Keystroke Analysis

Project for the Data Base and Data Mining course

Paulin Loïs, Badin de Montjoye Xavier, Gaziello Yannis, Menet Hugo

11/05/2017

Outline

Introduction

Keystrokes analysis for biometrics recognition. Why?

Introduction

Keystrokes analysis for biometrics recognition. Why?

- Keyboard.

Introduction

Keystrokes analysis for biometrics recognition. Why?

- Keyboard.
- Fast.

Introduction

Keystrokes analysis for biometrics recognition. Why?

- Keyboard.
- Fast.
- Relatively simple.

Introduction

Keystrokes analysis for biometrics recognition. Why?

- Keyboard.
- Fast.
- Relatively simple.
- Reliable.

Introduction

What is the scope of keystroke analysis?

- Fixed text.

Introduction

What is the scope of keystroke analysis?

- Fixed text.
- Free Text.

Introduction

What is the scope of keystroke analysis?

- Fixed text.
- Free Text.
- Different languages.

Introduction

What is the scope of keystroke analysis?

- Fixed text.
- Free Text.
- Different languages.
- Long time between learning and testing.

State of the Art

- Compute distances between two typing sessions
- Consider time sequences of keystroke as patterns
- Not applicable for small free texts (Curse of dimensionality)

Outline

Measures

What can we measure in typing?

- Delay between pair of letters.
- Time on a key.
- Mean.
- Standard deviation.

Observations

There is a difference in the means (with or without quartile) between the texts typed by two different users.

Example

X_l : learning, X_t : test

- Distance $(A_l, B_t) = 459$
- Distance $(B_l, B_t) = 240$
- Distance $(A_l, A_t) = 43$
- Distance $(B_l, A_t) = 153$

Outline

Data Set Constructed

Data Set

7 persons.

Sample

- 6 entry of a same fixed phrase.
- 2 free text of around 60 words.

Example

Je certifie que cette soumission est le fruit de mon propre travail effectu en accord avec la Charte Anti-Plagiat.

Protocols

Protocol 1: Fixed Phrases

Training on 3 fixed phrases. Test on the 3 remaining.

Protocols

Protocol 1: Fixed Phrases

Training on 3 fixed phrases. Test on the 3 remaining.

Protocol 2: Free Text

Training on 1 free text. Test on 1 free text.

Outline

for *all apparition of a pair of letter in the test text* **do**
  for *all profiles p* **do**
    μ training mean for this pair and profile;
    σ training standard deviation for this pair and profile;
    d time for this apparition of this pair in the test text;
    δ parameter of the algorithm, 1.5 here;
    if $d = \mu \pm \delta\sigma$ **then**
      $S_p = S_p + 1$ for this profile;
    end
  end
end
Return profile with highest score

Algorithm 1: Closest profile

Outline

Results: protocol 1

A correct and an incorrect result

Name	lois:	Yannis :
lois	28	22
Yannis	16	23
Xavier	15	19
Angele	13	19
hugo	20	24
Emile	23	23
red	21	24

Results: protocol 1

Variation with delta for protocol 1

delta	correct/21
1	12
1.4	15
2	16
2.2	13

Results: protocol 2

Few but encouraging results

- lois: lois
- hugo: hugo
- Emile: Emile
- red: Emile

Conclusion

- Encouraging results, need more testing and narrowing our approach

Conclusion

- Encouraging results, need more testing and narrowing our approach
- Outlook
  - If two score are close, we could rerun with a different δ (in front of the standard deviation).

Bibliography I


Gunetti, Daniele and Picardi, Claudia
Keystroke Analysis of Free Text
ACM Trans. Inf. Syst. Secur., 8(3):312–347, 2005.


Maas, Andrew and Heather, Chris and Do, Chuong (Tom) and Brandman, Relly and Koller, Daphne and Ng, Andrew
Offering Verified Credentials in Massive Open Online Courses: MOOCs and Technology to Advance Learning and Learning Research
Ubiquity, 2:1–2:11, 2014.