

# CAPITOLO 3 – Previsione

## 3.1 La previsione

I sistemi evoluti, che apprendono le regole di funzionamento attraverso l'interazione con l'ambiente, si rivelano una risorsa essenziale nella rappresentazione di domini complessi e mutevoli.

La necessità di reperire criteri guida attraverso i quali gestire le quantità di risorse in Internet, a quanto pare sempre meno sufficienti e adoperate spesso in modo inefficiente, e la disponibilità di grande potenza di calcolo a basso costo stanno determinando la maturazione e il passaggio di questi meccanismi dall'ambito della sperimentazione a quello dell'applicazione reale.

Le soluzioni basate su **information retrieval** estraggono o sintetizzano la conoscenza da una massa di informazioni caotiche e ridondanti. Il processo molto delicato di estrazione della conoscenza da banche dati di grandi dimensioni avviene tramite l'applicazione di algoritmi che individuano le associazioni, o *patterns*, "nascoste nei dati". L'obiettivo finale è quello di riuscire ad individuare le informazioni più significative e renderle disponibili e utilizzabili nel processo di *decision making*.

### 3.1.1 Sistemi predittivi

La maggior parte della ricerca sulle architetture è stata fatta sui sistemi di raccomandazione, ovvero che forniscono suggerimenti, basati su algoritmi di "*machine learning*", branca dell'intelligenza artificiale che si occupa della progettazione e dello sviluppo di algoritmi che consentono di sviluppare comportamenti sulla base di dati empirici, o "*data mining*", meccanismo che consente di creare modelli a partire dall'analisi di un set di dati con lo scopo di cercare tendenze e modelli specifici.

La parte più importante di uno schema di predizione è l'algoritmo predittivo che possono essere: point based o path based.

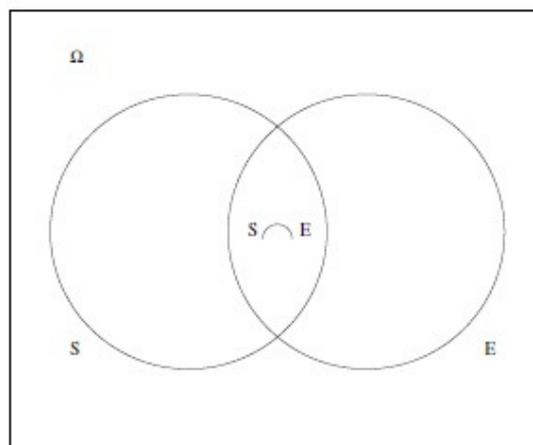
I modelli point based usano le azioni indicizzate nello stesso istante per predire la prossima azione basandosi su quella corrente. In altre parole usano la misura delle frequenze per predire le prossime richieste. Questo tipo di modelli considerano un numero relativamente piccolo di informazioni per ogni sessione ma, a loro svantaggio, gioca il fatto che tali previsioni possono essere inaccurate.

I modelli path based sono basati sul percorso di navigazione precedente dell'utente, essi realizzano un'analisi statistica delle azioni degli utenti e per questo possono avere una precisione relativamente alta.

Uno schema molto utilizzato nelle previsioni è quello basato sul **modello delle catene di Markov**, utilizzate soprattutto per predire i movimenti degli utenti mobili in modo da poter ottenere un sensibile miglioramento nelle prestazioni globali della rete, riducendo, per esempio, la latenza dell'utente.

### 3.1.2 Valutazione dei sistemi predittivi

Un modello di predizione, solitamente, viene validato attraverso la valutazione di due parametri: *accuracy* e *recall*. Queste due parametri sono stati definiti nell'ambito di tecniche di information retrieval e data mining. Il data mining è un processo di esplorazione ed analisi, con strumenti automatici o semi-automatici, di grandi quantità di dati al fine di scoprire strutture e regole significative. L'information retrieval è un insieme di tecniche che consentono un accesso mirato ed efficiente a grandi raccolte di dati.



**Figura 3.1.** Diagramma di precisione e applicabilità.

Il significato dei parametri introdotti precedentemente viene esplicito in maniera molto chiara ed evidente dal diagramma presente nella fig. 3.1, che prende il nome di diagramma di Venn, con:

- $\Omega$ : collezione da interrogare, ovvero l'informazione base.
- $E$ : sottoinsieme di  $\Omega$  con le soluzioni corrette dell'interrogazione
- $S$ : sottoinsieme di  $\Omega$  con le soluzioni date dal sistema.

Da questi sottoinsiemi è facile dare una definizione dei parametri di *recall*  $r$  e di *accuracy*  $p$ , dove il primo è pari al rapporto tra  $S \cap E$  con  $S$  mentre il secondo è pari al rapporto tra  $S \cap E$  con  $E$ .

Nell'ambito dei sistemi che tentano di prevedere, per esempio, la prossima cella di locazione di un utente mobile lo schema predittivo risulta essere più semplice perché non si richiede una collezione di soluzioni, ma vi è una sola soluzione che può essere corretta o non corretta. Utilizzando le definizioni fornite precedentemente in questo nuovo contesto sono stati definiti due parametri di precisione e applicabilità, dove *l'applicabilità* rappresenta la frequenza con cui il sistema è capace di predire e *la precisione* è la frequenza con cui queste predizioni sono corrette.

Da queste definizioni si può arrivare a dire che la precisione  $p$  è definita come il rapporto tra il numero di predizioni corrette e il numero di predizioni effettuate:

$$Precisione = p = \frac{N_{PredizioniCorrette}}{N_{PredizioniEffettuate}} \quad (3.1)$$

L'applicabilità  $r$  invece è definita come il rapporto tra il numero di predizioni effettuate sul numero di predizioni richieste:

$$Applicabilità = r = \frac{N_{PredizioniEffettuate}}{N_{PredizioniRichieste}} \quad (3.2)$$

La precisione  $p$  è pari a 1 quando tutte le predizioni proposte dal sistema sono corrette. L'applicabilità  $r$  è pari a 1 quando il sistema effettua sempre previsioni.

Un altro parametro di valutazione è l'***F-Measure***. Questo parametro è un indice di bontà dell'algorithmo predittivo ed è definito come una combinazione lineare dei due parametri precedenti, dando la possibilità, con un singolo valore, di poter descrivere la qualità del sistema predittivo.

La *F-Measure*  $f$  combina l'applicabilità  $r$  e la precisione  $p$  secondo la seguente formula:

$$F - Measure = f = \frac{2 \cdot p \cdot r}{r+p} \quad (3.3)$$

Tale parametro è una media armonica. Una media armonica  $H$  di  $n$  numeri  $x_1, x_2, \dots, x_n$  è definita come:

$$H = \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \quad (3.4)$$

Sfruttando questa formula per l'applicabilità  $r$  e la precisione  $p$ , si ottiene:

$$H = \frac{1}{\frac{1}{2} \left( \frac{1}{r} + \frac{1}{p} \right)} = \frac{2}{\frac{1}{r} + \frac{1}{p}} = \frac{2rp}{r+p} \quad (3.5)$$

Tale equazione rappresenta la *F-Measure*, che può variare tra 0 e 1, in cui 1 rappresenta il caso ottimale ovvero il caso in cui sia l'applicabilità che la precisione sono pari a 1.

### 3.1.3 Motivazioni

Le reti con utenti mobili, come le reti wireless di cui si è ampiamente parlato nei precedenti capitoli, introducono dei cambiamenti nella garanzia di QoS. Tali reti sono caratterizzate dal cosiddetto fenomeno di handover o handoff, ovvero quando un utente mobile si associa ad un nuovo AP lasciando, a causa della sua mobilità, il suo AP corrente entrando in uno appena vicino.

La maggior parte delle procedure che consentono di effettuare il passaggio da un AP ad un altro si basano tutti su algoritmi che sfruttano l'analisi e il confronto, con diverse tecniche di raffronto e con l'introduzione di parecchi parametri, tra la potenza del segnale dell'AP corrente e quella riguardante tutti gli AP vicini. Tali procedure, a causa della loro complessità e della grande quantità sia di tempo necessario all'analisi che di dati di confronto, rendono tutte queste tecniche molto lente e poco

adattabili in reti il cui fine primario è offrire velocità e un determinato livello di supporto di QoS.

Infatti, se le procedure di handoff sono lente si incappa nella perdita di pacchetti e di forti ritardi di trasmissione. Invece se si anticipa la conoscenza del prossimo AP a cui l'utente mobile dovrà associarsi, questo nuovo AP potrà prepararsi ad accogliere il nuovo utente, per esempio riservando le risorse necessarie per ospitarlo, riducendo i ritardi provocati dalla procedura di handoff.

Un altro motivo per cui vengono utilizzati gli algoritmi di previsione riguarda il tentativo di evitare, o almeno ridurre al minimo, il numero di applicazioni terminate a causa della scarsa disponibilità di risorse necessarie per l'esecuzione di tali applicazioni, provocando lo spiacevolissimo evento conosciuto come "call drop". Tra le applicazioni che possono essere vittime di tale evento ci sono le applicazioni multimediali in real-time, come una videoconferenza. Tali applicazioni richiedono, come già visto in precedenza, l'assicurazione di un minimo di banda ad essa associata. Se tale banda non può essere offerta a tale applicazione da parte della rete, la connessione, con l'applicazione annessa, cadrà, tale situazione è molto più sgradevole rispetto al caso in cui la chiamata venga bloccata sul nascere, in quanto è peggio dover rinunciare ad un servizio di cui si stava già facendo uso, piuttosto che rinunciare ad un servizio che, ancora, non ha avuto inizio [27].

Prima che un dispositivo passi al prossimo AP, prevedere la destinazione e il tempo di handoff aiutano la rete a prenotare le risorse sufficienti, come per esempio la banda, riducendo la probabilità di trovarsi di fronte ad un evento di call drop, migliorando così anche il supporto ai QoS.

Per tutti questi motivi sono ampiamente utilizzati vari tipi di algoritmi predittivi per le principali questioni di gestione della rete come:

- Prenotazioni delle risorse
- Routing
- Allocazione anticipata della banda.

Per riuscire ad acquisire tali capacità predittive vengono utilizzati sia algoritmi predittivi su base statistica, come per esempio si utilizzano le catene di Markov come accennato precedentemente, oppure si fa uso di uno strumento più potente ed efficace: **l'Intelligenza Artificiale (IA)**.

Di seguito verrà proposta una panoramica generale sull'IA e sui campi di diretta applicazione nelle reti wireless come:

- **Swarm intelligence**
- **Neural Network.**

### **3.2 Intelligenza Artificiale (IA)**

Il termine Intelligenza Artificiale, che fu coniato da Jhon McCarthy circa una trentina di anni fa, non ha una definizione precisa e formale, per questo motivo infatti il primo passo è quello in direzione di una definizione non ambigua dell'IA che, per sua natura, si presta a svariate interpretazioni e a qualche fraintendimento. L'IA è una moderna disciplina che negli anni recenti ha fornito un'importante contributo al progresso dell'intera informatica. Per questo motivo riveste un indubbio interesse per coloro che siano affascinati dalle potenzialità dell'informatica e degli effetti che il progresso di questa disciplina ha nei confronti dell'uomo. Inoltre l'IA è stata influenzata da numerose discipline fra le quali, oltre ovviamente all'informatica, la filosofia, la matematica, l'economia, le neuroscienze, la psicologia, la cibernetica, le scienze cognitive e la linguistica.

L'IA è quella disciplina, appartenente all'informatica, che studia i fondamenti teorici, le metodologie e le tecniche che permettono di progettare sistemi hardware e software capaci di fornire all'elaboratore elettronico delle prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana [28]. E' questa la definizione, accettata a livello internazionale, di un'area scientifica di notevole importanza strategica che si propone di superare le attuali frontiere dell'informatica, ponendosi come obiettivo scientifico quello di migliorare ed estendere le prestazioni offerte dall'elaboratore elettronico. Si osservi che l'IA è una disciplina sperimentale, nel senso che un sistema di IA soddisfa gli obiettivi richiesti solamente quando le prestazioni desiderate sono effettivamente misurabili.

L'IA ha come obiettivo quello di riprodurre o emulare l'intelligenza umana. Le emulazioni di prestazioni di intelligenza umana si ottengono grazie a meccanismi

propri delle macchine che replicano gli stessi risultati ottenuti in termini di qualità ma maggiori in termini di quantità.

L'IA è considerata sia una scienza che un'ingegneria [30]. È considerata una scienza nel momento in cui vengono simulati, con determinati sistemi artificiali, alcuni dei comportamenti intelligenti (con intelligente, qui e in seguito, ci si riferisce alla caratteristica propria dell'uomo, come già detto in precedenza), in questo modo l'uomo realizza l'obiettivo di modellare e sperimentare promettenti risultati che consentono di ottenere un indiscutibile progresso nello studio dell'intelligenza umana. È, invece, un'ingegneria in quanto, ottenendo dalle macchine prestazioni che simulano e che in precedenza erano ritenuti inaccessibili alle macchine, si fornisce un progresso al contributo che l'ingegneria offre al miglioramento della vita umana.

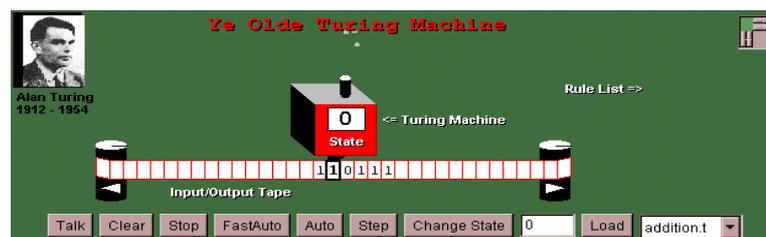
Nonostante tali definizioni ancora oggi rimane il problema aperto che nessuno ha definito in maniera oggettiva, o per lo meno ampiamente condivisa, che cosa sia l'intelligenza stessa. In generale, si intende per intelligenza la capacità di completare compiti e risolvere problemi nuovi (*problem solving*), eventualmente sorti durante il completamento del compito originale, nonché la capacità di adattarsi all'ambiente e ai suoi mutamenti. A tale definizione si vuole aggiungere anche la capacità di comprendere la realtà, qualunque essa sia.

Storicamente si sono distinte due tipi di intelligenza artificiale:

- **Intelligenza artificiale forte:** la quale ritiene che un computer opportunamente programmato possa veramente essere dotato di una genuina intelligenza, non distinguibile dall'intelligenza umana. Alla base di questa concezione vi è la teoria che anche la mente umana sia il prodotto di un complesso insieme di procedimenti di calcolo, eseguiti dal cervello. Resta naturalmente da stabilire rispetto a quale livello di descrizione cervelli e computer possano essere considerati la stessa cosa. La tecnologia che rappresenta tale tipologia di intelligenza nel miglior modo possibile è quella dei **sistemi esperti**, che, con estrema sintesi, è un programma capace di risolvere problemi complessi che rientrano in un particolare dominio con una efficienza paragonabile a quella di un esperto umano. Un IA di questo tipo è una intelligenza artificiale capace di pensare e avere una mente.
- **Intelligenza artificiale debole:** essa sostiene che un computer opportunamente programmato possa solamente *simulare* i processi cognitivi umani (o alcuni di essi), nello stesso senso in cui un computer può simulare il

comportamento di un evento atmosferico, o di una fissione nucleare. In questo caso dunque l'accento viene posto su ciò che un programma è in grado di fare, senza fare assunzioni sul fatto che il modo in cui lo fa coincida ad un qualche livello con i processi mentali umani. E d'altra parte non si pretende che un programma, per quanto complesso ed efficiente, possa veramente dirsi un esempio di mente. Questa tipologia di IA è considerata come un'intelligenza artificiale che agisce come se pensasse e avesse una mente.

Parlare di intelligenza artificiale implica, come minimo, il menzionare uno dei padri del calcolo automatico ovvero Alan Turing. Egli nel 1936 sviluppò le basi teoriche dell'informatica introducendo un modello di macchina calcolatrice programmabile, divenuta nota ai posteri con il nome di "macchina di Turing". Nel 1950 pubblicò la sua teoria con l'articolo "Computing Machinery and Intelligence" passato alla storia. Obiettivo di Turing era quello di risolvere alcuni complicati problemi logici che erano di grande interesse a quei tempi, ma ciò che fu veramente interessante fu il modo che Turing utilizzò. Egli immaginò una specie di macchina teorica, molto semplice, che facendo delle banali operazioni doveva essere in grado di compiere qualunque tipo di calcolo, di qualsiasi genere. Tale macchina era costituita da un dispositivo, una specie di testina, in grado sia di scrivere che di leggere simboli su un nastro. Il nastro era diviso in celle ognuna delle quali contiene un solo simbolo. Tale testina, oltre a scrivere, poteva spostarsi a destra o a sinistra una cella alla volta.



**Figura 3.2.** simulazione macchina di Turing.

Il suo comportamento avveniva secondo passi successivi, ogni passo era governato da regole che tengono conto del simbolo presente sulla cella e che la testina legge. Specificando tali regole la macchina di Turing era capace di calcolare moltissime cose, ovvero se esistono delle regole che indicano come manipolare tali simboli,