Ijesm
Consulting, help, relaxation INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & MANAGEMENT

DETECTING SEQUENTIAL IRREGULARITY THROUGH FILTERING AND HEDGING

SUDEEP SAMUELSON EZRA B(MTech Student),QCET,Nellore, sudeep.byravarapu@gmail.com.
SYED AKHTAR M.Tech Assistant Professor , akhtar.sab@gmail.com.
P.Babu, Associate professor, QCET,Nellore,babu123mca@gmail.com

ABSTRACT

This paper describes a methodology for detecting anomalies from sequentially observed and potentially noisy data. The proposed approach consists of two main elements: (1) filtering, or assigning a belief or likelihood to each successive measurement based upon our ability to predict it from previous noisy observations, and (2) hedging, or flagging potential anomalies by comparing the current belief against a time-varying and data-adaptive threshold. The threshold is adjusted based on the available feedback from an end user. Our algorithms, which combine universal prediction with recent work on online convex programming, do not require computing posterior distributions given all current observations and involve simple primal-dual parameter updates. At the heart of the proposed approach lie exponential-family models which can be used in a wide variety of contexts and applications, and which yield methods that achieve sublinear per-round regret against both static and slowly varying product distributions with marginals drawn from the same exponential family. Moreover, the regret against static distributions coincides with the minimax value of the corresponding online strongly convex game. We
also prove bounds on the number of mistakes made during the hedging step relative to the best offline choice of the threshold with access to all estimated beliefs and feedback signals. We validate the theory on synthetic data drawn from a time-varying distribution over binary vectors of high dimensionality, as well as on the Enron email dataset.

MOTIVATION

Anomaly is a deviation from a normal behavior. Anomaly detection techniques are used to detect unusual patterns in data. These patterns deviate from the spectrum of normal behaviors in the data, and typically they represent critical events that occurred in the monitored system. Anomaly detection can be used to identify sophisticated and targeted attacks like Advanced Persistent Threats, where standard security systems often fail to detect.

Anomaly detection is an important problem in intrusion detection. Intrusion

detection is the problem of detecting attacks on systems by examining various audit data of a system such as TCP packets or system logs and differentiating between normal users and intruders.

We explore the performance of online anomaly detection methods built on sequential probability feedback. We sequentially monitor the state of some system of interest. At each time step, we observe a possibly noise-corrupted version z^t of the current state x^t, and need to infer whether x^t is anomalous relative to the actual sequence x^{t-1} = (x_1, ... ,x_{t-1}) of the past states. This inference is encapsulated in a binary decision, $\hat{y}_t$ which can be either -1 (non-anomalous or nominal behavior) or +1 (anomalous behavior). After announcing our decision, we may occasionally receive feedback on the "true" state of affairs and use it to adjust the future behavior of the decision-making mechanism.

Existing System

The observations cannot be assumed to be independent, identically distributed, or even come from a realization of a stochastic process. In particular, an adversary may be injecting false data into the sequence of observations to cripple our anomaly detection system.

Observations may be contaminated by noise or be observed through an imperfect communication channel.

Declaring observations anomalous if their likelihoods fall below some threshold is a popular and effective strategy for anomaly detection, but setting this threshold is a notoriously difficult problem.

Obtaining feedback on the quality of automated anomaly detection is costly as it generally involves considerable effort by a human expert or analyst. Thus, if we have an option to request such feedback at any time step, we should exercise this option sparingly and keep the number of requests to a minimum. Alternatively, the times when we receive feedback may be completely arbitrary and not under our control at all — for instance, we may receive feedback only when we declare false positives or miss true anomalies.

Proposed System:

In this project, we propose a general methodology for addressing these challenges. With apologies to H.P. Lovecraft , we will call our proposed framework FHTAGN, or Filtering and Hedging for Time-varying Anomaly recognition. More specifically, the two components that make up FHTAGN are:

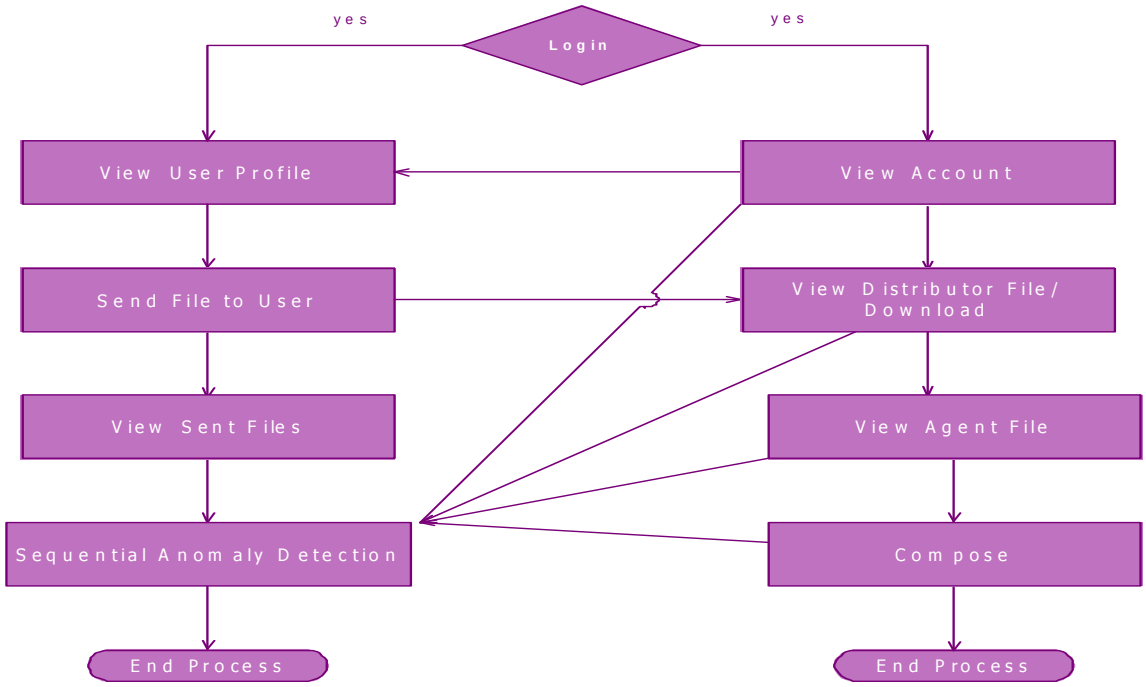
Filtering — the sequential process of updating beliefs on the next state of the system based on the noisy observed past. The term "filtering" comes from statistical signal processing and is intended to signify the fact that the beliefs of interest concern the unobservable actual system state, yet can only be computed in a causal manner from its noise-corrupted observations.

Hedging — the sequential process of flagging potential anomalies by comparing the current belief against a timevarying threshold. The rationale for this approach comes from the intuition that a behavior we could not have predicted well based on the past is likely to be anomalous. The term "hedging" is meant to indicate the fact that the threshold is dynamically raised or lowered, depending on the type of the most

recent mistake (a false positive or a missed anomaly) made by our inference engine.

Rather than explicitly modeling the evolution of the system state and then designing methods for that model (e.g., using Bayesian updates), we adopt an "individual sequence" (or "universal prediction") perspective and strive to perform provably well on any individual observation sequence in the sense that our per-round performance approaches that of the best offline method with access to the entire data sequence.

Fig 3.1: Context Diagram



ALGORITHMS

ONLINE CONVEX PROGRAMMING

The philosophy advocated in the present project is that the tasks of sequential probability assignment and threshold selection can both be viewed as a game between two opponents, the Forecaster and the Environment. The Forecaster is

continually predicting changes in a dynamic Environment, where the effect of the Environment is represented by an arbitrarily varying sequence of convex cost functions over a given feasible set, and the goal of the Forecaster is to pick the next feasible point in such a way as to keep the cumulative cost as low as possible. This is broadly formulated as the problem of online convex programming

FILTERING:SEQUENTIAL PROBABILITY ASSIGNMENT IN THE PRESENCE OF NOISE

The first ingredient of FHTAGN is a strategy for assigning a likelihood (or belief) $p_t(.\,|z^{t-1})$ to the clean symbol x_t based on the past noisy observations z^{t-1}. Alternatively, we can think of the following problem: if x_t is the actual clean symbol that has been generated at time t, then our likelihood $p_t = p_t(x_t|z^{t-1})$, though well-defined, is not accessible for observation. Thus, we would like to estimate it via some estimator $\hat{p}_t$, which will depend on the actual observed noisy symbol z_t, as well as on the previously obtained estimates $\widehat{p^{t-1}} = (\hat{p}_t, \ldots, \widehat{p_{t-1}})$. In the field of signal processing, problems of this kind go under the general heading of filtering; this term refers to any situation in which it is desired, at each time t, to obtain an estimate of some clean unobservable quantity causally based on noisy past observations.

HEDGING:SEQUENTIAL THRESHOLD SELECTION FOR ANOMALY DETECTION

In order to choose an appropriate τ_t, we rely on feedback from an end user. Specifically, let the end user set the label y_t as 1 if z_t is anomalous and -1 if z_t is not anomalous. However, since it is often desirable to minimize human intervention and analysis of each observation, we seek to limit the amount of feedback received. To this end, two possible scenarios could be considered:

At each time t, the Forecaster randomly decides whether to request a label from the end user. A label is requested with probability that may depend on f_t and τ_t.

At each time t, the end-user arbitrarily chooses whether to provide a label to the Forecaster; the Forecaster has no control over whether or not it receives a label.

As we will see, the advantage of the first approach is that it allows us to bound the average performance over all possible choices of times at which labels are received, resulting in stronger bounds. The advantage of the second approach is that is may be more practical or convenient in many settings. For instance, if an anomaly is by chance noticed by the end user or if an event flagged by the Forecaster as anomalous is, upon further investigation, determined to be non-anomalous, this information is readily available and can easily be provided to the Forecaster. In the

sequel, we will develop performance bounds for both of these regimes.

CONCLUSION & FUTURE ENHANCEMENT

We have proposed and analyzed a methodology for sequential (or online) anomaly detection from an individual sequence of potentially noisy observations in the setting when the anomaly detection engine can receive external feedback confirming or disputing the engine's inference on whether or not the current observation is anomalous relative to the past. Our methodology, dubbed FHTAGN for Filtering and Hedging for Time-varying Anomaly recognition, is based on the filtering of noisy observations to estimate the belief about the next clean observation, followed by a threshold test. The threshold is dynamically adjusted, whenever feedback is received and the engine has made an error, which constitutes the hedging step. Our analysis of the performance of FHTAGN was carried out in the individual sequence framework, where no assumptions were made on the mechanism underlying the evolving observations. Thus, performance was measured in terms of regret against the best offline (non sequential) method for assigning beliefs to the entire sequence of clean observations and then using these beliefs and the feedback (whenever available) to set the best critical threshold. The design and analysis of both filtering and hedging was inspired by recent developments in online convex programming.

REFERENCES:

[1] H. P. Lovecraft, "The call of Cthulhu," Weird Tales, vol. 11, no. 2, pp. 159–178, February 1928.
[2] A. Bain and D. Crisan, Fundamentals of Stochastic Filtering. New York: Springer, 2009.
[3] C. R. Shalizi, "Dynamics of Bayesian updating with dependent data and mispecified models," Electronic J. Statist., vol. 3, pp. 1039–1074, 2009.
[4] N. Merhav and M. Feder, "Universal prediction," IEEE Trans. Inform. Theory, vol. 44, no. 6, pp. 2124–2147, October 1998.
[5] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient descent," in Proc. Int. Conf. on Machine Learning, 2003, pp. 928–936.
[6] A. S. Nemirovsky and D. B. Yudin, Problem Complexity and Method Efficiency in Optimization. New York: Wiley, 1983.
[7] A. Beck and M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," Operations Res. Lett., vol. 31, pp. 167–175, 2003.
[8] M. Raginsky, R. Marcia, J. Silva, and R. Willett, "Sequential probability assignment via online convex programming using exponential families," in Proc. of IEEE International Symposium on Information Theory, 2009.
[9] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection - a survey," ACM Computing Surveys, vol. 41, no. 3, 2009.

[10] I. Steinwart, D. Hush, and C. Scovel, "A classification framework for anomaly detection," J. Machine Learn. Res., vol. 6, pp. 211–232, 2005.
[11] C. Scott and G. Blanchard, "Novelty detection: Unlabeled data definitely help," in Proc. 12th Int. Conf. on Artificial Intelligence and Statistics (AISTATS), D. van Dyk and M. Welling, Eds., 2009, pp. 464–471.
[12] A. B. Tsybakov, "On nonparametric estimation of density level sets," The Annals of Statistics, vol. 25, no. 3, pp. 948–969, 1997.
[13] P. Bartlett, E. Hazan, and A. Rakhlin, "Adaptive online gradient descent," in Adv. Neural Inform. Processing Systems, vol. 20. Cambridge, MA: MIT Press, 2008, pp. 65–72.
[14] J. Abernethy, P. L. Bartlett, A. Rakhlin, and A. Tewari, "Optimal strategies and minimax lower bounds for online convex games," in Proc. Int. Conf. on Learning Theory, 2008, pp. 415–423.
[15] P. R. Kumar and P. Varaiya, Stochastic Systems: Estimation, Identification, and Adaptive Control. Prentice Hall, 1986.
[16] N. Cesa-Bianchi and G. Lugosi, Prediction, Learning and Games. New York: Cambridge Univ. Press, 2006.
[17] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," SIAM J. Optim., vol. 19, no. 4, pp. 1574–1609, 2009.
[18] J.-B. Hiriart-Urruty and C. Lemar´echal, Fundamentals of Convex Analysis. Berlin: Springer, 2001. [19] L. M. Bregman, "The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming," Comput. Mathematics and Math. Phys., vol. 7, pp. 200–217, 1967.
[20] Y. Censor and S. A. Zenios, Parallel Optimization: Theory, Algorithms and Applications. Oxford, UK: Oxford Univ. Press, 1997.
[21] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, UK: Cambrdige Univ. Press, 2004.
[22] M. Raginsky, A. Rakhlin, and S. Y¨uksel, "Online convex programming and regularization in adaptive control," in IEEE Conf. on Decision and Control, Atlanta, GA, December 2010, pp. 1957–1962.
[23] S. Amari and H. Nagaoka, Methods of Information Geometry. Providence: American Mathematical Society, 2000.
[24] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," Foundations and Trends in Machine Learning, vol. 1, no. 1-2, pp. 1–305, 2008.
[25] A. R. Barron and C.-H. Sheu, "Approximation of density functions by sequences of exponential families," Ann. Statist., vol. 19, no. 3, pp. 1347–1369, 1991.
[26] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. New York: Wiley, 2006.
[27] K. S. Azoury and M. K. Warmuth, "Relative loss bounds for on-line density estimation with the exponential family of distributions," Machine Learning, vol. 43, pp. 211–246, 2001.
[28] T. Weissman and N. Merhav, "Universal prediction of individual binary sequences in the presence of noise," IEEE Trans. Inf. Theory, vol. 47, no. 6, pp. 2151–2173, 2001.
[29] B. S. Clarke and A. R. Barron, "Information-theoretic asymptotics of

Bayes methods," IEEE Trans. Inform. Theory, vol. 36, no. 3, pp. 453–471, May 1990.
[30] M. Herbster and M. K. Warmuth, "Tracking the best expert," Machine Learning, vol. 32, no. 2, pp. 151–178, 1998.
[31] ——, "Tracking the best linear predictor," J. Machine Learn. Res., vol. 1, pp. 281–309, 2001.