

Online aláírásfelismerés projekt

Bordé Sándor, Csernai Kornél

2011. december 11.

1. Feladat

Feladatunk egy online aláírásfelismerési verseny tanítóadataihoz versenyképes tanuló módszereket adni[1]. A verseny leírása a <http://forensic.to/webhome/afha/SigComp.html> oldalon található. Megoldásunkhoz csupán a "Holland online" adatbázist használhattuk.

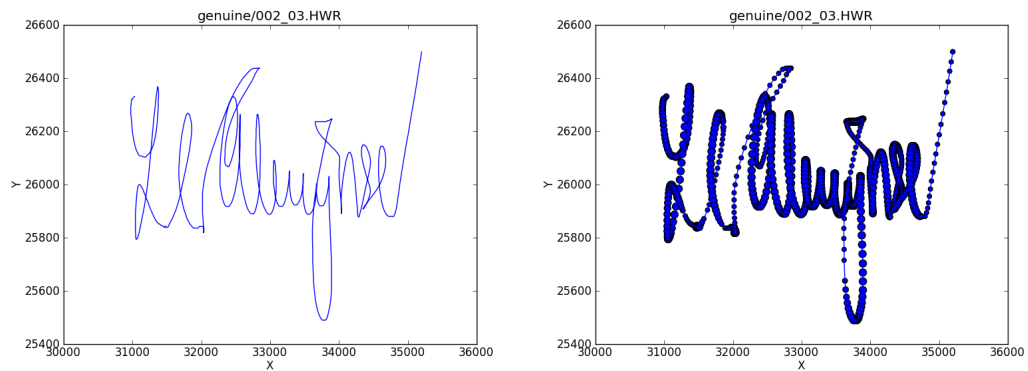
2. Az adatbázis

Az adatbázis WACOM Intuos3 A3 Wide USB Pen Tablet típusú eszközzel készült, amely x és y koordinátákat illetve nyomásértékeket rögzít, 200Hz frekvenciával.

Az adatokat letöltöttük, kicsomagoltuk, és áttekintettük. Az adatbázis tartalmaz egy valós aláírás adatbázist (*genuine*) szerzők egy csoportjától. Ezen szerzők aláírásait próbálták gyakorlott hamisítók reprodukálni (*forgery*).

TODO: hány szerző hány aláírás

Az aláírosakt vizualizáltuk, ezeket mutatják a következő ábrák.



1. ábra. Egy aláírás nyomásértékekkel és anélkül.

Tehát $3+3*3 = 12$ lokális attribútummal dolgoztunk. Impliciten jelen volt az idő értéke is, azaz, hogy hanyadik adatpontról van szó az aláírásmintán belül.

3.1.1. DTW

TODO: ez az évköziről lett kimásolva, frissíteni kell

- DTW algoritmust használtunk az összehasonlításra
- Interneten találtunk egy DTW implementációt
- A tanítóhalmaznak (a feladatkiírás ajánlása szerint) az aláírások felét választottuk (50-50)
- A tanítóhalmaz osztály meghatározza az elemek közötti távolságokat, előállítja a távolságmátrixot
- Eltárolja a legkisebb és legnagyobb távolságot, valamint kiszámítja az átlagos távolságot
 - globális tulajdonságok

TODO: itt már az ékezet is eltűnt...

- Az osztályozó ezután kiszámítja a kapott aláírástávolságot a tanító halmaz összes elemét.
- Itt is elmenti a min, max és átlagos távolságot
- Ez után osztályozza a teszt aláírást a tanítóhalmaz értékeihez hasonlítva
- Két típusú osztályozást neztünk meg
- Ha a teszt aláírástávolsága min, max, átlag értéke kisebb, mint a tanító halmazban található elemek távolságainak minimuma, maximuma, átlaga, akkor elfogadjuk
- Ha a fenti 3 tulajdonságból legalább kettő kisebb, mint a tanítóhalmaz értékei, akkor elfogadjuk

3.1.2. Improved DTW

3.2. Globális jellemzők

TODO: ezt előrébb vinni? mennyire használunk globális jellemzőt DTW-ben?

Minden egyes tanuló példán és bejövő teszt példán elvégezzük a lokális jellemzők aggregálását. Ez főképp a [2] alapján történt. A felhasznált jellemzők:

TODO: kisebb betűméret valahogy

Minták száma (1)	Magasság (2)	Szélesség (3)
1 / 3 (5)	Átlagos x sebesség (9)	Pozitív sebességű pontok száma / 14 (37)
Toll-lent minták száma (14)	Átlagos nyomás (20)	Maximum n sebességű pontok száma / 14 (37)
Maximum nyomás pontja (22)	Nyomás terjedeleme (28)	Maximum x sebességű pontok száma / 14 (37)
Átlagos x gyorsulás (30)	Maximum y sebesség (31)	Átlagos y sebesség (32)
Negatív sebességű pontok száma / 14 (34)	Pozitív sebességű pontok száma / 14 (37)	

1. táblázat. A felhasznált globális jellemzők. Zárójelben a [2] cikkben szereplő sorszám.

3.2.1. Gauss-görbék illesztésén alapuló módszer

Ezen módszer esetében a globális attribútumokat tartalmazó tanító vektorokra először normális eloszlást illesztünk:

$$\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)},$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)^2.$$

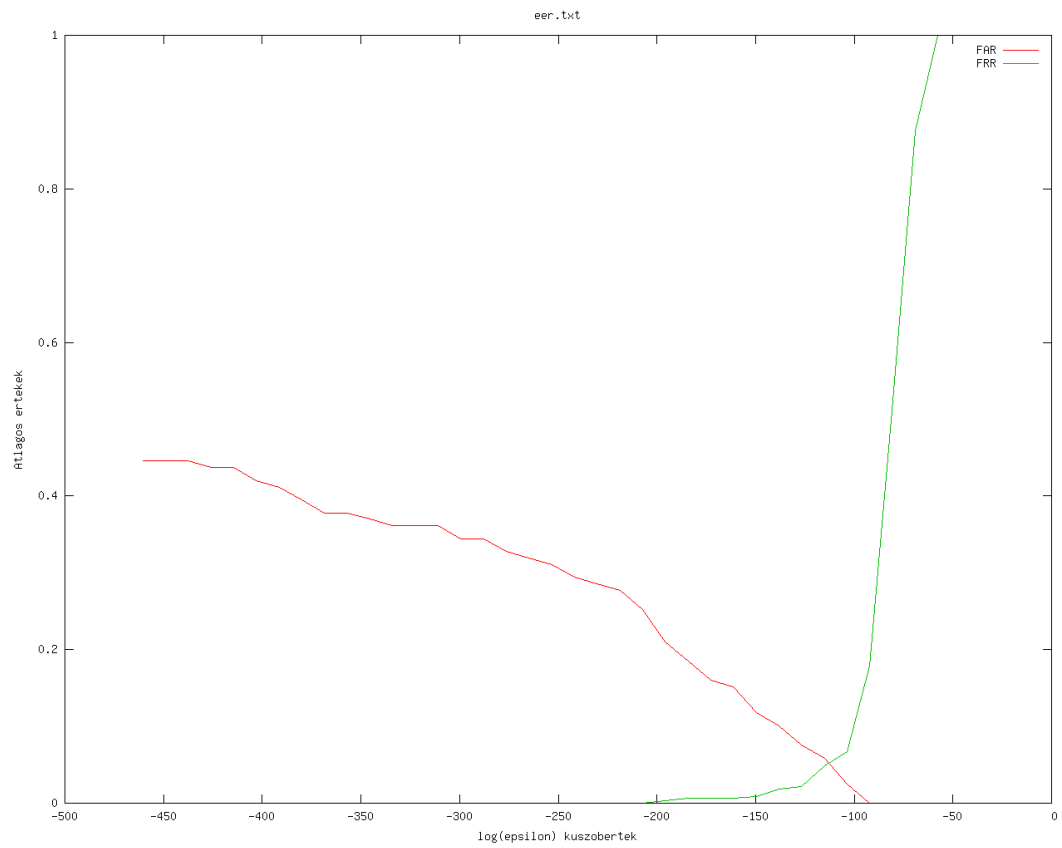
Ezután, ha egy ismeretlen teszt példányt kapunk, akkor elkészítjük annak globális jellemzővektorát (x), majd kiszámítjuk annak valószínűségét, hogy a komponensek külön-külön a megfelelő eloszlásba tartoznak:

$$p(x) = \prod_j^n p(x_j; \mu_j, \sigma_j^2) = \prod_j^n \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x_j - \mu_j)^2}{2\sigma_j^2}\right).$$

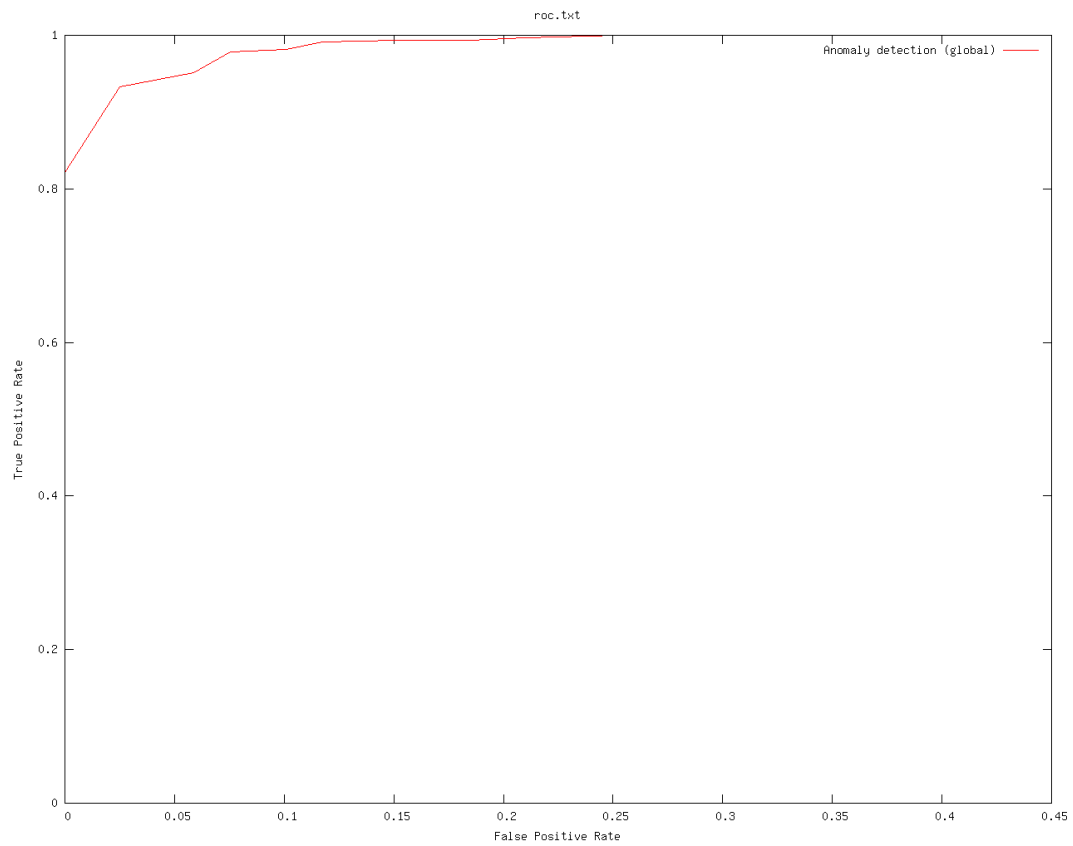
Választunk egy ϵ küszöbértéket, az x aláírást hamisítványnak tekintjük akkor és csak akkor, ha

$$p(x) < \epsilon.$$

Az ϵ érték kompromisszumot jelent, amely a FAR és FRR értékek közötti átmenetet meghatározza. A következő ábrákat keresztvalidációs kiértékeléssel kaptuk.



3. ábra. A globális módszer FAR és FRR görbéi.



4. ábra. A globális módszer ROC görbéje.

TODO: jó lenne közösen a lokális értékekkel

TODO: ábráról leolvasni az legjobb FRR&FAR és az EER értékeket; esetleg ROC görbe alatti területet számolni.

4. Tartalom

TODO: mindegyikről egy rövid leírás, fontosabb fileokról. használatról külön kellene a user doksiban.

- data/
- data-deriv/
- doc/
- Onsig/

- scripts/
- results/

5. Feladatmegosztás

5.1. Bordé Sándor

- DTW
- Improved DTW
- Dokumentáció
- TODO: bővíteni

5.2. Csernai Kornél

- Lokális jellemzők feldolgozása
- Globális jellemzők elkészítése
- Gauss-görbék illesztése
- FAR-, FRR-, és ROC görbék ábrázolása
- Dokumentáció
- TODO: bővíteni

Hivatkozások

- [1] Marcus Liwicki, Muhammad Imran Malik, C. Elisa van den Heuvel, Xiaohong Chen, Charles Berger, Reinoud Stoel, Michael Blumenstein, and Bryan Found. Signature verification competition for online and offline skilled forgeries (sigcomp2011). In *ICDAR*, pages 1480–1484. IEEE, 2011.
- [2] Jonas Richiardi, Hamed Ketabdar, and Andrzej Drygajlo. Local and global feature selection for on-line signature verification. In *In Proc. IAPR 8th International Conference on Document Analysis and Recognition (ICDAR 2005)*, pages 625–629, 2005.