

UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE



Corso di Laurea magistrale in
Informatica

IL TITOLO DELLA TESI

Relatore: Relatore 1
Correlatore: Correlatore 1

Tesi di Laurea di:
Lorenzo D'Alessandro
Matr. Nr. 939416

ANNO ACCADEMICO 2020-2021

Dedica

Ringraziamenti

Questa sezione, facoltativa, contiene i ringraziamenti.

Indice

Ringraziamenti	ii
Indice	iii
1 Introduzione	1
1.1 I contenuti	1
1.2 Organizzazione della tesi	1
2 Stato dell'arte	2
3 RS per dispositivi mobili e pervasivi	3
3.1 Architettura generale	3
3.2 Sistema di raccomandazione	6
3.2.1 Input	6
3.2.2 Struttura della rete neurale	6
3.2.3 Training e inferenza	9
3.2.4 Vantaggi e svantaggi	9
3.3 Informazioni di contesto	10
3.3.1 Contesto fisico	11
3.3.2 Contesto sociale	12
3.3.2.1 Ego Network	12
3.3.2.2 Modellare il contesto sociale dell'utente	13
4 Dataset	15
4.1 Frappe	15
4.1.1 Feature di contesto	16
4.1.2 Feedback	17
4.1.3 Feature degli oggetti	17
4.1.4 Feature degli utenti	19
4.2 My Digital Footprint	20
4.2.1 Negative sampling	20

4.2.2	Feature di contesto	20
4.2.3	Feature degli oggetti	22
5	Risultati	24
6	Conclusioni	25
6.1	Conclusioni	25
6.2	Sviluppi futuri	25
	Bibliografia	26

Capitolo 1

Introduzione

Introduzione...

1.1 I contenuti

Spiegazione problema...

1.2 Organizzazione della tesi

Organizzazione tesi...

Capitolo 2

Stato dell'arte

Capitolo 3

RS per dispositivi mobili e pervasivi

La maggior parte dei RS producono le loro raccomandazioni usando unicamente informazioni su utenti, oggetti, e le valutazioni che gli utenti hanno dato agli oggetti (rating). Meno comuni sono invece i sistemi di raccomandazione context-aware, che considerano nel processo di raccomandazione anche il contesto in cui un rating è stato generato. La maggior parte dei CARS sfruttano un insieme limitato di feature di contesto selezionate manualmente, in modo da evitare di dover gestire una matrice dei rating con dimensionalità elevata. Recentemente sono stati proposti i primi sistemi di raccomandazioni context-aware che integrano nel processo di raccomandazione un alto numero di feature contestuali, come può essere quello estratto dai sensori di un dispositivo mobile. Come spiegato in conclusione al Capitolo 2, le soluzioni proposte non sono adatte per essere implementate direttamente su dispositivo mobile, anche se dimostrano come un approccio deep learning possa essere un'ottima soluzione per gestire la dimensionalità del contesto. Per questi motivi in questo capitolo è descritto un nuovo sistema di raccomandazione context-aware basato su deep learning per sistemi mobili e pervasivi. Il sistema di raccomandazione è stato pensato per eseguire le fasi di training e inferenza su dispositivo mobile, e poter supportare un grande numero di feature di contesto. Per le caratteristiche appena elencate, questo nuovo RS è chiamato *moveCARS* (MOBILE pervasIVE Context-Aware Recommender System).

3.1 Architettura generale

Generare raccomandazioni sul dispositivo mobile non è sufficiente per realizzare un RS pervasivo che non dipenda da un server centralizzato. È necessario spostare anche le fasi di raccolta ed elaborazione dei dati su dispositivo mobile. Per questo

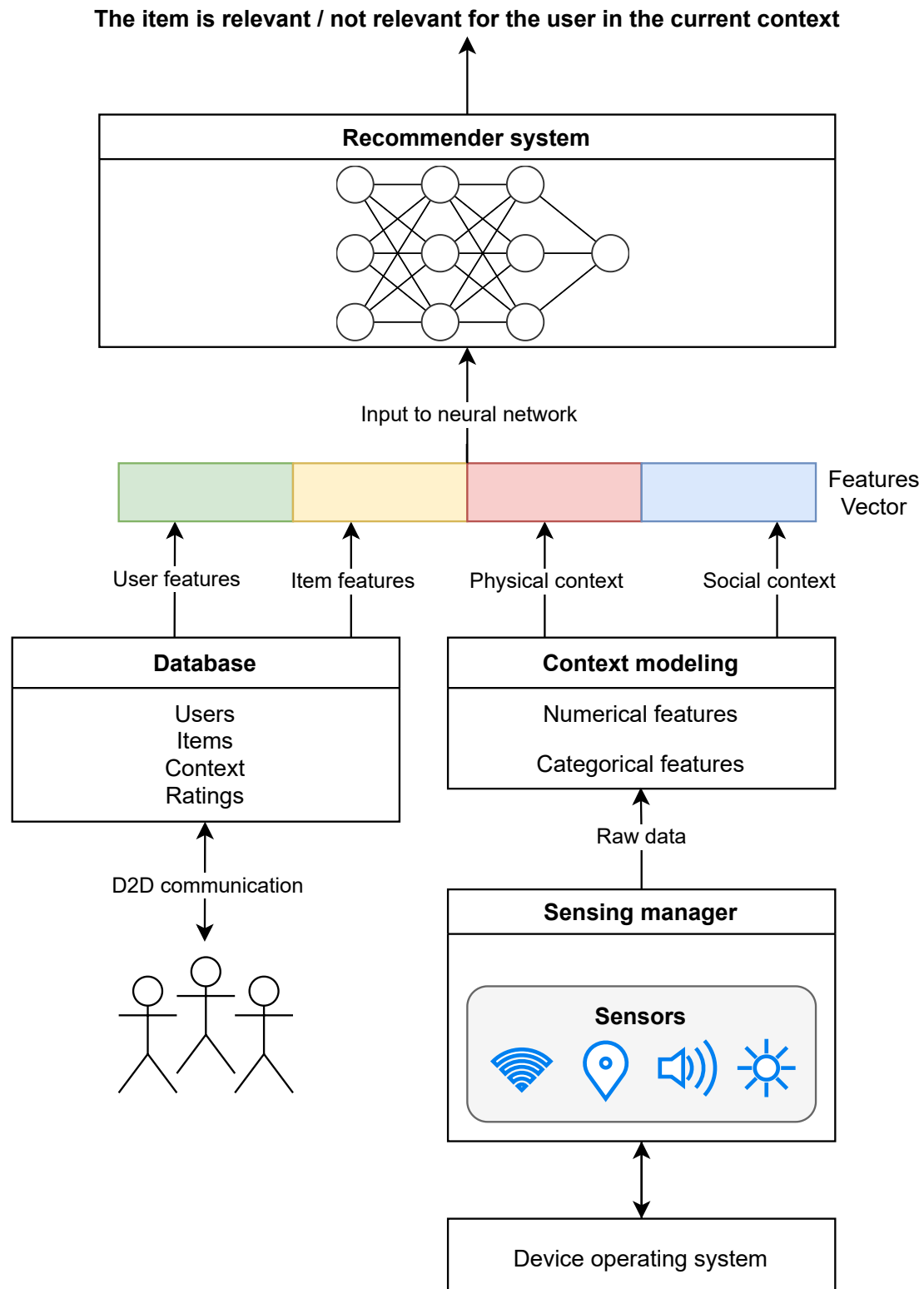


Figura 1: Architettura ad alto livello del sistema di raccomandazione

motivo il modello moveCARS è inserito in un architettura più complessa, che si occupa di generare i dati che saranno input per il RS. L'architettura ad alto livello è composta da quattro componenti:

1. *Sensing manager*. Il primo componente interagisce con il sistema operativo per raccogliere continuamente dati di contesto (come GPS e accelerometro), e dati che rappresentano informazioni sull'hardware e il software del dispositivo come lo stato del display e il livello di batteria. Tutte queste informazioni sono chiamate dati raw perchè non sono ancora state processate.
2. *Context modeling*. I dati raw prodotti dal sensing manager devono essere processati per inferire una rappresentazione più astratta del contesto dell'utente. A questo scopo, il componente context modeling (CM) raccoglie periodicamente gli ultimi dati disponibili del SM. Queste osservazioni sono processate per estrarre feature numeriche e categoriche che caratterizzano il contesto dell'utente locale (es. attività utente, luogo in cui si trova l'utente, temperatura). Il contesto è diviso in contesto fisico e sociale. Il contesto fisico è ottenuto dai sensori del dispositivo dell'utente, mentre il contesto sociale è ottenuto considerando gli individui in prossimità dell'utente. Dell'insieme di feature eterogenee ottenute viene fatto l'encoding per poi essere combinate in un singolo vettore di feature che rappresenta una fotografia del contesto corrente dell'utente. L'insieme di feature che compongono il contesto fisico e sociale dell'utente sono descritte nel dettaglio nella sezione 3.3.
3. *Database*. Interagendo con altri dispositivi tramite comunicazione D2D, il device dell'utente scopre nuovi utenti e oggetti che sono identificati con delle feature che li caratterizzano. Oltre a questo il dispositivo riceve anche i feedback che questi utenti hanno generato sugli oggetti in un certo contesto. Tutte queste informazioni sono date in input al sistema di raccomandazione durante la fase di apprendimento per imparare a prevedere i feedback dell'utente locale. A training terminato, le feature degli oggetti presenti nel database, e le feature dell'utente corrente sono usate per produrre raccomandazioni. Le fasi di training e inferenza sono descritte nella sottosezione 3.2.3.
4. *Sistema di raccomandazione*. Le feature di contesto c sono concatenate alle feature degli utenti u e degli oggetti i in un unico vettore. Questo vettore è dato in input ad una rete neurale, che restituisce valore 1 se per l'utente con feature u , l'oggetto con feature i è rilevante nel contesto c , 0 altrimenti. L'input, la struttura e l'output della rete sono descritti nella sezione 3.2.

3.2 Sistema di raccomandazione

In questa sezione è descritto il sistema di raccomandazione moveCARS. Nella prima parte è descritto l'input, e in che modo si differenzia dai sistemi di raccomandazione collaborative filtering e content-based. Nella seconda parte è descritta nel dettaglio la struttura e il training della rete neurale che genera le raccomandazioni context-aware. In conclusione sono descritti i vantaggi e gli svantaggi del modello.

3.2.1 Input

Solitamente l'input dei modelli collaborative filtering context-aware è composto da tuple (`user_ID`, `item_ID`, `rating`, `context`), in cui `user_ID` è l'utente che ha valutato l'oggetto `item_ID` con una valutazione `rating` in una situazione descritta dal contesto `context`. Al posto che identificare l'oggetto con un valore numerico intero `item_ID`, si possono usare delle feature che caratterizzano l'oggetto, esattamente nello stesso modo in cui sono solitamente descritti gli oggetti nei sistemi di raccomandazione content-based. Ad esempio, se si sta sviluppando un RS per consigliare ristoranti agli utenti, si può sostituire il valore `item_ID` che identifica il ristorante con delle feature che lo caratterizzano nel dettaglio come il tipo di cibo servito, il prezzo medio, l'atmosfera, se ha sedute all'aperto, etc. Allo stesso modo si può sostituire il valore `user_ID` con delle feature che descrivono l'utente. Queste possono essere feature non specifiche come età o genere, o feature specifiche per l'ambiente in cui il RS è implementato. Tornando all'esempio dei ristoranti, si potrebbe chiedere all'utente quanto è disposto a spendere per mangiare fuori e il tipo di cucina preferita. A feature di utente e oggetto si aggiungono le feature del contesto fisico e sociale generate dal modulo di Context modeling. Un'istanza di rating per il modello moveCARS è quindi una tupla (`user_features`, `item_features`, `rating`, `physical_context`, `social_context`).

3.2.2 Struttura della rete neurale

Il vettore di feature appena descritto è dato in input ad una rete neurale che deve prevedere se l'utente caratterizzato da `user_feature` è interessato all'oggetto caratterizzato da `item_feature` nei contesti fisici e sociali `physical_context` e `social_context`. Si tratta quindi di un problema di classificazione binaria. Nei problemi di classificazione, l'obiettivo è prevedere il valore di una variabile che può assumere diversi valori discreti. I problemi di classificazione in cui una variabile può assumere solo due valori possibili (come 0 o 1) sono chiamati problemi di classificazione binaria [1].

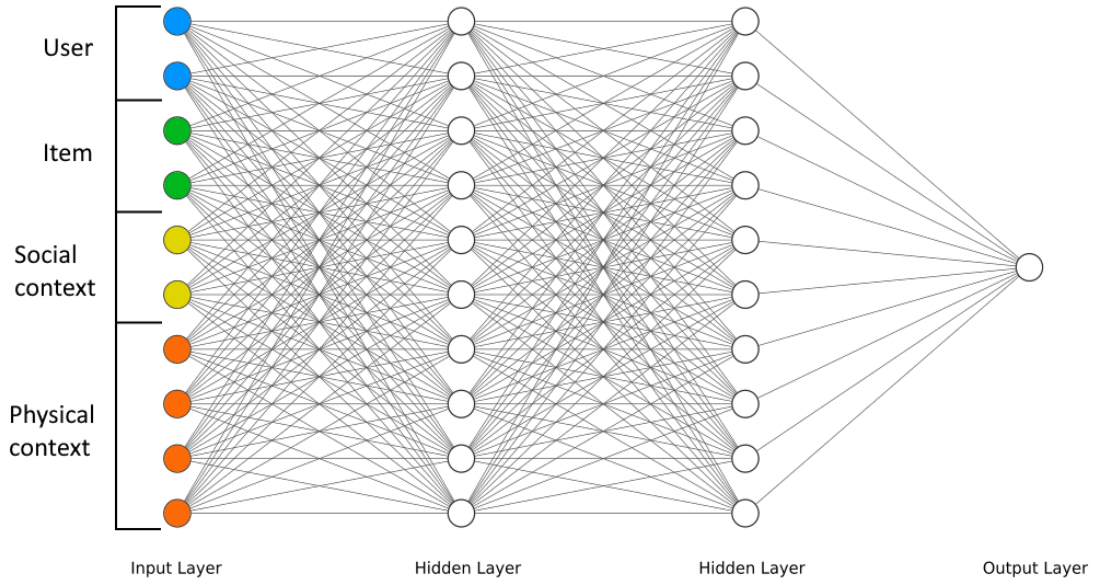


Figura 2: Schema di moveCARS

Layer e neuroni La rete neurale scelta rientra nella categoria feed-forward fully connected. Una rete feed-forward non contiene cicli nel suo grafo [2], fully connected indica che ogni neurone del layer i è connesso a tutti i neuroni del layer $i + 1$. La rete ha un layer di input, un layer di output e l layer nascosti. Il layer di input ha un numero di neuroni pari alle feature in ingresso (sommando user, item e context feature), il layer di output ha sempre un neurone, mentre il numero di neuroni nei layer nascosti l , e il numero di layer nascosti è calcolato facendo il tuning della rete tramite grid search, scegliendo la combinazione che ottiene i risultati migliori. In questa tesi è stato utilizzato lo stesso numero di neuroni in ogni layer nascosto, ma si può ad esempio adottare un design a torre in cui i layer più profondi contengono meno neuroni rispetto ai layer meno profondi.

Funzione di attivazione Una funzione di attivazione di un neurone definisce l'output di quel neurone in base all'insieme dei suoi input. Come funzione di attivazione dei layer nascosti ho scelto la funzione rectified linear unit (ReLU) definita come $f(x) = \max\{0, x\}$. La funzione ReLU è consigliata per la maggior parte delle reti feed-forward [2], e ha diversi vantaggi rispetto a funzioni di attivazione come sigmoide e tanh: è più plausibile biologicamente, non viene saturata (a differenza di tanh e sigmoide che hanno un output massimo uguale a 1), e incoraggiando l'attivazione sparsa dei neuroni rende più difficile che si verifichi l'overfitting del modello durante il training [3]. Come funzione di attivazione del layer di output

ho scelto la funzione sigmoide definita come

$$f(x) = \frac{1}{1 + e^{-x}}$$

che limita l'output della rete a valori tra 0 e 1, ed è quindi adatta per problemi di classificazione binaria [4].

Funzione di loss Una funzione di loss è una misura dell'errore tra il valore previsto dal modello e il valore effettivo. Come funzione di loss la scelta più comune per un classificatore binario è la funzione binary cross-entropy / log loss, definita come

$$C = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))$$

dove y è il valore reale del feedback di un utente su un oggetto (0 oppure 1), $p(y)$ è la probabilità predetta dalla rete che y abbia valore 1, e $1 - p(y_i)$ è la probabilità che y abbia valore 0 [5].

Ottimizzatore Un ottimizzatore è un algoritmo che modifica i pesi del modello in modo da minimizzare la funzione di loss e rendere le previsioni della rete più accurate possibile. Come ottimizzatore ho scelto Adam; nel paper in cui è introdotto viene dimostrato empiricamente di essere generalmente migliore rispetto ad altri algoritmi di ottimizzazione stocastici, e di risolvere in modo efficiente problemi di deep learning [6]. Adam ha diversi iperparametri configurabili, il più importante è il learning rate (chiamato α in Adam) che regola la velocità con cui il modello è adattato al problema. Gli altri parametri $(\beta_1, \beta_2, \epsilon)$ sono lasciati al valore di default della libreria Keras¹.

Epoche e batch size Epoche e batch size sono due parametri molto importanti da ottimizzare, il numero di epoche e la dimensione della batch size ideali sono calcolate tramite grid search nel Capitolo 5. La batch size corrisponde al numero di campioni processati prima di aggiornare i parametri del modello. Il numero di epoche indica quante volte viene presentato alla rete il training set prima di concludere il training.

In Figura 2 è rappresentata la struttura della rete neurale. In questo caso il modello ha due layer nascosti ed ogni layer contiene 10 neuroni, eccetto il layer di output che contiene un solo neurone.

¹<https://keras.io/api/optimizers/adam/>

3.2.3 Training e inferenza

Training

Il processo di training di una rete neurale si basa sul trovare un insieme di pesi nella rete che permettano di risolvere nel modo migliore possibile un problema specifico. Il processo di training è iterativo, il che significa che procede passo dopo passo con piccoli aggiornamenti nei pesi del modello ad ogni iterazione, migliorando le performance del modello. Il processo di training iterativo di una rete neurale risolve un problema di ottimizzazione per dei parametri, (i pesi del modello) che ha come risultato un errore minimo durante la valutazione degli esempi nel training dataset. In ambiente mobile e pervasivo inizialmente il numero di esempi nel training dataset è limitato, e non rappresenta la conoscenza globale su tutti gli utenti, oggetti e rating. Questo non è un problema per moveCARS che può iniziare il training sui sample disponibili, per poi riprenderlo in un secondo momento quando il device utente tramite comunicazione D2D avrà scoperto nuovi sample. Questo è possibile per la struttura dell'input della rete neurale. L'input è un vettore formato dalla concatenazione di `user_features`, `item_features`, `physical_context`, `social_context`, la cui dimensione è fissata. Infatti utenti, oggetto e contesto sono definiti da un insieme di feature che non cambia nel tempo. Non è quindi necessario ridefinire il modello ogni volta che il numero di utenti o oggetti cambia, cioè quando sono scoperti nuovi utenti od oggetti.

Inferenza

Il task per un sistema di raccomandazione che fa classificazione è determinare se un utente è interessato ad un oggetto in una situazione descritta dal contesto fisico e sociale. La rete restituisce valore 1 se l'oggetto è rilevante, 0 altrimenti. Nel caso di moveCARS utenti e oggetti non sono definiti con degli ID numerici, come nei metodi collaborative filtering, ma da feature che li caratterizzano. Il task di moveCARS quindi può essere riformulato in modo più specifico come prevedere se all'utente con feature `user_feature` interessa un oggetto con feature `item_feature`, nel contesto fisico `physical_context`, e nel contesto sociale `social_context`.

3.2.4 Vantaggi e svantaggi

Di seguito sono elencati vantaggi e svantaggi del modello moveCARS. Dato che può essere considerato un sistema ibrido che unisce caratteristiche degli approcci collaborative filtering e content-based, eredita alcuni vantaggi e svantaggi da entrambe le categorie di algoritmi.

Vantaggi:

1. *Nessuna conoscenza del numero di utenti e oggetti.* Utenti e oggetti sono rappresentati con delle feature e non con il loro ID, non è necessario conoscere a priori quanti utenti e oggetti sono presenti nel sistema.
2. *Consigliare nuovi oggetti.* Il modello può consigliare nuovi oggetti anche se non ci sono valutazioni fornite dagli utenti, a differenza dei metodi collaborative filtering.
3. *Feature di contesto.* Il modello permette di integrare una grande quantità di feature di contesto che possono migliorare sensibilmente la capacità di predizione, come dimostrato nel Capitolo 5.
4. *Serendipity:* il modello ha la capacità di fornire consigli fortuiti, il che significa che può consigliare elementi pertinenti per l'utente senza che il contenuto si trovi nel profilo dell'utente, a differenza dei metodi content-based.

Svantaggi:

1. *Feature di utenti e oggetti:* la precisione del modello dipende dall'insieme delle feature che descrivono gli utenti e gli oggetti. È necessario selezionare attentamente le feature più adatte che descrivono utenti e oggetti in un'applicazione specifica.
2. *Raccomandazioni multi-domain:* è difficile creare RS multi-domain perché è complicato definire un insieme di feature che valgano per contenuti di natura diversa.

3.3 Informazioni di contesto

In questa sezione sono descritte le informazioni di contesto fisico e sociale che vengono date in input al sistema di raccomandazione mobile. Nel Capitolo 5 è dimostrato che un insieme ampio di feature contestuali può portare a raccomandazioni molto più accurate, mentre un insieme di feature contestuali poco esteso ha un impatto decisamente minore.

3.3.1 Contesto fisico

Il contesto fisico è composto da tutte quelle informazioni rilevanti che possono essere utilizzate per caratterizzare la situazione di un utente. Le feature del contesto fisico sono ricavate dai sensori fisici dello smartphone di un utente (es. attività utente dall'accelerometro) e dal sistema operativo del telefono (es. stato display e livello batteria). A queste feature si vanno a integrare informazioni esterne come il meteo, la data e l'ora. Più nel dettaglio il contesto utente è caratterizzato dalle seguenti informazioni:

Posizione Informazioni relative alla posizione geografica che includono latitudine, longitudine, precisione della posizione e direzione del movimento. La posizione geografica può essere usata per capire il luogo in cui si trova l'utente (a casa, al lavoro, etc.) o per raccomandare punti di interesse nelle vicinanze.

Movimento utente Il movimento dell'utente include sia le attività svolte a piedi (correre e camminare), sia il movimento su un mezzo di trasporto (veicolo generico o bicicletta).

Applicazioni Applicazioni in esecuzione sul dispositivo.

Audio Informazioni relative alla configurazione audio dello smartphone, incluse la modalità audio (suono, vibrazione, silenzioso), il volume delle notifiche, e lo stato dell'altoparlante (acceso o spento). Anche l'audio può migliorare il riconoscimento del contesto, per esempio durante una riunione la modalità audio potrebbe essere impostata su silenzioso e l'altoparlante spento.

Batteria Informazioni relative alla batteria del telefono che includono il livello di carica e se la batteria si sta ricaricando.

Display Stato dello schermo dello smartphone (acceso o spento), e orientamento dello schermo (verticale od orizzontale).

Dati dei sensori fisici che includono sensori ambientali (es. temperatura dell'ambiente e luce), sensori di movimento (es. accelerometro e giroscopio) e sensori di posizione (es. rotazione e prossimità).

Celle di rete Lista delle celle di rete mobile rilevate dal dispositivo. Per ogni cella si identifica il tipo di tecnologia (es. GSM o LTE), l'ID della cella, e la forza del segnale. La rete mobile può migliorare l'identificazione della posizione dell'utente.

Wi-Fi Lista di tutti gli access point Wi-Fi disponibili in prossimità, e se l'utente è connesso ad uno di essi.

Meteo Informazioni relative alle condizioni meteo che includono il tempo in atto (es. nuvoloso, piovoso, soleggiato), la temperatura, l'umidità e la velocità del vento.

Data e ora Dalla data si possono generare feature come il giorno della settimana, la stagione, comprendere se è il fine settimana o un periodo di vacanza, etc. Dall'orario invece si può capire il momento della giornata (mattina, pomeriggio, sera, notte).

3.3.2 Contesto sociale

Il contesto sociale si riferisce all'insieme di persone con cui l'utente ha interazioni sociali durante la vita giornaliera, come lavorare con i colleghi o messaggiare con gli amici. È stato provato in letteratura che esiste una forte correlazione tra le attività umane e i dati sociali [7]. Questo implica che modellare una rete specifica per l'utente di relazioni sociali può contribuire a sottolineare le differenze tra i vari contesti in cui è coinvolto.

3.3.2.1 Ego Network

Una ego network è una rete sociale composta da un individuo chiamato ego, e dalle persone con cui l'ego ha un collegamento sociale, chiamati alter. I legami sociali in una ego network non hanno tutti la stessa importanza. Ogni individuo ha solo pochi collegamenti forti e molti collegamenti deboli, dovuti alla capacità umana di gestire un numero limitato di relazioni sociali. Una rappresentazione della ego network è mostrata in Figura 3: l'ego è il punto rosso al centro dei quattro cerchi concentrici chiamati layer, in cui gli alter sono distribuiti in base alla forza del legame sociale con l'ego. Il cerchio più interno (support clique) è il layer più piccolo, e contiene solo pochi alter che rappresentano le relazioni sociali più forti con l'ego. Il secondo layer (sympathy group) contiene le persone che possono essere considerati gli amici più cari. Il terzo cerchio (affinity group) è composto da amici e membri della famiglia meno vicini, mentre l'ultimo layer include persone con cui l'individuo ha interazioni sociali occasionali [7].

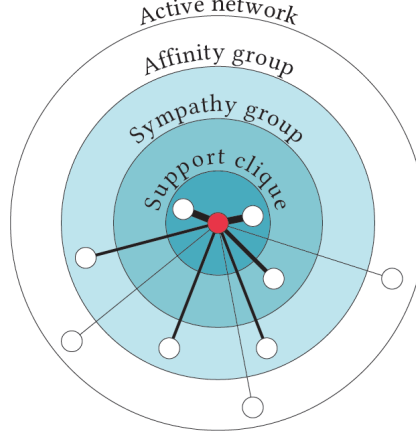


Figura 3: Ego network
[7]

3.3.2.2 Modellare il contesto sociale dell'utente

Per modellare il contesto sociale di un utente in ambiente mobile, si caratterizzano le interazioni sociali usando le seguenti sorgenti di dati: (i) chiamate telefoniche e log degli SMS, (ii) dati di prossimità, e (iii) attività svolte dall'utente sugli online social networks (OSN).

Il primo step per costruire l'ego network di un individuo è stimare la forza dei legami sociali con i suoi alter. Un buon indicatore della forza delle relazioni sociale tra due persone è data dal numero di interazioni che le due persone hanno avuto in passato. Basandosi su questa considerazione, per modellare la forza dei legami sociali dell'utente online, sono presi in considerazione diverse attività svolte dall'utente su OSN, inclusi commenti, reazioni (come “mi piace”) e persone menzionate. Formalmente, la forza dei legami sociali virtuali tra l'ego e ed uno dei suoi alter a , $\omega_{osn}(e, a)$ è calcolata nel modo seguente:

$$\omega_{osn}(e, a) = \sum_{v \in V} I_S(e, a) \quad (1)$$

dove V è l'insieme delle sorgenti di dati degli OSN nominate prima, e la funzione $I_S(e, a)$ calcola il numero di interazioni tra e ed a per una data sorgente di dati.

Per caratterizzare i link sociali fisici di un utente si calcola il numero di interazioni con altre persone basandosi su telefonate, SMS e contatti faccia a faccia inferiti usando tecnologie wireless disponibili sugli smartphone. In particolare sono considerate il Bluetooth (BT) e il Wi-Fi Direct (WFD), per scoprire persone che sono abbastanza vicine da aver un'interazione con l'utente locale. Sono filtrati i dispositivi che non si trovano in prossimità dell'utente, e sono selezionati solo i

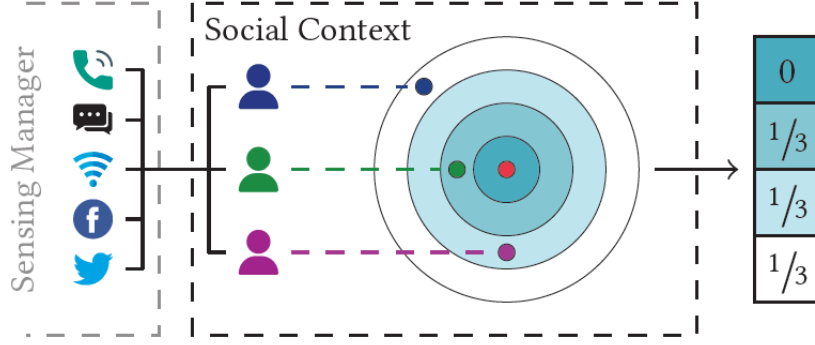


Figura 4: Riconoscimento del contesto sociale di un utente [7]

dispositivi personali dell'utente, in modo tale da non considerare stampanti, smart TV etc. In modo simile ai link sociali virtuali, si definisce la forza dei legami fisici sociali tra l'ego e e un alter a , $\omega_{phy}(e, a)$ come il numero delle loro interazioni tramite telefonate, SMS, e prossimità fisica come segue:

$$\omega_{phy}(e, a) = \sum_{p \in P} I_p(e, a) \quad (2)$$

dove P è l'insieme delle sorgenti fisiche considerate, e $I_p(e, a)$ rappresenta il numero di interazioni tra due utenti per una data sorgente di dati. Infine, la forza complessiva del collegamento sociale tra e ed a è data dalla combinazione lineare delle interazioni online e fisiche descritte prima:

$$\omega_s(e, a) = \lambda \cdot \omega_{osn}(e, a) + (1 - \lambda) \cdot \omega_{phy}(e, a) \quad (3)$$

con il parametro λ che regola l'importanza delle interazioni sociali e fisiche. Per ogni alter, solo l'ultimo peso calcolato è mantenuto in memoria, e viene aggiornato quando nuove interazioni sociali sono identificate. I link sociali tra l'utente locale e le altre persone sono raggruppate in base al peso calcolato nell'Equazione 3. L'output finale è un array di valori in cui ogni elemento rappresenta la percentuale di utenti attivi in ogni cerchio della ego network di un utente [7].

La Figura 4 mostra un esempio del processo di riconoscimento del contesto. Basandosi sui dati raccolti dal Sensing Manager, viene riconosciuta la rilevanza degli alter per l'utente locale in base a livello in cui essi sono posizionati nella ego network. L'output caratterizza il contesto sociale dell'utente, indicando chiaramente che sta interagendo con tre persone diverse che si trovano nel secondo, terzo e quarto layer rispettivamente.

Capitolo 4

Dataset

In questo capitolo sono descritti i dataset context-aware usati per valutare il modello moveCARS, e nel Capitolo 5, confrontarlo con altre soluzioni stato dell'arte. Uno dei problemi nella valutazione dei CARS, è la scarsità di dataset pubblici che contengono informazioni di contesto ad alta dimensionalità. Molti dataset infatti hanno informazioni di contesto limitate unicamente al timestamp dei feedback come Yelp¹ e nowplaying-rs². In generale è difficile trovare dataset che caratterizzano il contesto fisico e sociale dell'utente con una grande quantità di feature. Dato che il modello moveCARS si inserisce in un ambiente mobile e pervasivo, ho selezionato due dataset context-aware, in cui i feedback dell'utente corrispondono alle applicazioni in esecuzione sui dispositivi Android degli utenti: (i) My Digital Footprint³, (ii) Frappe⁴.

4.1 Frappe

Frappe [8] è un dataset di feedback impliciti pubblicamente disponibile collezionato da un sistema di raccomandazione context-aware di applicazioni Android. L'applicazione che monitora l'utilizzo degli smartphone, è stata installata da 957 utenti che hanno utilizzato un totale di 4082 applicazioni. Le informazioni raccolte descrivono la frequenza di utilizzo di un'applicazione da parte di un utente per un periodo di 2 mesi. Il numero totale di feedback presenti è 96203.

¹<https://www.yelp.com/dataset>

²<https://zenodo.org/record/3247476#.YK9FxqgzY2x>

³<https://github.com/contextkit/MyDigitalFootprint>

⁴<https://www.baltrunas.info/context-aware>

4.1.1 Feature di contesto

Le feature di contesto descrivono la situazione dell'utente nel momento in cui ha utilizzato un'applicazione Android. Frappe può essere considerato un dataset con un contesto a bassa dimensionalità, e non contiene informazioni raccolte dai sensori dei dispositivi Android. Di seguito sono descritte tutte le feature di contesto presenti nel dataset.

Daytime è il momento della giornata in cui un'applicazione è stata utilizzata. La giornata è divisa in sette momenti diversi: morning, noon, afternoon, evening, sunset, sunrise, night.

Weekday è il giorno della settimana in cui un'applicazione è stata utilizzata. I possibili valori sono ovviamente i sette giorni della settimana.

Isweekend indica se un'applicazione è stata utilizzata nel fine settimana oppure in un giorno lavorativo. Può assumere due valori diversi: weekend e workday.

Homework indica se l'utente si trova al lavoro o a casa. Può assumere tre diversi valori: work, home, unknown.

Weather descrive la situazione meteo nel momento in cui un'applicazione è stata utilizzata. Può assumere nove valori differenti: sunny, cloudy, foggy, stormy, rainy, snowy, drizzle, sleet, unknown.

Country indica la nazione in cui si trovava l'utente nel momento in cui ha utilizzato un'applicazione. Ci sono 80 stati diversi, ma il 55% dei feedback sono stati generati da USA, spagna e regno unito.

City è un valore numerico che rappresenta la città in cui si trovava l'utente nel momento in cui ha utilizzato un'applicazione. Ci sono 233 città diverse, ma per il 40% dei feedback la città è sconosciuta.

Delle feature di contesto appena descritte sono eliminate homework, country e city perché poco utili a definire il contesto dell'utente. In particolare city è stata eliminata perché contiene troppi valori sconosciuti, ed in corrispondenza di una città sconosciuta c'è spesso anche il valore della feature country sconosciuta. Anche homework è stato eliminato perché il numero di feedback che hanno la feature homework con valore unknown è pari al 78%. Country non viene considerata perché la maggior parte dei feedback utente sono associati a poche nazioni, e ci sono un

gran numero di nazioni a cui sono associati un numero non sufficiente di feedback. Le feature rimaste compongono il contesto: `daytime`, `weekday`, `isweekend`, `weather`. L'unica di queste feature che può assumere valore `unknown` è `weather`. Le righe del dataset in cui `weather` è `unknown` sono eliminate. Il risultato è un dataset con 78335 righe, 857 utenti e 3180 oggetti.

Per quanto riguarda l'encoding, essendo tutte variabili categoriche sono codificate con one-hot encoding. Il vettore del contesto risultato contiene 24 features: 7 per `daytime`, 7 per `weekday`, 8 per `weather` e 2 per `isweekend`.

4.1.2 Feedback

Qualsiasi dataset per sistemi di raccomandazione ha tre feature fondamentali: `user`, `item`, `rating`. `User` e `item` sono valori numerici che identificano univocamente gli utenti e gli oggetti. Il `rating` in Frappe è il numero di volte in cui un oggetto (un'applicazione) è stata utilizzata da un utente in un determinato contesto. Ad esempio, se una riga del dataset è composta da (`user:1`, `item:20`, `rating:50`, `daytime:morning`, `weekday:monday`), significa che l'utente 1, ha utilizzato l'applicazione 20, il lunedì mattina, 50 volte. Il numero di volte è ottenuto sommando tutti gli utilizzi durante il periodo di raccolta dei dati. Il valore minimo dei `rating` è 1, il valore massimo 21262, e la media 88.26. Questi `rating` vanno convertiti in feedback impliciti con valore 0 o 1 per essere compatibili con moveCARS, come spiegato nella sottosezione 3.2.1. Considerando il valore medio dei `rating`, e cercando di avere un dataset bilanciato, ho deciso di convertire tutti i `rating` con valore maggiore di 4 in feedback con valore 1, mentre i `rating` con valore 4 o minore in feedback negativi. Il risultato è un dataset con il 62% di feedback positivi (48604 data sample).

4.1.3 Feature degli oggetti

Le feature degli oggetti sono caratteristiche che descrivono le applicazioni usate dagli utenti di Frappe. Il dataset contiene tre feature rilevanti per il task di classificazione: la categoria dell'applicazione, la lingua, e il costo.

Category è la categoria delle applicazioni ottenuta dal Google Play Store. In Frappe le applicazioni sono divise in 23 categorie che descrivono la loro funzionalità principale (es. videogiochi, notizie, social). Come si può vedere in Figura 5, il numero di data sample per ogni categoria è molto sbilanciato: la categoria `communication` che comprende applicazioni di messaggistica come WhatsApp e Telegram ha più di 20k data sample. La seconda categoria più popolare è `social` che comprende applicazioni come Facebook, Instagram e Twitter.

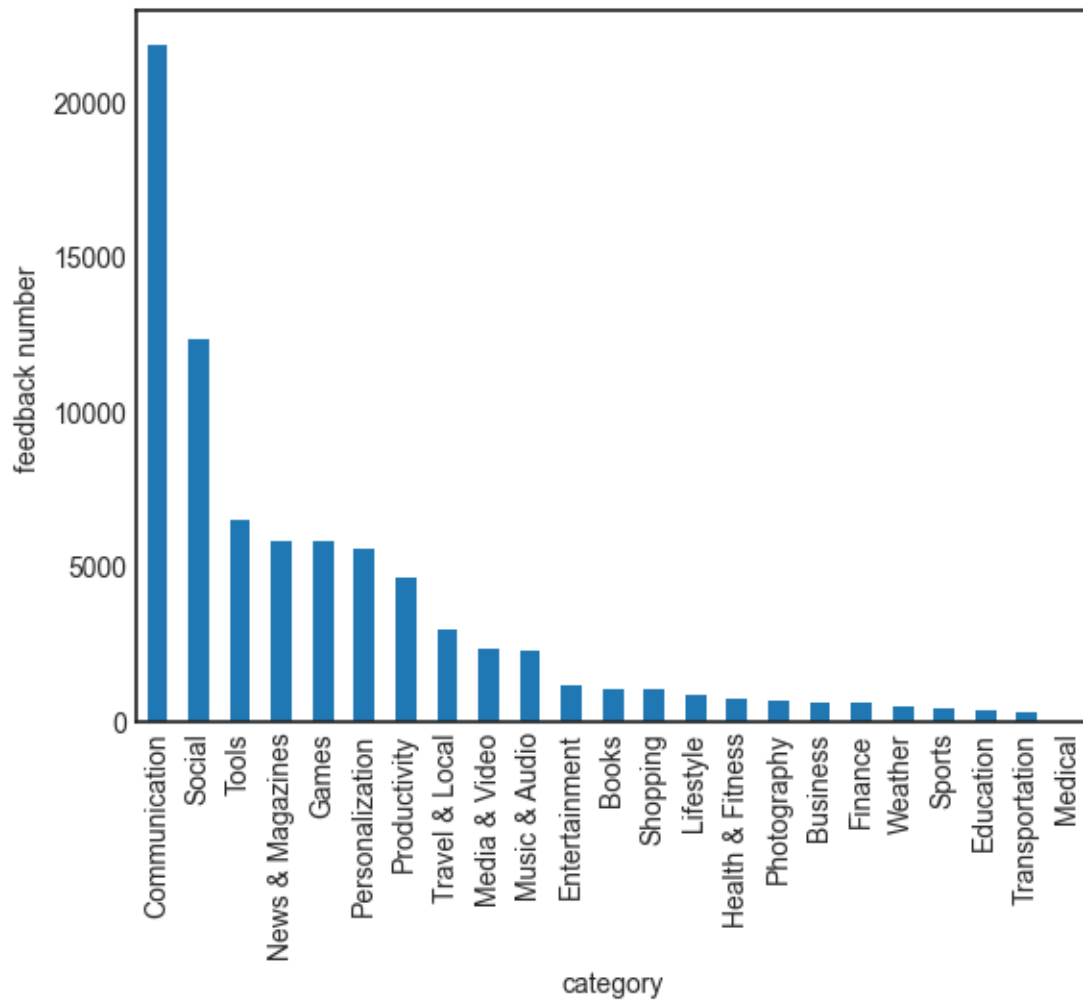


Figura 5: Numero di data sample per ogni categoria di applicazioni nel dataset Frappe

Language indica la lingua dell'applicazione ottenuta dal Google Play Store. Dato che il 96% dei data sample hanno applicazioni in lingua inglese, ho assegnato a tutte le altre lingue il valore `other`.

Cost indica se l'applicazione è gratuita o a pagamento.

Come per le feature di contesto, anche le feature degli oggetti sono variabili categoriche. Vengono codificate con one-hot encoding; il risultato è un vettore di 27 feature: 23 per la categoria, 2 per la lingua, e 2 per il costo.

4.1.4 Feature degli utenti

Il dataset Frappe non contiene nessuna informazione associata agli utenti oltre all'ID dell'utente. Per questo motivo, ho generato le feature degli utenti a partire dalle feature degli oggetti e di contesto.

User category indica la categoria di applicazioni più utilizzata dall'utente. Come era prevedibile dalla distribuzione delle categorie mostrata in Figura 5, le categorie preferite dagli utenti sono `communication` e `social`.

User weekday indica il giorno della settimana in cui l'utente ha generato il maggior numero di feedback. I giorni più popolari sono venerdì e sabato.

User daytime indica il momento della giornata in cui l'utente ha generato il maggior numero di feedback. I momenti della giornata più popolari sono la sera e il pomeriggio.

User weather indica la condizione meteo in cui l'utente ha generato il maggior numero di feedback. Le condizioni meteo più popolari sono `nuvoloso` e `soleggiato`.

User weekend indica se l'utente ha generato più feedback in settimana o nel weekend. La maggior parte degli utenti (85%), ha generato più feedback in settimana.

Uses paid apps indica se l'utente ha mai utilizzato un'applicazione a pagamento. Il 59% degli utenti ha utilizzato almeno una volta un'applicazione a pagamento.

Come le feature di contesto e degli oggetti, anche le feature degli utenti sono tutte variabili categoriche. Vengono codificate con one-hot encoding in un vettore di

47 feature: 22 per user category, 7 per user weekday, 7 per daytime, 7 per user weather, 2 per user weekend, 2 per uses paid apps.

Ricapitolando, il dataset Frappe processato contiene 78335 data sample, 857 utenti e 3180 oggetti. Oltre alle colonne user, item e feedback, il dataset contiene 47 feature degli utenti, 22 feature degli oggetti, e 24 feature di contesto.

4.2 My Digital Footprint

My Digital Footprint (MDF) [9] è un nuovo dataset composto dati di sensori di smartphone, informazioni di prossimità fisica, e interazioni sugli online social network. Il dataset include le misurazioni e le informazioni collezionate dai dispositivi personali di 31 volontari, nel loro ambiente naturale, senza limitare il loro comportamento usuale. Tutte queste informazioni compongono un insieme di informazioni completo per descrivere il contesto utente in ambiente mobile.

4.2.1 Negative sampling

MDF contiene l'utilizzo di applicazioni Android sul dispositivo utente per un periodo di due mesi. Per questo motivo esistono solo campioni positivi che indicano che un'applicazione è in esecuzione. Per eseguire il training di una rete neurale sono però necessari degli esempi negativi, i quali indicano che un'applicazione non era in uso da parte di un utente in una specifica situazione contestuale. Ogni data sample ha associata una label, che definisce il contesto dell'utente ad alto livello con i seguenti valori: `home`, `school`, `workplace`, `external school` (dove gli autori del dataset incontravano i volontari), `Free Time`, e `holiday`. Questa etichetta non è usata come feature di contesto, ma per fare negative sampling del dataset. L'algoritmo 1 mostra il procedimento: per ogni sample d nel dataset D vengono identificate le label in cui $d.item$ non è mai stato utilizzato (riga 2). Per ogni label n viene scelto in modo casuale un sample $\in D$ con label = n (riga 3). Di questo sample viene mantenuto solo il contesto $context_{neg}$ scartando user, item e feedback. Il sample negativo d_{neg} è ottenuto concatenando user e item di d , con 0 (il feedback negativo) e $context_{neg}$ (riga 6). d_{neg} è aggiunto al dataset D_{neg} che contiene solo sample negativi. In ultimo il dataset D_{neg} è unito al dataset D ed è eliminata la colonna corrispondente alle label.

4.2.2 Feature di contesto

Il dataset MDF contiene numerosi dati estratti dai sensori degli smartphone dell'utente. Dato che le feature sono davvero molte, di seguito sono elencate solo quelle selezionate che compongono il dataset finale.

Algoritmo 1 Negative sampling di MDF

Input: D - dataset**Output:** D_{neg} - dataset D with negative samples

```

1: for all  $d \in D$  do
2:    $labels_{neg} \leftarrow$  labels where  $d.item$  was never used
3:   for all  $n \in labels_{neg}$  do
4:      $context_{neg} \leftarrow$  context of a random data sample  $\in D$  with label =  $n$ 
5:      $feedback \leftarrow 0$ 
6:      $d_{neg} \leftarrow$  concatenate  $d.user$ ,  $d.item$ ,  $feedback$  and  $context_{neg}$ 
7:      $D_{neg} \leftarrow D_{neg} \cup d_{neg}$ 
8:   end for
9: end for
10:  $D_{neg} \leftarrow D \cup D_{neg}$ 
11:  $D_{neg} \leftarrow D_{neg}$  with label column dropped
12: Return  $D_{neg}$ 

```

Attività utente L'attività utente riconosciuta da Android Activity Recognition system⁵; include sia movimenti a piedi che su mezzi di trasporto. Le attività possibili sono `in vehicle`, `on bicycle`, `on foot`, `running`, `still`, `tilting`, `walking`, `unknown`. Ogni feature rappresenta la probabilità da 0 a 100 che l'utente stia facendo quell'attività specifica.

Modalità audio indica se la modalità audio del telefono è impostata su silenzioso, vibrazione o suono.

Volume Alarm volume, music volume, notification volume e ring volume, sono quattro feature con valore tra 0 e 1 che indicano il livello audio della sveglia, della musica, delle notifiche e della suoneria.

Musica `music active` è un valore booleano che indica se il dispositivo sta riproducendo della musica. Se la musica è riprodotta dallo speaker del telefono o dalle cuffie è identificato dalle feature `speaker on`, e `headset connected`.

Batteria Associate alla batteria ci sono due feature: `level` indica la carica della batteria (molto bassa, bassa, media, alta, carica) `charging` è un valore booleano che indica se la batteria si sta ricaricando oppure no.

⁵<https://developers.google.com/location-context/activity-recognition>

Schermo Associate al display ci sono due feature: `state` indica se il display è spento, acceso, o se si sta per spegnere. `Rotation` indica se l'utente sta usando il telefono in verticale o in orizzontale.

Meteo Il meteo è descritto da sei variabili diverse: temperatura, umidità, pressione atmosferica, velocità del vento, nuvole, e se ha piovuto nelle ultime tre ore.

Wifi La feature `connected` indica se il dispositivo dell'utente è connesso oppure no ad una rete Wi-Fi

Data e ora Dai timestamp dei feedback in formato `YYYY-MM-DD HH:MM:SS` sono calcolate come per il dataset Frappe le feature `daytime` (mattina, pomeriggio, sera, e notte) `weekday` e `isweekend`. Oltre a queste, con la libreria Python Holidays⁶, è calcolato se è un giorno di vacanza o no, in base al calendario delle festività italiano.

Feature sociali Le feature sociali sono ottenute dall'ego network descritta nella sottosezione 3.3.2. Ci sono quattro feature `social_c1`, `social_c2`, `social_c3` che corrispondono alle cerchie sociali della ego network, e indicano la percentuale di alter vicino all'utente nel momento in cui ha utilizzato un'applicazione. Il secondo gruppo di feature chiamate `layer_1`, `layer_2`, `layer_3`, `layer_4` indicano il layer dell'utente da cui è stata ricevuta la raccomandazione. Questo presuppone un dataset diverso per ogni utente, dato che l'ego network è personale per ogni utente (es. uno sconosciuto per l'utente 1, potrebbe essere un amico per l'utente 2).

Le feature categoriche sono codificate con one-hot encoding, mentre le feature numeriche sono normalizzate. Il risultato è un vettore che contiene 63 feature che descrivono il contesto fisico dell'utente, e 8 feature che descrivono il contesto sociale dell'utente.

4.2.3 Feature degli oggetti

Category è la categoria delle applicazioni ottenuta dal Google Play Store. In MDF le applicazioni sono divise in 26 categorie che descrivono la loro funzionalità principale (es. videogiochi, notizie, social). Come si può vedere dalla Figura 6, anche in MDF il numero di data sample per ogni categoria è molto sbilanciato: la

⁶<https://pypi.org/project/holidays/>

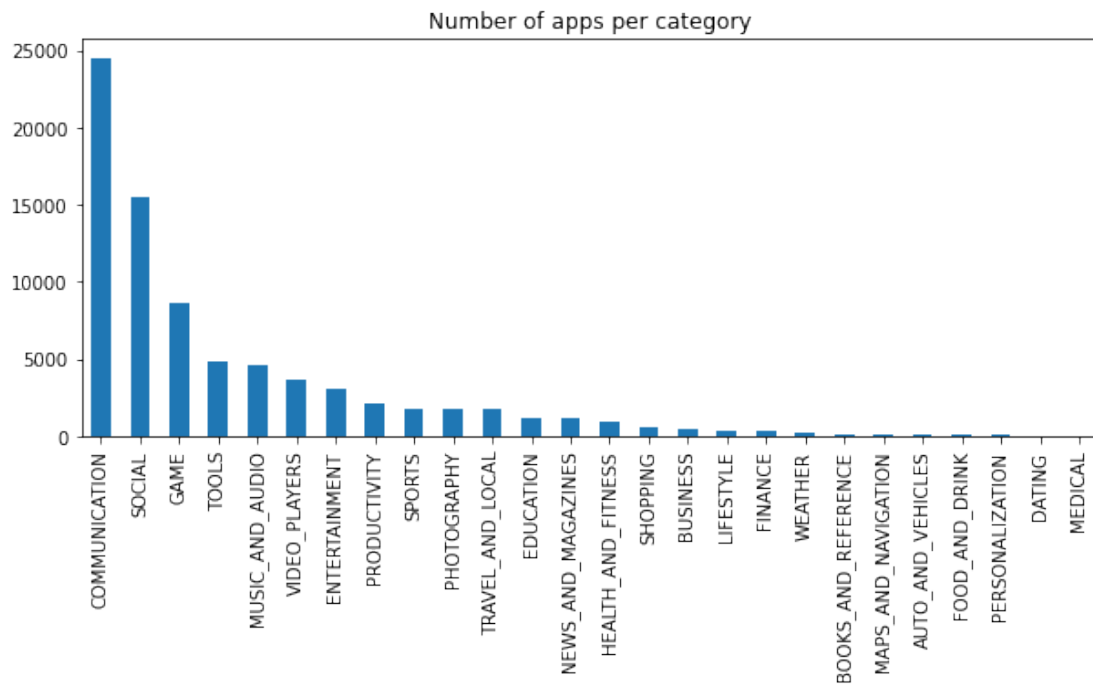


Figura 6: Numero di data sample per ogni categoria di applicazioni nel dataset My Digital Footprint

categoria communication che comprende applicazioni di messaggistica come Whatsapp e Telegram ha più di 20k data sample. La seconda categoria più popolare è social che comprende applicazioni come Facebook, Instagram e Twitter. La categoria è codificata con one-hot encoding.

4.2.4 Feature degli utenti

Come per Frappe, il dataset MDF non contiene nessuna informazione sugli utenti. Per questo motivo ho generato 4 feature utente dalle feature di contesto e degli utenti.

User category

User daytime

User weekday

User weekend

Capitolo 5

Risultati

Capitolo 6

Conclusioni

6.1 Conclusioni

Conclusioni...

6.2 Sviluppi futuri

Sviluppi futuri...

Bibliografia

- [1] Aurélien Géron. Hands-on machine learning with scikit-learn and tensorflow: Concepts. *Tools, and Techniques to build intelligent systems*, 2017.
- [2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [3] Xavier Glorot, Antoine Bordes, and Y. Bengio. Deep sparse rectifier neural networks. volume 15, 01 2010.
- [4] Jason Brownlee. How to choose an activation function for deep learning. <https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/>, 2021.
- [5] Daniel Godoy. Understanding binary cross-entropy / log loss: a visual explanation. <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a>, 2018.
- [6] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- [7] Unsupervised modelling of the user’s social context and visited locations at the edge of the internet.
- [8] Linas Baltrunas, Karen Church, Alexandros Karatzoglou, and Nuria Oliver. Frappe: Understanding the usage and perception of mobile app recommendations in-the-wild. 05 2015.
- [9] Mattia G. Campana and Franca Delmastro. Mydigitalfootprint: an extensive context dataset for pervasive computing applications at the edge. 2021.