

DATA MINING

Clustering e Reti di Kohonen

"Data Minig: clustering con Reti di Kohonen"

09/05/03



* Cos'è il Data Mining:

Traduzione letterale:

Data : informazione
To mine : scavare, estrarre

Processo di estrazione di conoscenza (nell'ambito del KDD, knowledge discovery in databases) che sfrutta diverse metodologie statistiche e informatiche, che in particolare permette di trattare GRANDI QUANTITA' DI DATI.

"Data Minig: clustering con Reti di Kohonen"

09/05/03



* DM a confronto con tecniche informatiche e statistiche "tradizionali":

Si affianca agli strumenti informatici di Query e Reporting -> Olap -> DATA MINING

Statistica classica: approccio top-down (analisi confermative) fatti conosciuti a priori

DM: approccio bottom-up (esplorativo) ricerca di informazioni utili spesso non note a priori, nascoste

"Data Minig: clustering con Reti di Kohonen"

09/05/03



Metodologie:

Analisi Associazioni: Basket analysis (Probabilità)

Modelli Previsivi: Regressione lineare (Statistica), Alberi Decisionali e Reti Neurali (Intelligenza Artificiale)

Clustering: Reti di Kohonen (IA) e K-medie (Statistica)

Altri: Algoritmi genetici, Reti bayesiane, Serie temporali...

"Data Minig: clustering con Reti di Kohonen"

09/05/03



* Le fasi di un processo di Data Mining:

- a) Definizione obiettivi
- b) Selezione, organizzazione e trattamento dei dati
- c) Analisi esplorativa dei dati
- d) Scelta del modello applicativo
- e) Elaborazione
- f) Valutazione ed interpretazione dei risultati ottenuti
- g) Ritorno a punto a)

SEMMA: *sample, explore, model, modify, assess.*

"Data Minig: clustering con Reti di Kohonen"

09/05/03



* Principali software suites commerciali:

IBM Intelligent Miner
SPSS Clementine
SAS Enterprise Miner
TERADATA Warehouse Miner

"Data Minig: clustering con Reti di Kohonen"

09/05/03



* Link:

<http://www.kdnuggets.com> (eng)
<http://www.dmreview.com> (eng)
[www://www.web-datamining.net/](http://www.web-datamining.net/) (fra,eng)
<http://open.cineca.it/datamining/dmCineca/> (ita)

"Data Minig: clustering con Reti di Kohonen"

09/05/03



* Testi:

Mastering data mining, *Berry, Linoff* (Wiley)
Knowledge discovery and DM, *Bramer* (Lee Books)
Data mining, *Berry, Linoff* (Apogeo)
Data mining, *Giudici* (Mcgraw-Hill)
DM Webmining e crm, *Tassinari, Camillo* (FrancoAngeli)

"Data Minig: clustering con Reti di Kohonen"

09/05/03



CSI-Piemonte Consorzio per il Sistema Informativo del Piemonte

"Data Minig: clustering con Reti di Kohonen"

09/05/03



La nascita del Data Warehouse della PA

Dal 1996 si è lavorato per proporre alla Pubblica Amministrazione piemontese un cambiamento di architettura nei sistemi informativi: l'obiettivo non è più costruire singoli programmi orientati alla produzione di report statistici, ma *strutturare uno strato informativo organico, storicizzato e periodicamente aggiornato sul quale predisporre funzioni di accesso flessibili e di semplice utilizzo.*

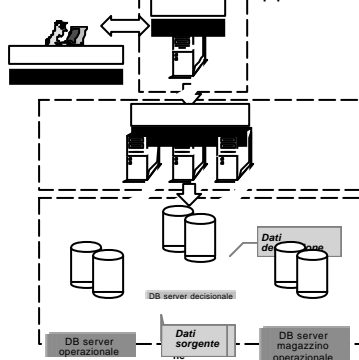
Nascono in questo modo le prime proposte di realizzazione dei Data Warehouse per i singoli Enti

"Data Minig: clustering con Reti di Kohonen"

09/05/03



Architettura informatica supporto decisionale



"Data Minig: clustering con Reti di Kohonen"

09/05/03



Base dati decisionale delle prescrizioni farmaceutiche della REGIONE PIEMONTE

"Data Minig: clustering con Reti di Kohonen"

09/05/03



Le domande a cui rispondere

Una base dati decisionali per estrarre quali informazioni ?

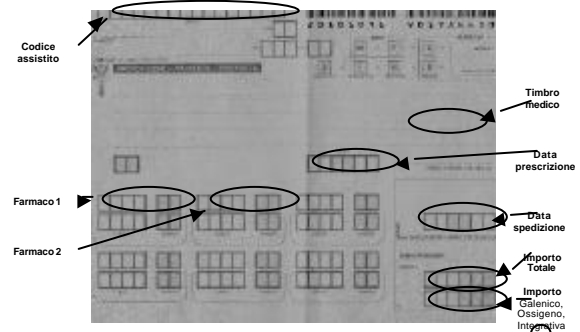
- ✓ 1- Come sta incrementato la spesa rispetto all'anno scorso ?
- ✓ 2- Quale profilo di popolazione contribuisce di più alla spesa ?
- ✓ 3- Qual è l'antibiotico più prescritto dai pediatri ? In estate e in inverno ?
- ✓ 4- Qual è la ripartizione sul territorio degli assistiti rispetto alla spesa ?
- ✓ 5- Qual è il principio attivo più prescritto ?
- ✓ 6- Come si diffondono i farmaci generici nel tempo?
- ✓ 7- Come varia la spesa media per assistito secondo la fascia di età ?
- ✓ 8- Per quali patologie abbiamo la spesa per assistito più elevata ?

"Data Minig: clustering con Reti di Kohonen" 09/05/03



L'universo di riferimento

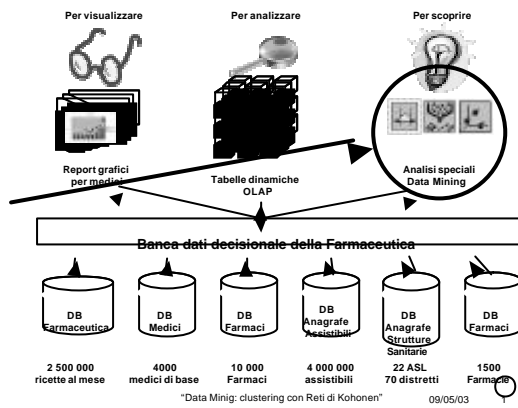
L'insieme delle ricette di prescrizioni farmaceutiche erogate nelle farmacie del Piemonte il cui pagamento dell'importo totale o parziale viene sostenuto dal SSN



"Data Minig: clustering con Reti di Kohonen" 09/05/03



Evoluzioni verso il Data mining



"Data Minig: clustering con Reti di Kohonen" 09/05/03



Data mining

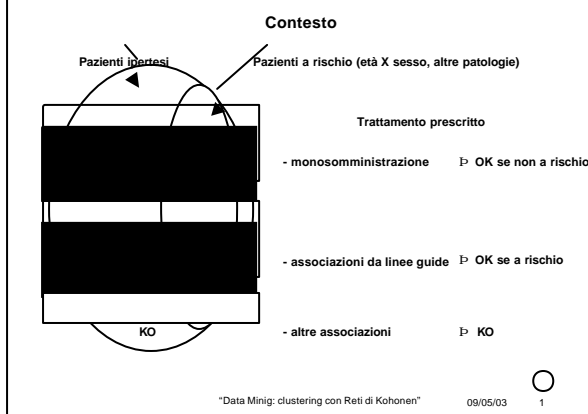
Analisi modelli prescrittivi piemontesi trattamento farmacologico dell'ipertensione

- **Obiettivo :**
 - raggruppare i medici con profili simili nel trattamento farmacologico prescelto (monotratamento, trattamenti combinati...) nel trattamento dell'ipertensione.
 - Verificare in quale modo le linee guida OMS per il trattamento dell'ipertensione sono rispettate.
- **Periodo di osservazione :**
 - primo semestre 2002
- **Soggetti :**
 - Medici di base generici

"Data Minig: clustering con Reti di Kohonen" 09/05/03



Data mining



"Data Minig: clustering con Reti di Kohonen" 09/05/03



Data mining

Metodologia seguita

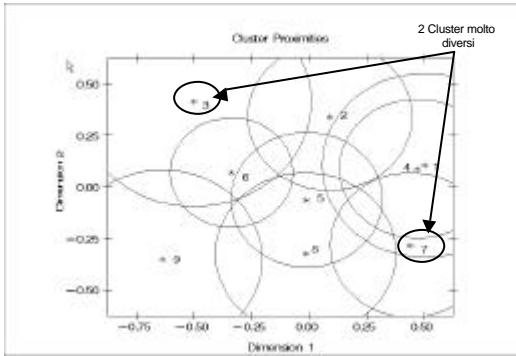
- **Fonte :** Base Table semestrale delle singole prescrizioni
- **Filtro :** prescrizione di Farmaci di principio attivo C02,C07,C08 ...
- **Aggregazione e calcolo di variabili ad hoc al livello del medico di base :**
 - N° pazienti ipertesi
 - N° pazienti curati con linea guida (associazione giusta di farmaci)
 - N° pazienti per ogni gruppo terapeutico
 - N° pazienti per ogni associazione di gruppo terapeutico
 - N° pazienti a rischio
 - N° confezioni / assistito per ogni principio attivo ...
- **Aggiunta di variabili supplementari**
 - N° assistiti totale
 - anni di anzianità, sesso
 - localizzazione geografica (ASL, campagna/città)
- **Filtro di medici outliers :** con meno di 30 assistiti, meno di 100 ricette...
- **Definizione di variabili Target :**
 - peso dei pazienti con comportamento giusto (buona condotta)
 - distinzione per pazienti a rischio o no

"Data Minig: clustering con Reti di Kohonen" 09/05/03



Data mining / profili dei medici

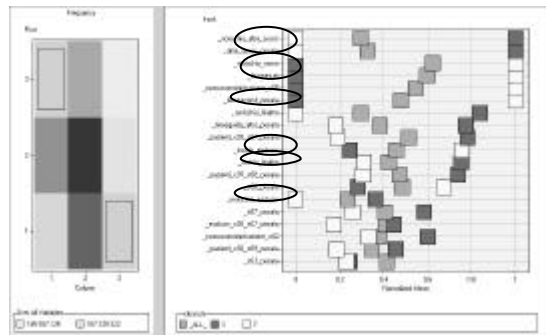
Risultati cluster analysis con Mappe di Kohonen



"Data Minig: clustering con Reti di Kohonen" 09/05/03 1

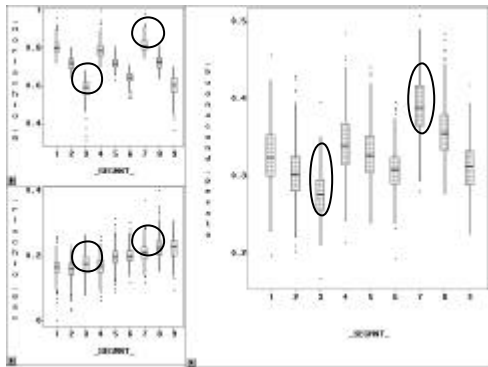
Data mining / profili dei medici

Risultati cluster analysis con Mappe di Kohonen



"Data Minig: clustering con Reti di Kohonen" 09/05/03 1

Data mining / profili dei medici



"Data Minig: clustering con Reti di Kohonen" 09/05/03 1

Data mining / profili dei medici

Cluster 7 (300 medici)

Comportamento : segue linee guida monosomministrazione per pazienti non a rischio, combinazioni esclusive per pazienti a rischio

Profilo medio del medico : medico di "campagna"

Cluster 3 (150 medici)

Comportamento : non segue linee guida, in particolare usa associazioni non consigliate, poco monosomministrazione

Profilo medio del medico : pochi anni attività 0-10, medico di "città"

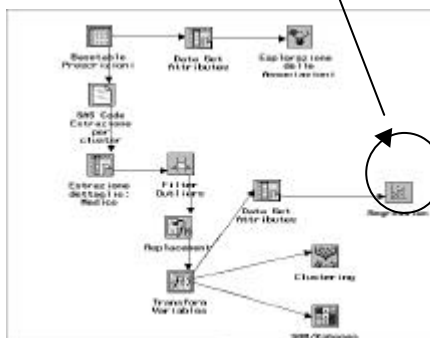
Tipo paziente : niente di particolare(età, numero pazienti ipertesi, ipertesi a rischio)

Azione di informazione specifica

"Data Minig: clustering con Reti di Kohonen" 09/05/03 1

Data mining

Prossimi passi : modello esplicativo



"Data Minig: clustering con Reti di Kohonen" 09/05/03 1