

UNIVERSITA' DEGLI STUDI DI CATANIA

FACOLTÀ DI ECONOMIA

DIPARTIMENTO DI ECONOMIA E METODI QUANTITATIVI

GIOVANNI PETRALIA

**L'APPROCCIO DEI ROUGH SETS BASATI
SULLA DOMINANZA APPLICATO ALLA
VALUTAZIONE DEL MERITO CREDITIZIO**

TESI DI DOTTORATO

COORDINATORE

Chiar.mo Prof. Benedetto Matarazzo

TUTOR

Chiar.mo Prof. Salvatore Greco

**DOTTORATO IN MATEMATICA PER LE DECISIONI
ECONOMICHE E FINANZIARIE XXII CICLO**

INDICE

INTRODUZIONE

Capitolo 1

MODELLI STATISTICI PER LA PREVISIONE DELLE INSOLVENZE

1	L'APPROCCIO UNIVARIATO	pag.1
2	L'APPROCCIO MULTIVARIATO	pag.5
2.1	L'ANALISI DISCRIMINANTE LINEARE	pag.6
2.1.1	IL MODELLO	pag.11
2.1.2	ESTENSIONI DEL MODELLO	pag.14
2.1.3	IL MODELLO Z SCORING	pag.19
2.2	LA REGRESSIONE LOGISTICA	pag.28
3	AMBITI DI APPLICAZIONE	pag.35
4	CONCLUSIONI	pag.36

Capitolo 2

L'APPROCCIO MULTICRITERIALE ALLE DECISIONI

1	I PROBLEMI DECISIONALI	pag.38
1.1	GENERALITÀ	pag.38
1.2	CLASSIFICAZIONI DEI PROBLEMI DECISIONALI	pag.42
1.3	AIUTO MULTICRITERIALE ALLA DECISIONE	pag.47
2	LE STRUTTURE DI PREFERENZA	pag.51
2.1	GENERALITÀ.....	pag.51

2.2	ALCUNE OSSERVAZIONI GENERALI SULLE RELAZIONI BINARIE	pag.54
2.3	SITUAZIONI ELEMENTARI DI PREFERENZA	pag.55
2.4	STRUTTURE DI PREFERENZA	pag.58
2.5	RELAZIONI DI PREFERENZA MULTIPLE	pag.64
2.6	RELAZIONE DI SURCLASSAMENTO A QUATTRO VALORI	pag.64
3	MODELLI DI AGGREGAZIONE DELLE PREFERENZE	pag.69
3.1	DOMINANZA	pag.69
3.2	PROCEDURE ELEMENTARI DI AGGREGAZIONE	pag.71
3.2.1	SOMMA PONDERATA	pag.72
3.2.2	MASSIMO	pag.73
3.2.3	MINIMO	pag.74
4	CARATTERISTICHE FONDAMENTALI DI UNA PROCEDURA DI AGGREGAZIONE MULTICRITERIALE	pag.75
4.1	PROCEDURE DI AGGREGAZIONE COMPENSATORIE E NON COMPENSATORIE	pag.75
4.2	DIFFERENTI TIPI DI SCALE	pag.79
5	I MODELLI MULTICRITERIALI	pag.81

Capitolo 3

L'APPROCCIO DEI ROUGH SETS ALL'ANALISI DELLE DECISIONI

1	INTRODUZIONE	pag.83
2	CLASSICAL ROUGH SETS APPROACH (CRSA)	pag.86
2.1	TAVOLA DELLE INFORMAZIONI E RELAZIONE DI INDISCERNIBILITÀ	pag.86
2.2	APPROSSIMAZIONI	pag.88

2.3	RIDOTTI E CORE	pag.91
2.4	TAVOLA DELLE DECISIONI E REGOLE DECISIONALI	pag.93
2.5	UN ESEMPIO PRATICO DI APPLICAZIONE DELLA METODOLOGIA CRSA	pag.95
2.6	CONFRONTO CON L'ANALISI STATISTICA	pag.100
2.7	GENERALIZZAZIONE DELLA RELAZIONE DI INDISCERNIBILITÀ	pag.102
3	I ROUGH SETS E LE DECISIONI MULTIATTRIBUTO	pag.105
3.1	PROBLEMI DI CLASSIFICAZIONE MULTIATTRIBUTO	pag.107
3.2	PROBLEMI DI CLASSIFICAZIONE MULTICRITERIALE	pag.107
4	DOMINANCE-BASED ROUGH SETS APPROACH (DRSA) ...	pag.109
4.1	APPROSSIMAZIONI BASATE SULLA DOMINANZA	pag.110
4.2	QUALITA' DELL'APPROSSIMAZIONE ED INSIEMI RIDOTTI	pag.113
4.3	PRINCIPALI VANTAGGI DELLA CLASSIFICAZIONE MULTICRITERIALE	pag.117

Capitolo 4

UN MODELLO DI SCORING BASATO SULL'APPROCCIO DEI ROUGH SETS

1.	INTRODUZIONE	pag.120
2.	LA METODOLOGIA	pag.120
3.	UN'APPLICAZIONE AD UN CASO CONCRETO	pag.130
3.1	CALCOLO DELLE REGOLE DECISIONALI	pag.136
3.2	ANALISI DEI RISULTATI	pag.140
4.	CONCLUSIONI	pag.143

CONCLUSIONE

BIBLIOGRAFIA

A mio padre

INTRODUZIONE

La pesante crisi che ha colpito i mercati finanziari globali, con maggiore enfasi nell'ultimo triennio a seguito dell'insolvenza dei mutui "subprime", ha posto ancora una volta l'accento sulle problematiche relative al rischio di credito. In particolare, sin dal primo accordo di Basilea del 1988, la regolamentazione internazionale ha richiesto agli istituti di credito l'osservanza di requisiti patrimoniali sempre più stringenti. Le motivazioni sottostanti il crescere dell'attenzione sul rischio di credito dipendono dalla circostanza che l'insolvenza di un intermediario finanziario produce dei costi sociali, in termini di oneri sostenuti dalla collettività attraverso l'assicurazione dei depositi, di contagio di altri operatori, di perdita di fiducia da parte del pubblico nei confronti dell'intero settore bancario. Pertanto, per un istituto di credito diventa centrale elaborare modelli per la previsione delle insolvenze e per il monitoraggio del proprio portafoglio clienti.

Nella letteratura sono presenti diverse metodologie per la previsione delle insolvenze: metodi statistici univariati, metodi della "sopravvivenza", analisi discriminante, modelli lineari di probabilità, analisi logit e probit, algoritmi di partizionamento recursivo, programmazione matematica, metodi multicriteriali di supporto alla decisione, sistemi esperti. Lo scopo del presente lavoro consiste nell'illustrare le caratteristiche di un approccio alla valutazione del merito

creditizio basato su logiche multicriteriali, in particolare mediante l'applicazione dei rough sets basati sulla dominanza.

Il presente lavoro si compone di 4 capitoli. Nel primo capitolo, vengono espone due delle principali metodologie statistiche utilizzate largamente nella pratica per la previsione delle insolvenze, l'analisi discriminante e la regressione logistica. Con riferimento alla prima metodologia, viene esposto il modello "Z-Scoring", metodologia introdotta negli anni 60 da Altman, che ha aperto la strada ad un nutrito filone di applicazioni in molti paesi ed in molti contesti.

Nel secondo capitolo, viene fornita una sintetica dissertazione sull'approccio multicriteriale alle decisioni. In particolare, vengono esposti i concetti di base che permettono di formulare un problema decisionale in termini di analisi multicriteriale, ed inoltre vengono introdotti alcuni dei principali modelli di aggregazione delle preferenze.

Nel terzo capitolo, viene introdotta la teoria dei rough sets. In particolare, vengono esposti l'approccio classico, proposto da Pawlak nel 1982, basato sulla relazione di indiscernibilità, ed il nuovo approccio proposto nel 1996 da Greco, Matarazzo e Slowinski, basato sulla relazione di dominanza.

Nel capitolo conclusivo, nella prima parte viene esposto l'approccio dei rough sets basati sulla relazione di dominanza per la valutazione del rischio di fallimento, approccio introdotto nel 1998 da Greco, Matarazzo e Slowinski, mentre nella seconda viene applicato tale approccio ad un campione di imprese fornito da un primario istituto bancario di italiano.

CAPITOLO 1

MODELLI STATISTICI PER LA PREVISIONE DELLE INSOLVENZE

1. L'APPROCCIO UNIVARIATO

L'approccio univariato esamina singolarmente i diversi indicatori cercando di percepirne gli elementi in grado di illustrare i punti deboli dell'impresa, lo stato attuale, i condizionamenti che gravano sullo sviluppo futuro. L'insieme degli indicatori viene organizzato in un sistema coerente di analisi, orientato alla particolare prospettiva con la quale si guarda l'impresa: è l'analista finanziario che, sulla base di ragionamenti, di confronti con dati di settore e con parametri di riferimento, dall'esame sistematico della serie di conti aziendali e degli indicatori ricavati raggiunge proprie conclusioni in merito alla situazione e alle prospettive dell'impresa; l'uso integrato di altre informazioni sui programmi aziendali e di natura qualitativa consente di confrontare le conclusioni precedenti e di comprendere meglio i meccanismi economici che stanno dietro i valori contabili. Pertanto, ciò che preme sottolineare è che l'analisi univariata considera gli

indicatori individualmente, o in sistema, ma non fa alcun tentativo di combinarli insieme in una misura quantitativa di sintesi.

In una famosa ricerca, Beaver (1966)¹, ha esaminato la capacità predittiva di alcuni singoli indicatori rispetto al fenomeno dell'insolvenza. Beaver ha utilizzato un campione di 79 imprese anomale; la definizione di anomalia comprende il fallimento, l'insolvenza nei confronti dei propri prestiti obbligazionari, l'esistenza di scoperti sui conti bancari o sconfinamenti, il mancato pagamento di dividendi sulle azioni privilegiate.

Il campione delle società sane, da confrontare con quelle anomale, è stato scelto estraendo casualmente un'impresa, per ogni società anomala, appartenente allo stesso settore e alla stessa classe dimensionale in termini di attivo netto totale: il campione delle società sane ha svolto una funzione di confronto omogeneo con quello delle società anomale, per facilitare l'individuazione delle caratteristiche distintive tra i due gruppi, neutralizzando, o riducendo, l'effetto di variabili quali l'appartenenza settoriale o la scala dimensionale.

Per ciascuna impresa Beaver ha calcolato una trentina di indicatori scelti tra quelli più citati e studiati nella letteratura, o dimostratisi più efficaci in studi precedenti; tali indicatori sono stati raggruppati in sei famiglie omogenee rispetto al significato economico.

Il confronto tra le medie degli indicatori dei due campioni ha confermato il risultato di studi precedenti risalenti agli anni trenta e quaranta, mettendo in luce

¹ Beaver William H (1966), *Financial Ratios As Predictors of failure*, Journal Of Accounting Research, Vol. 4, Empirical Research in Accounting: pp 71-111.

una sistematica differenza di livello e di andamento degli indicatori delle società anomale rispetto a quelli delle società sane.

Il paragone dei soli valori medi, tuttavia, è troppo limitativo e concentra l'intera distribuzione dei valori degli indicatori in un solo punto. Per ottenere una migliore valutazione delle capacità diagnostiche degli indicatori, Beaver ne ha esaminato la sovrapposizione delle distribuzioni calcolate separatamente sulle società sane e su quelle anomale, pervenendo sulla base di un test di classificazione dicotomica a individuare un punto ottimale di separazione (cut-off) per gli indicatori, in grado di ridurre al minimo gli errori di attribuzione delle società ai due insiemi (sane-anomale).

Sulla base di tali elaborazioni, Beaver ha trovato che il migliore indicatore per la previsione delle insolvenze è il rapporto tra Cash Flow² e i debiti totali che, nell'anno immediatamente precedente al momento dell'insolvenza o del fallimento (t-1), ha correttamente individuato l'87% delle società; negli anni precedenti al t-1 la performance, pur riducendosi, si mantiene su livelli molto buoni: cinque anni prima dell'insolvenza questo indicatore ha correttamente classificato il 78% delle società.

² Beaver definisce il *Cash Flow* come somma dell'utile netto e dei costi monetari, ma tale indicatore è incompleto, infatti nella prassi contabile questo valore viene definito come *Autofinanziamento*. In realtà, per giungere alla determinazione di una grandezza interpretabile come Cash Flow (senza leva finanziaria) occorrerà: aggiungere la somma algebrica tra gli interessi passivi e attivi relativi alla gestione finanziaria; sottrarre la variazione positiva o sommare quella negativa, registrata tra i bilanci del periodo t e (t-1), del Capitale Circolante Commerciale Netto; sommare (sottrarre) i valori rivenienti dal disinvestimento (investimento) in attività strumentali, . Il Capitale Circolante Commerciale (come d'altronde l'attività di investimento/disinvestimento), in altre parole, agisce come una "spugna" in grado di assorbire o espellere liquidità a seconda che subisca variazioni in aumento o in riduzione. Se poi si vuole determinare il Cash Flow (Levered) occorrerà sottrarre la somma algebrica tra interessi passivi e attivi relativi alla gestione finanziaria ed infine prendere in considerazione l'accensione/estinzione di finanziamenti e l'apporto/rimborso di capitale proprio. Brusa L. ,Zamprogna L, (1998), *Finanza D'Impresa*, Etaslibri, pag. 44.

Gli indicatori hanno messo in luce risultati inferiori, con percentuali di classificazione fortemente degradanti col procedere a ritroso dell'anno di osservazione delle variabili di bilancio. Gli indicatori con la minore capacità diagnostica sono risultati quelli connessi al circolante ed alla liquidità, che tradizionalmente erano in quell'epoca considerati i più efficaci nella valutazione della capacità di credito delle imprese.

L'analisi del comportamento nel tempo delle distribuzioni degli indicatori ha messo in luce andamenti assai consistenti con i risultati attesi: le distribuzioni delle società sane si sono mantenute stabili nel tempo, mentre quelle riguardanti le società anomale hanno avuto un progressivo spostamento verso la parte peggiore dei valori con l'avvicinarsi al momento dell'insolvenza, riducendo l'area della sovrapposizione con le imprese sane.

In conclusione lo studio di Beaver ha dimostrato che i dati contabili rappresentano una fonte in grado di fornire informazioni utili per l'identificazione precoce del rischio di insolvenza (o di fallimento). Non tutti gli indicatori hanno la stessa capacità diagnostica: le variabili legate alla capacità di generazione di cassa e alla struttura finanziaria hanno una rilevanza informativa migliore, sotto il profilo della insolvenza, rispetto alle variabili espressive della liquidità a breve termine.

2. L'APPROCCIO MULTIVARIATO.

Uno dei limiti della ricerca di Beaver consiste nell'uso individuale delle variabili economico- finanziarie, equivalente a considerare separatamente vari elementi dell'impresa: la redditività, la struttura finanziaria, la liquidità e così via. Il passo successivo non può che essere quello di cercare di combinare insieme tutti i segnali che arrivano dalle diverse variabili e cercare di ottenere un segnale complessivo che individui in misura sintetica lo stato attuale di salute dell'impresa dal punto di vista dei creditori: le varie prospettive con cui può essere esaminata un'impresa vengono così analizzate simultaneamente anziché essere valutate sequenzialmente.

L'obiettivo finale non è ovviamente la concentrazione in un'unica informazione della pluralità di segnali che arrivano dai diversi indicatori, quanto quello di gestire in modo coordinato i trade-off che si instaurano tra le varie componenti del sistema-impresa. Una società, ad esempio, può essere migliore di un'altra in termini di redditività, ma molto peggiore per quanto riguarda la struttura finanziaria e lievemente peggiore in termini di liquidità: nel complesso, la prima è preferibile alla seconda o le è inferiore, oppure sono considerate equivalenti?

La risposta non può essere facile, né evidenti possono essere le argomentazioni a sostegno. Un indicatore composito, sulla base di specifici rapporti di trade-off, combina i tre aspetti dell'esempio precedente e consente di ottenere un'unica

misura di sintesi in cui i fattori di superiorità e di inferiorità siano tra loro compensati nello stesso modo (con gli stessi criteri) per le due società.

Imprese profondamente differenti possono sotto questo aspetto essere giudicate complessivamente equivalenti. Il punto cruciale ovviamente risiede nel modo con il quale ricavare i pesi relativi, ovvero i fattori di scambio, con i quali ponderare i diversi indicatori.

Anche se l'approccio multivariato è un innegabile avanzamento rispetto all'analisi univariata, quest'ultima tuttavia rappresenta uno strumento prezioso. Infatti, l'analisi del comportamento individuale degli indicatori costituisce uno dei primi passi per la corretta messa a punto di un modello multivariato. Di seguito saranno analizzati due metodi alternativi con cui effettuare un'analisi multivariata, l'analisi discriminante e la regressione logistica.

2.1 L'ANALISI DISCRIMINANTE LINEARE

L'analisi discriminante lineare venne proposta per la prima volta, nel 1936, da Fischer³ ed è un metodo statistico che permette di classificare, col minimo errore, un insieme di unità statistiche in due o più gruppi, individuati a priori, sulla base di un insieme di caratteristiche note.

Pertanto, tale metodologia di analisi risulta utile in situazioni in cui si desidera creare un modello di previsione del gruppo di appartenenza in base alle

³ Fischer R.A. (1936), *The Use Of Multiple Measurement In Taxonomic Problems*, *Annals of Eugenics*, V. 7, p. 179-188.

caratteristiche osservate su ciascun oggetto. La procedura genera una funzione discriminante oppure, per più di due gruppi, un insieme di funzioni discriminanti, in base alle combinazioni lineari delle variabili stimatorie che forniscono la migliore discriminazione tra i gruppi. Le funzioni vengono generate da un campione di casi di cui è noto il gruppo di appartenenza; è quindi possibile applicare le funzioni ai nuovi casi con misurazioni per le variabili stimatorie, ma di cui non è noto il gruppo di appartenenza.

Nell'ambito dei modelli di previsione delle insolvenze l'analisi discriminata consente una valutazione automatica delle aziende sottoposte ad analisi fornendo per ognuna un numero (score) che ne individua in misura sintetica lo stato di salute dal punto di vista dei creditori. Pertanto, tale metodologia si presta sia per problemi di *scelta* (concedere o bocciare una richiesta di fido) che per problemi di *classificazione* del portafoglio crediti (ordinare i clienti in grado alla solvibilità).

Al fine di poter applicare l'analisi discriminante per decidere quali richieste di fido debbano essere accolte o rigettate (da parte di una banca) si dovranno seguire i seguenti passi:

- disporre di un campione significativo di clienti che in passato si sono rivelati affidabili o insolventi “training set”;
- individuare una combinazione di variabili (es. debt/equity, ROI etc..) che abbiano un contenuto informativo sufficiente a discriminare piuttosto nettamente tra il gruppo dei prenditori sani e quello degli insolventi;

- una volta scelti il campione e le variabili, ognuna viene inserita in una funzione discriminante (media ponderata) che contribuisce a determinare lo score per ogni cliente;
- in base al punteggio ottenuto, l'azienda sarà considerata più o meno rischiosa. Per esempio, se le imprese affidabili presentano punteggi più generalmente alti e quelle insolventi punteggi bassi, si ritiene che il cliente che abbia ottenuto uno score molto alto sarà “probabilisticamente” solvibile mentre quello con uno score basso si rileverà con una buona probabilità insolvente.;
- verificare il modello e gli eventuali errori di classificazione mediante l'uso di un campione di verifica “validation set”;
- una volta calcolati i punteggi e le relative fasce di rischio, è possibile che la banca stabilisca un valore soglia (cut-off score), al di sotto del quale le richieste di fido vengano respinte o sottoposte a revisione.

Pertanto, ogni volta che si presenterà una nuova richiesta di fido, occorrerà calcolare, utilizzando la funzione discriminante ottenuta, lo score per l'impresa ed applicare la regola decisionale di cui sopra.

Nella figura 1. viene data un'interpretazione geometrica del modello discriminante lineare, per il caso di due variabili discriminatorie e due popolazioni.

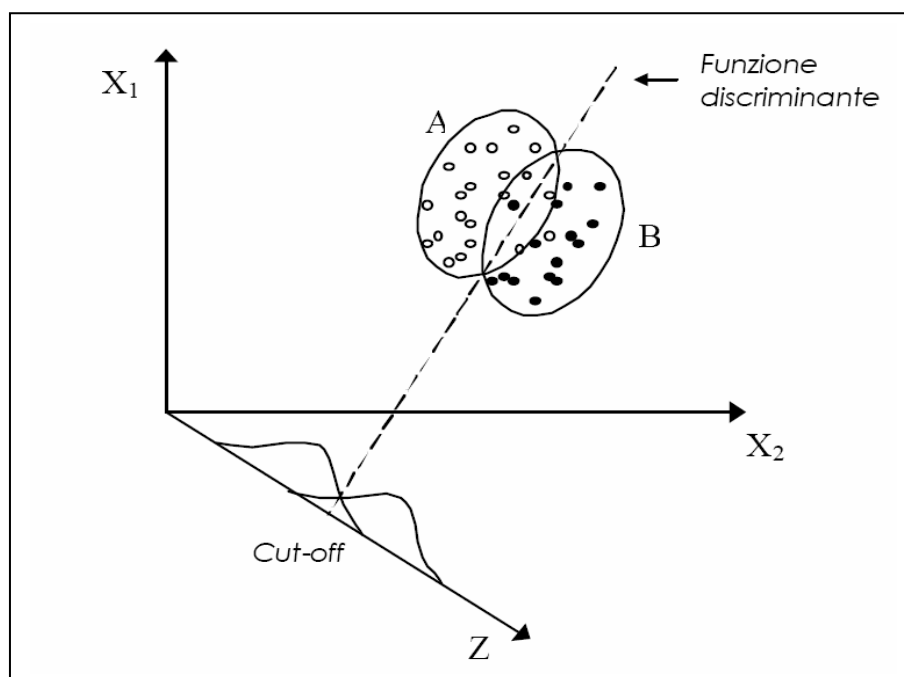


Figura 1 – Sintesi grafica dell’analisi discriminante lineare⁴

Sul piano X_1, X_2 sono riportate le imprese appartenenti ai due campioni delle popolazioni A e B. I due insiemi non sono nettamente separati, ma hanno una certa sovrapposizione: con l’analisi discriminante si individua quella funzione che meglio separa i due insiemi, ovvero che commette il minor numero di errori di attribuzione; tale retta ha la proprietà notevole che le proiezioni delle nuvole

⁴ Omacini C. (2001), *La previsione del rischio di insolvenza con modelli quantitativi*, in Resti A. “Misurare e gestire il rischio di credito nelle banche: una guida metodologica” ,Alpha Test, Milano, 2001.

dei punti sulla retta Z , perpendicolare ad essa, disegna delle distribuzioni con la minor area di sovrapposizione; la funzione discriminante rappresenta il luogo delle combinazioni lineari delle variabili, ovvero rappresenta la funzione discriminante ottima, date le caratteristiche X_1 e X_2 .

Come si vede, l'analisi discriminante lineare semplifica grandemente l'analisi delle distanze tra le imprese in un contesto multivariato, grazie alla riduzione della dimensione delle caratteristiche osservate, ovvero grazie al passaggio dello spazio da n dimensioni delle variabili a 1 dimensione della linea dei punti Z (la riduzione dimensionale in realtà passa da n a $g-1$ ove g è il numero delle popolazioni; nel caso in esame $g=2$).

Le imprese da classificare sono rappresentate da punti sulla funzione discriminante, sulla base dei quali è immediato e non equivoco effettuare degli ordinamenti. Si osservi anche che la scelta dei pesi (della funzione discriminante) non è effettuata soggettivamente dall'analista, ma è oggettiva e dipende dalle caratteristiche delle due popolazioni: l'elemento soggettivo dell'analista finanziario entra in gioco nella scelta delle variabili con le quali osservare le imprese.

Più è ampia l'area della sovrapposizione tra le due distribuzioni, maggiore è l'incertezza della classificazione; nel caso limite di due distribuzioni nettamente separate non vi sono errori, nell'altro caso di perfetta sovrapposizione vi è la massima incertezza di attribuzioni: le caratteristiche osservate non forniscono alcun elemento per individuare l'appartenenza di un oggetto alla popolazione.

Per utilizzare l'analisi discriminata come *metodo di classificazione* del portafoglio crediti di una banca, in luogo della regola decisionale basata sullo score di cut-off, occorrerà calcolare lo score di tutte le imprese presenti nel portafoglio, ordinarle in funzione decrescente dello score, ed al fine di creare le classi rating tradurre lo score in probabilità di Default.

2.1.1 IL MODELLO

Si supponga di avere un training set composto da due gruppi di imprese A (imprese insolventi) e B (imprese sane), di numerosità N_A e N_B , riguardanti le due popolazioni note a priori; inoltre, si supponga di aver trovato un vettore di n variabili X , che abbiano un contenuto informativo sufficiente a discriminare piuttosto nettamente tra il gruppo dei prenditori sani e quello degli insolventi, allora la funzione discriminante assumerà la seguente forma:

$$Z_j = a_1 X_{1j} + a_2 X_{2j} + \dots + a_i X_{ij} + \dots + a_n X_{nj}$$

Ovvero, in con notazione matriciale:

$$Z = \alpha^T X$$

Dove:

Z = valore discriminante calcolato per ogni singola impresa;

α = vettore degli n coefficienti di discriminazione (pesi);

X = vettore delle n variabili discriminatorie.

Il vettore dei pesi α viene trovato mediante un processo di ottimizzazione che ha come obiettivo quello di garantire che i valori discriminanti così ottenuti massimizzeranno la differenza complessiva tra i due gruppi di imprese (insolventi e sane). In altri termini, il vettore degli n pesi α_i , sarà quel vettore che renderà massimo il rapporto tra la distanza tra le medie degli score dei due gruppi (varianza tra i gruppi) e la varianza degli score all'interno dei due gruppi (varianza entro i gruppi). In pratica, se valutate attraverso i valori discriminanti, le imprese "buone" saranno il più possibile simili tra loro e il più possibile diverse dalle "cattive". Pertanto:

$$\alpha^T = (\bar{X}_A - \bar{X}_B)^T \Sigma^{-1}$$

dove \bar{X}_A e \bar{X}_B sono i vettori delle medie dei gruppi A e B calcolate sulle variabili discriminatorie X e Σ^{-1} rappresenta la matrice inversa di varianza e covarianza. Come accennato sopra, costruire la "migliore" funzione discriminante equivale a scegliere il vettore dei pesi α in modo che sia massima la distanza tra le medie dei due gruppi pesata per la varianza, cioè:

$$\max d = \frac{\alpha^T \bar{X}_A - \alpha^T \bar{X}_B}{\alpha^T \Sigma \alpha}$$

Calcolando la derivata prima di questa espressione e uguagliandola a zero si ottiene:

$$\frac{\delta d}{\delta \alpha} = \frac{2(\alpha^T \bar{X}_A - \alpha^T \bar{X}_B)(\bar{X}_A - \bar{X}_B)\alpha^T \Sigma \alpha - 2\Sigma \alpha (\alpha^T \bar{X}_A - \alpha^T \bar{X}_B)^2}{(\alpha^T \Sigma \alpha)^2} = 0$$

$$(\bar{X}_A - \bar{X}_B)\alpha^T \Sigma \alpha - \Sigma \alpha (\alpha^T \bar{X}_A - \alpha^T \bar{X}_B) = 0$$

$$(\bar{X}_A - \bar{X}_B) = \Sigma \alpha \left(\frac{(\alpha^T \bar{X}_A - \alpha^T \bar{X}_B)}{\alpha^T \Sigma \alpha} \right)$$

Con $\frac{(\alpha^T \bar{X}_A - \alpha^T \bar{X}_B)}{\alpha^T \Sigma \alpha}$ costante, di conseguenza si ottiene:

$$\alpha^T = (\bar{X}_A - \bar{X}_B)^T \Sigma^{-1}$$

Il valore di cut-off, ovvero il punteggio ottimo per la separazione tra i gruppi A (imprese insolventi) e B (imprese sane) sarà dato da:

$$Z_C = \frac{(\bar{X}_A - \bar{X}_B)^T \Sigma^{-1} \bar{X}_A + (\bar{X}_A - \bar{X}_B)^T \Sigma^{-1} \bar{X}_B}{2} = \frac{\bar{Z}_A + \bar{Z}_B}{2}$$

Pertanto l'utilizzatore del modello acquisisce gli score campionari ottenuti dal training set e decide l'assegnazione delle società da valutare in base al confronto tra essi e la soglia di cut-off trovata. Quindi la regola decisionale diventa: "assegna un'impresa al gruppo A se $Z < Z_C$, altrimenti assegnala al gruppo B (supponendo che $\bar{Z}_A < \bar{Z}_B$).

Se i gruppi non hanno uguale dimensione e si assume che siano rappresentativi delle proporzioni esistenti nelle due popolazioni, il punto di cut-off si ottiene come media ponderata dei centroidi.

$$Z_C = \frac{\bar{Z}_A N_A + \bar{Z}_B N_B}{N_A + N_B}$$

2.1.2 ESTENSIONI DEL MODELLO

Un'importante estensione nell'ambito dei modelli parametrici riguarda il caso della classificazione ricorrendo al criterio della massima verosimiglianza⁵: l'impresa j-esima viene attribuita alla popolazione h-esima tale che sia massima la probabilità ($p_h(x_j)$) che l'impresa sia generata da quella popolazione.

⁵ Zsegò G., Varetto F. (1999), *Il rischio creditizio*, Utet Torino.

Nel caso in cui le popolazioni siano multinormali, il criterio di classificazione conduce a una funzione discriminante quadratica; se, inoltre, le popolazioni hanno la stessa matrice di varianza covarianza il modello si semplifica notevolmente e si riduce al caso di funzione discriminante lineare.

Un'ulteriore estensione del modello⁶ base incorpora la conoscenza delle probabilità a priori delle diverse popolazione e i costi di errata classificazione. Siano $p_A(X)$ e $p_B(X)$ le probabilità (o densità di probabilità), note, che le due popolazioni generino l'impresa osservata; siano q_A e q_B le probabilità a priori che una generica impresa osservata provenga rispettivamente dalla popolazione A e dalla popolazione B.

Le probabilità a posteriori sono calcolabili ricorrendo al teorema di Bayes:

$$p(A|X) = \frac{q_A \cdot p_A(X)}{p(X)}$$

$$p(B|X) = \frac{q_B \cdot p_B(X)}{p(X)}$$

$$\text{dove } p(X) = \sum_{r=A}^B q_r \cdot p_r(X) = q_A \cdot p_A(X) + q_B \cdot p_B(X)$$

⁶ Varetto F. (1990), *Il sistema di diagnosi dei rischi di insolvenza della Centrale dei Bilanci*, Bancaria Editrice, Roma.

La probabilità a posteriori $p(A|X)$ indica la probabilità che, data l'osservazione delle caratteristiche X sull'impresa esaminata, quest'ultima sia generata dalla popolazione A. L'impresa viene quindi attribuita alla popolazione A se:

$$p(A|X) > p(B|X)$$

ovvero se

$$q_A \cdot p_A(x) > q_B \cdot p_B(x)$$

e cioè se

$$\frac{P_A(X)}{P_B(X)} > \frac{q_B}{q_A}$$

Rimanendo nel caso di distribuzioni multi normali, $p_A(X)$ e $p_B(X)$ sono definibili come funzioni di densità di probabilità normali nelle variabili X . Con le consuete semplificazioni delle matrici di varianza e covarianza uguali tra le popolazioni questo criterio converge ad una funzione discriminante lineare nella quale il valore critico (cut-off di attribuzione) è spostato della quantità $\ln(q_A/q_B)$, rispetto alla funzione originale di Fisher ($\ln =$ logaritmo naturale). Nel caso limite in cui le probabilità a priori siano uguali ($q_B=q_A$), con l'inserimento nel modello

di tali probabilità, conservando le altre semplificazioni, non si producono spostamenti nella funzione discriminante lineare.

I costi di errata classificazione complicano ulteriormente il modello, ma consentono di aumentare il grado di realismo: l'errore di decisione infatti è diverso se si tratta di una società sana rispetto ad una società anomala.

Gli errori che si possono commettere sono di due tipi: classificare sana un'impresa in realtà anomala e classificare anomala un'impresa in realtà sana.

Non c'è dubbio che il primo tipo di errore sia molto più costoso del secondo: nel primo caso il finanziatore va incontro alla perdita totale o parziale degli interessi e del capitale, oltre a dover sostenere oneri legali e amministrativi per la gestione dell'insolvenza (o del fallimento) e la riscossione di eventuali garanzie; nel secondo caso il costo è sostanzialmente dato dai redditi connessi all'opportunità di affari che si è persa considerando anomala la società.

Si consideri genericamente C_{AB} il costo di classificare nella popolazione B l'impresa in realtà proveniente dalla popolazione A e C_{BA} per il caso opposto.

Il criterio di decisione diventa pertanto quello di minimizzare il costo atteso degli errori di attribuire l'impresa esaminata alla popolazione A se

$$\frac{P_A(X)}{P_B(X)} > \frac{q_B C_{BA}}{q_A C_{AB}}$$

In cui $p_A(X)$ e $p_B(X)$ sono anche definibili in termini di funzioni di densità normali multivariate. Mantenendo l'ipotesi di uguaglianza tra le popolazioni

delle matrici di varianza e covarianza, la regola decisionale si semplifica in una funzione discriminante lineare il cui cut-off è spostato di una quantità pari a:

$$Ln \frac{q_B C_{BA}}{q_A C_{AB}}$$

rispetto alla funzione di Fisher.

Quindi si può pensare alla semplice discriminante lineare multivariata come un caso particolare di un criterio di classificazione più generale per il quale valgono le ipotesi di normalità multivariata delle distribuzioni delle variabili, uguaglianza delle matrici di varianza e covarianza tra le popolazioni, identiche probabilità a priori e costi di classificazione uguali (ovvero: nella messa a punto del modello vengono ignorate sia le probabilità a priori che i costi di errate classificazioni).

Se si rimuove l'ipotesi di uguaglianza delle matrici di varianza e covarianza la regola di classificazione si trasforma in una funzione discriminante quadratica. Le due grandezze q_B e q_A in termini semplicistici possono essere considerate come le proporzioni relative delle due popolazioni ovvero, quando non vengono specificate, come la dimensione relativa dei due campioni di società.

Per questo, nelle applicazioni pratiche, quando vengono utilizzati campioni di pari numerosità di imprese (come nel caso di campioni di sane e anomale tra loro “pareggiate” per anno, settore e classe dimensionale) e non vengono specificate le probabilità a priori né i costi di errate classificazioni, la funzione

discriminante che si ottiene ha un cut-off centrato sullo zero e la funzione converge verso la semplice funzione lineare di Fisher.

2.1.3 IL MODELLO Z SCORING

Altman⁷, alla fine degli anni sessanta, ha applicato con successo l'analisi discriminante lineare alla previsione delle insolvenze: anche se questa tecnica era stata impiegata in studi precedenti, la pubblicazione dell'articolo di Altman ha aperto la strada a un nutrito filone di critiche ed estensioni, oltre ad applicazioni in molti paesi.

Nella sua prima ricerca, pubblicata nel 1968, Altman ha applicato la versione più semplice della tecnica di analisi discriminante lineare a un campione di 33 imprese industriali fallite nel periodo tra il 1945 -1965 e a un campione "pareggiato" (per anno, settore e dimensione di attivo netto totale) di società sane, estratte casualmente dagli elenchi di Moody's e di altre fonti.

Il modello ottenuto, probabilmente il più citato nella letteratura in materia, è il seguente:

⁷ E. Altman (1968), *Financial Ratios, Discriminant Analysis and The Prediction of Corporate Bankruptcy*, The Journal Of Finance, Vol XXIII, No. 4

$Z = 0,012 * \text{capitale circolante/attivo netto}$

$+0,014 * \text{riserve da utili/attivo netto}$

$+0,033 * \text{utile ante interessi e tasse/attivo netto}$

$+0,006 * \text{valore di mercato del patrimonio netto/dediti totali}$

$+0,999 * \text{ricavi/attivo netto}$

La funzione include diverse componenti del sistema economico finanziario dell'impresa: la liquidità e l'equilibrio a breve termine, la redditività cumulata, la redditività corrente, la struttura finanziaria e l'efficienza complessiva. Tutti i coefficienti delle variabili discriminatorie hanno il segno atteso, in quanto le variabili discriminatorie create sono positivamente correlate allo stato di salute dell'impresa. Di seguito l'interpretazione delle cinque variabili discriminatorie prese in considerazione da Altman:

- “capitale circolante/attivo netto” esprime il valore delle attività liquide dell'azienda rispetto alla capitalizzazione totale. Risulta evidente che una società che va incontro a perdite operative consistenti avrà una forte riduzione delle attività correnti in relazione al totale delle attività.
- “riserve di utili/attivo netto” esprime la capacità che un'azienda ha avuto di reinvestire i propri utili. Un'azienda giovane avrà certamente un indice minore rispetto ad un'azienda di più antica costituzione; questo perché l'azienda giovane non ha avuto ancora il tempo di costituire le proprie

riserve e, pertanto può risultare penalizzata nella valutazione del rischio di fallimento. Ciò rappresenta proprio la situazione reale nella quale le società neo costituite hanno una probabilità di fallimento maggiore nei primi anni della loro vita.

- “utile ante interessi e tasse/attivo netto” misura la vera produttività delle attività di un’impresa, depurate da qualsiasi fattore di leva finanziaria o fiscale. Per tale motivo detto indice risulta particolarmente appropriato nella definizione della probabilità di insolvenza e successivo fallimento.
- “valore di mercato del patrimonio netto/dediti totali” mostra di quanto le attività di un’azienda si possono ridurre prima che le passività totali eccedano le attività e si creino le condizioni per il fallimento. Per esempio, una società con un patrimonio netto pari a 1.000 € e passività per 500 € può sopportare una perdita del valore di due terzi del proprio attivo prima di divenire insolvente. Invece, se la stessa azienda avesse un patrimonio netto pari a 250 € con lo stesso ammontare di passività, diverrebbe insolvente con una riduzione di solo un terzo del proprio attivo.
- “ricavi/attivo netto” evidenzia la capacità di un’azienda di generare ricavi con un determinato valore dell’attivo patrimoniale. Esso misura la capacità imprenditoriale di rapportarsi con la competitività del mercato di riferimento dell’azienda.

La capacità diagnostica di questa funzione valutata nell’anno immediatamente precedente all’insolvenza è stata molto buona: in media, il 95% delle imprese

sono state classificate correttamente, con diversa entità dei due tipi di errori; l'errore di primo tipo, consistente nel classificare come sana un'impresa anomala è stato del 6%, mentre l'errore di secondo tipo, riguardante la classificazione di un'impresa sana tra le anomale, è stato del 3%.

Valutata rispetto agli anni precedenti l'insolvenza dal secondo al quinto, la funzione mette in luce un progressivo deterioramento (peggiore di quello di Beaver) delle capacità diagnostiche, sia sul campione originale di stima sia sui campioni di controllo .

Naturalmente è giusto attendersi una riduzione dell'efficacia del modello nel classificare le imprese man mano che si risale indietro nel tempo rispetto al momento dell'insolvenza: infatti, salvo nei casi di prolungata crisi, misurati in vari anni prima dell'insolvenza, la distanza tra le società sane e quelle anomale tende a ridursi e le differenze tra i due insiemi si attenuano. La robustezza e stabilità del modello deve essere valutata anche alla luce della velocità di degrado della performance a ritroso.

La classificazione delle società nel modello di Altman avviene confrontando lo score calcolato sulle variabili dell'impresa e un intervallo di cut-off; l'autore infatti, in luogo di determinare lo score di cut-off come media dei centroidi dei due campioni, individua un'area grigia (o zona di ignoranza) nella quale gli errori di classificazione sono più elevati, corrispondente all'intervallo 1,81; 2,99 dello score, con 2,675 come valore puntuale di cut-off (per valori superiori alla soglia

dell'area grigia l'impresa è considerata sana). Altman ha successivamente modificato il modello originale per ampliarne il campo di applicabilità⁸.

In particolare:

1. l'indicatore di struttura finanziaria è stato calcolato utilizzando il valore contabile del patrimonio netto, anziché il valore di mercato, per rendere il modello (così ristimato) utilizzabile per le società non quotate;
2. il modello è stato ristimato senza l'indicatore di turnover (ultima variabile) per adattarlo alle società non industriali: questa variabile infatti incorpora in maggiore misura le influenze dell'appartenenza delle imprese ai settori industriali;
3. un ulteriore aggiustamento è stato effettuato per adattare il modello ai rischi di credito dei paesi emergenti (il Messico nel caso specifico), cercando di correlare il più possibile gli score della funzione discriminante con le classi di rating definite sulle obbligazioni statunitensi.

Nel 1977 Altman insieme ad altri autori⁹ ha messo a punto un nuovo modello dello Zeta, basato anche su alcune critiche alla ricerca iniziale Z ricevute da vari studiosi. Il nuovo modello, stimato su un campione "pareggiato" di 53 società fallite e 58 sane (5 società anomale non disponevano di dati sufficienti), composto quasi in ugual misura da imprese industriali e da imprese commerciali, ha posto maggiore attenzione a vari aspetti:

⁸ Caouette J., Altman E. e Narayann P. (1998), *Managing Credit Risk*, J. Wiley, New York.

⁹Altman E., Hadelman R. e Narayann P. (1977), *Zeta Analysis*, Journal Of Banking and Finance n.1.

- prima del calcolo degli indicatori sono stati condotti alcuni aggiustamenti ai dati di bilancio per renderli più espressivi della effettiva realtà aziendale: la correzione più importante è stata la capitalizzazione dei contratti di leasing operativo e finanziario, ma rettifiche sono state anche condotte sulle riserve, sul capitale di terzi minoritari, sul consolidamento delle consociate finanziarie, sulle attività immateriali, avviamenti e altre spese capitalizzate;
- è stata controllata l'eguaglianza della matrice di varianza-covarianza dei due campioni: accettata la diversità, è stata utilizzata l'analisi discriminante quadratica;
- l'analisi dell'importanza relativa dei diversi indicatori che compongono il modello è stata effettuata ricorrendo a 6 test diversi;
- sono state definite delle probabilità a priori e una stima dei costi di errata classificazione.

Il nuovo modello Zeta è stato realizzato con sette variabili discriminatorie:

1. ROA, misurato come rapporto tra utili ante interessi e tasse e l'attivo totale;
2. Stabilità degli utili, calcolata con una misura normalizzata dello scarto quadratico medio della stima intorno al trend decennale del ROA;
3. Servizio del debito, valutato come il rapporto tra utili ante interessi e tasse e gli oneri finanziari totali; per aumentare la normalità della distribuzione, questa variabile è stata trasformata con il logaritmo decimale;

4. Redditività cumulata, misurata dal rapporto tra riserve di utili e attivo netto;
5. Liquidità, calcolata in base al tradizionale indicatore di liquidità corrente;
6. Capitalizzazione, misurata col rapporto tra il valore di mercato del patrimonio netto (media dei prezzi delle azioni degli ultimi 5 anni) e del totale dell'indebitamento;
7. Dimensione, misurata dal logaritmo dell'attivo netto.

Sulla base degli esperimenti effettuati, Altman ha trovato che la discriminante quadratica e quella lineare hanno dato grosso modo gli stessi risultati; quest'ultima, in particolare, è risultata più accurata nella classificazione sul campione di test. Pertanto l'intero modello è stato stimato con la tradizionale metodologia lineare, benché dal punto di vista teorico fosse preferibile quella quadratica.

La funzione lineare dei 7 indicatori, la cui formula esatta è riservata, ha classificato, correttamente nell'anno t-1 il 96,2% delle società fallite e l'89,7% delle sane; risalendo al t-5 l'accuratezza complessiva del modello è dell'ordine del 70% circa.

Il punto ottimale di cut-off è stato definito assegnando le probabilità a priori e i costi degli errori di classificazione¹⁰:

¹⁰ Altman E. (1984), *A further empirical investigation of the bankruptcy cost question*, Journal Of Finance

$$\text{Zeta di cut off} = \ln \frac{q_A C_1}{q_S C_2}$$

In cui q_A e q_S sono le probabilità a priori che ha un'impresa di fallire o di essere sana e C_1 e C_2 sono i diversi costi dell'errore di primo e di secondo tipo (impresa anomala considerata sana e viceversa).

Il costo atteso dell'uso del modello Zeta ai fini decisionali è pertanto:

$$EC(\text{Zeta}) = q_A C_1 \cdot \frac{M_{A,S}}{N_A} + q_S C_2 \cdot \frac{M_{S,A}}{N_S}$$

ove N_A e N_S rappresentano la numerosità dei campioni delle imprese anomale e sane e $M_{A,S}$ e $M_{S,A}$ sono il numero delle imprese classificate erroneamente. Le probabilità a priori assegnate sono rispettivamente:

$$q_A = 2\% \text{ e } q_S = 98\%$$

I costi degli errori sono stati stimati pari al 70% per C_1 e 2% per C_2 . Pertanto il cut-off accettato nel modello ammonta a :

$$\text{Zeta di cut - off} = \ln \frac{0,02 \cdot 70}{0,98 \cdot 2} = -0,337$$

intorno al quale Altman ha condotto un'analisi di sensitività.

Lo spostamento del cut-off da zero (cut-off originale della funzione lineare in assenza di correzioni per le probabilità a priori e per i costi degli errori) al nuovo valore ha l'effetto di peggiorare il tasso di riconoscimento delle società anomale e di migliorare quello delle sane.

Come si vede in questo nuovo modello, le cui capacità diagnostiche sono migliori di quello iniziale, Altman ha separato la fase della messa a punto della funzione discriminante (funzione lineare con cut – off pari a zero, senza correzioni per probabilità a priori e costi d'errore) da quella dell'utilizzo decisionale della funzione stessa: quest'ultima fase, consistente nel semplice spostamento dell'intercetta della funzione, ovvero del valore di cut – off, può essere effettuata direttamente dall'utilizzatore sulla base delle proprie aspettative a priori e sui propri costi degli errori.

In tal modo viene separato il lavoro del ricercatore (produzione della funzione e degli score campionari) da quello dell'utilizzatore (analista del credito) che acquisisce gli score campionari e decide l'assegnazione della società in base al confronto tra essi e la soglia di cut – off definita sugli specifici elementi dell'utilizzatore stesso (probabilità e costi).

2.2 LA REGRESSIONE LOGISTICA

In questo paragrafo si dà una breve illustrazione dell'utilizzo della funzione logistica in alternativa all'analisi discriminante lineare (o quadratica)¹¹: infatti il modello logistico è stato applicato in un numero rilevante di studi, specie in anni recenti.

Come si è visto in precedenza, l'analisi discriminante lineare consiste nell'individuare la migliore combinazione lineare di indicatori in grado di distinguere al meglio due insiemi di società. Si può dimostrare che vi sono relazioni strette tra l'analisi discriminante lineare e la regressione lineare: i coefficienti della funzione lineare sono pari a quelli della regressione con i minimi quadrati ordinari a meno di un rapporto costante. In effetti anche la regressione multipla è stata utilizzata in alcune ricerche sul rischio di credito. Questo approccio consiste nello stimare un modello che ha come dipendente una variabile qualitativa (dicotomica) che descrive l'appartenenza all'insieme delle società sane o anomale:

$$Y = \begin{cases} 0 & \text{se impresa = sana} \\ 1 & \text{se impresa = anomala} \end{cases}$$

mentre gli indicatori di bilancio sono le variabili indipendenti.

¹¹ Lo A. (1986), *Logit versus discriminant analysis*, Journal of Econometrics.

Una versione particolare della regressione, il linear probability model, interpreta la y come probabilità di appartenenza al gruppo. Questo procedimento comporta alcune difficoltà: la varianza degli errori della stima non è costante, determinando un problema di eteroschedasticità; esso può essere risolto con una procedura a due stadi, ma ciò non risolve altre questioni, come la non-normalità degli errori. Un altro problema riguarda il fatto che la stima della y non determina valori compresi tra 0 ed 1, come sarebbe logico per interpretare i risultati in termini di probabilità: valori stimati negativi o molto maggiori di 1 creano evidentemente difficoltà interpretative in termini probabilistici. Più i valori stimati si allontanano dall'intervallo $[0;1]$, più gli errori della stima aumentano.

Invece, una caratteristica notevole del modello logistico (logit) consiste, contrariamente al modello lineare, nell'ottenere dei valori che appartengono tutti monotonamente all'intervallo $[0;1]$. Naturalmente il modello logistico non è l'unico in grado di produrre questi valori limitati, ma alcune particolarità matematiche lo rendono più facilmente manipolabile e quindi più adottato dagli studiosi.

L'idea che sta al di sotto del modello logistico consiste nel supporre che esista una relazione tra la probabilità di un'impresa di diventare insolvente (variabile inosservabile) ed una serie di grandezze osservabili che sono strettamente connesse con l'evento insolvenza¹². Ciò che si osserva nella realtà quindi non è la

¹² Lawrence E., Arshadi N. (1995), *A multinomial logit analysis of problem loan resolution choices in banking*, Credit and Banking.

probabilità di insolvenza (che può essere considerata come una variabile latente), ma una realizzazione dicotomica di tale probabilità.

Identificate con p le probabilità di insolvenza, con X il vettore delle variabili indipendenti e con α e β il termine costante ed i coefficienti del modello si ha:

$$p = F(\alpha + \beta X)$$

ove F identifica la funzione standard cumulativa logistica:

$$F(\alpha + \beta X) = \int_{-\infty}^{\alpha + \beta X} f(h) dh = \frac{1}{1 + e^{-(\alpha + \beta X)}}$$

ove $f(h)$ indica la funzione di densità logistica

$$f(h) = \frac{e^h}{(1 + e^h)^2}$$

L'ipotesi forte del modello riguarda quindi la forma della distribuzione di probabilità di insolvenza. Dalla relazione:

$$p = \frac{1}{1 + e^{-(\alpha + \beta X)}}$$

si ha:

$$e^{-(\alpha+\beta X)} = \frac{1-p}{p}$$

ovvero

$$e^{(\alpha+\beta X)} = \frac{p}{1-p}$$

in cui il termine di destra rappresenta l' "odd-ratio" (cioè il rapporto tra le probabilità dell'evento ed il suo complemento). Prendendo il logaritmo naturale si ottiene:

$$\ln \frac{p}{1-p} = \alpha + \beta X$$

Nel linear probability model è p ad essere messo in relazione con $a + bX$, mentre nel modello logistico, come si vede, è il logaritmo dell'odd-ratio: in entrambi i casi le variabili esplicative sono connesse alla dipendente con una funzione lineare. Quella relazione è equivalente a considerare:

$$\ln \frac{p_A(X)}{p_B(X)} = \alpha + \beta X$$

ove p_A e p_B sono le due densità di probabilità delle popolazioni A e B.

Pertanto, applicando il teorema di Bayes secondo le stesse linee esaminate per l'analisi discriminante, si attribuisce l'osservazione alla popolazione A se:

$$\ln \frac{p_A(X)}{p_B(X)} > \ln \frac{q_B}{q_A}$$

e quindi:

$$\alpha + \beta X > \ln \frac{q_B}{q_A}$$

Dalle relazioni precedenti emerge che quando $\alpha + \beta X = 0$, $e^{(\alpha + \beta X)} = 1$ e quindi $p = 1 - p$, ovvero $p = 0,5$: il valore di cut-off, nel caso più semplice, si ha quando vi è perfetta incertezza in termini probabilistici, cui corrisponde un valore nullo dell'esponente della funzione logistica cumulata.

Un modello assai simile a quello logistico è il probit. In quest'ultimo, l'ipotesi chiave riguarda la forma della distribuzione cumulata delle probabilità di insolvenza: invece di assumere che la forma della distribuzione sia la logistica cumulata, si ipotizza che essa sia la normale standardizzata cumulata:

$$F(\alpha + \beta X) = \int_{-\infty}^{\alpha + \beta X} f(h) dh = \frac{1}{\sqrt{2\pi}} e^{-\frac{h^2}{2}}$$

Benché le due distribuzioni siano diverse, i risultati sono tra di loro prossimi; la distribuzione normale tuttavia ha un grado di difficoltà nel trattamento matematico superiore alla logistica e pertanto nelle applicazioni è quest'ultima ad essere utilizzata in prevalenza. Prima di passare all'analisi critica dei diversi modelli proposti nella letteratura, è importante sottolineare la profonda diversità concettuale che separa l'analisi discriminante dal modello logistico.

L'analisi discriminante ipotizza implicitamente che le imprese osservabili siano tratte da due universi distinti dati; la rilevazione delle variabili di bilancio sulle imprese può essere di aiuto per trovare le caratteristiche rilevanti e per individuare da quale universo esse provengono effettivamente. L'analisi discriminante cerca pertanto di prevedere l'appartenenza a un gruppo, dopo aver osservato le variabili ritenute rilevanti per caratterizzare le diversità tra i due universi.

Il modello logistico (o probit ed altri simili), invece, come anche la regressione multipla, ipotizza che le imprese siano tratte casualmente da un unico universo cui appartengono e cercano di stimare una caratteristica specifica di tali imprese: il grado di salute ovvero la probabilità (logistica, normale, lineare) di insolvenza/fallimento. Tale caratteristica è immaginabile come una variabile latente continua, di cui sono osservabili solo due essenziali determinazioni estreme 0 e 1.

Questi modelli, quindi, ipotizzano che vi sia una relazione causale tra le variabili osservate sui dati contabili e la variabile dipendente; ciò significa che questi modelli implicitamente suppongono una relazione di causa-effetto tra i fenomeni economici sintetizzati dalle variabili di bilancio (e di cui queste ultime costituiscono delle proxy) e lo stato di salute dell'impresa. Questa famiglia di modelli, quindi, non stima l'appartenenza dell'impresa ad un gruppo, ma il grado dello stato di difficoltà economico- finanziaria in cui versa l'impresa.

Poiché i presupposti sono diversi, anche l'interpretazione del sistema di variabili e coefficienti che compongono i modelli è differente tra l'analisi discriminante e gli altri approcci.

Nell'analisi discriminante non si stima un modello esplicativo dell'insolvenza, ma si cerca di combinare insieme diverse variabili per avere un segnale unico, complesso, dell'appartenenza probabile ad un gruppo, dato a priori: gli indicatori che compongono la funzione discriminante (lineare o quadratica) vanno interpretati come segnali individuali che giustificano la loro presenza per il contributo marginale che danno al segnale complessivo.

Nella regressione multipla, come nella logistica, invece gli indicatori rappresentano le variabili esogene che sono funzionali a spiegare la situazione dell'impresa dal punto di vista del creditore: gli indicatori giustificano la loro presenza nella misura in cui concorrono ad individuare le varie componenti del modello economico- finanziario che spiega lo stato di crisi dell'impresa o/e la sua evoluzione o/e il suo deterioramento.

L'analisi discriminante tratta gli indicatori più come “segnali” di stato che come proxy quantitative di strutture e di risultati gestionali, in grado di sintetizzare i meccanismi economici sottostanti al sistema impresa.

3. AMBITI DI APPLICAZIONE

Oltre al tradizionale campo della valutazione del rischio di fallimento di imprese non finanziarie, l'analisi discriminante e la famiglia delle regressioni parametriche (lineare, logistica, probit) sono state applicate, con più o meno elevato successo:

- alla classificazione dei finanziamenti bancari¹³: i dati sui finanziamenti bancari, non essendo debiti trattati sul mercato finanziario, sono particolarmente difficili da ottenere in quantità tali da consentire stime applicabili dei modelli. Gli analisti di credito generalmente adottano una classificazione dei finanziamenti in base a grandi categorie di rischio (rating interno alla banca); le categorie sono più o meno numerose a seconda del grado di dettaglio e di analiticità voluti;
- alla classificazione dei prestiti personali: la tipologia dei problemi è simile a quella esaminata nel punto precedente, ma l'analisi di questa tipologia di

¹³ Altman E., Avery R., Eisenbeis R., Sinkey J. (1981), *Application of classification techniques in business, banking and finance*, Jai Press, NY

prestiti avviene tipicamente ricorrendo a questionari con valutazioni di tipo anagrafico, patrimoniale, occupazionale, e così via;

- all'analisi del Rating delle Obbligazioni: è questo uno dei campi tipici dell'applicazione delle tecniche di classificazione, con le quali si cerca di riprodurre, prevalentemente con dati di bilancio, lo schema concettuale adottato dalle società di rating per la valutazione del debito sia a breve termine sia a medio e lungo termine, negoziato sul mercato;
- alle caratteristiche delle insolvenze nelle obbligazioni ad alto rischio e alto rendimento.

4. CONCLUSIONI

Un primo problema di fondo, che vale per tutte le metodologie e non solo per l'analisi discriminante, riguarda la critica dell'assenza di una teoria di riferimento. I diversi modelli vengono messi a punto sostanzialmente in modo euristico, scegliendo gli indicatori e le altre variabili che forniscono i risultati più soddisfacenti in termini di performance e di stabilità dei risultati. Questa selezione è frutto di un processo di ricerca puramente empirico, con adattamenti che dipendono spesso dalle capacità individuali dei singoli ricercatori, e non si basa invece su una teoria dell'insolvenza (o della crisi) dell'impresa: lo testimonia la pluralità di modelli proposti nella letteratura, con variabili assai

spesso diverse tra loro (anche se spesso solo nel contenuto di calcolo) e nel diverso grado di importanza attribuito.

Il rischio di questa situazione consiste nel generare modelli che sono “sample-specific”, senza un’effettiva generalizzabilità: su quest’ultimo punto la terapia, parziale, in attesa di una teoria, consiste nel lavorare con campioni sempre più ampi e rappresentativi del reale comportamento della popolazione.

In secondo luogo, i modelli esaminati nel presente capitolo possono essere considerati come una black-box il cui risultato deve essere accettato perché è l’autorità di colui che produce la funzione e gli score campionari (ricercatore) a garantire che il risultato è quello giusto. Pertanto, al fine di soddisfare al meglio le esigenze del decisore (analista finanziario), occorrono metodologie più trasparenti in cui sia chiaramente mostrata la relazione tra le informazioni fornite all’analista e la raccomandazione finale. Un modello che possiede tali caratteristiche in termini di chiarezza e trasparenza viene definito glass-box, appartiene a questa categoria il modello dei Rough Sets, che sarà affrontato nel 4 capitolo.

CAPITOLO 2

L'APPROCCIO MULTICRITERIALE ALLE DECISIONI

1. I PROBLEMI DECISIONALI

1.1 GENERALITÀ

Un problema decisionale è un processo in cui uno o più decisori si trovano a dover effettuare delle scelte fra diverse alternative nel rispetto di determinati obiettivi e vincoli. La formulazione tradizionale di un problema di decisione è basata sui seguenti tre elementi:

- 1) Un insieme ben definito di alternative ammissibili: per esempio un insieme di possibili progetti d'investimento.
- 2) Un'unica "funzione obiettivo" a valori reali (detta anche "criterio"), che riflette le preferenze del decisore ("decision maker"): per esempio, il profitto, misurato in termini di valore attuale della differenza tra costi e ricavi dei progetti considerati, oppure il costo unitario, ecc..
- 3) Un problema matematico ben formulato descritto nei termini di una funzione obiettivo da massimizzare nel rispetto degli eventuali vincoli: la "soluzione"

del problema è pertanto l'alternativa che massimizza [o minimizza] la funzione obiettivo: nel nostro esempio il progetto che fornisce il massimo profitto.

Questa metodologia tradizionale, l'unica adoperata sino alla fine degli anni '60 per affrontare problemi di decisione, è definita *approccio monocriteriale*; essa riduce drasticamente la complessità della realtà modellizzandola su un'unica dimensione, una sola scala numerica esaustiva, spesso monetaria, riconducendo un complesso problema decisionale solamente ad un puro calcolo. L'analisi costi-benefici si inserisce in questo contesto, con le ulteriori complicazioni e forzature dovute alla presenza di effetti difficilmente quantificabili o valutabili in termini monetari, alla implicita ed assoluta compensazione tra effetti positivi e negativi, ecc.

L'approccio monocriteriale costituisce una forte astrazione dal comportamento reale. Esso, infatti, non permette di modellizzare la pluralità di obiettivi generalmente perseguiti dal decisore nei problemi della vita reale: per esempio, nella scelta di un progetto, un ente pubblico non considera solamente i possibili profitti del progetto, ma anche il suo impatto ambientale, le conseguenze economiche e sociali sul territorio, l'equilibrio finanziario, ecc.

Al fine di prendere in esplicita considerazione tutti questi aspetti, spesso conflittuali, è stato proposto un differente approccio ai problemi di decisione, basato su una appropriata riformulazione dei punti 2) e 3):

2') Un insieme di "obiettivi" rappresentati da funzioni a valori reali (criteri), aggregati per mezzo di una funzione di utilità che assegna una valutazione complessiva a ogni possibile alternativa, rendendone possibile il confronto sulla base del principio che maggiore è la valutazione complessiva, migliore è l'alternativa considerata.

3') Un problema matematicamente ben formulato, consistente nel trovare la (o le) alternative che massimizzano la funzione di utilità o funzione valore (soluzione di "compromesso").

Questa metodologia, detta delle decisioni multicriteriali (Multiple Criteria Decision Making - MCDM), pur rientrando ancora in un approccio normativo, rappresenta già un modo più realistico di trattare problemi di decisione, rendendo esplicite le preferenze pre-esistenti implicitamente nella mente del decisore. Essa rientra nella cosiddetta "ottimizzazione vettoriale" o programmazione matematica multi-obiettivo e viene usualmente chiamata Multiple Attribute Utility Theory (MAUT).

Tuttavia anche l'MCDM presenta alcune limitazioni (Roy 1990):

- l'insieme delle azioni ammissibili è spesso proposto in maniera non precisa;
- le preferenze del decisore non sono sempre ben stabilite, come si ipotizza utilizzando la funzione di utilità, che permette sempre di confrontare due alternative,
- i dati coinvolti nei problemi di decisione sono spesso incerti, a causa della casualità, della vaghezza e della granularità delle informazioni disponibili;

- la validazione della soluzione può non essere basata solamente su un modello matematico, senza considerare anche gli aspetti organizzativi e culturali del processo di decisione.

Sulla base di queste considerazioni è stata proposta una nuova formulazione del problema di decisione, che prende in considerazione l'intero processo decisionale. Essa si caratterizza per i seguenti punti (Roy, 1990):

1'') un insieme A non necessariamente stabile di azioni potenziali: le azioni considerate non sono necessariamente tutte ammissibili (realizzabili), perché anche alcune azioni "ideali" possono essere prese in considerazione durante il processo decisionale, per esempio come punti di riferimento per degli utili confronti. Inoltre, l'insieme delle azioni può evolvere durante il processo decisionale.

2'') un insieme G di criteri che rappresentino i differenti punti di vista dai quali studiare il problema di decisione: questi criteri dovrebbero prendere in considerazione anche le diverse fonti di incertezza e la loro modellizzazione dovrebbe inoltre consentire alcune forme di esitazione espresse dal decisore.

3'') un problema matematicamente non ben definito: in questo caso non esiste alcuna funzione da ottimizzare, bensì il supporto alla decisione mira a costruire un modello che permetta di confrontare le azioni potenziali sulla base dell'insieme di criteri G considerato al fine di affrontare coerentemente il problema decisionale affrontato.

Questa metodologia di supporto alla decisione è definita aiuto multicriteriale alla decisione (Multiple Criteria Decision Aid - MCDA). Durante una prima fase del processo, l'analista aiuta il decisore a costruire i propri convincimenti e ad ottenere una appropriata "raccomandazione" (recommendation) per il problema di decisione affrontato, lasciando al decisore medesimo la decisione finale.

1.2 CLASSIFICAZIONE DEI PROBLEMI DECISIONALI

I problemi decisionali affrontati nella realtà operativa sono diversi e di molteplice natura, sia con riferimento alla particolare problematica affrontata che al contesto che li caratterizza.

Le principali problematiche decisionali sono (Roy, 1985):

1) **Scelta** (choice): selezionare il più piccolo sottoinsieme di **A** (possibilmente una sola azione) che contenga le azioni considerate "migliori" o soddisfacenti con riferimento all'insieme di criteri **G**. Quindi lo scopo della decisione è quello di scegliere il migliore oggetto. Un esempio tipico è quello del processo decisionale che porta all'acquisto di un automobile, dove le automobili sono gli oggetti della decisione mentre caratteristiche come il prezzo, il colore, la velocità sono gli attributi.

2) **Classificazione** (classification): assegnare ogni azione ammissibile (se A è finito) ad una delle categorie predefinite (segmentazione), eventualmente preferenzialmente ordinate (sorting). In questo caso, lo scopo della decisione è quello di assegnare gli oggetti a classi predefinite. Problemi di questo tipo si riscontrano quando occorre assegnare un'impresa ad una classe predefinita di rischio (credit scoring), dove le imprese sono gli oggetti della decisione, mentre gli indicatori economici e finanziari sono gli attributi. Un altro esempio di decisione di questo tipo si ha quando si devono diagnosticare delle patologie ad un insieme di pazienti, dove i pazienti sono gli oggetti della decisione mentre i sintomi e i risultati dei test medici sono gli attributi.

3) **Ordinamento** (ranking): ordinare le azioni di A (se finito) dalla migliore alla peggiore in classi di equivalenza. In altre parole, lo scopo della decisione è quello di ordinare gli oggetti dal migliore al peggiore. L'esempio classico è quello delle graduatorie dei concorsi dove i candidati sono gli oggetti della decisione mentre i voti conseguiti nelle varie prove sono gli attributi.

Relativamente ai problemi di classificazione, questi possono essere a loro volta ripartiti in due sottocategorie: Tassonomici (Taxonomy), quando gli insiemi dei valori assunti dagli attributi e le classi predefinite a cui associare gli oggetti non sono ordinati da relazioni di preferenza: questo è il caso delle diagnosi mediche sopra esposte; problemi Classificazione Ordinale (multiple criteria sorting), quando gli insiemi dei valori assunti dagli attributi e le classi predefinite a cui

associare gli oggetti sono ordinati da relazione di preferenza, questo è il caso del Credit Scoring. Inoltre, se gli insiemi dei valori assunti dagli attributi sono ordinati da relazione di preferenza essi prenderanno il nome di criteri, altrimenti saranno chiamati semplicemente attributi. Per esempio, nelle decisioni che riguardano la selezione di un'automobile il prezzo dell'auto è un criterio perché, ovviamente, un prezzo basso è migliore di uno più alto, mentre il colore della auto non è un criterio perché in generale il colore rosso non è intrinsecamente migliore del colore verde. Tuttavia, anche il colore potrebbe diventare un criterio se, per esempio, il decisore considerasse il colore rosso migliore del colore verde. I modelli che analizzano i problemi decisionali fanno uso di processi che, in modo più o meno trasparente, legano le decisioni (output del modello) alle caratteristiche degli oggetti espresse dalle informazioni ottenute dagli attributi presi in considerazione (input del modello). Infatti, le informazioni ottenute dagli attributi, in merito agli oggetti da esaminare, molto spesso vengono elaborate secondo metodologie che non consentono al decisore di comprendere in modo chiaro le relazioni tra le informazioni che esso ha fornito (tramite gli attributi) e le raccomandazioni o i comportamenti consigliati dal modello decisionale. Per cui, quando il modello decisionale è poco trasparente, esso è percepito dal decisore come una black-box il cui risultato deve essere accettato perché è l'autorità dell'analista a garantire la sua validità. Pertanto, al fine di soddisfare al meglio le esigenze del decisore, occorrono metodologie più trasparenti in cui sia chiaramente mostrata la relazione tra le informazioni fornite all'analista e la

raccomandazione finale. Un modello che possiede tali caratteristiche in termini di chiarezza e trasparenza viene definito glass-box.

Si possono poi classificare e distinguere diversi tipi di problemi decisionali.

◆ Rispetto alle alternative:

- 1) discreti (numero finito di alternative),
- 2) continui (A insieme infinito).

◆ Rispetto alla natura delle informazioni:

- 1) soft (in presenza di informazioni solamente qualitative),
- 2) hard (informazioni solamente quantitative),
- 3) misti (disponibilità di informazioni qualitative e quantitative).

◆ Rispetto allo scenario:

- 1) in condizioni di certezza (informazioni deterministiche, perfettamente conosciute a priori),
- 2) in condizioni di rischio (conoscenza delle distribuzioni di probabilità delle informazioni),
- 3) in condizioni di incertezza (assenza di distribuzioni di probabilità),
- 4) in condizioni di incertezza competitiva (risultati che dipendono anche dalle decisioni adottate da altri soggetti, normalmente “avversari”, “game theory”).

◆ Rispetto alla distribuzione temporale degli effetti:

- 1) ad effetti immediati (conseguenze che si verificano e si esauriscono immediatamente),

2) ad effetti differiti (conseguenze lontane nel tempo o che si ripetono nel tempo).

◆ Rispetto alla misurazione delle conseguenze:

- 1) con conseguenze definite e nette (precise),
- 2) con conseguenze sfuocate (imprecisioni linguistiche, informazioni “fuzzy”),
- 3) quantitative (misurabili numericamente),
- 4) qualitative (descrivibili verbalmente).

◆ Rispetto al numero dei decisori:

- 1) singolo decisore (single person: unica persona fisica o istituzione),
- 2) pluralità di decisori (multi person: molteplicità di persone o di enti, spesso con interessi contrapposti),
 - concorrenti (più decisori con lo stesso grado di potere decisionale),
 - gerarchici (più decisori con poteri decisionali subordinati).

◆ Rispetto al numero dei punti di vista:

- 1) monocriteriali (in presenza di una sola funzione-obiettivo),
- 2) multicriteriali (preferenze espresse con riferimento esplicito a molteplici punti di vista).

◆ Rispetto alle fasi solutive:

- 1) single-step (risoluzione del problema in una sola fase)
- 2) multistep (necessità di affrontare il problema in fasi distinte e successive)

1.3 AIUTO MULTICRITERIALE ALLA DECISIONE

Secondo Roy (1993) si può definire l'MCDA come "l'attività di chi, in modi che noi definiamo scientifici, aiuta ad ottenere elementi di risposte a domande poste dagli attori coinvolti in processi di decisione, elementi che aiutano a chiarire questa decisione al fine di metter gli attori nelle condizioni più favorevoli per quel tipo di comportamento che aumenti la coerenza tra l'evoluzione del processo decisionale, da una parte, e gli obiettivi e/o il sistema di valori in cui questi attori si trovano reciprocamente ad operare." Pertanto, il fine dei problemi decisionali è quello di dare al decisore (Decision Maker) una raccomandazione, o di favorire un comportamento, riguardo ad un insieme di oggetti (chiamati anche alternative, soluzioni, atti, azioni, opzioni, candidati, etc...) valutati da diversi punti di vista considerati rilevanti per il problema stesso, chiamati attributi (o anche caratteristiche, variabili, criteri, etc..).

In un contesto MCDA una raccomandazione si ottiene come risultato finale di una procedura di quattro fasi (Roy, 1985):

- 1) la definizione delle *azioni* che devono essere prese in considerazione e la definizione e *formulazione del problema* di decisione: scelta, classificazione, ordinamento;

- 2) l'individuazione dei *punti di vista* da prendere in considerazione e la *modellizzazione delle preferenze* del decisore rispetto ad ognuno di questi punti di vista;
- 3) la sintesi delle informazioni disponibili in un *modello* complessivo che permette di aggregare le preferenze;
- 4) l'applicazione di una certa procedura al fine di ottenere una raccomandazione per il problema di decisione considerato.

Gli elementi di base dell'aiuto multicriteriale alla decisione sono quindi due: un insieme di azioni $A=\{a, b, \dots\}$ e una famiglia coerente di criteri $G=\{g_1, g_2, \dots, g_m\}$. Nel seguito si indicherà con F l'insieme degli indici dei criteri di G , cioè $F=\{1, 2, \dots, m\}$.

L'insieme di azioni A contiene l'insieme degli elementi (oggetti, progetti, candidati, ...) che devono essere analizzati durante il processo decisionale.

I differenti punti di vista considerati sono modellati per mezzo di attributi o criteri. Ogni **attributo** rappresenta uno o più punti di vista senza considerarne esplicitamente proprietà ordinali. Un **criterio**, invece, è una funzione $g_j: A \rightarrow \mathbf{R}$ tale che, $\forall a \in A$, $g_j(a)$ è la valutazione dell'azione a con riferimento al criterio g_j e, $\forall a, b \in A$, $g_j(a) \geq g_j(b)$ significa che “ a è almeno tanto buona quanto b con riferimento ai punti di vista rappresentati dal criterio g_j ”.

Nei problemi discreti le valutazioni delle azioni per mezzo dei criteri di G sono usualmente raccolte in una matrice, detta appunto **matrice delle valutazioni** multicriteriali o **impact matrix**.

L'insieme G dei criteri dovrebbe soddisfare alcune proprietà (Bouyssou 1990):

- 1) *leggibilità*, cioè l'insieme dei criteri G dovrebbe essere costituito da un numero di criteri sufficientemente piccolo in modo che essi possano costituire una base di discussione tra gli attori per permettere all'analista di ottenere le informazioni inter-criteriali necessarie per l'implementazione di una procedura di aggregazione,
- 2) *operatività*, cioè l'insieme di criteri G dovrebbe essere considerato come una base solida per continuare il processo di aiuto alla decisione.

Inoltre, l'insieme dei criteri G dovrebbe essere *coerente* (Roy e Bouyssou 1993), cioè dovrebbe rappresentare tutti i differenti aspetti del problema evitando ridondanze. Più precisamente, un insieme di criteri G è coerente se è:

- *esaustivo*, cioè contiene ogni punto di vista importante, cosicché, se $g_j(a)=g_j(b)$ per tutti i criteri di G , si deve concludere che a e b sono indifferenti;
- *monotono*, cioè le preferenze parziali che sono rappresentate per mezzo dei singoli criteri devono essere coerenti con la preferenza complessiva, cosicché se l'azione a è giudicata globalmente migliore dell'azione b , allora ogni azione c , che è almeno tanto buona quanto l'azione a su tutti i criteri di G , deve essere anch'essa giudicata migliore dell'azione b , $a,b,c \in A$;
- *minimale*, cioè non dovrebbe contenere nessun criterio ridondante, per cui la soppressione da G di qualsiasi criterio conduce ad un insieme di criteri che non soddisfa le due condizioni precedenti.

L'approccio multicriteriale, pertanto, si propone di aiutare il decisore nell'analisi del problema decisionale affrontato rispetto alle azioni ammissibili ed all'insieme dei criteri presi in considerazione. Esso:

1. migliora la trasparenza e la coerenza del processo decisionale,
2. definisce, precisa e mette in evidenza il peculiare ruolo del decisore,
3. usa tutte le informazioni che il decisore può, sa e vuole fornire per costruire un modello quanto più fedele possibile alle sue preferenze.

Il paradigma multicriteriale proprio dell'MCDA si caratterizza fondamentalmente per:

- pluralità di criteri o punti di vista esplicitamente presi in considerazione per condurre il sistema o guidarne la sua evoluzione,
- per la conflittualità, almeno locale, di questi criteri, per cui occorre ricercare un certo "compromesso" rispettando dei principi di coerenza,
- per l'obiettivo di questi compromessi che si prefiggono di conferire ai criteri dei valori compatibili con una certa forma di equilibrio, che in un contesto dinamico avrà necessariamente carattere transitorio.

2. LE STRUTTURE DI PREFERENZA

2.1. GENERALITÀ

Un approccio più realistico all'analisi multicriteriale delle decisioni deve prendere in considerazione la modellizzazione delle preferenze. Infatti, nell'approccio classico si dà per scontata la possibilità di rappresentare le preferenze per mezzo di una funzione di utilità $u:\mathbf{A}\rightarrow\mathbf{R}$ che assegna ad ogni azione $a\in\mathbf{A}$ un valore crescente con la preferibilità dell'azione considerata rispetto al punto di vista particolare - singolo criterio - o globale - relativamente alla sintesi di tutti i criteri del problema di decisione affrontato. In quest'ottica tanto maggiore è il valore $u(a)$ assegnato all'azione $a\in\mathbf{A}$, tanto più preferibile è l'azione a stessa rispetto alle altre azioni di \mathbf{A} , per cui $\forall a,b\in\mathbf{A}$ si ha che se $u(a)>u(b)$, allora a è preferita a b e se $u(a)=u(b)$, allora a e b sono indifferenti tra di loro. Questo approccio, anche se apparentemente molto neutro e naturale, ha delle conseguenze molto rilevanti dal punto di vista del tipo di preferenze rappresentate. Più in particolare, questo approccio implicitamente assume che, date due azioni, si riesca sempre a confrontarle tra di loro (infatti per ogni $a,b\in\mathbf{A}$ o $u(a)>u(b)$, e allora a è migliore di b , o $u(a)<u(b)$, e allora a è peggiore di b , o $u(a)=u(b)$, e allora a e b sono indifferenti). Inoltre, la presenza di una funzione di utilità implica la transitività dell'indifferenza (se a e b sono indifferenti e b e c sono pure indifferenti, allora anche a e c sono indifferenti) e la transitività della

preferenza (se a è preferito a b e b è preferito a c , allora anche a è preferito a c). Nei problemi reali però queste “conseguenze naturali” dell’esistenza di una funzione di utilità non sono sempre verificate. Infatti, non è sempre detto che si riescano a confrontare due azioni, e anzi molto spesso si sperimentano situazioni di esitazione nel valutare la preferenza tra due azioni. Inoltre, è abbastanza naturale riscontrare anche situazioni in cui non vale la transitività della preferenza o della indifferenza. Classico è l’esempio ispirato a un famoso paradosso dovuto a Condorcet e che in un certo modo sta alla base del teorema di impossibilità di Arrow. Si considerino tre alternative, per esempio tre impianti di depurazione dell’aria, e tre criteri, per esempio presenza nell’aria di tre agenti inquinanti. Indichiamo con a , b , e c le tre alternative e con g_1 , g_2 , e g_3 i tre criteri. L’ordine di preferenza delle tre alternative rispetto ai tre criteri sia quello rappresentato nella seguente Tabella 1. Pertanto, ad esempio rispetto al criterio g_1 l’alternativa a è la prima in ordine di preferenza, l’alternativa b è la seconda e c la terza.

Tabella 1. Ordinamento di preferenza delle tre alternative con riferimento ai tre criteri

Alternative\Criteri	g_1	g_2	g_3
A	1°	2a	3a
B	2°	3a	1a
C	3°	1a	2a

Si supponga ora che il decisore voglia ordinare globalmente le tre alternative seguendo questo principio: per tutte le coppie di alternative x e y , x è globalmente preferita a y se per la maggioranza dei criteri x è preferita a y . Con riferimento alle tre alternative a , b , e c si ha pertanto che a è preferita a b (infatti a è migliore di b con riferimento al criterio g_1 e al criterio g_2) e che b è preferita a c (infatti b è migliore di c con riferimento al criterio g_1 e al criterio g_2). Ci si aspetterebbe, quindi, dalla preferenza di a su b e di b su c anche la preferenza di a su c . Tuttavia si osserva che c è preferita ad a per i criteri g_2 e g_3 e, pertanto, abbastanza sorprendentemente, è c ad essere preferita ad a .

Queste ed altre osservazioni hanno spinto gli studiosi di MCDA ad abbandonare l'assunzione aprioristica dell'esistenza di una funzione di utilità (marginale o complessiva). Si è, invece, considerato come dato originario una relazione binaria di preferenza su \mathbf{A} , che non necessariamente soddisfi le proprietà di completezza e transitività che caratterizzavano l'esistenza di una funzione di utilità. In quest'ottica la funzione di utilità è solo una delle possibili rappresentazioni delle relazioni binarie di preferenza. Inoltre, essa esiste solo se alcune ben precise proprietà (o, se si vuole, requisiti tecnici) sono soddisfatte. Pertanto, l'attenzione si è spostata sulle proprietà delle relazioni binarie di preferenza e sulle conseguenti rappresentazioni numeriche. In questo contesto si riescono a rappresentare situazioni molto più variegata e realistiche (per esempio esitazioni, effetti soglia, preferenze sfumate, etc.) di quelle rappresentate dalla "classica" funzione di utilità.

2.2 ALCUNE OSSERVAZIONI GENERALI SULLE RELAZIONI

BINARIE

La modellizzazione delle preferenze è un passo fondamentale in economia, sociologia, psicologia, scelte sociali, etc. Essa è di fondamentale importanza per l'aiuto alla decisione¹⁴.

Al fine di introdurre le principali nozioni sulle strutture di preferenza, si introducono alcuni concetti generali sulle relazioni binarie. Sia X un dato insieme. Una relazione binaria R su X è un sottoinsieme del prodotto cartesiano $X \times X$. Se $(a, b) \in R$, allora si scrive anche aRb . Data una relazione binaria R , il complemento R^c , l'inverso R^{-1} e il duale R^d sono rispettivamente definiti come segue:

$$(a, b) \in R^c \Leftrightarrow (a, b) \notin R,$$

$$(a, b) \in R^{-1} \Leftrightarrow (b, a) \in R,$$

$$(a, b) \in R^d \Leftrightarrow (b, a) \notin R.$$

Una relazione binaria R definita su un insieme finito X può essere rappresentata da un grafo orientato (X, R) , dove X è l'insieme di nodi (vertici) e R è l'insieme

¹⁴ Figueira, J., Greco, S., Erghott, M. (2005), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, Berlin

di archi diretti. Esiste un arco dal nodo a al nodo b se e solo se aRb . aRa non si rappresenta con due distinti archi ma con uno solo che si chiama cappio.

Si ricordano le proprietà fondamentali delle relazioni binarie. Una relazione binaria è:

- riflessiva, se e solo se $aRa, \forall a \in X$,
- irriflessiva, se e solo se $aR^c a, \forall a \in X$,
- simmetrica, se e solo se $aRb \Rightarrow bRa, \forall a, b \in X$,
- antisimmetrica, se e solo se $[aRb \text{ e } bRa] \Rightarrow a=b, \forall a, b \in X$,
- asimmetrica, se e solo se $aRb \Rightarrow bR^c a, \forall a, b \in X$,
- completa, se e solo se aRb e/o bRa per $\forall a, b \in X$, con $a \neq b$,
- fortemente completa, se e solo se aRb e/o bRa per $\forall a, b \in X$,
- transitiva, se e solo se $[aRb \text{ e } bRc] \Rightarrow aRc, \forall a, b, c \in X$,
- negativamente transitiva, se e solo se $[aR^c b \text{ e } bR^c c] \Rightarrow aR^c c, \forall a, b, c \in X$,
- una relazione di Ferrer, se e solo se $[aRb \text{ e } cRd] \Rightarrow aRd$ e/o $cRb, \forall a, b, c, d \in X$,
- semitransitiva, se e solo se $[aRb \text{ e } bRc] \Rightarrow aRd$ e/o $dRc, \forall a, b, c, d \in X$.

2.2 SITUAZIONI ELEMENTARI DI PREFERENZA

Sia A un insieme di azioni e $a, b \in A$. Solitamente si suppone che confrontando due azioni un individuo possa reagire in uno dei seguenti tre modi:

- preferenza per una delle due azioni, per esempio a è preferita a b , indicata con aPb ,
- indifferenza tra le due azioni, indicata con aIb ,
- incomparabilità tra le due azioni, a causa di un rifiuto, di una incapacità o impossibilità di confrontare, indicata con aJb .

Talvolta (Vincke 1980, 1988, Roy e Vincke, 1984, 1987) si considera un'altra possibile situazione fondamentale:

- preferenza debole per una delle due azioni, per esempio l'azione a è debolmente preferita a b , indicata con aQb .

La preferenza debole caratterizza una situazione in cui si ha una esitazione tra la preferenza stretta e l'indifferenza.

Le relazioni binarie corrispondenti alle quattro situazioni fondamentali P, I, J, Q debbono soddisfare i seguenti requisiti:

- $aPb \Rightarrow$ non bPa , cioè P è asimmetrica,
- aIa , cioè I è riflessiva,
- $aIb \Rightarrow bIa$, cioè I è simmetrica,
- non aJa , cioè J è irriflessiva,
- $aJb \Rightarrow bJa$, cioè J è simmetrica,
- $aQb \Rightarrow$ non bQa , cioè Q è asimmetrica.

Le quattro relazioni binarie P, I, Q, J definite su un insieme di azioni potenziali A formano un sistema di relazioni di preferenza di base se esse costituiscono una partizione di $A \times A$, cioè:

1) esse sono esaustive, ossia per ogni coppia ordinata di azioni, vale almeno una delle quattro relazioni; formalmente si ha:

$$P \cup I \cup J \cup Q = A \times A,$$

2) esse sono mutualmente esclusive, cioè per ogni coppia ordinata (a,b) di azioni di A , vale al più una delle due relazioni; formalmente si ha che, per ogni $H, K \in \{P, I, J, Q\}$, $H \cap K = \emptyset$.

All'interno di ogni struttura di preferenza basata sulle tre situazioni fondamentali P , I , e J , queste possono essere completamente caratterizzate dalla relazione binaria S definita da

$$aSb \Leftrightarrow aPb \text{ e/o } alb \quad \forall a, b \in A, \text{ (ossia, } S = P \cup I).$$

Infatti, $\forall a, b \in A$ si ha:

$$aPb \Leftrightarrow aSb \text{ e non } bSa$$

$$alb \Leftrightarrow aSb \text{ e } bSa$$

$$aJb \Leftrightarrow \text{non } aSb \text{ e non } bSa.$$

All'interno delle strutture di preferenza basate sulle quattro situazioni fondamentali P , I , J e Q , si considera la seguente definizione:

$$aSb \Leftrightarrow aPb \text{ e/o } alb \text{ e/o } aQb \quad \forall a, b \in A, \text{ (ossia, } S = P \cup I \cup Q).$$

Tuttavia in questo caso una struttura di preferenza non può essere completamente caratterizzata utilizzando la sola relazione binaria S (Tsoukias e Vincke, 1998).

S viene definita relazione di surclassamento. Si osservi che $\forall a, b \in A$ “ aSb ” significa “ a è almeno tanto buona quanto b ” e che S è riflessiva, cioè aSa .

2.4 STRUTTURE DI PREFERENZA

Un preordine completo è una struttura di preferenza che soddisfa le seguenti condizioni $\forall a, b$ e $c \in A$:

- non aIb , cioè non ci sono situazioni di incomparabilità,
- $[aPb \text{ e } bPc] \Rightarrow aPc$, cioè P è transitiva,
- $[aIb \text{ e } bIc] \Rightarrow aIc$, cioè I è transitiva.

La relazione caratteristica S associata ad un preordine completo verifica le seguenti condizioni $\forall a, b$ e $c \in A$:

- aSb e/o bSa , cioè S è completa,
- $[aSb \text{ e } bSc] \Rightarrow aSc$, cioè S è transitiva.

Se A è un insieme finito o numerabile, allora esiste una funzione $g:A \rightarrow \mathbb{R}$ tale che

$$aPb \Leftrightarrow g(a) > g(b),$$

$$aIb \Leftrightarrow g(a) = g(b),$$

$$aSb \Leftrightarrow g(a) \geq g(b).$$

In altri termini, un preordine completo è la struttura di preferenza che corrisponde alla modellizzazione delle preferenze della teoria classica dell'utilità ordinale. La funzione g corrispondente viene chiamata anche vero-criterio.

In molte situazioni reali la transitività dell'indifferenza è una condizione troppo esigente, come evidenziato dal famoso paradosso di Luce (1956). Si considerino un certo numero di tazze di tè. La prima tazza di tè è senza zucchero, la seconda ha un solo milligrammo di zucchero, la terza ha due milligrammi di zucchero e così via. Naturalmente non si può esprimere qualsiasi preferenza tra due tazze consecutive di tè, tuttavia si può generalmente esprimere una preferenza tra una tazza di tè senza zucchero e un'altra con molto zucchero. Il semiordine è una struttura di preferenza che permette di rappresentare questo tipo di fenomeni, dovuti ad effetti di soglia, indebolendo la transitività sull'indifferenza.

Un semiordine è una struttura di preferenza che soddisfa le seguenti condizioni

$\forall a, b, c \text{ e } d \in A$:

- non aJb , cioè non ci sono situazioni di incomparabilità,
- $[aPb, bIc \text{ e } cPd] \Rightarrow aPd$,
- $[aPb, bPc \text{ e } aId] \Rightarrow dPc$.

La relazione caratteristica S associata ad un semiordine verifica le seguenti condizioni $\forall a, b, c \text{ e } d \in A$:

- $aSb \text{ e/o } bSa$, cioè S è completa,
- $[aSb \text{ e } cSd] \Rightarrow [aSd \text{ e/o } cSb]$, cioè S è una relazione di Ferrer,
- $[aSb \text{ e } bSc] \Rightarrow [aSd \text{ e/o } dSc]$, cioè S è semitransitiva.

Se A è un insieme finito o numerabile, allora esistono una funzione $g:A\rightarrow\mathbb{R}$ ed una soglia $q\in\mathbb{R}^+$, detta soglia di indifferenza, tale che:

$$aPb \Leftrightarrow g(a) > g(b) + q,$$

$$aIb \Leftrightarrow |g(a) - g(b)| \leq q.$$

Tale funzione g viene chiamata anche quasi-criterio. Nella rappresentazione del semiordine, la soglia q rappresenta una “piccola” differenza, non percepita dal decisore, che trasforma l’indifferenza da “puntuale” (come nell’approccio classico) a “segmentaria”. Tuttavia la soglia q è costante, mentre molto spesso la reazione a differenti valutazioni dipende anche dai valori assoluti delle quantità confrontate: per esempio una differenza di \$1000 non ha lo stesso significato quando si trattano migliaia di dollari o milioni di dollari. La seguente struttura di ordine di intervalli permette di introdurre una soglia di indifferenza variabile.

Un ordine di intervalli è una struttura di preferenza che soddisfa le seguenti condizioni $\forall a, b, c$ e $d \in A$:

- non aIb , cioè non ci sono situazioni di incomparabilità,
- $[aPb, bIc \text{ e } cPd] \Rightarrow aPd$.

La relazione caratteristica associata ad un ordine di intervalli soddisfa le seguenti condizioni $\forall a, b, c$ e $d \in A$:

- aSb e/o bSa , cioè S è completa,

- $[aSb \text{ e } cSd] \Rightarrow [aSd \text{ e/o } cSb]$, cioè S è una relazione di Ferrer.

Se l'insieme A è finito o numerabile, allora esistono una funzione $g: A \rightarrow \mathbb{R}$ ed una funzione $q: A \rightarrow \mathbb{R}^+$, tali che:

$$aPb \Leftrightarrow g(a) > g(b) + q(b),$$

$$aIb \Leftrightarrow g(a) \leq g(b) + q(b) \text{ e } g(b) \leq g(a) + q(a).$$

Recentemente gli ordini di intervalli sono stati estesi per considerare una soglia che dipende dalle valutazioni di entrambe le azioni confrontate, piuttosto che da una sola. Questo conduce a una struttura di preferenza (Matarazzo 1984, 1986, Abbas e Vincke 1993, Abbas, Pirlot e Vincke 1996, Fodor e Roubens 1996) in cui esistono una funzione $g: A \rightarrow \mathbb{R}$ e una funzione $Q: A \times A \rightarrow \mathbb{R}^+$, tali che, per ogni $a, b \in A$, si ha

$$aPb \Leftrightarrow g(a) > g(b) + Q(a, b),$$

$$aIb \Leftrightarrow |g(a) - g(b)| \leq Q(a, b).$$

Questa struttura di preferenza è definita ordine di co-comparabilità se, per ogni $a, b, c \in A$, è soddisfatta la seguente disuguaglianza triangolare

$$Q(a, b) \leq Q(a, c) + Q(c, b).$$

Si ricordi che le precedenti strutture di preferenza non prendono in considerazione la relazione di incomparabilità. Tuttavia in molte situazioni reali il decisore sperimenta la indisponibilità o l'impossibilità di confrontare alcune coppie di azioni, perché per esempio devono essere aggregate valutazioni fortemente conflittuali su differenti punti di vista (così, per esempio, è praticamente impossibile dire se è preferita una vettura familiare, molto economica, ma piuttosto lenta o, invece, una vettura sportiva, molto veloce ma anche piuttosto costosa). Una tipica struttura di preferenza che considera anche l'incomparabilità è il preordine parziale. Esso soddisfa le seguenti proprietà, $\forall a, b \text{ e } c \in A$:

- $[aPb \text{ e } bPc] \Rightarrow aPc$, cioè P è transitiva,
- $[aIb \text{ e } bIc] \Rightarrow aIc$, cioè I è transitiva,
- $[aPb \text{ e } bIc] \Rightarrow aPc$,
- $[aIb \text{ e } bPc] \Rightarrow aPc$,
- $P \cup I$ non è completa.

La relazione caratteristica S associata ad un preordine parziale verifica le seguenti condizioni, $\forall a, b \text{ e } c \in A$:

- aSa , cioè S è riflessiva,
- $[aSb \text{ e } bSc] \Rightarrow aSc$, cioè S è transitiva.

Se l'insieme A è finito o numerabile, allora esiste una funzione $g: A \rightarrow \mathbb{R}$ tale che

$$aPb \Rightarrow g(a) > g(b),$$

$$aIb \Rightarrow g(a) = g(b).$$

Infine, si ricorda lo pseudo-ordine, che è una tipica struttura di preferenza nella quale interviene anche la preferenza debole Q (Roy e Vincke 1984, 1987). Nella rappresentazione di questa struttura di preferenza ci sono due soglie: una soglia di indifferenza, q , all'interno della quale il decisore esprime una chiara indifferenza, e una soglia di preferenza, p , superata la quale il decisore è sicuro di una preferenza (stretta):

$$aPb \Leftrightarrow g(a) > g(b) + p(g(b)),$$

$$aQb \Leftrightarrow g(b) + p(g(b)) \geq g(a) > g(b) + q(g(b)),$$

$$aIb \Leftrightarrow \begin{cases} g(b) + q(g(b)) \geq g(a) \\ g(a) + q(g(a)) \geq g(b). \end{cases}$$

La funzione g corrispondente viene chiamata pseudo-criterio. Al fine di evitare alcune incoerenze, le funzioni di soglia devono soddisfare le seguenti condizioni:

$$g(a) > g(b) \Leftrightarrow g(a) + q(g(a)) > g(b) + q(g(b)),$$

$$g(a) > g(b) \Leftrightarrow g(a) + p(g(a)) > g(b) + p(g(b)).$$

2.5 RELAZIONI DI PREFERENZA MULTIPLE

Una struttura di relazioni di preferenza multipla (Roberts 1971, , Roubens e Vincke 1985, Doignon 1987) si ottiene utilizzando un insieme di relazioni di preferenza nidificate. Esse corrispondono a differenti “intensità” di relazioni di preferenza: preferenza molto debole, preferenza debole, preferenza forte, preferenza molto forte, preferenza quasi totale, preferenza totale, etc.

Una collezione di strutture di relazioni di preferenza nidificate è associata ad ogni struttura di preferenza multipla. Esse sono ottenute considerando, per ogni livello d'intensità, la corrispondente struttura di indifferenza tra due azioni se nessuna di esse è preferita all'altra con tale intensità.

Una rappresentazione valore-soglie di una struttura di relazioni di preferenza multiple consiste in una funzione valore g ed un vettore T di m funzioni soglia (t_1, t_2, \dots, t_m) . Se si tratta di una soglia superiore si ha $aP^k b$ se e solo se $g(a) > t^k(b)$. Se si tratta di una soglia inferiore si ha $g(b) < t^k(a)$.

2.6 RELAZIONE DI SURCLASSAMENTO A QUATTRO

VALORI

L'idea di base del modello di preferenza a quattro valori (Tsoukias e Vincke 1995) è legato alla ricerca di “ragioni positive” (ossia argomenti a favore) e

“ragioni negative” (ossia argomenti contrari) a supporto dell’ipotesi di verità della relazione di surclassamento per una coppia ordinata (x,y) di azioni. Le combinazioni di ragioni positive e negative creano allora quattro possibili situazioni di surclassamento:

- 1) surclassamento vero, che si indica con $xS^T y$, nel caso che esistano sufficienti ragioni positive per stabilire xSy e non si abbiano sufficienti ragioni negative per stabilire $xS^c y$;
- 2) surclassamento contraddittorio, indicato con $xS^K y$, se esistono sufficienti ragioni positive per stabilire xSy e sufficienti ragioni negative per stabilire $xS^c y$;
- 3) surclassamento incognito, indicato con $xS^U y$, se non esistono sufficienti ragioni positive per stabilire xSy e non esistono sufficienti ragioni negative per stabilire $xS^c y$;
- 4) surclassamento falso, indicato con $xS^F y$, se non esistono sufficienti ragioni positive per stabilire xSy ed esistono sufficienti ragioni negative per stabilire $xS^c y$.

La Tabella 2 riassume le quattro relazioni di surclassamento.

Tabella 2. Relazioni di surclassamento a quattro valori

	S^T	S^K	S^U	S^F
xSy	1	1	0	0
$xS^c y$	0	1	0	1

Combinando poi i quattro tipi di relazioni binarie di surclassamento prima ricordati con riferimento a ciascuna delle coppie ordinate (x,y) e (y,x) di azioni, la modellizzazione delle preferenze si arricchisce notevolmente, ottenendosi le seguenti dieci situazioni di preferenza per confrontare x e y :

- 1) preferenza stretta, che si indica con xPy , se x è strettamente migliore di y , cioè se $xS^T y$ e $yS^F x$;
- 2) preferenza, indicata con xHy , se x può essere migliore di y , ma non si è sicuri a causa di qualche evidenza contraria, cioè se $xS^T y$ e $yS^K x$;
- 3) semi preferenza, indicata con xJy , se x potrebbe essere migliore di y , ma non si è sicuri a causa della mancanza di tutte le necessarie informazioni, cioè $xS^T y$ e $yS^U x$;
- 4) preferenza semidebole, indicata con xLy , se x può essere migliore di y , ma si riscontrano informazioni contraddittorie e mancanza di informazioni necessarie, cioè $xS^K y$ e $yS^U x$;
- 5) indifferenza, indicata con xIy , se x e y sono strettamente equivalenti, cioè $xS^T y$ e $yS^T x$;
- 6) ambiguità, indicata con xKy , se x e y potrebbero essere indifferenti, ma esistono contraddizioni in tutte e due le direzioni, cioè $xS^K y$ e $yS^K x$;
- 7) ignoranza, indicata con xUy , se mancano le informazioni per stabilire la relazione che lega x e y , cioè $xS^U y$ e $yS^U x$;
- 8) incomparabilità, indicata con xRy , se x e y sono in opposizione forte, cioè $xS^F y$ e $yS^F x$;

9) incomparabilità debole, indicata con xQy , se x potrebbe essere incomparabile con y , ma ci sono informazioni contraddittorie, cioè xS^Ky e yS^Fx ;

10) semi incomparabilità, indicata con xVy , se x può essere in opposizione a y , ma non si è sicuri a causa della mancanza di tutte le necessarie informazioni, cioè xS^Uy e yS^Fx .

Le precedenti relazioni binarie possono essere raccolte in una matrice simmetrica di modellizzazione delle preferenze (Tabella 3).

Tabella 3. Le dieci situazioni di preferenza

	yS^Tx	yS^Kx	yS^Ux	yS^Fx
xS^Ty	xIy	xHy	xJy	xPy
xS^Ky	yHx	xKy	xLy	xQy
xS^Uy	yJx	yLx	xUy	xVy
xS^Fy	yPx	yQx	yVx	xRy

Si noti che nell'approccio classico del surclassamento vengono utilizzate solamente due relazioni (S^T e S^F), definite direttamente con riferimento alla coppia (x,y) e alla sua controparte simmetrica (y,x) . Così si ottengono solo quattro relazioni: preferenza (xPy , yPx), indifferenza (xIy) e incomparabilità (xRy), presenti ai quattro angoli della matrice di preferenza della Tabella 3.

Nella diagonale principale della matrice di preferenza, sono raggruppate quattro relazioni simmetriche: le già note indifferenza I ($xS^T y$ e $yS^T x$) ed incomparabilità R ($xS^F y$ e $xS^F y$) e le due nuove relazioni di ambiguità K ($xS^K y$ e $yS^K x$) e ignoranza U ($xS^U y$ e $yS^U x$).

Le due esitazioni tra preferenza e indifferenza sono tutte denominate come “preferenze” mentre le due esitazioni tra preferenza ed incomparabilità sono denominate come “incomparabilità”. Tutte queste relazioni potrebbero essere considerate come aventi un comune grado di preferenza tra la preferenza stretta e la relazione simmetrica. Inoltre, si usa “semi” solamente per esitazioni dovute alla non conoscenza e “debole” solamente per esitazioni dovute a situazioni contraddittorie. Così, si costruiscono altre cinque differenti (strettamente, semi, debolmente) relazioni asimmetriche ed un'altra (semi-debolmente) relazione simmetrica (Tabella 3).

Questo modo di rappresentare le preferenze consente di considerare tre differenti livelli di preferenza anziché solamente le due situazioni ottenute utilizzando l'approccio di surclassamento tradizionale (P,I,R) o il modello classico (P,I).

3. MODELLI DI AGGREGAZIONE DELLE PREFERENZE

3.1 DOMINANZA

Un concetto molto importante nel contesto dell'aiuto multicriteriale alla decisione è la relazione di dominanza. Date $a, b \in A$, si dice che a domina b , indicato con aDb , se e solo se $g_j(a) \geq g_j(b)$, $\forall g_j \in G$, dove almeno una delle diseguaglianze è stretta. In altri termini, l'azione a domina l'azione b se presenta valutazioni migliori o uguali a quelle di b su tutti i criteri considerati, con almeno una di esse strettamente migliore. Si osservi che la dominanza è una relazione oggettiva; essa è un concetto valido per tutti i decisori, perché non dipende dalla differente importanza soggettiva che diversi decisori possono attribuire ai criteri considerati, ma solamente dalle corrispondenti valutazioni (anche soltanto ordinali) delle azioni.

Strettamente legato al concetto di dominanza è il concetto di azione efficiente: l'azione $a \in A$ si dice efficiente se e solo se nessun'altra azione di A la domina. In altri termini, se l'azione a è efficiente, non è possibile trovare un'altra azione ammissibile b che sia migliore di a su (almeno) un criterio senza che sia peggiore su almeno un altro criterio. Spesso si indica con AE il sottoinsieme di A contenente le azioni efficienti (non dominate), dette anche di ottimo paretiano; la loro ricerca è anche chiamata problema di "ottimizzazione vettoriale". Ovviamente l'introduzione o l'eliminazione di un'azione o di un criterio possono

modificare le relazioni di dominanza e l'insieme delle azioni efficienti. La proprietà dell'efficienza è, infatti, una proprietà relativa, dipendente dalla composizione di A e di G e può mutare con l'alterazione di almeno una relazione di preferenza.

La dominanza di a su b rappresenta, dunque, l'unanimità dei punti di vista di G in favore di a rispetto a b . Perciò essa potrebbe essere un'informazione molto importante per fornire una soluzione al problema di decisione considerato. Per esempio, le possibili soluzioni per i problemi decisionali precedentemente elencati potrebbero essere le seguenti:

- 1) in un problema di scelta, si può selezionare e concentrare l'attenzione sull'insieme delle azioni efficienti trascurando le altre;
- 2) in un problema di classificazione dove le azioni di A dovrebbero essere divise nelle due categorie di "azioni buone" e "azioni cattive", una volta fissata un'azione c come "punto di riferimento medio", si può considerare buona ogni azione $a \in A$ tale che aDc e cattiva ogni azione $b \in A$ tale che cDb ,
- 3) in un problema di ordinamento si può considerare l'ordine stabilito dalla relazione di dominanza in A , cioè, $\forall a, b \in A$, l'azione a sarà ordinata meglio di (ossia precede) b se aDb .

Sfortunatamente la relazione di dominanza è generalmente abbastanza "povera", perché molto spesso solamente alcune coppie di azioni la soddisfano e talvolta addirittura nessuna coppia. Questo significa che le "semplici soluzioni" proposte

per i problemi 1), 2) e 3) potrebbero non essere applicabili nei problemi reali perché potrebbero funzionare solamente se:

- con riferimento al problema 1), si ha un piccolo numero di azioni efficienti (ma rimane il rischio di trascurare il “second best”, cioè un’azione che potrebbe scegliersi in subordine all’azione selezionata se questa dovesse risultare impraticabile; infatti il “second best” potrebbe risultare non efficiente),
- con riferimento al problema 2), per ogni azione $x \in A$ si ha $x D c$ o $c D x$,
- con riferimento al problema 3), per ogni coppia di azioni $a, b \in A$ si ha $a D b$ o $b D a$.

Perciò nei problemi di decisione reali deve essere utilizzata qualche procedura di aggregazione che, sulla base di un predefinito modello di preferenza, arricchisca la relazione di dominanza al fine di poter confrontare tutte le azioni di A con riferimento ai criteri di G . Per tutte queste procedure occorre, ancora, che il decisore fornisca alcune opportune informazioni preferenziali richieste dallo specifico modello di preferenza adottato.

3.2 PROCEDURE ELEMENTARI DI AGGREGAZIONE

Si indicano con questo nome dei metodi di aggregazione che vengono intuitivamente proposti quando si affronta un problema decisionale multicriteriale. Molte di queste procedure vengono applicate nella realtà per la loro semplicità ed immediata comprensione da parte del decisore. Purtroppo,

spesso, proprio per queste ragioni rischiano di affrontare il problema decisionale in maniera troppo semplicistica, ignorandone degli aspetti rilevanti e rappresentando in maniera molto approssimativa le preferenze del decisore. Nel seguito si considereranno le preferenze del decisore crescenti con i valori di g_j , $\forall g_j \in G$. Questa assunzione non lede la generalità perché laddove le preferenze dovessero essere decrescenti rispetto ai valori di g_j si potranno ricodificare i valori di g_j , per esempio sostituendovi il loro opposto. Le principali procedure sono: la somma ponderata, il massimo, il minimo, il metodo di Borda, il metodo di Condorcet, il metodo a livelli di aspirazione ed il modello lessicografico. Nei paragrafi successivi saranno accennati, a titolo esemplificativo, solo le prime tre procedure.

3.2.1 SOMMA PONDERATA

Il metodo più elementare è quello che associa ad ogni azione $x \in A$ la somma ponderata $W(x)$ (ossia la media ponderata) delle sue valutazioni quantitative $g_j(x)$, $j=1, 2, \dots, m$, con pesi λ_j che indicano i tassi di sostituzione tra i vari criteri, ossia $W(x) = \sum_{j=1}^m \lambda_j g_j(x)$. Talvolta i pesi λ_j vengono considerati dei coefficienti di importanza dei criteri corrispondenti. In questo caso le valutazioni rispetto ai diversi criteri devono essere opportunamente normalizzate, ossia ricondotte ad una comune unità di misura. Ovviamente tale procedura ammette una

compensazione totale tra scarti positivi e negativi nelle valutazioni rispetto ai differenti criteri. Si ha, $\forall a, b \in A$:

$$aPb \Leftrightarrow W(a) > W(b),$$

$$aIb \Leftrightarrow W(a) = W(b).$$

La struttura di preferenza $\{P, I\}$ così ottenuta costituisce un preordine totale. Tale metodo può essere adottato quando si tratta di aggregare grandezze sufficientemente omogenee, che rappresentano diversi aspetti di una stessa caratteristica (per esempio, i voti degli studenti in differenti materie). Ma la sua natura totalmente compensatoria lo rende particolarmente inaffidabile quando si devono confrontare azioni su criteri conflittuali e profondamente diversi. Tale inaffidabilità viene accresciuta nel caso in cui si opti per ridurre i criteri ad una comune unità di misura

3.2.2 MASSIMO

Con tale approccio si associa ad ogni azione $x \in A$ la massima valutazione da essa ottenuta con riferimento a tutti i criteri g_j considerati, ossia $M(x) = \max_{g_j \in G} g_j(x)$. Si

ha, $\forall a, b \in A$:

$$aPb \Leftrightarrow M(a) > M(b),$$

$$aIb \Leftrightarrow M(a)=M(b).$$

La struttura di preferenza $\{P, I\}$ così ottenuta costituisce un preordine completo.

Anche in tal caso le valutazioni rispetto ai diversi criteri devono essere espresse nella stessa unità di misura. In tale procedura ha effettiva rilevanza solamente la valutazione massima di ogni azione ed il metodo risulta parzialmente compensatorio. Tale approccio caratterizza ogni azione mediante la sua migliore performance (ossia premia azioni che presentano “picchi” di valutazioni rispetto ad azioni con valutazioni più uniformi), e prescinde dalle informazioni sugli altri criteri. Molte rilevanti informazioni vengono pertanto ignorate.

3.2.3. MINIMO

Tale approccio, che in un certo senso costituisce il simmetrico del precedente, associa ad ogni azione $x \in A$ la minima valutazione da essa ottenuta con riferimento a tutti i criteri g_j considerati, ossia $m(x) = \min_{g_j \in G} g_j(x)$. Si ha, $\forall a, b \in A$:

$$aPb \Leftrightarrow m(a) > m(b),$$

$$aIb \Leftrightarrow m(a) = m(b).$$

Anche in tal caso, la struttura di preferenza $\{P, I\}$ ottenuta costituisce un preordine completo e le valutazioni rispetto ai diversi criteri devono essere

espresse nella stessa unità di misura. In tale metodo ha effettiva rilevanza solamente la valutazione minima di ogni azione ed esso risulta parzialmente compensatorio. Tale approccio caratterizza ogni azione con la sua peggiore performance (e quindi premia le azioni che non presentano situazioni molto sfavorevoli rispetto a qualche criterio), prescindendo comunque dalle informazioni sugli altri criteri, per cui molte rilevanti informazioni vengono ancora ignorate.

4. CARATTERISTICHE FONDAMENTALI DI UNA PROCEDURA DI AGGREGAZIONE MULTICRITERIALE

4.1 PROCEDURE DI AGGREGAZIONE COMPENSATORIE E NON COMPENSATORIE

L'idea intuitiva di una procedura di aggregazione compensatoria è quella che un peggioramento di una valutazione di un'azione su un certo criterio possa essere compensata, ossia bilanciata, dal miglioramento di una sua valutazione rispetto ad uno (o più) differenti criteri tra quelli considerati. Tale concetto, pertanto, si basa sul ruolo cruciale che riveste l'intensità di preferenza espressa da ciascun criterio: quel che effettivamente conta ai fini della comparazione globale tra due azioni è di quanto un'azione sia preferita all'altra rispetto a ciascun criterio considerato, piuttosto che rispetto a quali criteri essa sia preferita.

Invece, l'idea di base di una procedura di aggregazione non compensatoria è che, ai fini della aggregazione delle preferenze, si tengono in considerazione solo informazioni di carattere ordinale sui singoli criteri. Più precisamente, una procedura di aggregazione multicriteriale è non-compensatoria se, date due azioni a e b tali che a è globalmente preferita a b , se l'insieme dei criteri per cui a è preferita a b si accresce e l'insieme dei criteri per cui b è preferita ad a si restringe, allora a continua ad essere preferita a b a prescindere dall'intensità di preferenza espressa dai diversi criteri. Formalmente una procedura di aggregazione è non-compensatoria (Fishburn, 1976) se $\forall a, b, c, d \in A$ si ha:

$$[\{gj \in G: aPjb\} \subseteq \{gj \in G: cPjd\} \text{ e } \{gj \in G: bPja\} \supseteq \{gj \in G: dPjc\}] \Rightarrow [aPb \Rightarrow cPd].$$

Alcune procedure di aggregazione considerano solo i criteri per cui a è almeno tanto buona quanto b . Più precisamente, una procedura di aggregazione multicriteriale è unilateralmente non-compensatoria se, date due azioni a e b tali che a è globalmente almeno tanto buona quanto b , se l'insieme dei criteri per cui a è almeno tanto buona quanto b si accresce, allora a continua ad essere almeno tanto buona quanto b a prescindere dall'intensità di preferenza espressa dai diversi criteri. Formalmente una procedura di aggregazione è unilateralmente non-compensatoria se $\forall a, b, c, d \in A$ si ha:

$$[\{gj \in G: aSjb\} \subseteq \{gj \in G: cSjd\}] \Rightarrow [aSb \Rightarrow cSd].$$

Più in generale, il termine non-compensazione rimanda all'idea che esistono delle situazioni in cui le intensità di preferenza di a su b non vengono prese in considerazione per stabilire la relazione di preferenza tra a e b . Ciò può accadere anche in presenza di qualche situazione di veto, che si ha quando la preferenza di b su a rispetto ad almeno un criterio è talmente forte da impedire che si possa dichiarare che globalmente a è almeno tanto buona quanto b . In un'applicazione ambientale questo potrebbe essere il caso di una situazione fortemente a favore di b su a in termini di presenza di un certo agente inquinante da far escludere che a possa essere dichiarato almeno tanto buono quanto b , qualunque sia l'insieme dei criteri in favore di a e contro b e qualunque sia l'intensità di preferenza di a su questi criteri. Formalmente, per definire un veto rispetto a un criterio $g_j \in G$ si introduce una soglia $v_j > 0$ tale che, $\forall a, b \in A$, si ha:

$$g_j(b) - g_j(a) \geq v_j \Rightarrow \text{non } aSb.$$

Si osservi che la presenza di un veto in una procedura di aggregazione impedisce di classificare tale procedura come non-compensatoria: infatti, il veto considera l'intensità di preferenza e anzi esso stesso è basato sull'idea che l'intensità della preferenza di a su b sia molto elevata. Una procedura di aggregazione con veti non si può neanche classificare come unilateralmente non-compensatoria: infatti può benissimo succedere che $\{g_j \in G: aS_j b\} \subseteq \{g_j \in G: cS_j d\}$ e si ha aSb ma non si

ha $c \succ d$ a causa della presenza di un veto perché su qualche criterio la preferenza di d su c è così forte da impedire di poter concludere che c è almeno tanto buono quanto d .

Infine, il termine non-compensazione può rimandare al caso in cui entro certi limiti variazioni molto rilevanti delle valutazioni dei criteri considerati non modificano le preferenze. E' questo il caso delle procedure di aggregazione multicriteriali basate sul minimo e sul massimo. Si consideri dapprima un semplice esempio relativo all'operatore di aggregazione "minimo". Si supponga di avere una azione a che ha le seguenti valutazioni sui tre criteri considerati nel problema di decisione affrontato: $g_1(a)=5$, $g_2(a)=7$, $g_3(a)=10$. La valutazione complessiva dell'azione a è pertanto $U(a)=5$, corrispondente alla valutazione data dal criterio g_1 . Si supponga ora che per una qualche ragione la valutazione dell'azione a rispetto al criterio g_1 si modifichi passando da $g_1(a)=5$ a $g_1(a)=4$. In questo caso la valutazione complessiva dell'azione a passa da $U(a)=5$ a $U(a)=4$. Si osservi che anche se le valutazioni di a rispetto a g_2 e a g_3 raddoppiassero, passando rispettivamente da $g_2(a)=7$ a $g_2(a)=14$ e da $g_3(a)=10$ a $g_3(a)=20$, la valutazione complessiva di a continuerebbe a rimanere $U(a)=4$. Vale a dire una diminuzione anche molto piccola sul criterio che ha la valutazione minima non può essere "compensata" da nessun incremento anche molto grande sugli altri criteri. Formalmente in questo caso si parla di non-sostituibilità totale (Sounderpandian, 1991). Una situazione simile si ha per l'aggregazione basata sull'operatore "massimo". In questo caso un incremento, anche molto piccolo sul

criterio che dà la massima valutazione, comporta un incremento nella valutazione globale, che non può essere intaccato nemmeno da decrementi molto grandi sugli altri criteri. Questa situazione si definisce di sostituibilità esclusiva totale (Sounderpandian 1991).

4.2 DIFFERENTI TIPI DI SCALE

Le funzioni di utilità possono essere rappresentate utilizzando differenti scale di misurazione. Queste scale si caratterizzano con riferimento alle cosiddette *trasformazioni ammissibili*, cioè trasformazioni che conducono da una scala accettabile ad un'altra senza alterarne il contenuto informativo. Ipotizzando che una scala assegni una valutazione $x \in \mathbf{R}$ all'oggetto misurato, si ha (Roberts, 1979):

- una **scala assoluta**, se le trasformazioni ammissibili sono della forma $\varphi(x)=x$ (identità): un tipico esempio di scala assoluta è il contare;
- una **scala di rapporti** (*ratio scale*), se le trasformazioni ammissibili sono della forma $\varphi(x)=\alpha x$, $\alpha > 0$ (*trasformazioni di similarità*, in cui esiste uno "zero" naturale): tipici esempi di scala di rapporti sono la massa (misurata in kg, libbre,...) ed i prezzi (misurati in valute diverse);
- una **scala di intervalli** (*interval scale*), se le trasformazioni ammissibili sono della forma $\varphi(x)=\alpha x + \beta$, $\alpha > 0$ (*trasformazioni lineari positive*): un esempio tipico di una scala di intervalli è la temperatura (misurata in gradi centigradi

ed in Fahrenheit, mentre quando si definisce uno zero assoluto, come nella scala Kelvin, si ha una scala di rapporti);

- una **scala ordinale**, se le trasformazioni ammissibili sono della forma $\varphi(x)$ ove $\varphi(\cdot)$ è una funzione strettamente crescente (*trasformazioni strettamente crescenti*): un esempio tipico di una scala ordinale è la scala di durezza di Mohs;
- una **scala nominale**, se le trasformazioni ammissibili sono della forma $\varphi(x)$ ove $\varphi(\cdot)$ è una qualsiasi funzione *iniettiva*: un tipico esempio di una scala nominale sono le numerazioni assegnate a progetti alternativi, che possono permutarsi arbitrariamente.

Le funzioni di utilità rappresentano preferenze utilizzando scale di rapporti, scale di intervalli o scale ordinali. Utilizzando una scala di rapporti, la scala è determinata a meno della scelta di una unità di misura; pertanto, è possibile effettuare confronti tra i *rapporti dei valori* di due azioni considerate. Utilizzando una scala di intervalli, la scala è determinata a meno della scelta di una unità di misura e di uno zero (origine); quindi, considerando quattro azioni a, b, c, d , è possibile effettuare confronti tra i *rapporti delle differenze dei valori* di una coppia di azioni (a, b) rispetto ad un'altra (c, d) , ossia possono misurarsi delle intensità di preferenza. Utilizzando una scala ordinale, la scala è determinata solamente in base a un ordinamento; non si può, pertanto, operare sui valori delle azioni, ma può solamente affermarsi se un'azione precede o segue un'altra.

Il contenuto informativo delle diverse scale si indebolisce passando da scale di rapporti a scale di intervalli ed a scale ordinali. Spesso nella realtà, si hanno valutazioni qualitative di tipo ordinale (per esempio, grado di inquinamento alto, medio, basso); in tal caso sarebbe un grave errore metodologico quello di “forzare” le informazioni, ossia attribuire un valore cardinale a dati puramente ordinali. Purtroppo si assiste spesso nella pratica applicazione di talune metodologie all’esecuzione di operazioni matematiche, anche elementari, su numeri che altro non sono che codificazioni numeriche di informazioni puramente ordinali.

In alcuni casi si richiede che una funzione di utilità rappresenti le *intensità di preferenza*. In questo caso per ogni $a, b, c, d \in A$ si ha che la preferenza globale di a su b è almeno uguale alla preferenza globale di c su d se e solo se

$$U(a) - U(b) \geq U(c) - U(d).$$

In questo caso, come sopra accennato, la funzione di utilità deve essere espressa su una scala di intervalli: infatti un’altra funzione di utilità $U'(\cdot)$ rappresenta la stessa struttura di preferenza se e solo se $U'(\cdot) = \alpha U(\cdot) + \beta$, con $\alpha \in \mathbf{R}^+$ e $\beta \in \mathbf{R}$.

5. I MODELLI MULTICRITERIALI

I tre principali modelli di aggregazione delle preferenze utilizzati sono:

1) il modello dell'utilità multicriteriale, detto anche MAUT (Multiattribute Utility Theory) o anche funzionale, applicato nella teoria dell'utilità multiattributo (Keeney e Raiffa 1976),

2) il modello della relazione di surclassamento (o modello outranking) detto anche relazionale, la cui rappresentazione più largamente conosciuta è nella forma di una relazione binaria di surclassamento (Roy 1985) e di una relazione fuzzy (Fodor e Roubens, 1994).

3) il modello interattivo, basato su un susseguirsi di fasi di discussione tra l'analista e il decisore e di fasi di calcolo.

A questi tre approcci, di recente, se ne è affiancato un quarto detto delle regole decisionali, nel quale le preferenze del decisore vengono rappresentate mediante un insieme di proposizioni "se..., allora..." (regole decisionali) (Greco, Matarazzo e Slowinski, 1999, 2001, 2005). A quest'ultimo approccio appartiene la metodologia nota come "Rough-Sets" che sarà affrontata nel prossimo capitolo.

CAPITOLO 3

L'APPROCCIO DEI ROUGH SETS ALL'ANALISI DELLE DECISIONI

1.INTRODUZIONE

La teoria dei rough sets (insiemi approssimati), introdotta da Pawlak (1982,1991), si è dimostrata spesso un eccellente strumento matematico per analizzare dati caratterizzati da imprecisione e vaghezza nella loro descrizione. Essa è fondata sull'assunzione che ad ogni oggetto dell'universo del discorso è associata qualche informazione (dati, conoscenza), espressa utilizzando opportuni attributi che descrivono gli oggetti considerati. Per esempio, se gli oggetti sono delle imprese che richiedono un affidamento bancario, le informazioni sono date dalle loro caratteristiche finanziarie, economiche e tecniche, che costituiscono la loro descrizione. Oggetti caratterizzati dalla stessa descrizione sono indiscernibili (similari) con riferimento alle informazioni disponibili. La relazione di indiscernibilità così generata costituisce il

fondamento matematico della teoria dei rough sets, i mattoni con cui si costruisce l'edificio della conoscenza della realtà.

Ogni insieme di oggetti indiscernibili si chiama insieme elementare e costituisce un granulo elementare (atomo) della conoscenza dell'universo¹⁵. Un qualunque sottoinsieme Y dell'universo può essere espresso in termini di granuli o in maniera precisa (unione di insiemi elementari) o solo approssimativamente. In quest'ultimo caso, il sottoinsieme Y può essere caratterizzato da due insiemi ordinari, chiamati approssimazione inferiore e superiore. Un rough set è definito mediante queste due approssimazioni, che coincidono nel caso di un insieme ordinario. L'approssimazione inferiore di Y è formata da tutti gli insiemi elementari inclusi in Y (i cui elementi, quindi, appartengono sicuramente a Y), mentre l'approssimazione superiore di Y è costituita da tutti gli insiemi elementari che hanno un'intersezione non vuota con Y (i cui elementi, quindi, possono appartenere a Y). Ovviamente, la differenza tra l'approssimazione superiore e quella inferiore costituisce la frontiera (boundary region) del rough set, i cui elementi non possono essere di conseguenza caratterizzati con certezza circa l'appartenenza a Y , usando le informazioni disponibili. Chiaramente, negli insiemi ordinari la frontiera è vuota. La cardinalità della frontiera esprime, ancora, in che misura è possibile esprimere Y in termini esatti, in base alle informazioni disponibili.

¹⁵ MATARAZZO B. (1997), "L'approccio dei rough sets all'analisi delle decisioni", Atti del XXI Convegno Annuale A.M.A.S.E.S., Appendice, Roma, pp. 77-111.

In tale approccio, quindi, due distinti oggetti possono apparire indiscernibili (similari) usando le informazioni che li caratterizzano, come conseguenza della granularità della conoscenza, peculiare dei rough sets. Pertanto, ogni concetto descritto solamente in maniera vaga, nella filosofia di tale approccio può essere rimpiazzato da una coppia di concetti precisi, le sue approssimazioni inferiore e superiore.

La teoria dei rough sets, che si propone di analizzare possibili relazioni di causa-effetto tra i dati imperfetti (caratterizzati da incertezza e vaghezza) disponibili, presenta talune intersezioni e si pone in alcuni casi come complementare a molte altre teorie matematiche che trattano l'incertezza e l'imprecisione: teoria della probabilità, analisi discriminante, etc..

Taluni importanti caratteristiche dell'approccio dei rough sets rendono tale strumento particolarmente interessante in numerose applicazioni a problemi concreti. Con riferimento all'input (informazioni richieste), è possibile trattare dati qualitativi (anzi, i dati quantitativi vanno in qualche maniera "discretizzati") e non è necessario effettuare alcuna analisi a priori circa la consistenza dei dati da analizzare. Con riferimento all'output (informazioni ottenibili), è possibile avere a posteriori informazioni circa il ruolo (l'importanza) che taluni attributi o loro sottoinsiemi hanno nell'analisi del problema affrontato (senza dover predefinire trade-offs, ecc.) e si ottengono risultati facilmente comprensibili nella forma di regole decisionali "se....., allora....." utilizzando gli attributi più rilevanti.

2. CLASSICAL ROUGH SETS APPROACH (CRSA)

2.1 TAVOLA DELLE INFORMAZIONI E RELAZIONE DI INDISCERNIBILITÀ

Le informazioni circa gli oggetti vengono fornite, per ragioni algoritmiche, sotto forma di una tavola, le cui righe si riferiscono ai distinti oggetti e le colonne ai diversi attributi considerati; ogni cella indicherà quindi la valutazione (quantitativa o qualitativa) dell'oggetto posto in quella riga tramite l'attributo della corrispondente colonna. Nel caso di valutazioni quantitative su un certo attributo q , il dominio di quest'ultimo viene opportunamente suddiviso in sottointervalli, che forniscono una buona descrizione del fenomeno studiato, e codificato conseguentemente. Il problema della discretizzazione dei dati quantitativi è abbastanza delicato, in quanto i risultati delle analisi possono dipendere dalla discretizzazione adottata.

Formalmente, una tavola delle informazioni è la 4-upla $S = \langle U, Q, V, f \rangle$ dove a ogni oggetto dell'universo U considerato, è associato un certa quantità di informazioni relative a una serie di attributi $Q = (q_1, q_2, q_3, \dots, q_m)$ tramite la funzione $f: U \times Q \rightarrow V$ con V insieme dei valori con i quali viene espressa

l'informazione. Pertanto, la tabella d'informazione contiene l'universo U degli oggetti considerati e l'insieme finito Q degli attributi.

Si indica con V_q l'insieme dei valori assunti da un attributo $q \in Q$ e pertanto

$V = \bigcup_{q \in Q} V_q$. La funzione di informazione f è una funzione definita in $U \times Q$ che ha

valori in V , tale che per ogni $q \in Q$ e $x \in U$ $f(x, q) \in V$. In parole semplici $f(x, q)$ restituisce il valore dell'attributo $q \in Q$ relativo all'oggetto $x \in U$.

Pertanto, ogni oggetto x di U sarà descritto da un vettore (stringa), ogni elemento del quale rappresenta il valore che il corrispondente attributo assume per x ; detto vettore è chiamato descrizione di x in termini delle valutazioni degli attributi di Q e denotato $\text{Des}_Q(X)$. Naturalmente può ottenersi una descrizione di $x \in U$ in termini di un qualunque sottoinsieme non vuoto $P \subseteq Q$.

Ad ogni sottoinsieme (non vuoto) di attributi P è associata una relazione di indiscernibilità su U , indicata con I_P :

$$I_P = \{(x, y) \in U \times U : f_q(x) = f_q(y), q \in P\}$$

Ovviamente, la relazione binaria di indiscernibilità così definita è una relazione di equivalenza (riflessiva, simmetrica e transitiva). La famiglia di tutte le classi di equivalenza della relazione I_P viene denotata con U/I_P e la classe di equivalenza contenente un elemento $x \in U$ con $I_P(X)$. Se $(x, y) \in I_P$, si dice che gli oggetti x e

y sono P-indiscernibili. Le classi di equivalenza della relazione I_p sono chiamate insiemi P-elementari. Se $P=Q$, gli insiemi Q-elementari sono chiamati atomi.

2.2 APPROSSIMAZIONI

Siano X un sottoinsieme non vuoto di U e $P \subseteq Q$. La P-approssimazione inferiore e P-approssimazione superiore di X sono definite rispettivamente da:

$$\underline{P}(X) = \{x \in U : I_p(X) \subseteq X\},$$

$$\overline{P}(X) = \{x \in U : I_p(X) \cap X \neq \emptyset\}.$$

In altri termini, gli elementi di $\underline{P}(X)$ sono tutti e solo gli $x \in U$ appartenenti a tutte le classi generate dalla relazione di indiscernibilità I_p e contenuti in X ; gli elementi di $\overline{P}(X)$ sono tutti e solo gli $x \in U$ appartenenti a tutte le classi generate dalla relazione di indiscernibilità I_p che hanno almeno un rappresentante appartenente ad X .

La frontiera di X , denotata con $Bn_p(X)$, è $Bn_p = \overline{P}(X) - \underline{P}(X)$ e vale la seguente relazione $\underline{P}(X) \subseteq X \subseteq \overline{P}(X)$. Pertanto se un oggetto x appartiene a $\underline{P}(X)$, esso è certamente anche un elemento di X , mentre se x appartiene a $\overline{P}(X)$, esso può appartenere all'insieme X . $Bn_p(X)$ costituisce la “regione del dubbio”

di x : nulla può dirsi con certezza circa l'appartenenza dei suoi elementi all'insieme X .

Se la frontiera di X è vuota $Bn_p(X) = \emptyset$, allora l'insieme X è un insieme ordinario (esatto) rispetto a P , ossia esso può esprimersi come unione di un certo numero di insiemi P -elementari; altrimenti se $Bn_p(X) \neq \emptyset$, l'insieme X è un insieme approssimato (rough) rispetto a P , caratterizzabile mediante le approssimazioni $\underline{P}(X)$ e $\overline{P}(X)$. La famiglia di tutti gli insiemi $X \subseteq U$ aventi le stesse approssimazioni inferiore e superiore si chiama rough set.

Si definisce accuratezza dell'approssimazione di X , $X \neq \emptyset$, mediante gli attributi P il rapporto:

$$\alpha_p = \frac{|\underline{P}(X)|}{|\overline{P}(X)|}$$

dove $|A|$ indica il cardinale di un insieme A , finito. Risulta naturalmente $0 \leq \alpha_p(X) \leq 1$; se $\alpha_p(X) = 1$, X è un insieme ordinario (preciso) rispetto a P ; se $\alpha_p(X) < 1$, X è un insieme rough (vago) rispetto a P .

Si definisce ancora qualità dell'approssimazione di X mediante gli attributi di P il rapporto:

$$\gamma_P = \frac{|P(X)|}{|X|}$$

Risulta $0 \leq \alpha_P(X) \leq \gamma_P(X) \leq 1$ e la qualità rappresenta la frequenza relativa degli oggetti correttamente classificati usando gli attributi di P.

Se si considera un concetto vago, ossia allorché gli elementi dell'universo non possono essere classificati con certezza come appartenenti al concetto, l'incertezza è collegata al grado di appartenenza degli elementi all'insieme. Allora, per discutere il problema dell'incertezza dal punto di vista dei rough sets, occorre definire la funzione di appartenenza $\mu_x^P(X)$ collegata al concetto di rough set (rough membership function). Utilizzando la relazione di indiscernibilità, si ottiene:

$$\mu_x^P(X) = \frac{|X \cap I_P(X)|}{|I_P(X)|}$$

Il valore $\mu_x^P(X)$ può essere interpretato in qualche caso come una probabilità condizionata, e può essere inteso come il grado di certezza (credibilità) con cui x appartiene a X.

Tra la rough membership function e le approssimazioni di X valgono le tre seguenti relazioni:

$$\underline{P}(X) = \{x \in U : \mu_X^P(X) = 1\}$$

$$\overline{P}(X) = \{x \in U : \mu_X^P(X) > 0\}$$

$$Bn_P(X) = \{x \in U : 0 < \mu_X^P(X) < 1\}$$

Nella teoria dei rough sets vi è, quindi, una stretta relazione tra vaghezza, insita negli insiemi e richiedente quindi le approssimazioni, ed incertezza, collegata agli elementi degli insiemi e per la quale è necessario introdurre il grado di appartenenza approssimativo. La peculiarità dei rough sets consiste nel trattare una rappresentazione imprecisa della realtà dovuta alla granularità della conoscenza, conseguenza della indiscernibilità tra oggetti aventi la stessa descrizione (“granuli”).

2.3 RIDOTTI E CORE

Un concetto molto importante per le applicazioni concrete è quello di dipendenza di attributi. Intuitivamente, un insieme di attributi $T \subseteq Q$ dipende totalmente da un insieme di attributi $P \subseteq Q$, notazione $P \rightarrow T$, se tutti i valori degli attributi di T sono unicamente determinati dai valori degli attributi di P , ossia se sussiste una dipendenza funzionale tra i valori assunti dagli attributi di T e di P . In altri termini, la partizione generata dagli attributi di P è più piccola di quella generata

dagli attributi di T, per cui è sufficiente adoperare gli attributi di T, per costruire la partizione U/I_T ; formalmente, T dipende totalmente da P se e solo se $I_P \subseteq I_T$.

Quindi, T è totalmente (parzialmente) dipendente da P se tutti (alcuni) elementi dell'universo U possono essere inequivocabilmente classificati come classi della partizione U/I_T , utilizzando solamente gli attributi di P.

Un'altra questione di grande rilievo per le applicazioni operative è quella concernente il problema dell'eventuale presenza di dati "superflui" in una tavola delle informazioni. I dati superflui, infatti, possono essere eliminati senza deteriorare le informazioni contenute nella tavola originale. Sia $P \subseteq Q$ e $p \in P$.

Si dice che l'attributo p è superfluo in P se $I_P = I_{P-\{p\}}$ altrimenti p è indispensabile in P. L'insieme P è indipendente (ortogonale) se tutti i suoi attributi sono indispensabili. Il sottoinsieme P' di P è un ridotto (reduct, notazione $\text{Red}(P)$) di P se P' è indipendente e $I_{P'} = I_P$.

Pertanto, un ridotto è un insieme di attributi che preserva le partizioni, cioè è un sottoinsieme minimale di attributi che consente di ottenere le stesse classificazioni, e quindi la stessa qualità dell'approssimazione, degli elementi di U ottenibili usando l'intero insieme di attributi P. In altri termini, gli attributi che non appartengono ad un ridotto sono superflui rispetto alle classificazioni degli elementi dell'universo.

Possono esistere più ridotti di P in una tavola delle informazioni. Dicesi nucleo (core) di P l'insieme contenente tutti gli attributi indispensabili di P, formalmente:

$$\text{Core}(P) = \bigcap \text{Red}(P)$$

Ovviamente, poiché il nucleo è l'intersezione di tutti i ridotti, esso è incluso in ogni ridotto di P , ossia il nucleo appartiene ad ogni ridotto. In altri termini, il nucleo è il più importante sottoinsieme di attributi di Q , in quanto nessuno dei suoi elementi può essere rimosso senza deteriorare la qualità della classificazione.

Il calcolo di tutti i ridotti è piuttosto complesso. Tuttavia, in molte applicazioni concrete non è necessario calcolare tutti i ridotti, ma solamente alcuni di essi. Ai fini operativi, dunque, è sufficiente prendere in considerazione solamente i più importanti attributi (ridotti) per l'analisi della tavola delle informazioni considerata.

2.4 TAVOLA DELLE DECISIONI E REGOLE DECISIONALI

Se in una tavola delle informazioni gli attributi di Q vengono distinti in attributi condizionali (insieme C) e attributi decisionali (insieme D), $C \cup D = Q$ e $C \cap D = \emptyset$, detta tavola è chiamata tavola delle decisioni. Gli attributi decisionali inducono delle partizioni di U dedotte dalla relazione di indiscernibilità I_D , in maniera assolutamente indipendente dagli attributi condizionali di C . Nelle applicazioni operative, si tende a ridurre gli attributi condizionali preservando la

dipendenza tra attributi condizionali e decisionali. In altri termini, si vuole usare il minor numero possibile di attributi condizionali senza deteriorare la qualità dell'approssimazione della classificazione indotta dagli attributi decisionali.

Poiché si tende a evidenziare la dipendenza funzionale tra gli attributi condizionali e quelli decisionali, una tavola delle decisioni può anche essere espressa come un insieme di regole decisionali. Queste sono delle proposizioni logiche (implicazioni) del tipo “se..., allora...”, ove l'antecedente riguarda valori assunti da uno o più attributi condizionali (descrizioni di insiemi C- elementari) ed il conseguente partizioni generate dagli attributi decisionali (descrizioni di insiemi D - elementari). Se queste ultime contengono le partizioni corrispondenti agli attributi condizionali considerati, la regola decisionale si dice esatta o certa; altrimenti si parla di regole decisionali approssimate o incerte. Formalmente, si ha una regola esatta se $I_C \subseteq I_D$, approssimata se $I_C \cap I_D = \emptyset$.

Il calcolo delle regole decisionali è spesso complesso ed esistono al riguardo numerosi algoritmi. Tuttavia, nelle applicazioni concrete spesso non è necessario conoscere tutte le regole decisionali, ma solamente l'insieme minimale di queste, che fornisce le stesse informazioni dell'insieme completo, ma è di dimensioni più ridotte e facilmente comprensibile ed applicabile.

2.5 UN ESEMPIO PRATICO DI APPLICAZIONE DELLA METODOLOGIA CRSA

L'esempio che segue è stato ideato da Pawlak nel 1997¹⁶. Dati sei magazzini descritti dai seguenti quattro attributi:

- A_1 , capacità del personale di vendita,
- A_2 , qualità percepita della merce,
- A_3 , localizzazione ad alto traffico,
- A_4 , utili o perdite del magazzino.

Tabella.1.

Magazzino	A_1	A_2	A_3	A_4
1	alta	buona	no	utile
2	media	buona	no	perdita
3	media	buona	no	utile
4	bassa	media	no	perdita
5	media	media	si	perdita
6	alta	media	si	utile

¹⁶ Pawlak, Z., Rough sets approach to Knowledge-based decision support, European Journal of Operational Research, 99, 1997, 48-57

Si ha perciò $U = \{1,2,3,4,5,6\}$, $Q = \{A_1, A_2, A_3, A_4\}$, $V_1 = \{alta, media, bassa\}$, $V_2 = \{buona, media\}$, $V_3 = \{no, si\}$, $V_4 = \{utile, perdita\}$, e la tabella rappresenta la funzione dell'informazione $f(x, q)$ (per esempio $f(1, A_1) = alta$, $f(1, A_2) = buona$, e così via).

Si osservi che ogni magazzino ha una descrizione differente nei termini degli attributi A_1, A_2, A_3, A_4 , cosicché possono essere distinti, cioè sono discernibili, per mezzo dell'informazione fornita dagli attributi considerati. Formalmente si ha $I_Q = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$ e perciò non esistono due distinti magazzini x e y tali che $(x, y) \in I_Q$. Tuttavia i magazzini 2 e 3 sono indiscernibili nei termini degli attributi di $P = \{A_1, A_2, A_3\}$, dal momento che con riferimento a questi attributi, essi hanno gli stessi valori. Formalmente $I_P = \{(1,1), (2,2), (2,3), (3,2), (3,3), (4,4), (5,5), (6,6)\}$ e perciò $(2,3) \in I_P$ (e ovviamente anche $(3,2) \in I_P$). Allo stesso modo i magazzini 1,2 e 3 da una parte, e 5 e 6 dall'altra, sono indiscernibili con riferimento agli attributi $P' = \{A_2, A_3\}$ e così via considerando tutti i possibili sottoinsiemi di attributi di Q .

Ogni $P \subseteq Q$ determina una partizione U/I_P , che raggruppa nelle corrispondenti classi di equivalenza gli oggetti aventi la stessa descrizione nei termini degli attributi di P : per esempio per $P' = \{A_2, A_3\}$ si ha $U/I_{P'} = \{\{1,2,3\}, \{4\}, \{5,6\}\}$ e perciò $\{1,2,3\}, \{4\}, \{5,6\}$ sono gli insiemi P' – elementari.

Si supponga di voler approssimare tramite l'insieme di attributi $P = \{A_1, A_2, A_3\}$ l'insieme X dei magazzini che hanno conseguito un utile, cioè $X = \{1,3,6\}$. Dal momento che $U/I_p = \{\{1\}, \{2,3\}, \{4\}, \{5\}, \{6\}\}$ si ha:

$$\underline{P}(X) = \{1,6\}$$

$$\overline{P}(X) = \{1,2,3,6\}$$

$$Bn_p(X) = \{2,3\}$$

Questi risultati danno una risposta alla domanda se si può descrivere X per mezzo dell'informazione fornita dagli attributi di P . La risposta a questa domanda non è univoca. Si osservi che esiste una frontiera $Bn_p(X)$ non vuota: essa è costituita dai magazzini 2 e 3 che hanno la stessa descrizione nei termini degli attributi considerati ma tali che il magazzino 3 ha conseguito un utile mentre il magazzino 2 ha conseguito una perdita. Tuttavia, anche l'approssimazione inferiore di X , $\underline{P}(X)$, è non vuota: essa è costituita dai magazzini 1 e 6 che hanno una descrizione nei termini degli attributi considerati differente da tutti i magazzini non appartenenti a X . Riassumendo, in termini intuitivi, si può dire che, sulla base delle informazioni fornite dagli attributi di P :

1. i magazzini 1 e 6, appartenenti alla approssimazione inferiore, sicuramente appartengono all'insieme X dei magazzini che hanno utili,

2. i magazzini 1,2,3 e 6, appartenenti alla approssimazione superiore, potrebbero appartenere all'insieme X, dei magazzini che hanno utili,
3. i magazzini 2 e 3, che appartengono alla frontiera, rappresentano i casi di appartenenza dubbia all'insieme X dei magazzini che hanno utili.

Si considerino ora i seguenti sottoinsiemi di Q: $P = \{A_1, A_2, A_3\}$, $R = \{A_1, A_2\}$, $S = \{A_1, A_3\}$, $T = \{A_2, A_3\}$. Si osservi facilmente che $I_R = I_P$, $I_S = I_P$, mentre $I_T \neq I_P$. Questo significa che R e S sono ridotti di P mentre non lo è T. In altri termini, questo significa che R e S sono dei sottoinsiemi minimali di attributi che consentono di ottenere le stesse classificazioni degli elementi di U ottenibili usando l'insieme di attributi P. Si ha anche che il nucleo di P è dato $R \cap S$, cioè dall'attributo A_1 che in un certo senso costituisce l'attributo più importante per descrivere il magazzino, mentre gli attributo R e S possono essere mutualmente scambiati.

Se dall'insieme degli attributi Q si distinguono attributi condizionali, $C = \{A_1, A_2, A_3\}$, e attributo decisionale $D = \{A_4\}$, allora la tavola dell'informazione può leggersi come una tavola delle decisioni, con l'intento di spiegare le valutazioni dell'attributo decisionale per mezzo delle valutazioni degli attributi condizionali.

In questo caso la tavola delle informazioni può anche essere interpretata come un insieme di regole di decisione. Per esempio con riferimento alla Tabella 1. si ha:

1. se $f(x, A_1) = alta$ e $f(x, A_2) = buona$ e $f(x, A_3) = no$, allora $f(x, A_4) = utile$
(o, in termini più elementari, “se la capacità del personale di vendita è alta e la qualità percepita è buona e la localizzazione non è ad alto traffico, allora il magazzino ha un utile”),
2. se $f(x, A_1) = media$ e $f(x, A_2) = buona$ e $f(x, A_3) = no$, allora
 $f(x, A_4) = perdita$,
3. se se $f(x, A_1) = media$ e $f(x, A_2) = buona$ e $f(x, A_3) = no$, allora
 $f(x, A_4) = utile$,
4. se $f(x, A_1) = bassa$ e $f(x, A_2) = media$ e $f(x, A_3) = si$, allora
 $f(x, A_4) = perdita$,
5. se $f(x, A_1) = media$ e $f(x, A_2) = media$ e $f(x, A_3) = si$, allora
 $f(x, A_4) = perdita$,
6. se se $f(x, A_1) = alta$ e $f(x, A_2) = media$ e $f(x, A_3) = no$, allora
 $f(x, A_4) = utile$.

Questo insieme di regole può allora essere opportunamente ridotto, ottenendo un insieme di regole più concise (nel senso di un minor numero di regole e di un utilizzo di un minor numero di attributi in ogni regola), per esempio:

- I. se $f(x, A_1) = alta$, allora $f(x, A_4) = utile$,
- II. se $f(x, A_1) = bassa$, allora $f(x, A_4) = perdita$,
- III. se $f(x, A_1) = media$ e $f(x, A_2) = media$, allora $f(x, A_4) = perdita$,
- IV. se $f(x, A_1) = media$ e $f(x, A_2) = buona$, allora $f(x, A_4) = utile$ o $perdita$.

Si osservi che le regole I , II e III hanno un conseguente univoco, e perciò esse sono regole esatte, mentre la regola IV non ha un conseguente univoco, e perciò essa è una regola approssimata.

2.6 CONFRONTO CON L'ANALISI STATISTICA

Come già accennato, esistono numerose relazioni tra la teoria dei rough sets e altre teorie matematiche che si propongono di trattare particolari tipi di “incertezza” o di analizzare dati “imperfetti”. Di seguito si riportano sinteticamente alcune brevi considerazioni sinottiche relative al confronto tra la teoria dei Rough Sets e la analisi statistica¹⁷.

Tabella 2.

Problema	Metodi Statistici	Rough Sets
Obiettivi	Identificazione e stima dei parametri delle equazioni strutturali	Riduzione degli attributi ridondanti, generazione di regole di decisione
Rappresentazione dei dati	Tavola a due entrate che rappresentano un campione	Tavola delle informazioni

¹⁷ Stefanosky J. (1992), *Rough Set theory and discriminant methods as tools for analysis of information systems. A comparative study*, Foundation of Computing and Decision Sciences, 17 (2), 81-98.

Tipi di attributi	Attributi quantitativi (almeno nel caso classico)	Attributi qualitativi; gli attributi quantitativi sono trasformati in qualitativi per mezzo di una opportuna discretizzazione
Requisiti dei dati	Il campione deve essere statisticamente significativo; distribuzione multivariata normale	Nessun requisito; possibilità di analizzare anche tavole delle informazioni di ridotte dimensioni
Operatori per l'aggregazione dei dati	Valori medi, matrice delle covarianze, test statistici	Nessun operatore; i dati vengono analizzati nella loro forma originaria
Riduzione dei dati	Selezione di attributi con il maggiore potere discriminante; tipo strumento: test statistici	Sottoinsiemi minimali di attributi che assicurano la stessa qualità di classificazione dell'intero insieme di attributi
Risultati finali	Rappresentazione	Regole di decisione nella

	funzionale	forma di proposizioni logiche
--	------------	----------------------------------

Spesso l'approccio dei rough sets non si pone come alternativo, ma come complementare ad altri approcci basati su teorie o tecniche differenti. Sono state effettuate diverse applicazioni concrete utilizzando differenti approcci; l'uso dei rough sets è risultato molto spesso particolarmente interessante, sia per le notevoli potenzialità applicative dovute alle sue peculiari proprietà (grande "povertà" di informazioni richieste) che per i peculiari risultati ottenuti (regole decisionali, rilevanza degli attributi).

2.7 GENERALIZZAZIONE DELLA RELAZIONE DI INDISCERNIBILITÀ

L'indiscernibilità, come osservato, implica la assoluta impossibilità di distinguere due oggetti che hanno la stessa descrizione in termini degli attributi di Q . Tale relazione induce su U delle classi di equivalenza, che costituiscono i granuli fondamentali delle conoscenza mediante l'indiscernibilità. Spesso, nella realtà, anche per l'imprecisione dei dati che descrivono gli oggetti, piccole differenze non sono considerate significative ai fini della distinzione e degli oggetti corrispondenti. Questa situazione può essere modellizzata formalmente introducendo delle relazioni di similarità o di tolleranza.

In generale, le relazioni di similarità R non generano delle partizioni su U ; le informazioni sulla similarità possono rappresentarsi usando delle classi di similarità per ogni oggetto $x \in U$. Precisamente, la classe di similarità di x , denotata con $R(x)$, è costituita dall'insieme degli oggetti che sono simili ad x :

$$R(x) = \{y \in U : yRx\}$$

È chiaro che un oggetto $z \in R(x)$ può essere simile ad un altro oggetto $y \in U, y \notin R(x)$. La relazione di similarità è ovviamente riflessiva (ogni oggetto è simile a se stesso. Slowinski e Vanderpooten (1997) hanno proposto una relazione di similarità che è solamente riflessiva, rilassando quindi le proprietà di simmetria e transitività. L'abbandono della transitività è facilmente giustificabile, ricordando, ad esempio, il paradosso delle tazzine di caffè di Luce (1956). Per la simmetria, gli autori fanno osservare che, yRx , che significa y (soggetto) è simile ad x (referente), è direzionale ed in generale non è equivalente alla proposizione “ x è simile a y ”. Ciò è abbastanza immediato quando si definisce la relazione di similarità in termini di differenza percentuale rispetto all'oggetto referente. Pertanto, la simmetria della relazione di similarità non deve essere imposta. In tali casi, gli autori ricordati propongono di considerare la relazione inversa di R , denotata R^{-1} , ove $x R^{-1}y$ significa ancora “ y è simile ad x ”; $R^{-1}(x), x \in U$, allora la classe degli oggetti referenti cui x è simile:

$$R^{-1}(X) = \{y \in U : xRy\}$$

Dato un sottoinsieme $X \subseteq U$, un oggetto $x \in U$ è, allora, detto non ambiguo in ciascuno dei due seguenti casi:

- x appartiene a X senza ambiguità, cioè:

$$x \in X \text{ e } R^{-1}(X) \subseteq X;$$

tali oggetti vengono chiamati “positivi”;

- x non appartiene ad X senza ambiguità, cioè

$$x \in U \setminus X \text{ e } R^{-1}(X) \subseteq U \setminus X; (\text{o } R^{-1}(X) \cap X = \emptyset)$$

tali oggetti vengono chiamati “negativi”.

Gli oggetti che non sono né positivi né negativi vengono definiti “ambigui”.

Può allora proporsi una definizione di approssimazione inferiore e superiore più generale. Sia $X \in U$ e R una relazione binaria riflessiva definita su U ; l'approssimazione inferiore di X , denotata con $\underline{R}(X)$ e l'approssimazione superiore di X , denotata con $\overline{R}(X)$, sono per definizione rispettivamente:

$$\underline{R}(X) = \{x \in U : R^{-1}(X) \subseteq X\}$$

$$\overline{R}(X) = \bigcup_{x \in X} R(x).$$

Può dimostrarsi che risulta ancora $\underline{R}(X) \subseteq X \subseteq \overline{R}(X)$ e che:

$$\underline{R}(X) = U - \overline{R}(U - X) \text{ e}$$

$$\overline{R}(X) = \{x \in U : R^{-1}(X) \cap X \neq \emptyset\}.$$

Inoltre, le definizioni proposte sono le uniche che caratterizzano propriamente l'insieme degli oggetti positivi (approssimazione inferiore) e l'insieme degli oggetti positivi o ambigui (approssimazione superiore) quando si usa una relazione di similarità riflessiva, ma non necessariamente simmetrica e transitiva.

3. I ROUGH SETS E LE DECISIONI MULTI ATTRIBUTO

Come accennato in precedenza, una tavola delle decisioni raccoglie tutte le informazioni relative ad un insieme di oggetti, descritti da un certo numero di attributi. Più precisamente, gli attributi condizionali forniscono una descrizione di ogni oggetto in termini di valutazioni su ciascuno di essi; gli attributi decisionali, uno o più, rappresentano uno stato della conoscenza di ciascun oggetto, basata su esperienze pregresse, su opinioni di esperti, su preferenze di decisori, ecc.. La tradizionale analisi di tale tavola mediante i rough sets consiste sostanzialmente nel confrontare le classificazioni degli oggetti di U indotte dagli attributi condizionali di C o di un sottoinsieme $P \subseteq C$, con quella dedotta dagli attributi decisionali D . Tali classificazioni sono, quindi, costruite l'una

indipendentemente dall'altra. Lo strumento che si utilizza per effettuare tali confronti sono le approssimazioni, inferiori e superiori, di ciascuna delle classi decisionali così ottenute, usualmente sulla base della classica relazione di indiscernibilità.

Tradizionalmente l'analisi decisionale condotta vuol dare una risposta alle due seguenti domande: spiegare la decisione in termini delle circostanze in cui essa è presa (analisi retrospettiva); fornire un aiuto al decisore (una prescrizione) su come prendere decisioni future (analisi prospettica). Quest'ultima si basa fondamentalmente sulle regole di decisione ottenute dalla tavola analizzata; la fase della spiegazione, quindi prepara quella della prescrizione, dandole utili informazioni per l'aiuto alle decisioni. Sotto tale aspetto, quindi l'approccio dei rough sets è simile ad un processo di apprendimento induttivo. Ancora, le regole decisionali generate vengono "ottimizzate", sia con riferimento agli attributi effettivamente adoperati (ridotti), consentendo un grande risparmio nella gestione delle informazioni (eliminazione dei dati superflui), che con riferimento alle regole effettivamente utilizzate (generazione di insiemi di regole decisionali minimali), facilitando la comprensione delle stesse da parte del decisore mediante l'eliminazione di regole "ridondanti".

3.1 PROBLEMI DI CLASSIFICAZIONE MULTIATTRIBUTO

I problemi di classificazione multi attributo, consistenti nell'assegnazione di ogni oggetto a delle categorie predefinite, rappresentano l'applicazione più naturale dei rough sets. Infatti, l'insieme degli esempi di classificazione può essere posto facilmente e direttamente nella tavola analizzata. Naturalmente, ogni problema decisionale considerato è suscettibile di molte interpretazioni possibili. Per esempio, gli attributi decisionali possono rappresentare diversi agenti coinvolti in una certa attività, oppure opinioni di uno o più decisori, risultati di studi precedenti o di casi analoghi, ecc. Il modello formale utilizzato, però, non è influenzato da tali differenti interpretazioni e rimane lo stesso per tutti i problemi di classificazione affrontati.

La teoria dei rough sets è stata applicata con successo a numerosi problemi reali di classificazione in differenti campi, quali medicina, farmacologia, ingegneria, gestione del credito, ricerche di mercato, analisi finanziarie, economia ambientale, linguistica, database e altri importanti settori.

3.2 PROBLEMI DI CLASSIFICAZIONE MULTICRITERIALE

Come evidenziato da Greco, Matarazzo e Slowinski (1996) l'approccio classico dei rough sets (CRSA), tuttavia, non considera problemi di classificazione multicriteriale, cioè basati su attributi con domini ordinati (criteri). Tuttavia, in

molti problemi reali è importante considerare proprietà ordinali degli attributi considerati. Per esempio, nelle valutazioni del rischio di fallimento, se l'indice di indebitamento (debiti totali/attività totali) dell'azienda A ha un valore modesto mentre lo stesso indice dell'azienda B ha un valore rilevante, all'interno dell'approccio dei rough sets le due aziende sono discernibili, ma nessuna preferenza è stabilita tra di esse con riferimento all'attributo "rapporto di indebitamento". Invece, dal punto di vista della valutazione del rischio di fallimento delle due aziende, sarebbe meglio considerare l'azienda A migliore dell'azienda B, e non semplicemente discernibile, con riferimento all'attributo in questione.

Sulla base di queste considerazioni, Greco Matarazzo e Slowinski (1997) hanno proposto un nuovo approccio dei rough sets per problemi di classificazione multicriteriale. Così come nell'analisi CRSA, l'approccio proposto è basato su approssimazioni di una partizione degli oggetti dell'universo in alcune classi predefinite sulla base della tavola delle informazioni. Tuttavia, a differenza dell'approccio originario dei rough sets, le approssimazioni sono costruite usando relazioni di dominanza invece che di indiscernibilità. Questo permette di prendere esplicitamente in considerazione le proprietà ordinali degli attributi (criteri) considerati.

4.DOMINANCE-BASED ROUGH SETS APPROACH (DRSA)

Il DRSA è un'evoluzione della teoria classica dei Rough Set (Classical Rough Set Approach - CRSA) che consente di applicare tale teoria a problemi di scelta multicriteriale.

Assunto dunque che tutti gli attributi condizionali siano dei criteri, sia \succeq_q la relazione di preferenza debole su U riferita al criterio $q \in Q$, dove $x \succeq_q y$ ha il significato “ x è almeno tanto buono quanto y rispetto al criterio q ”. Ciò presuppone che \succeq_q sia preordine completo, ovvero una relazione binaria riflessiva e transitiva, definita in U sulla base della valutazione $f(\cdot, q)$. L'insieme di attributi decisionali D (eventualmente un singoletto $\{d\}$) genera una partizione di U in un numero finito di classi, sia $Cl = \{Cl_t, t \in T\}$, $T = \{1, \dots, n\}$ con n numero delle classi, una classificazione di U , tale che ogni $x \in U$ appartiene ad un'unica classe $Cl_t \in Cl$.

Si suppone che le classi siano ordinate, ossia che per tutti gli $r, s \in T$ tali che $r > s$, allora gli oggetti di Cl_r saranno preferiti agli oggetti di Cl_s . Più formalmente se \succeq è in una relazione di preferenza debole su U , ovvero se per ogni $x, y \in U$, $x \succeq y$ allora “ x è almeno tanto buono quanto y ”:

$$[x \in Cl_r, y \in Cl_s, r > s] \Rightarrow [x \succeq y \text{ e non } y \succeq x]$$

Tale relazione di preferenza tra le classi di Cl costituisce la base concettuale dei problemi di classificazione multi-criteriale (multiple criteria sorting problem) (Greco et al., 2002a).

4.1 APPROSSIMAZIONI BASATE SULLA DOMINANZA

La ripartizione dell'universo in classi permette di definire nel rispetto della relazione di dominanza, delle unioni di classi, chiamate unioni ascendenti e unioni discendenti delle classi così definite:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$$

$$Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$$

con $t = \{1, 2, \dots, n\}$.

L'espressione $x \in Cl_t^{\geq}$ significa che “ x appartiene almeno alla classe Cl_t ”, mentre $x \in Cl_t^{\leq}$ significa che “ x appartiene al massimo alla classe Cl_t ”, $Cl_t \in Cl$. È da notare che $Cl_1^{\geq} = Cl_n^{\leq} = U$ e che $Cl_n^{\leq} = Cl_1$. Inoltre per $t = 2, \dots, n$ si ha: $Cl_t^{\geq} = U - Cl_{t-1}^{\leq}$.

L'idea chiave dei Rough Set è l'approssimazione della conoscenza espressa in termini di attributi decisionali desunta da una conoscenza espressa in termini di attributi condizionali. Nella teoria dei Dominance- based Rough Set Approach, (DRSA), nel momento in cui gli attributi condizionali sono criteri (ossia ordinati in funzione dell'attributo decisionale considerato) e le classi hanno un ordine di preferenza, la conoscenza approssimata è data da un insieme di unioni di classi (inferiori e superiori) e i “granuli di conoscenza” usati per l'approssimazione sono costituiti da insiemi di oggetti definiti utilizzando la relazione di dominanza (base del DRSA) invece che la relazione di indiscernibilità (tipica del CRSA).

Si dice che x domina y , cioè $x D_P y$ rispetto a $P \subseteq C$, se $x \succeq_q y$ per ogni $q \in P$.

Allora, considerato $P \subseteq C$, si definiscono per ogni $x \in U$ i “granuli di conoscenza” usati per le approssimazioni nel DRSA sono:

- l'insieme degli oggetti che dominano x , chiamati P-Dominanti (P-dominating set), $D_P^+(x) = \{y \in U : y D_P x\}$;
- l'insieme degli oggetti dominati da x , chiamati P-Dominati (P-dominated set), $D_P^-(x) = \{y \in U : x D_P y\}$.

Per ogni $P \subseteq C$ si dice che $x \in U$ appartiene alla classe Cl_t^{\geq} senza alcuna ambiguità se $x \in Cl_t^{\geq}$ e, per ogni oggetto $y \in U$ dominato da x rispetto a P , si ha $y \in Cl_t^{\geq}$, ossia $D_P^+(x) \subseteq Cl_t^{\geq}$. In altre parole, un'ambiguità relativa ad ogni oggetto x rispetto al criterio P riguarda il caso in cui ci sia almeno un altro oggetto che “non è peggiore di x ” per tutti i criteri considerati in P e tuttavia

assegnati ad una classe “peggiore”. Si dice che $y \in U$ può appartenere alla classe Cl_t^{\geq} con eventualmente qualche ambiguità se esiste almeno un oggetto $x \in Cl_t^{\geq}$ tale che y domini x rispetto all’insieme $P \subseteq C$, ovvero $y \in D_P^+(x)$. È da notare che dire che $y \in U$ potrebbe appartenere a Cl_t^{\geq} non necessariamente significa che vi appartenga.

Dunque rispetto a $P \subseteq C$, l’insieme di tutti gli oggetti appartenenti alla classe Cl_t^{\geq} senza alcuna ambiguità costituisce l’approssimazione inferiore di P in Cl_t^{\geq} (P-lower approximation of Cl_t^{\geq}), denotata con $\underline{P}(Cl_t^{\geq})$, mentre l’insieme di tutti gli oggetti che potrebbero appartenere alla classe Cl_t^{\geq} , eventualmente con qualche ambiguità, costituisce l’approssimazione superiore di P in Cl_t^{\geq} (P-upper approximation of Cl_t^{\geq}), denotata con $\overline{P}(Cl_t^{\geq})$:

$$\underline{P}(Cl_t^{\geq}) = \{x \in U : D_P^+(x) \subseteq Cl_t^{\geq}\}$$

$$\overline{P}(Cl_t^{\geq}) = \bigcup_{x \in Cl_t^{\geq}} D_P^+(x), t=1, 2, \dots, n.$$

Si osservi che $\underline{P}(Cl_t^{\geq}) \subseteq \overline{P}(Cl_t^{\geq})$ per ogni $P \subseteq C$ e per ogni $t=1, 2, \dots, n$.

Le frontiere rispetto a P di Cl_t^{\geq} e Cl_t^{\leq} (P-boundaries o P-doubtful regions) sono definite come:

$$Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq})$$

$$Bn_p(CI_t^{\leq}) = \overline{P}(CI_t^{\leq}) - \underline{P}(CI_t^{\leq})$$

per $t=1, 2, \dots, n$.

Più semplicemente la frontiera $Bn_p(CI_t^{\geq})$ è composta da tutti gli oggetti ambigui rispetto al set di criteri P e dall'unione superiore delle classi CI_t^{\geq} . Analogamente, la frontiera $Bn_p(CI_t^{\leq})$ è composta da tutti gli oggetti ambigui rispetto al set di criteri P e dall'unione inferiore delle classi CI_t^{\leq} . A causa della complementarità dell'approssimazione di tipo rough (Slowinski et al., 2005) seguono le seguenti proprietà:

$$Bn_p(CI_t^{\geq}) = Bn_p(CI_{t-1}^{\leq}) \text{ per } t=2, \dots, n \text{ e}$$

$$Bn_p(CI_t^{\leq}) = Bn_p(CI_{t+1}^{\geq}) \text{ per } t=1, 2, \dots, n.$$

4.2 QUALITÀ DELL'APPROSSIMAZIONE ED INSIEMI RIDOTTI

Come nel caso dell'approccio classico, si possono definire i parametri relativi alla qualità delle approssimazioni.

E' possibile definire per ogni $t \in T$ la qualità dell'approssimazione (quality of sorting) della classificazione Cl , in funzione del set di criteri $P \subseteq C$ come:

$$\gamma_P(Cl) = \frac{\text{card}\left(U - \left(\bigcup_{t \in T} Bn_P(Cl_t^{\geq})\right)\right)}{\text{card}U} = \frac{\text{card}\left(U - \left(\bigcup_{t \in T} Bn_P(Cl_t^{\leq})\right)\right)}{\text{card}U}$$

La qualità dell'approssimazione $\gamma_P(Cl)$ è il rapporto fra gli oggetti correttamente rappresentati tramite gli attributi di P e il numero di oggetti dell'universo, cioè la percentuale di oggetti per cui non c'è ambiguità. Bisogna osservare che ampliando l'insieme dei criteri considerati, la qualità dell'approssimazione non può decrescere, ma in generale può crescere. Infatti, prendendo in considerazione dei nuovi criteri, oggetti che erano ambigui possono diventare non ambigui.

Ogni sottoinsieme minimo di criteri $P \subseteq C$ tale che $\gamma_P(Cl) = \gamma_C(Cl)$ è definito ridotto di Cl (reduct) ed è denotato come RED_{Cl} . Un ridotto di P è un sottoinsieme minimo di criteri di C tale che oggetti ambigui possono diventare non ambigui se si considerano altri criteri; ciò significa che se P è un ridotto, gli oggetti ambigui rispetto a P lo saranno anche rispetto a C, e se qualche criterio non viene considerato allora almeno un oggetto diverrà ambiguo.

È da specificare che una tabella delle informazioni può avere più di un ridotto. L'intersezione di tutti i ridotti è detto core ed è denotato come $CORE_{Cl}$. Il core contiene tutti i criteri che non possono essere rimossi senza dare luogo a delle ambiguità che non sono presenti considerando tutti i criteri di C. Il risultato finale

di un'analisi effettuata con il DRSA è un insieme di regole decisionali, espresse in termini di proposizioni del tipo “se..., allora...” chiamate regole di decisione.

Le regole di decisione generate dall'approccio dei rough set non derivano direttamente dalla decision table, ma dalle approssimazioni inferiore e superiore delle unioni ascendenti e discendenti delle classi decisionali.

Per una data unione ascendente Cl_t^{\geq} o discendente Cl_s^{\leq} , le regole estratte nell'ipotesi che gli oggetti appartenenti a $\underline{P}(Cl_t^{\geq})$ o a $\overline{P}(Cl_s^{\leq})$ siano positivi e tutti gli altri negativi, suggeriscono, rispettivamente, una formulazione delle stesse del tipo “...allora x appartiene almeno alla classe Cl_t ” o “...allora x appartiene al massimo alla classe Cl_s ”. Le regole estratte invece nell'ipotesi che gli oggetti appartenenti all'intersezione delle approssimazioni superiori dell'unione delle classi $\underline{P}(Cl_s^{\leq}) \cap \underline{P}(Cl_t^{\geq})$ siano tutti positivi e i rimanenti negativi, suggeriscono invece che l'oggetto potrebbe appartenere ad una delle classi comprese fra Cl_t e Cl_s ($s < t$).

Più formalmente, assumendo per ogni $q \in C$, $V_q \subseteq R$ (con V_q quantitativo) e per ogni $x, y \in U$ che $f(x, q) \geq f(y, q)$ implica $x \succeq_q y$ (con V_q ordinato per preferenza), si possono considerare le tre seguenti tipologie di regole decisionali:

1. Regole decisionali D_{\geq} : forniscono indicazioni relative a limiti inferiori dei criteri per oggetti che appartengono all'unione ascendente delle classi Cl_t^{\geq} e assumono la forma generale:

se $f(x, q_1) \succeq r_{q_1}$ e $f(x, q_2) \succeq r_{q_2}$ e ... $f(x, q_p) \succeq r_{q_p}$ allora $x \in Cl_t^{\succeq}$

con $P = \{q_1, q_2, \dots, q_p\} \subseteq C$, $r_{q_1}, r_{q_2}, \dots, r_{q_p} \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}$ e $t \in T$;

2. Regole decisionali D_{\leq} : forniscono indicazioni relative a limiti superiori dei criteri per oggetti che appartengono all'unione discendente delle classi Cl_t^{\leq} e assumono la forma generale

se $f(x, q_1) \preceq r_{q_1}$ e $f(x, q_2) \preceq r_{q_2}$ e ... $f(x, q_p) \preceq r_{q_p}$ allora $x \in Cl_t^{\leq}$

Con $P = \{q_1, q_2, \dots, q_p\} \subseteq C$, $r_{q_1}, r_{q_2}, \dots, r_{q_p} \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}$ e $t \in T$;

3. Regole decisionali D_{\succeq} : forniscono indicazioni relative simultaneamente a limiti superiori e inferiori dei criteri per oggetti che appartengono all'unione di classi intermedie classi $Cl_s \cup Cl_{s+1} \cup \dots \cup Cl_{t-1} \cup Cl_t$ senza possibilità di distinguere quale, e assumono la forma generale:

se $f(x, q_1) \succeq r_{q_1}$ e $f(x, q_2) \succeq r_{q_2}$ e ... $f(x, q_k) \succeq r_{q_k}$ e $f(x, q_{k+1}) \succeq r_{q_{k+1}}$

e $f(x, q_p) \succeq r_{q_p}$ allora $x \in Cl_s \cup Cl_{s+1} \cup \dots \cup Cl_{t-1} \cup Cl_t$

con $O' = \{q_1, q_2, \dots, q_k\} \subseteq C$, $O'' = \{q_{k+1}, q_{k+2}, \dots, q_p\} \subseteq C$, $P = O' \cup O''$, e O' e O'' non necessariamente disgiunti, $r_{q_1}, r_{q_2}, \dots, r_{q_p} \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}$ e $s, t \in T, s < t$.

È possibile che $\{q_1, q_2, \dots, q_k\} \cap \{q_{k+1}, \dots, q_p\} \neq \emptyset$ nella parte condizionale di una regola decisionale D_{\leq} e che si possa avere $f(x, q) \succ r_q$ e $f(x, q) \succeq r'_q$ con $r_q \leq r'_q$ per $q \in C$. Inoltre se $r_q = r'_q$, le due condizioni divengono $f(x, q) = r_q$ e si avrebbe una situazione di indifferenza.

Un insieme di regole si definisce completo se permette di classificare tutti gli oggetti della decision table, e permette di ri-classificarli correttamente se non formano alcuna ambiguità con altri oggetti, oppure in insiemi di classi, comprendente quella corretta, se formano qualche ambiguità con altri oggetti. Un insieme di regole è minimo se è completo e non ridondante, ovvero se l'esclusione di una regola lo rende non completo.

4.3 PRINCIPALI VANTAGGI DELLA CLASSIFICAZIONE MULTICRITERIALE

Come accennato, l'analisi dei rough sets basata sulle approssimazioni mediante relazioni binarie di dominanza migliora, in generale, i risultati dei problemi di classificazione rispetto all'approccio classico basato sull'uso della relazione di

indiscernibilità. Nei problemi di classificazione, i vantaggi dell'approccio basato su relazioni di dominanza rispetto all'analisi dei CRSA, basata sulla relazione di indiscernibilità, possono sintetizzarsi come segue:

I. Si ottiene spesso un minor numero di ridotti ed un nucleo più grande.

Queste due caratteristiche sono generalmente riconosciute come delle proprietà desiderabili di una buona approssimazione.

II. La qualità dell'approssimazione ottenuta usando le relazioni di dominanza può essere inferiore a quella ottenuta approssimando con relazioni d'indiscernibilità. Ma questo apparente inconveniente mostra, in verità, un altro notevole vantaggio dell'approccio considerato. Infatti, l'approccio mediante approssimazioni basate su relazioni di dominanza mette spesso in luce delle inconsistenze nei risultati della classificazione, che non possono essere colte dall'approssimazione tramite indiscernibilità. Quest'ultima, infatti, classifica gli oggetti dell'universo distinguendoli solamente in conseguenza di loro descrizioni differenti in termini degli attributi/criteri considerati, ma non coglie assolutamente aspetti ordinali dei dati. Può pertanto accadere che due oggetti x e y siano classificati (da un esperto in esperienze passate, ecc.) in maniera tale che la valutazione globale di x sia peggiore di quella di y , mentre dalla valutazione degli stessi, evidenziata nella corrispondente tavola delle decisioni, risulta che x domina y . L'approccio considerato, è solamente esso, consente di

evidenziare questa inconsistenza, spiegando quindi anche la ragione dell'apparente peggioramento della qualità della classificazione.

III. Migliora la qualità dell'insieme delle regole decisionali ottenute dalle approssimazioni mediante relazioni di dominanza, che forniscono in generale una rappresentazione più sintetica della conoscenza contenuta nella tavola delle informazioni. Gli insiemi minimali di regole così ottenute hanno un minor numero di regole ed usano un minor numero di attributi e descrittori rispetto all'algoritmo di classificazione basato sulla classica analisi dei rough sets. Inoltre, l'applicazione di tali regole a nuovi oggetti da classificare fornisce in generale risultati migliori, talvolta, infatti, utilizzando l'algoritmo originario non si è in grado di classificare qualche nuovo oggetto.

CAPITOLO 4

UN MODELLO DI SCORING BASATO SULL'APPROCCIO DEI ROUGH SETS

1. INTRODUZIONE

Questo capitolo conclusivo ha come scopo quello di illustrare le caratteristiche e le potenzialità di un modello di scoring basato sull'approccio dei rough sets. Nel primo paragrafo saranno illustrati i principi generali sui quali si basa questo approccio multicriteriale alla valutazione del merito creditizio, problematica che è stata affrontata in modo approfondito da Greco, Matarazzo e Slowinski (1998). Nei paragrafi a seguire sarà analizzato un caso concreto applicato ad un campione di imprese fornito da un primario istituto di credito italiano.

2. LA METODOLOGIA

Vari metodi sono stati proposti nella letteratura specializzata per la valutazione del rischio di fallimento. Sulla scorta di Dimitras, Zanakis e Zopundis (1996) si ricordano i seguenti metodi: metodi statistici univariati, metodi della

“sopravvivenza”, analisi discriminante, modelli lineari di probabilità, analisi logit e probit, algoritmi di partizionamento recursivo, programmazione matematica, metodi multicriteriali di supporto alla decisione, sistemi esperti.

Un nuovo metodo per la valutazione del rischio di fallimento basato sull’approccio dei rough sets è stato introdotto nel 1995 da Slowinski e Zopounidis. Il concetto di rough sets introdotto da Pawlak (1982) si è mostrato uno strumento efficace per l’analisi di una tavola delle informazioni (tavola delle informazioni finanziarie) che descrive un insieme di oggetti (aziende) per mezzo di un insieme di attributi (indicatori finanziari e variabili qualitative).

Si ricorda che come evidenziato nel capitolo precedente, l’approccio classico dei rough sets (CRSA) non considera attributi ordinati (criteri). Cionondimeno, in molti problemi reali è importante considerare le proprietà ordinali degli attributi considerati. Questo problema ha rilevanza anche nelle valutazioni del rischio di fallimento. Per esempio, se il rapporto di indebitamento (debiti totali/totale fonti finanziarie) dell’azienda A ha un valore modesto mentre lo stesso indice dell’azienda B ha un valore rilevante, all’interno dell’approccio dei rough sets le due aziende sono discernibili, ma nessuna preferenza è stabilita tra di esse con riferimento all’attributo “rapporto di indebitamento”. Invece, dal punto di vista della valutazione del rischio di fallimento delle due aziende, sarebbe meglio considerare l’azienda A migliore dell’azienda B, e non semplicemente discernibile, con riferimento all’attributo in questione.

Pertanto, a seguito di tali considerazioni, Greco, Matarazzo e Slowinski nel 1998 hanno proposto un nuovo approccio basato sui rough sets per la valutazione del rischio di fallimento utilizzando la relazione di dominanza (DRSA), in luogo della relazione di indiscernibilità, usata nel CRSA.

Sono stati analizzati i dati relativi agli affidamenti di una banca d'affari greca, ETEVA, che finanzia aziende industriali e commerciali in Grecia. È stato selezionato un campione di 39 aziende. Con la collaborazione dei dirigenti finanziari dell'ETEVA, le aziende furono classificate in tre classi predefinite di rischio per l'anno 1988. Il risultato della classificazione è rappresentato dall'attributo decisionale d , il quale opera una tripartizione dell'insieme delle aziende considerate:

- $d=A$ significa “azienda accettabile”;
- $d=U$ significa “azienda incerta”;
- $d=NA$ significa “azienda non accettabile”.

La partizione è determinata da $Cl = \{Cl_A, Cl_U, Cl_{NA}\}$ e chiaramente la classe Cl_A è migliore della classe Cl_U che è migliore di Cl_{NA} . Le aziende sono state valutate in base ai seguenti 12 attributi:

- A_1 = EBIT/totale attività;
- A_2 = utile netto/capitale netto;
- A_3 = debiti totali/totale fonti finanziarie;
- A_4 = debiti totali/cash flow;
- A_5 = interessi passivi/vendite;

- A₆= spese generali ed amministrative/vendite;
- A₇= esperienza del management;
- A₈= posizione di mercato dell'azienda;
- A₉=strutture tecniche;
- A₁₀=organizzazione del personale;
- A₁₁=specifici vantaggi competitivi dell'azienda;
- A₁₂=flessibilità al mercato.

I primi sei attributi sono quantitativi (indici di bilancio) e gli ultimi sei sono qualitativi. I sei attributi qualitativi sono stati modellati secondo una scala ordinale (5 meglio di 4, 4 meglio di 3 e così via). Per gli attributi A₁, A₂ e per gli attributi da A₇ a A₁₂ la preferenza cresce all'aumentare del loro valore, mentre per gli attributi da A₃ a A₆ la preferenza decresce all'aumentare del loro valore.

Quindi l'analisi dei rough sets è stata condotta sulla tavola delle informazioni presentata di seguito:

Tabella 1. Tavola delle informazioni finanziarie													
Azienda	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	d
F1	16,4	14,5	59,82	2,5	7,5	5,2	5	3	5	4	2	4	A
F2	35,8	67	64,92	1,7	2,1	4,5	5	4	5	5	4	5	A
F3	20,6	61,75	75,71	3,6	3,6	8	5	3	5	5	3	5	A
F4	11,5	17,1	57,1	3,8	4,2	3,7	5	2	5	4	3	4	A
F5	22,4	25,1	49,8	2,1	5	7,9	5	3	5	5	3	5	A
F6	23,9	34,5	48,9	1,7	2,5	8	5	3	4	4	3	4	A
F7	29,9	44	57,8	1,8	1,7	2,5	5	4	4	5	3	5	A
F8	8,7	5,4	27,4	3,3	4,5	4,5	5	2	4	4	1	4	A
F9	25,7	29,7	46,8	1,7	4,6	3,7	4	2	4	3	1	3	A
F10	21,2	24,6	64,8	4,4	3,6	8	4	2	4	4	1	4	A
F11	18,32	31,6	69,3	0,7	2,8	3	4	3	4	4	3	4	A
F12	20,7	19,3	19	4,5	2,2	4	4	2	4	4	1	3	A
F13	9,9	3,5	53,1	9,4	8,5	5,3	4	2	4	4	1	4	A
F14	10,4	9,3	80,9	3,2	1,4	4,1	4	2	4	4	3	3	A
F15	17,7	19,8	52,8	1,3	7,9	6,1	4	4	4	4	2	5	A
F16	14,8	15,9	27,94	3,9	5,4	1,8	4	2	4	3	2	3	A
F17	16	14,7	53,5	3,9	6,8	3,8	4	4	4	4	2	4	A
F18	11,7	10,01	42,1	5,8	12,2	4,3	5	2	4	2	1	3	A
F19	11	4,2	60,8	6,5	6,2	4,8	4	2	4	4	2	4	A
F20	15,5	8,5	56,2	5,5	5,5	1,8	4	2	4	4	2	4	A
F21	13,2	9,1	74,1	11,21	6,4	5	2	2	4	4	2	3	U
F22	9,1	4,1	44,8	4,2	3,3	10,4	3	4	4	4	3	4	U
F23	12,9	1,9	65,02	6,9	14,01	7,5	4	3	2	2	1	2	U
F24	5,9	-27,7	77,4	-32,2	16,6	12,7	3	2	4	4	2	3	U
F25	16,9	12,1	60,1	5,2	5,6	5,6	3	2	4	4	2	3	U
F26	16,7	13,1	73,5	7,1	11,9	4,1	2	2	4	4	2	3	U
F27	14,6	9,7	59,05	5,8	6,7	5,6	2	2	4	4	2	4	U
F28	5,1	4,9	28,9	4,3	2,5	46	2	2	3	3	1	2	U
F29	24,4	22,3	32,8	1,4	3,3	5	2	3	4	4	2	3	U
F30	29,7	8,6	41,8	1,6	5,2	6,4	2	3	4	4	2	3	U
F31	7,3	-64,5	67,5	-2,2	30,1	8,7	3	3	4	4	2	3	NA
F32	23,7	31,9	63,6	3,5	12,1	10,2	3	2	4	4	1	3	NA
F33	18,9	13,5	74,5	10	12	8,4	3	3	4	4	3	4	NA
F34	13,9	3,3	78,7	25,5	14,7	10,1	2	2	4	4	3	4	NA
F35	-13,3	-31,1	63	-10	21,2	23,1	2	1	3	3	1	2	NA
F36	6,2	-3,2	46,1	5,1	4,8	10,5	2	1	3	3	2	3	NA
F37	4,8	-3,3	71,9	34,6	8,6	11,6	2	2	4	4	2	3	NA
F38	0,1	-9,6	42,5	-20	12,9	12,4	1	1	3	3	1	3	NA
F39	13,6	9,1	76	11,4	17,1	10,3	1	1	1	1	1	2	NA

Le principali domande a cui ha dovuto rispondere il processo di analisi sono le seguenti:

- le informazioni finanziarie contenute nella “tabella 1” sono consistenti?
- quali sono i ridotti degli attributi condizionali che assicurano la stessa qualità di approssimazione dell'intero insieme degli attributi condizionali rispetto agli attributi decisionali?
- quali sono le regole decisionali che possono essere ottenute dalla “tabella 1”?
- quali sono gli insiemi minimali di regole decisionali che coprono tutte le aziende della “tabella 1”?

Il primo risultato dell'approccio DRSA è stato la scoperta che la tavola delle informazioni finanziarie è consistente, ovvero l'accuratezza di tutte le approssimazioni è perfetta, cioè è uguale a 1.

La seconda scoperta è stata un insieme di 18 ridotti di attributi condizionali (criteri) che assicurano la stessa qualità di classificazione dell'intero insieme dei 12 attributi condizionali, di seguito si riportano i ridotti trovati:

$$RED_{CL}^1 = \{A_1, A_4, A_5, A_7\}; \quad RED_{CL}^2 = \{A_2, A_4, A_5, A_7\}; \quad RED_{CL}^3 = \{A_3, A_4, A_6, A_7\};$$

$$RED_{CL}^4 = \{A_4, A_5, A_6, A_7\}; \quad RED_{CL}^5 = \{A_4, A_5, A_7, A_8\}; \quad RED_{CL}^6 = \{A_2, A_3, A_7, A_9\};$$

$$RED_{CL}^7 = \{A_1, A_3, A_4, A_7, A_9\}; \quad RED_{CL}^8 = \{A_1, A_5, A_7, A_9\}; \quad RED_{CL}^9 = \{A_2, A_5, A_7, A_9\};$$

$$RED_{CL}^{10} = \{A_4, A_5, A_7, A_9\}; \quad RED_{CL}^{11} = \{A_5, A_6, A_7, A_9\}; \quad RED_{CL}^{12} = \{A_4, A_5, A_7, A_{10}\};$$

$$RED_{CL}^{13} = \{A_1, A_3, A_4, A_7, A_{11}\}; \quad RED_{CL}^{14} = \{A_2, A_3, A_4, A_7, A_{11}\}; \quad RED_{CL}^{15} = \{A_4, A_5, A_6, A_{12}\}$$

$$; RED_{CL}^{16} = \{A_1, A_3, A_5, A_6, A_9, A_{12}\}; \quad RED_{CL}^{17} = \{A_3, A_4, A_6, A_{11}, A_{12}\};$$

$$RED_{CL}^{18} = \{A_1, A_2, A_3, A_6, A_{11}, A_{12}\}$$

Il core è risultato vuoto, ciò significa che non sono stati trovati attributi indispensabili per l'approssimazione. Per selezionare un insieme ridotto su cui calcolare le regole di decisione è stata adottata la seguente procedura (Slowinski K., Slowinski R. e Stefanosky 1988). Un singolo attributo caratterizzato dalla più alta qualità di classificazione è stato aumentato di uno dei rimanenti attributi e la coppia che dava la più alta qualità di approssimazione è stata scelta. Quindi alla coppia selezionata si è aggiunto un altro attributo e la terna che dava la più alta qualità di classificazione è stata scelta, e così via finché la qualità è risultata essere pari a 1.

Tabella 2. Procedura di selezione dei migliori ridotti												
Attributo	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Qualità	0,128	0,154	0,077	0,077	0,18	0,25	0,28	0,103	0,1154	0,128	0,26	0,128
A7+	A1	A2	A3	A4	A5	A6	A8	A9	A10	A11	A12	
Qualità	0,59	0,718	0,667	0,615	0,77	0,692	0,41	0,59	0,59	0,462	0,59	
A5,A7+	A1	A2	A3	A4	A6	A8	A9	A10	A11	A12		
Qualità	0,923	0,923	0,795	0,897	0,923	0,872	0,95	0,846	0,795	0,795		
A5,A7,A9+	A1	A2	A3	A4	A6	A8	A10	A11	A12			
Qualità	I	I	0,945	I	I	0,949	0,949	0,949	0,949			

Pertanto, i migliori ridotti sono i seguenti: $RED_{CL}^8 = \{A_1, A_5, A_7, A_9\};$

$$RED_{CL}^9 = \{A_2, A_5, A_7, A_9\}; \quad RED_{CL}^{10} = \{A_4, A_5, A_7, A_9\}; \quad RED_{CL}^{11} = \{A_5, A_6, A_7, A_9\}.$$

La terza scoperta è stata l'insieme di tutte le regole decisionali. Sono state ottenute 74 regole che descrivono Cl_1^{\leq} (imprese inaccettabili) , 51 regole che descrivono Cl_2^{\leq} (imprese inaccettabili ed incerte), 75 regole che descrivono Cl_2^{\geq} (imprese inaccettabili e accettabili) e 79 regole che descrivono Cl_3^{\geq} (imprese accettabili). Di seguito saranno elencate le tre regole più forti per ciascuna unione considerata (per ciascuna regola saranno indicati tra parentesi i codici delle imprese che supportano la corrispondente regole; ovviamente più è alto il numero di imprese più alta è la forza della regola).

1) se $f(x, A_2) \leq 3,3$ e $f(x, A_7) \leq 2$, allora $x \in Cl_1^{\leq}$;

(F34, F35, F36, F37, F38);

2) se $f(x, A_4) \geq 10$ e $f(x, A_6) \leq 8,4$, allora $x \in Cl_1^{\leq}$;

(F33, F34, F37, F39);

3) se $f(x, A_3) \geq 67,5$ e $f(x, A_4) \geq -2,2$ e $f(x, A_6) \geq 8,7$ allora $x \in Cl_1^{\leq}$;

(F31, F34, F37, F39);

4) se $f(x, A_7) \leq 3$ allora $x \in Cl_2^{\leq}$;

(F21, F22, F24, F25, F26, F27, F28, F29, F30, F31, F32, F33, F34, F35, F36, F37, F38, F39);

5) se $f(x, A_2) \leq 12,4$ e $f(x, A_6) \geq 5,6$ allora $x \in Cl_2^{\leq}$;

(F22, F23, F24, F25, F27, F28, F30, F31, F34, F35, F36, F37, F38, F39);

6) se $f(x, A_6) \geq 5$ e $f(x, A_{12}) \leq 3$ allora $x \in Cl_2^{\leq}$;

(F21, F23, F24, F25, F28, F29, F30, F31, F32, F35, F36, F37, F38, F39);

- 7) se $f(x, A_1) \geq 5,9$ e $f(x, A_2) \leq -27,7$ e $f(x, A_9) \geq 4$ allora $x \in Cl_2^{\geq}$;
 (F1, F2 ,F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16,
 F17, F18, F19, F20, F21, F22, F24, F25, F26, F27, F29, F30);
- 8) se $f(x, A_1) \geq 5,9$ e $f(x, A_4) \leq 16,6$ e $f(x, A_9) \geq 4$ allora $x \in Cl_2^{\geq}$;
 (F1, F2 ,F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16,
 F17, F18, F19, F20, F21, F22, F24, F25, F26, F27, F29, F30);
- 9) se $f(x, A_6) \leq 8$ allora $x \in Cl_2^{\geq}$;
 (F1, F2 ,F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16,
 F17, F18, F19, F20, F21, F23, F25, F26, F27, F29, F30);
- 10) se $f(x, A_2) \geq 3,5$ e $f(x, A_7) \geq 4$ allora $x \in Cl_3^{\geq}$;
 (F1, F2 ,F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16,
 F17, F18, F19, F20);
- 11) se $f(x, A_7) \geq 4$ e $f(x, A_9) \geq 4$ allora $x \in Cl_3^{\geq}$;
 (F1, F2 ,F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16,
 F17, F18, F19, F20);
- 12) se $f(x, A_7) \geq 4$ e $f(x, A_{12}) \geq 3$ allora $x \in Cl_3^{\geq}$;
 (F1, F2 ,F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16,
 F17, F18, F19, F20).

Le regole decisionali sopra indicate rappresentano bene le relazioni tra attributi condizionali e attributi decisionali. Pertanto, i decisori possono scoprire, grazie a

queste regole, quali sono gli aspetti più importanti da prendere in considerazione nelle politiche di classificazione del merito creditizio.

La quarta scoperta è stata l'estrazione di un insieme minimale di regole decisionali. Diversi insiemi minimali sono stati trovati, uno di questi viene mostrato di seguito:

- 1) se $f(x, A_3) \geq 67,5$ e $f(x, A_4) \geq -2,2$ e $f(x, A_6) \geq 8,7$ allora $x \in Cl_1^{\leq}$;
(F31, F34, F37, F39);
- 2) se $f(x, A_2) \leq 3,3$ e $f(x, A_7) \leq 2$, allora $x \in Cl_1^{\leq}$;
(F34, F35, F36, F37, F38);
- 3) se $f(x, A_3) \geq 63,6$ e $f(x, A_7) \leq 3$ e $f(x, A_9) \leq 3$, allora $x \in Cl_1^{\leq}$;
(F34, F35, F36, F37, F38);
- 4) se $f(x, A_2) \leq 12,4$ e $f(x, A_6) \geq 5,6$ allora $x \in Cl_2^{\leq}$;
(F22, F23, F24, F25, F27, F28, F30, F31, F34, F35, F36, F37, F38, F39);
- 5) se $f(x, A_7) \leq 3$ allora $x \in Cl_2^{\leq}$;
(F21, F22, F24, F25, F26, F27, F28, F29, F30, F31, F32, F33, F34, F35, F36, F37, F38, F39);
- 6) se $f(x, A_2) \geq 3,5$ e $f(x, A_5) \leq 8,5$ allora $x \in Cl_2^{\geq}$;
(F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16, F17, F19, F20, F21, F22, F25, F27, F28, F29, F30);
- 7) se $f(x, A_7) \geq 4$ allora $x \in Cl_2^{\geq}$;

(F1, F2 ,F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16,
F17, ,E18, F19, F20,F23);

8) se $f(x, A_1) \geq 8,7$ e $f(x, A_9) \geq 4$ allora $x \in Cl_2^{\geq}$;

(F1, F2 ,F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16,
F17, ,E18, F19, F20,F21, F22, F23, F25, F26, F27, F29, F30);

9) se $f(x, A_2) \geq 3,5$ e $f(x, A_7) \geq 4$ allora $x \in Cl_3^{\geq}$;

(F1, F2 ,F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15, F16,
F17, F18, F19, F20);

L'insieme di regole è completo e composto da solo regole decisionali D_{\geq} e regole decisionali D_{\leq} . L'applicazione di queste regole alle 39 aziende consentirà il loro esatto assegnamento alle corrispondenti classi di rischio. Gli insiemi minimali di regole decisionali rappresentano la rappresentazione più compatta e non ridondante delle informazioni presenti nella tavola delle informazioni. L'insieme minimale di 9 regole decisionali visto sopra usa 8 attributi e 18 descrittori, circa il 3,85% dei descrittori presenti nella tavola delle informazioni.

3. UN'APPLICAZIONE AD UN CASO CONCRETO

Seguendo la metodologia introdotta nel paragrafo precedente è stato analizzato un campione di imprese fornito da un primario istituto bancario italiano. Lo scopo dell'analisi è stato quello di valutare le caratteristiche economiche e

finanziarie del campione di imprese, mediante l'analisi dei rough sets basati sulla relazione di dominanza (DRSA), al fine di ottenere un innovativo modello di scoring basato su logiche multicriteriali piuttosto che statistiche come l'analisi discriminante lineare.

Dal campione di imprese sono stati estratti due campioni di uguale dimensione (566 imprese: 500 sane – 66 insolventi): un Training set, per calcolare le regole decisionali ed i coefficienti della funzione discriminate; ed un Validation set, per verificarne l'affidabilità. Complessivamente il campione è costituito da 1.132 imprese di cui 132 sono state dichiarate insolventi tra 04/2006 e 01/2007. Delle imprese considerate sono stati forniti i bilanci civilistici relativi all'esercizio 2004. Tali prospetti poi sono stati riclassificati (il conto economico con il criterio del valore aggiunto mentre lo stato patrimoniale con il criterio finanziario) e sono stati calcolati i principali indici di bilancio. Al fine di comprendere il legame tra i dati contabili ed il verificarsi dell'insolvenza sono stati selezionati 10 indici di bilancio che di seguito vengono elencati:

-C₁= mol/fatturato;

-C₂=ROS;

-C₃=ATO;

-C₄=ROE;

-C₅=rapporto di capitalizzazione;

-C₆=debt/equity;

-C₇=quoziente di struttura secondario;

- C_8 =quoziente di liquidità;

- C_9 =turnover dei clienti;

- C_{10} =autofinanziamento/fatturato.

Gli indicatori sopra considerati offrono un'ampia panoramica circa le caratteristiche economiche e finanziarie desumibili dal bilancio d'esercizio.

Infatti vengono presi in considerazione aspetti che vanno dalla redditività all'equilibrio finanziario ed alla solidità patrimoniale dell'impresa. Di seguito si riporta l'interpretazione dei 10 indicatori prese in considerazione:

-“mol/fatturato” (dove mol sta per margine operativo lordo): tale indicatore permette di vedere chiaramente se l'azienda è in grado di generare ricchezza tramite la gestione operativa, escludendo quindi buona parte delle politiche di “window dressing” messe in atto dagli amministratori dell'azienda che non sempre danno una visione corretta dell'andamento aziendale;

-“ROS” (Return On Sales, *reddito operativo/fatturato*): indica sempre una informazione relativa alla ricchezza generata dalla gestione operativa, ma più debole, in quanto nel calcolo del numeratore vengono incluse alcune delle manovre contabili sopra accennate come le svalutazioni del circolante e gli accantonamenti;

-“ATO” (Asset Turn Over, *fatturato/capitale investito*): rappresenta un indicatore di efficienza, in quanto indica nel corso di una gestione produttiva il numero delle volte in cui il capitale investito si è rinnovato o ha ruotato per effetto dei ricavi delle vendite;

- “ROE” (Return On Equity, *reddito netto/patrimonio netto*): indica la redditività del capitale apportato in azienda dai soci;
- “rapporto di capitalizzazione” (*patrimonio netto/totale fonti finanziarie*): indica il livello delle risorse finanziarie apportate in azienda dai soci, mediante conferimento di beni o denaro, o dalla stessa gestione aziendale mediante il reinvestimento dei profitti conseguiti nei vari esercizi;
- “debt/equity” (*debiti totali/patrimonio netto*): indica quante volte i capitali di terzi superano il patrimonio netto dell’azienda, ovvero il grado di indebitamento rispetto al patrimonio aziendale;
- “quoziente di struttura secondario” (*(patrimonio netto + fonti permanenti)/totale immobilizzazioni*): indica il rapporto tra le fonti finanziarie permanenti ed il capitale immobilizzato, ed esprime pertanto un’indicazione sulla correlazione temporale tra fonti di finanziamento ed investimenti, ovvero indica il livello di equilibrio strutturale dell’azienda;
- “quoziente di liquidità” (*(liquidità immediate + liquidità differite)/passività a breve termine*): mettendo in relazione tra loro attività più o meno esigibili (disponibilità di c/c, titoli, crediti, etc.) a debiti da regolarsi nel breve e brevissimo termine (scoperti di c/c, debiti verso l’erario, debiti verso fornitori, etc.) esprime un segnale circa l’equilibrio finanziario a breve termine;
- “turnover dei clienti” (*(media crediti verso clienti/fatturato)* 365*): indica la velocità media in giorni con cui i crediti verso clienti vengono incassati, e quindi

fornisce informazioni importati circa le politiche commerciali dell'azienda e la qualità del portafoglio clienti;

-“autofinanziamento/fatturato”: esprime una misura delle risorse finanziarie generate dalla gestione reddituale; infatti, il numeratore (utile d'esercizio +/-costi e ricavi non monetari) esprime il contributo fornito dall'esercizio alla creazione o assorbimento di risorse finanziarie.

Si precisa che tutti gli attributi, ad eccezione di C_6 e C_9 , sono positivamente correlati allo stato di salute della società. Infatti, relativamente a questi due attributi risulta abbastanza chiaro che lo stato di salute di un'azienda peggiori all'aumentare del indebitamento (C_6) e all'aumentare del periodo di incasso dei crediti verso clienti (C_9).

Di seguito viene riportato uno stralcio di 40 aziende delle 566 che compongono il Training set, dove nelle righe vengono riportate le aziende (20 sane e 20 insolventi), mentre nelle colonne dalla seconda alla undicesima sono riportati gli indici di bilancio. Nell'ultima colonna viene riportata la variabile binaria “Default” relativa all'insolvenza (con valore pari a 1 in caso di Default, 0 altrimenti).

Tabella 3. Training set											
<i>Impresa</i>	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	D
A1	0,09	0,06	1,19	0,11	0,14	6,31	1,31	0,70	123,92	0,04	0
A2	0,13	0,07	0,85	0,02	0,20	4,09	1,49	0,75	172,50	0,04	0
A3	0,10	0,04	0,94	0,00	0,32	2,10	0,93	0,43	74,36	0,05	0
A4	0,03	0,03	2,24	0,22	0,15	5,62	1,58	0,88	106,11	0,02	0
A5	0,05	0,03	1,38	0,09	0,16	5,34	1,11	0,65	120,24	0,02	0
A6	0,09	0,07	0,76	0,05	0,08	11,88	4,15	0,73	202,61	0,01	0
A7	0,06	0,02	0,75	0,00	0,25	2,97	0,71	0,55	156,16	0,00	0
A8	0,05	-0,03	0,79	0,01	0,33	2,05	2,65	1,28	227,74	0,08	0
A9	0,04	0,04	0,95	0,01	0,08	11,11	2,77	0,77	163,06	0,07	0
A10	0,45	0,24	0,30	0,07	0,51	0,96	1,49	0,67	88,06	0,30	0
A11	0,09	0,02	1,03	-0,05	0,17	4,82	1,05	0,52	109,59	0,03	0
A12	0,02	0,01	1,83	0,00	0,29	2,40	1,16	0,63	65,37	0,01	0
A13	0,07	0,03	1,25	0,01	0,09	9,84	1,08	0,70	89,98	0,02	0
A14	0,06	0,04	0,74	0,02	0,13	6,78	1,53	0,82	219,13	0,02	0
A15	0,07	0,02	1,51	0,03	0,17	4,93	0,96	0,80	89,41	0,07	0
A16	0,07	0,03	1,32	0,03	0,07	13,84	0,88	0,77	168,55	0,02	0
A17	0,07	0,07	0,89	0,05	0,57	0,75	1,88	1,93	117,97	0,10	0
A18	0,04	0,02	1,04	0,01	0,15	5,78	1,30	0,93	164,65	0,01	0
A19	0,06	0,04	1,23	0,08	0,07	13,23	1,75	0,85	191,07	0,02	0
A20	0,13	0,05	0,94	0,02	0,11	7,73	1,92	0,62	174,38	0,08	0
.
A501	-0,10	-0,14	0,66	-3,43	0,04	27,13	0,74	0,32	67,21	-0,16	1
A502	0,10	0,04	0,66	0,03	0,04	22,09	1,92	0,80	165,71	0,05	1
A503	0,01	-0,04	1,12	0,03	0,06	17,12	0,85	0,78	203,46	0,02	1
A504	0,04	0,07	0,62	-0,66	0,03	34,31	0,81	0,46	217,09	0,11	1
A505	0,07	-0,01	0,72	-0,35	0,08	12,23	1,10	0,81	154,09	0,08	1
A506	0,11	0,05	0,72	0,02	0,18	4,53	0,55	0,43	155,51	0,06	1
A507	-0,20	-0,21	0,56	-0,56	0,16	5,28	0,75	0,71	215,61	-0,09	1
A508	0,07	0,07	0,59	0,04	0,05	18,71	0,83	0,79	326,04	0,04	1
A509	0,13	0,11	0,74	0,08	0,17	5,00	0,91	0,40	74,60	0,04	1
A510	0,13	0,07	0,73	0,01	0,08	11,78	0,91	0,93	257,01	0,18	1
A511	-0,01	-0,02	1,45	-0,82	0,10	9,26	1,18	0,47	76,32	-0,05	1
A512	0,00	0,00	1,03	0,14	0,37	1,67	1,62	0,95	126,50	0,06	1
A513	0,14	0,09	0,75	0,02	0,15	5,84	0,90	0,40	112,49	0,03	1
A514	0,05	0,03	0,67	0,01	0,18	4,64	1,52	0,71	230,56	0,04	1
A515	0,05	0,02	0,88	-0,27	0,04	23,85	0,83	0,54	0,00	-0,01	1
A516	0,02	0,01	2,11	0,07	0,04	25,99	1,12	0,98	149,42	0,01	1
A517	0,05	0,02	2,30	0,01	0,11	8,30	1,59	1,08	105,46	0,04	1
A518	0,12	0,10	0,51	0,05	0,11	7,70	1,14	0,74	378,86	0,06	1
A519	0,04	0,04	1,27	0,21	0,09	9,88	0,52	0,51	95,97	0,03	1
A520	0,05	0,03	2,10	0,01	0,09	9,74	1,08	0,90	98,47	0,01	1
.

3.1 CALCOLO DELLE REGOLE DECISIONALI

Al fine di applicare la metodologia dei rough sets al Training set, di cui nella tabella 3 viene riportato uno stralcio, gli indici di bilancio sono stati considerati come criteri condizionali, mentre la variabile dicotomica “Default” è stata definita come criterio decisionale. Pertanto, a fronte di 10 indici di bilancio (criteri condizionali) avremo due classi decisionali Cl_1 (imprese insolventi) e Cl_2 (imprese sane). Dall’analisi del Training set è stato possibile, mediante l’ausilio di un software creato appositamente, procedere al calcolo di regole e ridotti.

Nel caso specifico, è stato trovato solo un ridotto condizionale, composto dall’insieme di tutti i criteri condizionali ad eccezione del criterio C_5 “rapporto di capitalizzazione”, ritenuto ridondante in seno all’analisi eseguita.

$$RED_{Cl}^1 = \{C_1, C_2, C_3, C_4, C_6, C_7, C_8, C_9, C_{10}\}$$

Quindi, avendo ottenuto solo un ridotto questo coincide con il *Core*. Inoltre, è emerso che nel training set ci sono 26 oggetti ambigui (frontiera), ovvero aziende fallite che, rispetto agli attributi condizionali considerati, dominano aziende sane, o aziende sane dominate da aziende fallite.

L’insieme di tutte le regole decisionali è composto da 378 regole che descrivono le imprese insolventi Cl_1^{\leq} e 2339 regole che descrivono le imprese sane Cl_2^{\geq} con

una confidenza del 100%. Per l'applicazione al validation set si sono selezionate tra le regole che descrivono le aziende fallite (regole negative) le 16 che supportano almeno 10 aziende, e tra regole che descrivono le aziende sane (regole positive) le 29 che supportano almeno 135 aziende. Di seguito vengono elencate le cinque regole più supportate per ciascuna unione considerata (tra parentesi viene indicato il numero di imprese supportate dalla corrispondente regola; ovviamente tanto più è alto il numero di imprese supportato più alta è la forza della regola):

1. se $f(x, C_4) \leq -0,27$, allora $x \in Cl_1^{\leq}$ (supporto 17);
2. se $f(x, C_5) \leq 0,03$, allora $x \in Cl_1^{\leq}$ (supporto 14);
3. se $f(x, C_6) \geq 25,43$, allora $x \in Cl_1^{\leq}$ (supporto 18);
4. se $f(x, C_3) \leq 2,11$ e $f(x, C_5) \leq 0,04$ e $f(x, C_{10}) \leq 0,01$, allora $x \in Cl_1^{\leq}$ (supporto 16);
5. se $f(x, C_3) \leq 1,2$ e $f(x, C_5) \leq 0,06$ e $f(x, C_8) \leq 0,76$ e $f(x, C_{10}) \leq 0,02$, allora $x \in Cl_1^{\leq}$ (supporto 14);
6. se $f(x, C_1) \geq 0,06$ e $f(x, C_3) \geq 0,08$ e $f(x, C_5) \geq 0,12$ e $f(x, C_{10}) \geq 0,02$, allora $x \in Cl_2^{\geq}$ (supporto 190);
7. se $f(x, C_1) \geq 0,07$ e $f(x, C_3) \geq 0,85$ e $f(x, C_6) \leq 7,64$ e $f(x, C_{10}) \geq 0,02$, allora $x \in Cl_2^{\geq}$ (supporto 163);

8. se $f(x, C_1) \geq 0,08$ e $f(x, C_3) \geq 0,81$ e $f(x, C_6) \leq 7,98$ e $f(x, C_{10}) \geq 0,02$,

allora $x \in Cl_2^{\geq}$ (supporto 151);

9. se $f(x, C_1) \geq 0,06$ e $f(x, C_3) \geq 0,80$ e $f(x, C_6) \leq 7,08$ e $f(x, C_{10}) \geq 0,02$,

allora $x \in Cl_2^{\geq}$ (supporto 188);

10. se $f(x, C_1) \geq 0,05$ e $f(x, C_6) \leq 8,05$ e $f(x, C_7) \geq 0,99$ e

$f(x, C_9) \leq 163,21$ e $f(x, C_{10}) \geq 0,03$, allora $x \in Cl_2^{\geq}$ (supporto 166).

A maggior chiarimento di quanto esposto, se prendiamo in considerazione l'ultima regola, questa può leggersi in questo modo: se il rapporto mol/fatturato è almeno pari al 5% e, contemporaneamente, la leva finanziaria non è superiore di 8,05, il quoziente di struttura secondario è almeno pari a 0,99, il turnover dei clienti non è superiore 163,21 giorni e il rapporto autofinanziamento/fatturato è almeno pari almeno 0,03, allora l'azienda appartiene all'insieme delle imprese sane. Risulta chiaro che condurre un processo decisionale seguendo tale metodologia risulta molto trasparente, in quanto si può risalire alle singole caratteristiche (criteri condizionali) che hanno condotto alla valutazione positiva o negativa dell'impresa da esaminare. Pertanto, per verificare lo stato di salute di una nuova azienda, estranea al training set utilizzato per il calcolo delle regole decisionali, basterà verificare quante e quali regole decisionali sono rispettate.

Chiaramente, un giudizio in tal senso, potrà essere espresso solo dopo aver determinato, attraverso un insieme di regole decisionali, il grado di appartenenza dell'impresa all'insieme delle imprese sane piuttosto che quello delle insolventi.

Utilizzando l'analisi discriminante lineare, invece, il processo si sostanzia nel trovare il vettore dei coefficienti della funzione discriminante, che rende massimo il rapporto tra la distanza tra le medie degli score dei due gruppi (varianza tra i gruppi) e la varianza degli score all'interno dei due gruppi (varianza entro i gruppi). Applicando tale algoritmo di calcolo al training set esaminato è stata ottenuta la seguente funzione discriminante lineare:

$$Z = 0,102 \cdot C_1 + 0,023 \cdot C_2 + 0,238 \cdot C_3 + 0,051 \cdot C_4 + 0,121 \cdot C_5 - 0,719C_6 + \\ + 0,101 \cdot C_7 + 0,081C_8 - 0,236C_9 + 0,301C_{10}$$

I segni dei coefficienti coincidono con quelli attesi in relazione al segno della correlazione con lo stato di salute dell'impresa. Il valore di cut-off Z_c , ovvero il punteggio ottimo per la separazione tra i gruppi Cl_1^{\leq} (imprese insolventi) e Cl_2^{\geq} (imprese sane) è stato calcolato in - 40,25. Quindi la regola decisionale per classificare nuove imprese diventa: “assegna un'impresa i -esima al gruppo Cl_1^{\leq} se $Z_i < Z_c$, altrimenti assegnala al gruppo Cl_2^{\geq} ”. Pertanto, risulta abbastanza evidente che un processo decisionale impostato in tal modo si presenti al decisore come una “*black-box*”, in quanto quest'ultimo non è in grado di comprendere in modo trasparente la genesi della funzione discriminante e quindi del processo decisionale, fenomeno che come si è visto sopra non si verifica con il metodo dei rough sets che pertanto, per la chiarezza e per la trasparenza che li

contraddistingue, appartengono pienamente alla famiglia dei modelli “*glass-box*”.

3.2. ANALISI DEI RISULTATI

Al fine di verificare i risultati sopra ottenuti sono state applicate le regole decisionali ad un validation set composto sempre da 566 imprese (500 sane e 66 insolventi) ed è stata ottenuta la seguente matrice di confusione.

	Classificazione			Totale
	<i>Sane</i>	<i>Incerte</i>	<i>Insolventi</i>	
<i>Sane</i>	251	240	9	500
<i>Insolventi</i>	8	28	30	66

Le informazioni contenute nella tabella 4 esprimono la capacità diagnostica delle regole decisionali ottenute dal training set. In particolare, nella prima colonna viene riportato lo stato di salute delle imprese da analizzare, mentre nelle altre colonne viene riportato il numero di imprese classificate, mediante le regole decisionali trovate, come sane, incerte e fallite. Nella tabella 5, al fine di evidenziare i risultati, vengono riproposti i dati della tabella 4 in termini percentuali.

Tabella.5 Matrice di confusione delle regole decisionali %				
	Classificazione			Totale
	<i>Sane</i>	<i>Incerte</i>	<i>Insolventi</i>	
<i>Sane</i>	50,20%	48,00%	1,80%	100%
<i>Insolventi</i>	12,10%	42,40%	45,50%	100%

Pertanto, applicando le 29 regole decisionali positive alle imprese del validation set, verifico che queste nel 50,2% dei casi riescono a classificare correttamente le imprese sane, nel 48% non riescono a formulare un giudizio e nel 1,8% commettono l'errore di classificare come fallita un'impresa sana. Parallelamente, le 16 regole decisionali negative nel 45,5% dei casi riescono ad indovinare lo stato di insolvenza, nel 42,4% non riescono a formulare un giudizio e nel 12,1% classificano erroneamente un'impresa insolvente come sana. Pertanto, l'errore di I tipo è pari al 12,1% e l'errore di II tipo è pari al 1,8%. In realtà andrebbe aggiunta un'altra colonna alle matrici di confusione trovate, la colonna della contraddizione, ovvero dei casi in cui un'azienda viene classificata sana da almeno una delle 29 regole positive e contemporaneamente fallita da almeno una delle 16 regole negative, ma nell'analisi effettuata tale anomalia non si è verificata. Un'altra lettura, molto più interessante della tabella 4, può essere effettuata unendo i risultati ottenuti per le imprese sane con quelli ottenuti sulle imprese insolventi.

Tabella 6 Rielaborazione della matrice di confusione			
	<i>Sane</i>	<i>Incerte</i>	<i>Fallite</i>
Classificazione	259	268	39
Errore	8	-	9

In particolare:

- saranno considerate sane 259 imprese, con un errore del 3% (8 imprese);
- saranno considerate fallite 39 imprese, pari al 6,8% dei casi esaminati, con un errore del 23% (9 imprese);
- non si riuscirà ad esprimere un giudizio su 268 imprese, pari al 47% dei casi esaminati, chiaramente per tali imprese non si commetteranno errori in quanto non è stato possibile, tramite le 45 regole decisionali, effettuare l'assegnazione alla classe "sane" o "insolventi".

Se si vuole ridurre il numero di imprese incerte, sarà necessario andare a ricalcolare sul training set regole decisionali che supportano un numero minore di imprese, si ricorda che il supporto era stato fissato a 135 per le regole positive e 10 per quelle negative, ed applicare tali regole solo all'insieme delle imprese incerte. Ciò consentirà di poter formulare un giudizio anche su tali imprese ma con un errore di I e II tipo di maggiore entità.

4. CONCLUSIONI

Quindi, se un istituto di credito dovesse decidere di applicare un modello di scoring che utilizzi i rough sets basati sulla dominanza, dovrebbe articolare il processo di valutazione del merito creditizio secondo i seguenti passi:

1. predisporre un campione significativo di clienti che in passato si sono verificati affidabili o insolventi (“training set”);
2. individuare, con metodi statistici o multicriteriali, una combinazione di variabili (dati di bilancio, dati andamentali interni, dati relativi alla centrale rischi, etc.) che abbiano un contenuto sufficiente a descrivere le caratteristiche economiche e finanziarie delle imprese appartenenti al campione;
3. procedere al calcolo delle regole decisionali, fissando diversi supporti, in base alla numerosità del campione, sia per le regole positive che per quelle negative al fine di ottenere diversi insiemi di regole decisionali;
4. verificare attraverso l’analisi di un validation set la bontà delle regole trovate.

Pertanto, ogni volta che si presenterà un nuovo cliente per una richiesta di affidamento, mediante l’ausilio dell’insieme di regole decisionali associate al supporto (per le regole positive e negative) più elevato, il decisore sarà in grado di esprimere un parere favorevole o sfavorevole alla concessione dell’affidamento, con un certo margine di errore (errore I tipo e errore II tipo). Nel caso in cui l’insieme delle regole decisionali (positive e negative) associate

al supporto considerato non consenta di classificare il cliente nell'insieme dei clienti "sani" o "insolventi", occorrerà verificare se esistono regole decisionali associate a supporti inferiori che consentano di effettuare la classificazione in clienti sani o insolventi. Nel caso in cui nessun insieme di regole decisionali riesca a classificare il cliente da esaminare o le classificazioni ottenute siano associate a livelli di errore non sopportabili, il decisore sarà chiamato ad esprimere giudizio sul merito creditizio del cliente sulla base della propria esperienza in materia, eventualmente ricorrendo a maggiori informazioni e approfondimenti sullo stato di salute dell'azienda da valutare .

CONCLUSIONE

A conclusione di questo lavoro, si vogliono ancora un volta sottolineare le potenzialità di un approccio multicriteriale per la valutazione del merito creditizio. In particolare, partendo dai risultati ottenuti dall'applicazione dei rough sets basati sulla dominanza al credit scoring, è emerso che è possibile creare un modello di scoring innovativo, chiaro e trasparente. Innovativo in quanto i rough sets, rispetto all'analisi statistica, applicata largamente nella pratica, presentano una maggiore oggettività nei seguenti punti:

-non necessitano di alcuna procedura di identificazioni e stima dei parametri delle equazioni strutturali (funzione discriminante; funzione logistica; etc.), in quanto il principale processo di calcolo consiste nel determinare, dalle evidenze empiriche fornite dalla tavola delle informazioni (tavola che raccoglie le informazioni sugli oggetti da esaminare) le regole decisionali ed i ridotti;

-non occorre che i campioni da analizzare siano statisticamente significativi, pertanto è possibile analizzare anche tavole delle informazioni di ridotte informazioni;

-non necessitano di operatori per l'aggregazione dei dati (medie, varianze, matrice delle covarianze, etc.), in quanto i dati vengono analizzati nella loro forma originaria;

-il risultato del modello non è una rappresentazione funzionale, a volte difficile da interpretare, ma un insieme di regole decisionali sottoforma di proposizioni logiche del tipo “se....., allora...”.

Relativamente alla chiarezza ed alla trasparenza, è evidente che sottoporre ad un qualsivoglia decisore, nello specifico un organo decisionale di un istituto di credito (Comitato del Credito, Consiglio di Amministrazione, etc.), un problema sottoforma di proposizioni logiche facilmente comprensibili, in luogo di dati di sintesi la cui genesi è nota solo all’analista che li ha determinati, facilita enormemente la capacità di comprendere le problematiche relative al rischio di credito e rende più efficiente ed efficace il processo del credito:

- efficiente in quanto, grazie all’immediata comprensione dei risultati, consente di velocizzare la procedura relative alla concessione degli affidamenti;
- efficace in quanto, offrendo sempre informazioni di semplice interpretazione, consente all’istituto di credito, durante la negoziazione di un affidamento, di far comprendere al cliente le motivazioni sottostanti la delibera adottata e la connessa politica di pricing.

Possibili sviluppi potrebbero consistere nel costruire un modello di valutazione del rischio creditizio basato integralmente su logiche multicriteriali, il cui principale elemento sia rappresentato da un processo di scoring determinato dall’analisi dei rough sets basati sulla dominanza.

BIBLIOGRAFIA

1. Abbas M., Vincke P. (1993), *Preference structures and threshold models*, Journal of Multi-Criteria Decision Analysis, 2, 171-178.
2. Abbas M., Pirlot M., Vincke P. (1996), *Preference structures and co-comparability graphs*, Journal of Multi-Criteria Decision Analysis, 2, 81-98
3. Altman E. (1968), *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*, Journal of Finance.
4. Altman E. (1984), *A further empirical investigation of the bankruptcy cost question*, Journal of Finance.
5. Altman E., Avery R., Eisenbeis R, Sinkey J. (1981), *Application of classification techniques in business, banking and finance*, Jai Press, NY.
6. Altman E., Hadelman R., Narayanan P. (1977), *Zeta analysis*, Journal of Banking and Finance n. 1.
7. Beaver W. H. (1966), *Financial Ratios As Predictors of failure*, Journal Of Accounting Research, Vol. 4, Empirical Research in Accounting: pp 71-111.
8. Bouyssou D. (1990), *Building criteria: A prerequisite for MCDA*, in C.A. Bana e Costa (ed.), *Readings in Multiple Criteria Decision Aid*, Springer-Verlag, 58-80.
9. Brans J., P., Vincke P. (1985), *A preference ranking organization method*, Management Science, 31, 647-656.
10. Brusa L., Zamprognia L. (1998), *Finanza D'Impresa*, Etaslibri 1998,
11. Caouette J., Altman E., Narayanan P. (1998), *Managing Credit Risk*, J. Wiley, NY.
12. Dimitras A., Zanakis I., Zoupounidis C. (1996), *A survey of business failures with an emphasis on prediction methods and industrial applications*, European Journal of Operational Research, 90, 487-513.
13. Dimitras A., Slowinski R., Susmaga R., Zopounidis C. (1999), *Business failure prediction using rough sets*, European Journal of Operational Research.
14. Doignon J. P. (1987), *Threshold representation of multiple semiorders*, SIAM Journal on Algebraic and Discrete Methods, 8, 77-84.
15. Figueira J., Greco S., Ergthott M. (2005), *Multiple Criteria Decsion Analysis: State of the Art Surveys*, Springer, Berlin.
16. Fishburn P. C. (1991), *Nontransitive additive conjoint measurement*, Journal of Mathematical Psychology, 35, 1-40.
17. Fischer R. A. (1936), *The Use Of Multiple Measurement In Taxonomic Problems*, Annals of Eugenics, V. 7, p. 179-188.

18. Fodor J., Roubens M. (1996), *Parameterized Preference Structures and Some Geometrical Interpretation*, Institut de Mathématique Université de Liège, 96.008.
19. Greco S., Matarazzo B., Slowinski R. (1996), *Rough Approximation of Preference Relation by Dominance Relations*, ics research report 16/96, Warsaw University of Technology and European Journal of Operational Research, 117:63-83.
20. Greco S., Matarazzo B., Slowinski R. (1998), *A new rough set approach to evaluation of bankruptcy risk*, in Zopounidis C. ,“Operational tools in the management of financial risk” Kluwer A.P., Dordrecht.
21. Greco S., Matarazzo B., Slowinski R. (1999), *The use of rough sets and fuzzy sets in MCDM*, Chapter 14 in “Advances in Multiple Criteria Decision Making”, T.Gal, T.Stewart, T.Hanne (eds.),. Kluwer Academic Publishers, Boston, pp. 14.1-14.59.
22. Greco S., Matarazzo B., Slowinski R. (2001), *Rough sets methodology for multi-criteria decision analysis*, European Journal of Operational Research, vol. 129, pp. 1–47.
23. Greco S., Matarazzo B., Slowinski R. and Stefaniowski J. (2000), *Variable consistency model of dominance-based rough set approach*, in W. Ziarko, Y.Yao: *Rough Sets and Current Trends in Computing*, Lecture Notes in Artificial Intelligence, vol 2005, Springer-Verlag, Berlin, 2001, pp 170-181.
24. Greco S., Matarazzo B., Slowinski R. (2002a), *Rough sets methodology for sorting problems in presence of multiple attributes and criteria*, European Journal of Operational Research, vol. 138, pp. 247–259.
25. Greco S., Matarazzo B., Slowinski R. (2002b), *Rough approximation by dominance relations*, International Journal of Intelligent Systems, vol. 17 no. 2, pp. 153-171.
26. Greco S., Matarazzo B., Slowinski R. (2005), *Decision rule approach*. Chapter 13 [in]: J.Figueira, S.Greco and M.Ehrgott (eds.), “Multiple Criteria Decision Analysis: State of the Art Surveys”, Springer-Verlag, New York, pp. 507-562.
27. Keeney R. L., Raiffa H. (1976), *Decision with Multiple Objectives - Preferences and value Tradeoffs*, Wiley, New York.
28. Lawrence E., Arshadi N. (1995), *A multinomial logit analysis of problem loan resolution choices in banking* in Journal of Money, Credit and Banking.
29. Lo A. (1986), *Logit versus discriminant analysis*, Journal of Econometrics.
30. Luce R.D. (1956), *Semi-orders and a theory of utility discrimination*, Econometrica, 24, 178-191.
31. Matarazzo B. (1997), *L'approccio dei rough sets all'analisi delle decisioni*, Atti del XXI Convegno Annuale A.M.A.S.E.S., Appendice, Roma, pp. 77-111.

32. Pawlak Z. (1982), *Rough sets*, International Journal of information & Computer Sciences 11:341-356.
33. Pawlak Z. (1991), *Rough sets. Theoretical Aspects of Reasoning about data*, Dordrecht: Kluwer Academic Publishers.
34. Pawlak Z. (1997), *Rough set approach to knowledge-based decision support*, European Journal of Operational Research.
35. Resti A. (2001), *Misurare e gestire il rischio di credito nelle banche: una guida metodologica*, Alpha Test, Milano.
36. Roy B. (1985), *Méthodologie Multicritère d'aide à la Décision*, Economica, Paris.
37. Roy B. (1990), *Decision-aid and decision-making*, European Journal of Operational Research, 45, 324-331.
38. Roy B. (1993), *Decision science or decision aid science?*, European Journal of Operational Research, Special Issue on Model Validation in Operations Research, 66, 184-203.
39. Roy B., Bouyssou D. (1993), *Aide Multicritère à la Décision: Méthodes et Cas*, Economica, Paris.
40. Roy B., Vincke P. (1984), *Relational systems of preference with one or more pseudo-criteria: some new concepts and results*, Management Science, 30 (11), 1323-1335.
41. Roy B., Vincke P. (1987), *Pseudo-orders: definition, properties and numerical representation*, Mathematical Social Sciences, 14 (2), 263-274.
42. Roberts F. S. (1971), *Homogeneous families of semiorders and the theory of probabilistic consistency*, Journal of Mathematical Psychology, 8, 248-263.
43. Roubens M., Vincke P. (1985), *Preference Modelling*, Lectures Notes in Economics and Mathematical Systems, 250, Springer.
44. Slowinski K., Slowinski R., Stefanoski J. (1988), *Rough sets approach to analysis of data peritoneal lavage in acute pancreatitis*. Medical Informatics, 13, 145-159.
45. Slowinski R., Vanderpooten. D. (1997), *Similarity relation as a basis for rough approximations*. In P.P. Wang, editor, *Advances in Machine Intelligence and Soft-Computing*, vol.IV, pages 17--33. Duke University Press, Durham, NC.
46. Slowinski R., Zopounidis C. (1995), *Application of rough set approach to evaluation of bankruptcy risk*, International Journal of Intelligent Systems in Accounting, Finance and Management, March.
47. Sounderpandian J. (1991), *Value functions when decision criteria are not totally substitutable*, Operations Research, 39, 4, 592-600.
48. Stefanoski J. (1992), *Rough Set theory and discriminant methods as tools for analysis of information systems. A comparative study*, Foundation of Computing and Decision Sciences, 17 (2), 81-98.

49. Szegö G., Varetto F. (1999), *Il Rischio Creditizio Misura e Controllo*, Utet, Torino.
50. Tsoukias A., Vincke P. (1995), *A new axiomatic foundation of partial comparability*, Theory and Decision, 39, 79-114.
51. Tsoukias A., Vincke P. (1998), *Double Threshold orders: A new axiomatization*, Journal of Multi-criteria Decision Analysis, 7, 285-301.
52. Varetto F. (1990), *Il sistema di diagnosi dei rischi di insolvenza della Centrale dei Bilanci*, Bancaria Editrice, Roma.
53. Vincke P. (1980), *Vrais, quasi, pseudo et précritères dans un ensemble fini: propriétés et algorithmes*, Cahiers du Lamsade, 27, Université Paris-Dauphine.
54. Vincke P. (1988), *(P,Q,I)-preference structures*, in J. Kacprzyk e M. Roubens (eds), *Nonconventional preference relations in decision making*, Springer-Verlag, 301, 72-81.