A Signal Detection Theory Approach to Evaluating Oculometer Data Quality

Kara Latorella[1], William Lynn III[1], John S. Barry[2]
Lon Kelly[3], Ming-Yun Shih[3],
Charles T. Feigh[3], Dennis Frasca[3], Omar Scott[3], Julie Timmons[3]

[1]NASA Langley Research Center, Hampton VA
[2]Northrup Grumman Corporation, Hampton VA
[3]UNISYS Corporation, Hampton VA

Abstract

Currently, data quality is described in terms of spatial and temporal accuracy and precision [Holmqvist *et al.* in press]. While this approach provides precise errors in pixels, or visual angle, often experiments are more concerned with whether subjects' points of gaze can be said to be reliable with respect to experimentally-relevant areas of interest. This paper proposes a method to characterize oculometer data quality using Signal Detection Theory (SDT) [Marcum 1947]. SDT classification results in four cases: Hit (correct report of a signal), Miss (failure to report a), False Alarm (a signal falsely reported), Correct Reject (absence of a signal correctly reported). A technique is proposed where subjects' are directed to look at points in and outside of an AOI, and the resulting Points of Gaze (POG) are classified as Hits (points known to be internal to an AOI are classified as such), Misses (AOI points are not indicated as such), False Alarms (points external to AOIs are indicated as in the AOI), or Correct Rejects (points external to the AOI are indicated as such). SDT metrics describe performance in terms of discriminability, sensitivity, and specificity. This paper presentation will provide the procedure for conducting this assessment and an example of data collected for AOIs in a simulated flightdeck environment.

Method

The most common method for assessing the accuracy and precision of oculometer data is to ask a subject to look at certain points in the environment with known coordinates. The accuracy of the resulting point of gaze coordinate data is defined as the distance (in visual degrees) between the actual point and the measured gaze position [Holmqvist *et al.* in press]. While this approach provides precise errors in pixels, or visual angle, often experiments are more concerned with whether subjects' POGs can be said to be reliable with respect to experimentally-relevant AOIs. Signal Detection theory (SDT) [Marcum 1947] is a method for assessing the effectiveness of a classifier in a forced-choice, two-state decision problem, and provides a framework for this approach. SDT assumes a distribution of noise, and signals embedded within that noise, represented by another distribution (signal+noise) which is shifted by some distance, d' (Figure 1). The observer has some criterion (X_c) for distinguishing between a Signal and a Noise event. Events that are on the continuum to the right of X_c are considered Signals, whereas those to the left of X_c are considered Noise. For a given state of the world (there is a signal, or there isn't) and the classification decision (a signal is reported, or is not), there are four possible cases: Hit (an existing signal is reported), Miss (an existing signal isn't reported), False Alarm (a signal is reported when there isn't one), Correct Reject (no signal is indicated when there isn't one). Hit Rate (HR) is the ratio of Hits to actual signals existing (Hits+Misses), and is also defined as the *sensitivity* of the observer. False Alarm Rate (FAR) is the ratio of False Alarms to the cases in which there was no signal (False Alarms + Correct Rejects). S*pecificity* is defined as 1- FAR, or the Correct Reject Rate (CRR). The Receiver Operator Characteristic (ROC) curve illustrates how well different assessments distinguish signals from noise. A point on the ROC curve is defined by the False Alarm Rate (FAR), on the abscissa, and the corresponding Hit Rate (HR), on the ordinate, for a particular assessment; that is, an implementation of the decision criteria (X_c) (Figure 2).
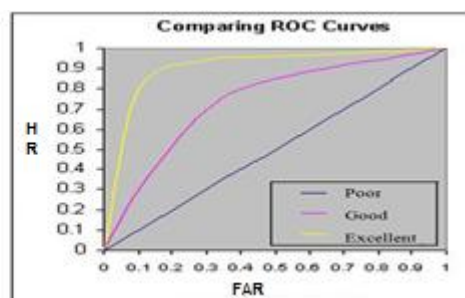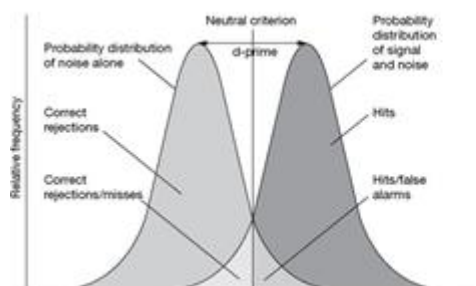
Figure 1 Signal & Noise Distributions [Oliver et al. 2008]   **Figure 2** ROC Curve [after Tape 2012]

An acceptable binary classifier must have both good sensitivity (high HR) and good specificity (low FAR). On the ROC, the line with slope = 1, is referred to as the "no discrimination" line (labeled "poor" in Figure 2), which describes a classifier that is of no use. In classical SDT, assuming Gaussian and equal-variance Signal+Noise and Noise distributions, d' is the difference between the z-score transforms of the HR and FAR, and is used as a measure of classifier *discriminability*. When the classifier is a better discriminator, d' scores are larger. The criterion, Xc, indicates the bias of a classifier and is calculated as the negative of the z-transform of the FAR. Non-parametric approaches to calculating discriminability have been developed when the normality and equal variance assumptions of SDT cannot be guaranteed. A distribution-free metric, A', has been used for discriminability (calculated as 1-0.25[{(1-HR)/(1-FAR)} + (FAR/HR)] [Macmillan and Creelman 1990]). The non-parametric equivalent of the criterion or bias metric, (c), is (-0.5)*-z- transformed (HR+FAR). A lenient bias (greater tendency to report signals), therefore, has c less than zero. A conservative bias (less tendency to report signals), has (c) greater than zero.

The proposed method is as follows. AOIs relevant to the hypotheses associated with a study are described, and modeled in the oculometer software for its classification of POGs. For each AOI, a number of internal assessment points are defined. Definition of these points should take into account the precision required for the inferences to be made in the study. For example, a four-inch square AOI with meaningful information clustered in the center should have sample points clustered in its center as well. The same AOI with meaningful information distributed throughout it, should have sample points more widespread. An equal number of "external assessment points" should be defined just outside of the AOI of interest (these may be within other AOIs). Subjects are asked to focus for a standard exposure time on each point. Data taken during each point is associated with a distinct event marker. The results can be classified according to SDT: Hit (internal point data that is classified in the AOI), Miss (internal point data that is classified outside the AOI), False Alarm (external point data that is classified in the AOI), Correct Reject (external point data that is classified outside the AOI). This process would be repeated for all AOIs of interest to the investigation.

Anticipated Results

At the first level of analysis, a ROC would be created for each AOI and subject, and discriminability calculated. These ROCs can be used to determine if the oculometer was a poor discriminator for any subjects – evidence that these subjects ought to be removed from further data analysis. After removing outlier subjects, the resulting data is collapsed over subjects to construct composite ROCs for each AOI. The oculometer performance for this set of subjects, for each AOI, can then be assessed in terms of discriminability, (and its parameters: sensitivity, and specificity). If external assessment points are located in other AOIs, one can construct a confusability matrix that addresses the probability of misinterpreting data. The presentation will demonstrate use of this method with data collected in a flight simulation environment.

Conclusion

This methodology is a complementary approach to characterizing data quality from oculometer systems based on basic SDT. While the described approach is based on a simple binary classifier, more sophisticated unsupervised machine learning classifier algorithms promise to provide a more holistic evaluation across multiple AOIs. This approach may be extended with the use of supervised classification methods to improve data quality as would be measured by these SDT metrics.

Acknowledgement

References

HOLMQVIST, K., NYSTRÖM, M., and MULVEY, F. 2012. Eye tracker data quality: what it is and how to measure it, Proceedings of the Symposium on Eye Tracking Research and Applications, p. 45-32, ACM.

MACMILLAN, N.A., AND CREELMAN, C.D. 1990. Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" measures. *Psychological Bulletin, 107*, 401-413.

MARCUM, J. I. 1947. A Statistical Theory of Target Detection by Pulsed Radar. *The Research Memorandum*: 90. *http://www.rand.org/pubs/research_memoranda/RM754/*. Retrieved 2011 November 25

OLIVER, R., BJOERTOMT, O., GREENWOOD, R., AND ROTHWELL, J. 2008. Noisy patients' - can signal detection theory help? Nature Clinical Practice Neurology 4, 306-316. *http://www.nature.com/nrneurol/journal/v4/n6/fig_tab/ncpneuro0794_F1.html*. Retrieved 2011 November 25.

TAPE, T.G. 2011. *Interpreting Diagnostic tests: The area under an ROC curve.* *http://gim.unmc.edu/dxtests/ROC3.htm*, Retrieved 2011 November 25.