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

Scopo del progetto



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

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



N-gram

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



FastText

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

Paper di riferimento di Bijeeta et alii



Si tiene conto della sequenza di tasti premuti



N-gram

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



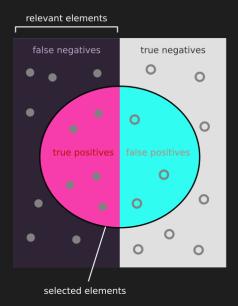
Euristiche

Pass2path per valutare il modello, basato su reti neurali ricorrenti

Implementazione



Valutazione dei risultati



How many selected items are relevant?

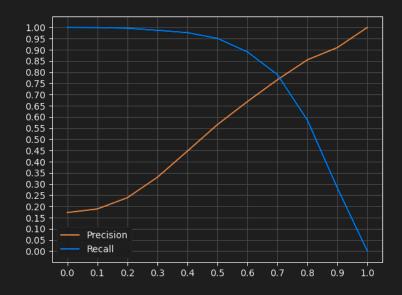
Precision =

Recall =

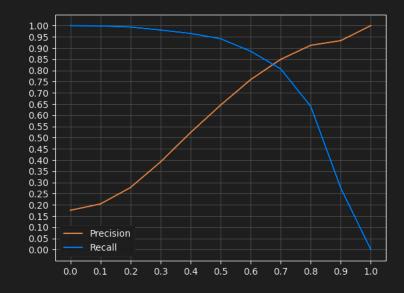
Recall =

- 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 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

