Evaluation

Recall & Precision

The traditional measure for Machine Learning/Natural Language experiments have been borrowed from Information Retrieval, where Precision reflects the accuracy of positive predictions, and Recall reflects the rate of return of real positives. These measures are totally independent of the number of negative cases that are truly predicted as negative (TN=D). The F-factor is a harmonic mean of Recall and Precision and thus also ignores these cases. The Rand Accuracy can be regarded as a weighted average of Recall and Inverse Recall, or of Precision and Inverse Precision (where the Inverse problem reverses positive and negative), and thus does reflect both. Often however, we do no know all the positive or negative cases, so we cannot calculate anything but Precision and derivatives of Recall based on the known subsets.

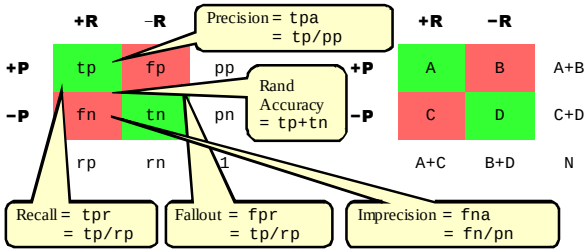
A second problem with all of these measures is that they are heavily influenced by two kinds of bias – Prevalence, the proportion of real positives, rp=A+C/N; and prediction Bias, the proportion of positive predictions, pp=A+B/N. A third issue is the confidence we have in these values, or the significance of the purported results.

The Bookmaker

Powers [1] derived an unbiased measure, Informedness, based on the idea of an "edge" in gambling, given knowledge of the underlying base probabilities and rewarding success and penalizing failures according fair odds (Bookmaker) for an arbitrary number of classes.

Dichotomous cases – Informedness, Markedness & Correlation

In the case of just two classes (+ and –), Informedness can be understood in terms of ROC analysis is either as the distance from a specific prediction system to the chance line, as the Unbiased Weighted Relative Accuracy (WRAcc for Skew=0), as twice the area between the curve and the chance line, or in terms of the Area Under the Curve (AUC) as 2AUC-1. In Psychology, Dichotomous Informedness corresponds to DeltaP'. DeltaP is identified empirically as the normative predictor of human associative judgements, as one concept primes or marks another. Whereas Informedness corresponds and DeltaP' and is the unbiased form Recall, DeltaP corresponds to Markedness, how much the actual class influences or marks the selected predictor, and relates to Precision, being dual regression coefficients whose produce is the Matthews Correlation, and associating these with Cramer's V gives us the corresponding χ^2 significance estimates. Similarly we can directly set confidence intervals [2].



Definition 1

Informedness quantifies how informed a predictor is for the specified condition, and specifies the probability that a prediction is informed in relation to the condition.

Informedness = Recall + Inverse Recall – 1 = `tpr-fpr` = `1-fnr-fpr`
= (Recall – Bias) / (1–Prevalence)

Confidence Interval CI = X.(1-|Informedness|/√[N-1]) where X= $-\Phi^{-1}(\alpha/\text{sides})$

Definition 2

Markedness quantifies how marked a condition is for the specified predictor, and specifies the probability that a condition is marked by the predictor.

Markedness = Precision + Inverse Precision – 1 = `tpa-fna` = `1-fpa-fna`
= (Precision – Prevalence) / (1– Bias)

Confidence Interval CM = CM = X.(1-|Markedness|/√[N-1]) where X= $-\Phi^{-1}(\alpha/\text{sides})$

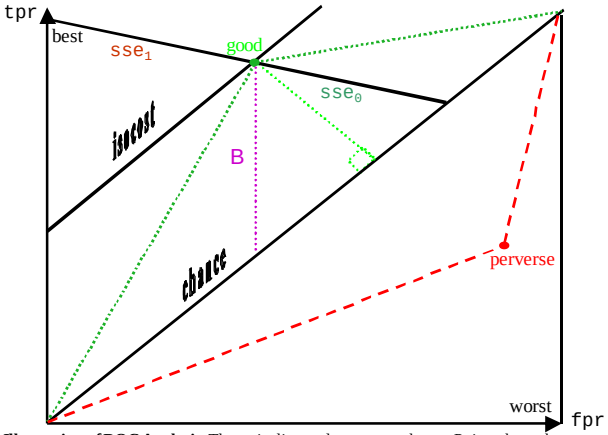


Illustration of ROC Analysis. The main diagonal represents chance. Points above the diagonal represent performance better than chance, those below worse than chance.

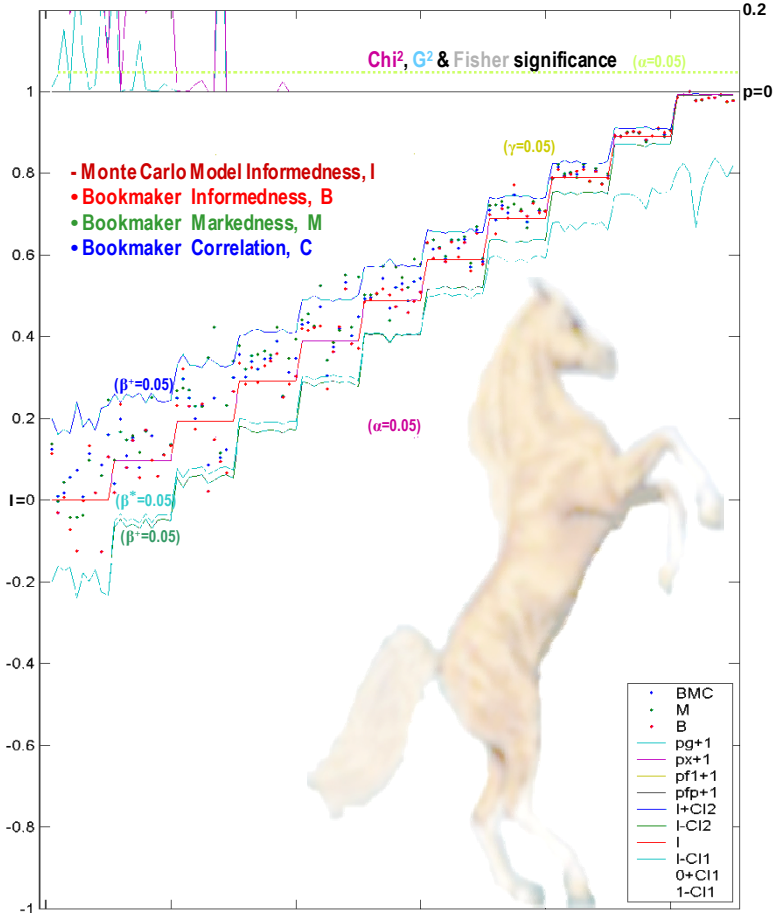


Illustration of Significance and Confidence. 110 Monte Carlo simulations with 11 stepped expected Informedness levels (red line) with Bookmaker-estimated Informedness (red dots), Markedness (green dot) and Correlation (blue dot), with significance (p+1) calculated using G^2, χ^2, and Fisher estimates, and confidence bands shown for both the theoretical Informedness and the B=0 and B=1 levels (parallel at B=0.18,0.82). The lower theoretical band is calculated twice, using both CI_{B1} and CI_{B2}. Here K=5, N=128, X=1.96 for two-sided α=β=0.05.

© 2008 David M. W. Powers

AILab, School of Computer Science,
Mathematics and Engineering
Flinders University of South Australia
G.P.O. Box 2100, Adelaide 5001
South Australia

Contact Details

http://www.csem.flinders.edu.au/people/pages/powers_david/

Copies of this poster, paper, spreadsheets, scripts may be obtained from:
David.Powers@flinders.edu.au or powers@acm.org or
http://david.wardpowers.info/BM

References

[1] Powers, David M. W. (2003), Recall and Precision versus the Bookmaker, *Proceedings of the International Conference on Cognitive Science* (**ICSC-2003**), Sydney Australia, 2003, pp. 529-534. http://david.wardpowers.info/BM/index.htm

[2] Powers, David M. W. (2007) *Evaluation*, Flinders InfoEng Tech Report SIE07001 http://www.infoeng.flinders.edu.au/research/techreps/SIE07001.pdf

David M. W. Powers is Associate Professor of Computer Science and Head of the AI Lab at Flinders University. He specializes in applications of unsupervised learning to language and speech processing. Dr Powers undertook his PhD in this area, as well as co-founding ACL's SIGNLL and CoNLL. He is also a trader and has a Diploma in Technical Analysis, being the study of how to find and exploit 'edges' in the financial markets.

European Conference on Artificial Intelligence, July 2008, Patras Greece