Gender Prediction from Handwriting (ICDAR 2013)

Aakansh Gupta, Utsav Bansal
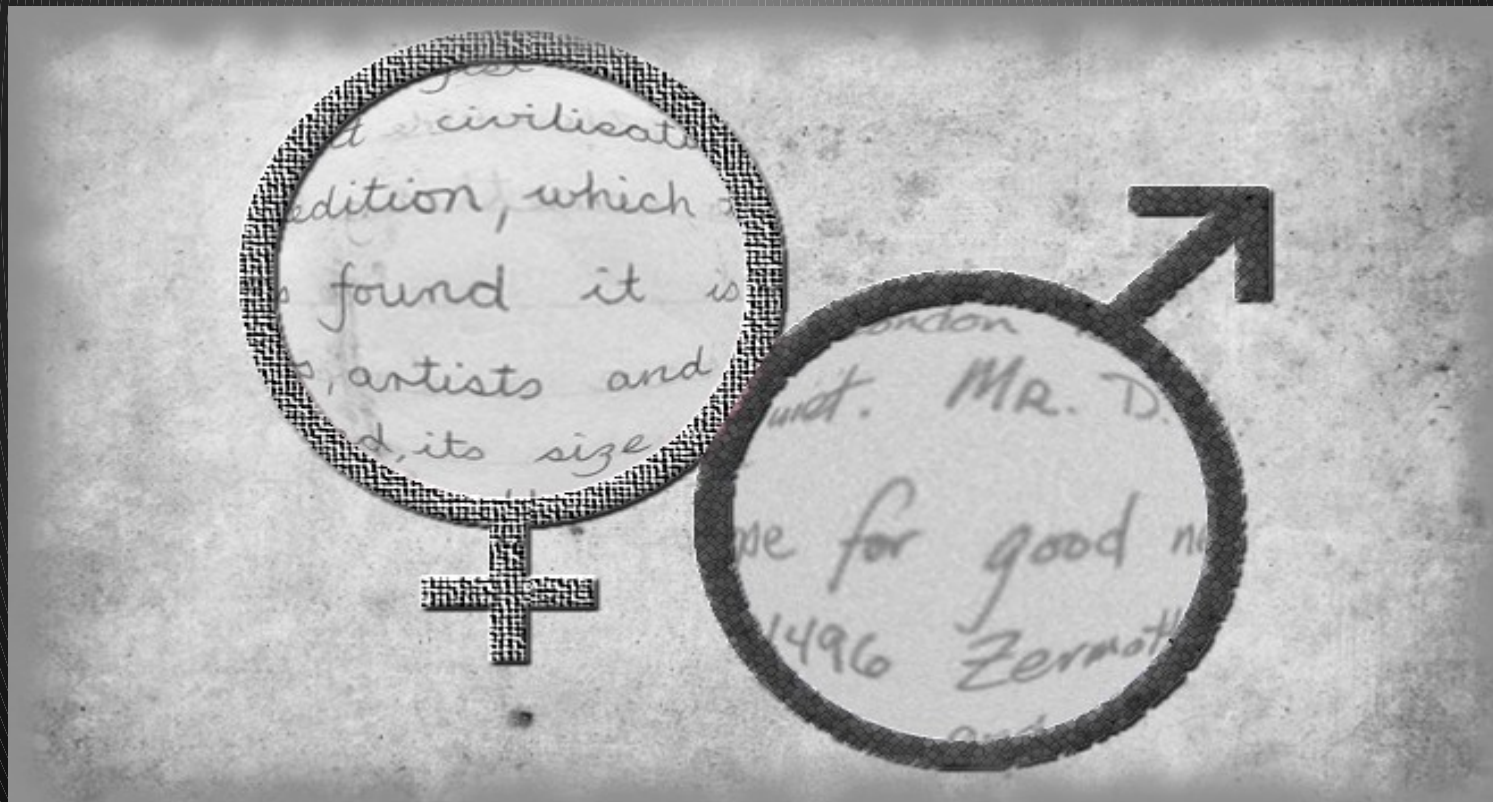
Dataset Images-282 writers

Arabic Page with different text

Arabic Page with same text

English Page with different text

English Page with same text

Features-7066/doc

Directions
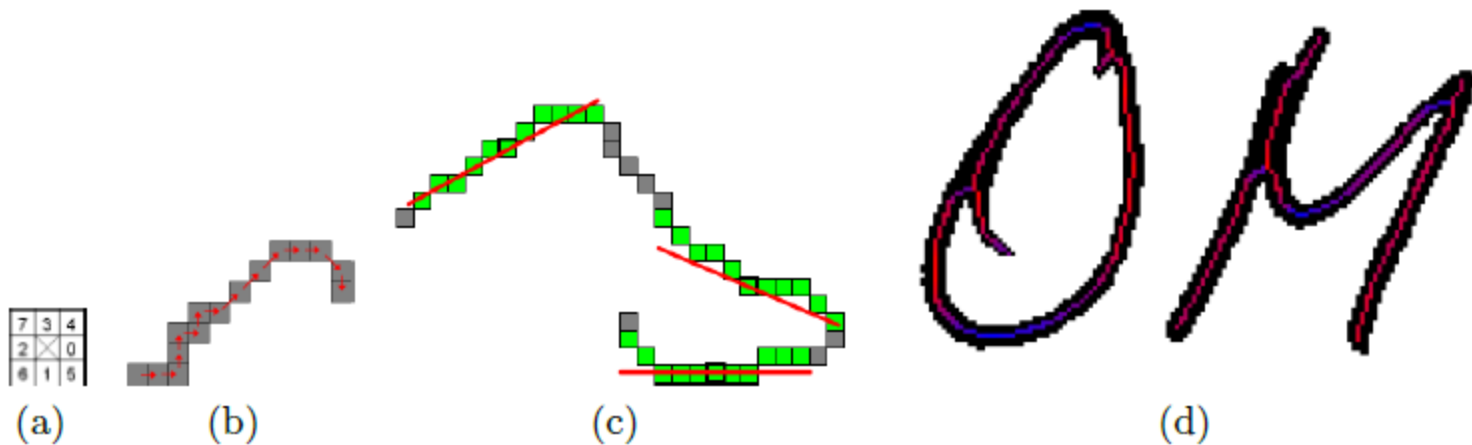
Fig. 1. Computing local directions. (a) The predefined order for traversing shapes. (b) Example of an ordered shape. (c) Estimating directions by computing linear regression of neighboring pixels. (d) Binary image and its corresponding Zhang skeleton: the red color corresponds to a $\frac{\pi}{2}$ tangent and the blue color corresponds to a zero tangent.

Curvatures
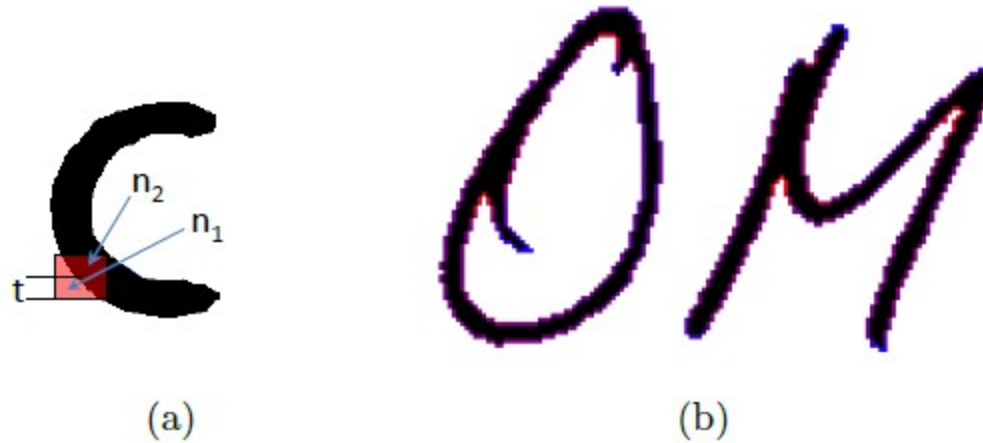
Fig. 2. (a) Computing curvatures. (b) Curvatures highlighted on binary image: red corresponds to the maximum curvature and blue corresponds to the minimum one.

Tortuosity

Fig. 3. Computing tortuosity (a) Longuest traversing segment for 4 different pixels. (b) Length of maximum traversing segment: red corresponds to the maximum length, blue to the minimum one.

Chain code features, edge based directional features, …

Implementation

- **Number of Features/Dimensions:** 7066
- **Training Data:** 200 people
- **Testing Data:** 82 people

  Training Results:

- Using LSSVM on raw data – 78% correct predictions

- Using 2 Layer feed forward network with 20 hidden neurons – 80% correct predictions. 0.54 log loss on Kaggle

Problems with results:

- Because of too many dimensions, but limited number of samples, our training methods were very prone to Over-Fitting.

- So, we went for dimension reduction and tried the following techniques:

- PCA, Fisher Reduction, Kruskal Wallis Analysis and Minimum Redundancy Maximum Relevance(MRMR) Feature selection.

Feature Reduction

This is how we fared:

- PCA: No improvement

- Fisher & Kruskal Wallis Analysis: Emphasis on *Curvature and Direction related features*, with moderate emphasis on Direction. No improve in accuracy by either SVMs or Neural Network method.

- MRMR: Strong Emphasis on *Chaincode and Curvature related features* , with moderate emphasis on Direction. Accuracy by SVMs and Neural Network Methods: 82%

Observations:

- Chaincode metrics tend to be higher for females and lower for males, same goes for Curvature.

- Direction metrics do not have a general variation trend but vary across individual parameters.

Thank You for Your Attention.