare con "proposizioni che non possono essere dimostrate vere o false in modo definitivo" (Van Horn, 2003, 2). Essi vanno dunque distinti dai "gradi

verità”, trattati dalla logica fuzzy, onde evitare “controversie non necessarie”. Per Colyvan “la probabilità è l’unica rappresentazione coerente dell’incertezza” (Colyvan, 2004, 72). Bisogna notare ancora come la plausibilità quale ragionamento su gradi di credenza si differenzia in modo essenziale dal ragionamento su gradi di verità, proprio della logica fuzzy. Mentre i primi, come osservato da Dubois e Prade (Dubois e Prade, 1994) sono indotti da stati d’incompletezza dell’informazione, i secondi si riferiscono a proprietà la cui soddisfazione è questione di gradi, sono verità parziali che tuttavia non costituiscono una forma di incertezza.

Come vedremo le concezioni non probabilistiche della plausibilità, come la teoria di Rescher, violano espressamente alcune delle condizioni i) – v), dando così luogo a visione della plausibilità tesa a trattare situazioni d’incertezza per le quali il calcolo probabilistico si rivela inefficace.

3.2.4 Il teorema di Bayes

Il teorema di Bayes (proposto dal matematico Thomas Bayes), deriva da due teoremi fondamentali della probabilità: il teorema della probabilità composta e il teorema della probabilità assolta.

Ci pone una questione che “dovrebbe essere necessariamente presa in considerazione da chiunque voglia dare una chiara descrizione della forza del ragionamento analogico o induttivo” (Bayes 1763, 373). Da un punto di vista formale il teorema di Bayes è una semplice conseguenza logica della definizione della probabilità condizionata:

(PC) $\Pr\{A\} \Pr\{B/A\} = \Pr\{B\} \Pr\{A/B\} = \Pr\{A,B\}$,
dove $\Pr\{A,B\}$, indica la probabilità congiunta di A e B.

Una sua classica rappresentazione, che si ottiene dividendo entrambi i membri di (PC) per $\Pr\{B\}$, è dunque,

$$\Pr\{A/B\} = \frac{\Pr\{B/A\}\Pr\{A\}}{\Pr\{B\}}$$

$\Pr\{B/A\}$ è nota come la funzione di probabilità per A dato un valore specifico di B;

Il teorema è così punto centrale della concezione bayesiana della probabilità, secondo la quale l’incertezza e i gradi di credenza possono essere misurati come probabilità.

3.2.5 Barriere e limiti

La teoria di Bayes è un esempio tipico di sistema monodimensionale a somma uno. Esso permette di gestire sistemi e stati d’informazione dalle caratteristiche piuttosto precise e limitate, che possono sostenere solo una visione limitata della probabilità e, di conoscenza, della plausibilità. La probabilità opera con la deduzione, la plausibilità con l’analogia e l’induzione.

3.3 La logica del plausibile di Polya

Lo studio di Polya della nozione di plausibilità prende inizio da una chiarezza che può essere considerata tuttora valida. Nel lessico attuale le seguenti le parole: probabile, plausibile, possibile, e credibile hanno significati molto simili.

Polya sviluppa in particolar modo un approccio al concetto di plausibilità e d'inferenza plausibile, direttamente ispirata alla concezione epistemologica della probabilità, fondata sul teorema di Bayes. Egli può essere così legittimamente considerato come uno dei principali teorici dell'approccio probabilistico alla plausibilità, secondo cui "la logica del plausibile dovrebbe condursi a coincidere con la logica del probabile (calcolo della probabilità) sia pure limitandosi – ove del caso – a considerare certe conclusioni qualitative ottenibili prescindendo dall'esatta determinazione numerica delle probabilità pur di supporre note delle disuguaglianze fra i loro valori" de Finetti 1949, 235).

3.3.1 L'inferenza plausibile

L'inferenza plausibile si va naturalmente a comporre di regole ben precise. Polya le formalizza con rigore matematico nel modo seguente:

1)

A fondamento dell'approccio probabilistico di Polya v'è innanzitutto il "modello induttivo fondamentale o, in qualche modo più brevemente, il metodo induttivo".

A implica B

B è vera

A è più credibile

Questo modello, pur "non dicendo nulla di sorprendente" formalizza una "credenza che nessuna persona ragionevole metterebbe in dubbio: la verifica di una conseguenza rende una congettura più credibile" (ibid).

Mentre l'argomento dimostrativo, come ad esempio il sillogismo aristotelico, è contraddistinto da proprietà precise: l'impersonalità, l'universalità, l'auto-sufficienza e la definitività, poiché esso, rispettivamente, non dipende dalla personalità di colui che compie il ragionamento; non dipende da alcun dominio di conoscenza specifico: non c'è "bisogno di nulla oltre le premesse per rendere valida la conclusione e nulla la può invalidare se le premesse rimangono solide"; se le premesse sono certe la conclusione è "definitiva".

L'argomento plausibile, così come definito da Polya, gode a sua volta delle prime tre proprietà, ma non della quarta – non è definitivo: esso produce dunque conclusioni che sono solo provvisorie, momentanee, transitorie. Il ragionamento plausibilistico non è perciò, in questo senso, soggettivo e psicologico: i modelli di ragionamenti sono impersonali, mentre è la forza delle conclusioni che è di natura soggettiva e quindi non rappresentabile per mezzo di quantità.

Al modello fondamentale induttivo sono legati altri principi del ragionamento plausibile, che in alcuni casi lo perfezionano. Come per il modello induttivo fondamentale, essi hanno un corrispondente nella logica dimostrativa.

1.1)

a)

A implica B_{n+1}

B_{n+1} è molto diversa dalle conseguenze di A

Precedentemente verificate B_1, B_2, \dots, B_n

B_{n+1} è vera

A è molto più credibile

b)

A implica B_{n+1}

B_{n+1} è molto simile alle conseguenze di A

precedentemente verificate B_1, B_2, \dots, B_n

B_{n+1} è vera

A è solo un poco più credibile

Queste due forme complementari di inferenza plausibile, che concernono la successiva verifica di diverse conseguenze, esplicitano un altro principio induttivo: "la verifica di una nuova conseguenza conta di più o di meno a seconda se la nuova conseguenza differisce di più o di meno dalle conseguenze precedentemente verificate".

1.2)

a)

A implica B

B è altamente improbabile

A è molto credibile

b)

A implica B

B è altamente probabile

A è solo un poco più credibile

Queste due forme complementari di inferenza plausibile, che concernono la verifica di conseguenze improbabili o meno, esplicitano ancora un altro principio induttivo: "la verifica di una conseguenza conta di più o di meno a seconda se la conseguenza è più o meno probabile di per sé".

2)

A è analogo a B

B è vera

A è più credibile

Questo modello, che esplicita l'inferenza per via analogica, esprime il principio di inferenza plausibile per la quale "una congettura diventa più credibile quando una congettura analoga si rivela essere vera".

L'inferenza analogica può inoltre assumere forme "più deboli o nascoste"; in particolare si ha:

2.1)

A è analogo a B

B è più credibile

A è più credibile

Questo modello esprime il principio di inferenza plausibile per il quale "una congettura diventa in qualche modo più credibile quando una congettura analoga diventa più credibile".

3)

A è implicato da B

B è falso

A è meno credibile

Questo modello, che valuta una possibile base – dove per base si intende una proposizione che implica la congettura in questione -, esprime il principio di inferenza plausibile per il quale "la nostra confidenza in una congettura può solo diminuire quando una possibile base della congettura viene inficiata".

4)

A è incompatibile con B

B è falso

A è più credibile

Questo modello, che valuta una congettura conflittuale, esprime il principio di inferenza plausibile per il quale "la nostra confidenza in una congettura può solo aumentare quando una congettura rivale incompatibile viene inficiata".

5.a)

A implica B

B meno è credibile

A è meno credibile

5.b)

A implica B

B è più credibile

A è in qualche modo più credibile

Questo insieme di modelli, che rappresenta una forma debole del modello fondamentale, esprime il principio di inferenza plausibile per il quale “la nostra confidenza in una congettura è influenzata dalla nostra confidenza in una delle sue conseguenze e varia nella stessa direzione”.

Esso è un altro esempio, insieme al principio 2.1, di modelli di ragionamento plausibile costruiti a partire dai modelli base, che permettono di produrre modelli più articolati, in particolare quelli che Polya definisce nascosti (“shaded”), in virtù “dell’indebolimento della seconda premessa: ‘meno credibile’ al posto di ‘falso’; ‘più credibile’ invece di ‘vero’ “. Polya osserva ancora come sia possibile formalizzare forme deboli di ragionamento plausibile per tutti i modelli base. Ad esempio, il modello, 4) – la valutazione di congetture conflittuali – ha la seguente forma debole:

A è incompatibile con B

B è meno credibile

A è in qualche modo più credibile

Questi modelli sono regole che sono dunque chiamate a stabilire qualche tipo di evidenza merita di essere presa in considerazione e sono di natura impersonale, mentre è personale e soggettivo stabilire “se una particolare prova appena sottoposta abbia un peso sufficiente o meno”.

3.3.2 Interpretazione probabilistica dei modelli d’inferenza plausibile

Punto principale e necessario per la concezione di Polya della plausibilità è il famoso teorema di Bayes. La credibilità di B e il grado di confidenza di A se B vera. Per teorema di Bayes abbiamo che la seguente relazione è valida.

$$Pr\{A\} Pr\{B/A\} = Pr\{B\} Pr\{A/B\}$$

Poiché A implica B è una condizione soddisfatta, si ha che $Pr\{B/A\} = 1$ e dunque sostituendo questo valore in (1) si ha

$$Pr\{A\} = Pr\{B\} Pr\{A/B\}$$

Questa equazione mostra dunque innanzitutto come “ $Pr\{A\}$ varia nella stessa direzione di $Pr\{B\}$ “. Si supponga ora, al fine di trattare propriamente il modello il modello induttivo, che si riesca a dimostrare che B, conseguenza di A, sia vera. Prima della prova di B, l’agente ideale ha un qualche grado di credibilità sia in B, espresso da $Pr\{B\}$, sia in A, espresso da $Pr\{A\}$ e può anche considerare di allocare un certo grado di confidenza in $Pr\{A/B\}$, la credibilità di A dopo la eventuale prova di B. Dopo la prova di B il valore $Pr\{B\}$ assume ovviamente il valore massimo pari a uno. Dunque otteniamo il nuovo valore di credibilità di A dopo la prova di B, che è pari a

$$\Pr \{A\} = \Pr\{A/B\}$$

Assumendo che il valore di $\Pr\{A/B\}$ rimanga ovviamente inalterato, e osservando che $0 < \Pr\{B\} < 1$, si ottiene

$$\Pr\{A\} < \Pr\{A/B\}$$

Poiché $\Pr\{A\}$ e $\Pr\{A/B\}$ rappresentano rispettivamente il grado di credibilità di A prima e dopo la prova di B, la disuguaglianza (4) esprime formalmente in termini probabilistici il principio secondo cui la verifica di una conseguenza rende una congettura più credibile. Essa è dunque la giustificazione probabilistica del modello induttivo. Tuttavia ciò non dice nulla di nuovo rispetto all'inferenza bayesiana: proprio come implicato dal teorema di Bayes si dimostra semplicemente come l'evidenza B conferma un'ipotesi A se solo se $\Pr \{A/B\} > \Pr\{A\}$.

Similmente, avvalendosi di opportune rappresentazioni in termini probabilistici, si possono giustificare gli altri modelli d'inferenza plausibile descritti da Polya.

3.3.3 Analogia e scoperta

Polya non attribuisce un ruolo di particolare rilevanza, tra i modelli d'inferenza plausibili, all'analogia, che egli definisce come "una forma di similarità". Il limite forte dell'analogia è presente nel fatto che essa non è in grado di offrire alcun supporto di natura logica o probabilistica alle sue conclusioni. Ha bisogno di un ulteriore sostegno, che può essere dato attraverso le regole impersonali del ragionamento plausibile esplicitata da Polya.

3.3.4 Analogia e Probabilità

Il problema che l'analogia pone all'intero approccio di Polya alla plausibilità è duplice. Da una parte riguarda la natura del supporto che le premesse conferiscono alla conclusione candidata; è infatti esso probabilistico o meglio non esprimibile in termini probabilistici. Dall'altra parte riguarda la giustificazione che può essere conferita ad una congettura mediante il processo di confronto con la conoscenza esistente. Intesa come mezzo di scoperta, l'analogia apre dunque una domanda critica e l'assenza di un adeguato approfondimento probabilistico della nozione di "analogo" e "analogia" rappresenta un vuoto essenziale dell'approccio di Polya al problema della plausibilità. L'analogia sfugge a un processo di analisi e riduzione alla teoria della probabilità. In realtà "né la teoria classica della probabilità né quella moderna sono state in grado di dare una descrizione soddisfacente e una giustificazione all'inferenza per analogia" (Carnap 1950, 569 – 570).

La logica e probabilità non sono riuscite a rendersi conto in modo totalmente efficace della complessità del processo d'inferenza analogica. Il maggior o minore grado di similarità (è questo il senso in cui Polya intende l'analogia), la maggiore o minore rilevanza delle proprietà messe in gioco nell'analogia, la presenza di "strutture determinanti" non rende più o meno probabile un'analogia. L'analogia è senza ombra di dubbio utile, fertile e feconda, ma non fornisce alla conclusione che candida alcuna probabilità. Essa rimane semplicemente una forma d'inferenza plausibile, dove plausibile non può essere ridotto e interpretato in termini probabilistici.

3.3.5 Il credibile

Il teorema di Bayes è il punto centrale intorno al quale Polya fissa con fermezza un'ulteriore distinzione fondamentale all'interno di una teoria della plausibilità: quella tra la nozione di *credibilità* e di *probabilità*. Alla luce della concezione epistemologica della probabilità adottata da Polya, tale distinzione riceve tuttavia un trattamento del tutto simile a quello riservato alla plausibilità: essa può essere in ultima analisi interamente trattata in termini probabilistici. Infatti, sebbene "credibilità e probabilità siano definite in modo del tutto differente" (Polya, 1954, II, 129), la credibilità può essere concepita come una probabilità

condizionata $Pr \{E/H\}$ e quindi “avere un valore numerico, uguale alla probabilità che un evento del tipo predetto da E accadrà, computata sulla base dell’ipotesi statistica H ”.

Per esempio supponiamo di analizzare un mercato rialzista, il cui prezzo dello strumento finanziario considerato sia improbabile perché molto alto, ad esempio superiore ad una media EMA 200 ed inferiore ad una media EMA 300 e comunque oltre le 3σ di deviazione standard dal prezzo di ingresso (improbabile quindi relativamente a tale ultima condizione). Se il prezzo è più vicino alla EMA 300 che non alla EMA 200 è più plausibile o credibile che si avvicina ad essa per rimbalzare e cominciare a scendere piuttosto che lo faccia al livello di prezzo attuale o corrente. Pertanto, secondo quanto espresso da Polya possiamo descrivere tale credenza come:

$P(EMA\ 300/EMA\ 200)$, cioè la probabilità che raggiunge il prezzo rappresentato dalla EMA 300 dato che ha già superato ampiamente il livello di prezzo fornito dalla EMA 200. Pertanto in generale, per fenomeni o dinamiche entro 3σ possiamo adottare la probabilità, oltre per descrivere la plausibilità e la credibilità adottiamo le probabilità condizionate.

La probabilità, nella sua interpretazioni epistemologica, funge così da fondamento sia per la nozione di credibilità sia per quella di plausibilità, conducendo ad una sostanziale sovrapposizione dei due concetti: quando abbiamo a che fare con la plausibilità di una congettura A , abbiamo infatti a che fare con “l’attendibilità di questa congettura A , la forza dell’evidenza a favore di A , la nostra fiducia in A il grado di credenza che dovremmo avere in A , in breve la *credibilità della congettura A* ”.

Così ovviamente è possibile procedere alla rappresentazione della credibilità in termini probabilistici, semplicemente riadattando il teorema di Bayes mediante la sostituzione della notazione della probabilità (Pr) con quella della credibilità (Cre).

3.3.6 *Barriere e limiti*

L’approccio sistematico di Polya alla nozione di plausibilità e le sue risposte alle domande cruciali per una teoria della plausibilità sono chiare e ben delineate. Afferma che la plausibilità è un concetto incaricato a trattare le forme di ragionamento euristico, per sua natura empirica e sperimentale. Successivamente afferma che un’ipotesi è plausibile quando è *credibile*, ovvero quando le sue conseguenze si accordano con la conoscenza esistente o con altra conoscenza plausibile. Il suo ambito di applicazione non è solo quello del discorso quotidiano, ma concerne la scienza in generale, spaziando dalla fisica sperimentale alla matematica. Infatti “matematici e fisici pensano in modo simile; essi sono guidati, e talvolta sviati, dagli stessi modelli di ragionamento plausibile” (Polya 1959, 384). Il ragionamento plausibile ha le sue proprie regole (i modelli 1-5), diverse da quella del ragionamento deduttivo, e una sua logica. La logica del ragionamento plausibile può essere espressa e fondata probabilisticamente: la plausibilità si riduce in ultima analisi ad opportuna versione ed interpretazione (quella qualitativa epistemologica) del calcolo probabilistico. Di fatto, plausibilità e probabilità sono indistinguibili, in quanto la prima può essere ridotta alla seconda.

La concezione probabilistica riformulata da Polya risente purtroppo di tutte le assunzioni e le limitazioni proprie della concezione bayesiana.

Essa tratta la plausibilità come un ragionamento su gradi di credenza che soddisfano gli assiomi del calcolo probabilistico, e si attiene al quadro classico trattando l’allocazione della credenza globale come una teoria mono-dimensionale *a somma uno*, escludendo così tutta una classe di fenomeni che ineriscono in modo essenziale al ragionamento plausibile. La concezione di Polya è dunque di natura fortemente riduzionistica e arriva a trattare il ragionamento plausibile come una forma di inferenza probabilistica. Se l’approccio probabilistico di Polya coglie un nodo critico della teoria della plausibilità quando dichiara la necessità di distinguere “plausibile” da “probabile” e “credibile” e mettere ordine all’interno di una costellazione di concetti e termini tra loro molto ambigui, non è possibile non rilevare, in ultima analisi, che “filosofi come

Russell (1948) hanno rimarcato come la probabilità matematica si applichi solo ad alcune istanze del ragionamento plausibile.

Tuttavia, pur nei suoi limiti, l'approccio di Polya rimane non solo un esempio paradigmatico della concezione probabilistica della plausibilità, ma un momento saliente della costruzione di una teoria generale della plausibilità. Come vedremo in dettaglio, gli altri approcci probabilistici possono essere guardati come semplici varianti del nucleo della teoria di Polya della plausibilità (l'uso della teoria della probabilità): esse si sviluppano modificando alcune condizioni iniziali del sistema, come ad esempio la monodimensionalità o la regola di composizione, ma non arrivano a sovvertire l'orizzonte concettuale fissato nella teoria di Polya.

3.4 La concezione di Dempster-Shafer

3.4.1 La teoria dell'evidenza

Viene considerata la più influente trattazione del ragionamento sotto condizione di incertezza, essa vanta un vasto numero di applicazioni nei domini più diversi.

La teoria dell'evidenza di Dempster- Shafer si pone come scopo principale il superamento dei limiti della concezione bayesiana. Per di più propone con eleganza matematica, un modello concettuale piuttosto accessibile e chiaro di cosa si debba intendere per plausibile.

La concezione bayesiana rappresenta il punto di inizio dello studio della nozione di plausibilità della teoria Dempster-Shafer: con essa condivide il concetto secondo il quale il ragionamento plausibile sia un qualcosa incerto in quanto condotto a partire da fonti che forniscono informazioni dotate di attendibilità, ma non di certezza. Tuttavia la concezione bayesiana è una teoria monodimensionale a somma una. Essa non consente una certa credenza ad una proposizione, senza però conciliare la rimanente credenza alla negazione della proposizione. Credere in una nota misura alla proposizione A non implica dunque credere alla sua negazione nella misura della rimanente credenza. Essa ammette dunque la possibilità di fare richiesta a due distinti valori per esprimere sia la credenza in una certa proposizione sia la credenza nella sua negazione.

3.4.2 Plausibilità e credenza

La teoria Dempster-Shafer si muove da quello che di solito si definisce come una struttura di discernimento o universo del discorso u , ovvero una serie di alternative mutuamente esclusive. La teoria Dempster-Shafer associa ai sottoinsiemi di u , che rappresentano proposizioni, un'assegnazione probabilistica di base, una funzione di credenza e una finzione di plausibilità. Un'assegnazione di probabilità di base, $m: 2^u \rightarrow [0,1]$, associa ad elementi di u numeri *reali* ed è tale che $m(\emptyset) = 0$ e $\sum_{A \subseteq u} m(A) = 1$. Essa rappresenta ed esprime numericamente la forza di una qualche evidenza, l'esatta credenza che un agente ha in una posizione A .

Mediante l'assegnazione di base di probabilità, la teoria Dempster-Shafer non solo permette di distinguere chiaramente tra la nozione di credenza e la nozione di plausibilità, ma ne offre una rappresentazione in termini di funzioni probabilistiche.

Una funzione $m: 2^u \rightarrow [0,1]$ è infatti chiamata funzione di credenza, Cre , quando soddisfa le seguenti condizioni:

$$Cre(\emptyset) = 0;$$

$$Cre(u) = 1;$$

e, per qualsiasi collezione A_1, \dots, A_n di sottoinsiemi di u , si ha che:

$$Cre(A_1, \dots, A_n) \geq \sum_{I \neq \emptyset, I \subseteq \{1, \dots, n\}} (-1)^{|I|+1} Cre\left(\bigcap_{i \in I} A_i\right).$$

Una funzione di credenza permette di allocare ad ogni sottoinsieme di u , ossia una proposizione, una certa porzione della credenza totale. Dunque assegnazione di base di probabilità e funzione di credenza sono legate in modo biunivoco: ad una funzione di credenza è associata una ed una sola assegnazione base di probabilità e viceversa ad una assegnazione di base di probabilità corrispondente una ed una sola funzione di credenza. Questo legame è espresso dalle seguenti relazioni:

$$Cre(A) = \sum_{B \subseteq A} m(B), \text{ per ogni } A \subseteq B,$$

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Cre(B).$$

Esse stabiliscono che sia l'assegnazione di base sia la funzione di credenza veicolano la stessa informazione. In particolare la nozione di credenza formalizzata dalla funzione Cre ha il compito di riassumere tutte le ragioni esplicite per credere in una proposizione A .

A questo punto è possibile definire la plausibilità. Innanzitutto bisogna osservare che "una proposizione è plausibile alla luce dell'evidenza fino al punto in cui l'evidenza non sostiene il suo contrario" (Shafer 1976, 37). Quindi si introduce la funzione dubbio, Dub , ponendola come uguale a $Cre(-A)$ dove $\neg A$ è il complemento di A in u e si definisce la funzione di plausibilità, Pl , come $Pl(A) = 1 - Dub(A) = 1 - Cre(-A)$.

La funzione di plausibilità esprime dunque quanto si dovrebbe credere nella proposizione A se tutti i fatti non conosciuti allo stato attuale sostenessero A . Essa è dunque un valore probabilistico limite: esprime il valore probabilistico massimo che si può allocare verso una proposizione A : in particolare "misura la massa totale di credenza che può essere mossa verso A " (Dezert 2002, 16). Il limite inferiore corrisponde invece alla funzione di credenza, e quindi vale la relazione $Cre(A) \leq Pl(A)$. La differenza tra credenza e plausibilità si riflette anche a livello funzionale: infatti la funzione di credenza "è spesso zero per tutte o quasi le proposizioni atomiche in domini complessi, a meno che un largo numero di prove siano disponibili. Invece la plausibilità, Pl , generalmente fornisce qualche discriminazione anche quando l'evidenza è sparsa: Quindi Pl , è una guida più robusta per prendere decisioni rispetto a Cre " (Barnett 1991, 599). La funzione di credenza può così rivelarsi uno strumento efficace per determinare e massimizzare le decisioni in situazioni in cui l'informazione è incompleta e frammentaria.

L'altra grande differenza tra teoria bayesiana e teoria DS viene sancita dalla modalità di combinazione dell'evidenza, ossia della regola d'aggiornamento della credenza alla luce della nuova evidenza. Alla regola bayesiana corrisponde infatti nella teoria DS la regola di combinazione, anche detta 'somma ortogonale'. Essa "fornisce un metodo per cambiare precedenti opinioni alla luce di nuove evidenze. Essa asserisce che date due assegnazioni di base di probabilità m_1 e m_2 sullo stesso universo u , la loro somma ortogonale $m = m_1 \otimes m_2$ godrà delle seguenti proprietà:

$$m(\emptyset) = 0$$

$$m(A) = K \sum_{X \cap Y} m_1(X) \otimes m_2(Y),$$

dove $K^{-1} = 1 - \sum_{X \cap Y} m_1(X) \otimes m_2(Y) = \sum_{X \cap Y \neq \emptyset} m_1(X) \otimes m_2(Y)$, per $A \neq \emptyset$.

Se $K^{-1} = 0$, allora la somma ortogonale non esiste e dunque le fonti sono in totale contraddizione. Questa operazione inoltre soddisfa la proprietà associativa $m_1 \otimes (m_2 \otimes m_3) = (m_1 \otimes m_2) \otimes m_3$, e commutativa, $m_1 \otimes m_2 = m_2 \otimes m_1$. La regola DS, come la regola bayesiana fornisce un "metodo per cambiare precedenti opinioni alla luce di nuova evidenza" (*ibid*), tuttavia differentemente da questa, "la principale caratteristica di questa regola è la simmetria con cui essa tratta sia la nuova sia la precedente evidenza sulla quale si basa l'opinione prima della prova, che invece ha un peso maggiore nella regola bayesiana" (*ibid*). La regola di DS combina le evidenze, la regola di Bayes le pone in rapporto di condizione.

Si supponga ora di avere il seguente universo u , e la seguente assegnazione di base di probabilità:

$$\begin{aligned} u &= \{A, B, C\}, \\ m(\{A\}) &= 0.3, \\ m(\{A, B\}) &= 0.2, \\ m(\{A, B, C\}) &= 0.5, \end{aligned}$$

e si proceda ora alla determinazione dei valori di alcune delle funzioni definite in precedenza:

$$\begin{aligned} Cre(\{A\}) &= 0.3 \\ Cre(\{A, B\}) &= 0.3 + 0.2 = 0.5, \\ Cre(\{A, C\}) &= 0.3 \\ Cre(\{A, B, C\}) &= 0.3 + 0.2 + 0.5 = 1, \end{aligned}$$

da cui otteniamo

$$\begin{aligned} Dub(\{A\}) &= Cre(\{-A\}) = Cre(\{B, C\}) = 0, \\ Dub(\{A, B\}) &= Cre(\{C\}) = 0, \\ Dub(\{A, C\}) &= Cre(\{B\}) = 0, \\ Dub(\{A, B, C\}) &= Cre(\{\emptyset\}) = 0, \end{aligned}$$

e dunque

$$\begin{aligned} Pl(\{A\}) &= 1 - Cre(\{B, C\}) = 1, \\ Pl(\{A, B\}) &= 1 - Cre(\{C\}) = 1, \\ Pl(\{A, C\}) &= 1 - Cre(\{B\}) = 1, \\ Pl(\{A, B, C\}) &= 1 - Cre(\{\emptyset\}) = 1 \end{aligned}$$

Per mostrare il funzionamento della regola di combinazione si consideri invece il seguente esempio.

Si abbia un universo $u = \{(C) \text{ rescita}, (D) \text{ iminuzione}\}$, relativo al comportamento di una borsa valori B per il giorno s , e di avere due distinte fonti d'informazione f_1 e f_2 che offrono due previsioni diverse, ossia due assegnazioni di probabilità di base diverse:

$$\begin{array}{lll} m_1(C) = 0.80 & m_1(D) = 0.12 & m_1(C \cup D) = 0.08 \\ m_2(C) = 0.90 & m_2(D) = 0.02 & m_2(C \cup D) = 0.08 \end{array}$$

Da esse, per applicazione della regola di combinazione, otteniamo che:

$$K = 1 - m_1(C) \cdot m_2(D) + m_2(C) \cdot m_1(D) = 1 - 0.108 - 0.016 = 0.876,$$

e quindi

$$\begin{aligned} m(C) &= m_1 \otimes m_2(C)/K = m_1(C) \cdot m_2(C) + m_1(C) \cdot m_2(C \cup D) + m_2(C) \cdot m_1(C \cup D) \\ D)/K &= 0.72 + 0.072 + 0.064/K = 0.977, \end{aligned}$$

$$\begin{aligned} m(D) &= m_1 \otimes m_2(D)/K = m_1(D) \cdot m_2(D) + m_1(D) \cdot m_2(C \cup D) + m_2(D) \cdot m_1(C \cup D) \\ D)/K &= 0.024 + 0.096 + 0.016/K = 0.016, \end{aligned}$$

e

$$\begin{aligned} m(C \cup D) &= m_1 \otimes m_2(C \cup D)/K = m_1(C \cup D) \cdot m_2(C \cup D)/K \\ &= 0.0064/K = 0.007 \end{aligned}$$

Come è lecito attendersi, la fusione delle fonti produce una previsione che rinforza la convinzione, già espressa da entrambe singolarmente, che il giorno s la borsa B chiuderà in rialzo (C), supposto che entrambe le fonti siano ugualmente attendibili.

3.4.3 Barriere e Limiti

La regola di Dempster-Shafe realizza una sorta di unione dell'evidenza che consente di costruire descrizioni del ragionamento probabile più modeste delle descrizioni bayesiane ma più fedeli al modo in cui gli esseri umani di fatto pensano. Tuttavia essa denuncia limiti strutturali alquanto ben noti: anche in situazioni piuttosto semplici essa può infatti generare risultati completamente inaspettati e contro-intuitivi, tali da limitare fortemente l'efficacia. Si prenda infatti in esame la seguente situazione (Zadeh 1979). Due medici esaminano lo stesso paziente e sulla base degli esami esso è affetto o da meningite (M) o da commozione cerebrale (C) o da tumore al cervello (T). Tuttavia essi elaborano due diagnosi completamente diverse: infatti, pur essendo d'accordo su una scarsa probabilità del tumore, essi divergono sulle altre due. In particolare il primo medico, m_1 , ritiene che si tratti quasi certamente di meningite, il secondo m_2 , ritiene che si tratti quasi sicuramente di commozione. Si rappresenti la situazione nel modo seguente:

$$m_1(M) = 0.99 \quad m_1(T) = 0.01 \quad m_2(C) = 0.99 \quad m_2(T) = 0.01;$$

dall'applicazione della somma ortogonale ricaviamo:

$$m(T) = m_1 \otimes m_2 (T) = \frac{0.0001}{1 - 0.0099 - 0.0099 - 0.9801} = 1,$$

che significa che il paziente è affetto con certezza da tumore. Una tale previsione, del tutto inaspettata, è il risultato del fatto che "le due prove (i medici) concordano che il paziente non soffre di tumore, ma sono quasi in totale contraddizione circa le altre cause della malattia" (Dezert 2002, 20). Quindi "questo semplice ma pratico esempio mostra le limitazioni dell'uso pratico della teoria DS nel ragionamento automatico" (*idib.*) e suggerisce che "alcune estreme precauzioni sul grado di conflitto delle fonti devono essere prese prima di effettuare una decisione definitiva basata sulla regola di combinazione di Dempster" (*ibid.*). Tuttavia l'elenco delle limitazioni della teoria DS non si esaurisce con l'esempio di Zadeh (ad esempio si veda Blackman 2000), con la conseguenza che la validità e l'efficacia di questa teoria ne risente in modo essenziale, nonostante goda di applicazioni in ogni settore scientifico.

La trattazione della plausibilità e dell'inferenza plausibile sviluppata dalla teoria DS permette dunque di superare alcuni dei limiti propri della visione basata sul bayesianesimo, mediante il potenziamento di alcune delle assunzioni del sistema probabilistico, come la monodimensionalità o la regola di fusione dell'evidenza. Tuttavia essa denuncia altrettanti limiti di natura sia interna sia esterna.

Da una parte infatti, anche in situazioni piuttosto semplici, dà luogo a comportamenti del tutto "abnormi", o comunque inaspettati e devianti proprio rispetto "al modo in cui gli esseri umani di fatto pensano" (Shafer 1979, 26), ossia riguardo uno degli aspetti che dichiara di voler migliorare rispetto alla visione bayesiana.

Dall'altra essa si concentra solo sul problema dell'incertezza delle premesse, e non prende in esame il problema dell'incertezza dei metodi cercando di render conto di processi inferenziali incapaci di trasmettere la verità. Di conseguenza anche il problema del ruolo della plausibilità nel processo di formazione delle ipotesi e della loro scoperta viene escluso dalla sua trattazione.

3.5 La concezione della teoria Dezert Smarandache

3.5.1 *La teoria del ragionamento plausibile*

La teoria nasce e si sviluppa, difatti come un possibile tentativo di superare alcune difficoltà presenti nella teoria Dempster-Shafer. Questo grazie al perfezionamento e raffinamento delle assunzioni e degli strumenti propri dei modelli concettuali che l'hanno preceduta.

3.5.2 *Plausibilità e credenza*

La teoria della plausibilità di Dezert-Smarandache sviluppa brillantemente il ragionamento plausibile come un modello concettuale per guidare la presa di decisioni in condizioni di incertezza, In particolare l'incertezza a cui si riferisce soprattutto Dezert estende quella concepita dalla teoria della probabilità e della teoria Dezert-Smarandache. Uno stato d'informazione, è "razionale" quando l'assegnazione di base m è a somma uno e la chiusura degli operatori \cap e \cup sugli elementi dell'universo del discorso è pari a 0; è "incerta" in senso stretto quando è a somma uno, e la chiusura dell'operatore \cup può essere diversa da zero; è "paradossale" quando è somma uno e la chiusura dell'operatore \cap può essere diversa da zero; è "incerta e paradossale" quando la chiusura sia di entrambi gli operatori \cap e \cup può ovviamente essere diversa da zero.

3.5.3 *Barriere e limiti*

L'approccio di Dezert alla teoria dell'inferenza plausibile è molto comprensibile e presenta svariato punti in comune con la teoria Dezert-Smarandache. Precisamente si ha che il ragionamento plausibile si rivolge per trattare stati e situazioni incerte, dove sono incerte anche le informazioni. Comprende anche situazioni paradossali, per fornire previsioni che permettano di prendere decisioni. Il concetto di plausibilità viene definito come una funzione all'interno di un contesto plausibilistico. La plausibilità non è altro che una forma di credenza probabilistica, in particolare essa esprime il massimo valore di credenza che può essere allocata nei confronti di una certa congettura o proposizione A . Il suo valore viene determinato all'interno di un quadro in cui la rappresentazione dell'incertezza è bidimensionale e per la quale non vale il vincolo della condizione della somma uno. La logica dell'inferenza plausibile può essere espressa e fondata probabilisticamente: la plausibilità è infatti incorporata dalla teoria come una funzione che veicola valori di probabilità ed è in ultima analisi espressione di una opportuna variante del calcolo probabilistico.

La modellizzazione della nozione di plausibilità sviluppata dalla teoria di Dezert è piuttosto restrittiva e dagli scopi limitati. Si può infatti sostenere che la sua utilità e rilevanza attiene alla possibilità che essa offre di illustrare una latente forma di sterilità dell'approccio probabilistico all'inferenza plausibile (e le sue possibili deformazioni), piuttosto che ad un suo effettivo contributo ad una teoria del ragionamento plausibile. Essa è infatti tutta rivolta al raffinamento e alla estensione di alcune assunzioni interne alla teoria D-S, che per quanto delicate e discutibili, finiscono per impedire una visione articolata e globale dell'inferenza plausibile. Al di là del dibattito aperto sulla sua effettiva utilità reale, questa teoria soffre di molti dei limiti già insiti nella teoria D-S e si limita ad agire su alcune delle assunzioni di base di questa teoria: modifica l'universo del discorso, rompe il vincolo della somma uno dell'allocatione della credenza, si avvale di conseguenza di una regola di combinazione che tenga conto di questi fattori. La modifica delle assunzioni produce ovviamente degli effetti sul modello concettuale sviluppato, ma non permette un superamento dei suoi limiti. Sebbene si candidi ad una trattazione più estesa del ragionamento plausibile, l'approccio di Dezert non fa altro che contribuire a lasciarlo ancorato ad una visione fortemente riduzionistica.

3.6 La concezione non probabilistica

3.6.1 Motivazioni e obiettivi

Mentre la concezione probabilistica della plausibilità, finisce per ridurre il problema della plausibilità a quello della ricerca di una sua adeguata rappresentazione in termini probabilistici, la concezione non probabilistica, richiamandosi ad una tradizione diversa e partendo da assunzioni diverse, mira a sviluppare un modello concettuale della plausibilità essenzialmente diverso dalla probabilità.

3.6.2 Le deduzioni di Rescher

Tra i molteplici approcci non probabilistici alla teoria della plausibilità all'approccio deduttivista del filosofo Rescher spetta un ruolo di primo piano. Nicholas Rescher afferma che la "teoria della plausibilità punta a fornire uno strumento relativamente razionale per trattare le dissonanze cognitive" (Rescher 1976, 1). Con la forte espressione "strumento razionale" il filosofo intende qualcosa di dettagliato: infatti. L'incipit della sua trattazione è la famosa questione "dell'*attendibilità o solidità probativa* delle fonti che rilasciano o autorizzano queste affermazioni". Una volta accettata anche in via provvisoria, sulla base delle fonti disponibili, una certa tesi, la plausibilità sistematizza "il ragionamento sulle affermazioni in base all'*attendibilità o solidità probativa* delle fonti o dei principi di supporto che 'stanno dietro' di loro" (*ibid.*). La "dissonanza cognitiva" che la teoria della plausibilità è chiamata a trattare altro non è una qualsiasi situazione d'incoerenza derivante da fonti di informazioni conflittuali, di "sovra-determinazione informazionale" (*informational over-determination*).

Rescher fornisce così da un lato, un'iniziale distinzione tra la nozione di plausibilità e quella di deduzione logica, e da un altro tra quella di plausibilità e di probabilità: nella prima "classifica le proposizioni secondo lo *status* delle fonti probatorie o dei principi validanti che garantiscono a loro favore", mentre "la probabilità *pesa* varie alternative e le valuta attraverso questo relativo peso contenutistico delle considerazioni che lo sostengono". Dunque "con la probabilità ci domandiamo 'quante alternative la tesi assorbe nel suo contenuto?'; con la plausibilità 'con quanta reputazione le fonti o i principi parlano a suo favore?'. Nel primo caso ci orientiamo verso il *contenuto* della tesi, nell'altro verso le sue *credenziali probative*" (*ibid.*).

In queste situazioni sia la logica formale che la probabilità non sono efficaci. Da una parte infatti "la logica ci dice molto semplicemente che la situazione così com'è non è fattibile", "essa ci informa del fatto *che* qualcosa deve essere tolto, ma non ci è di alcun aiuto riguardo al *cosa*" (*ibid.*), dall'altra la teoria della probabilità non è in grado di risolvere problemi di incoerenze date poiché "le probabilità non possono essere modellate secondo tesi auto contraddittorie" (*ibid.*).

Per poter sviluppare una teoria non probabilistica della plausibilità, il filosofo ricorre a una variante di una regola classica della logica conosciuta come *regola di Teofrasto*, in quanto formulata dal famosostudente di Aristotele.

3.6.2.1 Le inferenze plausibili

Un sistema di inferenze plausibili soddisfa sei assiomi preliminari P1 –P6:

(P1) *condizione metrica*: dato un insieme – p, ogni proposizione $p \in S$ riceve un valore di plausibilità k , dove $0 \leq k \leq 1$.

(P2) *condizione della massimalizzazione delle verità logiche*:

tutte le verità logiche hanno un valore di plausibilità massimo (ossia pari a 1),

(P3) *condizione di compatibilità*: tutte le proposizioni indicizzate con il valore di plausibilità 1 sono mutuamente compatibili.

(P4) *condizione di conseguenza*: se un gruppo di proposizioni in S implica un'altra proposizione in S , allora questa non può essere meno plausibile di quella meno plausibile tra quelle del gruppo. Tale condizione è dunque una riformulazione del principio di Teofrasto e formalmente asserisce che: Se $P_1, \dots, P_j \vdash Q$ e P_1, \dots, P_j sono elementi di S mutuamente coerenti e $Q \in S$, allora $\min_{1 \leq i \leq j} P_i \leq Q$.

(P5) *stipulazione d'incoerenza*: sia P sia $\neg P$ possono essere contemporaneamente altamente plausibili (per esempio avere un valore di plausibilità di 0.9). Ciò viola esplicitamente la condizione di somma uno, anche se non viene meno la monodimensionalità della teoria.

(P6) *condizione di preferibilità*: in caso di conflitto tra proposizioni di plausibilità differente, bisogna dare preferenza e precedenza alla tesi di più alta plausibilità.

3.6.2.2 Plausibilità e probabilità a confronto

La plausibilità proposta da Rescher è il vero e proprio prototipo di un approccio realmente non probabilistico alla nozione così importante di inferenza plausibile.

Il filosofo invece di utilizzare la teoria della probabilità per modellare il concetto plausibilità (come proposto da tutti coloro che utilizzano una concezione prettamente probabilistica della plausibilità), basa il suo approccio sul principio di Teofrasto, proponendo così un ragionamento molto più semplice di quello probabilistico.

3.6.3. L'Entimema

Un ulteriore punto fondamentale alla nozione di plausibilità utilizzato da Rescher è il legame che esso crea con l'Entimema (detto anche sillogismo retorico) definito da Aristotele nella sua "Retorica".

Generalmente si fa riferimento a un entimema (o ragionamento entimematico) anche quando si ha un sillogismo incompleto o ellittico, nel quale una o più premesse sono sottintese perché già conosciute.

Lo scopo di Rescher è di sviluppare una teoria della plausibilità in cui le proposizioni sono candidate alla verità e la teoria della plausibilità permetta di stabilire se queste proposizioni abbiano o meno le sufficienti credenziali per poter aspirare a diventare tali.

3.6.3 Barriere e limiti

Anche la concezione di Rescher presenta dei limiti:

1. La plausibilità tratta l'incertezza delle premesse e non l'incertezza dei metodi;
2. La logica del ragionamento plausibile è una versione della logica deduttiva, e schematizza la plausibilità proprio attraverso l'operazione di chiusura rispetto all'inferenza logica attraverso un concetto che Rescher definisce "ingrandimento";
3. Le sue applicazioni sono vaste e si estendono dalla trattazione dell'informazione fino all'inferenza induttiva e arrivando al ragionamento entimematico;

Con la concezione di Rescher siamo di nuovo di fronte ad una trattazione non completa della plausibilità: essa formalizza solo una piccola parte del ragionamento plausibile. Bisogna però dire che diversamente da altre concezioni e approcci, tiene conto di trattazioni dell'inferenza plausibile anche molto diverse tra loro, come ad esempio quella di Polya.

3.7 Plausibilità e non monotonia

3.7.1 Logiche non monotone

Un altro approccio di natura non probabilistica espressamente rivolto a trattare l'inferenza plausibile è quello delle logiche non monotone: "le logiche, cioè dedicate a catturare formalmente il ragionamento plausibile". Esse sono caratterizzate dalla violazione di una delle condizioni più forti della logica classica, nota appunto come la "monotonicità", ossia se la conclusione φ è conseguenza di un insieme di premesse Γ , allora rimane una conseguenza di qualsiasi insieme di premesse Δ che contenga Γ come suo sottoinsieme. Secondo questa condizione, dunque, una conclusione non può essere invalidata dall'aggiunta di nuova informazione: una volta che si è giunti ad una conclusione φ per via deduttiva partendo da Γ , questa rimane vera una volta per tutte, indifferentemente dalle premesse che possiamo aggiungere all'insieme delle premesse Γ . Così il numero delle proposizioni valide aumenta monotonicamente rispetto all'incremento delle premesse aggiunte a Γ .

La violazione della condizione di monotonicità all'interno di linee di ragionamento permette invece di avere a che fare con situazioni in cui una conclusione può essere rivista alla luce dell'aggiunta o dell'ingresso di nuove premesse (o informazioni). Se le proposizioni possono essere invalidate dall'aggiunta di nuove premesse il numero delle asserzioni valide non solo potrebbe non aumentare, ma potrebbe anche diminuire (e comunque presentare delle fluttuazioni). Quindi "l'insieme delle credenze accettate non cresce monotonicamente. Inizialmente la conoscenza di sfondo più un insieme di credenze accettate può implicare una conclusione s . Poco dopo aver appreso che la proposizione r è vera, la conoscenza di sfondo e le credenze accettate combinate con r potrebbero non implicare più s ". (Elio 2002, 4). In tale caso l'agente deve procedere a rimuovere s dall'insieme delle credenze accettate.

L'approccio della logica non monotona offre una descrizione piuttosto precisa dell'ambito della plausibilità ed è caratterizzato dall'idea che esistono delle regole formalizzabili di questa forma d'inferenza, e che sia quindi possibile produrre una logica della plausibilità. Il ragionamento plausibile, "più tipico della vita quotidiana (ma non solo)", viene innanzitutto pensato come legato a situazioni "in cui le conclusioni sono raggiunte solo *provvisoriamente*, riservandosi il diritto di ritrarle alla luce di ulteriori informazioni", (*ibid.*) in quanto "il soggetto 'salta alle conclusioni' in base a informazioni parziali, riservandosi il diritto di rivisitare tali conclusioni quando nuove informazioni siano disponibili" (*ibid.*)

L'approccio non monotono non mette dunque in discussione la certezza dei metodi attraverso i quali si giunge alla conclusione, né tanto meno le premesse. Esso intende rendere conto di un'idea della plausibilità quale forma debole d'incertezza dei metodi, che deriva semplicemente dalla possibile presenza o emergenza di eccezioni e controesempi, che possono invalidare una conclusione precedentemente ottenuta.

3.7.2 Barriere e limiti

Quest'ultimo approccio che abbiamo affrontato va a creare un quadro abbastanza completo del concetto di plausibilità:

- 1) la logica del ragionamento plausibile è una particolare variante della logica classica.
- 2) il ragionamento plausibile è deputato a trattare situazioni d'incertezza, dove la candidatura di un'ipotesi o una conclusione sono tali poiché possono essere inficiate o riviste alla luce dell'ingresso di una nuova informazione (e quindi la conoscenza è "time-sensitive", può cambiare nel corso del tempo)
- 3) Mentre il suo principale ambito di applicazione è quello del ragionamento esperto, basi di dati, tassonomia, etc.

Le logiche non monotone modellano una concezione davvero limitata, per quanto centrale, del ragionamento plausibile. Innanzitutto, non si può non osservare, in dominio in continua espansione come quello del ragionamento plausibile, come "la spinta a catturare un numero sempre maggiore di esempio

può portare a varianti del formalismo che non soddisfano le proprietà formali desiderate, e viceversa il desiderio di mantenere un formalismo matematicamente elegante può portare a volte a perdere i fenomeni che si vogliono rappresentare” (Antonelli, 2004, 248). Tuttavia questa difficoltà, legata al soddisfacimento dell’*adeguatezza materiale* e dell’*adeguatezza formale* della teoria, è solo la punta dell’iceberg delle difficoltà delle logiche non monotoniche.

Le forme d’inferenza modellate dalle logiche non monotone sono incerte in quanto producono conclusioni provvisorie che possono essere invalidate o riviste alla luce di una nuova informazione disponibile. Come visto, la provvisorietà della conclusione candidata è generata dalla parzialità, e più in generale, dall’incompletezza delle informazioni di cui si dispone, che ha l’effetto di minare la certezza della conclusione candidata. Non a caso il suo tipico dominio dell’applicazione – e ciò che a sua volta viene definito un “tipico esempio di ragionamento plausibile” (Antonelli, 2004, 230) – sono le tassonomie o i database. Qui infatti la non completa conoscenza (l’incertezza) dell’ordine tassonomico può generare “salti” inferenziali che possono essere rivisti quando si presenta un nuovo caso che non concorda con essi. L’efficacia della modellizzazione delle logiche non monotoniche è così legata ad una delle assunzioni più forti della logica classica, nota come *assunzione del mondo chiuso*, dove “l’uso di tale terminologia si basa su un’analogia con la fisica” (Cellucci, 2002, 54), che descrive i sistemi chiusi come non aventi alcun scambio di material, di energia e informazioni con altri sistemi. Tale assunzione ha diverse formulazioni, ma qui basterà dire che è caratterizzata dall’ipotesi che l’informazione cui si dispone al momento della candidatura dell’inferenza plausibile è completa, anche se rivedibile. Quindi l’inferenza non monotona non fa altro che esplicitare l’informazione già contenuta nel sistema (la tassonomia, la base di dati, etc.), e non porta alla luce nulla di veramente nuovo. Essa stabilisce solo che sulla base delle informazioni disponibili al momento, si può ritenere l’inferenza candidata come legittima (ovvero che non esiste, allo stato attuale della conoscenza, alcuna controindicazione che ne inibisca la candidatura).

Inoltre l’approccio non monotono concepisce il ragionamento plausibile come essenzialmente rivolto al trattamento di inferenze proprie dall’esperienza quotidiana. Così facendo, esso trascura tutta quella vasta regione dell’inferenza plausibile direttamente connessa alle discipline scientifiche, compresa la matematica, che, come teorizzato esplicitamente ad esempio sia da Polya (1954) sia da Cellucci (1998 e 2002), è ricca di ragionamenti plausibili e sperimentali (ad esempio quelli analogici ed induttivi) che la logica classica, pur nelle sue varianti non permette di trattare in modo efficace.

L’unico caso di ragionamento plausibile che le logiche non monotone sono candidate a trattare con efficacia e proprietà è dunque quello in cui se si conosce a e non vi è motivo per pensare che b sia falso, allora si può inferire c . Nel caso in cui b si rivela falso, l’inferenza viene invalidata o, meno drasticamente, si può procedere ad un suo raffinamento: la regola rimane valida, fatta eccezione per le istanze note di b che la falsificano (dunque attraverso l’esplicitazione di una “lista delle eccezioni”). Ma anche in questo essa non cattura una delle principali caratteristiche dell’inferenza plausibile.

3.8 Cognitivismo e plausibilità

3.8.1 L'approccio cognitivista

L'approccio cognitivo vuole render conto degli aspetti psicologici e soggettivi che intervengono nel processo di generazione d'inferenze e ipotesi plausibili: essa evidenzia e analizza come fattori legati alla psicologia dell'agente (anche se un agente di natura "ideale" o statistico"), impegnato ad effettuare valutazioni ed elaborare ipotesi in condizioni di incertezza, possano influenzare tale processo.

L'approccio cognitivo alla plausibilità intende dunque mostrare come fattori di natura soggettiva e psicologica (come la paura, l'emozione, le debolezze), possano giocare un ruolo decisivo nel processo che conduce ad una scelta comportamentale. Ad esempio lo studio della finanza, della borsa valori o delle strategie politiche sono casi particolarmente rilevanti di domini in cui sono prodotte inferenze plausibili sulla base della considerazione dell'influenza di fattori cognitivi e psicologici.

L'approccio cognitivo alla plausibilità trova una sua modellazione per molti versi esemplare nella teoria della plausibilità di Collins- Michalski (C-M). Essa intende "formalizzare le inferenze plausibili che ricorrono frequentemente nelle risposte delle persone a domande per le quali non hanno una risposta pronta" (Collins- Michalski Collins, 1989, 1). Quindi "gli obiettivi centrali della teoria sono la scoperta di modelli inferenziali generali del ragionamento plausibile umano e la determinazione dei parametri che affettano la certezza di queste conclusioni". Essa muove così da un approccio sperimentale, ossia da una collezione di protocolli inferenziali ricavati dalle risposte di soggetti posti di fronte a problemi, per giungere alla costruzione di una sorta di "tassonomia dei modelli d'inferenza plausibile" (*ibid.*).

Il ragionamento plausibile e la plausibilità sono dunque concepiti come nozioni deputate a trattare l'incertezza che nasce dalla mancanza di conoscenza soggettiva. Questo deriva da "alcuni parametri che minano la certezza di queste conclusioni" e uno dei primi compiti della teoria della plausibilità Collins-Michalski è quello di determinare quali siano questi parametri. Per far ciò la teoria "diversamente da altre teorie del ragionamento plausibile, combina aspetti semantici con aspetti parametrici catturati da stime di certezza numeriche o simboliche" e punta al raggiungimento di tre obiettivi principali:

- -l'elaborazione di "una rappresentazione formale dei modelli d'inferenza plausibile"
- -l'individuazione di parametri (*i parametri di certezza*) "che affettano la certezza delle risposte delle persone" (*ibid.*)
- -lo sviluppo "di un sistema che metta in relazione i differenti modelli d'inferenza plausibile con i diversi parametri di certezza" (*ibid.*)

3.8.2 Concetti e basi

Per raggiungere gli obiettivi, Collins utilizza il formalismo elaborato da Michalski per lo sviluppo della sua logica a valori multipli ed individua alcuni parametri che ricorrono nel corso dei protocolli d'inferenza plausibile. Il secondo punto si risolve nell'elencazione dei parametri che esprimono il grado di certezza relativo ad una certa espressione, che può essere espresso sia numericamente sia simbolicamente.

3.8.3 La rappresentazione della conoscenza

Il problema legato alla rappresentazione della conoscenza, viene risolto in due passaggi: adottando il sistema di rappresentazione della conoscenza proprio della logica e assumendo che la conoscenza umana sia in gran parte rappresentabile attraverso gerarchie dinamiche che sono costantemente aggiornate.

3.8.4 I modelli d'inferenza plausibile

I modelli di inferenza plausibile principali trattati dalla teoria C-M sono le deduzioni plausibili (ossia il principio induttivo di Polya), le analogie e le induzioni, anche se la teoria mira a incorporare altre forme di ragionamento plausibile come le inferenze meta-conoscitive, spaziali e temporali, Collins (1978 e 1989)

indica quattro principali regole del ragionamento plausibile: generalizzazione (GEN), specializzazione (SPEC), similarità (SIM), dissimilarità (DIS); essi si riferiscono a certi contesti (CX) e operano rispetto a certe caratteristiche.

L'ipotesi che fonda la teoria C-M- è che per le persone, quali agenti più o meno razionali, tenderanno a produrre queste forme d'inferenza plausibile fino "al grado in cui essi non hanno informazioni contrarie o fanno inferenze contrarie per annullare tali conclusioni". Dunque un'inferenza è plausibile fino al punto in cui non esistono espliciti motivi (evidenze dirette o inferenze che forniscono conclusioni) che contrastino con la conclusione da essa candidata. Questo ovviamente espone questa assunzione ad una prima ed alquanto ovvia critica; infatti "un argomento che può essere addotto contro la generalità di questi modelli d'inferenza è che le persone potrebbero trarre qualsiasi tipo di conclusione assurda se seguissero, nella maggior parte dei casi, questi modelli". In questo senso si possono costruire facili esempi di inferenze fallaci basate su questi modelli: ad esempio dall'osservazione che gli uccelli volano, per generalizzazione si potrebbe concludere che gli animali o gli esseri viventi in genere volano; oppure, mediante la specializzazione, che i pinguini o le ostriche volano; oppure, per dissimilarità, che la mosca o gli insetti non volano.

Tuttavia, secondo Collins, ciò non rappresenta propriamente una difficoltà, o un limite, ma una peculiarità e un punto di forza del ragionamento plausibile. Tale ragionamento plausibile: è facile che conduca a conclusioni corrette, quando si hanno sufficienti informazioni per andare avanti. L'importante è perciò disporre di informazioni strutturate sul contesto.

Anche il punto tre viene risolto adottando il sistema di calcolo logico a valori variabili elaborato da Michalski il quale evidenzia anche l'esistenza di uno spazio all'interno del quale le inferenze vengono condotte, in particolare uno spazio che Collins (1978) definisce a cinque dimensioni:

- 1) *inferenze sulla conoscenza – inferenze su meta – conoscenza*
 - ossia è possibile distinguere tra inferenze condotte basandosi sulle proprie conoscenze come l'analogia o la deduzione – e inferenze condotte basandosi sulla conoscenza della propria o altrui conoscenza (appunto meta-conoscenza).
- 2) *inferenze funzionali – inferenze insiemistiche*
 - per ogni tipo di inferenza è possibile distinguere una variante funzionale e una variante insiemistica. La seconda permette di trasferire la proprietà di un insieme su un altro insieme. La prima permette lo stesso trasferimento, sotto la condizione aggiuntiva che la proprietà in questione (la variabile dipendente) dipende a sua volta da un'altra proprietà (la variabile indipendente): essa è dunque una sorta di "correlazione direzionale" (*ibid.*)
- 3) *inferenze semantiche, spaziali, temporali*
 - per ogni tipo di inferenza Collins delinea tre possibili varianti: la semantica (ossia il "trasferimento di proprietà lungo uno spazio semantico"), la spaziale (ossia il trasferimento di una proprietà "lungo uno spazio euclideo"), la temporale (ossia il trasferimento di proprietà "lungo il tempo"). Questa distinzione viene operata in vista dell'elaborazione di una teoria computazionale, visto che le tre forme di inferenza implicano procedure per molti versi differenti.
- 4) *insiemi simili, superordinati-insiemi subordinati*
 - le inferenze sono le deduzioni, induzione, generalizzazione, analogia e abduzione, che implicano il trasferimento di proprietà da insiemi superordinati, simili o subordinati.
- 5) *inferenze negative- inferenze positive*

- Ogni inferenza ha una versione negativa e una positiva, a secondo del fatto se la conclusione candidata implica rispettivamente l'assenza o la presenza di una proprietà.

3.8.5 *Barriere e limiti*

- 1) La teoria dell'inferenza plausibile è solamente di natura descrittiva.
- 2) Il suo ambito di applicazione è quello del ragionamento quotidiano, condotto per via verbale.
- 3) La logica del ragionamento plausibile si articola attraverso la candidatura di conclusioni secondo i modelli d'inferenza plausibile (deduzioni plausibili, analogie e induzioni), fatto salvo che non esistono evidenze o conclusioni contrarie alla conclusione candidata.
- 4) Il ragionamento plausibile ha le sue regole che non possono essere ridotte al ragionamento probabilistico o a quello probabilistico.
- 5) Secondo Collins-Michalski si ha che Il ragionamento plausibile ha il compito di trattare situazioni d'incertezza, nelle quali la psicologia del soggetto chiamato a fare valutazione gioca un ruolo decisivo nel corso del ragionamento. L'incertezza deriva dunque dalla mancanza di conoscenza soggettiva, e viene espressa mediante opportuni parametri associati alla ipotesi.
- 6) Un' ipotesi è plausibile se è il risultato dell'applicazione di uno dei modelli d'inferenza plausibile nello spazio delimitati da "cinque dimensioni", fino al punto in cui non esistono espliciti motivi. Quindi si può generalmente sostenere che nella concezione di Collins "qualcosa è plausibile se è concettualmente sostenuta dalla conoscenza precedente" (Connell e Keane 2003, 264). Il fine esplicito della teoria C-M è l'elaborazione di una tassonomia dei modelli del ragionamento plausibile che sia descritta nel modo in cui si articola il ragionamento umano quotidiano in condizioni di incertezza (derivante da una conoscenza di base incompleta e dall'influenza di alcuni parametri cognitivi), sulla base della quale sviluppare un modello computazionale.

Il perseguimento di questo obiettivo avviene tuttavia attraverso un forte schematismo e una semplificazione dei modelli d'inferenza plausibile.

Nel caso delle illusioni psicologiche è provato (Tversky e Kahneman e Slovic 1982) che atteggiamenti emotivi che intervengono nel corso di inferenze e valutazioni, come la 'minimizzazione del rimorso', ne modificano il corso e si rivelano decisivi nella candidatura della scelta finale. Esiste una lunga casistica che mostra come ad esempio (Paulos 2003, 29) di fronte a scelte del tutto identiche, si decida di optare per una sulla base di fattori meramente emotivi e psicologici come l'avversione al rischio. In base a questo, il desiderio di ridurre al minimo il rimborso, ad esempio, nei riguardi di possibili perdite, genera processi decisionali e valutativi distorti, se non paradossali, negli investitori. Ancor più paradossali possono apparire altri fenomeni indotti dalle illusioni cognitive, come le credenze che si auto-avverano. Esse sono fenomeni che generano casi in cui l'illusione cognitiva, la convinzione legata ad una certa realtà, finisce per generare quella realtà. Ad esempio, nel caso del mercato azionario, se un numero sufficiente di investitori ad un certo punto crede nel potenziale di una data azione, ossia nel fatto che il suo prezzo stia per aumentare, allora il prezzo dell'azione salirà solo per quest'unica ragione, giustificando questa credenza. Infatti gli investitori inizieranno a comprare quelle azioni, credendo che stia per aumentare e quindi spinti dalla speranza di poterla poi rivedere quando il prezzo sarà alto al fine di trarne degli utili, e ciò farà lievitare il prezzo, proprio come la loro credenza supponeva. Certo non sempre questa credenza sarà soddisfatta, poiché esiste una "soglia" che deve essere raggiunta affinché il meccanismo si metta in moto (nell'esempio addotto sopra, la soglia è un "certo" numero di investitori).

Questi comportamenti paradossali della credenza nascono dal fatto che le inferenze che possono essere condotte sulle credenze coinvolgono non solo le credenze stesse, ma le credenze sulle altrui o proprie credenze, e così via, ovvero su ciò che si definisce la conoscenza comune. Affinché sia possibile che una conoscenza reciproca divenga una conoscenza comune, infatti, è necessaria la presenza di un "arbitrio indipendente" (ibid, 17) rispetto ai vari agenti che compiono inferenze e condividono informazioni (che nel

caso del mercato azionario può essere, ad esempio, la Sec, l'organo di regolamentazione e vigilanza della borsa statunitense).

Queste forme di ragionamento danno così luogo a modelli d'inferenza plausibile di larga diffusione che pur dipendendo da fattori cognitivi e psicologici non sono inclusi, o riconducibili, alla teoria di Collins-Michalsky. Essa finisce perciò per tralasciare un'intera area all'interno della regione del tipo ragionamento plausibile che intendeva esplicitamente trattare e modellare.

Il grado di incertezza della conclusione (CC) dipende dal grado di certezza delle informazioni di premessa (CP) e del grado di certezza dei modelli, cioè dalla certezza di inferenza CI; formalmente si ha che:

$$cc = f(CP, CI)$$

Con CC indichiamo il grado di incertezza della conclusione.

Con CP invece indichiamo il grado di certezza delle informazioni di premessa.

Le decisioni basate sulla plausibilità vivono in un iperspazio strutturato le cui iper-dimensioni sono: Fisico (dove, quando), logico (come), funzionali (perché).

3.9 Verso una teoria del ragionamento plausibile

3.9.1 Sistemi aperti e reti

Le teorie dell'inferenza plausibile nascono e si sviluppano in relazione ad un preciso dominio che intendono modellare: i sistemi esperti, il calcolo probabilistico, le tassonomie, le reti bayesiane, sono solo alcuni esempi di come di volta in volta l'idea di plausibilità, e dei suoi modelli, emergano da problemi e dinamiche che sono proprie di specifici ambiti. Tutti gli approcci che abbiamo analizzato sono perciò, in modo più o meno diretto, il risultato di una trattazione della plausibilità che dipende fortemente dal dominio di riferimento: la specificità dell'ambito al quale si richiamano finisce per influenzare e in molti casi determinare non solo i modelli d'inferenza plausibile, ma le assunzioni e gli assiomi della teoria stessa. Di conseguenza, i vari approcci non affrontano il problema della plausibilità in tutta la sua estensione e complessità, ma finiscono per proporre, più semplicemente, una propria visione della teoria dell'inferenza plausibile, nella fattispecie quella che si rivela più efficace nel trattare i problemi del dominio a cui si riferiscono.

La teoria dei sistemi concettuali aperti fornisce importanti indicazioni su come sia possibile costruire razionalmente ipotesi plausibili e come le sue dinamiche possano gettare nuova luce sui processi inferenziali propri della costruzione e formazione delle ipotesi. In quanto basati sull'assunzione del mondo aperto, questi sistemi hanno caratteristiche e strumenti completamente diversi, e nuovi, rispetto ai sistemi formali classici basati sull'assunzione del mondo chiuso. In molti casi essi hanno superiori capacità inferenziali rispetto al metodo assiomatico e permettono di affrontare razionalmente il problema della formulazione delle ipotesi e della scoperta.

La teoria delle reti, grazie ai suoi recenti sviluppi può essere guardata come una valida fonte di analisi e indagine dell'inferenza plausibile. La rappresentazione della conoscenza sotto forma di grafo non solo non è affatto inusuale, ma è in grado di rilevare aspetti essenziali circa la natura e le dinamiche delle forme d'inferenza plausibile. Tale rappresentazione gioca un ruolo decisivo nella teoria delle reti, i cui risultati mostrano come esistano semplici configurazioni e dinamiche topologiche che sono in grado di rilevare proprietà essenziali dei fenomeni indagati.

La rappresentazione della conoscenza (quella a reti) è molto semplice, ed è formata da due solo elementi costitutivi: i nodi (rappresentati convenzionalmente dal simbolo*) e i link (o collegamenti, rappresentati

convenzionalmente da segmenti o archi). I nodi rappresentano gli elementi base di un dominio indagato (ad esempio possono rappresentare ipotesi, proposizioni, agenti o formiche); i link rappresentano le relazioni esistenti tra nodi (ad esempio le relazioni di implicazione, affinità, parentela). Nodi e link danno luogo a grafi, che combinandosi creano le reti in senso stretto. Le reti, a loro volta, possono presentarsi secondo grafi di due tipi: i grafi orientati e i grafi non orientati. Essi si distinguono sulla base dei collegamenti tra i vari nodi che, rispettivamente, possono o non possono avere un verso.

La teoria delle reti ha permesso di cogliere l'esistenza d'alcune proprietà che ricorrono, in modo quasi ubiquitario, in domini del tutto eterogenei: il Web, Internet, le relazioni sociali, solo per fare alcuni esempi, presentano caratteristiche, dinamiche e comportamenti del tutto identici (come la legge di potenza, l'auto-similarità, etc.). Tuttavia, ogni dominio ha rilevato l'esistenza di proprietà peculiari, che hanno indirizzato la ricerca verso nuove direzioni. L'applicazione della teoria delle reti allo studio dell'inferenza plausibile può dunque stimolare la ricerca in almeno due direzioni: da una parte può mostrare come, e se, tali inferenze possano comportarsi secondo i principi della teoria delle reti; dall'altra, può mostrare come possano presentare dinamiche peculiari, in grado di fornire alla teoria delle reti stessa nuovi modelli e comportamenti su cui indagare.

Ai fini dello studio dell'inferenza plausibile, esistono alcune proprietà reticolari particolarmente rilevanti e interessanti. In particolare esse sono:

- la connettività
- La robustezza

La prima ha varie definizioni, adatte agli specifici domini applicativi rappresentati dai grafi; in generale possiamo dire che misura quanto l'intero grafo sia collegato, dove due nodi si dicono collegati se esiste una sequenza di archi che li collega, e un grafo è collegato se tutti i nodi sono collegati, ovvero se non esistono 'isole; per fare alcuni esempi, in un grafo non orientato il grafo si definisce connesso se da ogni nodo esiste un *cammino* (sequenza di archi) verso ogni altro nodo; in un grafo orientato si ha la connettività 'debole', che consiste nel considerare orientati gli archi e verificare se il grafo non orientato ottenuto è connesso, e la connettività 'forte', nella quale da ogni nodo deve esistere un cammino orientato (ovvero una sequenza di archi con lo stesso verso) verso ogni altro nodo.

La robustezza è una forma di resilienza, "una capacità di autoaggiustamento" (Barabasi 2002, 121), di tolleranza all'errore ed è definita, in funzione della definizione di 'connettività' e del numero di nodi che devono essere rimossi in modo casuale per *disconnettere* il grafo.

Queste proprietà permettono dunque di concepire, definire e analizzare la plausibilità in termini di connettività e di robustezza topologica. Si può infatti considerare un'ipotesi come 'plausibile' quando è connessa alla conoscenza esistente ed è dotata di una capacità di auto-aggiustamento, ossia di robustezza (resistenza all'errore): essa sarà tanto più plausibile quanto più alto è il suo grado di connessione e di resilienza rispetto alla conoscenza esistente. La connettività misurerà il suo grado di integrazione e di compatibilità con la conoscenza esistente, la robustezza quanto ne è supportata e sostenuta.

Tale costruzione richiede che i nodi e link non siano solo di natura "mono-dimensionali" (ossia rappresentanti entità appartenenti a classi omogenee), ma devono poter essere in grado di esprimere entità eterogenee.

I nodi che costituiscono la rete possono essere di natura eterogenea, ossia rappresentare elementi appartenenti a classi non omogenee (ad esempio possono rappresentare sia per proposizioni o teoremi noti, sia congetture o ipotesi dallo statuto provvisorio ed incerto); i nodi possono inoltre essere "pesati", come avviene nella trattazione di Rescher e Walton della plausibilità, ossia ricevere valori (convenzionalmente espressi da numeri reali compresi tra 0 e 1) che ne esprimono l'attendibilità;

Le relazioni che legano i nodi, i link, possono essere di natura eterogenea, e rappresentare simultaneamente questi diversi tipi di relazione (ad esempio: ipotesi e proposizioni – rappresentate da nodi

– possono essere legate da relazioni di implicazione, equivalenza, similarità, etc); anche i link, come i nodi, possono a loro volta essere “pesanti”.

3.9.2 Reti di inferenza plausibile

Bisogna osservare che il successo di un tale approccio dipende in modo essenziale da un’adeguata rappresentazione della conoscenza in forma reticolare, e in particolare dal *grado di granularità* scelto. Per grado di granularità si intende il dettaglio con il quale i singoli nodi modellano il dominio. All’aumentare del grado di dettaglio si ha, ovviamente, una maggiore precisione della descrizione del fenomeno indagati a scapito della leggibilità e maneggevolezza della rete.

Le inferenze plausibili, differentemente da quanto sostenuto in alcuni dei principali approcci di natura sia probabilistica sia non probabilistica, possono presentarsi in forme articolate e complesse, composte da molti passi e quale frutto dell’applicazione di regole diverse (deduzioni, induzioni, analogie), che danno luogo a grafi altrettanto articolati e complessi. L’analisi di questi modelli reticolari è destinata ad offrire un contributo rilevante ad un progresso della teoria della plausibilità.

La plausibilità di un nodo è una questione di natura duplice: è possibile infatti definire la plausibilità in termini di proprietà locali o globali. In particolare un nodo può essere tanto plausibile se ha numerosi archi incidenti; oppure può essere tanto plausibile se, a partire da nodi rappresentanti ipotesi plausibili, esistono diversi *cammini* che conducono ad esso. La questione ora è dunque quali siano le implicazioni di queste due diverse definizioni della plausibilità.

Si rende pertanto necessario l’utilizzo di strumenti di analisi standard della teoria delle reti, che sono in grado di individuare, ad esempio, l’insieme dei nodi più raggiungibili a partire da un insieme di premesse e i nodi con maggiore resilienza.

L’idea di utilizzare la teoria delle reti per render conto della plausibilità è quindi di sicuro interesse e promettente, visti anche i risultati raggiunti in ognuno dei domini ai quali è stata applicata: anche la semplice assenza di risultati, in questo senso, può già costituire di per sé un risultato che permetterebbe di evidenziare alcune proprietà specifiche e sorprendenti del ragionamento plausibile.

3.10 Decisione in Domini di info-Incertezza e/o di info-Incompletezza

3.10.1 Lo scenario

Nell'era della conoscenza e dell'intelligenza collettiva, così come in passato, siamo chiamati a prendere decisioni. Generalmente prendiamo decisioni per risolvere un dato problema (problem solving) ovvero per raggiungere un dato obiettivo (decision making).

E' di tutta evidenza, inoltre, che un dato obiettivo può essere raggiunto concatenando decisioni "elementari" successive. Ciò pone l'attenzione su una questione importante, ovvero che nei moderni sistemi di supporto alle decisioni che operano e sono impiegati in ambito strategico o comunque in ambiti dove è richiesta una alta complessità, diventa non solo importante l'individuazione degli obiettivi spesso fissati da un operatore/agente esperto, quanto l'individuazione del path decisionale, ovvero il susseguirsi di sotto obiettivi parziali la cui combinazione/concatenazione conduce all'obiettivo generale finale attraverso un percorso che definiamo ottimo.

Se fossimo in un contesto fisico in cui abbiamo un dato sistema materiale che deve transire da uno stato A a uno stato B collegabili attraverso n finiti stati intermedi ed m possibili traiettorie decisionali sarebbe facile comprendere e stimare subito la complessità del problema in studio e ridurlo attraverso e grazie ai principi di minimizzazione di variabili di interesse, come l'energia spesa per la transizione o il disordine generato nella transizione o ancora la minimizzazione del tempo impiegato.

Se fossimo in un contesto logistico dei trasporti, di merci o persone e l'obiettivo è lo spostamento dalla posizione A a quella B, tale obiettivo potrebbe raggiungersi massimizzando le quantità trasportate in un dato intervallo temporale ecc.

Questi semplici esempi e le loro possibili generalizzazioni o estensioni portano a concludere che una volta definito un problema schematizzabile in 1) condizioni iniziali, 2) vincoli, 3) obiettivo, 4) strategia di soluzione, 5) ottimizzazione della soluzione qualora vi siano più soluzioni possibili o anche una soluzione obiettivo raggiungibile attraverso diverse strategie competitive.

Una prima classificazione del problema è se esso è deterministico o se esso è stocastico. Ciò vuol dire che le variabili da cui dipende il problema in studio possono essere descritte attraverso leggi dinamiche definite o se invece ci sono dinamiche che sfuggono al controllo dell'analista.

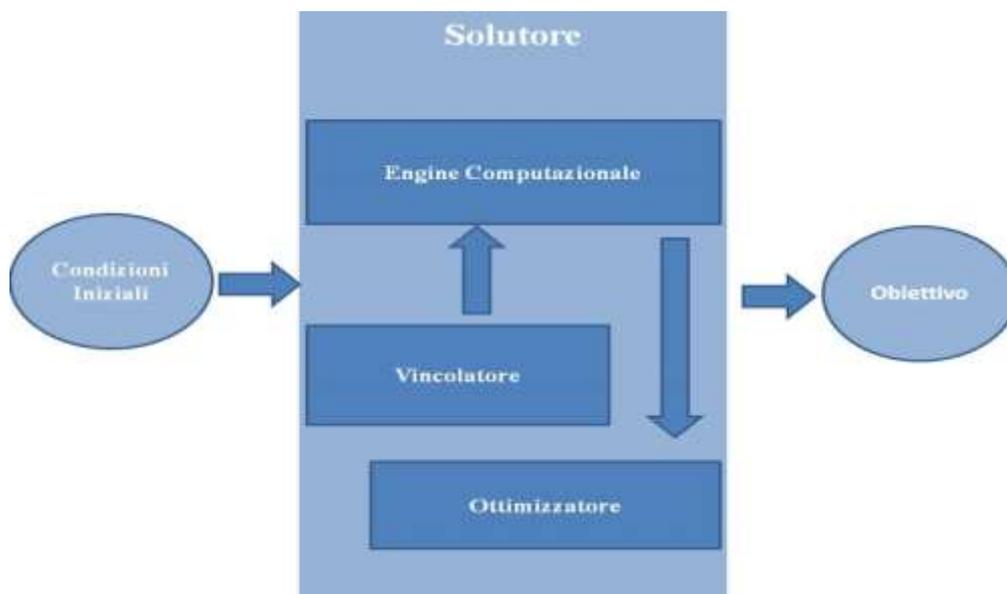


Figura 2: Grafico di esempio per dati incompleti

Tale indeterminazione può sia riguardare la singola variabile in sé sia il modo in cui essa influenza l'obiettivo. Nel primo caso abbiamo una indeterminazione diretta (indeterminazione della causa) cioè che riguarda proprio una variabile del problema, nel secondo caso invece abbiamo una indeterminazione derivata o di effetto (indeterminazione dell'effetto). A titolo esemplificativo si supponga che un utente debba recarsi dalla posizione A alla B utilizzando un autobus. Una indeterminazione di tipo A riguarda quando arriverà l'autobus, quindi abbiamo una indeterminazione di tipo temporale, ovvero sulla variabile input tempo.

E' di tutta evidenza però che il raggiungimento della posizione B può dipendere dal traffico che l'autobus incontrerà anche dopo aver raggiunto il passeggero e starà percorrendo il tratto di strada dal passeggero verso la posizione B, che rappresenta appunto una indeterminazione dell'effetto, data dal vincolo traffico o coefficiente di riempimento della strada. Per completezza dell'esempio è evidente il vantaggio di avere più percorsi possibili e la conoscenza dello stato di riempimento/traffico dei diversi percorsi al fine di minimizzare il tempo di raggiungimento della posizione B, magari pagando un costo di percorso, corrispondente ad un viaggio più lungo in termini di spazio percorso.

Un tale esempio seppur nella sua singolarità ci permette di comprendere come sia facile avere a che fare e quindi dover maneggiare problemi non deterministici a priori ma che magari possono essere ricondotti a deterministici grazie ad una serie di scelte (vincoli)

Il punto è come scegliere in tali condizioni? Come decidere di seguire un percorso piuttosto che un altro? La risposta a tale domanda conduce ad un altro dualismo, cioè info-completezza verso info-incompletezza. Infatti dall'esempio precedente è facile comprendere come la conoscenza del percorso con minor traffico permetta all'autista di decidere facilmente circa quale strada seguire per raggiungere la posizione B, obiettivo del viaggio, nel minor tempo possibile, trasformando così un problema non deterministico in uno deterministico grazie ad una scelta basata sulla conoscenza dello scenario di azione.

Abbiamo quindi introdotto un'altra caratterizzazione dei sistemi complessi e della teoria della decisione, ovvero la completezza - incompletezza informativa.

Da queste tre caratterizzazioni, cioè:

- i. determinismo, indeterminismo delle variabili,
- ii. determinismo indeterminismo del legame funzionale delle variabili e del relativo impatto sull'obiettivo,
- iii. completezza, incompletezza dello scenario di partenza (variabili di input), dello scenario di arrivo (variabili di output) e dei vincoli (modelli funzionali che legano gli input agli output), discende la stocasticità, ovvero il non determinismo di molti problemi nella maggior parte degli ambiti conoscitivi in cui si rende necessario prendere decisioni.

Venendo al contesto finanziario, considerato un dato strumento finanziario in cui l'obiettivo è il raggiungimento di un dato prezzo obiettivo, è evidente che l'indeterminatezza di quando ci sarà un tick, ovvero dell'istante della scala temporale in cui ci sarà un movimento del prezzo è una indeterminazione di una variabile del I tipo di quelle descritte, mentre la successione di tick è del II tipo, così come il non conoscere a quel dato tick quali volumi vengono scambiati e quindi di quanti punti varierà il prezzo è una incompletezza informativa, corrispondente alla III caratterizzazione suddetta.

Tutto ciò rende, quindi, stocastica la dinamica dei prezzi, dello strumento finanziario considerato.

Cosa fare?

Per rispondere a tale quesito è importante precisare che la parola stocastico, oggi non significa casuale, ciò poiché esistono o si possono costruire modelli che permettono di modellare la stocasticità e quindi di fare previsioni. In altre parole, traendo spunto dall'esempio precedente, non saremo in grado di dire quale sarà il prezzo esattamente che verrà raggiunto ma saremo in grado di dare una stima più o meno precisa in base ai vincoli noti o alle esigenze dell'analista, ovvero saremo in grado di prevedere esattamente il prezzo ma non ad un dato istante, bensì prevedendolo in un dato intervallo temporale. E' di tutta evidenza che tali

risposte risolvono molti dei problemi suddetti, grazie alla modellazione di sistemi dinamici stocastici, seppur non forniscano una risposta esatta o perfettamente deterministica.

Prima di rispondere alla domanda su “cosa fare”, resta da trattare un’ultima questione che riguarda la corretta posizione del problema su cui si vuole decidere ovvero per il quale si intende individuare una strategia decisionale. Tale problema riguarda la certezza dell’obiettivo. In riferimento agli esempi suddetti dobbiamo chiederci: il prezzo considerato come obiettivo è raggiungibile? La posizione A è normalmente servita dalla linea di autobus che stiamo prendendo in considerazione? ecc. Detto in altre parole l’obiettivo O ha la possibilità di accadere con certezza o è impossibile?

In tale considerazione appena fatta è emerso il dualismo certezza-incertezza dell’obiettivo/risultato con gli estremi, certo che vuol dire sicuro ed impossibile che vuol dire non raggiungibile a priori.

Di fatto, però, da un punto di vista linguistico sono emersi altri termini che necessitano di un approfondimento formale e che sono: possibile, plausibile, e possiamo aggiungere credibile, ed ultimo, ma non meno importante anzi forse quello più formalizzato, è il termine probabile.

In sintesi, il nostro studio sta evidenziando che i problemi di decisione nell’ambito della complessità sono generati dai dualismi:

determinismo, indeterminismo,
completezza, incompletezza,

ciò generano il dualismo:
certezza, incertezza

dell’obiettivo con i due estremi di soluzione evento certo, evento impossibile e la realtà che normalmente è nel mezzo, inteso come evento incerto.

Il dualismo dell’incertezza viene affrontato con lo studio e la quantificazione delle dicotomie:

plausibile – implausibile
credibile – incredibile
probabile – improbabile
possibile – impossibile
affidabile – inaffidabile
attendibile – inattendibile

Il nostro lavoro dovrà essere quindi di capire come utilizzare, quantificare e generalizzare tali dicotomie al fine di costruire, se possibile, un sistema di supporto alle decisioni non basato semplicemente sulla probabilità.

Qualcuno a tal punto potrebbe risultare confuso e chiedersi il perché di chiamare in causa tanti concetti diversi e non affidarsi semplicemente alla teoria della probabilità o ai processi bayesiani.

Il motivo è molto semplice perché alcuni fenomeni, la maggior parte di quelli che hanno a che fare con la complessità o i sistemi complessi non si distribuiscono in modo gaussiano o normale, o secondo altre distribuzioni note, ma piuttosto come si è soliti dire hanno delle code pesanti, cioè non gaussiane, nella distribuzione di probabilità o della data quantità che stiamo considerando.

Un esempio è la distribuzione dei prezzi, di un dato strumento finanziario, intorno ad un prezzo preso a campione in cui si osserva nella distribuzione dei prezzi intorno al campione una distribuzione non normale, ovvero con delle code pesanti, cioè più grandi delle code gaussiane o normali. Che vuol dire ciò? Ciò vuol dire che seppur improbabile alcuni prezzi di un dato strumento finanziario occorrono, cioè si verificano; ecco perché diventa importante in una teoria di stima dei prezzi chiedersi quanto quel prezzo sia plausibile, credibile, possibile ed infine quale sia l’affidabilità che siamo in grado di esprimere sulla stima o l’attendibilità della previsione che facciamo.

In questi termini il problema della decisione è emerso in tutta la sua complessità. Pertanto, si comprende perché spesso l'operatore, da razionale, cioè colui che decide secondo schemi predefiniti o preordinati diventi affettivo – emotivo, cioè lasci entrare le emozioni (il sentiment) nelle decisioni, soggettivando le scelte.

In altre parole l'info-incompletezza o l'indecisione dell'input generano emozioni che provocano credenze soggettivando le scelte nelle decisioni da prendere.

Lo scopo o il ruolo di un sistema di supporto alle decisioni basato sulla complessità deve essere quello di costruire scenari simulati, analoghi a quelli che possono ottenersi con la teoria dei giochi, grazie ai quali le diverse soluzioni/obiettivi e le diverse strategie decisionali competono, cooperano (nel caso multi step decision) o concorrono alla scelta migliore, dove la parola migliore si declina in base ad una serie di parametri di interesse del decisore (vincoli su come si desidera sia l'obiettivo finale considerandolo comunque già raggiungibile di per sé).

Con l'ausilio di tale metodologia e relative istanze tecnologiche, un operatore emotivo – affettivo sarà ricondotto a decidere razionalmente (operatore razionale indotto) grazie al supporto di scenari virtuali simulati che anticipano l'impatto delle soluzioni e permettono di decidere ora con la cognizione di quando si avrà raggiunto il risultato (decision forward in time).

Quindi il meccanismo di decisione da considerare per “rirendere” un operatore emozionale un operatore razionale, deve tenere in conto che l'incompletezza informativa, cioè la mancanza di conoscenza, l'indeterminazione delle informazioni/variabili o dei loro legami funzionali (cosiddetti meccanismi di azione) generano emozioni nel decisore e le emozioni inducono credenze.

Pertanto proprio la destrutturazione delle credenze – emozioni grazie allo studio che ci accingiamo a realizzare, ovvero altri approcci come la “sentiment analysis” o le analisi emozionali, possono permettere di gestire tali credenze (false realtà o realtà suggestive) e ricondurre l'operatore a essere di nuovo razionale ma su una base cognitiva più ampia (conoscenza espansa, conoscenza aumentata, iper-conoscenza).

Il momento di nuova razionalizzazione porterà allo stato che possiamo definire di ipercoscienza, cioè l'operatore di nuovo razionale è cosciente del nuovo scenario allargato, che chiamiamo iperscenario.

3.10.2 Il continuo verificarsi dell'improbabile possibile

Quante volte abbiamo assistito al verificarsi di eventi altamente improbabili?

Quante volte la probabilità di successo ci dava scientificamente vincitori e invece abbiamo perduto?

Quante volte era plausibile che un match calcistico si concludesse in un modo ma invece poi le cose sono andate diversamente?

E ancora quante volte la dinamica dei prezzi di uno strumento finanziario ci lascia credere di aver compreso la direzione dominante, ma poi al nostro ingresso il mercato ci mostra un comportamento diametralmente opposto da quel momento in poi?

Questi sono solo alcuni esempi di quanto sia complesso fare previsioni in condizioni di incertezza (incertezza delle premesse, dei modelli, dei risultati) o di incompletezza delle informazioni, che detta in termini diversi ma più realistici e forti potremmo definire “decidere in condizioni di ignoranza”, ovvero in condizioni di non conoscenza della realtà.

E' proprio in tali condizioni, cioè superato il determinismo ed entrati nel contesto dei fenomeni stocastici o dell'incertezza che troviamo i termini quali: probabile, plausibile, possibile, credibile.

In letteratura ci sono diverse linee di ricerca e centinaia, meglio dire migliaia di lavori atti a descrivere, distinguere e modellare tali termini, talvolta anche in modo conflittuale.

In questo studio, invece, presentiamo una visione organica graduata degli stessi, a partire dal concetto meglio modellato, ovvero quello probabile, ed utilizzando i metodi e i modelli della probabilità per meglio descrivere gli altri concetti laddove ciò sia possibile.

In altre parole, mentre autori come György Pólya⁶ o Arthur Dempster - Glenn Shafer⁷ hanno cercato di estrarre, ad esempio dalla sua componente descrivibile in termini probabilistici, in questa sede accettiamo la differenza concettuale tra tali termini, li gerarchizziamo in termini di forza e li descriviamo con i metodi utilizzati per descrivere la probabilità che tali metodi siano più generali, come di fatto sono, e non di esclusivo appannaggio della probabilità. OB 16/05

Nella nostra visione i concetti di:

- probabile (P_r)
- plausibile (P_l)
- credibile (C_r)
- possibile (P_o)

hanno una forza decrescente, seppur descrivono tutti eventi incerti, in altre parole un evento possibile non è detto che sia credibile, così come un evento credibile è certamente possibile, analogamente un evento credibile non è detto che sia plausibile, mentre è certamente possibile; ancora un evento plausibile non è detto che sia probabile, ma è certamente credibile e possibile; infine un evento probabile è sicuramente plausibile, credibile e possibile.

E' quindi necessario indicare una funzione caratteristica multipla per un dato processo o fenomeno, dove per multipla intendiamo ottenuta dalla composizione di più funzioni definite:

- F1 funzione di probabilità (o meglio distribuzione di probabilità)
- F2 funzione di plausibilità
- F3 funzione di credibilità
- F4 funzione di possibilità

dove le F sono funzioni monodimensionali a somma 1, ma multiscala.

Ciò vuol dire che se un evento è improbabile, ad esempio $P(E) \leq 1\%$, altamente improbabile $P(E) \leq 0,1\%$, assolutamente improbabile $P(E) \leq 0,01\%$, allorquando andiamo a considerare quanto esso sia plausibile, cioè valutiamo la plausibilità dell'evento E, $PI(E)$ inteso come quantità tale che $0 \leq PI(E) \leq 1$ la di probabilità data in precedenza ... improbabile, altamente improbabile o assolutamente improbabile diventa di scarsa rilevanza, poiché non ci muoviamo più in un regime probabilistico ma plausibilistico, in cui emerge una nuova scala di valutazione che è quella del plausibile di eventi improbabili.

Un ragionamento analogo può essere condotto per il concetto di credibilità e di possibilità.

Ciò può riassumersi nella seguente asserzione di provocazione concettuale:

“Quando l'improbabile è altamente possibile non c'è da meravigliarsi che si realizzi”.

Formulata alla stregua di una legge di Edward Murphy⁸ a tutti dovrebbe risultare quasi come una somma verità scientifica.

Il prezzo attuale $p(t)$ è una forma di uniformazione certa, mentre il trend $p(t-1) - p(t)$ è una fonte più o meno attendibile ma non certa.

Usare la teoria DS rispetto ad un modello bayesiano permette di superare il limite della multidimensionalità a somma uno.

Detto in altre parole, la plausibilità che il prezzo cresca con una certa probabilità soggettivamente stimata, non implica automaticamente che la negazione della crescita, ovvero la decrescenza abbia una

⁶La teoria di Pólya

⁷La teoria di Dempster e Shafer gestisce la distinzione tra incertezza e ignoranza e calcola l'evidenza di una proposizione, ossia la probabilità che essa sia supportata da prove: questa misura è chiamata funzione di credenza e si scrive $Bel(X)$.

⁸La Legge di Murphy è un insieme di paradossi pseudo-scientifici a carattere ironico e caricaturale realizzati da Arthur Bloch con frasi didascaliche in forma statistico-matematica.

probabilità pari al complemento a uno, cioè che la probabilità di un trend decrescente, o ribassista, sia uguale a uno meno la probabilità assegnata al trend rialzista.

Pertanto, grazie a DS è possibile dare all'evento E una probabilità e una negazione di E, cioè $P(E)$ e $P(\neg E)$, un'altra probabilità diversa da uno meno $P(E)$ con il vincolo comunque che $P(E) + P(\neg E) \leq 1$.

Nella teoria DS, $(E) \leq (PI)(E)$ cioè il concetto di credibile è meno forte di plausibile, cioè poiché la plausibilità è intesa il valore probabilistico massimo che si può allocare sulla proposizione E, mentre la credenza è il valore probabilistico minimo associabile ad E, ma anche e soprattutto perché con plausibile intendiamo qualcosa di credibile per più persone.

In questo lavoro intendiamo effettuare una analisi che parte da una prospettiva diversa e che emerge naturalmente dagli studi che si stanno effettuando nell'ambito dell'intelligenza artificiale allorquando si desidera estendere la capacità decisionale di un'intelligenza artificiale che operi in condizioni di incertezza, con capacità cognitive via via crescenti e tendenti a quelle del discernimento umano.

Con tale premessa è evidente che qualora si operi in condizioni di incertezza e/o di incompletezza di informazioni il verificarsi di eventi ritenuti poco probabili apra lo spazio delle emozioni che demarcano in modo significativo il passaggio da metodi dell'incertezza intesa in modo rigoroso, come è il caso della probabilità, al trattamento di eventi incerti sotto l'azione di effetti emotivi o affettivo- emotivi.

In questo scenario ad essere più vicino al concetto di probabilità, tra credenza e plausibilità è la plausibilità, poiché seppur entrambi concernono una teoria della stima sotto il condizionamento emotivo, mentre l'affettività tende a fornire una risposta che seppur soggettiva mira ad essere oggettivabile, cioè un dato evento è considerato tale o ad esso è associabile una probabilità tale in base all'opinione dei più, nel caso della credenza non è richiesto tale confronto, ovvero non vi è il processo mentale della oggettivazione della valutazione soggettiva dell'assegnazione di una probabilità ad un dato evento; ma piuttosto con la credenza si vuole proprio far emergere la valutazione personale e soggettiva ma magari da far condividere, ma non necessariamente condivisibile a patto che sia necessariamente possibile.

In altre parole, per la credenza è sufficiente che la proposizione a cui si vuole assegnare una probabilità soggettiva sia solo possibile.

Da tali considerazioni emerge che così come il concetto di probabilità è più stringente di quello di plausibilità, poiché nel primo assumiamo che seppur ci sia incertezza, non ci sia valutazione emotivamente driven, nel secondo concetto ovvero nella definizione di plausibilità abbiamo che all'incertezza si associa una componente emotiva, che però è tesa a rappresentare l'opinione dei più; tale emotività con la credenza diventa più condizionante facendo emergere l'ego, ovvero una visione personale e soggettiva che non necessita del confronto con la collettività, ma che necessita solo della possibilità.

Altro elemento di differenza della nostra teoria rispetto a quella DS è che mentre la teoria DS fornisce un metodo per cambiare precedenti opinioni alla luce di nuove credenze, trattando in modo simmetrico sia la nuova che la precedente evidenza, cioè equipensandola, la nostra esperienza propria dei sistemi dinamici ci porta a dare più peso alle nuove evidenze che alle vecchie, in un'ottica di raffinamento del processo cognitivo e di avvicinamento graduale alla soluzione.

Nel caso dei mercati finanziari ciò vuol dire che se effettuiamo una media dei prezzi per ipotizzare un prezzo futuro allora una media che pesa di più i prezzi recenti rispetto a quelli meno recenti sarà più indicata rispetto ad una media che pesa equamente i prezzi nell'intervallo in cui si è scelto di calcolare la media.

La teoria non probabilistica di Nicolas Rescher della plausibilità può essere utile allorquando si abbiano informazioni (notizie o indicatori) conflittuali di pari attendibilità.

A titolo di esempio si pensi che un insieme di news o indicatori ci dica che il prezzo di uno strumento finanziario in un dato intervallo di tempo sia destinato a salire, e un altro sia destinato a scendere.

Grazie alla teoria di Rescher, che non è a somma 1, possiamo associare un valore di credenza rialzista, ad esempio 0,9, ed uno ribassista dello stesso valore, cioè sempre 0,9, poiché non siamo vincolati dalla somma uguale a uno come accade per le teorie a somma uno come la probabilità.

Il risultato al netto delle considerazioni fatte è che certamente il prezzo in quel dato intervallo prospettico di tempo considerato non sarà stazionario, ovvero è da escludersi un comportamento laterale dello strumento finanziario considerato.

3.10.3 *Necessità di metodi d'inferenza per strumenti finanziari dominati dall'incertezza: interazione operatore finanziario-mercato*

L'operatore finanziario nell'entrare al mercato effettua un'analisi (o qualcuno lo fa per lui) al fine di decidere se entrare su un dato strumento finanziario, e in tal caso che tipo di posizione prendere, in termini di dimensioni/esposizioni e di direzioni al rialzo o al ribasso.

Da ciò ne consegue il suo ingresso a mercato: ciò vuol dire che non resta nella posizione di osservatore, come farebbe uno studioso, ma opera a mercato e quindi interagisce con esso, lo perturba.

Detto in altre parole, l'operazione di analisi finisce per essere un'interazione col mercato, una volta che l'operatore decide di entrare.

Tale ragionamento ricorda molto il ragionamento di Richard Feynman⁹ relativamente alla spiegazione del processo di misura di sistemi quantistici e il conseguente principio di indeterminazione di Werner Karl Heisenberg¹⁰.

A tale proposito Feynman per spiegare che le fluttuazioni o le incertezze che portano ad introdurre una descrizione in termini di valori di aspettazione e di probabilità lo stato di un dato sistema dinamico, propone il seguente esempio. Supponiamo che il sistema dinamico oggetto di studio sia un elettrone, o comunque una particella del mondo microscopico descritto attraverso la meccanica quantistica; supponiamo che su tale sistema dinamico si intende fare una misura della posizione e della velocità (o della quantità di moto). Effettuare una misura significa inviare uno o più fotoni (ad esempio un fascio) in modo da conoscere istante per istante dove si trova l'elettrone in studio e con quale velocità.

L'osservazione di Feynman è che nell'effettuare tale misura stiamo irreparabilmente cambiando lo stato di moto del sistema elettrone, cioè lo stiamo perturbando.

Ciò poiché il fotone che usiamo per la misura, per fornire l'informazione di dove si trova l'elettrone e con quale velocità, urtandolo per misurarlo cambierà la sua velocità e in generale il suo moto.

Pertanto, l'effetto della misura rappresenta a tutti gli effetti una perturbazione, da cui discende che più si è sicuri della posizione, cioè abbiamo inviato più fotoni per accertarci della posizione e meno possiamo dire sulla quantità di moto ovvero sulla velocità perché nel fare le diverse misure abbiamo perturbato se non stravolto il moto dell'elettrone e viceversa, da cui il noto principio di Heisenberg che nella misura congiunta di posizione e di quantità di moto è impossibile essere più precisi di un certo valore di soglia dato dalla costante di Planck¹¹ ridotta.

Nella realtà della scala macroscopica, si pensi al moto di una vettura, comprendiamo che quanto spiegato dalla Meccanica Quantistica non risponde al nostro senso comune; infatti se ancora una volta il processo di misura è legato ad un fascio di fotoni che urtano contro il veicolo per sapere dove si trova e con quale velocità, è evidente che se non è in particolarissime e non comuni condizioni (fascio di potenza ad alta energia) i fotoni, "assimilabili" a puntini di pulviscolo, non potranno perturbare lo stato di moto del veicolo. Pertanto, in tal caso l'osservatore/misuratore rimane estraneo al moto del veicolo stesso nell'atto della misura e non genera e non si genera alcuna incertezza.

Per tale motivo la realtà macroscopica viene descritta nella meccanica classica, a partire dai principi della dinamica (inerzia, proporzionalità, azione e reazione) di Isaac Newton.

Molti sistemi dinamici su scala macroscopica hanno un comportamento simile ai sistemi quantistici, prima descritti.

Se si pensa, infatti, alla dinamica del prezzo di uno strumento finanziario appare evidente come quanto spiegato di Feynman sia in completa analogia.

⁹Il ragionamento di Feynman sostiene che solo un sistema quantistico, in cui vige un regime di incertezza o peggio di ambiguità, può simulare il comportamento di un altro sistema quantistico.

¹⁰Il principio di *Heisenberg* stabilisce i limiti nella conoscenza, o determinazione, dei valori che grandezze fisiche coniugate, ovvero due grandezze associate a operatori che non commutano fra loro, assumono contemporaneamente in un sistema fisico.

¹¹La costante di Planck, anche detta quanto d'azione e indicata con h , è una costante fisica che rappresenta la quantità d'azione elementare, determinando che le grandezze fisiche fondamentali non evolvano in modo continuo, ma siano quantizzate, cioè possano assumere solo valori multipli di tale costante.

Infatti se sostituiamo la posizione dell'elettrone con il livello di prezzo del dato strumento finanziario, arriveremo a formulare un principio di indeterminazione dei mercati finanziari analogo al principio di Heisenberg affermando che è impossibile predire con certezza prezzo e volatilità di uno strumento finanziario in un dato istante futuro.

Ciò accade perché la presa di decisione da parte dell'operatore finanziario e quindi l'ingresso al mercato implica automaticamente la transizione del mercato da uno stato a un altro, che può essere più o meno diverso dallo stato precedente a seconda delle dimensioni dell'operatore, cioè della sua liquidità introdotta a mercato, ovvero da quando questi sia in grado di dare, in qualsiasi modo una direzione al mercato.

A tal proposito si pensi agli effetti provocati da market mover (come le analisi effettuate da FED, BCE, BOE etc...).

Quanto descritto spiega chiaramente la necessità di inquadrare i mercati finanziari nell'ambito della teoria della complessità e di descriverli con strumenti sempre più accurati in grado di guardare all'incertezza da diverse prospettive e non solo da quella probabilistica, plausibilistica, credibilistica o possibilistica.

Poiché in un dato istante tutte queste diverse visioni, seppur in misura diversa, concorrono alla formazione del prezzo.

Pertanto è interessante notare come lavorando in modo specifico sulla modellazione dell'incertezza, e su come i diversi modelli probabili, plausibili, credibili, possibili concorrano alla costruzione di una funzione di stato del mercato che assorba le singole funzioni di probabilità, plausibilità, credibilità, possibilità, e le armonizzi sia possibile permettere all'analista di effettuare delle inferenze sugli stati futuri del mercato e aiutare l'operatore a prendere delle decisioni sul se entrare a mercato, in che dimensioni, e se a rialzo o a ribasso.

3.10.4 *Trattamento dell'incertezza via fusione delle informazioni*

Probabile

chiamiamo fase di discernimento

Plausibile

$A = (\text{Buy}, \text{Sell}) = (A_1, A_2)$

Credibile

Chiamiamo probabilità di base la funzione

$P_r : 2^A \rightarrow [0, 1]$ tale che

$P_r(\emptyset) = 0, 0 \leq P_r(A) \leq 1, \sum P_r(B) = 1 \quad B \subseteq A$

Possibile

Totale

In analogia definiamo:

$P_l : 2^A \rightarrow [0,1]$ la plausibilità;

$C_r : 2^A \rightarrow [0,1]$ la credibilità;

$P_o : 2^A \rightarrow [0,1]$ la possibilità.

A tal punto la domanda diventa come modellare una funzione di aspettazione a uno degli elementi di A e il suo complemento, ovvero se operiamo in un contesto monodimensionale a somma 1 in cui la

$$\sum_{i=1}^2 a(A_i) = 1$$

oppure no.

L'esempio precedente riguarda i mercati finanziari, ma non solo; infatti, l'esempio è una tipica modellazione di una negoziazione che ad esempio potrebbe riguardare la transizione di dati da un punto ad un altro in una rete internet, piuttosto che lo scambio di dati peer to peer, ecc.

3.10.5 Information Fusion e modellazione della funzione di aspettazione

Definiamo funzione di aspettazione a la funzione $a: 2^A \rightarrow [0,1]$ ottenuta da una opportuna composizione funzionale in termini di P_r, P_l, C_r, P_o , in modo tale che $a(\emptyset) = 0$

e

$$\sum_{i=1}^2 a(A_i) = 1.$$

3.10.6 Modello della media

Se indichiamo con $P_1=P_r, P_2=P_l, P_3=C_r, P_4=P_o$ il modello più semplice di funzione di aspettazione che possiamo costruire è quello in termini della media, cioè formalmente

$$a_1 = a_1(P_1, P_2, P_3, P_4) = \frac{1}{4} \sum_{i=1}^4 P_i$$

Evidentemente un tale modello ha la caratteristica di dare la stessa importanza alle diverse distribuzioni di probabilità, plausibilità, credibilità e possibilità, cioè è sia un vantaggio che uno svantaggio a seconda di quanto sia esperto l'analista ed in funzione di quanto si desidera rendere esperto un sistema.

3.10.7 Modello del prodotto

Un altro modo di stimare l'aspettazione di un dato evento o comportamento è attraverso il prodotto delle P_i , cioè formalmente

$$a_2 = a_2(P_1, P_2, P_3, P_4) = \prod_{i=1}^4 P_i$$

dato che $P_i: 2^A \rightarrow [0,1]$

dove in generale

$$\prod_{i=1}^4 P_i \leq \sum_{i=1}^4 P_i$$

ma per poter dire che $a_2 \leq a_1$ si richiede uno studio più dettagliato che includa i modelli funzionali delle P_i , considerato che in a_1 abbiamo il fattore di normalizzazione che è $\frac{1}{4}$.

3.10.8 Modello della media pesata

Quale estensione del primo modello potremo assumere che non tutte le P_i hanno la stessa importanza nella costruzione della funzione di aspettazione, costruendo così un modello del tipo media pesata, ovvero

$$a_3 = \left(\sum_{i=1}^4 \alpha_i P_i \right) \Bigg| \sum_{i=1}^4 \alpha_i$$

Con α_i pesi relativi alle diverse P_i

Un esempio potrebbe prevedere di dare un peso del 30% a P_1 cioè

$\alpha_1 = 0,5$, dal 25% a P_2 , cioè $\alpha_2 = 0,25$ del 15% a P_3 , cioè $\alpha_3 = 0,15$, e del 10% a P_4 , $\alpha_4 = 0,1$,

in tal modo la funzione di aspettazione risulterebbe

$$a_3 = \frac{\sum_{i=1}^4 \alpha_i P_i}{\sum_{i=1}^4 \alpha_i} = 0,5 P_1 + 0,25 P_2 + 0,15 P_3 + 0,1 P_4$$

3.10.9 Modello del prodotto pesato

In analogia al caso precedente e con l'obiettivo di estendere il modello del prodotto per la funzione di aspettazione risulta di interesse la possibilità di costruire un modello di prodotto pesato per la a, cioè formalmente

$$a_4 = \frac{\prod_{i=1}^4 \alpha_i P_i}{\sum_{i=1}^4 \alpha_i}$$

Anche in questo modo avremo la possibilità di pesare in modo diverso i diversi contributi della probabilità, plausibilità, credibilità e possibilità, alla funzione di aspettazione.

Il vantaggio dei modelli 3 e 4 rispetto ad 1 ed a 2 è che essi permettono di introdurre ulteriore conoscenza all'interno del sistema grazie ai pesi, cioè potrebbe accadere o attraverso una conoscenza a priori dell'esperto o attraverso il backtesting con l'impiego di tecnologie proprie dell'intelligenza artificiale come il machine learning.

3.10.10 Modello basato sulle regole di composizione di Dempster

Un altro modo significativo per la costruzione della funzione di aspettazione è di usare la regola che Dempster ha usato per la plausibilità estendendola al caso di applicazione delle P_i .

3.10.11 Modello di sovrapposizione con shift basato sulla probabilità

Il modello 1 così come il 3 rappresentano modelli di sovrapposizione e sovrapposizione pesata dei diversi contributi P_i alla funzione di aspettazione.

Una ulteriore possibile generalizzazione potrebbe prevedere un uso gerarchizzato delle P_i .

Più nel dettaglio potremo vedere i concetti di probabilità, plausibilità, credibilità e possibilità come concetti a cui ricorrere in ordine decrescente di importanza via via che la loro di un fenomeno cresce.

A titolo esemplificativo potremo considerare

$$a_6 = \begin{cases} P_1 & \text{se } 0 \leq P_1 \leq 99\% \\ P_2 & \text{se } 1\% < P_1 \leq 0,1\% \\ P_3 & \text{se } 0,1 < P_1 \leq 0,01\% \\ P_4 & \text{se } 0,01 < P_1 \end{cases}$$

Ciò significherebbe che fino al 99% della distribuzione di probabilità ragioneremo solo con la probabilità classica; oltre tale probabilità definiremo l'evento considerato come un evento improbabile, ma con coefficiente di plausibilità tra 0% e 100% nell'intervallo in cui $1\% < P_1 \leq 0,1\%$; analogamente però $0,1 < P_1 \leq 0,01\%$ definiamo l'evento improbabile ed implausibile, ma con un coefficiente di credibilità tra 0% ed il 100%; infine per valori ancora inferiori della probabilità avremo eventi improbabili, implausibili, ma con un coefficiente di possibilità tra lo % ed il 100%.

Agli estremi di tali valutazioni troveremo gli eventi certo da un lato con $P_1 = 100\%$ ed impossibile dall'altro con $P_1 = 0\%$, ovvero $P_1: 2^A \rightarrow [0,1]$ equivalentemente.

3.10.12 Modello di sovrapposizione con shift basato sulle P_i considerate in modo gerarchico

Una alternativa al modello precedente che però non intende far svolgere alla probabilità un nodo cardine, ma intende da un lato generalizzare le P_i e dall'altro valutare con la propria importanza anche gli elementi più di carattere emotivo e cognitivo, porta alla formulazione di una funzione di aspettazione che può essere così definita

$$a_7 = \begin{cases} P_1 & \text{se } \bar{P}_{P_1} - 3 \sigma_{P_1} \leq p(t) \leq \bar{P}_{P_1} + 3 \sigma_{P_1} \\ P_2 & \text{se } \bar{P}_{P_2} - 3 \sigma_{P_2} \leq p(t) \leq \bar{P}_{P_2} + 3 \sigma_{P_2} \text{ e } p(t) > \bar{P}_{P_1} + 3 \sigma_{P_1} \\ P_3 & \text{se } \bar{P}_{P_3} - 3 \sigma_{P_3} \leq p(t) \leq \bar{P}_{P_3} + 3 \sigma_{P_3} \text{ e } p(t) > \bar{P}_{P_2} + 3 \sigma_{P_2} \\ P_4 & \text{se } \bar{P}_{P_4} - 3 \sigma_{P_4} \leq p(t) \leq \bar{P}_{P_4} + 3 \sigma_{P_4} \text{ e } p(t) > \bar{P}_{P_3} + 3 \sigma_{P_3} \end{cases}$$

Dove \bar{P} è il prezzo medio sul set di dati considerato e σ è la deviazione standard; il Pedice P_i indica rispetto a quale distribuzione si calcono i \bar{P} e le σ .

Dai modelli preposti si comprende quanto sia generale l'impianto proposto al tema della decisione in condizioni di incertezza e come sia facile attraverso la sostituzione della funzione di aspettazione costruire modelli per descrivere i processi di inferenza in condizioni di info-incompletezza ed incertezza delle premesse input.

In realtà la teoria del ragionamento proposto più che un modello concettuale è un meta modello grazie al quale definendo modelli specifici per la funzione di aspettazione è possibile costruire nuovi e più opportuni modelli operativi contestualizzati ad un dato dominio cognitivo, informativo e al dato specifico modello.

3.10.13 L'affidabilità

L'affidabilità, intesa come reliability, di un metodo, un elemento, un sistema è la probabilità che esso funzioni senza guastarsi ad un dato istante e/o per un certo periodo t , sotto fissate condizioni operative e ambientali; pertanto la sua tipica definizione è in termini frequentistici cioè

$$R = \frac{nf}{n}$$

con nf numero di volte in cui il metodo, elemento o sistema ha funzionato, ed n numero totale di prove e campioni utilizzati per il test.

Analogamente possiamo definire l'inaffidabilità come fault cioè

$$F = 1 - R$$

Pertanto, siamo in una logica monodimensionale a somma 1 e quindi R e F sono funzioni del tempo.

Detto in altre parole il modo in cui usiamo l'affidabilità è chiedendoci dopo un tempo t quanti successi/fallimenti si sono ottenuti.

Formalmente l'affidabilità R è una funzione del tempo a valori in $[0,1]$, cioè

$$R: T \rightarrow [0,1]$$

Definiamo inoltre tempo medio di funzionamento del metodo, elemento, sistema, ovvero *mean time to failure (mttf)* la seguente quantità

$$mHf = 1/\lambda$$

Con $\lambda(t)$ testo di guasto al tempo t , cioè numero di guasti / fallimenti nell'unità di tempo.

In questo studio abbiamo considerato $R = R(T)$; in generale $R = R(T, C, A)$ dove t è il tempo e il metodo univoco per giudicare i successi ed i fallimenti/guasti, A sono le condizioni ambientali.

Evidentemente fissato il criterio di verifica c'è le condizioni ambientali $A, R = R(T)$.

Il concetto di affidabilità così definito si presta bene ad essere applicato, quale stimatore, alle funzioni di probabilità, plausibilità, credibilità e possibilità, in cui facendo il backtesting, ovvero lavorando sulle serie storiche si verifichi come le diverse inferenze hanno funzionato.

Allor quando non ci riferiamo al metodo ma al risultato, abbiamo la necessità di introdurre un altro concetto che è quello di attendibilità; in altre parole, parleremo di affidabilità del modello e attendibilità del risultato.

3.10.14 Attendibilità

In alcuni contesti semplici o semplificati affidabilità e attendibilità, così come probabilità, plausibilità, credibilità e possibilità, possono essere considerati sinonimi, ma non è questo il caso di quei contesti, domini ed eventi dove si richiede la teoria della complessità.

L'attendibilità esprime la costanza di un insieme di risultati, fissate le condizioni di misura.

Con questa definizione comprendiamo che in un contesto di incertezza, non descritto solo in termini di probabilità ma anche di altre proprietà o distribuzioni dipendenti dallo stato emotivo e dalle credenze dell'analista, come la plausibilità e la credibilità, l'attendibilità può esprimere proprio come stesse dinamiche vengano considerate in modo diverso.

Ad esempio un trader a parità di trend in momenti diversi potrebbe sentire la necessità di agire/reagire in modo diverso.

Ecco quindi che l'attendibilità del risultato acquista un ruolo importante al pari dei concetti precedenti e considerato diverso dagli stessi.

L'attendibilità quindi non è sinonimo di affidabilità, accuratezza o validità di un dato; ciò che qui si vuole mettere in evidenza è che diremo che una stima è attendibile se i risultati rimangono costanti nel tempo.

Qualora ciò non accada diremo che la stima ovvero il risultato è inattendibile, il che implicherà che l'analista o il metodo utilizzato è inaffidabile.

Esistono diversi metodi per stimare l'attendibilità; noi utilizzeremo l'indice di correlazione di Pearson¹² che è alla base generalmente delle diverse stime di attendibilità.

In generale date due variabili statistiche x e y , l'indice di correlazione di Pearson è

$$r = \frac{\vartheta x y}{\vartheta x \vartheta y} \quad \text{con} \quad -1 \leq r < 1$$

dove $\vartheta x y$ la covarianza, e $\vartheta x \vartheta y$ le deviazioni standard delle variabili $x y$ rispettivamente.

¹²L'indice di correlazione di Pearson tra due variabili statistiche è definito come la loro covarianza divisa per il prodotto delle deviazioni standard delle due variabili ed esprime una eventuale relazione di linearità tra esse.

Generalmente distinguiamo i seguenti casi:

- $r > 0$: variabili correlate
- $r = 0$: variabili incorrelate
- $r < 0$: variabili anticorrelate
- $0 < r < 0.3$: variabili debolmente correlate
- $0.3 < r < 0.7$ variabili moderatamente correlate
- $r > 0.7$ variabili altamente correlate
-

ed analogamente per i casi di anticorrelazione.

Più in generale nel caso di più di due variabili avremo che l'indice di Pearson diventa

$$r = r_{ij}$$

Cioè abbiamo una matrice di correlazione di r righe ed r colonna simmetrica $r_{ij} = r_{ji}$ con coefficienti sulla diagonale uguali ad 1, dato che $r_{ii} = \vartheta_{ii} / \vartheta_i^2$.

Nel caso finanziario potremo così stimare l'attendibilità di due o più periodi di risultati ad esempio.

4 Conclusioni & Prospettive

Il presente documento rientra nell'ambito dell'obiettivo D.6a Piattaforma ICT per la gestione di smart districts che prevede la realizzazione di una piattaforma software in grado di raccogliere grandi moli di dati provenienti da distretti urbani. La gestione e l'estrazione delle informazioni dai dati possono aiutare nella creazione di servizi capaci di migliorare la qualità delle città e indirizzarle alla smartizzazione. Lo studio inerente al tema in esame è stato costituito da un rapporto tecnico che si è articolato in tre sezioni principali:

- 1) Metodologie multiscala in spazi trasformati;
- 2) Principali applicazioni presenti in letteratura relative alle metodologie analizzate in data fusion riguardanti i contesti urbani;
- 3) Trattamento delle informazioni incerte o incomplete in prospettiva probabilistica.

L'obiettivo della prima sezione è stato di offrire una panoramica prima generale e poi contestualizzata dell'analisi multiscala e degli spazi trasformati per il trattamento di Big Data in contesti complessi come le smart cities dove l'informazione più trovare un trattamento più efficace qualora la si relazioni alla scala. Tali tipi di analisi negli anni si sono rivelati particolarmente utili per lo studio di infrastrutture critiche e per ambiti ad alta complessità. La sezione in esame è stata articolata come segue: i) analisi delle principali tecniche di analisi in spazi trasformati, ovvero Fast Fourier Transform, la Trasformata di Walsh-Hadamard, la Trasformata di Hartley, la Trasformata di Laplace, la Trasformata di Hilbert; ii) la Wavelets Analysis, con specifico approfondimento sulla Trasformata Wavelet continua, la Trasformata Wavelet discreta: Frames, le Wavelet ortogonali e l'analisi multirisoluzione, le MultiWavelets, i Dizionari tempo-frequenza e i Wavelet Packets.

La seconda sezione ha riguardato l'information fusion ed i sistemi di supporto alle decisioni. I risultati indicati in essa sono i richiami a quanto presentato nel PAR 2016 di cui questo studio è da intendersi la naturale continuazione. Infatti lo studio del PAR2016 denominato "Soluzione Tecnologica di supporto alle decisioni ed Information Fusion per grandi moli di dati in contesto smart cities" approfondiva due temi fondamentali, ovvero la Definizione di metodologie, metodi e modelli della complessità per il decision making per la smart city e gli Esempi di procedure ed algoritmi per il decision making in smart cities context. Grazie all'analisi multiscala offerta nella prima sezione è stato possibile rendere quindi capace una soluzione tecnologica di estendere i risultati ottenuti con una analisi efficace che esponga diverse viste dell'informazione con una granularità diversa a seconda della scala a cui si è interessati ovvero dell'Ente/attore considerato.

La terza sezione ha completato il presente studio affrontando un ulteriore tema di interesse, ovvero la possibilità di gestire informazioni incerte o incomplete. Per raggiungere un tale obiettivo si è reso necessario operare in un contesto di probabilità estesa. Infatti, è consueto che in ambiti propri della complessità il concetto classico di probabilità non sia più sufficiente. Spesso si è ovviato alla costruzione di nuovi modelli, considerando il fenomeno in studio descritto attraverso una distribuzione gaussiana e modellando le discrepanze tra il modello ed i dati reali aggiungendo delle opportune code pesanti alla distribuzione normale. Data l'importanza del lavoro e la possibilità di impiego in molti altri contesti, in questo studio è stata costruita una nuova metodologia della attendibilità/aspettazione che coniuga in modo sinergico il concetto di probabilità, plausibilità, credibilità e possibilità all'interno di uno scenario che tratta informazione incerta ovvero incompleta. La sezione è quindi stata organizzata come segue: i) La plausibilità, ii) la definizione probabilistica, iii) la logica del plausibile di Polya, iv) il credibile, v) la concezione di Dempster-Shafer, vi) la concezione della teoria Dezert Smarandache, vii) la concezione non probabilistica, viii) Plausibilità e non monotonia, ix) Cognitivismo e plausibilità, x) Verso una teoria del ragionamento plausibile, xi) Decisione in Domini di info-Incertezza e/o di info-Incompletezza.

Questo studio ha mostrato quindi la capacità di costruire un sistema di supporto alle decisioni che si basi sull'information fusion per il trattamento di big data in contesti smart cities, prevedendo la possibilità di analisi multiscala ed il trattamento di informazione incerta e/o incompleta.

Quest'ultimo tipo di analisi effettuata per il trattamento di tale tipo di informazioni (incerte e/o incomplete) rientra in un'ottica più ampia ovvero di ampliamento di piattaforme smart cities come la suddetta con modelli di tipo blockchain così da ambire a dar vita ad un'economia di tipo circolare nei distretti urbani. Come prospettiva futura è interessante osservare la possibilità di introdurre informazioni in input che non siano solo dati relativi a misure in campo, ma anche informazioni non formattate e di natura linguistica, nonché di trasformare i risultati ottenuti con output non solo numerici, ma anche linguistici, ovvero più vicini alle capacità dialogiche umane. Ciò potrebbe essere ottenuto grazie all'impiego di innovative tecniche proprie dell'ambito di ricerca indicato col termine di Elastic Search. All'interno dell'analisi semantica, l'Elastic Search può essere usato per cercare qualsiasi tipo di informazione, fornendo un sistema di ricerca scalabile, quasi di tipo real-time, con supporto al multitenancy. L'Elastic Search è una soluzione distribuita, ciò significa che gli indici possono essere suddivisi in shard, ognuno con possibilità di replica. Ogni nodo contiene uno o più shard, ed è in grado di agire da coordinatore, delegando le operazioni necessarie allo shard (o agli shard) corretti. Inoltre, il Routing ed il rebalancing sono effettuati automaticamente. Ad oggi la soluzione più promettente di Elastic Search è basata sul framework semantico Lucene.

Lo scenario sembra molto interessante e si ritiene che i risultati conseguiti nel PAR2016 e PAR2017 meritino una ulteriore estensione al fine di accogliere anche aspetti semantici e di text analysis all'interno della soluzione di decision making proposta, a partire dalle tecniche di Information Fusion impiegate e coniugate già con tecniche di analisi multiscala e di probabilità estesa.

5 Riferimenti bibliografici

1. Gonzelez R.C. and Wintz P., *Digital Image Processing*, 2nd edition, Addison-Wesley, 1987.
2. S. J. Campanella and E. S. Robinson, A comparison of orthonormal transform for digital speech processing, *IEEE Trans. Commun. Technol*, Vol. COM-19, P. 1045-1050, Dec. 1971.
3. F.Y. Y. Shum, A. R. Elliot, and W. O. Brown, Speech processing with Walsh-Hadamard transform, *IEEE Trans. Audio Electro-acoustic*, Vol. AU-21, P. 174-179, June 1973.
4. W. K. Pratt, J. Kane, and H. C. Andrews, Hadamard Transform Image Coding, *Proc. IEEE*, Vol. 57, P. 58-68, Jan. 1969.
5. R. N. Bracewell, *The Hartley Transform*, England: Oxford University Press. 1985.
6. A. N. Belbachir, et al., *Image Compression using Hartley Transform*, Technical Report, PRIP-TR-087, Dec. 2003, Url: <http://www.prip.tuwien.ac.at/~nabil/pdf/>.
7. C. H. Paik, M. D. Fox, Fast Hartley Transform for Image Processing, *IEEE Trans. Med. Imag*, Vol. 7, P. 149-153, June 1988.
8. H. V. Malvar, Fast Computing of the Discrete Cosine Transform and the Discrete Hartley Transform, *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-35, N. 10, P. 1484-1485, Oct. 1987.
9. D.V. Widder, *The Laplace Transform*, Princeton, NJ: Princeton University Press, 1941.
10. D. H. Bailey, P. N. Swarztrauber, A Fast Method for the Numerical Evaluation of Continuous Fourier and Laplace Transforms, *SIAM Journal on Scientific Computing* Vol. 15, pp. 1105-1110, 1994.
11. S. L. Hahn, *Hilbert Transforms in Signal Processing*, Norwell, MA: Artech House, 1996.
12. Kai Schneider, Marie Farge, *Wavelets: theory*, *Encyclopedia of Mathematics Physics*, eds J.P. Francoise et al., Elsevier, 2006.
13. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.
14. S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 11, p. 674-693, July 1989.
15. M. Cotronei, L. B. Montefusco, and L. Puccio, "Multiwavelet analysis and signal processing," *IEEE Trans. Circuits Syst. II*, vol. 45, pp. 970-987, 1998.
16. Mariantonia Cotronei, Laura B. Montefusco, and Luigia Puccio, *Multiwavelet Analysis and Signal Processing*, *IEEE Transactions on Circuits and Systems: analog and digital signal processing*, vol. 45, no. 8, august 1998.
17. J. S. Geronimo, D. P. Hardin, and P. R. Massopust, Fractal function and wavelet expansions based on several scaling functions, *J. Approx. Theory*, vol. 78, pp. 373-401, 1994.
18. S. Mallat and Z. Zhang, *Matching Pursuits With Time-Frequency Dictionaries*, *IEEE Transactions on Signal Processing*, December 1993.
19. M.V. Wickerhauser, M. Farge, E. Goirand, E. Wesfreid, E. Cubillo, Efficiency Comparison of Wavelet Packet and Adapted Local Cosine Bases for Compression of a Twodimension Turbulent Flow, in *Wavelets: Theory, Algorithm and Application*, C.K. Chui, L. Montefusco, L. Puccio, Academic Press, Inc., 1994.
20. R. Coifman, Y. Meyer, S. Quake, and V.Wickerhauser, *Signal processing and compression with wave packets*, Numerical Algorithms Research Group, New Haven, CT Yale University, 1990.
21. M.V. Wickerhauser, Acoustic signal compression with wavelet packets, in *Wavelets: A Tutorial in Theory and Applications*, C. K. Chui, Ed. New York: Academic, pp. 679-700, 1992.
22. M.V. Wickerhauser, Comparison of Picture Compression Methods: Wavelet, Wavelet Packet, and Local Cosine Transform Coding, in *Wavelets: Theory, Algorithm and Application*, C.K. Chui, L. Montefusco, L. Puccio, Academic Press, Inc., 1994.
23. Dao-Qing Dai and Hong Yan, *Wavelets and Face Recognition*, url: <http://s.i-techonline.com/Book/Face-Recognition/ISBN978-3-902613-03-5-fr04.pdf>
24. Chien, J. T. & Wu, C. C., Discriminant waveletfaces and nearest feature classifiers for face recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No.12, pp. 1644-1649, 2002.
25. Feng, G. C., Yuen, P. C. & Dai, D. Q., Human face recognition using PCA on waveletsubband, *Journal of Electronic Imaging*, Vol. 9, No. 2, pp. 226-233, 2000.

26. Zhang, B. L., Zhang, H. H. & Ge, S. S., Face recognition by applying wavelet subband representation and kernel associative memory, *IEEE Trans. Neural Networks*, Vol. 15, No. 1, pp. 166–177, 2004.
27. Ekenel, H. K. & Sanker, B. (2005). Multiresolution face recognition, *Image and Vision Computing*, Vol. 23, (May 2005) pp. 469–477.
28. Zhang, Z. B., Ma, S. L. & Wu, D. Y., The application of neural network and wavelet in human face illumination compensation, *Proc. Advances in Neural Networks*, pp. 828–835, 2005.
29. Saito, N. & Coifman, R. R., Local discriminant bases and their applications, *J. Math, Imaging Vision*, Vol. 5, No. 4, pp. 337–358, 1995.
30. Saito, N., Coifman, R. R., Geshwind, F. B. & Warner, F., Discriminant feature extraction using empirical probability density estimation and a local basis library, *Pattern Recognition*, Vol. 35, pp. 2841–2852, 2002.
31. W. W. Boles and B. Boashash, A Human Identification Technique Using Images of the Iris and Wavelet Transform, *IEEE transactions on signal processing*, vol. 46, no. 4, 1998.
32. S. Mallat, Zero-Crossings of a Wavelet Transform, *IEEE Transactions on Information Theory*, vol. 37, p. 1019-1033, July 1991.
33. D. Donoho, Nonlinear Wavelet Methods for Recovery of Signals, Densities and Spectra from Indirect and Noise Data, *Different Perspectives on Wavelet*, *Proceeding of Symposia in Applied Mathematics*, vol. 47, 1 Daubechies ed. Amer. Math. Soc. Providence, R.I., 1993 pp. 173-205.
34. B. G. Sherlock and D. M. Monro, Psychovisually Tuned Wavelet Fingerprint Compression, *IEEE, Proceedings of International Conference on Image Processing*, Volume 1, pp. 585–588, 1996.
35. A. Ahmadyfard, M. S. Nosrati, A Novel Approach for Fingerprint Singular Points Detection Using 2D-Wavelet, *IEEE/ACS International Conference on Computer Systems and Applications*, pp. 688-691, 2007.
36. J. P. Leduc, F. Mujica, R. Murenzi and M. Smith, Spatio-Temporal Wavelet Transforms For Motion Tracking, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97)*, Volume 4, 1997.
37. M. Duval-Destin and R. Murenzi, Spatio-Temporal Wavelet: Application to the Analysis of Moving Patterns, in *Progress in Wavelets Analysis and Applications (Pro., Toulouse 1992)* Y. Meyer and S. Roques, Editors, Ed. Frontiers, Gif-sur-Yvette, 1993.
38. J. P. Leduc, J. R. Corbett, M. V. Wickerhauser, Rotational Wavelet Transforms for Motion Analysis, Estimation and Tracking, In *Proceedings of the 1998 IEEE International Conference on Image Processing (ICIP-98)*, Chicago, Illinois, pages 195–199, 1998.
39. A. Khare¹ and U. S. Tiwary, Daubechies Complex Wavelet Transform Based Moving Object Tracking, *Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing CIISP 2007*.
40. J. Ahmed, M.N. Jafri, J. Ahmad, Target Tracking In An Image Sequence Using Wavelet Features And A Neural Network. In *Proceedings of IEEE Region 10: Tencon'05 Conference*, Melbourne, 2005
41. W. T. Kuang and A. S. Morris, Using Short-Time Fourier Transform and Wavelet Packet Filter Banks for Improved Frequency Measurement in a Doppler Robot Tracking System, *IEEE Transactions On Instrumentation And Measurement*, Vol. 51, No. 3, June 2002.
42. M. Khansari, M. Asadi, P. Khadem Hamedani, M. Ghanbari, Adaptive Search Window for Object Tracking in the Crowds using Undecimated Wavelet Packet Features, *World Automation Congress (WAC)*, July 24-26, Budapest, Hungary, 2006.
43. C. Bagnoli, *La misurazione economica sfocata. Dal numero alla parola. Strumenti per la gestione della complessità*, 2007, Franco Angeli Editore.
44. A. Basso, C. Marchionni, *I modelli in economia*, n°11 Gennaio 2015, APhEx Portale Italiano di Filosofia Analitica http://www.aphex.it/public/file/Content20150130_APhEx11,2015TemiModelliEconomiaBasso-Marchionni.pdf
45. A. Bottani, *Verità e coerenza. Saggio sull'epistemologia coerentista di Nicholas Rescher*, 1989, Franco Angeli Editore.
46. P. Coretto, Una teoria della decidibilità: entropia e scelte in condizioni di incertezza, 2002, *Rivista Politica Economica* <http://www.rivistapoliticaeconomica.it/2002/nov-dic/coretto.pdf>

47. A.F.Dragoni, *Norme e metodi razionali per il cambiamento di stato cognitivo*, 1996, Dipartimento Ingegneria Informatica, Gestionale e dell'Automazione Università Politecnica delle Marche <http://leibniz.diiga.univpm.it/~dragoni/curriculum/pubblicazioni/SistemiIntelligenti.pdf>
48. F.Gil, *Prove. Attraverso la nozione di prova/dimostrazione*, 1990 Jaca Book.
49. E.Ippoliti, *Modello d'inferenza plausibile*, Tesi dottorale in Logica ed Epistemologia, Dipartimento di studi filosofici ed Epistemologici, 2014, Università La Sapienza di Roma <http://padis.uniroma1.it/bitstream/10805/679/1/IppolitiEmiliano52.pdf>
50. P.Terna, R.Boero, M.Morini, M.Sonnessa, *Simulazione, modelli ad agenti e scienze sociali*, 2005, Dipartimento di Informatica - Scienza e Ingegneria Università di Bologna http://www.cs.unibo.it/fioretti/pub_free/italiano/evidenza.pdf

Appendice: CV del gruppo di lavoro

Il prof. Gerardo IOVANE dopo essersi diplomato con il massimo dei voti alla Scuola Militare Nunziatella, ha conseguito la laurea con lode in Fisica Nucleare e Subnucleare, ha svolto poi un master di ricerca presso il CERN (Centro Europeo per la Fisica Nucleare) di Ginevra. Ha conseguito prima il Dottorato di Ricerca in Fisica ed in seguito il Dottorato di Ricerca in Matematica, essendosi poi dedicato al terzo dottorato in Ingegneria ed Economia dell'Innovazione. Rientrato in Italia è diventato Ricercatore Universitario in Fisica-Matematica e poi Professore Associato in Analisi Matematica presso la Facoltà di Ingegneria dell'Università degli Studi di Salerno.

Da oltre venti anni, Gerardo Iovane partecipa a diverso titolo a progetti di ricerca avanzata ed industriale in ambito internazionale, collaborando con rilevanti strutture su temi di ricerca scientifica e tecnologica. Oltre ad essere docente all'Università di Salerno collabora con diversi Atenei tra cui l'Università di Messina dove per diversi anni ha tenuto il corso di Tecnologie Avanzate per la Sicurezza, Biometria ed Identificazione al Corso di Laurea in Informatica della Facoltà di Scienze Matematiche Fisiche e Naturali, la Link Campus University dove ha tenuto il corso di Metodi Matematici per l'Economia.

In ambito Militare e Difesa è titolato IASD (Istituto Alti Studi Difesa), inoltre è consigliere scientifico di diversi organismi e strutture della Difesa sui temi della Sicurezza; per diversi anni ha diretto molteplici ricerche con il CeMiSS (Centro Militare Studi Strategici) del Ministero della Difesa.

Esperto Scientifico presso il Ministero dell'Università e della Ricerca Scientifica, quale valutatore di progetti industriali e grandi progetti strategici (gps), sui temi delle Metodologie Matematiche e Tecnologie Informatiche e presso il Ministero delle Politiche Agricole e Forestali, il Prof. Iovane è autore di oltre 200 lavori scientifici tra libri, saggi, articoli e contributi a conferenze nazionali ed internazionali.

Il prof. Gerardo Iovane è stato insignito di diverse onorificenze e premi, tra cui l'onorificenza di Cavaliere dell'Ordine al Merito della Repubblica Italiana, il Premio Speciale Leonardo per la Ricerca Scientifica per le attività sui numeri primi e la cifratura, il Premio Internazionale Giuseppe Sciacca, il Premio Best Application in Europe per le tecnologie per la sicurezza e la protezione delle infrastrutture critiche, la full professorship onoraria in Computer Science alla Costantinian University del Rhode Island State (USA), la Laurea Honoris Causa in Medicina Nucleare dell'Universitas Sancti Cyrilli – Malta per le sue ricerche sulla scrittura di informazioni su liquid support (progetto acqua informata), la Laurea Honoris Causa in Scienze Turistiche della Libera Facoltà di Scienze Turistiche di Caserta, per il suo contributo sulle tecnologie ICT applicate al settore turistico, la Laurea Honoris Causa in Affari Internazionali e Scienze Diplomatiche della Costantinian University per il suo contributo relazionale e diplomatico, quale segretario nazionale per l'Italia dello IODR (International Organization for Diplomatic Relations).