
INDICE

<i>INTRODUZIONE</i>	1
<i>Capitolo 1: RETI WIRELESS</i>	5
1.1 Reti Wireless.....	5
1.2 Lo standard IEEE 802.11	10
1.3 Le Vanet.....	14
1.4 LTE (Long Term Evolution).....	19
1.5 5G	21
<i>Capitolo 2: RETI CELLULARI</i>	24
2.1 Reti Cellulari	24
2.2 Lo standard GSM.....	28
2.3 Lo standard UMTS.....	33
2.4 Il WiFi	39
2.5 WiMax	45
<i>Capitolo 3: ALGORITMI DI PREVISIONE</i>	51
3.1 Introduzione.....	51
3.2 Processi Markoviani	51
3.3 Reti Neurali.....	53
3.4 Swarm Intelligence	57
3.5 Sequential Pattern Mining	59
3.6 Algoritmi per estrarre Pattern Sequenziali	61
3.6.1 <i>Spirit</i>	65
3.6.2 <i>cSpade</i>	66
3.6.3 <i>Prefix-Growth</i>	69

3.6.4	<i>GenPrefix-Growth</i>	70
3.6.5	<i>SLPMiner</i>	72
3.6.6	<i>Apriori-based</i>	75
3.6.7	<i>Pattern-Growth</i>	76
 Capitolo 4: DATA MINING		78
4.1	Introduzione.....	78
4.2	Il processo KDD	79
4.3	Data Warehouse.....	80
4.4	Selezione e pre-processing	83
4.5	Data Mining.....	84
4.5.1	<i>Regole di associazione</i>	84
4.5.2	<i>Classificazione</i>	87
4.5.3	<i>Clusterizzazione</i>	90
4.6	Workflow Mining	94
4.6.1	<i>Data Mining su grafi</i>	95
4.6.2	<i>Process Mining</i>	98
 Capitolo 5: SIMULAZIONE E RISULTATI		101
5.1	Introduzione.....	101
5.2	FreeSpan: Frequent Pattern-Projected Sequenzial Pattern Mining	102
5.3	PrefixSpan: Prefix-Projected Sequenzial Pattern Mining	103
5.4	Predizione di spostamento	108
5.5	Risultati.....	113
 CONCLUSIONI		122
 BIBLIOGRAFIA		124

Introduzione

L'obiettivo del seguente lavoro di tesi è quello di studiare ed analizzare tutte le tecnologie e le metodologie presenti in letteratura con lo scopo di sviluppare una tecnica basata su data mining in grado di eseguire una previsione delle mobilità utente e di fornire ,quindi, qualità del servizio all'interno di reti cellulari.

In primo luogo, abbiamo incentrato lo studio sulle tecnologie wireless soffermandoci, in particolar modo, sugli standard appartenenti alla famiglia IEEE 802.11 e sulle diverse tipologie di rete come ad esempio WLAN, WPAN, WAN, BWA.

La WLAN (Wireless Local Area Network) rappresenta la soluzione più diffusa e più semplice per configurare una rete LAN (Local Area Network) domestica, perché dà la possibilità, agli utenti, di muoversi in libertà all'interno del raggio di copertura della rete. Le attuali reti WLAN si basano sullo standard IEEE 802.11, noto commercialmente come Wi-Fi.

La WPAN (Wireless Personal Area Network) rappresenta, invece, una PAN (Personal Area Network), che consente la comunicazione tra i dispositivi, con l'ausilio di onde radio al posto del cablaggio fisico. Basata sullo standard 802.15, consente lo scambio di informazioni tra vari dispositivi, avendo però un raggio di azione di pochi metri. Le tecnologie utilizzate dalle WPAN sono essenzialmente: Bluetooth, Infrared Data Association e Wi-Fi.

La WAN (Wide Area Network) è caratterizzata da un'interconnessione di dispositivi a livello geografico, avendo quindi un'estensione territoriale pari a quello di una regione geografica. La più grande rete WAN mai realizzata è, appunto, la rete Internet.

La BWA (Broadband Wireless Access) è una tecnologia il cui scopo è quello di fornire accesso alla rete Internet ad alta velocità di trasmissione wireless, coprendo grandi aree. Le tecnologie usate per fornire il BWA includono LMDS, MMDS e WiMAX.

Lo studio, sugli standard della famiglia 802.11, è stato focalizzato sui quattro protocolli riguardanti la trasmissione dei dati (802.11, 802.1b, 802.11g, 802.11n), mentre per la sicurezza è stato descritto l'802.11i. Sono stati, inoltre, descritti gli standard 802.11c, 802.11d, 802.11e, 802.11f, 802.11h, i quali gestiscono estensioni dei servizi basilari ed introducono miglioramenti dei servizi già disponibili.

Successivamente alle reti wireless, sono stati introdotti gli studi riguardanti le reti a copertura cellulare quali ad esempio GSM, UMTS, LTE ed il futuro 5G, dando risalto alle tecniche di accesso al mezzo condiviso ed al concetto di cella e relativi cluster di celle con riuso dei canali.

Riguardo la copertura del territorio mediante la suddivisione dell'area geografica in celle ed agendo sulla loro dimensione, abbiamo visto come sia possibile variare la densità geografica degli utenti serviti, notando una diminuzione della densità geografica all'aumentare del raggio della cella R.

La descrizione sugli standard GSM è risultata essere particolarmente importante in quanto, a differenza dei sistemi precedenti come ad esempio il TACS, grazie all'introduzione della tecnologia digitale, consente l'attivazione di nuovi servizi come la trasmissione dei dati, posta elettronica, fax digitale, teleconferenza, trasferimento di chiamata ed il servizio SMS (Short Message Service), oltre a garantire una superiore densità geografica di utenze.

Per quanto concerne gli standard successivi al GSM, come ad esempio l'UMTS, sono state descritte le differenti classi di servizio disponibili, la maggiore velocità in fase di trasmissione dei dati, la possibilità di avere traffico simmetrico ed asimmetrico e la tecnica di accesso di tipo CDMA. L'introduzione di questa nuova tecnica wideband CDMA (W-CDMA), insieme alle tecniche di codifica variabile OVSF (Orthogonal Variable Spreading Factor), consentono di realizzare connessioni a velocità differenti per ciascun utente e, tramite l'utilizzo di servizi multipli contemporaneamente, garantiscono la qualità del servizio (QoS) in ogni momento della connessione.

A proposito degli standard successivi all'UMTS, sono state descritte le caratteristiche relative allo standard LTE (Long Term Evolution), in quanto si colloca come ponte tra la generazione 3G e la quarta generazione 4G. Quest'ultimo, risulta essere un arricchimento del precedente standard UMTS in termini di aumento di velocità trasmissiva dei dati, di utilizzo della modulazione OFDM (Orthogonal Frequency-Division Multiplexing) per quanto riguarda il downlink, e della modulazione FDMA (Frequency-Division Multiple Access) per l'uplink.

Successivamente allo studio sulle tecnologie wireless ed a copertura cellulare, abbiamo incentrato la nostra attenzione su algoritmi e metodologie riguardanti la previsione di eventi futuri. A tal proposito, sono stati analizzati approcci basati su processi Markoviani, Reti Neurali, Swarm Intelligence ed algoritmi basati sul concetto di sequential pattern mining, come ad esempio Spirit, cSpade, Prefix-Growth, SLPMiner, Pattern-Growth e Apriori-Based.

In particolare, un processo Markoviano è caratterizzato dall'aver una probabilità di transizione, che determina il passaggio da uno stato di sistema ad un altro, dipendente unicamente dallo stato di sistema immediatamente precedente, e non dalla storia passata di come si è giunti a tale stato. Un caso particolare, di questo tipo di processo, è rappresentato dalla catena di Markov, caratterizzata dall'aver uno spazio degli stati che possono assumere qualsiasi valore in uno spazio discreto. L'insieme S degli stati può essere finito o infinito, a tempo-continuo oppure a tempo-discreto.

Le Reti Neurali, invece, rappresentano interconnessioni di semplici elementi, chiamati nodi o neuroni, dove l'informazione in essi contenuta in un dato istante viene prodotta dal pattern di attivazione dei nodi presenti sulla rete. Uno dei più importanti algoritmi di apprendimento basati su reti neurali, è sicuramente l'algoritmo di apprendimento Hebbiano. Quest'ultimo si basa sul semplice principio che, se due neuroni si attivano contemporaneamente, la loro interconnessione deve essere rafforzata. Riguardo le reti neurali sono stati descritti, inoltre, i tre paradigmi di apprendimento, quali: apprendimento supervisionato (*supervised learning*), apprendimento non supervisionato (*unsupervised learning*) ed apprendimento per rinforzo (*reinforcement learning*).

Focalizzando l'attenzione sulla swarm intelligence, conosciuta anche come teoria dello sciame intelligente, abbiamo descritto come questa tecnica sia in grado di dare origine ad un'azione complessa come risultato di una piccola azione collettiva. L'adozione di tale teoria ha dato luogo, ai fini pratici, ad un progetto denominato I-Swarm, che rappresenta il primo tentativo di replicare uno sciame di robot millimetrici in grado di imitare ciò che avviene negli sciami in natura, ovvero auto-organizzarsi. L'applicazione di tale progetto, ha dato vita a tre algoritmi fondamentali a tale scopo: ACO (*Ant Colony Optimization*), PSO (*Particle Swarm Optimization*) e FSO (*Flock of*

Starlings Optimization), tutti replicanti ciò che avviene in natura in termini di sistemi organizzati come insetti, pesci ed uccelli.

Dopo aver studiato nel dettaglio le tecniche sopra elencate, la scelta sulla metodologia da applicare per la problematica relativa alla previsione della mobilità utente, è infine ricaduta sul Sequential Pattern Mining.

Questa tecnica consente di ottenere l'informazione più significativa da una base di dati, trovando le relazioni esistenti tra le occorrenze relative ad eventi sequenziali. In particolare, il database è visto come una sequenza di dati in cui ogni transazione è un insieme di items (itemset), ed ogni sequenza è una lista di transazioni ordinata nel tempo. La scoperta di relazioni sequenziali, o pattern presenti in tali dati, è particolarmente utile per quanto riguarda la descrizione dei dati, l'identificazione di regole sequenziali e per la previsione di eventi futuri. In particolare, nel nostro lavoro di tesi, abbiamo deciso di utilizzare l'algoritmo PrefixSpan, appartenente alla famiglia degli algoritmi Prefix-Growth. Quest'ultimo, si basa interamente sulla proprietà *prefix-monotone* e dai vincoli monotoni e anti-monotoni. Il formato dati utilizzato, è basato sulla proiezione del database rispetto una sequenza α , ovvero vi è la proiezione del database composto solo da tutte le sequenze che hanno come sotto-sequenza α . I dati prelevati in input sono il database di sequenze, la soglia di supporto minimo ed un vincolo C di tipo prefix-monotone, i quali si tradurranno in un output contenente l'insieme dei patterns sequenziali che soddisfano il vincolo C.

Riguardo il data mining, sono stati analizzati, inoltre, tutti i processi relativi a regole associative, classificazione, clusterizzazione, workflow mining e data mining su grafi. Quest'ultimo in particolar modo, per la tematica da noi usata, si è rivelato molto utile in quanto può essere utilizzato per la modellazione di molte strutture basate su pattern tra i dati. Infatti, dato un insieme di grafi $D = \{G_0, G_1, \dots, G_n\}$ con $supporto(g)$, che indica il numero di grafi contenuti in D nei quali g è sottografo, la soluzione riguardo il problema di estrarre sottografi frequenti è quella di trovare tutti i sottografi g il cui supporto deve essere maggiore di una determinata soglia minima: $supporto(g) \geq minSup$. L'approccio da noi utilizzato è stato proprio quello di creare un albero PrefixTree, derivante dai dati dell'algoritmo PrefixSpan, contenente tutte le sequenze relative agli spostamenti degli utenti mobili e relativi valori di supporto, in modo tale da poter fare previsioni di spostamento a partire dai dati presenti in esso.