DSP Final Project

劉義瑋
Blueswen Liu

資料挑戰賽

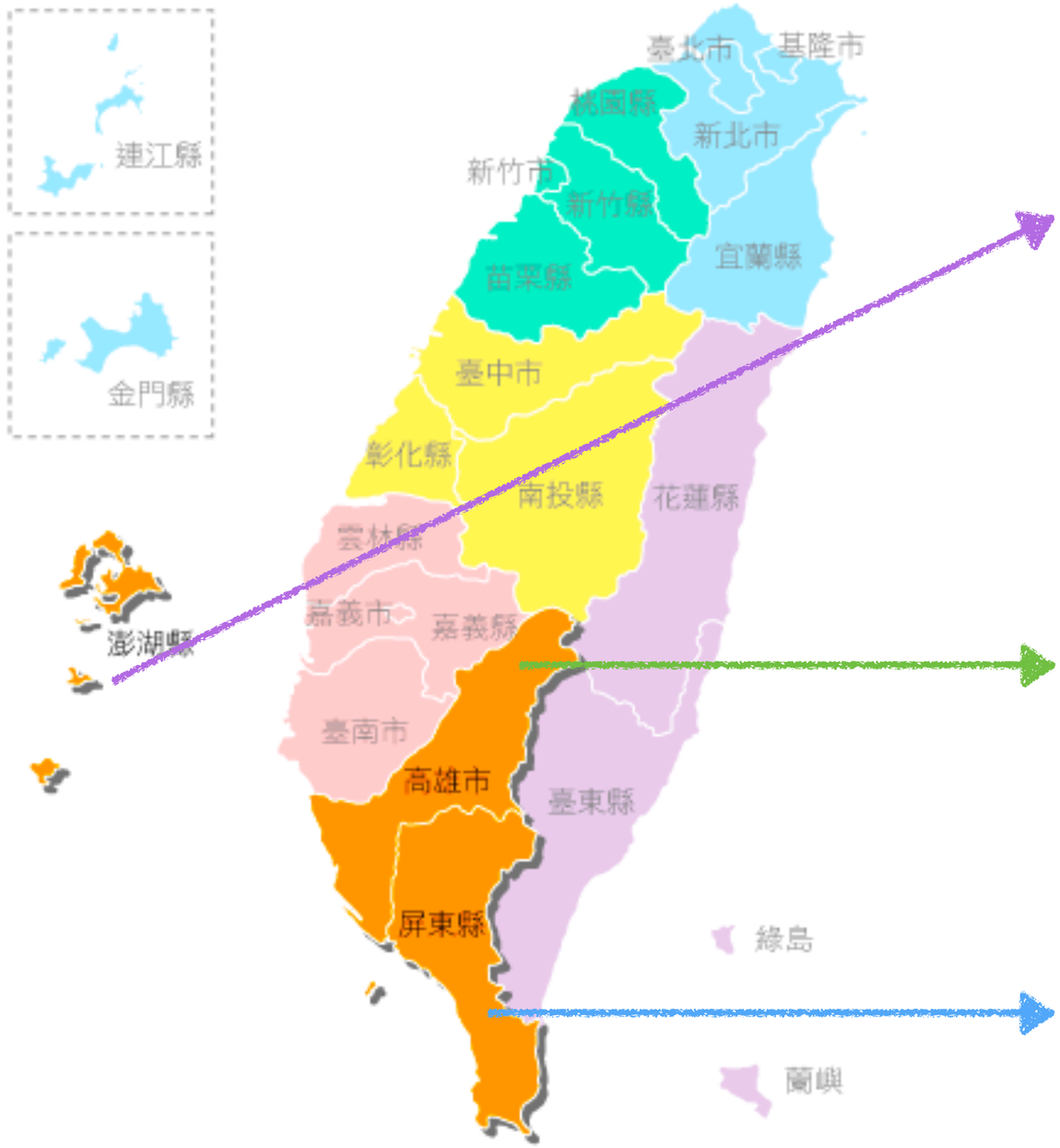
DSP創造新世代的數據力量

KAMERA Emergency Medicine Challenge

http://dc.dsp.im/main/content/KAMERA-Emergency-Medicine-Challenge

KAMERA

- 高屏緊急醫療應變聯盟
- Kaoping Area Medical Emergency Response Alliance
- KAMERA established in 2010. They want to combine the resources of hospital with first-aid capabilities and pre-hospital care to supply real time resource matching and support decision optimization. Main service area is the Kaoping area now, but in the future they want to expand to other area.



Penghu

Kaohsiung

Pingtung

Background

- Emergency department congestion is representative of a dysfunctional system
- There are many reasons, like the abusing of emergency medicine resource or dysfunctional information system
- KAMERA developed a trans-hospital emergency department congestion real time surveillance system. KAMERA provide some historical data from this system, and they hope they can move from surveillance to prediction.

Data

- Part of KAMERA's partners' emergency department's operating data
- Data Time: 2013/01/01 ~ 2013/12/31
- Number of Hospitals : 11
- Records: 21364 records
- Fields: 37 fields

Data

date	date of record
tz	time of record [1]
Hospital_PK	code of hospital
Level	level of hospital [2]
PDR	patient doctor ratio
PBR	patient bed ratio
total	total of all triage level
Light	emergency department congestion light

A01~A03	number of visits of Med., Surg., Ped.	A15	number of waiting bed of EM
A04~A06	waiting lists of Med., Surg., Ped.	A16	current longest waiting time for bed (mins)
A07~A10	number of 5 triage level	A17	number of respirator using
A12	number of appointment of EM	B01~B05	Empty bed of different department in ICU
A13	number of discharges by EM	B06	number of waiting ICU
A14	number of hospitalizations of EM	C01~C06	number of surgery

[1] time of record, separate one day to six part: [0,4), [4,8), [8,12), [12,16), [16,20), [20,24)
[2] the level of first-aid capabilities (1 represent first-rate capabilities, 2 represent middling first-aid capabilities)

Goal

- Predict
  - total of all triage level (total)
  - emergency department congestion light (light)
  - number of appointment of EM (A12)
  - number of waiting bed of EM (A15)
  - ~~current longest waiting time for bed (mins) (A16)~~

```
      A16
Min.    :  -837.0
1st Qu.:     0.0
Median :   695.5
Mean   : 11450.4
3rd Qu.:  3738.2
Max.   :774454.0
```

Method

- Tool: libSVM
- Training : Validation
  - 90:10, 80:20, 70:30, 60:40, 50:50
- K-Fold Cross Validation
  - 5 folds and 10 folds
- Measurement
  - Morisita-Horn Similarity
  - Adjusted R Squared

Morisita-Horn Similarity

$$C_M = \frac{\sum_{i=1}^{S} 2X_i Y_i}{\sum_{i=1}^{S} X_i^2 + \sum_{i=1}^{S} Y_i^2}$$

S: Number of Unique Species
Xi: Reference Value of Species i
Yi: Predict Value of Species i

- Morisita-Horn Similarity used to compare overlap among samples
- Morisita-Horn Similarity could consider normalize Euclidean distance to [0,1]
- Reference

1.Horn, H.S. (1966). “Measurement of “overlap” in comparative ecological studies”. American Naturalist, 100, 419-424.

2.Morisita, M. (1959). “Measuring of interspecific association and similarity between communities”. Memoirs of the Faculty of Science, Kyushu Univ., Series E (Biology), 3, 215-235.

Feature Selection

date	date of record
tz	time of record [1]
Hospital_PK	code of hospital
Level	level of hospital [2]
PDR	patient doctor ratio
PBR	patient bed ratio
total	total of all triage level
Light	emergency department congestion light

A01~A03	number of visits of Med., Surg., Ped.	A15	number of waiting bed of EM
A04~A06	waiting lists of Med., Surg., Ped.	A16	current longest waiting time for bed (mins)
A07~A10	number of 5 triage level	A17	number of respirator using
A12	number of appointment of EM	B01~B05	Empty bed of different department in ICU
A13	number of discharges by EM	B06	number of waiting ICU
A14	number of hospitalizations of EM	C01~C06	number of surgery

[1] time of record, separate one day to six part: [0,4), [4,8), [8,12), [12,16), [16,20), [20,24)
[2] the level of first-aid capabilities (1 represent first-rate capabilities, 2 represent middling first-aid capabilities)

Quick Test

- Random pick 3000 records to Train
- Random pick 3000 records to Test
- Predict light, total, A12, A15
- Explore which feature is important
- Feature
  - date -> month
  - tz -> number

Features

date	date of record
tz	time of record [1]
Hospital_PK	code of hospital
Level	level of hospital [2]
PDR	patient doctor ratio
PBR	patient bed ratio

Quick Test

A	B	C	D
target	type	removedfeature	MHS
total	test	date	0.9984
total	test	tz	0.9981
total	test	none	0.9980
total	test	PDR	0.9960
total	test	Hospital_PK	0.9920
total	test	Level	0.9919
total	test	PBR	0.9892

成績排行

排名	分數	參賽者	上傳時間(GMT +8)
1	99.07600	b10009047	2016/04/09-23:57:19
2	99.04100	cmchen	2016/04/10-18:15:12
3	98.87900	yi	2016/04/10-11:06:07

PDR	patient doctor ratio
PBR	patient bed ratio

Dummy Variable

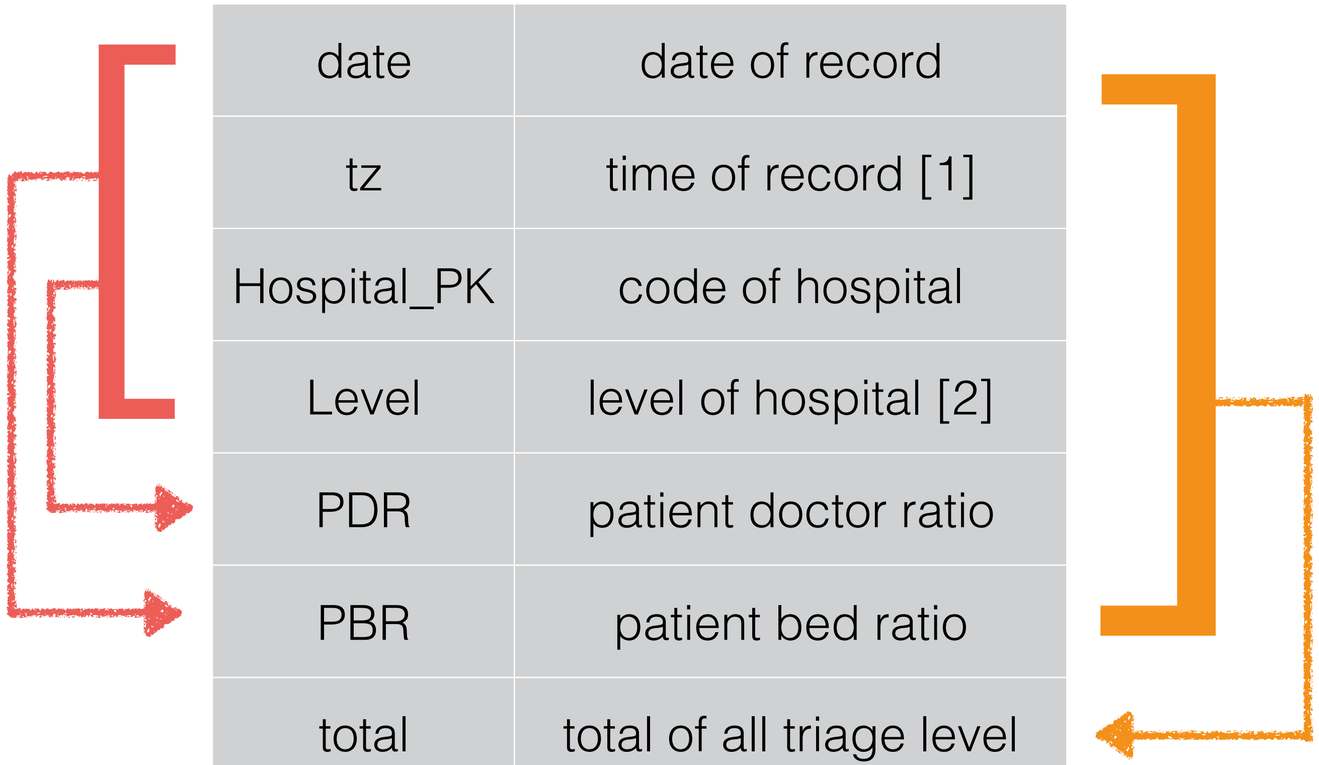
date	date of record
tz	time of record [1]
Hospital_PK	code of hospital

with R Package dummies

Dummy Coding

Level of race	race.f1 (1 vs. 2)	race.f2 (1 vs. 3)	race.f3 (1 vs. 4)
1 (Hispanic)	0	0	0
2 (Asian)	1	0	0
3 (African American)	0	1	0
4 (Caucasian)	0	0	1

Change Strategy



date	date of record
tz	time of record [1]
Hospital_PK	code of hospital
Level	level of hospital [2]
PDR	patient doctor ratio
PBR	patient bed ratio
total	total of all triage level

Result

target	type	note	MHS	AdjustedRSquared
PDR	test	remove none	0.9654	0.9450
PBR	test	remove none	0.9543	0.6971
total	test	remove none	0.9819	0.9559
total	test	add PDR,PBR	0.9813	0.9624

Quick Test with Dummy Variable

target	type	note	MHS	AdjustedRSquared
PDR	test	remove none	0.9599	0.8454
PBR	test	remove none	0.8767	0.3265
total	test	remove none	0.9718	0.9182
total	test	add PDR,PBR	0.9755	0.9421

Quick Test without Dummy Variable

Result

target	type	note	MHS	AdjustedRSquared
PDR	test	remove none	0.9654	0.9450
PBR	test	remove none	0.9543	0.6971
total	test	remove none	0.9819	0.9559
total	test	add PDR,PBR	0.9813	0.9624

Quick Test with Dummy Variable

target	type	note	MHS	AdjustedRSquared
PDR	test	remove none	0.9599	0.8454
PBR	test	remove none	0.8767	0.3265
total	test	remove none	0.9718	0.9182
total	test	add PDR,PBR	0.9755	0.9421

Quick Test without Dummy Variable

Result

target	TrainingPercent	folds	trainMHS	testMHS	calibrationMHS	AdjustedRSquared
total	0.8	5	0.9849	0.982	0.9836	0.9429
total	0.8	10	0.9877	0.9815	0.9828	0.9936
total	0.7	5	0.9859	0.9811	0.9826	0.9294
total	0.9	5	0.9851	0.9810	0.9822	0.9353
total	0.5	5	0.9856	0.9815	0.9822	0.9528

Without prediction PBR, PDR as features

target	TrainingPercent	folds	trainMHS	testMHS	calibrationMHS	AdjustedRSquared
total	0.8	5	0.9855	0.981	0.9812	0.9589
total	0.9	5	0.9867	0.9812	0.9803	0.9341
total	0.7	5	0.9852	0.9807	0.9801	0.9603
total	0.8	10	0.9885	0.9801	0.9794	0.9685
total	0.7	10	0.9882	0.9801	0.9792	0.9552

With prediction PBR, PDR as features

Result

target	TrainingPercent	folds	trainMHS	testMHS	calibrationMHS	AdjustedRSquared
total	0.8	5	0.9849	0.9820	0.9836	0.9429
total	0.8	10	0.9877	0.9815	0.9828	0.9936
total	0.7	5	0.9859	0.9811	0.9826	0.9294
total	0.9	5	0.9851	0.9810	0.9822	0.9353
total	0.5	5	0.9856	0.9815	0.9822	0.9528
total	0.6	5	0.9859	0.9818	0.9821	0.9484
total	0.6	10	0.9892	0.9809	0.9809	0.9461
total	0.9	10	0.9883	0.9802	0.9808	0.9274
total	0.7	10	0.9870	0.9802	0.9807	0.9096
total	0.5	10	0.9891	0.9800	0.9799	0.9060

排名	分數	參賽者	上傳時間(GMT +8)
1	99.07600	b10009047	2016/04/09-23:57:19

Result

target	type	note	MHS	AdjustedRSquared
PDR	test	remove none	0.9599	0.8454
PBR	test	remove none	0.8767	0.3265
total	test	remove none	0.9718	0.9182
total	test	add PDR,PBR	0.9755	0.9421

Quick Test without Dummy Variable

target	TrainingPercent	folds	trainMHS	testMHS	calibrationMHS	AdjustedRSquared
total	0.8	5	0.9849	0.9820	0.9836	0.9429

Final Result

Thanks for listening