

Hledání vícerozměrných vztahů mezi vedlejšími příznaky léčby a rychlostí metabolizace

Antonín Pavelka

2. března 2010

Seminar on knowledge discovery

Osnova

- “ Představení dat
- “ Cíle analýzy
- “ Metody
- “ Výsledky
- “ Závěr

Úvod

- “ pacienti s depresí, neurózou nebo bulimií
- “ léčení běžným léčivem (paroxetin)
- “ lék vyvolává řadu vedlejších příznaků
- “ rychlost odbourávání léčiva je individuální
 - . a tím i potřebná dávka

Rychlost odbourávání léčiva

- “ závisí na hodnotě Metabolic Ratio (MR)
 - . měřitelná, ale obtížně
 - . jen indikátor, měří metabolizace jiné látky, ne přímo léku
 - . hodnota je spojitá, ale lékaři pacienty rozdělili na PM (poor metabolizer) a EM (extensive metabolizer)
 - . před léčbou je často vyšší než při podávání léku
- “ pomalé odbourávání -> vyšší koncentrace -> menší dávky
- “ hodnota MR je ovlivněna dvojicí genů
 - . lze poměrně levně určit jejich varianty

Cíle

- “ **geny -?-> MR -?-> vedlejší příznaky**
- “ ví se: některé geny ovlivňují třídu pacienta (PM nebo EM)
- “ třída vedlejších příznaků (sexuální dysfunkce) souvisí s MR¹
 - . jednorozměrné statistické testy
 - . další jednorozměrné závislosti nebyly statisticky významné
- “ **Lze z genotypu předpovědět MR?**
- “ **Lze z kombinace ostatních atributů předpovědět MR?**
 - . **Souvisí MR s jinými vedlejšími příznaky?**

¹A. Žourková, et al. (2005). Clinical Impact of CYP2D6 Activity in Long-Term paroxetine Treatment. New Developments in Anxiety Disorder Research. Chapter 7

Datový soubor pro vedlejší příznaky

- “ 91 pacientů
- “ atributy
 - . 24 vedlejších příznaků
 - “ soustředění, únava, nausea, průjem, příbytek, pocení: {0, 1, 2}
 - “ ASEX ((Arizona Sexual Experience Scale)
 - . 5 kritérií
 - . každé hodnoceno 5 stupni
 - . cílový atribut: lékaři diskretizované MR při léčbě
 - “ EM – 42 pacientů
 - “ PM – 49 pacientů

Metody

Averaged One-Dependence Estimators (AODE)

- . slabší předpoklad o nezávislosti atributů, než naivnímu bayes
 - ” přidává závislost dvojic atributů (ODE)
 - ” výpočetně mírně náročnější
- . trénuje více ODE a průměruje jejich predikce
- . často dosahuje lepších výsledků než naivní bayes
- . **proč:** stále velmi rychlý, odolný proti přeučení
- . Weka implementace vyžaduje diskretizaci
 - ” bez učitele, 2 koše

G. Webb, J. Boughton, Z. Wang (2005). Not So Naive Bayes: Aggregating One-Dependence Estimators. Machine Learning. 58(1):5-24.

Výběr atributů - úvod

“ může zvýšit

- . správnost většiny učících algoritmů
- . interpretovatelnost naučeného konceptu
- . rychlost učení

I. H. Witten, E. Frank (2005). Data mining: Practical Machine Learning Tools and Techniques: 288-296.

Výběr atributů - přístupy

“ metoda filtru

- . nezávislá na učícím algoritmu
- . obecné charakteristiky dat

“ metoda obalu

- . učící algoritmus je vložen do procesu výběru

Výběr atributů – vybrané metody

“ **filtrové**

- . **množinové**: hodnotící kvalitu množiny atributů

- “ **CfsSubsetEval**

- . **atributové**: hodnotící přínos jednotlivých atributů

- “ **ChiSquaredAttributeEval**

“ **obalové**

- “ **WrapperSubsetEval**

CfsSubsetEval

- “ hodnocení množiny pomocí
 - . prediktivní síly jednotlivých atributů
 - . jejich vzájemnou závislost

ChiSquaredAttributeEval

- “ hodnotí atributy
 - . podle závislosti s cílovým atributem
 - . vygeneruje pořadí atributů
- “ nenáročný
 - . neprohledává se prostor podmnožin
 - . potřeba jen určit počet atributů

ClassifierSubsetEval

- “ kvalita množiny = správnost klasifikátoru na ní natrénovaném

WrapperSubsetEval

- “ kvalita množiny = správnost klasifikátoru na množině atributů určená křížovým ověřením

Výběr atributů – validace

- “ meta.AttributeSelectedClassifier
 - . vybere atributy zvolenou metodou
 - . navenek se chová jako normální klasifikátor
- “ 10-násobné křížové ověření
- “ náhodné umístění cílového atributu

Náhodné umístění cílového atributu

“ test, zda námi dosažená správnost může být nahodilá (a tedy nesvědčí o závislosti v datech)

proházet



nezávislý na A1, A2



A1	A2	T
5	3	B
1	2	A
6	7	B
8	6	B
1	1	A

skutečná data



A1	A2	T
5	3	A
1	2	B
6	7	B
8	6	A
1	1	B

randomizovaná data

Výsledky

Úplné prohledání 1

“ 70,8 % všech 25 atributů

“ všechny podmnožiny < 5

. 71,4 % {asexcelk, asex3}

 {asex1, asex3, asex4}

. 72,5 % {asex?, asex?, asex?, ?}

Úplné prohledání 2

- “ bez atributů ASEX (Arizona Sexual Experience Scale)
- “ všech 17 atributů
 - . ZeroR: 53.8 %
 - . AODE: 44,0 %
- “ výběr atributů:
 - . 62,6 % {anxieta, tenze}, {anxieta, sny}
 - . 68,1 {anxieta, soustr, sny, aka}
 - . randomizace: správnosti nejsou statisticky významné
- “ významné vícerozměrné vztahy kromě ASEX nenalezeny

CfsSubsetEval

křížové ověření:

	number of folds (%)	attribute
Acc = 68,1 %	10 (100%)	asex3
bez ASEX:	10 (100%)	asex1
Acc = 35,2 %	9 (90 %)	asexcelk
(< baseline)	7 (70 %)	prujem
	5 (50 %)	asex4
	5 (50 %)	soustr
	4 (40 %)	anxieta
	4 (40 %)	insomnie
	3 (30 %)	asex5
	2 (20 %)	poceni
	2 (20 %)	pribytek
	1 (10 %)	sucho

ChiSquaredAttributeEval

křížové ověření:

	average merit	average rank	attribute
Acc = 71,4 %	12.122 +- 1.728	1.5 +- 0.5	asexcelk
bez ASEX:	11.905 +- 2.287	2.4 +- 1.28	asex1
Acc = 44,0 %	10.893 +- 2.425	2.9 +- 1.04	asex3
(< baseline)	9.735 +- 2.063	4.1 +- 1.58	asex5
	9.357 +- 1.403	4.8 +- 0.75	dysfunkce
	8.162 +- 1.831	6.1 +- 0.94	asex2
	8.06 +- 1.656	6.2 +- 1.08	asex4
	3.606 +- 1.132	9 +- 1.41	soustr
	2.57 +- 0.662	10.9 +- 3.36	insomnie
	2.42 +- 0.469	11.5 +- 1.28	tenze
	2.517 +- 0.919	11.6 +- 3.56	anxieta

WrapperSubsetEval

křížové ověření:

Acc = 70,3 %

bez ASEX:

Acc = 53,8 % (= baseline)

number of folds (%)	attribute
9 (90%)	asexcelk
9 (90%)	asex3
9 (90%)	soustr
8 (80%)	asex4
5 (50%)	sny
4 (40%)	asex1
2 (20%)	anxieta
2 (20%)	asex5
2 (20%)	tenze
2 (20%)	insomnie
2 (20%)	sucho
2 (20%)	pribytek

Shrnutí

- “ Existuje vztah mezi některou podmnožinou vedlejších příznaků a rychlostí, s jakou pacient odbourává léčivo?
- “ 2 metody hodnocení množiny
 - “ CfsSubsetEval, WrapperSubsetEval
- “ 1 metoda hodnocení atributu
 - “ ChiSquareAttributeEval
- “ 2 způsoby prohledávání prostoru podmnožin
 - “ vyčerpávající prohledávání
 - “ BestFirst

Závěr

Data

- “ potvrzena známá závislost ASEX atributů a MR
- “ nenalezeny žádné jiné závislosti

Metody

- “ výběr atributů
 - . nezlepšil správnost predikce, ale ukázal důležité atributy (ASEX)
- “ použité metody hodnocení a prohledávání dávaly podobné výsledky
 - úplné prohledání
 - umožňuje zobrazit všechny kvalitní množiny
 - vestavěné heuristické metody Weky
 - snadná validace díky AttributeSelectionClassifier

Výhled

- “ predikovat MR z genotypu
- “ zkusit regresi

Děkuji za pozornost.

Vedlejší příznaky

- " gd-trs - deprese, neuróza a bulimie
- " CGI1 - clinical global impression-severity of illness scale, závažnost onemocnění
- " délka th - délka léčby v měsících
- " dávka - denní dávka paroxetinu, mg
- " Anxieta 0 - nepřítomna 2- středně silně přítomna
- " ASEX celk. 5-18 není sex. dysfunkce
- " ASEX 1 libido
- " ASEX 2 vzrušení
- " ASEX 3 erekce/lubrikace
- " ASEX 4 orgasmus
- " ASEX 5 satisfakce
- " dysfunkce (ASEX celk. binárně) 0- nepřítomna 1-přítomna
- " pohlaví 1-muž 2-žena
- " další vedl.účinky 0-nepřítomny 1-slabě 2-středně silně
- " soustředění
- " únava
- " somnol. - somnolence
<http://lekarske.slovníky.cz/pojem/somnolence>
- " tenze
- " insomnie
- " sny
- " tremor
- " aka - tak tohle nevím
- " parest. - parestezie, brnění
<http://lekarske.slovníky.cz/pojem/parestezie>
- " sucho
- " nausea
- " průjem
- " ortost. - možná pokles tlaku při vzpřímení
<http://lekarske.slovníky.cz/pojem/ortostaticky>
- " pocení
- " cefalgie - bolest hlavy
- " příbytek

ASEX

- “ zhodnocení sexuálních poruch psychiatrických pacientů a lidí se zdravotními problémy
- “ 1. Sexual drive
- “ 2. Arousal
- “ 3a. Penile erection;
- “ 3b. Vaginal lubrication
- “ 4. Ability to reach orgasm
- “ 5. Satisfaction from orgasm.

ASEX – Likertova stupnice

- “ 1 Strongly disagree
- “ 2 Disagree
- “ 3 Neither agree nor disagree
- “ 4 Agree
- “ 5 Strongly agree