

Detección inteligente de cambios y generación de alarmas tempranas en sistemas de monitoreo

Ricardo Coppo¹, Patricia Baldini¹, Héctor Bambill¹, Marina Jakomin², Guillermo Calandrini³, Javier Balmaceda¹, Andrés Fernandez¹, Santiago Tarnoski¹, María Tourret¹, Camila Silva Bustos y Julián Abdala¹.

¹Grupo de I+D en Sistemas y Tecnologías de la Información y las Comunicaciones (Grupo SiTIC).
Dpto. de Ingeniería Electrónica, Facultad Regional Bahía Blanca UTN, 11 de Abril 461, Bahía Blanca

²Dirección de cuencas, Dirección Nacional de Política Hídrica y Coordinación Federal, Secretaría de Infraestructura y Política Hídrica. Ministerio del Interior. Esmeralda 255, piso 11, CABA.

³Dpto. de Ingeniería Eléctrica y de Computadoras, Universidad Nacional del Sur, San Andrés 800, Bahía Blanca.

{*rcoppo, pnbaldi, hbambill*}@frbb.utn.edu.ar; *mjakobin@yahoo.com*; *calandri@criba.edu.ar*

RESUMEN

Un cambio o anomalía en un sistema de monitoreo se refiere al instante en que se produce una variación abrupta significativa de alguno de sus parámetros con respecto al patrón habitual de comportamiento. La detección de estos eventos son utilizadas en sistemas de alerta temprana (SAT) meteorológicas, ambientales, en mantenimiento de maquinarias, y el monitoreo de servidores informáticos, entre otros. Para la detección de cambios de estado se analizan los errores que resultan de la comparación entre los valores obtenidos mediante un modelo predictivo con los medidos por la red de monitoreo. Variaciones significativas de este error pueden ser atribuidas a cambios que disparen una alerta. El modelo puede ser estadístico, basado en el procesamiento adaptativo de señales y, en forma más reciente, en la inteligencia computacional.

En este proyecto se propone en una primera fase, el estudio de diferentes enfoques en el planteo de modelos predictivos y detectores de cambio orientados en particular al monitoreo del medio ambiente y de entornos industriales. En una segunda fase se plantea la evaluación e implementación de algoritmos de detección de cambio aplicados tanto a señales de origen simulado como provenientes de prototipos experimentales de laboratorio.

Palabras Clave: detección de eventos; sistemas de alerta temprana; procesamiento adaptativo; inteligencia computacional.

CONTEXTO

La línea de investigación presentada está en

ejecución desde enero de 2019, en el marco del proyecto acreditado y financiado por la Universidad Tecnológica Nacional titulado: “Detección inteligente de cambios y generación de alarmas tempranas en sistemas de monitoreo”.

Su ámbito de realización es el Grupo SiTIC (Grupo de I+D en Sistemas y Tecnologías de la Información y las Comunicaciones) Facultad Regional Bahía Blanca de la UTN. Se cuenta, además con la participación de un investigador del Departamento de Ingeniería Eléctrica y de Computadoras de la Universidad Nacional del Sur y con la asistencia técnica de una profesional de la Secretaría de infraestructura y política hídrica del Ministerio del Interior.

1. INTRODUCCION

Un cambio, evento, o anomalía en un sistema de monitoreo se refiere al instante en que el mismo sufre una variación abrupta de alguno de sus parámetros de funcionamiento. Estos cambios no son necesariamente de magnitud elevada, sino que representan alteraciones significativas del estado del mismo. Estos eventos se evidencian en series de tiempo, en secuencias de imágenes, y en datos heterogéneos no estructurados y son utilizados en SAT para señales sísmicas, meteorológicas y del medio ambiente [24], el monitoreo de calidad del agua y del aire [2][7][8][25][30], el reconocimiento de voz, aplicaciones médicas [5][30][34], en sistemas financieros, en mantenimiento predictivo de maquinarias [3][22][26][33], entre otras [35]. Todos estos sistemas, aunque diferentes en la forma de captura y presentación de la información,

poseen características funcionales similares: una etapa de adquisición, una de fusión y análisis de la información adquirida, una de selección y definición de un modelo predictivo adecuado, estocástico o determinístico, con el cálculo del error de predicción (residuales) y una de generación de una señal de alarma/anomalía al determinar la presencia de un cambio de estado significativo, [10][12][14][17][21][23][28].

Los enfoques tradicionales analizan una única señal y su función auto-correlación para determinar umbrales sobre los valores muestreados o alguno de sus parámetros (amplitud, frecuencia, potencia, media, varianza). Superado este valor límite el sistema emite la señal de alarma. En entornos reales estos límites pueden ser superados ocasionalmente por la presencia de valores atípicos producidos por la existencia de ruido en los sensores, en la transmisión de datos o en el sistema de control y supervisión SCADA, disparando una falsa alarma. Si estas se producen en forma frecuente el sistema de monitoreo pierde confiabilidad. Una simple modificación de los umbrales no siempre es una solución válida del problema, ya que si se los amplía en exceso el sistema pierde sensibilidad y aumenta el error de la "no detección" de eventos reales. Por otra parte, los diversos modelos predictivos permiten la detección en base a la comparación de cada conjunto de medidas con las predichas por el modelo adoptado generando un error residual o de predicción para la observación. Si este error no presenta una distribución probabilística similar a la del ruido blanco se presupone la existencia de un cambio o anomalía. Este enfoque ha permitido una extensa investigación en los campos de los filtros digitales y teoría de control adaptativo [6][12][13][15][32].

En forma más reciente, la disponibilidad de gran cantidad de sensores distribuidos de bajo costo, configurados en redes que monitorean el entorno a frecuencias de muestreo muy altas, permiten el tratamiento del problema de detección de anomalías mediante algoritmos de *BIG DATA* o de *datamining* temporal. Con este nuevo enfoque se puede adicionar, a las series de tiempo tradicionales, información heterogénea (bases de datos con campos no numéricos, documentos de texto, imágenes, páginas web) para su procesamiento [19][20].

Por otra parte, los avances en la inteligencia computacional, reconocimiento de patrones y de la minería de datos permiten actualmente plantear modelos predictivos no lineales basados en redes neuronales, máquinas de

soporte vectorial, aprendizaje profundo (*deep learning*), y modelos de grafos probabilísticos, por mencionar algunos. Estos predictores se apoyan en el reconocimiento del comportamiento del sistema observado de antemano (etapa de aprendizaje, aprendizaje supervisado) o sobre la inferencia de categorías de comportamiento similar (clusters, aprendizaje no supervisado). Los modelos obtenidos reducen el error de predicción permitiendo un análisis de inferencia de las características predominantes del sistema [3][9][16][18][19][26][27]. Los modelos de aprendizaje supervisados procesan vectores multi-paramétricos de atributos clasificados por un experto en forma previa. A partir de estos datos se construye un modelo, en la forma de una red neuronal, árbol de decisión, red Bayesiana, máquina de soporte vectorial, subespacio de estados, etc, que permite realizar la predicción para nuevas observaciones que se obtienen de la red de sensores. Un desvío significativo entre la predicción y la observación más reciente producirá la detección del evento o anomalía buscada [1].

En el aprendizaje no supervisado es innecesaria la participación de un experto previa al entrenamiento del sistema. Los algoritmos inteligentes reconocen clases o "estados" de equivalencia correspondientes a diferentes modos de operación. Al recibir una nueva observación si la clase asignada no es igual a la anterior puede interpretarse que hubo un cambio de comportamiento del sistema. Este enfoque, si bien se adapta a situaciones en las que no se conocen los posibles estados de operación del sistema, es de alta complejidad computacional y difícil de implementar en la práctica ya que requiere mantener toda la base de datos de entrenamiento a disposición del algoritmo para mejorar la robustez [1][11].

Los avances recientes en inteligencia computacional permiten formular el problema de detección de cambio en base a algoritmos de optimización con la selección de métricas adecuadas. Los conceptos de velocidad de detección del cambio, el tiempo medio entre falsas alarmas, la exactitud de la determinación del tiempo de cambio, y la sensibilidad general del algoritmo constituyen parámetros deseables de cuantificar y comparar.

En este proyecto se propone el planteo experimental de diferentes modelos predictivos y de algoritmos de detección de puntos de cambios. Se trabajará sobre series de tiempo simuladas y obtenidas de diversos sensores reales instalados sobre una plataforma de ensayo a fin de evaluar las características,

bondades y limitaciones de los mismos. Para finalizar, se debe señalar que pasar de la problemática de la detección de una anomalía a la generación de una alerta temprana y la decisión de poner en marcha un mecanismo de respuesta, posee consecuencias que superan el ámbito del proyecto y requiere de la formación de equipos multidisciplinarios.

2. LINEAS DE INVESTIGACION Y DESARROLLO

El problema de detección de puntos de cambio o anomalías en sistemas de monitoreo para alertas tempranas requiere del aporte de varias disciplinas en la formulación del riesgo, el sistema electrónico de monitoreo, el procesamiento computacional y las eventuales acciones de mitigación del problema.

En este proyecto, de carácter fundamentalmente experimental, se busca formar recursos humanos e informáticos para abordar los aspectos computacionales en las etapas de monitoreo electrónico y de selección, implementación y evaluación de algoritmos en la detección de las fallas o anomalías.

Debido a la amplitud y a las múltiples aplicaciones de la temática, se plantea su estudio inicial en el marco de dos ejes de interés para la región de influencia de la Facultad Regional Bahía Blanca. El primero asociado al medio ambiente y el otro referido al monitoreo de condición orientado al mantenimiento predictivo de maquinarias industriales. Se prestará especial atención a series provenientes de observaciones meteorológicas e hidrológicas en lo que hace a cantidad y calidad del agua para diversos usos. Por otra parte, la reciente instalación de un parque eólico cercano a la ciudad de Bahía Blanca ofrece la posibilidad de estudiar la conformación de SAT para mejorar los mecanismos de predicción de fallas posibles en su funcionamiento. En ambos casos se requiere de sistemas de monitoreo multisensor con capacidad para detección temprana de cambios de estado significativos.

Para cada eje se desarrollarán las siguientes actividades:

-Relevar y caracterizar los SAT, sistemas de monitoreo y el uso de algoritmos de detección de puntos de cambio. A su vez se procederá a la selección de casos factibles de simular y de probar en un banco de laboratorio adecuadamente diseñado a tal fin.

-Generación de secuencias de datos simuladas. Se usarán como series testigo de prueba y verificación para analizar los

algoritmos, provenientes tanto de bases de datos de acceso público como generadas por simulación mediante programas informáticos de desarrollo propio.

-Construcción de un banco de prueba con la red de sensores y el equipamiento necesarios para estudiar problemas planteados en los ejes mencionados. Esta es la etapa constructiva de hardware más significativa del proyecto. Se registrarán mediciones en condiciones de operación "normal" y con la imposición de cambios controlados. A partir de las series observadas se analizará el efecto del ruido de los sensores, la influencia de la disposición geométrica de los mismos y las dificultades electrónicas de sincronizar el muestreo de los diferentes sensores de la red.

-Ensayo de algoritmos y comparación de bondades y limitaciones de diferentes metodologías en base a criterios adecuados. Se espera que los mayores resultados cualitativos y cuantitativos de la investigación se produzcan en esta fase.

-Generación de un banco de software modular para facilitar la construcción de diferentes modelos. Se aplicará una metodología de desarrollo de software ágil basada en el paradigma de la programación extrema para enfatizar la adaptabilidad a los cambios de las especificaciones. El desarrollo de software se realizará en la modalidad *open-source* para facilitar su divulgación. Es de interés implementar un sistema modular que permita integrar diferentes series, probar e intercambiar algoritmos en un marco unificado.

-Transferencia de conocimientos y resultados al medio socio-productivo local y regional. Además de publicar los resultados obtenidos en eventos científicos, se contempla el dictado de cursos de capacitación y la posible asistencia para entidades interesadas en instalar o mejorar sus SAT.

3. RESULTADOS ESPERADOS

Se espera poder realizar contribuciones al avance científico y/o tecnológico en varios planos debido a la gran cantidad de aplicaciones que hacen uso de los sistemas de monitoreo y alerta temprana. La posible interacción con el grupo GEIA (Grupo de Estudio en Ingeniería Ambiental) de esta Facultad Regional permitirá encontrar aplicaciones en la prevención y mitigación de eventos que podrían afectar el medio ambiente, la salud poblacional o la calidad del aire o agua de un sistema. Por otro lado, la aplicación al monitoreo de condición en entornos

industriales, en conjunto con el Departamento de Ingeniería Mecánica podría traer como resultado la implementación de algoritmos que mejoren el esquema de mantenimiento preventivo de sus maquinarias.

En una etapa futura se espera poder relacionarse con organismos no universitarios interesados en la evaluación del cambio climático, alerta frente a catástrofes naturales de diversa índoles, y mantenimiento de plantas industriales entre otros.

El proyecto tiene como objetivo principal la generación de una biblioteca de software a publicar bajo la modalidad de *open-source* para fomentar el intercambio y transferencia entre investigadores a través de la Internet. El diseño de dicho software debe contemplar una estructura modular conformada por diversos algoritmos de detección y que permita el uso de diferentes bases de datos, tanto en un modo de análisis como el de una eventual implementación en un sistema productivo. Se espera también el desarrollo de diversas actividades de extensión y divulgación en la forma de cursos y seminarios sobre los aspectos teóricos del problema y análisis descriptivo de los algoritmos del software desarrollado. Así mismo, será de interés la participación de otros organismos, públicos y privados, con necesidad de SAT o detección temprana de fallas en la forma de proyectos de colaboración interinstitucional para el estudio de casos de aplicación real.

4. FORMACIÓN DE RECURSOS HUMANOS

La intención fundamental es crear un grupo de investigación que logre especializarse en los aspectos algorítmicos del tema de detección de anomalías (modelado, procesamiento, inteligencia computacional). El equipo buscará integrarse a grupos multidisciplinarios con los especialistas correspondientes a cada eje seleccionado. Al momento de iniciar el presente proyecto se cuenta con la participación en de cinco alumnos de grado y un graduado que se inician en actividades de I+D y transferencia al medio. Como objetivo fundamental se busca que se compenentren en la metodología de trabajo asociada a las tareas de investigación científica y adquieran destreza en el uso de herramientas computacionales, desarrollo de prototipos y sensado.

Se entiende que los nuevos conocimientos sumados al enfoque interdisciplinario del proyecto y la orientación a problemas del medio socio-productivo local representan un

valor agregado a su formación tradicional. La experiencia adquirida podrá definir la temática de sus respectivos proyectos finales de carrera. Además se buscará motivarlos para emprender estudios de posgrado en temáticas afines.

Al momento actual, están en proceso solicitudes de becas otorgadas por la Secretaría de Asuntos Estudiantiles (SAE) y por la Secretaría de Ciencia y Tecnología de la Facultad, a fin de formalizar la participación de alumnos y el graduado. Finalmente, se espera transferir algunos resultados al medio socio-productivo, mediante cursos y seminarios de capacitación y/o divulgación.

5. BIBLIOGRAFIA

- [1] S. Aminikhanghahi and D. Cook. A Survey of Methods for Time Series Change Point Detection. *Knowledge and Information Systems*, 51(2):339-367, May 2017.
- [2] V. Z. Antonopoulos, D. Papamichail, and K. Mitsiou. Statistical and Trend Analysis of Water Quality and Quantity Data for the Strymon River in Greece. *Hydrology and Earth System Sci.*, 5(4):679-692, 2001.
- [3] A. Arcos Jimenez, M. Muñoz Gomez, and F. García Marquez. Machine Learning for Wind Turbine Blades Maintenance Management. *Energies*, 11(1):13, Dec. 2017.
- [4] P. Baldini, G. Calandrini, P. Doñate and H. Bambill. PSO Algorithm-based Robust Design of PID Controller for Variable Time-delay Systems: AQM application. *Journal of Computer Science and Technology*, 15(2), Nov 2015.
- [5] I. Barnett and J.-P. Onnela. Change Point Detection in Correlation Networks. *Scientific Reports*, 6(18893), Jan 2016.
- [6] S. M. Bozic. *Digital and Kalman Filtering*. Edward Arnold Publishers Ltd. London, England, 2nd ed., 1979.
- [7] R. Coppo, L. M. Jakomin y C. Delrieux. Predicción de parámetros de calidad de agua mediante redes neuronales en la cuenca del río Pilcomayo - Argentina. IV Congreso Bianual PRODECA-PROIMCA, Bahía Blanca, Sep. 2017.
- [8] S. Dejus, A. Nescerecka, and T. Juhna. Online Drinking Water Contamination Event Detection Methods. 11th Environment, Technology, Resources Scientific and Practical Conf., v.1, 77-81, Latvia, 2017.
- [9] F. Desobry, M. Davy and C. Doncarli. AnOnline Kernel Change-Detection Algorithm. *IEEE Transactions in Signal Processing*, 53(8):2961-2974, Aug. 2005.
- [10] R. Domingues, M. Filippone, P. Michiardi.

- and J. Zouaoui. A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern Recognition*, 74:406–421, 2018.
- [11] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley and Sons, New York, 2nd ed., 2001.
- [12] F. Gustafsson. *Adaptive Filtering and Change Detection*. John Wiley & Sons Ltd., West Sussex, England, 2000.
- [13] L. B. Jackson. *Digital Filters and Signal Processing*. Kluwer Academic Publishers, Hingham, USA, 1986.
- [14] N. A. James and D. S. Matesson. ecp: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data. *Journal of Statistical Software, Articles*, 62(7):1-25, 2015.
- [15] Y. Kawara, Y. Yairi, and K. Machida. Change-point Detection in Time point Test for High-Dimensional Data. *Proceedings of the National Academy of Sciences*, 114(15):3873-3878, 2017.
- [16] N. Khan, S. McClean, S. Zhang, and C. Nugent. Optimal Parameter Exploration for Online Change-point Detection in Activity Monitoring using Genetic Algorithms. *Sensors*, 16(11):17-84, 2016.
- [17] D. Koller and N. Friedman. *Probabilistic Graphical Models - Principles and Techniques*. MIT Press, Cambridge, Massachusetts, USA, 2009.
- [18] K. K. Korkas and P. Fryzlewicz. Multiple Change-point Detection for Nonstationary Time Series using Wild Binary Segmentation. *Statistica Sinica*, 27(1):287-311, 2017.
- [19] L. I. Kuncheva and W. J. Faithful. PCA Feature Extraction for Change Detection in Multidimensional Unlabeled Data. *IEEE Transactions on Neural Networks Learning Systems*, 24(5):69-80, 2014.
- [20] Y. Li and R. Lund. Multiple Change-point Detection using Metadata. *Journal of Climate*, 28(10):4199-4216, 2015.
- [21] R. Lund, X. L. Wang, Q. Lu, J. Reeves, C. Gallagher and Y. Feng. Change-point Detection in Periodic and Autocorrelated Time Series. *Journal of Climate*, 20(20):5178-5190, 2007.
- [22] M. Morshedizadeh. *Condition Monitoring of Wind Turbines using Intelligent Machine Learning Techniques*. PhD thesis, University of Windsor, 2017.
- [23] H. V. Poor and O. Hadjiladis. *Quickest Detection*. Cambridge University Press, Cambridge, UK, 2nd ed., 2008.
- [24] J. Reeves, J. Chen, X L. Wang, R. Lund, and Q. Lu. A Review and Comparison of Change-point Detection Techniques for Climate Data. *Jour. Applied Meteorology and Climatology*, 46(6):900-915, 2007.
- [25] N. Sankary and A. Ostfeld. Inline Mobile Sensors for Contaminant Early Warning Enhancement in Water Distribution Systems. *Jour. Water Resources Planning and Management*, 143(2):5401-6073, 2017.
- [26] M. Schlechtingen and I. Santos. Comparative Analysis of Neural Network and Regression Based Condition Monitoring Approaches for Wind Turbine Fault Detection. *Mechanical Systems and Signal Processing*, 25(5):1849-1875, 2011.
- [27] X. Shi, Y. Wu and C. R. Rao. Consistent and Powerful Graph-based Change. point Test for High-Dimensional Data. *Proceedings of the National Academy of Sciences*, 114(15):3873-3878, 2017.
- [28] G. Shmueli. *Practical Time Series Forecasting - A Hands-on Guide*. Axelrod Schnall Publishers, USA, 3th ed., 2018.
- [29] A. Tartakovsky, I. Nikiforov and M. Basseville. *Sequential Analysis - Hypothesis Testing and Change-point Detection*. CRC Press - Taylor & Francis Group, Boca Raton, FL, USA, 2015.
- [30] S. Tinelli and I. Juran. Numerical Modeling of Early Bio-contamination in a Water Distribution System and Comparison with Laboratory Experiments. *International Conference on Sustainable Infrastructure 2017, American Society of Civil Engineers (ASCE)*, pages 258-269, New York, USA, Oct. 2017.
- [31] C. Truong, L. Oudre and N. Vayatis. A Review of Change Point Detection Methods. *The Computing Research Repository (CoRR)*, abs/1801.00718, 2018.
- [32] R. D. Turner. *Gaussian Processes for State Space Models and Change Point Detection*. PhD thesis, University of Cambridge, Cambridge, UK, July 2011.
- [33] A. Ukil and R. Zivanovic. The Detection of Abrupt Changes using Recursive Identification for Power System Fault Analysis. *Electric Power Systems Research*, 77(3):259-265, 2007.
- [34] P. Yang, G. Dumount and J. Ansermino. Adaptive Change Detection in Heart Rate Trend Monitoring in Anesthetized Children. *IEEE Trans. Biomedical Eng.*, 53(11):2211-2219, Nov. 2006.
- [35] X. Xu, H. Liu and M. Yao. *Recent Progress of Anomaly Detection*. Complexity. 2019. <https://doi.org/10.1155/2019/2686378>.