


H.R. POURREZA 1



سناریی ماشین

ارزیابی سیستم

ممدرفضا پوررضا



مقدمه

2

- آفرین گام در ساخت یک سیستم، ارزیابی آن است

H.R. POURREZA

تعریف برخی اصطلاحات

MV Lab

3

- **Degrees of Freedom (DOF):** این اصطلاح در سیستمهای ردیاب استفاده می‌شود و بنا به تعریف عبارتست از آزادی حرکت انتقالی یا دورانی به طوری که ردیابی کامل در مسیر حرکت صورت پذیرد.
- **Measurement Rate:** تعداد اندازه‌گیری‌های معتبر در واحد زمان.
- **RMS Error:** نمایشی از فضای اندازه‌گیری که بر اساس ریشه دوم میانگین مسابجی توان‌های دوم مولفه‌های خطا می‌باشد.
- **Resolution:** رزولوشن کوچکترین گامی است که می‌توانیم در اندازه‌گیری یک پارامتر اندازه بگیریم یا بیان کنیم. به عنوان مثال اگر یک مرات سنج دیجیتال، درجه مرارت را بصورت مثلا 21.3 درجه بیان کند، رزولوشن آن 0.1 درجه است. رزولوشن یک سیستم اندازه‌گیری ضرورتاً دقت قرائت نیست.

H.R. POURREZA

تعریف برخی اصطلاحات

MV Lab

4

- **Precision (دقت):** به این معنی است که در اندازه‌گیریهای متوالی از یک مقدار، چقدر مقادیر اندازه‌گیری شده به هم نزدیکند. اگر یک ساعت روزانه 15 ثانیه جلو برود، ولی این اتفاق هر روز بیافتد، Accuracy این ساعت کم است ولی Precision این ساعت فوق‌العاده زیاد است.



- **Bias:** در شکل زیر Precision زیاد ولی Accuracy کم است. بایاس تفاوت بین میانگین مقادیر اندازه‌گیری شده با مقدار واقعی است. بایاس زیاد به معنی کم بودن دقت است.



H.R. POURREZA

MV Lab

تعریف برخی اصطلاحات

5

Accuracy (صحت):

- به این معنی است که مقدار اندازه‌گیری شده بقدر به مقدار واقعی نزدیک است. برای Accuracy بالا بایستی Precision زیاد باشد ولی عکس این امر ضرورت ندارد. بالا بودن واریانس مقادیر اندازه‌گیری شده و بایاس هر دو به معنی Accuracy کم است. بنابراین مجموع مربع بایاس و واریانس به عنوان میزان Accuracy مورد استفاده قرار می‌گیرد.



Latency:

- زمان تأخیر بین لمذهای که یک تغییر ماصل می‌شود تا زمانی که مس می‌شود.

H.R. POURREZA

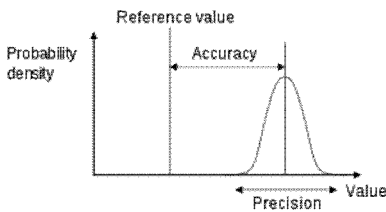
MV Lab

تعریف برخی اصطلاحات

6

Accuracy در مقابل Precision

- Precision پایین و Accuracy پایین
- Accuracy کم و Precision بالا



H.R. POURREZA



تعریف برخی اصطلاحات


7

□ **Sensitivity و Specificity:**

این دو پارامتر کمیت آماری برای ارزیابی یک کلاس بند باینری است. Sensitivity که به نام نرخ Recall هم از آن نام می برند عبارتست از کسری از جوابهای مثبت است که به درستی تشخیص داده شده اند (مثلا درصد افراد بیمار اعلام شده که بیماری را واقعا دارا هستند). در حالیکه Specificity عبارتست از کسری از جوابهای منفی است که به درستی تشخیص داده شده اند (مثلا درصد افراد سالم اعلام شده که بیماری را واقعا ندارند).

در بازایی اطلاعات از دو پارامتر Precision و Recall استفاده می شود. در این موزه برای استفاده از یک کمیت برای ارزیابی سیستم از F-measure که میانگین هارمونیک Precision و Recall است استفاده می شود: $F = 2 \times \frac{Precision \times Recall}{Precision + Recall}$

H.R. POURREZA




محاسبه می‌یارهای ارزیابی

8

□ برای یک کلاس بند باینری:

		وضعیت واقعی	
		Positive	Negative
نتیجه کلاس بند	Positive	True Positive (TP)	False Positive (FP) (خطای نوع 1)
	Negative	False Negative (FN) (خطای نوع 2)	True Negative (TN)

H.R. POURREZA



حسابی معیارهای ارزیابی

9

Recall یا Hit Rate . True Positive Rate . Sensitivity □
 $TPR = TP / P = TP / (TP + FN)$


Fall-out یا False Alarm Rate . False Positive Rate □
 $FPR = FP / N = FP / (FP + TN)$

Accuracy □
 $ACC = (TP + TN) / (P + N)$

True Negative Rate یا Specificity □
 $SPC = TN / N = TN / (FP + TN) = 1 - FPR$

Precision یا Positive Predictive Value □
 $PPV = TP / (TP + FP)$

H.R. POURREZA



حسابی معیارهای ارزیابی

10

Negative Prediction Value □
 $NPV = TN / (TN + FN)$

False Discovery Rate □
 $FDR = FP / (FP + TP)$

H.R. POURREZA

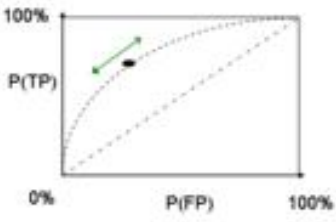
حسابی میاریابی ارزیابی

MV Lab

11

:Receiver Operating Characteristic (ROC) □

- رسم Sensitivity یا TPR بر مسب False Positive Rate (FPR) یا (1-Specificity) برای یک کلاس بند باینری
- ROC نمایش benefits (TPR) بر مسب Cost (FPR) است
- ممور افقی و عمودی ROC بین صفر تا یک رسم می شود.
- بهترین نقطه در این فضا گوشه بالا سمت چپ است که به معنی Sensitivity با مقدار 100% و Specificity با مقدار 100% است.



H.R. POURREZA

حسابی میاریابی ارزیابی

MV Lab

12

ضریب کاپا (kappa coefficient) □

برای ارزیابی توافق بین دو نظردهنده (که می تواند یکی ماشین باشد) مورد استفاده قرار می گیرد.

این ضریب از رابطه زیر محاسبه می شود


$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

که در آن $\Pr(a)$ درصد توافق مشاهده شده بین نظردهنده ها و $\Pr(e)$ احتمال فرضی شانس توافق است. $k=1$ به معنی توافق کامل و $k=0$ به معنی عدم توافق است.

فرض کنید که پرونده 50 مراجعه کننده به یک کلینیک توسط پزشک و ماشین مورد ارزیابی گرفته و در فصوص سالم و یا بیمار بودن آنها اعلام نظر شده است. مایلیم ارزیابی میزان توافق بین نظر پزشک و ماشین را انجام دهیم. فرض کنید نتیجه ارزیابی بصورت جدول زیر باشد

		نظر ماشین	
		سالم	بیمار
نظر پزشک	سالم	20	5
	بیمار	10	15

H.R. POURREZA



محاسبه‌ی میاریابی ارزیابی

13

□ ضریب کاپا (kappa coefficient)

□ بدین ترتیب درصد توافق مشاهده شده عبارت خواهد بود از $\Pr(a) = (20+15)/50 = 0.7$

□ اما در مورد تناسب $\Pr(e)$:

- پزشک در 50% موارد اعلام سلامت برای مراجعه کننده داشته
- در مالیکه ماشین در 60% موارد اعلام سلامت برای مراجعه کننده داشته است
- بنابراین احتمال اینکه بصورت اتفاقی هر دو نظر بر سلامت بدهند عبارتست از: $0.50 * 0.60 = 0.30$ و احتمال اینکه به صورت اتفاقی هر دو نظر به بیماری بدهند عبارتست از $0.50 * 0.40 = 0.20$. در این حال $\Pr(e) = 0.30 + 0.20 = 0.50$


□ بدین ترتیب ضریب کاپا مقدار زیر خواهد بود:

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} = \frac{0.7 - 0.5}{1 - 0.5} = 0.4$$

□ در دو تقسیم‌بندی مختلف:

- مقدار ضریب کاپای کمتر از 0.4 به معنی توافق کم، بین 0.41 و 0.75 به معنی توافق نسبی مطلوب تا خوب و بیشتر از 0.76 به معنی توافق عالی است.
- مقدار ضریب کاپای کمتر از صفر به معنی هیچگونه توافق، بین 0 تا 0.2 به معنی توافق ناپیچ، بین 0.21 تا 0.4 به معنی توافق نسبی، بین 0.41 تا 0.6 به معنی توافق متوسط، بین 0.61 تا 0.8 به معنی توافق قابل توجه و بین 0.81 تا 1 به معنی توافق عالی است

H.R. POURREZA



محاسبه‌ی میاریابی ارزیابی

14

□ ضریب همبستگی اسپیرمن (Spearman correlation coefficient)


□ برای ارزیابی رتبه‌بندی بکار می‌رود.

□ برای n دیتای مرتب شده، این ضریب که با ρ نمایش داده می‌شود، بصورت زیر محاسبه می‌شود

$$\rho = 1 - \frac{6 * \sum_{i=1}^k d_i^2}{n * (n^2 - 1)}$$

□ که در آن d_i فاصله‌ی رتبه بدست آمده تا رتبه واقعی برای داده i ام است.

H.R. POURREZA



محاسبه می‌یارهای ارزیابی

15

□ **ماتریس درهم‌ریختگی (confusion matrix)**

این جدول در کاربردهای یادگیری ماشین کارایی الگوریتم را به رویت می‌رساند. هر ستون ماتریس بیان‌کننده پیش‌بینی قرارگیری در کلاس‌ها و هر سطر ماتریس بیان‌کننده قرارگیری واقعی در کلاس‌ها است. مثلا فرض کنید که یک سیستم بینایی ماشین برای تشخیص ارقام 0، 1 و 2 از هم نوشته شده است. به این سیستم 30 تصویر شامل این ارقام شامل 10 تصویر از 0، 12 تصویر از 1 و 8 تصویر از 2 داده شده و نتیجه بصورت زیر ارائه شده است:

		کلاس پیش‌بینی شده		
		0	1	2
کلاس واقعی	0	9	0	1
	1	1	9	2
	2	0	1	7

H.R. POURREZA