CdL in Ingegneria Informatica

Progetto di sistemi basati su Deep Neural Network per la rilevazione di similarità tra password

Presentata da Karina Chichifoi

Relatore:
Marco Prandini

Correlatori:
Davide Berardi
Andrea Melis

Argomenti trattati

Perché usare password?

Garantire integrità e confidenzialità dei dati

Cambio Password

Il problema della scelta di password simili tra loro p.es. *password* vs *P@$$w0rd*

Scopo del progetto

Fornire un sistema di valutazione della similarità tra password basato sui Deep Neural Network

Valutazione progetto

Confronto dei risultati ottenuti con paper di riferimento di Bijeeta et alii

Come determinare la similarità tra password?

Analisi dei tasti premuti

Traduzione di maiuscole e caratteri speciali in sequenze di tasti premuti
p.es. *PASSword!* → *<c>pass<c>word<s>1*

FastText

Modello di word embedding che consente di capire la sfera semantica e la sintassi delle parole

N-gram

Analisi delle password in base alle sottostringhe della parola
p.es. *ciao* → *{ci, ia, ao}*

Paper di riferimento di Bijeeta et alii

Word2Keypress

Si tiene conto della sequenza di tasti premuti

N-gram

Numero minimo di n-gram a 1, massimo a 4

Euristiche

Pass2path per valutare il modello, basato su reti neurali ricorrenti

Implementazione

Password tra 4 e 30 caratteri

Rimozioni di bot

Rimozione di HEX e HTML non decodificati

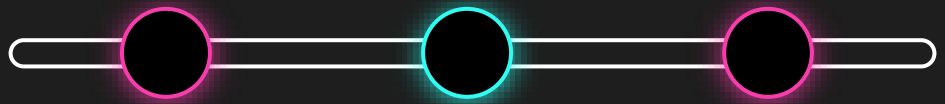
01 Filtraggio Dataset

02 Allenamento di FastText

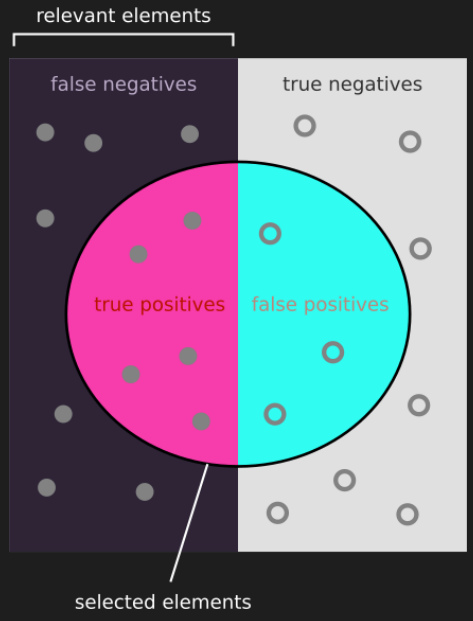
Allenato 5 modelli con variazioni di iperparametri

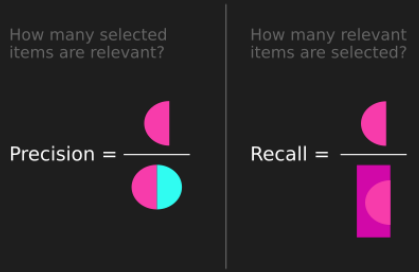
Maggiore portabilità per ambienti distribuiti

03 Compressione del modello



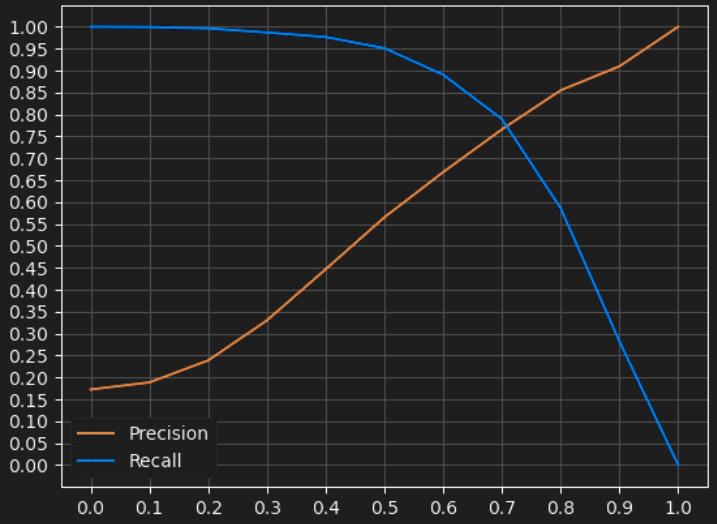
Valutazione dei risultati



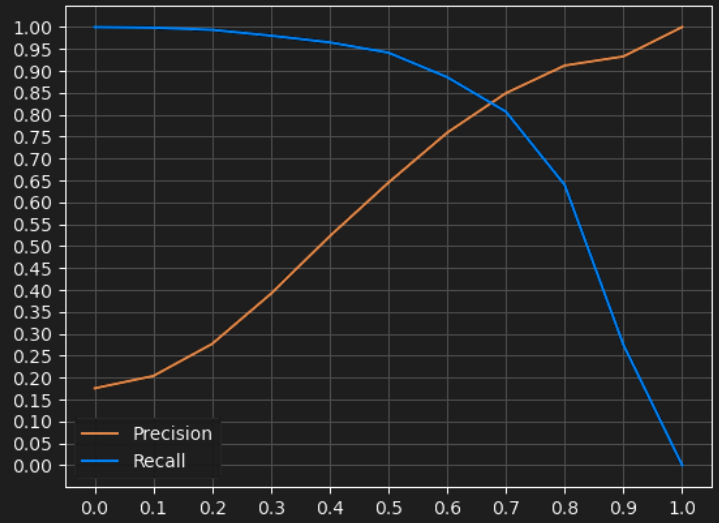


- Precision basso: imprecisa distinzione tra password simili e password non simili.
- Recall basso: comporta molte password simili non rilevate come tali.

Risultati



Paper di riferimento
Precision: 67%
Recall: 89%



Modello creato:
Precision: 77%
Recall: 89%

Conclusioni

Il miglioramento:

- Rimozione di word2keypress
- Analisi del numero ottimale di n_mingram
- Euristica diversa per il calcolo di precision e recall

Grazie a queste modifiche il valore di precision è stato migliorato del 10% rispetto al modello proposto, con valori analoghi di recall.

Sviluppi futuri

Architettura C/S

- Client: elaborazione locale password proposta
- Server: invio al client del modello
- DB: password rappresentata come word embedding

Confronto con Bloom Filter

Confronto prestazionale di precision e recall tra algoritmo probabilistico e modelli di word embedding