Asymptotic Upper Bound on False Alarm Rate

Theorem 1. *The false alarm rate of the proposed algorithm is asymptotically (as $M_2 \to \infty$) upper bounded by*

$$FAR \leq e^{-\omega_0 h}, \tag{1}$$

where h is the decision threshold, and $\omega_0 > 0$ is given by

$$\omega_0 = v_m - \theta - \frac{1}{\phi}\mathcal{W}\left(-\phi\theta e^{-\phi\theta}\right), \tag{2}$$
$$\theta = \frac{v_m}{e^{v_m d_\alpha^m}}.$$

In (2), $\mathcal{W}(\cdot)$ is the Lambert-W function, $v_m = \frac{\pi^{m/2}}{\Gamma(m/2+1)}$ is the constant for the m-dimensional Lebesgue measure (i.e., $v_m d_\alpha^m$ is the m-dimensional volume of the hyperball with radius d_α), and ϕ is the upper bound for δ^t.

Proof.

In (Basseville & Nikiforov, 1993)[page 177], for CUSUM-like algorithms with independent increments, such as the proposed detector with independent δ^t, a lower bound on the average false alarm period is given as follows

$$E_\infty[T] \geq e^{\omega_0 h},$$

where h is the detection threshold, and $\omega_0 \geq 0$ is the solution to $E[e^{\omega_0 \delta^t}] = 1$.

To analyze the false alarm period, we need to consider the nominal case. In that case, since there is no anomalous object at each time t, the selection of object with maximum kNN distance in $\delta^t = (\max_i\{d_i^t\})^m - d_\alpha^m$ does not necessarily depend on the previous selections due to lack of an anomaly which could correlate the selections. Hence, in the nominal case, it is safe to assume that δ^t is independent over time.

We firstly derive the asymptotic distribution of the frame-level anomaly evidence δ^t in the absence of anomalies. Its cumulative distribution function is given by

$$P(\delta^t \leq y) = P((\max_i\{d_i^t\})^m \leq d_\alpha^m + y).$$

It is sufficient to find the probability distribution of $(\max_i\{d_i^t\})^m$, the mth power of the maximum kNN distance among objects detected at time t. As discussed above, choosing the object with maximum distance in the absence of anomaly yields independent m-dimensional instances $\{x^t\}$ over time, which form a Poisson point process. The nearest neighbor ($k = 1$) distribution for a Poisson point process is given by

$$P(\max_i\{d_i^t\} \leq r) = 1 - \exp(-\Lambda(b(x^t, r)))$$

where $\Lambda(b(x^t, r))$ is the arrival intensity (i.e., Poisson rate measure) in the m-dimensional hypersphere $b(x^t, r)$ centered at x^t with radius r (Chiu et al., 2013). Asymptotically, for a large number of training instances as $M_2 \to \infty$, under the null (nominal) hypothesis, the nearest neighbor distance $\max_i\{d_i^t\}$ of x^t takes small values, defining an infinitesimal hyperball with homogeneous intensity $\lambda = 1$ around x^t. Since for a homogeneous Poisson process the intensity is written as $\Lambda(b(x^t, r)) = \lambda|b(x^t, r)|$ (Chiu et al., 2013), where $|b(x^t, r)| = \frac{\pi^{m/2}}{\Gamma(m/2+1)} r^m = v_m r^m$ is the Lebesgue measure (i.e., m-dimensional volume) of the hyperball $b(x^t, r)$, we rewrite the nearest neighbor distribution as

$$P(\max_i\{d_i^t\} \leq r) = 1 - \exp\left(-v_m r^m\right),$$

where $v_m = \frac{\pi^{m/2}}{\Gamma(m/2+1)}$ is the constant for the m-dimensional Lebesgue measure.

Now, applying a change of variables we can write the probability density of $(\max_i\{d_i^t\})^m$ and δ^t as

$$\begin{aligned} f_{(\max_i\{d_i^t\})^m}(y) &= \frac{\partial}{\partial y}\left[1 - \exp\left(-v_m y\right)\right], \\ &= v_m \exp(-v_m y), \\ f_{\delta^t}(y) &= v_m \exp(-v_m d_\alpha^m)\exp(-v_m y) \end{aligned} \tag{3}$$

Using the probability density derived in (3), $E[e^{\omega_0 \delta^t}] = 1$ can be written as

$$\begin{aligned} 1 &= \int_{-d_\alpha^m}^{\phi} e^{\omega_0 y} v_m e^{-v_m d_\alpha^m} e^{-v_m y} dy, \\ \frac{e^{v_m d_\alpha^m}}{v_m} &= \int_{-d_\alpha^m}^{\phi} e^{(\omega_0 - v_m)y} dy, \\ &= \left.\frac{e^{(\omega_0 - v_m)y}}{\omega_0 - v_m}\right|_{-d_\alpha^m}^{\phi}, \\ &= \frac{e^{(\omega_0 - v_m)\phi} - e^{(\omega_0 - v_m)(-d_\alpha^m)}}{\omega_0 - v_m}, \end{aligned} \tag{4}$$

where $-d_\alpha^m$ and ϕ are the lower and upper bounds for$\delta^t = (\max_i\{d_i^t\})^m - d_\alpha^m$. The upper bound ϕ is obtained from the training set.

As $M_2 \to \infty$, since the mth power of $(1-\alpha)$th percentile of nearest neighbor distances in training set goes to zero, i.e., $d_\alpha^m \to 0$, we have

$$e^{(\omega_0 - v_m)\phi} = \frac{e^{v_m d_\alpha^m}}{v_m}(\omega_0 - v_m) + 1.$$

We next rearrange the terms to obtain the form of $e^{\phi x} = a_0(x + \theta)$ where $x = \omega_0 - v_m$, $a_0 = \frac{e^{v_m d_\alpha^m}}{v_m}$, and $\theta = \frac{v_m}{e^{v_m d_\alpha^m}}$. The solution for x is given by the Lambert-W function (Scott et al., 2014) as $x = -\theta - \frac{1}{\phi}\mathcal{W}(-\phi e^{-\phi\theta}/a_0)$, hence

$$\omega_0 = v_m - \theta - \frac{1}{\phi}\mathcal{W}\left(-\phi\theta e^{-\phi\theta}\right).$$

Finally, since the false alarm rate (i.e., frequency) is the inverse of false alarm period $E_\infty[T]$, we have

$$FAR \leq e^{-\omega_0 h},$$

where h is the detection threshold, and ω_0 is given above.

Although the expression for ω_0 looks complicated, all the terms in (2) can be easily computed. Particularly, v_m is directly given by the dimensionality m, d_α comes from the training phase, ϕ is also found in training, and finally there is a built-in Lambert-W function in popular programming languages such as Python and Matlab. Hence, given the training data, ω_0 can be easily computed, and based on Theorem 1, the threshold h can be chosen to asymptotically achieve the desired false alarm period as follows

$$h = \frac{-\log(FAR)}{\omega_0}. \tag{5}$$

References

Basseville, M. and Nikiforov, I. *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall, Englewood Cliffs, 1993.

Chiu, S. N., Stoyan, D., Kendall, W. S., and Mecke, J. *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

Scott, T. C., Fee, G., and Grotendorst, J. Asymptotic series of generalized lambert w function. *ACM Communications in Computer Algebra*, 47(3/4):75–83, 2014.