Imbalanced Domains and Rare Event Detection

Performance Evaluation

An Example from Regression

Forecasting Stock Market Returns

- Very high or low returns (% variations of prices) are interesting
- Near-zero returns are very common but uninteresting for traders - unable to cover transaction costs
- Examples:
  - Forecasting a future return of 3% and then it happens -5% is a very bad error!
  - Forecasting a return of 3% and then it happens 11% has the same error amplitude but it is not a serious error
  - Forecasting 0.2% for a true value of 0.4% is reasonably accurate but irrelevant!
  - Forecasting -7.5% for a true value of -8% is a good an useful prediction
- Because near 0 returns are very common a model that always forecasts 0 is hard to beat in terms of Mean Squared Error. But this model is useless!

Metrics and the Available Information

- Different applications may involve different type of information on the user preferences
- This may have an impact on the metrics you can and/or should calculate
- Independently, there are two classes of metrics: scalar and graphical

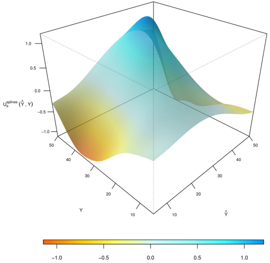
Evaluation with Full Utility Information

Utility Matrices

Table where each entry specifies the cost (negative benefit) or benefit of each type of prediction

		Pred.		
		c_1	c_2	c_3
Obs.	c_1	$B_{1,1}$	$C_{1,2}$	$C_{1,3}$
	c_2	$C_{2,1}$	$B_{2,2}$	$C_{2,3}$
	c_3	$C_{3,1}$	$C_{3,2}$	$B_{3,3}$

- Models are then evaluated by the total utility of their predictions, i.e. the sum of the benefits minus the costs.
- Similar setting for regression using Utility Surfaces (Ribeiro, 2011)



R. Ribeiro (2011). "Utility-based Regression". PhD on Computer Science, Univ. Porto.

The Precision/Recall Framework

Classification

- Problems with two classes
- One of the classes is much less frequent and it is also the most relevant

		Preds.	
		Pos	Neg
Obs.	Pos	True Positives (TP)	False Negatives (FN))
	Neg	False Positives (FP)	True Negatives (TN)

The Precision/Recall Framework

Classification - 2

		Preds. P	Preds. N
Obs.	P	TP	FN
	N	FP	TN

- *Precision* - proportion of the signals (events) of the model that are correct

$$Prec = \frac{TP}{TP + FP}$$

- *Recall* - proportion of the real events that are captured by the model

$$Rec = \frac{TP}{TP + FN}$$

The F-Measure

Combining Precision and Recall into a single measure

- Useful to have a single measure - e.g. optimization within a search procedure
- Maximizing one of them is easy at the cost of the other (it is easy to have 100% recall - always predict "P").
- What is difficult is to have both of them with high values

The F-Measure

Combining Precision and Recall into a single measure

- Useful to have a single measure - e.g. optimization within a search procedure
- Maximizing one of them is easy at the cost of the other (it is easy to have 100% recall - always predict "P").
- What is difficult is to have both of them with high values
- The F-measure is a statistic that is based on the values of precision and recall and allows establishing a trade-off between the two using a user-defined parameter (β),

$$F_\beta = \frac{(\beta^2 + 1) \cdot Prec \cdot Rec}{\beta^2 \cdot Prec + Rec}$$

where β controls the relative importance of *Prec* and *Rec*. If $\beta = 1$ then F is the harmonic mean between *Prec* and *Rec*; When $\beta \to 0$ the weight of *Rec* decreases. When $\beta \to \infty$ the weight of *Prec* decreases.

The G-Mean and Adjusted G-Mean

$$Gm = \sqrt{\frac{TP}{TP+FN} \times \frac{TN}{TN+FP}} = \sqrt{sensitivity \times specificity}$$

$$AGm = \begin{cases} \frac{Gm+Specificity \times N_n}{1+N_n} & sensitivity \geq 0 \\ 0 & sensitivity = 0 \end{cases}$$

where N_n is the proportion of majority class examples in the data set.

M. Kubat and S. Matwin. "Addressing the curse of imbalanced training sets: one-sided selection." In Proc. of 14th Int. Conf. on Machine Learning, 1997, Nashville, USA, pp.179-186

R. Batuwita and V. Palade. "A new performance measure for class imbalance learning. Application to bioinformatics problems." In ICMLA'09, pp.545–550. IEEE, 2009.

Metrics for Multiclass Imbalance Problems

- $\phi(i)$ is the relevance of class i.
- Different ways to obtain $\phi()$ depending on the available domain information (Branco, 2017).

$$Rec^{\phi} = \frac{1}{\sum_{i=1}^{C} \phi(i)} \sum_{i=1}^{C} \phi(i) \cdot recall_i \qquad Prec^{\phi} = \frac{1}{\sum_{i=1}^{C} \phi(i)} \sum_{i=1}^{C} \phi(i) \cdot precision_i$$

$$F_{\beta}^{\phi} = \frac{(1+\beta^2) \cdot Prec^{\phi} \cdot Rec^{\phi}}{(\beta^2 \cdot Prec^{\phi}) + Rec^{\phi}} \qquad AvF_{\beta}^{\phi} = \frac{1}{\sum_{i=1}^{C} \phi(i)} \sum_{i=1}^{C} \frac{\phi(i) \cdot (1+\beta^2) \cdot precision_i \cdot recall_i}{(\beta^2 \cdot precision_i) + recall_i}$$

$$CBA^{\phi} = \sum_{i=1}^{C} \phi(i) \cdot \frac{mat_{i,i}}{max\left(\sum_{j=1}^{C} mat_{i,j}, \sum_{j=1}^{C} mat_{j,i}\right)}$$

P. Branco, L. Torgo, and R. Ribeiro. "Relevance-based evaluation metrics for multi-class imbalanced domains." PAKDD. Springer, Cham, pp.698-710 (2017).

The Precision/Recall Framework

Regression

For forecasting rare extreme values, the concepts of Precision and Recall were also adapted to regression (Torgo and Ribeiro, 2009; Branco, 2014),

$$prec^{\phi} = \frac{\sum_{\phi(\hat{y}_i) > t_R} (1 + U(\hat{y}_i, y_i))}{\sum_{\phi(\hat{y}_i) > t_R} (1 + \phi(\hat{y}_i))}$$

$$rec^{\phi} = \frac{\sum_{\phi(y_i) > t_R} (1 + U(\hat{y}_i, y_i))}{\sum_{\phi(y_i) > t_R} (1 + \phi(y_i))}$$

L. Torgo and R. P. Ribeiro (2009). "Precision and Recall for Regression". In: Discovery Science'2009. Springer.

P. Branco (2014). "Re-sampling Approaches for Regression Tasks under Imbalanced Domains". MSc on Computer Science, Univ. Porto.

Summary of Scalar Metrics for Imbalanced Domains

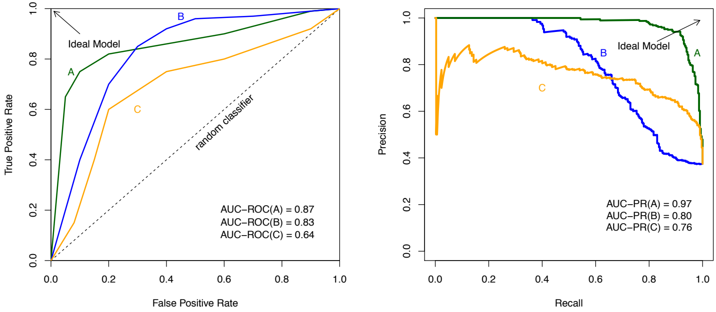
Metric type	Task type		Metric	Main References
Scalar	Classification	binary	$TP_{rate}(recall\ or\ sensitivity)$, $TN_{rate}(specificity)$, FP_{rate}, FN_{rate}, $PP_{value}(precision)$, NP_{value}, F_β, $G-Mean$, $dominance$, $IBA_\alpha(M)$, CWA, $balanced\ accuracy$, $optimized\ precision$, $adjusted\ G-Mean$, B_{42}	Rijsbergen [1979], Kubat et al. [1998], Estabrooks and Japkowicz [2001], Cohen et al. [2006], Ranawana and Palade [2006], García et al. [2008, 2009], Batuwita and Palade [2009], Brodersen et al. [2010], García et al. [2010], Thai-Nghe et al. [2011], Batuwita and Palade [2012]
		multiclass	$recall(c)$, $precision(c)$, $F_\beta(c)$, Rec_μ, $Prec_\mu$, Rec_M, $Prec_M$, MF_β, $MF_{\beta\mu}$, $MF_{\beta M}$, $MAvA$, $MAvG$, CWA, $Prec^{Prev}$, Rec^{Prev}, F_β^{Prev}, CBA^{Prev}, $Prec^{TO}$, Rec^{TO}, F_β^{TO}, CBA^{TO}, $Prec^{PO}$, Rec^{PO}, F_β^{PO}, CBA^{PO}, $Prec^\phi$, Rec^ϕ, F_β^ϕ, CBA^ϕ	Sun et al. [2006], Ferri et al. [2009], Sokolova and Lapalme [2009], Branco et al. [2017b]
	Regression		NMU, $precision^u$, $recall^u$, $precision^\phi$, $recall^\phi$	Torgo and Ribeiro [2007, 2009], Ribeiro [2011], Branco [2014]

Adapted from:
P. Branco, L. Torgo and R. Ribeiro. “A Survey of Predictive Modeling on Imbalanced Domains”. In: ACM Comput. Surv. 49-2, 1–31 (2016).

P. Branco (2018). "Utility-based Predictive Analytics". PhD on Computer Science, Univ. Porto.

ROC curve and Precision-Recall Curve

Classification

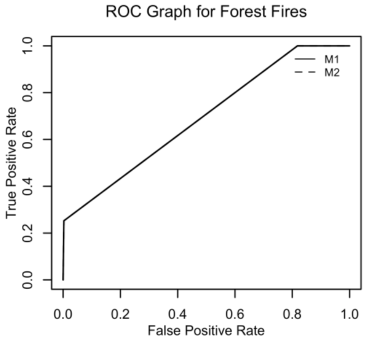


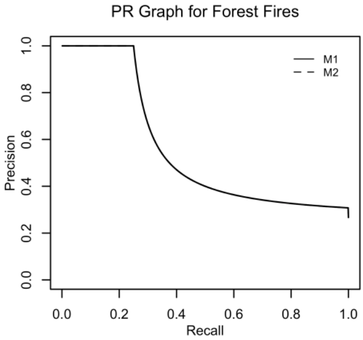
Taken from:

P. Branco (2018). "Utility-based Predictive Analytics". PhD on Computer Science, Univ. Porto.

ROC curve and Precision-Recall Curve

Regression





Taken from:

R. Ribeiro (2011). "Utility-based Regression". PhD on Computer Science, Univ. Porto.

Summary of Graphical Metrics for Imbalanced Domains

Metric type	Task type		Metric	Main References
Graphical	Classification	binary	$ROC\ curve$, AUC, $ProbAUC$, $ScoredAUC$, $WAUC$, $PR\ curve$, $Cost\ curve$, $Brier\ curve$,	Egan [1975], Metz [1978], Bradley [1997], Provost and Fawcett [1997], Provost et al. [1998], Drummond and Holte [2000a], Ferri et al. [2005], Davis and Goadrich [2006], Fawcett [2006b], Wu et al. [2007], Weng and Poon [2008], Hand [2009], Ferri et al. [2011b,a]
		multiclass	$ROC\ surface$, $AUNU$, $AUNP$, $AU1U$, $AU1P$, $SAUC$, $PAUC$	Mossman [1999], Ferri et al. [2009], Alejo et al. [2013], Sánchez-Crisostomo et al. [2014]
	Regression		$AUC-ROC^{\phi}$, $AUC-PR^{\phi}$, $AUC-ROCIV^{\phi}$, $AUC-PRIV^{\phi}$, $REC\ surface$	Torgo [2005], Ribeiro [2011]

Adapted from:

P. Branco, L. Torgo and R. Ribeiro. "A Survey of Predictive Modeling on Imbalanced Domains". In: ACM Comput. Surv. 49-2, 1–31 (2016).

P. Branco (2018). "Utility-based Predictive Analytics". PhD on Computer Science, Univ. Porto.