Università degli Studi Mediterranea di Reggio Calabria

DiGiES Dipartimento di GIURISPRUDENZA ECONOMIA e SCIENZE UMANE

DIPARTIMENTO di ECCELLENZA

La Teoria dei Rough Sets

(Insiemi approssimati)

BUSINESS ANALYTICS AND DECISIONS THEORY

Prof. MASSIMILIANO FERRARA

(A cura della D.ssa Merenda Domenica Stefania)

La Teoria dei Rough Sets

La teoria degli *Rough Sets* è una metodologia di estrazione della conoscenza, da raccolte di **dati** spesso riguardanti ***problemi decisionali.***

I dati rappresentano gli *oggetti* che vengono descritti da un insieme di **attributi** o criteri (ossia proprietà o caratteristiche) e da una loro **valutazione globale**.

La Teoria dei Rough Sets

Per esempio:

Considerando dei dati relativi all'affidamento bancario;

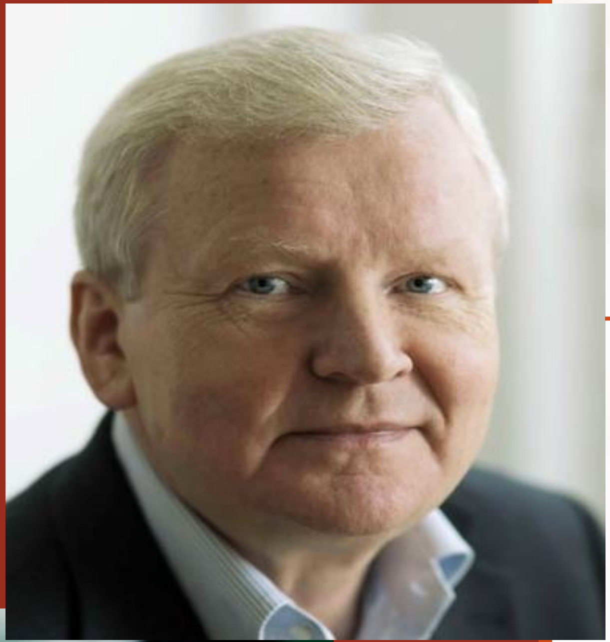
- gli *oggetti* saranno le aziende,
- gli *attributi* saranno alcuni dati di bilancio,
- la *valutazione globale* sarà la decisione finale.

La Teoria dei Rough Sets

Questa metodologia di estrazione della conoscenza è basata sui *c.d.* ***Rough Sets*.**

I risultati forniti dall'applicazione sono utili per:

- decisioni passate,
- analoghe situazioni future.



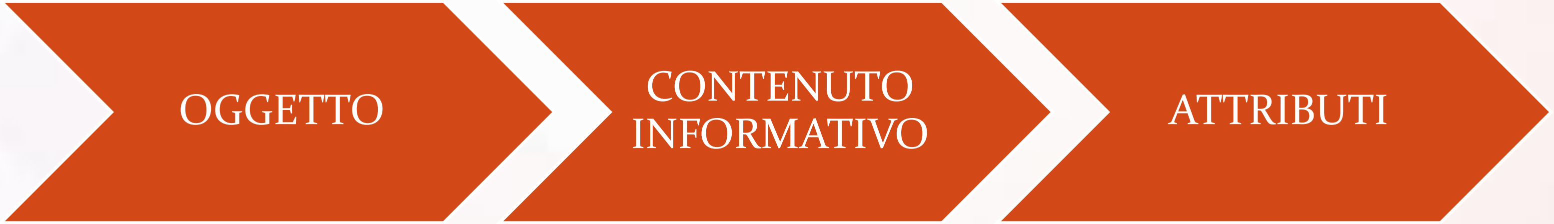
La Teoria dei Rough Sets

La teoria dei Rough Sets fu introdotta agli inizi degli anni ottanta, da Pawlak & Slowinski (1982) come:

«Lo strumento matematico per l'analisi di oggetti descritti in maniera vaga, ossia in presenza di informazioni che possono anche essere inconsistenti o ambigue».

La Teoria dei Rough Sets

Ad ogni oggetto *dell'Universo U* di interesse, sono associate delle informazioni espresse tramite un insieme di attributi che lo descrivono.



La Teoria dei Rough Sets:
Relazioni di Indiscernibilità

Relazioni di indiscernibilità: relazioni di similarità fra gli oggetti caratterizzati dallo stesso contenuto informativo.

Atomo o ***granulo elementare di conoscenza:*** insieme di uno o più oggetti tra loro indiscernibili.

La ***relazione binaria di indiscernibilità*** è una relazione di *equivalenza*: partizione dell'Universo in **classi di equivalenza.**

La Teoria dei Rough Sets

Ogni sottoinsieme *X* di *U* può essere espresso in termini di *insiemi elementari:*

- in maniera *precisa*, come loro unione;
- *Approssimativamente.*
  - L'***approssimazione inferiore***: elementi che appartengono ad *X* con <u>*certezza.*</u>
  - *L'**approssimazione superiore:*** elementi <u>*possono*</u> appartenere ad *X*.

La Teoria dei Rough Sets

La *differenza* tra le due approssimazioni costituisce la ***frontiera dei Rough Sets***, chiamata anche *boundary region*, i cui elementi non possono essere caratterizzati con certezza come appartenenti o no ad *X*.

Pertanto l'***informazione*** è quindi *inconsistente* o *ambigua*.

La Teoria dei Rough Sets

In termini formali, indichiamo con:

- $I(X) \rightarrow$ la **classe di equivalenza** di $x \in U, \forall\, X \subseteq U$;

- $\underline{I}(X) \rightarrow$l'***approssimazione inferiore***

$$\boxed{\underline{I}(X) = \cup_{I(x) \subseteq X} I(x)}$$

- $\overline{I}(X) \rightarrow$ l'***approssimazione superiore***

$$\boxed{\overline{I}(X) = \cup_{I(x) \cap X \neq \emptyset} I(x)}$$

La Teoria dei Rough Sets

Input: sia **dati quantitativi** che **qualitativi.**

Output: informazioni circa la rilevanza degli attributi, rispetto alla qualità dell'approssimazione considerata.

Rapporti di causa-effetto descritti sotto forma di semplici regole, ad esempio attraverso proposizioni logiche del tipo:

«if….then…»

in tal modo si possono ottenere dei modelli di conoscenza *certi* e *possibili nella forma di regole decisionali*.

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA)

Sia **S** una ***tabella di dati***, le cui righe si riferiscono ai diversi *oggetti* e le colonne ai vari *attributi* considerati: ogni sua cella contiene una *valutazione* cioè una descrizione quantitativa o qualitativa di un oggetto tramite l'attributo corrispondente.

La *data table* formalmente rappresenta una quadrupla:

$$\boxed{S=U, Q, V, f}$$

Dove:

- U è l'Universo del discorso,
- Q={q1 ,…, qm} è l'insieme degli *m attributi*,
- $V = U_{q \in Q} V q$ *è l'insieme dei domini degli attributi*
- $f : U \times Q \to V$ è una *funzione di informazione* tale che f(x , q) ∈Vq per ogni q ∈ Q, x ∈U

Descrizione Formale Dei Rough Sets: tabella delle informazioni

ATTRIBUTI

OGGETTI

	q_1	q_2	…	…	q_m
x_1	$f(x_1, q_1)$	…	…	…	
x_2	…	$f(x_2, q_2)$			
x_3	…	…			
…	…		…		
…	…			…	
x_n	…				$f(x_n, q_m)$

Valutazioni

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA)

Ogni oggetto x di U è descritto da un *vettore stringa,* chiamato *descrizione di x in termini di attributi di Q e che contiene le informazioni disponibili su x:*

$$\boxed{Des_{Q(x)} = [f(x, q_1), f(x, q_2), \dots \dots \dots f(x, q_m)]}$$

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA)

Ad ogni sottoinsieme non vuoto di attributi $(P \subseteq Q)$ si associa una *relazione di indiscernibilità* su U, denotata *Ip* e così definita:

$$\boxed{Ip = \{(x, y) \in U \times U : f(x, q) = f(y, q), \quad \forall\, q \in P\}}$$

Se $(x, y) \in Ip$, si dice che gli oggetti *x* e *y* sono *P-indiscernibili.*

La *classe di equivalenza* contenente l'elemento x $\in U$ è denominata *Ip(x)* e le classi di equivalenza della relazione sono chiamate *P-insiemi elementari* o ***granuli di conoscenza*** determinati da P.

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA)

La *P-approssimazione inferiore* di X (denotata da $\underline{P}(X)$) e la *P-approssimazione superiore* di X (denotata $\overline{P}(X)$) in S sono definite rispettivamente:

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

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

ovvero *l'approssimazione inferiore* si compone di tutti e solo quegli ***oggetti che appartengono alle classi di equivalenza*** generate dalla relazione di indiscernibilità contenute in *X* .

L'approssimazione superiore, invece, comprende tutti e solo quegli **oggetti che appartengono alle classi di equivalenza** che **contengono almeno un oggetto** appartenente ad *X*.

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA)

La *Frontiera* di X in S, denotata $B_{np}(X)$ è definita come differenza tra gli insiemi approssimazione:

$$\boxed{B_{np}(X) = \overline{P}(X) - \underline{P}(X)}$$

- $B_{np}(X)$ -> *regione del dubbio*
- Se $B_{np}(X) = \emptyset$ −> *insieme ordinario*

- Se $B_{np}(X) \neq \emptyset$ −> *Insieme Rough* = approssimazioni $\overline{P}(X)$ *e* $\underline{P}(X)$.

Il *Rough Set* non sono altro che la famiglia di tutti gli insiemi $X \subseteq U$ aventi le stesse approssimazioni.

La Teoria dei Rough Sets

Accuratezza dell'approssimazione:

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

- $se\ \alpha_p(X) = 1 \rightarrow insieme\ ordinario$
- $se\ \alpha_p(X) < 1\ \rightarrow insieme\ Rough$

La Teoria dei Rough Sets

Qualità dell'approssimazione:

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

- la **qualità** rappresenta la frequenza relativa.

La Teoria dei Rough Sets

Grado di appartenenza degli elementi all'insieme:

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

Rough Membership Function

La Teoria dei Rough Sets

Relazioni tra *rough memebership function* e con le approssimazioni:

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

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

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

La Teoria dei Rough Sets

Ridotto:

sottoinsieme di attributi privo di attributi sono superflui

Red (P)

Nucleo:

l'insieme contenente tutti gli attributi indispensabili.

$$Core(P) = \cap \operatorname{Re} d(P)$$

La Teoria dei Rough Sets

Tavola delle decisioni

Tavola delle informazioni dove gli attributi di Q vengono distinti in:

- attributi condizionali (insieme C)
- attributi decisionali (insieme D)

Può anche essere espressa come un **insieme di regole decisionali**.

La Teoria dei Rough Sets: Esempio

Deve essere espresso un giudizio globale, ***criterio decisionale***, su alcuni studenti di una scuola, sulla base delle loro *valutazioni* ottenute in Matematica e Fisica.

Si supponga di voler approssimare gli insiemi di studenti con **valutazione complessiva** Scarso, Medio e Buono indicati rispettivamente con S, M e B; mediante le valutazioni in Matematica e Fisica.

La Teoria dei Rough Sets: Esempio

Per la costruzione del modello si utilizzano valutazioni indicate nella tabella:

STUDENTE	MATEMATICA	FISICA	VALUTAZIONE GLOBALE
S1	BUONO	MEDIO	**BUONO**
S2	MEDIO	MEDIO	**MEDIO**
S3	MEDIO	MEDIO	**MEDIO**
S4	BUONO	BUONO	**BUONO**
S5	BUONO	MEDIO	**BUONO**
S6	BUONO	BUONO	**BUONO**
S7	SCARSO	SCARSO	**SCARSO**
S8	SCARSO	SCARSO	**SCARSO**

La Teoria dei Rough Sets: Esempio

L'**Universo del discorso** è dato dall'insieme di tutti gli studenti:

$$U = \{S1, S2, S3, S4, S5, S6, S7, S8\}$$

Si considera la ***relazione di indiscernibilità*** definita come segue:

due studenti generici *x* e *y* si dicono indiscernibili rispetto alle loro valutazioni matematiche e fisiche ($\boldsymbol{xI_{MF}y}$), se le valutazioni di *x* in Matematica e Fisica sono le stesse di quelle di *y*.

Quindi nel nostro esempio si avrà:

S1 $\boldsymbol{I_{MF}}$ S1 -S1 $\boldsymbol{I_{MF}}$ S5, - S2 $\boldsymbol{I_{MF}}$ S2,- S2 $\boldsymbol{I_{MF}}$ S3, -S3 $\boldsymbol{I_{MF}}$ S3,- S3 $\boldsymbol{I_{MF}}$ S2,

-S4 $\boldsymbol{I_{MF}}$ S4, -S4 $\boldsymbol{I_{MF}}$ S6, -S5 $\boldsymbol{I_{MF}}$ S5, -S5 $\boldsymbol{I_{MF}}$ S1 -S6 $\boldsymbol{I_{MF}}$ S6,- S6 $\boldsymbol{I_{MF}}$ S4,

-S7 $\boldsymbol{I_{MF}}$ S7, -S7 $\boldsymbol{I_{MF}}$ S8, - S8 $\boldsymbol{I_{MF}}$ S8;- S8 $\boldsymbol{I_{MF}}$ S7.

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA)

Ogni oggetto x di U è descritto *descrizione di x in termini di attributi di Q:*

$$\boxed{Des_{Q(x)} = [f(x, q_1), f(x, q_2), \dots \dots \dots f(x, q_m)]}$$

Riprendendo il nostro esempio, ogni oggetto di x è descritto come segue:

$Des_Q(S1) = [buono, medio, buono]$

$Des_Q(S2) = [medio, medio, medio]$

……………………………………………

$Des_Q(S8) = [scarso, scarso, scarso]$

La Teoria dei Rough Sets: Esempio

Le ***classi di equivalenza***, $I_{MF}(x)\ con\ x \in U$, cioè gli insiemi elementari saranno, pertanto composti dagli studenti che hanno le stesse valutazioni rispetto a Matematica e Fisica, come segue:

$$I_{MF}(S1) = I_{MF}(S5) = \{S1, S5\}$$
$$I_{MF}(S2) = I_{MF}(S3) = \{S2, S3\}$$
$$I_{MF}(S4) = I_{MF}(S6) = \{S4, S6\}$$
$$I_{MF}(S7) = I_{MF}(S8) = \{S7, S8\}$$

A questo punto siamo in grado di calcolare l'***approssimazione inferiore*** e ***superiore*** degli studenti con **valutazione globale** Scarso, Medio e Buono.

La Teoria dei Rough Sets: Esempio

$$SCARSO \begin{cases} \underline{I}(S) = U_{I(x) \subseteq S} I(x) = I(S7) \cup I(S8) = \{S7, S8\} \\ \bar{I}(S) = U_{I(x) \cap S \neq \emptyset} I(x) = I(S7) \cup I(S8) \cup I(S1) \cup I(S5) = \{S7, S8\} \end{cases}$$

$$MEDIO \begin{cases} \underline{I}(M) = U_{I(x) \subseteq M} I(x) = I(S2) \cup I(S3) = \{S2, S3\} \\ \bar{I}(M) = U_{I(x) \cap M \neq \emptyset} I(x) = I(S2) \cup I(S3) = \{S1, S2, S3, S5\} \end{cases}$$

$$BUONO \begin{cases} \underline{I}(B) = U_{I(x) \subseteq B} I(x) = I(S4) \cup I(S6) = \{S4, S6\} \\ \bar{I}(B) = U_{I(x) \cap B \neq \emptyset} I(x) = I(S4) \cup I(S6) \cup I(S1) \cup I(S5) = \{S1, S4, S5, S6\} \end{cases}$$

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA): Esempio

Sulla base di queste definizioni se consideriamo P=$\{Matematica, Fisica\}$ come $P \subseteq Q$, si possono calcolare le *P-approssimazioni inferiori* e *superiori* degli insiemi degli studenti con valutazione globale S, M, B:

$$\underline{P}(S) = \{x \in U: Ip(x) \subseteq S\} = \{S7, S8\}$$

$$\overline{P}(S) = \{x \in U: Ip(x) \cap S \neq \emptyset\} = \{S7, S8\}$$

$$\underline{P}(M) = \{x \in U: Ip(x) \subseteq M\} = \{S2, S3\}$$

$$\overline{P}(M) = \{x \in U: Ip(x) \cap M \neq \emptyset\} = \{S1, S2, S3, S5\}$$

$$\underline{P}(B) = \{x \in U: Ip(x) \subseteq B\} = \{S4, S6\}$$

$$\overline{P}(B) = \{x \in U: Ip(x) \cap B \neq \emptyset\} = \{S1, S4, S5, S6\}$$

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA):Esempio

P=$\{Matematica, Fisica\}$

Le *P-frontiere* degli studenti con valutazione globale S, M,B sono:

$B_{np}(S) = \overline{P}(S) - \underline{P}(S) = \{S7, S8\} - \{S7, S8\} = \emptyset$

$B_{np}(M) = \overline{P}(M) - \underline{P}(M) = \{S1, S2, S3, S5\} - \{S2, S3\} = \{S1, S5\}$

$B_{np}(B) = \overline{P}(B) - \underline{P}(B) = \{S1, S4, S5, S6\} - \{S4, S6\} = \{S1, S5\}$

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA):Esempio

Pawlak 1997

Dati sei magazzini descritti dai seguenti quattro attributi:

- A1, capacità del personale di vendita,

- A2, qualità percepita della merce,

- A3, localizzazione ad alto traffico,

- A4, utili o perdite del magazzino.

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA):Esempio

TAVOLA
DELLE
INFORMAZIONI

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

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA):Esempio

Si ha perciò:

$$U\ \{1,2,3,4,5,6\}$$

$$Q\ =\ \{A1, A2\ , A3\ , A4\}$$

V1 ={alta, media, bassa},

V2= {buona, media} ,

V3={no, si} ,

V4 ={utile, perdita},

Tabella =funzione dell'informazione f (x, q) es. f (1, A1) = alta , f (1/ A2) = buona..

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA):Esempio

Considerando tutti gli attributi:

$$Q = \{A1, A2, A3, A4\}$$

Tutti gli elementi sono discernibili.

Formalmente si ha:

$$\boxed{I_Q = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}}$$

RELAZIONI DI INDISCERNIBILITA'

I magazzini 2 e 3 sono indiscernibili $\rightarrow P = \{A1, A2, A3\}$

Formalmente

$$I_P = \{(1,1), (2,2), (2,3), (3,2), (3,3), (4,4), (5,5), (6,6)\}$$

I magazzini 1,2 e 3 e 5 e 6 sono indiscernibili $\rightarrow P = \{A2, A3\}$

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA):Esempio

Ogni $P \subseteq Q$ determina una partizione U/IP:

raggruppa oggetti aventi la stessa descrizione nei termini degli attributi di P.

ES: $P' = \{A2\ , A3\ \}$ si ha U/IP= {{1,2,3},{4},{5,6}} P' = {1,2,3},{4},{5,6} sono gli insiemi P' – elementari.

Descrizione Formale Dei Rough Sets Basati Sulla Indiscernibilità (IRSA):Esempio

Approssimazione attraverso $P = \{A1, A2 , A3\}$ dei magazzini che hanno conseguito un utile, cioè $X = \{1,3,6\}$ considerando che $U/I_P = \{\{1\},\{2,3\},\{4\},\{5\},\{6\}\}$ P = si ha:

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

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

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

- 1 e 6 sicuramente appartengono all'insieme «utili»
- 1, 2, 3, e 6 potrebbero appartenere all'insieme «utili»
- 2 e 3 appartenenza dubbia.

La Teoria dei Rough Sets:
da IRSA a DRSA

La Teoria Rough Set Basata Sui Rapporti Di Dominanza (DRSA)

Il ***nuovo approccio*** alla teoria dei *Rough Set,* che integra al suo interno i rapporti di preferenza espressi dal decisore, viene introdotto da Greco et al. (1997; 1999; 2001).

Alla base vi è la sostituzione della *relazione di indiscernibilità* con la *relazione di dominanza* *(Dominance-based Rough Set Approach: DRSA)* nella definizione delle approssimazioni.

La Teoria Rough Set Basata Sui Rapporti Di Dominanza (DRSA)

Un attributo il cui dominio risulta ordinato secondo le preferenze del decisore, si dice ***criterio.***

L'oggetto $x \in U$ *domina* l'oggetto y $\in U$ se «*x è almeno tanto buono quanto y*», rispetto a tutti i criteri considerati.

Si definisce così una ***relazione binaria di dominanza:*** un profilo domina un altro profilo se presentano entrambi le stesse valutazioni rispetto ai valori sugli attributi e le valutazioni dell'uno sono peggiori di quelli del secondo rispetto alle valutazioni sui criteri.

Il Metodo Rough Set

Il metodo *Rough Set* ha trovato molteplici applicazioni come supporto alla decisione in diversi casi pratici. In particolare, per quelli in cui si richiede di individuare la corretta assegnazione di un insieme di oggetti, descritti da un certo insieme di attributi, rispetto ad una o più categorie predefinite, ovvero quello che, nella letteratura della teoria delle decisioni, viene definito come classificazione *multi-attributo*.