

Data Mining - Il processo KDD

Federico Calò

Contents

1	Lezioni	4
1.1	1 - Introduzione Sistemi basati su conoscenza	4
2	Libro	6

Intro

Il sito a cui si riferisce il professore è Artificial Intelligence: Foundations of Computational Agents, second edition, Cambridge University Press. Del libro vi è presente una versione online presente al seguent [link](#). Gli appunti sono organizzati in due macro sezioni: gli appunti delle lezioni e gli appunti derivanti dal libro.

L'obiettivo del corso è quello di acquisire le competenze necessarie alla specifica e progettazione i sistemi intelligenti basati su conoscenza, attraverso la padronanza degli aspetti teorici modelli logico e matematico-statistici e la capacità implementative, di valutazione e miglioramento di sistemi esistenti nei diversi domini applicativi. Inoltre ci si approccerà a problemi complessi mediante modelli di rappresentazione e ragionamento:

- formalismi per la rappresentazione della conoscenza di dominio (proposizionale, multi-relazionale)
- forme i ragionamento automatico
- tecniche di acquisizione della conoscenza (modelli per la predizione) basate sull'apprendimento automatico
- tecniche di valutazione
- modelli di ragionamento/acquisizione in presenza di incertezza: approccio probabilistico esteso anche a rappresentazioni multi-relazionali

Per quanto riguarda i capitoli del libro di cui si andrà a prendere spunto per le trattazioni sono:

- Introduzione ai sistemi basati su conoscenza [1-2]
- Fondamenti - Problemi e Ricerca delle soluzioni [4]
- Rappresentazione della conoscenza (Proposizionale) [5]
- Rappresentazione di individui e relazioni [13]
- Ontologie e BAsi di conoscenza distribuite [14]
- Ragionamento in presenza di incertezza [8]
- Apprendimento Supervisionato [7]
- Apprendimento e incertezza [10]
- Modelli probabilistici relazionali [15]

1 Lezioni

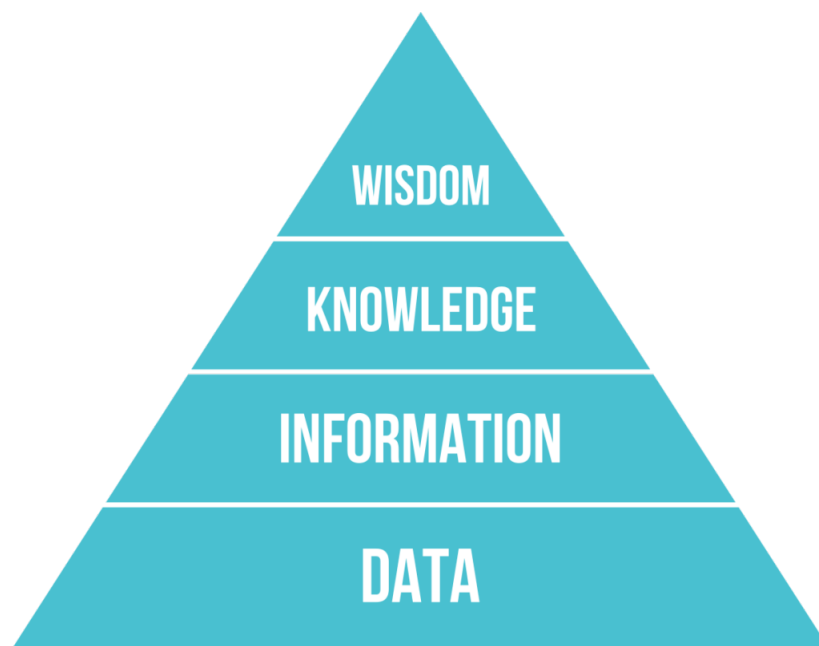
1.1 1 - Introduzione Sistemi basati su conoscenza

Vi sono diverse definizioni dell'ingegneria della conoscenza, una è la seguente:

”L'ingegneria della conoscenza (KE) si riferisce a tutti quegli aspetti tecnici, scientifici e sociali che sono coinvolti nella costruzione, mantenimento e uso dei sistemi basati su conoscenza”

Tipicamente l'informazione è definita in termini di dati, conoscenza in termini di informazioni e saggezza in termini di conoscenza. Molto importante è la **piramide DIKW** (gerarchia della conoscenza) che consiste in una piramide formata da 4 livelli, dal più basso:

- Dati (Data)
- Informazioni (Information)
- Conoscenza (Knowledge)
- Sapere (Wisdom)



I **dati** sono costituiti da segni o simboli che rappresentano simboli o segnali che rimangono inutili fino quando non sono messi in una qualche forma. Possono essere *universali* se sono prodotti da osservazioni, oppure *soggettivi* costituite dalle osservazioni stesse. Possono essere costituiti da *fatti*, cioè da osservazioni discrete, oggettive, non organizzate o elaborate che mancano di contesto interpretativo; da *segnali*, cioè da stimoli sensoriali o letture di segnali attraverso sensi o sensori; o *simboli*, cioè da insiemi di segni che rappresentano le percezioni di proprietà di oggetti, eventi dell'ambiente e che vengono registrati al fine della comunicazione.

Le **informazioni** sono dati dotati di significato e scopo, ottenute per descrizione e distinte dai dati per la loro utilità. Vengono inferite dai dati rispondendo a specifiche domande o rendendoli utili a prendere decisioni.

La **conoscenza** è definita come informazione elaborata, organizzata, o altrimenti applicata, messa in atto attraverso una commistione di esperienza sistematizzata, valori, informazioni contestuali, comprensione profonda e ben fondata. Fornisce un *ambiente* e una struttura per la valutazione e l'acquisizione di nuove esperienze e informazioni nei singoli agenti, nei quali origina informazione e viene applicata a livello mentale, e nelle organizzazioni, spesso incorporata non solo attraverso documenti anche in senso esteso, e sistemi di memorizzazione, ma anche nelle procedure organizzative. Si possono definire inoltre diversi tipi di conoscenza:

- **conoscenza elaborata** costituita dalla sintesi di più sorgenti di informazioni nel tempo, dalla comprensione, esperienza e apprendimento generati dall'organizzazione ed elaborazione dell'informazione. L'informazione è connessa attraverso relazioni di contesto, valori, esperienza e regole.
- **conoscenza procedurale** definita come conoscenza raggiunta attraverso un'esperienza pratica, quindi attraverso azioni e non attraverso delle descrizioni. In questo tipo di conoscenza si applicano dati e informazioni.
- **conoscenza proposizionale** descritta come strutturazione delle credenze e internalizzazione. In termini proposizionali la conoscenza può diventare a sua volta informazione, invece soggettivamente la conoscenza può essere costituita da un pensiero caratterizzato dalla credenza, che essa sia vera o meno, empirica o non empirica (logica, matematica, filosofia).

A questo punto possiamo dare una definizione di **Intelligenza Artificiale (AI)**:

"Disciplina che mira a studiare e comprendere i principi che rendono possibile un comportamento intelligente in sistemi artificiali."

quindi rendere vera l'equazione: ragionamento \approx computazione. Un'ipotesi collegata è quella legata alla tesi di Church-Turing in cui nel quale il livello di astrazione del ragionamento equivale alla manipolazione dei simboli descritti come azioni/decisioni di un sistema spiegate in termini dei suoi input. Si mira a metodi per la progettazione di artefatti SW intelligenti, utili a scopi precisi.

Intelligenza dei sistemi è diversa da quella umana. Una comunità organizzata può esibire un comportamento intelligente, in quest'ambito si parla di **visione olistica** quando organizzazioni complesse, costituite da singole unità non particolarmente intelligenti, ma che nel loro complesso riescono a esibire un comportamento più intelligente.

L'AI come scienza mira a comprendere i principi del ragionamento secondo il metodo scientifico, attraverso il quale si dovrebbero creare e verificare teorie sulla soluzione algoritmica di problemi d'interesse supportate da implementazioni, attraverso anche la verifica sperimentale.

L'AI vista come disciplina ingegneristica, tesa a costruire tecnologie/sistemi che risolvano specifici problemi come quello di creare e testare sistemi software intelligenti basati su conoscenza.

2 Libro

s