

Capitolo 3

Telecomunicazioni e reti neurali

3.1 Introduzione

Nel campo delle telecomunicazioni il contributo delle reti neurali è molto importante e va ad interessare svariati settori. Molte tematiche da affrontare sono riconducibili, infatti, a problemi di ottimizzazione, i quali trovano nelle reti neurali un efficace strumento per raggiungere risultati in tempi ridotti. Le ricerche svolte si occupano dello sviluppo di hardware e di tecniche software che migliorano le performances di sistemi già esistenti. Le prime applicazioni commerciali delle reti neurali interessarono la qualità della trasmissione vocale sulla linea telefonica e risalgono alla metà del secolo scorso (filtri adattivi di Widrow, 1959). Da allora esse hanno trovato impiego nei sistemi satellitari (ad esempio per il tracking nei sistemi GPS, ricevitori per trasmissioni a bit rate variabili con satelliti LEO), nelle reti cablate per risolvere problemi di routing, di allocazione di risorse e di previsione di mobilità e nei sistemi di telefonia senza fili, in cui è importante realizzare componenti che permettano l'equalizzazione della voce, l'eliminazione del rumore ambientale e, per sistemi 3G, la trasmissione video secondo gli standard (ad esempio, MPEG) effettuando compressioni e codifiche già all'interno del telefono, operazioni che eseguite con altri metodi richiederebbero una quantità di calcoli molto elevata ed un elevato assorbimento energetico. Nei prossimi paragrafi saranno analizzati problemi di telecomunicazioni risolti con tecniche neurali, mostrando quanto queste tecniche siano vantaggiose.

3.2 Reti neurali per l'individuazione della posizione di un utente in una rete cellulare

Le chiamate di emergenza che provengono da una rete radiomobile devono permettere di individuare con un certo grado di accuratezza la posizione del chiamante per permettere ai soccorsi di rintracciarlo in tempo utile per prestare i primi soccorsi. Già nel 1996 la commissione americana FCC (Federal

Communication Commission) stabilì che un requisito delle reti di telefonia è quello di determinare la posizione del chiamante con una tolleranza di 125 metri in almeno il 67% dei casi. Questa richiesta fu modificata nel 1999 dando specifiche diverse per i due possibili modi in cui si può gestire il problema:

- L'approccio basato sul terminale (handset based), in cui il telefono stima la sua posizione basandosi sui segnali ricevuti dalle stazioni radiomobili, deve avere una accuratezza di 50 m nel 67% dei casi. Per questo metodo il telefono dell'utente deve disporre di circuiteria di tipo GPS.
- L'approccio basato sulla rete (network based) in cui la rete, in base ai segnali ricevuti, individua la posizione del chiamante. Tale metodo deve avere una accuratezza di 100 m nel 67% dei casi. In questo caso le modifiche da effettuare riguardano tutte l'apparato hardware/software delle stazioni radio e non gli apparecchi dell'utenza.

I metodi tradizionali stimano la posizione basandosi su alcune tecniche quali AOA (Angle Of Arrival), TOA (Time Of Arrival) e RSS (Received Signal Strength). Zari-Jafarian, Mirsalehi et al., hanno proposto nel 2003 un metodo di posizionamento dell'utente basato su reti neurali gerarchiche basato sull'idea di dividere la copertura di ogni BTS in differenti aree parzialmente sovrapposte tra di loro e combinando i metodi di AOA e TOA. In figura 3.1 è proposto lo schema a blocchi utilizzato:

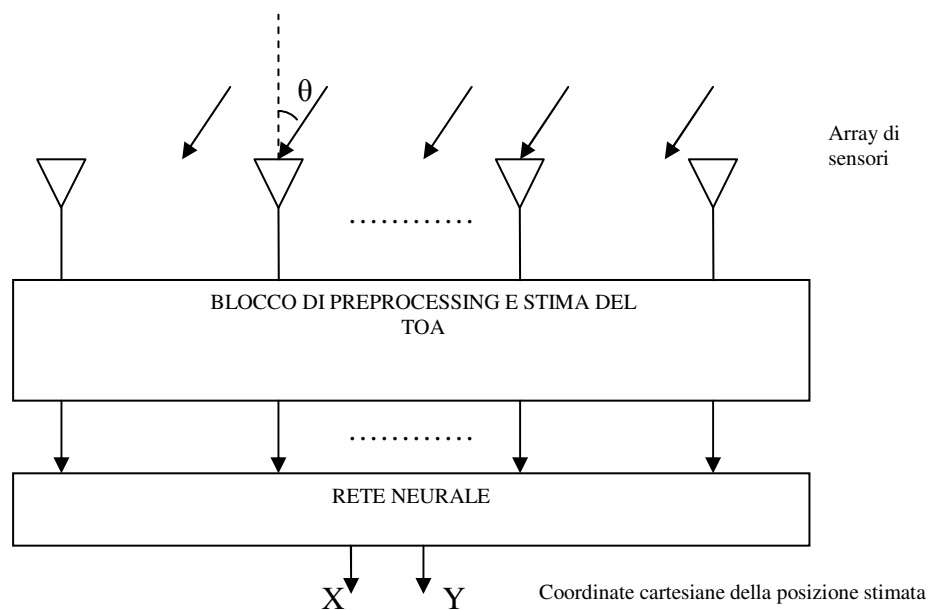


Figura 3.1. Diagramma a blocchi dello schema proposto.

In ogni stazione di trasmissione (BS) è presente un array lineare di antenne che permette di trovare le informazioni correlate con l'AOA. In input al blocco di preprocessing ci sarà un vettore di segnali ricevuti del tipo:

$$U(k) = [u_1(k), u_2(k), \dots, u_M(k)]'. \quad (3.1)$$

Indicando con $s(k)$ il segnale trasmesso dal telefono al tempo k , l'elemento m -esimo dell'array riceverà un segnale definito come:

$$u_m(k) = s(k) \exp\left(-j2\pi f_0 \frac{(m-1)d \sin(\theta)}{c}\right) + n_m(k) \quad m = 1 \dots M, \quad (3.2)$$

Dove con f_0 si indica la frequenza della portante, d la distanza tra gli elementi dell'array, c la velocità della luce, n_m il rumore additivo presente al sensore considerato e θ è l'angolo con cui il segnale incide sul sensore. Il blocco di preprocessing calcola la matrice di autocorrelazione del segnale ricevuto basandosi su N campioni del segnale.

$$R_u = E[U(k)U(k)^H] \cong \frac{1}{N} \sum_{k=1}^N U(k)U(k)^H \quad (3.3)$$

Dove si indica con $U(k)^H$ il trasposto complesso coniugato di $U(k)$. La matrice R_u è una matrice Hermitiana e contiene le informazioni sulla direzione di arrivo del segnale; di essa vengono considerati solo gli elementi della parte triangolare superiore o inferiore come input per la rete neurale. Per quanto riguarda il tempo di arrivo (TOA) è anch'esso individuato usando le informazioni sulla ricezione del segnale da parte dell'antenna. Il numero totale di input per la rete è pari a $(M^2 + M + 2)/2$ di cui uno è correlato al TOA e gli altri all'AOA. Utilizzando questo metodo una sola BS riesce a stimare la posizione (x,y) dell'utente. Nell'esempio proposto si considera un array di antenne di 5 elementi, una rete di perceptron con 16 input e 7 strati nascosti ognuno di 10 neuroni e uno strato di output con 2 unità. La funzione di attivazione usata è la tangente iperbolica mentre i pesi delle connessioni vengono calcolati tramite l'algoritmo di backpropagation. La tecnica utilizzata consiste nell'andare a "zoomare" la ricerca dell'utente in zone della cella sempre più piccole usando una o più BS. L'area di copertura di una BS è quindi divisa in più aree o celle. Nella struttura gerarchica considerata ad ogni livello n ci sono celle che al livello $n+1$ saranno divise in aree più piccole. La dimensione e il numero di tali aree sono determinate basandosi sulla stima dell'errore al livello immediatamente precedente. L'errore, oltre che alla dimensione dell'area considerata, dipende anche

dalla complessità dell'area stessa, cioè dalla presenza o meno, all'interno di essa, di strutture (ad esempio edifici) che provocano fenomeni di riflessione di parte del segnale. Al primo livello della struttura si stima la posizione, addestrando la rete neurale su tutta l'area coperta dalla BST (livello zero) e si sceglie la frazione il cui centro ha distanza minima dal segnale stimato; di tale area, al livello successivo si sceglie la sotto-area il cui centro è maggiormente vicino al segnale rilevato e si procede così fino alla stima finale della posizione. Se si dispongono di stime provenienti da celle diverse la posizione finale verrà quantificata al livello 1 della gerarchia come:

$$(x_l, y_l) = \arg \min_{(x_l, y_l)} \sum_{b=1}^B \{(x - x_{l,b})^2 + (y - y_{l,b})^2\} \quad (3.4)$$

In cui B rappresenta il numero di BST che danno informazioni. In figura 3.2 si mostra lo schema di funzionamento nel caso in cui si consideri il segnale ricevuto da due BST.

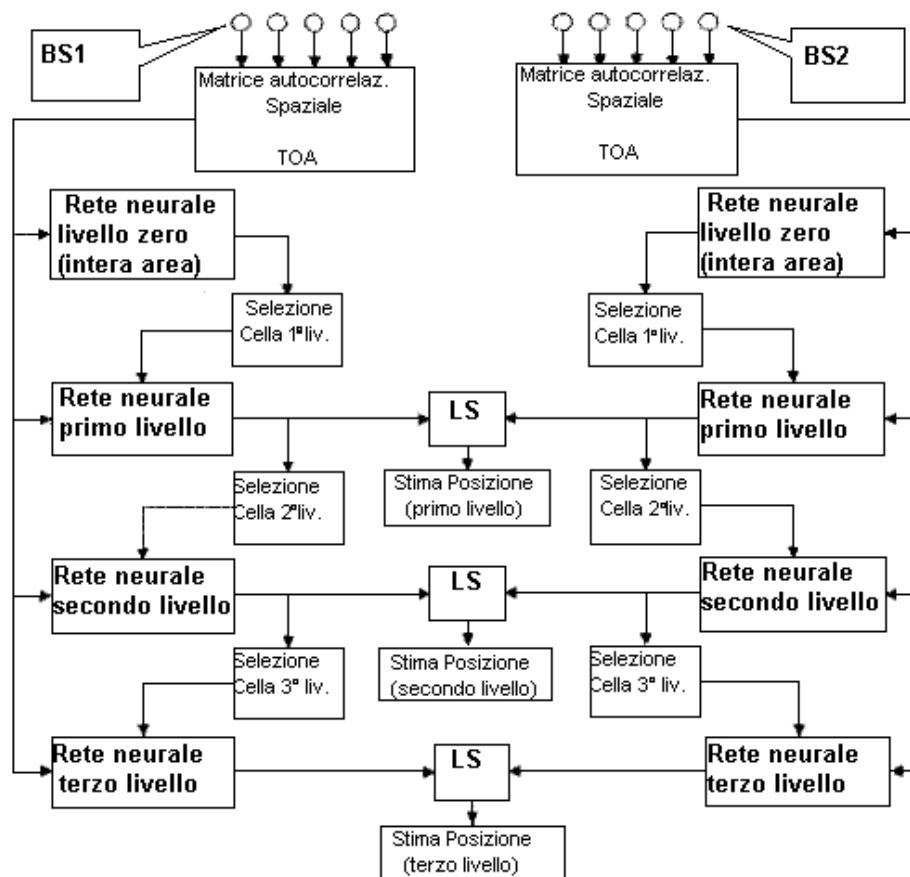


Figura 3.2. Diagrammi a blocchi del sistema usato per la ricerca della posizione.

Le simulazioni sono state effettuate considerando un ambiente urbano con due BST distanti tra loro 1000 metri. I raggi scelti per le sottoaree delle celle sono di dimensioni pari a $r_1= 300\text{m}$, $r_2=220\text{m}$ e $r_3=170\text{m}$ rispettivamente per il primo, secondo e terzo livello di “zoom” mentre al livello zero il raggio è quello dell’intera area. Le distanze dei centri tra le celle del primo livello è di 200m, del secondo livello è 150m e per il terzo livello di 110m.

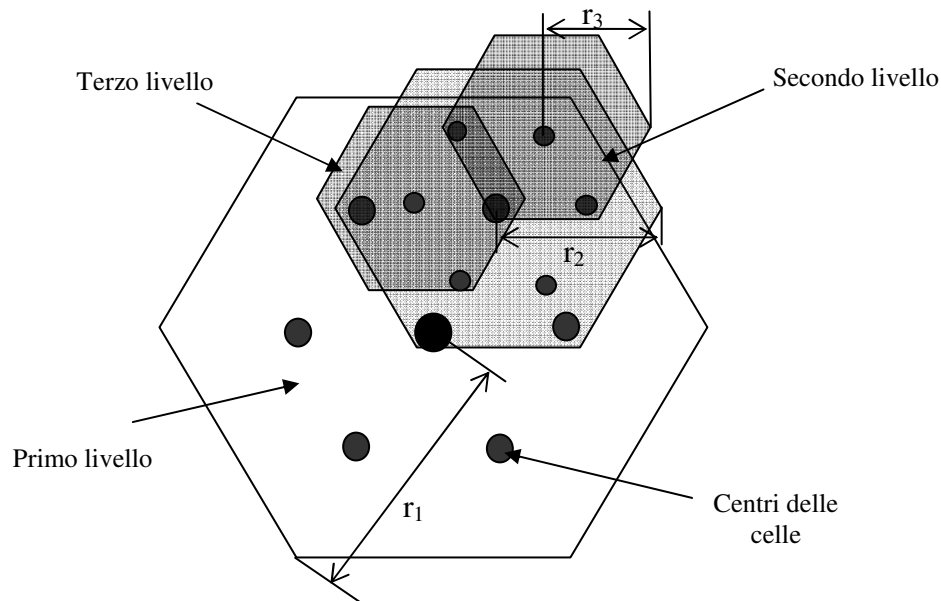


Figura 3.3. Suddivisione in sotto-aree di una cella.

Per quanto riguarda il numero di campioni di addestramento, le reti di primo livello sono state addestrate con 150 patterns, quelle di secondo livello con 140 patterns e, infine, quelle di terzo livello con 130 patterns. I risultati ottenuti sono soddisfacenti perché, combinando tecniche di AOA e TOA, facendo poi confluire i dati nella rete l’accuratezza della stima raggiunge valori prossimi all’80% per stime in aree di raggio pari a 100m.

3.3 QoS e reti neurali

La QoS (Quality of Service) indica l’insieme dei parametri che vengono utilizzati per permettere di caratterizzare i servizi della rete, sia in termini di risorse che in termini di qualità (perdita di pacchetti, ritardo). La QoS può essere caratterizzata andando a distinguere le tipologie di servizio che una rete di trasmissione a pacchetto può offrire in:

1. Guaranteed service (servizio garantito). Questa classe fornisce al flusso di dati generati dall'applicazione cliente un limite stretto sul ritardo di trasmissione end-to-end del generico pacchetto, inoltre viene garantita anche la banda che sarà fornita al flusso. Il servizio garantito, quindi, rappresenta sicuramente quello con le garanzie massime possibili ed ha caratteristiche tali da permettere di emulare una comunicazione su circuito dedicato.
2. Controlled load service (Servizio a carico controllato). Il servizio a carico controllato può essere considerato come una sorta di "controlled best-effort", nel senso che ha le stesse caratteristiche di un servizio best-effort (quindi nessuna garanzia) su di una rete poco carica; tuttavia si cerca di fare in modo che il servizio non cambi anche al crescere del carico della rete. In altre parole l'utente che richiede un servizio a carico controllato ottiene un servizio best-effort che però risulta essere poco sensibile alle variazioni del carico della rete. È chiaro, quindi, che per riuscire a mantenere queste promesse è fondamentale il ruolo del Call Admission Control (CAC).
3. Best effort è la tipologia offerta ad esempio da Internet, in cui viene garantito il migliore servizio possibile per le condizioni di traffico relative a quel determinato istante temporale. Non si hanno garanzie né in termini di banda disponibile né in termini di ritardi.

Nelle reti wireless è difficile garantire il servizio in mobilità, perché durante gli spostamenti intercella cambia il percorso dei dati e potrebbe non essere disponibile la quantità di banda necessaria per poter permettere ad un utente di utilizzare i servizi richiesti: il nuovo percorso dati potrebbe infatti attraversare rami della rete con una capacità ridotta che farebbero da collo di bottiglia influenzando negativamente tutta la capacità di trasmissione; inoltre si deve considerare che la distribuzione degli utenti non è uniforme e alcune aree della rete possono risultare più congestionate di altre, sia in funzione della loro posizione (ad esempio centro cittadino o zona periferica) sia dall'intervallo temporale in cui è richiesto il servizio (una rete cellulare ad esempio sarà molto più utilizzata nelle ore diurne); permettere al maggior numero di utenti di poter utilizzare la rete è reso ancora più difficile dal bisogno di gestire classi di utenza con esigenze di servizio differenti. Tecniche di controllo di accesso e di previsione degli spostamenti assumono un'importanza fondamentale. Una buona previsione della mobilità permetterà una migliore gestione

delle risorse mentre previsioni poco attente faranno in modo che la quantità di banda allocata ma non effettivamente utilizzata sia una percentuale molto alta della banda totale, provocando una certa degradazione delle prestazioni della rete. Per evitare ciò è possibile agire monitorando i movimenti all'interno della cella dividendola in due settori usando la "two-tier cell structure":

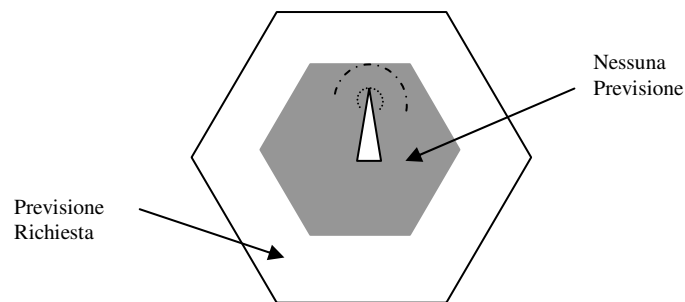


Figura 3.4. Struttura 2-tier della cella.

Quando l'utente si trova in prossimità della BS ("zona grigia" in figura 3.4) si prevede che l'utente permarrà per un certo tempo t ancora all'interno della stessa cella e nessuna prenotazione viene effettuata nelle celle adiacenti. Ad intervalli regolari la BST emetterà dei segnali (Beacon signals) che permetteranno di individuare la posizione di ogni utente che essa serve. La PSS (Pilot Signal Strength) degrada quando cresce la distanza tra la BST e l'utilizzatore; ciò permette di determinare quando l'utilizzatore si sposta nella "zona bianca" della cella in cui risulta necessario effettuare una prenotazione di banda su altre celle, in quanto la probabilità che l'utente si sposti agganciandosi a un'altra BST è alta. Conoscendo la direzione degli spostamenti di un utente si potrebbe effettuare una prenotazione come in figura 3.5, in cui l'utente spostandosi nella cella B permette un rilascio di risorse legate a celle adiacenti ad A ma non a B.

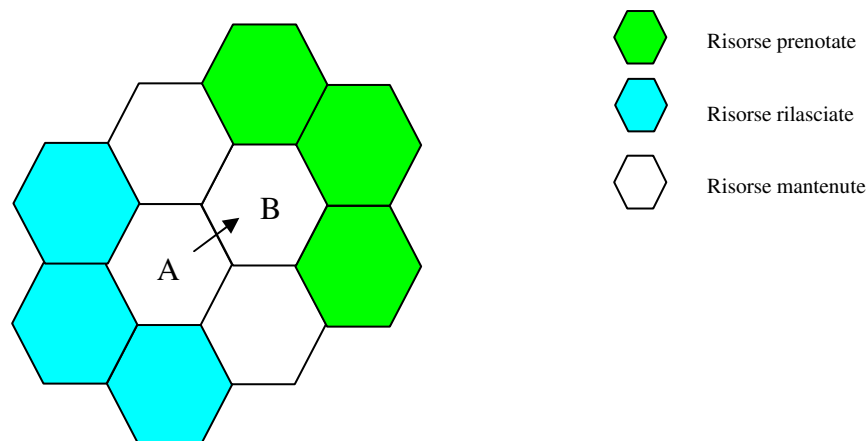


Figura 3.5. Procedura di prenotazione di banda.

Questa procedura, avendo a disposizione banda illimitata in ogni cella, potrebbe rappresentare la soluzione del problema, ma nelle situazioni reali in cui la banda disponibile è limitata l'aumento degli utenti provocherebbe una forte degradazione delle prestazioni. Per far fronte a ciò sono state studiate procedure di prenotazione basate su concetti quali lo Shadow Cluster o modelli di mobilità costruiti su stime statistiche riguardanti sia il tempo di permanenza per ogni cella di un utente (CST - Cell Stay Time) sia la determinazione delle celle adiacenti visitate con più frequenza in un periodo di osservazione della rete creando profili diversi per ogni utente o considerando intervalli temporali ben definiti. In letteratura molti schemi basati su reti neurali sono stati proposti soprattutto per quanto riguarda la previsione "One step" (previsione di una sola cella).

3.3.1 Schema Three Time Resource Reservation (TTRR)

Liou S. e Lu H. hanno proposto un modello di previsione basato su una rete addestrata con algoritmo di backpropagation chiamato "Three Time Resource Reservation scheme (TTRR)". La rete è addestrata andando a considerare tutte le coppie di valori (x,y) che rappresentano lo spostamento di un utente; la previsione della posizione al tempo t dipenderà da quelle precedenti di cui sarà tenuto conto grazie all'inserimento di elementi di ritardo D nella rete. Schematicamente è possibile rappresentare la rete neurale come in figura 3.6.

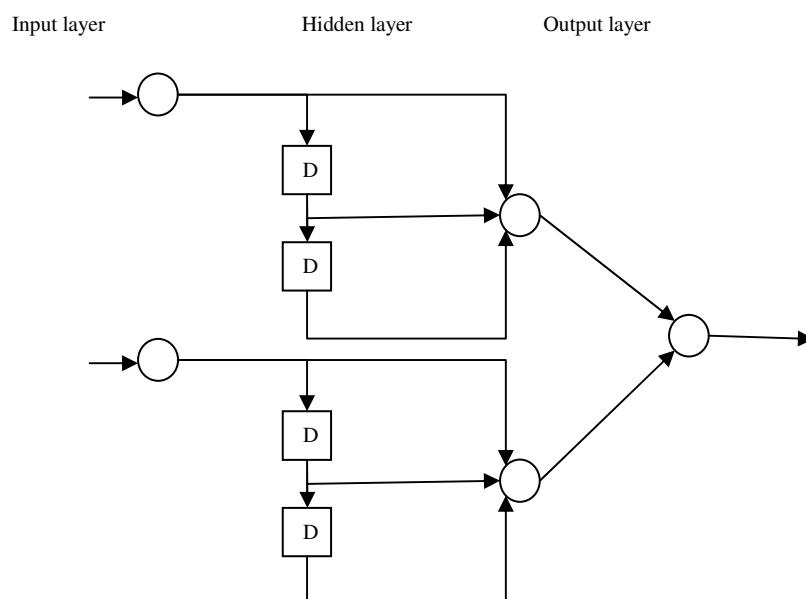


Figura 3.6.

Architettura neurale per il modello TTRR.

La tecnica consiste nel provare ad effettuare la prenotazione della banda richiesta per tre volte con una probabilità di successo molto più alta di quella che si avrebbe effettuando un solo tentativo (alcune risorse potrebbero essere rilasciate nell'intervallo tra una richiesta di prenotazione e la successiva). Una volta effettuato l'addestramento, fornendo alla rete i valori dello spostamento di un utilizzatore, si ottengono previsioni con un tasso di accuratezza medio che è dell'ordine del 72%. I passi svolti dall'algoritmo TTRR sono visualizzabili tramite il seguente diagramma di flusso:

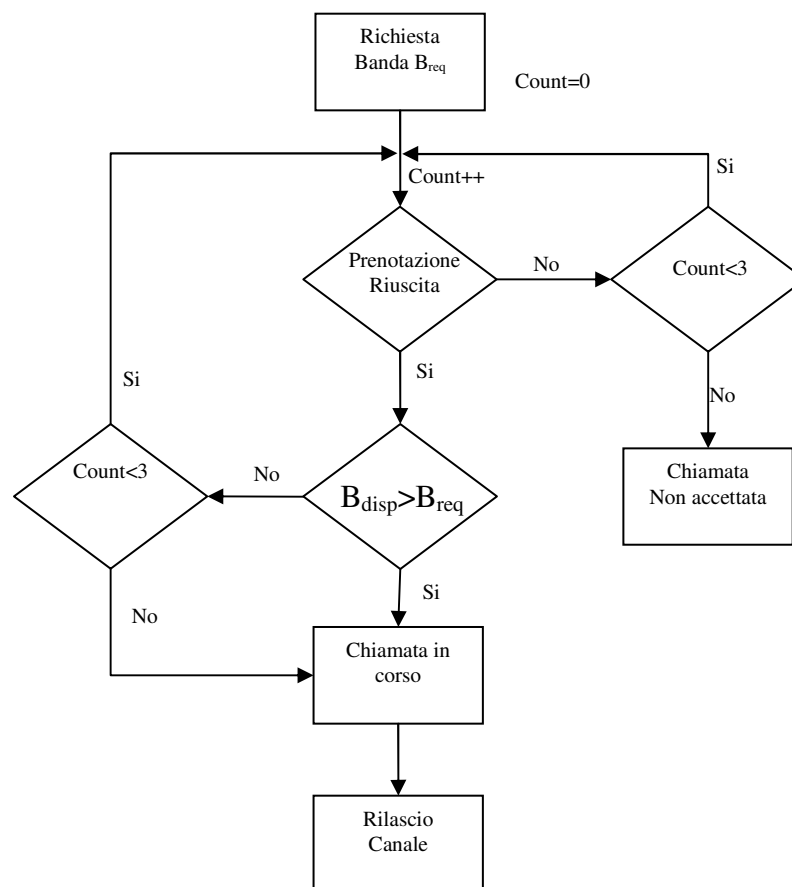


Figura 3.7. Diagramma di flusso del TTRR.

3.3.2 Gestione delle prenotazioni tramite NLP (Neural Location Predictor)

Un modello con risultati migliori in termini di accuratezza rispetto a quello presentato in 3.3.1 è stato proposto da Liou e Huang nel 2005. Il movimento intercella è considerato inizialmente come una funzione di posizione (p), velocità (v), accelerazione (a) e direzione (d). In un secondo momento si semplifica il