

Adataalapú megoldások

Döntési fák

és

Jósági indikátorok osztályozásnál

Osztályozás

- Videógyűjtemény
Spam-videó
(osztályozási feladat)
- *Ismétlés: osztályozás
vs. klaszterezés*

<i>Felügyelt tanulás</i>	Független változó magyarázó	Függő változó cél v.
Tanuló állomány	ismert	ismert
Cél állomány	ismert	?

Döntési fák

- Classification (Decision) Tree: kategória célvált.
- Regression (Decision) Tree: folytonos célvált.
csomóponton belüli variancia minimalizálása
- ID3: Iterative Dichotomiser
- C4.5
- C5.0
- CART: Classification And Regression Tree
- CHAID: Chi-squared Automatic Interaction Detector

Döntési fák általános működése

Iteratívan minden csomópontnál:

1. Kiválasztják a legjobb változót
tisztasági mérték alapján
1. Ez alapján felbontják a sokaságot
2. Megvizsgálják, hogy vége van-e

ID3 algoritmus

- Információs entrópia alapján:
- $I_E = -\text{SUM}(p * \log_2 p)$
- $I_{\text{split}} = \text{SUM} [(n_i/n) * I_E]$
- $\text{Gain} = I_E - I_{\text{split}}$

Az az attribútum lesz választva, amelyiknek az információs nyeresége (gain) a legnagyobb.

C4.5 algoritmus

ID3-hoz képest:

- Folytonos vált. is (diszkretizálva)
 - Hiányos adatokra is.
 - Fanyesés
 - Információs nyereség arány alapján:
 - $I_{\text{partition}} = -\text{SUM}((n_i/n) * \log_2(n_i/n))$
 - $\text{GainRatio} = \text{Gain} / I_{\text{partition}}$
- Attribútum: Max információs nyereség arány

C5.0 algoritmus

C4.5-hez képest:

- Gyorsabb
- Kevesebb memória
- Kisebb fa
- Különböző osztályozási hibákhoz különböző súlyok rendelhetők

CART (C&RT) algoritmus

- Szétválasztás a Gini index alapján:
- $Gini() = 1 - \sum (p_i * p_i)$
- Eredetileg folytonos változóra, de kategória, bináris változóra is.

Attribútum: Min Gini index

IBM Intelligent Miner használja

CHAID algoritmus

- χ^2 alapján
- SAS, SPSS használja
- Nem bináris fa -> szélesebb, mintha bináris lenne
- Nincs utólagos metszés

Kiértékelés osztályozási feladatoknál

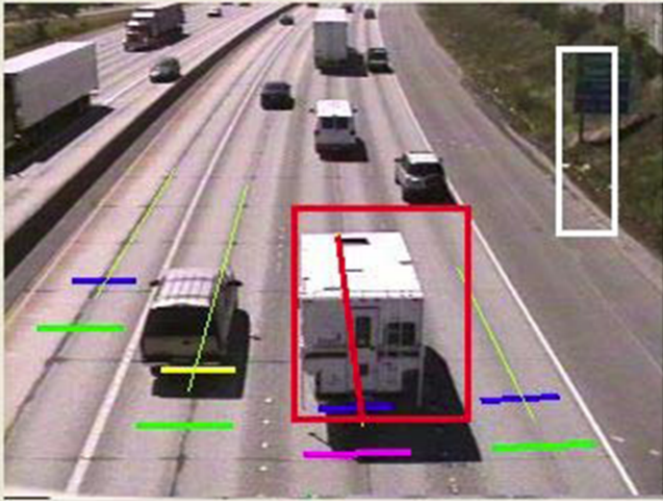
Kiértékelési indikátorok: EER, ROC görbe, AUC, Lift, stb.

Járművek osztályozása

Video-Based Vehicle Detection and Classification System

File View Tools Help

Image Sequence
C:\Documents and Settings\guohui\Desktop\I-5_145\I-5_145_00



Vehicle Data | Image Frames

Line No.	Trucks	Total
1	0	7
2	4	41
3	16	37
4	3	25
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
Total	23	110

Trucks Detected: 23
Vehicles Detected: 110
Elapsed Time: 0

Realtime Application | Static Image Stream Application

Open < > Run Stop Reset Close

Detecting Vehicles...

Osztályozási feladat: arc detektálás + szegmentálás



Város/tájkép osztályozási feladat

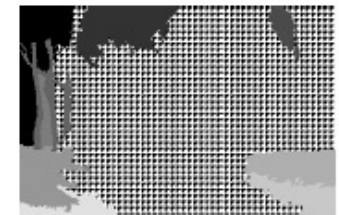
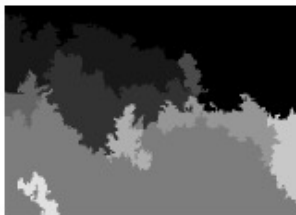
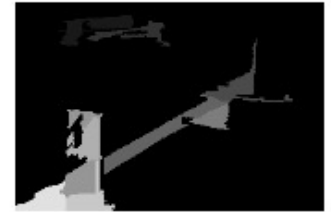
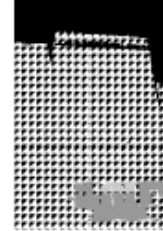
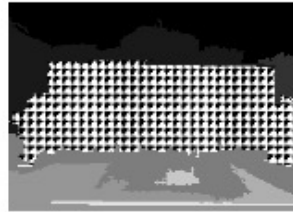
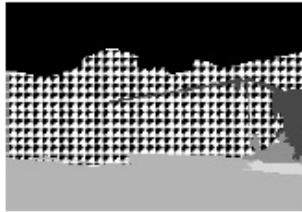
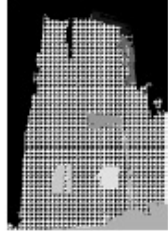


Figure 2: Sample results of city/landscape image classification experiments

Feladat: ég detektálás + szegmentálás

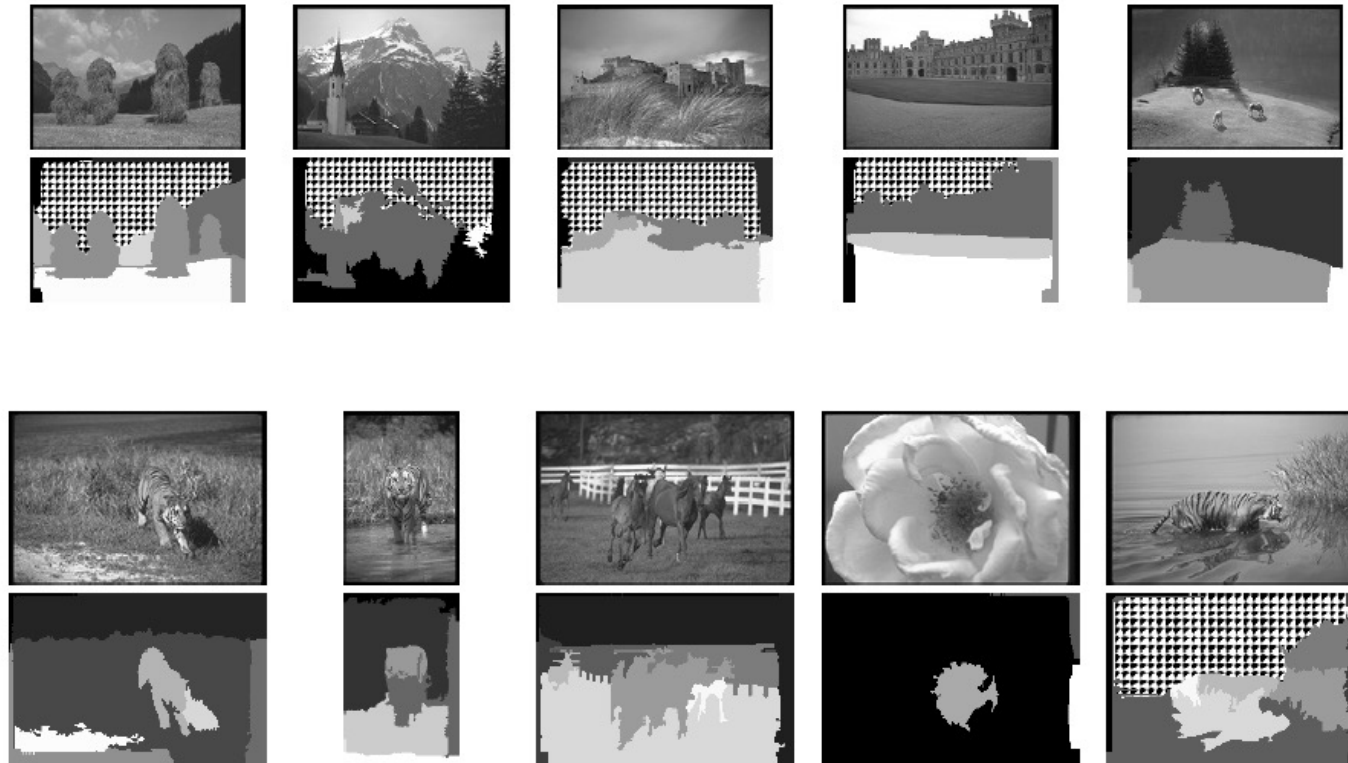


Figure 3: Sample results of classification experiments for images containing/not containing sky

Kiértékelési indikátor: IOU (metszet és unió hányadosa)

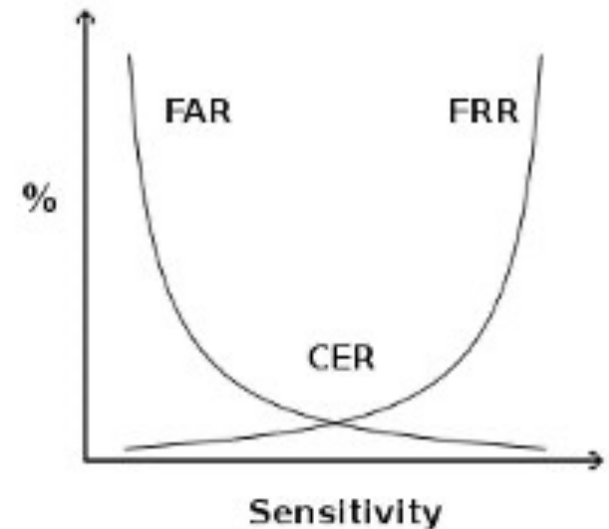
Osztályozók kiértékelő indikátorai

- Fedés (recall)
- Pontosság (precision)
- Pontosság (accuracy)
- F_1 mérték
- EER
- ROC görbe, AUC
- Lift

EER

EER (Equal Error Rate) = Crossover Error Rate (CER)

- FAR (False Accept Rate)
- FRR (False Reject Rate)



ROC

ROC: Receiver Operating Characteristic

Bináris osztályozás jóságának mérésére alkalmas

- vízszintes: $FPR = FP / (FP + TN)$
- függőleges: $TPR = TP / (TP + FN)$

Választunk egy küszöbértéket az elemek T és F szétválasztására. Egy küszöbértékre kapunk egy pontot az FPR-TPR síkon. Az összes küszöbértékhez tartozó pontokat összekötött vonal a ROC görbe.

AUC

- Csak bináris osztályozás esetén.
- Minden elemhez rendeljünk olyan 2 pontszámot (1-et az egyik, 1-et a másik osztályhoz), mely jelzi az osztályba tartozás erősségét.
- Számítsuk ki a 2 pontszám arányát (*likelihood ratio*) úgy, hogy $r = \text{Score}(C_1) / \text{Score}(C_2)$.
- r alapján állítsuk csökkenő sorrendbe őket.
- **Rangsorolás jóságát** mutatja a köv. indikátor:
1-es osztályból véletlenszerűen kivéve egyet (x_i),
2-es osztályból egyet (y_k), mi lesz a valószínűsége,
hogy $r(x_i) > r(y_k)$?
- $\text{AUC} = P(r(x_i) > r(y_k))$

Lift görbe

bizonyosság alapján csökkenő sorrendbe helyezett elemek adott részhalmazára megadja, hogy hányszorosára nőtt a pozitívak aránya

- vízszintes: osztályozandó elemek részaránya
- függőleges: lift érték (hányszorosára nőtt az teljes populáció arányához képest)

Covid teszt



ÉRVÉNYTELEN: A kontroll sávban (C) sem látható színes vonal. Az elégtelen mintamennyiség vagy a nem megfelelő eljárás a legvalószínűbb okai ennek. Olvassa el még egyszer a használati utasításban a tesztelés menetét, és ismételje meg azt egy új teszttel, vagy lépjen kapcsolatba az orvosával, vagy forduljon COVID-19 tesztközpontoz.

KORLÁTOZÁSOK

1. A tesztelési lépések be nem tartása pontatlan eredményhez vezethet.
2. A COVID-19 antigén gyorsesz (nyáteszt) kizárólag in vitro - külsőleges - diagnosztikai önellenőrzésre szolgál.
3. A teszteredményt más laboratóriumi vizsgálatokból és kiértékelésekből származó egyéb klinikai leletekkel együtt kell figyelembe venni.
4. Ha a teszteredmény negatív vagy reakció nélküli, miközben a klinikai tünetek fennállnak, ez azért lehetséges, mert a nagyon korai vírusfertőzés még nem mutatható ki. Ajánlott 1-2 nappal később újratestelni egy új teszttel, vagy kórházba menni a fertőzés kizárása érdekében.
5. COVID-19 pozitív eredmény adódhat nem-SARS-CoV-2 koronavírus törzsekkel való fertőzés vagy egyéb zavaró tényezők miatt is.

TELJESÍTMÉNYJELLEMZŐK

Klinikai teljesítmény. Klinikai értékelést végeztek, amelyben összehasonlították a COVID-19 antigén gyorsesz segítségével kapott eredményeket az RT-PCR vizsgálati eredménnyel. A klinikai vizsgálatban 406 nyálmintát vizsgáltak. Az eredmények 99,3%-os specifikitást és 90,1%-os szenzitivitást mutattak ki, 97,0%-os összesített pontosságot.

	PCR-rel megerősített minta	Helyesen beazonosított	Arány
Pozitív minta	101	91	90,1% (Szenzitivitás)
Negatív minta	305	303	99,3% (Specifitás)
Összesen	406	394	97,0% (Összesített pontosság)

90,1% szenzitivitás: Az összesen 101 PCR-rel megerősített pozitív mintából 91-et igazolt vissza helyesen pozitívnak a COVID-19 antigén gyorsesz, és 10-et tévesen negatívnak.

99,3%-os specifikitás: Az összesen 305 PCR-rel megerősített negatív mintából 303-at igazolt vissza helyesen negatívnak a COVID-19 antigén gyorsesz, és csak 2-t tévesen pozitívnak.

97%-os pontosság: Az összesen 406 PCR-rel megerősített mintából 394 esetben a COVID-19 antigén gyorsesz eredménye megegyezett a PCR-teszt eredményével.

A megfigyelt pontosság a vírus populációban való előfordulásának mértékétől függően változhat.

Megnevezés	Koncentráció
Adenovirus type 3	$3,16 \times 10^4$ TCID ₅₀ /ml
Adenovirus type 7	$1,58 \times 10^4$ TCID ₅₀ /ml
Human coronavirus OC43	1×10^5 TCID ₅₀ /ml
Human coronavirus 229E	5×10^5 TCID ₅₀ /ml
Human coronavirus NL63	1×10^5 TCID ₅₀ /ml
Human coronavirus HKU1	1×10^5 TCID ₅₀ /ml
Influenza A H1N1	$3,16 \times 10^4$ TCID ₅₀ /ml
Influenza A H3N2	1×10^5 TCID ₅₀ /ml
Influenza B	$3,16 \times 10^4$ TCID ₅₀ /ml
Parainfluenza vírus 2	$1,58 \times 10^4$ TCID ₅₀ /ml

	PCR-rel megerősített minta	Helyesen beazonosított	Arány
Pozitív minta	101	91	90,1% (Szenzitivitás)
Negatív minta	305	303	99,3% (Specifitás)
Összesen	406	394	97,0% (Összesített pontosság)