

ASSOCIATIVE CLASSIFICATION.

L'idea di base è quella di cercare forti associazioni fra frequent pattern e classe labeled.

Una regola associativa è spesso rappresentata

$$P_1 \text{ AND } P_2 \dots \text{ AND } P_n \Rightarrow \text{CLASS } C$$

Dove con $P_1 \dots P_n$ si intendono le coppie Attributo-Valore che in congiunzione portano alla classe C.

Questo metodo permette di sfruttare gli attributi insieme e non uno per volta come nel DT.

CBA - CLASSIFICATION BASED ON ASSOCIATIONS

Usa l'approccio di identificare le regole associative nella forma $P_1 \text{ AND } P_2 \dots \text{ AND } P_n \Rightarrow \text{CLASS } C$.

1) Costituzione del modello

Metodo euristico in cui le regole vengono ordinate per ordine decrescente di confidenza e poi di supporto

2) Classificare una nuova tupla

Si sceglie la prima regola che la copre.

Regola di default come ultima spiaggia.

CMAR - CLASSIFICATION BASED ON MULTIPLE ASSOCIATION RULES

Variante di FP-Growth

1) Costituzione del modello

Usa una variante di FP-Growth per trovare un set completo di regole secondo min-sup e min-conf.

2) Pruning

Prese due regole R_1 e R_2 si va a valutare:

- L'antecedente della regola (la parte IF) R_1 è più generale rispetto a R_2 ?
- $\text{Conf}(R_1) \geq \text{Conf}(R_2)$?

In queste condizioni si elimina R_2 .

L'algoritmo passa anche se l'antecedente della regola e la classe della regola sono positivamente correlate, usando χ^2 -test

3) Classificare una nuova tupla

CMAR cerca di trovare il set di regole che soddisfa la nuova tupla. Può capitare che più regole associative soddisfino i criteri, ma che abbiano classe diversa.

A questo punto per sceglierne una si usa un test statistico di indipendenza, weighted χ^2 :

- Le regole vengono divise in gruppi secondo la loro classe
- Si cerca il gruppo che ha correlazione maggiore con la classe (Usando il weighted χ^2)

- Si classifica con la classe del gruppo con χ^2 maggiore

CPAR - CLASSIFICATION BASED ON PREDICTIVE ASSOCIATION RULES

FOL-like algorithm

1) Costruzione del modello

Basare la creazione delle regole su un algoritmo di costruzione delle regole (FOL-like)

L'unica differenza è che in FOL una volta generata la regola, la tuple in questione vengono eliminate, mentre per CPAR vengono solo pesate meno

2) Classificare una nuova tuple

Le regole vengono divise in gruppi per classe etichetta. Le migliori K regole, per accuratezza, per ogni gruppo vengono usate per predire la classe.

Si sceglie la classe che corrisponde al gruppo con l'accuratezza media maggiore.